Université de Montréal

Évolution des roses (*Rosa* : Rosaceae) indigènes de la section *Cinnamomeae* à l'est des montagnes Rocheuses

par

Simon Joly

Département de Sciences biologiques

Université de Montréal

Thèse présentée à la Faculté des études supérieures en vue de l'obtention du grade de Ph.D. en Sciences biologiques.

Mai 2006

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée :

Évolution des roses (*Rosa* : Rosaceae) indigènes de la section *Cinnamomeae* à l'est des montagnes Rocheuses

présentée par :

Simon Joly

a été évaluée par un jury composé des personnes suivantes :

…….
François-Joseph Lapointe, président rapporteur

…….
Anne Bruneau, directrice de recherche

…….
Luc Brouillet, membre du jury

…….
Christopher Eckert, examinateur externe

…….
représentant du doyen de la FES

Résumé

Cette thèse porte sur le complexe *Rosa* sect. *Cinnamomeae* à l'est des montagnes Rocheuses. Bien que le nombre d'espèces reconnues dans ce groupe a grandement fluctué, les espèces les plus souvent acceptées comprennent cinq diploïdes (*R. blanda*, *R. foliolosa*, *R. nitida*, *R. palustris* et *R. woodsii*), trois tétraploïdes (*R. arkansana*, *R. carolina* et *R. virginiana*) et un hexaploïde (*R. acicularis*, qui ne sera pas étudié ici). Ces espèces sont très polymorphiques et s'hybrident fréquemment en nature, ce qui complexifie leur identification et remet leur statut d'espèce en question. Les relations phylogénétiques au sein du complexe sont aussi inconnues. Les objectifs de cette thèse sont donc de délimiter les espèces dans ce complexe et de reconstruire leur évolution. Des analyses multivariées de données morphologiques et moléculaires (AFLP) ont permis de distinguer sept groupes d'individus pouvant être considérés comme des espèces. En s'appuyant sur les espèces reconnues auparavant, ces groupes correspondent à *R. arkansana*, *R. blanda* – *R. woodsii*, *R. carolina*, *R. foliolosa*, *R. nitida*, *R. palustris* et *R. virginiana*. Il est impossible de distinguer le *R. blanda* du *R. woodsii* en fonction des caractères utilisés et ce groupe d'individus devrait être reconnu sous l'épithète *blanda* étant donné qu'il a priorité sur *woodsii*. L'évolution des diploïdes a été reconstruite en utilisant trois gènes nucléaires à l'aide de deux méthodes : l'une utilise les individus et l'autre les espèces comme unités terminales de l'analyse phylogénétique. Les résultats suggèrent que les espèces diploïdes du complexe forment un clade distinct des espèces diploïdes de l'ouest-américain. *Rosa nitida* et *R. palustris* sont des espèces sœurs, mais les analyses diffèrent relativement à la position de *R. blanda* (incl. *R. woodsii*) et *R. foliolosa*. L'évolution des polyploïdes a ensuite été reconstruite à l'aide du gène nucléaire *GAPDH* et des analyses multivariées morphologiques et moléculaires. Les résultats suggèrent que *R. arkansana* dérive de *R. blanda* (incl. *R. woodsii*), que *R. carolina* a évolué à partir d'un croisement entre *R. blanda*

(incl. *R. woodsii*) et *R. palustris* et que *R. virginiana* provient de *R. nitida*, de *R. palustris*, ou d'un croisement entre ces deux diploïdes. Les résultats suggèrent aussi que les espèces polyploïdes sont apparues à plusieurs reprises.

Mots clés : *Rosa*, délimitation d'espèces, polyploïdie, hybridation, phylogénie, gènes nucléaires à copie unique, arbres de gènes vs arbre d'espèces, systématique, taxonomie, évolution.

Abstract

This thesis discusses the evolution and systematics of species in the *Rosa* sect. *Cinnamomeae* complex east of the Rocky Mountains. Although the number of species recognized in this complex has fluctuated, generally five diploids (*R. blanda*, *R. foliolosa*, *R. nitida*, *R. palustris* and *R. woodsii*), three tetraploids (*R. arkansana*, *R. carolina* and *R. virginiana*) and one hexaploid (*R. acicularis*, not studied here) are accepted. These species are highly polymorphic and often hybridize in nature, which makes their identification laborious and raises the question of species integrity. Moreover, the evolutionary history of the group is poorly known. The objectives of this thesis are therefore to delimit species boundaries in the complex and to reconstruct their evolution. Multivariate analyses of morphological and molecular (AFLPs) data identified seven distinct groups of individuals that could be considered as distinct species. According to previous taxonomic species delimitation, these groups consist of *R. arkansana*, *R. blanda – R. woodsii*, *R. carolina*, *R. foliolosa*, *R. nitida*, *R. palustris* and *R. virginiana*. It was impossible to distinguish *R. blanda* from *R. woodsii* using morphological and molecular data and this group of individuals should be recognized as a single species under *R. blanda*. The evolution of diploids is reconstructed using three single-copy nuclear gene sequences and two methods of analysis: one uses individuals and the other species as terminal units of the analysis. Results suggest that diploid species of the complex form a clade distinct from western diploid species. *Rosa nitida* and *R. palustris* are sister species, but the analyses differed relative to the position of *R. blanda* (incl. *R. woodsii*) and *R. foliolosa*. The evolution of diploids is also assessed with the *GAPDH* nuclear gene and with morphological and molecular multivariate analyses. Results suggest that *R. arkansana* evolved from *R. blanda* (incl. *R. woodsii*), that *R. carolina* evolved from a hybrid between *R. blanda* (incl. *R. woodsii*) and *R. palustris*, and that *R. virginiana* evolved from either *R. nitida*, *R. palustris*, or from a

hybrid between the two. The results also show that all three polyploid species have evolved multiple times.

Key words: *Rosa*, species boundaries, polyploidy, hybridization, phylogeny, single-copy nuclear genes, gene trees vs species tree, systematics, taxonomy, evolution.

Table des Matières

Liste des tableaux

Liste des figures

Liste des abréviations

ADN	acide désoxyribonucléique
AFLP	Amplified fragment length polymorphism
CTAB	cetyltrimethylammonium bromide
DNA	deoxyribonucleic acid
GAPDH	glyceraldehyde 3-phosphate déshydrogénase
Incl.	inclus, incluant
M	molar
mM	mili molar
MS	malate synthase
PC	principal components
PCA	principal component analysis
PCR	polymerase chain reaction
PEG	polyethylène glycol
PCoA	principal coordinate analysis
RDA	redundancy analysis
TPI	triose phosphate isomerase

Remerciements

Mes premiers remerciements vont à ma directrice de recherche, Anne Bruneau, qui m'a donné le support, l'encouragement et la liberté qui m'ont permis d'effectuer cette thèse dans le plus grand bonheur. Anne m'avait dit avant de me lancer dans cette aventure que le doctorat développait avant tout une façon de penser et je suis convaincu que grâce à ses conseils, j'ai grandi dans ce sens.

Je dois aussi remercier les personnes qui m'entourent à l'Institut de recherche en biologie végétale ainsi qu'à l'Université de Montréal. Les professeurs Luc Brouillet, François-Joseph Lapointe et Bernard Angers ont particulièrement été une source de motivation pour moi et m'ont permis de toujours pousser ma réflexion plus loin. Je ne peux pas mentionner tous les professeurs, botanistes, étudiants et le personnel qui ont fait que ces années passées à l'Université de Montréal ont été si agréables et ont passé si rapidement ; il y a trop de personnes importantes. Cependant, sachez que vous avez tous été importants pour moi.

Je voudrais remercier aussi les gens qui partagent mon quotidien. Isabelle, tu as toujours cru en moi et tu m'as toujours supporté. Même les moments de stress qui accompagnent immanquablement un doctorat ont paru faciles grâce à ta compréhension. Raoul, je voulais aussi te remercier pour la joie que tu m'apportes et pour m'avoir forcé à m'organiser comme jamais, ce qui m'a d'ailleurs permis d'écrire cette thèse dans un laps de temps que je n'aurais pas cru possible au départ. Merci aussi à toute ma famille qui a toujours cru en moi et qui a donc rendu ces nombreuses années d'études beaucoup plus facile.

Je suis aussi extrêmement reconnaissant envers tous ceux qui ont récolté des échantillons de roses où qui ont trouvé des populations et qui ont donc contribué à améliorer la qualité de cette thèse : Walter Lewis, Barbara Ertter, Luc Brouillet, Elisabeth Dickson, Jeff Saarela, Félix Hébert, Bob Dirig, Frédéric Coursol, Alain Meilleur et Annie Archambault. Je dois avoir une pensée spéciale ici pour Julian Starr, avec qui j'ai effectué des longs voyages de récolte, et avec qui j'ai pu partager mes passions pour les roses, la musique et le hockey.

Mon expérience au doctorat a été enrichie de plusieurs conférences internationales qui n'auraient pu être possibles sans l'apport financier de l'Institut de recherche en biologie végétale via les bourses Jacques-Rousseau et des fonds de recherche de ma directrice (CRSNG, FQRNT). Finalement, cette thèse a été rendue possible grâce à l'apport financier du CRSNG, du FQRNT, de la fondation des Muséums nature de Montréal et de la Faculté des études supérieures de l'Université de Montréal.

Avant-propos

Le sujet de cette thèse s'est décidé par un soir pluvieux à l'automne 2001. À ce moment, je cherchais intensément un sujet de doctorat qui saurait me motiver et m'intéresser pendant plusieurs années. Alors que j'évaluais plusieurs possibilitées, Julian Starr, qui travaillait alors sur la phylogénie du genre *Rosa* pour son stage post-doctoral dans le laboratoire d'Anne Bruneau, m'avait suggéré d'étudier les roses est-américaines. Cela constituait un bon sujet de doctorat selon lui. Par chance, Walter Lewis, actuellement le plus grand taxonomiste des roses en Amérique de Nord, devait se rendre à Ottawa ce même automne. Dans les années 1950s, Walter Lewis avait effectué sa thèse sur les roses nord-américaines situées à l'est des montagnes Rocheuses. Ainsi, il n'y avait personne de mieux placé que lui pour me parler de ce groupe. Moi et Julian Starr sommes donc allés rencontrer Walter Lewis en cette soirée qui fut déterminante pour mon doctorat. J'ai alors rencontré un botaniste passionné qui a su me transmettre sa passion des roses. Au terme de cette soirée, mon idée était faite et il ne faisait plus aucun doute que ma thèse porterait sur le même sujet que celui abordé par Walter Lewis cinquante ans plus tôt dans sa thèse.

CHAPITRE 1

Introduction

Depuis Linné (1753), les taxonomistes ont été confrontés à un problème de taille en ce qui a trait à la délimitation des espèces au sein du genre *Rosa*, problème qui est en majeure partie causé par le grand polymorphisme intra-spécifique des roses. Mais si plusieurs personnes ont vu dans cette confusion l'effet du polymorphisme, d'autres ont préféré y voir plusieurs espèces. C'est ce qui explique que le nombre d'espèces reconnues a tant varié dans le genre, fluctuant de 14 pour tout le genre (Linné, 1753) à 4266 en Asie et en Europe seulement (Gandoger, 1881). Le même phénomène a été observé en Amérique du Nord : Watson (1885), Crépin (1896) et Erlanson MacFarlane (1966) ont tour à tour décrit 18, 13 et 22 espèces sur le continent, alors que Rydberg (1920) a décrit 129 espèces uniquement pour le nord-est de l'Amérique du Nord.

Cependant, le polymorphisme intra-spécifique n'est pas la seule source de confusion taxonomique chez *Rosa* puisque l'hybridation est aussi considérée comme un phénomène important depuis longtemps (Crépin, 1894). Linné (1753) avait notamment reconnu son importance dans son introduction au *Species plantarum* : "the species of *Rosa* are with difficulty to be distinguished, with even greater difficulty to be defined; nature seems to me to have blended several or by way of sport to have formed several from one" (traduit par Stearn, 1957:158). Des études biosystématiques du début du 20^{e} siècle ont d'ailleurs confirmé l'importance de l'hybridation dans le genre (Erlanson, 1929, 1934; Ratsek *et al.*, 1939; 1940). Aussi au début du 20^{e} siècle, des études cytologiques ont montré que la polyploïdie était fréquente dans le genre *Rosa* alors que six des dix sections possèdent des polyploïdes (Täckholm, 1922). La polyploïdie pouvait donc aussi

contribuer aux problèmes taxonomiques des roses. Par exemple, *Rosa arkansana*, une espèce nord-américaine, avait été considérée par Crépin (1896) comme un écotype de *R. blanda*. Cependant, on a démontré plus tard que *R. arkansana* était en fait une espèce tétraploïde et que *R. blanda* était diploïde (Erlanson, 1929), ce qui venait résoudre ce problème taxonomique. En dépit de l'information apportée par la cytologie et par les études sur l'hybridation des roses, le désordre reste important dans plusieurs groupes tels la section *Canninae* (Ritz *et al*., 2005) ou la section *Cinnamomeae* en Amérique du Nord (Lewis, 1957c).

La présente thèse se penche sur la section *Cinnamomeae* à l'est des montagnes Rocheuses. Ce complexe d'espèces est représentatif du genre pour le polymorphisme intra-spécifique des espèces ainsi que pour la présence d'hybridation et de polyploïdie. Le nombre d'espèces généralement reconnues dans ce groupe est de 9 (Erlanson MacFarlane, 1966; ce nombre exclut une espèce d'origine hybride), mais jusqu'à 23 ont déjà été décrites à l'est de la rivière Mississippi (Rydberg, 1920). Ce complexe d'espèces a été l'objet de plusieurs études portant sur la cytologie (Erlanson, 1929; Lewis, 1957b, 1966), la morphologie (Crépin, 1896; Erlanson, 1930, 1934; Lewis, 1957b, 1957c, 1958, 1962), ainsi que sur des croisements expérimentaux (Erlanson, 1929; Ratsek *et al*., 1939; 1940; Lewis et Basye, 1961). Malgré ces nombreuses études, il est toujours aussi difficile de distinguer les espèces dans ce complexe.

La délimitation d'espèces dans ce groupe reste donc précaire. Les espèces sont tellement polymorphiques qu'aucun caractère ne permet à lui seul de distinguer une espèce des autres. De plus, le portrait est grandement complexifié par la présence d'espèces polyploïdes qui sont particulièrement polymorphiques et qui comblent les minces différences morphologiques présentes entre les diploïdes. D'ailleurs, ces espèces polyploïdes sont reconnues pour s'hybrider dans les zones de contact, rendant leur identification plus difficile, mais soulevant aussi la question de savoir si ces espèces sont véritablement distinctes. De plus, nous ne connaissons rien de leur évolution. Il y a donc un besoin certain de mieux comprendre la délimitation des espèces dans ce groupe ainsi

que pour comprendre leur évolution. D'ailleurs, les deux concepts vont de pair puisqu'une bonne délimitation des espèces permettra de reconstruire l'évolution des espèces, mais l'évolution de certaines espèces pourrait aussi s'avérer informative pour délimiter les espèces. Avant de définir clairement les objectifs de recherche, il est important de présenter certaines notions relatives au groupe à l'étude et aux problématiques abordées dans cette thèse.

1.1 Cadre taxonomique

1.1.1 Le genre *Rosa*

Le genre *Rosa* L. est constitué d'une centaine d'espèces distribuées dans les régions tempérées et sub-tropicales de l'hémisphère Nord (Krüssmann, 1981). Il est composé d'espèces vivaces ligneuses, aux feuilles pennées (rarement simples) qui possèdent des stipules adnées et des folioles serrées, et qui possèdent des fleurs aux étamines et pistils nombreux, ces derniers étant insérés à la base d'un hypanthium bien développé en forme d'urne (cynorhodon en fruit). Les espèces du genre *Rosa* sont inégalement réparties en quatre sous-genres : *Rosa*, qui comprend 10 sections et la majorité des espèces, et *Hulthemia* (Dumort.) Focke, *Platyrhodon* (Hurst) Rehd. et *Hesperhodos* Cockerell, constitués d'une ou deux espèces chacun (Rehder, 1940; Clarke, 1980).

Deux sous-genres sont représentés en Amérique du Nord, soit *Hesperhodos* (2 espèces) et *Rosa* (20 espèces), ce dernier étant représenté par les sections *Cinnamomeae* et *Synstylae* (Tableau 1). La majorité des espèces de roses nord-américaines font partie de la section *Cinnamomeae*.

Présentement, le type du genre est *R. gallica* L. (Wissemann, 2003). Cependant, ce type n'est pas adéquat (Lewis, 1957a; Rowley, 1976) étant donné que la première espèce du genre qui ait été décrite est *R. cinnamomea* L. (Linné, 1753). Donc, le genre *Rosa* doit

Tableau 1.1 Espèces du genre *Rosa* en Amérique du Nord

Sous-genre	Sections	Espèces [1]	Ploïdie (2n =) [2]	Répartition
Rosa	*Cinnamomeae*	*R. acicularis* Lindl.	2x, 4x, 6x, 8x	Est, ouest, Eurasie
		R. arkansana Porter	4x	Est
		R. blanda Ait.	2x	Est
		R. californica Cham. & Schltdl.	4x	Ouest
		R. carolina L.	4x	Est
		R. durandii Crép.	4x	Ouest
		R. foliolosa Nutt.	2x	Est
		R. gymnocarpa Nutt. ex Torr. & A. Gray	2x	Ouest
		R. nitida Willd.	2x	Est
		R. nutkana C. Presl	6x	Ouest
		R. palustris Marsh.	2x	Est
		R. pisocarpa A. Gray	2x	Ouest
		R. spithamea S. Watson	4x	Ouest
		R. virginiana Mill.	4x	Est
		R. woodsii Lindl.	2x	Est, ouest
	Synstylae	*R. setigera* Michx.	2x	Est
Hesperhodos		*R. minutifolia* Engelm.	2x	Ouest
		R. stellata Wooton	2x	Est

[1] Selon Erlanson (1966) pour la section *Cinnamomeae*.
[2] Le nombre chromosomique de base chez *Rosa* est x = 7.

être lectotypifié pour corriger cette erreur, ce qui implique que la section *Cinnamomeae* deviendrait la section *Rosa*. Cependant, ce changement n'a pas encore été accepté par l'International Botanical Congress et *Cinnamomeae* reste le nom de section valide. C'est donc cette nomenclature qui sera utilisée ici même si elle n'est pas adéquate.

1.1.2 Les espèces à l'étude

Des espèces de la section *Cinnamomeae* présentes à l'est des Rocheuses, soient celles qui sont à l'étude, cinq sont diploïdes (*R. blanda*, *R. foliolosa*, *R. nitida* et *R. palustris* ; voir la répartition de chaque espèce à la Fig. 1.1) et trois sont tétraploïdes (*R. arkansana*, *R. carolina* et *R. virginiana* ; Fig. 1.2). L'espèce hexaploïde *R. acicularis*, présente dans l'est et

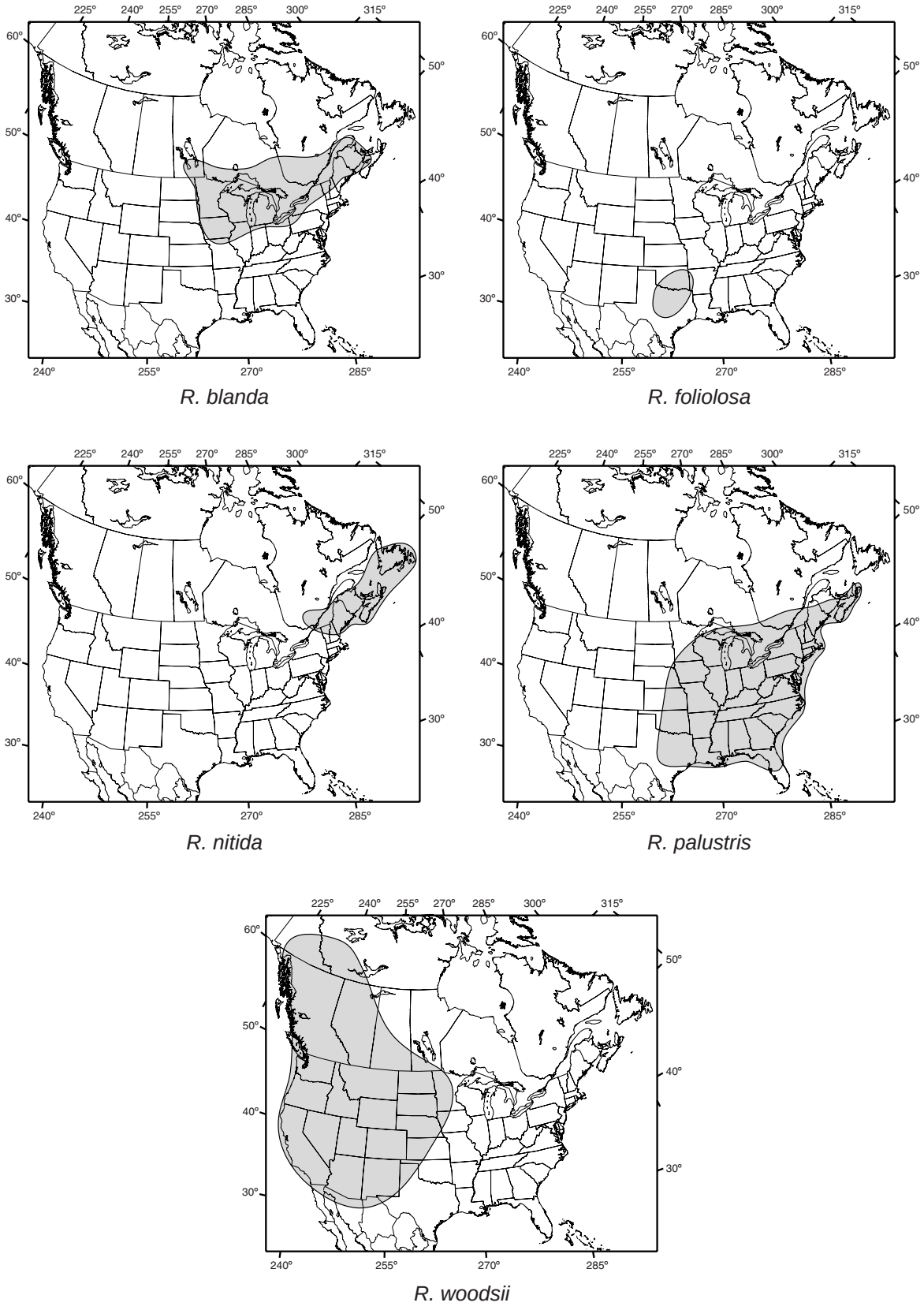


Figure 1.1 Répartition des espèces indigènes de roses diploïdes de la section *Cinnamomeae* à l'est des montagnes Rocheuses (d'après Lewis, 1957b; S. Joly, observations personnelles).

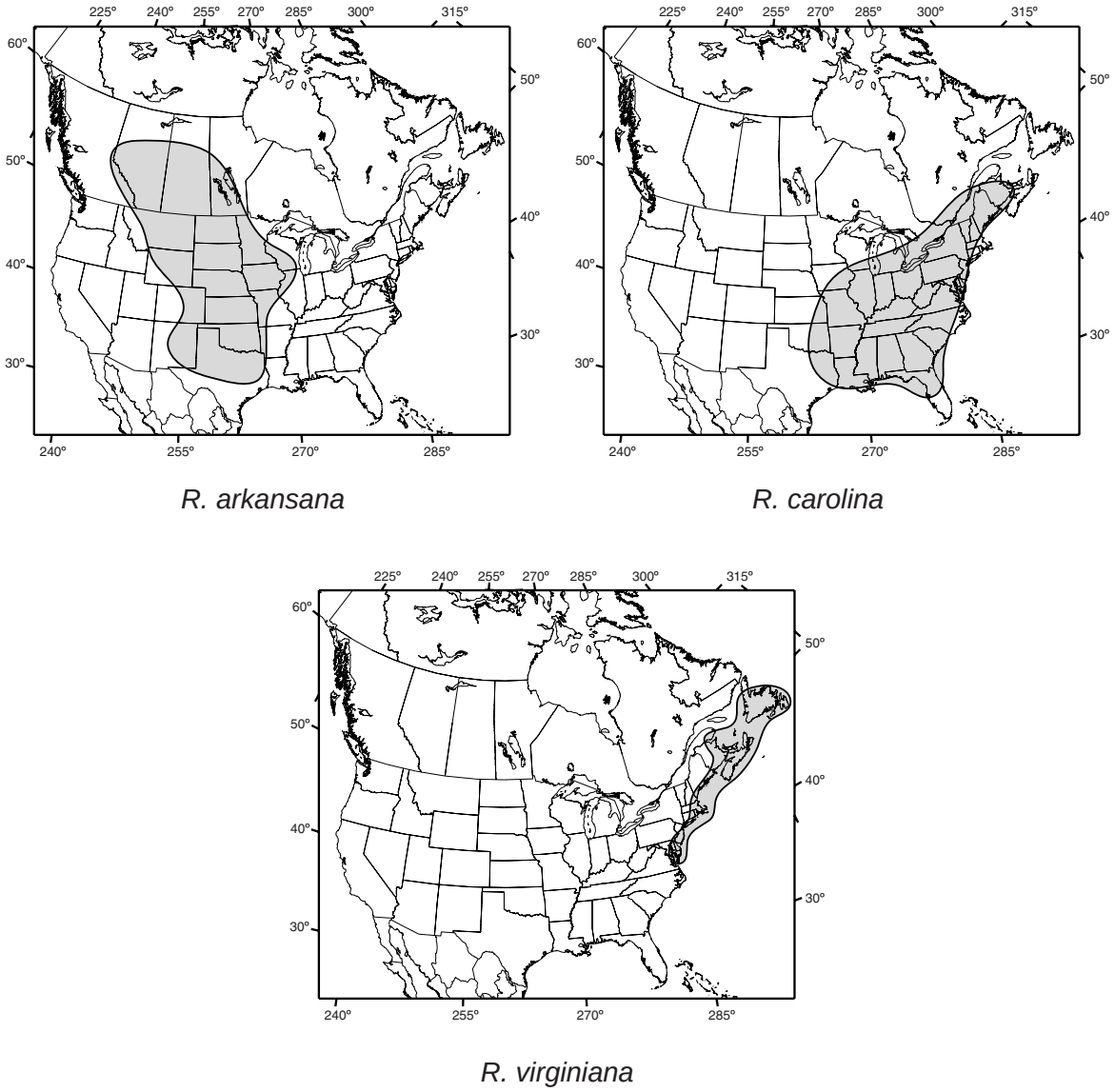


Figure 1.2 Répartition des espèces indigènes de roses tétraploïdes de la section *Cinnamomeae* à l'est des montagnes Rocheuses (d'après Lewis, 1957b; S. Joly, observations personnelles).

l'ouest de l'Amérique du Nord mais aussi en Eurasie, ne sera pas incluse dans la présente étude. Son importante répartition (Lewis, 1959) implique que ses ancêtres pourraient inclure des espèces européennes ou asiatiques et l'étude de ses origines nécessiterait un échantillonnage beaucoup plus important que celui qui a été effectué pour cette thèse. Les espèces étudiées dans cette étude ont souvent été divisées en plusieurs espèces dans le passé (e.g., Rydberg, 1920) et les espèces placées en synonymie sont énumérées ailleurs (Erlanson, 1934).

1.1.3 La position phylogénétique des espèces nord-américaines

Les études phylogénétiques du genre *Rosa* donnent très peu d'information relativement à l'évolution du genre (Bruneau *et al.*, submitted). Ceci est d'une part causé par la faible variabilité des marqueurs moléculaires (e.g., *mat*K: Matsumoto *et al.*, 1998), mais aussi parce que les études antérieures avaient inclus très peu d'espèces (Millan *et al.*, 1996; Matsumoto *et al.*, 1998). L'étude de Wissemann et Ritz (2005), avec 61 espèces échantillonnées, fut le premier effort véritable vers une reconstruction phylogénétique complète du genre. Par contre, les deux gènes utilisés (la région ITS de l'ADN ribosomique nucléaire et l'espaceur chloroplastique *atp*B-*rbc*L) ne possèdent pas une très bonne variation et la phylogénie obtenue n'est donc pas très robuste. La majorité des espèces nord-américaines se retrouvent dans un clade dans cette étude, mais quelques espèces de *Cinnamomeae* asiatiques s'y trouvent aussi. Cependant, toutes les espèces nord-américaines de cette étude ont été récoltées dans des jardins botaniques, ce qui représente une source potentielle d'erreur. En effet, étant donné la propension des rosiers pour l'hybridation et la proximité des rosiers dans les jardins botaniques, on peut douter de la pureté des espèces utilisées dans l'analyse. D'ailleurs, la séquence de l'espaceur *atp*B-*rbc*L de *R. palustris* s'est retrouvée dans un clade autre que celui qui comprenait tous les autres taxons nord-américains (Wissemann et Ritz, 2005). Ceci montre bien le danger d'inclure des rosiers provenant de jardins botaniques dans des analyses phylogénétiques.

Dans une autre analyse phylogénétique du genre utilisant deux gènes chloroplastiques plus variables que ceux utilisés par Wissmann et Ritz (2005) et dont les espèces nord-américaines ont presque toutes été récoltées en nature (Bruneau *et al.*, submitted), les espèces d'Amérique du Nord se trouvaient toutes dans un seul clade, lequel contenait aussi un clade asiatique. Donc, bien que les roses nord-américaines de la section *Cinnamomeae* ne semblent pas monophylétiques, elles sont à tout le moins paraphylétiques et peuvent être étudiées sans craindre que d'autres espèces aient influencé leur évolution.

Même si quelques analyses phylogénétiques ont inclus des espèces nord-américaines (Matsumoto *et al.*, 1998; 2000; Ritz *et al.*, 2005), la faible variation des marqueurs utilisés n'a pas permis d'obtenir une résolution suffisante pour pouvoir reconstruire leur phylogénie. Donc, la question de l'évolution des espèces nord-américaines reste toujours sans réponse.

1.2 Délimitation et concepts d'espèce

La question des concepts d'espèce a occupé une place importante de la littérature en biologie évolutive au cours des cinquante dernières années (Hull, 1997). Cependant, il n'y a toujours pas de consensus relativement à ce que les biologistes considèrent comme une espèce, ni même à quoi le concept devrait correspondre. Par exemple, on ne s'entend pas au sujet des processus qui sont importants pour délimiter les espèces. Alors que certains croient que les espèces devraient être définies en fonction du flux génique (Mayr, 2000), d'autres définissent l'espèce en fonction de leur mécanismes reproductifs (Paterson, 1985), en termes de compétition (Ghiselin, 1975), de destin évolutif (Wiley et Mayden, 2000) ou même de distinctions morphologiques et génétiques (Mallet, 1995). De plus, tous les biologistes ne s'entendent pas sur ce qu'une espèce devrait représenter. Certains voient l'espèce comme un niveau taxonomique particulier qui représente l'unité à laquelle l'évolution prend place (Mayr, 1969; Ghiselin, 1975; Hull, 1976), alors que d'autres voient l'espèce comme une unité taxonomique qui n'est pas différente des autres (e.g., Mallet, 1995) ou comme l'unité minimale qui devrait être incluse dans une analyse phylogénétique (Mishler et Theriot, 2000; Wheeler et Platnick, 2000).

Toutes ces considérations théoriques sont importantes, mais elles n'aident pas à l'identification des espèces. D'ailleurs, pendant longtemps, le débat sur les différents concepts d'espèces était divisé entre les concepts engagés théoriquement (qui font référence aux processus biologiques impliqués dans la spéciation) et les concepts opérationnels ou pratiques (Luckow, 1995; Hull, 1997). Certains (par exemple Mallet,

1995) prétendent qu'un concept ne devrait pas être fondé sur des prémisses évolutives afin que celles-ci soient testées *a posteriori*, mais un concept d'espèce sans aucun fondement théorique est pratiquement impossible. De plus, de tels concepts n'expliquent pas pourquoi l'on retrouve des espèces en nature. De façon similaire, un concept purement théorique et non-applicable n'est pas plus utile aux biologistes.

Plus récemment, certains auteurs ont fait un effort pour rendre leurs concepts davantage applicables (e.g., Davis et Nixon, 1992; Templeton, 2001). De plus, en se basant sur les prémisses des différents concepts, il est possible de développer des moyens pour définir des espèces. Par exemple, l'importance accordée au flux génique par plusieurs concepts a permis le développement de méthodes permettant de déterminer où le flux génique cesse entre populations (e.g., Davis et Nixon, 1992; Doyle, 1995; Sites Jr. et Marshall, 2003). D'autres méthodes ont aussi été développées récemment afin de délimiter les espèces le plus objectivement possible (Wiens et Servedio, 2000; Puorto *et al.*, 2001; Sites Jr. et Marshall, 2003, 2004).

Pour chaque étude voulant délimiter les espèces dans un groupe donné, il est important d'énoncer la définition d'espèce qui sera utilisée ainsi que le moyen qui sera employé pour délimiter celles-ci. Dans le cadre de cette thèse, l'espèce est considérée comme étant l'unité de l'évolution, en ce sens qu'elle représente un taxon singulier. Les caractéristiques qui sont considérées comme importantes pour définir l'espèce comme l'unité de l'évolution sont sa continuité spatio-temporelle, avec une possibilité d'évolution, ainsi qu'une cohésion entre les organismes qui la composent (Hull, 1997). Bref, une espèce peut être considérée comme le groupe le plus inclusif d'organismes qui a le potentiel de maintenir une cohésion et qui évolue indépendamment d'autres groupes d'organismes. Une telle définition prévoit que différentes espèces devraient devenir morphologiquement ou génétiquement distinctes avec le temps. En effet, parce que la différentiation morphologique et génétique peuvent être atteintes à différent moments, une espèce ne doit pas nécessairement être morphologiquement et génétique distincte. La nature des mécanismes responsables de la cohésion des espèces est intentionnellement

omise de cette définition puisque c'est principalement sur ce sujet que les différents concepts d'espèce diffèrent, en partie parce que les mécanismes importants pour la cohésion peuvent varier d'un groupe à l'autre. Par contre, peu importe si ces mécanismes impliquent le flux génique, les systèmes reproductifs, l'écologie, la compétition ou une combinaison de facteurs, tous devraient permettre de définir des espèces en fonction de la présente définition.

1.3 L'hybridation

L'hybridation a depuis longtemps été considérée comme un processus évolutif important chez les roses, tout comme chez les angiospermes en général (Arnold, 1997; Rieseberg, 1997; Rieseberg *et al.*, 2003). Dans le groupe à l'étude, l'hybridation a été rapportée à plusieurs reprises tant au niveau diploïde (Rydberg, 1920; Erlanson, 1929; Lewis, 1962) que polyploïde (Fernald, 1922; Lewis, 1957b; Rhoads et Klein Jr., 1993; Haines et Vining, 1998; Hinds, 2000). Elle est d'ailleurs tenue responsable des problèmes d'identification entre les espèces polyploïdes (Lewis, 1957b) et entre certaines espèces diploïdes (Lewis, 1962). L'hybridation pourrait aussi provoquer la fusion de deux espèces ou être impliquée lors d'une spéciation polyploïde impliquant deux espèces parentales. Cependant, nous ne savons pas si l'hybridation a vraiment joué un rôle important dans l'évolution de ce complexe.

Dans le cadre de cette thèse, deux concepts relatifs à l'hybridation doivent être présentés pour comprendre l'influence de l'hybridation dans cette étude : les zones d'hybridation et le problème des hybrides en analyse phylogénétique.

1.3.1 Les zones d'hybridation

Les zones d'hybridation ont souvent été considérées comme un système permettant d'étudier l'évolution (e.g., Dobzhansky, 1937; Mayr, 1942). Différents modèles de zones

d'hybridation existent (Arnold, 1997), mais chaque modèle peut se former de deux façons différentes : par contact primaire ou par contact secondaire (Endler, 1977; Barton et Hewitt, 1985). Dans le cas des zones d'hybridation secondaires, les populations qui s'hybrident ont d'abord évolué en allopatrie avant d'entrer en contact. Au contraire, les populations s'hybridant dans les zones d'hybridation primaires ont divergé en parapatrie. Ainsi, les populations de zones d'hybridation secondaires peuvent être considérées comme des espèces distinctes puisqu'à un moment donné, leur évolution était indépendante. On ne peut pas en dire autant des zones d'hybridation primaire puisque dans ce cas les populations n'ont jamais vraiment eu une évolution indépendante, bien que la zone d'hybridation puisse éventuellement mener à la spéciation.

Il est souvent difficile de distinguer ces deux types de zones (Barton et Hewitt, 1985), mais cela peut être plus facile dans certains cas comme lorsque les espèces qui s'hybrident ne sont pas des espèces-sœurs ou lorsque qu'elles sont des polyploïdes qui ont des origines distinctes. Il importe donc de faire une distinction entre ces deux types de zones d'hybridation, lorsque c'est possible, puisque cela peut avoir un impact sur la délimitation des espèces. Ceci est particulièrement important dans le groupe à l'étude étant donné que des zones d'hybridation ont été décrites entre certaines espèces (Lewis, 1957b, 1962).

1.3.2 Les hybrides en analyse phylogénétique

Les méthodes de reconstruction phylogénétique traditionnelles ne peuvent pas adéquatement positionner les hybrides dans une phylogénie. Même si McDade (1990; 1992) a montré que l'introduction d'hybrides dans les analyses phylogénétiques avait peu d'impact sur les relations entre les espèces non hybrides, il n'en demeure pas moins que les relations des taxons hybrides ne sont pas reconstruites de façon adéquate. Une approche fréquemment utilisée est d'enlever les hybrides avant de procéder à l'analyse phylogénétique. Les méthodes de détection deviennent ainsi importantes pour identifier

les organismes hybrides. L'alternative est d'utiliser des méthodes de reconstruction phylogénétique qui permettent de représenter l'évolution des espèces hybrides (e.g., Xu, 2000; Nakhleh *et al.*, 2004; Huson *et al.*, 2005). Cependant, ces méthodes sont en développement et sont encore peu utiles pour résoudre les problèmes liés à l'hybridation (Linder et Rieseberg, 2004). D'autres méthodes phylogénétiques, la plupart utilisant des distances, tentent d'expliquer l'incongruence présente dans les jeux de données et sont parfois utilisées pour étudier l'hybridation (Lapointe, 2000; Buntjer *et al.*, 2002; Hodkinson *et al.*, 2002; Legendre et Makarenkov, 2002; Bryant et Moulton, 2004; Holland *et al.*, 2004). Cependant, ces méthodes souffrent de différents problèmes (Lapointe, 2000; Bryant et Moulton, 2002; Legendre et Makarenkov, 2002; Bryant et Moulton, 2004) et peuvent, au mieux, être utilisées pour suggérer la présence d'hybridation.

1.4 La polyploïdie

La polyploïdie est un processus évolutif important chez les eucaryotes, étant rencontrée très fréquemment chez les plantes à fleurs (Stebbins, 1950; Grant, 1981; Masterson, 1994), les fougères (Wagner et Wagner, 1980), les mousses (Crosby, 1980), les algues (Nichols, 1980) et les animaux (Ohno, 1970; McLysaght *et al.*, 2002). Chez les angiospermes, la polyploïdie est associée à 6% des événements de spéciation (Otto et Whitton, 2000). D'ailleurs, il semble que certains organismes aient passé par plusieurs événements de polyploïdisation dans leur passé, comme en témoigne l'étude d'*Arabidopsis thaliana* qui, malgré son petit génome ($2n = 2x = 10$), aurait été dupliqué à trois reprises (Barnes, 2002; Blanc *et al.*, 2003; Simillion *et al.*, 2003; Ziolkowski *et al.*, 2003). Les données génomiques suggèrent même que la polyploïdie pourrait être responsable de la radiation des angiospermes et des eudicotylédones (De Bodt *et al.*, 2005). Un tel succès évolutif pourrait bien être relié aux changements écologiques (Favarger, 1967; Ehrendorfer, 1980; Lewis, 1980; Petit et Thompson, 1999), phénotypiques (Comai *et al.*, 2000; Ramsey et Schemske, 2002), physiologiques (Tal, 1980; Levin, 1983; Thompson *et al.*, 1997b; Levin, 2002), génomiques (Song *et al.*, 1995; Liu *et al.*, 1998a; 1998b; Ozkan *et al.*, 2001; Shaked *et al.*, 2001) et protéomiques (Lee et Chen, 2001; Lewis et Pikaard, 2001;

Pikaard, 2001; Adams *et al.*, 2003; Osborn *et al.*, 2003; Pontes *et al.*, 2003; Riddle et Birchler, 2003; Joly *et al.*, 2004) souvent associés à la polyploïdie.

Tel que mentionné plus haut, la polyploïdie est fréquente chez les roses, tout comme chez les Rosaceae (Vamosi et Dickinson, 2006). D'ailleurs, certain complexes sont devenus des classiques dans ce domaine d'étude (section *Caninae*: Stebbins, 1950; Grant, 1981). La polyploïdie est aussi importante en Amérique du Nord et dans le groupe à l'étude (Tableau 1.1), alors que trois des espèces étudiées sont tétraploïdes: *R. arkansana*, *R. carolina* et *R. virginiana*. Comparativement aux diploïdes, les polyploïdes occupent généralement des habitats plus xériques que les diploïdes. Un tel fait a été observé dans plusieurs groupes de plantes (pour une revue de littérature, voir Levin, 2002), mais il existe cependant des contre-exemples (Ehrendorfer, 1980).

Deux facettes de l'évolution des polyploïdes sont fréquemment abordées dans les études qui reconstruisent l'évolution des polyploïdes : la nature des polyploïdes et la possibilité d'origines récurrentes.

1.4.1 Nature des polyploïdes

En général, deux types de polyploïdes sont reconnus : les autopolyploïdes et les allopolyploïdes (Ramsey et Schemske, 1998, 2002). Cependant, deux définitions sont fréquemment utilisées pour chacun de ces termes. Selon la définition cytologique, les autopolyploïdes sont issus d'individus dont les parents forment des hybrides fertiles et sont caractérisés par la formation de multivalents à la méiose (Stebbins, 1950, 1980; Levin, 2002). Selon cette même définition, les allopolyploïdes sont issus d'un croisement entre deux individus dont l'hybride a une fertilité réduite (Stebbins, 1950, 1980; Levin, 2002). Dans ce cas, le dédoublement du génome est nécessaire pour réinstaurer la fertilité de l'hybride et l'allopolyploïde ainsi formé sera caractérisé par la formation de bivalents à la méiose, où chaque chromosome s'apparie avec le chromosome provenant du même individu parental. Les prédictions respectives de formation de multivalents et de

bivalents pour les auto- et les allopolyploïdes ne sont généralement vraies que pour les polyploïdes récents. En effet, avec le temps, les polyploïdes tendent à « diploïdiser » leur génome et même les autopolyploïdes finissent par former des bivalents avec le temps (Wendel, 2000).

Les premières définitions d'autopolyploïdes considéraient que ceux-ci évoluaient strictement par dédoublement de chromosomes (Darlington, 1928), ce qui explique pourquoi plusieurs auteurs ont cru que les autopolyploïdes n'avaient pas beaucoup de potentiel évolutif (voir Stebbins, 1980). Cependant, nous savons maintenant que si ce mode de formation est biologiquement possible, il n'est pas le plus fréquent. En fait, il semble que la plupart des polyploïdes soient formés par le biais de gamètes non-réduits, souvent via un « pont » triploïde (Harlan et deWet, 1975; deWet, 1980; Ramsey et Schemske, 1998; Burton et Husband, 2000). Ceci explique la diversité présente dans certaines lignées polyploïdes clonales, incompatible avec l'hypothèse d'un simple dédoublement chromosomique (Haufler *et al.*, 1985; Joly et Bruneau, 2004).

Une deuxième définition, taxonomique celle-ci, est souvent utilisée (Ramsey et Schemske, 1998, 2002). Selon cette définition, les autopolyploïdes ont évolué à partir d'une seule espèce alors que les allopolyploïdes ont évolué à partir de deux espèces. Si elle semble simple, parce qu'on n'a pas besoin d'étudier la cytologie des espèces parentales et de leurs hybrides, cette définition a le désavantage de ne pas permettre de faire des prédictions sur le comportement méiotique des polyploïdes. Or, de telles prédictions sont importantes lorsque l'on veut retracer l'évolution des polyploïdes. En effet, selon la définition cytologique, les allopolyploïdes devraient conserver des traces de leurs deux parents puisque les chromosomes provenant de chacun ne ségrègent pas et sont transmis de génération en génération. La même chose n'est pas vraie des autopolyploïdes puisqu'à cause de la formation de multivalents, tous les chromosomes peuvent ségréger à la méiose et l'origine des parents du polyploïde peut donc être perdue.

Pour ce qui est de l'origine des polyploïdes du complexe, nous avons présentement très peu d'indices permettant de formuler des hypothèses sur leur nature ou sur leurs progéniteurs. *Rosa arkansana* pourrait provenir d'un croisement entre *R. blanda* et *R. woodsii* (Erlanson, 1934), alors que la grande ressemblance morphologique des *R. carolina* et *R. virginiana* avec les espèces diploïdes *R. foliolosa*, *R. nitida* et *R. palustris* laisse supposer des liens de parenté possibles avec celles-ci, plus particulièrement avec *R. palustris* (Lewis, 1957b). Cependant, aucune étude ne s'est sérieusement penchée sur l'évolution des polyploïdes et leurs espèces parentales restent à déterminer.

1.4.2 Origines multiples

L'application des données moléculaires à l'étude des polyploïdes a démontré que la plupart des taxons polyploïdes étaient formés à plusieurs reprises (Soltis et Soltis, 1993, 1995; 2000). Si de tels exemples provenaient initialement d'allopolyploïdes (Soltis et Soltis, 1993, 1999), des recherches récentes montrent que les autopolyploïdes n'échappent pas à la règle (Soltis *et al.*, 1989; Van Dijk et Bakx-Schotman, 1997; Segraves *et al.*, 1999; Takamiya *et al.*, 2001; Joly et Bruneau, 2004). Ces évolutions multiples sont importantes pour comprendre le succès évolutif des polyploïdes puisqu'elles permettent d'introduire de la variabilité génétique dans les taxa polyploïdes (Joly et Bruneau, 2004). Il est donc probable que les espèces polyploïdes du complexe aient évolué à plusieurs reprises, mais aucune étude n'a encore exploré cette question.

1.5 Objectifs

Malgré les nombreuses études biosystématiques qui ont été effectués dans ce complexe, il a été impossible de démontrer que les espèces sont vraiment distinctes. Le problème est particulièrement criant au niveau polyploïde puisque ceux-ci sont particulièrement variables et qu'ils peuvent s'hybrider. Mais l'hybridation est aussi présente au niveau diploïde et ceci laisse planer des doutes sur la validité des espèces

présentement reconnues. D'autre part, le manque de variabilité des marqueurs moléculaires utilisés dans les analyses phylogénétiques antérieures ne permettent pas de tirer de conclusion relativement à l'histoire évolutive du complexe. En fonction de ces observations, cette thèse a deux objectifs majeurs.

1) **Délimiter les espèces dans la sect. *Cinnamomeae* à l'est des Rocheuses.**

   Afin de ne pas biaiser les résultats, il est important d'utiliser une méthode objective, qui ne se base pas sur des identifications faites *a priori*, dans le but de délimiter les espèces dans ce groupe. Des méthodes multivariées seront utilisées ici étant donné l'échec des méthodes univariées employées dans le passé. Aussi, en plus d'utiliser la morphologie, cette thèse utilisera des marqueurs génétiques qui n'étaient pas disponibles au temps des dernières études sur le complexe (Lewis, 1957b).

2) **Reconstruire l'évolution des espèces de la sect. *Cinnamomeae* à l'est des Rocheuses**

   Retracer l'évolution de ces espèces permettra de mieux comprendre l'impact de l'hybridation et de la polyploïdie dans ce groupe. Cependant, la polyploïdie complexifie la reconstruction phylogénétique dans ce complexe à cause de la présence potentielle d'allopolyploïdes et de la complexité de reconstruire l'évolution d'espèces hybrides (Linder et Rieseberg, 2004). Ainsi, l'étude de l'évolution des espèces du complexe se fera en deux étapes :

   a) *Reconstruire l'évolution des diploïdes*. Il est en effet important de bien connaître les relations entre les diploïdes avant de pouvoir espérer déterminer celles entre les polyploïdes. Des marqueurs moléculaires plus variables que ceux utilisés dans le passé seront développés pour reconstruire l'évolution des espèces diploïdes.

   b) *Retracer l'évolution des polyploïdes*. Lorsque les relations entre les espèces diploïdes seront bien établies, il sera possible de reconstruire l'évolution des polyploïdes.

Bien connaître l'évolution des polyploïdes permettra, entre autres, de statuer sur la validité de certaines espèces.

La première question abordée dans cette thèse sera celle de la délimitation des espèces (Chapitre 2). Une fois les espèces définies, l'évolution des espèces diploïdes sera étudiée (Chapitres 3 et 4), puis celle des espèces tétraploïdes (Chapitre 5). Il sera ensuite possible de conclure sur le statut des espèces du complexe et d'en reconstruire l'évolution du complexe (Chapitre 6).

CHAPITRE 2

Delimiting species boundaries in *Rosa* sect. *Cinnamomeae* (*Rosa*ceae) in eastern North America[1]

2.1 Résumé

Cette étude se penche la délimitation d'espèces dans un complexe polyploïde, celui de *Rosa* sect. *Cinnamomeae* à l'est des montagnes Rocheuses. Ce complexe est caractérisé par un important polymorphisme intra-spécifique qui est en partie une conséquence de l'hybridation et de la polyploïdie. Des analyses multivariées, telles que des analyses de groupement, de composantes principales, de coordonnées principales et de redondance, ont été utilisées sur 25 caractères morphologiques quantitatifs et sur des marqueurs AFLP (*Amplified fragment length polymorphisms*) afin de délimiter les frontières d'espèces dans ce complexe. L'étude de polyploïdes peut être problématique puisque comme ils sont potentiellement d'origine hybride, ils peuvent occuper des positions intermédiaires entre leurs progéniteurs et peuvent donc obscurcir les distinctions d'espèces dans les analyses multivariées. Pour cette raison, les polyploïdes ont été séparés *a priori* des diploïdes en utilisant la longueur des cellules de garde des stomates afin qu'ils puissent être analysés indépendamment. Quatre espèces ont été trouvées au niveau diploïde : *R. blanda* – *R. woodsii*, *R. foliolosa*, *R. nitida* et *R. palustris*. Il est impossible de distinguer le *R. blanda* du *R. woodsii* selon les données morphologiques et moléculaires. Au niveau polyploïde, trois espèces ont été identifiées, *R. arkansana*, *R. carolina* et *R. virginiana*, même s'il y a hybridation entre celles-ci. La similarité des individus des espèces polyploïdes avec les

---

[1] Joly, Simon et Anne Bruneau. Article non publié

individus des différentes espèces diploïdes, ainsi que les analyses d'ordination combinant les diploïdes et les polyploïdes, ont permis d'identifier les parents potentiels aux polyploïdes. *Rosa arkansana* proviendrait de *R. blanda* – *R. woodsii*, *R. carolina* d'un croisement entre *R. blanda* – *R. woodsii* et *R. palustris*, et *R. virginiana* de *R. nitida*, de *R. palustris*, ou d'un croisement entre ces deux espèces diploïdes.

