

DISSERTATION

DNA marker development for plant genetic analysis

Mag. Silvia Fluch, MSc

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2010

Studienkennzahl lt. Studienblatt: A 091-441

Dissertationsgebiet lt. Studienblatt: Biologie; Genetik und Mikrobiologie

Betreuerin / Betreuer: Ao Prof. Mag. Dr. Mary R. Samuel

Acknowledgement:

first of all I want to thank my parents for making it possible for me to be here and do this work; then all my colleagues and supervisors at ait/arc/fzs kornel burg and Josef Schmidt for believing in me; thanx to all my friends and especially to doris, siegfried and dieter who forced me to finish the work in one or the other way.

WAS WIR HEUTE TUN (UND DENKEN)
ENTSCHEIDET DARÜBER,
WIE DIE WELT MORGEN AUSSIEHT
Marie von Ebner-Eschenbach

CONTENTS:

ABSTRACT	5
ZUSAMMENFASSUNG	7
1 INTRODUCTION	9
1.1 MOLECULAR MARKERS – WHAT IS IT?	9
1.2 PROPERTIES OF THE DNA USED FOR MARKER GENERATION	10
1.3 TYPES OF MOLECULAR MARKERS	13
1.3.1 Non-PCR-based techniques	14
1.3.2 PCR-based techniques	19
1.3.2.1 Anonymous Markers	19
1.3.2.2 Sequence Specific Marker Types:	23
1.4 MOLECULAR MARKERS IN PLANT GENOME ANALYSIS	28
2 SPECIFIC MARKER DEVELOPMENT AND APPLICATIONS FOR SELECTED RESEARCH TOPICS – THESIS WORK	32
2.1 MARKERS IN THE NUCLEAR GENOME	32
2.1.1 SSR Markers Development for <i>Rhododendron</i>	36
2.1.1.1 Microsatellite markers for <i>Rhododendron ferrugineum</i> and <i>Rhododendron hirsutum</i> , sympatric alpine shrubs	37
2.1.1.2 Genetic structure of relict populations of <i>Rhododendron ferrugineum</i> L. in the Northern Apennines and their relationships with the nearest populations from the Alps	42
2.1.2 SSR Markers Development for Norway spruce	57
2.1.2.1 Characterization of variable EST SSR markers for Norway spruce [<i>Picea abies</i> L.) Karst.]	59
2.1.2.2 Genetic diversity in managed subpopulations of Norway spruce [<i>Picea</i> <i>abies</i> (L.) Karst.]	66
2.1.3 Nuclear markers for oak (<i>Quercus</i> ssp)	78
2.1.3.1 Detection of microsatellite instability during somatic embryogenesis of oak (<i>Quercus robur</i> L.)	85
2.1.4 SSR isolation for <i>Paeonia</i>	96
2.1.4.1 Microsatellite markers in tree peony, <i>Paeonia suffruticosa</i> (<i>Paeoniaceae</i>)	97
2.2 THE ORGANELLAR GENOMES	105
2.3 THE CHLOROPLAST DNA – SPECIFIC WORK	107
2.3.1 Introduction	107
2.3.2 Cp DNA marker development and application	109
2.3.2.1 A Chloroplast marker system for studying genetic variation in <i>Populus</i> <i>nigra</i> L.	110

2.3.2.2	Postglacial migration of <i>Populus nigra</i> L.: lessons learnt from chloroplast DNA	117
2.4	THE MITOCHONDRIAL DNA – SPECIFIC WORK	142
2.4.1	Introduction	142
2.4.2	Mitochondrial DNA marker development and application in Norway spruce	144
2.4.2.1	Development of mitochondrial markers for population genetics of Norway spruce [<i>Picea abies</i> (L.) Karst.]	146
2.4.2.2	Population differentiation of Austrian Norway spruce [<i>Picea abies</i> (L.) Karst.] populations assayed by mitochondrial markers	152
3	DISCUSSION & CONCLUSION	165
4	APPENDIX	177
5	APPENDIX II	181
6	REFERENCES	195
7	CURRICULUM VITAE	215

Abstract

Genetic markers use short DNA sequences in order to investigate and label a particular locus in the genome of a given individual. Whenever this particular locus shows sequence differences between two individuals, this region can be used as molecular marker. The development of a variety of DNA-based markers in the past twenty years has revolutionized biological, forestry and agricultural science. Molecular markers have demonstrated their usefulness for plant genotyping and population genetic analysis. Genome characterization by DNA-fingerprinting has generated valuable information on population structure and molecular genetic diversity. Depending on the compartmental origin of the genomic information (organelle, nucleus), molecular markers can be used for answering different questions. While organellar markers could serve for postglacial migration studies, nuclear markers can be used for studying biodiversity and paternity analysis on the level of the single individual.

With the work presented here, we contributed to the marker development in different areas of plant genomics, and proved the applicability of the different marker types targeting the three different genomes (nuclear, mitochondrial, chloroplast) available in plants.

In order to investigate the genetic variation present in Alpine *Rhododendron* populations, nuclear SSRs were developed for *Rhododendron ferrugineum* and *Rhododendron hirsutum* by generating genomic libraries enriched for dinucleotide repeats. These newly generated markers subsequently were used to investigate relict populations of the alpine rose in Italy and Austria. The same method was applied to isolate SSR regions for the identification of *Paeonia suffruticosa* varieties available in a germ plasm collection in Schönbrunn. In the latter case, also an in silico approach was used in order to identify SSRs present in publicly available ESTs (sequence information on Expressed Sequence Tags). By this approach, 5 additional SSR markers for *Paeonia suffruticosa* as well as 24 EST-SSRs for *Picea abies* could be generated in relatively short time. These EST-SSRs have added value as they are part of expressed sequences and thus can be used for measuring the functional diversity in Norway spruce as well as *Paeonia*.

For *Quercus robur* L. and *Quercus petraea* L. SSRs as well as VNTRs were isolated by a standard wet lab approaches and used for diversity measurements in Austrian populations of oak. In addition, these markers, together with RAPDs, RFLPs and SCARs were mapped on a genetic map of a French controlled oak cross. Due to their hypervariable nature and sensitivity to mutations, nuclear microsatellites for oak were used to test for somaclonal variation occurring in tissue culture material over time. With this approach we could prove the usefulness of the SSRs for testing the genome stability under stress.

Maternally inherited markers were developed for *Populus nigra* L. and *Picea abies*. In order to assess the genetic diversity in *Populus nigra* L. sampled throughout Europe, a chloroplast DNA based marker system was developed and applied. For Norway

spruce on the other hand, the maternally inherited mitochondrial genome was targeted for marker development. Using these new mt markers for Norway spruce it could be shown that there is moderate genetic variation in Austrian Norway spruce strands, not detected before.

The newly developed nuclear as well as organellar markers are valuable tools for the assessment of genetic diversity in natural populations as well as plant selection programs. Using this DNA based molecular information it is possible to analyze the genetic structure of populations, to estimate gene flow between populations and, as in the case of poplar, to identify naturally occurring clonal material. With these new markers, specific population genetic as well as phylogeographical questions can be answered, and, as in the case of rhododendron, decisions on conservational measures taken.

Zusammenfassung

Als Molekulare Marker werden kurze DNA Abschnitte herangezogen, die man für die Untersuchung und Markierung einer bestimmten Stelle im Genom eines Individuums verwendet. Immer wenn in dieser Region zwischen zwei Individuen Unterschiede auftreten, dann kann die Region als molekularer Marker verwendet werden. In den letzten zwanzig Jahren hat die Entwicklung solcher genetischen Marker die biologische Forschung in Botanik, Forst- und Landwirtschaft revolutioniert. Molekulare Marker fanden Anwendung in Populationsgenetik und Genotypisierung, in Züchtung und Taxonomie. Die Untersuchung von Individuen mittels genetischem Fingerprint ergab wertvolle Einblicke in Populationsstrukturen und die bestehende genetische Vielfalt. Je nachdem, welches Genom man untersucht (Organellen, Nucleus), haben die angewandten molekularen Marker unterschiedliche Aussagekraft und es können verschiedene Fragestellungen bearbeitet werden. Während mit Organellenmarkern beispielsweise postglaziale Wanderungsrouten von Spezies rekonstruiert werden können, dienen Kern-Marker unter anderem zur Untersuchung von Verwandtschaftsverhältnissen bei Einzelindividuen.

Die gegenständliche Arbeit trägt mit ihren Entwicklungen zur Identifizierung neuer Marker aus den drei verschiedenen in Pflanzen vertretenen Genomen bei, wodurch verschiedene Pflanzengenetische Fragestellungen behandelt werden können.

Um die genetische Vielfalt im, in den Alpen vorkommenden, Almrausch (*Rhododendron ferrugineum* und *Rhododendron hirsutum*) untersuchen zu können, wurden Microsatellitenregionen mittels aus dem Kerngenom mittels, für Dinukleotide angereicherte Genbanken isoliert. Die so neu identifizierten Kern-Marker wurden in weiterer Folge für populationsgenetische Untersuchungen in Österreichischen und Italienischen Reliktbeständen der Alpenrose verwendet. Die gleiche Anreicherungsmethode wurde angewendet, um SSR Regionen für die Unterscheidung von *Paeonia suffruticosa* Sorten einer Sammlung in Schönbrunn zu entwickeln. Da die Anreicherungsmethode nicht die gewünschten Ergebnisse brachte, wurde für *Paeonia* die Datenbankanalyse von ESTs zur Anwendung gebracht, bei der bereits publizierte Gensequenzen für die Identifikation von SSRs in exprimierten Genen herangezogen werden. Dadurch konnten weitere SSRs identifiziert werden, mit denen eine Unterscheidung der *Paeonia* Individuen möglich war. Mit der gleichen Methode wurden auch 24 EST-SSRs für *Picea abies* charakterisiert, die durch den Umstand, dass sie in exprimierten Genen liegen, auch zusätzliche Information über funktionelle Variation in den untersuchten Probensets geben können.

Für *Quercus robur* L. und *Quercus petraea* L. wurden mittels standard Labormethoden sowohl SSRs als auch VNTRs isoliert welche für die Untersuchung der genetischen Vielfalt Österreichischer Eichen als auch, gemeinsam mit anderen molekularen Markern, für die Herstellung einer genetischen Karte einer Französischen Eichenkreuzung verwendet wurden. Microsatelliten sind hypervariable DNA Regionen im Genom, die leicht mutieren. Durch die Untersuchung von Pflanzenmaterial, das

mittels Gewebekultur vermehrt wurde, konnte gezeigt werden, dass sich Microsatelliten sehr gut für die Detektion von somaklonaler Variation, hervorgerufen durch Stress, eignen.

Maternal vererbte Markersysteme wurden sowohl für Pappel (*Populus nigra* L.) als auch für Fichte (*Picea abies*) entwickelt. Um die bestehende genetische Vielfalt in Europäischen Schwarzpappelherkünften zu untersuchen, wurde ein Chloroplastenbasiertes Markersystem entwickelt, das dazu diente, das postglaziale Wanderungsverhalten der Pappel zu untersuchen. Bei der Fichte untersucht das neu entwickelte maternale Markersystem Unterschiede in der mitochondrialen DNA der zu untersuchenden Proben. Mit dem neuen Markersystem konnte für Fichte moderate genetische Variation in 37 österreichischen Herkünften identifiziert werden.

Die neu entwickelten nukleären sowie Organellen-Marker stellen wertvolle Werkzeuge zur Untersuchung der genetischen Vielfalt in natürlichen Populationen sowie für Zucht- und Auswahlprogramme dar. Durch die Verwendung von DNA basierten molekularen Daten kann die genetische Struktur von Populationen und auch der Genfluss zwischen Populationen gemessen werden. Die Marker dienen auch, so wie im Fall der Pappel, dazu, natürlich auftretendes klonales Material zu identifizieren. Mit den neuen Markern ist es möglich, spezifische populationsgenetische als auch phylogeographische Fragestellungen zu beantworten, und, so wie bei Rhododendron, Entscheidungen über Konservierungsmaßnahmen getroffen werden.

1 Introduction

1.1 Molecular markers – what is it?

Differences at the DNA level between individuals are, to a large extent, the basis and cause for the phenotypic differences between them. In order to understand and use phenotypic alterations i.e. diseases, resistance, quantitative traits, researchers want to understand the underlying genetic cause and the alteration on the genomic level. The differences in the genome are known as genetic variations and are caused by various types of alterations at various levels, such as nucleotide substitutions, insertion/deletion of nucleotides, differences in the copy number of repetitive elements or even of genomic segments up to parts of whole chromosomes and the combinations of these changes. A subset of these genetic variations is accepted as genetic polymorphisms when they are observed at a frequency of 0,1% in a certain population (Nakamura 2009).

In order to detect genetic variations between two genomes, various techniques have been established throughout the years starting in 1980 with Botstein (Botstein *et al.*, 1980) using RFLPs for genetic mapping in humans. In general, when variations on DNA level are detected and described, they are termed '*molecular markers*'. Therefore a molecular marker is defined as a rather short DNA fragment of up to several hundred or thousand base pairs. This specific region of the genome, called locus, shows variation between two or more investigated individuals and provides landmarks in the genome. Variation could be detected as presence or absence of a particular fragment or sequence, as length variation between two fragments or as sequence difference as small as one base pair. The detected variation arises from an alteration in the investigated genomic locus. It could be associated with a known location on a chromosome and linked to a particular gene or trait, or that marker region is still of unknown function and location in the genome. Detection of the variants of a locus of interest can be performed using various molecular techniques depending on the resolution power needed to detect the variation.

There are two major classes of molecular markers available, depending on the a priori knowledge available for the application of the respective marker system. In a situation where genomic sequences describing a marker are available, we speak about a sequence specific marker, in contrast to anonymous markers. Anonymous markers are applied in situations where no knowledge on the genomic composition is available at the time of use. They rely on fact that a genome can be treated in bulk by either restriction/ligation/selection techniques or PCR based filtration/amplification techniques to generate individual-specific fingerprints. Differences in these fingerprints are then used for downstream applications.

Due to their structural state, sequence specific molecular markers to detect genetic variations can be classified into major categories: RFLP (restriction fragment length polymorphism), VNTR (variable number of tandem repeat) and STR (short tandem repeat or microsatellite), SNP (single-nucleotide polymorphism) as well as

CNV (copy-number variation). All these markers are available in the nuclear as well as the organellar genome of an individual. Therefore these markers can help answering different questions related to population genetics or trait related questions.

1.2 Properties of the DNA used for marker generation

Genomic DNA in general can be subdivided in to categories – repetitive DNA and single copy regions, and both of them exist in the nuclear as well as the organellar genomes. For a long time, the repetitive part of the genome was the major target for the development and application of molecular markers. Another possibility to generate molecular marker information based on intrinsic properties of the DNA is the use of expressed regions of the genome. This allows filtering of the vast genomic information to just focus on the functional part represented in the mRNA (i.e. expressed parts). This approach is limited by the fact that gene expression is tissue specific as well as it is influenced by environmental factors. In recent years, yet another feature of the DNA has been exploited for marker development – the epigenetic information being added to the basic DNA sequence information. Such variation involves information encoded in biochemical marks on the DNA or chromatin like differential histone modification states, histone distribution, and DNA methylation. Much of the work on stable transmissible epigenetic variation has exploited experimental models, such as mice, maize, and Arabidopsis, and most of these studies have examined either phenotypically silent epigenetic variation or epigenetic variation causing dramatic developmental phenotypes (Richards 2009). However, also in the disciplines of evolutionary biology, ecology, and breeding epigenetics mutations are gaining interest as primarily complex polygenic traits and phenotypic variation in natural contexts are affected by epigenetic changes (Bossdorf *et al.*, 2008; Richards 2008).

Repetitive DNA.

The discovery that up to 90% of the genome of virtually all the species is constituted by regions of repetitive DNA, which are highly polymorphic in nature, was a big step forward in genetic identification (Charlesworth *et al.*, 1994). Repetitive DNA regions play an important role in absorbing mutations in the genome. Thus repetitive DNA and mutational forces together form the basis of a number of marker systems that are useful for various applications in plant genome analysis. Markers belonging to this class can either be hybridization-based or PCR-based. Repetitive DNA as compared to single copy number DNA (scnDNA) regions show a much higher mutation rate. VNTRs for example mutate 100–1000 times than of scnDNA (Nishant *et al.*, 2009).

These repetitive DNA regions can be used as loci for genetic marker development comprising several hundred alleles, differing from each other with respect to length, sequence or both. They are interspersed in tandem arrays ubiquitously throughout the genome. Repetitive DNA can be found in clusters such as telomers and centromers as well as interspersed between single copy regions throughout the

genomes, in the heterochromatin. The heterochromatin plays an important role in the establishment and maintenance of the centromeric, telomeric and subtelomeric regions, which are essential for proper chromosome segregation. If the heterochromatin contains functional genes, their gene expression state negatively influenced (Dimitri *et al.*, 2005).

It is generally accepted that repeated sequences follow an evolutionary pattern known as 'concerted evolution' (Moyzis *et al.*, 1988; Palomeque & Lorite 2008). The spreading of the new variants of repeats leads to variant homogenization and takes place by means of a variety of mechanisms, such as nonreciprocal DNA transfer within and between chromosomes (i.e. gene conversion, unequal crossing over, slippage replication, transposition, RNA-mediated exchange). The consequence of the 'concerted evolution' is that sequences are homogenized within a repeat family followed by their subsequent fixation in the sexual population.

The physical ends of chromosomes are protected and stabilised by telomeres. Telomeres are nucleoprotein structures, located at the ends of chromosomes and are subject to shortening at each cycle of cell division. They prevent chromosomal ends from being recognized as double strand breaks and protect them from end to end fusion and degradation. Telomeres consist of stretches of repetitive DNA with a high GC content (Moyzis *et al.*, 1988; Pich *et al.*, 1996). The sequence of telomeric DNA normally consists of a simple repeat unit that is conserved in many organisms. Most plants examined up to now seem to possess Arabidopsis-type telomeres consisting of many repeat copies of the sequence 5'-TTTAGGG-3' (Adams *et al.*, 2000).

Repetitive DNA is also ubiquitous and abundant in centromeric regions of higher eukaryotes. In all flowering plants investigated so far, different centromeres are generally composed of the same types of DNA component, mainly large arrays of centromeric satellite repeats and centromeric retrotransposons (Ma *et al.*, 2007). In general, these satellite repeats are specific to centromeric regions, with few copies and only short arrays found elsewhere in the genome.

Yet another repeat category is the Satellite DNA (satDNA) which is the major DNA component of heterochromatin. It is located mainly in the pericentromeric and/or telomeric regions of chromosomes (Charlesworth *et al.*, 1994). It is generally formed by long tandem arrays in which the repeat units are arranged in direct repeats. This molecular organization gives rise to a characteristic multiband-ladder-pattern after agarose gel electrophoresis of genomic DNA digested with the appropriate restriction endonucleases. These different patterns are caused by copy number differences of satellite repeats which could change over time by unequal crossing-over events, although other processes, such as rolling-circle replication or conversion-like mechanisms could also be involved [reviewed by (Charlesworth *et al.*, 1994)].

While many repetitive sequence families have no clear known function and are often species-specific, a few have well-defined functions and are found universally. The ribosomal rRNA genes for 5S, 5.8S, 18S and 25S rRNA are present in long arrays of tandemly repeated gene units including their intergenic transcribed spacers (ITS). Because of their high repetition, they provide useful markers to identify

chromosomes by in situ hybridization (Leitch & Heslop-Harrison 1993) and can be used to combine genetic and chromosomal maps (Schondelmaier *et al.*, 1997). The number and locations of sites may vary between species and therefore can be used to investigate chromosomal homologies and genome evolution (Boonruangrod *et al.*, 2009). Another repeat class with assigned function are the transposable elements. They are ubiquitous in the plant kingdom and share many common features with mobile elements from other eukaryotes - on the structural as well as mechanistic level. Transposition of these elements can influence plant genes and genomes in many ways. One of the earliest observations in this respect was made by McClintock (McClintock 1950), when she discovered transposable elements and found that, in addition to their movement, these elements had the ability to cause cytologically detectable chromosome rearrangement.

It was also shown that transposable element derived sequences can contribute up to 80% of plant genomes thus being very significant factors in plant evolution. A large number of transposable repeat elements have been studied in plants; however, only a restricted number has been exploited as molecular markers (Kumar & Bennetzen 1999; Ruas *et al.*, 2008). In evolutionary terms, they have contributed to genetic differences between species and individuals by promoting unequal crossing over. By their capability of moving around in genome, transposable elements mediate different types of genome alterations, inducing mutations, chromosomal rearrangements and gene expression regulations (Agarwal *et al.*, 2008c). Retrotransposon-mediated fingerprinting has been shown to be an efficient fingerprinting method for detection of genetic differences between different species. The two most important techniques to be mentioned here are SSAP (Sequence Specific Amplification Polymorphism) (Berenyi *et al.*, 2002; Forneck *et al.*, 2009; Waugh *et al.*, 1997a) and IRAP (Inter Repeat Amplification Polymorphism (Heslop-Harrison & Schwarzacher 2007) both targeting retrotransposon insertion sites in the genome.

Within the transposable elements, different categories are being distinguished, retrotransposons and DNA transposable elements. Retrotransposons comprise class-I and class-II transposons, the former mobilized via an RNA intermediate and the classII via a DNA cut-and-paste mechanism. In the retrotransposon category, SINEs, LINEs and LTR retrotransposon families are being distinguished. The term LTR means long terminal repeat and reflects the structural fact that these elements have identical boarder sequences of inverted orientation which are generated when the element is integrated into the genome (Kumar & Bennetzen 1999). These transposons that move via DNA intermediates and thus are called Type II elements. The TY1 copia (Heslop-Harrison *et al.*, 1997) as well as the Gypsy elements form the largest group with the retrotransopsons. They can be used as basis for marker development (see SSAP) as well as for chromosome painting to identify phylogenetic differences between closely related species (Ruas *et al.*, 2008).

A prominent representative of the SINE repeat class is the Alu-repeat family. Alu repeats are a class of randomly repeated interspersed DNA elements, revealing

considerable levels of polymorphism when used as hybridization probes on Alu I digested human genomic DNA. They are representatives of short interspersed nuclear elements known as SINES. Alu elements are approximately 300 bp in size have been studied largely in humans for the enrichment of genetic maps (Roy *et al.*, 1999), as fingerprinting probes (Stallings *et al.*, 1990) and to detect chromosomal rearrangements (Sirivatanauksorn *et al.*, 2002) while their presence in plants remains largely unexplored.

1.3 Types of Molecular Markers

Nucleic acid based molecular markers offer numerous advantages over conventional phenotype based measurements which have been established by Gregor Mendel in the nineteenth century, and used since then for gemplasm evaluation and breeding. DNA based markers are stable and detectable in all tissues regardless of growth, differentiation, development, or defence status of the plant cell and are not influenced by the environment. An ideal molecular marker should exhibit a variety of criteria: (i) be polymorphic and evenly distributed throughout the genome, (ii) generate multiple, independent, easy scoreable and reliable signals (iii) its application should be simple, quick and inexpensive (iv) need small amounts of tissue and DNA samples; (v) suitable for high throughput applications (vi) be linked to distinct phenotypes or traits. Out of the variety of different molecular markers available to date, there is no perfect marker as such - techniques differ from each other with respect to features such as genomic abundance, level of polymorphism detected, locus specificity, reproducibility, technical requirements and cost.

In general, two classes of marker types can be distinguished: (1) Non PCR based techniques and (2) PCR based techniques. Another way to categorize markers is their mode of inheritance, which is depending on the genome they are located in – in case of markers located on the organellar genome (i.e. chloroplast (cp) or mitochondria (mt)) they show a uniparental way of inheritance. In contrast, markers detected in the nuclear genome show Mendelian inheritance, meaning that 50% of the loci are contributed by each parent respectively. The third way of classification is the potential of molecular markers to discriminate locus specific alleles independently thus allowing to identify the parental contribution. Marker types, which allow this allelic discrimination, are called ‘codominant’ markers. Markers, where the presence or absence of a signal or detected fragment can not directly be correlated with the presence or absence of a specific allelic variant, and their origin is due to ‘unknown’ changes in the investigated genome, are termed ‘dominant’ markers.

Table 1 shows an overview on the marker types available. Depending on the scientific question as well as the availability of DNA sequences for the respective species, different markers can be applied.

Abbreviations	Marker name	type	technique
AFLP	Amplified fragment length polymorphism	d	PCR based
CAPS	Cleaved amplified polymorphic sequence	co	PCR+RFLP
cDNA display	AFLP like technique using mRNA/cDNA as template	d	PCR based
CNV	Copy number variation	co	Hybridization
DAF	DNA amplification fingerprinting	d	PCR based
DART	Diversity array technology	co	Hybridization
IMP	Inter-MITE polymorphism (MITES Miniature inverted repeat transposable elements)	d	PCR based
IRAP	Inter-retrotransposon amplified polymorphism	d	PCR based
ISSR	Inter SSR amplification	d	PCR based
ITS	Internal transcribed spacer of rDNA	co	PCR based
HRM	High resolution melting	co	PCR based + Hybridization
RAMP	Randomly amplified microsatellite polymorphisms	d	PCR based
RAPD	Random amplified polymorphic DNA	d	PCR based
REMAP	Retransposon-microsatellite amplified polymorphism	d	PCR based
RFLP	Restriction fragment length polymorphism	co	Hybridization, PCR-RFLP
SCAR	Sequence characterized amplified region	co	PCR based
SNP	Single nucleotide polymorphism	co	PCR based (RT-PCR/resequencing, SNP array, Hybridization)
eSNP	SNPs detected in silico using resequencing data generated by next generation sequencing technologies (454, Solexa)	co	Sequencing
SRAP	Sequence-related amplified polymorphism	d	PCR based
S-SAP	Sequence-specific amplification polymorphism	d	PCR based
SSCP	Single strand conformation polymorphism	co	PCR based
SSR	Simple sequence repeats	co	PCR based
TD	Transposable display	d	PCR based
TRAP	Target region amplification polymorphism	d	PCR based

Table 1 available marker types and their abbreviations, as well as type [d: dominant; co: codominant] and technical information for generating the marker profile.

1.3.1 Non-PCR-based techniques

There are various techniques available to detect the marker types described above. When using hybridization assays to detect variations in the genome, the traditional way was membrane based Southern (E.Southern 1975) hybridization. To analyze whole genomes, total genomic DNA is extracted in large quantities, subject to restriction endonuclease digest, and subsequent gel blotting to transfer the size separated fragments onto a nylon membrane. The membrane bound DNA can then be probed with a radioactively labelled DNA fragment, which will detect all available

genomic copies representing the same sequence as the probe. When using RNA instead of DNA, the same procedure is termed Northern blot (Alwine *et al.*, 1977), which is being used to detect transcribed copies of genes using a target molecule representing that specific gene.

Restriction Fragment Length Polymorphism (RFLP)

RFLP is the oldest, hybridization based, molecular marker type established in the pre-PCR era. In RFLP analysis, DNA polymorphism is detected by hybridizing a chemically/radioactively labelled DNA probe specific for a certain locus to total genomic DNA (gDNA) which was digested by restriction endonucleases and subsequently separated on a gel to produce a Southern blot. This fragmented, size separated DNA is hybridized with the labelled DNA fragment resulting in differential DNA fragment profile. The restriction fragment length polymorphism (RFLP) derives from three types of mutations: (i) the presence or absence of a specific restriction enzyme recognition site in the flanking regions of the targeted locus; (ii) DNA rearrangements like insertions/deletions in the flanking sequence of the respective DNA fragment and the special case (iii) where the target region contains a repetitive element, which exhibits variability repeat number at this locus. A special situation are minisatellites or VNTRs (variable number of tandem repeats) which are detected by hybridization of multi-locus probes on Southern blots of restriction-enzyme-digested genomic DNA. A common 'core' repeat sequence shared between different minisatellite loci allows the detection of many different minisatellites loci simultaneously, producing a multiband, barcode-like, pattern known as 'DNA fingerprints' (Tamaki & Jeffreys 2005).

The RFLP markers show a relatively high degree of polymorphism, depending on the locus specific probe used in the RFLP assay. They are co-dominantly inherited and highly reproducible. Their presence throughout the plant genome, the availability of RFLP type markers in the nuclear as well as the organellar genome made this marker type reliable and widely used marker system in plant genomics for breeding (Tanksley *et al.*, 1992) as well as population genetic analysis (Barreneche *et al.*, 1998b; Garcia-Mas *et al.*, 2000). The method allows to simultaneously screen several tens of samples to detect related genomic regions in parallel. A special application of the RFLP technique is the genomic fingerprinting (Jeffreys *et al.*, 1985) where repetitive elements which are widely spread throughout the genome, and are available in several hundred copies, are used as hybridization probes.

Due to the fact, that the technique needs relatively high amounts of very pure DNA, it is not widely used anymore. In addition, it is time consuming, expensive and involves radioactive/toxic reagents which nowadays is hindering its application due to the requirement of special lab facilities and equipment. The requirement of prior sequence information for probe generation increases the complexity of the methodology. The sequential nature of gel-based marker systems inherently has low throughput and high costs of analysis. These limitations led to the conceptualization

of a new set of less technically complex methods, the PCR-based techniques (see 1.3.2) as well as highly parallel hybridization assays (i.e. microarrays).

Microarray applications

In recent years, Southern and Northern blot analysis has largely been competed out by the technology of *microarrays or DNA chips*. Microarrays are miniaturized hybridization setups on activated surfaces, only a few square centimetres in size, allowing highly parallel hybridization experiments. The surface of a typical microarray can hold up to 600.000 individual DNA fragments of various length (20bp – <2000bp) representing either genes or genomic regions. Microarrays holding gene information can be used for expression profiling which help to detect genes differentially expressed in healthy versus affected tissue (Barakat *et al.*, 2009; Rinaldi *et al.*, 2007), between tissues of different developmental stages (Fluch *et al.*, 2008) or between two individuals showing different susceptibility to stress (Darrigues *et al.*, 2008; Deluc *et al.*, 2009; Pestsova *et al.*, 2008; Qin *et al.*, 2008; Zombori *et al.*, 2008). In the area of marker related microarray applications, two recently developed methodologies should be mentioned: i) DArT (Diversity Array Technology) and ii) whole genome resequencing arrays.

In the plant kingdom where the availability of whole genome sequences is scarce, the DArT[®] Technology (<http://www.diversityarrays.com>) is a good alternative. Especially with large genomes like *Norway spruce* with $1,8 \times 10^{10}$ bp or pine with $2,7 \times 10^{10}$ bp as compared to *A. thaliana* with only $1,6 \times 10^8$ bp or in the case of high ploidy levels (hexaploid wheat: $1,6 \times 10^{10}$ bp) the genome filtration which is included in the DArT[®] technology reduces the genome complexity to a level where marker analysis is getting feasible. DArT[®] filters a DNA sample to obtain just a 'representation' of that sample. The method of complexity-reduction relies on a combination of restriction enzyme digestion and adapter ligation, followed by PCR amplification (Wenzl *et al.*, 2004). DArT operates on the principle that the genomic 'representation' contains two types of fragments: i) Constant fragments, found in any 'representation' prepared from a DNA sample from an individual belonging to a given sample set, and ii) polymorphic fragments to be used as molecular markers, only found in some but not all of the individual 'representations' in that sample set. The variable fragments are informative because they reflect sequence variation present in the different individuals analyzed. These variable fragments are called the DArT[®] markers. Their presence vs. absence in a genomic 'representation' is assayed by hybridizing the 'representation' to a DArT[®] array consisting of a library of the species to be investigated. The polymorphisms scored are presence versus absence of hybridisation to individual array elements. They reflect DNA sequence variation that determines which genomic sequences are present in the genomic representations. In principle, the markers could be located anywhere in the genome, but preferably should provide broad coverage. DArT[®] maps for barley (Wenzl *et al.*, 2004) and wheat (Akbari *et al.*, 2006) confirm that this is the case.

Whenever the whole genome sequence of a species is available, microarrays can be designed and produced that represent the whole genome in very small subsections, called oligonucleotides or oligos (25 mers to 60 mers) that represent the genome in a specific way. Methods for whole-genome polymorphism discovery has been a focus in the life sciences (Shendure *et al.*, 2004). The generation of high-density oligonucleotide microarrays suitable for whole-genome variation detection have been developed (Hinds *et al.*, 2005; Patil *et al.*, 2001) and are called 'tiling arrays' or 'whole genome resequencing arrays' (Zeller *et al.*, 2008). This type of arrays is currently only available for 3 model systems: *Arabidopsis thaliana*, *Drosophila melanogaster* and *Homo sapiens*. They contain oligos representing a 1bp tiling path (i.e. oligos that overlap by just 1bp difference) to query bases relative to a known reference sequence. Using these special arrays, the whole genome can be scanned systematically for the presence of larger insertions or deletions on the chromosomes, or for larger duplications of genomic areas of interest. The latter assay is called CNV (copy number variation). Copy-number variation is therefore defined as a form of genomic structural variation and refers to differences in the number of copies of a particular genomic region (Albert *et al.*, 2007; Bansal *et al.*, 2007). CNVs are considered to influence the quantity of the gene products because of the different copy number of certain genes available, or if a functional unit (including the regulatory region) of a particular gene is duplicated or multiplied (Zhang *et al.*, 2008). The involvement of copy-number alterations in human phenotypes was first reported as 'genomic disorders', which are often caused by de novo structural alterations (Inoue & Lupski 2002) and so far dozens of human genomic disorders have been reported (Lupski 2007).

Sequencing - Single Nucleotide Polymorphisms (SNP)

Sequencing - determining the order of the nucleotide bases in a molecule of DNA, was time consuming and expensive. Since 1975, when Sanger (Sanger *et al.*, 1992; Sanger & Coulson 1975) invented the chain-termination method, there were continuous improvements to the technique in order to achieve larger sample throughputs. During the human genome project, the biggest technical step has been made by the development of automated sequencers, allowing to automatically processing 384 samples at a time. Until then, sequencing mostly has been applied for de novo sequence generation, whereas resequencing efforts (i.e. looking for sequence differences on the genomic level) were limited to a few percent of the applications only.

With the Pyrosequencing technology (Ronaghi *et al.*, 1996), which introduced a new detection methodology including online detection, the door was opened for a new generation of highly parallel sequencing technologies that followed. Highly parallel sequence read outs are currently changing biological research (Hall 2007). These technologies, called 'Next Generation Sequencing' (NGS) have different applications such as resequencing of known genes or whole genomes and de novo sequence generation of unknown genomes. The three most widely used

technologies are 454 sequencing (Roche) where 400.000 fragments of up to 450 bp can be generated in parallel (Margulies *et al.*, 2005), SOLiD (Applied Biosystems), using sequencing by ligation and generating 20 gigabases per single run in approx. 400 Mio 50 bp fragments. Solexa, the third technology produces 100 bp reads for a total of over 20 Gb of paired-end data per run in approx. 1Mio reads. These NGS technologies are being widely used for transcriptome analysis, gene regulation analysis as well as resequencing of whole genomes for CNV analysis, the detection of somatic mutations or genome rearrangements and for SNP detection (Ellegren 2008b).

Single Nucleotide Polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide in the genome differs between members of a species or between homolog chromosomes in an individual. SNPs may be caused by changed (substitution), removed (deletions) or added (insertion) single bases in a given DNA sequence. Single-nucleotide polymorphisms may fall within coding or non-coding regions of genes, or in the intergenic regions between genes. SNPs are used as molecular markers that might have multiple effects: When occurring in coding regions they can lead to functional changes in the protein structure encoded in this gene. SNPs in non-protein-coding regions may still have consequences for gene splicing or transcription factor binding. The abundance of SNPs in genes ranges from 1 mutation in 30 bp –1000 bp, depending on the species and gene investigated (Ching *et al.*, 2002; Ramos *et al.*, 2009). In the human genome, SNPs are the most frequent type of variation. There are around 10 million SNPs that have been identified in the human genome (Sherry *et al.*, 2001). SNPs are usually biallelic and thus easily assayed. Following the association hypothesis, SNPs are considered a useful tool for population genetics. Particularly for the identification of genes or gene variants, that confer susceptibility to diseases (Nakamura 2009). Currently one of the key applications of next-generation sequencing is studying genetic variation between individuals using whole-genome or target region resequencing for SNP detection.

Once SNPs in the genome of a species are identified, large scale SNP genotyping tools are available to screen several thousand SNPs in parallel on one array. SNP arrays are setups which allow the detection of polymorphisms within a sample set in a highly parallel manner using either DNA hybridization and fluorescence microscopy, or more sophisticated detection methods like primer extension (Gunderson *et al.*, 2006), or an oligonucleotide ligase assay. The most advanced method currently is the Illumina bead array assay using primer extension for the analysis of over 100.000 SNPs in parallel (Li *et al.*, 2009). Another high throughput method is the PCR based Taqman assay (McGuigan & Ralston 2002).

SNP genotyping can also be performed using 'dynamic allele-specific hybridization' (DASH) or High resolution melting analysis (HMR) which exploits the fact that mismatched base pairs generate differences in the melting temperature of DNA (Vaughn & Elenitoba-Johnson 2004). Here the fact is used that SNPs usually cause a shift in the melting temperature of several degrees when being present in heteroduplex configuration. The differences in the melting temperature can be

monitored when using intercalating dyes which are being released from the dissociating double strand thus leading to the ceasing of a detectable light signal (Wittwer *et al.*, 2003).

As SNPs are highly conserved throughout evolution and within a population, SNPs serve as an excellent genotypic marker. SNP arrays are useful tools to study the whole genome. As each individual has different single nucleotide polymorphisms that together create its unique DNA sequence, SNP-based genetic linkage analysis can be used to map loci for traits of interest, and hence determine susceptibility genes for an individual. A SNP array can also be used to determine the copy number of each SNP on the array which even can be used then align the SNPs in chromosomal order (John *et al.*, 2004).

1.3.2 PCR-based techniques

With the advent of the Polymerase chain reaction (PCR) invented in 1983 by K. Mullis (Mullis & Faloona 1987; Saiki *et al.*, 1985) a large number of applications for molecular markers based on PCR were developed, turning this technology to the basic principle of today's molecular genetics toolbox. PCR promoted the development of a range of molecular assays which allow the detection of polymorphisms at the DNA level as a versatile and cheap alternative to hybridization based techniques.

Other than previous techniques, the PCR reaction allows to conduct a single assay with very little DNA (10ng) of rather crude quality, and these assays can easily be conducted in parallel (up to 384 or 1536 reactions/assay). In combination with other techniques like restriction digest and ligation, the direct substitution of RFLPs was possible. PCR usually is applied using known sequences (primers) for the specific amplification of regions of interest. These specific primers allow the use of total genomic DNA extracts to target either the nuclear or organellar genomes for analysis. Even when no 'a priori' sequence information is available for a species of interest, using either arbitrary primers, or restriction digest in combination with adaptor ligation can be applied in order to generate PCR based molecular markers or individual specific fingerprints.

1.3.2.1 Anonymous Markers

Other than sequence targeted PCR techniques, where the DNA sequence of the investigated locus is known, anonymous markers are generating multiband patterns targeting regions which are 'presumably present' and distributed throughout the genome. The detected regions are hit either by arbitrary primers, or primers representing sequences of repetitive nature present in the genome. Another possibility is to add, via ligation, known adaptor sequences to fragments generated from the genome of interest by restriction digest. These adaptor sequences can subsequently be used as primer binding site for the differential amplification of a subset of the generated ligation fragments.

Random amplified polymorphic DNA markers (RAPD).

In 1990 (Welsh & McClelland 1990) developed a new PCR-based genetic assay namely randomly amplified polymorphic DNA (RAPD). The basis of this technique is the differential PCR amplification of genomic DNA. It detects nucleotide sequence polymorphisms in DNA using a single short primer of arbitrary nucleotide sequence of 10 bp in length (Williams *et al.*, 1990b). In a typical RAPD reaction, the primer anneals to the genomic DNA at two different sites on complementary strands of DNA template, representing inverted sequence stretches of the same 10 bp. If these priming sites are within an amplifiable range of each other, a PCR product is generated. The used primer amplifies a number of discrete loci in the genome, thus generating a multi band pattern of amplification products being visible when analyzed on agarose gels. The technique therefore can be used to investigate nucleotide sequence polymorphism between individuals.

RAPD represents a dominant marker type, meaning that size variants i.e. alleles belonging to one single locus can not be discriminated from one another. Therefore there is no possibility to distinguish homo- from heterozygous individuals. Also the dosage effect being caused by multiple copies of the respective locus (i.e. in polyploidy plants) would not be visible. Another uncertainty in analysing RAPD profiles is the fact that the primers used also anneal to, and therefore amplify, organellar regions. To which extent the bands represented in a RAPD profile originate from chloroplast or mitochondrial DNA as compared to nuclear DNA is unclear. Still, due to the speed and efficiency of RAPD analysis, it was used in high-density genetic mapping in many plant species like faba bean (Torress *et al.*, 1993) or apple (Hemmat *et al.*, 1994), for diversity analysis in poplar cultivars (Rajora & Rahman 2003; Sabatti *et al.*, 2001) or oilseed rape varieties (Ma *et al.*, 2000), to compare wild to cultures *Amaranthus* (Chan & Sun 1997) or testing of genetic stability (Sánchez *et al.*, 2008). The popularity of RAPD application in genotyping has decreased due to the fact that reaction conditions strongly influence the resulting RAPD profile, more or less preventing the direct comparison of results between laboratories (Garcia-Mas *et al.*, 2000; Milbourne *et al.*, 1997).

DNA amplification fingerprinting (DAF).

An independently developed, yet RAPD related technique, is the DAF (Caetano-Anolles & Bassam 1993). Here, single arbitrary primers of only 5 bp are used to generate fingerprint like banding patterns by PCR amplification. The amplicons obtained are analyzed using polyacrylamide gels along with silver staining to achieve better resolution and sensitivity. Like RAPD, DAF is a codominant marker system that found its application in fingerprinting bacterial populations (Bassam *et al.*, 1992) or the characterization of the origin of Bermudagrass (Caetano-Anolles *et al.*, 1997).

Amplified fragment length polymorphism (AFLP)

The AFLP technology, also an 'arbitrary' PCR technique, helped to overcome the limitation of reproducibility associated with RAPD (Vos *et al.*, 1995). It is a technique

based on the detection of genomic restriction fragments by PCR amplification and can be used for DNAs of any origin or complexity. Genomic fingerprints can be produced, without any prior knowledge of the sequence, using a limited set of generic primers. The number of fragments detected in a single reaction can be adjusted by the selection of specific primer sets.

AFLP is a combination of RFLP with the PCR-mediated amplification of sequences and mainly involves 3 steps: First the generation of DNA fragments by restriction endonuclease digest using two different enzymes (*MseI* or *PstI* and *EcoRI*). These restriction fragments are in a second step ligated to adaptors fitting either restriction site, allowing selective amplification steps by using the adaptors as primer sites in PCR reactions. By altering the number of selective nucleotides which are added to the amplification primers, as well as by adjusting the primer combination, the number of bands generated during an AFLP analysis can be influenced. In a third step, the amplicons are analyzed on an automated sequencer. As only one of the two primers used for the PCR amplification is fluorescently labelled, only fragments will be detected, that were generated by this labelled primer.

The AFLP fragments also represent dominant markers and correspond mostly to unique positions on the genome. An AFLP band of a certain length represents a presence allele (scored 1) at such an AFLP locus. Individuals not having a band, instead have an absence allele (scored 0). The presence and absence character of the data cannot provide complete genotypic information for diploid organisms. Most population genetic analyses based on AFLP data assume that the absent allele really is absent from the data caused by either mutations in the restriction sites for the enzymes or in the sequence corresponding to the additional bases in the amplification steps. However, other types of mutations (*inDel*) may result in a DNA fragment of a different length and therefore a band at a different position, scored as a separate locus.

AFLP markers found a wide field of applications especially in plant genetics. AFLP markers are widely used when attempting to locate and characterize regions in the genome that affect the expression of quantitative traits when trying to improve the efficiency and precision of plant breeding programs. They have been used to determine the genetic architecture of economically important traits such as disease resistance in melon, barley or apricot (Budahn *et al.*, 2009; Lalli *et al.*, 2008; Teixeira *et al.*, 2008) and resistance breeding (Kim *et al.*, 2008; Parh *et al.*, 2008), productivity (Barbosa *et al.*, 2003; Tullu *et al.*, 2008) or malting quality (Faccioli *et al.*, 1999) but also to infer the history of crop domestication (Allaby & Brown 2003). To localize genetic regions that code for quantitative traits, so-called quantitative trait loci (QTL), it is essential to have access to a linkage map. AFLP, together with microsatellites, is the most common marker used when developing new linkage maps (Pakull *et al.*, 2009; Travis *et al.*, 1998). AFLP has proven to be especially suitable for enrichment of existing genetic maps (Blas *et al.*, 2009) and for producing consensus linkage maps (Isobe *et al.*, 2009). AFLP can also provide insights into the evolutionary relationships between organisms, patterns of introgression or hybridization (Gedil *et*

et al., 2009; Paun *et al.*, 2007; Travis *et al.*, 2004; Wilson *et al.*, 2005) and genome-wide diversity measures (Leitch *et al.*, 2009).

