Université de Montréal

Phylogénie moléculaire du genre *Salix* L. (Salicaceae) en Amérique du Nord

Par Aurélien Lauron-Moreau

Département de sciences biologiques, Faculté des arts et sciences

Mémoire présenté à la Faculté des arts et sciences en vue de l'obtention du grade de maitrise en sciences biologiques

Juin 2013

Résumé

La culture de saules (*Salix* sp.) est une pratique courante en Europe et en Amérique du Nord pour produire de la biomasse végétale. Cependant, le développement d'outils moléculaires est très récent. De plus, la phylogénie des saules est incomplète. Il y a un manque d'information pour les programmes de sélection d'espèces indigènes et pour la compréhension de l'évolution du genre. Le genre *Salix* inclut 500 espèces réparties principalement dans les régions tempérées et boréo-arctique de l'hémisphère nord. Nous avons obtenu l'ensemble des espèces retrouvées naturellement en Amérique (121 indigènes et introduites). Dans un premier temps, nous avons développé de nouveaux outils moléculaires et méthodes : extraction d'ADN, marqueurs microsatellites et gènes nucléaires. Puis, nous avons séquencé deux gènes chloroplastiques (*matK* et *rbcL*) et la région *ITS*. Les analyses phylogénétiques ont été réalisées selon trois approches : parcimonie, maximum de vraisemblance et Bayésienne. L'arbre d'espèces obtenu a un fort support et divise le genre *Salix* en deux sous-genres, *Salix* et *Vetrix*. Seize espèces ont une position ambiguë. La diversité génétique du sous-genre *Vetrix* est plus faible. Une phylogénie moléculaire complète a été établie pour les espèces américaines. D'autres analyses et marqueurs sont nécessaires pour déterminer les relations phylogénétiques entre certaines espèces. Nous affirmons que le genre *Salix* est divisé en deux clades.

Mots clés : BEAST; Classification; *ITS*; *matK*; Phylogénie moléculaire; *rbcL*; Salicaceae; *Salix*

Abstract

Fast growing willows (*Salix* sp.) are increasingly used in Europe and North America for biomass production and other environmental applications. However, the development of molecular tools is recent. The phylogeny of willows is incomplete, which slows down the selection of suitable native species and the development of improvement programs. The genus *Salix* includes approximately 500 species worldwide, and these are mainly located in temperate and cold regions of the Northern Hemisphere. We gathered leaf material from all 121 willows of North America (species native and introduced). We developed three molecular tools-methods: DNA extraction, SSR markers, and nuclear genes. We sequenced two chloroplast genes *matK* and *rbcL* and the *ITS* region. Phylogenetic analyses were carried out using parsimony, maximum likelihood and Bayesian approaches. The species tree provides strong support for a division of the genus into two subgenera, *Salix* and *Vetrix*. Sixteen species have ambiguous positions. A complete molecular phylogeny of American willows has been established. It needs to be confirmed and further resolved using other molecular data. Nonetheless, the genus clearly has two clades.

Key words: BEAST; Classification; *ITS*; *matK*; Molecular phylogeny; *rbcL*; Salicaceae; *Salix*

Table des matières

AVANT-PROPOS VIII

1. INTRODUCTION 1

1.1 OUTILS MOLECULAIRES DEVELOPPES CHEZ LE SAULE 1
1.2 STRUCTURE TAXONOMIQUE DE LA FAMILLE DES SALICACEAE ET DU GENRE *SALIX* 2
1.2.1 FAMILLE DES SALICACEAE 2
1.2.2 GENRE *SALIX* 2
1.3 ÉTUDES PHYLOGENETIQUES AU SEIN DE *SALIX* 4
1.4 OBJECTIFS ET HYPOTHESE DE RECHERCHE 6

2. CHAPITRE 1 : MICROSATELLITE MARKERS OF WILLOW SPECIES AND CHARACTERIZATION OF ELEVEN POLYMORPHIC MICROSATELLITES FOR *SALIX ERIOCEPHALA* (SALICACEAE), A POTENTIAL NATIVE SPECIES FOR BIOMASS PRODUCTION IN CANADA 8

2.1 INTRODUCTION 9
2.2 RESULTS AND DISCUSSION 10
2.3 EXPERIMENTAL SECTION 14
2.4 CONCLUSIONS 17

3. CHAPITRE 2 : PHYLOGENETIC RELATIONSHIPS OF AMERICAN WILLOWS (*SALIX* L., SALICACEAE) 18

3.1 INTRODUCTION 19
3.2 MATERIALS AND METHODS 21
3.2.1 PLANT MATERIAL 21
3.2.2 DNA EXTRACTION, AMPLIFICATION AND SEQUENCING 22
3.2.3 SEQUENCE ALIGNMENT 22
3.2.4 PHYLOGENETIC ANALYSES: GENE TREES 23
3.2.5 PHYLOGENETIC ANALYSES: SPECIES TREE 24
3.3 RESULTS 25
3.3.1 SUCCESS RATE OF THE AMPLIFICATIONS AND DNA SEQUENCES 25
3.3.2 POLYMORPHISMS 25
3.3.3 PHYLOGENETIC ANALYSES: GENE TREES 26
3.3.4 PHYLOGENETIC ANALYSES: SPECIES TREE 27
3.4 DISCUSSION 27
3.4.1 SEQUENCE POLYMORPHISM 27
3.4.2 PHYLOGENETIC RELATIONSHIPS BETWEEN AMERICAN WILLOWS SPECIES 28
3.4.3 BIOGEOGRAPHY OF *SALIX* 33
3.4.4 CLASSIFICATION OF *SALIX* 34

4. CONCLUSION ET PERSPECTIVES 42

4.1 OUTILS MOLECULAIRES DEVELOPPES DURANT LE PROJET 42
4.2 PHYLOGENIE MOLECULAIRE DU GENRE *SALIX* 44
4.3 PERSPECTIVES 45

SOURCES DOCUMENTAIRES (BIBLIOGRAPHIE) 46

Liste des tableaux

TABLEAUX CHAPITRE 1

TABLEAU 1. CHARACTERISTICS OF 24 MICROSATELLITE MARKERS USED FOR FIVE WILLOW SPECIES, OF WHICH 11 SUCCESSFULLY AMPLIFIED FOR *SALIX ERIOCEPHALA* 11
TABLEAU 2. RESULTS OF INITIAL PRIMER SCREENING IN SIX POPULATIONS OF *SALIX ERIOCEPHALA* 15
TABLEAU 3. GENETIC DIVERSITY STATISTICS 15

TABLEAUX CHAPITRE 2

TABLEAU 1. SUMMARY OF MOLECULAR PHYLOGENIES OF THE GENUS *SALIX* 36
TABLEAU 2. PRIMERS AND PCR CYCLE CHARACTERISTICS FOR THE THREE GENES USED IN THE STUDY 37

Liste des figures

FIGURE CHAPITRE 1

FIGURE 1. THE LOCALIZATION OF POPULATIONS USED IN THIS STUDY FROM EASTERN CANADA 16

FIGURES CHAPITRE 2

FIGURE 1. RELATIONSHIPS BETWEEN THE FIVE SUBGENERA OF *SALIX* IN NORTH AMERICA BASED ON MORPHOLOGICAL CHARACTERS (ARGUS, 1997) 38
FIGURE 2. BEAST GENE TREE OF *MATK* AND *RBCL* 39
FIGURE 3. BEAST GENE TREE OF *ITS* 40
FIGURE 4. *BEAST SPECIES TREE GENERATED WITH *ITS*, *MATK* AND *RBCL* 41
44

Liste des annexes

ANNEXE 1. EXTRACTION ADN PAR LA METHODE CTAB **X**
ANNEXE 2. METHODOLOGIE POUR CARACTERISER DE NOUVEAUX MARQUEURS NUCLEAIRES EN SIMPLE COPIE **XII**

Liste des sigles et abréviations

ADN (DNA) : acide désoxyribonucléique (deoxyribonucleic acid)

AFLP : amplified fragment length polymorphism

cp : chloroplastique (chloroplastic)

CTAB : cetyltrimethylammonium bromide

ITS : internal transcribed spacer

matK : maturase K

PCR : polymerase chain reaction

PVP : polyvinylpyrrolidone

QTL : quantitative trait loci

RFLP : restriction fragment length polymorphism

SNP : single-nucleotide polymorphism

SSR : simple sequence repeats

Avant-propos

Le développement des bioénergies est croissant dans le monde entier (Committee on Biobased Industrial Products, 2000). Des préoccupations à l'égard de l'approvisionnement et de la demande en énergie ont amené de nombreux pays à rechercher la sécurité énergétique et encouragé l'émergence d'un intérêt envers les énergies alternatives (Peart et al., 1993). La production de biomasse est l'une des énergies vertes les plus prometteuses. La matière organique issue des algues, animaux, champignons et plantes est utilisée directement comme source d'énergie par combustion (Demirbas, 2001; Gallagher et al., 2003). Plus récemment, de nouvelles applications sont apparues pour transformer la matière organique en biogaz ou en biocarburant (Ledin et Alriksson, 1992; Mitchell, 1994; Karp et al., 2011) où plusieurs arbres se sont révélés intéressants (ex. peuplier, saule et pin).

Les saules (*Salix* L., Salicaceae) intéressent de plus en plus les pays de l'hémisphère nord (Kuzovkina et al., 2008; Karp et al., 2011). Certaines espèces arbustives de ce genre ont une croissance rapide et une production de biomasse intéressantes pour les producteurs et utilisateurs (Stott, 1984; Karp et al., 2001; Volk et al., 2006). Depuis 30 ans, des programmes de sélection et d'amélioration ont été entrepris en Suède, au Royaume-Uni, et plus récemment aux États-Unis et au Canada (Kuzovkina et al., 2008; Karp et al., 2011). L'objectif est d'accroître la croissance, la biomasse, et la résistance aux ravageurs et aux maladies des espèces ou hybrides sélectionnés (Kuzovkina et al., 2008; Karp et al., 2011). Depuis peu, des outils moléculaires sont utilisés pour accélérer le processus de sélection (Karp et al., 2011). Ce projet de recherche a pour objectif de mettre en avant plusieurs types d'outils moléculaires chez le saule et d'accroitre les connaissances phylogénétiques du genre.

1. Introduction

Les outils moléculaires chez le saule se développent de plus en plus. Ils sont principalement mis en place dans une perspective d'utilisation dans les programmes de sélection. Également, les outils moléculaires ont une place importante pour appuyer des hypothèses de recherche. La classification des saules par les taxonomistes est sans cesse en changement. Plusieurs études moléculaires ont été publiées pour évaluer ces classifications. La synthèse des classifications basées sur les caractères morphologiques et des premières études de phylogénie moléculaire a permis d'élaborer notre hypothèse de recherche sur la classification des saules américains.

1.1 Outils moléculaires développés chez le saule

Les cultivars commercialisés aujourd'hui sont issus de programmes de sélection traditionnels où la méthodologie employée est analogue à celle utilisée pour le peuplier (Dickmann, 2001). Depuis peu, le groupe de recherche de *Rothamsted Research* (Royaume-Uni, http://www.rothamsted.ac.uk/) a innové en incorporant des outils moléculaires aux programmes de sélection. Un brevet s'intitulant «Method for improving biomass yield» fut d'ailleurs déposé en 2010. Ce brevet utilise principalement les données des cartes génétiques et des QTLs développés sur *Salix viminalis* (Tsarouhas et al., 2002, 2003, 2004; Rönnberg-Wästljung et al., 2005, 2006; Weih et al., 2006; Berlin et al. 2010, 2011; Brereton et al., 2010; Hanley et al., 2011; Karp et al., 2011; Samils et al., 2011).

D'autres travaux moléculaires ont été effectués avec pour objectifs de fournir de nouvelles informations en vue de la sélection et d'accroître les connaissances théoriques sur le genre.

Par exemple, Barker et al. (2003) ont caractérisé 46 microsatellites chez neuf espèces. D'autre part, le séquençage du génome complet du *Salix purpurea* est en réalisation par Lawrence Smart (Université Cornell).

Les travaux de L. Zsuffa (Université de Toronto) et de A. Mosseler (Agriculture Canada) ont mis en évidence le potentiel de plusieurs espèces nord-américaines pour la production de biomasse (com. pers.). Cet intérêt est récent, comme l'est le développement d'outils moléculaires associés à ces espèces.

1.2 Structure taxonomique de la famille des Salicaceae et du genre *Salix*

1.2.1 Famille des Salicaceae

La famille des Salicaceae est composée de 55 genres, incluant les saules (*Salix* L.) et les peupliers (*Populus* L.) (The Angiosperm Phylogeny Group, 2009). Les espèces appartenant aux genres *Chosenia* Nakai (Nakai, 1920) et *Toisusu* Kimura (Kimura, 1928) connurent différentes affiliations. Les premiers auteurs les considéraient comme des genres de la famille des Salicaceae. Cependant, certaines études morphologiques les écartèrent de cette famille, car de nombreuses différences furent observées dans la structure des grains de pollen (Kuprianova, 1965; Rouleau, 1970). Depuis 2009, cependant, les espèces de *Chosenia* et de *Toisusu* sont reconnues comme des Salicaceae et considérées comme des membres du genre *Salix* (The Angiosperm Phylogeny Group, 2009).

1.2.2 Genre *Salix*

Les saules sont un genre diversifié d'environ 500 espèces, les estimations variant de 300 à 550 (Skvortsov, 1968; Fang, 1999; Argus, 1997, 2010). Les saules sont répartis dans tout

l'hémisphère nord. La classification des saules est controversée depuis leur description par Linné en 1753. Argus (1997) a résumé l'histoire des classifications des espèces du genre *Salix*. Ces dernières sont basées sur différents caractères morphologiques, variant selon l'auteur (ex. Mulhenberg, 1806; Fries, 1825; Moss, 1914; Raup, 1959). Ce n'est que récemment qu'un certain consensus est survenu entre plusieurs auteurs concernant l'organisation de *Salix* en sous-genres et sections (Skvortsov, 1968; Dorn, 1976; Argus, 1997).

En 1968, Skvortov fut le premier à publier une révision des saules indigènes de l'ex-Union Soviétique et de l'Europe. Il classa les espèces en trois sous-genres (*Chamaetia*, *Salix* et *Vetrix*) incluant 15 sections (Skvortov, 1968). Cependant, il indiqua que la séparation des sous-genres *Chamaetia* et *Vetrix* n'était pas claire. Dorn (1976) publia la première révision des saules nord-américains, dans laquelle il utilisa deux des sous-genres décrits par Skvortov (*Salix* et *Vetrix*). Il ajouta de nouvelles sections (ex. *Floridanae*) pour classer les 89 espèces identifiées dans la flore. En Chine, les saules furent décrits et organisés en 37 sections (Wang et Fang, 1984; Fang, 1999), mais ces deux études n'identifièrent pas de sous-genres. Les travaux de Argus permirent de mieux caractériser les espèces de *Salix* trouvées en Amérique du Nord. Il publia de nombreuses descriptions à des échelles locales (Argus, 1964, 1965, 1869, 1973, 1983, 1986a, 1986b, 1991, 1993, 1995; Argus et Steele, 1979; Argus et McJannet, 1992). De plus, il s'intéressa à l'hybridation interspécifique (Argus, 1974) et au niveau de ploïdie (Suda and Argus, 1968) de plusieurs espèces nord-américaines. Ses différents travaux et des efforts de synthèse des données permirent d'établir une classification des 103 espèces nord-américaines (Argus, 1997, 2010). Il établit une liste de 197 caractères morphologiques en se basant sur plus de vingt ans d'observations personnelles et de travaux antérieurs

(Crovello, 1968; Skvortsov, 1968). Il obtint une classification en cinq sous-genres en utilisant la méthode de regroupement ISS FLEX (Dallwitz, non publié; Van den Borre and Watson, 1994; Argus, 1997). La classification de Argus est la plus récente et la plus complète. Les espèces nord-américaines se classent en cinq sous-genres: *Chamaetia* (8 sections), *Longifoliae* (1 section), *Protitea* (2 sections), *Salix* (8 sections) et *Vetrix* (15 sections). En 2007, Argus décrivit la répartition géographique des espèces de saules natives et introduites en Amérique du Nord et au nord du Mexique. Argus (com. pers.) travaille actuellement à une classification mondiale des espèces du genre *Salix* en se basant principalement sur les descriptions de Skvortsov (1968, 1999), Argus (1997, 2010), Fang et al. (1999), et Ohashi (2000). Cette classification organiserait le genre *Salix* en 7 sous-genres et 54 sections. Des différences s'observent entre les classifications des saules nord-américains de Dorn et Argus.