2.2 Abstract

This study investigates species boundaries in the polyploid complex of *Rosa* sect. *Cinnamomeae* east of the Rocky Mountains. This complex is characterized by extensive intra-specific polymorphism that is the consequence, in part, of hybridization and polyploidy. Multivariate analyses such as cluster analyses, principal component analyses, principal coordinate analyses and redundancy analyses were used on 25 quantitative morphological characters and on amplified fragment length polymorphisms (AFLPs) markers to delimit species boundaries in this complex. Because polyploid individuals are morphologically extremely polymorphic and fill the gaps between diploid species, they were discriminated *a priori* using stomata guard cell lengths in order to investigate species boundaries at each ploidy level separately. Four distinct species are found at the diploid level: *R. blanda* – *R. woodsii*, *R. foliolosa*, *R. nitida* and *R. palustris*. According to the morphological and molecular data, *R. blanda* and *R. woodsii* are indistinguishable and should be considered as a single species. Three species are also identified at the polyploid level, *R. arkansana*, *R. carolina* and *R. virginiana*, albeit with evidence of hybridization between them. The similarity between individuals of the polyploid species and those of the different diploid species, and the ordination analyses of diploids and polyploids together, allowed us to identify the possible parents of the polyploid species. *Rosa arkansana* likely originated from *R. blanda* – *R. woodsii*, *R. carolina* from a hybrid between *R. blanda* – *R. woodsii* and *R. palustris*, and *R. virginiana* from either *R. nitida*, *R. palustris*, or a hybrid between these two species.

2.3 Introduction

Species delimitation in the genus *Rosa* has always been challenging. François Crépin summarized well the historical taxonomic work on this genus: "species [of the genus *Rosa*] become more obscure and less recognizable as the work upon them has multiplied" (Crépin, 1896). This taxonomic complexity is for the most part the consequence of the important polymorphism of *Rosa* species. Similar problems can be found in eastern North America among the native species of section *Cinnamomeae* that are also characterized by extensive intra-specific polymorphism. But where some people saw polymorphism, others saw distinct species. This explains why Rydberg (1920) documented 23 native species in section *Cinnamomeae* in eastern North America whereas Erlanson (1966) and Lewis (1957b) recognized only nine species in the same region (all excluding hybrid species).

This study focuses on this group, the *Rosa* section *Cinnamomeae* east of the Rocky Mountains, which forms a polyploid species complex. It comprises five native diploid species, *R. blanda* Ait, *R. foliolosa* Nutt., *R. nitida* Wild., *R. palustris* Marsh. and *R. woodsii* Lindl., and three tetraploid species, *R. arkansana* Porter, *R. carolina* L. and *R. virginiana* Mill. The ninth species of *Rosa* sect. *Cinnamomeae* in eastern North America recognized by Erlanson (1966), *R. acicularis* Lindl. ($2n$ = 6x, 8x), is morphologically distinct from the other species (Lewis, 1957b) and will not be included in this study. Previous studies on this complex have focussed on morphology (Erlanson, 1930, 1934; Lewis, 1957b, 1958, 1959, 1962), cytology (Erlanson, 1929; Lewis, 1957b, 1966) and experimental crosses (Erlanson, 1934; Ratsek *et al.*, 1939; Ratsek *et al.*, 1940; Lewis and Basye, 1961). Although this important biosystematic work has greatly stabilized the number of species accepted, and though some species recognized by Rydberg (which are still used in some flora, e.g., The Flora of Canada – Scoggan, 1978) were shown indistinguishable from one of the nine species recognized by Erlanson (e.g., Bruneau *et al.*, 2005), species delimitation in this complex remains problematic. Taxonomic problems are known at the diploid level where some species hybridize and are morphologically difficult to tell apart (e.g., *R. blanda* Ait.

and *R. woodsii* Lindl.; Lewis, 1962), but the problem is particularly acute at the polyploid level. *Rosa carolina*, a widespread species that occurs east of the Mississippi river, is known to hybridize with *R. arkansana*, a species of the prairies, in the western portion of its distribution (Lewis, 1957b; Erlanson MacFarlane, 1966), but also in the east with *R. virginiana* (Fernald, 1922; Lewis, 1957b), which occurs along the Atlantic coast. The polyploid species are morphologically similar to each other, but they are also similar to the diploid species. As a result, no single morphological character can be used to distinguish one species from another. Moreover, although the polyploids were suggested to have independent origins (Joly *et al*., 2006), it is still not clear whether these represent distinct evolutionary entities.

To delimit species boundaries, it is important to first determine what a species is and how it is to be recognized in nature. The position taken in the present paper is that species are unique and that they are distinct from other hierarchic classification levels (Ghiselin, 1975; Hull, 1976). "Species are the real units of evolution" (Mayr, 1969). It is at the species level that adaptations, the end product of natural selection, are fixed and allowed to be passed to sibling species. Characteristics that are important for making species the unit of evolution are their spatiotemporal continuity with potential for evolution and their cohesiveness (Hull, 1976). In this sense, species can be viewed as the most inclusive group of organisms that has the potential of maintaining cohesion and that evolve independently from other such groups. These characteristics allow predictions to be made in order to identify species in nature. Indeed, if a species is a cohesive group of organisms, and if they evolve independently from other such groups, then they should eventually become morphologically or genetically distinct with time. Because genetic and morphological differentiation may be reached at different times, a group of individuals will be considered to form a distinct species if it is morphologically or genetically distinct. Intentionally, the nature of the mechanisms that lead to speciation and that are responsible for maintaining cohesion within species is not mentioned. It is principally on this topic that most species concepts differ, in part because the relevant mechanisms vary among groups of organisms. However, whether the cohesive

mechanisms imply gene flow, reproductive systems, competition, ecology, etc., all will eventually create genetically and morphologically distinct species with time.

Even though the delimitation of species in any group showing extensive morphological polymorphisms is never simple, investigation of species boundaries in a polyploid species complex is even more problematic. Polyploid individuals are generally reproductively isolated from their diploid parents and therefore form distinct species. Yet, a recent autopolyploid that resembles its parental species will be difficult to distinguish from it. Also, because hybrid individuals commonly show intermediate morphological (and genetic) characteristics relative to their parents (Schilling and Heiser Jr., 1976; Neff and Smith, 1978; McDade, 1997), and because recombination can create any intermediate between the parents (Anderson and Hubricht, 1938; Jensen and Eshbaugh, 1976; Jensen *et al.*, 1993), including allopolyploids with their diploid ancestors in an analysis could easily blur the species boundaries of both diploids and polyploids. Therefore, the best strategy for delimiting species in polyploid complexes is to discriminate diploids and polyploids *a priori* and to analyse them independently.

In this study, species boundaries were defined objectively without using *a priori* taxonomic identifications, the null hypothesis being that a single species is present in the complex. Taxonomic identifications were only discussed *a posteriori*. To define species in this complex, we investigated quantitative morphological characters and amplified fragment length polymorphisms (AFLP) as molecular markers. Because AFLP amplify random fragments in the genome, results obtained from them should be less biased by selection than individual genetic markers (e.g., single-copy nuclear genes) and should more accurately reflect the overall similarity between individuals. Species boundaries were investigated by looking for gaps in the morphological and molecular variation of organisms using several multivariate methods including ordinations, clustering methods and canonical analyses. The goals of the present study were threefold: (1) identify the species boundaries in this complex; (2) identify the morphological characters that can

help differentiate the species; (3) investigate potential parental relationships for the polyploid species.

2.4 Methods

2.4.1 Morphological analyses

A total of 178 individuals were investigated for the morphological analyses (Table 1). For each species, individuals sampled covered its entire geographic distribution in order to have a good representation of its full extent of variation. In most cases, one individual per population was investigated, eventhough more than one individual was generally available, but there were some cases for *R. palustris* and *R. nitida* where more than one individual per population was sampled.

Henceforth, individuals will be referred to by the *a priori* species name followed by the collector's number (e.g., *nitida*1010-1). The characters selected had to be available on most herbarium specimens in order to be as useful as possible to a botanist wanting to identify these species. Consequently, few floral characters were investigated because these are present only 2-3 weeks on an individual over the whole vegetative season. This should not have biased the results as most characters that have been thought to be important in identifying rose species meet this criterion (Erlanson, 1934; Lewis, 1957b).

Twenty-five quantitative morphological characters were examined and measured using herbarium specimens (characters listed in Table 2). The character values used for each individual were the mean of four to five measurements per specimen, except for the number of leaflets per leaf that was estimated from ten leaves. The number of measurements per specimen was sometimes less if material was insufficient. All length measures were taken using electronic callipers with a precision of 0.01 mm, using a dissection microscope when necessary. Length measures were log transformed (using the

natural logarithm) before the analyses. The distribution of all characters was verified to ascertain that none included extreme outliers.

Table 2.1 List of the specimens included in the multivariate analyses of morphological and molecular data. Specimens are ordered by their *a priori* taxonomic identifications.

Species	Collector[1] (Herbarium)	Provenance	Analyses
R. arkansana	Moss 82 (US)	Canada, Alberta	(mopho)
R. arkansana	Joly 730 (MT)	Canada, Manitoba	(mopho, molecular)
R. arkansana	Joly 738 (MT)	Canada, Manitoba	(mopho, molecular)
R. arkansana	Ryan 3 (MT)	Canada, Saskatchewan	(mopho, molecular)
R. arkansana	Ryan 7 (MT)	Canada, Saskatchewan	(mopho, molecular)
R. arkansana	Ryan 8 (MT)	Canada, Saskatchewan	(mopho, molecular)
R. arkansana	Cary 66 (US)	U.S.A., Colorado	(mopho)
R. arkansana	Allen (1873) (US)	U.S.A., Idaho	(mopho)
R. arkansana	Hayden 11581 (US)	U.S.A., Iowa	(mopho)
R. arkansana	Joly 601 (MT)	U.S.A., Iowa	(mopho, molecular)
R. arkansana	Lewis 15792-2 (MO)	U.S.A., Kansas	(mopho, molecular)
R. arkansana	Joly 655 (MT)	U.S.A., Minnesota	(mopho, molecular)
R. arkansana	Joly 663 (MT)	U.S.A., Minnesota	(mopho, molecular)
R. arkansana	Joly 673 (MT)	U.S.A., Minnesota	(mopho, molecular)
R. arkansana	Standley 17604 (US)	U.S.A., Montana	(mopho)
R. arkansana	Arsène 17732 (US)	U.S.A., New Mexico	(mopho)
R. arkansana	Joly 763 (MT)	U.S.A., North Dakota	(mopho, molecular)
R. arkansana	Ruth 726 (US)	U.S.A., Texas	(mopho)
R. arkansana	Joly 605 (MT)	U.S.A., Wisconsin	(mopho, molecular)
R. arkansana	Tweedy 3224 (US)	U.S.A., Wyoming	(mopho)
R. blanda	Joly 699 (MT)	Canada, Manitoba	(mopho, molecular)
R. blanda	Joly 722 (MT)	Canada, Manitoba	(mopho, molecular)
R. blanda	Joly 409 (MT)	Canada, New Brunswick	(mopho, molecular)
R. blanda	Joly 962 (MT)	Canada, New Brunswick	(mopho, molecular)
R. blanda	Joly 988 (MT)	Canada, New Brunswick	(mopho, molecular)
R. blanda	Joly 993 (MT)	Canada, New Brunswick	(mopho, molecular)
R. blanda	Joly 935 (MT)	Canada, Nova Scotia	(mopho, molecular)
R. blanda	Joly 582 (MT)	Canada, Ontario	(mopho)
R. blanda	Joly 788 (MT)	Canada, Ontario	(mopho, molecular)
R. blanda	Bruneau 1214 (MT)	Canada, Québec	(mopho, molecular)
R. blanda	Bruneau 1219 (MT)	Canada, Québec	(mopho, molecular)
R. blanda	Bruneau 1236 (MT)	Canada, Québec	(mopho, molecular)
R. blanda	Bruneau 1239 (MT)	Canada, Québec	(mopho)
R. blanda	Drouin 98016 (MT)	Canada, Québec	(mopho, molecular)
R. blanda	Joly 1011-1 (MT)	Canada, Québec	(mopho, molecular)
R. blanda	Joly 784 (MT)	U.S.A., Michigan	(mopho, molecular)
R. blanda	Joly 657 (MT)	U.S.A., Minnesota	(mopho, molecular)
R. blanda	Joly 678 (MT)	U.S.A., Minnesota	(mopho, molecular)
R. blanda	Joly 692 (MT)	U.S.A., Minnesota	(mopho, molecular)
R. blanda	Joly 770 (MT)	U.S.A., Minnesota	(mopho, molecular)
R. blanda	Joly 622 (MT)	U.S.A., Wisconsin	(mopho, molecular)

Species	Collector[1] (Herbarium)	Provenance	Analyses
R. blanda	Joly 636 (MT)	U.S.A., Wisconsin	(mopho, molecular)
R. blanda	Joly 780 (MT)	U.S.A., Wisconsin	(mopho, molecular)
R. carolina	Joly 967 (MT)	Canada, New Brunswick	(mopho, molecular)
R. carolina	Joly 576 (MT)	Canada, Ontario	(mopho, molecular)
R. carolina	Joly 580 (MT)	Canada, Ontario	(mopho, molecular)
R. carolina	Pollard 89 (US)	U.S.A., Alabama	(mopho)
R. carolina	Scully 1326 (US)	U.S.A., Arkansas	(mopho)
R. carolina	Palmer 35234 (US)	U.S.A., Florida	(mopho)
R. carolina	Allard 101 (US)	U.S.A., Georgia	(mopho)
R. carolina	Lewis 15843-1 (MO)	U.S.A., Kansas	(mopho, molecular)
R. carolina	Braun 3117 (US)	U.S.A., Kentucky	(mopho)
R. carolina	Thieret 22907 (US)	U.S.A., Louisiana	(mopho)
R. carolina	Joly 460 (MT)	U.S.A., Massachusetts	(mopho, molecular)
R. carolina	Joly 651 (MT)	U.S.A., Minnesota	(mopho, molecular)
R. carolina	McDougall 1620 (US)	U.S.A., Mississippi	(mopho)
R. carolina	Lewis 15779 (MO)	U.S.A., Missouri	(mopho, molecular)
R. carolina	Lewis 15783 (MO)	U.S.A., Missouri	(mopho)
R. carolina	Lewis 15844 (MO)	U.S.A., Missouri	(mopho, molecular)
R. carolina	Joly 502 (MT)	U.S.A., New Jersey	(mopho, molecular)
R. carolina	Joly 491 (MT)	U.S.A., Pennsylvania	(mopho, molecular)
R. carolina	McDougall 1635 (US)	U.S.A., Tennessee	(mopho)
R. carolina	Lewis 2064 (US)	U.S.A., Texas	(mopho)
R. carolina	Joly 523 (MT)	U.S.A., Virginia	(mopho, molecular)
R. carolina	Joly 524 (MT)	U.S.A., Virginia	(mopho, molecular)
R. carolina	Joly 545 (MT)	U.S.A., West Virginia	(mopho, molecular)
R. carolina	Joly 620 (MT)	U.S.A., Wisconsin	(mopho, molecular)
R. carolina	Joly 775 (MT)	U.S.A., Wisconsin	(mopho, molecular)
R. foliolosa	Erlanson 9529 (MO)	U.S.A., Arkansas	(mopho)
R. foliolosa	Emig 614 (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Emig 758 (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Engelmann (1897) (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Griffith 3484-1 (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Hill 11782 (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Houghton 3968 (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Lewis 15846 (MO)	U.S.A., Oklahoma	(mopho, molecular)
R. foliolosa	Merrill 783 (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Palmer 13079 (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Palmer 42016 (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Palmer 8306 (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Waugh 125 (MO)	U.S.A., Oklahoma	(mopho)
R. foliolosa	Butler 11074 (MO)	U.S.A., Texas	(mopho)
R. foliolosa	Eggert (1899) (MO)	U.S.A., Texas	(mopho)
R. foliolosa	Erlanson 9526 (MO)	U.S.A., Texas	(mopho)
R. foliolosa	Heller 4184 (MO)	U.S.A., Texas	(mopho)
R. foliolosa	Lindheimer 608 (MO)	U.S.A., Texas	(mopho)
R. foliolosa	Lundell 13902 (MO)	U.S.A., Texas	(mopho)
R. foliolosa	O'Kennon 19069A (MT)	U.S.A., Texas	(mopho, molecular)
R. nitida	Joly 941 (MT)	Canada, New Brunswick	(mopho, molecular)

Species	Collector[1] (Herbarium)	Provenance	Analyses
R. nitida	Joly 943 (MT)	Canada, New Brunswick	(mopho, molecular)
R. nitida	Joly 944 (MT)	Canada, New Brunswick	(mopho, molecular)
R. nitida	M.-Victorin 46572 (MT)	Canada, New Brunswick	(mopho)
R. nitida	Brouillet 03-55-1 (MT)	Canada, Newfoundland	(mopho, molecular)
R. nitida	Brouillet 03-55-2 (MT)	Canada, Newfoundland	(mopho, molecular)
R. nitida	Brouillet 03-55-3 (MT)	Canada, Newfoundland	(mopho, molecular)
R. nitida	Joly 1016-1 (MT)	Canada, Newfoundland	(mopho, molecular)
R. nitida	Joly 1016-2 (MT)	Canada, Newfoundland	(mopho, molecular)
R. nitida	Joly 1016-3 (MT)	Canada, Newfoundland	(mopho, molecular)
R. nitida	Joly 1018-1 (MT)	Canada, Newfoundland	(mopho, molecular)
R. nitida	Joly 1018-4 (MT)	Canada, Newfoundland	(mopho, molecular)
R. nitida	Joly 1018-5 (MT)	Canada, Newfoundland	(mopho, molecular)
R. nitida	Smith 8288 (MT)	Canada, Nova Scotia	(mopho)
R. nitida	Fernald 7664 (MT)	Canada, Prince Edward Island	(mopho)
R. nitida	Bergeron 81-39 (MT)	Canada, Québec	(mopho)
R. nitida	Cinq-Mars 66-226 (MT)	Canada, Québec	(mopho)
R. nitida	Hamel 12486 (MT)	Canada, Québec	(mopho)
R. nitida	Joly 1010-1 (MT)	Canada, Québec	(mopho, molecular)
R. nitida	Joly 1010-2 (MT)	Canada, Québec	(mopho, molecular)
R. nitida	Joly 1010-3 (MT)	Canada, Québec	(mopho, molecular)
R. nitida	M.-Victorin 49425 (MT)	Canada, Québec	(mopho)
R. nitida	LeGallo 460 (MT)	France, St.-Pierre et Miquelon	(mopho)
R. palustris	Joly 417 (MT)	Canada, New Brunswick	(mopho, molecular)
R. palustris	Joly 418 (MT)	Canada, New Brunswick	(mopho, molecular)
R. palustris	Joly 573 (MT)	Canada, Ontario	(mopho, molecular)
R. palustris	Taylor 2141 (US)	Canada, Ontario	(mopho)
R. palustris	Bowers 2182 (MT)	Canada, Québec	(mopho)
R. palustris	M.-Victorin 2362 (MT)	Canada, Québec	(mopho)
R. palustris	R.-Germain 7114 (MT)	Canada, Québec	(mopho)
R. palustris	R.-Germain 7115 (MT)	Canada, Québec	(mopho)
R. palustris	Raymond (1947) (MT)	Canada, Québec	(mopho)
R. palustris	Joly 476 (MT)	U.S.A., Connecticut	(mopho, molecular)
R. palustris	Small 8652 (US)	U.S.A., Florida	(mopho)
R. palustris	Duncan 6222 (US)	U.S.A., Georgia	(mopho)
R. palustris	Lewis 2406 (US)	U.S.A., Illinois	(mopho)
R. palustris	Joly 587 (MT)	U.S.A., Michigan	(mopho, molecular)
R. palustris	Joly 588 (MT)	U.S.A., Michigan	(mopho, molecular)
R. palustris	Palmer 6159 (US)	U.S.A., Missouri	(mopho)
R. palustris	Lewis 2305 (US)	U.S.A., Ohio	(mopho)
R. palustris	Lewis 2156 (US)	U.S.A., New Hampshire	(mopho)
R. palustris	Joly 548 (MT)	U.S.A., Pennsylvania	(mopho)
R. palustris	Joly 549 (MT)	U.S.A., Pennsylvania	(mopho, molecular)
R. palustris	Joly 560 (MT)	U.S.A., Pennsylvania	(mopho, molecular)
R. palustris	Joly 561 (MT)	U.S.A., Pennsylvania	(mopho, molecular)
R. palustris	Godfrey 734 (US)	U.S.A., South Carolina	(mopho)
R. palustris	McDougall 1362 (US)	U.S.A., Tennessee	(mopho)

Species	Collector[1] (Herbarium)	Provenance	Analyses
R. palustris	Allard 11491 (US)	U.S.A., West Virginia	(mopho)
R. palustris	Joly 644 (MT)	U.S.A., Wisconsin	(mopho, molecular)
R. virginiana	Hébert 6 (MT)	Canada, New Brunswick	(mopho)
R. virginiana	Joly 431 (MT)	Canada, New Brunswick	(mopho, molecular)
R. virginiana	Joly 946 (MT)	Canada, New Brunswick	(mopho, molecular)
R. virginiana	Joly 973 (MT)	Canada, New Brunswick	(mopho, molecular)
R. virginiana	Brouillet 03-57-1 (MT)	Canada, Newfoundland	(mopho, molecular)
R. virginiana	Brouillet 03-60-1 (MT)	Canada, Newfoundland	(mopho, molecular)
R. virginiana	Joly 1017-1 (MT)	Canada, Newfoundland	(mopho, molecular)
R. virginiana	Joly 1019-1 (MT)	Canada, Newfoundland	(mopho)
R. virginiana	Joly 1015-1 (MT)	Canada, Nova Scotia	(mopho, molecular)
R. virginiana	Joly 924 (MT)	Canada, Nova Scotia	(mopho, molecular)
R. virginiana	Joly 928 (MT)	Canada, Nova Scotia	(mopho, molecular)
R. virginiana	Fernald 7667 (MT)	Canada, Prince Edward Island	(mopho)
R. virginiana	Joly 997 (MT)	Canada, Québec	(mopho, molecular)
R. virginiana	Joly 474 (MT)	U.S.A., Connecticut	(mopho, molecular)
R. virginiana	Joly 444 (MT)	U.S.A., Maine	(mopho, molecular)
R. virginiana	Joly 517 (MT)	U.S.A., Maryland	(mopho, molecular)
R. virginiana	Joly 520 (MT)	U.S.A., Maryland	(mopho, molecular)
R. virginiana	Joly 454 (MT)	U.S.A., Massachusetts	(mopho, molecular)
R. virginiana	Bartram 3668 (MT)	U.S.A., New Jersey	(mopho)
R. virginiana	Joly 496 (MT)	U.S.A., New Jersey	(mopho, molecular)
R. virginiana	Collins (1920) (MT)	U.S.A., Rhode Island	(mopho)
R. woodsii	Dickson 2008 (MT)	Canada, Alberta	(mopho, molecular)
R. woodsii	Lewis 15848-1 (MO)	Canada, British Columbia	(mopho, molecular)
R. woodsii	Lewis 15850-2 (MO)	Canada, British Columbia	(mopho, molecular)
R. woodsii	Joly 741 (MT)	Canada, Manitoba	(mopho, molecular)
R. woodsii	Porsild 16664 (MT)	Canada, Northwest Territories	(mopho)
R. woodsii	Joly 750 (MT)	Canada, Saskatchewan	(mopho, molecular)
R. woodsii	Joly 754 (MT)	Canada, Saskatchewan	(mopho, molecular)
R. woodsii	Ryan 1 (MT)	Canada, Saskatchewan	(mopho)
R. woodsii	Ertter 17989 (JEPS)	U.S.A., California	(mopho, molecular)
R. woodsii	Ertter 18307 (JEPS)	U.S.A., California	(mopho, molecular)
R. woodsii	Joly 1005-1 (MT)	U.S.A., Colorado	(mopho, molecular)
R. woodsii	Joly 1005-2 (MT)	U.S.A., Colorado	(mopho, molecular)
R. woodsii	Joly 1008-1 (MT)	U.S.A., Colorado	(mopho, molecular)
R. woodsii	Ertter 18005 (JEPS)	U.S.A., Idaho	(mopho, molecular)
R. woodsii	Hitchcock 13164 (MT)	U.S.A., Montana	(mopho)
R. woodsii	Ertter 17525 (JEPS)	U.S.A., Nevada	(mopho, molecular)
R. woodsii	Spellenberg 12555 (MT)	U.S.A., New Mexico	(mopho, molecular)
R. woodsii	Joly 758 (MT)	U.S.A., North Dakota	(mopho, molecular)
R. woodsii	Ertter 17990 (JEPS)	U.S.A., Oregon	(mopho, molecular)
R. woodsii	Ertter 18289c (JEPS)	U.S.A., Utah	(mopho, molecular)

[1] Only the principal collector is given when there was more than one collector. When no collection number was indicated on a specimen, the collection date is given in parentheses.

2.4.1.1 Estimation of missing data

Of the 178 individuals sampled, six had missing data for the four floral characters. Because these individuals were also surveyed for the molecular analyses, these missing characters were estimated to avoid removing these individuals from the analyses. For four individuals (*palustris*417, *nitida*1016-1, *virginiana*520, *virginiana*444), other individuals were available from their population and missing data were estimated by taking the population mean for each character. Because it was not possible to use a population mean for the two remaining individuals (*carolina*502, *carolina*15783), the missing data were estimated by multiple linear regression using all non floral characters as independent variables. The conclusions reached by the present study did not change when these individuals were removed from the analyses.

2.4.1.2 Ploidy level determination

To determine the ploidy level of the individuals studied, the length of the stomata guard cells was measured. This character was shown to be strongly correlated to the ploidy level of individuals (Lewis, 1957b) and to be useful in discriminating diploid and tetraploid roses (Joly *et al.*, 2006). The average length of the stomata guard cells for each individual was estimated from 20 stomata as described in Joly *et al.* (2006). A K-means analysis of two clusters was used to objectively discriminate the diploids and polyploids. The K-means analysis was performed using the "kmeans" function in R (R Development Core Team, 2005). The results were compared to the ploidy level expected from the *a priori* taxonomic identifications. When these disagreed, we used the pollen size as another way to evaluate the ploidy level. Lewis (1957b) has shown that pollen size also is a robust indicator of the ploidy level of individuals. The size of the pollen was measured using a microscope with a 63 × objective. Because the length of the stomata guard cells was used *a priori* on the individuals to classify them (diploid or polyploid), this character was not used in the multivariate analyses, except when explicitly mentioned.

Table 2.2 Description of the morphological characters used for delimiting species boundaries in *Rosa* sect. *Cinnamomeae* east of the Rocky Mountains.

Abbreviation	Description
Leaf characters	
NLFT	Number of leaflets per leaf
NSER	Number of leaflet serrations on one side of the terminal leaflet, including the terminal serration
N2SER	Number of double serrations on one side of the terminal leaflet
LLFT	Length of the terminal leaflet, from the base to the extremity of the limb of the leaflet, in mm
L1SER	Length of the terminal leaflet along the primary vein from the base of the limb to the point perpendicular to the first serration on either side of the leaflet, in mm
WLFT	Width of the leaflet at the widest point of the leaflet, in mm
LWLFT	Length along the primary vein of the terminal leaflet from the base of the limb to the point perpendicular to the greatest width of the leaflet, in mm
LTEET	Depth of the teeth sinuses near the middle of the leaflet, in mm
PULFT	Number of hairs on the adaxial surface of the terminal leaflet in a 1.71 × 1.71 mm area
P1VEIN	Number of hairs along 1 cm of the primary vein on the abaxial side of the terminal leaflet
P2VEIN	Number of hairs along 1 cm of a secondary vein on the abaxial side of the terminal leaflet
PBLFT	Number of hairs on the abaxial surface of the terminal leaflet in a 1.71 × 1.71 mm area
G1VEIN	Number of glands along 1 cm of the primary vein on the abaxial side of the terminal leaflet
GBLFT	Number of glands on the abaxial surface of the terminal leaflet in a 1.71 × 1.71 mm area
Stipule characters	
LSTP	Length of the part of the stipule that is adnate to the petiole, in mm
LAUR	Length from the point where the stipule diverges from the petiole to the extremity of the auricle, in mm
WAUR	Width of the auricle, from the middle of the petiole to the portion of the stipule that is furthest from the rachis, in mm
WSTP	Width of the stipule at half length, in mm, from the middle of the petiole to the blade of the stipule, at the middle of the portion of the stipule attached to the rachis
GSTP	Number of glands along the border of the stipule on one side of the petiole
Thorn and bristle characters	
B1YW	Number of bristles on 1 cm of current-year stem
T1YW	Proportion of leaves below which there are infrastipular thorns on branches of the current year. Infrastipular thorns are always in pairs on either side of the leaf and immediately below its point of attachment
Floral characters	
NFLW	Number of flowers per inflorescence
GHYP	Number of glands found on one side of the hypanthium
GPED	Number of glands on one side of the pedicel over its whole length
LPED	Length of the pedicel, in mm

2.4.2 Molecular analyses

Of the 178 individuals included in the morphological analysis, 108 were also investigated at the molecular level using AFLPs (Table 1). This represents the specimens that were collected in the field and for which we had leaves preserved in silica gel. The sampling for the molecular analysis is therefore not has thorough as that for the morphological analyses. Species that were most affected by this reduced sampling are *R. foliolosa*, for which only two populations were sampled, and *R. carolina* and *R. palustris* for which we lack populations from the south of their distribution. The impact of this limited sampling on the results is discussed later.

DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method of Doyle and Doyle (1987) as modified in Joly *et al.* (2006). Amplified fragment length polymorphisms (AFLPs; Vos *et al.*, 1995) were used to characterize the genetic constitution of individuals at the genomic level. We followed the protocol of Applied Biosystems (Foster City, California, USA) for large genomes with the modifications described by Bruneau *et al.* (2005). From a preliminary screening of 18 different primer combinations, four combinations were selected that showed the greatest number of polymorphic bands: EcoRI-AAG + MseI-CAC, EcoRI-ACA + MseI-CAC, EcoRI-AAG + MseI-CTG, and EcoRI-ACA + MseI-CTG. Two primer combinations (differentiated by distinctive chromophores) were run simultaneously on an ABI3100-*avant* sequencer (Applied Biosystems). Unambiguous bands were scored using the Genographer software (Benham, 2001).

2.4.3 Analysis of the diploids

In order to identify distinct groups of diploid individuals, cluster analyses were performed on the morphological dataset, on the molecular dataset, and on a combined dataset. Because it makes sense to think of species as a group of individuals forming a hyperspherical cluster, i.e. spherical in multidimensional space, Ward's minimum

variance method was used (Ward, 1963). Ward's phenograms were obtained from the "agnes" function of the "cluster" package in R (Maechler, 2005). The Euclidean distance from standardized variables was used for the morphological matrix. For the molecular dataset, the Jaccard distance (Jaccard, 1900) was used to avoid considering the shared absence of a band as a similarity. Ward's phenogram were also obtained from the combined dataset. The distance matrix used for the combined analysis was the mean of the morphological and the molecular matrices, recomputed to include only the individuals that had both morphological and molecular information. To give the two datasets approximately the same weight, morphological and molecular matrices were scaled so that the maximum distance in each matrix was 1.

The differentiation of the groups also was assessed by Principal Component Analysis (PCA) for the morphological data and by Principal Coordinate Analysis (PCoA) for the molecular and the combined datasets. The PCA was performed using the "prcomp" function in R from the correlation matrix, scaling the character vector lengths to 1. The PCoA analysis was performed using the "cmdscale" function in R from the Jaccard distance matrix for the molecular dataset. For the combined analysis, the PCoA was performed on the same matrix that was used for the combined cluster analysis.

It was impossible to distinguish the species defined *a priori* as *R. blanda* and *R. woodsii* in the present analyses. Because a hybrid zone has been previously described between these species, and because they are mainly distributed along an east-west transect, the variation for each morphological character was plotted as a function of longitude to determine whether the observed pattern is in agreement with a hybrid zone hypothesis. A pattern typical of a hybrid zone should show fixed morphological differences in individuals far from the hybrid zone and a gradual change from the values of one species to those of the other near the hybrid zone, and this for all morphological characters. Only the characters that showed a significant variation among the two *a priori* species according to an ANOVA were investigated; character values standardized using the data of the two species were plotted. The hybrid zone was assumed to range from

western North Dakota to western Minnesota, approximately from longitudes 104°W to 95°W (Lewis, 1962).

To investigate further the lack of differentiation between *R. blanda* and *R. woodsii*, the amount of variation among individuals of these species that can be explained by the *a priori* species identification was evaluated using Redundancy Analysis (RDA). Also, because these two species are widely distributed, a partial RDA was performed to investigate the proportion of the variance explained by the species identifications while controlling for the geographic distribution. The RDA is a direct extension of multiple regression when the response variable is multivariate. Therefore, it allows to quantify the proportion of the variation that can be explained by one or more variable (e.g., species identification), or in the case of the partial RDA, by one or more variable while controlling for another one (e.g., geographic distribution). It is also possible to test whether the proportion of variation explained by one variable is significant using permutations (Legendre and Legendre, 1998; Peres-Neto *et al.*, in press). These analyses are similar to the Mantel test (Mantel, 1967), although they are more powerful because they are performed on the raw data matrices rather than on distance matrices obtained from the data. The analyses were done with the standardized morphological matrix and with the molecular matrix on which a Hellinger transformation (Legendre and Legendre, 1998) was applied to avoid the double-zero problem in RDA. The matrix of species consisted in a column of "0" and "1" to differentiate the two species and the matrix of geographic coordinates consisted of latitude and longitude columns. These analyses were performed with the "vegan" package (Oksanen *et al.*, 2005) in R; adjusted R^2 values (Peres-Neto *et al.*, in press) were used in all cases. The significance of the adjusted R^2 values obtained by RDA were tested by permutation (Legendre and Legendre, 1998; Peres-Neto *et al.*, in press).

2.4.4 Analysis of the polyploids

The position of the polyploid individuals relative to each other and to the diploid species was assessed using PCA for the morphological dataset and PCoA for the molecular dataset. The same parameters as in the analyses of the diploids were used. We also compared the overall morphological and molecular similarity of the individuals of each polyploid with the individuals of each diploid. The distribution of the pairwise distances of each polyploid species with each diploid species was based on the standardized matrix of Euclidean distances for the morphological data and on the Jaccard distance matrix for the molecular dataset. For each polyploid, the mean distance obtained with the different diploid species were tested using Tukey's HSD test (5% level - Tukey, 1953; Kramer, 1956) to determine whether the difference was significant.

2.4.5 Classification tree

Although distinct groups of individuals were found in the analysis, the results showed that it is impossible to differentiate any species from the others by using a single character (see results, section 2.5). Yet, it may be still possible to properly identify species if a hierarchic method is used, such as classification trees. Classification trees aims at recovering only pure species by using a hierarchic key, which can also be helpful for differentiating species and to identify characters that are most useful in delimiting species or groups of species. Classification trees were therefore constructed using the species identified in the present study (see Results, section 2.5). Two trees were constructed: both included all characters, but one also included the length of the stomata guard cells. The minimum number of individuals at a node and in a newly formed group was set to ten and five, respectively. The analysis was performed using the "tree" package in R (Ripley, 2005).

2.5 Results

2.5.1 Ploidy level determination

The distribution of the lengths of the stomata guard cells showed two modes that also correspond to the K-means clusters (Fig. 2.1). The K-means cluster with the smallest length consisted of 114 individuals and had a mean length of 17.43 μm, whereas the other cluster comprised 64 individuals and had a mean length of 22.86 μm. These two clusters were assumed to include diploid and polyploid individuals, respectively. According to the *a priori* taxonomic identifications, only 6 individuals (3.4 %) were misclassified: *foliolosa*4184, *woodsii*2008, *carolina*967, *virginiana*454, *virginiana*03-57-1, and *virginiana*6. Pollen grains were available for two individuals and they invalidated the results of the stomata guard cells (*virginiana*03-57-1: 36.3μm ± 4.1, *virginiana*454: 35.1 μm ± 4.6; see Lewis, 1957b for mean grain size for diploids and polyploids). For both these individuals, the result of the pollen grain size was used because it also agreed with the *a priori* taxonomic identification. For the other specimens that lacked pollen information, the ploidy level obtained from the K-means analysis was used.

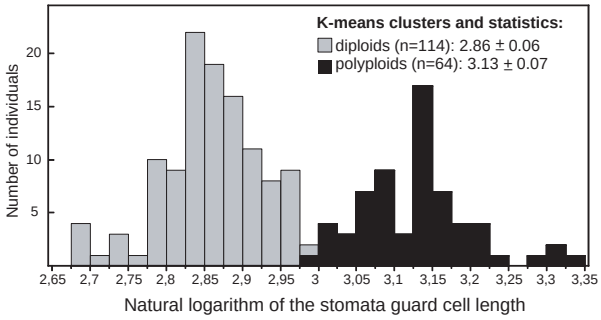


Figure 2.1 Distribution of the natural logarithm of the length of the stomata guard cell in the *Rosa* individuals surveyed. The k-means clusters and statistics that correspond to the diploid and polyploid individuals are given.

2.5.2 Relationships among diploids

2.5.2.1 Morphology

The cluster analysis (Fig. 2.2a) and the principal component analysis (PCA; Fig. 2.3) give complementary information for determining the number of distinct groups of individuals present in the data. The cluster analysis defines groups of individuals from all the variables included in the study, whereas the PCA illustrates the dispersion of individuals in the few dimensions that explain the greatest amount of variation. PCA are therefore useful to identify gaps in the variation and they also can give an idea of the characters that are the most important to discriminate among groups of individuals.

The phenogram obtained from the morphological dataset suggested that there were six relatively well differentiated clusters (Fig. 2.2a). The *a priori* taxonomic identifications closely matched the clusters found: cluster 1 and 2 both contained *R. blanda* and *R. woodsii* individuals, cluster 3, 4, and 5 respectively contained *R. nitida*, *R. foliolosa*, and *R. palustris* individuals with only a few misclassifications (5 out of 70), and cluster 6 consisted of three *R. woodsii* individuals.

According to the Kaiser-Guttman criterion, only the first three principal components of the PCA (Fig. 2.3) are meaningful, as only these had eigenvalues above the mean of all eigenvalues (data not shown; Legendre and Legendre, 1998) (the correlation matrix used in the PCA is given in appendix 1). Because the length of each character vector was scaled to 1 and because 25 characters were included in the analysis, a variable is considered to contribute significantly to a principal component (PC) if its projection on the PC is greater than 0.2. The first PC was strongly influenced by size, as all length characters showed strong negative loadings on this axis (character loadings are given in appendix 2). However, the first PC was also influenced by other variables such as pubescence on the primary and secondary veins. Therefore, the first axis was retained for interpretation to avoid eliminating the discriminating effect of some characters.

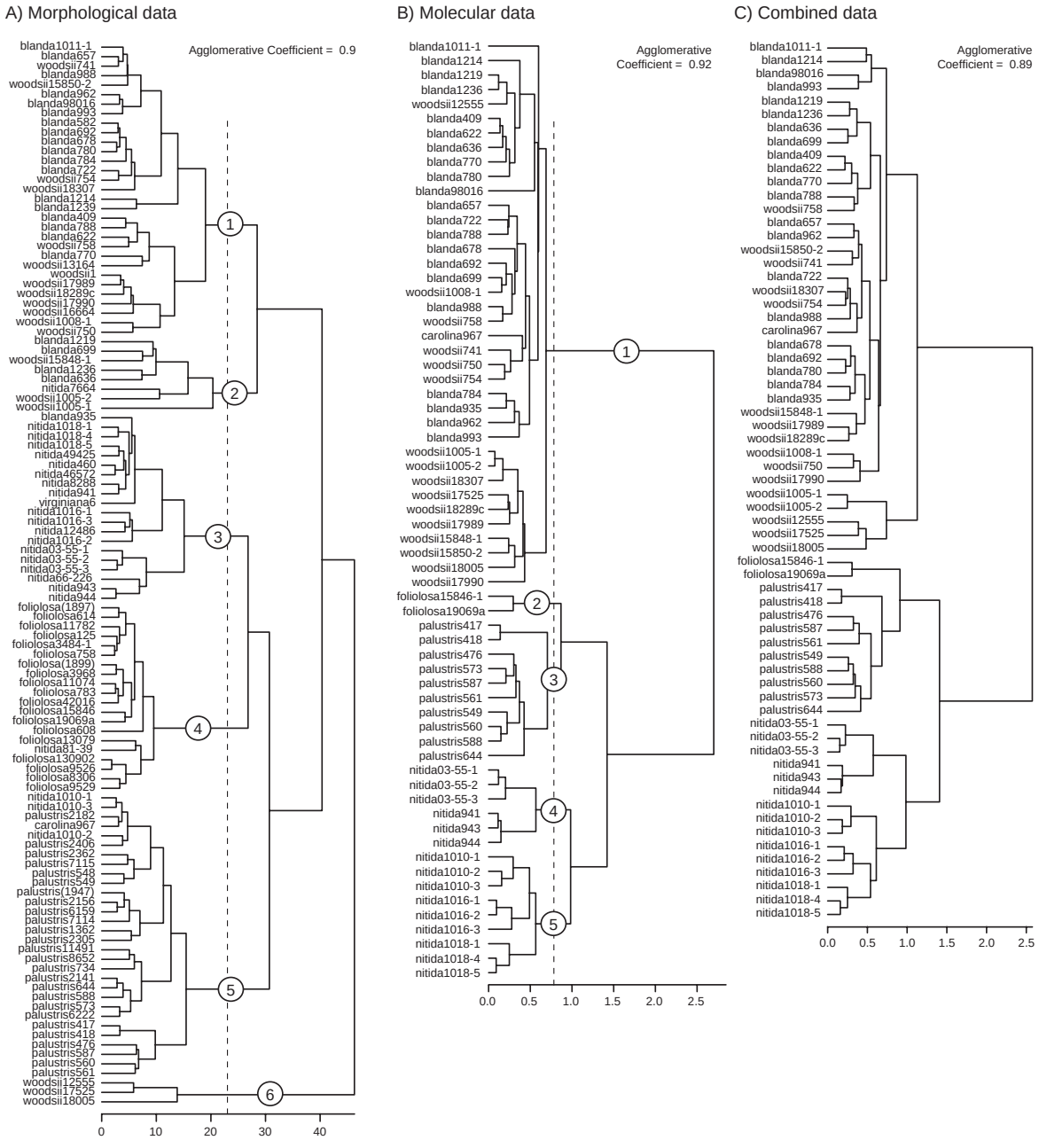


Figure 2.2 Phenograms of diploid individuals obtained using Ward's minimum variance method from the morphological (A), molecular (B), and combined (C) datasets. The *a priori* taxonomic identification of individuals is shown at the left of each phenogram. For the morphological and molecular datasets, the major groups are numbered in order to facilitate reference with the ordination analyses.

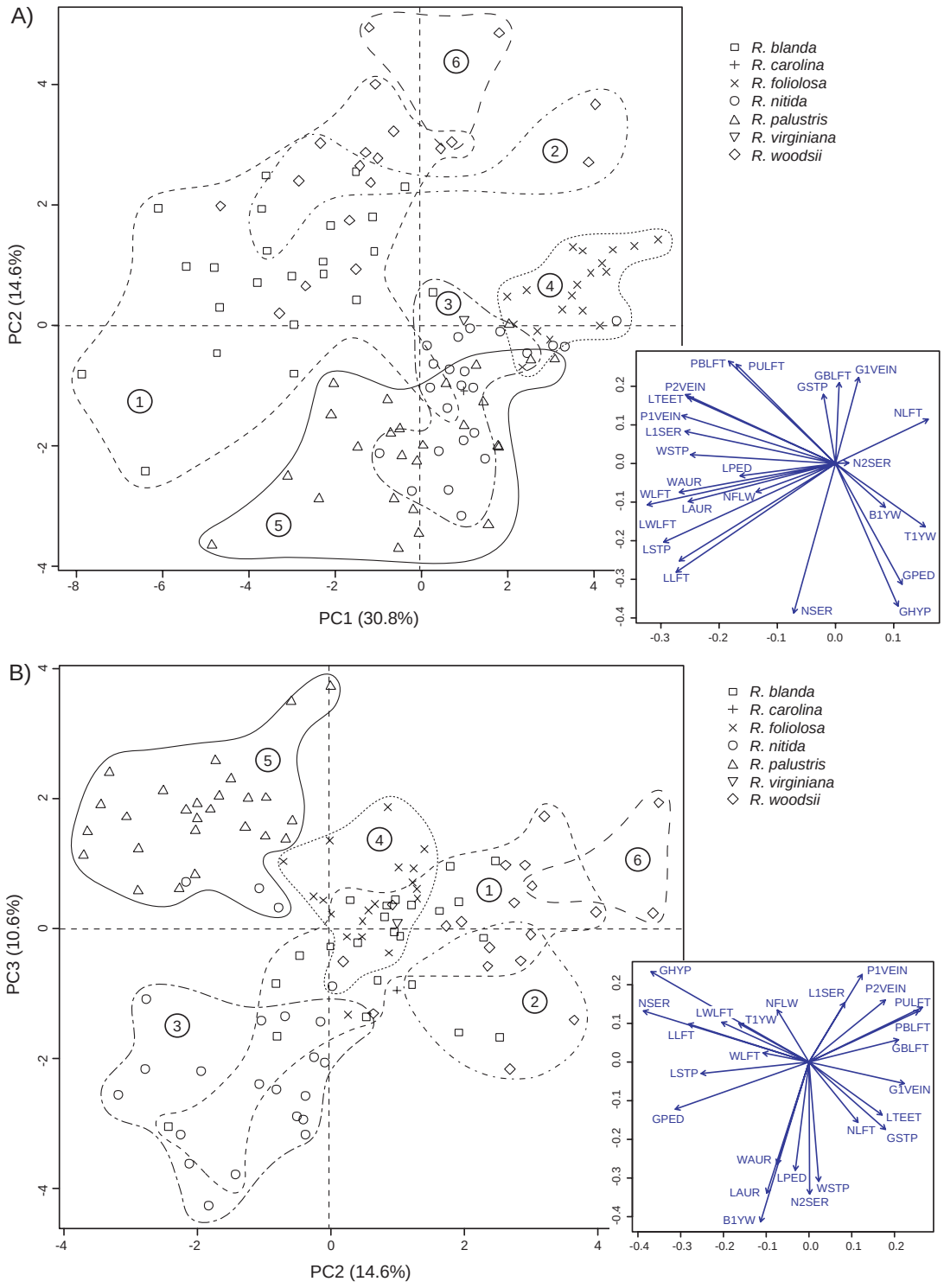


Figure 2.3 Principal component analysis of the morphological data for the diploid individuals. The insets represent the character vectors, which were scaled to 1 in the analysis and therefore are orthogonal to each other in multidimensional space (loadings in appendix 2). The percentage of the variance explained by each axis is noted. The *a priori* species identifications are indicated by symbols. The outlines and their associated numbers represent the major groups found in the cluster analysis of the morphological data (Fig. 2a). The *R. carolina* and the *R. virginiana* individuals that were considered to be diploids (see text) were not included in the outlines.

Ordination on the first and second PC identified two distinct groups of individuals. The group on the upper left consisted of the species *R. blanda* and *R. woodsii* from *a priori* taxonomic identification, and the lower right group comprised species *R. foliolosa*, *R. nitida*, and *R. palustris* (Fig. 2.3a). The characters that contributed the most to separate these two groups were the pubescence of leaflets (more abundant for *R. blanda* and *R. woodsii* than for *R. foliolosa*, *R. nitida* and *R. palustris*) and the glands on the hypanthium and on the pedicels (generally absent in *R. blanda* and *R. woodsii*, and present in *R. foliolosa*, *R. nitida* and *R. palustris*). Most of the length characters pointed toward the lower-left, so the length characters explained an important portion of the variation within these two groups. On the ordination of the second and third PC, three other groups of individuals were distinguished (Fig. 2.3b). These species were *R. foliolosa*, *R. nitida* and *R. palustris*, as identified *a priori*. The character vectors showed that the group that consisted of *R. palustris* individuals was characterized by more glands on the hypanthium and a higher number of serrations on the terminal leaflet than in other species. The group that comprised *R. nitida* individuals was differentiated by numerous bristles on branches, by long and wide auricles, more secondary serrations than in other species in general, and fewer hairs on the first vein of the leaflet. Finally, *R. palustris* and *R. nitida* tended to have more glands on pedicels than the other species.

In general, the groups found in the cluster analysis (Fig. 2.2a) also appeared to be distinct in the PCA (Fig. 2.3). Cluster 3, 4, and 5, which corresponded generally to *R. foliolosa*, *R. nitida* and *R. palustris* (see above), were distinct in the ordination of the second and third PCs. One individual notwithstanding (*nitida*7664, which creates an incursion of cluster 1 into the lower-right part of the ordination of the first two PC), these three groups were also distinct from clusters 1, 2 and 6 as shown on the ordination of PC 1 and 2. In contrast, clusters 1, 2 and 6 did not form distinct groups in the PCA. Therefore, it is more conservative to treat clusters 1, 2 and 6 as a single group of diploid individuals. This large group consisted of all but one *R. blanda* and all *R. woodsii* individuals.

2.5.2.2 Molecular data

A total of 107 AFLP bands were scored. Within diploids, eight bands were constant in all individuals and three were unique. The phenogram obtained with Ward's minimum variance method identified five major groups (Fig. 2.2b). According to the *a priori* taxonomic identifications, one group consisted of *R. blanda* – *R. woodsii*, a second of *R. foliolosa*, another of *R. palustris* and the last two clusters consisted of groups of *R. nitida* individuals. With the exception of the splitting *R. nitida* individuals into two groups, there were no misclassifications among these groups. The PCoA also delimited five distinct groups, which were the same as those found in the cluster analysis (Fig. 2.4). The two groups of *R. nitida* individuals were close to each other on the first two principal coordinates, but they were discriminated by the third principal coordinate (Fig. 2.4). In both the cluster analysis and the PCoA, the *R. carolina* individual that was considered to be a diploid based on its stomata guard cell length was placed with the *R. blanda* – *R. woodsii* group.

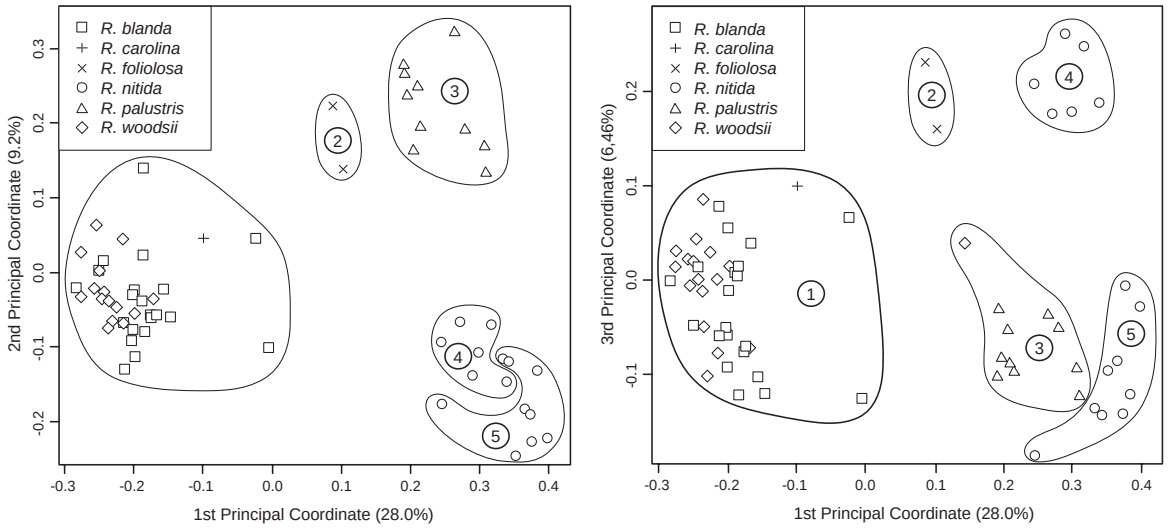


Figure 2.4 Principal coordinate analysis of the molecular data for the diploid individuals. The *a priori* species identifications are shown by symbols and the outlines (with their associated numbers) represent the major groups found in the cluster analysis of the molecular data (Fig. 2b). The percentage of the variance explained by each axis is shown.

