

## **Population genetic structure and unravelling hybridisation of riparian softwood *Salix* species in parts of Europe: A case study on *Salix alba* and the *S. alba* – *S. euxina* complex.**

**Karolien Van Puyvelde**

### **SUMMARY**

*Salix* species acts as pioneers and are the main vegetation type in alluvial and European softwood riparian ecosystems. An important issue in the dynamics of an ecosystem is the understanding of changes in the genetic structure of populations. Many factors such as mutation, gene flow, random genetic drift, hybridisation, introgression and selection interact through time to create the genetic structure of a species flock seen today.

*Salix alba* and *S. euxina* are closely related tetraploid species with a sympatric distribution. They possess many features which have an influence on their population structure: they are dioecious obligate outcrosser; they are capable of vegetative reproduction; they are wind- and insect-pollinated and wind-dispersal. Hybridisation between species with possible gene introgression has been reported which complicates a straightforward overview of their genetic structure.

The first part of the thesis deals with population genetic structure at different scales (from river stretch to different river systems across Europe) on basis of genetic diversity of *S. alba*. The fine-scale spatial genetic structure of different *S. alba* populations was assessed using ISSR markers. These populations were sampled along an altitudinal gradient in two rivers – Adige and Rhine - on a southern and northern slope of the Alps. To estimate genetic structure at population and region level of European *S. alba* we investigated the population structure from Western/Central Europe to Southern Europe using nuclear dominant and chloroplast markers. Standard parameters for the calculation of genetic diversity and several statistical analyses were used to investigate the phylogeographic structure, the geographical patterns and genetic population structure. Although the existence of hybridisation and introgression between *S. alba* and *S. euxina* (as *S. fragilis*) in Europe was proven, the degree of this hybridisation and introgression level remained unclear.

Therefore the second part of the thesis focuses on hybridisation of the *S. alba* – *S. euxina*-complex. The identification of genotypes of *S. alba* and *S. euxina* using codominant microsatellite data is challenging since we are dealing here with allotetraploid species. This additional difficulty of having to cope with up to four allelic positions for partial heterozygotes, obliged us to design a new software program ATETRA to analyse microsatellite data sets obtained from tetraploid species. ATETRA can handle an infinite number of partial heterozygotes and calculates several genetic variables. The dynamics of hybridisation and introgression of the *S. alba* – *S. euxina*-complex in tributaries of the Mosel (Luxemburg) and Rhine river (Germany) were analysed through microsatellite markers using two Bayesian clustering methods (STRUCTURE and NEWHYBRIDS) and chloroplast markers. Here we first identified species since the taxon *S. "fragilis"* was often inadvertently used for both the pure species *S. euxina* (formerly known as *S. fragilis*) and for its hybrid *S. × fragilis* (formerly known as *S. × rubens*) with *S. alba*. We then investigated the type, extent and spatial distribution of hybrids as well as the nature and factors responsible for the observed patterns of hybridisation. By using simulated genotypes the efficiency and accuracy of the two clustering methods were assessed. It also allowed us to choose the threshold value for the posterior probability to discriminate pure from admixed individuals.

Our findings for the first part of this thesis indicate a pattern of genetic structure for both river stretches in the Alps. This corresponds to the model of isolation by distance (IBD) at both intra- and interpopulation levels. ISSR markers proved to be very useful in estimating isolation by distance at both individual and population level for *S. alba* populations. At the European level however, we have found no strong barriers to gene flow between populations of *S. alba*. However there was a fair amount of geographical clustering of populations at any given geographical scale (river, catchment and regions). At a European scale we detected a strong geographic but no phylogeographic structure for *S. alba* with three haplotypes based on chloroplast markers. The results suggest three possible glacial refugia for *S. alba* in the Balkan and Iberian Peninsulas as well as in north-eastern France. At a European scale a weak but clear IBD was found based on ISSR markers. This IBD was more pronounced within Western/Central Europe and Southern Europe when considered separately. Total gene

diversity and genetic differentiation were higher for the Western/Central European regions than for the Southern European regions. Nuclear dominant markers provided more insight in the geographical genetic structure in some regions and river catchments. The stronger genetic differentiation in Western/Central Europe suggests refugia outside the traditional southern refugia.