1.3 Études phylogénétiques au sein de *Salix*

Cinq études moléculaires ont été effectuées chez les saules entre 1999 et 2011 (Leskinen et Alstrom-Rapaport, 1999; Azuma et al., 2000; Chen et al., 2010; Hardig et al., 2010; Abdollahzadeh et al., 2011).

1) L'étude de Leskinen et Alstrom-Rapaport (1999) s'est intéressée à la région *ITS* de 13 espèces du genre *Salix*, quatre espèces de *Populus*, et *Chosenia bracteosa*. L'objectif principal était d'observer la monophylie au sein du genre *Salix* et de voir la position taxonomique du *Chosenia*. Trois points majeurs ont été soulevés : A) La variabilité de la région *ITS* est relativement faible en comparaison aux autres Angiospermes (Baldwin et al., 1995; Wen et Zimmer, 1996); les auteurs ont suggéré que le genre *Salix* a connu une divergence récente ou

une évolution lente. B) L'analyse phylogénétique place le *C. bracteosa* parmi les *Salix*; ils ont conclu qu'il devrait être considéré comme un saule. C) Les genres *Populus* et *Salix* sont des groupes frères.

2) L'étude d'Azuma et al. (2000) a porté sur les relations phylogénétiques de 19 espèces du genre *Salix* en se basant sur le gène chloroplastique *rbcL*. *Chosenia arbutifolia* et *Toisusu urbaniana* ont été inclus dans l'échantillonnage. Les auteurs ont confirmé la monophylie des genres *Salix*, *Chosenia* et *Toisusu* avec une valeur de support de 100%. Les auteurs ont comparé leur topologie avec les classifications morphologiques, mais aucune ne supporte les relations obtenues dans l'analyse moléculaire.

3) L'étude de Hardig et al. (2010) a utilisé le gène chloroplastique *matK* et la région *ITS*. Cette étude a incorporé 25 espèces du genre *Salix* en plus de *Chosenia arbutifolia* et *Populus deltoides*. Cette étude a illustré la présence de variabilité intra-spécifique sur les deux régions. Les auteurs ont attribué ce phénomène à des évènements d'hybridation inter-spécifique et d'introgression.

4) Chen et al. (2010) ont généré une phylogénie moléculaire en utilisant trois marqueurs chloroplastiques: le gène *rbcL*, et les régions inter-géniques *atpB-rbcL* et *trnD-T*. Les auteurs ont utilisé 31 saules, principalement asiatiques, ainsi que *Chosenia arbutifolia*. De plus, ils ont intégré les données de Azuma et al. (2000). Les auteurs ont apporté de nouvelles réponses et des confirmations : A) le genre *Salix* forme un groupe monophylétique, incluant *Chosenia* et *Toisusu*. B) L'arbre phylogénétique distingue deux clades majeurs; le premier clade intègre les espèces du sous-genre *Salix* (sauf les sections *Triandrae* et *Urbanianae*), le second toutes

les espèces des sous-genres *Chamaetia* et *Vetrix*, plus *Chosenia*, *Toisusu* et les sections *Triandrae* et *Urbanianae* du sous-genre *Salix*.

5) Abdollahzadeh et al. (2011) se sont intéressés à la région *ITS* pour réaliser une étude de phylogénie moléculaire de 26 espèces iraniennes. Ils ont intégré les séquences de Leskinen et Alstrom-Rapaport (1999) et de Hardig et al. (2010). Le support de l'arbre présenté n'est pas optimal. Selon les auteurs, les espèces polyploïdes de la flore iranienne provoquent des incongruences dans leurs données. Les auteurs concluent que le genre *Salix* est monophylétique et que le sous-genre *Salix* est probablement mal structuré ou polyphylétique.

Les arbres phylogénétiques basés sur les gènes chloroplastiques donnent une structure en deux clades majeurs, mais la résolution à l'intérieur des clades est faible (Azuma et al., 2000; Chen et al., 2010; Hardig et al., 2010). Pour la région *ITS*, une grande polytomie est observée (Leskinen and Alstrom-Rapaport, 1999; Hardig et al., 2010; Abdollahzadeh et al., 2011). Ces études moléculaires sont préliminaires, car l'échantillonnage est incomplet. Cependant, des divergences semblent être présentes entre les données moléculaires et les classifications récentes basées sur les caractères morphologiques (Skvortsov, 1968 ; Dorn, 1976; Argus, 1997, 2010). Ces interrogations nous ont permis d'élaborer une hypothèse de recherche pour notre étude, dans laquelle toutes les espèces américaines sont incluses.

1.4 Objectifs et hypothèse de recherche

Notre principale hypothèse de recherche est la suivante : la classification selon les données moléculaires des espèces américaines sera corrélée à la classification en cinq sous-genres de

Argus (1997, 2010): cinq clades seront présents et les espèces appartenant à chaque sous-genre formeront un groupe monophylétique. Le premier objectif est de développer de nouveaux outils moléculaires chez le saule. Le second objectif est d'obtenir une phylogénie moléculaire des saules nord-américains.

Ce projet de recherche se divise en deux parties. La première présentera les outils moléculaires qui ont été mis au point pour caractériser les saules nord-américains. Trois outils moléculaires sont présentés : 1) Le protocole de l'extraction d'ADN génomique à haut débit (Annexe 1). 2) Douze marqueurs microsatellites ont été caractérisés sur cinq espèces nord-américaines possédant un potentiel en production de biomasse (Chapitre 1). 3) La méthodologie pour développer de nouveaux marqueurs nucléaires en simple copie est présentée dans l'Annexe 2. La deuxième partie du projet a permis de caractériser les affinités phylogénétiques entre tous les saules trouvés en Amérique (Chapitre 2). Trois marqueurs moléculaires ont été utilisés sur 213 spécimens (121 espèces) et ils ont permis d'apporter de nouveaux éléments pour supporter une nouvelle structure au sein du genre *Salix*.

2. Chapitre 1 : Microsatellite markers of willow species and characterization of eleven polymorphic microsatellites for *Salix eriocephala* (Salicaceae), a potential native species for biomass production in Canada

Plants **2013**, *2*, 203-210; doi:10.3390/plants2020203

OPEN ACCESS

plants

ISSN 2223-7747

www.mdpi.com/journal/plants

Short Note

Aurélien Lauron-Moreau, Frédéric E. Pitre, Luc Brouillet and Michel Labrecque

Received: 21 January 2013; in revised form: 8 March 2013 / Accepted: 20 March 2013 / Published: 27 March 2013

Abstract: Biomass produced from dedicated plantations constitutes a source of renewable energy and is expected to play an important role in several countries in the coming decades. The cultivation of woody crops such as willows therefore raises several environmental issues. In North America, several native willows are potentially interesting for biomass producers. Willow trees are diverse but few species used for environmental applications have been the object of molecular genetic studies. Based on the sequenced poplar genome, 24 microsatellite markers were assayed on five native North American willow species: *Salix amygdaloides*, *S. discolor*, *S. eriocephala*, *S. interior* and *S. nigra*. Polymorphic microsatellite markers were used to characterize the allele variation on the shrub *Salix eriocephala*, a North American species with economic potential. Eleven markers amplified and confirmed the potential of this species. Analysis of samples from six populations in eastern Canada showed that all markers were variable and polymorphic in at least one population. The number of alleles per locus ranged from 1 to 9 (mean 2.95) and showed that these microsatellite markers can be used to assess genetic diversity of North American willow species.

Keywords: allele data; microsatellite markers; Salicaceae; *Salix eriocephala*; willow

2.1 Introduction

Genus *Salix* L. (Salicaceae) includes about 500 species worldwide, concentrated in temperate and cold regions of the Northern Hemisphere. Although it presents greater diversity in areas such as China (around 275 species; Fang, 1999), more than 100 species are found in North America (Argus, 2010). Willow species are cultivated worldwide for the production of biomass and for environmental applications (Karp et al., 2011) using short-rotation intensive culture (SRIC). Some willow species and cultivars have been favored due to their high biomass yield (Labrecque and Teodorescu, 2005; Volk et al., 2011). In North America, most of the cultivars used for biomass production are of exotic origin. The use of native species could represent several advantages to stimulate the deployment of bioenergy plantations. In addition to their higher level of social acceptability, it is believed that native species will be better adapted to the pedoclimatic conditions characterizing North American regions. They are also characterized by a higher genetic diversity within their geographical distribution than introduced cultivars and they could be exploited for diverse applications (Lin et al., 2009).

The North American willow species, *Salix amygdaloides*, *S. discolor*, *S. eriocephala*, *S. interior*, and *S. nigra*, have morphological traits that suggest high biomass potential. These shrubs can grow rapidly to 6–8 meters and are widespread in eastern North America (Argus, 2010). Among these, *Salix eriocephala* is of particular interest because of its extensive North-South distribution and adaptation to a range of climatic conditions. Knowledge of the genetic diversity of populations is a prerequisite to breeding programs aimed at improving biomass production. At a broader scale, given that half the current range of this species was once covered by Quaternary glaciers, genetic information may provide clues to the species

population structure following recolonization of glaciated areas (e.g. Godbout et al., 2003). Barker et al. (2003) developed 46 microsatellite markers in *Salix*, of which 17 were polymorphic (2–22 alleles per locus). Two studies derived microsatellite markers for *Salix* from the related *Populus trichocarpa* (Salicaceae) genome (Tuskan et al., 2006) to evaluate genetic diversity: Puschenreiter et al. (2010) successfully amplified 73% of the *Populus* SSRs assayed in *S. caprea* (only 10% were polymorphic), each of which was functional on a variable number of species among the four species assayed; Lin et al. (2009) used eight microsatellite markers, from both *Populus* and the Barker study (PMGC2709, PMGC2889, Karp_SB24, Karp_SB199, Karp_SB493, Karp_W293, Karp_W504 and Karp_W784) to compare the genetic diversity of *S. eriocephala* to that of *S. purpurea* (introduced willow) at a regional level.

Here, we assayed 24 microsatellite markers on five willow species and characterized 11 polymorphic microsatellite markers for *S. eriocephala* in widely distributed populations in eastern Canada. Our objectives were: (1) to determine whether these markers were present in the North American *Salix* species studied, and (2) to evaluate polymorphisms within the species *S. eriocephala*.

2.2 Results and Discussion

In total, twelve microsatellites were amplified with *Salix* species of this study and five markers were shared in these willows (Table 1). In addition, WPMS_7 amplified in *S. amygdaloides*, *S. discolor*, *S. interior* and *S. nigra*; ORPM_207 and PMGC_2658 in *S. amygdaloides*, *S. eriocephala* and *S. nigra*; PMGC_2315 in *S. amygdaloides*, *S. eriocephala*

and *S. interior*; and WPMS_18 in *S. discolor* and *S. eriocephala*. Two markers (ORPM_21 and ORPM_349) were present only in *S. eriocephala*.

Table 1. Characteristics of 24 microsatellite markers used for five willow species, of which 11 successfully amplified for *Salix eriocephala*. Information on each primer pair includes locus name, expected size (bp) in poplar, forward and reverse sequences, repeat motif, successful amplification with populations of *S. amygdaloides* (Sa), *S.discolor* (Sd), *S.eriocephala* (Se), *S. interior* (Si) and *S. nigra* (Sn), and number of cycles (Ca).

Locus[a]	Expected size (bp) in poplar	Primers sequences (5' - 3')	Repeat motif	Successful amplification with populations of					Number of cycles C_a
				Sa	Sd	Se	Si	Sn	
GCPM_1011	221	F: ATGAAATAATCGTTTGGTGC R: CACCCGAGTTTATCTCACTC	$(AT)_{11}$	+	+	+	+	+	27
GCPM_1037	123	F: ATGAAATTCGCAAAGTCAGT R: AAAAGAGGAAATTACGGTCC	$(TA)_{11}$	+	+	+	+	+	33
GCPM_1043	154	F: TTTCCATGTAGTATTACTCCTTTCT R: ATGCGTACCTTAGTGGAAGA	$(AT)_{21}$	-	-	-	-	-	33
ORPM_16	238	F: GCAGAAACCACTGCTAGATGC R: GCTTTGAGGAGGTGTGAGGA	$(CTT)_{15}$	-	-	-	-	-	33
ORPM_21	230	F: GGCTGCAGCACCAGAATAAT R: TGCATCCAAAATTTTCCTCTTT	$(AG)_4$	-	-	+	-	-	25
ORPM_23	197	F: ATTCCATTTGGCAATCAAGG R: CCCTGAAAGTCACGTCTTCG	$(AT)_6(AG)_6$	-	-	-	-	-	33
ORPM_29	206	F: TGGTGATCCAGTTTTGGTGA R: GTCCTTGCAAGCCATGAA	$(AC)_{11}$	-	-	-	-	-	33
ORPM_127	200	F: TCAATGAGGGGTGCCATAAT R: CTTTCCACTTTTGGCCCTTT	$(TG)_8$	-	-	-	-	-	33
ORPM_203	209	F: CCACCAGGCATGAGATATGA R: TCAAACCGAAAGGTCAACAA	$(TA)_4$(A/T rich region)	-	-	-	-	-	33
ORPM_206	196	F: CCGTGGCCATTGACTCTTTA R: GAACCCATTTGGTGCAAGAT	$(GCT)_7$	-	-	-	-	-	33
ORPM_207	199	F: TGCATATTTCACGTGCCTTT R: CAAAGTGAGGAAGCGTCAGA	$(TC)_8$	+	-	+	-	+	25
ORPM_349	202	F: GAGCATGAAGCATGAGCAGA R: TTTTCAGAACCAGGGGAAAA	$(AC)_{16}$	-	-	+	-	-	33
PMGC_223	170	F: CGATGAGGTTGAAGAAGTCG R: ATATATGTACCGGCACGCCAC	$(CTT)_n$	+	+	+	+	+	25
PMGC_2015	160	F: TTTTGGCATTCAAAGACTTGGC R: AGTTGATTCCATGTCGTGTCC	$(GA)_n$	-	-	-	-	-	33

PMGC_2315	143	F: CTGTGGTATTTGTGCAATGTG R: CAACAGAGCAAACTTGAGTCG	$(GA)_n$	+	-	+	+	-	33
PMGC_2392	192	F: AAGAGAGATAGCATCACCAAG R: TATGTCGAGGAAATCCTTAGC	$(GA)_n$	-	-	-	-	-	33
PMGC_2531	140	F: TAAGAGAATTGGGAGAGCAAC R: TTTTATCTTTTCCAGTTGTCTAC	$(GA)_n$	-	-	-	-	-	33
PMGC_2610	114	F: AACACGCAAGAACATACATAAG R: GATTAACATGTTTCGCTACGC	$(GA)_n$	-	-	-	-	-	33
PMGC_2647	129	F: CTCGTTAATTAGAGTCGAATTAG R: TTGTTATCCACTGCCAGTGC	$(GA)_n$	-	-	-	-	-	33
PMGC_2658	251	F: GCCCTTGAATACCATGAGCG R: ACCTTCAGTAGATCAGGTTAGTG	$(GA)_n$	+	-	+	-	+	33
WPMS_7	230	F: ACTAAGGAGAATTGTTGACTAC R: TATCTGGTTTCCTCTTATGTG	$(GT)_{24}$	+	+	-	+	+	33
WPMS_15	193	F: CAACAAACCATCAATGAAGAAGAC R: AGAGGGTGTTGGGGGTGACTA	$(CCT)_n$	+	+	+	+	+	25
WPMS_16	145	F: CTCGTACTATTTCCGATGATGACC R: AGATTATTAGGTGGGCCAAGGACT	$(GTC)_n$	+	+	+	+	+	25
WPMS_18	245	F: CTTCACATAGGACATAGCAGCATC R: CACCAGAGTCATCACCAGTTATTG	$(GTG)_{13}$	-	+	+	-	+	33

[a] from The International *Populus* Genome Consortium

n = unknown

F = forward primer, R = reverse primer

+ = amplification, - = no amplification

For *Salix eriocephala*, population genetic statistics are summarized in Tables 2 and 3, which shows the number of alleles (A), observed (Ho) and expected (He) heterozygosities, p-values for the Hardy-Weinberg Expected test, and details of each population. Eleven of the 24 markers were successfully amplified in the six populations of *S. eriocephala* (Table 1) and they were single markers. The values of A varied between 1 and 9 (mean 2.95) (Table 2). All 11 markers were polymorphic in at least one population (Table 2). The means of Ho and He ranged from 0.52 to 0.68 and from 0.43 to 0.56, with mean values of 0.59 and 0.48,

respectively (Table 3). These results reveal a new set of polymorphic microsatellite markers in which the genetic diversity within the *S. eriocephala* populations is expressed clearly.