2.5.2.3 Combination of the morphological and molecular analyses

The phenogram of the combined analysis was similar to that of the separate analysis in terms of the groups delineated: all *R. blanda* and *R. woodsii* individuals grouped together, and *R. foliolosa*, *R. palustris* and *R. nitida* individuals all formed distinct groups (Fig. 2.2c). However, the distances among individuals within these groups were different, so that if four groups were to be defined at the same distance threshold, *R. foliolosa* would group with *R. palustris* and the *R. blanda* – *R. woodsii* group would be split in two. The PCoA of the combined dataset distinguished the same four groups as the morphological analysis: *R. blanda* – *R. woodsii*, *R. foliolosa*, *R. nitida*, and *R. palustris* (data not shown). The two groups of *R. nitida* individuals that were found in the molecular analysis were not distinguishable in the combined analysis.

In general, the morphological and molecular analyses closely agreed in terms of the distinct groups of individuals identified (Fig. 2.2, 2.3, 2.4). For individuals included in both analyses, the sole incongruence involved the placement of the *R. carolina* 967 individual that occurred in the *R. nitida* cluster in the morphological analysis, but with the *R. blanda* – *R. woodsii* cluster in the molecular analyses.

The general agreement among analyses allows the delimitation of four distinct groups of individuals, which were used in the subsequent analyses: *R. blanda* – *R. woodsii*, *R. foliolosa*, *R. nitida*, and *R. palustris*. The two groups of *R. nitida* found in the molecular analysis will not be considered to represent distinct species because this result was not found in the morphological analysis and did not hold in the combined analysis (further discussed later). When the datasets disagreed regarding the position of some individuals, the grouping found in the combined analyses was used. For example, the *R. carolina* 967 individual was considered to be part of the *R. blanda* – *R. woodsii* group (Fig. 2.2). Individuals that were not included in the molecular analysis and for which the *a priori* identification of individuals did not agree with the multivariate analyses had their identification changed to reflect that of the multivariate analyses.

2.5.2.4 Rosa blanda and R. woodsii

Individuals identified as *R. blanda* and *R. woodsii* could not be distinguished by the morphological or molecular datasets. However, individuals from one species are often closer with conspecific individuals than with individuals of the other species (Figs. 2.2, 2.3, 2.4). To investigate this further, the standardized values of the 16 morphological characters that showed differences among the two *a priori* species were plotted against the longitude (Fig. 2.5). No morphological character showed fixed differences among the individuals of the two *a priori* species, even when only individuals far from the hybrid zone were considered (Fig. 2.5). Some individuals in one species always had values that fell within the range of the other species. Several characters showed a gradual change from west to east (LLFT, L1SER, WLFT, LWLFT, PBLFT, LSTP, LAUR, WAUR, Fig. 2.5), but these changes were not associated with the hybrid zone but rather showed a gradual change from the west to the east. Perhaps the only character for which the values changed near the hybrid zone is the presence of infrastipular thorns on first year wood (T1YW).

The amount of variation found in the morphological and the molecular datasets that can be explained by the *a priori* classification between *R. blanda* and *R. woodsii* and by the geographic distance was also investigated (Fig. 2.6). For the morphological dataset, the *a priori* taxonomy explained 12% of the variation, which was significant ($p < 0.001$; Fig. 2.6). Yet once the geographic distance is controlled for, the variation explained was 2% ($p > 0.05$). Similarly for the molecular dataset, the *a priori* classification accounted for 5% of the molecular variation ($p < 0.001$), but when the geographic distance was accounted for, the variation explained changed to 1%, which was still marginally significant ($p < 0.05$).

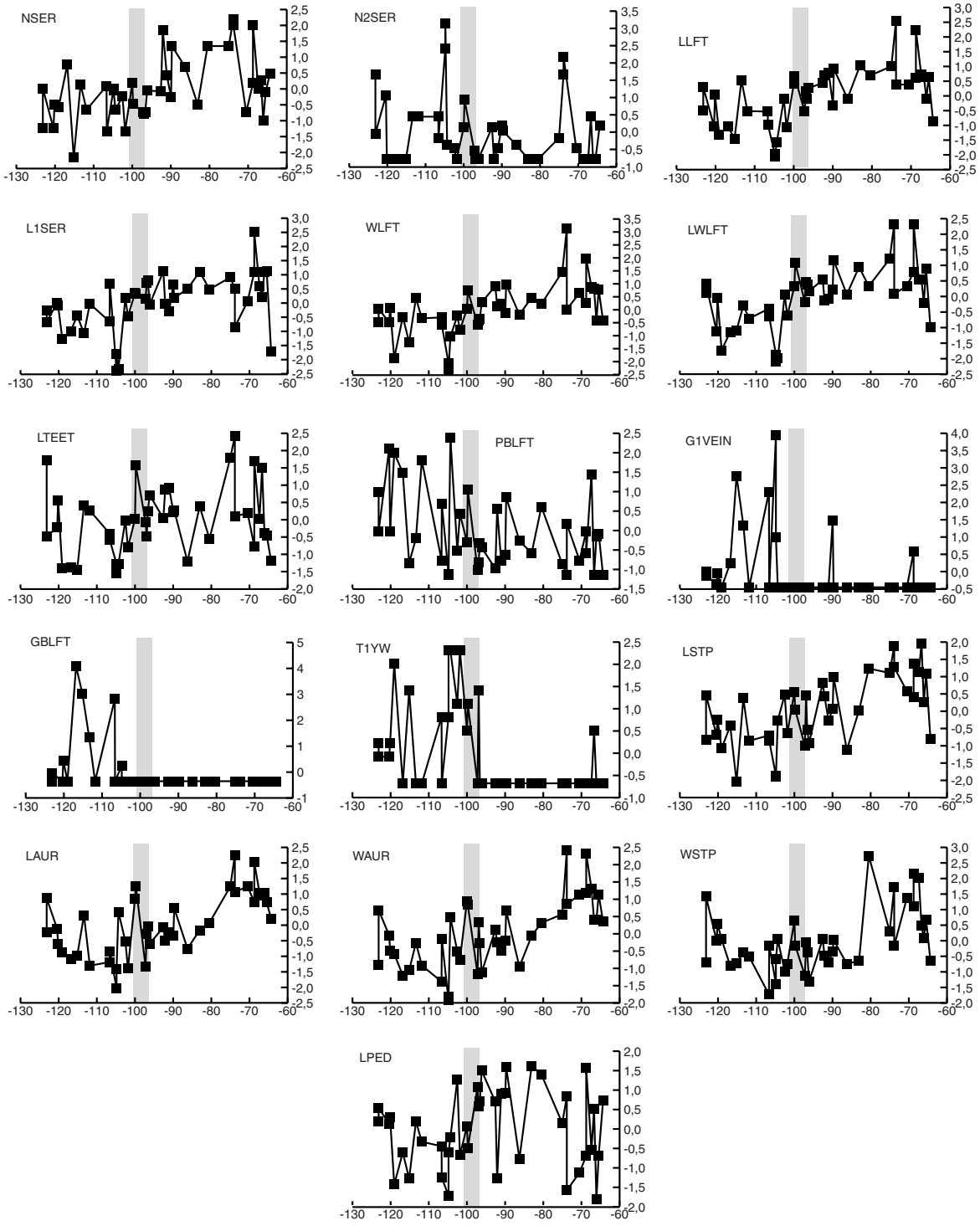


Figure 2.5 Morphological character variation in the *R. blanda* – *R. woodsii* group as a function of the longitude of the individuals sampled, where each square represents an individual. The standardized character values are plotted. The gray boxes represent the hybrid zone area between *R. blanda* and *R. woodsii* according to Lewis (1962), with *R. woodsii* and *R. blanda* located west and east of this zone, respectively.

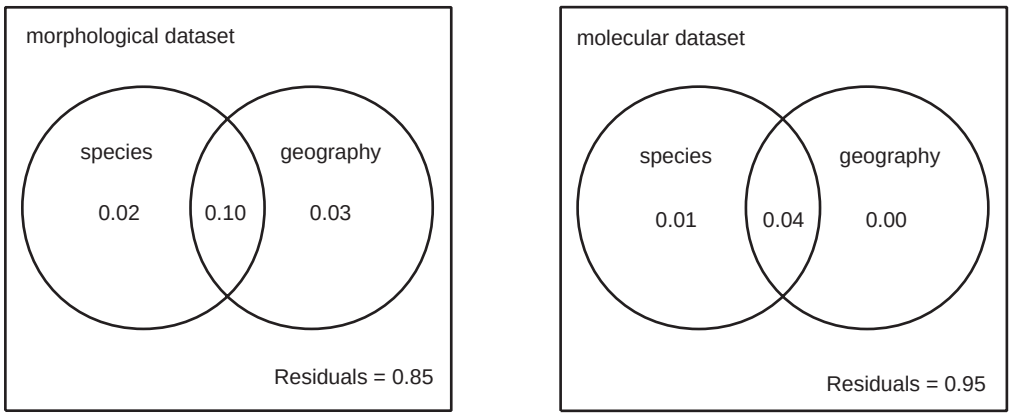


Figure 2.6 Diagrams showing the proportion of variance explained by the *a priori* classification (species), the geographic distance (geography), and these two factors combined for the morphological and the molecular dataset. The values shown are adjusted R^2 values calculated by redundancy analysis. The size of the circles are not proportional to the percentage of the variance explained.

2.5.3 Polyploids

2.5.3.1 Morphology

The PCA for the morphological data with both diploid and polyploid individuals is shown in figure 2.7 (the correlation matrix used in the PCA is given in appendix 3). According to the Kaiser-Guttman criterion, only the first three PCs were meaningful (data not shown). The length and relative directions of the character vectors (Fig. 2.7; character loading in appendix 4) were similar to those obtained in the analysis of the diploid individuals alone (Fig. 2.3). This may be explained by the fact that the variation in the polyploid individuals is included within that of the diploid individuals. Again, a variable is considered to contribute significantly to a PC if its projection on the PC is greater than 0.2.

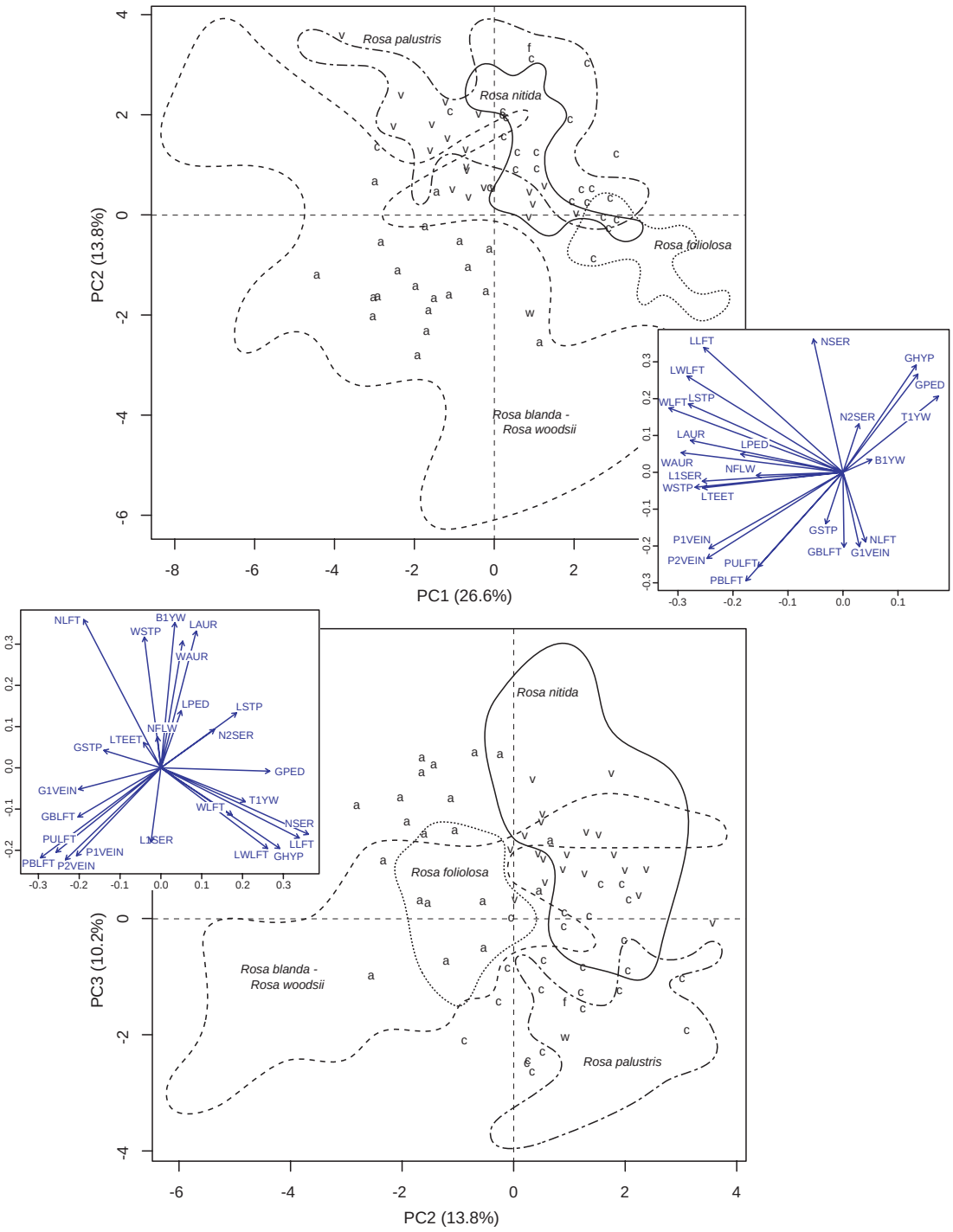


Figure 2.7 Principal component analysis (PCA) of all individuals. The dispersion of each diploid group found in the analyses of diploids alone (see Results) are illustrated by outlines. The *a priori* taxonomic identifications of polyploid individuals are represented by a letter: a – *R. arkansana*, c – *R. carolina*, v – *R. virginiana*. The insets on the figure show the representation of the vector of characters in the dimensions illustrated (loadings in appendix 4). The vectors are perpendicular to each other in the multidimensional space. The individuals that were identified *a priori* as *R. woodsii* and *R. foliolosa* but that were classified as polyploids (stomata size) are identified by a "w" and a "f", respectively. The percentage of the variance explained by each axis is shown.

A few extremes notwithstanding, the polyploid individuals were divided into two groups. One of these was to the lower-left of the origin on the ordination of the first two PCs and to the left of the origin on the ordination of the second and third PCs, and the second group was positioned opposite the first. When these groups were compared with the *a priori* taxonomic identifications, the first group consisted of *R. arkansana* and the second of *R. carolina* and *R. virginiana*. The characters that are most important to discriminate these groups are generally mostly the same that distinguished the *R. blanda* – *R. woodsii* diploid group from the other groups of diploids: pubescence of the leaflets and glands on the hypanthium and pedicels. The PCA did not reveal any clear discontinuity within the *R. carolina* – *R. virginiana* group. However, the third PC almost completely discriminated these two species identified *a priori*, with *R. carolina* having values below zero and *R. virginiana* above zero on this PC. The characters that are important in explaining this PC are the overall size of the stipules (larger for *R. virginiana*) and the pubescence of the leaflets (greater for *R. carolina*). The proportion of bristles is also important on the third axis but this character does not help to distinguish *R. carolina* and *R. virginiana* (see below).

It is also possible to compare the relative similarity of the diploid species identified earlier (see *Diploids* section) to the three polyploid species. *Rosa arkansana* is clearly morphologically more similar to the *R. blanda* and *R. woodsii* group and to *R. nitida* than to the other diploid species (Fig. 2.7, 2.9), *R. carolina* is equally similar morphologically to *R. foliolosa*, *R. nitida* and *R. palustris*, and *R. virginiana* is morphologically more similar to *R. nitida* than to the other species.

2.5.3.2 Molecular data

For the combined dataset of diploids and polyploids, eight bands (of the 107 bands that were scored) were present in all individuals and three were unique. Three bands were found only among diploids and none was restricted to polyploids.

The PCoA of all the individuals for which molecular information was available is shown in Figure 2.8. The first three axes of the PCoA represented 19.8 %, 6.8 %, and 5.1 % of the total variance. The dispersion of the different groups of diploids is illustrated by curves, although the exact points are not shown to facilitate the interpretation of the figure. When only polyploids were considered, two distinct groups can be differentiated, based on *a priori* taxonomic identifications: one that consisted of all *R. arkansana* plus one

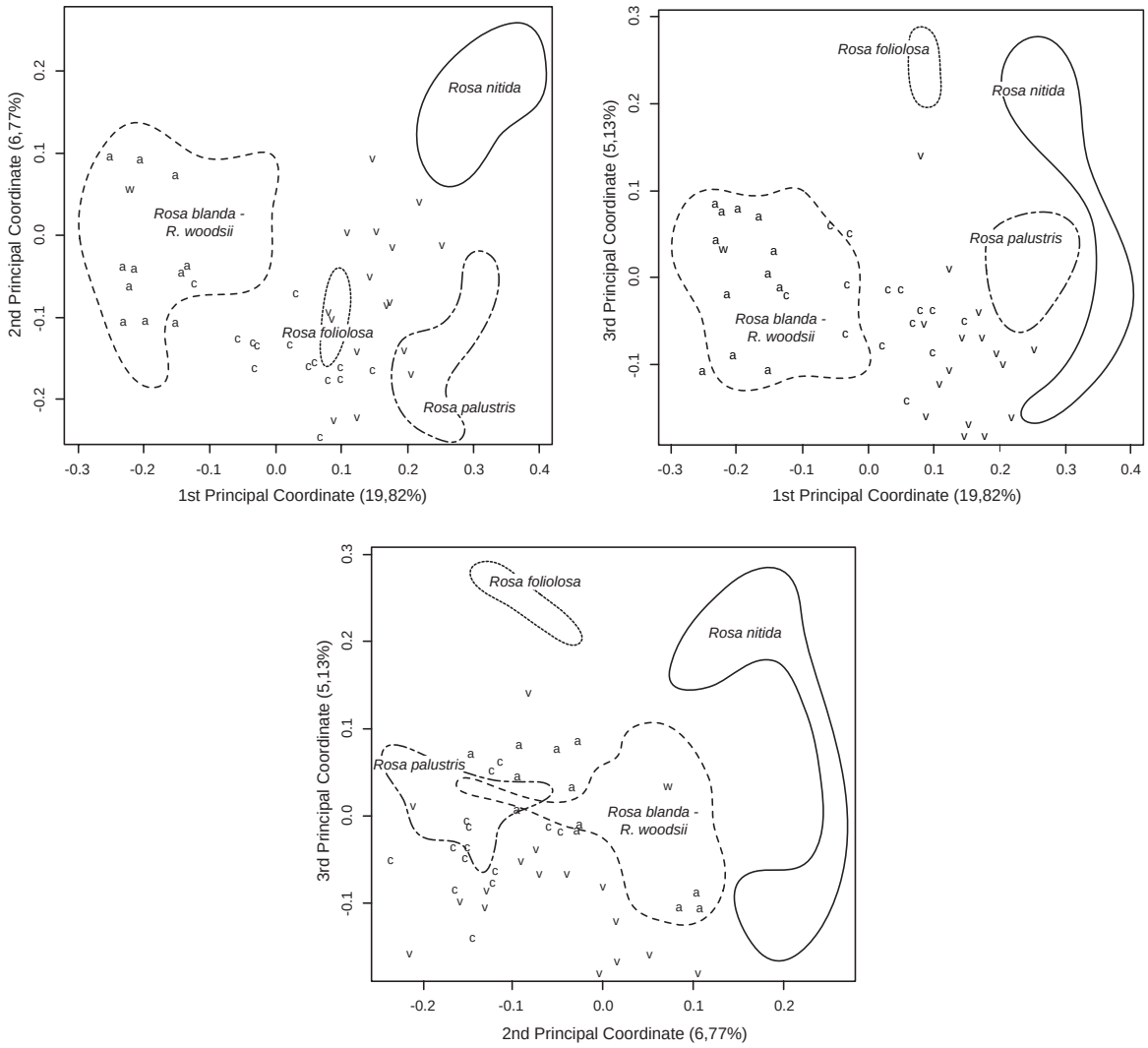


Figure 2.8 Principal coordinate analysis (PCoA) of all individuals. The dispersion of each diploid group found in the analyses of diploids alone (see Results) are illustrated by outlines. The *a priori* taxonomic identification of polyploid individuals are represented by a letter: a – *R. arkansana*, c – *R. carolina*, v – *R. virginiana*. The individual that was identified *a priori* as *R. woodsii* but that was classified as a polyploid (stomata size) is identified by a "w". The percentage of the variance explained by each axis is shown.

R. carolina individual, and one that comprised *R. carolina* and *R. virginiana* individuals. Within the group of *R. carolina* and *R. virginiana*, there was no clear distinction between the species, but we can see that they are nevertheless nearly all discriminated by the first and the second principal coordinates, with *R. carolina* having lower values than *R. virginiana* on both axes.

Considering that the three *a priori* polyploid species represent three distinct entities, their molecular similarities with the distinct groups of diploids identified can be evaluated (Fig. 2.9). *Rosa arkansana* was more similar to the *R. blanda* – *R. woodsii* group than to the other diploid groups at the molecular level and R. carolina was most similar to *R. blanda* – *R. woodsii* and *R. palustris* (Figs. 2.8, 2.9). Finally, *R. virginiana* was genetically more similar to *R. palustris* than to the other species (Fig. 2.9), although the PCoA also suggested affinities with *R. nitida*, especially with the individuals that formed group 5 in the phenogram (Figs. 2.2b, 2.4).

2.5.4 Classification tree

To construct the classification tree, the four distinct groups of diploid individuals found in the previous analyses were used, as well as the three polyploid species. The individuals *foliolosa*4184 and *woodsii*2008 that were considered to be polyploids (stomata size) were treated as *R. carolina* and as *R. arkansana* in the classification tree, respectively, based on the PCA. The tree constructed including the stomata guard cell length consisted of 9 groups, with only 11 misclassifications of the total 178 individuals (6.1 %; Fig. 2.10a). The first division of the tree was based on the stomata guard cell length, therefore dividing the individuals into almost pure diploid and polyploid groups. If the stomata character is omitted from the analysis, the number of misclassifications obtained was

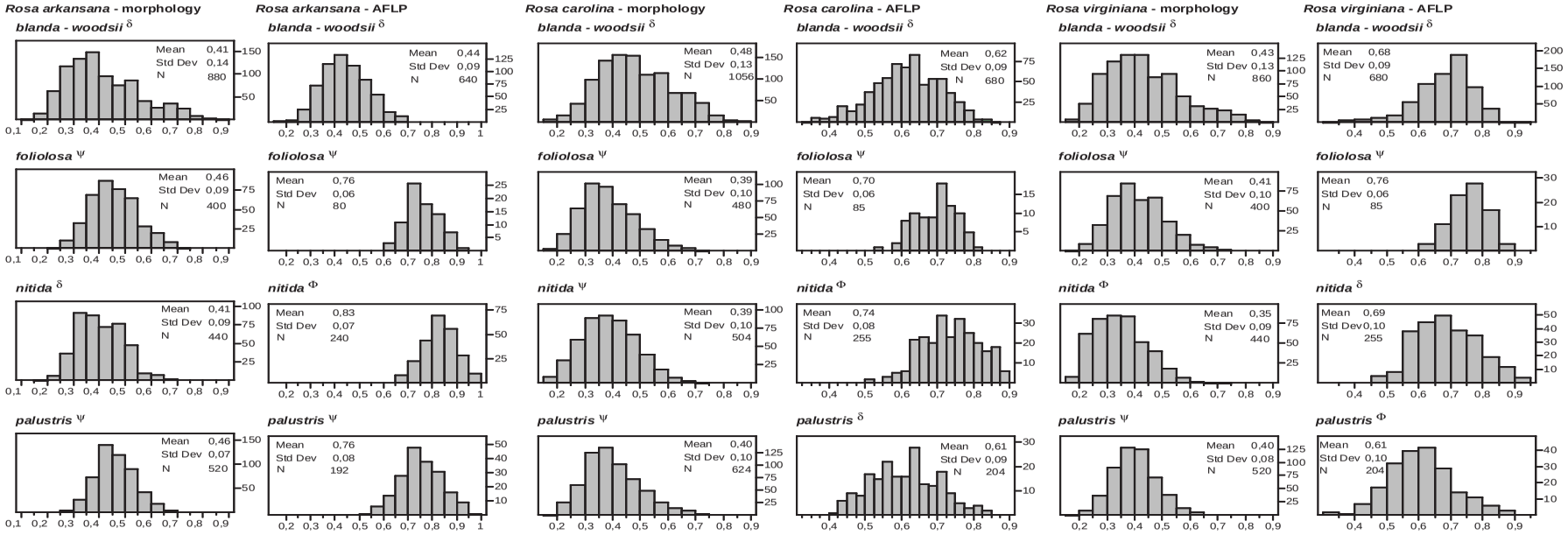


Figure 2.9 Histograms of the pair-wise distance between individuals of each polyploid species and that of each diploid group identified in the analyses. For each column, different Greek letters following the species indicate that their means are significantly distinct according to Tukey's HSD test (5% level; Tukey, 1953; Kramer, 1956).

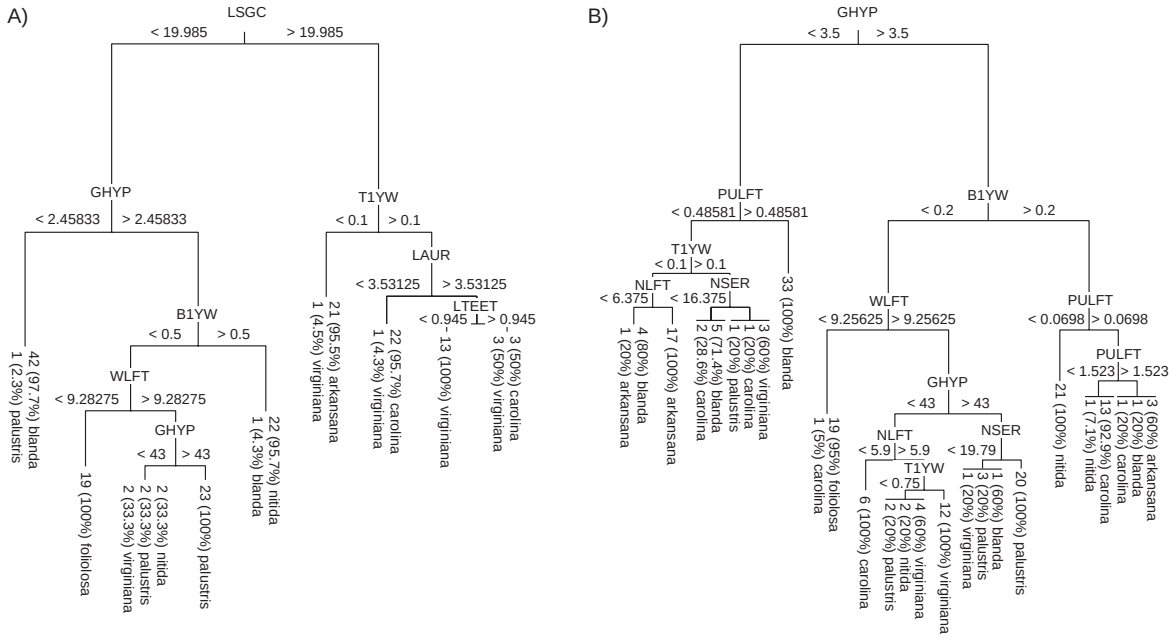


Figure 2.10 Classification trees obtained from the species objectively described in this paper. The two trees represent the best solution when the length of the stomata guard cell length (LSGC) is included (A) and excluded (B) from the analysis. For each terminal group, the number of individuals of each species (and its percentage in the group) is indicated. The lengths of the edges are proportional to the amount of deviance (impurity) resolved by the split at the node above. At each split, the group characterised by the smallest values for the character is always situated to the left of the node.

only slightly higher than when it was included (15 of 178, for 8.4%; Fig. 2.10b). To help visualize the main morphological differences among the species, boxplots for all the characters evaluated are shown in Fig. 2.11.

2.6 Discussion

2.6.1 How many species?

2.6.1.1 Diploid species

The morphological and the molecular analyses generally revealed the same groups of individuals, although the gaps among these were more pronounced in the molecular

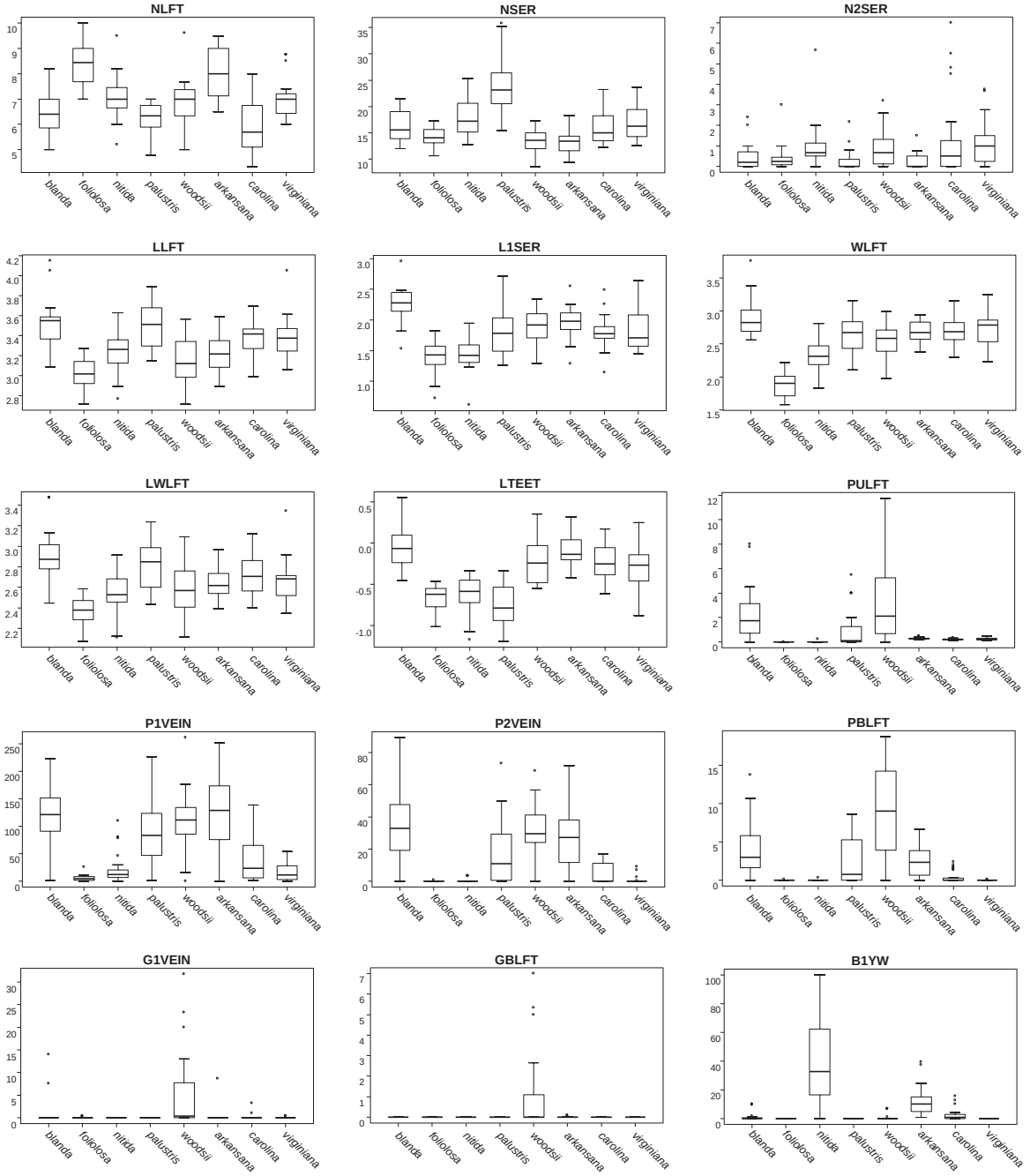


Figure 2.11 Boxplots illustrating the variation of each species for each morphological character. Even though *R. blanda* and *R. woodsii* may not form distinct species, boxplots are given for each taxon following the *a priori* taxonomic identification in order to illustrate the slight differences in variation among the two *a priori* species. All length characters are given on a logarithmic scale. Abbreviations are found in Table 2.

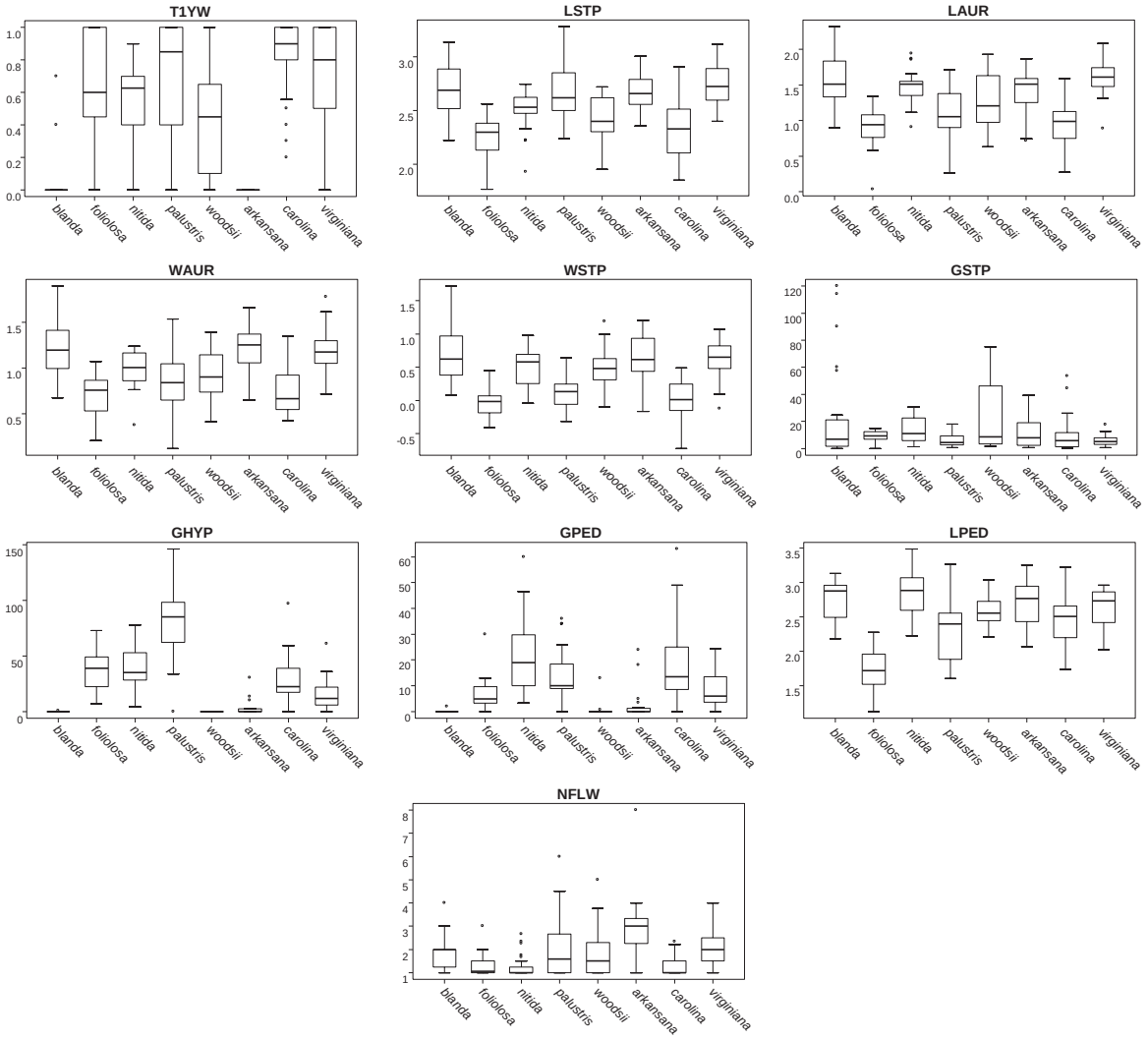


Figure 2.11 … *continued*

analyses. A point of discordance between the morphological and the molecular analyses has to do with the two distinct groups of individuals that were found within *R. nitida* in the molecular analyses. Because the two groups were not distinct in the morphological and combined analysis, they were not considered to represent distinct species. Yet, this is a conservative decision and it is also possible that these groups represent cryptic species (species that are morphologically indistinguishable). Considering that individuals *a priori* identified as *R. nitida* forms a single species, four distinct diploid species were found in *Rosa* sect. *Cinnamomeae* east of the Rocky Mountains. Three of these are equivalent to the taxonomic species *R. foliolosa*, *R. nitida* and *R. palustris*, and the fourth consists of *R. blanda*

and *R. woodsii* as a single group. According to both the morphological and the molecular analyses, these four species can be divided into two groups: one consists of the *R. blanda* – *R. woodsii* group and the other of the species *R. foliolosa*, *R. nitida*, and *R. palustris*. These two groups of diploids were once classified as separate sections (Crépin, 1889): *R. foliolosa*, *R. nitida* and *R. palustris* being placed in section *Carolinae* and *R. blanda* – *R. woodsii* being placed in section *Cinnamomeae*. Although recognizing section *Carolinae* makes section *Cinnamomeae* paraphyletic (Wissemann and Ritz, 2005; Joly *et al.*, 2006) and although section *Carolinae* itself is polyphyletic because of the inclusion of allopolyploid species (Joly *et al.*, 2006), this distinction nonetheless reflects important differentiations at the morphological level. The principal characteristics that differentiate these two groups are the length of the leaflet teeth (larger for *R. blanda* – *R. woodsii*) and the glands on hypanthiums and pedicels (absent in *R. blanda* – *R. woodsii*). These two groups and their species will be investigated in more details below, starting with the close allies *R. foliolosa*, *R. nitida* and *R. palustris*.

Rosa foliolosa is the species with the most restricted geographic distribution: it grows in mesic prairies of northeastern Texas, western Arkansas, and Oklahoma (Lewis, 1958). It is also the species with the smallest individuals of the group, which is reflected here by their small terminal leaflets, small stipules, and small pedicels (Fig. 2.11). The characters that mostly differentiate *R. foliolosa* from the other species are the narrow terminal leaflets and the short pedicels (Fig. 2.10, 2.11; Lewis, 1957b). Even if only two individuals of *R. foliolosa* were sampled for the molecular study, the inclusion of more individuals should not change the conclusion of this study. Indeed, *R. foliolosa* is probably the most distinct species in the molecular analyses, but also in the morphological analyses where the sampling covered the entire range of the species. Moreover, its restricted geographic distribution limits the effects of isolation by distance and decreases the potential for important intra-specific genetic variation.

Rosa nitida and *R. palustris* are the only two species studied that grow in bogs and poorly drained soils. *Rosa nitida* is mostly characterized by the presence of numerous

bristles on the stem (Fig. 2.11, 2.3, 2.10) that are normally reddish in colour (S. Joly, pers. obs.). *Rosa nitida* also has longer auricles, wider stipules, and more double serrations than its close allies *R. foliolosa* and *R. palustris*. The two distinct groups of individuals revealed by the molecular analyses suggest that there is intra-specific differentiation at the molecular level among populations of *R. nitida*. Yet, this differentiation is not reflected at the morphological level and is not related to the geographic distance among individuals: one group (#4; Figs. 2.2b, 2.4) consists of one Newfoundland and one New Brunswick population, whereas the other (#5) consists of two Newfoundland and one Québec population. *Rosa nitida* populations are not frequent in nature and are difficult to find due to its habitat. For this reason, only five populations have been sampled for the molecular study. Therefore, it is possible that the observed intra-specific genetic differentiation within *R. nitida* is caused by this reduced sampling and that it would disappear if more populations are analysed. Alternatively, it may also represent two phylogeographic lineages, which can be the result of distinct glacial refugia or two cryptic species. Further investigations are needed to clarify this.

As the third member of this trio, *R. palustris* individuals are generally taller than individuals of the first two species. *Rosa palustris* is characterized by a greater number of serrations on the leaflet and more glands on the hypanthium than the other species, in general (Fig. 2.11, 2.3). It is also differentiated from its close allies *R. foliolosa* and *R. nitida* by large leaflets and the presence of more leaflet hairs (Fig. 2.11, 2.3). Although southern populations of *R. palustris* were not sampled for the molecular dataset, its clear distinctiveness in the molecular analyses, as well as in the morphological analyses, suggest that it would remain genetically distinct with an increased sampling.

The group that consists of *R. blanda* and *R. woodsii* is very polymorphic and there is no clear evidence that there is more than one species present in this group according to our species definition. The redundancy analyses have shown that although a significant amount of variation was explained by the *a priori* taxonomy, this is more likely to be an effect of the geographical distance between individuals. The importance of the

geographic distance is further supported by the graphical representation of the morphological variation as a function of the longitude of the individuals sampled (Fig. 2.5). The gradual change in morphological variation from one ocean to the other observed in several characters, rather than an abrupt change at a single geographic position for many characters, argues against the presence of a hybrid zone and suggests that geographic distance is important in shaping the morphological variation within this group. This is important because evidence of a hybrid zone would have implied a differentiation between the two hybridizing populations. Even though other sources of evidence advocate in favour of a hybrid zone, such as the presence of a characteristic short pair of chromosomes in *R. woodsii* individuals in the hybrid zone between *R. woodsii* and *R. blanda* (Lewis, 1962), the present results clearly show that isolation by distance is the most important factor for structuring the morphological and molecular variation in this group. Indeed, given the wide distribution of these species, it is not surprising that the geographic distance has an important impact in explaining the morphological and molecular variation. *Rosa woodsii* is found west and east of the Rocky Mountains from the south of the USA to northern British Columbia in Canada and it reaches Manitoba in the northern prairies (north of USA and Canada). In Manitoba, *R. woodsii* is gradually replaced by *R. blanda*. The latter reaches Nova Scotia in a range of latitude that goes from James Bay to the state of Missouri. The wide distribution of these taxa may also be responsible for the important morphological variation in this group because they are exposed to various environmental conditions.

Other information tends to suggest that *R. blanda* and *R. woodsii* do not form distinct species. For example, both species grow in similar habitats, mesic soils along woods and rivers (Erlanson, 1934; S. Joly, pers. obs.). Moreover, both species give fertile hybrids when crossed (Erlanson, 1934) and an investigation of three single-copy nuclear genes has failed to find a distinction between *R. blanda* and *R. woodsii* (Joly and Bruneau, in press). Therefore, given that *R. blanda* and *R. woodsii* are indistinguishable at the molecular and morphological levels, that they give fertile hybrids and that the morphological data argue against the presence of a hybrid zone, *R. blanda* and *R. woodsii*

should be considered a single species. This would imply placing *R. woodsii* Lind. as a synonym of *R. blanda* Ait. because of priority. But this group is complex and warrants further study. For example, varieties have been described within *R. woodsii* and these are sometimes considered distinct species (e.g., *R. woodsii* var. *ultramontana* (S. Watson) Jepson, *R. woodsii* var. *glabrata* (Parish) Cole). Other studies with a denser sampling in the *R. woodsii* area are needed to address these taxonomic problems and more critically assess the distinctiveness of *R. woodsii* and *R. blanda*.

2.6.1.2 Polyploid species

Hybrids, when included in phenograms, can distort the true species relationships (Heiser Jr. *et al.*, 1965; Jensen and Eshbaugh, 1976; McDade, 1997). At best, the phenogram will identify one of the two parental species of the allopolyploid as it cannot group them with the two parents at the same time. Because the same thing can be said of allopolyploids, and because allopolyploids are thought to be much more frequent than autopolyploids in nature (Soltis and Soltis, 1993), using clustering methods to represent the relationships of polyploids may be problematic. Therefore, the polyploid individuals were analysed only with ordination methods.

Among polyploid species, there were two clearly distinct groups of individuals present in the dataset. One of these groups is equivalent to the species *R. arkansana*, whereas the other comprises individuals that belong to the taxonomic species *R. carolina* and *R. virginiana*.

Rosa arkansana grows in the prairies, approximately from the Rocky Mountains to the Mississippi River. It is differentiated from the other polyploid species by a greater number of leaflets per leaf, by the presence of more hairs on the veins and on the limb of the terminal leaflet, by more bristles and by the absence of thorns on stems, by very few (or no) glands on the hypanthia and on the pedicels, and by more numerous flowers per inflorescence than in other species (Figs. 2.7, 2.10, 2.11).

In the other group, there was no clear gap in either the morphological or the molecular datasets. However, when the *a priori* taxonomic identifications were considered, the species *R. carolina* and *R. virginiana* were almost completely discriminated in both the morphological and the molecular analyses, and the individuals that were intermediate between the two species occurred in or near the zone of sympatry. This suggests that *R. carolina* and *R. virginiana* form distinct groups that are connected by a hybrid zone. However, it is critical for proper species delimitation to determine whether this hybrid zone represents a secondary hybrid zone, implying that hybridizing populations evolved separately prior to contact, or a primary hybrid zone that has been established by parapatric divergence (Endler, 1977; Barton and Hewitt, 1985). Although it is difficult to distinguish these two types of hybrid zones (Barton and Hewitt, 1985), a strong argument in favour of a secondary hybrid zone would be to demonstrate that *R. carolina* and *R. virginiana* have distinct origins. Joly *et al.* (2006) investigated this question in their study of the single-copy nuclear gene *GAPDH*. They found that *R. carolina* and *R. virginiana* likely have distinct origins based on to the *GAPDH* alleles found in the two species. Those analyses suggested that *R. carolina* would be an allopolyploid with one parent from the *R. blanda* – *R. woodsii* group and another from a group formed by *R. nitida* and *R. palustris*. The exact parents were impossible to identify because of a lack of discriminating power of the marker at the diploid level. In contrast, *R. virginiana* most likely evolved from within the *R. nitida* – *R. palustris* group. Because the present study also supports distinct origins for *R. carolina* and *R. virginiana* (see below), the hybrid zone between these species is most likely secondary. Therefore, because *R. carolina* and *R. virginiana* have evolved independently before the formation of the hybrid zone and because the two species maintain cohesion, as reflected by their discrimination in the ordination analyses, *R. carolina* and *R. virginina* should be maintained as distinct species.

Regardless, hybridization takes place between these two species, and differentiating the two is likely to remain difficult in nature. Yet, the present study identifies some characters that are useful in differentiating *R. carolina* and *R. virginiana*.

Morphologically, *R. carolina* is often slightly more pubescent, particularly on the secondary veins of the leaflet, and it sometimes has bristles. *Rosa carolina* individuals also have fewer flowers per inflorescence in general. However, the length and width of the stipules and the length of the auricules seem to be the most useful characters. It is of interest to note that the pubescence of the leaflets has never previously been reported as diagnostic for these species (Lewis, 1957b) and it may increase our ability to identify them. A character useful in differentiating these species is thorn morphology: *R. carolina* usually has strait, not especially broad-based infrastipular thorns, whereas *R. virginiana* has stout thorns that often are broad-based (Lewis, 1957b). This character was not included in this study because many individuals lack infrastipular thorns and including such characters would have resulted in missing data that are not allowed with some of the analyses performed. Another difference between *R. carolina* and *R. virginiana* is their distribution. *Rosa carolina* occurs widely east of the Mississippi River. In contrast, *R. virginiana* is found along the Atlantic coast, principally on dunes and in salt marshes, and probably does not spread very far inland.

In addition to the apparent hybrid zone between *R. carolina* and *R. virginiana*, there is also a putative hybrid zone between *R. arkansana* and *R. carolina*. However, because few specimens from the hybrid zone were included in the present study, *R. arkansana* and *R. carolina* were clearly distinct in the analyses. In another study where specimens from the hybrid zone were included in a morphological analysis of *R. arkansana* and *R. carolina*, the individuals from the hybrid zone clearly occurred between the allopatric individuals of the two species in a principal component analysis (A. Fishbein and W.H. Lewis, unpublished manuscript, Washington University). Other evidences of a hybrid zone between *R. arkansana* and *R. carolina* include the presence of aborted fruits (S. Joly, pers. obs.), pollen abortion (Lewis, 1957c), and cytological irregularities (Lewis, 1957c, 1966) in individuals from the zone of sympatry. Therefore, although it is easy to identify *R. arkansana* and *R. carolina* individuals from allopatric regions, identification will be complex when specimens come from the region of sympatry between the two species. But as for *R. carolina* and *R. virginiana*, *R. arkansana* and *R. carolina* also have distinct

origins (present study; Joly *et al.*, 2006), which further supports their status as distinct species.

2.6.2 Delimiting species in a polyploid complex

The present study clearly shows that the analysis of individual characters is not sufficient to identify all species in the complex (Fig. 2.11). None of the characters studied can completely discriminate one species from another; the range of variation always overlaps. Only with a multivariate approach can species boundaries be identified by finding gaps in the morphological variation. However, these efforts would have been vain if diploids and polyploids had not been previously distinguished. The ordinations show that if all individuals were included in a single morphological analysis without prior knowledge of the ploidy level of the individuals, the result would be a single cloud of points (Fig. 2.7). The position of the polyploids between those of the diploids blurs the species boundaries of diploids and reciprocally also those of polyploids. The same is true for the molecular analysis, to a lesser extent. To solve this problem, diploids were differentiated *a priori* from polyploids in order to be able to delimit species boundaries at both ploidy levels separately. If this approach could be useful in polyploid complexes where polyploids are indistinguishable from diploids (e.g., Vanderhoeven *et al.*, 2002), it may not be in complexes where polyploids are easily distinguished (e.g., Suda and Lysák, 2001; Perný *et al.*, 2005). Consequently, it is difficult to draw a general analysis scheme that can be applied to all polyploid complexes.

It is sometimes possible to differentiate polyploid and diploid individuals using morphological characters. For example, polyploid often differs from diploids by their greater size (Stebbins, 1971; Levin, 2002), but sometimes even by qualitative characters such as in *Solanum* (Schilling and Heiser Jr., 1976). However, besides the length of the stomata guard cells and le pollen size, no character seems useful in differentiating diploids and polyploids in this complex. In such a situation, one may think that it is almost impossible to accurately identify species in this group using macro-morphology

only. This is of concern to botanists that may wish to identify an individual in the field, because measuring stomata guard cell length or pollen size requires a microscope. Yet, the classification tree from which the length of the stomata guard cell lengths was removed suggests that it is still possible to identify species with only limited misclassifications when diploids and polyploids are not *a priori* differentiated. This is even more encouraging as the classification tree discriminates species using a single character per node. Thus, it may be possible to increase the accuracy of identifications by using the information of all characters at the same time. Clearly, using a hierarchic classification makes it possible to differentiate species even where no distinctions are evident when all individuals are considered at the same time (see Fig. 2.11).

2.6.3 Origins of polyploid species

In this study, the overall similarity of polyploids to the diploid species and their position in the ordinations were used to investigate the origins of the polyploid species with the assumption that hybrids are expected to be intermediate between their parental species (Neff and Smith, 1978; McDade, 1997). This is more straightforward with the molecular data, as polyploid species must have acquired markers that are still present in their diploid progenitors. In contrast, more caution is required in interpreting morphological data, especially for allopolyploids. Even though hybrid individuals are often thought to be morphologically intermediate between their parents, it is not always the case. Hybrids may also show morphological characteristics that are more extreme than those of either parent, a phenomenon called transgressive hybridization (Rosenthal *et al.*, 2002). A hybrid also may be closer to one parent than to the other because of introgressive hybridization (Whiffin, 1973), dominance of characters (Ramon, 1968), or because they represent backcross generations (Knops and Jensen, 1980), thereby potentially hiding the hybrid origin of the species. Also, a hybrid between two distant individuals also may be more similar to a non-parental species that has intermediate morphological characteristics relative to those of the two parental species (McDade, 1997). Finally, morphological characters may be more influence by environmental

variation than molecular markers (but see Rieseberg *et al.*, 2003). In roses, although some transplant experiments have been achieved (Lewis, 1957b), it is still unknown to what extent the environmental conditions can modify the morphology of individuals.