Results for the second part of the study indicated that both clustering methods to analyse hybridisation and introgression with any given threshold, correctly classified all F1 hybrids consistently as hybrids and never as pure species. There was a moderate accuracy in distinguishing backcrosses hybrids but a low accuracy in distinguishing F2. This study clearly confirms three entities - *S. alba*, *S. euxina* and their hybrids - in Europe. Evidence of extensive hybridisation was found in both tributaries and was dominated by *S. euxina* and F1 hybrids suggesting a hybrid swarm formation. Both nuclear and chloroplast markers confirmed hybridisation with limited introgression through backcrosses with both parental species. Both genetic barriers and genetic-environment associations could explain the absence of F2 ( $F1 \times F1$ ).

Overall, the results presented in this thesis deliver further insight in the larger framework on dispersal ecology of wind dispersed species from European softwood riparian ecosystems.

## **SAMENVATTING**

Wilgen (*Salix* spp.) vormen het belangrijkste vegetatietype in Europese alluviale rivierecosystemen en treden er vaak op als pioniersoorten. Het begrijpen van veranderingen in de genetische structuur van populaties is een belangrijk onderwerp in het veld van de ecosysteemdynamieken. Vele factoren zoals genmutaties, 'gene flow,' willekeurige genetische drift, kruisingen, 'introgressions' and selectie interageren alle doorheen de tijd met elkaar om de actuele groepen van nauw verwante soorten te vormen in een bepaald gebied.

De Schietwilg (*Salix alba*) en de Kraakwilg (*S. euxina*) zijn nauw verwante en tetraploïde soorten die samen in eenzelfde gebied kunnen voorkomen. Ze bezitten bepaalde eigenschappen die duidelijk een invloed hebben op hun populatiegenetische structuur: het zijn tweehuizige planten en daardoor verplichte kruisbestuivers; ze kunnen zich lokaal klonaal vermeerderen; ze gebruiken zowel wind als insecten voor hun bestuiving maar zaden worden enkel via de wind verspreid. Opvallend is dat er bij wilgen zich vaak kruisingen - mogelijk zelfs met 'introgression' van genen - tussen verschillende soorten kunnen voordoen, wat natuurlijk een klaar en duidelijk overzicht van hun genetische structuur bemoeilijkt.

Het eerste deel van dit doctoraat behandelt de genetische structuur van populaties op verschillende schalen (van slechts een stuk rivier tot verschillende rivierbekkens over heel Europa) aan de hand van de genetische diversiteit van de Schietwilg (*S. alba*). De genetische structuur van verschillende Schietwilgpopulaties op een lokale schaal werd met behulp van ISSR markers onderzocht. Deze Schietwilgpopulaties werden langs een altitudinale gradiënt in twee rivieren - Adige en Rijn - op een zuidelijke en noordelijke helling van de Alpen ingezameld. Om de genetische structuur van Schietwilgen op populatie- en regionale schaal in te schatten, onderzochten we de populatiestructuur van Schietwilgen in West-, Centraal- en Zuid-Europa door middel van zowel dominante als chloroplast merkers. Gangbare parameters voor de berekening van de genetische diversiteit en verschillende statistische

methoden werden gebruikt om zowel de fylogeografische structuur, de geografische patronen als de genetische structuur te onderzoeken. Hoewel het bestaan van kruisingen en 'introgression' tussen *S. alba* en *S. euxina* (vroeger als *S. fragilis* bestempeld) werd bewezen in Europa, blijft de grootteorde van deze kruising en 'introgression' onduidelijk.