Using AFLPs to investigate the level of genetic diversity reveals information about historical population sizes and structure (Ortiz *et al.*, 2009; Rahimmalek *et al.*, 2009; Terrab *et al.*, 2008; Terrab *et al.*, 2009). For managing rare and threatened species, knowledge of intraspecific genetic variation may help to assess extinction risks and evolutionary potential in a changing world. Many molecular ecologists study the genetic structure of populations in order to estimate historical events (Terrab *et al.*, 2008). In a review about genetic structures in plants, Nybom (Nybom 2004) demonstrated that F_{ST} values (and F_{ST} analogues) obtained from dominant markers (AFLP and RAPD) were overall similar to estimates obtained from microsatellites and allozymes.

cDNA-AFLP

A quick and quite reliable method to study differences at the gene expression level between two individuals from one species as well as between closely related species is the cDNA AFLP or differential display. Here, mRNA populations from two individuals are used as starting material for the analysis. When following the AFLP process on cDNA generated from the initial mRNA, differences in the expression pattern can be observed, being visible as differences in the observed banding pattern. This method found restricted application in plant molecular analysis, although interesting results could be generated (Simoes-Araujo *et al.*, 2002; Dubos & Plomion; 2003).

AFLP based DNA methylation detection

Methylation of cytosine is a common epigenetic modification of DNA in most eukaryotic organisms. The methylation patterns of DNA are relatively stable over cell generations, but can also be modified by environmental influences (Bender 2002). Variation in methylation has been found to influence gene expression and genomic imprinting. Patterns of DNA methylation can be retrieved by a modification of the original AFLP protocol using isoschizomeric restriction enzymes which cut DNA at the same sequence motif with differential sensitivity to DNA methylation. Methylation-sensitive AFLP has been used to study vernalization in winter wheat (Sherman & Talbert 2002) and ecotype differences in *Arabidopsis thaliana* (Cervera *et al.*, 2002).

Inter simple sequence repeat markers (ISSR)

ISSR is a technique where primers based on microsatellites (see 1.3.2.2.) are utilized to amplify inter-SSR regions of the genome (Zietkiewicz *et al.*, 1994). Like in the RAPD technique, just one primer is used in the PCR reaction. Here, the primer consists of a microsatellite repeat unit (i.e. (TG)₉). When combining this technique with the AFLP approach, or with the RAPD technique (Wu *et al.*, 1994), a new type of anonymous dominant markers can be generated, the RAMP (Random amplified Microsatellite Polymorphism). Microsatellite-AFLP is a fingerprinting method for simultaneous amplification of microsatellite- and AFLP markers. Microsatellite-AFLP

uses the combination of a randomly amplified microsatellite polymorphism (RAMP) primer and a selective AFLP primer to amplify restriction fragments containing simple sequence repeat (SSR) motif sequences for diversity studies (Yang *et al.*, 2006).

SSAP: Sequence Specific Amplified Polymorphism

SSAP is a multi banding marker technology which first was described by Waugh *et al.* (Waugh *et al.*, 1997b) and describes a method where differences in the insertion sites of transposable elements are targeted to generate polymorphic markers. Due to their conserved nature, different elements of the ClassII retrotransposons can be used to perform an SSAP assay, combining AFLP with a site specific PCR reaction targeting a specific retrotransposable element. Two approaches have been pursued i) using the LTR (long terminal repeat) region as primer anchoring site for the PCR amplification step, or ii) using the RNaseH-LTR region in combination with a restriction site at the other end (Berenyi *et al.*, 2002). The parallel analysis of retrotransposon insertion site polymorphisms has been used for linkage and diversity analysis in various species like pea, vine or oil palm (Ellis *et al.*, 1998; Forneck *et al.*, 2009; Labra *et al.*, 2004; Price *et al.*, 2003).

1.3.2.2 Sequence Specific Marker Types:

Whenever sufficient knowledge on the genome or parts of the genome of interest is available, PCR reactions can be targeted to specific sites. Markers generated using sequence specific primers have the advantage of allowing the discrimination of single alleles of the same locus, thus exhibiting codominant nature. With these markers homo- and heterozygous individuals can be discriminated easily, allowing paternity analysis. They are widely used for anchoring genetic maps that have been generated using anonymous markers. Further more they have wide applications in plant breeding, diversity and population genetic analysis. With the growing amount of sequence information available in public databases, often anonymous or former hybridization based markers can be turned over to Sequence tagged Sites (STS). In this way, repetitive DNA (SSR, repeat classes i.e. transposon), VNTR, telomeric, centromeric regions as well as single copy regions (genes and intergenic regions, SNPs and InDels) can be investigated.

Sequence-tagged sites (STS)

Anonymous markers like RAPD, AFLP or cDNA AFLP bands as well as RFLP are very often of interest for further analysis. Especially when these markers are linked to a trait of interest, then these genomic regions can serve as markers linked to QTLs. In this case, there is strong interest to get the exact sequence information for this locus, and to convert the probes into PCR-based STS markers based on their nucleotide sequence (Olson *et al.*, 1989). Therefore a STS is a short DNA sequence that occurs only once in a genome or chromosome. With the obtained specific amplicons, variations at the genomic level can be specifically investigated. There is a

wide palette of reports for anonymous markers being converted to informative STS markers like in rice, where RAPD markers have been transformed to STS markers for variety testing (Monna *et al.*, 1994) or by converting AFLP markers to STS as markers for male sterility in radish (Murayama *et al.*, 2003). When these markers are linked to specific traits, they can be easily integrated into plant breeding programmes for marker-assisted selection of the trait of interest.

Microsatellites (SSR, simple sequence repeats)

Within the past decade microsatellites have developed into one of the most widely used source of nuclear as well as organellar markers for a wide range of applications, from genome mapping to forensic testing from population genetic studies to phylogeographical investigations. Microsatellites were first described by (Litt & Luty 1989), and represent a ubiquitous class of simple repetitive DNA sequence. These repetitive elements have been described in all eukaryotes analyzed so far. Features such as hypervariability i.e high degree of polymorphism and ubiquitous occurrence explain their widespread use. Microsatellites, comprising tandemly repeated short nucleotide sequences of 2 to 6 bp can be easily PCR typed and follow, when located in the nuclear genome, Mendelian co-dominant inheritance. Studies have been conducted in order to explain the origin of this high degree of variability (Schlotterer 2000). Mutations within microsatellites are frequent, altering their overall length by insertion or deletion of a small number of repeat units which is thought to be caused by replication slippage. In humans, the rate of contraction mutations increases exponentially with allele size, whereas the rate of expansion mutations is constant across the entire allele distribution (Xu *et al.*, 2000).

Microsatellites have been shown to be associated with human disease, not only as markers of risk but also directly with diseases like fragile X or Huntington's disease (Koreth *et al.*, 1996). It has been observed, that in plants the overall frequency of microsatellites is inversely related to genome size and to the proportion of repetitive DNA, still remaining constant in the transcribed portion of the genome (Morgante *et al.*, 2002a). The microsatellite frequency is higher in transcribed regions, especially in the untranslated portions showing a significant association with the low-copy fraction of plant genomes (Galindo *et al.*, 2009).

Expressed sequence tag markers (EST).

When working with complex genomes like in the plant kingdom, there is interest in getting rid of the repetitive part of the genome which sometimes adds up to more than 90% of the total genomic DNA. One approach is to just work on the expressed part of the genome which is represented in the mRNA population of a tissue. When extracting mRNA and converting it to cDNA and further to cDNA libraries in plasmid vectors, the expressed part of the genome can directly be analyzed. The EST term describes the fact, that when isolating and sequencing expressed genes starting from mRNA, often only partial sequence information for one gene can be generated due to technical reasons. In addition the sequences generated on randomly picked clones

usually are generated by just one sequencing run into one direction (single pass) generating sequence reads of about 700–800 bp at the maximum, being limited by the Sanger sequencing chemistry. This is why the sequence information generated serves as a tag for the full gene. These single pass sequence reads are of rather poor quality which has to be kept in mind when working with ESTs from public data bases.

A new approach towards more efficient generation of EST information as well as whole transcriptome analysis is the use of NextGen sequencing technologies, where an isolated mRNA population is directly subjected to sequencing without the time consuming library construction. This generates a vast amount of data and is currently applied to species, where a lot of knowledge on the transcriptome is already available (Wall *et al.*, 2009). As the generated sequences are rather short (up to 450 bp max, mostly 50 bp), alignments to already characterized sequences is recommended. In order to assign a function to these short tags, also longer reference sequences are needed.

The 3' end of an EST can serve as a source for the development of STSs because of their likelihood of being unique to a particular species. They have the additional benefit of being directly associated to an expressed gene. Once generated, they are useful in cloning specific genes of interest and synteny mapping of functional genes (Mochida *et al.*, 2008). Comparative analysis as well as application as markers across closely related species (COS markers) are applications for markers targeting ESTs. They are used in full genome sequencing projects to facilitate genome annotation. ESTs are frequently used for SNP detection, and for several species like for rice (Feltus *et al.*, 2004); <http://irfgc.irri.org/>) wheat (<http://wheat.pw.usda.gov/SNP/snpdb.html>) or barley (Rostoks *et al.*, 2005)[http://bioinf.scri.ac.uk/ barley_snpdb](http://bioinf.scri.ac.uk/barley_snpdb)) these SNPs are described in public data bases wherein thousands of functional cDNA clones are being converted into SNP marker regions. Also for non-model species like melon, ESTs serve as a valuable source for SNP marker development applicable for cultivar identification (Deleu *et al.*, 2009a).

Single strand conformation polymorphism (SSCP)

Apart from DNA sequencing, SSCP is a way to discover new DNA polymorphisms. SSCP uses amplicons of a specific locus that have been produced in a preceding PCR reaction. SSCP analysis is an useful application for the detection of mutations based on the ability of a single nucleotide change in a given sequence to alter the electrophoretic mobility of that DNA single strand under non-denaturing conditions (Orita *et al.*, 1989). Other than for double stranded DNA, the mobility of single strands (ssDNA) in electrophoretic gels is affected by changes in sequence, leading to instable intra-strand base pairing, resulting in loops and folds that give the single strand a unique secondary structure. During SSCP analysis dsDNA undergoes denaturation to form ssDNA followed by subsequent folding, generating a unique conformational state based on its sequence. A single nucleotide change therefore dramatically affects the mobility of the ssDNA strand through a gel by altering the

conformation. By this conformational changes, SSCP can identify heterozygosity of DNA fragments of the same molecular weight and even detect changes of a few nucleotide bases as the mobility of the single-stranded DNA changes with change in its GC content. This is a rapid technique for allele analysis particularly for detection of point mutations as well detection of DNA polymorphisms and mutations at multiple sites in DNA fragments. Apart from regular mutation detection in genes of interest, SSCP has been applied in genotyping (Kuhn *et al.*, 2008) as well as in molecular ecology (Sunnucks *et al.*, 2000) or mutation detection (Sato & Nishio 2003).

SNP detection using PCR based assays

In the amplification refractory mutation system (ARMS) strategy a PCR primer is designed in a way that it is possible to discriminate among templates that differ by only one single nucleotide (Marefat *et al.*, 2006). ARMS has also been termed allele-specific PCR. The main advantage of ARMS is that the amplification step and the diagnostic steps are combined, so that the presence of an amplified product indicates the presence of a particular allele whereas the absence of a PCR product shows the absence of a particular allele/SNP. Here also quantitative PCR analysis as well as direct detection of positive PCR results by intercalating dyes or TaqMan probes can be applied to discriminate between alleles (Lyon *et al.*, 2009).

HRM and its combinations

A new technology that can be seen as a further improvement of SSCP is high resolution melting analysis (HRM). First described in 2003 (Wittwer *et al.*, 2003) it detects new sequence variants in double stranded amplicons with unknown variations. As the location of a potential sequence variation is not known, sequence-specific analysis cannot be done, unless sequencing is applied. High resolution melting uses special dyes that fluoresce only in the presence of double stranded DNA (dsDNA) as they have the ability to intercalate into the double strand structures (SybrGreen, LC Green; (Wittwer *et al.*, 2003). When the sample is heated to high temperatures, the dsDNA gradually denatures forming ssDNA which can not accommodate the dye anymore. Thus the fluorescent color fades away as the dsDNA strands separate. This process can be monitored online in a Lightcycler, resulting in a specific melting curve of the respective fragment. Unknown sequence variations in diploid organisms become visible in a heterozygous state due to the presence of heteroduplex DNA which is formed between the two heterologous alleles in the analysis process. Due to a shift in melting temperature of the heteroduplexes, the shape of the melting curve is altered as compared to the curves produced by homologous strands. The process can be used for specific genotyping, comparing sequence identity between two DNA samples (Vaughn & Elenitoba-Johnson 2004), and scanning for any sequence variant within a locus. Recent studies applied HRM as a cost efficient method that can be employed as an initial screen for the identification of genes or genomic regions harbouring point mutation/SNPs. The

results of these initial screens can serve as the basis for in depth sequencing of the putative polymorphic regions for subsequent SNP assay development.

1.4 Molecular markers in plant genome analysis

The fast development and widespread use of molecular markers during the last decades shows that these tools provide easy and fast access to relevant information on the genomes of interest for scientists as well as breeders (Agarwal *et al.*, 2008a; Angaji 2009; Dwivedi *et al.*, 2007). Genome analysis based on molecular markers is generating a vast amount of information which can be used for manifold purposes. Molecular markers are used as tools for a large number of applications ranging from localization of a gene to the improvement of plant varieties by marker-assisted selection. From the estimation of post glacial migration of species to population genetic analysis as well as plant systematics and taxonomy molecular markers are being applied. They have also become popular tools for phylogenetic analysis opening new insights to develop evolutionary theories, and are used to develop a new view on the 'tree of life'. Genetic diversity of plant species plays a major role in the sensitivity, stability, and dynamics of ecosystems, and is one of the most important determinants in conferring the adaptive potential of populations or species. Appropriate conservation programs cannot be envisaged without detailed information on the amount and distribution of the genetic resources of a species.

Population genetics and diversity studies

Early theories of evolution were based on morphological and geographical variations between organisms. However, nowadays molecular biology techniques provide detailed information about the genetic structure of natural populations. RFLP, DNA sequencing, and a number of PCR-based markers are being used extensively for reconstructing phylogenies of various species up to the barcode of life (Song *et al.*, 2008). Furthermore, these studies hold a great promise for revealing more about the pattern of genetic variation within species (Allan *et al.*, 2008).

Genotyping provides a measurement of the genetic variation between members of a given population or sample set. Using molecular markers to analyze the DNA of all individuals of interest provides highly accurate tools for assessing population size, sex ratio, mortality, reproduction, effective population size, inbreeding, and migration and therefore are very useful in monitoring population growth or decline (Fraser *et al.*, 2007). Additionally, dispersal and migration patterns inferred from family relationships of fragmented populations are useful for evaluating the population's status and its putative history.

The goal of population genetics is the study of the allele frequency distribution and change under the influence of various forces like selection and adaptation, gene flow, mutation or genetic drift that result in evolutionary changes in a given species over time. In order to judge on conservation needs of endangered species, the putative effects of loss of genetic diversity and changes in the genetic structure on the long-term survival need to be assessed. Using molecular tools, the effective population size and the degree of inbreeding in a fragmented or declining population

can be effectively estimated. It also takes account of population subdivision and population structure in space and measures gene flow between populations as well as between species and therefore, it attempts to explain adaptation and speciation.

Molecular markers in taxonomy

Molecular systematics is an essentially cladistic approach: it assumes that classification must correspond to a phylogenetic descent. First analysis were done by with chemotaxonomy which made use of proteins, enzymes or carbohydrates which were separated and characterized using techniques such as chromatography. These have been largely replaced in recent times by DNA sequencing which produces the exact sequences of nucleotides of mitochondrial, nuclear or chloroplast DNA.

These are generally considered superior for evolutionary studies since the actions of evolution are ultimately reflected in the genetic sequences where the molecular clock assumption is used that quantitative similarity of genotype is a sufficient measure of genetic divergence.

In a molecular systematic analysis, the haplotypes are determined for a defined area of genetic material for a sufficiently big number of individuals of the target species or taxon. Delimitations of species are crucial for correct and precise identification of taxa. The application of molecular techniques is rapidly transforming the study of plant systematics due to the precision they offer. Thus they enable researchers to classify plants that have not been subject to rigorous classification before and thus molecular tools allow to obtain a clearer picture of evolutionary relationships.

The impact of molecular phylogenetics on the angiosperms has been particularly interesting. Inferences from three genes (*rbcl*, *atpB*, 18S rDNA) agree in the major features of angiosperm phylogeny and have resulted in a reclassification of the angiosperms (Soltis & Soltis 2000). Phylogenetic information has also contributed substantially to new perspectives on the evolution of polyploid genomes. For example in the area of peat mosses (Sphagaceae: Bryophyta) it was found that the evolutionary history of this group is more complex than previously thought. Based on nuclear SSRa and plastid markers, allopolyploidy and homoploid hybridization could be detected in populations of the *Sphagnum subsecundum* complex (Ricca & Shaw 2010).

Studying the genetic variation in plants using molecular markers revealing patterns of intra- or inter-specific variations is applied to understand their evolution from their wild progenitors and to classify them into appropriate groups. The taxonomic classification is an essential first step to determine whether any germplasm is a part of the primary, secondary or tertiary gene pool of the system concerned. This is especially important in cases where morphological markers can prove to be inaccurate and misleading. Examples are the reports of Muir *et al.*, 2000 (Bacilieri *et al.*, 1995) on oak, studies on musa (Boonruangrod *et al.*, 2009) or eucalyptus (McDonald *et al.*, 2009).

DNA barcoding serves as tool for rapid species identification based on DNA sequences and offers new answers to questions previously beyond the reach of

traditional disciplines. DNA barcodes consist of a standardized short sequence of DNA (400–800 bp) that in principle should be easily generated and characterized for all species (Fazekas *et al.* 2009; Hollingsworth *et al.* 2009; Kress & Erickson 2008; Pennisi 2007). DNA barcoding aims to use the information of one or a few gene regions to identify all species on earth. DNA barcoding will allow users to efficiently recognize known species based on their DNA barcode which will be accessible via a standardized on-line digital library of barcodes (Kress & Erickson 2008).

Molecular tools in plant breeding

Plant improvement, either by natural selection or through the efforts of breeders, has always relied upon creating, evaluating and selecting the right combination of alleles. Breeders have traditionally improved plant varieties by selection and crossings based on their phenotypic appearance. Now various molecular markers serve as tools to generate linkage maps which provide a more direct method for selecting desirable genes via their linkage to easily detectable markers. A genetic linkage map indicates the relative locations of specific markers along a chromosome. Thus, applying these markers allows dissecting the genome into genetic linkage groups representing the chromosomal units (Rae *et al.* 2007). To generate linkage maps, a suitable mapping population derived from a cross (full sib family) of two individuals exhibiting different phenotypic traits as well as a large set of markers providing sufficient variable alleles for map saturation are needed. For the construction of a linkage map of for example durum wheat, co-inheritance of allelic variants of different marker types applied have been analyzed (Mantovani *et al.* 2008). Due to meiotic recombination, two markers originally located on the same chromosome can become separated. The smaller the distance between the loci, the lower the probability that a crossing over will occur in the space between them, and thus the stronger is their linkage (Williams *et al.*, 1995). The first genome map in plants was reported in maize, followed by rice and *Arabidopsis* (Beavis & Grant 1991; Causse *et al.*, 1994; Hauge *et al.*, 1993) using RFLP markers. Maps have since then been constructed for many other crops like potato (Jacobs *et al.*, 1995), barley (Ramsay *et al.*, 2000), papaya or apple (Blas *et al.*, 2009; Celton *et al.*, 2009). Once the framework maps are generated, a large number of markers derived from various techniques are used to saturate the maps as much as possible. Comparative mapping between closely related species reveals synteny regions between genomes that have been conserved throughout species evolution (Ahn & Tanksley 1993; Barreneche *et al.*, 2004a; Ellegren 2008a; Gale & Devos 1998).

Markers have proved their importance as helpful landmarks in locating and manipulating quantitative trait loci (QTL). A QTL is a region in the genome that affects the trait or traits of interest and which can be allocated to a specific site on a genetic linkage map. When having a QTL co-located to a known marker, cloning of the gene responsible for the trait is getting possible. Markers linked to genes involved in several traits like leaf rust resistance in coffee (Herrera *et al.*, 2009), citrus leprosis resistance (Bastianel *et al.*, 2009) or leaf length in grasses (Barre *et al.*, 2009) will

allow map based cloning. Allele-specific associated primers help in genotyping of allelic variants of loci that result from size differences or point mutations like the waxy gene locus in rice (Sato & Nishio 2003).

The availability of such linkage maps facilitated the development of marker-based gene tags, map-based cloning of agronomically important genes like malting quality in barley (Schmalenbach & Pillen 2009; Williams *et al.*, 1995), tomato fruit size (Brewer *et al.*, 2007; Chaim *et al.*, 2003; Van Der Knaap & Tanksley 2003; Williams *et al.*, 1995) or male sterility in melon (Causse *et al.*, 2004; Park *et al.*, 2009; Schmalenbach & Pillen 2009; Williams *et al.*, 1995), synteny mapping (Devos *et al.*, 1995; Gale & Devos 1998; Grant *et al.*, 2000) as well as marker-assisted selection of desirable genotypes or traits (Lecomte *et al.*, 2004; Van Der Knaap & Tanksley 2003; Williams *et al.*, 1995).

2 Specific marker development and applications for selected research topics – thesis work

Depending on the taxonomic group, species-, population- or individual-level, different marker systems for discrimination experiments are available. Arbitrary, semi-arbitrary methods, sequencing data and other multi-locus profiling techniques are used to provide information for answering questions at the species level. In addition to the various marker types described above, these markers are available in the 3 different DNA containing compartments of a plant cell: i) the nucleus, ii) the chloroplast (cp) and iii) the mitochondria (mt). The organellar DNA (cp, mt) follows uniparental inheritance whereas the nuclear genome is bi-parentally inherited according to the Mendelian laws. In the uniparentally inherited organellar DNA, no recombination is occurring, in contrast to the nuclear genome, thus different questions can be answered when targeting the different genomes.

2.1 Markers in the Nuclear Genome

Introduction

The size of the nuclear genome of plants varies considerably (Leitch *et al.*, 2005; Soltis *et al.*, 2003) as genomes of eukaryotic organisms vary over 200,000-fold in size. The first genome size estimates were conducted in the late 1940s. Nuclear DNA content data for >10 000 species of plants, animals and fungi are currently made freely available through three independent databases (Gregory *et al.*, 2007). Data on plant genome sizes can be found at the Plant DNA C-values Database (<http://www.kew.org/genomesize>) Among angiosperms, genome sizes range from 108 Mb for *Fragaria viridis* (Bennett & Leitch 2005) to >130,000 Mb in some members of the Liliaceae like *Fritillaria assyrica* (Leitch *et al.*, 2005; Leitch *et al.*, 2009). *Arabidopsis thaliana*, the smallest recorded flowering genome has a size of 157 Mb (Arabidopsis Genome Initiative 2001) whereas the amoeba *Polychaos dubium* ("Amoeba" dubia) possesses the largest genome known so far with 670,000 Mb (Parfrey *et al.*, 2008). The first tree genome sequenced was the one of *Populus trichocarpa* with 480 Mb (Tuskan *et al.*, 2006). This shows that huge variation in genome size is common among distantly related organisms, but it also is unexceptional even among closely related species like in vicia, where genome sizes range approximately sixfold among members of the genus. As laid out by (Bennetzen 2002) some portion of this genome size variation may be ascribed to differences in gene, chromosome segment, and whole genome duplications, as well as to gene loss. Nevertheless, >90% of plant genes possess close homologs within other plant species, indicative of highly conserved gene content (Conte *et al.*, 2008).

The plant genome includes both genes and non-coding sequences. In plants, amplification and insertion of newly activated long terminal repeat (LTR) retrotransposable elements appear to be major contributors to genome size

expansion. For example, 70% of the maize nuclear genome is composed of LTRretrotransposons (Kumar & Bennetzen 1999). In the span of a few million years, the maize genome doubled in size due to transposable element (TE) activity. These TEs are often found in nested arrangements located between “gene islands” and often are associated with centromeres. Other mechanisms proposed to be responsible for genome size expansion include intron size expansion of tandemly repetitive DNA sequences (Meyers *et al.*, 2001), segmental duplication (Bancroft 2001; Flagel & Wendel 2009; Grant *et al.*, 2000; Soltis *et al.*, 2007), accumulation of pseudogenes (Guo *et al.*, 2009; Karro *et al.*, 2007; Podlaha & Zhang 2009), and transfer of organellar DNA to the nucleus (Millen *et al.*, 2001; Shahmuradov *et al.*, 2003).

Some organisms have multiple copies of chromosomes i.e. diploid (2), triploid (3), tetraploid (4) and so on. During meiosis, the chromosome set is divided into two so that the gamete receives half the number of chromosomes of the somatic cell. This typically happens when gametes are produced (pollen and ovary cells) and the genetic information is being transferred to the F1 generation according to Mendel's laws.

Eukaryotic genome data are becoming increasingly important both as the basis for comparative research into genome evolution as well as serving bioinformatics approaches for marker development. In addition, genome sequences of model organisms are very helpful for expanding research on non-model organisms, as orthologous sequences can easily be fished from the unknown genomes when using sequence information based on already known closely related species (COS markers) (Liewlaksaneeyanawin *et al.*, 2009).

SSRs

Microsatellite markers were first described by Weber and May for humans in 1989 (Weber & May 1989). SSR ('Simple sequence repeats' or 'microsatellites') are short segments of two or more (usually up to six) base pairs repeated tandemly. The typical characteristics of microsatellites, such as their presence in the genomes of all living organisms, high level of allelic variation, co-dominant mode of inheritance and potential for automated analysis make them an excellent tool for a number of approaches like genotyping, mapping and positional cloning of genes. Although microsatellites are present both in non-coding and coding regions, their frequency is higher in transcribed regions, especially in UTRs (Morgante *et al.*, 2002b). The microsatellites in coding and non-coding regions may also differ in their composition. For seven eukaryotic clades, Metzgar (Metzgar *et al.*, 2000) reported that tri- and hexanucleotide repeats were present in both coding and noncoding sequences, but other repeat types like di- or tetra -were much less frequent in coding regions. This might be attributed to the fact that genes and thus mRNA is translated to protein sequences by translating triplets of the sequence to an amino acid code, and thus variation in coding regions is only viable when 3 bases are concerned i.e. a triplet is added or deleted.

SSRs are considered the most efficient markers, but their use is still limited because of the long and laborious steps to develop them. With the availability of freely accessible genome information on various genomes, the *in silico* identification of SSR regions for marker applications has become possible. Other than in previous days, when enriched libraries (Edwards *et al.*, 1996; Hamilton *et al.*, 1999; Milbourne *et al.*, 1998) had to be generated in order to isolate SSRs from a genome of interest (Zane *et al.*, 2002) the vast amounts of public genomic sequences, EST and BAC sequence resources can be utilized for marker development (Kofler *et al.*, 2007). In rice, Cho *et al.*, (Cho *et al.*, 2000) showed that microsatellites derived from genomic libraries detected a higher level of polymorphism than those derived from ESTs contained in the GenBank database (83.8% vs. 54.0%). As has been shown by Ritland (Rungis *et al.*, 2004), EST data from pines could be efficiently used to generate SSR markers that work across a set of closely related species. Using ESTs as a source for marker development has the advantage that expressed sequences i) are more conserved across species, which eases transferability of markers and ii) are located in functional units i.e. expressed genes, giving them a putative correlation to phenotypic variation (Varshney *et al.*, 2002; Varshney *et al.*, 2005).

Hypervariability of SSRs allows using them for numerous purposes such as the construction of molecular maps. EST derived microsatellite loci were detected and mapped in many species, such as barley (Becker & Heun 1995), maize or rice. Their hypervariable nature made them attractive for breeders who could utilize them for marker assisted selection (MAS) and map construction (Peleg *et al.*, 2008) as molecular markers tightly linked to a target gene of agronomic importance. Zhou (Zhou *et al.*, 2003) showed that MAS for the major scab resistance QTL in wheat was much more effective when using SSRs in combination with phenotypic traits. Because of the possibility to detect several alleles at a high frequency, SSRs also are an ideal tool for identifying genetic diversity between individuals and the genotyping of cultivars. Even in self-pollinated species like wheat or barley just a few tens or SSRs were enough to discriminate between 48 genotypes of wheat (Prasad *et al.*, 2003). It was demonstrated that microsatellites in plants could even be up to ten-fold more variable than other markers; thus, they are highly suited for genetic diversity analysis. Due to their instability and sensitivity to stress conditions during replication, SSRs were even used for the detection of somaclonal variation in poplar (Rahman & Rajora 2001) or oak (Wilhelm *et al.*, 2005). SSRs are useful for quantification of genetic diversity, characterization of accessions in plant germplasm collections and taxonomic studies. Microsatellite markers have also been advantageous in pedigree analysis as they represent a single locus. The multi allelic state of this marker facilitates variability detection across a wide range of germplasm and allows individuals to be genotyped, so that gene flow (Marques *et al.*, 2002) and paternity can be established. One interesting application of this markers has been shown in sex identification of papaya where the microsatellite probe (GATA)_n is found to reveal sex-specific differences in Southern hybridization and thus can be used as a

diagnostic marker in this system where male and female plants do not show any sex-specific morphological difference until flowering (Parasnis *et al.*, 1999).



2.1.1 SSR Markers Development for *Rhododendron*

Introduction

Rhododendron hirsutum, *Rhododendron ferrugineum*, *Rhododendron intermedium* are three *Rhododendron* varieties of the alpine rose, endemic to alpine areas in Austria, Slovenia, Italy and Switzerland. It is a species of ecological and economical interest. Due to its natural habitat on alpine slopes it can be used as plant cover for skiing slopes, where it prevents erosion. On the other hand these species are of interest for gardeners as there is a lot of interest on these small rhododendrons to be planted in rock gardens. Little is known on the genetic diversity of this species. *Rhododendron ferrugineum* L. is an acidophilous evergreen shrub growing on arenaceous soils with more or less podsol features being highly acidic. It is found in the Pyrenees, the Jura Mountains, the Alps and Carpathians with large populations at elevations of 1600–2200 m. *Rhododendron hirsutum* on the other hand grows at somewhat lower elevations on more basic soils. *R. intermedium* is thought to be a hybrid of *R. hirsutum* and *R. ferrugineum* showing an intermediate phenotype. The species reproduce both sexually and asexually by selfed and out-crossed seeds (Escaravage *et al.*, 1997; Escaravage & Wagner 2004) as well as vegetatively through layering (Pornon *et al.*, 2000; Pornon & Escaravage 1999).

Molecular markers so far applied RAPD (Random Amplified Polymorphic DNA) (Wolf *et al.*, 2004) and AFLPs (Amplified Fragment Length Polymorphisms) (Escaravage *et al.*, 1998) revealed low values of genetic variation in Alpine *R. ferrugineum* populations- and SSR markers have been developed for *R. meternichii* (Naito *et al.*, 1998). In order to have a better set of primers, (heterologous primer sets were reported to exhibit reduced variability), new SSR markers for *Rhododendron* were developed for population genetic analysis

2.1.1.1 Microsatellite markers for *Rhododendron ferrugineum* and *Rhododendron hirsutum*, sympatric alpine shrubs

Fluch S¹, Hristoforoglu K.², Kopecky D.¹, Berenyi M¹

¹ Austrian Institute of Technology GmbH – AIT, Bioresources, A-2444 Seibersdorf

² Höhere Bundeslehr- und Versuchsanstalt für Gartenbau, Grünbergstraße 24, A-1131 Wien

Keywords: *Rhododendron ferrugineum*, *Rhododendron hirsutum*, SSR, genetic diversity

Submitted to Molecular Ecology Notes

Abstract

We report on the characterization and use of 27 SSR markers for *Rhododendron ferrugineum* and *Rhododendron hirsutum*. These markers were obtained from a genomic library enriched for SSR motifs. 17 of the 27 designed primers revealed variable fragments in the chosen test set, but only 11 of these regions were producing a nice clear banding pattern and could be used for further analysis. These 11 primer pairs revealed 3–24 alleles per locus, 110 alleles in the over 230 samples tested so far.

A subset of 4 primer pairs additionally was used for variability tests on a set of 227 individuals of *R. ferrugineum*, *R. hirsutum* and *R. x intermedium* from 11 natural populations from Austria, Switzerland, Italy and Slovenia revealing 65 different alleles. Estimation of genetic diversity as well as the identification of clonal material in the populations was possible. These newly developed SSR markers are therefore well suited for population genetic analysis in Alpine *Rhododendron* species.

Rhododendron hirsutum, *Rhododendron ferrugineum* and *Rhododendron x intermedium* sometimes called *alpenrose* are evergreen subalpine shrubs that grow just above the tree line in mountainous areas in Austria, Slovenia, Italy and Switzerland. *R. ferrugineum* L. (the *rusty-leaved alpenrose* due to its brown colored leaves) is an acidophilous evergreen shrub growing on arenaceous soils with more or less podsol features being highly acidic (Milne, R. I. and Abbott, R. J.; 2008). It is found in the Pyrenees, the Jura Mountains, the Alps and Carpathians with large populations at elevations of 1600–2200 m. This is in contrast to *Rhododendron hirsutum*, which has hairy edges on the leaves and grows over limestone on in somewhat lower elevations on more basic soils. Where the two species co-occur, hybrids between the acid-loving species *R. ferrugineum* and the basic soil species *R. hirsutum* occur on soils of intermediate pH in the European Alps, being termed *R. x intermedium* exhibiting an intermediate phenotype. These *Rhododendron* species reproduce both sexually and asexually by selfed and out-crossed seeds (Escaravage, N. and Wagner, J. 2004) as well as vegetative through layering (Pornon, A. and Escaravage, N. 1999); (Pornon, A. *et al.*, 2000). The alpine rose is a species of ecological and economical interest. Due to its natural habitat on alpine slopes it can be used as plant cover for skiing slopes, where it prevents erosion. On the other hand these species are of interest for gardeners as there is a lot of interest on these small *Rhododendrons* to be used as ornamentals in rock gardens.

Little is known on the genetic diversity of this species. Molecular markers so far applied RAPD (Random Amplified Polymorphic DNA) (Wolf, P. G. *et al.*, 2004), (Milne, R. I. and Abbott, R. J. 2008) and AFLPs (Amplified Fragment Length Polymorphisms) (Milne, R. I. and Abbott, R. J. 2000, Pornon, A. and Escaravage, N. 1999) revealed low values of genetic variation in Alpine *R. ferrugineum* populations. Codominant marker systems like SSR markers so far only have been developed for *R. meternichii* in 1998 (Naito, K. *et al.*, 1998). In order to have a suitable set of primers, new SSR markers for the alpine *Rhododendron* species were established.

Genomic DNA of four unrelated individuals of the two species of interest from three different geographical origins was extracted using 100 mg fresh leaf material per plant (Qiagen Plant Minikit). Construction of an enriched library based on the protocol established by Refseth (Refseth, U. H. *et al.*, 1997) was performed on Sau3A partial digested DNA pools. After adaptor ligation (Sau3AI A Oligo, 5'-GGCCAGAGACCCCAAGCTTCG-3' Sau3AI B Oligo, 5'-PHO-GATCCGAAGCTTGGGGTCTCTGGCC-3') fragments were PCR amplified with Sau3A adaptor fitting primers, and PCR of 400–1000 bp isolated from an preparatory agarose gel. Magnetic streptavidin coated beads (Dynal, Dynabeads) were functionalized by (AG)₁₀ and (CA)₁₀ synthetic oligo nucleotides. These beads were then used for capturing SSR containing fragments present in the Sau3A amplification product mix. The captured fragments were subsequently used for library construction, cloning the isolated fragments into TopoTA cloning vector (Invitrogen). From the generated library, 176 clones were randomly selected and sequenced from both

ends. Only 27 fragments eventually showed microsatellite sequences of considerable length. The other 149 sequence had only small repeat units.

For 25 sequences primer design was possible. These primers were tested for variability on a restricted test set of 8 *Rhododendron* individuals from the whole geographical sampling range, containing *R. hirsutum*, *R. x intermedium* and *R. ferrugineum* respectively. A standard PCR amplification was performed at the appropriate locus specific T_m (see tab.1) in 25 µl reaction volume using 30–50 ng genomic DNA, 15 pmol of the respective primers (one always FAM labelled), 150 µM dNTPs, and 1U *Hotstar* Taq polymerase (Qiagen). SSR loci banding pattern was assessed using an automated sequencer (AB 3100) and Genotyper software (Applied Biosystems).

Out of the 25 tested primer pairs, 17 loci revealed variable fragments in the chosen test set, but only 11 of these regions were producing a nice clear banding pattern and could be used for further analysis (see tab1). These 11 loci together revealed 69 different alleles, with 3–9 variants per locus respectively in the 8 initially tested individuals.

Locus	Repeat motif	Na	N (ind)	size range bp	AccNr	Tm	Fwd (5'-3')	Rev (5'- 3')
6AG013	(AG)11	6	8	130-182	FJ823230	49	GACAAAATAGCTCCTCTAA	GTTCCTATTAATGGGGATT
6AG023	(GT)8(CT)8	9	8	105-137	FJ823252	56	TATGCATACATACAGACA	TCATTAGCATAACTGC
6AG180	(GA)15	3	8	141-152	FJ823235	65	CTGTCAGCTTTTGGGTTAC	CAGGTGGGTTCTCACA
6AG185	(GA)8	6	8	207-227	FJ823236	65	GCACAAGGGAGGTATCAAG	GTGGCTCCTGAGACATTTCT
6AG190	(GA)17(GT)12	5	8	105-117	FJ823238	50	GGACCGATGATTACTTCT	GGACCGATGATTACTTCT
6AG201	(AG)14	7	8	129-169	FJ823239	50	ACTTAAAGAATGCCTCTC	CCCGTTTAGTATGGATG
6AG218	GT)11(GA)19	5	8	109-121	FJ823243	50	TTCTTCTGATTTTCGTTGC	TTTCTTAGGACCGATGATT
6AG219	(AG)22	6	8	92-121	FJ823244	49	GATCACATGCTTATATGT	AGCTGCCATTGTGA
6AG230	(A)7(CTTGTT	7	8	290-312	FJ823249	56	GATTGGAAGTGACAAGTC	AGGGTGGTTCGAATG
6AG238	(AG)21	6	8	64-90	FJ823251	59	CCTGTGCTACAGTTGG	GCAGGTTACCAAGTAAAC
6CA139	(TG)8(GA)20	9	8	61-137	FJ823254	56	TATCTACAAGCCCAATG	AGCACTCGAGTATGTTG

Table 1: Repeat motifs, number of alleles (Na), number of individuals investigated, size range of alleles detected in bp, NCBI accession number, annealing temperature (T_m) as well as forward (Fwd) and reverse (Rev) primer sequences for 11 variable SSR loci characterized in *R. ferrugineum* and *R. hirsutum*.

In order to test four of these loci for their applicability for population genetic analysis, we investigated 227 individuals from 11 populations from Austria (8), Slovenia (1), Italy (1) and Switzerland (1), at least 20 individuals per population. Six of these populations consisted of *R. hirsutum*, three populations were made up by *R. ferrugineum*, and in 2 populations *R. ferrugineum*, *R. hirsutum* as well as *R. x intermedium* were present.

Investigating the 4 selected loci (see tab. 2) we could identify 110 different alleles with 9 to 24 allelic variants per locus. This high degree of variability of the selected loci allowed population genetic analysis of the 11 populations sampled all over the distribution range of alpine *Rhododendron*. Using PopGene 1.31 (<http://www.ualberta.ca/~fyeh/>; Francis C. Yeh;) we analyzed the heterozygosity

state of the sampled populations. Measures show moderate decrease of heterozygosity and thus deviation from the Hardy Weinberg equilibrium in the populations. This can be attributed to the relatively high degree of selfing and the possibility of vegetative propagation reported for this species (Escaravage, N. *et al.*, 2001). Investigation of 50 seedlings derived from several seed capsules from one mother plant revealed a high degree of selfing (over 90%, data not shown) when analyzing genomic DNA of germinated plantlets with the 4 selected SSR regions.

Locus	bp	N ind	N _a	Het _o	Het _e
6AG013	122-188	216	18	0.6852	0.8660
6AG218	107-143	210	14	0.6394	0.8572
6AG219	74-130	221	24	0.5023	0.9004
6AG230	278-312	206	9	0.5534	0.7037

Table 2: Population genetic measurements of four SSR loci on 227 individuals of *R. hirsutum*, *R. ferrugineum* and *R. intermedium*,

giving the allelic size range in bp, where N_{ind} is the number of individuals investigated, N_a the number of alleles detected in the sample set. Observed heterozygosity (Het_o) and expected heterozygosity (Het_e) are listed for the 4 loci as well.

Using the new SSR markers, we were able to detect clonal material within the 20 individuals sampled per population. In some of the populations, we identified 2 to 9 individuals exhibiting the same genotype; the samples from Chur/CH having the most limited variation in genotypes (20 individuals and only 11 genotypes – data not shown). Differences in relatedness of individuals within populations also have been described for *R. meternichii* (Kameyama, Y. *et al.*, 2001). A cluster analysis using UPGMA to construct a dendrogram of relatedness between the populations showed that all *R. hirsutum* populations clustered together well separated from *R. ferrugineum* in a separate cluster. The populations containing *R. intermedium* as well as the two other species were co-located in either of the two species specific clusters showing the distinctness of the two sympatric occurring species in the alpine region (Milne, R. I. *et al.*, 2003).

Reference List

- Botstein,D., White,R.L., Skolnick,M., and Davis,R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32:314-331.
- Escaravage,N., Flubacker,E., Pornon,A., Doche,B., and Till-Bottraud,I., I (2001) Stamen dimorphism in *Rhododendron ferrugineum* (Ericaceae): development and function. *Am.J.Bot.* 88:68-75.
- Escaravage,N. and Wagner,J. (2004) Pollination effectiveness and pollen dispersal in a *Rhododendron ferrugineum* (Ericaceae) population. *Plant Biol.(Stuttg)* 6:606-615.
- Kameyama,Y., Isagi,Y., and Nakagoshi,N. (2001) Patterns and levels of gene flow in *Rhododendron meternichii* var. hondoense revealed by microsatellite analysis. *Mol.Ecol.* 10:205-216.
- Milne,R.I. and Abbott,R.J. (2000) Origin and evolution of invasive naturalized material of *Rhododendron ponticum* L. in the British isles. *Mol.Ecol.* 9:541-556.

- Milne, R.I. and Abbott, R.J. (2008) Reproductive isolation among two interfertile *Rhododendron* species: low frequency of post-F1 hybrid genotypes in alpine hybrid zones. *Mol.Ecol.* 17:1108-1121.
- Milne, R.I., Terzioglu, S., and Abbott, R.J. (2003) A hybrid zone dominated by fertile F1s: maintenance of species barriers in *Rhododendron*. *Mol.Ecol.* 12:2719-2729.
- Naito, K., Isagi, Y., and Nakagoshi, N. (1998) Isolation and characterization of microsatellites of *Rhododendron metternichii* Sieb. et Zucc. var. *hondoense* Nakai. *Molecular Ecology* 7:927-928.
- Pornon, A. and Escaravage, N. (1999) Genotypic structure in clonal *Rhododendron ferrugineum* L. (Ericaceae) populations: Origin and maintenance. *Plant Ecology* 141:145-150.
- Pornon, A., Escaravage, N., Thomas, P., and Taberlet, P. (2000) Dynamics of genotypic structure in clonal *Rhododendron ferrugineum* (Ericaceae) populations. *Mol.Ecol.* 9:1099-1111.
- Refseth, U.H., Fangan, B.M., and Jakobsen, K.S. (1997) Hybridization capture of microsatellites directly from genomic DNA. *Electrophoresis* 18:1519-1523.
- Wolf, P.G., Doche, B., Gielly, L., and Taberlet, P. (2004) Genetic structure of *Rhododendron ferrugineum* at a wide range of spatial scales. *J.Hered.* 95:301-308.