In spite of these potential pitfalls, the similarities of the polyploids with the diploids are clearly an indication of their origin. For instance, *R. arkansana* is genetically more similar to the group of *R. blanda* and *R. woodsii* than it is to the other groups. In terms of morphology, the pair-wise distances indicate that *R. arkansana* is on average as close to the *R. blanda* – *R. woodsii* group as to *R. nitida*, although the PCA clearly associated *R. arkansana* with the *R. blanda* – *R. woodsii* group. Given that *R. nitida* is the species most distant from *R. arkansana* based on genetic distances, the similarity at the morphological level is more likely the consequence of convergence rather than of parental relationship. Thus, the present study suggests that *R. arkansana* has evolved from within the *R. blanda* and *R. woodsii* group. This supports the conclusions reached by the study of the *GADPH* nuclear gene (Joly *et al.*, 2006) and by earlier morphological considerations (Erlanson, 1929).

Rosa carolina is genetically equally close to *R. blanda* – *R. woodsii* than it is to *R. palustris* (Fig. 2.8, 2.9). Considering the overall morphological similarity (Fig. 2.9), *R. carolina* is equally similar to *R. foliolosa*, *R. nitida* and *R. palustris*, but is less similar to *R. blanda* – *R. woodsii*. However, the PCA suggests that it is not as similar to *R. foliolosa* as it is to *R. nitida* and *R. palustris*. The morphological and molecular information are therefore in contradiction for *R. carolina*. The molecular data suggest that *R. carolina* could be an allopolyploid derivative between the *R. blanda* – *R. woodsii* group and *R. palustris*. This is in agreement with the study of the GAPDH nuclear gene that suggested a similar relationship (Joly *et al.*, 2006), although it was impossible to clearly identify *R. palustris* as the second parent because of the lack of resolution among diploids. In contrast to the molecular results, the morphology suggests that *R. carolina* takes its origin from within a group that consists of *R. foliolosa*, *R. nitida* and *R. palustris*. Because previous studies have shown that *R. blanda* – *R. woodsii* form a distinct genetic group from that comprised of *R.*

foliolosa, *R. nitida* and *R. palustris* (Joly *et al.*, 2006; Joly and Bruneau, in press), the close genetic similarity of *R. carolina* to both *R. palustris* and *R. blanda – R. woodsii* is unlikely to be the result of shared ancestral polymorphisms between these diploids. Therefore, the molecular data probably better reflects the evolutionary history than morphology does. Considering this, *R. carolina* probably evolved from a cross between *R. blanda – R. woodsii* and *R. palustris*. The closer morphological similarities of *R. carolina* with the *R. foliolosa – R. nitida – R. palustris* group may be explained by morphological convergence. Because several examples are known of hybrids that remain morphologically closer to their maternal parent (reviewed in Levin, 2003), another hypothesis would be that *R. palustris* was the maternal parent in the evolution of *R. carolina*.

Finally, *R. virginiana* is genetically closest to *R. palustris* and morphologically closest to *R. nitida*. Again, both sources of information are in contradiction. The morphological similarities with *R. nitida* may be biased by the choice of the characters used in this study. Indeed, from a general morphological aspect, *R. virginiana* looks more similar to *R. palustris* than to *R. nitida* (S. Joly, pers. obs.). Alternatively, because *R. virginiana* also shares genetic affinities to one of the subgroups of *R. nitida*, it may be argued that both species were involved in its evolution. Even though the origin of *R. virginiana* is not completely solved, it is at least clear that *R. foliolosa* and *R. blanda – R. woodsii* were not involved in it.

Although the similarities between polyploid and diploid individuals have already been used to interpret the origins of hybrid or polyploid taxa (Heiser Jr. *et al.*, 1965; Schilling and Heiser Jr., 1976), it has never been applied to species as a whole to our knowledge. The common technique is to compare the position of polyploids relative to that of diploids in ordinations (Perný *et al.*, 2005). Yet, the similarity approach as several advantages. First, it makes use of the full amount of information contained in the data, and not just the amount of information kept in the reduced spaced of the ordination (rarely more than two or three axes). More importantly, if applied to species rather than to individuals, the similarity approach also allows testing statistically which diploid

species it is closer rather than aving rely on qualitative comparisons. Consequently, such an approach should be use more often in these types of investigations.

In conclusion, the present study has shown that it is possible to use an objective approach without *a priori* taxonomic identification to circumscribe species in a polyploid complex characterized by extensive intra-specific polymorphisms and where each character exhibits a continuum of variation. In addition of being useful for finding gaps in the morphological and molecular variation when all characters are analysed simultaneously, the multivariate approaches used have been informative in identifying potential origins for the polyploid species.

CHAPITRE 3

Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America[2]

3.1 Résumé

La variation allélique présente au sein des individus contient de l'information phylogénétique qui pourrait être mise à profit pour reconstruire l'évolution des organismes. Par contre, peu d'efforts ont été investis pour développer cette idée dans le passé. Les arbres d'haplotypes représentent une solution intéressante lorsqu'un seul gène est étudié, mais il n'existe pas de méthode qui permette de combiner plusieurs gènes. Dans cet article, nous présentons un algorithme qui permet de convertir une matrice de distances alléliques en matrice de distances entre organismes, ce qui permettra d'incorporer l'information allélique des individus dans la reconstruction phylogénétique d'organismes pour un ou plusieurs gènes. Cette méthode est utilisée ici pour reconstruire l'histoire évolutive des sept espèces indigènes de *Rosa* sect. *Cinnamomeae* en Amérique du Nord. Les gènes codants pour la glycéraldéhyde 3-phosphate déshydrogénase (*GAPDH*), la triose phosphate isomerase (*TPI*) et la malate synthase (*MS*) ont été séquencés pour 40 individus provenant de ces sept espèces de roses. Les trois gènes étaient peu variables et la plupart des espèces montraient du triage de lignées incomplet (*incomplete lineage sorting*), suggérant une origine récente pour celles-ci. Malgré ces difficultés, les réseaux (NeighborNet) d'organismes reconstruits à l'aide de l'algorithme décrit dans cet article

---

[2] Joly, Simon et Anne Bruneau. 2006. Syst. Biol. Article sous presse

ont donné des groupes qui concordaient davantage avec les espèces taxonomiques que les arbres d'haplotypes. Le réseau reconstruit à l'aide des trois gènes montre que les espèces de l'ouest américain, *R. gymnocarpa* et *R. pisocarpa*, forment des groupes exclusifs et suggère une distinction des espèces ouest- et est-américaines. Dans l'est, trois groupes distincts ont été identifiés : *R. nitida* – *R. palustris*, *R. foliolosa*, *R. blanda* – *R. woodsii*. Ces groupes sont en accord avec les données morphologiques et écologiques des espèces. La méthode décrite est aussi utile pour reconstruire les relations phylogénétiques des individus hybrides lorsque les relations phylogénétiques sont reconstruites à l'aide d'un réseau phylogénétique.

3.2 Abstract

Allelic variation within individuals holds information regarding the relationships of organisms, which is expected to be particularly important for reconstructing the evolutionary history of closely related taxa. However, little effort has been devoted to incorporate such information for reconstructing the phylogeny of organisms. Haplotype trees represent a solution when a single non-recombinant marker is considered, but there is no satisfying method to combine multiple genes. In this paper, we propose an algorithm that converts a distance matrix of alleles to a distance matrix among organisms. This algorithm allows the incorporation of allelic variation in reconstructing the phylogeny of organisms from one or more genes. The method is applied to reconstruct the phylogeny of the seven native diploid species of *Rosa* sect. *Cinnamomeae* in North America. The glyceralgehyde 3-phosphate dehydrogenase (*GAPDH*), the triose phosphate isomerase (*TPI*) and the malate synthase (*MS*) genes were sequenced for 40 individuals from these species. The three genes had little genetic variation and most species showed incomplete lineage sorting, suggesting they have a recent origin. Despite these difficulties, the networks (NeighborNet) of organisms reconstructed from the matrix obtained with the algorithm described in this paper recovered groups that more closely match taxonomic boundaries than did the haplotype trees. The combined network of individuals shows that species west of the Rocky Mountains, *Rosa gymnocarpa* and *R.*

pisocarpa, form exclusive groups and that together they are distinct from eastern species. In the east, three groups were found: *R. nitida* – *R. palustris*, *R. foliolosa*, and *R. blanda* – *R. woodsii*. These groups are congruent with the morphology and the ecology of these species. The method is also useful for reconstructing the phylogenetic relationships of hybrid individuals when the relationships are reconstructed using a phylogenetic network.

3.3 Introduction

Allelic variation at autosomal loci holds information regarding the relationships of organisms. Indeed, using two alleles instead of one can give better estimations of phylogenetic relationships because twice the amount of information is provided. This is especially true of closely related taxa for which incomplete lineage sorting is likely (Rosenberg, 2002, 2003; Degnan and Salter, 2005). In addition, allelic variation allows the detection of hybrid individuals with a single marker, whereas at least two are required when only one allele per locus is sampled. But despite the amount of data contained in allelic variation, little effort has been devoted to date at incorporating such information for reconstructing the phylogenetic relationships of organisms.

One solution when a single non-recombinant marker is considered is to utilize haplotype trees, which frequently are used in evolutionary studies of closely related species (Schaal and Olsen, 2000). At present, however, no phylogenetic method can easily incorporate allelic variation for more than one gene when reconstructing the evolutionary history of individuals. Yet the importance of investigating several markers for reconstructing the phylogeny of species is widely recognized as any single gene can be incongruent with the evolutionary history of species (Pamilo and Nei, 1988; Takahata, 1989; Wu, 1991; Doyle, 1992; Maddison, 1997; Nichols, 2001; Rosenberg, 2002, 2003; Degnan and Salter, 2005).

Most current approaches used for reconstructing phylogenies from multiple markers, either using a total evidence (e.g., Kluge, 1989; Yang, 1996; Seo *et al*., 2005) or a consensus approach (e.g., de Queiroz, 1993), cannot incorporate allelic variation for multiple genes because they use haplotypes as the terminal units of the analysis. Because it makes no sense to concatenate alleles from different loci that segregate in natural populations, such methods are limited to using a single haplotype per individual. If individuals, rather than alleles, were the terminals in the analysis, it would be possible to combine information from different genes.

In this paper, we propose an algorithm that incorporates allelic variation for reconstructing the phylogeny of organisms. The proposed algorithm converts a distance matrix of alleles into a distance matrix of organisms so that individuals become the terminals of the analysis. The matrix of organisms for one marker can either be used alone or in combination with other matrices obtained from independently evolving markers to reconstruct a phylogeny of organisms.

The algorithm is applied to reconstruct the evolutionary history of the seven native diploid species of *Rosa* sect. *Cinnamomeae* in North America using allelic variation at three nuclear loci for 40 individuals. Very little is known of the phylogenetic relationships of these rose species, mostly because of the incomplete species sampling of previous phylogenetic studies (e.g., Millan *et al.*, 1996; Matsumoto *et al.*, 1998). Moreover, the little molecular variation found among North American species (Wissemann and Ritz, 2005; Joly *et al.*, 2006) limits our understanding of their relationships and suggests that these species are of recent origin. Consequently, incomplete lineage sorting (or deep coalescence) could be an important issue in this group as it is expected to be most severe among recently diverged species (Rosenberg, 2002, 2003; Degnan and Salter, 2005). Hybridization also could be a confounding evolutionary process because of the propensity of these roses to hybridize (Erlanson, 1934; Ratsek *et al.*, 1939; Ratsek *et al.*, 1940; Lewis and Basye, 1961). Therefore, this group represents a good case study to test

the proposed algorithm because of the potentially important additional information that allelic variation can provide.

3.4 The POFAD algorithm

The POFAD (for Phylogeny of Organisms From Allelic Data) algorithm starts with a distance matrix of alleles for a given marker. The algorithm will be illustrated using a hypothetical example with five individuals (*A*-*E*) for which we have a haplotype distance matrix (Fig. 3.1a) that can be represented by a haplotype tree (Fig. 3.1b). In the example, letters are used to distinguish individuals: capital and lower-case letters represent individuals and alleles, respectively. Alleles within an individual are set apart by a number (1 or 2).

3.4.1 Calculating the distance between organisms

Let $d(A,B)$ be the distance between individuals A and B and $d(a,b)$ be the distance between alleles a and b. Moreover, let min[x; y] be the minimum of the values of x and y. When evaluating the distance between two diploid individuals at a locus, three situations can be encountered:

(1) Both individuals have a single allele.—In this situation, the distance between individuals is equal to the allelic distance. If A and B are two individuals with one allele,

$$d(A,B) = d(a,b)$$

In the hypothetic example, $d(A,B) = 3$.

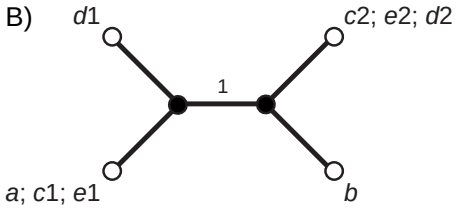
(2) One individual has one allele and the other has two alleles.—If A is an individual with one allele (a) and C is an individual with two alleles ($c1$, $c2$), then

$$d(A,C) = \frac{d(a,c1) + d(a,c2)}{2}$$

A)

	a	*b*	*c*1	*c*2	*d*1	*d*2	*e*1	*e*2
a	0							
b	3	0						
*c*1	0	3	0					
*c*2	3	2	3	0				
*d*1	2	3	2	3	0			
*d*2	3	2	3	0	3	0		
*e*1	0	3	0	3	2	3	0	
*e*2	3	2	3	0	3	0	3	0

B)



C)

	A	*B*	*C*	*D*	*E*
A	0				
B	3	0			
C	1.5	2.5	0		
D	2.5	2.5	1	0	
E	1.5	2.5	0	1	0

D)

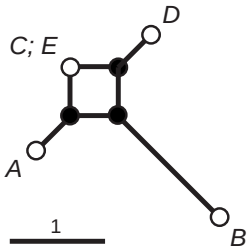


Figure 3.1 Example illustrating the distinction between a haplotype tree (B) and a network of organism incorporating allelic variation (D), for the same individuals. Letters distinguish individuals: capital and lower-case letters represent individuals and alleles, respectively, and alleles within an individual are distinguished by a number (1 or 2). (A) Hypothetic haplotype distance matrix, and (B) the unrooted haplotype tree obtained from it. (C) Matrix of distances between organism obtained from the haplotype distance matrix using the POFAD algorithm, and (D) the NeighborNet network reconstructed from it.

If we apply this to the theoretical example, $d(A,C) = (0+3)/2 = 1.5$.

(3) Both individuals have two alleles.—The two individuals, *D* and *E*, each have two alleles (*d*1, *d*2 and *e*1, *e*2). There are two pairs of allelic distances possible among these

individuals: $d(d1, e1)$ and $d(d2, e2)$ or $d(d1, e2)$ and $d(d2, e1)$. The distance between these individuals is the mean of the shortest pair of distances:

$$d(D,E) = \frac{\min\left[d(d1,e1) + d(d2,e2); d(d1,e2) + d(d2,e1)\right]}{2}$$

This minimizes the distance between the two pairs of alleles compared. It also compares the allele in one individual with the allele in the other with which it shares a most recent common ancestor. Taking individuals *D* and *E* from the hypothetic example (Fig. 3.1b), allele $d1$ will be compared with allele $e1$ (that are distant by two mutations) and allele $d2$ with $e2$ (that are identical) because the mean distance for this pair of comparisons, one mutation, is less than the mean distance of three mutations obtained when $d1$ is compared to $e2$ and $d2$ to $e1$. Therefore, in this theoretical example, $d(D,E) = [d(d1,e1) + d(d2,e2)]/2 = [2+0]/2 = 1$. This distance is preferable to using the mean of the four different allelic distances, as the latter option can result in a non-zero distance for two identical individuals. To illustrate this with the hypothetic example, take individuals *C* and *E* that have identical alleles. The sum for each pair of allelic distances is 0 for $[d(c1,e1) + d(c2,e2)]$ and 6 for $[d(c1,e2) + d(c2,e1)]$. Taking the mean of all four comparisons would give a distance of (0+6)/4=1.5 for the distance between *C* and *E*, which would not make sense because they are genetically identical. In contrast, taking the mean of the pair with the shortest allelic distance gives a distance of 0.

3.4.2 Combining information from different genes

The matrix of organisms obtained from one marker can either be used alone or be combined with matrices obtained from other markers. For this study, each gene matrix is re-weighted so that each gene makes an equal contribution to the combined phylogeny. This is accomplished by dividing each distance by the largest distance of the matrix, for each gene matrix. By attributing the same weight to each gene, every gene is considered to represent an independent estimation of the phylogeny. To fulfil this requirement, there needs to be no recombination within markers. In the presence of recombination, more than one evolutionary history is present in one marker and consequently the non-

recombining portions of a recombinant gene would be down-weighted. It is therefore recommended to test for recombination before combining different genes (see Posada and Crandall, 2001; Wiuf *et al.*, 2001; Posada, 2002; Bruen *et al.*, 2006).

When combining multiple gene matrices, the final distance between two individuals is the mean of the distances between these individuals in the individual matrices. If *M* and *N* are two individuals, then the mean distance between them will be:

$$d(M,N) = \frac{1}{n}\sum_{i=1}^{n} d_n(M,N) \big/ d^n_{\max}$$

where *n* is the number of datasets and $d^n_{\max}$ is the maximum distance in matrix *n*. Once the final matrix is obtained, one can reconstruct the phylogeny of the organisms with any phylogenetic or network method that uses distances. The program POFAD, written in C++, implements these algorithms and is available at www.irbv.umontreal.ca/pofad.htm.

In our theoretical example, the relationship of individuals was reconstructed from the matrix of organisms (Fig. 3.1c) using the NeighborNet method (Bryant and Moulton, 2004; Fig. 3.1d).

3.5 Material and methods

3.5.1 Plant material

Forty individuals from all seven North American diploid species of *Rosa* sect. *Cinnamomeae* were investigated (Table 3.1). *Rosa gymnocarpa* Nutt. and *R. pisocarpa* Gray are found exclusively west of the Rocky Mountains, *R. blanda* Ait., *R. foliolosa* Nutt. ex Torr. & A. Gray, *R. nitida* Willd. and *R. palustris* Marsh. occur strictly east of the Rockies and *R. woodsii* Lindl. can be found on both sides of these mountains. Two diploid species of section *Synstylae* found in North America, *R. setigera* Michx. (native) and *R. multiflora*

Table 3.1 Individuals included in this study with their collectors and locality. The number of alleles found for the different genes is indicated and the number of clones sequenced for each species and for each gene is showed in brackets. A hyphen in brackets indicates that there were two alleles that differed by a single mutation and that cloning was not necessary.

Species	Accession	Collector	Province/ State [a]	Lat. – Long.	Nb. alleles		
					GAPDH	*TPI*	*MS*
R. blanda	160	Joly and Starr 409	N.B.	45°57'43.7''N, 67°22'26.1''W	2 [3]	2 [4]	2 [5]
R. blanda	326	Joly and Starr 582	Ont.	42°15'29.7''N, 83°02'58.8''W	2 [3]	2 [4]	2 [5]
R. blanda	365	Joly and Starr 622	Wis.	42°39'07.5''N, 89°43'32.4''W	2 [4]	2 [4]	2 [5]
R. blanda	421	Joly and Starr 678	Minn.	48°06'36.3''N, 96°09'16.0''W	2 [4]	2 [2]	2 [5]
R. blanda	462	Joly and Starr 722	Man.	50°00'59.3''N, 96°55'35.2''W	2 [4]	2 [3]	2 [-]
R. blanda	528	Joly and Starr 788	Ont.	46°28'15.4''N, 80°29'27.2''W	1	2 [-]	2 [-]
R. blanda	567	Joly 921	N.Y.	-	2 [4]	1	2 [-]
R. blanda	652	Joly *et al.* 993	Que.	48°02'58.8''N, 65°28'43.6''W	2 [4]	1 [3]	2 [6]
R. blanda	1214	Bruneau *et al.* 1214	Que.	45°31'18'' N, 73°50'02'' W [b]	1	1	2 [5]
R. blanda	1219	Bruneau *et al.* 1219	Que.	45°30'18''N, 73°50'02''W [b]	1	2 [-]	2 [5]
R. blanda	1236	Bruneau *et al.* 1236	Que.	48°21'36''N, 68°45'36''W [b]	2 [4]	2 [4]	2 [1]
R. blanda	98016	Drouin 98-016	Que.	47°26'27''N, 70°30'18''W [b]	2 [-]	1	2 [5]
R. foliolosa	699	Lewis 15846-3	Okla.	34°24'N, 96°00'W	2 [3]	2 [4]	1
R. foliolosa	795	O'Kennon and McLemore 19069A	Tex.	33°24'32.2''N, 97°30'22.0''W	2 [-]	1	2 [5]
R. gymnocarpa	543	Ertter 18001	Idaho	-	1	2 [4]	1
R. gymnocarpa	751	Lewis 15852-1	B.C.	49°02'N, 118°13'W	2 [3]	2 [4]	1
R. gymnocarpa	767	Ertter 18293a	Idaho	-	1	1	2 [4]
R. multiflora	302	Joly and Starr 558	Pa.	42°08'48.4''N, 80°08'00.1''W	2 [4]	2 [3]	2 [5]
R. nitida	570	Meilleur s.n.	Que.	-	2 [4]	2 [2]	1
R. nitida	604	Joly *et al*. 941	N.B.	45°56'29.2''N, 64°52'07.3''W	2 [3]	2 [4]	2 [-]
R. nitida	675	Brouillet 03-55-1	Nfld.	-	2 [-]	2 [4]	2 [5]
R. nitida	812	Joly 1010-1	Que.	46°22'45.3''N, 75°00'20.6''W	2 [4]	2 [1]	1
R. palustris	168	Joly and Starr 417	N.B.	45°33'43.2''N, 67°25'31.2''W	2 [4]	1	1
R. palustris	304	Joly and Starr 560	Pa.	42°09'32.9''N, 80°07'10.7''W	2 [4]	2 [4]	1
R. palustris	317	Joly and Starr 573	Ont.	42°19'41.0''N, 82°18'49.0''W	2 [4]	2 [4]	1
R. palustris	331	Joly and Starr 587	Mich.	42°19'32.0''N, 84°29'51.2''W	1	2 [3]	1
R. palustris	386	Joly and Starr 644	Wis.	44°01'30.6''N, 89°43'13.1''W	1	1	2 [2]
R. palustris	581	Joly 912	N.Y.	-	1	2 [-]	2 [5]
R. pisocarpa	774	Ertter 18303a	Calif.	-	2 [4]	2 [4]	2 [4]
R. pisocarpa	847	Ertter 18428	Calif.	41°09.2'N, 123°49.2'W	2 [4]	2 [4]	2 [4]
R. setigera	298	Joly and Starr 554	Pa.	42°08'48.4''N, 80°08'00.1''W	1	1	2 [5]
R. woodsii	4	Spellenberg 12555	N.Mex.	-	1	2 [1]	2 [6]
R. woodsii	492	Joly and Starr 752	Sask.	49°12'35.3''N, 101°50'46.1''W	1	2 [-]	2 [4]
R. woodsii	498	Joly and Starr 758	N.Dak.	48°21'09.6''N, 99°47'07.5''W	2 [-]	2 [-]	1
R. woodsii	700	Saarela 266-1	Alta.	-	2 [-]	2 [4]	2 [5]
R. woodsii	733	Dickson 2017	Alta.	-	2 [-]	2 [-]	1
R. woodsii	741	Lewis 15848-1	B.C.	49°45'N, 120°50'W	2 [3]	2 [3]	2 [5]
R. woodsii	800	Joly 1005-1	Colo.	40°12'23.4''N, 104°49'54.0''W	1	2 [3]	1
R. woodsii	807	Joly 1008-1	Colo.	40°38'36.8''N, 104°20'32.0''W	1	2 [-]	2 [-]

Notes: a – Abbreviations follow the nomenclature of Flora of North America (Flora of North America Editorial Committee, 1993); b – Approximate coordinates that were not determined by GPS.

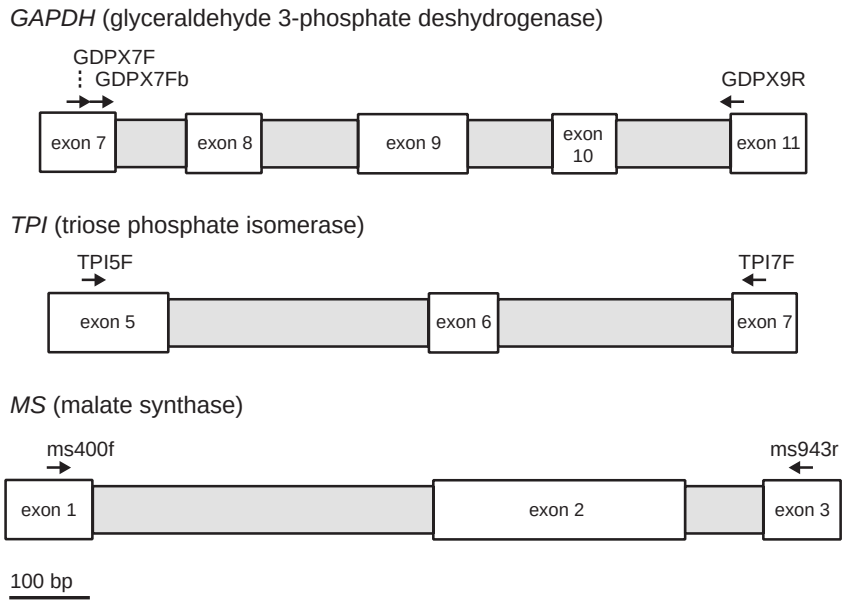


Figure 3.2 Scheme of the loci used in the study of North American *Rosa*. Primers are not to scale and their positions are approximate. Introns are in gray.

Thunb. [introduced and now a noxious invasive (Meiners *et al.*, 2001; Hunter and Mattice, 2002)], were included as outgroup taxa. DNA was extracted using the CTAB method of Doyle and Doyle (1987) as modified in Joly *et al.* (2006).

3.5.2 Gene sequencing and allele sampling

Three nuclear genes were used: the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), the triose phosphate isomerase (*TPI*), and the malate synthase (*MS*). The *GAPDH* sequences are from Joly *et al.* (2006; GenBank DQ091014-027, 030-035, 038-057, 061-069, 072-086, 172-174). *TPI* was amplified and sequenced using forward primer TPI5F (5'-AAGGTGATCGCCTGTGTTGG-3') and reverse primer TPI7R (Strand *et al.*, 1997) located in the fifth and seventh exon of the gene, respectively (Fig. 3.2). The *MS* gene was amplified and sequenced using primers ms400f and ms943r (Lewis and Doyle, 2001); the amplified region covers the first two introns of the gene (Fig. 3.2). The PCR conditions

were as in Joly *et al.* (2006) except that annealing temperatures were 52 °C and 48 °C for *TPI* and *MS*, respectively, and that a manual hotstart was used for *TPI* (i.e., the *Taq* was included after the sample reached 95 °C). PCR purification and sequencing followed Joly *et al.* (2006). Allele recovery was achieved using the procedure described in Joly *et al.* (2006). In short, individuals with no polymorphic peaks in direct sequencing were considered to be homozygous. Alleles of individuals that showed a single polymorphic site were easily extrapolated, but individuals that showed more than one polymorphic site or that had indels among its alleles needed to be cloned. Three to four clones were sequenced per individual to allow the detection of PCR-induced mutations and of in vitro recombinants. The cloning procedure is described in Joly *et al.* (2006).

3.5.3 Analyses

3.5.3.1 Recombination

For each gene, recombination was tested using the homoplasy test (Maynard Smith and Smith, 1998), the neighbor similarity score (Jakobsen and Easteal, 1996), the Max Chi-Squared (χ2; Maynard Smith, 1992), and the pairwise homoplasy index statistic (Φ; Bruen *et al.*, 2006). These methods were selected because they were demonstrated to perform well in datasets of low divergence (Posada and Crandall, 2001; Posada, 2002; Bruen *et al.*, 2006). The homoplasy test was performed without an outgroup using Maynard Smith's program (1998) under conservative ($S_E = 0.6S$) and liberal ($S_E = S$) conditions, where S_E is the effective number of sites and S is the total number of sites in the dataset. The three other methods were implemented in a program written by Bruen (2005), testing the significance of the statistics using 1000 permutations. The χ^2 test used a sliding window of size corresponding to the number of polymorphic sites divided by 1.5 and the Φ test used a relative window size (w) of 100.

3.5.3.2 Phylogenetic analyses

For each gene, gaps were coded using the simple gap-coding method (Simmons and Ochoterena, 2000) implemented in GapCoder (Young and Healy, 2003). Haplotype trees were obtained with PAUP* (ver. 4.10b, Swofford, 2002) by heuristic parsimony analysis with 10 random addition sequence replicates, each retaining a maximum of 1000 trees, TBR branch swapping and saving all minimal trees during branch swapping.

Two methods were used for obtaining allelic distance matrices from sequences. The first used allelic distances corrected using the appropriate evolutionary model, according to the Akaike Information Criterion (AIC, Akaike, 1974) calculated in Modeltest (ver. 3.7, Posada and Crandall, 1998) from a neighbor-joinning tree using the matrices without the gaps recoded and treating gaps as missing data. The second used the uncorrected distance of PAUP* to recover allelic distances from the matrices with gaps coded as presence/absence characters.

The matrices of organisms were obtained from POFAD for each gene individually and for the three genes in combination. The phylogeny of organisms was reconstructed using the NeighborNet algorithm (Bryant and Moulton, 2004) implemented in SplitsTree (Huson and Bryant, 2006).

3.6 Results

Sequences for the genes *TPI* and *MS* were deposited in GenBank (DQ200986-DQ201120) and matrices used for the analyses are available from TreeBase (study accession number S1444). All gene regions have a greater proportion of intron than exon positions in the aligned matrix, with *TPI* having a greater proportion of intron positions than the other genes for the regions under study (Table 3.2). Of the three genes, *MS* is the most variable, particularly in the exons where it has a higher number of both synonymous and non-synonymous mutations (Table 3.2). Indeed, *GAPDH*, *TPI* and *MS*

have 1, 1 and 8 variable amino acid positions, respectively. All datasets have several indels, which are all located in the intron, except one that resulted in the removal of two amino acids in the *MS* gene.

3.6.1 Recombination

Of the four methods used for detecting recombination, only the homoplasy test showed evidence of recombination, returning a positive result for all three datasets (Table 3.3). This discrepancy between methods could be the consequence of the presence of rate variation among sites in the datasets (see Table 3.2) because the homoplasy test has been shown to give false evidence of recombination in the presence of rate heterogeneity (Posada and Crandall, 2001; Posada, 2002). Therefore, it is more likely that there has been no recombination in the three datasets. Visual inspection of homoplasies on haplotype trees (see Templeton *et al.*, 1992) also did not reveal evidence of recombination, further supporting an absence of recombination in each of the three datasets.

3.6.2 Haplotype trees

Because no recombination was detected in the datasets, it is appropriate to use haplotype trees to represent the genealogy of the haplotypes for each gene. The haplotype trees differ with respect to which taxa form a clade for the different genes (Figs. 3.3a, 3.4a, 3.5a). Haplotypes of *R. gymnocarpa* form a clade with *GAPDH* and *MS*, but not with *TPI*. Haplotypes of *R. pisocarpa* only group together with *GAPDH* and none of the other species have their alleles in a single clade, yet this is sometimes the consequence of one or few incongruent haplotypes. Although haplotypes are more often closer to haplotypes of its species than to those of haplotypes from other species, the overall pattern is a lack of differentiation of species for any single gene. Despite the little information available regarding species relationships, some species are found in different

Table 3.2 Characteristics of the portions of the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), triose phosphate isomerase (*TPI*) and malate synthase (*MS*) genes used for inferring the phylogenetic relationships of *Rosa* sect. *Cinnamomeae* in North America. The best model of evolution and the gamma shape (α) are provided for each gene. The mean pair-wise divergence per site is also indicated for synonymous (d*S*) non-synonymous (d*N*) and intron positions (d(intron)).

Dataset	Length	Ratio exon/intron [a]	Variable characters [b]	Informative characters [b]	Indels	Model of evolution [c]	α [c]	d*S* [b]	d*N* [b]	d(intron) [b]
GAPDH	739-755	0.75	64	37	10	TrN + Γ	0.222	0.0084	0.0009	0.0120
TPI	806-808	0.26	54	31	7	HKY + Γ	0.263	0.0015	0.0007	0.0110
MS	995-1045	0.74	60	27	17	HKY + Γ	0.536	0.0206	0.0037	0.0144

Notes: a – Calculated from the aligned sequences; b – Excluding indels; c – Calculated with the outgroup.

Table 3.3 Recombination inference for the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), triose phosphate isomerase (*TPI*) and malate synthase (*MS*) of *Rosa* in North America. Methods used are the Homoplasy test (Homo), the neighbor similarity score (NSS), the Max Chi-squared (χ^2) and the Phi statistic (Φ) (see text). The probability for the null hypothesis of no recombination is shown for all methods.

Dataset	Mean diversity	S_E (=0.6S) [a]	P(Homo) [b]	P(NSS)	P(χ^2)	P(Φ)
GAPDH	0.9 %	325	0.000	0.637	0.973	0.922
TPI	1.1 %	422	0.000	0.205	0.304	0.139
MS	1.2 %	478	0.004	0.101	0.486	0.428

Notes: a –The effective number of sites (S_E) is calculated from the total number of sites excluding the 1st and 2nd codon positions (S); b – Only the results with the conservative conditions are shown as these are all significant.

positions in the haplotype trees. For instance, *R. gymnocarpa* is sister to all remaining North American species of sect. *Cinnamomeae* for *GAPDH* but not according to the other genes.

3.6.3 Organism trees

The two ways of recovering allelic distances – the uncorrected distance using gap information and the corrected distance according to the appropriate evolutionary model – gave similar results although including gaps gave a slightly better resolution (data not shown). For this reason, only the results obtained with the uncorrected distance are shown. This choice is further motivated by the presence of several indels in the datasets. Indels are frequent among closely related species or individuals (Britten *et al.*, 2003) and provide phylogenetic information (Kelchner, 2000) that should not be overlooked in phylogenetic studies. Moreover, because of the low divergence among species, it is less important to correct for multiple hits when calculating the distances.

The gene networks of organisms were more often congruent with the taxonomic boundaries than the haplotype trees (Fig. 3.3b, 3.4b, 3.5b). The haplotypes trees for the genes *GAPDH*, *TPI* and *MS* resolved one, zero, and one species as monophyletic, respectively, whereas the network of organisms for the same genes had three, one, and three species resolved by splits. For example, *R. foliolosa* individuals are resolved by a

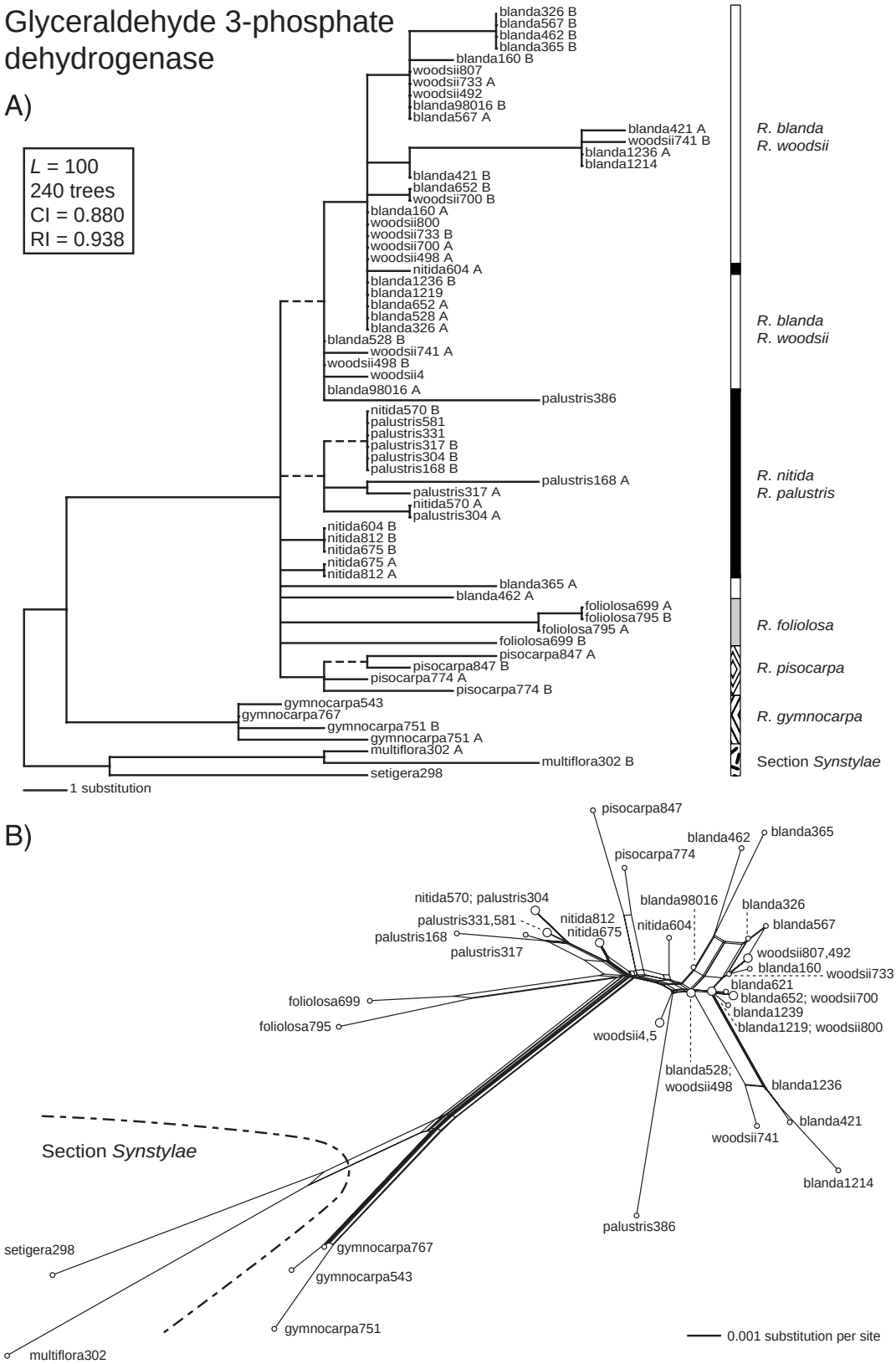


Figure 3.3 Analyses of the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) dataset. (A) One of the 240 most parsimonious haplotype trees. Dashed lines indicate branches that are not found in the strict consensus tree. (B) Phylogenetic network (NeighborNet) of the organisms.

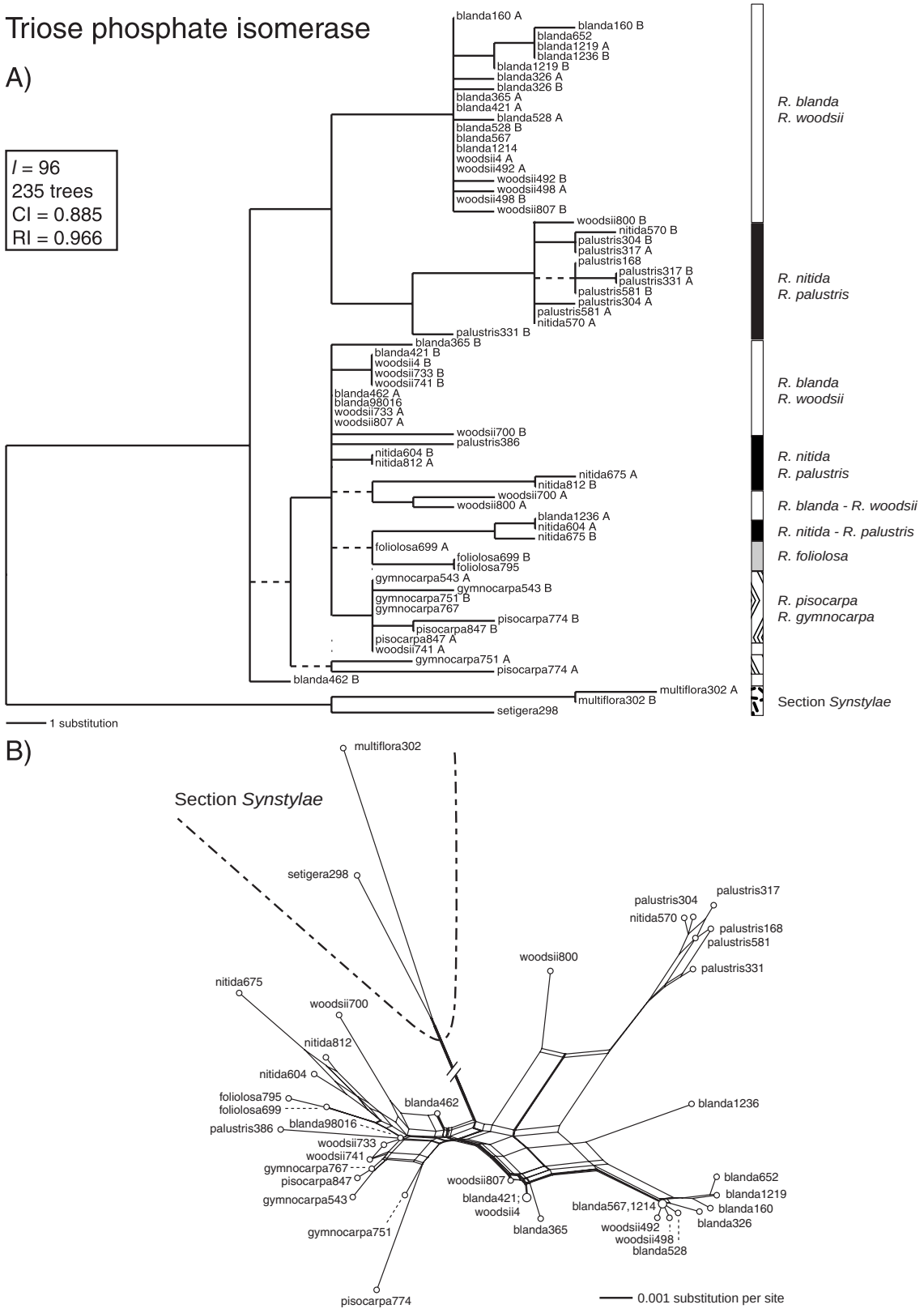


Figure 3.4 Analyses of the triose phosphate isomerase (*TPI*) dataset. (A) One of the 235 most parsimonious haplotype trees. Dashed lines indicate branches that are not found in the strict consensus tree. (B) Phylogenetic network (NeighborNet) of the organisms. The length of the branch connecting the outgroup to the ingroup is 0.016.

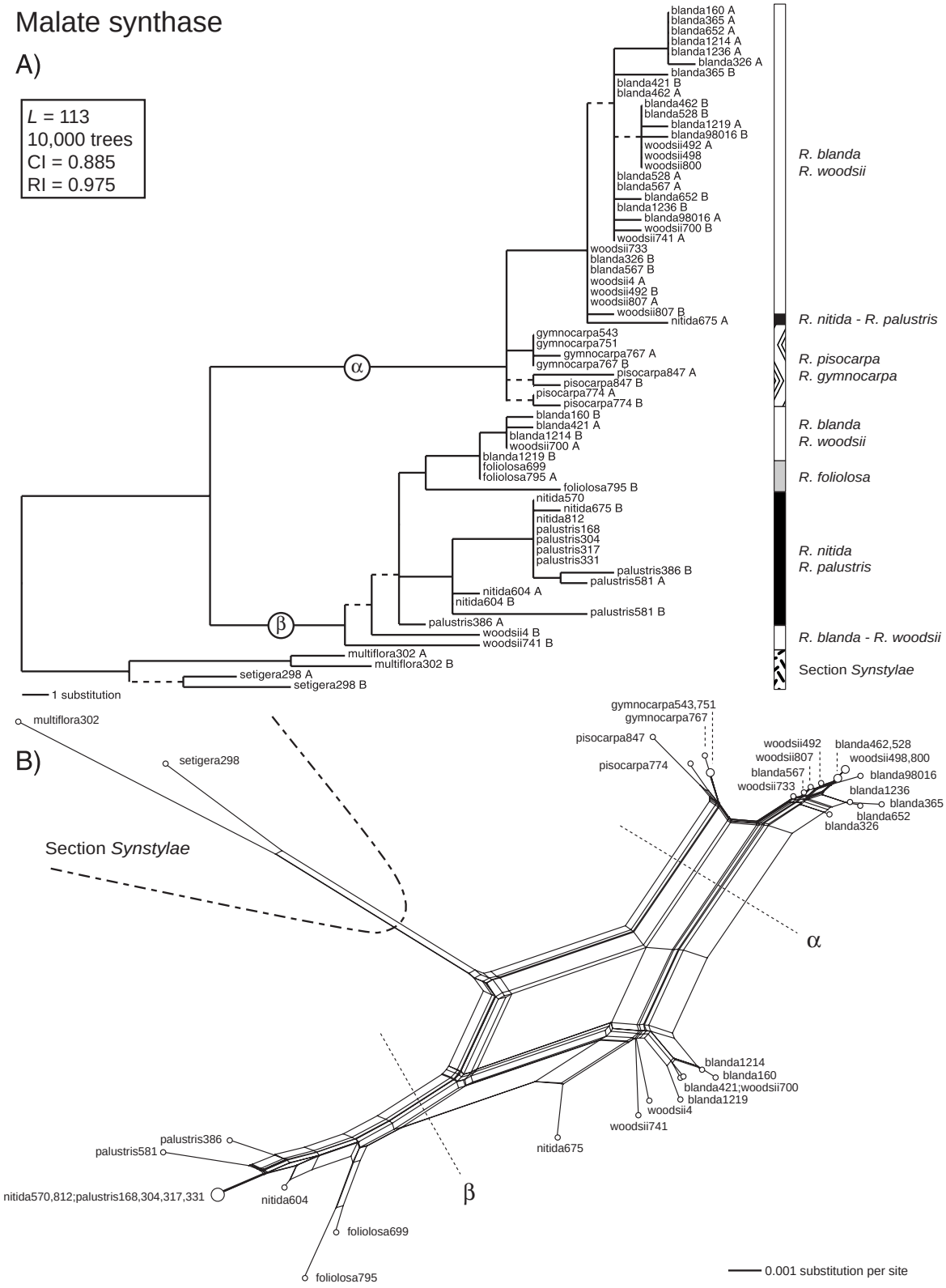


Figure 3.5 Analyses of the malate synthase (*MS*) dataset. (A) One of the 10,000 most parsimonious haplotype trees. Dashed lines indicate branches that are not found in the strict consensus tree. (B) Phylogenetic network (NeighborNet) of the organisms. And α and β indicate two genetically distinct groups of alleles (A) or individuals (B) (see text).

split in all three genes and *R. pisocarpa* individuals group together with *GAPDH* and *MS*. Similarly, individuals of *R. nitida* and *R. palustris* together are resolved by splits with *GAPDH* and *MS*, with few exceptions. Finally, *R. blanda* and *R. woodsii* individuals together are resolved by a split with *GAPDH*, although this group also includes individual *palustris*386.

The networks of organisms appear to appropriately represent intermediate individuals. For example, many individuals (*blanda*[160, 421, 1214, 1219], *woodsii*[4, 700, 741], *nitida*675) have *MS* haplotypes that occur in each of the two major clades on the haplotype tree (α and β; Fig. 3.5a). Their intermediate status is clearly represented in the network of organisms as these individuals are positioned between the clusters corresponding to the two clades in the haplotype trees (α and β; Fig. 3.5b). Similar examples are found with the other genes.

The phylogenetic network obtained when the three nuclear genes are combined (Fig. 3.6) is more resolved and relationships are clearer than when genes are analysed individually. The network clearly shows that individuals of *R. gymnocarpa* are supported by a split as are individuals of *R. pisocarpa*. However, the relationship of these western species to the eastern ones is not clear. For example, one split suggests that *R. gymnocarpa* is sister to all remaining North American species, whereas another suggests that it is closer to *R. pisocarpa* and some individuals of *R. blanda* and *R. woodsii*. Neither *R. blanda* nor *R. woodsii* are exclusive in the combined analysis, but these two species together are resolved by a weak split (a bipartition on the network for which there is another bipartition of similar or greater length that is in contradiction with), which groups all individuals except *woodsii*700. The species *R. nitida*, *R. foliolosa* and *R. palustris* are resolved as a group on the network, being supported by a weakly contradicted split. Of these three species, *R. foliolosa* individuals are clearly distinct and are strongly resolved by a split. *Rosa nitida* and *R. palustris* are not distinguished from one another but are grouped together by a weak split on the network (Fig. 3.6).

3.7 Discussion

3.7.1 The POFAD algorithm

The results obtained with the networks of organisms more closely match taxonomic boundaries than those obtained from the haplotype trees. This is probably because the proposed method increases the amount of information included per terminal by incorporating allelic variation for reconstructing the evolutionary history of organisms.

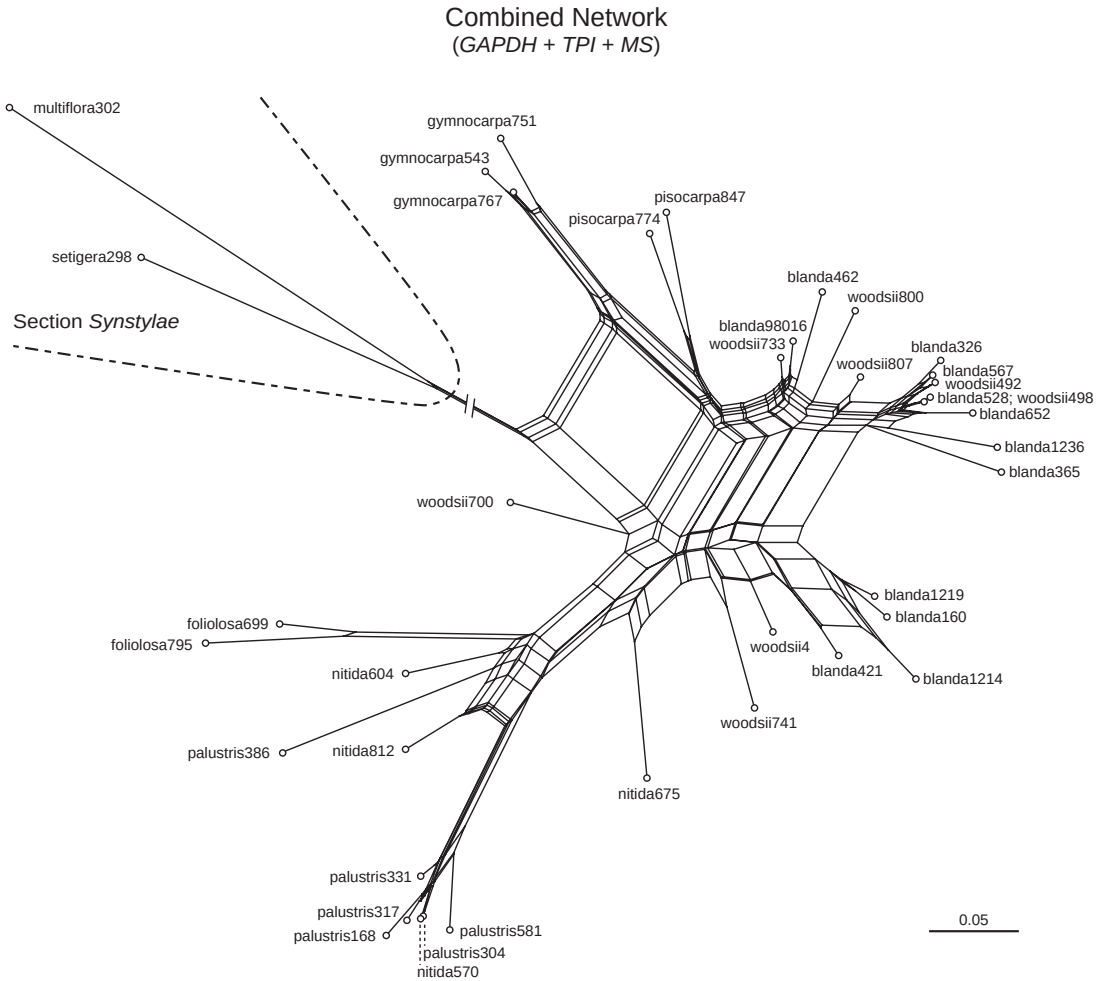


Figure 3.6 Phylogenetic network (NeighborNet) representing the relationships of the organisms obtained from the combined analysis of the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), triose phosphate isomerase (*TPI*) and malate synthase (*MS*) loci. The length of the branch connecting the outgroup to the ingroup is 0.298. The scale only provides a relative indicator of distance because the matrices were standardized.