Tuskan et al. (2006), using the complete genome of *Populus trichocarpa*, characterized microsatellites markers and predicted that 30–50% of these might transfer to the related *Salix*. Puschenreiter et al. (2010) successfully amplified 73% of microsatellite markers assayed, 10% of which were polymorphic. Out of 28 *Populus* loci assayed, Lin et al. (2009) used four loci in their willow study that proved polymorphic (no data were available on the percentage that amplified in willow but were monomorphic). Our amplification success of *Populus* microsatellites was 50% (polymorphic in at least one species), which falls within the range predicted by Tuskan et al. (2006); the difference in the percentage of polymorphic amplified loci in the Puschenreiter et al. (2010) study and ours may stem from the fact that the former used fewer individuals per species to characterize their loci. In *Salix eriocephala*, Lin et al. (2009) obtained a higher number of alleles (8–13 vs. 1–9) for *Populus*-derived loci, but their population size was much larger than that of populations in our study, which may account for some of the difference. Furthermore, some of our populations were at the extreme northern edge of the species range of *S. eriocephala* and may have lost genetic diversity while migrating northward after glaciations (Godbout et al., 2010). In all these studies, a majority of markers amplified and were polymorphic for many species, with a number being variably useful depending on the species.

2.3 Experimental Section

Six populations of *S. eriocephala* were selected to reflect the wide distribution of this species (Table 1 and Figure 1). These populations were collected at Matagami, QC (BEL, 10 trees); Blainville, QC (BAL, 8 trees); Montmagny, QC (MON, 7 trees); Radisson, QC (RAD, 12 trees); Shepody Creek, NB (SHE, 8 trees) and Val d'Or, QC (VDO, 13 trees). Two are part of a provenance trial maintained at the Montreal Botanical Garden, and four were collected in the wild in 2010–2011. For *S. amygdaloides*, *S. discolor*, *S. interior* and *S. nigra*, we used three populations each from Eastern Canada (Table 1 and Figure 1). The samples of *S. amygdaloides* were collected at Cobden lake, ON (4 trees); Hanlon Marsh, ON (4 trees); and Richmond Fen, ON (4 trees). The samples of *S. discolor* were collected at Levis, QC (10 trees); Montmagny, QC (8 trees); and Norton, NB (10 trees). The samples of *S. interior* were collected at Lafarge Pit, ON (5 trees); Limerick Forest, ON (5 trees); and Long Sault, ON (5 trees). The samples of *S. nigra* were collected at Gagetown, NB (4 trees); Pembroke, ON (4 trees); and Westmeath, ON (6 trees).

Genomic DNA was extracted from 25 mg of leaf material using a modified CTAB method (Doyle and Doyle, 1987). Leaf tissue was ground for 60 s in a TissueLyser II (QIAGEN) equipped with a 3 mm tungsten ball. 400 µL of 2× CTAB (2% (w/v) hexaldecyltrimethylammonium bromide), 8 µL of mercaptoethanol and 1% (w/v) polyvinylpyrrolidone were added to the powder; the mixture was vortexed and incubated at 65 °C for 90 min. An equal volume of isoamyl alcohol:chloroform (1:24) was then added. Following centrifugation at 5788 g for 30 min at 4 °C, the aqueous phase was collected and the DNA was precipitated by the addition of 1,5 volumes of isopropanol; it was conserved at

−20 °C overnight. Then, after centrifugation at 5788 g for 30 min at 4 °C, the pellet was washed with 500 µL of 70% ethanol and centrifuged at 5788 g for 10 min at 4 °C. The DNA pellet was suspended in 100 µL TE (10 mN Tris-HCl pH 8.5, 1 mM EDTA) and stored at −20 °C until analysis.

Table 2. Results of initial primer screening in six populations of *Salix eriocephala*. Two populations (MON and SHE) are part of a provenance trial maintained at the Montreal Botanical Garden (45°33'41'' N, 73°34'7'' W) and four were collected in nature (BEL, BLA, RAD and VDO). The localization of each population is 49°45.456' N, 77°36.921' W (BEL); 45°41.160' N, 73°51.908' W (BLA); 46°94' N, 70°60' W (MON); 53°41.441' N, 78°06.596' W (RAD); 45°71' N, 64°77' W (SHE) and 48°01.770' N, 77°45.937' W (VDO). Data on markers alleles. The range of allele sizes includes 23 bp of M13 sequence.

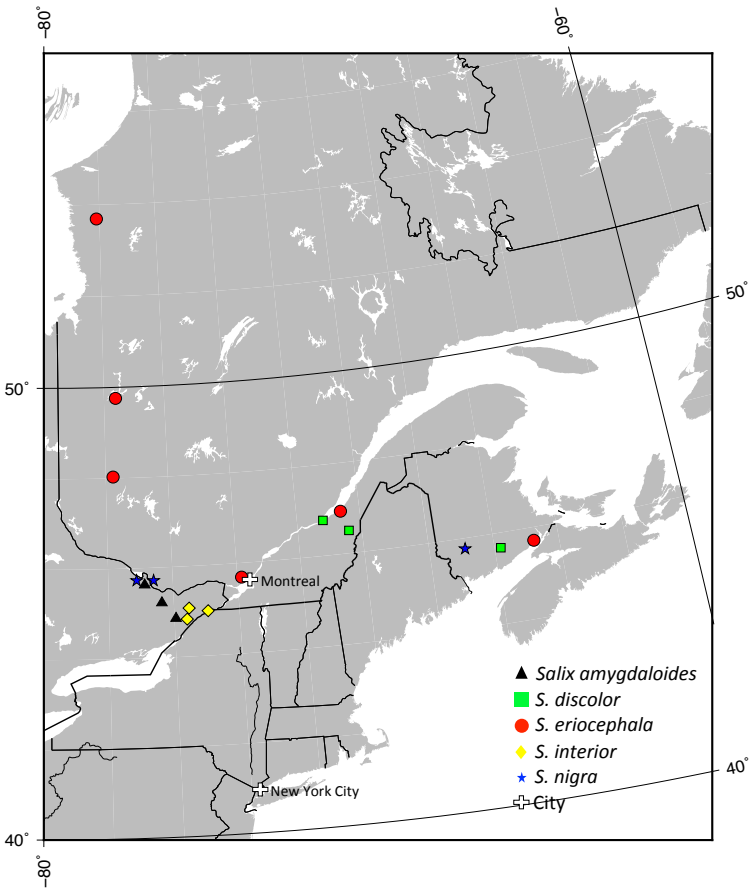
Locus	Range of allele sizes (bp)	BEL	BLA	MON	RAD	SHE	VDO	Average
GCPM_1011	207–229	3	4	1	4	1	4	2.83
GCPM_1037	93–188	3	3	5	3	5	5	4.00
ORPM_21	201–229	1	2	3	4	1	2	2.17
ORPM_207	186–221	4	1	1	7	1	4	3.00
ORPM_349	127–133	2	2	2	1	2	2	1.83
PMGC_223	181–219	3	2	2	2	2	5	2.67
PMGC_2315	142–182	3	2	3	2	1	3	2.33
PMGC_2658	205–242	9	6	5	8	4	6	6.33
WPMS_15	161–229	4	2	2	5	1	2	2.67
WPMS_16	137–185	2	3	3	3	3	5	3.17
WPMS_18	245–248	1	1	2	1	2	2	1.50

Table 3. Genetic diversity statistics. N = number of individuals sampled in the population.

Population	N	A	*Ho*	*He*	*p*-HWE
BEL	10	3.18	0.60	0.44	0.51
BLA	8	2.45	0.68	0.46	0.15
MON	7	2.82	0.64	0.49	0.40
RAD	12	3.64	0.57	0.43	0.38
SHE	8	2.09	0.54	0.47	0.46
VDO	13	3.55	0.52	0.56	0.15

A = number of alleles; *He* = expected heterozygosity; *Ho* = observed heterozygosity; and *p*-HWE: *p*-values for the Hardy-Weinberg Expected test.

Figure 1. The localization of populations used in this study from Eastern Canada. Triangle for *Salix amygdaloides*, square for *S. discolor*, circle for *S. eriocephala*, lozenge for *S. interior* and star for *S. nigra*.



The International *Populus* Genome Consortium identified 4,200 *Populus* SSRs (Van der Schoot et al., 2000; Smulders et al., 2001; Tuskan et al., 2004). A selection of 24 microsatellite markers (Table 1) was initially assayed to visualize the amplifications. The protocol developed by Schuelke (2000) was used to amplify the microsatellite regions. This method fluoresces the PCR products at 700 nm or 800 nm for detection by laser. In this protocol, three primers are necessary: a specific forward primer with a M13 tail at its 5' end, a specific reverse primer, and the fluorescent-labeled M13 primer (5'-AGGGTTTTCCCAGTCACGACGTT-3'). PCRs

were carried out in a 10 µL solution containing 0.5 µL of genomic DNA (approximately 50 ng), 0.75× of PCR buffer (BIO BASIC), 0.10 µM of forward primer, 0.25 µM of reverse primer, 0.15 µM of fluorescent-labeled M13 primer (Integrated DNA Technologies), 0.25 mM of dNTPs, 2.25 mM of $MgCl_2$, 1 U Taq DNA polymerase (GenScript). An Eppendorf Mastercycler® pro Thermal Cyclers (Eppendorf) was used under the following cycling parameters: initial denaturation at 94 °C for 3 min followed by 25–33 cycles (Table 1) of 30 s at 94 °C; 30 s at 52 °C, 45 s at 72 °C and followed by a final extension at 72 °C for 5 min was used. The successful amplification was visualized on agarose gel (1%); the results are summarized in Table 1. For *S. eriocephala*, amplification products were separated on a 6.5% KBPlus Gel Matrix and visualized using a 4,300 DNA Analyzer from LI-COR. The SagaGT v3.2 software was used to calculate band sizes. For each locus, we calculated standard population genetic statistics (*Ho*, *He* and *p*-HWE) using GENEPOP v.4.2 (Raymond and Rousset, 2005; Rousset, 2008).

2.4 Conclusions

In conclusion, we have added 11 new polymorphic microsatellite markers for *S. eriocephala*. If the loci used by Lin et al. (2009) are added, a total of 19 loci have been identified and have shown to be polymorphic in this species. The twelve markers characterized in this study are available for further genetic analyses with *S. eriocephala*, *S. amygdaloides*, *S. discolor*, *S. interior* and *S. nigra*. Studies of *Salix* microsatellites increased the molecular tools available for the investigation of willow species and the knowledge of this genus.

3. Chapitre 2 : Phylogenetic relationships of american willows (*Salix* L., Salicaceae)

Aurélien Lauron-Moreau, Frédéric E. Pitre, Luc Brouillet and Michel Labrecque

Abstract

Premise of the study: *Salix* L. is the largest genus in the family Salicaceae (450 species). Several classifications have been published, but taxonomic subdivision has been under continuous revision. Our goal is to establish the phylogenetic structure of the genus using molecular data on all American willows, using three DNA markers.

Methods: Material was obtained for the 121 native and introduced willow species of America. Sequences were obtained from the *ITS* (ribosomal nuclear DNA) and two chloroplastic regions, *matK* and *rbcL*. Phylogenetic analyses (parsimony, ML, MrBayes, BEAST) were performed on the data and the BEAST species tree is presented. Species distributions were mapped onto the tree.

Key results: The species tree provides strong support for a division the genus into two subgenera, *Salix* and *Vetrix*. Subgenus *Salix* comprises temperate species from the Americas and Asia. Subgenus *Vetrix* is composed of boreo-arctic species of the Northern Hemisphere. Sixteen species have ambiguous positions; genetic diversity is lower in subg. *Vetrix*.

Conclusion: A complete molecular phylogeny of American willows was established, although it needs to be confirmed and further resolved using other molecular data. Nonetheless, the genus clearly has two clades.

Key words: America; BEAST; *ITS*; *matK*; Molecular phylogeny; *rbcL*; Salicaceae; *Salix*

3.1 Introduction

Salix L. is the largest genus of family Salicaceae with about 450 species (Newsholme, 1992; Fang et al., 1999; Argus, 1997, 2010). The genus is distributed across the temperate and arctic regions of the Northern Hemisphere, entering tropical regions along montane ranges; willows also have been introduced worldwide. Over half the willow species, 275, are found in China (Fang et al., 1999), 107 in the former Soviet Union (Skvortsov, 1999), 65 in Europe (Argus, 1997), and 103 in North America (Argus, 2010). In Canada, 30% of the woody species are willows (Labrecque, com. pers.). Willows are mostly shrubs that play an important role in riparian habitats, wetlands and in shrub tundra and have contributed socially and economically to human societies (Sneader, 2000; Kuzovkina et al., 2008; Karp et al., 2011). During the last century, interest in environmental applications has grown, notably for biomass production (Stott, 1984; Lindegaard and Barker, 1997; Volk et al., 2006; Karp et al., 2011; Guidi et al., 2013).

Chase et al. (2002) characterized the relationships among the genera of an extended family Salicaceae, and Alford et al. (2009) studied more closely the relationships of *Salix* and *Populus*. *Salix* and *Populus* form a monophyletic group. In Chase et al. (2002), *Itoa* and *Poliothyrsis* are successive sister to *Salix* and *Populus*. In Alford et al. (2009), the genera *Idesia, Bennettiodendron* and *Olmediella* are sister to *Populus* and *Salix*, with *Itoa, Poliothyrsis, Carrierea* and *Macrohasseltia* sister to this clade.

Several classifications of *Salix* have been published and the subdivision of the genus has been under continuous revision. Argus (1997) reviewed the history of *Salix* classifications and showed that the genus has been divided into 35 genera since its description by Linnaeus,

each author using different morphologic characters to justify these divisions. For instance, some Asian treatments recognized the genera *Chosenia* Nakai (1920) (*Salix arbutifolia*) and *Toisusu* Kimura (1928) (*Salix cardiophylla*). Argus (1997) showed that subgenus *Chosenia* (including both species above) is sister to subgenus *Salix* (Fig. 1) and the Angiosperm Phylogeny Group (2009) included *Chosenia* and *Toisusu* within *Salix.* Several subgeneric classifications of *Salix* have been proposed. Most recently, Skvortsov (1968) divided the species of the former Soviet Union and Asia into three subgenera, *Salix, Chamaetia* and *Vetrix*. Dorn (1976) divided the American species into two subgenera, *Salix* and *Vetrix*. Argus (1997, 2010) suggested five subgenera for American willows (*Longifoliae*, *Protitea*, *Salix, Chamaetia* and *Vetrix*); Figure 1 illustrates the relationships between these subgenera.

Five molecular phylogenies have addressed the relationships between willow species (Leskinen and Alström-Rapaport, 1999; Azuma et al., 2000; Chen et al., 2010; Hardig et al., 2010; Abdollahzedeh et al., 2011). Table 1 summarizes the number of species and the molecular markers used in these studies. They confirmed the monophyly of *Salix,* membership of *Chosenia* (*S. arbutifolia*) and *Toisusu* (*S. cardiophylla*) to *Salix* and the presence of two major clades within the genus. These studies all included a small number of willow species relative to the total number of species in the genus.

The chloroplastic genes *matK* and *rbcL,* and the nuclear ribosomal *ITS* region have been used extensively in molecular phylogenetic studies (e.g., Hilu et al., 2003; Davis and Anderson, 2010; Vijaykumar et al., 2010). Their widespread use and ease of amplification has led to their selection as the main DNA regions in the barcoding program (Chase et al., 2007;

Ausubel, 2009; Yao et al., 2010), to be used for the identification of plants (e.g., Kuzmina et al., 2012).

Our objectives are to determine the phylogenetic relationship among all American *Salix* species (107 species), using *ITS*, *matK* and *rbcL,* in order to evaluate willow classifications and distribution patterns. We show that *Salix* is subdivided into two major clades, the first composed of temperate and the second of boreo-arctic species.

3.2 Materials and Methods

3.2.1 Plant material

This study includes all *Salix* species from America (107 species) plus introduced species in North America (14 species: 7 from Europe, 5 from Eurasia, and 2 from Asia) (Argus, 2010). The specimens were obtained from G. Argus' personal collection, Marie Victorin Herbarium (MT), the living collection of the Montreal Botanical Garden, the Canadian Museum of Nature (CAN), the Herbarium of the University of Texas (TEX), the University of Arizona Herbarium (ARIZ) and the Herbarium of the Missouri Botanical Garden (MO). *Chosenia arbutifolia* (= *Salix arbutifolia*) was sampled in the living collection of the Montreal Botanical Garden. A total of 211 specimens (121 species) of *Salix* were used in this study, with 1 to 3 specimens per species. The identity of a majority of specimens has been confirmed by G. Argus. We verified other specimens using Argus (2010). We downloaded sequences from GenBank for *Toisusu cardiophylla* (= *Salix cardiophylla*) and two outgroup genera, *Idesia* and *Populus* (Chase et al., 2002; Alford et al., 2009; The Angiosperm Phylogeny Group, 2009).