2.1.1.2 Genetic structure of relict populations of *Rhododendron ferrugineum* L. in the Northern Apennines and their relationships with the nearest populations from the Alps

Ilaria Bruni¹, Fabrizio De Mattia², Massimo Labra^{2*}, Fabrizio Grassi³, Silvia Fluch⁴, Maria Berenyi⁴, Carlo Ferrari⁵

¹Research Centre on Environmental Sciences, Vegetation Ecology Laboratory, University of Bologna, Via S. Alberto 163, I- 48100 Ravenna, Italy

²Department of Bioscience and Biotechnology, University of Milano-Bicocca, Pizza della Scienza 2, I-20126 Milan, Italy

³Botanical Garden, Department of Biology, University of Milan, Via Celoria 26, I-20133 Milan, Italy

⁴Austrian Research Centers GmbH - ARC, Department Health & Environment, Bioresources/PICME, 2444 Seibersdorf, Austria

⁵Department of Experimental Evolutionary Biology, University of Bologna, Via Irnerio 42, I-40126 Bologna, Italy

submitted to: Plant Biosystems in March, 2010

Abstract

Background and Aims

Rhododendron ferrugineum L. (Ericaceae) is a subalpine shrub found throughout the Pyrenees and Alps above the timberline. One of the southernmost limits of its geographic range is on the Northern Apennines. Local populations are very few and scattered over three mountain groups in the chain. In this work we examined the genetic structure of Apennine populations and their relationships with different Alpine populations in order to evaluate the genetic erosion risk in response to environmental changes.

Methods

The genetic structure and relationships among 12 *Rhododendron ferrugineum* populations were examined by using 7 microsatellite markers. Three of these populations represent the relict population of Apennines, 3 were collected from Maritime Alps, and 6 came from the Central-Eastern Alps.

Key Results

Genetic variations across six microsatellite markers revealed that three Apennine populations showed the lowest genetics diversity values (values ranged from 0.57 to 0.74) as well as heterozygosity deficiency in comparison to the Alpine populations. Genetic relationships among populations (Nei's genetic distance) showed two main clusters which separately contain populations from the Maritime Alps (AM1, AM2 and AM3) and populations from other Alps. The Apennine populations (AP1, AP2, AP3)

do not cluster together. Our data suggest a low gene flow between Apennine and Alpine populations, as well as a different genetic structure of the three Apennine populations.

Conclusions

The genetic structure of the three Apennine populations and the low gene flow between the Apennine and Alpine *R. ferrugineum* populations suggest a clear extinction risk for these small populations owing to the ongoing climate changes and subsequent decrease in suitable sites, and the need for a conservation strategy based on a germplasm bank.

Key words

Apennines, Maritime Alps, microsatellites, nuclear SSR, pseudo-rarity, *Rhododendron ferrugineum* L.,

Introduction

Rhododendron ferrugineum L. - the Alpenrose - is an acidophilous evergreen shrub which reproduces sexually and asexually by selfed and out-crossed seeds (Escaravage *et al.*, 1997; Escaravage and Wagner, 2004), as well as vegetatively through layering (Pornon and Escaravage, 1999; Pornon *et al.*, 2000). This species is found in the Pyrenees, Jura mountains, Alps and Carpathians with large populations at 1600–2200 m elevations. Conversely, in the Northern Apennines, it is rare, with small relict populations, often reduced to few individuals (Ferrarini, 1973; 1974; 1979). On this mountain chain, like many other species belonging to South-European orophytes (e.g. *Pinus uncinata* Miller, *Antennaria carpatica* (Wahlenb.) Bluff & Fingerh, *Artemisia umbelliformis* Lam.), Eurosiberian (e.g. *Pinus sylvestris* L., *Empetrum hermaphroditum* Hagerup), or arctic alpine species (e.g. *Lychnis alpina* L., *Salix herbacea* L.), *R. ferrugineum* can be found on the southernmost limits of its geographic distribution in Europe (Ferrarini, 1979; Pignatti, 1982; Conti *et al.*, 2005).

Its small local populations scattered over very few sites offer an example of “pseudo-rarity” (Rabinowitz, 1981) or “peripheral rarity” (Hengeveld and Haeck, 1982; Brown, 1984; Nathan *et al.*, 1996). Wood charcoals of *Rhododendron* dating back to about 2000 years BC were found at 1740 m a.s.l. at a site near Mt. Cusna, where the species no longer exists today (Rossi and Orrico, 1996). It is likely that from the late glacial age up to about 2000 years BC, *R. ferrugineum* was relatively abundant above the Northern Apennine timberline, despite the heavy human interference (Lowe *et al.*, 1994).

At present, populations of *R. ferrugineum* can be found only on three mountain groups in the Northern Apennines: Mt. Libro Aperto, Mt. Prado-Vecchio, Mt. Nuda-Cima Belfiore (Alessandrini and Bonafede, 1996), within open ericaceous dwarf shrublands on some northern and north-western facing slopes (Ferrari and Piccoli, 1997). The geological substratum consists of thick bedded turbiditic sandstones and marlstones dating back to the upper Oligocene-lower Miocene (Carton and Panizza, 1988). The soils have more or less podzolic features and are highly acidic (Downgiallo *et al.*, 1998). On such slopes the snow cover often attains favourable conditions for *R. ferrugineum*: a snow cover lasting about 200 days from November to early June (Turner, 1961) with a depth suitable for the protection of buds from frost (Körner, 2003). Strong winds, (above 4 ms⁻¹) to which the Northern Apennine slopes are particularly exposed, have a negative effect (up to 60%) on the transpiration and growth of *R. ferrugineum* (Tranquillini, 1979). This can be inferred from the morphological traits characterizing the three Apennine populations: small plant size, presence of few buds per plant, and close-to-the-ground plant profile.

In general, the geographically isolated populations offer significant case studies on the relative effects of reproductive isolation, genetic drift, and selection on the genetic structure of populations. For the *R. ferrugineum* populations, the analysis must take into account their remarkable clonal propagation (Körner, 2003). In the Northern Apennine populations, clonality may have been the only reproductive strategy for many centuries, leading to a decline in their genetic variability. The

genetic structure of *R. ferrugineum* Northern Apennine populations was analyzed to estimate the genetic erosion risk in response to environmental changes. At the same time the genetic relationship between these populations and the closest ones in the Maritime and Central-Eastern Alps was investigated. Such information may provide a better understanding of the genetic origin of the Northern Apennine populations and will be crucial both for evaluating their biological value and for planning the best conservation strategy.

Considering the low values of genetic variation detected in the *R. ferrugineum* populations from the Alps by RAPD (Random Amplified Polymorphic DNA) (Wolf *et al.*, 2004) and AFLP (Amplified Fragment Length Polymorphisms) studies (Pornon and Escaravage, 1999), our analysis focused on hyper-variable DNA regions such as Simple Sequence Repeats (SSRs). SSRs or microsatellites have become the genetic marker of choice for population genetic analysis (Gugerli *et al.*, 2007; Reisch *et al.*, 2007; González-Pérez *et al.*, 2009) and mapping (Lalli *et al.*, 2008, Mantovani *et al.*, 2008) in many plant species. They are PCR-based, highly reproducible, polymorphic, show codominant inheritance with multi-allelic nature, and are abundant in plant genomes (Powell *et al.*, 1996). A specific set of these markers developed by Fluch S. *et al.*, (unpubl. res) was used in our study.

Site	Geographical name	Mountain chain	Latitude (N)	Longitude (E)	Elevation (m)	Nr. of samples
AP1	Mt. Libro Aperto	Northern Apennines	44°11'	10°43'	2100	21
AP2	Mt. Prado – Vecchio	Northern Apennines	44°13'	10°27'	1906	10
AP3	Mt. Nuda – Belfiore	Northern Apennines	44°17'	10°14'	1793	19
AM1	Mt. Saccarello	Maritime Alps	44°03'	7°43'	2100	9
AM2	Mt. Grai – Val Nervia	Maritime Alps	43°59'	7°40'	1800	11
AM3	Colle delle Finestre	Maritime Alps	44°07'	7°21'	2125	12
AL1	Col Merende – Casera Razzo	Alps	46°28'	12°36'	1805	16
AL2	Campitello di Fassa	Alps	46°28'	11°44'	1860	14
AL3	Passo Rolle	Alps	46°10'	11°50'	1950	10
AL4	Mt. Padrio	Alps	46°10'	10°13'	1858	20
AL5	Gromo	Alps	45°57'	09°55'	1730	20
AL6	Paglio	Alps	46° 03'	09°21'	1795	19

Table 1. Geographical information on the sampled populations. The site symbols and the no. of samples are also indicated.

Materials and methods

Plant material and DNA analysis

Twelve populations of *R. ferrugineum* and a total of 181 individuals were sampled. The list of selected sites, their names, geographical coordinates, elevations, and sample sizes are shown in Table 1.

In the Northern Apennines *R. ferrugineum* grows at three different sites (Mt. Libro Aperto, Mt. Prado - Vecchio, Mt. Nuda - Cima Belfiore). Only on Mt. Libro Aperto (AP1) one large population can be found on a north-western facing slope. On site AP2 (Mt. Prado - Vecchio) there is a small population on its north-facing slope, which is fragmented, and depressions in the slope ground. On site AP3 (Mt. Nuda - Cima Belfiore) Alpenrose individuals are mainly clustered in few small groups – each occupying less than 50 m² – in open *Vaccinium myrtillus* and *V. gaultherioides* dwarf shrublands, slightly above the timberline.

Three populations from the Maritime Alps (AM1, AM2 and AM3 – close to the French boarder) and six from the Central-Eastern Alps (AL1, AL2, AL3, AL4, AL5 and AL6 – along the Austrian and Slovenian boarder) were sampled for analysis.

The same plant sampling strategy was used for all populations. It was designed to avoid collection of individuals closer than 2 metres from one another, thus minimizing the collection of asexually (clonal) reproduced plants. A significant number of individuals for each population (more than 10) were sampled. Young leaf samples were collected. When possible, the sampled leaves were frozen or dried immediately after the sampling and stored for subsequent DNA extraction.

DNA extraction

Young leaves were frozen in liquid nitrogen and ground to a fine powder with a mortar and pestle. Total genomic DNA was extracted and purified using the DNeasy™ Plant Mini Kit (Qiagen, Italy) from approximately 20 mg of dry leaves or 100 mg of fresh leaves, according to the manufacturer's handbook. Quality and quantity of the isolated DNA were determined using agarose gel and DNA previously quantified using the Qubit fluorimeter (Invitrogen Life Technologies, Milan, Italy).

SSR analysis

A total of 10 SSR loci were tested on a subset of *R. ferrugineum* samples. Seven of these primer combinations showed reproducible amplification products as well as variability in the test samples. Details of each locus (RodA, RodC, RodD, RodE, RodG, RodH and RodI) are described in Table 2.

Nuclear DNA of 181 individuals corresponding to the 12 populations was analyzed using the 7 SSR loci. Amplification was performed with a PCR-bead Ready-to-go Kit (Amersham- Bioscience, Italy) in a volume of 10 µl starting from 10 ng of the total DNA and 10 pmol of forward and reverse primers.

Amplification reactions were multiplexed based on the annealing temperature (T_a) of the corresponding primer pairs: RodA was combined with Rod C at T_a 56 °C

while RodD was combined with RodE, RodG and RodH at T_a 49 °C. In the multiplex PCRs, the forward or reverse primer of each combination was labelled with 6-FAM, NED, VIC and PET fluorescent dyes (Applied Biosystems, UK) respectively (Table 2). Rod I having a T_a of 65 °C was analyzed in single amplifications.

The PCR thermal profile was: 3 min at 94 °C, 35 cycles of 50 sec at 94 °C, 50 sec at annealing temperature (T_a), 50 sec at 72 °C with a final extension step of 30 min at 72 °C.

PCR product (1 µL) was mixed with 12 µL of formamide and 0,2 µL of LIZ-500 size standard (Applied Biosystems, UK) and then analyzed on an ABI 3130 automated sequencer (Applied Biosystems, UK). The allele sizes for each SSR locus were identified using Genemapper 4.0 software (Applied Biosystems, UK).

SSR loci	Original name	AccNr	Forward Primer (5'-3')	Reverse Primer (5'-3')	Dye	Size (bp)	SSR loci	Na	Ho	He	Fis
Rod A	Rho_ARC_238	FJ823251	CCTGTGCTACAGTTGG	GCAGGTTACCACTAAAC	6-FAM (F)	64-84	RodA	8	0.59	0.722	-0.174
Rod C	Rho_ARC_230	FJ823249	GATTGGAAGTCAAGTC	CCCGTTTAGTATGGATG	NED (R)	291-310	RodC	7	0.26	0.727	0.248
Rod D	Rho_ARC_201	FJ823239	ACTTAAAAGAATGCCTCTC	CCCGTTTAGTATGGATG	6-FAM (R)	127-181	RodD	20	0.29	0.838	0.462
Rod E	Rho_ARC_218	FJ823243	TTCTTCTGATTGTC	TTTCTTAGGACCGATGATT	NED (R)	82-145	RodE	14	0.19	0.819	0.580
Rod G	Rho_ARC_190	FJ823238	GGACCGATGATTCTCT	GGACCGATGATTCTCT	VIC (F)	102-140	RodG	11	0.29	0.610	0.264
Rod H	Rho_ARC_013	FJ823230	GACAAAATAGCTCCTCTAA	GTTCTATTAATGGGGATT	PET (F)	165-188	RodH	15	0.46	0.667	0.182
Rod I	Rho_ARC_185	FJ823236	GCACAAGGGAGGTATCAAG	GTGGCTCCTGAGACATTC	VIC (R)	207-234	RodI	11	0.41	0.822	0.249

Table 2 SSR primer pairs used for DNA amplification. The values of the genetic diversity parameters are shown. Na, average number of alleles; Ho, observed heterozygosity; He, expected heterozygosity; Fis, coefficient of inbreeding. Details in the text.

Data analysis

SSR profiles were defined for each analyzed accession. The amplified SSR product representing different alleles was scored as different genotypes because of the co-dominant nature of the SSR markers.

The genetic diversity was measured by both Nei's index (Nei, 1973) and Shannon's diversity index (Shannon and Weaver, 1949). In order to assess the genetic relationships among the members of the three Apennine populations as individual plants, a Principle Component Analysis (PCA) was performed, based on the SSR patterns using GenAlEx6 (Peakall and Smouse, 2006).

A cluster analysis of populations was made by measuring the proportion of shared alleles among the individuals, based on Nei's genetic distance (Nei *et al.*, 1983). For measuring the genetic structure of the populations from which samples had been taken, the total amount of genetic variation detected was partitioned by AMOVA (Excoffier *et al.*, 1992).

Genetic parameters such as the observed average allele number (N_a), the percentage of polymorphic loci (P%), observed heterozygosity (H_o), expected heterozygosity (H_e), Shannon's diversity index (I), and the number of private alleles (PA) were calculated in order to estimate the level of genetic diversity of each locus and population. F-statistics were calculated for polymorphic loci both to test for the deviation from Hardy-Weinberg equilibrium and to estimate the genetic differentiation among populations (Wright, 1978). All of the calculations were performed using POPGENE version 3.2 (Yeh *et al.*, 1999), and MSA (Dieringer and Schlötterer, 2002). The divergence among populations was also estimated by Nei's (1972) genetic distances calculated for all reciprocal pairs of populations. A dendrogram based on these distances was constructed using the Unweighted Pair Group Method with Arithmetic Averages (UPGMA).

A hierarchical analysis of molecular variance (AMOVA) of *R. ferrugineum* populations was calculated in order to partition genetic diversity within and among regions and populations by GenAlEx6 (Peakall and Smouse, 2006). The gene flow (number of migrants per generation) among regions was estimated indirectly from F_{st} as $N_m = 0.25(1/F_{st}-1)$ (Wright, 1951).

Results

Genetic variability in the sampled populations of R. ferrugineum.

All selected SSR loci were polymorphic (Table 2), revealing a total of 86 different alleles in 181 individuals from 12 different populations, with an average of 12.3 alleles for each locus. A total of 23 private alleles were detected (about 12.5%). Table 2 summarizes the results at each SSR locus showing the range of allele size (bp) and the number of alleles (N_a). All the loci studied are highly polymorphic: the number of detected alleles per locus across all the populations ranges from 7 (locus RodC) to 20 (locus RodD) (Table 2). A high genetic diversity, as measured by Nei's index of expected heterozygosity (H_e), was found (Table 2). The average H_e shows a more restricted range, with values ranging from 0.610 (locus RodG) to 0.838 (locus RodD). These results prove the usefulness of the seven selected SSR markers for the study of the genetic variability of *R. ferrugineum* natural populations.

The Hardy-Weinberg equilibrium status was investigated for all the loci by testing the deviation of F_{is} from zero under the null hypothesis. Deviations from Hardy-Weinberg equilibrium were associated with F_{is} positive values in all the tested SSR loci with the exception of RodA which showed negative F_{is} value.

The mean Shannon's diversity index value was 0.94 (Table 3). The three Apennine populations (AP1, AP2 and AP3) showed the lowest values in genetics

diversity (*I* values ranged from 0.57 to 0.74), while the populations from the Maritime Alps showed a very high Shannon's diversity index (Table 3). The mean value of observed heterozygosity (*H_o*) in the analyzed samples was 0.42. Heterozygosity deficiency was detected in all the analyzed populations; however AP1 and AP3 showed the lowest *H_o* values (0.18 and 0.13, respectively). In accordance with the detected heterozygosity deficiency, significant high values of *F_{is}* coefficients were found in AP1 (0.47) and AP3 (0.70) populations.

Site	Na	P%	I	Ho	He	Fis	PA
AP1	2.34	71.43	0.57 (0.47)	0.18 (0.23)	0.34(0.27)	0.47	1
AP2	2.50	100.00	0.59 (0.19)	0.33 (0.21)	0.35(0.14)	0.06	2
AP3	3.20	85.71	0.74 (0.43)	0.13 (0.14)	0.42(0.23)	0.70	2
AM1	5.00	85.71	1.22 (0.69)	0.44 (0.29)	0.62(0.32)	0.30	4
AM2	5.33	100.00	1.35 (0.48)	0.34 (0.19)	0.68(0.18)	0.51	5
AM3	4.78	100.00	1.29 (0.41)	0.40 (0.12)	0.67(0.15)	0.41	3
AL1	3.31	100.00	0.92 (0.29)	0.48 (0.22)	0.53(0.16)	0.09	1
AL2	3.94	100.00	1.02 (0.43)	0.45 (0.21)	0.54(0.23)	0.17	1
AL3	3.21	85.71	0.86 (0.45)	0.47 (0.28)	0.50(0.26)	0.07	0
AL4	3.40	100.00	0.92 (0.31)	0.38 (0.27)	0.51(0.18)	0.27	1
AL5	3.14	100.00	0.89 (0.27)	0.39 (0.31)	0.52(0.14)	0.25	0
AL6	3.21	100.00	0.88 (0.24)	0.38 (0.29)	0.51(0.16)	0.27	3

Table 3: Genetic diversity of the 12 populations of *Rhododendron ferrugineum* L. from Apennines (AP), Maritime Alps (AM) and Central-Eastern Alps (AL), based on polymorphisms of the 7 SSR loci. Standard deviations in brackets. Na, average number of alleles; P, percentage of polymorphic loci; I, Shannon diversity index; *H_o*, observed heterozygosity; *H_e*, expected heterozygosity; *F_{is}*, coefficient of inbreeding and PA, private allele. Site symbols as in Table 1.

Genetic relationships among populations

The clustering of the 12 populations based on Nei's genetic distance (Nei *et al.*, 1983) is shown in Figure 2. According to their geographical distribution, populations from the Central-East Alps (AL1, AL 2, AL 3, AL 4, AL5 and AL 6) and the three populations from the Maritime Alps (AM1, AM2 and AM3) cluster together in 2 separate groups. The three Apennine populations (AP1, AP2, AP3) do not cluster together. The AP1 population is clustering close to the populations from the Maritime Alps, while the other two Apennine samples show a clear genetic distance from the alpine populations.

The total amount of genetic variations was partitioned by AMOVA into components, which well agree with geographic regions as well as among populations within regions. Northern Apennines, Maritime Alps and Central-Eastern Alps contain populations which are clearly characterized by different genetic structures. Table 4 shows that the within-population variation accounts for 51 % of the total variance.

About 27 % ($P \leq 0.010$) of genetic variation is the result of differences among inner populations versus 22 % among populations from different regions ($P \leq 0.010$).

Source of variation	df	SS	MS	% Tot	Est. Var.	<i>P</i>
Among Regions ^a	2	288.920	144.460	22 %	1.935	0.010
Among Populations	9	366.772	40.752	27 %	2.417	0.010
Within Pops	169	765.976	4.532	51 %	4.532	0.010

d.f., degrees of freedom; SS, sum of squares; MS, mean squared deviations; *P*, the significance of the variance components after 999 random permutations. ^aRegions: Northern Apennines, Maritime Alps and Central Eastern Alps.

Table 4: Analysis of molecular variance (AMOVA) based on the 7 SSR loci of 181 *Rhododendron ferrugineum* individuals collected from three different geographical regions: Northern Apennines, Maritime Alps and Central Eastern Alps. A high proportion of variations is “within population”. All values are considered highly significant ($P \leq 0.010$).

Both the estimated *N_m* values among the three geographical regions (Table 5) and the presence of a high number of private alleles (Table 3) indicate reduced gene flow among the analyzed geographic regions. The only exceptions are the largest Northern Apennine population (AP1) and the populations from the Maritime Alps. The AP1 population shows values of *N_m* of 1.286; 1.735, 1.304 for the AM1, AM2 and AM3, respectively (data not shown).

MC 1	MC 2	Sample - MC1	Sample - MC2	Fst	Nm
AP	AM	50	32	0.054	4.386
AP	AL	50	99	0.131	1.657
AM	AL	32	99	0.173	1.197
AP1	AP2	21	10	0.449	0.307
AP1	AP3	21	19	0.419	0.347
AP2	AP3	19	10	0.318	0.537

Table 5. Number of individual migrants per mountain chain and per Apennine population (*N_m*) based on *F_{st}*. Codes for the populations as in Tables 1. MC: Mountain chain; AP: Northern Apennine, AM: Maritime Alps, AL: Central Eastern Alps.

Genetic structure in the Northern Apennine populations

Results of PCA ordination are shown in Fig.3. The first two main components (Axis 1 and Axis 2) in PCA explain 49 % and 25 % of the variation, respectively. There is an apparent separation of the three Apennine populations with a clear genetic structure at each site. These results are in agreement with the high *F_{st}* values registered among these populations (Table 5). They indicate a high level of differentiation and a low gene flow (*N_m* = 0.347; 0.307; 0.537).

The PCA results suggest several cases of clonal propagation events in the Apennine populations. As for AP1 samples, the individuals *o* and *t*, *i* and *l*, *p* and *v*, *b*,

q and *u* showed the same SSR profiles (data not shown). The same situation was detected for the individuals *a* and *c* in the AP2 population and *m* and *n* in the AP3 population.

Discussion

The protection of viable populations is a major goal of conservation biology, and the loss of genetic diversity is thought to potentially lead to a decrease in the ability of a species to survive environmental changes and demographic fluctuations (Ellstrand and Elam, 1993; Milligan *et al.*, 1994). Our results suggest that in the analyzed Apennine *R. ferrugineum* populations, the clonal propagation and inbreeding events induced a decrease in genetic variability and, consequently, a genetic drift. As described by Escaravage *et al.*, (1998) and Pornon *et al.*, (2000) this species tends to adopt a clonal phalanx growth. The balance between sexual and asexual reproduction is not easy to elucidate (Escaravage *et al.*, 1998); our analyses, however, suggest that the Alpine populations showed consistent intrapopulation genetic variability and heterozygosity. One previous work (Wolf *et al.*, 2004) showed that *R. ferrugineum* usually adopts clonal propagation within a small spatial scale but the genetic diversity in a population is maintained by sexual reproduction and seed diffusion.

In small-sized populations, characterized by individuals with few buds and flowers, outcrossing is clearly reduced, whereas both self-pollination and clonal phalanx growth strategy increase. In the Northern Apennine populations, our molecular results confirm the prevailing self-pollination and clonal propagation, with very low heterozygosity values in all three populations studied, which register scattered growth patterns. In addition, both the isolation of the three populations from one another and the short flowering season (from mid June to early July) prevent gene flow among different sites (Kudo, 1993). As this situation characterizes all the three Apennine sites, we suggest that the vulnerability of these Alpenrose populations is not related to the size of their geographical range, but to the availability and decline of their habitats, and therefore to the genetic effects of the pseudo-rarity condition.

The potential ecological impact of the ongoing climate change has been much discussed in recent years (Theurillat and Guisan, 2001). Mountain ecosystems in high elevations were identified early on as potentially very sensitive areas. Many species may suffer a severe reduction in the area of their preferred habitats, which may in turn affect biodiversity patterns (e.g. Ozenda and Borel, 1995; Theurillat and Guisan, 2001). Such problems seem to be crucial for the species conservation in their peripheral areas, where species rarity corresponds to rarity of available habitats. This is not the case of *R. ferrugineum* on the Alps, where habitats characterized by a long-lasting and thick snow cover are still widely available. Several studies on Alpine populations confirm our results and suggest that different Alpine populations attain high genetic diversity through different mechanisms: microsite heterogeneity, which promotes the coexistence of clones through diversifying selection (Ellstrand and

Roose, 1987), occasional seedling recruitment (Soane and Watkinson, 1979), and symmetric competition among clones (Hartnett and Bazzaz, 1985).

In the near future, it is likely that climatic changes will induce a further decline of the *R. ferrugineum* habitats on the Northern Apennines, with an increasing vulnerability of its populations. A less lasting snow cover will support denser *Vaccinium gaultherioides* and *Vaccinium myrtillus* populations, significantly reducing favourable sites for *R. ferrugineum* seed germination (Bianco and Boulard, 1974; Ferrari and Piccoli, 1997).

The question of how important such populations are for the gene diversity conservation of species is linked to the genetic relationship among these peripheral populations and the ones which correspond to the species distribution core. Such an analysis was performed for *Pinus sylvestris* only (Labra *et al.*, 2006). The relationship between the genetic structure of the Northern Apennine populations and Alpine ones showed that these populations had a common origin but, mainly as a result of geographic isolation and small-size population, Northern Apennine populations have a lower genetic variability than the Alpine ones. The genetic diversity observed in the Northern Apennine populations and their comparison with the ones from the Maritime and Central-Eastern Alps confirm that a population's genetic diversity is mainly correlated with population size (Reisch *et al.*, 2003). To this, it is necessary to add the genetic effects of clonality. Putatively, the Apennine populations of *R. ferrugineum* derive from a Pleistocene expansion of this species from the Maritime Alps: our results provide some suggestions on this matter. It is very likely that the genotypic structure in Apennine populations was established by seedling recruitments only at an early colonizing stage (Pornon and Escaravage, 1999), while present populations are relicts of wider clones.

Our results show that only the largest Apennine population (Mt. Libro Aperto; AP1) has a genetic diversity comparable to the diversity of large-sized populations from the Maritime Alps. The other two Apennine populations, which are very small, are clearly genetically distant from the populations in the Alps, even with a different genetic structure at each site: a fact which can be explained by the reduction in gene flow and genetic drift.

For conservation purposes, the main conclusion that can be drawn from our results is that Northern Apennine populations show rare genotypes. As they are small and isolated from one another, they are highly vulnerable. The first aim of a conservation strategy should be a germplasm bank for the Apennine populations. An *ex situ* conservation of the Northern Apennine populations may be envisaged, using a strategy similar to that of the well-known Millennium Seed Bank Project (MSBP), an international conservation project coordinated by the Royal Botanic Gardens, Kew (see e.g. Linington, 2000; Guerrant *et al.*, 2004). Such a project might usefully apply not only to the Apennine Alpenrose but also to the Northern Apennine species characterized by peripheral rarity.

Figure 2 - Classification dendrogram (UPGMA) based on polymorphisms of 7 SSR loci in the 12 *R. ferrugineum* populations, using Nei's unbiased genetic distance coefficients. Population codes as in Table 1 and Table 3.

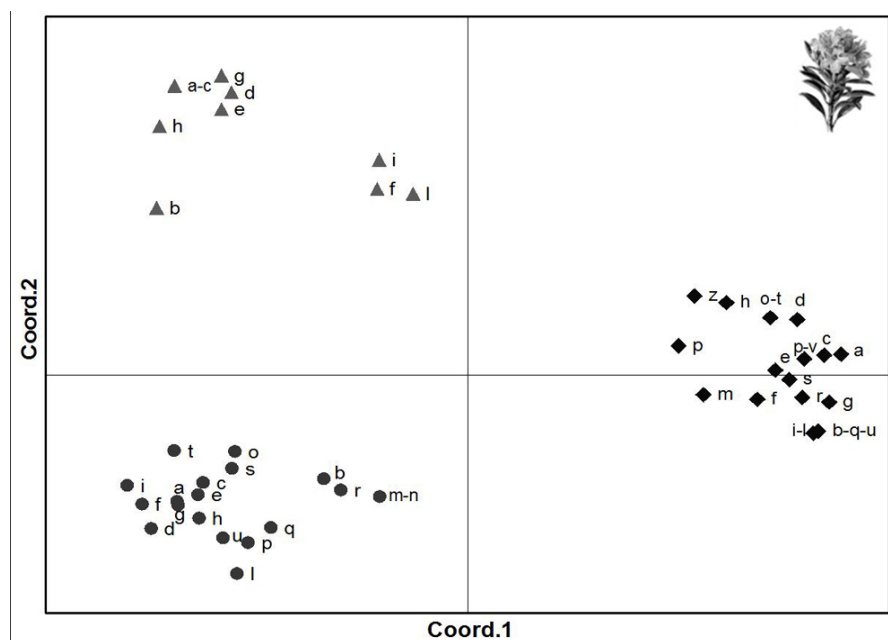


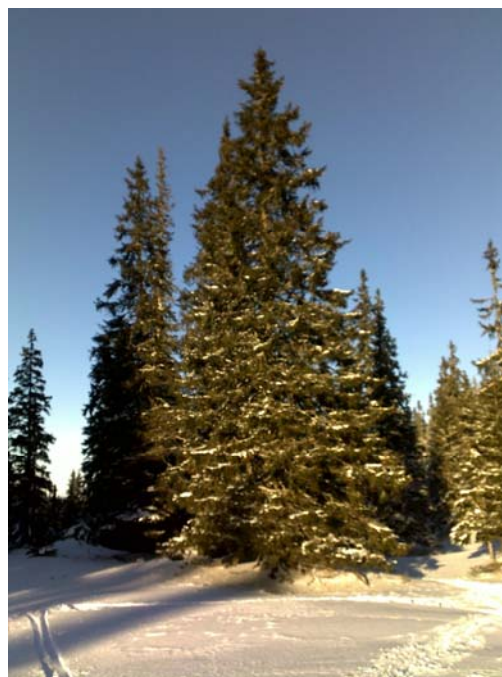
Figure 3 - Scattergram of the three Northern Apennine populations of *R. ferrugineum*. The ordination method is PCA. Coord. 1= 48,64% Coord. 2= 24,68%. ♦AP1 (Mt. Libro Aperto) ▲AP2 (Mt. Prado – Vecchio) ●AP3 (Mt. Nuda – Cima Belfiore). Individuals are indicated by small letters. Clonal individuals are separated by hyphens.

Literature Cited

- Alessandrini A, Bonafede F. 1996. Atlante della Flora protetta della Regione Emilia-Romagna, Regione Emilia-Romagna: Bologna.
- Bianco J, Bulard C. 1974. Etude de la germination des graines de *Rhododendron ferrugineum* L. et de *Tofieldia calyculata* L. Wahnlnb. Trav. Sci. Parc Nat. Vanoise 5: 121-130.
- Brown JH. 1984. On the relationship between abundance and distribution of species. *American Naturalist* 124: 255-279.
- Carton A., Panizza M. 1988. Il paesaggio fisico dell'Alto Appennino emiliano. Grafis Edizioni, Bologna.
- Conti FD, Abbate G, Alessandrini A, Blasi C. 2005. An annotated checklist of the Italian Vascular Flora. Palombi: Roma.
- Dieringer D, Schlötterer C. 2002. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* 3: 167-169.
- Downgiallo G, Rossi G, Tomaselli M. 1998. Vegetation and soil diversity in the areas above timberline of the Tuscan-Emilian Apennines (N. Italy). *Écologie* 29: 159-162.
- Ellstrand NC, Elam DR. 1993. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology & Systematics* 24: 217-242.
- Ellstrand NC, Roose ML. 1987. Patterns of genotypic diversity in clonal plant species. *American Journal of Botany* 74: 123-131.
- Escaravage N, Pornon A, Doche B, Till-Bottraud I. 1997. Breeding system in an alpine species: *Rhododendron ferrugineum* L. (Ericaceae) in the French northern Alps. *Canadian Journal of Botany* 75: 736-743.
- Escaravage N, Questiau S, Pornon A, Doche B, Taberlet P. 1998. Clonal diversity in a *Rhododendron ferrugineum* L. (Ericaceae) population inferred from AFLP markers. *Molecular ecology* 7: 975-982.

- Escaravage N, Wagner J. 2004. Pollination effectiveness and pollen dispersal in a *Rhododendron ferrugineum* (Ericaceae) population. *Plant Biol.* 6: 606-615.
- Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.
- Ferrari C, Piccoli F. 1997. The ericaceous dwarf shrublands above the Northern Apennine timberline (Italy). *Phytocoenologia* 27: 53-76.
- Ferrarini E. 1973. *Rhododendron ferrugineum* in fitocenosi relitte dell'Appennino settentrionale. *Giorn. Bot. Ital.* 107: 143-156.
- Ferrarini E. 1974. Altre cenosi e stazioni relitte reperite sull'Appennino settentrionale (ad Antennaria carpatica, a *Rhododendron ferrugineum*, a *Salix herbacea*, a *Saussurea discolor*). *Webbia* 29: 411-417.
- Ferrarini E. 1979. Note floristiche sull'Appennino settentrionale (dal Passo della Cisa al Passo delle Radici). *Webbia* 33: 235-267.
- Fluch S, Berenyi M, Hristoforoglu K, Kopecky D, Burg K. (subm.) Development of a microsatellite marker system for the Alpine species *Rhododendron ferrugineum* and *Rhododendron hirsutum*.
- González-Pérez MA, Sosa PA, Rivero E, González-González EA, Naranjo A. 2009. Molecular markers reveal no genetic differentiation between *Myrica rivas-mortinazii* and *M. faya* (Myricaceae). *Annals of Botany* 103: 79-86.
- Guerrant EO, Havens K, Maunder M. 2004. *Ex Situ Plant Conservation: Supporting Species Survival in the Wild*. Island Press, Covelo, U.S.A.
- Gugerli F, Walser JC, Dounavi K, Holderegger R, Finkeldey R. 2007. Coincidence of small-scale spatial discontinuities in leaf morphology and nuclear microsatellite variation of *Quercus petraea* and *Q. robur* in a mixed forest. *Annals of Botany* 99: 713-722.
- Hartnett DC, Bazzaz FA. 1985. The regulation of leaf, ramet and genet densities in experimental populations of the rhizomatous perennial *Solidago canadensis*. *Journal of Ecology* 73: 429-443.
- Hengeveld R, Haack J. 1982. The distribution of abundance. I. Measurements. *Journal of Biogeography* 9: 303-316.
- Körner C. 2003. *Alpine plant life: functional plant ecology of high mountains ecosystems*. 2nd edn. Springer: Berlin-Heidelberg.
- Kudo G. 1993. Relationship between flowering time and fruit-set of the entomophilous alpine shrub, *Rhododendron-aureum* (Ericaceae), inhabiting snow patches. *American Journal of Botany* 80: 1300-1304.
- Labra M, Grassi F, Sgorbati S, Ferrari C. 2006. Distribution of genetic variability in southern populations of Scots pine (*Pinus sylvestris* L.) from the Alps to the Apennines. *Flora* 201: 468-476.
- Lalli DA, Abbott AG, Zhebentyayeva TN, Badenes ML, Damsteegt V, Polák J, Krška B, Salava J. 2008. A genetic linkage map for an apricot (*Prunus armeniaca* L.) BC1 population mapping plum pox virus resistance. *Tree Genetics & Genomes* 4: 481-493.
- Linington S. 2000. The Millenium Seed Bank Project. In B.S. Rushton, P. Hackney and Tyrie CR, eds. *Biological Collections and Biodiversity*. Linnean Society of London Special Publication 3.
- Lowe JJ, Davite C, Moreno D, Maggi R. 1994. Holocene pollen stratigraphy and human interference in the woodlands of the northern Apennines, Italy. *The Holocene* 4: 153-164.
- Mantovani P, Maccaferri M, Sanguineti MC, Tuberosa R, Catizone I, Wenzl P, Thomson B, Carling J, Huttner E, DeAmbrogio E, Kilian A. 2008. An integrated DArT-SSR linkage map of durum wheat. *Molecular Breeding* 22: 629-648.
- Milligan BG, Leebensmack J, Strand AE. 1994. Conservation genetics – beyond the maintenance of marker diversity. *Molecular ecology* 3: 423-435.
- Nathan R, Shmida A, Fragman O. 1996. Peripherality and regional rarity are positively correlated: quantitative evidence from the upper Galilea Flora (North Israel). In Y. Steinberger editor. *Preservation of Our World in the Wake of Change*. Jerusalem, Israel, Vol. VI A/B ISEEQS Pub: 561-564.
- Nei M. 1972. Genetic distance between populations. *American Naturalist* 106: 283-392.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* 70: 3321-3323.
- Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol.* 19: 153-170.

- Ozenda P, Borel JL. 1995. Possible response of mountain vegetation to a global climatic change: the case of the Western Alps. In Guisan A, Holten JI, Spichiger R, Tessier L. Eds. Potential Ecological Impacts of Climate Change in the Alps and Fennoscandian Mountains Ed. Conserv. Jard. Bot. Genève: 137–144.
- Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.
- Pignatti S. 1982. Flora d'Italia. Edagricole, Bologna.
- Pornon A, Escaravage N. 1999. Genotypic structure in clonal *Rhododendron ferrugineum* L. (Ericaceae) populations: origin and maintenance. Plant Ecology 141: 145-150.
- Pornon A, Escaravage N, Thomas P, Taberlet P. 2000. Dynamics of genotypic structure in clonal *Rhododendron ferrugineum* (Ericaceae) populations. Molecular Ecology 9: 1099–1111.
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingley S, Rafalski A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed 2: 225–238.
- Rabinowitz D. 1981. Seven forms of rarity. In H. Synge (ed.), The Biological Aspects of Rare Plant Conservation : 205-217. John Wiley & Sons Ltd., New York.
- Reisch C, Poschlo P, Wingender R. 2003. Genetic variation of *Saxifraga paniculata* Mill. (Saxifragaceae): molecular evidence for glacial relict endemism in central Europe. Biological journal of the Linnean society 80: 11-21.
- Reisch C, Schurm S, Poschlo P. 2007. Spatial genetic structure and clonal diversity in an Alpine population of *Salix herbacea* (Salicaceae). Annals of Botany 99: 647-651.
- Rossi G, Orrico D. 1996. Peto-anthracological studies on the evolution of the upper treeline in the Tuscan-Emilian Apennines: preliminary data. Arch.Geobot. 2: 173-174.
- Shannon CE, Weaver W. 1949. The mathematical theory of communication. Univ. of Illinois Press, Urbana.
- Soane ID, Watkinson AR. 1979. Clonal variation in populations of *Ranunculus repens*. New Phytologist 82: 557–573.
- Theurillat JP, Guisan A. 2001. Potential impact of climate change on vegetation in the European Alps: a review. Climatic Change 50: 77-109.
- Tranquillini W. 1979. Physiological Ecology of the Alpine Timberline. Tree Existence at High Altitudes with Special Reference to the European Alps. (Ecological Studies, 31). New York and Berlin: Springer Verlag.
- Turner H. 1961. Ökologische Untersuchungen in der subalpinen Stufe. 8. Die Niederschlags und Schneebeziehungen. Mittl. Forstl. Bundesversuchsanst. Mariabrunn 59: 265-315.
- Wolf PG, Doche B, Gielly L, Taberlet P. 2004. Genetic Structure of *Rhododendron ferrugineum* at a Wide Range of Spatial Scales. Journal of Heredity 95: 301-308.
- Wright S. 1951. The genetical structure of populations. Annals of Eugenics 15: 395-420.
- Wright S. 1978. Variability within and among natural populations. Chicago: University of Chicago Press.
- Yeh FC, Yang RC, Boyle T. 1999. POPGENE version 1.31: Microsoft Window-Based freeware for population genetic analysis. University of Alberta and Centre for International Forestry Research, Alberta, Canada



2.1.2 SSR Markers Development for Norway spruce

Introduction

Picea abies (Norway spruce) belongs to the family *Pinaceae* L. and order *Pinales*. Spruces are evergreen trees with horizontal branches and pendulous branchlets. They are 40–50 m tall with diameters up to 150 cm and conic crown shape. Early growth of Norway spruce is slow, increasing to maximal rates from 20 to 60 years of age. Within its native range Norway spruce remains healthy up to 200 years and lives up to 300 to 400 years at the northern limits of its range.

Natural populations of Norway spruce are found from northern and western Europe outside permafrost areas down south to northern Greece, westwards to the Massif Central (France) and east to the Ural Mountains. *Picea abies* is only growing in mountains above 400–500 m and ascends close to 2000 m in the Alps. In Austria, Norway spruce is widely distributed. The Alps, mountains of the Bohemian massif [Mühlviertel (Upper Austria) and Waldviertel (Lower Austria)], and higher elevations of the alpine foreland (Hausruck, Kobernauerwald) are naturally covered by Norway spruce. The eastern natural limits are defined by a connection line starting from the cities of St. Pölten over Wiener Neustadt to Graz and then south along the eastern slope of the Koralpe.

Studies on genetic variation based on allozymes (Bergmann & Ruetz 1991; Lagercrantz & Ryman 1990) have shown that, over the whole range of *Picea abies*, genetic differentiation among populations is low. Previous studies of the genetic structure of *P. abies* showed pronounced differentiation between a north-east boreal distribution and disjunct areas in the central European mountains (Acher *et al.*, 2005)

supporting the hypothesis of two distinct main glacial refugia postulated from pollen data (Tollefsrud *et al.*, 2008).

The limited sampling of Norway spruce from the eastern Alps in a previous study on mitochondrial tandem repeats (Gugerli *et al.*, 2001; Scotti *et al.*, 2008), however, did not allow more detailed inferences about the genetic structure of populations in the eastern Alps. In an earlier study on *P. abies* in Austria based on isozyme allele frequencies Geburek (1999) found an increase in genetic variation towards the east.

Informative, multi-allelic markers, such as SSRs, are, however, especially desirable (Kirst *et al.*, 2004) in conifers. Initial reports on the occurrence of microsatellites in Radiata pine (Smith & Devey 1994), Scots pine (Kostia *et al.*, 1995) or Norway spruce (Pfeiffer *et al.*, 1997) have shown that several DNA fragments in addition to the expected ones are frequently amplified when using primers flanking microsatellite regions. This can be attributed to the very large size of the conifer genomes, with Norway spruce having 18.300 Mb in the diploid genome of $2n=24$. Additionally the high proportion of repetitive DNA in the conifer genomes is thought to cause problems when isolating or amplifying SSR regions. These repetitive elements and pseudogenes frequently produce complex amplification products from multiple loci that are difficult to score (Echt *et al.*, 1996; Scotti *et al.*, 1998; Soranzo *et al.*, 1998). To overcome this problem, Scotti *et al.*, (2000) isolated SSRs from cDNA libraries of conifer species. They have investigated the usefulness of the molecular marker obtained for their ability to determine paternity in a Norway spruce breeding programme.

SSR markers so far have been used in population genetic analysis of Norway spruce (Geburek *et al.*, 1998b) as well as in testing of clonal variation of in vitro propagated individuals (Helmersson *et al.*, 2004).