For example, if an individual has an allele that is closer to alleles of another species because of deep coalescence, the individual could still group with its species depending on the other allele. This is indeed what happens with *R. foliolosa* that is resolved by a split in all networks of organisms but that is not monophyletic in any of the haplotype trees.

The incorporation of allelic data using the POFAD algorithm also potentially allows the detection and the representation of hybrid individuals if the phylogeny is reconstructed using a reticulate phylogenetic method. For instance, some individuals have malate synthase alleles that fall into two distinct clades in the haplotype tree; these individuals were represented as intermediate between individuals belonging to these two clades in the network of organisms (see Results and Fig. 3.5). Using both alleles instead of one for autosomal loci allows the detection of hybrid individuals with a single marker, whereas a minimum of two markers is required when only one allele per individual is sampled. The power of detecting and representing hybrid individuals in phylogenies increases as more genes are investigated because of the additional information it provides (Linder and Rieseberg, 2004), and the information contained in allelic variation should similarly increase our ablity to reconstruct the evolution of hybrid individuals when it is included in the analysis.

These examples demonstrate the importance of incorporating allelic variation whenever possible in phylogenetic analyses. Using allelic variation effectively doubles the number of lineages sampled. This increases the probability of sampling ancestral lineages within species that provide independent tests of the relationships among species (Rosenberg, 2002). With more ancestral lineages, there is an increased probability of sampling at least one lineage that will have a most recent interspecific coalescent event with its sister species, thereby improving the probabilities of recovering the species phylogeny. This is particularly important for recently diverged species where haplotypes have had less time to coalesce within the species (Rosenberg, 2002).

3.7.1.1 Combining multiple genes

The greatest interest of the POFAD algorithm certainly is its ability to incorporate allelic variation when reconstructing the phylogenetic history of organisms from multiple datasets. Because any single gene can be incongruent with the species tree, it is important to sample multiple independently evolving markers to be confident in the resulting phylogeny. When analysing multiple markers, one approach is to combine the datasets first and then to analyse of the concatenated dataset (Kluge, 1989; Yang, 1996; Seo *et al.*, 2005). This approach suffers from the fact that alleles are the terminal units of the analysis, hence hindering the concatenation of alleles from different loci because they segregate in natural populations. One solution would be to use a consensus sequence of alleles for each individual (see Howarth, 2005), therefore making the organisms the terminals units of the analysis. However, this would result in a loss of information because ambiguities are optimized so as to minimize the number of evolutionary changes in phylogenetic analyses. To illustrate this, consider a sequence that differs at a single site between two diploid individuals. Then suppose that one individual is coded as R (A or G) at the site (which means that it has one allele with an A and one with a G) and that the other individual has an A. These two individuals would then be treated as if they were identical even if the first individual has two alleles, of which one differs from the alleles of the second individual.

The alternative to the total evidence approach is the "gene as character" approach that consists of combining the trees from each marker analysed independently, either by using consensus trees (e.g.,de Queiroz, 1993), reconciled trees (Page and Charleston, 1997; Slowinski *et al.*, 1997) or supertree methods (e.g., Doyle, 1992; Sanderson *et al.*, 1998; Bininda-Emonds, 2004; Wilkinson *et al.*, 2005). As in the total evidence approach, these methods use haplotypes as terminal units and cannot incorporate allelic variation in phylogenetic analyses of multiple genes, with the exception of reconciled trees. Reconciled trees, however, differ from the POFAD method in that species, rather than individuals, are the terminal units of the analysis. Indeed, one assumption of this method

is that gene transmission is strictly vertical among the terminal units of the analysis (Page and Charleston, 1997).

Because of such problems with existing methods, studies that have used allelic variation from multiple markers have either compared the topologies of the different haplotype trees (Hare and Avise, 1998), used allelic consensus sequences for individuals in a concatenated matrix (Howarth, 2005), found concordant signals among gene trees to identify non-recombining groups of individuals (Koufopanou *et al.*, 1997), or compared the demographic events that were found to have affected each genealogy (Templeton, 2002). The method proposed in this paper provides an alternative to these options by reconstructing a single phylogeny of organisms from multiple datasets that contain allelic information.

3.7.1.2 Applicability

The POFAD method should be useful whenever haplotype trees are used such as at the intraspecific level or at the species interface among closely related species. At the intraspecific level, it could be useful for phylogeographic studies that wish to draw conclusions from more than one nuclear gene. The use of nuclear genes for phylogeographic studies is becoming more frequent (e.g., Olsen and Schaal, 1999; Hare, 2001; Antunes *et al.*, 2002; Joly and Bruneau, 2004) and some studies have already used multiple nuclear gene trees (Hare and Avise, 1998; Templeton, 2002). The proposed method could also be useful for studies at the species interface where it could help delimit species. Because alleles at nuclear loci segregate in natural populations due to sexual reproduction (gene segregation and recombination), relationships within species should be reticulate (tokogenetic) whereas they should be hierarchic (phylogenetic) among species. Tokogenetic relationships result in the sharing of alleles among individuals, which in turn tends to make individuals within species more similar to each other than to individuals of other species. This also implies that there should be no shared phylogenetic patterns among genes within species. In contrast, strong

phylogenetic signals shared by a majority of genes should correspond to the speciation event (Koufopanou *et al.*, 1997). These speciation events should therefore result in strong splits in the combined network of organisms, if inter-specific hybridization does not occur.

3.7.2 Phylogeny of North American diploid roses

Little is known of phylogenetic relationships among rose species in North America. Previous studies provided little information because of the low resolution of molecular markers and of poor species sampling (Millan *et al.*, 1996; Matsumoto *et al.*, 1998; Wissemann and Ritz, 2005). In contrast, the three nuclear genes sequenced for several individuals per species in this study allow an assessment of phylogenetic relationships among North American species but also provide information regarding species boundaries.

Firstly, the diploid species of *Rosa* in North America appear to be of recent origin according to the low levels of genetic variation found in haplotype trees. Yet, it is also possible that the long generation time, which is typical for shrubs, could accentuate this trend. A rapid radiation is also supported by the lack of monophyly observed for most species. Indeed, recently diverged species are not expected to be reciprocally monophyletic and incomplete lineage sorting is expected to be frequent among such species (Rosenberg, 2002, 2003; Degnan and Salter, 2005). Nevertheless, polyphyletic species could also be the consequence of inter-specific gene flow that is indicative of poorly defined species boundaries. Or course, the phenomenon responsible for non-monophyletic species is likely to be different from one species to another. But despite the low levels of genetic variation and the absence of monophyly for most species for one or more of the genes studied, the combined analysis of individuals remains informative regarding the phylogenetic relationships of North American species.

Botanists have generally treated the western and eastern North American rose species as distinct entities (Lewis, 1957c; Erlanson MacFarlane, 1966). Yet, the hypothesis that western and eastern species form distinct phylogenetic groups has never been tested. The combined network suggests that a distinction between the west and the east may exist, although it is only supported by a weak split. Relative to the outgroup species of section *Synstylae*, one strong split suggests that *R. gymnocarpa* is sister to all remaining North American species, a signal mostly contributed by the *GAPDH* gene. The alternative solution, which is supported by a split of similar strength contributed mostly by the *MS* gene, groups *R. gymnocarpa* with *R. pisocarpa* and some individuals of *R. blanda* and *R. woodsii*. Congruent with the latter solution, a split on the network supports the monophyly of western species, but this split is rather weak. Because of the incongruence regarding the exact position of the western species among the genes studied, more genes will have to be investigated to determine the exact branching pattern and to confirm the distinction between western and eastern diploid species. Individually, however, both western species *R. gymnocarpa* and *R. pisocarpa* form exclusive groups of individuals suggesting there is little or no genetic exchange between them. Thus even if the sampling is limited for these species, the results suggest that these species are distinct.

In the east, the combined network shows that species are divided into two clear groups: one consists of *R. blanda* and *R. woodsii*, and the other of *R. foliolosa*, *R. nitida* and *R. palustris*. In the former group, individuals of *R. blanda* and *R. woodsii* cannot be distinguished from one another. However, both species together form a genetically variable group that is supported by a split in the combined analysis, with the exception of the *woodsii*700 individual. The high genetic diversity observed in this group may be explained in part by the widespread distribution of these species that could reduce the homogenizing effect of gene flow. *Rosa woodsii* ranges from California and British Columbia to the eastern Great Plains, whereas *R. blanda* is distributed from Manitoba and Minnesota in the west to New Brunswick and Maine in the east.

Several clues suggest that the lack of differentiation between *R. blanda* and *R. woodsii* is caused by ongoing gene flow. These species are indeed ecologically (they grow in mesic soils along woods and rivers) and morphologically similar, and are difficult to tell apart (Lewis, 1962). Moreover, hybrids between these species have been shown to be highly fertile (Erlanson, 1934; Ratsek *et al.*, 1939) and in the area where the two species overlap, Lewis (1962) described a hybrid zone. Clearly, the species status of these taxa needs to be reassessed.

The second eastern group revealed by the combined network consists of *R. foliolosa*, *R. nitida* and *R. palustris*. This group is congruent with morphological data because these species share many characters that distinguish them from other North American species. Indeed, they represent all the diploid species that were once placed in sect. *Carolinae* (Crépin, 1889).

Within this group, *R. foliolosa* distinguishes itself from the other species by having its two individuals clearly resolved as a group on the network. Although only two individuals were investigated for *R. foliolosa*, the network suggests that it is genetically distinct from the other species. *Rosa foliolosa* is also distinct from the other species morphologically, being characterized by narrow leaflets and short pedicels (Lewis, 1957b, 1958). This species is also peculiar for having the smallest geographic distribution of all species of sect. *Cinnamomeae* in North America, as it occurs only in Oklahoma, Texas and western Arkansas (Lewis, 1958).

Individuals of the last two species, *R. nitida* and *R. palustris*, cannot be distinguished from one another on the network but together are supported as a group, albeit by a weak split. If we consider that *R. foliolosa* individuals are clearly distinct from individuals of these species, then *R. nitida* and *R. palustris* together form a rather cohesive group. A close relationship between these species is not surprising as both have narrow stipules, hypanthium glands, and a preference for bogs and poorly drained soils. In contrast with *R. blanda* and *R. woodsii*, however, *R. nitida* and *R. palustris* are clearly distinct

morphologically (Lewis, 1957c, 1957b). This suggests that the lack of genetic distinction between these species is the consequence of a recent origin rather than of poorly defined species boundaries. Although the prevalence of incomplete lineage sorting among species suggests that little time has occurred since the formation of species, the often small populations of these roses and the patchiness of populations over wide geographic areas could also contribute to the retention of ancient polymorphisms. For example, the *palustris*386 individual is from the western extremity of the distribution of *R. palustris*, where few populations are found. This could explain why this individual has retained alleles that are more closely related to *R. blanda* and *R. woodsii* haplotypes for the *GAPDH* and *TPI* genes.

3.7.3 Gene trees and species tree and individual sampling within species

In agreement with most phylogenetic studies investigating multiple markers, incongruence was observed among gene trees obtained from the three loci investigated (Chen and Li, 2001; Cronn and Wendel, 2003; Doyle *et al.*, 2003; Rokas *et al.*, 2003; Jennings and Edwards, 2005). Although some of the incongruence results from the relative position of species among gene phylogenies (e.g., *R. gymnocarpa*), most of the incongruence observed in this study was caused by the lack of monophyly of the species. Such incongruence could be the result of paralogy, incomplete lineage sorting, or hybridization. No sign of gene duplication was noted in this study so paralogy does not seem to be the cause of the lack of species monophyly. Incomplete lineage sorting is more likely to be the cause of incongruence when the incongruent allele is distant from alleles of other species and when their divergence is basal (Holder *et al.*, 2001; Funk and Omland, 2003; Joly *et al.*, 2006). This appears to be case for the allele *palustris*386 that falls within the group of *R. blanda* and *R. woodsii* individuals in the *GAPDH* haplotype tree. In contrast, hybridization should cause an incongruent haplotype to have diverged recently and to be similar to alleles of another species (Holder *et al.*, 2001; Funk and Omland, 2003; Joly *et al.*, 2006). For example, hybridization could explain the position of allele A of *nitida*604 in the *GAPDH* haplotype tree, which is located in an otherwise exclusively *R.*

blanda and *R. woodsii* clade. It is not always obvious how to distinguish the two processes, however, and it may be often impossible to be confident of the process that caused incongruence (Holder *et al.*, 2001; Joly *et al.*, 2006).

Incongruence caused by non-monophyletic species demonstrates the importance not only of sampling many genes but also of sampling many individuals per species when reconstructing the phylogenetic history of closely related species. Rosenberg (2002) indeed showed that enhanced haplotype sampling increases the probability that the gene tree is topologically concordant with the species tree, in particular for recent radiations as in North American diploid roses. Maddison and Knowles (2006) arrive at the same conclusion in a simulation study demonstrating that given limited resources, it is more advantageous to sample more individuals per species for a single gene than to sequence few individuals for more genes if the species have diverged recently. As discussed above in the context of allelic variation, sampling more individuals increases the probability of sampling ancestral lineages and gives a better chance of accurately reconstructing the phylogenetic history of species, particularly when recently diverged (Rosenberg, 2002).

Studies that assess the gene tree – species tree problem often sample a single individual per species and highlight incompatibilities among the phylogenies obtained from different genes. In these studies, a gene can only be congruent or incongruent with the species tree. Yet, it is probably more frequent that for any particular gene there will be some haplotypes that agree with the species tree and some that will be incongruent with it. As noted by Rosenberg (2003), without an appropriate sampling of individuals within species, one could conclude that a gene has coalesced within the species when it has not. Such incorrect inferences could result in biased conclusions concerning the evolutionary processes involved in speciation (Funk and Omland, 2003).

The algorithm present in this paper gives an overview of the importance of incorporating allelic variation into the phylogenetic analysis of organisms. Consequently, we hope that this study will be a starting point for the development of other methods and

that it will stimulate other studies that will make use of allelic data into phylogenetic analyses as it represents an important source of information that too often is neglected.

CHAPITRE 4

Évolution des roses diploïdes : une approche spécifique

4.1 Abstract

This chapter aims at reconstructing the phylogeny of diploid species of *Rosa* in North America. The method of Madison (Syst. Biol. 1997. 46:523), applied for the first time on morphological data, was used to find the species tree from several gene trees. According to a parsimony criterion, the best species tree minimizes the number of deep coalescences when gene trees are forced to evolve in it. This method uses species as terminal units of the analysis and therefore assumes that species are well characterised. Consequently, the species identified earlier were used for the analysis (Chapter 2). The species tree for *Rosa* species was reconstructed from the *GAPDH*, *TPI* and *MS* gene trees. The search resulted in one species tree that implied a total of 41 deep coalescences for the three genes. It suggest that the western species *R. gymnocarpa* and *R. pisocarpa* are sister species and that they are sister to all eastern species. In the east, *R. foliolosa* is ancestral and *R. nitida* and *R. palustris* are sister species. The position of *R. pisocarpa* is not stable as a tree with only two additional deep coalescences placed *R. pisocarpa* nested in the eastern species. This method focussing on species relationships represents a complementary analysis to that based on organisms (as in Chapter 3).

4.2 Introduction

Dans le chapitre précédent (Chapitre 3), la phylogénie des roses diploïdes de la section *Cinnamomeae* de l'Amérique du Nord a été reconstruite en utilisant les individus

comme unités terminales de l'analyse phylogénétique. Une autre approche est possible, cependant, qui consiste à utiliser les espèces comme unité terminale et à utiliser les arbres géniques pour déterminer quelle phylogénie est la meilleure. C'est le problème classique des arbres d'espèces *vs* les arbres géniques (Pamilo et Nei, 1988; Takahata, 1989; Wu, 1991; Doyle, 1992; Hudson, 1992; Maddison, 1997; Page et Charleston, 1997; Rosenberg, 2002; Degnan et Salter, 2005). Tel que discuté précédemment (Chapitre 3), les populations de deux espèces nouvellement formées partagent souvent plusieurs allèles et il est possible que ces allèles ne soient pas éliminés, même après plusieurs générations. En fonction des allèles échantillonnés pour l'analyse phylogénétique, ces polymorphismes ancestraux pourraient donner un arbre de gènes incongruent avec l'arbre des espèces. Ces rétentions de polymorphismes ancestraux, aussi appelées coalescences profondes (*deep coalescences*), sont plus fréquentes lorsque les espèces sont récentes et que la taille effective des populations sont grandes (Kingman, 1982a, 1982b; Rosenberg, 2002, 2003). Cependant, le nombre de gènes en accord avec la « vraie » phylogénie des espèces sera toujours plus grand que le nombre de gènes appuyant des phylogénies alternatives (Takahata, 1989; Rosenberg, 2002).

Dans son important article sur cette question, Maddison (1997) a décrit une méthode basée sur le principe de parcimonie pour trouver l'arbre des espèces qui représente le mieux un ensemble d'arbres géniques. La méthode vise à choisir l'arbre génique qui minimise le nombre de coalescences profondes pour l'ensemble des gènes étudiés. Pour évaluer le nombre de coalescences profondes, on doit forcer un arbre de gène dans un arbre d'espèces et regarder pour chaque branche le nombre de lignées qui évoluent parallèlement. Prenons par exemple l'arbre des espèces de la figure 4.1a. Si l'on force l'arbre génique de la figure 4.1b dans la phylogénie des espèces (Fig. 4.1c), deux lignées doivent obligatoirement coexister le long de la branche menant au groupe (C,D) pour expliquer l'arbre génique. Le nombre de coalescences profondes pour une branche correspond au nombre de lignées coexistantes moins 1. Dans l'exemple de la figure 4.1, une coalescence profonde est nécessaire pour expliquer l'évolution du gène dans l'arbre des espèces.

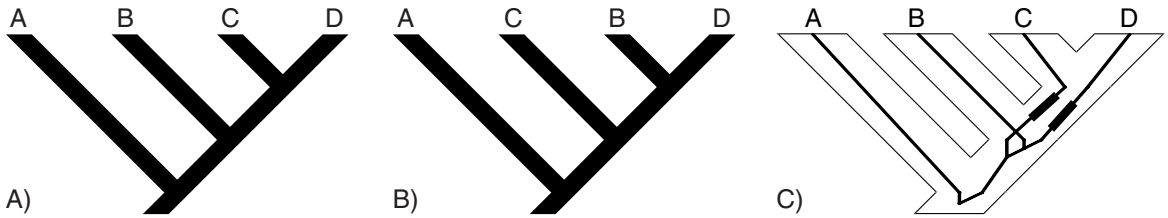


Figure 4.1 Exemple démontrant l'évaluation du nombre de coalescences profondes entre un arbre d'espèces (A) et un arbre de gènes (B). Lorsque l'arbre génique est forcé sur l'arbre des espèces (C), le nombre de coalescences profondes pour chaque branche est le nombre de lignées qui coexistent (représentées par les deux rectangles noirs) moins 1, ce qui donne une coalescence profonde pour cet exemple.

Étrangement, cette méthode n'a pas encore été utilisée sur des données empiriques, du moins à ma connaissance. Une raison pouvant expliquer cela serait le manque de jeux de données pouvant être analysés ainsi. Cependant, avec le nombre croissant d'études reconnaissant l'importance de l'échantillonnage intra-spécifique pour les analyses phylogénétiques (Rosenberg, 2002; Funk et Omland, 2003; Rosenberg, 2003; Ayoub *et al.*, 2005), l'utilisation de ce type d'analyse devrait devenir plus fréquent. Néanmoins, une récente analyse de simulations a démontré que cette méthode donnait de bons résultats et ce même lorsque les espèces sont d'origine récente et que peu de gènes sont utilisés (Maddison et Knowles, 2006).

Un inconvénient de cette méthode est que les espèces doivent être définies *a priori*, ce qui n'est pas requis lorsque les individus sont utilisés comme unité terminale. Autrement dit, si l'on n'est pas certain des frontières entre les espèces dont la phylogénie est reconstruite, cette méthode ne permet pas de statuer sur leur validité. Dans le cas présent, les espèces délimitées précédemment (Chapitre 2) permettent d'utiliser cette méthode. Ainsi, il sera possible de tester quelle phylogénie des espèces est la meilleure selon un critère minimisant les incongruences entre les arbres de gènes et l'arbre des espèces.

4.3 Méthodes

Les espèces de roses diploïdes dont on veut reconstruire la phylogénie sont les espèces diploïdes de la section *Cinnamomeae* de l'Amérique du Nord. Pour ces espèces, les groupes d'espèces identifiées comme distincts au Chapitre 2 seront utilisés : *R. blanda* – *R. woodsii*, *R. foliolosa*, *R. nitida* et *R. palustris*. Bien que la délimitation des espèces de l'ouest n'ait pas été étudiée en profondeur, les espèces *R. gymnocarpa* et *R. pisocarpa* seront tout de même incluses dans l'analyse. Finalement, les deux espèces de la section *Synstylae* présentes en Amérique du Nord, *R. setigera* et *R. multiflora*, serviront d'extra-groupe. Celui-ci semble être important puisque si la méthode de Maddison (1997) permet d'identifier la racine d'un arbre non enraciné, l'enracinement obtenu n'est pas toujours bon (Maddison et Knowles, 2006).

Les gènes étudiés sont les mêmes trois gènes utilisés pour l'analyse phylogénétique qui utilisait les individus comme taxons terminaux (Chapitre 3), soit la glycéraldehyde 3-phosphate déshydrogénase (*GAPDH*), la triose phosphate isomerase (*TPI*) et la malate synthase (*MS*). Nous savons donc déjà qu'il n'y a pas de recombinaison à l'intérieur de ces marqueurs et qu'ils sont à copie unique (Chapitre 3). Les arbres ont été obtenus par analyse de parcimonie heuristique de la même façon que pour l'obtention des arbres d'haplotypes au Chapitre 3. Lorsque plusieurs arbres ont été obtenus pour un même gène, le consensus strict a été utilisé pour la reconstruction de l'arbre des espèces.

L'arbre des espèces minimisant les coalescences profondes a été reconstruit par analyse de parcimonie heuristique à l'aide du module Coalescence (Maddison, 2005) du logiciel Mesquite (Maddison et Maddison, 2005). Une addition de séquences « as is » a été utilisée, suivi d'un réarrangement des branches par SPR (la seule option disponible) en sauvant 1 arbre à chaque étape. Certaines hypothèses évolutives alternatives ont été testées en forçant les arbres de gènes dans des topologies prédéfinies. Parmi les hypothèses testées figuraient le monophylétisme des espèces ouest-américaines (*R.*

gymnocarpa et *R. pisocarpa*), le monophylétisme des espèces est-américaines et le monophylétisme du clade de *R. foliolosa*, *R. nitida* et *R. palustris*.

4.4 Résultats

Le nombre d'individus échantillonnés ainsi que le nombre d'allèles inclus dans l'analyse pour chaque gène et chaque espèce sont indiqués au tableau 4.1. Les matrices qui ont servi à l'analyse phylogénétique des gènes ainsi que les consensus stricts obtenus suite à l'analyse de parcimonie (voir Fig. 3.3, 3.4, 3.5) ont été soumis à TreeBase (numéro d'accession de l'étude : S1444). Les arbres de gènes sont discutés en détails au chapitre 3.

L'analyse phylogénétique de l'arbre des espèces qui visait à minimiser le nombre de coalescences profondes pour l'ensemble des gènes a résulté en une seule topologie (Fig. 4.2). Un total de 41 coalescences profondes est nécessaire pour forcer l'évolution des arbres de gènes à l'intérieur de cet arbre d'espèces. Cet arbre considère que les deux espèces présentes à l'ouest des Rocheuses, *R. gymnocarpa* et *R. pisocarpa*, forment un groupe monophylétique qui est le groupe frère d'un autre clade comprenant toutes les espèces est-américaines. Dans le groupe est-américain, *R. foliolosa* diverge en premier, suivi du groupe de *R. blanda* – *R. woodsii*, et *R. nitida* et *R. palustris* sont groupes-frères.

Tableau 4.1 Nombre d'individus et d'allèles qui ont été utilisés pour chaque gène pour la reconstruction phylogénétique des arbres géniques.

Espèce	Nb. d'individus	Nb. d'allèles		
		GAPDH	*TPI*	*MS*
R. blanda – R. woodsii	20	33	36	37
R. foliolosa	2	4	3	3
R. gymnocarpa	3	4	5	4
R. multiflora	1	2	2	2
R. nitida	4	8	8	6
R. palustris	6	9	10	8
R. pisocarpa	2	4	4	4
R. setigera	1	1	1	2

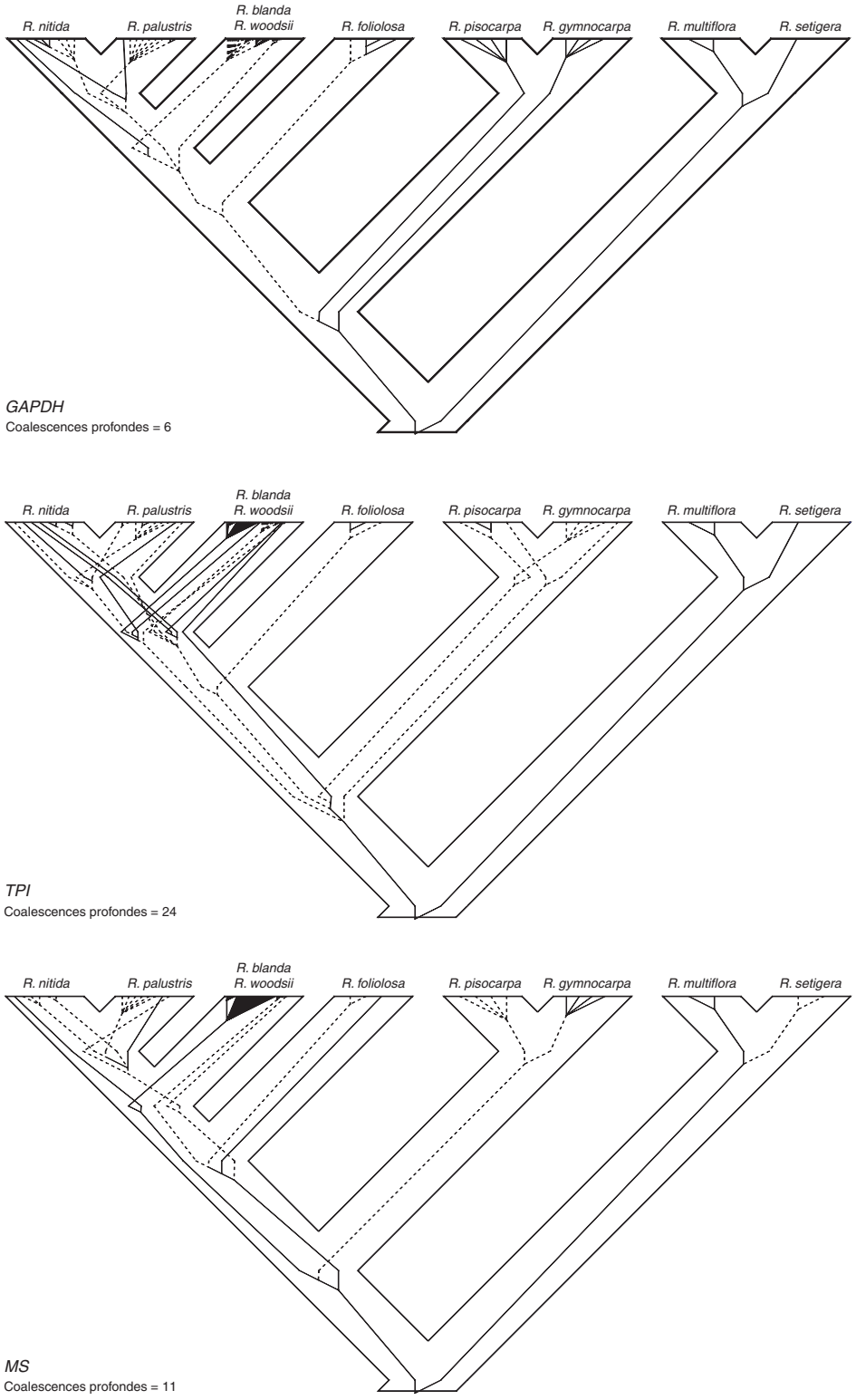


Figure 4.2 Arbre des espèces de roses diploïdes de la section *Cinnamomeae*, contenant les arbres des gènes *GAPDH*, *TPI* et *MS*. Les lignes pointillées représentent des branches qui proviennent de polytomies dans les arbres géniques.

Lorsque les arbres de gènes sont contraints dans la topologie de l'arbre obtenu, le gène *TPI* est celui qui a le plus grand nombre d'incongruences (24 coalescences profondes), suivit de *MS* (11) et de *GAPDH* (6) (Fig. 4.2). Même si l'analyse a résulté en une seule topologie, d'autres topologies ont reçu des scores très près de l'arbre optimal (Fig. 4.3). Les arbres qui ont obtenu des scores presque identiques à l'arbre optimal variaient seulement relativement à la position du *R. pisocarpa* sur l'arbre. D'ailleurs, les incongruences entre gènes relativement à la position du *R. pisocarpa* sont évidentes lorsque l'on compare les topologies des différents gènes dans l'arbre des espèces (Fig. 4.2). Les topologies examinées montrent aussi que contraindre *R. nitida*, *R. palustris* et *R. foliolosa* dans un groupe monophylétique donne de relativement moins bons résultats (Fig 4.3D, E).

4.5 Discussion

Bien que les phylogénies d'organismes fournissent beaucoup d'information sur l'évolution d'un groupe et qu'elles permettent dans certains cas de statuer sur la validité de certaines espèces (Chapitre 3), c'est davantage l'évolution des espèces qui intéresse la plupart des biologistes. Il est souvent important de savoir dans quelle séquence les espèces ont divergé et quelles sont les espèces ancestrales dans un groupe donné. En

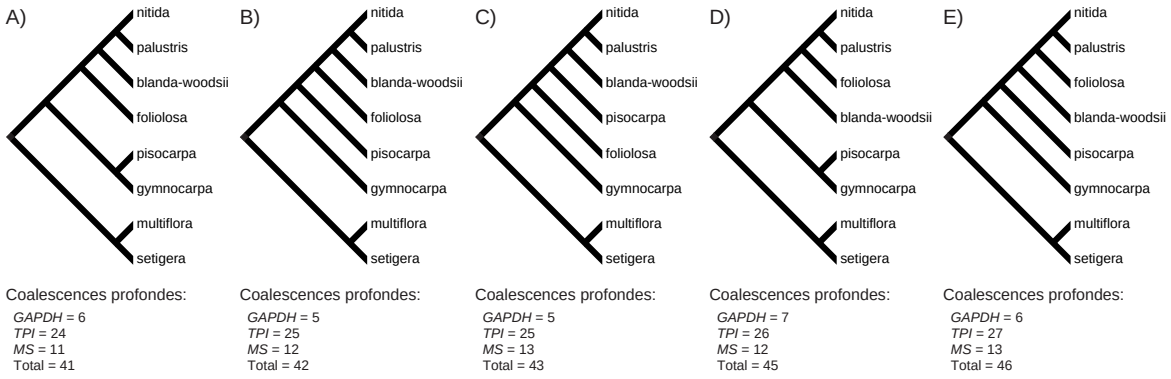


Figure 4.3 Comparaison du meilleur arbre d'espèces obtenu (A) avec des alternatives légèrement moins parcimonieuses (B-E).

fonction de cet aspect, l'approche spécifique de Maddison (1997) est intéressante et fournit une alternative à l'approche fondée sur les organismes (Chapitre 3).

La méthode de Maddison (1997) est aussi intéressante puisqu'elle permet de s'attaquer directement au problème d'incompatibilité entre les arbres de gènes et les arbres d'espèces. Elle n'est cependant pas la seule méthode disponible pour pallier à ce problème. Par exemple, certains auteurs ont proposé que l'arbre des espèces devrait correspondre à la topologie rencontrée le plus fréquemment lorsque plusieurs arbres géniques sont considérés (e.g., Jennings et Edwards, 2005). Cependant, ce critère, contrairement à la méthode de Maddison, n'intègre pas directement la sélection de l'arbre des espèces dans une analyse phylogénétique. De plus, pour donner de bons résultats, une telle analyse doit se baser sur plusieurs arbres de gènes. Ce dernier pré-requis n'est pas nécessaire pour la méthode de Maddison (e.g., Maddison et Knowles, 2006).

4.5.1 La phylogénie des roses diploïdes d'Amérique du Nord

Les résultats diffèrent des résultats obtenus lorsque les individus étaient les taxons terminaux de l'analyse (Chapitre 3). La principale différence réside dans la position du *R. foliolosa*, qui se groupait avec les *R. nitida* et *R. palustris* lorsque les organismes étaient les unités terminales de l'analyse (Fig. 3.6), alors qu'ici le *R. foliolosa* se retrouve groupe-frère de toutes les espèces est-américaines (Fig. 4.2). Cette différence est peut-être causée par une différence méthodologique. La méthode POFAD (Chapitre 3) est basée sur des distances, et la phylogénie groupera préférentiellement les individus qui se ressemblent le plus. Par contre, la méthode minimisant les coalescences profondes est historique et se fie principalement sur des différences topologiques entre les phylogénies géniques et non sur la similarité générale entre les séquences. Ainsi, il se peut que la phylogénie des organismes (Chapitre 3) soit plus sensible à des phénomènes de convergence que la méthode utilisée ici.

Cette nouvelle hypothèse au sujet de la position du *R. foliolosa* est intéressante. En effet, il aurait semblé naturel que le *R. foliolosa* forme un clade avec les *R. nitida* et *R. palustris* étant donné leur ressemblance morphologique et génétique (Chapitre 2). Cependant, la présente analyse montre que cette hypothèse est moins parcimonieuse puisqu'une telle contrainte requiert au moins quatre coalescences profondes supplémentaires pour expliquer l'évolution des gènes étudiés (Fig. 4.3). Il est toutefois important de noter que seulement deux individus du *R. foliolosa* ont été inclus dans cette analyse et que les résultats pourraient changer avec l'ajout d'autres individus. De plus, d'autres gènes devront être étudiés pour confirmer ces résultats puisque seulement trois gènes ont été investigués ici et il se peut que ceux-ci ne reflètent pas l'évolution des espèces.

La phylogénie obtenue ici suggère que les espèces de l'est et de l'ouest forment deux clades distincts. Cependant, la position réelle du *R. pisocarpa* n'est peut-être pas complètement certaine si l'on considère les arbres qui ont obtenu des scores similaires au meilleur arbre (Fig. 4.3). En effet, ces topologies qui sont presque aussi bonnes que l'arbre le plus parcimonieux ne varient que pour la position du *R. pisocarpa*, qui avec seulement deux coalescences profondes supplémentaires, pourrait se retrouver à l'intérieur du clade est-américain (Fig. 4.3). Tout comme pour le *R. foliolosa*, seulement deux individus du *R. pisocarpa* ont été échantillonnés pour l'étude et l'ajout d'individus, tout comme l'ajout d'autres gènes, permettrait sans doute d'obtenir des résultats plus fermes.

Une hypothèse n'a pas été testée ici et c'est la possibilité que l'arbre des espèces possède des polytomies réelles (*hard polytomies*). Même si une telle hypothèse n'est pas à exclure biologiquement, elle n'est pas incluse dans l'analyse pour une raison bien évidente. En effet, pour chaque branche impliquant des incongruences, il sera toujours plus parcimonieux de créer une polytomie car la branche supprimée réduira automatiquement le nombre total de coalescences profondes observées pour l'arbre. Ces

hypothèses de polytomies seront donc toujours plus parcimonieuses, mais moins informatives.

4.5.2 Coalescence profonde ou hybridation ?

Une limitation de la méthode utilisée pour reconstruire l'arbre des espèces est qu'elle considère tous les types d'incongruence entre un arbre génique et un arbre d'espèce comme des coalescences profondes (Maddison, 1997). Cela peut représenter un problème puisqu'un seul événement d'hybridation entre deux espèces éloignées pourrait être comptabilisé comme plusieurs coalescences profondes. En fait, plus les espèces impliquées dans l'événement d'hybridation seront éloignées, plus le nombre de coalescences profondes sera grand puisque le nombre de branches le long desquelles les lignées géniques doivent coexister sera aussi plus grand. Ainsi, ce qui représente un seul événement évolutif pourrait être comptabilisé comme plusieurs coalescences profondes. Ce problème n'est pas facile à résoudre, en partie parce qu'il est très difficile de distinguer ces deux processus évolutifs (Holder *et al.*, 2001; Joly *et al.*, 2006). Et même s'il était possible de distinguer ces processus, il reste le problème associé à la pondération de ces différents événements évolutifs (Maddison, 1997); un événement d'hybridation vaut-il plus qu'un événement de coalescence profonde?

En général, il semble y avoir peu d'incongruences qui soient dues à l'hybridation chez les roses diploïdes de la section *Cinnamomeae* en Amérique du Nord pour les gènes utilisés (Chapitre 3 et 5). Donc, bien que ces deux processus évolutifs n'aient pas été pris en compte dans la présente analyse, le fait de ne pas considérer les processus d'hybridation ne devrait pas trop affecter les résultats.

4.5.3 Validation et stabilité

Un autre inconvénient de l'implémentation actuelle de la méthode de Maddison (1997) est qu'il est impossible d'obtenir des valeurs de support pour les différents

groupes monophylétiques. Une méthode adéquate pour évaluer la stabilité des groupements impliquerait sans doute un ré-échantillonnage des allèles de chaque gène, suivi d'une analyse phylogénétique pour chaque gène et d'une analyse phylogénétique pour trouver l'arbre des espèces minimisant le nombre de coalescences profondes. Une autre possibilité, plus rapide et plus simple, serait de ré-échantillonner les allèles des arbres géniques et de refaire seulement l'analyse phylogénétique pour trouver l'arbre des espèces. Cette dernière méthode suppose toutefois qu'enlever des allèles ne devrait pas influencer la topologie des arbres géniques. Comme de telles méthodes ne sont pas encore disponibles dans Mesquite, il est possible de tester des phylogénies alternatives en forçant les arbres de gènes sur différents arbres d'espèces comme dans le cas présent. Comme Mesquite ne permet pas présentement d'avoir des contraintes dans l'analyse phylogénétique, l'évaluation des topologies alternatives doit se faire manuellement. Cela peut fonctionner lorsqu'un petit nombre d'espèces est étudié, mais une telle démarche s'avérerait laborieuse pour un grand nombre d'espèces. Une autre alternative serait d'utiliser le programme GeneTree (Page, 2001), qui implémente la même méthode et qui permet d'utiliser des contraintes lors de l'analyse phylogénétique. Cependant, GeneTree accepte seulement des arbres géniques complètement résolus, ce qui peut causer des problèmes lorsque ces derniers ont d'importantes polytomies comme dans la présente étude. Cette alternative devrait donc être utilisée seulement lorsque les arbres géniques sont presque entièrement dichotomiques.

Il est aussi souhaitable d'augmenter la stabilité de la phylogénie obtenue. Comme la théorie de la coalescence prédit que plus de gènes devraient être concordant avec la vraie phylogénie des espèces qu'avec d'autres topologies (Kingman, 1982a, 1982b; Pamilo et Nei, 1988; Takahata, 1989; Maddison, 1997; Rosenberg, 2002), cette méthode devrait donner une meilleure estimation de la « vraie » phylogénie des espèces à mesure que le nombre de gènes étudiés augmente. Cependant, bien que le nombre de gènes est important pour obtenir une bonne phylogénie, le nombre d'individus échantillonnés par espèce semble tout aussi important (Rosenberg, 2002; Maddison et Knowles, 2006). Ceci est d'ailleurs particulièrement vrai pour les espèces d'origine récente (Rosenberg, 2002).

Étant donné le nombre de coalescences ancestrales entre les espèces étudiées, il semble bien que les espèces de roses diploïdes de la section *Cinnamomeae* en Amérique du Nord soient d'origine récente. Donc, il serait important d'augmenter à la fois le nombre de gènes échantillonnés et le nombre d'individus étudiés par espèce dans les études futures. Une attention particulière devrait d'ailleurs être accordée aux espèces *R. pisocarpa* et *R. foliolosa*, dont la position dans la présente phylogénie semble être la moins stable.

CHAPITRE 5

Polyploid and hybrid evolution in roses east of the Rocky Mountains[3]

5.1 Résumé

Cet article se penche sur l'impact de l'hybridation et de la polyploïdie dans l'évolution des espèces de roses nord-américaines. Ces processus sont explorés dans le complexe du *Rosa carolina* (section *Cinnamomeae*), qui consiste en cinq espèces diploïdes et trois tétraploïdes. Afin de clarifier le statut et les origines des diploïdes et des polyploïdes, un réseau d'haplotypes (parcimonie statistique) du gène nucléaire codant pour la glycéraldéhyde 3-phosphate déshydrogénase a été reconstruit pour les polyploïdes et les diploïdes de la section *Cinnamomeae* en Amérique du Nord. Une approche généalogique a permis de reconstruire l'histoire évolutive des polyploïdes en dépit du bruit créé par l'hybridation, le triage incomplet de lignées (*incomplete lineage sorting*) et la ségrégation allélique chez les polyploïdes. Au niveau diploïde, les espèces à l'ouest des montagnes Rocheuses se distinguent des espèces est-américaines. Dans l'est, deux groupes diploïdes ont été trouvés : l'un comprend *R. blanda* et *R. woodsii* et l'autre *R. foliolosa*, *R. nitida* et *R. palustris*. Seules les espèces est-américaines ont été impliquées dans l'évolution des polyploïdes. *Rosa arkansana* a évolué à partir du groupe *blanda-woodsii*, *R. virginiana* du groupe *foliolosa-nitida-palustris*, et *R. carolina* d'un hybride entre ces deux groupes. En conclusion, les origines distinctes de ces polyploïdes appuient l'hypothèse qu'elles représentent des espèces distinctes.

---

[3] Joly, Simon, Julian R. Starr, Walter H. Lewis et Anne Bruneau. 2006. Amer. J. Bot. 93(3):412-425.

5.2 Abstract

This study investigates the impact of hybridization and polyploidy in the evolution of eastern North American roses. We explore these processes in the *Rosa carolina* complex (section *Cinnamomeae*), which consists of five diploid and three tetraploid species. To clarify the status and origin of polyploids, a haplotype network (statistical parsimony) of the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) nuclear gene was estimated for polyploids and diploids of section *Cinnamomeae* in North America. A genealogical approach helped to decipher the evolutionary history of polyploids despite the noise created by hybridization, incomplete lineage sorting, and allelic segregation. At the diploid level, species west of the Rocky Mountains are distinct from eastern species. In the east, two groups of diploids were found: one consists of *R. blanda* and *R. woodsii*, and the other of *R. foliolosa*, *R. nitida* and *R. palustris*. Only eastern diploids are involved in the origins of the polyploids. *Rosa arkansana* is derived from the *blanda–woodsii* group, *R. virginiana* from the *foliolosa–nitida–palustris* group, and *R. carolina* from a hybrid between the two diploid groups. The distinct origins of these polyploid taxa support the hypothesis that they are separate species.

5.3 Introduction

Wild species of roses are characterized by extensive morphological variation, which has resulted in a notoriously complex taxonomy. For instance, Linnaeus wrote in *Species Plantarum*: "the species of *Rosa* are with difficulty to be distinguished, with even greater difficulty to be defined; nature seems to me to have blended several or by way of sport to have formed several from one" (Stearn, 1957:158). North American roses are no exception; Crépin (1896), Watson (1885), Rydberg (1920), and Erlanson MacFarlane (1966) described 13, 18, 129, and 22 *Rosa* species on this continent, respectively. Hybridization has long been considered to be one of the major causes of taxonomic confusion

(Linnaeus, 1753; Crépin, 1894, 1896) and artificial crosses have shown that in fact most diploids are interfertile (Erlanson, 1934; Ratsek *et al.*, 1939; 1940; Lewis and Basye, 1961). Cytological studies in the early 20th century demonstrated that polyploidy is frequent in *Rosa* (Täckholm, 1922; Hurst, 1925) and that it could represent another source of variation. The present research explores issues related to hybridization and polyploidy, two important processes in plant evolution (Arnold, 1997; Otto and Whitton, 2000), that may explain the difficulty in recognizing species in wild roses.

This study focuses on the North American *Rosa carolina* L. complex of section *Cinnamomeae*, a group that epitomizes the complexity of the genus. Indeed, Lewis (1957c:126) considered the group to be "[…] the most difficult taxonomic problem in our North American *Rosa*". The complex consists of five diploid and three tetraploid species, almost entirely located east of the Rocky Mountains. The diploids *R. blanda* Ait., *R. foliolosa* Nutt., *R. nitida* Wild., *R. palustris* Marsh., and *R. woodsii* Lindl. (the sole species of the complex also found west of the Rocky Mountains) are relatively well circumscribed (Lewis, 1957c; Erlanson MacFarlane, 1966), but natural interspecific hybrids have been reported (Erlanson, 1929, 1934; Lewis, 1962) and some have been given species status (Rydberg, 1920; Erlanson, 1934) (see Fig. 1.1 for the distribution of the diploid species). In contrast, the tetraploid taxa *R. arkansana* Porter, *R. carolina* L. and *R. virginiana* Mill. are characterized by extensive continuous morphological variation that blurs their limits with each other and with their putative diploid ancestors in the *R. carolina* complex (Erlanson, 1934; Lewis, 1957b). Despite the important biosystematic investigations involving cytology and morphology in this complex (Erlanson, 1929, 1934; Lewis, 1957b), the limits and origins of the polyploid taxa are still unclear. The broad polymorphism of polyploid species may be caused by hybridization given that it frequently has been reported in areas of contact between *R. carolina* and *R. arkansana* in the west (e.g., R. × rudiuscula Greene: Lewis, 1957b; Erlanson MacFarlane, 1966; A. Fishbein and W. H. Lewis, Washington University, unpublished manuscript; Fig. 5.1) and between *R. carolina* and *R. virginiana* in the east (Fernald, 1922; Lewis, 1957b; Fig. 5.1). Yet, it is also possible that these taxa represent a single polymorphic species rather than three distinct taxa.

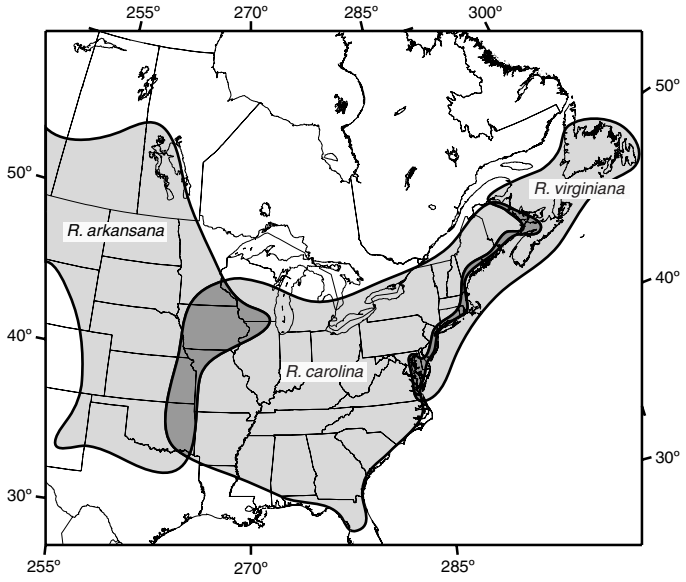


Figure 5.1 Approximate distributions of the polyploid taxa *Rosa arkansana*, *R. carolina* and *R. virginiana*. Areas where species overlap are in dark grey. The distributions are based upon Lewis (1957b) and personal collections.

Therefore, reconstructing the origins of the polyploids is a logical first step toward a global understanding of the *R. carolina* complex as it could be relevant to solving the species status of the polyploids if these are shown to have evolved independently.

Several factors can impair our ability to determine from which species polyploids evolved and whether they have evolved by autopolyploidy (from a single species) or by allopolyploidy (from more than one species – Grant, 1981; Ramsey and Schemske, 1998). For instance, introgression of foreign alleles into an autopolyploid can hide its real origin by making it look like an allopolyploid. Other problems can result from irregularities in chromosome segregation. Allopolyploids are expected to have disomic segregation where chromosomes only pair with their homoeologues (bivalent formation) (Stebbins, 1950, 1971; Levin, 2002), thus guaranteeing the preservation of homoeologous loci inherited from the parental species. However, these predictions are not always met and allopolyploids may have occasional polysomic segregation via multivalent formation. This could lead to the fixation of alleles from a single parental species in the genome of

the allopolyploid and hide its reticulate origin. The challenge when investigating polyploid evolution is thus to extract the true signal from the noise created by these confounding events in order to adequately reconstruct the evolutionary history of polyploids.

Investigation of polyploid origin must be done within a sound phylogenetic framework. To date, phylogenetic studies of *Rosa* have not included a good sampling of North American roses (e.g., Millan *et al.*, 1996; Matsumoto *et al.*, 1998), leaving their relationships obscure. Reconstruction of the diploid relationships could be further complicated by the recent origin of the complex, which is suggested by the low variation of ribosomal (Ritz *et al.*, 2005) and chloroplast markers (Wissemann and Ritz, 2005). Recent origin of species may result in incomplete lineage sorting of several molecular markers for the diploids (Pamilo and Nei, 1988; Rosenberg, 2002, 2003), which in turn could hamper our ability to accurately identify the species that were involved in the origins of polyploids. These potential problems need to be addressed prior to investigating polyploid evolution.

A genealogical approach using a single-copy nuclear gene is used to address the relationship of diploids and to investigate the origins of the polyploids. A genealogical approach has major advantages over a genotyping method (i.e., microsatellites, ALFPs, isozymes, etc.) because it places the data in a historical perspective: it relates who is ancestral to whom rather than who is similar to whom. This is particularly important in order to discern some of the confounding events mentioned above from our principal goal – reconstructing polyploid evolution. The use of nuclear genes is particularly useful in this regard because non-haploid organisms (except for clonal and apomict taxa) receive one chromosome copy from each parent. Thus, nuclear genes can retain information about the reticulate history of organisms, which is impossible for maternally or paternally transmitted markers. Such an approach has been successful in reconstructing the polyploid origins of other taxa (Doyle *et al.*, 2002; Senchina *et al.*, 2003;

Smedmark *et al.*, 2003; Helfgott and Mason-Gamer, 2004; Joly and Bruneau, 2004; Mason-Gamer, 2004; Petersen and Seberg, 2004; Evans *et al.*, 2005).