3.2.2 DNA extraction, amplification and sequencing

Genomic DNA was extracted from herbarium specimens or fresh leaves dried in silica gel. The CTAB method (Doyle and Doyle, 1987) was used, as modified in Lauron-Moreau et al. (2013). Three molecular regions were used in this study: *ITS*, *matK* (partial) and *rbcL* (partial). They were amplified using the specific primers detailed in Table 2. PCRs were carried out in a 20 µL solution containing 1 µL of genomic DNA (approximately 50-70 ng), 0.75X of PCR buffer (BIO BASIC, Markham, ON, Canada), 0.25 µM of each primer, 0.25 mM of dNTPs, 2.25 mM of $MgCl_2$, 1 U Taq DNA polymerase (BIO BASIC). PCRs were performed using an Eppendorf Mastercycler® pro Thermal Cyclers (Eppendorf Canada, Mississauga, ON, Canada) under the following cycling parameters: initial denaturation at 94°C for 3 min followed by 33-35 cycles (C_a) of 30 s at 94°C, 30 s at 52°C, 45-70 s at 72°C; followed by a final extension at 72°C for 5 min. The sequencing of PCR products were realized by the group McGill University and Génome Québec Innovation Centre. Over half the sequences of *matK* and *rbcL* were obtained with the help of the Barcode of Life Data Systems (BOLD) following standard protocols at the Canadian Centre for DNA barcoding (CCDB) for plants, as described in Kuzmina et al. (2012).

3.2.3 Sequence alignment

Sequences were assembled in contigs using Geneious Pro version 4.8.5 created by Biomatters (http://www.geneious.com). Alignments were done in SeaView version 4.2.6 (Gouy et al., 2010) using Muscle (Edgar, 2004), followed by manual correction. Parsimony, maximum likelihood (ML), and Bayesian (BA) analyses were performed to determine the phylogenetic relationships on four datasets: *ITS*, *matK*, *rbcL*, and the concatenated *matK-rbcL*

sequences. The program jModelTest2 (Guindon and Gascuel, 2003; Darriba et al., 2012) was used to select the best model of sequence evolution for ML and BA analyses. Data matrices from this study were deposited on TreeBase (website: http://treebase.org) under accession number study ID14313.

3.2.4 Phylogenetic analyses: gene trees

Parsimony analyses were performed using PAUP version 4.0b10 (Swofford et al., 2003). We selected the optimal trees using a heuristic search using these parameters: 100 random additions of sequences followed by tree bisection and reconnection (TBR) branchswapping, retaining at most 100 trees at each replicate. Branch support was estimated using 10,000 bootstrap replicates with the same heuristic settings.

Maximum likelihood analyses were performed using PhyML 3.0 (Guindon et al., 2010). ML heuristic searches and bootstrap analysis (10,000 replicates) were conducted to obtain the best trees under the parameters of evolution model selected. The adequate evolution models were GTR+G+I for *ITS*, GTR+G for *matK* and *matk-rbcL*, and K80+I for *rbcL*.

Bayesian analyses were performed using MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003) and BEAUti/BEAST v1.7.5 (Drummond et al., 2012). In MrBayes, two independent runs were performed, each consisting of four parallel Markov chain Monte Carlo (MCMC) of 100 million generations (the average standard deviation of split frequencies being lower than 0.01). Trees were sampled every 10,000 generations. The evolution models used were identical with those in the ML analyses. Tree parameters reached stationarity after a burn-in period of 250,000 generations. Optimal trees were then sampled every 1,000 generations to

obtain the final consensus tree and associated posterior probabilities. For the BEAST analysis, each molecular region was analyzed separately and the gene tree was estimated concurrently. Two independent runs of 100 millions generations were performed, each with a sampling at every 10,000 generations. We used the same evolution models as above, with four gamma categories, a coalescent tree prior and a strict clock model for each partition. After analysis, the software Tracer (Rambaut and Drummond, 2007) was used to evaluate the convergence after the first 20% of generations had been discounted as burn-in. The software TreeAnnotator v1.7.5 (available in BEAST package) was used to estimate the maximum-clade-credibility using the Bayesian posterior probabilities.

3.2.5 Phylogenetic analyses: species tree

Species tree was obtained using BEAUti/*BEAST (Drummond et al., 2012). The parameters used were identical with those in the BEAST analyses and we added the species tree ancestral reconstruction option (developed by Heled and Drummond, 2010). The chloroplast data are concatenated in this analysis. On the *BEAST species tree, we illustrated the main native area of each species following Argus (2007; for North America), using seven zones: four in North America (western temperate, western boreo-arctic, eastern temperate, eastern boreo-arctic) and three representing Europe, Asia and Mexico (including Central and South America).

3.3 Results

3.3.1 Success rate of the amplifications and DNA sequences

We extracted genomic DNA from 212 specimens and did PCR amplifications on *ITS*, *matK* and *rbcL* on these. We obtained 210 sequences for *ITS* (including 2 partial sequences) and 212 sequences for *matK* and *rbcL* (all sequences are available in GenBank). For the *ITS* region, amplification of two specimens of *S. atrocinerea* was not a successful, and it was partial for *S. jaliscana* and *S. prolixa*. New DNA extractions and a modification to the PCR protocol did not give better results. Fifteen sequences were downloaded from GenBank and aligned with our data. We did not find intra-species variation within our data. The alignment of *ITS*, *matK* and *rbcL* resulted in 608, 874 and 553 aligned nucleotides, respectively (Table 2). Including the GenBank data, we had 215 sequences for *ITS*, and 217 sequences for *matK* and *rbcL*. The *ITS* region had a higher proportion of polymorphic sites (21%) when compared with *matK* (12%) and *rbcL* (8%) (Table 2).

3.3.2 Polymorphisms

We observed many polymorphic sites in the *ITS* region. Sixteen species (*S. arbusculoides*, *S. arctica*, *S. arctophila*, *S. barclayi*, *S. cana*, *S. columbiana*, *S. discolor*, *S. exigua*, *S. famelica*, *S. floridana*, *S. humboldtiana*, *S. jejuna*, *S. monticola*, *S. raupii*, *S. richardsonii*, *S. rotundifolia*) had polymorphisms at 17 nucleotide sites (1-5 polymorphic sites per species). We also found polymorphisms in the chloroplastic genes. *Salix aeruginosa* and *S. jaliscana* are polymorphic at four (34, 367-368, 398) and two sites (80, 514), respectively, in *matK*. Six species had polymorphic sites on *rbcL*: *S. jaliscana* (396-397); *S. pedicellaris*, *S. pseudomyrsinites* (285); and *S. argyrocarpa*, *S. cascadensis*, *S. orestera* (286).

3.3.3 Phylogenetic analyses: gene trees

We compared the resolution and branch support of four analytical approaches (PhylML, MrBayes, BEAST and PAUP) on all datasets. The topologies were similar and we are presenting the results from BEAST because its support values were greater (Figs. 2-3). Nonetheless, on the trees shown, Bayesian posterior probabilities and ML bootstrap values are provided.

The phylogenetic trees obtained for *matK* and *rbcL* were identical except for the position of *Salix petrophila,* and we are presenting the consensus tree of these two chloroplastic genes (Fig. 2). Two major clades are apparent on the cp DNA tree. Clade A1 includes 32 *Salix* species and clade A2, 88. The two clades are well supported. The relationship within each clade is not well resolved, however, and branches with a posterior probability lower than 0.7 were collapsed. Clade A1 comprises the majority of species from subgenera *Longifolia*, *Protitea* and *Salix*. Clade A2 includes most species of the subgenera *Chamaetia* and *Vetrix*. *Salix arbutifolia* and *S. cardiophylla* belong to clade A2.

In the *ITS* tree, four different clades are supported (Fig 3). Clade B1 comprises most species of subgenus *Protitea*, clade B2 most species of subgenus *Salix*, clade B3 most species of subgenus *Longifoliae*, and clade B4 most species of subgenera *Chamaetia* and *Vetrix*. Overall, the species of clades B1, B2 and B3 (Fig. 3) are present in clade A1 of the cp trees (Fig 2), while B4, *S. arbutifolia* and *S. cardiophylla* are in clade A2. Fourteen taxa have incongruent positions on the two trees, however.

3.3.4 Phylogenetic analyses: species tree

Figure 4 presents the species tree of the three markers and using BEAUti/*BEAST. We concatenated the chloroplast data (*matK* and *rbcL*). The tree exhibits two major clades, C1 and C2. Five subclades are shown in clade C1 and in clade C2. Subclades C9 or C10 have low support. The distribution of species is presented on the tree.

3.4 Discussion

3.4.1 Sequence polymorphism

As Leskinen and Alström-Rapaport (1999), we observed intra-individual polymorphic nucleotide sites in the *ITS* region. Ribosomal sequences are present in thousands of copies in the nuclear genome (Rogers and Bendich 1987). Usually, sequences within individuals are uniformized due to concerted evolution (Alvarez and Wendel, 2003; Bailey et al., 2003). However, in cases of recent hybridization or homoploid speciation, sequence homogenization is often not achieved in the short period of time involved (Alvarez and Wendel, 2003; Bailey et al., 2003). For instance, Leskinen and Alström-Rapaport (1999) hypothesized that *S. schwerinii* could result from homoploid speciation after hybridization between *S. viminalis* and a second, unidentified species. In our study, *S. exigua*, a diploid species, shows polymorphic sites. It could be due either to introgression from recent hybridization or the species could be the result of homoploid speciation. Other species, however, are polyploid and polymorphism may merely result from a lack of homogenization, particularly in allopolyploid taxa. Conversely, *Salix alba,* a tetraploid, is without polymorphic sites in the *ITS* region. Data are currently insufficient to explain the presence of polymorphic sites in the *ITS* of American *Salix*.

Polymorphic nucleotide sites in plastid sequences may seem surprising but are not new. Few studies have reported this (McCauley et al., 1996; Olivier et al., 2010). One hypothesis to explain such polymorphism would be the inclusion of cp DNA fragments in nuclear DNA (Guo et al., 2000; Olivier et al., 2010). A second hypothesis would be an error occurring during plastidic division (Renzaglia et al., 1994). Our data are insufficient to determine what mechanism is acting. It would require, among others, detailed population and genomic studies of the species concerned.

3.4.2 Phylogenetic relationships between American willows species

Chloroplastic trees

The chloroplastic tree (Fig. 2), using *Populus* and *Idesia* as outgroups, confirms the monophyly of genus *Salix* and the inclusion of *S. arbutifolia* (*Chosenia*) and *S. cardiophylla* (*Toisusu*) within the genus, as was shown by Chen et al. (2010). This tree also shows the separation of American *Salix* species into two major clades, as was also found by Azuma et al. (2000) and Chen et al. (2010) on Asian species (see Table 1). As in our study, one clade (equivalent to our clade A2) included species of subgenera *Chamaetia* and *Vetrix* (our A2 clade), and the other (our clade A1) subgenus *Salix* (no representative of subg. *Longifolia* and *Protitea* were included). Twenty species were shared between our study and that of Chen et al. (2010), 18 of which are found in the same clade in both analyses (shown by black stars in Fig. 2). Two taxa, *S. discolor* and *S. maccaliana*, were found in different clades, however. This could be explained by the fact that the two species are polyploid, 4x and 10x, respectively (Argus 2010); it could also be the result of intra-specific variability or chloroplast capture

following hybridization, or of an error of identification or manipulation. Despite differences possibly caused by the taxonomy used (see below), Hardig et al. (2010), working on American species (Table 1), also retrieved two similar clades. Resolution within clades is low and poorly supported, resulting in polytomies. Low rates of evolution of the plastid genome in willow species could explain this. The positions of *S. petrophila* (clade A1) and *S. lasiandra* (clade A2) are surprising, both being early divergent in each clade; this cannot be readily explained with current data.

ITS tree

The *ITS* region shows more variation than the cp DNA markers, but resolution of the tree was not greatly improved. *ITS* trees in Leskinen and Alström-Rapaport (1999), Hardig et al. (2010) and Abdollahzedeh et al. (2011), built respectively using parsimony, maximum likelihood and MrBayes, were similar to our analyses (not shown) carried out with the same approaches: a large polytomy is retrieved, with a single small clade comprised of species belonging to subg. *Salix*, *Longifolia* and *Protitea*. Our BEAST tree (Fig. 3), however, provided greater resolution, identifying four clades: species of subgenus *Protitea* in clade B1, subgenus *Salix* in B2, and subgenus *Longifoliae* in B3, with subgenera *Chamaetia* and *Vetrix* species intermixed in the large clade B4. All North American species used by Hardig et al. (2010) were also included in our study (shown by black stars in Fig. 3). Differences were observed in the placement of a few species. For instance, Hardig et al. included specimens of *S. eriocephala* and *S. lucida* from Idaho, species that Argus (2010) do not report for this area; this may result from the taxonomy used, since varieties sometimes attributed to *S. eriocephala* in western North America, for instance, are considered distinct species by Argus.

Incongruence between the chloroplastic and ITS trees

Overall, clades A2 (cp) and B4 (*ITS*) include species of subg. *Chamaetia* and *Vetrix*, while the species of clade A1 (cp) coincide with those in clades B1, B2 and B3 (*ITS*) (Figs. 2-3). The global structure of the trees is similar. There is significant incongruence however, for 14 taxa. Ten species of clade A1 in the cp tree (Fig. 2) were found in clade B4 on the *ITS* tree (Fig. 3). Conversely, four species of clade A2 (cp) were retrieved in clade B4. These taxa are highlighted in our trees. Five hypotheses could explain these incongruences. Firstly, plastid or ribosome capture following hybridization could have occurred. For instance, *S. pellita* (clade B4) can form natural hybrids in nature with *S. alaxensis* (clade B1) (Argus, 2010). Secondly, part of the chloroplast genome of one parent could have migrated to the nucleus in allopolyploid taxa, a rare but not impossible phenomenon (Timmis et al., 2004). Thirdly, horizontal gene transfer from another species is possible (Stegemann et al., 2012). Fourth, chloroplast fusion may occur, though it is rarely documented (Birky, 2001). And finally, field or laboratory errors could have happened, which seems improbable given the number of taxa involved.

Species tree

The species tree (Fig. 4) results from the simultaneous *BEAST analysis of the cpDNA and *ITS* datasets. More resolution is apparent on this tree. The topology confirms the presence of two major clades (C1 and C2), such as described above. Within clade C1, species are mostly grouped according to the subgenera where they are assigned by Argus (2010), i.e., subg. *Longifoliae* (subclade C5), *Protitea* (subclade C3), and *Salix* (subclade C7, also retrieved in Chen et al., 2010), which are well supported. A few species (discussed below) appear in novel

positions with respect to Argus (2010). Subgenera *Chamaetia* and *Vetrix* (clade C2) form one group, which corresponds to the observations of Chen et al. (2010). There is no clear pattern of subgeneric segregation in clade C2. All subclades within C2 have low support and cannot be substantiated. Skvortsov (1968), discussing Russian material, also indicated that the distinction between these subgenera was difficult, while Dorn (1976) only recognized subg. *Vetrix*. The branching order and groupings observed in this analysis are similar to those obtained in a morphology-based, numerical analysis of *Salix* by Argus (1997), if one excepts the position of subg. *Chosenia*, which groups in C2 in our analysis and the equivalent of C1 in that of Argus.

Fourteen species had different positions in the *ITS* and cp DNA analyses. They are indicated by a ! in figures 2-4. Seven species, *Salix cascadensis*, *S. discolor*, *S. eriocephala*, *S. humilis*, *S. microphylla*, *S. petiolaris* and *S. sericea*, form a subclade (C4) within clade C1 in the analysis. Yet, the morphology of these species is heterogeneous (Argus, 2010) and no morphologic character appears to support such a group. Also in clade C1, *S. jaliscana* and *S. nigra* are grouped, both of which belong to subg. *Protitea* (Argus 2010), which would make subg. *Protitea* paraphyletic to subg. *Longifoliae* and the artificial subclade C4. All these species were in clade A1 in the cp tree and in clade B2 in the *ITS* tree. The grouping of species in subclade C4 suggests a random grouping of species with similar behaviors. Subclade C4 appears artificial. *Salix petrophila* (subg. *Chamaetia*) appears to be an early diverging branch of clade C1. This species occupied different positions in the *matK* and *rbcL* trees (not shown). Similarly, clade C6, comprised of *S. cana*, *S. lasiandra*, and *S. pellita*, form an early diverging group sister to clade C2. *Salix setcheliana* is an early diverging branch within clade C2. All

these species were in clade A2 in the cp tree and in clade B1 in the *ITS* tree. In all instances, it appears as if the position in the species tree is determined primarily by the cp DNA.