2.1.2.1 Characterization of variable EST SSR markers for Norway spruce [*Picea abies* L.] Karst.]

Fluch S.¹, Burg A.¹, Vendramin G.², Kopecky D.¹

¹ AIT Austrian Institute of Technology GmbH, Health & Environment Department, Bioresources, Plant Genomics, 2444 Seibersdorf, Austria

² National Research Council, Plant Genetics Institute, Via Madonna del Piano 10, 50019 Sesto Fiorentino (Firenze), Italy

Keywords: *Picea abies*, EST-SSR, population genetics, Austria

Submitted to Molecular Ecology Notes

Abstract

We report on the characterization and use of 29 SSR markers for *Picea abies* L. (Norway spruce), an alpine forest tree species. These markers were obtained screening public as well as private EST collections for SSR motifs. All together, 15.178 sequences were analyzed to generate a non redundant fragment collection which subsequently was screened for the presence of microsatellite motifs. 92 SSR containing 'unigenes' were derived from 36 clusters and 56 singletons. In this unigene set, 48 different repeat motifs were identified. 29 of the 60 designed primers revealed variable fragments in the chosen test set of 96 individuals. 25 of the loci revealed two alleles per locus per individual, whereas 4 of the loci produced up to four bands. The 29 variable primer pairs revealed 3–16 alleles per locus, 116 alleles in the 112 individuals tested so far.

Due to their moderate to high degree of variability, these newly developed SSR markers are well suited for the analysis of genetic variation present in the Alpine Norway spruce populations.

Natural populations of *Picea abies* (Norway spruce) are found from northwestern Europe outside permafrost areas down to northern Greece, westwards to the Massif Central (France) and east to the Ural Mountains. *Picea abies* is only growing above 400–500 m and ascends close to 2000 m in the Alps. Studies on genetic variation based on allozymes have shown that *Picea abies* genetic differentiation among populations is low over its whole distribution range (Lagercrantz and Ryman 1990; Bergmann and Ruetz 1991). Previous studies on the genetic structure of *P. abies* showed pronounced differentiation between a north-east boreal origin as compared to areas in the central European mountains (Vendramin *et al.*, 2000; Sperisen *et al.*, 2001), supporting the hypothesis of two distinct main glacial refugia as postulated from pollen data (Huntley B. and Birks H.J.B., 1983).

Initial reports on the occurrence of microsatellites in conifers such as Radiata pine (Smith and Devey 1994), Scots pine (Kostia *et al.*, 1995) or Norway spruce (Pfeiffer *et al.*, 1997) have shown that marker development for such complex genomes is difficult. Frequently several DNA fragments in addition to the expected ones are amplified when using primers flanking microsatellite regions. This can be attributed to the very large size of the conifer genomes, with Norway spruce having 18.300 Mb in the diploid genome of $2n=24$ as well as the high proportion of repetitive DNA in the conifer genomes (Dhillon 1987). These repetitive elements as well as pseudogenes frequently produce complex amplification products from multiple loci (Echt *et al.*, 1996; Soranzo *et al.*, 1998). To overcome this problem, Scotti *et al.*, (2000) isolated 10 di-nucleotide SSRs from cDNA libraries of conifer species relying on the fact that expressed genes are less likely to be of repetitive nature in the genome.

Following this line of exploiting the advantages of expressed sequences as sources for marker development (Varshney *et al.*, 2002) as described also by Ritland *et al.*, (2007) for conifers, 15,178 Norway spruce EST sequences were used for in silico filtering of SSR regions which could serve for marker development. The screening using the software SciRoKo (Kofler *et al.*, 2007) revealed 158 sequences containing various repeat motifs. Clustering of these sequences produced 92 SSR containing 'unigenes' which fell into 36 clusters and 56 singletons. In these 92 unigenes, 48 different repeat motifs were identified, with tri-nucleotide repeats being the most abundant repeat type (50%) with (CGG)_n and (CAA)_n being the most abundant, followed by penta- (23%) and hexanucleotide repeats (10,4%). Di- and tetranucleotide repeats being the least abundant with 8,3% each. The high abundance of trinucleotide repeats can be explained by the fact, that sequence variation by 3 bp between alleles doesn't lead to a frame shift in the reading frame of the respective gene whereas di- or tetranucleotide variations lead to a distortion in reading frame and thus possibly to the abortion of the coded protein.

Out of the 92 EST fragments 59 were chosen randomly from each category for primer planning and PCR testing on a set of 16 Norway spruce individuals from all over Austria. Out of this primer set, 16 (27%) primer pairs gave no or bad amplification, 14 (23%) were generating a monomorphic banding pattern in the used test set of individuals, and 29 (49%) of the loci proved to be polymorphic. Six of these

variable loci generated 3–4 bands per PCR amplification, thus targeting two loci in the genome. Most of the variable loci were identified for trinucleotide repeats (65%) where the conversion rate from the unigene sequence to variable banding pattern was 68%. The tetra- (2) and pentanucleotide repeats (3) represented a conversion rate close to 90%, revealing 2 variable loci each. Dinucleotide repeats contributed only 8% to the total set of variable regions, also having a conversion rate of 100%. For the hexanucleotide repeats, no primers could be planned.

To further test the applicability of the newly generated EST-SSR markers, a test set of additional 96 individuals from 2 distinct Norway spruce populations (Gusswerk, Mayerhofen; 48 individuals each) were tested using Pa_28 variable SSR regions, 4 of them showing more than 2 alleles per locus (see table 1). The 24 'normal' loci revealed 2–16 alleles per locus and a total of 146 alleles in the 96 individuals from the 2 populations. This moderate to high degree of variability of the selected loci allowed population genetic analysis of the genetic diversity of the 2 Austrian Norway spruce populations using PopGene 1.31 (<http://www.ualberta.ca/~fyeh/>; Francis C. Yeh;). Population genetic analysis showed a somewhat enhanced heterozygosity in the investigated populations as compared to expected heterozygosity values (table 2), showing that populations are in Hardy Weinberg equilibrium. It also could be observed, as described previously for Norway spruce, that within-population diversity is higher than between-population diversity.

Table2 Analysis of the quality and information content of the 24 newly described SSR loci

Marker	N	k	i	H _o	H _e	PIC	f
NS_EST_SSR_28	95	8	15	0,621	0,745	0,712	0,172
NS_EST_SSR_47	96	3	3	0,521	0,414	0,370	-0,254
NS_EST_SSR_52	96	5	6	0,948	0,537	0,432	-0,762
NS_EST_SSR_51	96	4	4	0,365	0,305	0,270	-0,190
NS_EST_SSR_50	93	4	5	0,226	0,221	0,205	-0,015
NS_EST_SSR_49	96	2	2	0,990	0,500	0,375	-0,979
NS_EST_SSR_43	93	4	4	0,129	0,141	0,135	0,092
NS_EST_SSR_42	95	5	6	0,126	0,157	0,150	0,201
NS_EST_SSR_41	93	5	8	0,258	0,333	0,301	0,231
NS_EST_SSR_53	96	2	2	0,010	0,010	0,010	0,000
NS_EST_SSR_57	95	1	1	0,000	0,000	0,000	n. def.
NS_EST_SSR_59	96	3	4	0,094	0,110	0,106	0,149
NS_EST_SSR_29	95	5	7	0,568	0,442	0,369	-0,281
NS_EST_SSR_36	95	5	5	0,158	0,149	0,144	-0,057
NS_EST_SSR_55	96	3	3	0,927	0,502	0,381	-0,845
NS_EST_SSR_44	96	7	9	0,188	0,263	0,253	0,293
NS_EST_SSR_33	95	9	12	0,968	0,611	0,536	-0,581
NS_EST_SSR_60	96	5	5	0,594	0,444	0,377	-0,332
NS_EST_SSR_25	94	8	7	0,074	0,241	0,231	0,694
NS_EST_SSR_22	82	6	7	0,683	0,504	0,419	-0,350
NS_EST_SSR_12	84	2	2	0,976	0,500	0,375	-0,953
NS_EST_SSR_5	87	7	7	0,609	0,446	0,378	-0,361
NS_EST_SSR_48	94	17	27	0,713	0,757	0,722	0,064
NS_EST_SSR_56	96	6	11	0,740	0,715	0,670	-0,029

N	Number of samples
k	Number of alleles
i	Number of genotypes
H_o	Observed Heterozygosity
H_e	Expected Heterozygosity
PIC	Primer Information Content
f	Inbreeding coefficient

Table3 population genetic data for the 2 populations at Gusswerk (pop1) and Mayerhofen (pop2) as revealed by 24 SSR loci. Values generated using PopGene v 1.31

Locus		Size	Obs_ Hom	Obs_H et	Exp_ Hom*	Exp_ Het*	Nei**	Ave_H et
Pop1	Mean	93	0.5580	0.4420	0.6484	0.3516	0.3478	0.3717
	<i>St. Dev</i>		<i>0.3509</i>	<i>0.3509</i>	<i>0.2407</i>	<i>0.2407</i>	<i>0.2381</i>	<i>0.2192</i>
Pop2	Mean	92	0.4840	0.5160	0.6000	0.4000	0.3956	0.3717
	<i>St. Dev</i>		<i>0.3597</i>	<i>0.3597</i>	<i>0.2203</i>	<i>0.2203</i>	<i>0.2179</i>	<i>0.2192</i>
Total	Mean	186	0.5224	0.4776	0.6208	0.3792	0.3772	0.3717
	<i>St. Dev</i>		<i>0.3425</i>	<i>0.3425</i>	<i>0.2211</i>	<i>0.2211</i>	<i>0.2199</i>	<i>0.2192</i>

* Expected homozygosity and heterozygosity were computed using Levene (1949)

** Nei's (1973) expected heterozygosity

Comparing allelic frequency within the two investigated populations, for 7 loci out of the 24 biallelic loci, 'private alleles' could be observed, being present in one population, but not in the other (fig1). Blast results for the SSR containing sequences revealed functional annotations for 14 sequences. Annotations were produced with the EST2Uni pipeline (Forment et al 2008) using BLASTX against three databases, NCBI NR, Arabidopsis TAIR7 and Swissprot/Uniprot. Region 28, being a putative defensin gene, reveals private alleles in the population of Gusswerk. This allele, which is not present in the other investigated population might have adaptation potential by conferring microbial resistance. In general, variation in the coding region could go along with functional diversity in these genes, thus influencing the adaptation potential of the individuals carrying these allelic variants.

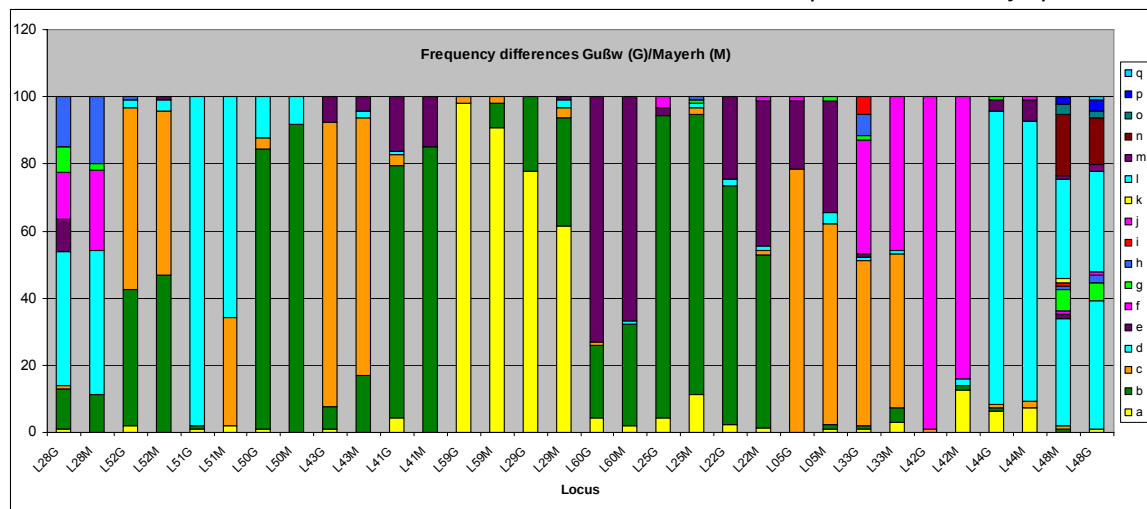


Figure 1: Allelic frequencies as revealed by SSR analysis investigating 96 Norway spruce individuals from the populations Gusswerk (G) and Mayerhofen (M) (48 each). Always 2 columns represent one locus, one for G, the other for M. colours represent the relative abundance/frequency of different alleles per locus.

References:

- Bergmann F. and Ruetz W. (1991) Isozyme genetic variation and heterozygosity in random tree samples and selected orchard clones from the same Norway spruce populations. *Forest Ecology and Management*, 46(1-2): 39-48.
- Dhillon S.S. (1987) DNA in tree species. In *Cell and Tissue Culture in Forestry*, Vol. 1. Eds. J.M. Bonga and D.J. Durzan. Nijhoff, Dordrecht, The Netherlands. 298-313.
- Echt C.S., May-Marquardt P., Hsieh M., Zahorchak R. (1996) Characterization of microsatellite markers in eastern white pine. *Genome*, 39: 1102-1108.
- Forment J., Gilabert F., Robles, A., Conejero V., Nuez F., Blanca J.M. (2008) EST2uni: an open, parallel tool for automated EST analysis and database creation, with a data mining web interface and microarray expression data integration. *BMC Bioinformatics*, 9:5.
- Huntley B. and Birks H.J.B. (1983) *An Atlas of Past and Present Pollen Maps for Europe: 0-13000 years ago*. Cambridge University Press, Cambridge.
- Kostia S., Varvio S.L., Vakkari P., Pulkkinen P. (1995) Microsatellite sequences in a conifer, *Pinus sylvestris*. *Genome*, 38: 1244-1248.
- Lagercrantz U. and Ryman N. (1990) Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution*, 44: 38-53.
- Pfeiffer A., Olivieri A.M., Morgante M. (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome*, Aug., 40(4): 411-9.
- Smith D.N. and Devey M.E. (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. *Genome*, 37: 977-983.
- Scotti I., Troggio M., Soranzo N., Vendramin G.G., Bucci G. (1998) A new set of PCR-based, locus-specific markers for *Picea abies* (L.) Karst. *Molecular Ecology*, 7: 789-792.
- Scotti, I. Magni, F. Fink, R. Powell, W. Binelli, G. Hedley, P.E. Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences (2000) *Genome* vol 43p.41-46
- Soranzo N., Provan J., Powell W. (1998) Characterisation of microsatellite loci in *Pinus sylvestris* L. *Molecular Ecology*, 7: 1260-1261.
- Sperisen C., Büchler U., Gugerli F., Mátyás G., Geburek T.h., Vendramin G.G. (2001) Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. *Molecular Ecology*, 10 (1): 257-263.

- Vendramin G.G., Anzidei M., Madaghiele A., Sperisen C., Bucci G. (2000) Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). *Genome*, Feb., 43(1): 68-78.
- Varshney,R.K. Thiel,T. Stein,N. Langridge,P Graner,A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species (2002) *Cellular & Molecular Biology Letters* vol 7 p:537- 546

Table 1: variable EST SSR primers for Norway spruce

locus name	NCBI RefSeq	forward	reverse	repeat	nr of repeats	number of alleles in GEM	size range	annotation	ref seq	e value
Pa 5	GT887224	AGGTGGATATGTTCAAC	CGCATACTCATCCCTAAC	(CCG) _n	7	8	136-158	no homology		
Pa 12	AM171475/AM170164	GACCGAGAACCCCTTGTT	CAAAAGCAGGAAAGAAGAAC	(AAG) _n	6	2	136-138	brassinosteroid signalling positive regulator-related	AT4G18890.1	5.0e-12
Pa 16	ES226936	GCCACCAGTCTACAGC	ACCCCTCTTTCTCTTACCAG	(AT) _n	12	7*	121-159	leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis thaliana] receptor protein kinase-like protein [Arabidopsis thaliana]	gi 15240263 ref NP_200956.1	3.0e-27
Pa 22	GT886979	TCACTGGCCACAGTTTATCG	ATGAGGCCCAAGAGGAAGAC	(TTC) _n	10	6	169-188	no homology		
Pa 24	GT887224	TCATGAACCAATTCGGAAGC	ATCGCGCATACTCATCCCTA	(CCG) _n	7	4*	137-149	no homology		
Pa 25	GT887533	TGATTGAAATGATGGCTGCT	CATGTACGGTGCTCCTCCTC	(CTG) _n	19	8	131-165	no homology		
Pa 28	GT887962	GGCCGAAAGTGCTACTGCTA	TGCTCCAGAAGAACAACCTACA	(TCG) _n	7	8	148-162	defensin	gi 134141945 gb ABO61348.1	1.0e-31
Pa 29	GT884785	ACAACAGCAACAGCAGCAAC	CGGGCTGAAGAATTTGTTGT	(CAA) _n	16	4	97-108	heterogeneous nuclear ribonucleoprotein A2/B1-like	gi 115438550 ref NP_001043567.1	6.0e-23
Pa 33	GT884592	GGTCGAGGAGGAGGAGGTAG	CACCGCTAGTGCAGTCTCTG	(CGG) _n	17	8	91-104	putative glycine-rich protein [Picea glauca]	gi 76577745 gb ABA54143.1	
Pa 34	GT885056	AGTGGACAAGGCCACATAGG	AATTGCTCGCCGTAGTTGTT	(CAA) _n	31	5*	123-145	no homology	gi 4704605 gb AAD28176.1 AF109917_1	4.0e-38
Pa 36	GT887107	CGGCAGGAACATCACTGTTA	ACCGTAACCTCCCTACCCAC	(CGG) _n	20	4	178-200	glycine-rich RNA-binding protein [Picea glauca]		
Pa 40	AM169587	GCTCCGATGGTGGTAACCTA	GGAGCCACATAAGCACCAGT	(TGG) _n	20	7*	160-183	no homology		
Pa 41	AM169657	CGAAGAGAAAGCGAGGATG	CAGGCTGCGAGAATCCTCTA	(CTG) _n	9	5	176-183	no homology		
Pa 42	AM169827	CAATGCAATGGCTCCTATC	TAGACAGCAGCAGCTCTCTG	(CCG) _n	15	5	160-163	cysteine protease Cp5 [Actinidia deliciosa]	gi 146216002 gb ABQ10203.1	9.0e-36
Pa 43	AM169948	AACCGCCGGAGTCTGTAAA	TGCTTCTGTCTGACCAGGTG	(GAA) _n	7	4	168-175	zinc finger (C2H2 type) protein (WIP2) [Arabidopsis thaliana] WIP2 protein [Arabidopsis thaliana] zinc finger-like protein [Arabidopsis thaliana] At3g57670 [Arabidopsis thaliana]	gi 15230370 ref NP_191326.1	3.0e-19
Pa 44	AM171132	AAGGCAGCCAAAGTGAAGAA	CTTGGCATTCCCTAGTGAGC	(GGA) _n	9	7	274-293	hypothetical protein	gi 55733882 gb AAV59389.1	3.0e-17
Pa 47	AM172144	ATCAATTGCCCTACCAGCAC	TGCTCAATTTCTCGCATCTG	(CAG) _n	19	3	106-118	HAP5B (Heme activator protein (yeast) homolog 5B); DNA binding / transcription factor [Arabidopsis thaliana] HAP5B (Heme activator protein (yeast) homolog 5B) [Arabidopsis thaliana] Nuclear transcription factor Y subunit C-2 (AtNF-YC-2) (Transcriptional activator HAP5B) transcription factor, putative [Arabidopsis thaliana] putative transcription factor [Arabidopsis thaliana]	gi 15223482 ref NP_176013.1	1.0e-34
Pa 48	AM172164	ATTGCACAAGAGCGCAACCTT	CCAGCACCAAAATCACCCAG	(CGG) _n	32	16	131-195	no homology	no homology	
Pa 49	AM173566	GAAAGCGCAAGCAAACTG	CCGTGTTTTTGAGGGAAGAA	(ATG) _n	14	2	117-128	no homology		
Pa 50	AM174647	CCTCTCAGCCAAATGCAAT	GTCTCCGAGAGCCATTACT	(CAG) _n	8	4	124-129	R2R3-MYB transcription factor MYB1 [Picea glauca]	gi 147744702 gb ABQ51217.1	5.0e-47
Pa 51	AM175215	CAGATGTGGGCACTTGTTTG	TGTCATGTTGGTGTTCAT	(CCA) _n	19	4	119-140	HMA4 (Heavy metal ATPase 4); cadmium-transporting ATPase [Arabidopsis thaliana] Putative cadmium/zinc-transporting ATPase 2 putative cadmium-transporting ATPase [Arabidopsis thaliana] putative heavy metal transporter [Arabidopsis thaliana] At2g19110/T20K24.12 [Arabidopsis thaliana] putative heavy metal transporter [Arabidopsis thaliana]	gi 15224717 ref NP_179501.1	5.0e-25
Pa 52	AM175251	GTAGTGTACCGGCTCCTTT	CTGGAGGATGATGGTGGTG	(CCA) _n	14	5	135-145	no homology		
Pa 53	ES290836	CCCAATGAAGATCTCGATGC	GGCTTTGCAGGAACAAGAA	(TCG) _n	6	2	176-178	hypothetical protein		
Pa 55	GT887161	TGAGCAACGAGAAATCAGC	TCTCAACCCCTTTGTTTCG	(GAA) _n	13	3	148-158	no homology		
Pa 56	GT884695	ATCGTCTGCATTGCATTAC	CTTCGTTCTCTCTGATCCA	(AGGTG) _n	6	6	113-128	FED A (FERREDOXIN 2); electron carrier/ iron ion binding [Arabidopsis thaliana] Ferredoxin-1, chloroplast precursor At1g60950 [Arabidopsis thaliana] putative ferredoxin precursor protein [Arabidopsis thaliana] ferredoxin precursor; 39650-40096 [Arabidopsis thaliana] putative ferredoxin precursor protein [Arabidopsis thaliana] ferredoxin precursor [Arabidopsis thaliana] ferredoxin A ferredoxin precursor [Arabidopsis thaliana] ferredoxin precursor [Arabidopsis thaliana] ferredoxin precursor	gi 15219837 ref NP_176291.1	1.0e-39
Pa 57	GT885232	TAGAGCAGGGAATGGGTTTG	AACAACAGAGCCATTACGG	(CAGGC) _n	7	1	117	no homology		
Pa 59	AM169794	TAGACACGTTGTGCTGAAA	GCGACCTTTGCAATTAAC	(CTCTG) _n	5	3	90-112	no homology		
Pa 60	GT885001	CGCCGTATCCATTCCCAAGC	CCCAGCCCAAGTTGAGTTTGC	(GGCTG) _n	8	5	241-264	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein [Arabidopsis thaliana] protease inhibitor/seed storage/lipid transfer protein (LTP) family protein [Arabidopsis thaliana] putative	gi 15232384 ref NP_190966.1	8.0e-27

* more than 2 alleles/locus

2.1.2.2 Genetic diversity in managed subpopulations of Norway spruce [*Picea abies* (L.) Karst.]

F. Maghuly^a, W. Pinsker^b, W. Praznik^c, S. Fluch^{a,*}

^a Austrian Research Center Seibersdorf (ARCS) Research GmbH, Biogenetics/Natural Resources, 2444 Seibersdorf, Austria ^b Museum of Natural History Vienna, 1st Zoological Department, Molecular Systematics, Burgring 7, 1014 Vienna, Austria ^c University of Natural Resources and Applied Life Sciences, Department of Chemistry, Muthgasse 18, 1190 Vienna, Austria

Forest Ecology and Management 2006, vol 222, p266-271

Keywords: *Picea abies*; Nuclear microsatellites; Chloroplast microsatellites; Mitochondrial DNA; Genetic diversity; Population differentiation; Norway spruce; Austria

Abstract

Abstract Uniparentally inherited chloroplast simple sequence repeats (cpSSRs), mitochondrial markers (mt), as well as biparentally inherited nuclear simple sequence repeats (SSRs) were used to analyse genetic diversity in different subpopulations of Norway spruce in Austria. For that purpose Norway spruce individuals of different age classes, old (70–100 years) versus young (6–10 years) trees, were fingerprinted to elucidate the genetic effects caused by human management. Three populations from different elevations were selected in a restricted area. Four hundred and fifty individuals, 50 old and 100 young trees from each population, were analysed for this study. Seven paternally inherited cpSSRs revealed no variation within and among populations. In contrast, analysis of three maternal mitochondrial markers showed variation in one locus in individuals originating from the high elevation sample set. Analysis of five polymorphic nuclear SSR loci revealed 135 different alleles. Nuclear SSRs results showed that there is generally more genetic variation and heterozygosity within populations than among populations. Populations from high elevation exhibit more genetic variation than samples from middle and low elevations.

1. Introduction

Norway spruce is widely distributed throughout the temperate zone of Scandinavia and northern Europe. It is one of the economically most important tree species used for the production of pulp and timber (Koski *et al.*, 1997). Genetic diversity is the basis for adaptability, stability and evolution of species and forest tree populations (Müller-Starck *et al.*, 1992). Populations with a narrow genetic basis are widely

thought to be more sensitive to environmental changes or disease, leading to a decrease in productivity (Oleksyn *et al.*, 1994), although several other studies provide contradictory results (e.g. Chen *et al.*, 2001; Vargas *et al.*, 2002). Forest management based on partial or total cutting followed by natural or artificial regeneration can significantly impact the genetic diversity and population genetic structure of forest populations (Rajora, 1999; Rajora *et al.*, 2000). In order to determine genetic variation, different methods are available. Morphological markers (Langner, 1953), isoenzymes (Bartels, 1971; Morgante and Vendramin, 1991; Krutovskii and Bergmann, 1995; Müller-Starck, 1995; Geburek, 1999) or random amplified polymorphic DNAs (RAPDs) (Collignon and Favre, 2000) have been used to assess the population genetic structure and gene dispersal. More recently, nuclear sequence-tagged sites (STS; Scotti *et al.*, 2000), cpDNA microsatellites (Vendramin *et al.*, 2000), or mtDNA microsatellites (Gugerli *et al.*, 2001; Sperisen *et al.*, 2001) have been employed. Among all available markers, nuclear simple sequence repeats (SSRs or microsatellites) are a widely used class of genetic markers for population studies and genetic diversity assessment (for review see: Tautz and Schlotterer, 1994; Fossati *et al.*, 2003). But also this marker system has its benefits and restrictions due to its biparental mode of inheritance. Organellar DNA, being uniparentally inherited, offers a good alternative to this. Plant mitochondrial DNA is almost exclusively maternally inherited allowing the investigation of seed dispersal. Chloroplast DNA (cpDNA) is paternally inherited in gymnosperms, thus being a good tool for monitoring pollen migration (Birky, 1995; Neale and Sederoff, 1989; Vendramin and Ziegenhagen, 1997; Vendramin *et al.*, 2000). The combination of three marker systems investigating different parts of the available plant genomes (mt, cp, nt) appears as an appropriate way to determine the genetic status of the investigated individuals as they highlight different aspects of genetic diversity. The present study was carried out to assess genetic diversity in subpopulations of Norway spruce accessions from Kalwang (Austria), sampled at different elevations and age classes. The goal was to monitor the impact of human forestry management with respect to existing local genetic diversity on the resulting genetic profile along a centre timescale. We got data on the genetic generation structure of subpopulations of old trees as putative progenitors of naturally regenerated seedling patches. Natural regeneration is a common reforestation practice in Austria.

2. Materials and methods

2.1. Plant material and DNA isolation

The material for this study originated from the Kalwang area 80 km northeast of Graz (Austria). Plant material was taken at three different elevations to investigate the influence of altitude, as it was expected that populations from higher elevations in this area might be autochthonous, whereas those from lower elevations might have been replanted during the last 100–150 years (personal communication—Forstverwaltung Liechtenstein/ LIECO). Three populations were sampled: population A from low

elevation (960 m), population B from middle elevation (1260 m), and population C from high elevation (1520 m). Population C is most probably autochthonous, for populations A and B the status is uncertain. Populations A and B are from the same valley (distance 2 km), whereas population C is located in another valley, 5.5 km from A and B. The local altitudinal limits for the Norway spruce are between 960 and 2000 m. From each population sets of 50 old and 100 young trees were sampled. The young trees were 6–10 years old seedlings from naturally regenerated seedling patches, surrounded by 70–100 years old trees, which were sampled as putative mother trees, based on average seed shed distance (Geburek *et al.*, 1998). From each sample 100 mg of needles were first cooled in liquid nitrogen and disrupted in a Retsch Mil MM2000. Total genomic DNA was extracted using the DNeasy1 Plant Mini Kit (QIAGEN) following the supplier's instructions.

2.2. Molecular markers

Five nuclear primer pairs for nuclear dinucleotide repeat SSR loci (SpAGC1, SpAGC2, SpAC1F7, SpAGG3, and SpAC1H8) (Pfeiffer *et al.*, 1997), seven chloroplast microsatellite primers (PCP30277, PCP71978, PCP1289, PCP41131, PCP45071, Pt63718, and Pt71936; Vendramin *et al.*, 1996; Provan *et al.*, 1998), and three primer pairs for mitochondrial markers (mt23-D02: 50-CACCCTTGGGTAGACTGG-30 and 50-GGTTCACGCAGTGCTTCT-30; mt15-D02: 50-TATCTGACTTGCCCTTATC- 30 and 50-ATCCGAATACATACACC-30; mt1H01: 50-AAGATGGATCGCCCTT ACGC-30 and 50-GAGGAGGAGGCTTCGTCGTC- 30) newly developed for *P. abies* (GenBank accession numbers: AY897578, AY897577, and AY897576) were chosen for this study. PCR amplifications were conducted in a total volume of 25 µl using 1× PCR buffer (provided by the manufacturer QIAGEN), 2 mM MgCl₂, 0.2 mM dNTPs, 4 pmol of each primer, 0.6 units HotStarTaq polymerase (QIAGEN HotStar- Taq™ PCR), 20–30 ng of total genomic DNA. PCR-cycling conditions consisted of an initial denaturation step of 95 °C for 15 min, followed by 35 cycles of 50 s at 95 °C, 50 s at annealing temperature, and 1 min at 72 °C. A final extension step of 10 min at 72 °C ended the cycle. The fluorescently labelled microsatellite (cp, nuclear) fragments were analysed on an ABI 373 automated sequencer. Fragment sizing was performed using the ABI Genotyper software. After PCR amplification, mitochondrial fragments were analysed by PCR-RFLP. PCR products were digested using the restriction endonuclease HinfI. The digested DNA fragments were separated on a 8% non-denaturing polyacrylamide gel, followed by silver-staining for band detection.

2.3. Data analysis

Population genetic parameters were used to evaluate genetic polymorphisms and to analyse the population genetic structure of the three populations. Allele frequencies at each locus, number of alleles, inbreeding coefficient (F_{ST}), gene flow (N_m), observed (H_o) and expected (H_e) heterozygosity were measured as described by Nei and Roychoudhury (1974), and Nei (1978). Gene flow was estimated from F_{ST} using

the formula $N_m = 0.25(1/F_{ST} - 1)$ (where N is the effective population size and m is the fraction of migrants per generation). Genetic distances were computed according to Nei (1978). Computation analysis was conducted using the software POPGENE (vers. 1.32; Yeh and Boyle, 1997; <http://www.ualberta.ca/~fyeh/index.htm>).

3. Results

3.1. Organellar DNA analysis

Seven cpSSRs primers were employed to investigate the samples originating from three Austrian populations (A, B, C). No variation was detected in the 45 individuals analysed. Three mitochondrial markers were used (mt15-D02, mt23-D02, and mt1H01) for the analysis of 450 individuals. With two markers (mt23-D02 and mt1H01) monomorphic bands were obtained in all samples. With marker mt15-D02 variation was detected in three individuals from the high elevation population C. These three individuals belonged to the subpopulation of young trees.

3.2. Nuclear microsatellite analysis

With all five selected nuclear primers polymorphic bands were obtained. Locus SpAC1H8 was the most polymorphic among the five loci (41 alleles), while locus SpAC1F7 was the least polymorphic (15 alleles) among the 450 analysed samples. In total, 135 different alleles were detected across all the five loci. In the three populations, taking old and young trees together, 102 alleles were detected in population A, 109 alleles in population B, and 116 in population C. Population C (high elevation) had the highest number of alleles (35 alleles) at locus SpAC1H8 (see Table 1). Out of the 135 different alleles detected in three populations, 26 occurred only once in the investigated samples. These alleles, each found in only one subpopulation, are called private alleles (Hartl and Clark, 1997; Scotti *et al.*, 1999). In the young subpopulations some private alleles were found, which were not present in the putative mother trees (data not shown). In population C the highest number of private alleles (15) could be identified (see Table 1). Population C had the highest expected heterozygosity (H_e) value (0.8289) compared to populations A and B (Table 2). In all three sampled populations H_e was slightly higher in old trees compared to the young trees. All populations showed high levels of variability (H_e : 0.76–0.85). In all populations, the observed heterozygosity (H_o) was lower than the expected heterozygosity (Table 2). F_{ST} values between old and young trees in the three populations, as well as for pairwise comparisons among the three populations, ranged from 0 to 0.03. The inter-population difference was higher between population C and each of the other two populations (Table 3). The gene flow N_m between populations A and B was estimated at 1020, a value three times higher than between populations A–C and B–C. Genetic similarity among populations was quantified using Nei (1972) genetic distance and genetic identity (Table 4). Comparing the old subpopulations, it turned out that A–C are very similar with identity values between

90% and 92%. The genetic identity between old and young trees of the same population was generally lower (78–88%). Based on genetic distances a UPGMA dendrogram was constructed. Two main clusters can be seen (Fig. 1), one containing the young trees of population A and young trees of population B and the other comprising the rest of the samples. Within the latter cluster the young trees of population C splits off first, whereas the old trees of populations A–C cluster together.

4. Discussion

In order to assess the change of genetic diversity occurring via natural regeneration the existing genetic diversity in three subpopulations originating from three elevations within a restricted area was assessed. The aim was to investigate the genetic differentiation between stands and to detect changes of the gene pools by comparisons of old and young trees from the same stand. The results were expected to provide information on the influence of forest management practice on the genetic profile of geographically closely located subpopulation of Norway spruce. The fact of private alleles occurring in the old trees but not being present in the young trees may be ascribed to planting or to natural selection against certain genotypes. On the other hand, new alleles in the young trees might indicate gene flow from outside, either from natural sources or through planting, and cannot be explained by natural causes, i.e. selection, drift or isolation. The genetic structure of subpopulations was studied at both intra- and inter-population level with three different marker types, which were employed to estimate the extent of genetic diversity in Norway spruce from three Kalwang populations.

4.1. Organellar data analysis maternally inherited mitochondrial markers are of great value to clarify the relative importance of pollen versus seed dispersal (Sperisen *et al.*, 1999). In general, low degrees of variation were found within Norway spruce populations throughout Europe, when using maternal markers for population studies (Sperisen *et al.*, 2001). When concentrating on the alpine area (Austria, Swiss, Italy, France, and Germany) Gugerli *et al.*, (2001) as well as Maghuly *et al.*, (2005) found variation in few selected populations. In this study, the three analysed mitochondrial markers (mt15-D02, mt23-D02, and mt1H01) showed variation at one locus in just three individuals originating from the high elevation population (marker mt15- D02). These data can be taken as a hint that the less managed high altitude population might be autochthonous. Chloroplasts are paternally inherited in Norway spruce (Vendramin *et al.*, 1996; Sperisen *et al.*, 1999) and thus migrate through both seeds and pollen. In Norway spruce the genetic diversity detected using cpSSR markers is lower than reported for other analysed conifer species, i.e. *Pinus halepensis*, *P. brutia* (Bucci *et al.*, 1998), *Pinus pinaster* or *Abies alba* (Vendramin *et al.*, 1998, 1999). From the whole sample set, 1% of the probes were used for cpSSR analysis. For each of the three populations (A–C) 5 old trees and 10 young trees were selected randomly for analysis. Given the fact that variation in cpSSR has turned out to be low in previous studies covering a much larger geographic range (Schuster *et al.*, 1989; Vendramin *et al.*, 2000), this sample

size appears sufficient for the restricted sampling area of the present investigation. Our analysis of cpSSRs in this subset of 45 individuals revealed no variation in seven investigated cp regions. On a larger geographical scale (IUFRO provenance), Vendramin *et al.*, (2000) reported that cpSSR variation in 97 investigated Norway spruce populations was lower than the variation obtained using mitochondrial markers. This was expected given the different mode of transmission of the two genomes in conifers. As the pollination strategy of Norway spruce is based on long-range pollen dispersal by wind, the cp haplotypes might become uniform over longer distances (Schuster *et al.*, 1989; Vendramin *et al.*, 2000). Another reason for the lack of polymorphisms in our study could be the small size of the study area. Due to the limited size of our sample we may also have missed some rare variants, but major polymorphisms should have been detected.

4.2. Nuclear SSRs analysis

In previous studies (Geburek *et al.*, 1998; Maguire *et al.*, 2000; Rajora *et al.*, 2000; Perera *et al.*, 2000; Mariette *et al.*, 2001; Fossati *et al.*, 2003) the use of nuclear SSR has proven to be successful in detecting genetic variation in populations. In our study all five nuclear microsatellites were polymorphic in all three populations. In general, the number of alleles detected was lower in old trees than in young trees, although old trees possessed more private alleles in all populations. He proved to be higher in old trees as compared to young trees investigated. Population C showed the overall highest degree of variation. Two aspects were elucidated by nuclear SSR analysis of our samples. On one hand, the human impact on genetic diversity due to forestry management practices was evaluated in different elevations, which were differentially managed. On the other hand, changes in the genetic profile of young trees in relation to the existing genetic diversity was assessed.

4.3. Variation between populations

In each of the three investigated populations heterozygosity as well as allele number was high, indicating a high degree of genetic diversity. Comparing heterozygosity among the three populations revealed no significant difference. The occurrence of private alleles in the young tree samples demonstrates that the old trees are not the only ancestors of the young populations, which initially was anticipated due to the sampling strategy. This finding can be attributed to wind mediated pollen migration and seed dispersal in Norway spruce on one hand, and to the fact that our sampling did not cover the whole area of putative mother trees. Seed dispersal is another important component for gene flow in most plant species. Especially the animal or wind mediated dispersal of seed may be a more important component for gene flow than pollen transfer, at least across elevation gradients (Schuster *et al.*, 1989). Therefore we have to take into account that a genetic contribution to our young populations from individuals outside the sampling area occurred. According to Wright (1951), F_{ST} values between 0 and 0.05 indicate little genetic differentiation (Hartl and Clark, 1997). As the F_{ST} values between the populations from different elevations

were between 0.002 and 0.0068 (Table 3), the genetic differentiation among populations as well as between old and young subpopulations (F_{ST} between 0.012 and 0.029) is low as compared to values of $F_{ST} = 0.053$ reported by Lagercantz and Ryman (1990) for 70 populations using 22 allozyme markers. However, this may be explained by the fact that in those previous studies sampling was performed on a much larger geographical scale. In contrast to these data and in accordance with our findings, Geburek (1999) reported $F_{ST} = 0.012$ for 29 alpine populations using 17 isoenzyme loci, as well as Scotti *et al.*, (2000) who described $F_{ST} = 0.118$ in for eight populations along the Alpine range using seven SCAR (sequence characterized amplified region) loci. In Norway spruce, low levels of genetic differentiation between populations could be explained by the mating system and by high migration rates (Taberlet *et al.*, 1998). Under the assumption that the populations are autochthonous, our data would indicate high inter-population gene flow (as described by N_m) between middle and low elevations but restricted gene flow between low and high elevation as well as between middle and high elevation (Table 3). However, taking into account that the populations from middle and low elevations may have been replanted during the last 100–150 years, the low gene flow with respect to the high elevation population may be simply ascribed to the fact that the replanted populations have a different origin. Another explanation could be the different geographical position of origin of the high elevation population C.

4.4. Impact of human selection

Comparing the genetic profile of old trees in the three populations to the profile of young trees we found that old trees were slightly more variable than the young trees although the number of individuals investigated is lower. Also more private alleles, these are alleles present in only one of the investigated subsets, occurred in the old trees. By using Nei's genetic distance and genetic identity measures, it was shown that the genetic identity values between the selected young trees and the corresponding populations of old trees are lower than the genetic identity values between different populations. In the dendrogram (Fig. 1) the old trees of populations A–C form a cluster, whereas the young tree populations are separated. As possible interpretation for this finding we can assume that the old trees represent rather autochthonous stands which, although grown at different elevations, have not undergone significant differentiation. In contrast, the gene pools of the young trees (in particular populations A and B) may have experienced influences from outside, i.e. introduction of new genotypes through planting or from natural sources. Furthermore, the comparison of observed versus expected heterozygosity (Table 2) reveals a lack of heterozygotes in all populations. This result suggests a selective advantage of homozygotes which may be favoured in less heterogenous environments or environments with one dominant stress factor (Gregorius *et al.*, 1985). The deviation in favour of the homozygotes is less pronounced in young trees of populations A and B. Yazdani *et al.*, (1985) compared the genetic structure of 100 years old adult and 10–20 years old young trees by isoenzyme analysis in *Pinus silvestris*. In their study the observed heterozygosity was also higher in young than in adult trees. Our results indicate that higher genetic variation is found in populations from high elevations as

revealed by both nuclear microsatellites and mitochondrial markers. Reduced variation in the low and middle elevations (populations A and B) might exemplify a case of potential human impact. Our study for the first time analysed genetic differentiation and possible human impact at the subpopulation level using molecular tools.

Acknowledgements

The authors wish to thank Dr. Heinze of the Bundesforschungs- und Ausbildungszentrum für Wald, Naturgefahren und Landschaft (BFW) for helpful comments and suggestions on the manuscript, and LIECO for providing plant material for this study. The Österreichischer Austauschdienst (ÖAD) supported this study.

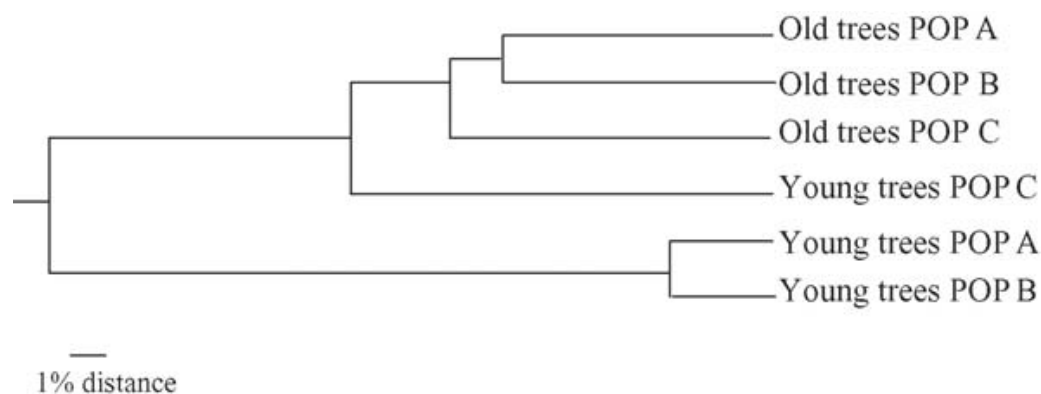


Fig. 1. UPGMA dendrogram constructed from genetic distances (Table 4) between the three populations and subpopulations calculated from the nuclear SSR data. While the old tree populations A–C cluster together, the young tree populations are clearly distinct from their corresponding old trees.

Table 1 Total number of different alleles at each nuclear microsatellite locus in the three populations

	SpAGC1	SpAGC2	SpAC1F7	SpAGG3	SpAC1H8	P over all 5 loci	P p.a. over all 5 loci
POP A old	8	16	12	15	19		
POP A young	7	19	17	16	23		
POP A	8	21	20	22	31	102	7
POP B old	9	16	16	18	25		
POP B young	8	19	19	19	23		
POP B	9	21	22	24	33	109	4
POP C old	13	17	15	19	29		
POP C young	12	21	20	20	32		
POP C	15	20	22	24	35	116	15

The total number over all five loci and the total number of private alleles (p.a.) are also given.