5.4 Material and methods

5.4.1 Sampling

Because it was more important to assess the extent of genetic variation within species rather than within populations, a single individual per population was investigated. Populations were sampled to represent the geographical range of each species of the complex (Table 5.1). Diploid roses of section *Cinnamomeae* west of the Rocky Mountains, *R. gymnocarpa* Nutt. and *R. pisocarpa* Gray, were included because they could be involved in the origins of the eastern polyploids. Diploid roses of section *Synstylae* found in North America, *R. setigera* (native to North America) and *R. multiflora* (introduced from China and now a noxious invasive in eastern North America (Meiners *et al.*, 2001; Hunter and Mattice, 2002)), were included as outgroup taxa. Only one species of *Rosa* section *Cinnamomeae* occurring east of the Rocky Mountains was not investigated here, *R. acicularis* Lindl., a circumboreal species that has both hexaploid and octoploid populations (Lewis, 1959). Investigation of its origin would require a broader taxonomic sampling at the diploid level, which is beyond the scope of the present study.

5.4.2 Ploidy level determination

Lewis (1957b) showed that the length of the stomatal guard cells can discriminate diploid and tetraploid roses of the complex. Twenty-five guard cells per individual were measured for all specimens of eastern species for which we had material. Nail polish was used to fingerprint the abaxial surface of one dried terminal leaflet. The length of stomatal guard cells was measured with a Leitz (type: 307-107.002) microscope using a

Table 5.1 Accessions included in this study of eastern North American roses. For each accession, voucher information, locality data, the number of alleles at the *GAPDH* locus that were found and the number of clones that were sequenced are indicated. The stomata guard cell length is given for eastern species.

Species	Accession	Collector(s)	Province/ State[a]	Lat. – Long.	Clones	Nb. alleles	Stomata guard cell length (μm)
R. arkansana	345	Joly and Starr 601 (MT)	Iowa	43°00'22.7"N, 89°58'44.6"W	12	4	27.72
R. arkansana	406	Joly and Starr 663 (MT)	Minn.	43°43'37.1"N, 95°03'53.1"W	12	2	23.35
R. arkansana	416	Joly and Starr 673 (MT)	Minn.	46°33'27.5"N, 96°13'07.9"W	12	4	22.40
R. arkansana	470	Joly and Starr 730 (MT)	Man.	50°04'55.7"N, 97°08'17.1"W	13	4	23.65
R. arkansana	503	Joly and Starr 763 (MT)	N. Dak.	47°58'09.9"N, 97°46'35.3"W	19	4	23.16
R. arkansana	665	Lewis 15837-1 (MO)	Kansas	38°32'N, 94°55'W	11	3	-
R. arkansana	692	Lewis 15792-2 (MO)	Kansas	-	14	4	23.10
R. arkansana	848	Ryan 3 (MT)	Sask.	52°05.479N, 106°41.857W	11	4	26.78
R. blanda	160	Joly and Starr 409 (MT)	N.B.	45°57'43.7"N, 67°22'26.1"W	3	2	14.65
R. blanda	326	Joly and Starr 582 (MT)	Ont.	42°15'29.7"N, 83°02'58.8"W	3	2	16.15
R. blanda	365	Joly and Starr 622 (MT)	Wis.	42°39'07.5"N, 89°43'32.4"W	4	2	15.99
R. blanda	421	Joly and Starr 678 (MT)	Minn.	48°06'36.3"N, 96°09'16.0"W	4	2	14.71
R. blanda	462	Joly and Starr 722 (MT)	Man.	50°00'59.3"N, 96°55'35.2"W	4	2	17.54
R. blanda	528	Joly and Starr 788 (MT)	Ont.	46°28'15.4"N, 80°29'27.2"W	-	1	16.96
R. blanda	567	Joly 921 (MT)	N.Y.	-	4	2	-
R. blanda	621	Joly *et al.* 962 (MT)	N.B.	47°22'32.5"N, 66°04'30.5"W	4	2	15.63
R. blanda	652	Joly *et al.* 993 (MT)	Que.	48°02'58.8"N, 65°28'43.6"	4	2	14.87
R. blanda	1214	Bruneau *et al.* 1214 (MT)	Que.	45°31'N, 73°50'W [b]	-	1	16.60
R. blanda	1219	Bruneau *et al.* 1219 (MT)	Que.	45°30'N, 73°50'W [b]	-	1	16.14
R. blanda	1236	Bruneau *et al.* 1236 (MT)	Que.	48°21'N, 68°45'W [b]	4	2	16.26
R. blanda	1239	Bruneau *et al.* 1239 (MT)	Que.	48°21'N, 68°45'W [b]	4	2	17.02
R. blanda	98016	Drouin 98-016 (MT)	Que.	47°26'N, 70°30'W [b]	-	2	15.44
R. carolina	268	Joly and Starr 523 (MT)	Va.	38°21'29.8"N, 79°04'54.1"W	13	4	27.54
R. carolina	289	Joly and Starr 545 (MT)	W.Va.	38°41'N, 80°00'W [b]	10	4	22.07
R. carolina	320	Joly and Starr 576 (MT)	Ont.	42°15'29.7"N, 83°02'58.8"W	13	4	27.97
R. carolina	395	Joly and Starr 651 (MT)	Minn.	43°48'03.6"N, 92°29'21.6"W	13	4	23.16
R. carolina	553	Lewis 15783-3 (MO)	Mo.	38°31'23"N, 90°40'36"W	14	4	22.80

Species	Accession	Collector(s)	Province/ State[a]	Lat. – Long.	Clones	Nb. alleles	Stomata guard cell length (μm)
R. carolina	576	Joly 906 (MT)	N.Y.	-	12	4	-
R. carolina	626	Joly *et al.* 967 (MT)	N.B.	47°22'36.2''N, 66°04'42.2''W	11	2	19.30
R. carolina	671	Lewis 15844 (MO)	Okla.	36°92'N, 94°88'W	10	4	24.47
R. foliolosa	699	Lewis 15846-3 (MO)	Okla.	34°24'N, 96°00'W	3	2	19.18
R. foliolosa	795	O'Kennon and McLemore 19069A (MT)	Tex.	33°24'32.2''N, 97°30'22.0''W	-	2	17.02
R. gymnocarpa	543	Ertter 18001 (JEPS)	Idaho	-	-	1	-
R. gymnocarpa	751	Lewis 15852-1 (MO)	B.C.	49°02'N, 118°13'W	3	2	-
R. gymnocarpa	767	Ertter 18293a (JEPS)	Idaho	-	-	1	-
R. multiflora	302	Joly and Starr 558 (MT)	Pa.	42°08'48.4''N, 80°08'00.1''W	4	2	-
R. nitida	570	Meilleur s.n. (MT)	Que.	-	4	2	14.68
R. nitida	604	Joly *et al.* 941 (MT)	N.B.	45°56'29.2''N, 64°52'07.3''W	3	2	17.97
R. nitida	675	Brouillet 03-55-1 (MT)	Nfld.	-	-	2	19.09
R. nitida	812	Joly 1010-1 (MT)	Que.	46°22'45.3''N, 75°00'20.6''W	4	2	17.05
R. palustris	168	Joly and Starr 417 (MT)	N.B.	45°33'43.2''N, 67°25'31.2''W	4	2	15.63
R. palustris	255	Joly and Starr 510 (MT)	N.J.	38°56'02.8''N, 74°57'29.5''W	4	2	15.96
R. palustris	304	Joly and Starr 560 (MT)	Pa.	42°09'32.9''N, 80°07'10.7''W	4	2	16.81
R. palustris	317	Joly and Starr 573 (MT)	Ont.	42°19'41.0''N, 82°18'49.0''W	4	2	17.24
R. palustris	331	Joly and Starr 587 (MT)	Mich.	42°19'32.0''N, 84°29'51.2''W	-	1	16.14
R. palustris	386	Joly and Starr 644 (MT)	Wis.	44°01'30.6''N, 89°43'13.1''W	-	1	17.12
R. palustris	581	Joly 912 (MT)	N.Y.	-	-	1	-
R. pisocarpa	774	Ertter 18303a (JEPS)	Calif.	-	4	2	-
R. pisocarpa	784	Ertter 18305c (JEPS)	Oreg.	42°05.7'N, 123°41.0'W	4	2	-
R. pisocarpa	847	Ertter 18428 (JEPS)	Calif.	41°09.2'N, 123°49.2'W	4	2	-
R. setigera	298	Joly and Starr 554 (MT)	Pa.	42°08'48.4''N, 80°08'00.1''W	-	1	-
R. virginiana	182	Joly and Starr 431 (MT)	N.B.	45°05'00.4''N, 67°03'01.1''W	14	3	21.92
R. virginiana	195	Joly and Starr 444 (MT)	Maine	44°30'56.7''N, 68°11'14.6''W	10	4	21.83
R. virginiana	225	Joly and Starr 474 (MT)	Conn.	41°20'43.0''N, 71°54'14.2''W	13	3	23.29
R. virginiana	246	Joly and Starr 496 (MT)	N.J.	38°55'57.4''N, 74°57'28.5''W	12	4	20.16
R. virginiana	262	Joly and Starr 517 (MT)	Md.	38°14'08.2"N, 75°08'15.7''W	13	4	21.49

Species	Accession	Collector(s)	Province/ State[a]	Lat. – Long.	Clones	Nb. alleles	Stomata guard cell length (μm)
R. virginiana	587	Joly and Edelist 924 (MT)	N.S.	45°43'09.7''N, 61°53'56.3''W	11	4	23.44
R. virginiana	656	Joly *et al.* 997 (MT)	Que.	48°02'58.8''N, 65°28'43.6''W	10	3	21.16
R. virginiana	684	Brouillet 03-60-1 (MT)	Nfld.	-	14	3	23.68
R. woodsii	4	Spellenberg 12555 (MT)	N.Mex	-	-	1	16.54
R. woodsii	492	Joly and Starr 752 (MT)	Sask.	49°12'35.3''N, 101°50'46.1''W	-	1	15.20
R. woodsii	498	Joly and Starr 758 (MT)	N.Dak.	48°21'09.6''N, 99°47'07.5''W	-	2	18.00
R. woodsii	700	Saarela 266-1 (MT)	Alta.	-	-	2	-
R. woodsii	733	Dickson 2017 (MT)	Alta.	-	-	2	-
R. woodsii	741	Lewis 15848-1 (MO)	B.C.	49°45'N, 120°50'W	3	2	-
R. woodsii	800	Joly 1005-1 (MT)	Colo.	40°12'23.4''N, 104°49'54.0''W	-	1	14.59
R. woodsii	807	Joly 1008-1 (MT)	Colo.	40°38'36.8''N, 104°20'32.0''W	-	1	15.72

Notes: a – Abbreviations follow the nomenclature of Flora of North America (Flora of North America Editorial Committee, 1993); b – Approximate coordinates not determined by GPS

63× objective. A K-mean analysis for two groups was performed in R (R Development Core Team, 2005) to see whether diploids and polyploids could be differentiated without previous knowledge of the ploidy level of individuals.

5.4.3 Molecular methods

DNA was extracted using a modified version of the CTAB extraction of (Doyle and Doyle, 1987). Modifications involved scaling the protocol for a total CTAB volume of 600 µL; adding 12 µL of 0.5M EDTA pH 8.0 per 600µL of CTAB and 1% of polyvinylpyrrolidone (PVP) to the extraction buffer prior to extraction; adding 20 µg of RNAse A to the CTAB buffer prior to incubation at 65°C; performing two chloroform-isoamyl alcohol (24:1) extractions and precipitating the DNA with 1.5 volumes of 100% ethanol.

5.4.3.1 Gene selection

North American roses are particularly uniform at the DNA level. For example, sequences of the internal transcribed spacer of the 18S-5.8S-26S ribosomal gene family showed few variations among North American rose species sampled by Ritz *et al.* (2005), even if this marker is generally considered to be highly variable in many plant taxa (Baldwin *et al.*, 1995). Similarly, only five variable characters were found between *R. woodsii*, *R. blanda*, and *R. palustris* among 4318 bp from seven chloroplast gene spacers or introns (unpublished data). Because of this, introns of single-copy nuclear genes became the alternative for providing sufficient variation. Initial screening (data not shown) of several nuclear genes, *LEAFY* (Frohlich and Meyerowitz, 1997; Archambault and Bruneau, 2004), *GBSSI* (Evans *et al.*, 2000), *rpb*2 (Denton *et al.*, 1998; Pfeil *et al.*, 2004), and *GAPDH* (Strand *et al.*, 1997; Olsen and Schaal, 1999), identified *GAPDH* as the most variable region.

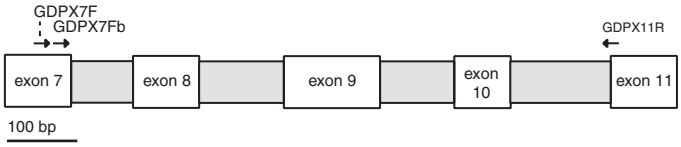


Figure 5.2 The glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) locus in North American *Rosa*. Primers are not to scale and their positions are approximate. The first exon (7) is numbered according to *Arabidopsis thaliana*, but the amplified region between primers GDPX7F and GDPX11R contains two introns not present in *A. thaliana*.

5.4.3.2 Gene amplification

The cytosolic glyceraldehyde 3-phosphaste dehydrogenase (*GAPDH*) gene was amplified from the end of exon seven (according to the *Arabidopsis thaliana* sequence; GenBank locus tag: At3g04120) to the beginning of exon eleven (which is exon 9 in *A. thaliana*; Fig. 5.2). The 5′ end of the forward primer GPDX7F (5′-GATAGATTTGGAATTG TTGAGG-3′) (Strand *et al.*, 1997) starts 52 bp upstream of the intron in the seventh exon, whereas the GPDX11R primer (5′-GACattgaatgagataaacc-3′; lower-case letters represent intron nucleotides) spans the junction between exon eleven and the previous intron. Polymerase chain reactions (PCR) in final volumes of 50 µl contained 1× PCR reaction buffer (Roche Diagnostics, Laval, Québec; for a total $MgCl_2$ concentration of 1.5 mM), 0.05% of Tween 20, 5 µg of BSA, 1 mM of each primer, 200 µM of each dNTP, two units of Taq polymerase, and ca. 300 ng of genomic DNA. PCR conditions included an initial denaturation step of 3 min at 94 °C, followed by 40 cycles of denaturation (30 sec at 95 °C), annealing (30 sec at 48 °C), and elongation (2 min at 72 °C), with a final extension step of 10 min at 72 °C. A long elongation time was used and reactions were performed in triplicate to reduce the potential for PCR recombinants (Judo *et al.*, 1998; Cronn *et al.*, 2002). The triplicate reactions also reduced the possibility of finding the same *Taq*-induced mutation in many different clones. PCR products were purified with polyethylene glycol (PEG; M.W. 8000) according to the following procedure. PCR reactions were mixed with an equal volume of PEG solution (20% PEG, 2.5M NaCl), incubated 15 min at 37 °C, and centrifuged 15 min at 12,000 ×g. The supernatant was

removed and the pellet was washed twice with 80% ethanol (spinning five minutes at 12,000 ×g before ethanol removal). The pellet was dried 2 min in a vacuum centrifuge (no heat) and was resuspended in $TE_{0.1}$ (20 mM Tris-HCl, 0.1 mM EDTA, pH 8.0).

Sequencing was performed using the reverse primer GPDX11R and the forward primer GPDX7Fb (5′-CTTATGACTACCGTGCACTC-3′; Fig. 2). The 5′ end of GPDX7Fb is located 28 bp upstream of the intron in exon seven. Sequencing reactions were performed with BigDye terminator chemistry (ver.1.1; Applied Biosystems, Foster City, CA) following the manufacturer's protocols and were run on a 3100-*avant* automated sequencer (Applied Biosystems). Sequences were assembled and edited in Sequencher (ver. 4.1; GeneCodes, Inc., Ann Arbor, MI).

5.4.3.3 Allele sampling

In order to derive firm conclusions on the origin of polyploids, it is important to sample all alleles in each individual. The approach used to achieve this objective differed for diploids and polyploids. Diploids that did not show polymorphic nucleotides in direct sequencing (from the total PCR reaction) were assumed to be homozygous and were not cloned. Such an assumption is valid because two equally frequent templates should be equally visible on chromatograms if there is no strong PCR bias in the reactions (Rauscher *et al.*, 2002). When a single polymorphic nucleotide was found for an individual, no cloning was necessary because the alleles can easily be distinguished. In contrast, individuals that showed more than one polymorphic site or that had indels among its alleles were cloned. In these cases, three to four clones were sequenced to retrieve allelic sequences. More than one clone was sequenced to eliminate the possibility of sampling a PCR recombinant with a single clone.

All tetraploids were cloned because it is easier to miss polymorphic sites on direct sequences when four alleles may be present in the genome. Assuming no PCR bias

between alleles (Wagner *et al.*, 1994), the binomial distribution predicts that the probability of sampling all alleles in an individual is:

$$P = \left[1 - \left(\frac{t-1}{t}\right)^{n}\right]^{t}$$

where t is the number of alleles in the individual and n is the number of clones sequenced. If there were four alleles in a tetraploid, 15 clones would be required in order to obtain a 95 % probability that all alleles have been sampled. With three alleles, 11 clones are needed. On average, 11 to 15 clones were sequenced per individual (Table 5.1), with additional clones sequenced in all cases where the alleles resulting in polymorphisms detected in direct sequencing were not recovered.

For both diploids and tetraploids, *Taq*-induced PCR errors were identified and removed from analyses by comparing the sequence of cloned amplicons to one another and to the initial sequences obtained from direct sequencing. Henceforth, it will be assumed that all alleles were retrieved from each individual even if there is a non-zero probability that some alleles were not sampled in some individuals. PCR products were cloned with the TOPO TA cloning kit (Invitrogen, Burlington, Ont.). Plasmids containing the gene were extracted from *E. coli* using the QIAprep miniprep kit (Qiagen, Mississauga, Ont.) and were sequenced as described above. Alleles from both diploids and tetraploids were aligned with ClustalX (Thompson *et al.*, 1994; 1997a) with a gap opening penalty of 25 and a gap extension penalty of 6. The resulting alignment did not need further manual corrections.

5.4.4 Testing recombination

Two different methods were used to detect recombination: the homoplasy test (Maynard Smith and Smith, 1998), which works best when divergence between sequences is low (less than 5%; Maynard Smith and Smith, 1998; Posada and Crandall, 2001), and a parsimony network approach (Templeton *et al.*, 1992). The homoplasy test

was performed using Maynard Smith's Qbasic program under conservative (S_E = 0.6S) and liberal (S_E = S) conditions, where S_E is the effective number of sites and S is the total number of sites in the dataset. First and second codon positions in exons were removed from the analysis because they are evolutionary constrained (Maynard Smith and Smith, 1998) and the analysis was performed only on ingroup taxa. With the parsimony network approach, recombination was inferred only when it could explain at least two homoplasies and when the homoplasies corresponding to the parental alleles were physically clustered on the recombinant allele (Aquadro *et al.*, 1986; Templeton *et al.*, 1992).

5.4.5 Network construction

GapCoder (Young and Healy, 2003) was used to code indels under the simple gap coding method of Simmons and Ochoterena (2000). The resulting matrix was used to estimate the gene genealogy of the *GAPDH* locus by statistical parsimony (Templeton *et al.*, 1992) as implemented in the TCS software (ver. 1.18; Clement *et al.*, 2000). The statistical limit of parsimony was evaluated on the matrix with the gaps recoded (although estimating it without the gaps gave the same result) and the final network was constructed so that all the haplotypes could be united in a single network.

5.4.6 Statistical distinction of diploid species

Diploid species boundaries were tested by permutations using an analysis of molecular variance (AMOVA; Excoffier *et al.*, 1992). An uncorrected P distance matrix among haplotypes was calculated in PAUP* (ver. 4.10b, Swofford, 2002), and the partitioning of haplotype variance in different groups (species) was tested in Arlequin ver. 2 (10,000 permutations; Schneider *et al.*, 2000).

5.4.7 Origins of the polyploids

To reconstruct the evolutionary history of the polyploid taxa, the closest diploid haplotype for each allele of each polyploid individual was identified to determine which diploid species contributed to polyploids. Because alleles can mutate in polyploids, simply counting the number of haplotypes in a polyploid species will overestimate the number of origins (Doyle *et al.*, 2004). A conservative way of evaluating the likelihood that the polyploid species evolved recurrently is to estimate the number of "polyploid haplotype groups" that comprise all polyploid haplotypes that have a most recent common diploid haplotype (or expected diploid haplotype) ancestor (Fig. 5.3; see also Doyle *et al.*, 2004). At formation, a tetraploid can acquire up to four different alleles from diploids. Independent polyploid origins can involve one or more identical diploid alleles, yet it is impossible to detect this if there is segregation in polyploid populations. To be conservative, it was therefore assumed that for one polyploid species, a polyploid haplotype group can only be involved in one origin and that each origin always involved four polyploid haplotype groups. So if there are n polyploid haplotype groups in one polyploid species (n = 4 in Fig. 5.3), there needs to be at least $n/4$ (rounded to the upper unit) distinct origins to account for this variability (one distinct origin in the simplified example given in Fig. 5.3).

5.5 Results

5.5.1 Sequences and alleles

The number of alleles found and the number of clones sequenced for each individual is indicated in Table 5.1. The phylogenetic analysis used the portion of the *GAPDH* gene that starts immediately after exon seven and stops at the GPDX11R primer, 17 bp downstream of exon eleven. The length of this aligned region is 759 bp and includes fifteen indels. Multiple alleles in an individual were distinguished by a letter

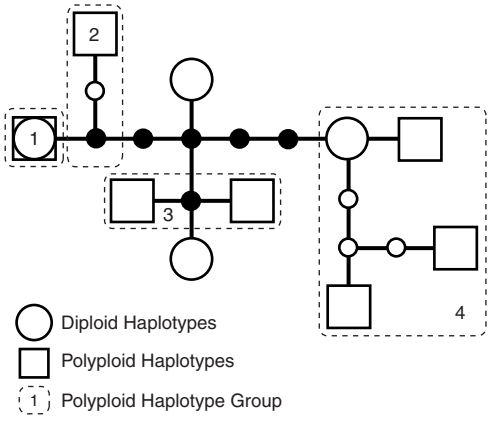


Figure 5.3 Network illustrating how to evaluate the minimum number of haplotypes contributed by diploids to polyploids when estimating the number of independent origins of a polyploid species. Large white circles (diploids) and squares (polyploids) represent sampled haplotypes. Small black circle represent unsampled interior haplotypes inferred to have occurred in diploids, whereas small white circles represent interior unsampled haplotypes that cannot be inferred to have occurred in the diploids. Polyploid haplotype groups (broken lines) are comprised of polyploid haplotypes that have a most recent common diploid or expected diploid haplotype ancestor.

(i.e., A, B, etc.) following the species name and accession number (GenBank accession numbers DQ091014-DQ091057, DQ091060-DQ091174).

Of all alleles recovered, one was obviously a pseudogene: the *carolina*289.A allele. This allele has a deletion of 1 bp in exon 10 that causes a frame shift and introduces a stop codon. Because the indel was visible in the direct sequences, and therefore present in relatively high proportions in the PCR products (Rauscher *et al.*, 2002), and because the reactions were performed in triplicate, it is unlikely that this mutation is the result of a PCR error.

5.5.2 Length of stomatal guard cells

Based on the taxonomic identifications, diploids and polyploids had disjoint distributions for their mean stomatal guard cell length (Fig. 5.4) and the difference

between the two groups is statistically significant (two-way student *t* test: $\nu = 50$, $t = -4.061$, $p < 0.001$; homoscedasticity hypothesis accepted: Levene $F = 3.949$, $p = 0.53$). The mean lengths of diploids and polyploids were under 19.18 μm and over 19.30 μm, respectively (Fig. 5.4, Table 5.1). The gap is more important when making abstraction of the *carolina*626 individual, without which all polyploids would have a mean length over 20.16 μm.

The mean lengths of the two clusters recovered by a K-mean analysis were 16.60 μm and 23.75 μm. Only two assignments (out of 52) disagreed with taxonomic identifications: individuals *carolina*626 (19.30 μm) and *virginiana*246 (20.16 μm) fell in the shorter cluster otherwise constituted of only diploid species. To confirm the ploidy level of these individuals, it is helpful to consider the number of alleles found. For example, *virginiana*246 has four alleles (Table 5.1), which is strong evidence of polyploidy, and it will hereafter be treated as a polyploid. In contrast, *carolina*626 only has two alleles, which is inconclusive as to its ploidy level. This latter individual will be treated as a polyploid based on its morphology and on its stomatal guard cell length that is longer than that of any diploid (Table 5.1).

The stomatal cell lengths reported are about 1.3 times smaller than those obtained by Lewis (1957b; 1958; 1959) for both diploids and polyploids. These discrepancies are caused by differences in methodology.

5.5.3 Network

One of the premises of tree-like phylogenetic methods is that all characters have the same evolutionary history. Recombination can violate this assumption for nuclear loci and it is important to test for its presence when using such markers. The homoplasy test was significant under both the conservative and liberal conditions ($p < 0.001$), suggesting that recombination is present in the dataset. In contrast, no clear recombinants were

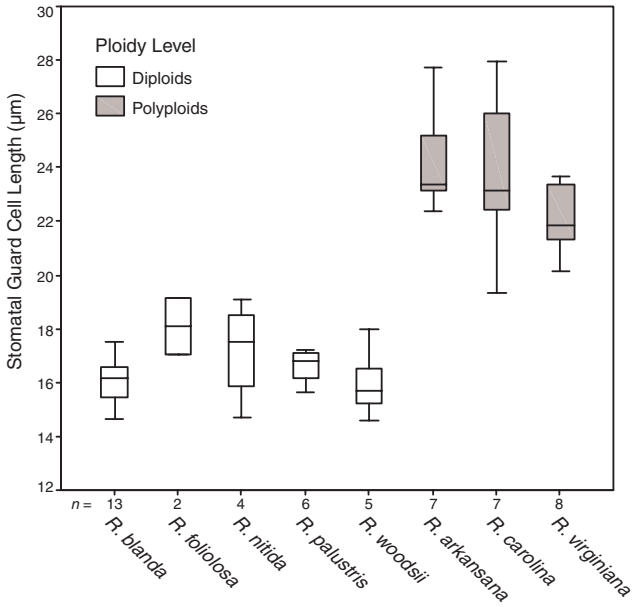


Figure 5.4 Boxplots showing length variation of the stomatal guard cells for each species, contrasting results from diploids and polyploids. The boxplots were constructed using the mean length per individual and the number of individuals assessed per species is indicated (*n*). The mean length of stomatal guard cells for each individual can be found in Table 5.1.

detected using the network approach. Even within the loops, there was always one alternative that required only one homoplasy. The discrepancy between these results could be due to the presence of homoplasious sites in the dataset: a standard parsimony analysis gave a consistency index of 0.83. Even if allelic variation ranges from 0 to 3.4% of variation among ingroup taxa, this level of homoplasy may be high enough to violate the homoplasy test's assumption of low levels of variation, which could bias the test towards a conclusion for recombination. Such behavior of the homoplasy test has previously been reported (Posada and Crandall, 2001; Posada, 2002). Because no clear recombination events were identified on the network, the evidence for recombination in the data is equivocal at best and the dataset was analyzed as if there were no recombination.

Haplotypes with a distance of more than 12 steps (parsimony limit) from all other haplotypes were not statistically supported and their relationship to the rest of the haplotypes should be viewed as if estimated by standard parsimony procedures (Fig. 5.5). However, only section *Synstylae* was not connected to the rest of the network with

this limit; the two sub-networks were 13 steps away. Henceforth, haplotypes will be referred to by the number of the box in which they occur on the network and by their specific letter (e.g., I-a represents the haplotype of allele *multiflora*302.A of section *Synstylae* ; Fig. 5.5).

5.5.4 Diploids

Relative to the outgroup species *R. multiflora* and *R. setigera* (section *Synstylae*), alleles of *R. gymnocarpa* are monophyletic. The other western species, *R. pisocarpa*, is either paraphyletic or polyphyletic depending upon how the loop involving the *R. pisocarpa* haplotypes is resolved on the network (Fig. 5.5). In the presence of ambiguity, one hypothesis can be favored over others because it has been shown that a loop is more likely to be broken beside the most recent haplotype of the loop (Crandall and Templeton, 1993). Coalescent theory predicts that an old haplotype is more frequent (Donnelly and Tavaré, 1986) and that more lineages are related to it (Crandall and Templeton, 1993; Castelloe and Templeton, 1994). These predictions can be used to determine "outgroup weights" that are correlated to the age of the haplotype (Castelloe and Templeton, 1994). According to these outgroup weights and the predictions of Crandall and Templeton (1993), the most likely hypothesis is the one that links the *R. pisocarpa* III-a haplotype to the other *R. pisocarpa* haplotypes (solid line on the network; the alternative solution is shown by a broken line; Fig. 5.5). Consequently, the paraphyletic option for *R. pisocarpa* is more likely than the polyphyletic one. This also suggests a division between western (boxes II, III) and eastern (IV, V, VI) diploid species of section *Cinnamomeae* (Fig. 5.5).

Regarding the diploid species east of the Rocky Mountains, two main groups can be distinguished on the network (Fig. 5.5). The first group includes all alleles of diploid species *R. blanda* and *R. woodsii* (the *blanda/woodsii* or BW group, box IV in Fig. 5.5), whereas the other contains most alleles of *R. foliolosa*, *R. nitida* and *R. palustris* (the *foliolosa/nitida/palustris* or FNP group, box V). These groups are not monophyletic, but

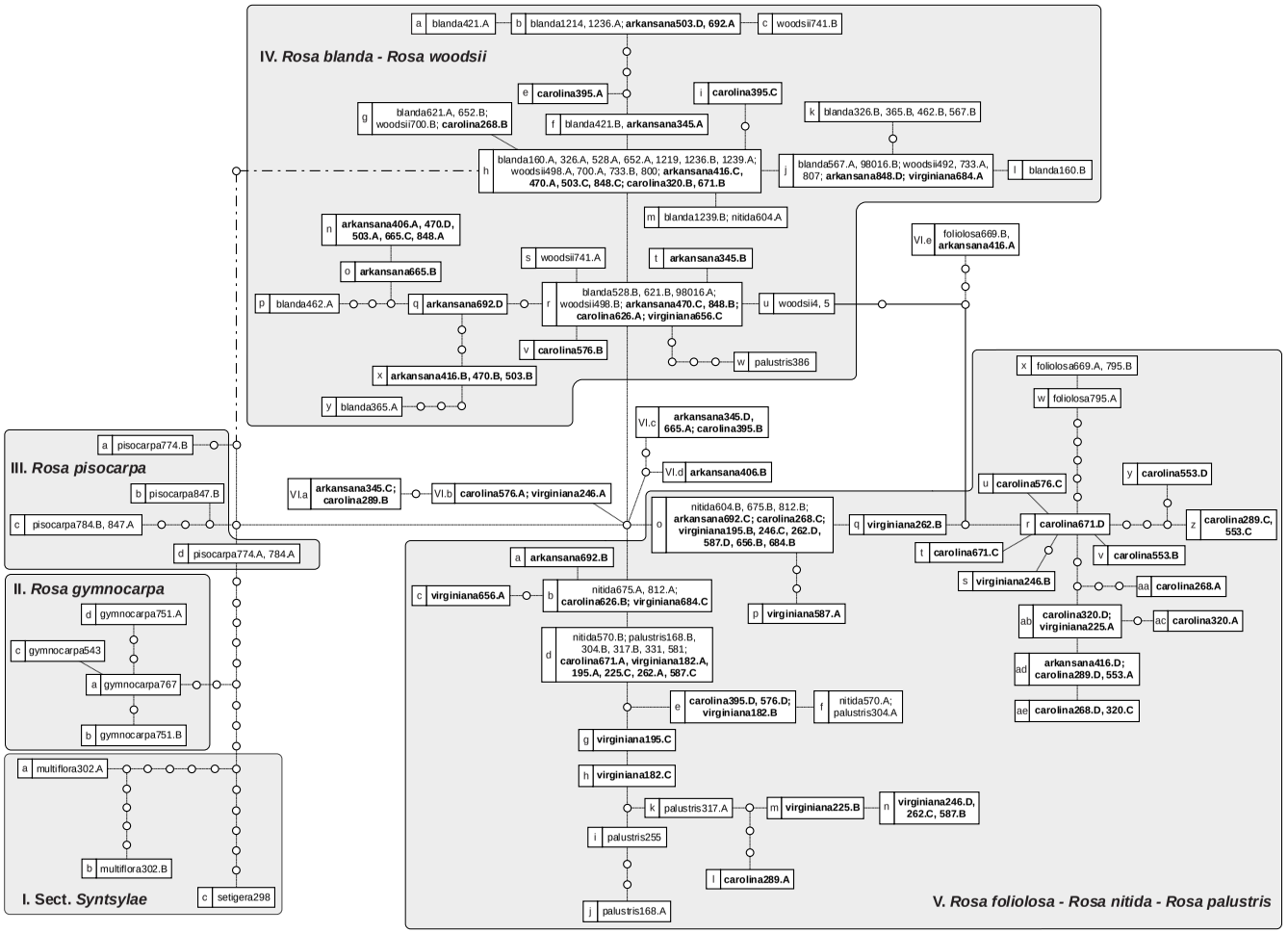


Figure 5.5 Network obtained by statistical parsimony analysis of the *GAPDH* gene. The limit of parsimony is of 12 steps. Individuals within species are identified by their specific epithet and accession number, and letters following accession numbers are used to differentiate multiple alleles within an individual. Diploids are in normal font, whereas polyploids are in bold. Each box represents a haplotype; small white circles between boxes represent unsampled (i.e., inferred) haplotypes. The grey boxes represent the principal diploid groups that are discussed in the text. The broken line indicates an alternative branching scenario that is less likely for *R. pisocarpa* (see text).

they are nevertheless almost exclusive. There are two exceptions: one allele each of *R. palustris* and *R. nitida* occur in the BW group. Even with these, the AMOVAs showed that the distinction between the BW and the FNP groups is significant ($p < 0.001$; Table 5.2). Neither the AMOVAs nor the network found a distinction between *R. blanda* and *R. woodsii*. Within the FNP group, AMOVAs suggest that *R. foliolosa* is significantly distinct from *R. nitida* and *R. palustris* ($p < 0.001$) and also that the differentiation between *R. nitida* and *R. palustris* is marginally significant ($p < 0.05$; Table 5.2). The network is ambiguous regarding these distinctions, however, and *R. nitida* and *R. palustris* do not clearly form distinct groups (Fig. 5.5). Moreover, only two individuals of *R. foliolosa* were investigated, limiting the significance of the distinction found with AMOVAs. In addition, the *R. foliolosa* alleles have *R. nitida* alleles as ancestors. Therefore, *R. foliolosa*, *R. nitida* and *R. palustris* are considered to form a single group in the following analyses.

5.5.5 Polyploids

Polyploid haplotypes on the network are exclusively related to eastern diploids. Most polyploid alleles can be clearly attributed to either the BW or the FNP diploid groups, and only a limited number of alleles have an ambiguous relationship (those that could not be placed in either group; VI, a-e in Fig. 5.5). All polyploid species have haplotypes that belong to both the FNP and the BW group (Fig. 5.5, 5.6), but not all individuals of each species have alleles from both diploid groups. All eight *R. arkansana* individuals studied have alleles that belong to the BW diploid group (Fig. 5.6a). Three of them have exclusively such alleles, two also have one allele that has an ambiguous relationship, and three have one allele from the FNP group. Six of eight *R. carolina* individuals have alleles that are from both the BW and FNP diploid groups (Fig. 5.6b). There are only two exceptions and one of these has an allele of unknown relationship. Finally, five individuals of *R. virginiana* have exclusively FNP-related alleles, two have haplotypes related to both eastern diploid groups, and one has haplotypes from the FNP group and of unknown origin (Fig. 5.6c).

Table 5.2 Partition of variance (AMOVAs) within and among different species or groups of species.

Groups tested	d.f.	% variance
R. blanda, *R. woodsii* vs. *R. nitida*, *R. palustris* vs. *R. foliolosa*		
Among groups	2	43.26***
Within group	57	56.74
R. blanda, *R. woodsii* vs. *R. nitida*, *R. palustris*, *R. foliolosa*		
Among groups	1	32.66***
Within group	58	67.34
R. blanda, *R. woodsii* vs. *R. nitida*, *R. palustris*		
Between groups	1	37.14***
Within group	54	62.86
R. blanda, *R. woodsii* vs. *R. foliolosa*		
Between groups	1	60.06***
Within group	40	39.94
R. nitida, *R. palustris* vs. *R. foliolosa*		
Between groups	1	49.54***
Within group	20	50.46
R. blanda vs. *R. woodsii*		
Between groups	1	-3.06^{ns}
Within group	36	103.06
R. nitida vs. *R. palustris*		
Between groups	1	11.45*
Within group	16	88.55

Note: *** = P(observed value ≥ random value) ≤ 0.0001; * = P(obs. ≥ rand.) ≤ 0.05; ns = not significant.

The number of polyploid haplotype groups was 13, 11, and 12 in *R. arkansana*, *R. carolina* and *R. virginiana*, respectively (see appendix 5, 6 and 7). This requires a minimum of three distinct polyploid origins to explain the observed genetic diversity in all three polyploid taxa.

5.6 Discussion

5.6.1 Diploid species boundaries

Three evolutionary processes can result in non-monophyletic species within a genealogical framework: hybridization, incomplete lineage sorting (or deep coalescence)

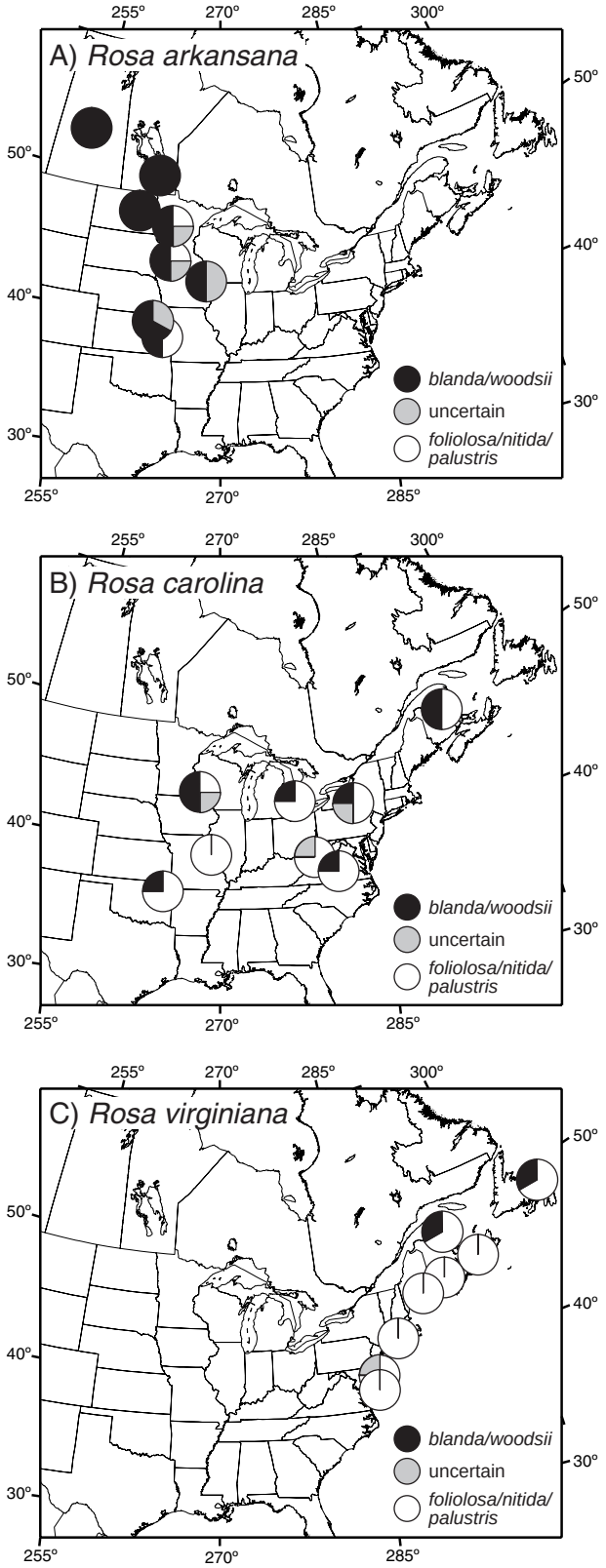


Figure 5.6 Genetic constitutions of individuals sampled for the tetraploids, *R. arkansana* (A), *R. carolina* (B) and *R. virginiana* (C). The genetic constitution of each individual is represented by a pie chart where the shades of grey represent the proportion of alleles from each of the diploid groups (Fig. 5.5). The total number of alleles for each accession is given in Table 5.1.

and gene duplication (Maddison, 1997; Funk and Omland, 2003). Among these processes, gene duplication is the least likely to be problematic at low phylogenetic levels. Because no evidence of gene duplication was found, this process will not be discussed further.

5.6.1.1 Incomplete lineage sorting and hybridization

Attempts have been made to distinguish between incongruence due to incomplete lineage sorting and that due to hybridization in gene trees (Sang and Zhong, 2000), but they mostly have been unfruitful (Holder *et al.*, 2001). However, in some circumstances it is possible to discriminate between the two processes by using the full amount of information contained in branch lengths (Holder *et al.*, 2001). Take the hypothetic example of a lineage that splits into two distinct species at time T_S, where one incongruent haplotype happens to be more closely related to the haplotypes of its sister species than it is to its own (Fig. 5.7). Note that the time of speciation is independent of the gene lineages and corresponds to the time when gene flow ceased among sibling species (Holder *et al.*, 2001). With incomplete lineage sorting, the most recent common ancestor of the incongruent haplotype and the haplotypes of the sister species must have have been present in the common lineage before the speciation event (Fig. 5.7a). Therefore, the time since the divergence of the incongruent allele and the alleles of the sister species (T_{LS}) must be at least as old as the time of divergence of the two species ($T_{LS} \geq T_S$). On a hypothetic genealogy, the incongruent allele should branch near the split between the two species relative to an outgroup taxon and should be quite divergent from the alleles of the sister species because it has evolved independently from the other sister species alleles for a time T_{LS} (Fig. 5.7b).

In contrast, the time of divergence between an incongruent haplotype caused by hybridization and haplotypes of its sister species (T_H) can be younger that the speciation event (e.g., Fig. 5.7c), which would result in an incongruent allele connected on the network far from the root and similar the contemporary alleles of its sister species (Fig. 5.7d). However, because the incongruent allele could also coalesce with alleles of the

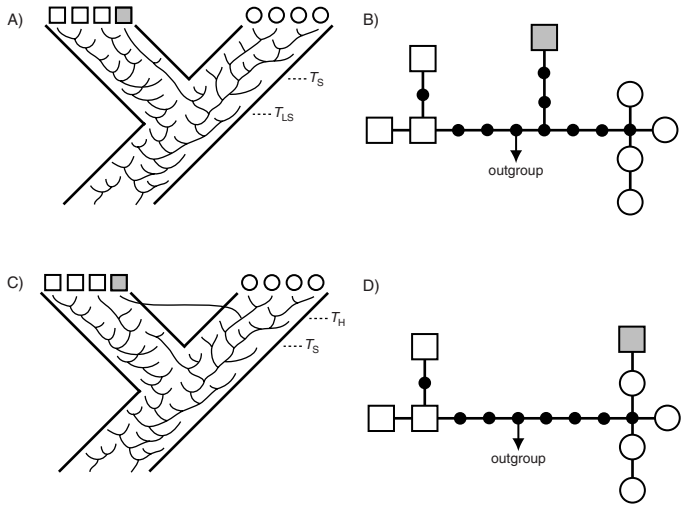


Figure 5.7 Hypothetical evolutionary networks that illustrate expected patterns of incongruence due to incomplete lineage sorting or hybridization. (A) Two hypothetical species between which there is incomplete lineage sorting and (B) the expected network for this scenario. (C) Hypothetical species between which there has been a hybridization event and (D) the expected network for this scenario. The grey box represents the incongruent allele. T_S = time to speciation, T_{LS} = time to divergence of the incongruent allele and the other species alleles in the situation of lineage sorting; T_H = time to the hybridization event.

other species before the speciation event, hybridization could result in a pattern identical to that expected from incomplete lineage sorting (e.g., Fig. 5.7b). Therefore, it should be possible to identify an hybridization event when the pattern observed is similar to the one in Fig. 5.7d, but in the presence of a pattern such as that of Fig. 5.7b, it would impossible to discriminate between both hypotheses (Holder *et al.*, 2001).

The *GAPDH* haplotype network may give us examples of both hybridization and incomplete lineage sorting between the *blanda*/*woodsii* and the *foliolosa*/*nitida*/*palustris* diploid groups. First, a hybridization event is probably the cause of the position of the *nitida*604.A allele (haplotype IV-m) in the *blanda*/*woodsii* group (Fig. 5.5). The hybridization hypothesis is supported because the haplotype connects to the network three steps away from the node separating the two diploid groups on the network and also because it is found in a contemporary *R. blanda* individual. This shows that the divergence between the incongruent haplotype IV-m and the other species' allele is

recent relative to the separation between the two diploid groups. The other incongruent allele, *palustris*386 (haplotype IV-w), is more likely to be caused by incomplete lineage sorting because it diverges from its ancestor one step away from the node delimiting the two diploid groups on the network (i.e., the split is relatively old) and because it is five steps away from the closest contemporary alleles of the *blanda/woodsii* group, which is plausible if it has evolved independently from these alleles for some time. As discussed above, however, it is impossible to completely reject the hypothesis of hybridization for this incongruence. It is also plausible that contemporary *blanda/woodsii* alleles closer to this allele exist but were not sampled.

5.6.1.2 Testing species boundaries

Hybridization is more frequent among closely related species. The same is true of incomplete lineage sorting, which is particularly important for nuclear genes because their effective population size is greater than for chloroplast or mitochondrial genes (Moore, 1995; Wollenberg and Avise, 1999; Rosenberg, 2003). If we consider that species are ecologically, morphologically, and (or) genetically cohesive groups of populations that evolve independently from other such groups, then nuclear genes may fail to identify recently derived species if a criterion of monophyly [e.g., the genealogical species concept (Baum and Shaw, 1995), the monophyletic species concept (Mishler and Theriot, 2000; Wheeler and Platnick, 2000)] is applied (Hudson and Coyne, 2002). Templeton (2001) has proposed using nested clade analysis as a way to test "cohesive" species boundaries (i.e., Templeton, 1989), therefore allowing some incongruence between the species tree and the gene tree. Unfortunately, this method requires extensive population sampling, which is a laborious task for single copy nuclear genes because of the extensive cloning effort necessary to properly sample alleles. As an alternative, AMOVAs were used to evaluate the genetic variation due to within species (or groups of species) variation as compared to among species variation, and to test whether the latter variance is greater than that expected by chance. This method also allows some alleles to be incongruent with the species tree.

The network suggests that *R. gymnocarpa* is sister to all other North American *Rosa* species of section *Cinnamomeae*. The distinctiveness of this species has already been reported based on morphological characters (Watson, 1885; Crépin, 1896), but its phylogenetic position was uncertain. *Rosa pisocarpa*, although non-monophyletic, is distinct from diploid species of the *R. carolina* complex on the network and its position suggests that eastern diploid species are monophyletic.

Among the largely eastern taxa of the complex, AMOVAs identified two major groups of diploids: *blanda/woodsii* and *foliolosa/nitida/palustris*. This shows that the incongruence found among groups (and discussed above) is not significant and that these groups could be considered as distinct. In the *blanda/woodsii* group, no distinction was found between *R. blanda* and *R. woodsii*. Indeed, these species cannot be distinguished using morphological and molecular (AFLP) characters (Chapter 2). Moreover, hybrids between *R. blanda* and *R. woodsii* have been shown to be highly fertile (Erlanson, 1934; Ratsek *et al.*, 1939) and a hybrid zone appears to exist in the area where the two species overlap (Lewis, 1962). Given this, the status of these species certainly needs to be addressed. In the *foliolosa/nitida/palustris* group, analyses of molecular variance suggested that *Rosa foliolosa* was distinct, although no strong conclusions regarding this species are drawn because of limited sampling. Yet, the distinction of *R. foliolosa* from other eastern diploid species is supported by morphology, this species being peculiar for its narrow leaflets and short pedicels, among other characters (Lewis, 1957b, 1958). The AMOVAs also suggest a weak distinction between *R. nitida* and *R. palustris* even if the network clearly shows that they do not form distinct groups. The species status for these two taxa is different from that of *R. blanda* and *R. woodsii* because they are clearly distinct morphologically (Lewis, 1957a, 1957b). *Rosa nitida* has numerous red bristles, is generally less than one meter tall and has no distinct infrastipular thorns, whereas *R. palustris* lacks bristles, is greater than one meter and almost always has curved infrastipular thorns. Therefore, the absence of reciprocal monophyly between *R.*

nitida and *R. palustris* for the *GAPDH* marker may be a consequence of their recent divergence.

5.6.2 Origin of the polyploids

The identification of genetically distinct groups of diploids in section *Cinnamomeae* in North America allows the evaluation of different evolutionary hypotheses concerning the origin of the polyploids. Yet, it can be difficult to determine if a polyploid is an autopolyploid or an allopolyploid in the event of conflicting signals produced by hybridization among polyploid species, gene flow between diploids and polyploids, or allelic segregation in polyploids. Both homoploid hybridization among polyploid species and gene flow from diploids to polyploids can introduce haplotypes in a polyploid that were not originally involved in its formation and can cause an autopolyploid to look like an allopolyploid. However, gene flow also can cause an allopolyploid to look like an autopolyploid if alleles from a diploid species are fixed in the allopolyploid due to recurrent gene flow. A further confounding factor is allelic segregation. Allopolyploids are expected to maintain alleles from both parental species in their genomes by disomic segregation due to bivalent formation at meiosis. This is to be expected in northeastern American polyploid *Rosa* species because individuals from the three polyploid species investigated show bivalent formation (Erlanson, 1929; Lewis, 1957b). Nonetheless, occasional pairing between homoeologous chromosomes (from the different diploid species) at meiosis could cause tri- or tetravalent formation. Indeed, trivalents and tetravalents have been observed in these polyploids (W.H. Lewis, unpublished data), but these and other meiotic irregularities such as lagging chromosomes and interlocked ring bivalents are rare and are only known of individuals from the zone of sympatry between *R. arkansana* and *R. carolina* (Lewis, 1966). Such multivalent formation leads to multisomic segregation that could bias the expected 1:1 ratio of parental alleles in an individual. Eventually this could lead to the fixation of alleles that come from a single diploid parent, resulting in a situation where an allopolyploid might look like an autopolyploid.

Inspection of the *GAPDH* network shows that polyploids are of recent origin because many polyploid haplotypes are also found in contemporary diploids. The presence of shared haplotypes among diploids and polyploids makes the determination of the type of polyploid formation more difficult for each species. This is because it is harder to eliminate hypotheses of hybridization among polyploid species and of gene flow between diploids and polyploids when diploids and polyploids share the same haplotypes. Of these confounding processes, gene flow between ploidy levels seems unlikely for many reasons. First, very few triploids have been reported in wild roses (Erlanson, 1929) and crosses between diploids and tetraploids give triploids that are highly sterile (Erlanson, 1934). Second, diploid and tetraploid species of *Rosa* are often separated both in space and in time of flowering, with diploids flowering before the tetraploids, except for *R. palustris* that flowers after all other species (Erlanson, 1930). Polyploids more often grow in dry soils, either in sandy soils (*R. carolina* and *R. virginiana*; although *R. virginiana* also grows in salt marshes) or in upland prairies (*R. arkansana*), whereas diploids grow in bogs (*R. nitida* and *R. palustris*) or in mesic soils along woods and rivers (*R. blanda* and *R. woodsii*). Therefore, we consider that the probability of gene flow between ploidy levels is low. For the other conflicting processes, hybridization at the polyploid level and allele segregation in the polyploids, the recent origin of the complex allows us to make some assumptions about the expected results.