Subgenus attribution of three species

The species tree (Fig. 4) shows that three species *Salix floridana* (subg. *Protitea*), *S. maccalliana* and *S. triandra* (both subg. *Salix*), probably are assigned to the wrong subgenus. Our data suggest that *S. floridana* belongs to subgenus *Salix*, where it would be sister to the other species. Chen et al. (2010) also found a similar position for *S. floridana*. The composition of subg. *Protitea* (*S. amygdaloides*, *S. bonplandiana*, *S. caroliniana*, *S. gooddingii*, *S. humboltiana*) has been discussed repeatedly (Schneider, 1919; Rehder, 1949; Ball, 1961), without consensus. Dorn (1976) proposed the exclusion of *S. floridana* from this subgenus, placing it instead in either subg. *Salix* or *Vetrix*. He hypothesized that the morphological similarities (bud scales distinct, flowers with 3 to 7 stamens) of this species to subgenus *Protitea* was the result of hybridization (Dorn, 1976). Argus (1986a, 2010) classified species of subgenus *Protitea* together because they share many morphological traits. The branching of the species tree (Fig. 4) suggests that the morphological similarities underlined by Dorn could be symplesiomorphic and not the result of hybridization. Chmelar (1978) proposed that ovule number could be taxonomically significant. *Salix floridana* and *S. babylonica* have 2 ovules per carpel, *S. alba* 3 to 6, and *S. amygdaloides*, *S. caroliniana* and *S. nigra* 6 to 9 (Argus, 1986a). Low ovule number could be a feature of sect. *Salix*.

Our data suggest that *S. maccalliana* and *S. triandra* belong in a large subgenus *Vetrix* (see below). *Salix maccalliana* is decaploid or dodecaploid (Argus, 2010), which indicates a complex origin. Its morphology is similar to that of *S. lucida* (subg. *Salix*) (Argus, 2010). The

staminate flowers with abaxial nectaries and tawny, persistent bracts, and the villous ovaries, however, suggest relationship with *S. glauca* (subg. *Chamaetia*). Dorn (1976) placed this species in subgenus *Vetrix*. Chen et al. (2010) included *S. maccalliana* in subgenus *Salix*. The provenance of their sample appears geographically suspect, however. In the case of *S. triandra*, both our study and those of Leskinen and Alström-Rapaport (1999) and Chen et al. (2010) that it belongs with subg. *Vetrix*. The latter fully discussed this issue and indicated that *S. triandra* could be considered to belong to a distinct subgenus. In our tree (Fig. 4), however, *S. triandra* falls fully within subg. *Vetrix*.

Placement of Salix ballii *and* Salix irrorata

Salix ballii and *S. irrorata* are assigned to subgenus *Vetrix* by Argus (2010). Our data, however, show that *S. ballii* is related with subg. *Protitea*, and *S. irrorata* with subg. *Salix*. Two distinct specimens were sequenced for each species with the same result. One hypothesis would be the capture of plastid or ribosomal DNA after hybridization. A laboratory error cannot be excluded. At this time, data are insufficient to explain these placements.

3.4.3 Biogeography of *Salix*

Formal biogeographic analyses (DEC, Ree and Smith, 2008, or DIVA, Nylander et al., 2008) could not be carried out with our tree due notably to the lack of resolution in clade C2 (Fig. 4). Doing a calibrated datation was not feasible at this time due to a lack of verified *Salix* fossil material that could be accurately placed on the tree topology. Nonetheless, the patterns observed on figure 4 allow the formulation of biogeographic hypotheses.

Overall, most species of clade C1 are found in temperate regions of both North America (western and eastern) and Eurasia, while and those of clade C2 include mostly species from boreo-arctic regions or montane areas southward. Globally, distribution patterns in *Salix* reflect well the biogeographic regions delimited by Takhtajan (1986) within the Holarctic kingdom, Boreal subkingdom: C1 taxa are predominantly in the Eastern Asiatic, North American Atlantic and Rocky Mountain regions (= subg. *Salix*, see below), and C2 taxa in the Circumboreal region (= subg. *Vetrix*, see below).

3.4.4 Classification of *Salix*

Skvortsov (1968) divided the species of the former Soviet Union and Asia into three subgenera, Dorn (1976) the American species into two subgenera, and Argus (1997, 2010) American taxa into five subgenera (*Longifoliae*, *Protitea*, *Salix, Chamaetia* and *Vetrix*). Our molecular phylogenetic study and that of Chen et al. (2010) show a primary subdivision of *Salix* into two clades (Fig. 4), the latter pointing out that the number of subgenera proposed for *Salix* was too high. The studies by Leskinen and Alström-Rapaport (1999), Azuma et al. (2000), Hardig et al. (2010), and Abdollahzedeh et al. (2011) also suggest such a division. We are proposing to divide Salix into two subgenera, *Salix* and *Vetrix*. Three sections may be recognized within subgenus *Salix*: *Salix, Protitea,* and *Longifoliae*, the latter American only. Within subgenus *Vetrix,* lack of resolution prevents the definition of sections at this time. *Salix arbutifolia* (*Chosenia*) and *S. cardiophylla* (*Toisusu*) are definitely members of subg. *Vetrix*, possibly as an early branch including other willows, a group that may deserve

sectional recognition. Another unresolved issue is the definitive position of the 16 problematic species discussed above, notably the clade within subg. *Salix*.

The three molecular regions used in our study are the markers selected for the barcoding of plants. The degree of variation of these molecular markers in *Salix*, however, is insufficient to provide species identification in subg. *Vetrix*, and as other studies have shown (Hollingsworth, 2011; Percy and Graham, 2011), other regions will need to be developed for full barcoding of willows.

We present the first complete phylogeny of willows for the Americas, based on three molecular markers from chloroplastic and nuclear ribosomal DNA. We are confirming the subdivision of genus *Salix* species into two clades that correspond to two subgenera proposed earlier on the basis of morphologic and molecular studies. Nonetheless, relationships among species remain tentative due to a lack of resolution within subg. *Vetrix* and to the unusual relationships exhibited by a 16 problematic species. Further phylogenetic analyses using low-copy nuclear genes should help address this lack of resolution and membership issues, and help in obtaining a tree that could be the object of formal biogeographic analyses. The challenge presented in this genus by hybridization and polyploidy may be resolved by phylogeographic analyses of species complex, such as was done by Tsai and Carstens (2013). Our phylogenetic analysis provides a framework to interpret data from other fields of study, such as eco-physiology and the development of willows for economic usages, such as biomass production.

Table 1. Summary of molecular phylogenies of the genus *Salix*. The molecular regions investigated are indicated. The number of species has been evaluated for each study (total number used and number specifically sequenced). The main native area of the species involved is indicated.

Reference	Molecular region	Number of species		Main native area
		Used in study	Sequenced for study	
Leskinen and Alström-Rapaport (1999)	*ITS* region	13	13	Europe
Azuma et al. (2000)	*rbcL* gene	19	19	Asia
Chen et al. (2010)	*rbcL* gene; *atpB-rbcL* spacer; *trnD-T* spacer	46[a]	32[b]	Asia
Hardig et al. (2010)	*ITS* region; *matK* gene	25	25[c]	North America
Abdollahzadeh et al. (2011)	*ITS* region; *trnL-F* region	57[d]	26[e]	Iran
This study (2013)	*ITS* region; *matK* gene; *rbcL* gene	123	122	North America

a : 45 species used with *rbcL* analysis. 31 species used in strict consensus of combined *rbcL*, *atpB-rbcL* and *trnD-T* analysis. *Salix babylonica* f. *rokkaku* was excluded from the count.

b : For 4 species, the sequences of rbcL were from Azuma et al. (2000) and the spacer region was sequenced from the same specimens.

c : The sequences of *matK* for *S. exigua* and *S. interior* were from Brunsfeld et al. (2007). The *ITS* sequences of *S. arctica* and *S. discolor* are not available in GenBank.

d : *trnL-F* only for 14 species. *Salix alba* f. alba and *Salix* sp. (unidentified) were excluded from the count.

e : For *ITS*; only 6 species were sequenced for *trnL-F* (no analysis published). *Salix* sp. (unidentified) was excluded from the count.

Table 2. Primers and PCR cycle characteristics for the three genes used in the study, indicating: source of primers; number of sequences obtained and their length (bp); percentage of polymorphic sites (including and excluding outgroup).

Gene	Source	Primer name	Sequence 5'-3'	Number of cycles (C_a)	Elongation time (s)	Number of sequences	Alignment length (bp)	Polymorphic sites	
								With outgtoup	Without outgroup
ITS	this paper	ALM-P001	F : CGTAACAAGGTTTCCGTAGG	35	60	210	608	125 (21%)	95 (16%)
		ALM-P002	R : TGCTTAAACTCAGCGGGTAG						
matK	Ford et al., 2009	matK X	F : TAATTTACGATCAATTCATTC	33	70	212	874	106 (12%)	50 (6%)
	Kew Barcoding	matK_Equisetum	R : GTACTTTTATGTTTACGAGC						
rbcL	Levin et al., 2003	P1630	F : ATGTCACCACAAACAGAGACTAAAGC	33	45	212	553	44 (8%)	20 (4%)
	Kress et al., 2009	rbcLa-R	R : GTAAAATCAAGTCCACCRCG						

Figure 1

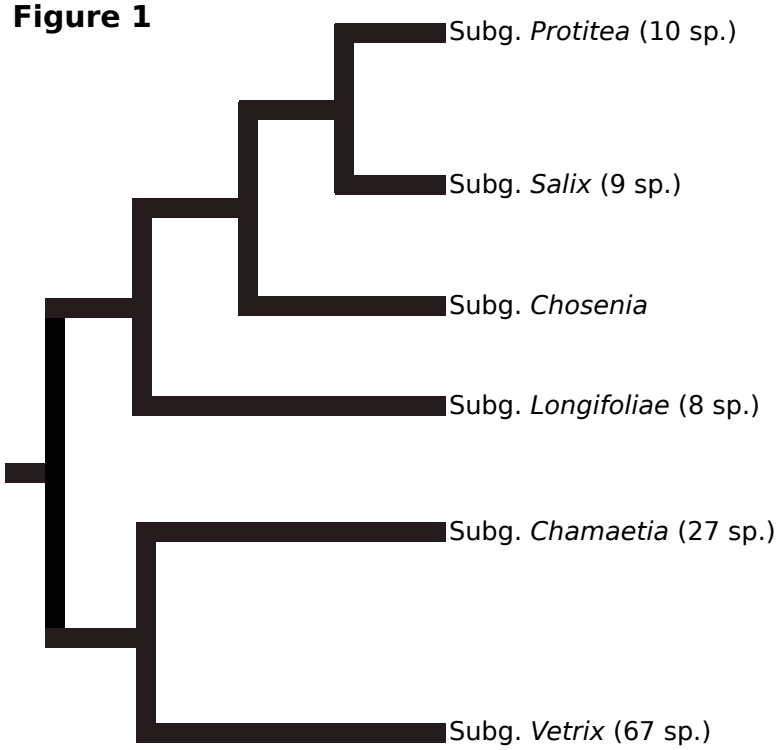


Figure 1. Relationships between the five subgenera of *Salix* in North America based on morphological characters (Argus, 1997). The number of species per subgenus is indicated. The Asian subgenus *Chosenia* was added to show its position.

Figure 2

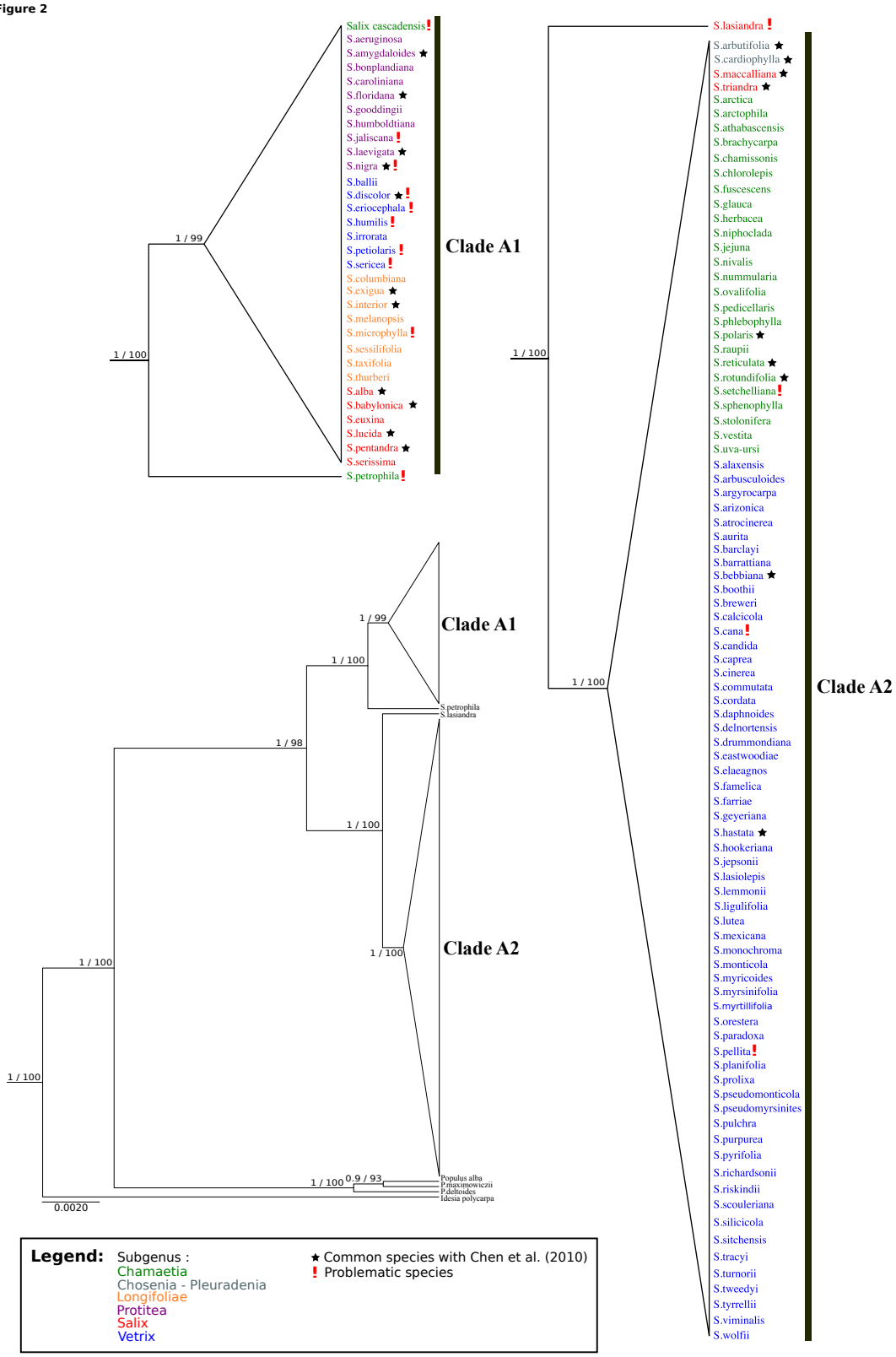


Figure 2. BEAST gene tree of *matK* and *rbcL*. Branch support is Bayesian posterior probabilities and ML bootstrap values; subgenera are identified using colors; *Idesia* and *Populus* are outgroups.