Table 2 Observed (Ho) and expected (He) heterozygosity across the loci in the three old and young populations

POP	Ho	He
POP A young trees	0.6019	0.7596
POP B young trees	0.5360	0.7658
POP C young trees	0.4490	0.8106
POP A old trees	0.4366	0.8173
POP B old trees	0.4077	0.8488
POP C old trees	0.4692	0.8356
POP A old and young trees	0.5452	0.8006
POP B old and young trees	0.4919	0.8145
POP C old and young trees	0.4561	0.8289

Table 3 Pairwise estimation of genetic differentiation (F_{ST}) and gene flow (N_m = mi-grants per generation) for the three populations A–C

Pairwise comparison of sample sets	Mean (?ve loci) Sample size	F_{ST}	N_m
POP A old–young	288	0.0292	83.89 196
POP B old–young	284	0.0270	
POP C old–young	298	0.0125	
	572	0.0024	1020
POP A–B POP A–C	586	0.0068	365
POP B–C	586	0.0068	365

Table 4 Genetic identity (above diagonal) and genetic distance (below diagonal) estimates between the three populations using the nuclear SSR data

POP ID	POP A old	POP A young	POP B old	POP B young	POP C old	POP C young
POP A old		0.7823	0.9258	0.8105	0.9224	0.9057
POP A young	0.2456		0.7693	0.9718	0.8081	0.8793
POP B old	0.0771	0.2623		0.7848	0.9043	0.8803
POP B young	0.2101	0.0286	0.2423		0.8367	0.8837
POP C old	0.0808	0.2130	0.1006	0.1783		0.8840
POP C young	0.0990	0.1286	0.1275	0.1237	0.1233	

References

- Bartels, H., 1971. Genetic control of multiple esterases from needles and macrogametophytes of *Picea abies*. *Planta* 99, 238–289.
- Birky, C.W., 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. U.S.A.* 92, 11331–11338.
- Bucci, G., Anzi dei, M., Madaghiele, A., Vendramin, G.G., 1998. Detection of haplotypic variation and natural hybridization in halepensis-complex pine species using chloroplast (SSR) markers. *Mol. Ecol.* 7, 1633–1643.
- Chen, Z., Kolb, T.E., Clancy, K.M., Hipkins, V.D., DeWald, L.E., 2001. Allozyme variation in interior Douglas-fir: association with growth and resistance to western spruce budworm herbivory. *Can. J. For. Res.* 31, 1691–1700.
- Collignon, A.M., Favre, J.M., 2000. Contribution to the postglacial history at the western margin of *Picea abies*' natural area using RAPD markers. *Ann. Bot.* 85, 713–722.
- Fossati, T., Grassi, F., Sala, F., Castiglione, S., 2003. Molecular analysis of natural populations of *Populus nigra* L. intermingled with cultivated hybrids. *Mol. Ecol.* 12, 2033–2043.
- Geburek, T., Mottinger-Kroupa, S., Burg, K., 1998. Genetic variation of Norway spruce (*Picea abies* [L.] Karst.) populations in Austria. II. Microspatial patterns derived from nuclear sequence tagged microsatellite sites. *For. Genet.* 5, 231–237.
- Geburek, Th., 1999. Genetic Variation of Norway Spruce (*Picea abies* [L.] Karst.) populations in Austria. III. Macrospace allozyme pattern of high elevation populations. *For. Genet.* 6, 201–211.
- Gregorius, H.R., Hattemer, H.H., Bergmann, F., Müller-Starck, G., 1985. Umweltbelastung und Anpassungsfähigkeit von Baumpopulation. *Silvae Genet.* 34, 230–241.
- Gugerli, F., Sperisen, C., Büchler, U., Magni, F., Geburek, T., Jeandroz, S., Senn, J., 2001. Haplotype variation in a mitochondrial tandem repeat of Norway spruce (*Picea abies*) populations suggests a serious founder effect during postglacial re-colonization of the western Alps. *Mol. Ecol.* 10, 1255–1263.
- Hartl, D.L., Clark, A.G., 1997. Principles of Population Genetics, 2nd ed. Sinauer Associates, Sunderland, MA.
- Koski, V., Skroppa, T., Paule, L., Wolf, H., Turok, J., 1997. Technical Guidelines for Genetic Conservation of Norway Spruce (*Picea abies* (L.) Karst.) International Plant Genetic Resources Institute, Rome, 42 pp.
- Krutovskii, K.V., Bergmann, F., 1995. Introgressive hybridization and phylogenetic relationships between Norway, *Picea abies* (L.) Karst., and Siberian, *P. obovata* Ledeb., spruce species studied by isoenzyme loci. *Heredity* 74, 464–480.
- Lagercantz, U., Ryman, N., 1990. Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution* 44, 38–53.
- Langner, W., 1953. Eine Mendelspaltung bei Aurea-Formen von *Picea abies* (L.) Karst. als Mittel zur Klärung der Befruchtungsverhältnisse im Walde. *Zeitschrift für Forstgenetik und Forstpflanzenzüchtung* 2, 49–51.
- Maghuly, F., Praznik, W., Pinsker, W., Fluch, S., 2005. Population differentiation of Austrian Norway spruce (*Picea abies*) Populations assayed by mitochondrial markers. *Tree Genet. Genomes*, submitted for publication.
- Maguire, T.L., Saenger, P., Baverstocks, P., Henry, R., 2000. Microsatellite analysis of genetic structure in the mangrove species *Avicennia marina* (Forsk.) Vierh (Avicenniaceae). *Mol. Ecol.* 9, 1853–1862.
- Mariette, S., Chagne, D., Lezier, C., Pastuszka, P., Raffin, A., Plomion, C., Keremer, A., 2001. Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers. *Heredity* 86, 469–479.
- Morgante, M., Vendramin, G.G., 1991. Genetic variation in Italian populations of *Picea abies* (L.) Karst. and *Pinus leucodermis* Ant. In: Müller-Starck, G., Ziehe, M. (Eds.), Genetic Variation in European Populations of Forest Trees. J.D. Sauerländer's Verlag, Frankfurt am Main, pp. 205–227.
- Müller-Starck, G., Baradat, Ph., Bergmann, F., 1992. Genetic variation within European tree species. *New For.* 6, 23–47.
- Müller-Starck, G., 1995. Genetic variation in high-elevated populations of Norway spruce (*Picea abies* (L.) Karst.) in Switzerland. *Silvae Genet.* 44, 356–362.
- Neale, B., Sederoff, R.R., 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in Loblolly pine. *Theor. Appl. Genet.* 77, 212–216.

- Nei, M., 1972. Genetic distance between populations. *Am. Nat.* 106, 283–292.
- Nei, M., Roychoudhury, A.K., 1974. Sampling variances of heterozygosity and genetic distance. *Genetics* 76, 379–390.
- Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89, 583–590.
- Oleksyn, J., Prus-Glowacki, W., Giertych, M., Reich, P.B., 1994. Relation between genetic diversity and pollution impact in a 1912 experiment with East European *Pinus sylvestris* provenances. *Can. J. For. Res.* 24, 2390–2394.
- Perera, L., Russell, J.R., Provan, J., Powell, W., 2000. Use of microsatellite DNA markers to investigate the level of genetic diversity and population genetic structure of coconut (*Cocos lucifera* L.) *Genome* 43, 15–21.
- Pfeiffer, A., Olivieri, A.M., Morgante, M., 1997. Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.) *Genome* 40, 411–419.
- Provan, J., Soranzo, N., Wilson, N.J., McNicol, J.W., Forrest, G.I., Cottrell, J., Powell, W., 1998. Gene pool variation in Caledonian and European Scots pine (*Pinus sylvestris* L.) revealed by chloroplast simple sequence repeats. *Proc. R. Soc. Lond. Ser. B* 265, 1697–1705.
- Rajora, O.P., 1999. Genetic biodiversity impacts of silvicultural practices and phenotypic selection in white spruce. *Theor. Appl. Genet.* 99, 954–961.
- Rajora, O., Rahman, M.H., Buchert, G.P., Dancik, B.P., 2000. Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. *Mol. Ecol.* 9, 339–348.
- Schuster, W.S., Alles, D.L., Mitton, J.B., 1989. Gene flow in limber pine: evidence from pollination phenology and genetic differentiation along an elevational transect. *Am. J. Bot.* 76, 1395–1403.
- Scotti, I., Paglia, G., Magni, F., Morgante, M., 1999. Microsatellite markers as a tool for the detection of intra and inter population genetic structure. In: Gillet, E.M. (Ed.), Which DNA Marker for Which Purpose? Final Compendium of the Research Project “Development, Optimisation and Validation of Molecular Tools for Assessment of Biodiversity in Forest Trees” in the European Union DGXII Biotechnology FW IV Research Programme Molecular Tools for Biodiversity (Chapter 8). <http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>.
- Scotti, I., Vendramin, G.G., Matteotti, L.S., Scarponi, C., Sari-Gorla, M., Binelli, G., 2000. Postglacial recolonization routes for *Picea abies* K. in Italy as suggested by the analysis of sequence-characterized amplified region (SCAR) markers. *Mol. Ecol.* 9, 699–708.
- Sperisen, C., Büchler, U., Ma'tya's, G., Ackzell, L., 1999. Mitochondrial DNA variation provides a tool for identifying introduced provenances: a case study in Norway spruce. In: Gillet, E.M. (Ed.), Which DNA Marker for Which Purpose? Final Compendium of the Research Project “Development, Optimisation and Validation of Molecular Tools for Assessment of Biodiversity in Forest Trees” in the European Union DGXII Biotechnology FW IV Research Programme “Molecular Tools for Biodiversity” (Chapter 10). <http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>.
- Sperisen, C., Büchler, U., Gugerli, F., Ma'tya's, G., Geburek, T., Vendramin, G.G., 2001. Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. *Mol. Ecol.* 10, 257–263.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* 7, 453–464.
- Tautz, D., Schlötterer, C., 1994. Simple sequences. *Curr. Opin. Genet. Dev.* 4, 832–837.
- Vargas, C.F., López, A., Sánchez, H., Rodríguez, B., 2002. Allozyme analysis of host selection by bark beetles in central Mexico. *Can. J. For. Res.* 32, 24–30.
- Vendramin, G.G., Lelli, L., Rossi, P., Morgante, M., 1996. A set of primers for the amplification of 20 chloroplast microsatellites in Pinacea. *Mol. Ecol.* 5, 595–598.
- Vendramin, G.G., Ziegenhagen, B., 1997. Characterisation and inheritance of polymorphic plastid microsatellites in *Abies*. *Genome* 40, 857–864.
- Vendramin, G.G., Anzidei, M., Madaghiele, A., Bucci, G., 1998. Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. *Theor. Appl. Genet.* 97, 456–463.

- Vendramin, G.G., Degen, B., Petit, R.J., Anzidei, M., Madaghiele, A., Ziegenhagen, B., 1999. High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Mol. Ecol.* 8, 1117–1126.
- Vendramin, G.G., Anzidei, M., Madaghiele, A., Sperisen, C., Bucci, G., 2000. Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.) *Genome* 43, 68–78.
- Wright, S., 1951. The genetical structure of populations. *Ann. Hum. Genet.* 15, 323–354.
- Yazdani, R., Muona, O., Rudin, D., Szmidt, A.E., 1985. Genetic structure of *Pinus sylvestris* L. seed-tree stand and naturally regeneration understory. *For. Sci.* 31, 430–436.
- Yeh, F.C., Boyle, T.J.B., 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belg. J. Bot.* 129, 156–159.



2.1.3 Nuclear markers for oak (*Quercus* ssp)

Quercus petraea and *Qercus robur* are diploid species with a karyotype of $2n=24$ chromosomes and a genome size of about 900 Mb. DNA content has been measured as 1,88pg/2C. Pedunculate and sessile oaks belong to *Quercus* taxonomic section (white oaks) (Manos *et al.*, 1999). Interspecific genetic exchange i.e. hybridization within the genus *Quercus* is typical and widespread. Even though the species are genetically compatible, their morphology and ecological preferences stay distinct as ecological factors seem to prevent or limit gene flow. There have been several attempts to discriminate the two sympatric highly interfertile white oak species *Q. robur* and *Q. petraea* on DNA level. Genetic markers like microsatellites (Steinkellner *et al.*, 1997a; Streiff *et al.*, 1998a), 18S, 25S, 5S ribosomal DNA (Chokchaichamnankit *et al.*, 2008), random amplified polymorphic DNA (RAPD) (Williams *et al.*, 1990a), sequence characterized amplified regions (SCARs) (Bodenes *et al.*, 1997), isozymes (Müller-Starck *et al.*, 1992; Samuel 1999; Zanetto *et al.*, 1996a) and amplified fragment length polymorphism (AFLP) have been developed in oak. *Quercus petraea* and *Q. robur* hybridize frequently, but still form two discrete genetic entities, which could be underpinned by microsatellite analysis (Muir *et al.*, 2000).

A study of interspecific gene flow in a mixed oak stand revealed high unidirectional hybridisation: *Q. petraea* pollen contributed highly to *Q. robur* progeny, while *Q. petraea* itself was preferably fertilized by its own species (Bacilieri *et al.*, 1996c; Lexer *et al.*, 2000b). Asymmetric hybridisation was also observed in controlled interspecific crosses of *Q. petraea* and *Q. robur* (Bacilieri *et al.*, 1996a; Bacilieri *et al.*, 1996b; Kremer *et al.*, 2002a; Kremer *et al.*, 2002b; Zanetto *et al.*, 1996b). Thus by means of pollen flow *Q. petraea* could be able to colonize, after successive frequent backcrossing, new, formerly *robur*-specific habitats. Nevertheless long-distance pollen flow, heterogeneity of the pollen dispersal and variability of mating success certainly contribute to keep high levels of diversity in oak stands (Lexer *et al.*, 2000a; Lexer *et al.*, 2006b; Petit *et al.*, 2004a; Streiff *et al.*, 1998b).

The two interfertile species exhibit phenotypic differences, but show only extremely low genetic differentiation for isoenzymes or rDNA and generally share the same chloroplast genotypes in the same geographic areas. Thus it has been assumed that genomic regions, where significant interspecific differences occur, are rare. Recently, numerous genetic markers (AFLP, SSR) were revisited and applied to natural mixed oak stands in Western Europe. The distribution of markers relevant in species differentiation yielded an “L-shaped” curve indicating a small number of markers with a higher index of species differentiation. (Petit *et al.*, 2004b)

VNTR Isolation

For the isolation of VNTR markers as being used for mapping in Barreneche (Barreneche *et al.*, 1998a) RAPD has been used to isolate a genomic fragment of 2.3 kb in size generated by Operon primer D7 on genomic oak DNA from an Austrian accession (NW1 – Niederweiden1, NÖ). Using this fragment in Southern hybridization experiments on HaeIII digested genomic DNA, it revealed a multi band hybridization profile (see fig.2). Due to the nature of RAPD fragments target the repetitive region of the genome as the method relies on primer binding in closely situated inverted repeat regions. The sequence of the region is found in the Appendix 0. The banding pattern exhibits no regular sized banding, but rather generates an individual-specific fingerprint profile. Sequence comparison using BLAST revealed homology of the region 130–740 bp *Vitis vinifera* HSC70-1 (heat shock cognate 70 kDa protein 1) with $7e^{-96}$ whereas the rest of the sequence revealed no sequence homology.

Restriction analysis of the 2,3 kb fragment revealed 4 cloneable subfragments that have been used for further hybridization experiments in order to elucidate which proportion of the fragment reveals the fingerprint like banding pattern, indicating the presence of a highly conserved, highly abundant repetitive element.

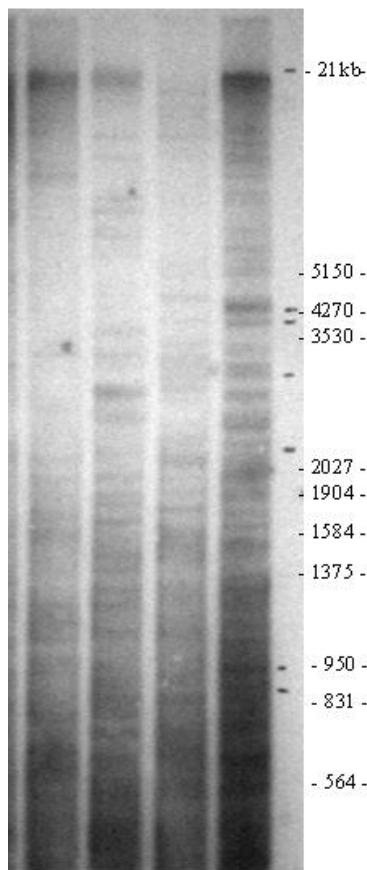


Fig 2: Southern hybridization autoradiogram of HaeIII digested genomic oak DNA of 4 individuals hybridized with the 2.3 kb RAPD fragment (ST8, GU8, GS2, GR4). The gel separation run was performed on at 2V/cm on a 0.7% agarose gel.

Sequence analysis of this fragment revealed the presence of several short and longer repetitiv units. Three different tandem repeat units (Omin 1,2,3) could be identified within the fragment (see fig 3). These are minisatellite regions of different length and different numbers of repeats as well as different orientation towards each other. Omin3 is a tandem direct repeat unit of 36 and 37 bp (RevCore1 and 2) respectively, following each other without insertion. 1000 bp upstream of this region (1131– 1245 bp), another direct tandem repeat can be found, termed Omin1. It consists of 4 direct repeat units of 14–16 bp in length being separated from each other by 4, 6 and 14 bp inserts. These repeats all start with a Mnl1 restriction enzyme cut site (mnl core 1–4). Fig 5 shows the hybridization pattern generated on Mnl1 digested genomic DNA when using the Omin1 repeat region as target. The repetitive nature of the element in the genome can nicely be seen as banding pattern with an average distance between the bands of approx. 150 bp. This region has been mapped on the oak genome in Barreneche T. et al (1997).

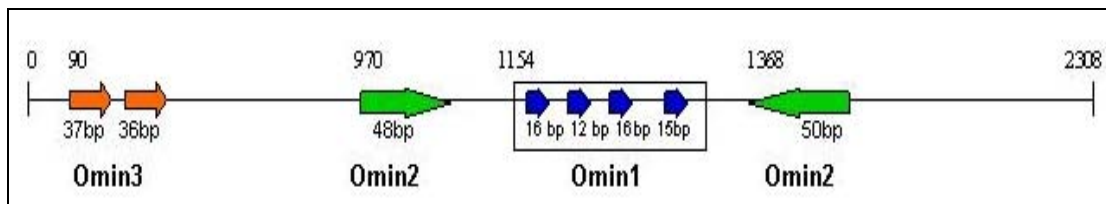


Fig 3: Scheme of the 2.3 kb fragment and its repetitive elements

Multiple Sequence Alignment				
MN1	CCCTCTATG	<u>TTTG</u>	TAA	16 (1131-1146)
MN3	CCCTCTATG	<u>TTTG</u>	GAA	16 (1181-1196)
MN4	CCCTCTATC	<u>TTTG</u>	-AA	15 (1207-1221)
MN2	- - -TCTATG	<u>TTTG</u>	-AA	12 (1153-1164)
	*****	*****	**	
Consensus-Sequence: 16bp; Identity: 11 (68.8%)				
‘*’ 100% conserved sequences				

Fig 4: Sequence alignment of the MnlCore regions of Omin1

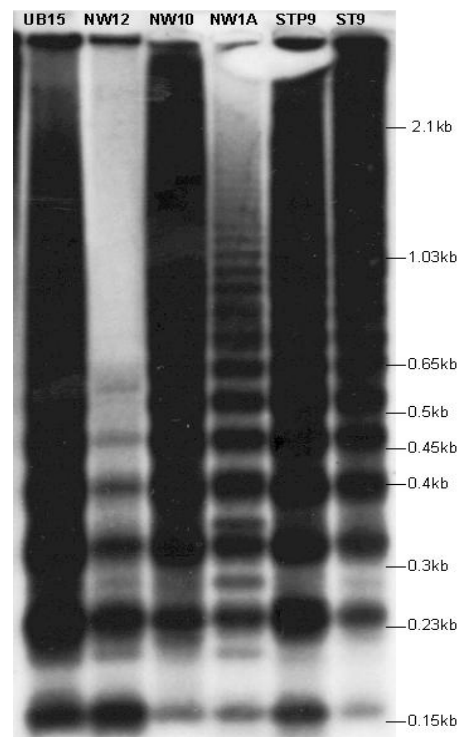


Fig 5: Southern hybridization autoradiogram of MnlI digested genomic oak DNA of 6 individuals (UB15, NW12, NW10, NW1A, STP9, ST9). The gel separation run was performed at 2V/cm on a 3% agarose gel. Hybridization with the 221 bp Omin1 fragment amplified by PCR.

Outside of the Omin1 region, another set of repeat elements could be identified – inverted repeats of 48 and 50 bp respectively (Omin2) in a distance of about 100 bp from the Mnl repeats. When using this region as hybridization probe, also a repetitive banding pattern could be generated (fig 7) with distances between the repeats of about 350 bp.

a)

F1C	-	AAAATATTGAACTTTAGGGGAGCAAAATT-AAAATTA--CCCC-AAAGCTT	-48
F2CINV	-	AAA-TTTTGAACTTTAGGG-AGTAAAAGTCAAAAATAATCCCCTAAAGCTT	-50
		.*.**.***.****.*.***.****.****.****.*	
Identity: 42 bases (87.50%)			

b)

Sequence comparison		REVC1 (90-127), REVC2 (128-164)	
REVC1	-	AAAGAATCCTCTTTGCTTTCTCACAAGCAGTTCTTAA	-37
REVC2	-	AAAGAATCCTCTTTGCTT-CTCACAAGCAGTTCTTAA	-36
		*****.*****	
Identity: 36 bases(100.00%)			

Fig 6 a, b: Sequence alignment of the MnlCore regions of Omin1 (a) and Omin3 (b)

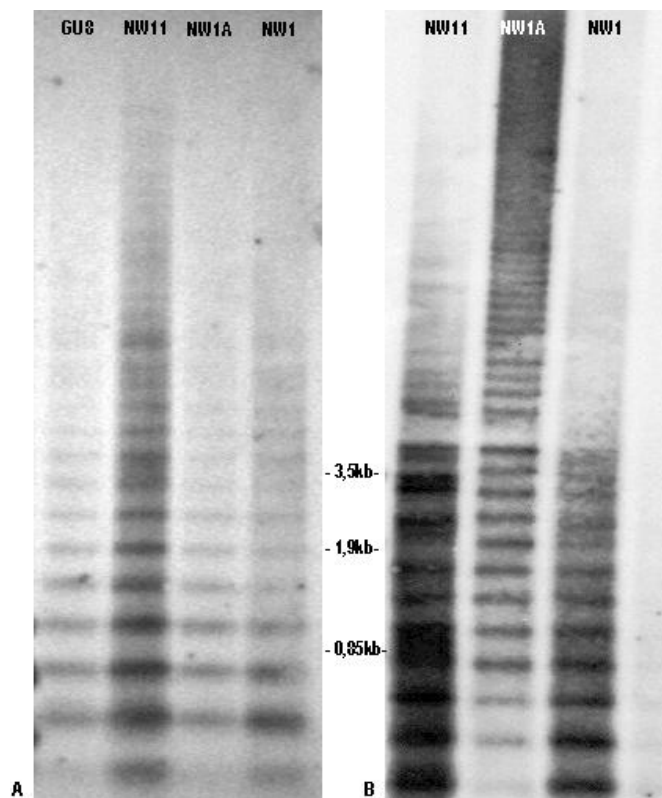


Fig 7: Southern hybridization autoradiogram of HaellI digested genomic oak DNA of 3 and 4 individuals (GU8, NW11, NW1A, NW1). The gel separation run was performed on a 2,5% agarose gel. Hybridization was performed with Omin2 A and B fragments amplified by PCR (150 bp each).

The fragment outside the upstream inverted repeat (1425–1986 bp) however revealed, when used as hybridization probe on *Hae*III digested DNA, the initially generated, individual specific banding pattern. This region showed no repetitive nature, neither homology to any known sequence (see figure 8)

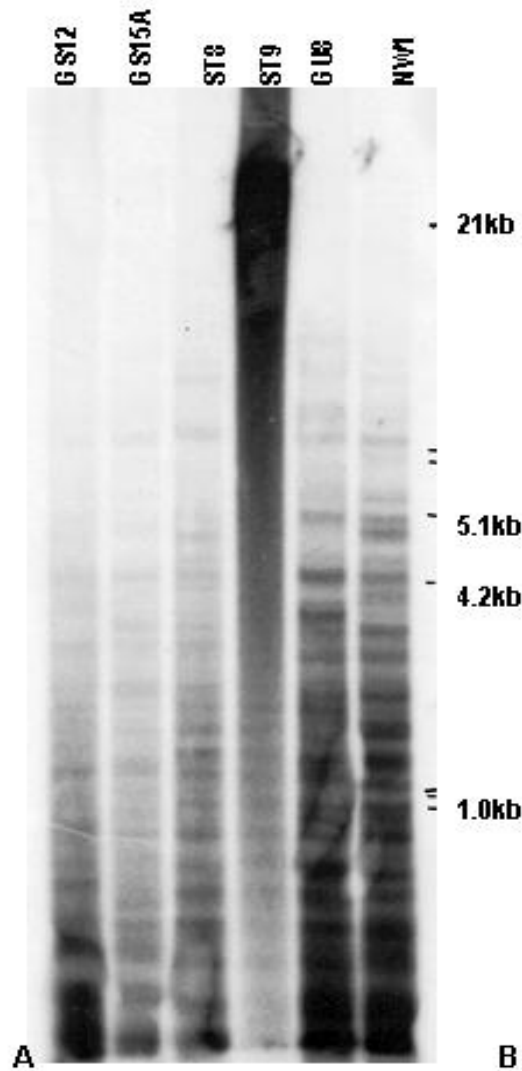


Fig 8: Southern hybridization autoradiogram of *Hae*III digested genomic oak DNA of 6 individuals (GS12, GS15A, ST8, ST9, GU8, NW1). The gel separation run was performed on a 0,7 % agarose gel. Hybridization with 221 pb Omin2 A and B fragments amplified by PCR (350 bp fragment 1631–1986 bp)

BLAST analysis of the 2.3 kb fragments revealed no significant sequence similarities to known gene sequences. The repetitive regions within the sequence though show homology to transposable elements like MULEs – Mutator Like elements from maize or the Stowaway family (MITE) from *Arabidopsis* (Fig 9).

```

800AATGCACAAGAAGCATTATATATTGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACGAGTGTACTTTT
CATTCTGAATTTAAGAAATTTGAAGCGCAACATTCAGTAGCTCATGACCTAATAAGATAGTCATAAAAG
AGAGAGGACTAAGGAAAAATCAAGATTAAATTATAAATTACACTCAAGCTTTAGGGGATTATTTTTTGAC
TTTTACTCCCTAAAGTTTCAAAATTTCAATATGCTCTCATAACATAAAAGTCTCATGAGTAGTCAATCCC
TTCCAACCTTTGTGTTGTGCCTAGCTGTCTTAAGCGTATAAGATTGGGCCGCAGTGCTGCCCTCTATGTT
TGTAACCTTACATCTATGTTTGAAATTTGATTGATTTGAACCCCTCTATGTTTGGAACCTGTCATACACCCCT
CTATCTTTGAAATTTGTTTTTAATTGATTTTCGAACTTAACCAGCTAACCAGTCAATTACAAGCCCTCTC
CTCAGGGAAGGATTTTTTTTTTATAAGGCTTCACACATTGTATGGGAAGCAAATCCAAATGATGATACCT
GAAGTTAGGAGAGCCAATCAAAATATTGAACTTTAGGGGAGCAAAATTTAAATTTACCCCAAAGCTTTA
TATATATATAGATGGGATTGACGAGAACATGT 1400

```

Fig. 9: Region 800–1450 bp of the 2.3 kb RAPD fragment. Different colours indicate transposable representing sub-elements: MULE-element starting with a 8 bp direct repeat as insertion site (light blue). In between underlined the elements of a Stowaway family, starting at the TA target site.

The above described region Omin3 (RevC 1 and 2) was used as sequence characterized region for anchoring a genetic linkage map of *Q. robur* L. (see Appendix II: A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers Barreneche T, Bodenes C, Lexer C, Trontin J F, Fluch S, Streiff R, Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glössl J, Kremer A)

Micosatellites loci in *Quercus* ssp

Microsatellites are most informative polymorphic markers which have proved to be highly conserved among oak species (Lexer *et al.*, 2006a; Steinkellner *et al.*, 1997b) and thus can be applied in a wide range of oak as well as in related Fagaceae species in comparative studies (e.g. comparative genetic mapping (Barreneche *et al.*, 2004b). Microsatellites have already been used for answering the question of species status in *Q. robur* and *Q. petraea* (Muir *et al.*, 2000; Muir *et al.*, 2004) using twenty loci in 10 European sessile and pedunculate oak populations. Relatedness on the basis of shared alleles was visualized using an UPGMA approach and revealed a clustering of populations according to species rather than geographic origin. Streiff *et al.*, (1999) used six hypervariable SSR loci in parentage analysis for tracing pollen dispersal in a mixed oak stand of *Q. robur* and *Q. petraea*.

2.1.3.1 Detection of microsatellite instability during somatic embryogenesis of oak (*Quercus robur* L.)

E. Wilhelm¹, K. Hristoforoglu², S. Fluch¹ and K. Burg¹

(1) Division of Environment and Life Sciences, Department of Biotechnology, ARC Seibersdorf Research, 2444 Seibersdorf, Austria

(2) Horticultural College and Research Institutes, Grünbergstrasse 24, 1131 Vienna, Austria

Published: 2005 Plant Cell Reports 23: 790-795

Abstract Five microsatellite loci (QpZAG1/5, QpZAG9, QpZAG36, MSQ4, MSQ13) were used to test for genetic stability of three somatic embryogenic culture lines of *Quercus robur* L. and plantlets derived therefrom. DNA variation was detected among somatic embryos within all embryogenic lines, whereas no genetic instability was found among the regenerated plants. Two microsatellite loci revealed variation, and a locus-dependent instability was observed. The most polymorphic and useful microsatellite locus for detecting genetic variation was QpZAG9, with 28.5% of the investigated loci being variable.

Keywords Simple sequence repeats - Somaclonal variation - Somatic embryogenesis - In vitro culture - *Quercus* ssp

Abbreviations AFLP Amplified fragment length polymorphism - RAPD Random amplified polymorphic DNA - SE Somatic embryo - SSR Simple sequence repeat

1 Introduction

Cultures of isolated plant cells and tissues in artificial nutrient media have become enormously important for fundamental and practical applications, particularly in the

micropropagation industry. In order to utilize the advantages of micropropagation, trueness-to-type reproduction is highly desirable. However, in vitro culture has revealed itself as being mutagenic, and plants regenerated from organ cultures, calli, protoplasts and via somatic embryogenesis often show phenotypic and DNA variation, a phenomenon which has been termed somaclonal variation (Larkin and Scowcroft 1981). Reliable means for assessing the genetic stability of a genotype throughout a given tissue culture would therefore be of a tremendous advantage for micropropagation and genetic engineering.

Various methods are available for characterizing mutations induced by tissue cultures, including phenotypic identification, cytological methods and DNA analysis techniques (Jain *et al.*, 1998; Rani and Raina 2000), and many investigators working with tree species have dealt with the assessment of somaclonal variation using various molecular markers (Isabel *et al.*, 1996; Fourré *et al.*, 1997; DeVerno *et al.*, 1999; Vendrame *et al.*, 1999). Both dominant and co-dominant marker systems have been successful in detecting genetic instability induced by tissue culture. Among them, RAPDs (random amplified polymorphic DNAs) and AFLP (amplified fragment length polymorphism) have resulted both in the detection of variation as well as the lack of variation. In *Quercus serrata* (Thakur *et al.*, 1999), *Q. suber* (Gallego *et al.*, 1997) and *Q. robur* (Sanchez *et al.*, 2003) somatic embryos (SEs) have been monitored with regard to their genetic stability using RAPD markers. These studies exhibited no aberration within the RAPD banding pattern among the samples tested. No differences in AFLP patterns could be detected among SEs of immature zygotic embryos of *Q. suber*, whereas SEs derived from leaves of a mature cork oak tree showed AFLP polymorphisms (Hornero *et al.*, 2001). In contrast to RAPDs and AFLPS, simple sequence repeats (SSRs) or microsatellites are co-dominant markers and show a high rate of mutability in all of the species studied to date (Lagercrantz *et al.*, 1993). Due to their relatively high degree of variability, microsatellite markers have been widely used for assessing genetic diversity (Powell *et al.*, 1996), but their application in the detection of genetic instability induced during in vitro culture has been reported only for a limited number of plant species, such as SEs of Norway spruce, *Picea abies* (Burg *et al.*, 1999), microcuttings of trembling aspen, *Populus tremuloides* (Rahmann and Rajora 2001) and adventitious regenerants of kiwi fruit, *Actinidia deliciosa* A. Chev. (Palombi and Damiano 2002).

For several years we have been establishing a system of somatic embryogenesis for pedunculate oak (*Q. robur* L.) with the aim of demonstrating the potential use of this vegetative propagation technique for tree improvement programmes (Wilhelm 2000) as well as for studying aspects of tree physiology (Sunderlikova and Wilhelm 2002; Prewen *et al.*, 2004). In the course of the study reported here, we evaluated the applicability of SSRs for assessing genetic variability in SEs and derived plantlets of pedunculate oak (*Q. robur* L.). Comparisons were made among and within three embryogenic culture lines and the plants derived from them.

2 Materials and methods

2.1 Plant material

Two embryogenic culture lines (E4.31H9; P28H9) of pedunculate oak (*Quercus robur* L.) were generated from zygotic embryo explants in 1995 and 1997 and maintained as described (Endemann and Wilhelm 1999). The culture line C++2P was obtained from Ireland (David Thompson, Coillte, Ireland), while the others were of Austrian origin. Briefly stated, excised immature zygotic embryos or longitudinal sections of small acorns were placed on an induction medium composed of P24 medium (Teasdale 1992) supplemented with 5 μ M 2,4-dichlorophenoxyacetic acid and 0.5 μ M 6-benzyladenine (BA), 3% sucrose and 0.8% agar (Daishin) for 3 weeks. The cultures were subsequently maintained on P24 medium with 0.9 μ M BA at $24\pm 1^\circ\text{C}$ under a 16/8-h (light/dark) photoperiod with light provided at white light (TLD 36W/33; Philips, The Netherlands) at an intensity of 65 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Under these conditions the cultures remained repetitively embryogenic.

In order to test variation within SE culture lines, we removed 24 individual SEs at the cotyledonary stage (length: greater than 5 mm) from each one of the three SE culture lines. The screening period took place in 2000. Leaf tissue from the regenerated plantlets of the oak lines that had been grown in the greenhouse was used to test the genetic profile of the regenerated plants.

2.2 DNA isolation and quantification

Plant material from both single SEs and the leaf tissue (30–50 mg) was cooled in liquid nitrogen and disrupted using a grinding mill (model MM301; Retsch, Haan, Germany). DNA was extracted using the DNeasy Plant Mini kit (Qiagen, Valencia, Calif.). For eluting the purified DNA from the DNeasy spin columns we used 50 μ l AE buffer (Qiagen) for the SE tissue and 100 μ l of the buffer for the leaf material.

2.3 Microsatellite analysis

Five polymorphic microsatellite loci isolated from *Q. petraea* and *Q. macrocarpa* (Dow *et al.*, 1995; Steinkellner *et al.*, 1997) and previously mapped on a linkage map for *Q. robur* L. (Barreneche *et al.*, 1998) were selected—QpZAG36, located on linkage group 2; MSQ4, on linkage group 4; MSQ13, on linkage group 6; QpZAG9 and QpZAG1/5, on linkage group 7—separated by 23 cM. PCR amplifications were performed using HotStar Taq polymerase (Qiagen) following standard protocols. MgCl_2 concentrations had to be set for each locus (Table 1).

Amplifications were performed on PTC 100 thermocyclers (MJ Research, Waltham, Mass.) using the following amplification profile: one cycle of 15 min at 94°C , 35 cycles of 45 s at 94°C , 45 s at the primer T_m and 45 s at 72°C and one 10-min cycle at 72°C . Forward primers were fluorescently labeled for automated fragment detection using an ABI 373 DNA Sequence Analyzer (PE Applied Biosystems, Foster City, Calif.). Fragment sizes were calibrated using the genescan-

Rox 350 molecular weight standard (Perkin Elmer), and data scored and compiled using genotyper ver. 2.5 (PE Biosystems). The genescan and genotyper software provides estimates of fragment sizes reliable to two decimal places. The distribution of fragment sizes for each microsatellite locus was not continuous but displayed “natural” discontinuities or breaks which were used to define sets of bins, i.e. alleles. For example, fragment sizes 159.80–160.54 bp might be designated as allele “160”, in the case that relatively large breaks separate them from neighbouring bins. We applied two criteria in defining bins: (1) the range of fragment sizes within a bin should not exceed 1 bp; (2) the gaps between bins should be substantially greater than the gaps between fragment sizes within bins. Apart from a few exceptions for which some judgment was required, the fragment sizes sorted naturally into well-defined bins or alleles.

2.4 Determination of the SSR allele frequencies

Q. robur L. is a diploid species ($2n=2x=24$), and a high degree of heterozygosity at many loci from zygotic embryos can be expected. In particular, using microsatellite markers QpZAG1/5, QpZAG9 and QpZAG36, Lexer *et al.*, (2001) calculated the expected heterozygosity to be between 0.82 and 0.85 from 36 oak seedlings originating from one population. Since the cell lines had been maintained in tissue culture for up to 5 years, the genotype of the starting material (zygotic embryo) was no longer available. Therefore, the most frequent genotype was assumed to represent the initial genotype from the beginning of the cell culture. The allele sizes of the putative initial genotypes were thus used as references to ascertain whether a sample contained altered alleles. The allelic variation frequency was calculated by dividing the number of alleles of the mutated size by the total number of alleles determined for the particular locus.

3 Results and discussion

The genetic variability of three embryogenic cell lines of oak (*Q. robur* L.) along with the regenerated plantlets was assessed using five microsatellite markers (QpZAG1/5, QpZAG9, QpZAG36, MSQ4 and MSQ13). All of the primer pairs produced polymorphic banding patterns with primer pair QpZAG9 producing the maximum number—five initial alleles plus three additional alleles—and primer pair QpZAG1/5 yielding the minimum number—three alleles (Table 1). The average number of alleles was calculated to be 4.8 per locus. Microsatellite DNA variation was observed among SEs within all lines at two SSR loci (QpZAG9 and MSQ13), whereas no variation was found among the regenerated plants within the clonal lines (Tables 2, 3). The five loci analysed revealed 35 spontaneous variations affecting 42 alleles in 35 SEs, while the remainder of the SEs analysed belonged to non-altered genotypes. This corresponds to a variation rate of one per embryo and 0.2 variations per locus within the altered SEs (35 variant alleles/35 variant SEs/five loci), and of 0.05 (35 variant alleles/144 alleles analysed/five loci), as compared to all of the SEs

analysed. It appears that microsatellite instability is a frequent phenomenon among oak SEs. Thus, 48.6% (35 of 72 SEs) of the SEs studied possessed variable microsatellite alleles. All of the altered embryos showed variation in only one locus (Tables 2, 3). In most cases the allelic fragment decreased in length, but increases were also observed (Table 2). The altered genotypes could be explained by single mutations following the stepwise mutation model (Zhu *et al.*, 2000).

3.1 Genetic stability of the oak cell lines

The instability of SEs within the cell lines varied from 29.2% (line E4.31H9) to 62.5% (line C++2P), indicating a genotypic effect (Tables 2, 3). Genotypic dependency of somaclonal variation has been shown among SEs from other broad-leaved tree species such as *Coffea arabica* (Etienne and Bertrand 2001) and *Prunus persica* (Hashimi *et al.*, 1997). AFLP analysis of pecan SEs originating from the same line also resulted in polymorphic patterns, and no relationship between the occurrence of embryos exhibiting a greater degree of genetic diversity and culture age could be found (Vendrame *et al.*, 1999).

Considerable variability in the sizes of the microsatellite alleles was found among SEs, but no change from the putative initial genotype was detected among the regenerated plants. To explain these observations two hypotheses can be put forward. The first is that the variation among SEs is pre-existing or due to de novo induced mutational processes. Pre-existing variation among tissues from the donor plant (Hornero *et al.*, 2001) was attributed to DNA instability detected among AFLPs in cork oak SEs initiated from leaves of adult trees. Pre-existing variation may be also found in the multicellular origin of the SE culture derived from several fertilized ovules. The ovary of the oak flower consists of three locules with two pendulous ovules in each locule. Only one ovule per ovary gives rise to a seed, whereas all ovules may be fertilized. In the present investigation we explanted single excised immature zygotic embryos or longitudinally halved acorns to induction medium to initiate somatic embryogenesis. In vitro conditions may have facilitated their outgrowth, resulting in a mixture of several genotypes. The putative initial genotypes was easily determined for all three culture lines, P28H9, E4.31H9 and C++2P, and verifies that these lines had been derived from one single zygotic embryo explant. In addition, the regenerated plantlets display the putative initial genotypes, therefore we exclude the hypothesis that variation among SEs is pre-existing.

The second possibility is de novo-induced mutations, which have been observed during differentiation-de-differentiation processes in both in vivo and in vitro cell and tissue cultures (Jain *et al.*, 1998; Rani and Raina 2000). Significant DNA-sequence rearrangements were detected during tissue culture processes in carrot (Arnholdt-Schmitt 1995), tomato (Bogani *et al.*, 2001) and cauliflower (Leroy *et al.*, 2001), which the authors discussed within the framework of physiological changes resulting either from the hormonal composition of the media used or from an undifferentiated cell step. Arnholdt-Schmitt (1995) assumed that during the growth and de-differentiation

of cultured tissues, plant genomes might undergo significant modifications, which can be considered as physiological instability. Physiological DNA variability as a sign of genome plasticity may be an inherent and reversible event but may preferentially provide an opportunity for stress-induced heritable changes in the genome. The physiological genome instability caused by the in vitro culture may explain why no variability was detected among the regenerated plants. Similar results were observed for embryogenic tissue of white spruce (*Picea glauca*) (DeVerno *et al.*, 1999), in which case different RAPD fragment patterns as well as altered mitochondrial DNA RFLP patterns were found in *Larix* embryogenic tissue but not in the regenerated plants (DeVerno *et al.*, 1994). These authors suggested that the somatic embryogenic tissue may be composed of a mixture of genetically altered and unaltered cells and altered genomes, with only the unaltered genomes having the capacity to regenerate into plants. Instability induced by in vitro culture was also detected in cauliflower (*Brassica oleracea*) calli based on an inter-simple sequence repeat (ISSR)-PCR approach (Leroy *et al.*, 2001). Similar to our results, no genetic differences were reported for plants regenerated from cauliflower by means of somatic embryogenesis.

3.2 Genetic stability of the microsatellite loci

With respect to allelic instability (Table 4), 42 of 720 alleles (5.8%) from the oak SEs revealed variations, which corresponds to a mutation rate of 4.0% (42/1,055 alleles) when the regenerated plants are included. The five loci showed different degrees of variability, indicating that a specific locus is particularly susceptible to DNA rearrangements (Tables 3, 4). Loci QpZAG1/5, QpZAG 36 and MSQ4 proved to be the more stable ones, with no alleles being altered within all cell lines studied, while locus QpZAG9 showed the highest instability, with 41 allelic variations (28.5%; Table 4). The variability detected with SSR markers in our oak embryogenic lines is much higher than those measured with AFLP and RAPD markers in other oak cell lines in which no variations were found (Hornero *et al.*, 2001; Sanchez *et al.*, 2003). It is possible that a variation occurred at an early stage of embryogenesis because some of the SEs carry the same variation. In addition, we also found different variations in embryogenic cells. Thus, independent instability events must have taken place (Table 2).

The presence of loci with a high rate of mutation promoted by in vitro conditions has also been found in other species (Burg *et al.*, 1999; Linacero *et al.*, 2000; Leroy *et al.*, 2001). It was assumed that some of the DNA instability detected by RAPDs in rye plants regenerated from immature embryos and inflorescences could be attributed to hypervariable regions of the rye genome (Linacero *et al.*, 2000). Hypervariable sequences are in many cases repetitive or highly repeated regions such as microsatellites. Single-strand slippage replication effects have been proposed as one of the main mutational mechanisms accounting for the high mutation rates observed in microsatellite loci (Levinson and Gutman 1987). Plant

tissues in vitro are exposed to high levels of oxidative stress (Cassells and Curry 2001) and reactive oxygen species (ROS) are known to cause DNA damage, including microsatellite instability (Jackson *et al.*, 1998). Stress activation of plant retrotransposons is a factor in somaclonal variation (Hirochika 1993; Hirochika *et al.*, 1996). Several long terminal repeat sequences of Ty1-copia retrotransposon elements have been identified in oak as markers for population studies (Berenyi *et al.*, personal communication). *Q. robur* L. has the largest physical genome among the woody species investigated to date, but the size of its genome is not greater than that of tomato (Barreneche *et al.*, 1998). Although pedunculate oak is a diploid species, B chromosomes (Besendorfer *et al.*, 1996) and also fertile triploid individuals have been reported (Naujoks *et al.*, 1995). Increased genetic flexibility might facilitate evolutionary adaptation of plant populations to stressful environments, such as in the case of rapidly dividing oak cells in tissue culture conditions.

We have been able to show that de novo DNA mutations occurred during oak somatic embryogenesis but that the altered SEs were unable to regenerate. Therefore, we conclude that microsatellites are highly useful and sensitive markers suitable for monitoring the quality of SEs at early developmental stages in order to regulate the tissue culture procedure.

Acknowledgements The work was supported financially by the Austrian Ministry of Agriculture, Forestry Environment and Water Management.