Given that each polyploid species has evolved recurrently (see below), the recent origin of polyploids gives little time for between population genetic homogenization within polyploid species. Thus, if we have many recent formations of the polyploid species, we expect that individuals from several separate populations retain information of their origin. In other words, hybridization and allele segregation should only affect a limited number of populations in each species. Therefore, the expectation for an autopolyploid species is that most individuals will have alleles from a single diploid species even if a few may have acquired alleles from another diploid species via introgression. Moreover, individuals bearing introgressed alleles should be geographically close to individuals (or species) from which the allele is derived

(Rieseberg, 1998). In a similar way, it is unlikely that parental alleles in allopolyploid individuals will segregate in all populations and even less likely that the segregation will always be toward the same parental alleles (unless there is selection). Therefore, we expect that most individuals of an allopolyploid species will possess alleles from two diploid species even if some individuals could have fixed alleles from a single diploid species or have segregated toward a ratio of parental diploid alleles that deviates from the expected 1:1 ratio. In a further attempt to limit the potential impact of hybridization on the topology of our network, we avoided sampling individuals in areas where the distribution of polyploid species overlapped. The only exception is *R. arkansana* for which a few individuals were sampled from the zone of sympatry with *R. carolina*; potential impacts on the conclusions are discussed below.

Of the eight *R. arkansana* individuals sampled, all have alleles in the *blanda/woodsii* group, five lacking alleles from the *foliolosa/nitida/palustris* diploid group. Moreover, the three individuals with alleles from the latter group come from the region of sympatry between *R. arkansana* and *R. carolina* (Figs. 5.1 and 5.6). This suggests that *R. arkansana* evolved from within the *blanda/woodsii* group and that the presence of alleles from the *foliolosa/nitida/palustris* group in some individuals could be the result of introgression from *R. carolina*. Indeed, a hypothesis of introgression from *R. carolina* to *R. arkansana* is supported by cytological (Lewis, 1966) and morphological (A. Fishbein and W. H. Lewis, unpublished manuscript) evidence suggesting hybridization between these species. Because the relationships within the *blanda/woodsii* group are unresolved using the *GAPDH* marker, it cannot be stated whether *R. arkansana* is an auto- or an allopolyploid using a taxonomic definition (Grant, 1981; Ramsey and Schemske, 1998). Yet, some prefer to define autopolyploidy in a cytological context (Stebbins, 1980; Levin, 2002), according to which autopolyploids evolve from parents that are inter-fertile at the diploid level whereas allopolyploids are formed from a hybrid that has reduced fertility. This definition predicts multivalent formation in autopolyploids and bivalent formation in allopolyploids, at least in the first stages of their evolution. According to the cytological definition, *R. arkansana* would probably be an autopolyploid because *R. blanda* and *R.*

woodsii produce highly fertile hybrids and because they are morphologically and genetically similar.

Rosa carolina is different from *R. arkansana* in that all individuals investigated except two have alleles from both the *blanda/woodsii* and the *foliolosa/nitida/palustris* diploid groups. Given the wide geographic distribution of the individuals sampled, we can affirm that *R. carolina* is an allopolyploid with one parent from the *blanda/woodsii* diploid group and the other from the *foliolosa/nitida/palustris* group. The deviation from a 1:1 ratio of parental alleles expected for allopolyploids observed in some individuals is probably the result of either segregation of homoeologous chromosomes or introgression.

Finally, individuals of *R. virginiana* were found to possess only alleles that were exclusive to the *foliolosa/nitida/palustris* diploid group, except for two individuals that also have a *blanda/woodsii* allele and one that has an allele of ambiguous origin. Therefore, the most likely hypothesis for the origin of this polyploid species is that it originated from within the *foliolosa/nitida/palustris* diploid group. Again, we cannot be certain whether *R. virginiana* is an auto- or an allopolyploid due to the lack of resolution within the *foliolosa/nitida/palustris* group. It is highly likely that *R. foliolosa* was not involved in the evolution of this species, however, because no *R. virginiana* allele was closely related to those sampled from *R. foliolosa*. The situation is also different from that for *R. arkansana* because we have no information on the fertility of hybrids between *R. palustris* and *R. nitida*. Hence, any conclusion regarding the polyploid origin of *R. virginiana* must await further data.

To summarize, *R. arkansana* evolved from the *blanda/woodsii* group, *R. virginiana* from the *foliolosa/nitida/palustris* group and *R. carolina* from a cross between these two eastern diploid groups. These results allow an evaluation of different hypotheses that have been proposed concerning the origins of eastern polyploids. Erlanson (1929) proposed that *R. arkansana* originated from a cross between *R. blanda* and either *R. macounii* Greene or *R. fendleri* Crépin, two species now considered synonymous with *R.*

woodsii (Erlanson, 1934). This hypothesis is compatible with the present findings, although our results cannot confirm that two taxonomic species were involved. For *R. carolina*, Erlanson (1929) first proposed that *R. virginiana* would have crossed with *R. palustris* and that the hybrid eventually would have given a tetraploid that would have backcrossed to *R. virginiana* to give *R. carolina*. This hypothesis is improbable according to the present results because it would imply that the genetic diversity of *R. carolina* is a subset of *R. virginiana*. Because several *R. carolina* individuals lack *R. virginiana* haplotype ancestors, our data disagree with such an evolutionary scenario. A few years later, Erlanson (1938) suggested that *R. blanda* and *R. woodsii* gave rise to all three eastern tetraploid species as well as to *R. foliolosa*, *R. nitida*, and *R. palustris*. Her hypothesis regarding the evolution of *R. foliolosa*, *R. nitida* and *R. palustris* seems improbable in light of the present data because these species do not appear to be derived from *R. blanda* and *R. woodsii*. Her hypothesis regarding the evolution of *R. carolina* and *R. virginiana* from *R. blanda* and *R. woodsii* alone is also likely inaccurate because the *foliolosa/nitida/palustris* diploid group was certainly involved in the origin of these two tetraploid species.

The results clearly show that the western diploid species were not involved in the origins of the eastern polyploid species. It is indeed improbable that a western species would have been involved in the origin of the polyploids without leaving a trace given that several polyploid individuals from a wide geographic range were sampled. A general pattern of evolution within section *Cinnamomeae* in North America thus emerges from these results: diploids west and east of the Rocky Mountains seem to form distinct groups and eastern polyploids evolved from eastern diploids following the diversification of diploids.

5.6.3 Multiple origins of polyploidy

The number of polyploid origins was estimated using "polyploid haplotype groups" (Fig. 5.3), which estimates the genetic diversity of polyploids that is contributed by diploids. When working with haploid markers, each polyploid haplotype group can

be interpreted as a distinct polyploid origin (e.g., Soltis *et al.*, 1989; Doyle *et al.*, 1990; Segraves *et al.*, 1999; Sharbel and Mitchell-Olds, 2001). Similarly for autosomal markers, a specific combination of polyploid haplotype groups in individuals can sometimes be considered to represent a distinct origin. This is true of selfing allopolyploids that are homozygous at each homoeologuous locus (as in *Glycine*; Doyle *et al.*, 2004) and of clonal taxa (Joly and Bruneau, 2004). More often alleles at nuclear loci will segregate in polyploids, however, and this can create any possible combination of alleles. Hence, interpreting each genotype as an independent origin would seem to overestimate the true number of polyploid origins. For this reason it was assumed that each tetraploid formation involved four distinct polyploid haplotype groups and that each independent formation always involved polyploid haplotype groups that were not involved in other polyploid origins. These assumptions are clearly overly conservative. For example, there may be unsampled diploid haplotypes that would increase the number of polyploid haplotype groups and a tetraploid formation can involve less than four alleles. Yet, the approach is legitimate if the objective is to evaluate the likelihood that species evolved recurrently rather than to estimate the true number of polyploid origins.

According to these conservative assumptions, all polyploid species must have evolved at least three times to explain the observed diversity. This estimate makes many simplifications such as the absence of gene flow between ploidy levels that would tend to overestimate the number of independent origins. Yet, the impact of gene flow between ploidy levels is probably limited in North American roses (see above). Hybridization between polyploid species is another way by which polyploids acquire genetic variability that is not due to multiple origins. It is harder to account for hybridization because polyploids are known to hybridize and because they have a recent origin; this is why individuals mostly were sampled from outside the zones of sympatry between polyploids. The only exception is *R. arkansana* from which we sampled five individuals that are considered near or in the sympatric zone with *R. carolina* (Figs. 5.1, 5.6). But even with these individuals removed (accessions 345, 406, 416, 665, and 692), there are still

seven polyploid haplotype groups represented and two independent origins of *R. arkansana* are needed to explain such a diversity.

Interestingly, polyploids have been able to acquire most of their genetic diversity at the diploid level; almost all diploid haplotypes were also found in one or more polyploid species (Fig. 5.5). This further supports the hypothesis of independent origins of the polyploid species, but above all it shows that polyploids possess a high degree of genetic variation. In the end, it is this genetic diversity that is most important, not how it was acquired. This variability, coupled with recombination and mutation in polyploid species, is likely to allow polyploid species to create adaptive genotypes that will be fitter and have more evolutionary potential in certain environments.

5.6.4 Taxonomic consequences

The rose species investigated here have sometimes been divided into sections *Cinnamomeae* (*R. arkansana*, *R. blanda*, *R. woodsii*) and *Carolinae* (*R. carolina*, *R. foliolosa*, *R. nitida*, *R. palustris*, *R. virginiana*) based on strictly basal placentation (*Carolinae*) versus basilo-parietal placentation (*Cinnamomeae*), presence (*Carolinae*) versus absence (*Cinnamomeae*) of hypanthium glands, and deciduous (*Carolinae*) versus persistent (*Cinnamomeae*) sepals after fruit maturation (Crépin, 1889). The present data suggest that the separation of these two sections is artificial. First, it makes section *Cinnamomeae* paraphyletic and second, the reticulate origin of *R. carolina* also renders section *Carolinae* unnatural. Therefore, the best solution would be to treat section *Carolinae* as synonymous with section *Cinnamomeae*. This was previously proposed by (Erlanson, 1934) and (Lewis, 1957a) based on the unreliability of the morphological characters that were used to separate these sections, and also supports investigations of biochemical (Grossi *et al.*, 1998) and molecular characters (Wissemann and Ritz, 2005). Yet, this taxonomy still is used in the most recent comprehensive flora treatments in the United States (generic Flora of the southeastern United States: Robertson, 1974) and in Europe (Tutin *et al.*, 1968), perhaps because Rehder's (1940) classification, which uses section *Carolinae*, is still

the most widely cited taxonomic treatment of *Rosa*. We suggest that section *Carolinae* be synonymized in further taxonomic treatments.

The present study also sheds light upon the species status of the three polyploid taxa of the *R. carolina* complex. The results suggest that *R. arkansana*, *R. carolina* and *R. virginiana* have distinct evolutionary histories, although it will certainly be important to confirm this with more markers. This also suggests that these polyploids should be considered distinct species. These species are highly polymorphic probably in part owing to their recurrent origins, and their identification will remain difficult, especially in regions of sympatry where the extensive variation is best explained by hybrid zones. Yet, the present results suggest that these are secondary hybrid zones (Endler, 1977; Barton and Hewitt, 1985) that were formed after polyploid speciation. Of course, distinct evolutionary histories do not guarantee that species will always remain distinct and the extent of gene flow in these secondary hybrid zones will be determinant for the future of these polyploids.

In conclusion, the genealogical approach based on single-copy nuclear genes has allowed reconstructing the evolutionary history of polyploid species even in the presence of incomplete lineage sorting among diploids, of hybridization among diploids and polyploids, and of allelic segregation in polyploids. This study therefore gives a conceptual framework that may be used to unveil the evolutionary history of other species complexes where hybridization and polyploidy are important.

CHAPITRE 6

Conclusion

Cette thèse s'est penchée sur les problèmes taxonomiques et sur l'évolution des roses indigènes de la section *Cinnamomeae* à l'est des montagnes Rocheuses, un complexe d'espèces représentatif du genre *Rosa* en termes de polymorphisme, de polyploïdie, d'hybridation et surtout de problèmes taxonomiques. Les études antérieures basées sur la morphologie, la cytologie et les croisements expérimentaux n'avaient pas pu clairement démontrer l'existence de groupes d'organismes distincts pouvant être reconnus en tant qu'espèces. De plus, les résultats ambigus obtenus par les analyses phylogénétiques récentes, à cause de la faible variabilité génétique des roses, laissaient planer un doute relativement à la possibilité de reconstruire l'évolution de ce groupe.

Toutefois, cette thèse a montré qu'en utilisant des méthodes multivariées appropriées, il est possible de définir des espèces dans ce groupe à l'aide de données morphologiques et moléculaires. De plus, en utilisant des gènes nucléaires à copie unique, il a d'abord été possible de reconstruire l'évolution des diploïdes, puis celle des polyploïdes. Bref, nous avons maintenant une meilleure connaissance de la délimitation ainsi que de l'évolution des roses de la section *Cinnamomeae* à l'est des montagnes Rocheuses.

6.1 Combien d'espèces ?

Contrairement aux dires de certains botanistes qui aiment prétendre qu'il y a une seule espèce de rose en Amérique du Nord, cette thèse montre clairement qu'il est

possible d'identifier plusieurs espèces chez les roses de la section *Cinnamomeae* à l'est des montagnes Rocheuses. En utilisant une approche objective pour définir les espèces, l'analyse multivariée des données morphologiques et moléculaires a permis de confirmer que presque toutes les espèces reconnues par Erlanson (1966) étaient bel et bien distinctes. Il y a cependant deux exceptions majeures, et les décisions à prendre relativement à celles-ci diffèrent.

En premier lieu, il y a le problème des *R. blanda* et *R. woodsii*. Ces deux espèces sont généralement différenciées par la présence ou l'absence d'aiguillons infrastipulaires chez *R. woodsii* et *R. blanda*, respectivement. Bien que d'autres distinctions ont été proposés pour différencier ces espèces (voir Lewis, 1962), ce caractère évident semblait le plus utile pour ce faire. Or, la présente thèse a montré qu'il est impossible de distinguer le *R. blanda* du *R. woodsii*, que ce soit par des analyses multivariées de données morphologiques ou moléculaires (Chapitre 2), ou par l'analyse de gènes nucléaires à copie unique (Chapitre 3). Même les aiguillons infrastipulaires ne semblent pas toujours différencier ces espèces (Chapitre 2). Mais si ce constat est nouveau, le problème entourant ces deux espèces ne l'est pas. En effet, une zone d'hybridation avait été décrite entre les deux espèces (Lewis, 1962), ce qui donnait alors une lecture différente du problème puisque la difficulté à différencier les *R. blanda* et *R. woodsii* était considérée comme une conséquence de l'hybridation entre eux. Par contre, les résultats présentés ici semblent contredire la présence d'une zone d'hybridation et favorisent davantage l'hypothèse d'une seule espèce. En effet, s'il y avait une zone d'hybridation, on s'attendrait à ce que les individus d'une espèce soient différenciés de ceux de l'autre à l'exception des individus qui se trouvent près de la zone d'hybridation. Cependant, ce n'est pas ce que l'on observe puisque les individus des *R. blanda* et *R. woodsii* se chevauchent considérablement, encore plus au niveau morphologique que génétique (Chapitre 2). De plus, les caractères morphologiques varient graduellement d'un océan à l'autre, suggérant que la variation observée entre les populations de l'est et de l'ouest est due à un isolement par distance plutôt qu'à une zone d'hybridation (Chapitre 2). Donc, étant donné que les *R. blanda* et *R. woodsii* ne peuvent être distingués morphologiquement et génétiquement (Chapitre 2 et

3), qu'ils produisent des hybrides fertiles (Erlanson, 1929) et que les données morphologiques n'appuient pas la présence d'une zone d'hybridation (Chapitre 2), la meilleure solution présentement serait de considérer *R. blanda* et *R. woodsii* comme une seule espèce. Ceci est vrai en fonction de la définition d'espèce utilisée dans cette thèse (Chapitre 1), mais une conclusion identique résulterait aussi de l'utilisation de plusieurs autres concepts d'espèce (e.g., phénétique - Sneath et Sokal, 1973; reconnaissance - Paterson, 1985; cohésive - Templeton, 1989; généalogique - Baum et Shaw, 1995; génotypique - Mallet, 1995; biologique - Mayr, 2000; phylogénétique - Wheeler et Platnick, 2000). Ainsi, *R. woodsii* Lindl. devrait être réduit en synonymie du *R. blanda* Ait. parce que ce dernier a priorité sur *R. woodsii*.

En second lieu, les données morphologiques et moléculaires (Chapitre 2) n'ont pas clairement distingué les tétraploïdes *R. carolina* et *R. virginiana*. Contrairement au problème des *R. blanda* et *R. woodsii*, il est toutefois possible de discriminer presque complètement le *R. carolina* du *R. virginiana* dans les analyses morphologiques et génétiques. D'ailleurs, les individus qui sont intermédiaires sont aussi ceux qui sont le plus près de la zone de sympatrie entre ces deux espèces, ce qui correspond à un patron typique d'une zone d'hybridation (Arnold, 1997). De plus, l'analyse des similarités génétiques et morphologiques avec les diploïdes ainsi que l'analyse du gène *GAPDH* suggèrent des origines distinctes pour ces deux espèces, ce qui supporte l'hypothèse d'une zone d'hybridation secondaire. Bref, les résultats présentés dans la thèse militent en faveur de reconnaître ces deux espèces tétraploïdes. Cependant, la zone d'hybridation entre elles pourrait venir perturber leur cohésion. Si l'introgression est trop importante, ces espèces n'évolueront plus de façon indépendante et la cohésion des espèces risque d'être compromise. Donc, le constat actuel que les *R. carolina* et *R. virginiana* sont deux espèces distinctes ne garantit pas leur évolution distincte dans l'avenir.

Pour conclure, les espèces de la section *Cinnamomeae* présentes à l'est des montagnes Rocheuses sont les suivantes :

Diploïdes : *R. blanda* Ait., *R. foliolosa* Nutt., *R. nitida* Wild. et *R. palustris* Marsh.

Tétraploïdes : *R. arkansana* Porter, *R. carolina* L. et *R. virginiana* Mill.

6.2 Évolution de *Rosa* sect. *Cinnamomeae* en Amérique du Nord

La faible variabilité génétique trouvée dans les études phylogénétiques antérieures des roses nord-américaines (par exemple, Ritz *et al.*, 2005) laissait planer un doute relativement à la possibilité de reconstruire l'évolution de ces espèces. Cependant, l'étude de trois gènes nucléaires a permis d'obtenir la variabilité nécessaire pour acquérir des phylogénies bien résolues et de formuler de bonnes hypothèses évolutives pour les espèces du complexe.

L'évolution des polyploïdes a d'abord été étudiée à l'aide de trois gènes nucléaires. Deux approches ont été utilisées. Dans un premier cas, une méthode a été développée afin d'incorporer la variation allélique dans la reconstruction phylogénétique d'individus à partir de plusieurs gènes (Chapitre 3). En plus d'être informative relativement aux relations phylogénétiques des espèces, cette méthode permet aussi de déterminer si différentes espèces semblent génétiquement distinctes en fonction des marqueurs utilisés. La seconde approche visait à reconstruire l'évolution des espèces à partir de plusieurs arbres géniques en utilisant un principe de parcimonie qui minimise le nombre de coalescences profondes (*deep coalescences*) (Chapitre 4).

Les résultats suggèrent que les espèces diploïdes de la section *Cinnamomeae* à l'est et à l'ouest des montagnes Rocheuses forment deux clades distincts (Figure 6.1). Bien que la position du *R. pisocarpa* ne semble pas certaine (Chapitre 3 et 4), cette solution est la plus parcimonieuse selon un critère de minimisation des coalescences profondes (Chapitre 4). Ceci implique que l'évolution des diploïdes à l'ouest et à l'est des Rocheuses a été indépendante. Elle implique aussi que la présence actuelle du *R. blanda* (syn. *R. woodsii*) à l'ouest des Rocheuses est la conséquence d'une migration récente de *R. blanda* depuis l'est américain. Dans l'est, *R. palustris* et *R. nitida* sont clairement groupes-frères (Fig. 6.1).

Cependant, les approches diffèrent relativement au placement de *R. foliolosa*. Selon l'approche basée sur les organismes, *R. foliolosa* se rapproche plus des *R. nitida* et *R. palustris* (Chapitre 3), alors que l'approche basée sur les espèces suggère que le *R. foliolosa* est le groupe-frère de toutes les autres espèces est-américaines (Chapitre 4). Étant donné l'incongruence entre ces analyses, ces relations sont ici considérées comme ambigües (Fig. 6.1). Plus d'individus et plus de gènes devront donc être utilisés afin de déterminer la position phylogénétique exacte du *R. foliolosa*.

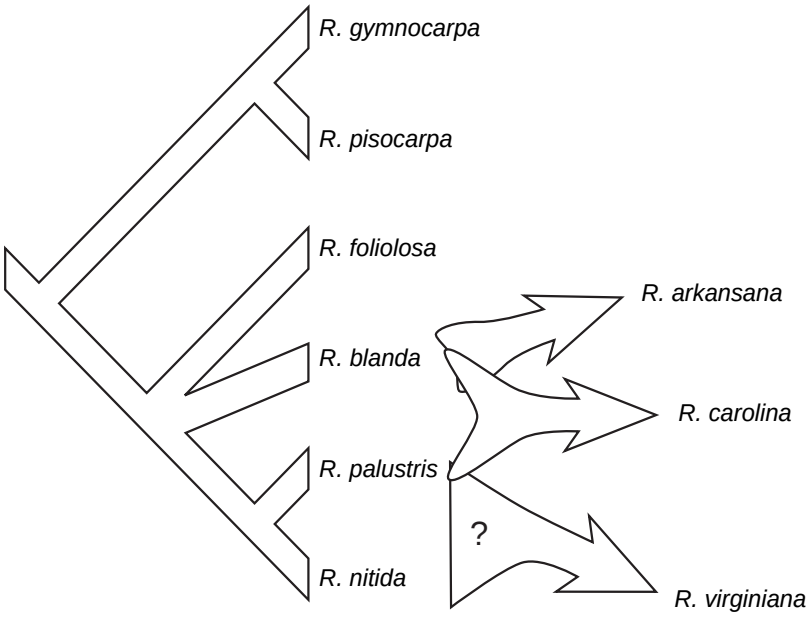


Figure 6.1 Schéma représentant l'évolution probable de la section *Cinnamomeae* à l'est des montagnes Rocheuses. Les flèches représentent les événements de spéciation polyploïdes et le point d'interrogation indique qu'il n'est pas clair qu'elle espèce a été impliquée dans l'évolution. La polytomie au niveau diploïde reflète les différentes topologies obtenues dans les analyses.

Pour ce qui est des polyploïdes, le gène *GAPDH* a été séquencé pour plusieurs individus par espèce afin de déterminer leur origine (Chapitre 5). De plus, la comparaison des similarités morphologiques et génétiques des polyploïdes avec les diploïdes ainsi que le positionnement des polyploïdes dans les analyses d'ordination ont aidé à déterminer l'origine des polyploïdes (Chapitre 2). Les résultats démontrent que

seules les espèces diploïdes à l'est des montagnes Rocheuses ont été impliquées dans l'évolution des polyploïdes du complexe.

Pour le *R. arkansana*, le scénario le plus probable implique une évolution à partir du *R. blanda* (incluant *R. woodsii*) (Fig. 6.1). Ceci est d'ailleurs en accord avec les hypothèses antérieures concernant son origine (Erlanson, 1929, 1934). En fonction de ces résultats, le *R. arkansana* serait donc un autopolyploïde à la fois selon les définitions cytologique et taxonomique.

Pour le *R. carolina*, les résultats présentés ici suggèrent une évolution à partir d'un croisement entre le *R. blanda* (incluant *R. woodsii*) et le *R. palustris* (Fig 6.1). Si l'analyse du gène *GAPDH* ne permettait pas de déterminer qui du *R. palustris* ou du *R. nitida* avait été impliqué dans l'évolution du *R. carolina* à cause d'un manque de résolution (Chapitre 5), l'analyse des données morphologiques et des AFLPs ont montré que le *R. carolina* était plus près du *R. palustris* (Chapitre 2). *Rosa palustris* avait d'ailleurs déjà été suggéré comme un géniteur potentiel du *R. carolina* (Erlanson, 1929; Lewis, 1957b). D'autres scénarios d'évolution avaient été proposés pour le *R. carolina* (Erlanson, 1929; Erlanson, 1938), mais ceux-ci se sont avérés improbables en fonction des données actuelles (voir Chapitre 5). Selon une définition taxonomique, *R. carolina* serait donc un allopolyploïde. Par contre, la même conclusion ne serait pas nécessairement obtenue en fonction d'une définition cytologique puisque des croisements effectués entre les *R. blanda* et *R. palustris* ont donné des hybrides fertiles (Erlanson, 1934). Donc on peut s'attendre à ce que le *R. carolina* exhibait des multivalents à la méiose peu après sa formation, même si aujourd'hui ce sont principalement des bivalents qui sont observés (Lewis, 1966).

Finalement, le *R. virginiana* a évolué à partir du *R. nitida*, du *R. palustris* ou d'un croisement entre ces deux diploïdes (Figure 6.1). L'impossibilité de distinguer ces deux espèces diploïdes à l'aide du gène *GAPDH* (Chapitre 5) et l'incongruence obtenue entre les similarités génétiques et morphologiques (Chapitre 2) ne permettent pas d'obtenir d'hypothèse évolutive plus précise. Dans le cas du *R. virginiana*, deux hypothèses avaient

été avancées relativement à son évolution. L'une proposait le *R. blanda* comme espèce ancestrale (Erlanson, 1938), l'autre suggérait que le *R. palustris* devait avoir été impliqué dans son évolution (Lewis, 1957b). Les résultats présentés rejettent la première hypothèse alors que la seconde est probable en fonction des conclusions tirées ici. L'incertitude quant à l'origine exacte du *R. virginiana* ne permet pas de conclure s'il s'agit d'un autopolyploïde ou d'un allopolyploïde selon la définition taxonomique. De plus, nos connaissances minimes relativement aux croisements entre les *R. nitida* et *R. palustris* (voir Erlanson, 1934) ne permettent pas non plus de catégoriser le *R. virginiana* en fonction de la définition cytologique.

Finalement, l'analyse du gène nucléaire *GAPDH* a démontré que toutes les espèces polyploïdes avaient évolué à plusieurs reprises. Ainsi, les espèces tétraploïdes du groupe appuient les données moléculaires récentes qui suggèrent que la plupart des polyploïdes évoluent de façon répétée (Soltis et Soltis, 2000; Joly et Bruneau, 2004).

Ces résultats montrent bien l'importance de la polyploïdie et de l'hybridation dans ce groupe. La polyploïdie est responsable de trois événements de spéciation sur sept dans le complexe. De plus, elle implique à la fois des phénomènes d'autopolyploïdie (*R. arkansana*) et d'allopolyploïdie (*R. carolina*, du moins selon la définition taxonomique). Pour ce qui est de l'hybridation, les résultats montrent qu'elle a clairement été impliquée dans l'évolution du *R. carolina*, et peut-être aussi dans celle du *R. virginiana*. De plus, les processus évolutifs œuvrant dans la zone d'hybridation entre les *R. carolina* et *R. virginiana* seront déterminants pour l'avenir de ces deux espèces.

D'un point de vue général, il est intéressant de noter que bien peu d'études sur des complexes polyploïdes se sont penchées sur la problématique de délimitation d'espèces (voir cependant Suda et Lysák, 2001; Vanderhoeven *et al.*, 2002; Perný *et al.*, 2005) si l'on compare avec le nombre d'études qui ont reconstruit l'histoire évolutive d'espèces polyploïdes. La question de délimitation des espèces est néanmoins importante pour bien comprendre l'évolution d'un groupe. Si les délimitations d'espèces ne sont pas claires,

l'interprétation de l'évolution des individus peut être faussée. Ainsi, afin de bien comprendre l'évolution d'un complexe polyploïde, il est important d'étudier à la fois la délimitation et l'évolution des espèces. Il est à souhaiter que plus d'études dans le futur attaqueront ces deux problèmes de front, puisque cela ne peut qu'aider la compréhension globale des complexes d'espèces polyploïdes.

6.3 L'avenir…

Si l'évolution de *Rosa* sect. *Cinnamomeae* est maintenant plus claire, beaucoup d'autres expériences peuvent encore nous éclairer sur l'évolution du groupe et sur les processus évolutifs actuellements importants dans celui-ci. Par exemple, il serait important d'étudier avec plus de profondeur les zones d'hybridation identifiées ici. Ceci permettrait entre autres d'évaluer l'importance réelle de l'introgression entre espèces et dans certains cas de voir à quel point les frontières d'espèces sont menacées par l'hybridation. Il serait aussi intéressant de se pencher sur la problématique du flux génique entre les niveaux de ploïdie. Si celui-ci semble faible, des études dans d'autres groupes de plantes (résumé dans Ramsey et Schemske, 1998) montrent qu'il doit exister et qu'il pourrait être plus important qu'on le croît. Déterminer l'importance de ce flux génique permettrait certainement de mieux comprendre ce groupe. D'un côté plus écologique, il serait important de déterminer l'importance de l'environnement dans l'évolution des espèces (Rieseberg *et al.*, 2003) ainsi que sont influence sur la variation morphologique. Finalement, il serait intéressant d'étudier la protéomique et la génomique des espèces polyploïdes, afin d'étudier ces changements dans un contaxte évolutif.

6.4 Clé d'identification

Dans le but d'aider à l'identification des espèces de roses de la section *Cinnamomeae* à l'est des montagnes Rocheuses, une clé d'identification artificielle est formulée plus bas.

Les unités terminales de la clé sont les espèces identifiées dans cette thèse. Si certains sont intéressés à distinguer les *R. blanda* et *R. woodsii*, il est toujours possible de se référer à l'indice d'hybridation de Lewis (1962). Seule la macromorphologie a été utilisée dans cette clé. Bien sûr, s'il est possible de mesurer la taille des stomates ou la taille des grains de pollen, cela faciliterait l'identification en différenciant les diploïdes des tétraploïdes. L'arbre de classification reconstruit au chapitre 2 a servi de base pour construire la clé et d'autres caractères ont été ajoutés à chaque nœud en fonction de ceux qui différenciaient le mieux les groupes (Chapitre 2). Cette clé ne garantit pas que les espèces seront toujours bien identifiées, mais en tenant compte de l'ensemble des informations de la clé ainsi que des résultats présentés au Chapitre 2, les chances d'erreur dans l'identification devraient être minimisées. Les erreurs les plus probables surviendront entre les *R. palustris*, *R. carolina* et *R. virginiana*. Pour différencier ces espèces, l'ensemble des caractères étudiés dans cette thèse devraient être pris en compte (e.g., Fig. 2.11).

Clé d'identification de *Rosa* sect. *Cinnamomeae* à l'est des montagnes Rocheuses

1 Hypanthiums glabres

2 Aiguillons infrastipulaires trapus à base large.....*R. virginiana*

2 Aiguillons infrastipulaires absents ou fins et à base étroite

3 Généralement moins de 2 poils par mm^2 sur la surface abaxiale de la feuille ; aiguillons infrastipulaires toujours absents sur les nouvelles branches ; acicules toujours présents sur les nouvelles branches ; folioles au nombre de 6,5 à 9 par feuille .....*R. arkansana*

3 Généralement plus de 2 poils par mm^2 sur la surface abaxiale de la feuille ; aiguillons infrastipulaires présents ou absents sur les nouvelles branches ; acicules généralement absents sur les nouvelles branches ; folioles au nombre de 5 à 7,5 par feuille .....*R. blanda*

1) Hypanthiums glandulaires

2) Acicules présents sur les nouvelles branches

3) Poils présents sur la surface abaxiale de la foliole terminale

4) Aiguillons infrastipulaires présents .....*R. carolina*

4) Aiguillons infrastipulaires absents .....*R. arkansana*

3) Poils absents de la surface abaxiale de la foliole terminale .....*R. nitida*

2) Acicules absents sur les nouvelles branches

3) Largeur de la foliole terminale inférieure à 9 mm ..... *R. foliolosa*

3) Largeur de la foliole terminale supérieure à 9 mm

4) Hypanthium typiquement muni de plus de 86 glandes ; marge de la foliole terminale généralement avec plus de 20 fines dents par côté, foliole terminale oblongue ..... *R. palustris*

4) Hypanthium typiquement muni de moins de 86 glandes ; marge de la foliole généralement avec moins de 20 dents par côté, foliole terminale ovée, elliptique ou obovée

5) Acicules absents des nouvelles branches ; auricules plus longues que 3,82 mm ; Stipule plus large que 1,1 mm ; aiguillons infrastipulaires trapus et à base large.....*R. virginiana*

5) Acicules parfois présents sur les nouvelles branches ; auricules plus courtes que 3,82 mm ; stipule d'une largeur de moins de 1,62 mm ; aiguillons infrastipulaires fins et pas nécessairement à base large ....*R. carolina*

Bibliographie

Adams, K. L., R. Cronn, R. Percifield et J. F. Wendel. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. USA 100:4649-4654.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Autom. Contr. 19:716-723.

Anderson, E. et L. Hubricht. 1938. Hybridization in *Tradescantia*. III. The evidence for introgressive hybridization. Amer. J. Bot. 25:396-402.

Antunes, A., A. R. Templeton, R. Guyomard et P. Alexandrino. 2002. The role of nuclear genes in intraspecific evolutionary inference: genealogy of the *transferin* gene in the brown trout. Mol. Biol. Evol. 19:1272-1287.

Aquadro, C. F., S. F. Desse, M. M. Bland, C. H. Langley et C. C. Laurie-Ahlberg. 1986. Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. Genetics 114:1165-1190.

Archambault, A. et A. Bruneau. 2004. Phylogenetic utility of the *LEAFY/FLORICAULA* gene in the Caesalpinioideae (Leguminosae): gene duplication and a novel insertion. Syst. Bot. 29:609-626.

Arnold, M. L. 1997. Natural hybidization and evolution. Oxford University Press, New York.

Ayoub, N. A., S. E. Riechert et R. L. Small. 2005. Speciation history of the North American funnel web spiders, Agelenopsis (Araneae: Agelenidae): Phylogenetic inferences at the population-species interface. Mol. Phylog. Evol. 36:42-57.

Baldwin, B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell et M. J. Donoghue. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Mo. Bot. Gard. 82:247-277.

Barnes, S. 2002. Comparing *Arabidopsis* to other flowering plants. Curr. Opin. Plant Biol. 5:128-133.

Barton, N. H. et G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113-148.

Baum, D. A. et K. L. Shaw. 1995. Genealogical perspectives on the species problem. Pages 289-303 *dans* Experimental and molecular approaches to plant biosystematics (P. C. Hoch, et A. G. Stephenson, éds.). Monographs in systematic botany from the Missouri Botanical Garden, St. Louis, MO.

Benham, J. J. 2001. Genographer, version 1.6.0. Montana State University, Bozeman, Montana, USA. http://hordeum.oscs.montana.edu/genographer/.

Bininda-Emonds, O. R. 2004. The evolution of supertrees. Trends Ecol. Evol. 19:315-322.

Blanc, G., K. Hokamp et K. H. Wolfe. 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome Res. 13:137-144.

Britten, R. J., L. Rowen, J. Williams et R. A. Cameron. 2003. Majority of divergence between closely related DNA samples is due to indels. Proc. Natl. Acad. Sci. USA 100:4661-4665.

Bruen, T. C. 2005. PhiPack: PHI test and other tests of recombination, McGill University, Montréal, Québec, Canada. www.mcb.mcgill.ca/~trevor/.

Bruen, T. C., H. Philippe et D. Bryant. 2006. A simple robust statistical test for detecting the presence of recombination. Genetics 172:2665-2681.

Bruneau, A., S. Joly, J. R. Starr et J.-N. Drouin. 2005. Molecular markers indicate that the narrow Québec endemics *Rosa rousseauiorum* and *Rosa williamsii* are synonymous with the widespread *Rosa blanda*. Can. J. Bot. 83:386-398.

Bruneau, A., J. R. Starr et S. Joly. submitted. Phylogenetic relationships in the genus *Rosa*: new evidence from chloroplast DNA sequences and appraisal of current knowledge. Syst. Bot.

Bryant, D. et V. Moulton. 2002. NeighborNet: An agglomerative method for the construction of planar phylogenetic networks *dans* Workshop in Algorithms for Bioinformatics.

Bryant, D. et V. Moulton. 2004. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21:255-265.

Buntjer, J. B., M. Otsen, I. J. Nijman, M. T. R. Kuiper et J. A. Lenstra. 2002. Phylogeny of bovine species based on AFLP fingerprinting. Heredity 88:46-51.

Burton, T. L. et B. C. Husband. 2000. Fitness differences among diploids, tetraploids, and their triploid progeny in *Chamerion angustifolium*: mechanisms of inviability and implications for polyploid evolution. Evolution 54:1182-1191.

Castelloe, J. et A. R. Templeton. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylog. Evol. 3:102-113.

Chen, F.-C. et W.-H. Li. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am. J. Hum. Genet. 68:444-456.

Clarke, D. L. 1980. Rosa, the species. Pages 37-48 *dans* W. J. Bean, Trees and shrubs hardy in the British Isles (D. L. Clarke, et G. Taylor, éds.). John Murray, London.

Clement, M., D. Posada et K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657-1659.

Comai, L., A. P. Tyagi, K. Winter, R. Holmes-Davis, S. H. Reynolds, Y. Stevens et B. Byers. 2000. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. Plant Cell 12:1551-1567.

Crandall, K. A. et A. R. Templeton. 1993. Empirical test of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics 134:959-969.

Crépin, F. 1889. Sketch of a new classification of roses. J. Royal Hort. Soc. 11:217-228.

Crépin, F. 1894. *Rosae hybridae*: études sur les roses hybrides. Bulletin de la Société Royale de Botanique de Belgique, 1ière partie (Mémoires) 33:52-55.

Crépin, F. 1896. *Rosae Americanae*. I. Observations upon the genus *Rosa* in North America. Bot. Gaz. 22:1-34.

Cronn, R., M. Cedroni, T. Haselkorn, C. Grover et J. F. Wendel. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. Theor. Appl. Genet. 104:482-489.

Cronn, R. et J. F. Wendel. 2003. Cryptic tryst, genomic mergers, and plant speciation. New Phyt. 161:133-142.

Crosby, M. R. 1980. Polyploidy in bryophytes with special emphasis on mosses. Pages 193-198 *dans* Polyploidy: biological relevance (W. H. Lewis, éd.) Plenum Press, New York.

Darlington, C. D. 1928. Studies in *Prunus*, I and II. Journal of Genetics 19:213-256.

Davis, J. I. et K. C. Nixon. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. Systematic Biology 41:421-435.

De Bodt, S., S. Maere et Y. Van de Peer. 2005. Genome duplication and the origin of angiosperms. Trends Ecol. Evol. 20:591-597.

de Queiroz, A. 1993. For consensus (sometimes). Systematic Biology 42:368-372.

Degnan, J. H. et L. A. Salter. 2005. Gene tree distributions under the coalescent process. Evolution 59:24-37.

Denton, A. L., B. L. McConaughy et B. D. Hall. 1998. Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. Mol. Biol. Evol. 15:1082-1085.

deWet, J. M. J. 1980. Origins of polyploids. Pages 3-15 *dans* Polyploidy: biological relevance (W. H. Lewis, éd.) Plenum press, New York.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia University Press, New York.

Donnelly, P. et S. Tavaré. 1986. The ages of alleles and a coalescent. Adv. Appl. Prob. 18:1-19.

Doyle, J. J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst. Bot. 17:144-163.

Doyle, J. J. 1995. The irrelevance of allele tree topologies for species delimitation, and a non-topological alternative. Syst. Bot. 20:574-588.

Doyle, J. J. et J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11-15.

Doyle, J. J., J. L. Doyle, A. H. D. Brown et J. P. Grace. 1990. Multiple origins of polyploids in the *Glycine tabacina* complex inferred from chloropast DNA polymorphism. Proc. Natl. Acad. Sci. USA 87:714-717.

Doyle, J. J., J. L. Doyle, A. H. D. Brown et R. G. Palmer. 2002. Genomes, multiple origins, and lineage recombination in the *Glycine tomentella* (Leguminosae) polyploid complex: histone H3-D gene sequences. Evolution 56:1388-1402.

Doyle, J. J., J. L. Doyle, J. T. Rauscher et A. H. D. Brown. 2003. Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus *Glycine*). New Phyt. 161:121-132.

Doyle, J. J., J. L. Doyle, J. T. Rauscher et A. H. D. Brown. 2004. Evolution of the perennial soybean polyploid complex (*Glycine* subgenus *Glycine*): a study of contrasts. Biol. J. Linn. Soc. 82:583-597.

Ehrendorfer, F. 1980. Polyploidy and distribution. Pages 45-60 *dans* Polyploidy: Biological relevance (W. H. Lewis, éd.) Plenum press, New York.

Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton University Press, Princeton, New Jersey.

Erlanson, E. W. 1929. Cytological conditions and evidences for hybridity in North America wild roses. Bot. Gaz. 87:443-506.

Erlanson, E. W. 1930. The phenological procession in North American wild roses in relation to the polyploid series. Papers Mich. Acad. Sci. Arts and Letters 11:137-150.

Erlanson, E. W. 1934. Experimental data for a revision of the North American wild roses. Bot. Gaz. 96:197-259.

Erlanson, E. W. 1938. Phylogeny and polyploidy in *Rosa*. New Phyt. 37:72-81.

Erlanson MacFarlane, E. W. 1966. The old problem of species in *Rosa* with special reference to North America. American Rose Annual 51:150-160.

Evans, B. J., D. B. Kelley, D. J. Melnick et D. C. Cannatella. 2005. Evolution of RAG-1 in polyploid clawed frogs. Mol. Biol. Evol. 22:1193-1207.

Evans, R. C., L. A. Alice, C. S. Campbell, E. A. Kellogg et T. A. Dickinson. 2000. The granule-bound starch synthase (GBSSI) gene in the Rosaceae: multiple loci and phylogenetic utility. Mol. Phylog. Evol. 17:388-400.

Excoffier, L., P. Smouse et J. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Favarger, C. 1967. Cytologie et distribution des plantes. Biol. Rev. 42:163-206.

Fernald, M. L. 1922. Notes on the Flora of western Nova Scotia 1921. Rhodora 24:157-180.

Flora of North America Editorial Committee. 1993. Flora of North America, vol. 1. Oxford University Press, New York.

Frohlich, M. W. et E. M. Meyerowitz. 1997. The search for flower homeotic gene homologs in basal angiosperms and Gnetales: a potential new source for data on the evolutionary origin of flowers. International Journal of Plant Sciences 158:131-142.

Funk, D. J. et K. E. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 34:397-423.

Gandoger, M. 1881. Tabulae Rhodologicae Europaeo-Orientales Locupletissimae, Paris, France.

Ghiselin, M. T. 1975. A radical solution to the species problem. Syst. Zool. 23:536-544.

Grant, V. 1981. Plant speciation, 2nd edition. Columbia University Press, New York.

Grossi, C., O. Raymond et M. Jay. 1998. Flavonoid and enzyme polymorphisms and taxonomic organisation of *Rosa* sections: *Carolinae*, *Cinnamomeae*, *Pimpinellifoliae* and *Synstylae*. Biochemical Systematics and Ecology 26:857-871.

Haines, A. et T. F. Vining. 1998. Flora of the Maine: a manual for identification of native and naturalized vascular plants of Maine. V. F. Thomas Co., Bar Harbor, ME.

Hare, M. P. 2001. Prospects for nuclear gene phylogeography. Trends Ecol. Evol. 16:700-706.

Hare, M. P. et J. C. Avise. 1998. Population structure in the american oyster as inferred by nuclear gene genealogies. Mol. Biol. Evol. 15:119-128.

Harlan, J. R. et J. M. J. deWet. 1975. On Ö. Winge and a prayer: the origins of polyploidy. Bot. Rev. 41:361-390.

Haufler, C. H., M. D. Windham, D. M. Britton et S. J. Robinson. 1985. Triploidy and its evolutionary significance in *Cystopteris protrusa*. Can. J. Bot. 63:1855-1863.

Heiser Jr., C. B., J. Soria et D. L. Burton. 1965. A numerical taxonomic study of *Solanum* species and hybrids. Am. Nat. 99:471-488.

Helfgott, D. M. et R. J. Mason-Gamer. 2004. The evolution of North American *Elymus* (Triticeae: Poaceae) allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. Syst. Bot. 29:850-861.

Hinds, H. R. 2000. Flora of New Brunswick, 2nd edition. Unversity of New Brunswick, Fredericton, NB.

Hodkinson, T. R., M. W. Chase, C. Takahashi, I. J. Leitch, M. D. Bennet et S. A. Renvoize. 2002. The use of DNA sequencing (ITS and *trnL-F*), AFLP, and fluorescent in situ hybridization to study allopolyploid *Miscanthus* (Poaceae). Amer. J. Bot. 89:279-286.

Holder, M. T., J. A. Anderson et A. K. Holloway. 2001. Difficulties in detecting hybridization. Systematic Biology 50:978-982.

Holland, B. R., K. T. Huber, V. Moulton et P. J. Lockhart. 2004. Using consensus networks to visualize contradictory evidence for species phylogeny. Mol. Biol. Evol. 21:1459-1461.

Howarth, D. G. 2005. Genealogical evidence of homoploid hybrid speciation in an adaptive radiation of *Scaevola* (Goodeniaceae) in the hawaiian islands. Evolution 59:948-961.

Hudson, R. R. 1992. Gene trees, species trees and the segregation of ancestral alleles. Genetics 131:509-512.

Hudson, R. R. et J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56:1557-1565.

Hull, D. L. 1976. Are species really individuals? Syst. Zool. 25:174-191.

Hull, D. L. 1997. The ideal species concept - and why we can't get it. Pages 357-380 *dans* Species: the units of biodiversity (M. F. Claridge, H. A. Dawah, et M. R. Wilson, éds.). Chapman and Hall, London.

Hunter, J. C. et J. A. Mattice. 2002. The spread of woody exotics into the forests of a northeastern landscape, 1938-1999. J. Torrey Bot. Soc. 129:220-227.

Hurst, C. C. 1925. Chromosomes and characters in *Rosa* and their significance in the origin of species. Experiments in Genetics 37:534-550.

Huson, D. H. et D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254-267.

Huson, D. H., T. Klöpper, P. J. Lockhart et M. A. Steel. 2005. Reconstruction of reticulate networks from gene trees *dans* Recomb 2005:233-249.

Jaccard, P. 1900. Contribution au problème de l'immigration post-glaciaire de la flore alpine. Bull. Soc. Vaudoise Sci. Nat. 36:87-130.

Jakobsen, I. B. et S. Easteal. 1996. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. CABIOS 12:291-295.

Jennings, W. B. et S. V. Edwards. 2005. Speciational history of Australian grass finches (*Poephilla*) inferred from thirty gene trees. Evolution 59:2033-2047.

Jensen, R. J. et W. H. Eshbaugh. 1976. Numerical taxonomic studies of hybridization in *Quercus*. I. Populations of restricted areal distribution and low taxonomic diversity. Syst. Bot. 1:1-10.

Jensen, R. J., S. C. Hokanson, J. G. Isebrands et J. F. Hancock. 1993. Morphometric variation in oaks of the Apostle Islands in Wisconsin: evidence of hybridization between *Quercus rubra* and *Q. ellipsoidalis* (Fagaceae). Amer. J. Bot. 80:1358-1366.

Joly, S. et A. Bruneau. 2004. Evolution of triploidy in *Apios americana* (Leguminosae) revealed by the genealogical analysis of the histone H3-D gene. Evolution 58:284-295.

Joly, S. et A. Bruneau. in press. Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. Systematic Biology.

Joly, S., J. T. Rauscher, S. L. Sherman-Broyles, A. H. D. Brown et J. J. Doyle. 2004. Evolutionary dynamics and preferential expression of homeologous 18S-5.8S-26S nuclear ribosomal genes in natural and artificial *Glycine* allopolyploids. Mol. Biol. Evol. 21:1409-1421.

Joly, S., J. R. Starr, W. H. Lewis et A. Bruneau. 2006. Polyploid and hybrid evolution in roses east of the Rocky Mountains. Amer. J. Bot. 93:412-425.

Judo, M. S. B., A. B. Wedel et C. Wilson. 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res. 26:1819-1825.

Kelchner, S. A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann. Mo. Bot. Gard. 87:482-498.

Kingman, J. F. C. 1982a. The coalscent. Stochastic Processes and their Applications 13:235-248.

Kingman, J. F. C. 1982b. On the genealogy of large populations. J. Appl. Prob. A 19:27-43.

Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38:7-25.

Knops, J. F. et R. J. Jensen. 1980. Morphological and phenolic variation in a three species community of red oaks. Bull. Torrey Bot. Club 107:418-428.

Koufopanou, V., A. Burt et J. W. Taylor. 1997. Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis*. Proc. Natl. Acad. Sci. USA 94:5478-5482.

Kramer, C. Y. 1956. Extension of multiple range tests to group means with unequal numbers of replications. Biometrics 12:309-310.

Krüssmann, G. 1981. The complete book of roses. Timber Press, Portland, OR.

Lapointe, F.-J. 2000. How to account for reticulation events in phylogenetic analysis: a comparison of distance-based methods. Journal of Classification 17:175-184.

Lee, H.-S. et Z. J. Chen. 2001. Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. Proc. Natl. Acad. Sci. USA 98:6753-6758.

Legendre, P. et L. Legendre. 1998. Numerical Ecology, second english edition. Elsevier, Amsterdam.

Legendre, P. et V. Makarenkov. 2002. Reconstruction of biogeographic and evolutionary networks using reticulograms. Systematic Biology 51:199-216.

Levin, D. A. 1983. Polyploidy and novelty in flowering plants. Am. Nat. 122:1-25.

Levin, D. A. 2002. The role of chromosomal change in plant evolution. Oxford University Press, Oxford, UK.

Levin, D. A. 2003. The cytoplasmic factor in plant speciation. Syst. Bot. 28:5-11.

Lewis, C. E. et J. J. Doyle. 2001. Phylogenetic utility of the nuclear gene malate synthase in the palm family (Arecaceae). Mol. Phylog. Evol. 19:409-420.

Lewis, M. S. et C. S. Pikaard. 2001. Restricted chromosomal silencing in nucleolar dominance. Proc. Natl. Acad. Sci. USA 98:14536-14540.

Lewis, W. H. 1957a. An introduction to the genus *Rosa* with special reference to *R. acicularis*. Va. J. Sci. 8:197-202.

Lewis, W. H. 1957b. A monograph of the genus *Rosa* in North America east of the Rocky Mountains. Ph.D. thesis, University of Virginia. 443 pages.

Lewis, W. H. 1957c. Revision of the genus *Rosa* in eastern North America: a review. American Rose Annual 42:116-126.

Lewis, W. H. 1958. A monograph of the genus *Rosa* in North America. II. *R. foliolosa*. Southwest. Nat. 3:145-153.

Lewis, W. H. 1959. A monograph of the genus *Rosa* in North America. I. *R. acicularis*. Brittonia 11:1-24.

Lewis, W. H. 1962. Monograph of the genus *Rosa* in North America. IV. *R.* x *dulcissima*. Brittonia 14:65-71.

Lewis, W. H. 1966. Chromosome numbers of phanerogams. 1. Ann. Mo. Bot. Gard. 53:100-101.

Lewis, W. H. 1980. Polyploidy in species populations. Pages 103-144 *dans* Polyploidy: biological relevance (W. H. Lewis, éd.) Plenum press, New York.

Lewis, W. H. et R. E. Basye. 1961. Analysis of nine crosses between diploid *Rosa* species. Proceedings of the American Society for Horticultural Science 78:573-579.

Linder, C. R. et L. H. Rieseberg. 2004. Reconstructing patterns of reticulate evolution in plants. Amer. J. Bot. 91:1700-1708.

Linnaeus, C. 1753. *Species Plantarum*. Impensis Laurentii Salvii, Stockholm.