Figure 3

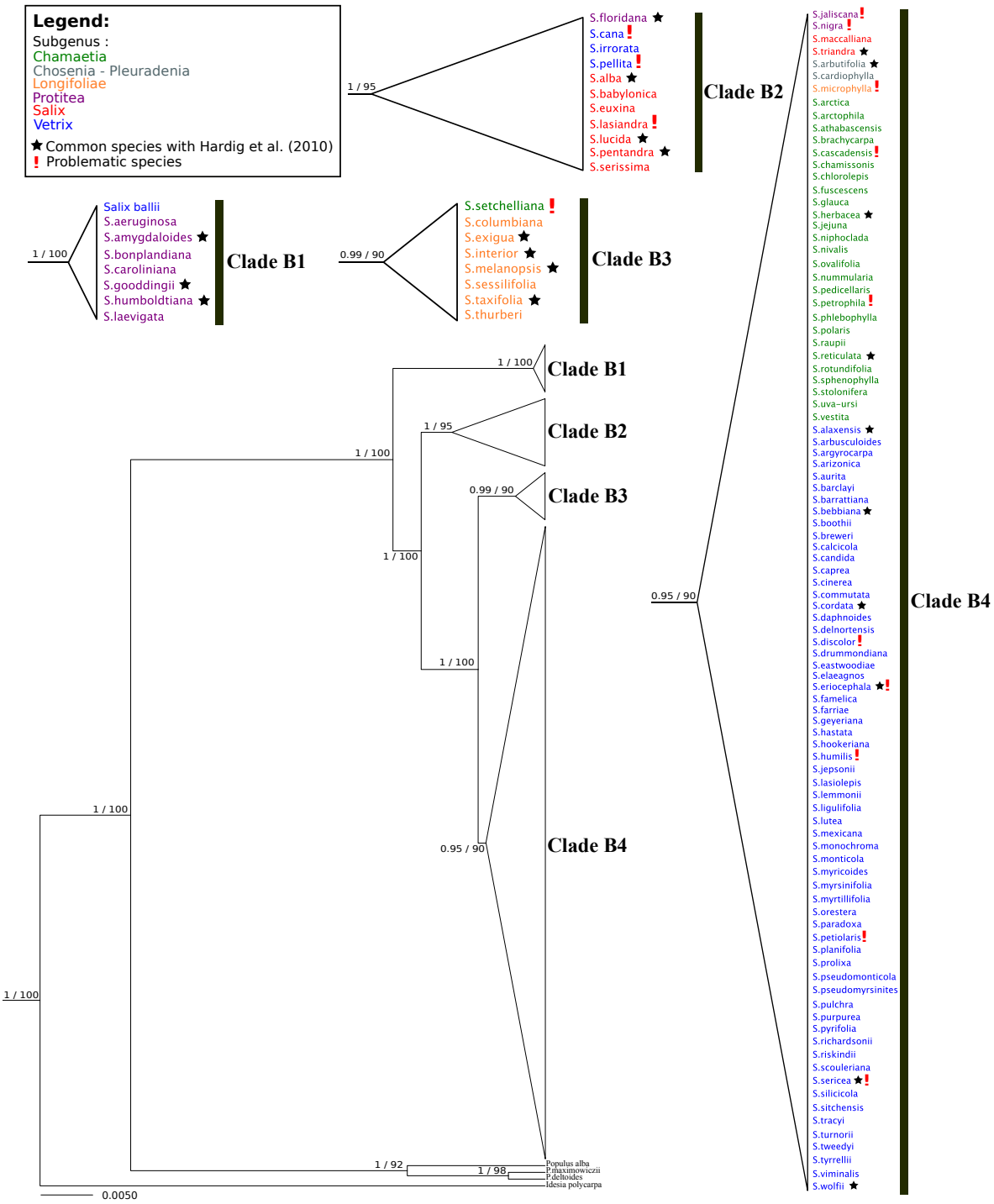


Figure 3. BEAST gene tree of *ITS*. Branch support is Bayesian posterior probabilities and ML bootstrap values; subgenera are identified using colors; *Idesia* and *Populus* are outgroups.

Figure 4

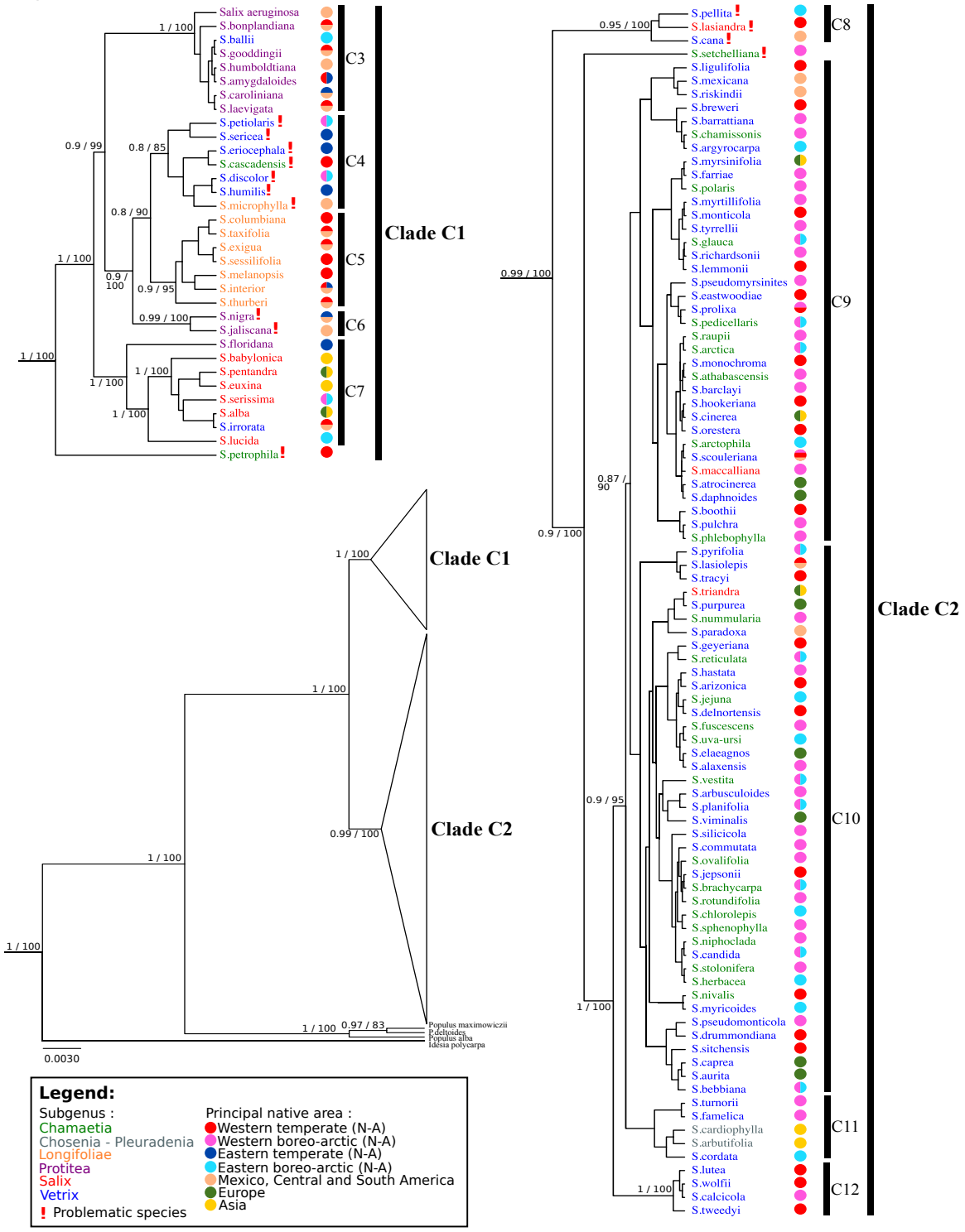


Figure 4. *BEAST species tree generated with *ITS*, *matK* and *rbcL*. Branch support is Bayesian posterior probabilities and ML bootstrap values; subgenera are identified using colors; native areas (Argus 2007) are indicated by colored circles; *Idesia* and *Populus* are outgroups.

4. Conclusion et perspectives

Ce projet de recherche a été mené à terme et les résultats majeurs s'inscrivent dans deux publications (Chapitre 1 et Chapitre 2). Nous revenons ici brièvement sur les conclusions qui se dégagent de ces deux études.

4.1 Outils moléculaires développés durant le projet

La méthode d'extraction d'ADN génomique a été optimisée au début de ce projet de recherche. Ngantcha (2010) a souligné que l'extraction d'ADN chez le saule est une étape critique en raison du fort taux de polysaccharides dans les parties foliaires. Nous avons adapté la méthode CTAB avec un broyage mécanique et en incorporant du β-mercaptoethanol et PVP lors de l'extraction (Annexe 1). Les avantages de ce protocole sont sa reproductibilité, son faible coût et le grand nombre d'échantillons traités par manipulation (jusqu'à 192). Cette technique a été utilisée dans les expériences décrites aux chapitres 1 et 2 (Lauron-Moreau et al., 2013; Lauron-Moreau et al., *en préparation*) avec des échantillons provenant de spécimen d'herbier ou de feuilles fraiches séchées dans du gel de silice.

Douze nouveaux marqueurs microsatellites sur cinq espèces du genre *Salix* (*S. amygdaloides*, *S. discolor*, *S. eriocephala*, *S. interior* et *S. nigra*) ont été caractérisés dans ce projet. Les espèces ont été sélectionnées pour leur potentiel en production de biomasse au Canada (Mosseler, com. pers.). Les résultats de cette étude permettront aux futurs projets de bénéficier d'éléments techniques pour réaliser des études en génétique des populations sur ces cinq espèces. Également, ces marqueurs pourront être utilisés indirectement dans les programmes de sélection. Les marqueurs microsatellites permettent d'obtenir des cartes

génétiques fines (ex. Berlin et al., 2010) sur lesquelles des QTLs peuvent être positionnés (ex. Rönnberg-Wästljung et al., 2008). Au *Rothamsted Research*, de nouvelles lignées élites de saule ont été sélectionnées par la présence de QTL d'intérêt (Karp et al., 2011). La sélection d'espèces canadiennes est une voie intéressante, car des espèces avec une forte production de biomasse sont identifiées et de nombreux outils moléculaires ont fait leur preuve dans les programmes de sélection européens.

Les résultats du chapitre 2 ont également permis de déterminer les relations phylogénétiques entre les saules américains en utilisant trois marqueurs moléculaires (*ITS*, *matK* et *rbcL*). Cependant, les affinités génétiques entre les espèces du sous-genre *Vetrix* doivent être approfondies. L'ajout de marqueurs chloroplastiques n'est pas à prioriser, car l'étude de Chen et al. (2010) n'a pas obtenu plus de résolution en utilisant deux autres régions intergéniques (*atpB-rbcL* et *trnD-T*). Les gènes nucléaires en simple copie représentent une alternative intéressante. L'étude de Duarte et al. (2010) a mis en évidence une dizaine de gènes nucléaires possédant un potentiel en phylogénie. Cinq de ces régions ont fait l'objet d'une étude préliminaire pour évaluer leur potentiel chez le saule. La démarche méthodologique est présentée en Annexe 2. Les travaux sont au stade de développement, mais les premiers résultats sont encourageants. Le polymorphisme est supérieur à celui de l'*ITS* ou des deux gènes chloroplastiques (*matK* et *rbcL*). De plus, des espèces pourraient être distinguées par des SNPs spécifiques (Annexe 2). La caractérisation de ces cinq nouveaux gènes nucléaires offre des perspectives pour augmenter la résolution de la phylogénie des saules.

4.2 Phylogénie moléculaire du genre *Salix*

Les résultats mis en évidence au chapitre 2 nous ont permis de répondre à l'hypothèse du projet. Nous avions proposé que la classification des espèces américaines, selon les données moléculaires, se divisaient en cinq clades. Ces clades correspondraient aux cinq sous-genres (*Chamaetia*, *Longifoliae*, *Protitea*, *Salix* et *Vetrix*) décrits par Argus (1997, 2010).

Au chapitre 2, nous avons montré la présence de deux clades principaux. Le premier se divise en trois sous-clades. La structure interne du second clade n'est pas suffisamment supportée pour différencier avec certitude des sous-clades. Par conséquent, la classification d'Argus (1997, 2010) est imprécise dans sa hiérarchisation en cinq sous-genres. Cependant, le regroupement des espèces prédit par Argus est respecté au sein du premier clade.

Nous proposons une réorganisation pour le genre *Salix* en deux sous-genres, *Salix* et *Vetrix*. Celle-ci est basée sur les espèces américaines, mais nous pouvons généraliser à l'ensemble du genre car les résultats des précédentes études tendent à soutenir une telle structure (Azuma et al., 2000; Leskinen et Alstrom-Rapaport, 1999; Chen et al., 2010; Hardig et al., 2010; Abdollahzadeh et al., 2011). Dans le sous-genre *Salix*, nous avons identifié trois sections (*Longifolia*, *Protitea* et *Salix*). Globalement, la répartition des espèces dans ces trois sections respecte la classification de Argus (1997, 2010). Le sous-genre *Vetrix* comprend la majorité des espèces assignées à *Chamaetia* et *Vetrix* (Argus, 1997, 2010). Ce sous-genre ne montre pas de structure avec nos données moléculaires et des analyses supplémentaires seront nécessaires pour y identifier des sections. La classification proposée dans nos travaux se rapproche de celles proposées par Skvortsov (1968) et Dorn (1976).

À l'aide des répartitions des espèces fournies par Argus (2007), nous avons proposé des hypothèses expliquant les événements biogéographiques au sein des deux sous-genres, *Salix* et *Vetrix*. Cependant, une datation avec des fossiles est nécessaire pour vérifier ces hypothèses. De plus, des efforts dans le développement de nouveaux marqueurs devraient se poursuivre afin d'améliorer la résolution des relations entre espèces du sous-genre *Vetrix*.

4.3 Perspectives

Ce projet de recherche a permis de développer de nouveaux outils moléculaires chez les saules, en particulier autour des espèces nord-américaines. De plus, une phylogénie moléculaire globale a été développée et cette étude présente un fort atout pour comprendre et élaborer de nouvelles hypothèses au sein du genre *Salix* et de la famille des Salicaceae. À long terme, les éléments du projet pourront être utilisés pour continuer à accroitre les connaissances au sein de genre *Salix* et à caractériser les cultivars commercialisés. De plus, les données génétiques publiées par ce projet sont des éléments supplémentaires pour évaluer la corrélation entre les liens génétiques et d'autres paramètres (ex. données écologiques ou morphologiques). Par exemple, l'étude de Bell et al. (*en soumission*) a utilisé les données de la région *ITS* sur onze cultivars pour déterminer si les flores microbienne et fongique du sol vont s'associer avec certains groupes de saule.

Sources documentaires (bibliographie)

ABDOLLAHZADEH, A., K. OSALOO, AND A. A. MAASSOUMI. 2010. Molecular phylogeny of the genus *Salix* (Salicaceae) with an emphasize to its species in Iran. *The Iranian Journal of Botany* 17(2): 244-253.

ALFORD, M. H., R. J. BRANTLEY, C. L. HERNANDEZ, AND T. SAMARAKOON. 2009. What are the closest relatives of *Salix* and *Populus*? Botany and Mycology 2009, Snowbird, Utah, USA [online abstract : http://2009.botanyconference.org/engine/search/index.php?func=detail&aid=786].

ALVAREZ, I., AND J. F. WENDEL. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417-434.

ARGUS, G. W. 1964. Preliminary reports on the flora of Wisconsin. No. 51. *Salicaceae*, The genus *Salix* - willows. *Wisconsin Academy of Sciences, Arts and Letters* 53: 217-272.

ARGUS, G. W. 1965. The taxonomy of the *Salix glauca* L. complex in North America. *Contributions from the Gray Herbarium of Harvard University* 196: 1-142.

ARGUS, G. W. 1969. Some new combinations in *Salix*. *Canadian Journal of Botany* 47: 795-801.

ARGUS, G. W. 1973. The Genus *Salix* in Alaska and the Yukon. Ottawa. *National Museum of Canada* vol. 2.

ARGUS, G. W. 1974. An experimental study of hybridization and pollination in *Salix* (willows). *Canadian Journal of Botany* 52: 1613-1619.

ARGUS, G. W. 1983. *Salix*. In E.H. Moss, revised by J.G. Packer [eds.]. Flora of Alberta. University of Toronto Press, ON, Canada.

ARGUS, G. W. 1986a. The genus *Salix* in the southeastern United States. *Systematic Botany Monographs* 9: 1-170.

ARGUS, G. W. 1986b. Studies in the *Salix lucida* Muhl. and *S. reticulata* L. complexes in North America. *Canadian Journal of Botany* 64: 541-551.

ARGUS, G. W. 1991. *Salicacae*, The vascular plants of British Colombia. *Victoria: B.C. Ministry of Forests, Special Report Series* 3: 55-67.

ARGUS, G. W. 1993. *Salix*. In J.C. Hickman [ed.]. The Jepson manual. Higher plants of California, Berkeley, University of California Press, CA, USA.

ARGUS, G. W. 1995. Arizona Salicaceae: Willow family. Part two: *Salix*. *Journal of Arizona-Nevada Academy of Sciences* 29: 39-62

ARGUS, G. W. 1997. Infrageneric classification of New World *Salix* L. (Salicaceae). *Systematic Botany Monographs* 52: 1-121.

ARGUS, G. W. 2007. *Salix* (Salicaceae) distribution maps and a synopsis of their classification in North America, north of Mexico. *Harvard Papers in Botany* 12 (2): 335-368.

ARGUS, G. W. 2010. *Salix*. *In* Flora North American Editorial Committee [eds.], Flora of North America: Magnoliophyta: Salicaceae to Brassicaceae, vol. 7, 23-51. Oxford University Press, United Kingdom.

ARGUS, G. W., AND C. L. MCJANNET, 1992. A taxonomic reconsideration of*Salix taxifolia sensu lato* (Salicaceae). *Brittonia* 44: 461-474.