Het tweede deel van dit doctoraat spitst zich daarom toe op de kruisingen van het *S. alba* - *S. euxina*-complex. De identificatie van de genotypes van *S. alba* en *S. euxina* door middel van codominante microsatellietgegevens is een hele uitdaging aangezien we hier te maken hebben met allotetraploïde soorten. Deze extra moeilijkheid van rekening te moeten houden met maximaal vier allelposities voor partiële heterozygoten, verplichtte ons om een nieuw softwareprogramma ATETRA te ontwikkelen om microsatellietgegevens van tetraploïde soorten te analyseren. ATETRA is in staat een oneindig aantal partiële heterozygoten te verwerken en verschillende genetische variabelen te berekenen. De kruising- en 'introgression'-dynamiek van het *S. alba* – *S. euxina*-complex in zijrivieren van de Moezel (Luxemburg) en de Rijn (Duitsland) werden onderzocht met behulp van zowel microsatelliet en twee Bayesiaanse clustermethodes (STRUCTURE en NEWHYBRIDS) als als chloroplast merkers. Eerst werden de verschillende soorten geïdentificeerd aangezien het taxon *S. "fragilis"* vaak foutief werd gebruikt om zowel de zuivere soort *S. euxina* (vroeger *S. fragilis* genoemd) en haar kruising met hybride *S. alba* (zijnde *S. X fragilis*) aan te duiden. Dan werden naast de bepaling van het type, omvang en de ruimtelijke verspreiding van hybriden ook de aard en factoren verantwoordelijk voor de waargenomen hybridisatie patronen onderzocht. De efficiëntie en nauwkeurigheid van de twee clustermethodes kon beoordeeld worden door gebruik te maken van gesimuleerde genotypes. Dit liet ons ook toe de drempelwaarde voor de 'posterior probability' te bepalen om pure van gemengde individuen te kunnen onderscheiden.

Onze bevindingen voor het eerste deel van deze thesis geven aan dat er een patroon van genetische structuur voor beide alpiene riviertrajecten. Dit komt goed overeen met het model van 'isolation by distance' (IBD) op zowel intra- als interpopulatie-niveau. ISSR merkers voor *S. alba* bleken zeer geschikt te zijn voor het bepalen van IBD op zowel individueel- als populatie-niveau. Op Europees niveau werden echter geen sterke barrières voor 'gene flow'

tussen *S. alba*-populaties gevonden maar er was toch enige geografische clustering van populaties op eender welke geografische schaal we beschouwden (rivier, stroombekken, regio's en continent). Op Europees niveau ontdekten we een sterke geografische structuur maar geen fylogeografische structuur met drie haplotypen voor *S. alba* waargenomen op basis van chloroplastmerkers. Drie mogelijke glaciële refugia voor *S. alba* konden aangewezen worden in het Iberisch en Balkanschiereiland, en in het noordoosten van Frankrijk. Een zwakke maar duidelijke IBD volgens ISSR merkers werd gevonden op Europese schaal. Deze IBD was nog meer uitgesproken binnen West/Centraal-Europa en Zuid-Europa wanneer ze afzonderlijk beschouwd werden. Totale gendiversiteit en genetische differentiatie waren hoger voor de West/Centraal-Europese regio's dan voor de Zuid-Europese regio's. Nucleaire, dominante merkers bieden meer inzicht in de geografische genetische structuur voor sommige regio's en stroomgebieden. De sterkere genetische differentiatie gevonden met behulp van de chloroplast en nucleaire merkers suggereren refugia buiten de traditionele zuidelijke refugia.

Resultaten van het tweede deel van deze thesis geven aan dat beide clustermethodes in de kruising- en 'introgression'-studie alle F1-hybriden altijd correct als hybriden en nooit als zuivere soorten classificeerden, wat ook de ingestelde drempelwaarde mocht zijn. Er was een matige nauwkeurigheid bij het onderscheiden terugkruisingen, maar een lage nauwkeurigheid bij het onderscheiden van F2. Deze studie bevestigt duidelijk dat er drie entiteiten *S. alba*, *S. euxina* en hun hybriden in Europa zijn. Uitgebreide kruisingen werden in beide zijrivieren gevonden. Deze waren gedomineerd door *S. euxina* en F1-hybriden wat wijst op de formatie van een hybridezwerf. Zowel nucleaire- en chloroplastmerkers bevestigden kruisingen met beperkte 'introgression' door terugkruisingen met beide ouderlijke soorten. Zowel genetische barrières als het verband tussen gen en omgeving konden de afwezigheid van F2 ( $F1 \times F1$ ) uitleggen.

De hier aangetoonde resultaten leverden inzicht in het grotere kader voor verspreidings-ecologie van soorten met windverspreiding in Europese rivierbosecosystemen.