References

- Arnholdt-Schmitt B (1995) Physiological aspects of genome variability in tissue culture. II. Growth phase-dependent quantitative variability of repetitive *Bst*NI fragments of primary cultures of *Daucus carota* L. *Theor Appl Genet* 91:816–823
- Barreneche T, Bodenes C, Lexer C, Trontin JF, Fluch S, Streiff C, Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glössl J, Kremer A (1998) A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. *Theor Appl Genet* 97:1090–1103
- Besendorfer V, Zoldos V, Peskan T, Krsnik-Rasol M, Littvay T, Papes D (1996) Identification of potential cytogenetical and biochemical markers in bioindication of common oak forests. *Phyton* 36:139–146
- Bogani P, Simoni A, Lio P, Germinarlo A, Buiatti M (2001) Molecular variation in plant cell populations evolving in vitro in different physiological contexts. *Genome* 44:549–558
- Burg K, Hristoforoglu K, Fluch S, Hohl K, Burg A, Schmidt J (1999) Analysis of the fidelity of DNA replication in embryogenic cultures of Norway spruce. In: Espinel S, Ritter E (eds) *Proc Applic Biotechnol For Genet. Biofor 99*, Vitoria-Gasteiz, pp 231–235
- Cassells AC, Curry RF (2001) Oxidative stress and physiological, epigenetic and genetic variability in plant tissue culture: implications for micropropagators and genetic engineers. *Plant Cell Tissue Organ Cult* 64:145–157
- DeVerno LL, Charest PJ, Bonen L (1994) Mitochondrial DNA variation in somatic embryogenic cultures of *Larix*. *Theor Appl Genet* 88:727–732
- DeVerno LL, Park YS, Bonga JM, Barret JD (1999) Somaclonal variation in cryopreserved embryogenic clones of white spruce (*Picea glauca* (Moench) Voss.). *Plant Cell Rep* 18:948–953
- Dow BD, Ashley MV, Howe HF (1995) Characterization of highly variable (GA/CT)_n microsatellites in the bur oak, *Quercus macrocarpa*. *Theor Appl Genet* 1:137–141

- Endemann M, Wilhelm E (1999) Factors influencing the induction and viability of somatic embryos of *Quercus robur* L. *Biol Plant* 42:499–504
- Etienne H, Bertrand B (2001) Trueness-to type and agronomic characteristics of *Coffea arabica* trees micropropagated by the embryogenic cell suspension technique. *Tree Physiol* 21:1031–1038
- Fourré JL, Berger P, Niquet L, André P (1997) Somatic embryogenesis and somaclonal variation in Norway spruce: morphogenetic, cytogenetic and molecular approaches. *Theor Appl Genet* 94:159–169
- Gallego FJ, Martinez I, Celestino J, Toribio M (1997) Testing somaclonal variation using RAPDs in *Quercus suber* L. somatic embryos. *Int J Plant Sci* 158:563–567
- Hashimi G, Huettel R, Meyer R, Krusberg L, Hammerschlag F (1997) RAPD analysis of somaclonal variants derived from embryo callus cultures of peach. *Plant Cell Rep* 16:624–627
- Hirochika H (1993) Activation of tobacco retrotransposons during tissue culture. *EMBO J* 12:2521–2528
- Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. *Proc Natl Acad Sci USA* 93:7783–7788
- Hornero J, Martinez I, Celestino C, Gallego FJ, Torres V, Toribio M (2001) Early checking of genetic stability of cork oak somatic embryos by AFLP analysis. *Int J Plant Sci* 162:827–833
- Isabel N, Boivin R, Lefasseur C, Charest PM, Bousquet J, Tremblay FM (1996) Occurrence of somaclonal variation among somatic embryo-derived white spruces (*Picea glauca*, Pinaceae). *Am J Bot* 83:1121–1130
- Jackson AL, Chen R, Loeb LA (1998) Induction of microsatellite instability by oxidative DNA damage. *Proc Natl Acad Sci USA* 95:12468–12473
- Jain SM, Brar DS, Ahloowalia B (1998) Somaclonal variation and induced mutation in crop improvement. Kluwer, Dordrecht
- Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res* 21:1111–1115
- Larkin PJ, Scowcroft WR (1981) Somaclonal variation—a novel source of variability from cell cultures for plant improvement. *Theor Appl Genet* 60:167–214
- Leroy XJ, Leon K, Hily JM, Chaumail P, Benchard M (2001) Detection of in vitro culture-induced instability through inter-simple sequence repeat analysis. *Theor Appl Genet* 102:885–891
- Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4:203–221
- Lexer C, Heinze B, Gerber S, Macalka-Kampfer S, Steinkellner H, Ziegenhagen B, Kremer A, Glössl J (2001) Microsatellite analysis of small anonymous seedlot samples from pedunculate oak (*Quercus robur*): a promising approach to monitor the number of different seed parents and pollen donors. In: Müller-Starck G, Schubert R (eds) Genetic response of forest systems to changing environmental conditions. *Forest science*, vol 70. Kluwer, Dordrecht, pp 239–250
- Linacero R, Freitas Alves A, Vazquez AM (2000) Hot spots of DNA instability revealed through the study of somaclonal variation in rye. *Theor Appl Genet* 100:506–511
- Naujoks G, Hertel H, Ewald D (1995) Characterization and propagation of an adult triploid pedunculate oak (*Quercus robur* L.). *Silvae Genet* 44:282–286
- Palombi MA, Damiano C (2002) Comparison between RAPD and SSR molecular markers in detecting genetic variation in kiwifruit (*Actinidia deliciosa* A. Chev). *Plant Cell Rep* 20:1061–1066
- Powell W, Gordon C, Machray J, Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1:215–222
- Prewin C, Vagner M, Wilhelm E (2004) Changes in water status and proline and abscisic acid concentrations in developing somatic embryos of pedunculate oak (*Quercus robur* L.) during maturation and germination. *Tree Physiol* 24:1251–1257
- Rahmann MH, Rajora OP (2001) Microsatellite DNA somaclonal variation in micropropagated trembling aspen (*Populus tremuloides*). *Plant Cell Rep* 20:531–536

- Rani V, Raina SN (2000) Genetic fidelity of organized meristem-derived micropropagated plants: a critical reappraisal. *In Vitro Cell Dev Biol Plant* 36:319–330
- Sanchez MC, Martinez MT, Valladares S, Vieitez AM (2003) Maturation and germination of oak somatic embryos originated from stem explants: RAPD markers for genetic analysis of regenerants. *J Plant Physiol* 160:699–707
- Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A, Burg K, Glössl J (1997) Identification and characterization of (GA/CT)_n—microsatellite loci from *Quercus petraea*. *Plant Mol Biol* 33:1093–1096
- Sunderlikova V, Wilhelm E (2002) High accumulation of legumin and *lea*-like mRNAs during maturation is associated with increased conversion frequency of somatic embryos from pedunculate oak (*Quercus robur* L.). *Protoplasma* 220:97–103
- Teasdale R (1992) Formulation of plant culture media and applications therefore. International Publ No.: Wo 92/07460. Forbio PTY, Queensland. Patent No. Europe: 92902531.0
- Thakur RC, Gota S, Ishii K, Jain SM (1999) Monitoring genetic stability in *Quercus serrata* Thunb. somatic embryogenesis using RAPD markers, *J For Res* 4:157–160
- Vendrame WA, Kochert G, Wetzstein HY (1999) AFLP analysis of variation in pecan somatic embryos. *Plant Cell Rep* 18:853–857
- Wilhelm E (2000) Somatic embryogenesis in oak (*Quercus* spp.). *In Vitro Cell Dev Biol Plant* 36:349–357
- Zhu Y, Strassmann E, Queller DC (2000) Insertions, substitutions, and the origin of microsatellites. *Genet Res* 76:227–236

Table 1 Microsatellite primer sequences developed for *Quercus petraea* and *Q. macrocarpa* (Dow *et al.*, 1995; Steinkellner *et al.*, 1997) with the PCR conditions, length of alleles and number (n) of initial and additional alleles

Locus and microsatellite sequence	Forward/reverse primers (5'–3')	Annealing temperature (°C)	MgCl ₂ (mM)	Length of alleles (bp)	Initial alleles (n)	Additional alleles (n)
ssrQpZAG9 (AG) _n	GCA ATT ACA GGC TAG GCT GG/GTC TGG ACC TAG CCC TCA TG	65	1.75	185–199	5	3
ssrQpZAG36 (AG) _n	GAT CAA ATT TGG AAT ATT AAG AGA G/ACT GTG GTG GTA GTC TAA CAT GTA G	52	1.75	213–224	5	0
ssrQpZAG1/5 (AG) _n	GCT TGA GAG TTG AGA TTT GT/GCA ACA CCC TTT AAC TAC CA	55	2.25	161–170	3	0
MSQ4 (AG) _n	CCC TTC TGA CAT TGC ATA TTC GA/CCA ATT CGA CAA TTT CTT AGT GCA	51	1.75	205–215	4	0
MSQ13 (TC) _n	TGG CTG CAC CTA TGG CTC TTA G/ACA CTC AGA CCC ACC ATT TTT CC	52	1.75	190–230	4	0
					Σ ₂₁	Σ ₃

Table 2 Allele sizes, number of mutated samples/alleles at five SSR loci of three embryogenic oak (*Q. robur* L.) lines and derived plantlets thereof

SSR locus	QpZAG9				QpZAG36				QpZAG1/5				MSQ4				MSQ13			
Cell line	Number of samples	Putative initial genotype, allele size (bp)	Number of mutated samples/alleles	Allele size (bp)	Number of samples	Putative initial genotype, allele size (bp)	Number of mutated samples/alleles	Allele size (bp)	Number of samples	Putative initial genotype, allele size (bp)	Number of mutated samples/alleles	Allele size (bp)	Number of samples	Putative initial genotype, allele size (bp)	Number of mutated samples/alleles	Allele size (bp)	Number of samples	Putative initial genotype, allele size (bp)	Number of mutated samples/alleles	Allele size (bp)
P28H9	24	187/197	6/6	187/195	24	213/221	0/0	–	24	168/170	0/0	–	24	207/213	0/0	–	24	218/224	0/0	–
			5/5	185/197																
			2/4	185/195																
C++2P	24	186/196	6/6	186/194	24	219/224	0/0	–	24	168/170	0/0	–	24	205/205	0/0	–	24	190/224	0/0	–
			5/10	184/194																
			4/4	184/196																
E4.31H9	24	187/199	3/3	199/199	24	215/224	0/0	–	24	161/170	0/0	–	24	215/215	0/0	–	24	218/230	1/1	218/218
			2/2	187/197																
			1/1	185/199																
Regenerated plantlets																				
P28H9	24	187/197	0/0	–	24	213/221	0/0	–	24	168/170	0/0	–	24	207/213	0/0	–	24	218/224	0/0	–
C++2P	32	186/196	0/0	–	32	219/224	0/0	–	32	168/170	0/0	–	32	205/205	0/0	–	32	190/224	0/0	–
E4.31H9	11	187/199	0/0	–	11	215/224	0/0	–	11	161/170	0/0	–	11	215/215	0/0	–	8	218/230	0/0	–

Table 3 Genotype variation frequency (%) of embryogenic cell lines and number of samples (*n*) of regenerated plantlets of *Quercus robur* L. analysed

Cell line	Total			P28H9			C++2P			E4.31H9		
	Σ	Embryos (<i>n</i>)	Plants (<i>n</i>)	Σ	Embryos (<i>n</i>)	Plants (<i>n</i>)	Σ	Embryos (<i>n</i>)	Plants (<i>n</i>)	Σ	Embryos (<i>n</i>)	Plants (<i>n</i>)
Genotypes analysed	139	72	67	48	24	24	56	24	32	35	24	11
Genotypes altered	35	35	0	13	13	0	15	15	0	7	7	0
Variation frequency (%)	25.2	48.6	0.0	27.1	54.2	0.0	26.8	62.5	0.0	20.0	29.2	0.0

Table 4 Total number of alleles analysed per locus, number of altered alleles (*n*), variability (%) and number of new alleles (+) of five microsatellite loci in three oak (*Q. robur* L.) somatic embryogenic cell lines

Cell lines	Alleles analysed per locus	QpZAG1/5			QpZAG36			QpZAG9			MSQ4			MSQ13			Average per cell line		
	Σ	<i>n</i>	%	+	<i>n</i>	%	+	<i>n</i>	%	+	<i>n</i>	%	+	<i>n</i>	%	+	<i>n</i>	%	Σ +
P28H9	48	0	0.0	–	0	0.0	–	15	31.3	2	0	0.0	–	0	0.0	–	15	6.3	2
C++2P	48	0	0.0	–	0	0.0	–	20	41.7	2	0	0.0	–	0	0.0	–	20	8.3	2
E4.31H9	48	0	0.0	–	0	0.0	–	6	12.5	3	0	0.0	–	1	2.1	1	7	2.9	4
Average	48		0.0			0.0			28.5			0.0			0.7		14	5.8	2.6



2.1.4 SSR isolation for *Paeonia*

Wild Tree Peonies (*Paeonia* sect. *Moutan*) are endemic to the mountainous regions of China and Tibet and became more and more popular in Europe from the end of the 18th century. Due to their woody stems they are able to resist extreme temperatures. However the over collection of their roots for medical purposes and the cultivation as garden flowers has almost caused the complete destruction of wild tree peony sites. Crossings were conducted over centuries, producing many new hybrids where little or no information on the crossing partners is available, leaving many of the new varieties in a state of 'unknown origin'. At the Botanical Garden (HBLFA) of Schönbrunn/Vienna, a unique tree peony collection is available, which has been established over 55 years ago. Nothing is known on the origine of these plants. Some of the plants exhibit a rather unique phenotype which can not been found elsewhere. Like in other collections, crossings have been made, and seedlings have been grown, but also here, the records are scarce, so seedlings can not be traced back to their mother plants.

To clarify the genetic relationship of the individuals we have developed a microsatellite marker system by isolating microsatellite regions (SSRs, simple sequence repeats). Starting from plant material derived from tissue culture, genomic DNA has been extracted, and fragments containing SSR regions were isolated using an enrichment method for [(AG)₁₀, (CA)₁₀ and (TA)₁₀] containing sequences.

Thirteen primer pairs have been designed for SSR containing clones and their functionality has been tested on 77 different *Paeonia* samples. Three polymorphic loci have been identified and applied to generate a phylogenetic tree for 70 plants. Relationships between the genotypes have been shown as well as changes caused by tissue culture and the imperfectness of present taxonomy.

2.1.4.1 Microsatellite markers in tree peony, *Paeonia suffruticosa* (Paeoniaceae)

Andreas Homolka, Maria Berenyi, Kornel Burg, Dieter Kopecky and Silvia Fluch²

²Health & Environment Department –Bioresources, AIT Austrian Institute of Technology GmbH, 2444 Seibersdorf, Austria

²Author for correspondence: silvia.fluch@ait.ac.at

Submitted to American Journal of Botany *Primer Notes & Protocols* 20.04.2010

The authors thank the HBLFA Schönbrunn and Dr. Hristoforoglu, the Botanical Garden of the University of Vienna and Alpengarten Belvedere for providing *Paeonia* samples.

ABSTRACT

- *Premise of the study:* Microsatellite primers were developed in tree peony, *Paeonia suffruticosa*, to perform paternity tests as well as variety assignments in special Austrian collections.
- *Methods and results:* Using SSR enriched libraries and EST-mining, 8 polymorphic primer sets were identified in Austrian collections of *Paeonia* sect. Moutan DC. The primers amplified di- and trinucleotide repeats with 2-10 alleles per locus. All primers also amplified in *P. ostii*, *P. pontaninii* var *trolloides*, *P. delavaii* and *P. lutea* and in the herbaceous species *P. peregrina* and *P. tenuifolia* (*Paeonia* Sect. Paeon).
- *Conclusions:* These results show the usefulness of primers in *P. suffruticosa* for population genetic studies and their ability to cross amplify in related taxa across the genus.

Key words: EST-mining; enriched library; microsatellite; Paeoniaceae, *Paeonia*; population genetics

INTRODUCTION

Tree peonies (*Paeonia* L. sect. Moutan DC.) develop woody stems resistant to extreme differences in temperature. Due to plant breeding, there exist hundreds of ornamental hybrids (Rieck and Hertle, 2002). Endemic to China and playing an important role as national flower, tree peonies were first discovered by Europeans in 1683 and imported to Europe in 1789. In order to perform paternity tests as well as variety assignment in a set of *Paeonia* ssp. plants available in a special Austrian ornamental cultivar collection at Schönbrunn, a SSR marker system needed to be developed. The collection consists of 17 *Paeonia* cultivars each over 60 years old and of unknown origin but similar to *Paeonia suffruticosa* in morphology, and 23 intraspecific hybrids produced from them over the last decades. We created a *P.*

suffruticosa sample set of 18 individuals by combining the 17 cultivars and one *P. suffruticosa* species from the Botanical Garden of Vienna. From the hybrids, four individuals were selected for the microsatellite analysis. These plants exhibit very special leaf morphology or floral types which are of special interest for further breeding. Additional three individuals from a 'gardener's collection' of unknown origin were added to the hybrid set (N=7). In addition to these plants, a reference set (N=7) of five different *Paeonia* sect. Moutan species (*P. ostii*, *P. pontaninii* var *trolloides*, 2x *P. delavaii* and *P. lutea*) and two *Paeonia* sect. Paeon species (*P. tenuifolia* and *P. peregrina*) that could be used for subsequent species assignment of the plant material in question was also examined. These species were greenhouse grown and are cultivated in botanical gardens in Vienna.

Microsatellites are known under several pseudonyms, e.g. simple sequence repeats (SSRs), simple sequence length polymorphisms (SSLPs), short tandem repeats (STRs), simple sequence motifs (SSMs) or sequence target microsatellites (STMs), and serve as highly variable marker system from diversity studies to paternity tests (Menz *et al.*, 2004). In order to identify variable microsatellite regions for *Paeonia*, we pursued two approaches – we used a SSR enrichment method according to Refseth *et al.*, (1997), and we also incorporated an in silico approach mining publicly available EST sequences for identification of SSR regions.

METHODS AND RESULTS

20 µg genomic DNA from *Paeonia suffruticosa* tissue culture material (HBLFA Schönbrunn) extracted with DNeasy kit (QIAGEN, Düsseldorf, Germany) were digested with 20 U *Sau3A* according to Refseth *et al.*, (1997) and fragments between 1.7 and 0.9 kb were isolated. *Sau3A* adaptors [A: 5'TTCCATAACCCCAATCTTCT3' and B: 5'TATCCTAATCTTTTTTCTCTTTCC3'] were ligated to the ends of the fragments. 20 ng DNA adaptor ligated fragments were amplified in the presence of 50 pmol *Sau3A* oligo A primer and 2 U *DyNazym II*. One µg of each of the amplified fragments was hybridized with 125 pmol 5' biotinylated oligos (AT)₁₀, (CA)₁₀ and (TA)₁₀ respectively. Streptavidin coated Dynabeads (approx. 3.75x10⁸ beads) were used to capture SSR containing fragments. Captured fragments were released from the biotinylated oligos by heat-denaturation. Isolated fragments were amplified with *Sau3A* oligo A and resulting amplicons were cloned into pCR®4-TOPO® vector (INVITROGEN, Carlsbad, California, USA). 1728 clones were analyzed using M13 primers to amplify the complete microsatellite-containing insert. A third primer (AT₁₀, CA₁₀ or TA₁₀) was used to specifically amplify microsatellite regions. SSR containing clones were selected as positives when two (complete insert and microsatellite) or more bands (complete insert and compound microsatellite) were visible on a 1.5 % agarose gel after PCR.

96 positive clones were sequenced using an ABI 3130xl Genetic Analyzer (ABI), and sequence analysis was performed with Sequencher 4.5 (GeneCodes

Corporation, Ann Arbor, Michigan, USA). Fifty-six clones (58% of clones sequenced) contained SSR stretches. Within the (CA)_n library, 41 clones showed SSR motifs of appropriate length, but many were redundant (38) reducing the number of valid clones to four unique sequences. Out of 30 clones picked from the (AG)_n library, 15 contained an SSR motif, nine sequences could be further used for primer planning. There was no SSR motif found in clones picked from the (TA)_n library. Thirteen primer pairs were designed using OLIGO (Molecular Biology Insights, Cascade, Colorado, USA). To evaluate polymorphism of these markers, PCR reactions were performed in 20 µl volumes containing 30 to 50 ng genomic DNA, 1X PCR Buffer (QIAGEN), 0.25 mM each dNTP, 1 pmol FAM-labelled forward primer, 1 pmol reverse primer, 1.5 mM MgCl₂, and 2 U *Taq* polymerase (QIAGEN). The PCR program consisted of 5 min. initial denaturation at 95 °C followed by 30 cycles denaturation at 95 °C for 45 sec., a primer-specific annealing-temperature (Table 1) for 30 sec., extension at 72 °C for 50 sec., and a final extension at 72 °C for 8 minutes. 13 primer pairs (Table 1) amplified in the *P. suffruticosa* set (N=18) but only three were polymorphic across the set. All primer pairs amplified in the “hybrid-” (N=7) and the “reference set” (N=7) with the same results regarding polymorphism.

To exploit the advantages of ESTs as source for marker development (Thiel, *et al.*, 2002; Rungis, *et al.*, 2007), 2,024 *Paeonia suffruticosa* ESTs from the NCBI dbEST database were used for in silico filtering of SSR regions. Screening with the SciRoKo software program (Kofler, *et al.*, 2007) revealed 726 sequences containing various repeat motifs. Clustering with EST2uni (Forment, *et al.*, 2008) produced 437 SSR containing ‘unigenes’ separating into 85 clusters and 352 singletons. 598 microsatellite motifs could be identified. Di-nucleotide repeats were most abundant (47%), followed by tri-nucleotide repeats (30%) and tetra-nucleotide repeats (11%). Penta- and hexa-nucleotide repeat frequency was low (3%). Twenty-five primer pairs designed with OLIGO were analysed according to the protocol from Schuelke (2000). Forward primers were designed with an M13 tag compatible to a FAM-labelled M13 universal primer. 30 ng DNA was amplified in 25 µl volumes with 1x PCR Buffer (Solis BioDyne, Tartu, Estonia), 0.25 mM each dNTP, 0.25 pmol forward primer, 1 pmol reverse and universal labelled primer, 2.5 mM MgCl₂ and 2 U Hot FirePol (Solis BioDyne). PCR conditions were set to 15 min. initial denaturation at 95 °C followed by 30 cycles denaturation at 95 °C for 30 sec., a primer-specific annealing-temperature (Table 1) for 45 sec., extension at 72 °C for 45 sec., followed by eight cycles of 30 sec. at 95°C, 45 sec. at 53 °C, 45 sec. at 72 °C, and a final extension at 72 °C for 10 minutes. Twelve primer pairs did not reveal an amplification product, eight were not polymorphic, and five showed varying degrees of variability (Table 1) across the *P. suffruticosa* set (N=18). These results were also observed across “hybrid-” (N=7) and the “reference set” (N=7).

Fragments derived with both methods were analyzed on an ABI 3130xl Genetic Analyzer. Sizes were extracted using PeakScanner 1.0 (Applied Biosystems, foster City, California, USA) and allele calls were performed in Excel using the summary

statistics function of PopTools – PlugIn (Hood, 2005) which produces a frequency histogram. This histogram was further used for manually verifying the alleles. Observed heterozygosity was calculated using PowerMarker (Liu and Muse, 2005). Table 2 shows the characteristics of eight polymorphic SSR-markers amplified in 32 *Paeonia* individuals from three different Austrian collections. The collections also included two individuals from *Paeonia* sect. *Paeon*. Number of alleles ranged between two and six with values for observed heterozygosity from 0 to 1. Due to the fact that we were studying collections and not natural populations, parameters like Hardy Weinberg Equilibrium and expected heterozygosity were not calculated.

CONCLUSIONS

We have isolated eight highly polymorphic microsallelite markers for *P. suffruticosa* with the ability to cross amplify in *P. ostii*, *P. pontaninii* var *triloides*, *P. delavaii*, *P. lutea* and two herbaceous species *P. tenuifolia* and *P. peregrina*. The available SSR loci could prove useful for future genetic studies not only in *P. suffruticosa* but in other members of the Paeoniaceae family. The ability of the present marker set to cross amplify in related taxa provides a good basis for ornamental peony identification.

LITERATURE CITED

- FORMENT, J., GILABERT, F., ROBLES, A., CONJERO, V., NUEZ, F. AND BLANCA, J. M. 2008. EST2uni: an open, parallel tool for automated EST analysis and database creation, with a data mining web interface and microarray expression data integration. *BMC Bioinformatics* 9: 5. DOI 10.1186/1471-2105-9-5
- HOOD, G. M. 2005. PopTools version 2.7.1. Website <http://www.cse.csiro.au/poptools>
- KOFER, R., SCHLÖTTERER, C. AND LELLEY, T. 2007. SciRoKo: A new tool for whole genome microsatellite search and investigation. *Bioinformatics* 23: 1683-1685. DOI 10.1093/bioinformatics/btm157
- LIU, K. AND MUSE, S. V. (2005) PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics* 21:2128-2129. DOI 10.1093/bioinformatics/bti282
- MENZ M. A., KLEIN R. R., UNRUH N. C., ROONEY W. L., KLEIN P. E. AND MULLET, J. E. 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. *Crop Science* 44: 1236-1244
- REFSETH U. H., FANGAN, B. M. AND JAKOBSEN, K. S. 1997. Hybridization capture of microsatellites directly from genomic DNA. *Electrophoresis* 18: 1519-1523. DOI 10.1002/elps.1150180905
- RIEK, I. AND HERTLE, F. 2002. Strauch-Pfingstrosen, 1st ed .Verlag Eugen Ulmer, Stuttgart, Germany
- RUNGIS S, D., BÉRUBÉ, Y., ZHANG, J., RALPH, S., RITLAND, C. E., ELLIS, B. E., DOUGLAS, C., BOHLMANN, J. AND RITLAND, K. 2004. Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics* 109: 1283-1294. DOI 10.1007/s00122-004-1742-5

- Schuelke M. 2000. An economic method for the fluorescent labelling of PCR fragments. *Nature Biotechnology* 18: 233-234
- THIEL, T., MICHALEK, W., VARSCHNEY, R. K. AND GRANER, A. 2002. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106: 411-22. DOI 10.1007/s00122-002-1031-0

Tables

Table 1. Characteristics of eight poly- and 18 monomorphic microsatellite primers developed in *Paeonia suffruticosa*. Shown for each primer pair are the forward and reverse sequence, repeat type, size of the original fragment (bp), annealing temperature when run individually (*T_a*) and the GenBank accession number. All values are based on 32 samples representing Austrian collections located in Vienna (N = 7-18 for each). An asterisk indicates primers developed from enriched libraries, underlined primers are polymorphic.

Primer	Sequence	Repeat	Size	T _a	GeneBank
<u>PAG1*</u>	F: AGTGGTGGAAGATTGGAC R: AAATACTCCGTCTTAGTGTGAA	(AG) ₂₄	211	55	EU678295
<u>PAG16A*</u>	F: TCAATGATTACTCTCGTAT R: TCCTTAACTCATCTTTACTC	(TC) ₁₆	322	50	EU678299
<u>PCA1*</u>	F: TAGTCAGTCGTAGCTAGCATAGGCA R: GATGGCCACCTATAGAAAAGAATCA	(GT) ₂₀	142	60	EU678305
PAG4*	F: CCTACTATTACCCAGCTGC R: CTTGAGCCACCGCTCAGCC	(AG) ₂ (AG) ₄	102	60	EU678296
PAG4C*	F: CCCTTATACTTCATTCTTGCT R: AACCTTAATTAGTGCCAACC	(TC) ₁₄ (CT) ₃	192	52	EU678297
PAG6*	F: CACTTGGAAGGGTTAAAGAGAG R: GCCAATCATGTCCTTTAATCCAT	(AG) ₁₇ (AG) ₆	205	55	EU678298
PAG22*	F: TGGGAGTAAGTCCCCCTCTCTC R: GTCTTTTTTCTCCCCTAAGC	(TC) ₃ (TC) ₆ (TC) ₄	243	55	EU678300
PAG237C*	F: TTGAATATGGGCCAAACTTTA R: GAGGGAGTCTGACGAAGATG	(TC) ₁₀	124	57	EU678301
PAG24*	F: ATTTTCATCGATGGCATATTCCC R: TCTCAATCAAAATGGGCACC	(TC) ₁₀	238	56	EU678302
PAG28*	F: ATTCCAACCCCAATATGGAGG R: ACTTGACCGCTCCTGACTTGC	(GA) ₃ (AG) ₅	139	58	EU678304
PCA19*	F: CATTGCCATGTCTACGACA R: AACTCCTATCCTGTAAATG	(GT) ₃ (GT) ₃	261	52	EU678306
PCA38A*	F: GAAGGCCAAAAAGGTATTATT R: TACTACCCAAACCCTACTACA	(GT) ₄ (GT) ₉	349	52	EU678307

SSR Isolation for Paeonia ssp

PCA46*	F: GATATGCTTGTGATAAGTG R: TCATTAAGATGGCACTAA CAA	(GT) ₁₃	124	54	EU678308
<u>AT8051</u>	F: GGTATCAATCCGTGTGC R: GCGAAAATTTAGATGAGTGT	(AT) ₅	190	50	FE528051
<u>AG8073</u>	F: TCAGCTAATATGGGTGTTTC R: ATCAAAGTGGAAGTTCTACAGT	(AG) ₁₀	226	50	FE528073
<u>ATG9706</u>	F: TGCACCCAAGGAGGAG R: CCATGATAAAAAACACCCAAAG	(ATG) ₅	182	58	FE529706
<u>ACG9797</u>	F: GGTCGCTGGGTGTTC R: GTTGGAGTTGGAGAGGTTG	(ACG) ₈	192	55	FE529797
<u>C-8918</u>	F: AAGAAGGGGAGGAGGAG R: GCGAGAAGAAAAGGCCAAAAT	(AAG) ₅ (AGG) ₄	201	53	FE528918
ACC2C1	F: TGCACCCAAGGAGGAG R: CCATGATAAAAAACACCCAAAG	(AC) ₅	244	58	FE529049
ACC8347	F: GAGGGGAGCAAGAAGAT R: TAAACAGACCAAATCCAAAG	(ACC) ₄	208	54	FE528347
AAAAG9656	F: GCAACGGAAGAAAAGAAAAG R: GTTTGGGTCCATATTACACTCT	(AAAAG) ₃	238	56	FE529656
AAATG8335	F: GGGGGAGTTCAGATGTGTGG R: CACGGCGAGGAGACGATT	(AAATG) ₃	199	50	FE528335
3AGAG9896	F: GGAATGGCGGCAACT R: TGGCAATACTACTGGACAGG	(AAA GAG) ₃	200	66	FE529896
C-C6C2	F: GGGAGAAGGGTGAGAAGAA R: TTTTATTTTCGCTTGGACA	(AAG) ₅ (AGC) ₄	197	58	FE528494
C-C78C1	F: GGAAGACGACGAAGAGG R: AATTCAAGGCCATAAAGAGA	(ACG) ₅ (ACG) ₄ (AAG) ₄	275	60	FE529660
C-C9808	F: TCTCCGTCATCTGCTGT R: TTCATCACTCTCCTCCTCT	(ACC) ₄ (AGC) ₄	165	53	FE529808

Table 2. Results of initial primer screening in collections of *Paeonia* Sect. Moutan and Paeon. Shown for each polymorphic primer pair are the number of alleles (A), and mean values of observed (*Ho*) heterozygosity. The sample size for each population is shown in parentheses.

	No. of alleles	Observed heterozygosity
Suffruticosa collection (N=18)		
PAG1	6	0.81
PAG16A	6	0.40
PCA1	6	0.81
AT8051	2	0

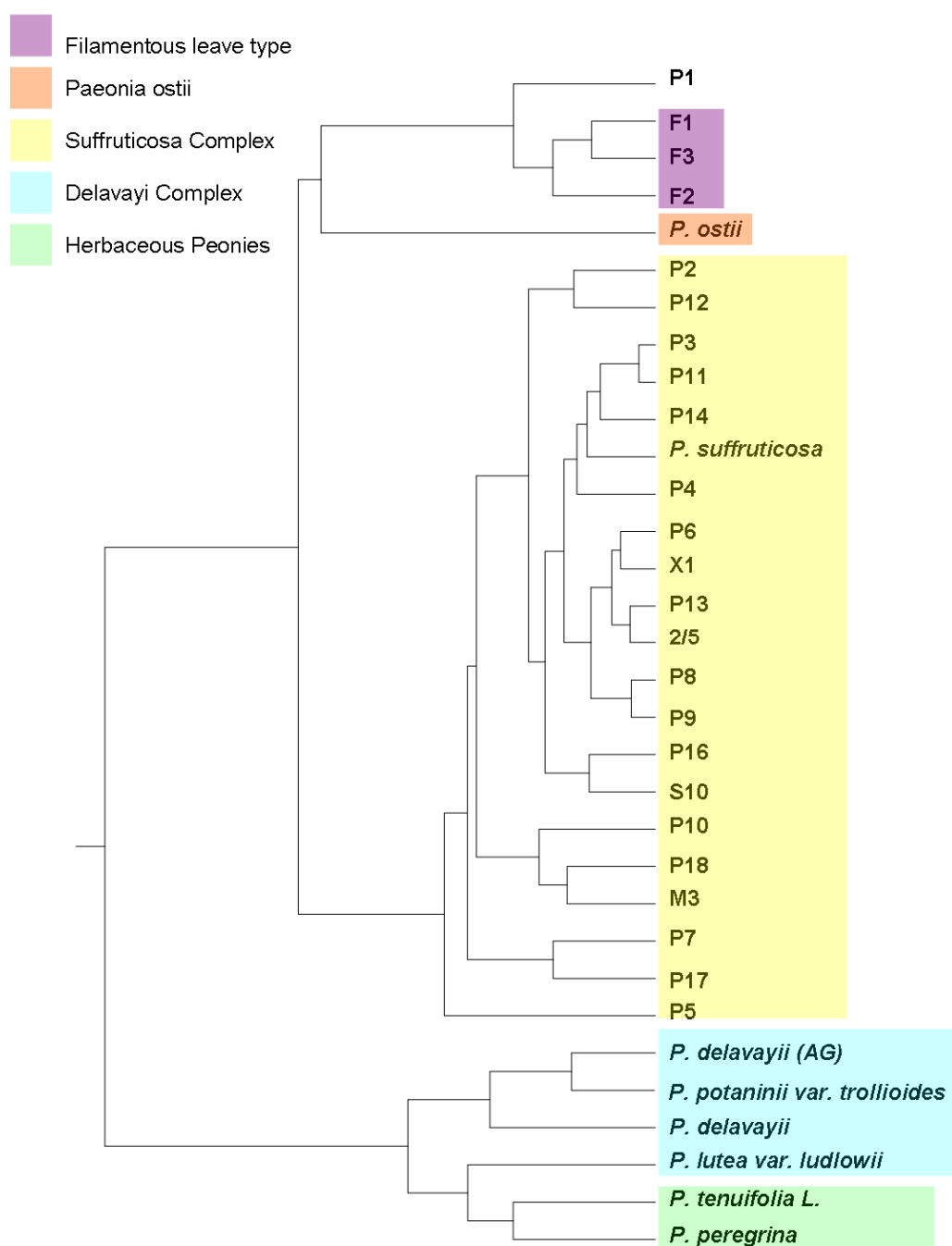
AG8073	3	0.33
ATG9706	2	0.11
ACG9797	4	0.44
C-8918	2	0.06
Hybrid collection (N=7)		
PAG1	4	0.50
PAG16A	7	0.50
PCA1	3	0.75
AT8051	2	0
AG8073	2	0
ATG9706	2	0.57
ACG9797	3	0.29
C-8918	3	0.14
Reference collection (N=7)		
PAG1	2	0
PAG16A	5	0.80
PCA1	3	0
AT8051	5	0.29
AG8073	5	1.00
ATG9706	2	0.14
ACG9797	4	0.29
C-8918	2	0

APPENDIX Taxa used in this study. all specimens cultivated in Botanical Gardens in Vienna. Specimens are deposited in the following herbaria: Botanical Garden. University of Vienna = BG. Alpengarten. Belvedere = AG

Taxon; *Voucher number*; Collection locale; Herbarium.

Paeonia. Suffruticosa; WU 23164; W. Austria; AG. *P. ostii*; W. Austria; AG. *P. pontaninii var trolloides*; W. Austria; AG. *P. delavaii*; WU 23162; W. Austria; AG. *P. delavaii*; WU 33864; W. Austria; BG. *P. lutea*; W. Austria; BG. *P. tenuifolia*; WU 23161; W. Austria; BG. *P. peregrina*; W. Austria; BG

Figure 1. Cladogram constructed with the UPGMA method, established with frequency based distances according to Nei 1972.



2.2 The organellar genomes

Plant cells contain different DNA containing subcellular compartments (organelles: chloroplasts and mitochondria), which serve distinct physiological functions. One of these organelles, the mitochondrion, provides most of the non-photosynthetic energy required in the cell. Mitochondria contain their own genome, which is 10–150 times larger in plants than in animals, but encodes only a limited number of vital gene products. During evolution, many mitochondrial genes have been transferred to the nuclear genome (Song *et al.*, 2008) leading to a situation that, when using mt markers in population genetic analysis, one has to test the uniparental mode of inheritance and true localization of the DNA fragments used in the analysis.

According to the endosymbiont theory not only mitochondria but also chloroplasts originate from unicellular organisms that have been acquired by the plant cells in early evolutionary developmental stages (Moustafa *et al.*, 2009). Chloroplasts serve as the green power plants of plant cells capturing light and converting CO₂ and water into carbohydrates as energy source for the plant. Like mitochondria, chloroplasts have their own DNA encoding for several tens of genes. Other than in mitochondria, the chloroplast genome is quite conserved in sequence as well as structure across all plant species. Chloroplasts as well as mitochondria show uniparental inheritance, mostly both being transferred from the mother to the progeny (F1) generation, with some percent of leakage in the cp as well as the mt genome transfer (McCauley *et al.*, 2007; Pearl *et al.*, 2009).

In plants, gene migration occurs through pollen as well as through seeds. Their impact on genetic population structure and variation differs depending on the mode of pollen dispersal (wind or insect) as well as the way seeds are dispersed. However, they are important to the adaptation and adaptability of a species (Ellstrand & Schierenbeck 2000; Kane & Rieseberg 2008; Slatkin 1987). For example, in forest management, knowledge of dispersal is of practical importance because foreign pollen introgression can reduce the genetic gain in seed orchards and breeding programs or alter the composition of open-pollinated families (Ellstrand 1992; Nason *et al.*, 1992).

Due to the uniparental inheritance of both chloroplast and mitochondrial DNA, they are a source of unique markers that are very useful for phylogenetic and population genetic studies (Karp *et al.*, 1997; Petit *et al.*, 1993; Petit *et al.*, 1997c). Both cp and mtDNA are conserved in their evolution, and evolve slower than nuclear DNA in angiosperms. Advances in the development of uni-parentally inherited markers for different plant species have opened a new way to assess the different impacts of seed and pollen dispersal. Such approaches have revolutionised the application of genetic markers to the study of both ecological and evolutionary processes. In animal studies, the development of techniques for the study of mt DNA has proved to be of particular importance, owing to the relatively rapid rate of base substitution and non-recombining maternal inheritance of the mtDNA molecules (Lynch *et al.*, 2006; Palmer 1990). These characteristics enable the construction of

phylogenetic relationships among haplotypes, adding a historical perspective to population genetic studies that traditionally were more difficult to obtain from frequency data (e.g. allozymes).

Mitochondrial DNA (mtDNA) has been less applied for phylogeographic studies in plants, because of relatively low rate of nucleotide substitution in the molecule and frequent rearrangements of the genome (Palmer 1990) leading to the fact that it isn't conserved in structure (Hanson & Folkerts 1992) which makes marker development and transfer of markers between species time consuming (Bereiter-Hahn *et al.*, 2008). Therefore, most phylogeographic studies in plants have assessed variation in chloroplast DNA (cpDNA) as it evolves slowly and exhibits little structural variation below the species level. Based on known cpDNA sequences, universal primers for cpDNA have been developed that target non-coding regions within the cp genome where mutation rates are relatively high (Ferris *et al.*, 1995a; Taberlet *et al.*, 1991). CpDNA markers have been largely used for phylogenetic inferences (Olmstead & Palmer 1994) and to some extent, for within-species genetic studies (Latta 2004; Latta & Mitton 1997; Soltis *et al.*, 1997). Comparison of patterns of variation between paternally inherited and nuclear-inherited markers allow inferences about pollen movement and paternity (Latta *et al.*, 1998; McCauley 1995).

2.3 The Chloroplast DNA – specific work

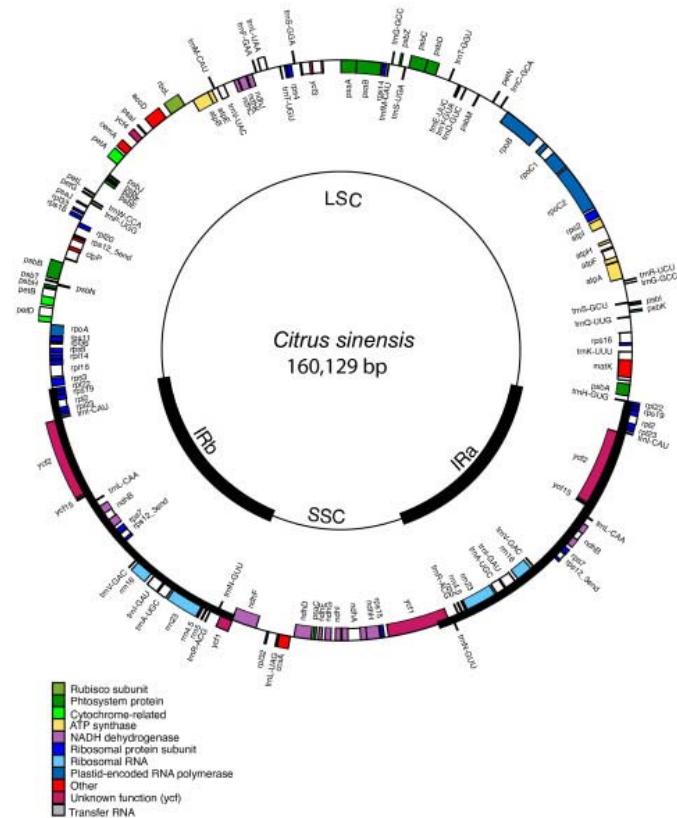


Fig 10: The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple' (from (Bausher *et al.*, 2006))

2.3.1 Introduction

Chloroplasts are organelles found in plant cells and other eukaryotic organisms that house the photosynthetic apparatus. Chloroplasts capture light energy to conserve free energy in the form of ATP and reduce NADP to NADPH through a complex set of processes called photosynthesis. In addition to photosynthesis, important metabolic activities take place within chloroplasts including the production of starch, certain amino acids and lipids, vitamins and several key aspects of sulfur and nitrogen metabolism.

Chloroplasts are thought to originate from independent living cyanobacteria (Keeling 2004) that have been acquired by the plant cells. The endosymbiotic theory was first articulated by the Russian botanist Konstantin Mereschkowski in 1905. Ever since there has been more and more prove that this hypothesis might be correct (Moustafa *et al.*, 2009).

Chloroplasts possess their own genome and a full complement of transcriptional and translation machinery to express their genetic information. In all 170 angiosperms and land plants investigated so far (Cui *et al.*, 2006), the size of the cpDNA ranges from 120–160 kb with three exceptions: (*N. accuminati* 171kb, Duckweed 180kb, *Geranium* 217kb).

All cpDNA molecules are circular. The presence of nucleic acids within chloroplasts was first reported in 1963 by Sager R and Ishida. This subsequently led to the selection of cpDNA as one of the first candidates for complete genome sequencing (Sugiura 2003). Studies of the organization and evolution of chloroplast genomes have been rapidly expanding due to the availability of the number of completely sequenced genomes published in the past decade. Comparative studies indicate that chloroplast genomes of land plants are highly conserved in both gene order and gene content (Petit *et al.*, 1997b) possessing on average 100 genes per cp genome.