ARGUS, G. W., AND J. W. STEELE. 1979. A reevaluation of the taxonomy of *Salix tyrrellii*, a sand dune endemic. *Systematic Botany Monographs* 4: 163-177.

AUSUBEL, J. H. 2009. A botanical macroscope. *Proceedings of the National Academy of Sciences of the United States of America* 106: 12569-12570.

AZUMA, T., T. KAJITA, J. YOKOYAMA, AND H. OHASHI. 2000. Phylogenetic relationships of *Salix* (Salicaceae) based on *rbcL* sequence data. *American Journal of Botany* 87: 67-75.

BARKER, J. H. A., AND A. KARP. 1998. Detecting genetic diversity of willow (*Salix* spp.) using amplified fragment length polymorphism (AFLP) and microsatellite marker analyses. *Life Science News* I: 1-3.

BARKER, J. H., A. PAHLICH, S. TRYBUSH, K. J. EDWARDS, AND A. KARP. 2003. Microsatellite markers for diverse *Salix* species. *Molecular Ecology Notes* 3(1): 4-6.

BAILEY, C. D., T. G. CARR, S. A. HARRIS, AND C. E. HUGHES. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* 29: 435-455.

BALDWIN, B. G., M. J. SANDERSON, J. M. PORTER, M. F. WOJCIECHOWSKI, C. S. CAMPBELL, AND M. J. DONOGHUE. 1995. The Its Region of Nuclear Ribosomal DNA - a Valuable Source of Evidence on Angiosperm Phylogeny. *Annals of the Missouri Botanical Garden* 82: 247-277.

BALL, C. R. 1961. *Salix*. In Lundell C. L. et al. [eds.], *Flora of Texas*, vol. 3(6): 369-392. Renner, Texas, USA.

BERLIN, S. U. LAGERCRANTZ, S. VON ARNOLD, T. OST, AND A. C. RONNBERG-WASTLJUNG. 2010. High-density linkage mapping and evolution of paralogs and orthologs in *Salix* and *Populus*. *BMC Genomics* 11:129.

BERLIN, S., J. FOGELQVIST, M. LASCOUX, U. LAGERCRANTZ, AND A. C. RONNBERG-WASTLJUNG. 2011. Polymorphism and Divergence in Two Willow Species, *Salix viminalis* L. and *Salix schwerinii* E. Wolf. *G3: Genes, Genomes, Genetics*, 1(5): 387-400.

BIRKY, C. W. 2001. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review of Genetics* 35: 125-148.

BOUCHER, L. D., S. R. MANCHESTER, AND W. S. JUDD. 2003. An extinct genus of Salicaceae based on twigs with attached flowers fruits, and foliage from the Eocene Green River Formation of Utah and Colorado, USA. *American Journal of Botany* 90: 1389-1399.

BRERETON, N. J. B., F. E. PITRE, S. J. HANLEY, M. J. RAY, A. KARP, AND R. J. MURPHY. 2010. QTL Mapping of Enzymatic Saccharification in Short Rotation Coppice Willow and Its Independance from Biomass Yield. *Bioenergy Resources* 3: 251-261.

BRUNSFELD, S. J., P. S. SOLTIS, D. E. SOLTIS, P. A. GADEK, C. J. QUINN, D. D. STRENGE, AND T. A. RANKER. 1994. Phylogenetic-Relationships among the Genera of *Taxodiaceae* and *Cupressaceae* - Evidence from Rbcl Sequences. *Systematic Botany* 19: 253-262.

BRUNSFELD, S. J., T. R. MILLER, AND B. C. CARSTENS. 2007. Insights into the biogeography of the Pacific Northwest of North America: Evidence from the phylogeography of *Salix melanopsis*. *Systematic Botany* 32: 129-139.

CHASE, M. W., S. ZMARZTY, M. D. LLEDO, K. J. WURDACK, S. M. SWENSEN, AND M. F. FAY. 2002. When in doubt, put it in Flacourtiaceae: a molecular phylogenetic analysis based on plastid *rbcL* DNA sequences. *Kew Bulletin* 57: 141-181.

CHASE, M. W., R. S. COWAN, P. M. HOLLINGSWORTH, C. VAN DEN BERG, S. MADRINAN, G. PETERSEN, O. SEBERG, T. JORGSENSEN, K. M. CAMERON, M. CARINE, N. PEDERSEN, T. A. J. HEDDERSON, F. CONRAD, G. A. SALAZAR, J. E. RICHARDSON, M. L. HOLLINGSWORTH, T. G. BARRACLOUGH, L. KELLY, AND M. WILKINSON. 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon* 56: 295-299.

CHEN, J. H., H. SUN, J. WEN, AND Y. P. YANG. 2010. Molecular phylogeny of *Salix* L. (Salicaceae) inferred from three chloroplast datasets and its systematic implications. *Taxon* 59: 29-37.

CHMELAR, J. 1978. Taxonomic importance of bud scale in the *Salix* genus. *Folia Dendrologica*, 4: 5-21.

COMMITTEE ON BIOBASED INDUSTRIAL PRODUCTS. 2000. Biobased Industrial Products: Priorities for Research and Commercialization. *National Research Council*, Washington, DC, USA.

CROVELLO, T. J. 1968. A numerical taxonomic study of the genus *Salix*, section *Sitchenses*. *University of California Botany* 44: 1-61.

DARRIBA, D., G. L. TABOADA, R. DOALLO, AND D. POSADA. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772-772.

DAVIS, C. C., AND W. R. ANDERSON. 2010. A Complete Generic Phylogeny of Malpighiaceae Inferred from Nucleotide Sequence Data and Morphology. *American Journal of Botany* 97: 2031-2048.

DEMIRBAS, A. 2001. Biomass resource facilities and biomass conversion processing for fuelds and chemicals. *Energy Conversion and Management* 42: 1357-1378.

DICKMAN, D. I. 2001. An overview of the genus Populus. *In* D. I. Dickman, J. G. Isebrands, J. E. Eckenwalder, and J. Richardson [eds.], Poplar culture in North America. NRC Research Press, Ottawa, ON, Canada.

DORN, R. D. 1976. A synopsis of American *Salix*. *Canadian Journal of Botany* 54(24): 2769-2789.

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

DRUMMOND, A. J., M. A. SUCHARD, D. XIE, AND A. RAMBAUT. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969-1973.

DUARTE, J. M., P. K. WALL, P. P. EDGER, L. L. LANDHERR, H. MA, J. C. PIRES, AND J. LEEBENS-MACK. 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* 10(1), 61.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792-1797.

FANG, Z. F., S. D. ZHAO, AND A. K. SKVORTSOV. 1999. Salicaceae. In Z. Y. WU and P. H. Raven [eds.], Flora of China, vol. 4, 139-274. Missouri Botanical Garden Press, St. Louis, Missouri, USA.

FEAU, N., A. LAURON-MOREAU, D. PIOU, B. MARCAIS, C. DUTECH, AND M.-L. DESPREZ-LOUSTAU. 2012. Niche partitioning of the genetic lineages of the oak powdery mildew complex. *Fungal Ecology* 5: 154-162.

FORD, C. S., K. L. AYRES, N. TOOMEY, N. HAIDER, J. V. STAHL, L. J. KELLY, N. WIKSTROM, P. M. HOLLINGSWORTH, R. J. DUFF, S. B. HOOT, R. S. COWAN, M. W. CHASE, AND M. J. WILKINSON 2009. Selection of candidate coding DNA barcoding regions for use on land plants. *Botanical Journal of the Linnean Society* 159: 1-11.

FRIES, E. 1825. Uppställining af de Svenska Pilaterna. *Physiogr. Sällsk. Arsb*. 1: 30-37.

GALLAGHER, P. W., M. DIKEMAN, J. FRITZ, E. WAILES, W. GAUTHIER, AND H. SHAPOURI. 2003. Supply and Social Cost Estimates for Biomass from Crop Residues in the United States. *Environmental and Resource Economics* 24: 335-358.

GAUT, B. S. 1998. Molecular clocks and nucleotide substitution rates in higher plants. *Evolutionary Biology,* 30: 93-120.

GOUY, M., S. GUINDON, AND O. GASCUEL. 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution* 27: 221-224.

GUIDI, W., PITRE F.E., AND LABRECQUE M. 2013. Short-rotation coppice of willows for the production of biomass in Eastern Canada. In M. D. Matovic [eds.], Biomass Now – Sustainable Growth and Use, 17: 421-448. Tech Open Science.

GUINDON, S., J. F. DUFAYARD, V. LEFORT, M. ANISIMOVA, W. HORDIJK, AND O. GASCUEL. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* 59: 307-321.

GUINDON, S., AND O. GASCUEL. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696-704.

GUO, X., S. RUAN, W. HU, D. CAI, AND L. FAN. 2008. Chloroplast DNA insertions into the nuclear genome of rice: the genes, sites and ages of insertion involved. *Functional & integrative genomics* 8(2): 101-108.

HANLEY, S. J., M. H. PEI, S. J. POWERS, C. RUIZ, M. D. MALLOTT, J. H. BARKER, AND A. KARP. 2011. Genetic mapping of rust resistance loci in biomass willow. *Tree Genetics & Genomes* 7: 597-608.

HARDIG, T. M., C. K. ANTTILA, AND S. J. BRUNSFELD. 2010. A phylogenetic analysis of *Salix* (Salicaceae) based on *matK* and ribosomal DNA sequence data. *Journal of Botany*, vol. 2010, 1-12.

HELED, J. AND A. J. DRUMMOND. 2010. Bayesian inference of species trees from multilocus data. *Molecular biology and evolution*, 27(3): 570-580.

HILU, K. W., T. BORSCH, K. MULLER, D. E. SOLTIS, P. S. SOLTIS, V. SAVOLAINEN, M. W. CHASE, M. P. POWELL, L. A. ALICE, R. EVANS, H. SAUQUET, C. NEINHUIS, T. A. B. SLOTTA, J. G. ROHWER, C. S. CAMPBELL, AND L. W. CHATROU. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758-1776.

HOLLINGSWORTH, P. M. 2011. Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* 108: 19451-19452.

KARP A.,W. J. MACALPINE, AND I. F. SHIELD. 2001. Willow has advanced as an energy crop but is the UK reaping the benefits ? *Royal Agricultural Society of England* 171 :55-61.

KARP, A., S. J. HANLEY, S. O. TRYBUSH, W. MACALPINE, M. PEI, AND I. SHIELD 2011. Genetic Improvement of Willow for Bioenergy and BiofuelsFree Access. *Journal of Integrative Plant Biology* 53: 151-165.

KIMURA, A. 1928. Über Toisusu—eine neue Salicaceen-Gattung und die systematische Stellung derselben. *Botanical Magazine (Tokyo)* 42: 287-290.

KRESS, W. J., D. L. ERICKSON, F. A. JONES, N. G. SWENSON, R. PEREZ, O. SANJUR, AND E. BERMINGHAM. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences of the United States of America* 106: 18621-6.

KUPRIANOVA, L. A. 1965. The palynology of the Amentiferae. The Academy of Sciences of the USSR, Moscow, vol. 214.

KUZOVKINA, Y. A., M. WEIH, M. ABALOS ROMERO, J. CHARLES, S. HURST, I. MEIVOR, A. KARP, S. TRYBUSH, M. LABRECQUE, AND T. I. TEODORESCU. 2008. *Salix*: botany and global horticulture. *Horticulture Review* 34: 447-489.

KUZMINA, M. L., K. L. JOHNSON, H. R. BARRON, AND P. D. N. HEBERT. 2012. Identification of the vascular plants of Churchill, Manitoba, using a DNA barcode library. *BMC Ecology* 12: 25.

LAURON-MOREAU, A., F. E. PITRE, L. BROUILLET, AND M. LABRECQUE. 2013. Markers of Willow Species and Characterization of 11 Polymorphic Microsatellites for *Salix eriocephala* (Salicaceae), a Potential Native Species for Biomass Production in Canada. *Plants* 2(2): 203-210.

LINDEGAARD, K. N., AND J. H. A. BARKER. 1997. Breeding willows for biomass. *Aspects of Applied Biology* 49: 155-162.

LEDIN, S., AND A. ALRIKSSON. 1992. Handbook on How to Grow Short Rotation Forests. *IEA Bioenergy Agreement Task IV*, Swedish University of Argicultural Sciences, Section of Short Rotation Forestry, Uppsala, Sweden.

LESKINEN, E., AND C. ALSTROM-RAPAPORT. 1999. Molecular phylogeny of Salicaceae and closely related Flacourtiaceae: evidence from 5.8 S, *ITS* 1 and *ITS* 2 of the rDNA. *Plant Systematics and Evolution* 215: 209-227.

LEVIN, R. A., W. L. WAGNER, P. C. HOCH, M. NEPOKROEFF, J. C. PIRES, E. A. ZIMMER, AND K. J. SYTSMA. 2003. Family-level relationships of Onagraceae based on chloroplast rbcL and ndhF data. *American Journal of Botany* 90: 107-115.

McCAULEY, D. E., J. E. STEVENS, P. A. PERONI, AND J. A. RAVEILL. 1996. The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). *American Journal of Botany* 83: 727-731.

MEIER, B., Y. SHAO, R. JULKUNEN-TIITTO, A. BETTSCHART, AND O. STICHER. 1992. A chemotaxonomic survey of phenolic compounds in Swiss willow species. *Proceedings of the Royal Society of Edinburgh, Section B, Biological Sciences* 98: 229-232.

MOSS, C. E. 1914. The Cambridge British Flora, vol. 2. Cambridge University Press, England.

MITCHELL, C. P. 1994. Developments in the production and supply of wood for energy. *Renewable Energy 5*(5): 754-761.

MUHLENBERG, G. H. 1806. On North America willows. *Annals of Botany* 2: 62-69.

NAKAI, T. 1920. *Chosenia*, a new genus of Salicaceae. *Botanical Magazine (Tokyo)* 34: 66-69.

NEWSHOLME, C. 1992. Willows: the genus *Salix*. Timber Press Inc., Portland, Oregon, USA.

NGANTCHA, A. C. 2010. DNA fingerprinting and genetic relationships among willow (*Salix* spp.). University of Saskatchewan, Rapport de maitrise.

NYLANDER, J. A. A., U. OLSSON, P. ALSTROM, AND I. SANMARTIN. 2008. Accounting for phylogenetic uncertainty in biogeography: A Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves : Turdus). *Systematic Biology* 57: 257-268.

OHASHI, H. 2000. A systematic enumeration of Japanese *Salix* (*Salicaceae*). *Journal of Japan Botany* 75: 1-41.

OLIVER, M. J., A. G. MURDOCK, B. D.MISHLER, J. V. KUEHL, J. L. BOORE, D. F. MANDOLI, K. DE EVERETT, P. G. WOLF, A. M. DUFFY, AND K. G. KAROL. 2010. Chloroplast genome sequence of the moss *Tortula ruralis*: gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. *BMC genomics* 11: 143.

PEART, R., R. CURRY, J. JONES, K. BOOTE, AND L. ALLEN. 1993. Reducing the rate of carbon dioxide build up with biomass fuel under climate change. In *Proceeding of the first biomass conference of the Americas*, Burlington, VT, USA.

PREGITZER, K. S., AND A. L. FRIEND. 1996. The structure and function of *Populus* root systems. *In* R. E. Stettler, H. D. Bradshaw, P. E. Heilman, and T. M. Hinckley [eds.], Biology of *Populus* and its implications for management and conservation, 331-354. NRC Press (National Research Council of Canada), Ottawa, Ontario, Canada.

RAMBAUT, A., AND A. J. DRUMMOND. 2007. Tracer v1.4, website: http://beast.bio.ed.ac.uk/Tracer [accessed 15 January 2013].

RAUP, H. M. 1959. The willows of boreal western America. *Contributions of the Gray Herbarium of Harvard University* 185: 3-95.

REE, R. H., AND S. A. SMITH. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4-14.

REHDER, A. 1949. Bibliography of Cultivated Trees and Shrubs hardy in the cooler temperate regions of the northern hemisphere. Jamaica Plain, Massachusetts USA.

RENZAGLIA, K. S., A. R. MADEN, J. G. DUCKETT, AND D. P. WHITTIER. 1994. Monoplastidy in Spermatogenesis of Lycopodium-Obscurum. *Canadian Journal of Botany-Revue Canadienne De Botanique* 72 : 1436-1444.

ROGERS, S. O., AND A. J. BENDICH. 1987. Ribosomal-Rna Genes in Plants - Variability in Copy Number and in the Intergenic Spacer. *Plant Molecular Biology* 9: 509-520.