Linné, C. 1753. *Species Plantarum*. Impensis Laurentii Salvii, Stockholm.

Liu, B., J. M. Vega et M. Feldman. 1998a. Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. Genome 41:535-542.

Liu, B., J. M. Vega, G. Segal, S. Abbo, M. Rodova et M. Feldman. 1998b. Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy noncoding DNA sequences. Genome 41:272-277.

Luckow, M. 1995. Species concepts: assumptions, methods, and applications. Syst. Bot. 20:589-605.

Maddison, W. P. 1997. Gene trees in species trees. Systematic Biology 46:523-536.

Maddison, W. P. 2005. Coalescence package for Mesquite, version 1.06. http://mesquiteproject.org.

Maddison, W. P. et L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. Systematic Biology 55:21-30.

Maddison, W. P. et D. R. Maddison. 2005. Mesquite: a modular system for evolutionary analysis, version 1.06. http://mesquiteproject.org.

Maechler, M., based on S original by Peter Rousseeuw, Anja.Struyf@uia.ua.ac.be, Mia.Hubert@uia.ua.ac.be and initial R port by Kurt.Hornik@R-project.org. 2005. cluster: Cluster analysis extended Rousseeuw et al., version 1.10.2.

Mallet, J. 1995. A species definition for the modern synthesis. Trends Ecol. Evol. 10:294-299.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.

Mason-Gamer, R. J. 2004. Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. Systematic Biology 53:25-37.

Masterson, J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264:421-424.

Matsumoto, S., M. Kouchi et H. Fukui. 2000. Phylogenetic analyses of the subgenus Eurosa using the ITS nrDNA sequence. Acta Hort. 521:193-202.

Matsumoto, S., M. Kouchi, J. Yabuki, M. Kusunoki, Y. Ueda et H. Fukui. 1998. Phylogenetic analyses of the genus *Rosa* using the *mat*K sequence: molecular evidence for the narrow genetic background of modern roses. Sci. Hortic. 77:73-82.

Maynard Smith, J. 1992. Analysing the mosaic structure of genes. J. Mol. Evol. 34:126-129.

Maynard Smith, J. 1998. Homoplasy test: datain and exph programs, version 3, University of Sussex, Brighton, UK. www.lifesci.sussex.ac.uk/home/John_Maynard_Smith/.

Maynard Smith, J. et N. H. Smith. 1998. Detecting recombination from gene trees. Mol. Biol. Evol. 15:590-599.

Mayr, E. 1942. Systematics and the origin of species from the viewpoint of a zoologist. Columbia University Press, New York.

Mayr, E. 1969. Principles of systematic zoology. McGraw-Hill Book Compagny, New York, USA.

Mayr, E. 2000. The biological species concept. Pages 17-29 *dans* Species concepts and phylogenetic theory (Q. D. Wheeler, et R. Meier, éds.). Columbia University Press, New York.

McDade, L. A. 1990. Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implication for cladistic analysis. Evolution 44:1685-1700.

McDade, L. A. 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. Evolution 46:1329-1346.

McDade, L. A. 1997. Hybrids and phylogenetic systematics. III. Comparison with distance methods. Syst. Bot. 22:669-683.

McLysaght, A., K. Hokamp et K. H. Wolfe. 2002. Extensive genomic duplication during early chordate evolution. Nature Genet. 31:200-204.

Meiners, S. J., S. T. A. Pickett et M. L. Cadenasso. 2001. Effect of plant invasions on the species richness of abandoned agricultural land. Ecography 24:633-644.

Millan, T., F. Osuna, S. Bobos, A. M. Torres et J. I. Cubero. 1996. Using RAPDs to study phylogenetic relationships in *Rosa*. Theor. Appl. Genet. 92:273-277.

Mishler, B. D. et E. C. Theriot. 2000. The phylogenetic species concept (*sensu* Mishler and Theriot): monophyly, apomorphy, and phylogenetic species concepts. Pages 44-54 *dans* Species concepts and phylogenetic theory (Q. D. Wheeler, et R. Meier, éds.). Columbia University Press, New York.

Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49:718-726.

Nakhleh, L., T. Warnow et C. R. Linder. 2004. Reconstructing reticulate evolution in species - theory and practice *dans* Eighth annual international conference on research in computational molecular biology (RECOMB'04):337-346.

Neff, N. A. et G. R. Smith. 1978. Multivariate analysis of hybrid fishes. Syst. Zool. 28:176-196.

Nichols, H. W. 1980. Polyploidy in algae. Pages 151-161 *dans* Polyploidy: biological relevance (W. H. Lewis, éd.) Plenum Press, New York.

Nichols, R. 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16:358-364.

Ohno, S. 1970. Evolution by gene duplication. Springer, New York.

Oksanen, J., R. Kindt, P. Legendre et R. B. O'Hara. 2005. vegan: community ecology package, version 1.7-82. http://cc.oulu.fi/~jarioksa/.

Olsen, K. M. et B. A. Schaal. 1999. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. Proc. Natl. Acad. Sci. USA 96:5586-5591.

Osborn, T. C., J. C. Pires, J. A. Birchler, D. L. Auger, Z. J. Chen, H.-S. Lee, L. Comai, A. Madlung, R. W. Doerge, V. Colot et R. A. Martienssen. 2003. Understanding mechanisms of novel gene expression in polyploids. Trends Genet. 19:141-147.

Otto, S. P. et J. Whitton. 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34:401-437.

Ozkan, H., A. A. Levy et M. Feldman. 2001. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-triticum*) group. Plant Cell 13:1735-1747.

Page, R. D. M. 2001. GeneTree, version 1.0. University of Glasgow, Glasgow, Scotland. http://taxonomy.zoology.gla.ac.uk/rod/genetree/genetree.html.

Page, R. D. M. et M. A. Charleston. 1997. From gene to organismal phylogeny: reconciled trees and the gene tree / species tree problem. Mol. Phylog. Evol. 7:231-240.

Pamilo, P. et M. Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568-583.

Paterson, H. E. H. 1985. The recognition concept of species. Pages 21-29 *dans* Species and speciation (E. S. Vrba, éd.) Transvaal Museum, Pretoria.

Peres-Neto, P., P. Legendre, S. Dray et D. Borcard. in press. Variation partitioning od species data matrices: estimation and comparison of fractions. Ecology.

Perný, M., A. Tribsch, T. F. Stuessy et K. Marhold. 2005. Allopolyploid origin of *Cardamine silana* (Brassicaceae) from Calabria (southern Italy): karyological, morphological and molecular evidence. Bot. J. Linn. Soc. 148:101-116.

Petersen, G. et O. Seberg. 2004. On the origin of the tetraploid species *Hordeum capense* and *H. secalinum* (Poaceae). Syst. Bot. 29:862-873.

Petit, C. et J. D. Thompson. 1999. Species diversity and ecological range in relation to ploidy level in the flora of the Pyrenees. Evol. Ecol. 13:45-66.

Pfeil, B. E., C. L. Brubaker, L. A. Craven et M. D. Crisp. 2004. Paralogy and orthology in the Malvaceae *rpb*2 gene family: investigation of gene duplication in *Hibiscus*. Mol. Biol. Evol. 21:1428-1437.

Pikaard, C. S. 2001. Genomic change and gene silencing in polyploids. Trends Genet. 17:675-677.

Pontes, O., R. J. Lawrence, N. Neves, M. Silva, J.-H. Lee, Z. J. Chen, W. Viegas et C. S. Pikaard. 2003. Natural variation in nucleolar dominance reveals the relationship between nucleolus organizer chromatin topology and rRNA gene transcription in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 100:11418-11423.

Posada, D. 2002. Evaluating methods for detecting recombination from DNA sequences: empirical data. Mol. Biol. Evol. 19:708-717.

Posada, D. et K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817-818.

Posada, D. et K. A. Crandall. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA 98:13757-13762.

Puorto, G., M. da Graça Salomao, R. D. G. Theakston, R. S. Thorpe, D. A. Warrell et W. Wüster. 2001. Combining mitochondrial DNA sequences and morphological data to infer species boundaries: phylogeography of lanceheaded pitvipers in the Brazilian Atlantic forest, and the status of *Bothrops pradoi* (Squamata: Serpentes: Viperidae). J. Evol. Biol. 14:527-538.

R Development Core Team. 2005. R: A language and environment for statistical computing, version 2.2.1. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ramon, S. 1968. A numerical taxonomic study of certain taxa of *Haplopappus*, section *Blepharodon*. Univ. Kansas Sci. Bull. 47:863-900.

Ramsey, J. et D. W. Schemske. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu. Rev. Ecol. Syst. 29:467-501.

Ramsey, J. et D. W. Schemske. 2002. Neopolyploidy in flowering plants. Annu. Rev. Ecol. Syst. 33:589-639.

Ratsek, J. C., W. S. Flory Jr. et S. H. Yarnell. 1940. Crossing relations of some diploid and polyploid species of roses. Proceedings of the American Society for Horticultural Science 38:637-654.

Ratsek, J. C., S. H. Yarnell et W. S. Flory, Jr. 1939. Crossing relations of some diploid species of roses. Proceedings of the American Society for Horticultural Science 37:983-992.

Rauscher, J. T., J. J. Doyle et A. H. D. Brown. 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. Mol. Ecol. 11:2691-2702.

Rehder, A. 1940. Manual of cultivated trees and shrubs hardy in North America, 2nd ed edition. Macmillan, New York, USA.

Rhoads, A. F. et W. M. Klein Jr. 1993. The vascular flora of Pennsylvania : annotated checklist and atlas. American Philosophical Society, Philadelphia, PA.

Riddle, N. C. et J. A. Birchler. 2003. Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. Trends Ecol. Evol. 19:597-600.

Rieseberg, L. H. 1997. Hybrid origins of plant species. Annu. Rev. Ecol. Syst. 28:359-389.

Rieseberg, L. H. 1998. Molecular ecology of hybridization. Pages 243-265 *dans* Advances in Molecular Ecology (G. R. Carvalho, éd.) IOS Press, Burke, VA.

Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan et C. Lexer. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211-1216.

Ripley, B. 2005. tree: classification and regression trees, version 1.0-20.

Ritz, C. M., H. Schmuths et V. Wissemann. 2005. Evolution by reticulation: european dogroses originated by multiple hybridization across the genus *Rosa*. J. Hered. 96:4-14.

Robertson, K. R. 1974. The genera of Rosaceae in the southeastern United States: tribe Roseae. Journal of the Arnold Arboretum 55:611-621.

Rokas, A., B. L. Williams, N. King et S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798-803.

Rosenberg, N. A. 2002. The probability of topological concordance of gene trees and species trees. Theor. Popul. Biol. 61:225-247.

Rosenberg, N. A. 2003. The shape of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution 57:1465-1477.

Rosenthal, D. M., A. E. Schwarzbach, L. A. Donovan, O. Raymond et L. H. Rieseberg. 2002. Phenotypic differentiation between three ancient hybrid taxa and their parental species. Int. J. Plant Sci. 163:387-398.

Rowley, G. D. 1976. Typification of the genus *Rosa* L. Taxon 25:181.

Rydberg, A. 1920. Roses of northeastern America. Bull. Torrey Bot. Club 47:45-66.

Sanderson, M. J., A. Purvis et C. Henze. 1998. Phylogenetic supertrees: assembling the trees of life. Trends Ecol. Evol. 13:105-109.

Sang, T. et Y. Zhong. 2000. Testing hybridization hypotheses based on incongruent gene trees. Systematic Biology 49:422-434.

Schaal, B. A. et K. M. Olsen. 2000. Gene genealogies and population variation in plants. Proc. Natl. Acad. Sci. USA 97:7024-7029.

Schilling, E. E. et C. B. Heiser Jr. 1976. Re-examination of a numerical taxonomic study of *Solanum* species and hybrids. Taxon 25:451-462.

Schneider, S., D. Roessli et L. Excoffier. 2000. Arlequin: a software for population genetics data analysis, version 2. Genetic and Biometry Laboratory, University of Geneva, Switzerland.

Scoggan, H. J. 1978. The flora of Canada. National Museums of Canada, Ottawa, Canada.

Segraves, K. A., J. N. Thompson, P. S. Soltis et D. E. Soltis. 1999. Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia*. Mol. Ecol. 8:253-262.

Senchina, D. S., I. Alvarez, R. C. Cronn, B. Liu, J. Rong, R. D. Noyes, A. H. Paterson, R. A. Wing, T. A. Wilkins et J. F. Wendel. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol. Biol. Evol. 20:633-643.

Seo, T.-K., H. Kishino et J. L. Thorne. 2005. Incorporating gene-specific variation when inferring and evaluating optimal evolutionary tree topologies from multilocus sequence data. Proc. Natl. Acad. Sci. USA 102:4436-4441.

Shaked, H., K. Kashkush, H. Ozkan, M. Feldman et A. A. Levy. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 13:1749-1759.

Sharbel, T. F. et T. Mitchell-Olds. 2001. Recurrent polyploid origins and chloroplast phylogeography in the *Arabis holboellii* complex (Brassicaceae). Heredity 87:59-68.

Simillion, C., K. Vandepoele, M. C. E. Van Montagu, M. Zabeau et Y. Van de Peer. 2003. The hidden duplication past of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 99:13627-13632.

Simmons, M. P. et H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. Systematic Biology 49:369-381.

Sites Jr., J. W. et J. C. Marshall. 2003. Delimiting species: a renaissance issue in systematic bioloby. Trends Ecol. Evol. 18:462-470.

Sites Jr., J. W. et J. C. Marshall. 2004. Operational criteria for delimiting species. Annu. Rev. Ecol. Evol. Syst. 35:199-227.

Slowinski, J. B., A. Knight et A. P. Rooney. 1997. Inferring species trees from gene trees: a phylogenetic analysis of the Elapidae (Serpentes) based on the amino acid sequences of venom proteins. Mol. Phylog. Evol. 8:349-362.

Smedmark, J. E. E., T. Eriksson, R. C. Evans et C. S. Campbell. 2003. Ancient allopolyploid speciation in Geinae (Rosaceae): evidence from nuclear granule-bound starch synthase (GBSSI) gene sequences. Systematic Biology 52:374-385.

Sneath, P. H. A. et R. R. Sokal. 1973. Numerical Taxonomy. W. H. Freeman and Compagny, San Francisco.

Soltis, D. E. et P. S. Soltis. 1993. Molecular data and the dynamic nature of polyploidy. Crit. Rev. Plant Sci. 12:243-273.

Soltis, D. E. et P. S. Soltis. 1995. The dynamic nature of polyploid genomes. Proc. Natl. Acad. Sci. USA 92:8089-8091.

Soltis, D. E. et P. S. Soltis. 1999. Polyploidy: recurrent formation and genome evolution. Trends Ecol. Evol. 14:348-352.

Soltis, D. E., P. S. Soltis et B. D. Ness. 1989. Chloroplast-DNA variation and multiple origins of autopolyploidy in *Heuchera micrantha* (Saxifragaceae). Evolution 43:650-656.

Soltis, P. S. et D. E. Soltis. 2000. The role of genetic and genomic attributes in the success of polyploids. Proc. Natl. Acad. Sci. USA 97:7051-7057.

Song, K., P. Lu, K. Tang et T. C. Osborn. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc. Natl. Acad. Sci. USA 92:7719-7723.

Stearn, W. T. 1957. An introduction to the *Species Plantarum* and cognate botanical works of Carl Linnaeus, Carl Linnaeus - *Species Plantarum* - A facsimile of the first edition 1753. The Ray Society, London.

Stebbins, G. L. 1950. Variation and evolution in plants. Columbia University Press, New York.

Stebbins, G. L. 1971. Chromosomal evolution in higher plants. Edward Arnold, London.

Stebbins, G. L. 1980. Polyploidy in plants: unsolved problems and prospects. Pages 495-520 *dans* Polyploidy: biological relevance (W. H. Lewis, éd.) Plenum Press, New York.

Strand, A. E., J. Leebens-Mack et B. G. Milligan. 1997. Nuclear DNA-based markers for plant evolutionary biology. Mol. Ecol. 6:113-118.

Suda, J. et M. A. Lysák. 2001. A taxonomic study of the *Vaccinium* sect. *oxycoccus* (Hill) W.D.J. Koch (Ericaceae) in the Czech Republic and adjacent territories. Folia Geobot. 36:303-320.

Swofford, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer Associates, Sunderland, MA.

Täckholm, G. 1922. Zytologische studien über die Gattung Rosa. Acta Horti Bergiana 7:97-381.

Takahata, N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics 122:957-966.

Takamiya, M., N. Ohta, Y. Yatabe et N. Murakami. 2001. Cytological, morphological, genetic, and molecular phylogenetic studies on intraspecific differentiations within *Diplazium doederleinii* (Woodsiaceae: Pteridophyta). Int. J. Plant Sci. 162:625-636.

Tal, M. 1980. Physiology of polyploids. Pages 61-75 *dans* Polyploidy: biological relevance (W. H. Lewis, éd.) Plenum Press, New York, USA.

Templeton, A. R. 1989. The meaning of species and speciation: a genetic perspective. Pages 3-27 *dans* Speciation and its consequences (D. Otte, et J. A. Endler, éds.). Sinauer Associates, Sunderland, MA.

Templeton, A. R. 2001. Using phylogeographic analyses of gene trees to test species status and processes. Mol. Ecol. 10:779-791.

Templeton, A. R. 2002. Out of Africa again and again. Nature 416:45-51.

Templeton, A. R., K. A. Crandall et C. F. Sing. 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.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin et D. G. Higgins. 1997a. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24:4876-4882.

Thompson, J. D., D. G. Higgins et T. J. Gibson. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.

Thompson, J. N., B. M. Cunningham, K. A. Segraves, D. M. Althoff et D. Wagner. 1997b. Plant polyploidy and insect/plant interactions. Am. Nat. 150:730-743.

Tukey, J. 1953. A problem of multiple comparisons. Pages 396, Princeton University.

Tutin, T. G., V. H. Heywood, N. A. Burges, D. M. Moore, D. H. Valentine, S. M. Walters et D. A. Webb. 1968. *Flora europaea*. Cambridge University Press, London, UK.

Vamosi, J. C. et T. A. Dickinson. 2006. Polyploidy and diversification: a phylogenetic investigation in Rosaceae. Int. J. Plant Sci. 167:349-358.

Van Dijk, P. et T. Bakx-Schotman. 1997. Chloroplast DNA phylogeography and cytotype geography in autopolyploid *Plantago media*. Mol. Ecol. 9:345-352.

Vanderhoeven, S., O. Hardy, X. Vekemens, C. Lefèvre, M. de Loose, J. Lambinon et P. Meerts. 2002. A morphometric study of populations of the *Centaurea jacea* complex (Asteraceae) in Belgium. Plant Biol. 4:403-412.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van De Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper et M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414.

Wagner, A., N. Blackstone, P. Cartwright, M. Dick, B. Misof, P. Snow, G. P. Wagner, J. Bartels, M. Murtha et J. Pendleton. 1994. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. Systematic Biology 43:250-261.

Wagner, W. H., Jr. et F. S. Wagner. 1980. Polyploidy in pteridophytes. Pages 199-214 *dans* Polyploidy: biological relevance (W. H. Lewis, éd.) Plenum Press, New York.

Ward, J. H. 1963. Hierarchical grouping to optmize an objective function. J. Am. Stat. Assoc. 58:236-244.

Watson, S. 1885. Contributions to American botany. I. A history and revision of the roses of North America. Proceedings of the American Academy of Arts and Sciences 20:324-352.

Wendel, J. F. 2000. Genome evolution in polyploids. Plant Mol. Biol. 42:225-249.

Wheeler, Q. D. et N. I. Platnick. 2000. The phylogenetic species concept. Pages 55-69 *dans* Species concepts and phylogenetic theory (Q. D. Wheeler, et R. Meier, éds.). Columbia University Press, New York.

Whiffin, T. 1973. Analysis of a hybrid swarm between *Heterocentron elegans* and *H. glandulosum*. Taxon 22:413-423.

Wiens, J. J. et M. R. Servedio. 2000. Species delimitation in systematics: inferring diagnostic differences between species. Proc. R. Soc. Lond. B 267:631-636.

Wiley, E. O. et R. L. Mayden. 2000. The evolutionary species concept. Pages 70-89 *dans* Species concepts and phylogenetic theory (Q. D. Wheeler, et R. Meier, éds.). Columbia University Press, New York.

Wilkinson, M., J. A. Cotton, C. Creevey, O. Eulenstein, S. R. Harris, F.-J. Lapointe, C. Levasseur, J. O. Mcinerney, D. Pisani et J. L. Thorley. 2005. The shape of supertrees to come: tree shape related properties of fourteen supertree methods. Systematic Biology 54:419-431.

Wissemann, V. 2003. Conventional taxonomy (wild roses). Pages 111-117 *dans* Encyclopedia of rose science (A. Roberts, T. Debener, et S. Gudin, éds.). Elsevier, London.

Wissemann, V. et C. M. Ritz. 2005. The genus *Rosa* (Rosoideae, Rosaceae) revisited: molecular analysis of nrITS-1 and *atp*B-*rbc*L intergenic spacer (IGS) versus conventional taxonomy. Bot. J. Linn. Soc. 147:275-290.

Wiuf, C., T. Christensen et J. Hein. 2001. A simulation study of the reliability of recombination detection methods. Mol. Biol. Evol. 18:1929-1939.

Wollenberg, K. et J. C. Avise. 1999. Sampling properties of genealogical pathways underlying population pedigrees. Evolution 52:957-966.

Wu, C.-I. 1991. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. Genetics 127:429-435.

Xu, S. 2000. Phylogenetic analysis under reticulate evolution. Mol. Biol. Evol. 17:897-907.

Yang, Z. 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42:587-596.

Young, N. D. et J. Healy. 2003. GapCoder automates the use of indel characters in phylogenetic analysis. BMC Bioinformatics 4:6.

Ziolkowski, P. A., G. Blanc et J. Sadowski. 2003. Structural divergence of chromosomal segments that arose from successive duplication events in the *Arabidopsis* genome. Nucleic Acids Res. 31:1339-1350.

Annexes

Annexe 1

Matrice des corrélations usitlisée pour l'analyse en composantes pricipales des individus diploides et polyploides (voir Fig. 2.3)

	NLFT	NSER	N2SER	LLFT	L1SER	WLFT	LWLFT	LTEET	PULFT	P1VEIN	P2VEIN	PBLFT	G1VEIN
NLFT	1,0000	-0,3029	0,0221	-0,4654	-0,4695	-0,5365	-0,4810	-0,1482	0,0131	-0,3795	-0,2680	-0,1471	-0,1103
NSER	-0,3029	1,0000	0,0293	0,5903	-0,1111	0,3651	0,4454	-0,2518	-0,1308	0,1393	0,0828	-0,1356	-0,1881
N2SER	0,0221	0,0293	1,0000	-0,1900	-0,3072	-0,0552	-0,1567	0,1210	-0,1371	-0,1227	-0,0051	0,0090	0,2613
LLFT	-0,4654	0,5903	-0,1900	1,0000	0,5498	0,8427	0,9445	0,4113	0,0502	0,4012	0,3355	0,0363	-0,2553
L1SER	-0,4695	-0,1111	-0,3072	0,5498	1,0000	0,6702	0,6608	0,5686	0,2912	0,5464	0,4641	0,3155	0,0174
WLFT	-0,5365	0,3651	-0,0552	0,8427	0,6702	1,0000	0,8606	0,6306	0,2926	0,5879	0,5171	0,3050	-0,0491
LWLFT	-0,4810	0,4454	-0,1567	0,9445	0,6608	0,8606	1,0000	0,5039	0,1190	0,4775	0,4241	0,1112	-0,1568
LTEET	-0,1482	-0,2518	0,1210	0,4113	0,5686	0,6306	0,5039	1,0000	0,3226	0,3861	0,4287	0,3439	0,0778
PULFT	0,0131	-0,1308	-0,1371	0,0502	0,2912	0,2926	0,1190	0,3226	1,0000	0,6464	0,6592	0,7610	-0,0402
P1VEIN	-0,3795	0,1393	-0,1227	0,4012	0,5464	0,5879	0,4775	0,3861	0,6464	1,0000	0,8808	0,6929	-0,0480
P2VEIN	-0,2680	0,0828	-0,0051	0,3355	0,4641	0,5171	0,4241	0,4287	0,6592	0,8808	1,0000	0,7365	-0,0324
PBLFT	-0,1471	-0,1356	0,0090	0,0363	0,3155	0,3050	0,1112	0,3439	0,7610	0,6929	0,7365	1,0000	-0,0229
G1VEIN	-0,1103	-0,1881	0,2613	-0,2553	0,0174	-0,0491	-0,1568	0,0778	-0,0402	-0,0480	-0,0324	-0,0229	1,0000
GBLFT	-0,0646	-0,1312	-0,0447	-0,1662	0,0335	0,0308	-0,1116	0,0362	0,0884	0,1128	0,0997	0,1365	0,5717
B1YW	0,1257	-0,0124	0,2906	-0,2280	-0,3410	-0,1963	-0,3016	-0,2140	-0,2051	-0,3148	-0,2790	-0,2465	-0,0704
T1YW	0,1853	0,2423	0,0243	-0,1948	-0,4575	-0,3223	-0,2527	-0,4352	-0,1467	-0,2365	-0,3231	-0,2092	-0,0062
LSTP	-0,3120	0,5630	-0,0086	0,7904	0,3455	0,7134	0,7618	0,3111	0,1578	0,4189	0,3865	0,1487	-0,2725
LAUR	-0,1175	0,1438	0,2196	0,5016	0,2361	0,5903	0,5125	0,5144	0,1631	0,2919	0,3184	0,1785	-0,1628
WAUR	-0,0757	0,0877	0,0836	0,5346	0,3686	0,5982	0,5738	0,5052	0,2304	0,3485	0,3620	0,2319	-0,1778
WSTP	-0,1534	-0,0516	0,1136	0,3605	0,3418	0,5169	0,4056	0,5106	0,2926	0,3330	0,3881	0,3302	-0,0516
GSTP	0,0983	-0,1014	0,3077	-0,1161	0,0241	0,0127	-0,0401	0,2029	0,1225	0,0071	0,1246	0,0458	0,4800
GHYP	-0,1482	0,5343	-0,1684	0,1775	-0,2256	-0,1095	0,0717	-0,5497	-0,3144	-0,1795	-0,3114	-0,3105	-0,2275
GPED	-0,0175	0,3222	0,1363	-0,0060	-0,3893	-0,1514	-0,0716	-0,4276	-0,3343	-0,3206	-0,3684	-0,3481	-0,1197
LPED	-0,3116	0,0239	0,1006	0,3036	0,3374	0,4572	0,3190	0,3557	0,1256	0,2024	0,1615	0,1404	0,0108
NFLW	-0,0863	0,3020	-0,1491	0,3638	0,2254	0,3430	0,3384	0,1460	0,1342	0,2534	0,2344	0,1882	-0,0462

	GBLFT	B1YW	T1YW	LSTP	LAUR	WAUR	WSTP	GSTP	GHYP	GPED	LPED	NFLW
NLFT	-0,0646	0,1257	0,1853	-0,3120	-0,1175	-0,0757	-0,1534	0,0983	-0,1482	-0,0175	-0,3116	-0,0863
NSER	-0,1312	-0,0124	0,2423	0,5630	0,1438	0,0877	-0,0516	-0,1014	0,5343	0,3222	0,0239	0,3020
N2SER	-0,0447	0,2906	0,0243	-0,0086	0,2196	0,0836	0,1136	0,3077	-0,1684	0,1363	0,1006	-0,1491
LLFT	-0,1662	-0,2280	-0,1948	0,7904	0,5016	0,5346	0,3605	-0,1161	0,1775	-0,0060	0,3036	0,3638
L1SER	0,0335	-0,3410	-0,4575	0,3455	0,2361	0,3686	0,3418	0,0241	-0,2256	-0,3893	0,3374	0,2254
WLFT	0,0308	-0,1963	-0,3223	0,7134	0,5903	0,5982	0,5169	0,0127	-0,1095	-0,1514	0,4572	0,3430
LWLFT	-0,1116	-0,3016	-0,2527	0,7618	0,5125	0,5738	0,4056	-0,0401	0,0717	-0,0716	0,3190	0,3384
LTEET	0,0362	-0,2140	-0,4352	0,3111	0,5144	0,5052	0,5106	0,2029	-0,5497	-0,4276	0,3557	0,1460
PULFT	0,0884	-0,2051	-0,1467	0,1578	0,1631	0,2304	0,2926	0,1225	-0,3144	-0,3343	0,1256	0,1342
P1VEIN	0,1128	-0,3148	-0,2365	0,4189	0,2919	0,3485	0,3330	0,0071	-0,1795	-0,3206	0,2024	0,2534
P2VEIN	0,0997	-0,2790	-0,3231	0,3865	0,3184	0,3620	0,3881	0,1246	-0,3114	-0,3684	0,1615	0,2344
PBLFT	0,1365	-0,2465	-0,2092	0,1487	0,1785	0,2319	0,3302	0,0458	-0,3105	-0,3481	0,1404	0,1882
G1VEIN	0,5717	-0,0704	-0,0062	-0,2725	-0,1628	-0,1778	-0,0516	0,4800	-0,2275	-0,1197	0,0108	-0,0462
GBLFT	1,0000	-0,0728	-0,0540	-0,1681	-0,1210	-0,1445	-0,0554	0,1967	-0,1758	-0,1373	0,0131	0,0429
B1YW	-0,0728	1,0000	-0,0170	-0,1119	0,2129	0,0501	0,1512	0,0083	0,0455	0,4781	0,2983	-0,1589
T1YW	-0,0540	-0,0170	1,0000	-0,1059	-0,2372	-0,2152	-0,3219	-0,1594	0,4038	0,2882	-0,2610	-0,0235
LSTP	-0,1681	-0,1119	-0,1059	1,0000	0,6728	0,6998	0,5359	-0,0696	0,0876	-0,0160	0,2854	0,3168
LAUR	-0,1210	0,2129	-0,2372	0,6728	1,0000	0,8994	0,7879	0,0615	-0,2776	-0,0283	0,4374	0,2296
WAUR	-0,1445	0,0501	-0,2152	0,6998	0,8994	1,0000	0,7880	0,0406	-0,2805	-0,1177	0,3708	0,2262
WSTP	-0,0554	0,1512	-0,3219	0,5359	0,7879	0,7880	1,0000	0,0958	-0,3978	-0,1339	0,4304	0,2240
GSTP	0,1967	0,0083	-0,1594	-0,0696	0,0615	0,0406	0,0958	1,0000	-0,2397	-0,0734	0,0931	0,0352
GHYP	-0,1758	0,0455	0,4038	0,0876	-0,2776	-0,2805	-0,3978	-0,2397	1,0000	0,5560	-0,2642	0,0450
GPED	-0,1373	0,4781	0,2882	-0,0160	-0,0283	-0,1177	-0,1339	-0,0734	0,5560	1,0000	0,1047	-0,0939
LPED	0,0131	0,2983	-0,2610	0,2854	0,4374	0,3708	0,4304	0,0931	-0,2642	0,1047	1,0000	-0,0078
NFLW	0,0429	-0,1589	-0,0235	0,3168	0,2296	0,2262	0,2240	0,0352	0,0450	-0,0939	-0,0078	1,0000

Annexe 2

Projection des vecteurs de caractères sur les trois premiers axes de l'analyse en composantes principales des individus diploïdes (voir Fig. 2.3).

	PC1	PC2	PC3
NLFT	0,159041357	0,11947808	-0,14123200
NSER	-0,073978023	-0,38871645	0,12719911
N2SER	0,022529169	0,04720675	-0,34490698
LLFT	-0,273711656	-0,28420965	0,08471353
L1SER	-0,259040775	0,07368759	0,15212640
WLFT	-0,323620150	-0,10474560	0,02491336
LWLFT	-0,295160251	-0,20904109	0,08364899
LTEET	-0,252044945	0,17719253	-0,12758567
PULFT	-0,170198835	0,25129392	0,14964192
P1VEIN	-0,263147307	0,11644278	0,23227666
P2VEIN	-0,257361008	0,17329344	0,16502519
PBLFT	-0,183058599	0,26071778	0,15384496
G1VEIN	0,040244443	0,22647015	-0,05126395
GBLFT	0,006554349	0,20689191	0,07168567
B1YW	0,084842517	-0,09499894	-0,41710326
T1YW	0,153527512	-0,16582356	0,09426939
LSTP	-0,268775727	-0,24811106	-0,03976743
LAUR	-0,252470490	-0,08436425	-0,34993764
WAUR	-0,267992128	-0,06373933	-0,26923789
WSTP	-0,249147320	0,03642573	-0,30315073
GSTP	-0,020829967	0,18926184	-0,16432991
GHYP	0,106628963	-0,37712852	0,21342789
GPED	0,114357661	-0,30947164	-0,15885046
LPED	-0,166466963	-0,01689472	-0,26744092
NFLW	-0,135903808	-0,08140520	0,12280803

Note : une variable contribute significativement à une composante principale si sa projection sur celle-ci est plus grande que 0.2.

Annexe 3

Matrice des corrélations usitlisée pour l'analyse en composantes pricipales des individus diploides et polyploides (voir Fig. 2.7)

	NLFT	NSER	N2SER	LLFT	L1SER	WLFT	LWLFT	LTEET	PULFT	P1VEIN	P2VEIN	PBLFT	G1VEIN
NLFT	1,0000	-0,2506	-0,1628	-0,3979	-0,2858	-0,3758	-0,3914	-0,0682	0,0201	-0,0267	-0,0316	-0,0287	-0,0900
NSER	-0,2506	1,0000	0,2221	0,5805	-0,1567	0,3095	0,4384	-0,2104	-0,0578	0,0233	0,0312	-0,0867	-0,1473
N2SER	-0,1628	0,2221	1,0000	0,0541	-0,2810	0,0990	0,0069	0,1874	-0,1373	-0,2464	-0,1745	-0,1218	0,0929
LLFT	-0,3979	0,5805	0,0541	1,0000	0,4851	0,7983	0,9289	0,4231	0,0248	0,1860	0,1923	-0,0059	-0,2181
L1SER	-0,2858	-0,1567	-0,2810	0,4851	1,0000	0,5800	0,6127	0,4691	0,2285	0,4222	0,3848	0,2466	0,0095
WLFT	-0,3758	0,3095	0,0990	0,7983	0,5800	1,0000	0,8064	0,6676	0,1783	0,3901	0,3689	0,1932	-0,0660
LWLFT	-0,3914	0,4384	0,0069	0,9289	0,6127	0,8064	1,0000	0,4838	0,0984	0,3104	0,3253	0,0884	-0,1434
LTEET	-0,0682	-0,2104	0,1874	0,4231	0,4691	0,6676	0,4838	1,0000	0,1530	0,2461	0,2730	0,1735	0,0315
PULFT	0,0201	-0,0578	-0,1373	0,0248	0,2285	0,1783	0,0984	0,1530	1,0000	0,5179	0,6026	0,7539	-0,0109
P1VEIN	-0,0267	0,0233	-0,2464	0,1860	0,4222	0,3901	0,3104	0,2461	0,5179	1,0000	0,8713	0,6533	0,0295
P2VEIN	-0,0316	0,0312	-0,1745	0,1923	0,3848	0,3689	0,3253	0,2730	0,6026	0,8713	1,0000	0,7488	0,0098
PBLFT	-0,0287	-0,0867	-0,1218	-0,0059	0,2466	0,1932	0,0884	0,1735	0,7539	0,6533	0,7488	1,0000	0,0225
G1VEIN	-0,0900	-0,1473	0,0929	-0,2181	0,0095	-0,0660	-0,1434	0,0315	-0,0109	0,0295	0,0098	0,0225	1,0000
GBLFT	-0,0613	-0,0928	-0,0414	-0,1287	0,0158	0,0148	-0,0959	0,0079	0,1049	0,1129	0,1125	0,1650	0,5944
B1YW	0,2044	0,0090	0,1179	-0,1986	-0,2931	-0,1725	-0,2582	-0,1614	-0,1632	-0,1260	-0,1554	-0,1584	-0,0601
T1YW	-0,1048	0,2346	0,1900	-0,0537	-0,3419	-0,2183	-0,1591	-0,3202	-0,1378	-0,3872	-0,4111	-0,2644	-0,0503
LSTP	0,0682	0,4114	-0,0288	0,5905	0,3038	0,5682	0,5706	0,2768	0,1011	0,3286	0,2873	0,1040	-0,1866
LAUR	0,1214	0,1313	0,0699	0,3988	0,2301	0,4942	0,4039	0,4314	0,1194	0,2120	0,2296	0,1293	-0,1078
WAUR	0,1785	0,0416	-0,0220	0,3926	0,3542	0,5170	0,4313	0,4568	0,1411	0,2884	0,2688	0,1536	-0,1291
WSTP	0,1576	-0,0604	-0,0450	0,2303	0,2917	0,4108	0,2784	0,3753	0,2122	0,2836	0,3045	0,2597	-0,0247
GSTP	0,0823	-0,0240	0,2269	-0,0569	-0,0198	0,0085	0,0023	0,1749	0,1314	0,0310	0,1302	0,0846	0,4646
GHYP	-0,1561	0,4917	-0,1150	0,1332	-0,2110	-0,1643	0,0530	-0,5405	-0,2010	-0,1563	-0,2387	-0,2103	-0,1823
GPED	-0,1869	0,2362	0,1373	0,0130	-0,2679	-0,1025	-0,0457	-0,3225	-0,2556	-0,2855	-0,3058	-0,2839	-0,1180
LPED	-0,1272	0,0155	0,0654	0,2742	0,3413	0,4223	0,2994	0,3250	0,0731	0,2094	0,1639	0,1054	0,0168
NFLW	0,1849	0,1257	-0,1734	0,2128	0,1692	0,2759	0,2104	0,1463	0,0491	0,2787	0,2284	0,1563	-0,0610

	GBLFT	B1YW	T1YW	LSTP	LAUR	WAUR	WSTP	GSTP	GHYP	GPED	LPED	NFLW
NLFT	-0,0613	0,2044	-0,1048	0,0682	0,1214	0,1785	0,1576	0,0823	-0,1561	-0,1869	-0,1272	0,1849
NSER	-0,0928	0,0090	0,2346	0,4114	0,1313	0,0416	-0,0604	-0,0240	0,4917	0,2362	0,0155	0,1257
N2SER	-0,0414	0,1179	0,1900	-0,0288	0,0699	-0,0220	-0,0450	0,2269	-0,1150	0,1373	0,0654	-0,1734
LLFT	-0,1287	-0,1986	-0,0537	0,5905	0,3988	0,3926	0,2303	-0,0569	0,1332	0,0130	0,2742	0,2128
L1SER	0,0158	-0,2931	-0,3419	0,3038	0,2301	0,3542	0,2917	-0,0198	-0,2110	-0,2679	0,3413	0,1692
WLFT	0,0148	-0,1725	-0,2183	0,5682	0,4942	0,5170	0,4108	0,0085	-0,1643	-0,1025	0,4223	0,2759
LWLFT	-0,0959	-0,2582	-0,1591	0,5706	0,4039	0,4313	0,2784	0,0023	0,0530	-0,0457	0,2994	0,2104
LTEET	0,0079	-0,1614	-0,3202	0,2768	0,4314	0,4568	0,3753	0,1749	-0,5405	-0,3225	0,3250	0,1463
PULFT	0,1049	-0,1632	-0,1378	0,1011	0,1194	0,1411	0,2122	0,1314	-0,2010	-0,2556	0,0731	0,0491
P1VEIN	0,1129	-0,1260	-0,3872	0,3286	0,2120	0,2884	0,2836	0,0310	-0,1563	-0,2855	0,2094	0,2787
P2VEIN	0,1125	-0,1554	-0,4111	0,2873	0,2296	0,2688	0,3045	0,1302	-0,2387	-0,3058	0,1639	0,2284
PBLFT	0,1650	-0,1584	-0,2644	0,1040	0,1293	0,1536	0,2597	0,0846	-0,2103	-0,2839	0,1054	0,1563
G1VEIN	0,5944	-0,0601	-0,0503	-0,1866	-0,1078	-0,1291	-0,0247	0,4646	-0,1823	-0,1180	0,0168	-0,0610
GBLFT	1,0000	-0,0646	-0,0747	-0,1231	-0,0823	-0,1208	-0,0421	0,2230	-0,1401	-0,1204	0,0076	0,0055
B1YW	-0,0646	1,0000	-0,1151	-0,0460	0,1897	0,0796	0,1657	0,0503	0,0395	0,2947	0,2763	-0,0284
T1YW	-0,0747	-0,1151	1,0000	-0,1698	-0,2495	-0,2610	-0,3359	-0,1552	0,3320	0,3156	-0,2054	-0,1676
LSTP	-0,1231	-0,0460	-0,1698	1,0000	0,7188	0,7190	0,6003	-0,0422	0,0073	-0,1523	0,3230	0,3737
LAUR	-0,0823	0,1897	-0,2495	0,7188	1,0000	0,8957	0,7829	0,0657	-0,2597	-0,1279	0,3828	0,2610
WAUR	-0,1208	0,0796	-0,2610	0,7190	0,8957	1,0000	0,7930	0,0297	-0,2931	-0,2125	0,3611	0,2990
WSTP	-0,0421	0,1657	-0,3359	0,6003	0,7829	0,7930	1,0000	0,0597	-0,3400	-0,2279	0,3825	0,3337
GSTP	0,2230	0,0503	-0,1552	-0,0422	0,0657	0,0297	0,0597	1,0000	-0,1684	-0,0375	0,0917	-0,0084
GHYP	-0,1401	0,0395	0,3320	0,0073	-0,2597	-0,2931	-0,3400	-0,1684	1,0000	0,5005	-0,2285	-0,0807
GPED	-0,1204	0,2947	0,3156	-0,1523	-0,1279	-0,2125	-0,2279	-0,0375	0,5005	1,0000	0,0485	-0,2086
LPED	0,0076	0,2763	-0,2054	0,3230	0,3828	0,3611	0,3825	0,0917	-0,2285	0,0485	1,0000	0,0954
NFLW	0,0055	-0,0284	-0,1676	0,3737	0,2610	0,2990	0,3337	-0,0084	-0,0807	-0,2086	0,0954	1,0000

Annexe 4

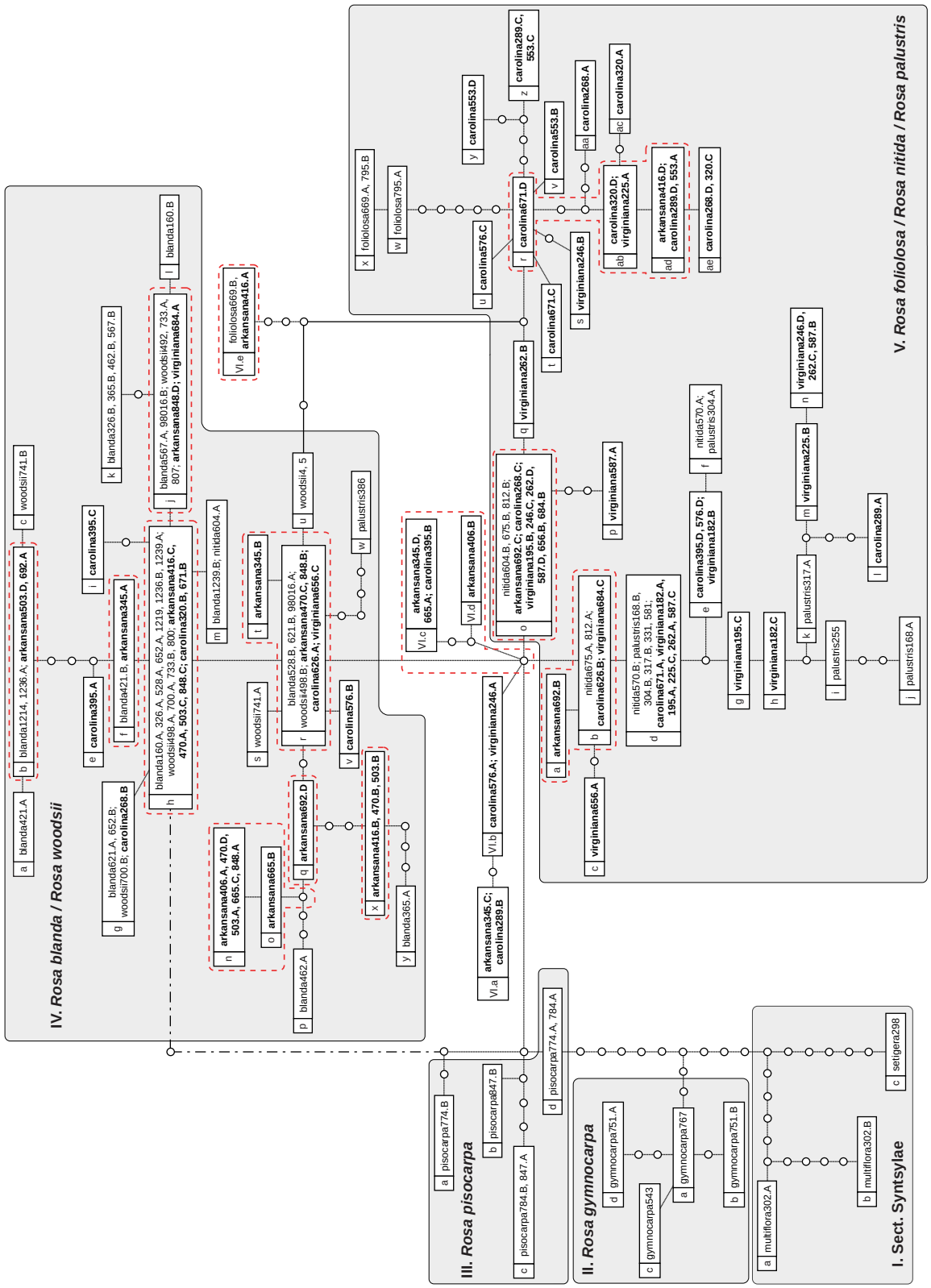
Projection des vecteurs de caractères sur les trois premiers axes de l'analyse en composantes principales des individus diploïdes et polyploïdes (voir Fig. 2.7).

	PC1	PC2	PC3
NLFT	0.041271882	-0.189055536	0.359501089
NSER	-0.053685093	0.361478746	-0.161226518
N2SER	0.029086303	0.131466913	0.093093980
LLFT	-0.252992611	0.338919996	-0.170046088
L1SER	-0.255662902	-0.024186462	-0.180078424
WLFT	-0.316430465	0.174729201	-0.114933982
LWLFT	-0.283246117	0.261378377	-0.195619757
LTEET	-0.254909610	-0.042251852	0.061307032
PULFT	-0.154632608	-0.256764183	-0.204646760
P1VEIN	-0.243266696	-0.206652248	-0.213190812
P2VEIN	-0.247063860	-0.233732185	-0.223015575
PBLFT	-0.176472390	-0.294603077	-0.218197449
G1VEIN	0.029855050	-0.201595985	-0.052026099
GBLFT	0.001951433	-0.203022532	-0.119151871
B1YW	0.052123171	0.034403192	0.351940771
T1YW	0.173423223	0.206746578	-0.082154000
LSTP	-0.280985384	0.185508410	0.133354718
LAUR	-0.277095861	0.086695902	0.331515811
WAUR	-0.294026168	0.053388481	0.307140695
WSTP	-0.268926905	-0.040611475	0.316852615
GSTP	-0.031169184	-0.139632317	0.042452429
GHYP	0.133177311	0.291215323	-0.195478680
GPED	0.136158480	0.266632502	-0.008119913
LPED	-0.185750956	0.049859058	0.138531386
NFLW	-0.156988689	-0.008282247	0.075091625

Note : une variable contribute significativement à une composante principale si sa projection sur celle-ci est plus grande que 0.2.

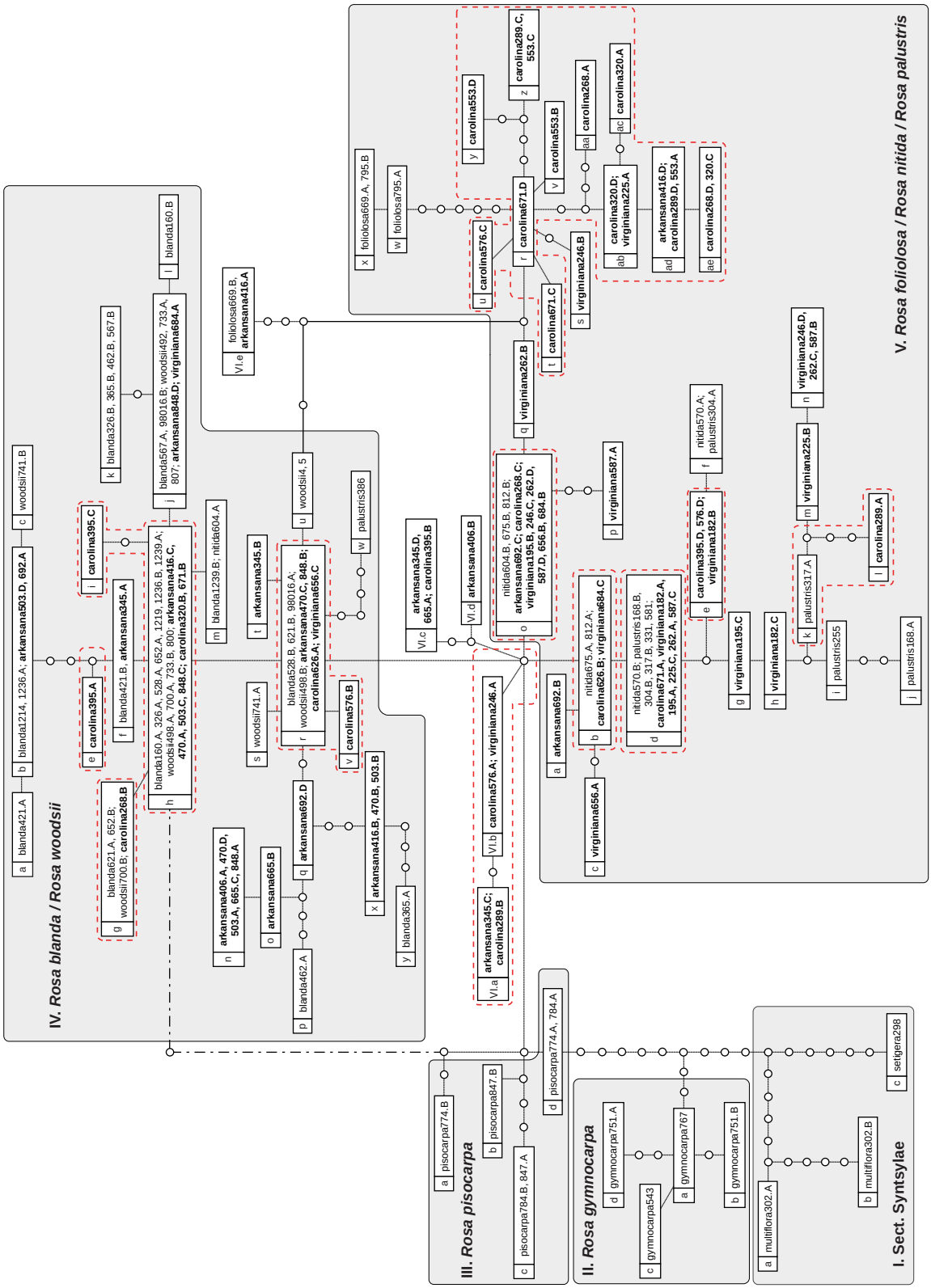
Annexe 5

Groupes d'haplotypes polyploïdes pour *Rosa arkansana* pour le gène GAPDH (Chapitre 5)



Annexe 6

Groupes d'haplotypes polyploïdes pour *Rosa carolina* pour le gène GAPDH (Chapitre 5)



Annexe 7

Groupes d'haplotypes polyploïdes pour *R. virginiana* pour le gène GAPDH (Chapitre 5)