Cp DNA is maternally inherited in most angiosperms, with a few exceptions like biparental inheritance in *Epilobium hirsutum*, *Medicago sativa*, *Pelargonium* and paternal inheritance in *Daucus* and *Oryza* (Reboud & Zeyl 1994). The population of chloroplast DNA in a plant species is generally homogeneous. Generally it has a highly conserved organization (Olmstead & Palmer 1994; Palmer *et al.*, 1983; Palmer & Stein 1986b), with most land plant genomes composed of a single circular chromosome with a structure that includes two copies of a long inverted repeat (IR) that separates large and small single copy regions (LSC and SSC; see Fig 1) (Bausher M. *et al.*, 2006, Daniell H *et al.*, 2006). Data from genome sequencing projects have demonstrated that fragments of the chloroplast DNA are present in both the nuclear and mitochondrial genomes (Cullis *et al.*, 2009a; Cummings *et al.*, 2003; Huang *et al.*, 2003). This DNA is still being actively transferred from the chloroplast to the nucleus in an ongoing and frequent process (Cullis *et al.*, 2009b; Huang *et al.*, 2003).

Recently, chloroplasts have caught attention by developers of genetically modified plants. In most flowering plants, chloroplasts are inherited from the female parent (Ruf *et al.*, 2007) although in plants such as pines, chloroplasts are inherited from the male parent (Powell *et al.*, 1995). Chloroplast bioengineering offers a number of advantages over nuclear transformation including high levels of transgene expression and gene containment as chloroplasts cannot be disseminated by pollen (Daniell *et al.*, 2004). Transgene integration into the chloroplast genome occurs exclusively by homologous recombination of chloroplast DNA flanking sequences. Therefore, chloroplast genome sequence analysis is crucial for identification of regions into which transgenes should be ideally integrated. This makes plastid transformation a valuable tool for the creation and cultivation of genetically modified plants that are biologically contained, thus posing significantly lower environmental risks.

2.3.2 Cp DNA marker development and application

Two DNA features of the chloroplast DNA have widely been exploited for marker applications in population genetic studies – SSRs and InDel variation in the LSC region. SSR markers in the chloroplast genome have largely been applied as informative markers for phylogeographic studies and to study conifer genetic diversity to detect spatial genetic differentiation related to the possible recolonization processes in the postglacial period (Cato & Richardson 1996; Troupin *et al.*, 2006; Vendramin *et al.*, 2000b).

Several authors have described the use of "universal" primers for PCR from chloroplast DNA, i.e., primers that amplify DNA fragments from many species in a genus, family, or even higher taxonomic groupings (Dumolin-Lapegue *et al.*, 1999b; Grivet *et al.*, 1999; Hamilton *et al.*, 1999; Taberlet *et al.*, 1991). Usually, such primers are based on conserved DNA sequence in exons. Therefore the majority of the published primers are located in the LSC region, in genes directed towards the intergenic regions or spacers which are more variable than the exons (Dumolin-Lapegue *et al.*, 1999a; Heinze 2007). Thus, using these primers one can amplify the non-coding regions via PCR for evolutionary studies and for identifying intraspecific genetic markers (Taberlet *et al.*, 1991). The PCR-RFLP as an analytical method allows detection of variation within the amplified fragments as it detects mutations mediating gain or loss of a particular restriction site without sequencing. This high resolution technique has led to the wide spread use of these markers. The technique involves the digestion of amplified PCR product by restriction enzymes, and the visualization of the products on ethidium bromide stained agarose gels or silver stained polyacrylamide gels (Karp *et al.*, 1997; Petit *et al.*, 2002). In order to ease access to sequence and primer information, a comprehensive database has been set up by Heinze B. (Heinze 2007) compiling all necessary information on primers for cpDNA investigations reported by various authors.

The chloroplast information has been utilized to investigate plant evolution as well as diversity and phylogenetic relationships between different species. Phylogenetic studies as well as population studies (Burnier *et al.*, 2009; Chappell *et al.*, 2008) have been conducted, phylogeography and post glaciation colonization of species has been assessed (Feleke *et al.*, 2006; Ferris *et al.*, 1995b; Ikeda *et al.*, 2008; Johnk & Siegmund 1997; Kim *et al.*, 2009; Marchelli *et al.*, 1998; Neale *et al.*, 1988; Petit *et al.*, 1997a), and the markers based in the LSC region of the cp genome have been used for the identification of hybrids (Kim *et al.*, 2009; Lexer *et al.*, 2004).



2.3.2.1 A Chloroplast marker system for studying genetic variation in *Populus nigra* L.

Fluch S., Krystufek V., Burg K.

Austrian Research Centers, Dept of Biotechnology, A-2444 Seibersdorf, AUSTRIA

Published in: van Dam, B.C., Bordacs, S. (Eds.), Genetic Diversity in River Populations of European Black Poplar-implications for Riparian Eco-system Management.

Proceedings of the International Poplar Symposium, Szekzard, Hungary, May 2001

Abstract

Populus nigra L. , which is a unique pioneer species of the riparian ecosystem, is widely distributed throughout Europe and Asia. Due to the loss of its natural habitat (drainage and management of rivers and riverbanks) this species has become highly endangered in its diversity. However, *P. nigra* is not only a valuable species of ecological importance, but it is also of economical relevance, since it serves as crossing partner for commercially used, highly productive hybrid poplars. Therefore, conservation strategies have to be established based on firm scientific data.

For the estimation of the existing genetic variation among natural populations of *P. nigra*, we chose a marker system based on the chloroplast DNA sequence from tobacco for primer planning. Using a subset of universal primers (B. Demesure *et al.*, 1995; S. Dumolin-Lapegue *et al.*, 1997) as well as other regions partly described in literature, we performed PCR-RFLP analysis detecting the fragments by silverstained native polyacrylamide gels. In order to clarify the nature of the detected variations, the regions of interest were cloned and sequenced. Sequence analysis showed that different types of mutations are present in the analysed fragments. Very frequently variation in polyA regions is causing the observed variation in the banding pattern. Sometimes also insertions and deletions were detected as a source for variation.

Comparing sequence data with the banding pattern produced by PCR-RFLP it turned out that with the system chosen, not only fragment length variation, but also conformational polymorphisms could be detected.

Introduction

Riparian ecosystems are valuable elements for controlling rivers hence due to extensive river regulations, these natural habitats and their most important plants are endangered. *Populus nigra* L. is one of the main trees of such systems, germinating on poor sandy soils along the riverbanks. In order to work out strategies for conservation as well as for reforestation, it is necessary to know more about the genetic diversity still existing in natural populations.

There are different ways of estimating genetic diversity on DNA level using molecular techniques. Maternally inherited chloroplast (cp)DNA proved to be an appropriate tool (Petit et al 2002) for measuring the genetic variation on a larger geographical scale. Due to the uniparental inheritance of the cp DNA, there is no recombination occurring in this genome, making it appropriate for phylogenetic studies as well as inter species hybridisation analysis.

The organisation of the organellar cpDNA is highly conserved among higher plants, offering the opportunity of applying a set of universal primers on various plant species (B. Demesure *et al.*, 1995; S. Dumolin-Lapegue *et al.*, 1997, Grivet *et al.* 2001) for amplifying relatively long variable stretches of intergenic regions by PCR. The sequence of the tobacco cp genome serves as a basis for primer planning.

Material and Methods:

For the determination of different cytotypes in the cpDNA of black poplar, universal PCR primers planned for the chloroplast genome of *Nicotiana tabacum* from different sources (Heinze 1998, Samuel *et al.*, 1997, B. Demesure *et al.*, 1995; S. Dumolin-Lapegue *et al.*, 1997, Weising 1999), were given a trial on total DNA extracted from leaf material of 35 *Populus nigra* L. individuals of the EUFORGEN core collection. In total 5 primer combinations proved successful for the amplification of different regions of the chloroplast genome (Table 1). The primers DT, TF, ORFM and 17/20 were working without any modifications and the generated PCR fragments could be used for further investigation with restriction enzymes right away. The primers CD produced a weak PCR product with high background smear. To develop a better primer pair for the CD region, the specific fragment was cloned and partly sequenced in order to plan a black poplar specific primer pair.

Table 1: Primer sequences

cpDNA locus	fragment length	Primer sequence 5'-3' fwd	Primer sequence 5'-3' rev
DT	approx. 800bp	ACC AAT TGA ACT ACA ATC CC	CTA CCA CTG AGT TAA AAG GG
CD*	approx. 2300 bp	TTC CCG GTA GCA CAT ACA CA	TTG GGC TGC TTT GAT GGT AG
TF	approx. 1500 bp	CAT TAC AAA TGC GAT GCT CT	ATT TGA ACT GGT GAC ACG AG
ORFM**	approx. 1040 bp	CTT GCT TTC CAA TTG GCT GT	CAT AAC CTT GAG GTC ACG GG
17+20***	approx. 800pb	GAA GTA GTA GGA TTG ATT CTC	CCC TAC AAC TCA TGA ATT AAG

* newly developed poplar specific primer pair

** Amplified by the primers Orf62-P and trnfM (Heinze 1998)

*** Amplified by the primers atpBsam and rbcLsam (Samuel *et al.*, 1997)

In order to detect variation within the different amplified chloroplast regions a PCR-RFLP assay was set up, testing up to 10 different restriction enzyme per region in order to select the enzymes most suitable for the subsequent population analysis (Table 2). PCR-fragment digests were subsequently analysed on 35cm long, 8% non denaturing polyacrylamide gels, run at 18 V/cm and 15°C, followed by silver staining of the DNA fragments.

Table 2: Selected fragment – enzyme combinations

Fragment	Enzymes
CD	TaqI, EcoRI
DT	MnII,
TF	TaqI, MnII
ORF-M	EcoRI,
17+20	Hinf I

Sequence analysis of some of the different cytotypes detected in the investigated samples was performed after cloning the PCR fragments (TA-cloning kit, Invitrogen) using an ABI 373 automated sequencer. Sequence alignments were performed using DNASIS software version 2.5.

Results and discussion

To identify primer combinations suitable for amplifying intergenic regions of the *P. nigra* chloroplast genome a total of 20 primer combinations out of 7 different regions of the LSC (Large Single Copy) region have been tested. Out of these, 5 primer pairs were selected for the determination of different cytotypes in the cpDNA of black poplar, which were subsequently digested with 1 or 2 restriction enzymes each.

Investigating 35 trees from the European wide collection (EUFORGEN core collection) of black poplar, we were able to identify 20 different haplotypes (Table 3), using the fragment/ enzyme combinations described above. Among the detected variation, two trees always showed a distinct different banding pattern from the other *P. nigra* individuals. Comparing the banding pattern of these two Romanian samples (ICAS 3, ICAS 4) with fragments detected in other species of poplar, it turned out that the observed pattern resembles the one found in *P. deltoides*. Therefore we assume that these 2 individuals might be hybrids of *P. deltoides*, pollinated with *P. nigra*. Investigating samples originating from accessions all over Europe, we found that there is little overall variation on cpDNA level. Especially accessions from central Europe show little variation compared to samples from south Eastern Europe with Spain having the highest variability present in its natural population based on cpDNA analysis.

In order to characterise the molecular basis of the detected variability in the chloroplast genome, 17 fragments from different haplotypes identified within these 35 trees, have been cloned and sequenced (Table 4). From the sequence analysis of the various fragments we conclude that many of the variations detected on the silverstained PAGE were due to only one or a few basepair changes (deletion, insertion, single base pair mutation) which led to a change in the conformation of the detected fragments. Therefore, the fragments do not migrate to the positions on the gel corresponding to their actual size, but rather to a position reflecting their base pair composition. Very often variation was found in poly A stretches within the fragments, where of A_n was up to 3bp different between the investigated haplotypes. Only fragment ORF-M showed an insertion/deletion situation within a region of 200bp length (Fig.1).

From our studies, we can conclude that regions of the cp genome of *P. nigra*, as already proven in studies with other species (Demesure 1997 (not in ref.list), Heinze 1998, Lagercrantz 1997?, King 1998), are applicable for biodiversity studies. Furthermore this system is capable of detecting naturally occurring hybridisation as it is clearly marking the maternal genome, even after several events of backcrossing. Especially fragment DT proved to be informative for this purpose, as it shows a relatively big difference in size among different species of poplar. In addition we proved by sequencing, that a lot of the detected variation using the native PAGE system, is due to point mutations as well as microsatellite (polyA) variations in the investigated regions (Echt *et al.*, 1998, Vendramin *et al.*, 1999). In various studies in

conifers, these variable regions were used for population genetic studies by only amplifying the microsatellite region.

Table 4: Cloned and sequenced fragments

fragment	Trees from Euforgen collection
TF	SEEFAR_PAZARDZIK_N1, FBS_87/65_OFFENBURG_, IZT_NS002
ORF-M	IBW_N009, SEEFAR_PAZARDZIK_N1, FBS_87/65_OFFENBURG_
DT	IBW_N009, FBS_87/65_OFFENBURG_, FCRA_HUNTINGDON, FF_V336, IZT_NS001
17+20	IBW_N009, SEEFAR_PAZARDZIK_N1; FBS_87/65_OFFENBURG_; SIA_LUC2; INRA_71017-401; KAE N90.013;

Fig. 1 Sequence alignment of the region 400-650 bp of the ORF fragment

		10	20	30	40	50
IBW-N009	1	TTCCATTTT	AATAAATTCT	AAATGAACTA	AATGAATATA	AATTTGAATT
Seefar-p	1	TTCCATTTT	AATAAATTCT	AAATGAA---	-----TATA	AATTTGAATT
SIA-PA1	1	TTCCATTTT	AATAAATTCT	AAATGAA---	-----TATA	AATTTGAATT
		60	70	80	90	100
IBW-N009	51	CTATTCAAAT	AAAATGGAAT	AAAAAATTAA	ATATA-----	--TAAATAAT
Seefar-p-N1		CTATTCAAAT	AAAATGGAAT	AAAAAATTAA	ATATA-----	--TAAATAAT
SIA-PA1	51	CTATTCAAAT	AAAATGGAAT	AAAAAATTAA	ATATAAAATA	TATAAATAAT
		110	120	130	140	150
IBW-N009	101	ATAAATATAT	TATATAAATT	-----	---ATAAATT	AAATGTATTA
Seefar-p-N1		ATAAATATAT	TATATAAATT	-----	---ATAAATT	AAATGTATTA
SIA-PA1	101	ATAAATATAT	TATATAAATT	ATTTATATAA	ATTATAAATT	AAATGTATTA
		160	170	180	190	200
IBW-N009	151	TATAAATCTA	TAAATCTAAA	TGTATTAGAT	GTAATCTATT	AATTAAATAT
Seefar-p-N1		TATAAATCTA	-----AA	TGTATTAGAT	GTAATCTATT	AATTAAATAT
SIA-PA1	151	TATAAATCTA	-----AA	TGTATTAGAT	GTAATCTATT	AATTAAATAT
		210	220	230	240	250
IBW-N009	201	AATATAATTA	AATATAATTA	AATATAATAG	GAATATATTA	AATAGGTAAT
Seefar-p-N1		AATATAATTA	AATATAAT--	-----AG	GAATATATTA	AATAGGTAAT
SIA-PA1	201	AATATAATTA	AATATAATTA	AATATAATAG	GAATATATTA	AATAGGTAAT

Table 3: Haplotypes detected in the EUFORGEN core collection

Clone name	country	CD/ EcoRI	CD/ Taq	TF/ Taq	DT/ MnII	Orf-M /EcoR	17/20 HinfI	haplotype
IBW_N004	BEL	a	a	a	a	a	a	A
IBW_N009	BEL	a	a	a	a	a	a	A
SEEFAR_PAZARDZIK_N1	BGR	b	d	b	b	b	b	B
SEEFAR_SVICTOV_N2	BGR	a	b	a	a	a	b	C
FBS_215/63_JUGENHEIM	GER	a	a	a	a	a	a	A
FBS_87/65_OFFENBURG	GER	c	a	d	c	c	c	E
SIA_PA1	ESP	c	a	c	c	c	d	F
SIA_LUC2	ESP	c	a	d	c	c	d	G
INRA_71017-401	F	c	c	d	c	c	e	D
INRA_92510-1	F	c	a	d	c	c	e	H
FCRA_HUNTINGDON	GBR	a	a	a	a	a	a	A
FCRA_HOBSONS_CONDUIT	GBR	a	a	a	c	c	e	J
FF_V336	HRV	a	a	a	d	a	a	K
FF-V408	HRV	a	a	a	a	a	a	A
ERTI_33-3-1	HUN	a	a	a	a	a	a	A
ERTI-33-3-2	HUN	a	a	a	a	a	a	A
ISP_N068	ITA	a	a	a	a	a	a	A
ISP_N351	ITA	d	a	e	-	d	-	
IBN_1238	NLD	c	a	a	a	a	b	M
IBN1792	NLD	a	a	a	a	a	a	A
POL_TORUN_B	POL	a	a	a	a	a	a	A
POL_KORNIK	POL	a	a	a	a	a	a	A
LVU_BAKA	SVK	a	a	a	a	a	a	A
LVU_IVACHNOVA	SVK	a	a	a	a	a	b	O
KAE_N90.013	TUR	b	d	b	f	b	f	P
IZT_NS002	YUG	c	c	f	c	c	e	Q
IZT_NS001	YUG	e	e	a	e	a	a	R
FBVA_LH_HL55	A	a	a	a	a	a	a	A
VULHM_88044	CZE	a	a	a	a	a	a	A
VULHM_88045	CZE	j	a	a	a	a	a	I
ICAS_3	ROM	f	f	g	g	e	g	S
ICAS_4	ROM	f	f	g	g	e	g	S
ICAS_5	ROM	g	a	a	e	a	a	T
ICAS_6	ROM	a	a	a	a	a	a	A
UKR_B12	UKR	a	a	a	a	a	a	A
Σ 35 individuals								
variations detected		7	6	7	7	5	7	Σ 20

References

- Demesure, B., N. Sodzi and R.J. Petit 1995; A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* 4: 129-131.
- Dumolin-Lapegue, S., B. Demesure, S. Fineschi, V. LeCorre and R.J. Petit 1997. Phylogeographic structure of white oaks throughout the European continent. *Genetics* 146(4): 1475-1487.
- Echt, C., L. DeVerno, M. Anzidei and G. Vendramin 1998. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa*. *Mol. Ecology*, 7: 307-316.
- Grivet, D., B. Heinze, G. Vendramin and R. Petit 2001 Genome walking with consensus primers: application to the Large Single Copy region of the chloroplast DNA. submitted
- Heinze, B. 1998 PCR-based chloroplast DNA assays for the identification of native *P. nigra* and introduced poplar hybrids in Europe;. *Forest genetics* 5: 31-38.
- King, R. and C. Ferris 1998. Chloroplast DNA phylogeography of *Alnus glutinosa*. *Molecular Ecology* ? : 1151-1161.
- Lagercrantz, U., J. Putterill, G. Coupland and D. Lydiate 1997. Comparative mapping in *Arabidopsis* and *Brassica*: fine scale genome colinearity and congruence of genes. *Plant Journal* 9: 13-20.
- Petit, Csaikl, Bordacs, Burg, Coart, Cottrell, Deans, Dumolin.Lapegue, Fineschi, Finkelday, Gillies, Giocoechea, Jensen, König, Lowe, Madsen, Matyas, Oledska, Pemonge, Popescu, Slade, Turchini, VanDam, Ziegenhagen, Kremer; 2002 Chloroplast DNA variations in European white oaks: synthesis based on data from over 2500 populations. *Forest Ecology and Management* (in press)
- Samuel, R., W. Pinsker and M. Kiehn 1997. Phylogeny of some species of *Cyrtandra* inferred from the atpB/rbcL cpDNA intergene region. *Botanica Acta* 110:503-510.
- Vendramin, G., M. Anzidei, A. Madaghiele, C. Sperisen and G. Bucci 1999. Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce. *Genome* 43: 68-78.
- Weising, K. and R. Gardener 1999. A set of conserved PCR primers for the detection of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* 42: 9-19.

2.3.2.2 Postglacial migration of *Populus nigra* L.: lessons learnt from chloroplast DNA

J.E. Cottrell^{a,*}, V. Krystufek^b, H.E. Tabbener^a, A.D. Milner^a, T. Connolly^a, L. Sing^a, S. Fluch^b, K. Burg^b, F. Lefevre^c, P. Achard^c, S. Bordacs^d, K. Gebhardt^e, B. Vornam^f, M.J.M. Smulders^g, A.H. Vanden Broeck^h, J. Van Slycken^h, V. Stormeⁱ, W. Boerjanⁱ, S. Castiglioneⁱ, T. Fossati^j, N. Alba^k, D. Agundez^k, C. Maestro^l, E. Notivol^l, J. Bovenschen^m, B.C. van Dam^m

^aForest Research, Northern Research Station, Roslin, Midlothian, Scotland EH25 9SY, UK

^bAustrian Research Centre Siebesdorf (ARCS), A-2444 Siebersdorf, Austria

^cINRA, Av. A. Vivaldi, 8400 Avignon, France

^dNational Institute for Agricultural Quality Control, Forestry Department, H-1024 Keleti K. u.24, Budapest, Hungary

^eHessian Forest Centre, Prof. Oelkers-Str. 6, D-34346 Hann. Münden, Germany

^fInstitute of Forest Genetics and Tree Breeding, Büsgenweg 2, 37077 Göttingen, Germany

^gPlant Research International, Wageningen, The Netherlands

^hInstitute for Forestry and Game Management (IFG), Research Station of the Flemish Community, Gaverstraat 4, B-9500 Geraardsbergen, Belgium

ⁱDepartment of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, Technologiepark 927, B-9052 Gent, Belgium

^jUniversità degli Studi di Milano, Dipartimento di Biologia (UVM), Sezione di Botanica Generale, Via Celoria 26, Torre CI-20133, Milano MI, Italy

^kNational Institute for Agrarian and Food Research (Instituto Nacional de Investigación Agraria y Alimentaria I.N.I.A.), Ctra de la Coruna, km 7.5, 28040 Madrid, Spain

^lSIA-DGA, Zaragoza, Spain

^mP.O. Box 47, 6700 AA Wageningen, The Netherlands

Forest Ecology and Management 206 (2005) 71–90

Abstract

Eleven laboratories have collaborated to study chloroplast DNA (cpDNA) variation in black poplar (*Populus nigra* L.) across Europe in order to improve our understanding of the location of glacial refugia and the subsequent postglacial routes of recolonisation. A common analysis based on the restricted fragments produced by five primer pairs was used to determine the cpDNA haplotype of 637 samples obtained from genebank collections established in nine European countries. Haplotype 2 was particularly common and was found in 46% of the non-hybrid samples. A total of 81 non-hybrid chloroplast variants were detected. Three haplotypes (from four trees believed to originate from Eastern Europe) clustered

together and were very different from the rest of the samples. The remaining samples were divided into two groups, one of which had a largely eastern distribution and samples from the other group were mostly located in the west. This, along with the fact that Spain in the southwest and Austria and Italy in the southeast had high diversity, suggest that there were ice age refugia of black poplar in both southwestern (Spain) and southeastern Europe (Italy and/or Balkan). Results also indicate that the Pyrenees formed a significant barrier, since only 7 of the 45 haplotypes in Spain exist elsewhere in Europe.

Keywords: Postglacial colonisation; Refugia; *Populus nigra* L.; Black poplar

1 Introduction

European black poplar (*Populus nigra* L.) is a pioneer species of riparian ecosystems. Flowers are wind pollinated and seed is largely wind dispersed. Vegetative propagules are disseminated both by water and by human activity. Regeneration of *P. nigra* by seed takes place through colonisation of newly perturbed sites and scattered trees are more frequent than huge stands. The species has a natural distribution ranging from North Africa and Ireland in the west, across to Russia and China in the east (Zsuffa, 1974). The abundance of black poplar is threatened due to the loss of its natural habitat by urbanisation, drainage of wetlands for agricultural use and canalisation of rivers for flood prevention. At one time its wood was highly prized because it is lightweight, resistant to fire and has excellent shock absorbing qualities. As a result, it has a long history of use for clogs, fruit baskets, furniture, flooring sheep hurdles and wagons. However, human mediated propagation of the species declined in the 19th century when the faster growing hybrid *P. x euramericana* was introduced to northern Europe. *P. nigra* is now recognised to be endangered and it has, therefore, been listed as one of the important species in need of conservation in the Strasbourg resolution of 1990 for the protection of forest trees in Europe (Anonymous, 1990; Arbez and Lefeuvre, 1997). Black poplar is an important component of interspecific poplar breeding programmes and both conservationists and tree breeders are aware of how important it is to protect the species.

In an attempt to conserve the genetic diversity that remains within this endangered species several European countries have independently set up ex situ genebanks in which cuttings of native black poplars from within each country are grown. Since 1994, the co-ordination of national conservation initiatives at the European level has been achieved through the EUFORGEN *Populus nigra* Network (Lefeuvre *et al.*, 1998). Among its different tasks for ex situ conservation, the Network has developed a European database of national genebanks (www.ipgri.cgiar.org) and a core collection of *P. nigra* clones has been established, which is representative of the whole geographic range of the species. This paper reports the results of a survey

of chloroplast DNA (cpDNA) variation in the genebanks in seven European countries. The original locations of the genotypes that were contained in the genebanks were used to determine the distribution of the various cpDNA haplotypes in Europe. The objective of this paper is to use this information to locate the postglacial refugia of black poplar and to determine the postglacial routes of colonisation of this species. Information regarding the distribution of cpDNA diversity is valuable in developing conservation policy as the colonising effect is known to have a major influence on the structure and distribution of existing diversity within a species. The amount of diversity within collections with nuclear markers (isozymes, AFLP, microsatellites is reported elsewhere (Storme *et al.*, 2004)). The genebanks provided an excellent source of material based on collections made by people who were both knowledgeable regarding the locations of natural populations and able to distinguish *P. nigra* from hybrid material on the basis of morphology. This saved the time and expense of travelling throughout Europe to collect samples. The collections did however, have the disadvantage that only one sample per population was usually included and this prevented any comparison of within and between population diversity from being made. Also, in those genebanks, which had been set up as a breeding rather than a conservation resource, there had been no attempt to make an equal geographic sampling of the material which existed within each country. However, these aspects do not preclude an analysis of the distribution of chloroplast haplotypes in order to understand the recolonisation of Europe by *P. nigra* after the last ice age.

Ice ages occur at regular intervals of 100,000 years with warm interglacial periods lasting 15–20,000 years as a result of instabilities in the earth's climate caused by Milankovitch cycles (Bennett, 1990). Many trees common in northern Europe today survived these glacial periods as small, low-density populations in refugia in deep valleys between the mountains of southern Europe (Bennett *et al.*, 1991). Fossil pollen maps of European deciduous oaks indicate refugia in southern Spain, southern Italy and the Balkan peninsula (Huntley and Birks, 1983; Bennett *et al.*, 1991). However, unlike other species, pollen records are of little value in determining the colonisation routes of *P. nigra*. This is partly because *Populus* pollen has been found to be present in very low quantities in sampled pollen cores and also because it is impossible to distinguish the pollen of *P. nigra* from that of *P. alba* and *P. tremula*. Therefore, there is, in poplar, a particular requirement for an alternative method to determine the postglacial routes of colonisation. This is now available in the form of cpDNA methodology, which offers a significant technological advance that can be used to provide the first insight into the postglacial colonisation routes of black poplar in Europe.

Many universal primers have been developed for the chloroplast genome of *Nicotiana tabacum* (Heinze, 1998a,b; Samuel *et al.*, 1997; Demesure *et al.*, 1995; Dumolin-Lapegue *et al.*, 1997; Weising and Gardner, 1999; Petit *et al.*, 2002a). These regions of the chloroplast genome have provided useful information for

biodiversity in a range of species (Demesure *et al.*, 1995; Heinze, 1998a,b; Lagercrantz *et al.*, 1997; King and Ferris, 1998; Petit *et al.*, 2002a,b). In these species, the uniparental mode of inheritance, the absence of recombination and the low mutation rate make this genome an appropriate source of markers for phylogenetic studies as well as for the study of postglacial routes of colonisation. Maternally inherited genomes are more geographically structured, due to limited seed dispersal compared to pollen movement (El Mousadik and Petit, 1996). The cpDNA genome has been extensively studied in poplar using both the RFLP and the PCR approach (Smith and Sytsma, 1990; Mejnartowicz, 1991; Rajora and Dancik, 1992, 1995a,b,c; Sabsch, 1992; Vornam *et al.*, 1994; Heinze, 1998a,b). Initial interest was directed at the development of species-specific markers and whereas Smith and Sytsma (1990) found no interspecific variation, several others demonstrated interspecific variation between *P. nigra* and *P. deltoides* (Vornam *et al.*, 1994; Rajora and Dancik, 1995a,b,c; Heinze, 1997; Krystufek, 2001; Krystufek *et al.*, 2002). Intraspecific variation was also detected in *P. nigra*, as well as in several other poplar species (Sabsch, 1992; Rajora and Dancik, 1995a; Heinze, 1998a; Krystufek, 2001). In many cases this intraspecific variation was dependent on the geographic origin of the maternal line of the tested material. The maternal inheritance of cpDNA in poplar was first demonstrated in controlled crosses by Mejnartowicz (1991), and this mode of inheritance was later confirmed by Rajora and Dancik (1992). However, in a later paper Rajora and Dancik (1995c) questioned the finding that cpDNA is entirely maternally inherited in poplar. The variants found in the *P. - euramericana* hybrids that were studied had not been detected in any *P. deltoides*, which had previously been studied. This finding did not fit with the belief that this species had acted as the maternal parent of these hybrids. Rajora and Dancik (1995a) suggest that this represents evidence that cpDNA is not entirely maternally inherited in poplar and that there may be parental recombination in *P. - euramericana* hybrid clones. This is an important consideration, as it would render the cpDNA molecule unsuitable for phylogenetic studies in poplar. This phenomenon has been detected in other species but not in any other studies involving poplar. Heinze (1998b) is sceptical that these results do indeed present evidence of cpDNA paternal leakage. Instead, he suggests that the results may be due to the existence of undetected cpDNA variants in *P. deltoides* or to probe contamination. In addition, Heinze (1998b) points out that as these results are confined to one particular probe which covers a region which is known to be a mutation hotspot in other species and as these results were detected in hybrids they may reflect sequence instability in hybrids rather than evidence of paternal leakage of cpDNA. Therefore, taken in its entirety, the balance of evidence indicates that cpDNA in *P. nigra* is maternally inherited and therefore variation in the cpDNA molecule has been used in this paper to study postglacial routes of colonisation. Although no detailed data exist on the distribution of cpDNA variation in black poplar such information exists for *Alnus glutinosa* L. (black alder), which is also a wind pollinated tree species of riparian and waterlogged habitats (King and Ferris,

1998). Alder is common in Europe and the Mediterranean and extends as far as the mountains of Turkey and North Africa. The cpDNA results show that most of northern Europe was colonised from a refuge in the Carpathian region (Hungary and Romania), although two further refugia in Spain and Turkey are suggested. One of the objectives of the current study was to determine whether current populations of *P. nigra*, a species, which has much in common with black alder in ecological terms, originate from the same refugia. Additionally, this study aimed to determine how the diversity in cpDNA haplotypes is distributed across Europe.

2. Materials and methods

2.1. Material

A consortium of 11 laboratories in 9 European countries participated in this EU funded project. One of the project objectives was to develop appropriate cpDNA markers that could be used to assess the diversity, which had been captured in the genebanks set up by each country for ex situ conservation of *P. nigra*. A total of 637 samples from genebanks in seven countries were analysed (Table 1). Detailed information about sampling locations is available on the EUFORGEN web pages.

2.2. DNA extraction and analysis

DNA was extracted from young leaves using QIAGEN plant Dneasy mini kits according to the manufacturer's instructions (<http://www.qiagen.com>). Details of the genebanks and the number of samples analysed by each laboratory are listed in Table 1. The five primer pairs that were used in the study are described by Fluch *et al.*, (2002) and are listed in Table 2. The optimised cpDNA primers and appropriate restriction enzymes (Table 3) were optimised for *P. nigra* in the Austrian laboratory and distributed to the other participating laboratories with detailed instructions on how to perform the analysis. The EUFORGEN Core Collection (Vietto, 2000) was used to provide reference samples against which banding patterns from the genebank samples could be compared. Several hybrid and non-hybrid *P. nigra* poplar samples were also included as references. Novel bands and new haplotypes, which were discovered in the course of the genebank survey, were checked by the Austrian laboratory before being included in the database. This precaution was implemented to reduce misscoring errors that might otherwise have arisen because the gels were run and scored in several different laboratories. The gels showing the banding patterns obtained for each primer/restriction enzyme combination and the results obtained for the EUFORGEN core collection and a range of interspecific crosses are presented by Krystufek (2001). Unfortunately, in the current survey, some of the primer/restriction enzyme combinations failed in certain laboratories. In particular, the ORFMEcoRI failed to work in the Italian laboratory, the DT MnlI failed with the Spanish samples and 17 + 20 HinfI failed in France.

PCR reactions were set up in 25 ml aliquots using the following components: 15 ml sterile distilled water, 2.5 ml 10 $\times$ reaction buffer, 0.8 ml 50mM MgCl₂, 0.4 ml W-1

(1%), 0.4 ml BSA (20 mg/ml), 0.365 ml dNTP (10 mM), 0.024 ml primer 1 (100 mM), 0.024 ml primer 2 (100 mM), 0.3 ml Taq polymerase (5 U/ml) and 5 ml DNA (5:95 dilution). The primers were supplied by Amersham Pharmacia Biotech and all other PCR components by Gibco BRL. The PCR conditions for DT, CD and TF were as follows: 95 °C for 15 min, followed by 40 cycles of (93 °C for 45 s, 55 °C or 58 °C (see Table 2) for 45 s and 72 °C for 2 min) followed by 72 °C for 10 min. For ORF-M there was an initial step of 95 °C for 15 min followed by 30 cycles of (95 °C for 50 s, 55 °C for 50 s and 70 °C for 1 min) followed by 70 °C for 10 min. The PCR conditions for 17 + 20 were 95 °C for 15 min followed by 20 cycles of (95 °C for 50 s, 58 °C for 50 s with a reduction of 0.5 °C per cycle and 72 °C for 1 min 45 s) followed by a further 20 cycles of (95 °C for 50 s, 50 °C for 50 s and 70 °C for 1 min 45 s) followed by 72 °C for 10 min.

The PCR products were then digested using the restriction enzymes shown in Table 3. The 15 ml digestion mixture contained 2 ml 10× buffer, 0.5 ml restriction enzyme (10 U/ml) and 12.5 ml sterile distilled water. 5 ml of amplified PCR product was added to 15 ml of digestion mixture. Following digestion, 5 ml of bromophenol blue was added to 12.5 ml of digestion product and each sample was loaded on to a 8% polyacrylamide gels. After electrophoresis at 250 V for 4.5 h the bands were visualised using a silver staining system.

2.3. Phylogenetic analysis

The data were scored as multistate, unordered characters in which each restriction fragment was a character and the multistates were the different sizes of each fragment. The most common fragment size was assigned as 1 and other size fragments were numbered in order of decreasing frequency. Two methods of analysis were applied to the dataset of 94 unique haplotypes. The first method involved the construction of a genetic distance matrix based on Jaccard similarity coefficients. All the points were then connected in a minimum spanning tree using the software package Genstat (Payne *et al.*, 1993). This procedure connects haplotypes by direct links, which have the smallest possible total length (Prim, 1957). The data were also analysed using the Fitch algorithm in the PHYLIP 3.5 computer package (Felsenstein, 1993) in order to construct a phylogenetic tree. The Fitch analysis estimates phylogenies from distance matrix data using the 'additive tree model' where the distances are expected to equal the sums of branch lengths between taxa (Petit *et al.*, 2002a).

3. Results

3.1. Classification and relatedness of cpDNA haplotypes

The classification into individual haplotypes was complicated by the fact that data were missing for certain primer/restriction enzyme combinations in several countries. A strategy was therefore adopted where unique haplotypes were identified based on complete sets of data at all primer/restriction enzyme combinations. Samples with

incomplete data were then compared with haplotypes based on complete data sets. If a sample with an incomplete data set did not differ from a haplotype with a complete dataset it was considered to be the same haplotype but was given a suffix to indicate that this was based on an incomplete data set. For example, if a sample with a complete data set was called haplotype 1, a sample which matched this sample but had missing data at a given primer/restriction enzyme was called haplotype 1a. Another sample which also matched haplotype 1 but had missing data at a different primer/restriction enzyme was called haplotype 1b. Samples which had missing data but did not match a sample with a complete dataset were allocated a unique haplotype number. The list of haplotypes along with their banding patterns is presented in Table 4. A total of 94 unique haplotypes, based on six primer/restriction enzyme combinations, was detected in the 637 trees analysed in this study.

The two methods used to assess relatedness, namely the minimum spanning tree and the Fitch analysis, largely agreed with one another (Fig. 1a and b). The haplotypes could be divided into four groups. With the exception of haplotype 90, both analyses clearly separated the samples that were suspected hybrids (on the basis of morphology) from the nonhybrid *P. nigra* samples. Haplotype 90, a suspected hybrid, had missing data and this may account for its unexpected position in the relatedness trees (Table 4). Alternatively, although it was suspected of being a hybrid it may in fact have been a pure black poplar. The hybrid haplotypes were called Group I. In total, this represented only 22 samples consisting of 11 haplotypes, which indicates that less than 4% of the genebank trees are hybrid in origin. The majority (73%) of these hybrid trees came from the Belgian and Dutch genebanks. They consistently differed from the non-hybrids at CD EcoRI-1 and CD EcoRI-2 (Table 4). Their haplotypes were similar but not identical to those of *P. deltoides* x *P. nigra* controlled crosses tested by Krystufek (2001) using the same primer/enzyme combinations. This, combined with their morphology indicates that they were probably *P. x euramericana* hybrids. These 11 haplotypes were excluded from subsequent analysis for the determination of postglacial routes of colonisation.

The relatedness analysis (particularly the Fitch) also clearly distinguished haplotypes 79, 80 and 81 as a separate group (Group II). This group consisted of three samples from Germany and one from Hungary. The original locations from which these samples had been taken was not recorded for the three German samples although when the records relating to the German genebanks were examined in more detail it emerged that the trees which were haplotypes 79 and 80 were fastigiate in form. They were poplars that had been grown as ornamentals and were not native to Germany but were believed to be of eastern European origin. The distribution of the remaining samples is illustrated in Fig. 2a–d. These samples could be divided into two groups consisting of Group III, which had a largely eastern distribution and Group IV, which exhibited a more western distribution (Figs. 2 and 3). The division of the samples into Group III and Group IV shown in Table 4 is based on the minimum spanning tree analysis and the Fitch analysis largely, though not

entirely agrees with this division. The majority of samples belonged to Group IV. In the Fitch analysis the haplotypes from Italy (haplotypes 5, 13, 22, 23, 24 and 25) are located in an intermediate position between Group III and Group IV. In the minimum spanning tree some of the Italian haplotypes (haplotypes 23 and 24) cluster with the eastern European samples and the others (haplotypes 5, 13, 20, 22 and 25) group with samples originating from the west. The Italian samples have unique bands for several of the TF TaqI-1 and DT Mnl-1 fragments (Table 4). It is therefore difficult to interpret how closely related these Italian samples are to the rest of the material that was analysed.

Table 5 shows the number of non-hybrid haplotypes that were present in latitude categories of two degrees. The number of samples was different in each latitude category. However, a chi square test showed that there is statistically significant evidence ($p < 0.01$) to reject the hypothesis that the number of haplotypes depends on the number of samples regardless of latitude. The southern latitudes contained a higher diversity with fifty-one of the haplotypes being restricted to latitudes below 46°N. This large number is disproportional to the number of samples taken from this region and is consistent with the notion of glacial refugia in southern Europe.

3.2. Distribution of haplotypes

In total, the genebanks that were sampled contained 83 non-hybrid haplotypes. From this total, 50 were only represented by a single sample and a further 16 were only present in two or three samples. The distribution of the majority of haplotypes (85%) was restricted to a single country and only 11 haplotypes occurred in more than one country (Table 6). Thirty-eight haplotypes were restricted to Spain. In Austria, Italy, France and Germany between five and seven haplotypes occurred which were unique to one particular country. Northern countries also had some unique haplotypes with four in Belgium, three in the Netherlands and one in Britain. Hungary only had three haplotypes of which two only occurred there (Table 6).

The most common haplotype was haplotype 2, which belonged to Group III. This was by far the most common haplotype in the whole survey and was represented by 284 (46%) of the 615 non-hybrid samples that were analysed. It was abundant in Italy, Austria, Germany, Netherlands, Belgium and Britain whereas it was extremely rare in Spain and France and entirely absent from Hungary (Fig. 2a). It is impossible to know whether the samples from Italy truly belong to haplotype 2 as they have data missing at locus ORF-M so caution is needed in interpreting the presence and distribution of haplotype 2 in Italy. Haplotypes 35 and 52, which were borderline between Groups III and IV were restricted to the Netherlands. The other haplotypes which belonged to Group III (haplotypes 1, 3, 4, 6, 10, 11, 16, 17, 23, 24, 26, 28, 29, 30, 32 and 77) had an almost entirely eastern distribution. An area along the Danube in northeastern Austria was particularly rich in diversity with eight haplotypes from Group III growing within a very small area (Fig. 2a). The majority of samples from Hungary belonged to haplotype 10 and this haplotype was restricted to Hungary with

the sole exception of one sample from Austria. If haplotype 2 and 52 are excluded, haplotypes, which belonged to Group III, only occurred in seven samples outside Italy, Austria and Hungary (Fig. 2a).

Spain had the highest diversity with a total of 45 haplotypes of which 38 were unique to Spain. Of the remaining seven haplotypes only one (haplotype 2) occurred further east than the Netherlands and two (haplotypes 7 and 58) were restricted to France and Spain. Haplotypes 14 and 53 were common haplotypes, which occurred both within and beyond Spain. The 40 samples that were haplotype 14 were found in Spain, France and Germany and the 50 samples of haplotype 53 occurred in Spain, France, the Netherlands and Britain.

Italy also showed quite a high diversity with nine haplotypes of which six were unique to Italy. The countries with the lowest diversity were Hungary with three haplotypes and Britain with three haplotypes. In Britain and the Netherlands, which are at the northwestern limit of the species, haplotypes from Groups III and IV grew in close proximity to one another (Fig. 2).

4. Discussion

4.1. Postglacial refugia and routes of colonisation

The phylogenetic analyses of all the European genebank samples clearly divided the non-hybrid haplotypes, which had a western distribution from those located in the east (Fig. 3). The Italian group was located between the eastern and the western samples. Haplotypes from the west were more similar to each other than to those found in the east and vice versa. This, along with the fact that Spain in the southwest and Austria and Italy in the southeast had high diversity, suggest that there were refugia of black poplar in both southwestern and southeastern Europe. The Iberian Peninsula is considered to have been a glacial refugium for other tree species such as *Quercus* (Jimenez *et al.*, 1999; Goicoechea and Aguñdez, 2000) and *Pinus* (Salvador *et al.*, 2000) and Alba *et al.*, (2002) suggested, on the basis of the high diversity detected in black poplar in Spain, that this may have acted as a refuge for *P. nigra*. The current analysis provides further evidence to support the idea that western refugia existed in Spain. The results indicate that the Pyrenees acted as an effective barrier to migration because only 6 of the 44 non-hybrid haplotypes detected in Spain existed elsewhere in Europe. The exact number and locations of Spanish refugia remains unclear. In an analysis of the Spanish results Alba *et al.*, (2002) commented that the cpDNA diversity in the Ebro valley was much higher than in the Douro and Tagus valleys. The comparison of black poplar populations on these three rivers showed that 41 of the 45 haplotypes found were unique to the Ebro valley (Alba *et al.*, 2002). The Ebro valley runs parallel to the Pyrenees in eastern Spain and eventually flows into the Mediterranean. In contrast, the other two rivers flow in a westerly direction across Spain and eventually into the Atlantic. The higher diversity in the Ebro valley may indicate that this area included some of the more important

glacial refugia of the species. Uncertainty regarding the exact location of Iberian refugia also exists in other tree species. For example, palynological evidence led Brewer *et al.*, (2002) to propose that the Spanish refugia for oak were restricted to the extreme south of the country. However, Olande *et al.*, (2002) point out that in Iberia, altitude and topography override the latitudinal effects on climate, bringing moisture to southern regions and mild temperatures to northern ones. Vegetation could have survived the glacial period in basal mountain valleys and deep gorges outside southern Iberia. They quote evidence to support the idea that other species, such as *Corylus* (Sa'nchez Gon'i and Hannon, 1999), may have survived in refugia located in deep gorges in the river Ebro basin. The results, therefore, highlight the need to perform a more structured sampling of Spain, perhaps for a number of tree species, if the exact locations of the refugia are to be identified. Such a survey would identify the regions of greatest diversity, which merit the most effort in terms of conservation.