RONNBERG-WASTLJUNG, A. C., C. GLYNN, AND M. WEIH. 2005. QTL analyses of drought tolerance and growth for a *Salix dasyclados x Salix viminalis* hybrid in contrasting water regimes. *Theoretical and Applied Genetics* 110: 537-549.

RONNBERG-WASTLJUNG, A. C., I. AHMAN, C. GLYNN, AND O. WIDENFALK. 2006. Quantitative trait loci for resistance to herbivores in willow : field experiments with varying soils and climates. *Entomologia Experimentalis et Applicata* 118: 163–174.

RONQUIST, F., AND J. P. HUELSENBECK. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.

ROULEAU, E. 1970. Guide to Index *Kewensis* and its supplements *i-xiv*. Montréal: published by the author.

SAMILS, B., A. C. RONNBERG-WASTLJUNG, AND J. STENLID. 2011. QTL mapping of resistance to leaf rust in *Salix*. *Tree Genetics & Genomes* 7: 1219-1235.

SCHNEIDER, C. 1919. Notes on American willows. IV. Species and varieties of section *Longifoliae*. *Botanical gazette* 67(4): 309-346.

SKVORTSOV, A. K. 1968. Willows of the URSS. Nauka, Moscow [In Russian].

SKVORTSOV, A. K. 1999. Willows of Russia and adjacent countries. Joensuu: University of Joensuu, 307pp.

SNEADER, W. 2000. The discovery of aspirin: a reappraisal. *British Medical Journal* 321: 1591-1594.

STEGEMANN, S., M. KEUTHE, S. GREINER, AND R. BOCK. 2012. Horizontal transfer of chloroplast genomes between plant species. *Proceedings of the National Academy of Sciences of the United States of America* 109: 2434-2438.

STOTT, K. G. 1984. Improving the biomass potential of willow by selection and breeding. *In* K. L. Perttu [eds.], Ecology and Management of Forest Biomass Production Systems, 233-260. Swedish University of Agricultural Sciences, Uppsala, Sweden.

SUDA, Y., AND G. W. ARGUS. 1968. Chromosome numbers of some North American *Salix*. *Brittonia* 20: 191-197.

SWOFFORD, D. L. 1993. Paup - a Computer-Program for Phylogenetic Inference Using Maximum Parsimony. *Journal of General Physiology* 102: A9-A9.

TAKHTAJAN, A., AND T. J. CROVELLO. 1986. Floristic regions of the world. University of California press, Berkeley, California, USA.

THE ANGIOSPERM PHYLOGENY GROUP 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105-121.

TIMMIS, J. N., M. A. AYLIFFE, C. Y. HUANG, AND W. MARTIN. 2004. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* 5: 123-U16.

TRYBUSH, S., S. JAHODOVA, W. MACALPINE, AND A. KARP. 2008. A Genetic Study of a *Salix* Germplasm Resource Reveals New Insights into Relationships Among Subgenera, Sections and Species. *Bioenergy Research* 1: 67-79.

TSAI, Y. H. E., AND B. C. CARSTENS. 2013. Assessing model fit in phylogeographical investigations: an example from the North American sandbar willow Salix melanopsis. *Journal of Biogeography* 40: 131-141.

TSAROUHAS, V., U. GULLBERG, AND U. LAGERCRANTZ. 2002. An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in *Salix*. *Theoretical and Applied Genetics* 105: 277-288.

TSAROUHAS, V., U. GULLBERG, AND U. LAGERCRANTZ. 2003. Mapping of quantitative trait loci controlling timing of bud flush in *Salix*. Hereditas 138 :172-178.

TSAROUHAS, V., U. GULLBERG, AND U. LAGERCRANTZ. 2004. Mapping of quantitative trait loci (QTLs) affecting autumn freezing resistance and phenology in *Salix*. *Theoretical and Applied Genetics* 108: 1335-1342.

TUSKAN, G. A., S. DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV, U. HELLSTEN, N. PUTNAM, S. RALPH, S. ROMBAUTS, A. SALAMOV, J. SCHEIN, L. STERCK, A. AERTS, R. R. BHALERAO, R. P. BHALERAO, D. BLAUDEZ, W. BOERJAN, A. BRUN, A. BRUNNER, V. BUSOV, M. CAMPBELL, J. CARLSON, M. CHALOT, J. CHAPMAN, G. L. CHEN, D. COOPER, P. M. COUTINHO, J. COUTURIER, S. COVERT, Q. CRONK, R. CUNNINGHAM, J. DAVIS, S. DEGROEVE, A. DEJARDIN, C. DEPAMPHILIS, J. DETTER, B. DIRKS, I. DUBCHAK, S. DUPLESSIS, J. EHLTING, B. ELLIS, K. GENDLER, D. GOODSTEIN, M. GRIBSKOV, J. GRIMWOOD, A. GROOVER, L. GUNTER, B. HAMBERGER, B. HEINZE, Y. HELARIUTTA, B. HENRISSAT, D. HOLLIGAN, R. HOLT, W. HUANG, N. ISLAM-FARIDI, S. JONES, M. JONES-RHOADES, R. JORGENSEN, C. JOSHI, J. KANGASJARVI, J. KARLSSON, C. KELLEHER, R. KIRKPATRICK, M. KIRST, A. KOHLER, U. KALLURI, F. LARIMER, J. LEEBENS-MACK, J. C. LEPLE, P. LOCASCIO, Y. LOU, S. LUCAS, F. MARTIN, B. MONTANINI, C. NAPOLI, D. R. NELSON, C. NELSON, K. NIEMINEN, O. NILSSON, V. PEREDA, G. PETER, R. PHILIPPE, G. PILATE, A. POLIAKOV, J. RAZUMOVSKAYA, P. RICHARDSON, C. RINALDI, K. RITLAND, P. ROUZE, D. RYABOY, J. SCHMUTZ, J. SCHRADER, B. SEGERMAN, H. SHIN, A. SIDDIQUI, F. STERKY, A. TERRY, C. J. TSAI, E. UBERBACHER, P. UNNEBERG, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596-1604.

VAN DER BORRE, AND L. WATSON. 1994. The infrageneric classification of Eragrostis (Poaceae). *Taxon* 43: 383-422

VIJAYKUMAR, A., A. SAINI, AND N. JAWALI. 2010. Phylogenetic Analysis of Subgenus Vigna Species Using Nuclear Ribosomal RNA *ITS*: Evidence of Hybridization among Vigna unguiculata Subspecies. *Journal of Heredity* 101: 177-188.

VOLK, T. A., L. P. ABRAHAMSON, C. A. NOWAK, L. B. SMART, P. J. THARAKAN, AND E. H. WHITE 2006. The development of short-rotation willow in the northeastern United States for bioenergy and bioproducts, agroforestry and phytoremediation. *Biomass & Bioenergy* 30: 715-727.

WANG C., AND C.-F. FANG. 1984. *Salicaceae*. Flora Reipublicae Popularis Sinicae, vol. 20(2): 1-403. Beijing: Science Press, China.

WEIH, M., A. C. RONNBERG-WASTLJUNG, AND C. GLYNN. 2006. Genetic basis of phenotypic correlations among growth traits in hybrid willow (*Salix dasyclados x S. viminalis*) grown under two water regimes. *New Phytologist* 170: 467-477.

WEN, J., AND E. A. ZIMMER. 1996. Phylogeny and biogeography of Panax (the gingseng genus, *Araliaceae*): inferences from *ITS* sequences of nuclear ribosomal DNA. *Molecular Phylogenetic Evolution* 6: 167-177.

YAO, H., J. SONG, C. LIU, K. LUO, J. HAN, Y. LI, X. PANG, H. XU, Y. ZHU, P. XIAO, AND S. CHEN. 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One* 5.

Annexes

Annexe 1. Extraction ADN par la méthode CTAB

Avant propos :
-Ce protocole utilise une plaque de 96 microtubes de Qiagen (numéro de catalogue 19560). Les étapes sont transférables aux autres formats (ex. tube individuel de 1.5 mL).
-Vérifier la présence des solutions et du matériel. Allumer le bain marie au préalable.

Étape préalable:
Lyophilisation des échantillons (non nécessaire si les échantillons sont secs (ex. gel de silice)
1) Prélever le matériel végétal frais, de préférence la partie foliaire.
2) Disposer dans les puits (quantité à ajuster).
3) Recouvrir la plaque avec du papier aluminium.
4) Percer des trous dans le papier d'aluminium (un trou par puits).
5) Lancer le lyophilisateur pendant 12 heures.

Jour 1, Étape 1 :
Broyage du matériel végétal
6) Disposer une bille de tungstène de 3 mm par puit contenant l'échantillon (Qiagen, numéro de catalogue 69997).
7) Mettre les capuchons (Qiagen, numéro de catalogue 19566).
8) Positionner les adaptateurs du TissueLyser II et placer l'ensemble dans la machine. Remarque : a) Sélectionner les adaptateurs au bon format. b) Équilibrer la machine.
9) Lancer le broyage durant 30 s à une fréquence de 30 Hz. Remarque : un bon broyage laisse apparaitre un halo vert sur le capuchon, si nécessaire cette étape peut-être réalisée à nouveau.

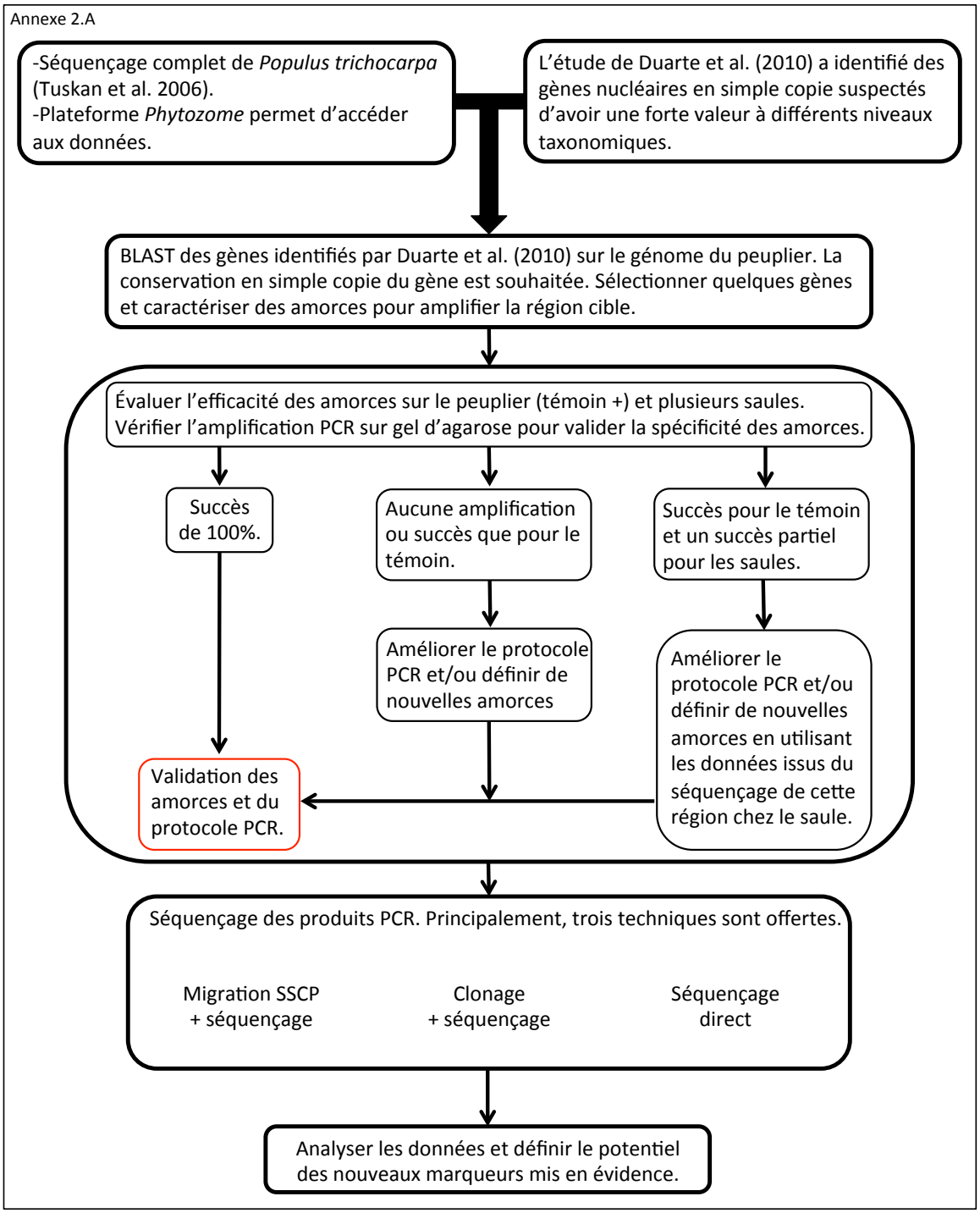
Jour 1, Étape 2 :
Extraction (les volumes indiqués correspondent à 1 échantillon)
10) Préparer le mélange suivant: 400 μ L de CTAB (2X) + 0.004 g de polyvinylpyrrolidone (PVP).
11) Placer le mélange à 65°C pendant 10min pour dissoudre le PVP.
12) Ajouter 0,8 μ L de beta-mercaptoethanol dans le mélange.
13) Transférer 400 μ L du mélange dans chaque puits contenant le broyat.
14) Placer les tubes (contenant le broyat et le mélange) à 65°C pendant 1h30. Homogénéiser régulièrement (15-20 min). Remarque : les capuchons peuvent sauter; il faut mettre une plaque au dessus pour soutenir les capuchons.
15) Ajouter 400 μ L de chloroform:isoamylalcool (24:1) dans chaque puits.
16) Centrifuger 30 min à 4°C à 5788 g.

17) Transférer la phase supérieure (environ 300 μ L) dans un nouveau tube de même format.
18) Ajouter 200 μ L d' isopropanol 100%.
19) Laisser précipiter l'ADN toute la nuit à -20°C.

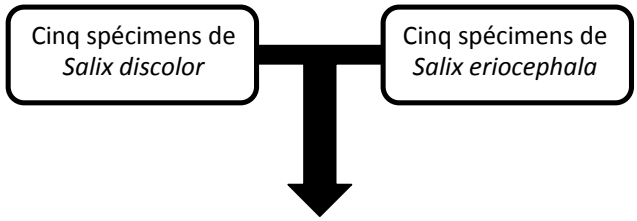
Jour 2, Étape 2 (suite):
Extraction (les volumes indiqués correspondent à 1 échantillon)

20) Centrifuger 30 min à 4°C à 5788 g.
21) Vider le surnageant (culot visible).
22) Ajouter 500 μ L d'éthanol 70%.
23) Centrifuger 10 min à 4°C à 5788 g.
24) Vider le surnageant.
25) Sécher le culot d'ADN sous la hotte ou avec un speed-vac.
26) Suspendre le culot dans 100 μ L du tampon TE (ou eau stérile) (volume modifiable).
27) Conserver les échantillons à -20°C.

Annexe 2. Méthodologie pour caractériser de nouveaux marqueurs nucléaires en simple copie. A) Description théorique des étapes à réaliser en partant du génome du peuplier et des données de Duarte et al. (2010). B) Comparaison du polymorphisme entre *Salix discolor* et *S. eriocephala* sur cinq nouveaux gènes nucléaires et sur *ITS*, *matK* et *rbcL*.



Annexe 2.B



matK	Aucune variabilité en les spécimens de *Salix discolor* et *S. eriocephala*
rbcL	Aucune variabilité en les spécimens de *Salix discolor* et *S. eriocephala*
ITS	Aucune variabilité intra-espèce 1 SNP différenciant *Salix discolor* de *S. eriocephala*
Gènes nucléaires en simple copie	Faible variabilité intra-espèce 7 SNPs différenciant *Salix discolor* de *S. eriocephala* sur le gène 2 et 10 SNPs sur le gène 4 (gènes 1, 3 et 5 ont été amplifiés mais non séquençés).

Détail des cinq gènes nucléaires. Le numéro POPTR fait référence au génome du peuplier

POPTR_0001s16120	Gène 1	Forward	GGAGTTTGTTGCGGTTGATT
		Reverse	ATTGATGCCAGATTCCCTTG
POPTR_0011s02930	Gène 2	Forward	GTATTGGCTCTGGGGGATCT
		Reverse	CCATGGCATAGTCCTAACTTGA
POPTR_0014s15330	Gène 3	Forward	TCCTGAGGTTGATGCTGTTG
		Reverse	GGGTCATCCATGTTGTGAATC
POPTR_0016s07110	Gène 4	Forward	GGAAAGATGATCGCCAAGAA
		Reverse	GCAGGAATAGCTCCATCGTG
POPTR_0018s03160	Gène 5	Forward	TGACAATGGAGGATCTGTCG
		Reverse	CGATGGGCTATCAGCAAAAT