The exact locations of the eastern refugia are also difficult to identify accurately. The Italian peninsula contained many unique haplotypes and this lends support to the idea that this area acted as a refugium. The Alps appear to have acted as a barrier to the escape of most of the Italian haplotypes. There were also several unique haplotypes in eastern Austria and Hungary so the data also support the idea of a refugium somewhere to the east of Italy. The high diversity detected by Bordacs *et al.*, (2002a) in a population growing along the Danube River in Hungary is further support for this. It will be necessary to obtain additional material from locations south and east of Hungary to confirm the existence and location of this putative refugium. The presence of haplotype 2 throughout Italy and also in Austria but almost entirely absent from France is interesting. This pattern of distribution suggests that the Alps formed a very effective barrier to the movement of this haplotype into France but appeared to allow migration into Austria. Alternatively, this haplotype may have existed in more than one refugium, possibly in Italy and also in an area east of the Adriatic Sea in the Balkan region. The Balkan material could then have migrated northwards to colonise Austria without having to cross any major mountain ranges. A similar situation was found for an oak haplotype which occurs in both southern Italy and the Balkans (Bordacs *et al.*, 2002b; Csaikl *et al.* 2002; Fineschi *et al.*, 2002; Petit *et al.*, 2002b). They suggested that the presence of a haplotype in both these refugial areas may have resulted from its migration across the Adriatic Sea in a previous interglacial period. This would have been possible because of the land bridge, which existed at this time.

The phylogenetic analysis showed that three haplotypes (79, 80 and 81), which constituted Group II, were very different from the other haplotypes that were detected in the study. Bordacs *et al.*, (2002a) found that some of the Hungarian samples, which had a similar cpDNA fingerprint to these samples, produced the banding pattern typical of the 'Thevestina' clone when analysed using the Heinze (1998a,b) cpDNA marker. Haplotypes 79, 80 and 81 might therefore be distinct because they originate from the same area as 'Thevestina', which itself originates from the Central

Asian and Black Sea region and has been widely used as an ornamental tree. This highlights the need to extend the sampling to regions further east.

The hypothesis that there were at least three refugia for black poplar is consistent with findings for other tree species (Taberlet *et al.*, 1998). For example, in oak, refugia in Spain, Italy and the Balkans have been proposed (Dumolin-Lape`gue *et al.*, 1997; Ferris *et al.*, 1998; Petit *et al.*, 2002b). The three refugia are all thought to have provided material for postglacial colonisation of countries to the north. In contrast, beech and alder, share a refuge in the Carpathian region from whence most of Europe was recolonised. Both species also have unique haplotypes in southern Italy, which apparently did not spread across the Alps. The higher levels of cpDNA diversities in the southern compared to the northern populations found in *P. nigra* agree with previous phylogeographic results in other tree species such as *Fagus*, *Quercus* and *Alnus* (Demesure *et al.*, 1996; Dumolin-Lapegue *et al.*, 1997; King and Ferris, 1998). For example, twelve of the thirteen alder haplotypes detected by King and Ferris (1998) occur south of 45°N latitude. Similarly in oak, of the 32 haplotypes that were detected 12 were restricted to regions below 45°N, 3 occurred only above 45°N and 17 occurred in both northern and southern regions (Petit *et al.*, 2002a). This is also true of *P. nigra* in which 50 of the 83 nonhybrid haplotypes are restricted to latitudes below 46°N. Two possible reasons for this thinning of haplotypes from south to north have been suggested in the literature. Firstly, the large mountain ranges such as the Pyrenees, Alps and Carpathian mountains are all in Southern Europe and these may have acted as barriers preventing the majority of haplotypes from migrating northwards during postglacial recolonisation. Secondly, this reduction in diversity during expansion from refugia has been predicted for species that undergo leptokurtic (long distance) as opposed to normal dispersal (Hewitt, 1996).

In her initial survey of cpDNA variation in the EUFORGEN Core Collection (based on a complete data from the seven loci used in the current study), Krystufek (2001) found that haplotype 2 occurred in all the countries tested in the Europop project except for France and Spain. This haplotype was also present in samples from as far east as Romania and the Ukraine. The great abundance of haplotype 2 may be a reflection of both the difficulty that black poplar experienced in escaping the refugia after the ice age and the subsequent rapid rates of colonisation. The viability of black poplar seeds is known to be extremely short lived and it has very exacting germination requirements found only along river valleys, which are flooded in winter (Barsoum and Hughes, 1998). Indeed, Imbert and Lefe`vre (2003) deduced from the isolation by distance pattern of the distribution of diversity within a single river system that seed dispersal in black poplar is not effective over a very long distance. This may have made it difficult for the haplotypes that had escaped across the mountain ranges to find land immediately beyond which was suitable for germination. The fact that black poplar readily propagates vegetatively may have assisted it to migrate across mountain barriers. The occasional transport of twigs over long distances by water, birds, beavers or man may have played an important role in the expansion of

the species from refugia. Haplotype 2 may have been the first to escape over the physical mountain barriers to establish founding populations beyond. The seed produced by these first generation trees would have had the opportunity to be spread rapidly by wind across the flat plains of Europe in the absence of other haplotypes. Although there are no palynological data for black poplar, data for other wind or water dispersed species such as *Betula*, *Pinus* and *Alnus* indicate rapid colonisation rates of 500–2000 m year⁻¹ (Huntley and Birks, 1983). Other species such as oaks, which rely on birds for dispersal, spread at a slower rate of 350–500 m year⁻¹.

Despite the fact that haplotype 2 predominated, there was relatively high diversity of cpDNA haplotypes in northern Europe. Presumably, many of the haplotypes, which escaped across the physical mountain barriers, successfully migrated northwards, albeit in lower density than haplotype 2. This pattern is quite different from that of oak, where no unique haplotypes are found in northwestern Europe (Ibrahim *et al.*, 1996). For species such as oak, the hypothesis is that rare, long distant dispersal events allowed single haplotypes to colonise large areas. Once these large, single haplotype areas were established it was impossible for other haplotypes to gain a foothold because appropriate sites had already been colonised. This led to a loss of diversity as the species moved northwards. The ecology of black poplar is very different from that of oak in several ways and this may account for the different distribution patterns of the haplotypes. First, oak is a climax species whereas black poplar is a non-climax pioneering species, which colonises disturbed riverine sites. As a consequence, black poplar tends to have a patchy distribution of populations concentrated along linear river valleys. Secondly, the black poplar populations are in a much more dynamic state than those of the oaks, they are constantly being re-established due to changing river dynamics as old sites with mature trees are lost and new ones are founded from immigrant seed. As a result, new colonisation sites are constantly being made available behind the initial colonisation front. Thirdly, the size of the catchment area for recruited seeds and vegetative propagules may be relatively large. As a result the following scenario can be envisaged. Haplotypes, which arrive after the first colonisation front, can establish themselves as new sites become available when rivers flood and change their courses. They, therefore, form small populations dotted in among populations of the original colonising haplotype. These small populations form steppingstones to the northerly progression of the haplotypes, which are later in arriving. Some of these steppingstones are lost when river dynamics change resulting in the ebb and flow of loss and establishment of populations. The dynamic renewal of colonisation sites would also explain why haplotypes from both eastern and western refugia coexist in Britain. One of the lineages will have arrived first but this will not have prevented the lineage, which arrived later from colonising, because new sites would constantly have been becoming available. This may also explain why some of the northern populations contain unique haplotypes, which are absent further south. The route taken by some of the *P. nigra* haplotypes may have been obliterated when populations were lost

when rivers changed their course. Alternatively, if the haplotypes are rare, the sampling intensity may just have been too low to enable the route to be traced.

4.2. Human influence

The hypothesis outlined above assumes that the observed pattern of cpDNA haplotypes has been established naturally. Movement of material by man and human mediated planting of one of the lineages could be an alternative explanation for the coexistence of eastern and western lineages in Britain. Cottrell *et al.*, (2002) showed that although eastern and western haplotypes were present in Britain in similar numbers the situation was very different in terms of numbers of clones. Britain has a very high degree of clonal duplication (Cottrell *et al.*, 1997, 2002) and there were only three different clones derived from the western lineage compared with twenty separate clones from the eastern lineage. It is therefore possible that the three clones from the western lineage were introduced, propagated and distributed by man. This would mean that Britain had only been naturally colonised by the eastern lineage.

This example illustrates the difficulty in interpreting cpDNA data for black poplar. The fact that it can be vegetatively propagated, coupled with the high level of human interference that the species has experienced, makes it impossible to be certain that the distribution of the various cpDNA haplotypes in Europe today is a true reflection of natural pattern of postglacial colonisation.

One of the other effects of human interference on *P. nigra* in Europe has been the widespread occurrence of *P. x euramericana* and other poplar hybrids throughout northern Europe. Artificially produced hybrids involving *P. nigra* and several other poplar species including *P. deltoides*, *P. trichocarpa*, *P. maximowiczii* and *P. laurifolia* have been planted throughout Europe. Such crosses have been made with *P. nigra* acting as either the male or the female parent. In addition, natural hybridisation and backcrossing of hybrids with black poplar is known to occur. It should be borne in mind that failure to recognise a hybrid or a backcross in which the maternal line is a species other than *P. nigra* could lead to misinterpretation of the current results for the construction of colonisation routes. The interpretation of the current results assumes that any hybrids that were sampled were detected and removed before consideration of postglacial colonisation routes. This assumption is probably correct because the Fitch analysis positioned all but one of the suspected hybrid haplotypes into a well-defined group of their own. This, coupled with the fact that all the morphology of all the trees included in the genebanks was examined prior to inclusion to check that they were typical of non-hybrid *P. nigra* specimens gives confidence that the postglacial colonisation routes defined in this paper are based on non-hybrid *P. nigra* trees.

4.3. Future work

The current work only sampled countries as far south as Spain and Italy and as far east as Hungary. A large area of the distribution range of black poplar in North

Africa, Russia and China remains unsampled. Group II provides an initial suggestion that diversity in Eastern Europe may be very different from that further west. In order to trace the colonisation routes more clearly there is a need for a further sampling effort in the east. Along with a more structured sampling of Spain, this will identify material in the regions where the diversity was highest and which merit the greatest conservation effort. An analysis of the Italian samples at locus ORF-M EcoRI would also clarify the relationship of these trees to those elsewhere in Europe.

Acknowledgements

This study has been carried out with financial support from the Commission of the European Communities, Agriculture and Fisheries (FAIR) specific RTD programme, PL97-3386, 'Genetic diversity in river populations of European Black Poplar for evaluation of biodiversity, conservation strategies, nature development and genetic improvement'. It does not necessarily reflect its views and in no way anticipates the Commission's future policy in this area. The Hungarian studies were financed by a Hungarian governmental grant (OTKA T32404). We thank Glen Brearley for his expert assistance in the preparation of the figures and maps.

References

- Alba, N., Maestro, D., Agundez, D., Novitól, E., 2002. Advances in the preservation of genetic resources in *Populus nigra* L. in Spain. In: van Dam, B.C., Bordacs, S. (Eds.), Genetic Diversity in River Populations of European Black Poplar-implications for Riparian Ecosystem Management. Proceedings of an International Symposium, Szekzard, Hungary, May 2001. Anonymous, 1990. Ministerial Conference Strasbourg. Actes de la Conférence Ministérielle pour la Protection des Forêts en Europe, 18 Décembre 1990. Strasbourg. Ministère de L'Agriculture et des Forêts, France. 255 pp.
- Arbez, M., Lefeuvre, F., 1997. Towards a European genetic resource programme: objectives and general conception—a case study concerning black poplar (*Populus nigra* L.). *Boccone* 7, 389–398.
- Barsoum, N., Hughes, F.M.R., 1998. Regeneration response of black poplar to changing river levels. In: Wheeler, H., Kirby, C. (Eds.), *Hydrology in a Changing Environment*. pp. 397–412.
- Bennett, K.D., 1990. Milankovitch cycles and their effects on species in ecological and evolutionary time. *Paleobiology* 16, 11–21.
- Bennett, K.D., Tzedakis, P.C., Willis, K.J., 1991. Quaternary refugia of north European trees. *J. Biogeography* 18, 103–115.
- Bordacs, S., Borovics, A., Bach, I., 2002a. Genetic diversity of natural populations and genebank of black poplar in Hungary. In: van Dam, B.C., Bordacs, S. (Eds.), Genetic Diversity in River Populations of European Black Poplar-implications for Riparian Ecosystem Management. Proceedings of an International Symposium, Szekzard, Hungary, May 2001.
- Bordacs, S., Popescu, F., Slade, D., Csaikl, U.M., Lesur, I., Borovics, A., Kezdy, P., Kónig, A.O., Górnó, D., Brewer, S., Burg, K., Petit, R.J., 2002b. Chloroplast DNA variation of white oaks in northern Balkans and in the Carpathian Basin. *For. Ecol. Manag.* 156, 197–210.
- Brewer, S., Cheddadi, R., de Beaulieu, J.L., Reille, M., 2002. The spread of deciduous *Quercus* throughout Europe since the last glacial period. *For. Ecol. Manag.* 156, 27–48.
- Cottrell, J.E., Forrest, G.I., White, I.M.S., 1997. The use of RAPD analysis to study diversity in British black poplar (*Populus nigra* L. subsp. *betulifolia* (Pursch) W. Wettst. (Salicaceae)) in Great Britain. *Watsonia* 21, 305–312.

- Cottrell, J.E., Tabbener, H.E., Forrest, G.I., 2002. Distribution of variation in British black poplar: role for human management. In: van Dam, B.C., Bordacs, S. (Eds.), Genetic Diversity in River Populations of European Black Poplar-implications for Riparian Eco-system Management. Proceedings of an International Symposium, Szekzard, Hungary, May 2001.
- Csaikl, U.M., Burg, K., Fineschi, S., Kořnig, A.O., Mařtya's, G., Petit, R.J., 2002. Chloroplast DNA variation in white oaks in the alpine region. *For. Ecol. Manag.* 156, 131–147.
- Demesure, B., Sodzi, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* 4, 129–131.
- Demesure, B., Comps, B., Petit, R., 1996. Phylogeography of common beech (*Fagus sylvatica* L.) in Europe inferred by restriction studies of PCR amplified chloroplast DNA fragments. *Evolution* 50, 2515–2520.
- Dumolin-Lapegue, S., Demesure, B., Fineschi, S., Le Corre, V., Petit, R., 1997. Phylogeographic structure of white oaks throughout the European continent. *Genetics* 146, 1475–1487.
- El Mousadik, A., Petit, R.J., 1996. Chloroplast DNA phylogeography of the argan tree of Morocco. *Mol. Ecol.* 5, 547–555.
- Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package) Version 3.5c. Department of Genetics, University of Washington, Seattle, WA.
- Ferris, C., King, R.A., Vainola, R., Hewitt, G.M., 1998. Chloroplast DNA recognises three refugial sources of European oaks and shows independent eastern and western immigrations to Finland. *Heredity* 80, 584–593.
- Fineschi, S., Turchini, D., Grossini, P., Petit, R.J., Vendramin, G.G., 2002. Chloroplast DNA variation in Italy. *For. Ecol. Manag.* 156, 103–114.
- Fluch, S., Krystufek, V., Burg, K., 2002. A chloroplast marker system for studying genetic variation in *P. nigra*. In: van Dam, B.C., Bordacs, S. (Eds.), Genetic Diversity in River Populations of European Black Poplar-implications for Riparian Eco-system Management. Proceedings of an International Symposium, Szekzard, Hungary, May 2001.
- Goicoechea, P.G., Agundez, D., 2000. Robles y Hayas en Espan'a. Conservacio'n de Recursos Geneticos. Investigacio'n Agraria: Sistemas y Recursos Forestales. Fuera de Serie 2, 125–142.
- Heinze, B., 1997. A PCR marker for a *Populus deltoides* allele and its use in studying introgression with native European *Populus nigra* L. . *Belg. J. Bot.* 129, 123–130.
- Heinze, B., 1998a. PCR-based chloroplast DNA assays for the identification of native *Populus nigra* L. and introduced poplar hybrids in Europe. *For. Genet.* 5, 31–38.
- Heinze, B., 1998b. Biochemical and molecular genetic methods available for the characterization of *Populus nigra* L. In: Turok, J., Lefevre, F., de Vries, S., Alba, N., Heinze, B., Van Slycken, J. (Compilers), *Populus nigra* Network. Report of the fourth meeting, 3–5 October 1997, Geraardsbergen, Belgium. International Plant Genetic Resources Institute, Rome.
- Hewitt, G.M., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* 58, 247–276.
- Huntley, B., Birks, H.J.B., 1983. An Atlas of Past and Present Pollen Maps of Europe, 0–13,000 Years Ago. Cambridge University Press, Cambridge.
- Ibrahim, K.M., Nichols, R.A., Hewitt, G.M., 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* 77, 282–291.
- Imbert, E., Lefevre, F., 2003. Dispersal and gene flow of *Populus nigra* L. (Salicaceae) along a dynamic river system. *J. Ecol.* 91, 447–456.
- Jimenez, P., Agundez, D., Alis, D., Gil, L., 1999. Genetic variation in Central and Marginal population of *Quercus suber* L. *Silv. Genet.* 48 (6), 278–284.
- King, R., Ferris, C., 1998. Chloroplast DNA phylogeography of *Alnus glutinosa*. *Mol. Ecol.* 7, 1151–1161.
- Krystufek, V., 2001. Population genetic analysis of *Populus nigra* L. in Austria using nuclear and Chloroplast DNA markers, Ph.D. Thesis, University of Vienna, pp. 47–52.
- Krystufek, V., Fluch, S., Burg, K., 2002. Artificial yet natural: colonisation by black poplar of an artificial island in the river Danube in Vienna. In: van Dam, B.C., Bordacs, S. (Eds.), Genetic Diversity in River Populations of European Black Poplar-implications for Riparian

- Eco-system Management. Proceedings of an International Symposium, Szekzard, Hungary, May 2001.
- Lagercrantz, U., Putterill, J., Coupland, G., Lydiate, D., 1997. Comparative mapping in *Arabidopsis* and *Brassica*: fine scale genome colinearity and congruence of genes. *Plant J.* 9, 13–20.
- Lefevre, F., Legionnet, A., de Vries, S., Turok, J., 1998. Strategies for conservation of a pioneer tree species, *Populus nigra* L., in Europe. *Genet. Sel. Evol.* 30 (Suppl. 1), 181–196.
- Mejnartowicz, M., 1991. Inheritance of chloroplast DNA in *Populus*. *Theor. Appl. Genet.* 82, 477–480.
- Olande, M., Herran, A., Espinel, S., Goicoechea, P.G., 2002. White oaks phylogeography in the Iberian Peninsula. *For. Ecol. Manag.* 156, 89–102.
- Payne, R.W., Lane, P.W., Digby, P.G.N., *et al.*, 1993. *GenstatTM 5 Release 3 Reference Manual*. Clarendon Press, Oxford.
- Petit, R.J., Csaikl, U.M., Bordacs, S., Burg, K., Coart, E., Cottrell, J., van Dam, B., Deans, J.D., Dumolin-Lapegue, S., Fineschi, S., Finkeldey, R., Gillies, A., Glaz, I., Goicoechea, P.G., Jensen, J.S., Konig, A.O., Lowe, A.J., Madsen, S.F., Matyas, G., Munro, R.C., Oledska, M., Pemonge, M.-H., Popescu, F., Slade, D., Tabbener, H., Turchini, D., de Vries, S.G.M., Ziegenhagen, B., Kremer, A., 2002a. Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *For. Ecol. Manag.* 156, 5–26.
- Petit, R.J., Brewer, S., Bordacs, S., Burg, K., Cheddadi, R., Coart, E., Cottrell, J., Csaikl, U.M., van Dam, B., Deans, J.D., Espinel, S., Fineschi, S., Finkeldey, R., Glaz, I., Goicoechea, P.G., Jensen, J.S., Konig, A.O., Lowe, A.J., Madsen, S.F., Matyas, G., Munro, R.C., Popescu, F., Slade, D., Tabbener, H., de Vries, S.G.M., Ziegenhagen, B., de Beaulieu, J.-L., Kremer, A., 2002b. Identification of refugia and postglacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *For. Ecol. Manag.* 156, 49–74.
- Prim, R.C., 1957. Shortest connection networks and some generalizations. *Bell Syst. Tech. J.* 36, 1389–1401.
- Rajora, O.P., Dancik, B.P., 1992. Chloroplast inheritance in *Populus*. *Theor. Appl. Genet.* 84, 280–285.
- Rajora, O.P., Dancik, B.P., 1995a. Chloroplast DNA variation in *Populus* 1. Intraspecific restriction fragment diversity within *Populus deltoides*, *P. nigra* and *P. maximowiczii*. *Theor. Appl. Genet.* 90, 317–323.
- Rajora, O.P., Dancik, B.P., 1995b. Chloroplast DNA variation in *Populus* 2. Interspecific restriction fragment polymorphisms and genetic relationships among *Populus deltoides*, *P. nigra*, *P. maximowiczii* and *P. canadensis*. *Theor. Appl. Genet.* 90, 324–330.
- Rajora, O.P., Dancik, B.P., 1995c. Chloroplast DNA variation in *Populus* 3. Novel chloroplast variants in natural *Populus* - *canadensis* hybrids. *Theor. Appl. Genet.* 90, 331–334.
- Sabsch, M., 1992. Untersuchungen u̇ber inter- und intraspezifische variation der cpDNA in der Gattung *Populus*. Thesis Dissertation, University of Göttingen.
- Salvador, L., Alia, R., Agundez, D., Gil, L., 2000. Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait) in the Iberian Peninsula. *Theor. Appl. Genet.* 100, 89–95.
- Samuel, R., Pinsker, W., Kiehn, M., 1997. Phylogeny of some species of *Cyrtandra* inferred from the *atpB/rbc1* cpDNA intergene region. *Botanica Acta* 110, 503–510.
- Smith, R.L., Sytsma, K.J., 1990. Evolution of *Populus nigra* L. (Sect. *Aigeiros*): introgressive hybridization and the chloroplast contribution of *Populus alba* (Sect. *Populus*). *Am. J. Bot.* 77, 1176–1187.
- Sanchez Goni, M.F., Hannon, G.E., 1999. High altitude vegetational pattern on Iberian Mountain Chain (north-central Spain) during the Holocene. *Holocene* 9, 39–57.
- Storme, V., Boejeran, W., Vanden Broeck, A.H., Ivens, B., Halfmaerten, D., Van Slycken, J., Castiglione, S., Grassi, F., Fossati, T., Cottrell, J.E., Tabbener, H.E., Lefevre, F., Imbert, E., Fluch, S., Krustufek, V., Burg, K., Bordacs, S., Gebhardt, K., Vornam, B., Pohl, A., Alba, N., Bovenschen, J., van Dam, B., van der Schoot, J., Vosman, B., Smulders, M.J.M., 2004. Ex situ conservation of black poplar in Europe: duplications and diversity in nine genebank collections and their use for conservation and nature development. *Theor. Appl. Genet.* 108, 969–981.

- Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17, 1105–1109.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G., Cousson, J.-F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* 7, 453–464.
- Vietto, L. 2000. Update of the EUFORGEN core collection and the database of clones. In: Borelli, S., de Vries, S., Lefe`vre, F., Turok, J. (Eds.), *Report of the Sixth Meeting of the Populus nigra L. Network*, pp. 26–28.
- Vornam, B., Herzog, S., Preisig-Müller, R., Hattemer, H.H., 1994. Restriction fragment length polymorphisms of a chloroplast photosystem II gene from poplar and their use for species identification. *Genome* 37, 747–750.
- Weising, K., Gardner, R., 1999. A set of PCR primers for the detection of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* 42, 9–19.
- Zsuffa, L., 1974. The genetics of *Populus nigra* L. *Ann. Forestales (Zagreb)* 6/2:, 29–53.

Table 5 The effect of latitude on the number of haplotypes present. The exact location of haplotype 91 in Hungary is unknown

Latitude	Identity and total number of non-hybrid haplotypes	
<42°N	11, 15, 18, 20, 23, 24, 25, 27, 33, 36, 37, 38, 39, 40, 41, 42, 45, 46, 48, 49, 50, 54, 55, 58, 61, 62, 63, 66, 69, 73, 74, 75, 76, 82, 83, 14a, 2a, 51a, 53a, 7a	41
42–44°N	5, 12, 23, 24, 34, 40, 43, 44, 56, 58, 65, 67, 68, 70, 71, 72, 14a, 2a, 53a, 7a	21
44–46°N	7, 13, 19, 21, 23, 24, 58, 59, 64, 14a, 16a, 2a, 53c, 8a	14
46–48°N	2, 8, 10, 19, 81, 14a, 2a	7
48–50°N	2, 3, 6, 9, 10, 14, 16, 17, 28, 29, 30, 32	13
50–52°N	1, 2, 8, 22, 26, 31, 51, 52, 53, 57, 40	11
>52°N	2, 8, 35, 47, 51, 52, 60, 40a, 53b	9

Table 6 The presence of non-hybrid haplotypes in each country

Country	Number of non-hybrid haplotypes	Non-hybrid haplotype composition according to country and Group													
		Group II	Group III					Group IV							
			2	4	10	16	U	53	7	8	14	58	40	51	U
Austria	9	1	X	X	X	X	5								
Belgium	5		X				2								
Hungary	3				X		1								
Italy	9		X	X		X	2								
Britain	3	2	X					X							
Germany	10		X				2			X	X				
France	12		X				1	X	X	X	X				
Netherlands	7		X				2	X					X	X	
Spain	45		X					X	X		X	X	X	X	3

The presence of a haplotype is represented by X and the number of haplotypes which are unique to a given country are listed in the column labelled U. The total number of non-hybrids per country per group are presented.

Table 1 Details of the location and establishment date of each genebank along with the number of samples from individual genebank and the laboratory responsible for the cpDNA analysis

Country	Genebank date of establishment	Number of ramets	Number of trees	Laboratory for analysis	Location of genebank
Netherlands	1986	240	79	Alterra, Wageningen	De Moerhoek, Dronten
France	>1970	500	64	INRA, Avignon	National Commission for the Conservation of Forest Genetic Resources
Spain	1980	110	75	Austrian Research Centre, Siebersdorf	SIA-DGA Zaragoza
	1990	45	25	Austrian Research Centre, Siebersdorf	INIA, Madrid
Belgium	1958–2003	250	84	VIB, Gent, Belgium	IFG, Geraardsbergen,
Austria	1998	180	100	Austrian Research Centre, Siebersdorf	Austrian Federal Office and Research Centre for Forests (BFW)
Italy		400	74	Università degli Studi di Milano, Milan	Casale Monferrato
UK	1993–1995	100	87	Forest Research, Roslin	Talybont, Fineshade and Downham Market
Germany	1970	370	22	Institute of Forest Genetics and Forest Tree Breeding, University of Göttingen	Hessen-Forst Hann. Muenden
Hungary	1950		31	OMMI, Department of Forestry, Budapest	Forestry Gene Bank Sárvár, managed by Hungarian Forest Research, established by Ferenc Kopecky

Table 3 The restriction enzyme, restriction temperature and number of fragments that were obtained for each cpDNA region that was studied

Table 3
The restriction enzyme, restriction temperature and number of fragments that were obtained for each cpDNA region that was studied

CpDNA fragment	Restriction enzyme	Restriction temperature (°C)	Restriction time (h)	Number of fragments (I–V) and number of variants per fragment (1–4) in the genebanks
CD	<i>EcoRI</i>	37	3	I(4), II(3), III(3), IV(4), V(3)
CD	<i>TaqI</i>	65	3	I(4), II(2), III(2), IV(5), V(3)
TF	<i>TaqI</i>	65	3	I(3), II(3), III(4)
DT	<i>MnII</i>	37	3	I(6), II(4), III(4)
ORF-M	<i>EcoRI</i>	37	3	I(4), II(6)
17 + 20	<i>HinfI</i>	37	3	I(7), II(2), III(3)

Fig. 1. (a) The minimum spanning tree for 94 hybrid and non-hybrid *P. nigra* genebank samples based on the molecular data presented in Table 4. The division of the individual haplotypes into Groups I–IV is shown. (b) The phylogenetic tree of 94 hybrid and non-hybrid *P. nigra* genebank samples obtained by the Fitch algorithm. The molecular data for each numbered haplotype are presented in Table 4. The large circles encompass Group I and II haplotypes. The uncircled numbers indicate haplotypes, which belong to Group III, and circles around single or, in rare cases, two haplotypes indicate membership of Group IV.

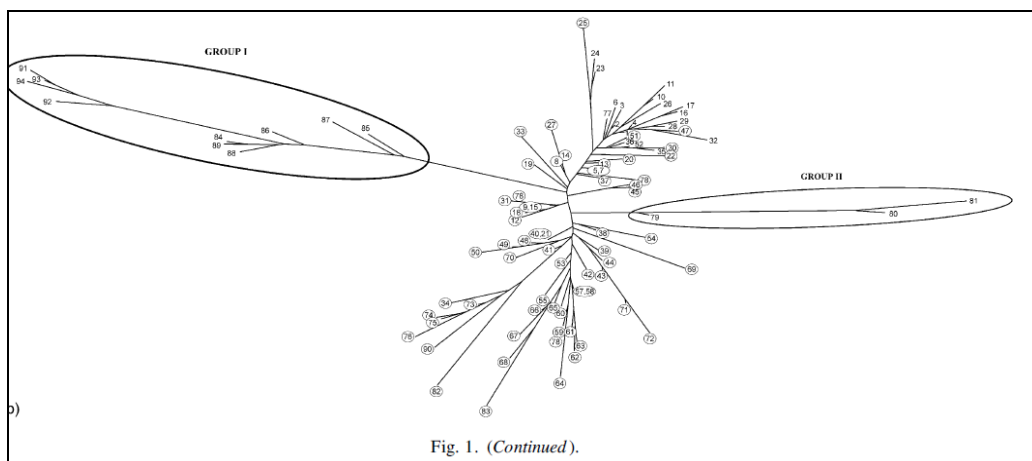
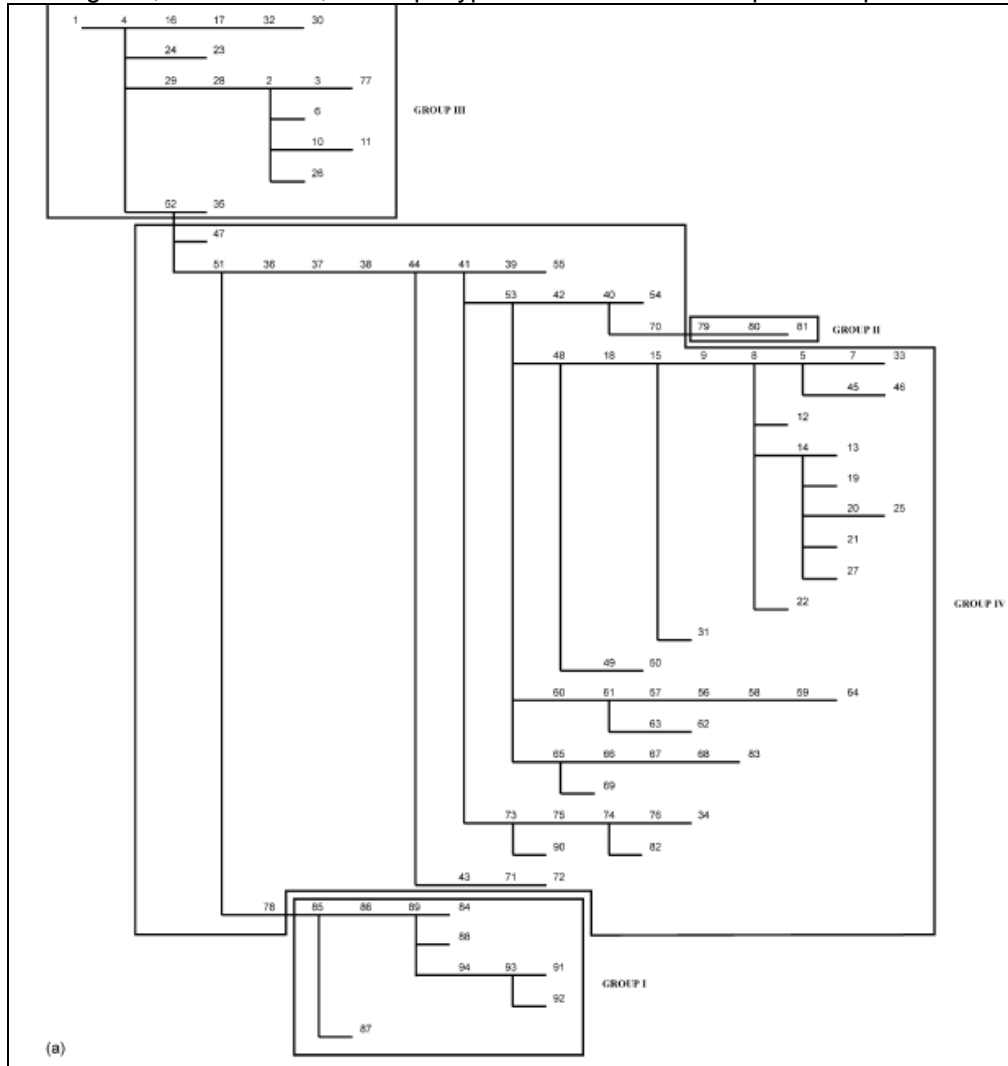


Fig. 1. (Continued).



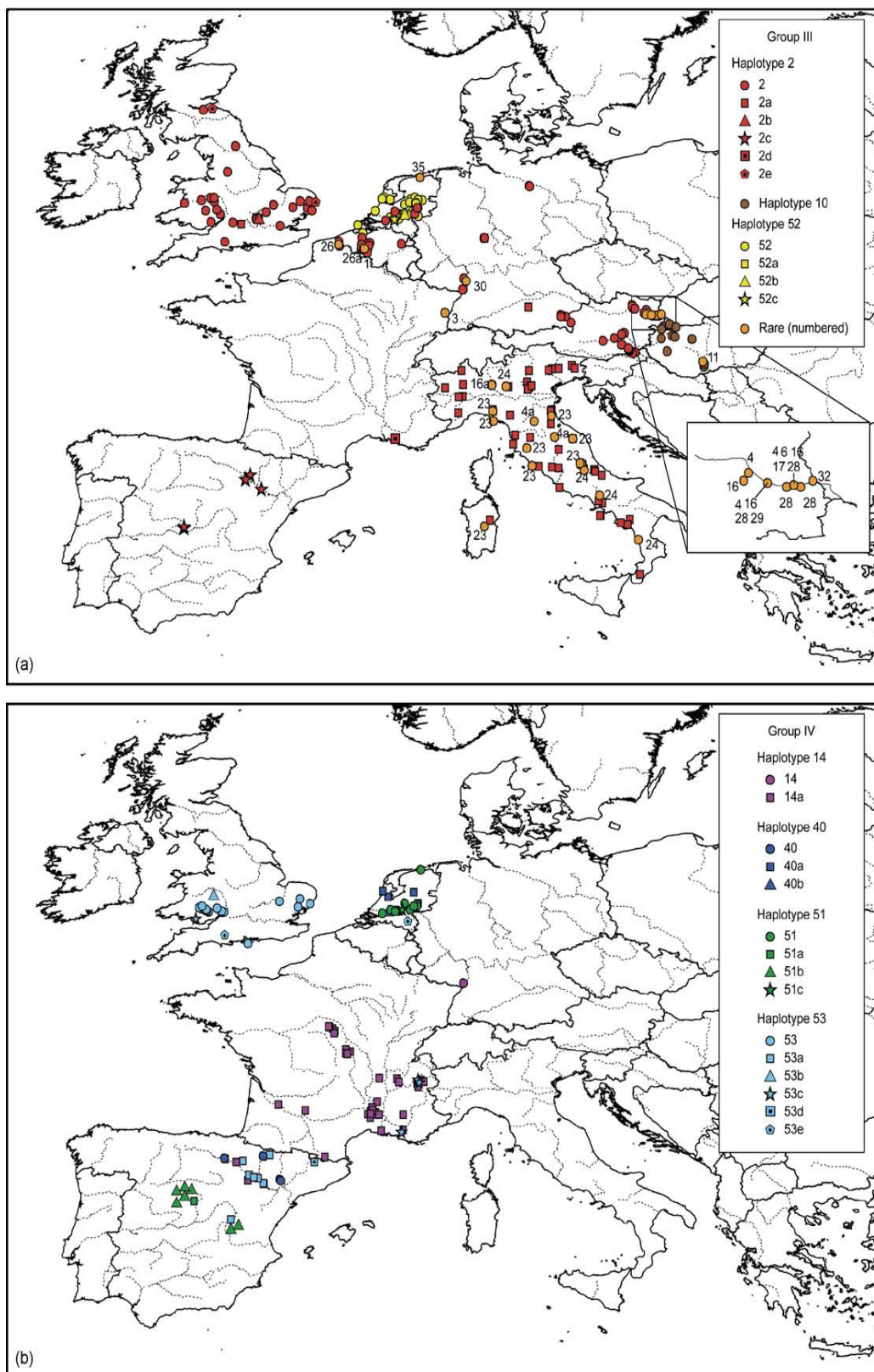


Fig. 3. Summary map showing distribution of haplotypes from Groups III and IV and suggested routes of colonisation.

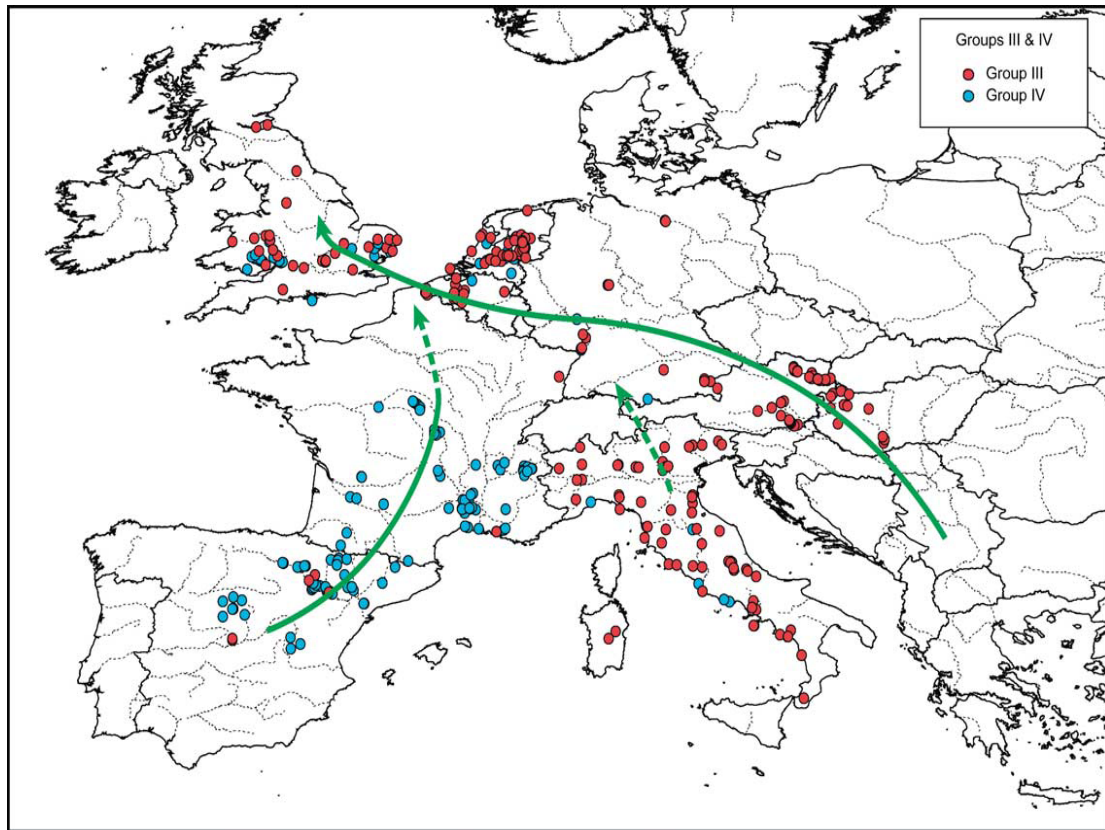


Table 2 Details of the size of each PCR fragment, primer sequence and annealing temperature for each of the cpDNA regions that were studied

CpDNA fragment	Approximate size of fragment(bp)	Primer sequence 5' → 3' forward	Primer sequence 5' → 3' reverse	Annealing temperature (°C)	Number of PCR cycles	Reference
CD	2300	TTCCCGGTAGCACATACACA	TTGGGCTGCTTTGATGGTAG	58	40	Fluch <i>et al.</i> , (2002)
TF	1500	CATTACAAATGCGATGCTCT	ATTTGAACTGGTGACACGAG	58	40	Taberlet <i>et al.</i> , (1991)
DT	800	ACCAATTGAACTACAATCCC	CTACCACTGAGTTAAAAGGG	55	40	Dumolin-Lapeague <i>et al.</i> , (1997)
ORF-M	1050	CTTGCTTTCCAATTGGCTGT	CATAACCTTGAGGTCACGGG	55	30	Heinze (1998a,b)
17 + 20	800	GAAGTAGTAGGATTGATTCTC	CCCTACAACCTCATGAATTAAG	55	20	Samuel <i>et al.</i> , (1997)

Table 4 The banding pattern of each haplotype produced with the six primer/restriction enzyme combinations

Haplotype	Group	CD / EcoRI-1	CD / EcoRI-2	CD / EcoRI-3	CD / EcoRI-4	CD / EcoRI-5	CD / TaqI-1	CD / TaqI-2	CD / TaqI-3	CD / TaqI-4	CD / TaqI-5	TF / TaqI-1	TF / TaqI-2	TF / TaqI-3	DT / MnlI-1	DT / MnlI-2	DT / MnlI-3	ORF-M / EcoRI-1	ORF-M / EcoRI-2	17 + 20 / HinfI-1	17 + 20 / HinfI-2	17 + 20 / HinfI-3
1	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
2	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2a	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	1	1	1
2b	III	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	1	1	1	1	1
2c	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	1	1
2d	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-
2e	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-
3	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
4	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
4a	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	2	1	1
16	III	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	2	1	1
16a	III	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	1	2	1	1		
17	III	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	1	1	1	2	1	1
26	III	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1
26a	III	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	-	-	1	1	1
10	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
11	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	1	1	1	1
52	III	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
52a	III	1	2	1	1	1	1	1	1	1	1	1	-	1	1	1	1	1	1	2	1	1
52b	III	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	1	1
52c	III	1	2	1	1	1	1	1	1	1	1	-	-	-	1	1	1	1	1	2	1	1
35	III	1	2	1	1	1	1	1	1	1	1	-	-	-	1	3	1	1	1	2	1	1
77	III	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-
32	III	1	1	1	1	1	2	1	1	1	1	1	1	1	2	3	1	1	1	1	1	1
28	III	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
29	III	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
30	III	1	1	1	1	1	2	1	1	1	1	1	1	1	1	3	1	1	1	1	1	2
23	III	1	1	1	1	1	1	1	1	1	1	3	1	0	1	1	1	-	-	1	1	1
24	III	1	1	1	1	1	1	1	1	1	1	3	1	0	1	1	1	-	-	2	1	1
6	III	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8	1	1	1
5	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	3	1	1
13	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	-	-	1	1	1

cpDNA marker development and application in black Poplar

20	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	1	-	-	1	1	1
22	IV	1	1	1	1	1	1	1	1	1	1	1	1	4	4	1	1	-	-	1	1	1
25	IV	1	1	1	1	1	1	1	1	1	1	3	1	0	3	3	1	-	-	1	1	1
7	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	-	-	-
7a	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	2	
33	IV	1	1	1	1	1	3	1	1	1	1	1	1	1	1	-	1	2	2	1	1	2
8	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	1	1	1
8a	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	-	-	-	-
9	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	1	1	2
12	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	2	3	-	-	-
14	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	2	3	1	1	1
14a	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	3	1	1	1	1
15	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	3	1	1	2
18	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	2	2	3	1	1	2
19	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	1	2	3	-	-	-
21	IV	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	2	3	-	-
27	IV	1	1	1	1	1	1	1	1	1	5	1	1	1	1	1	-	1	2	3	1	1
31	IV	1	1	1	1	1	2	1	1	1	1	1	1	1	1	3	1	2	3	1	1	2
34	IV	1	1	2	3	2	1	1	1	1	1	1	1	1	1	-	1	2	3	1	1	1
36	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	3	1	1
37	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	3	1	2
38	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	1	3	1	2
47	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	-	2	1	1	2	1	1
51	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
51a	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
51b	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
51c	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	1	1	-	1
54	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	5	-	1	2	1	2	1	2
39	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	2	3	1	2
55	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	3	1	-	1	2	2	3	1
40	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	2	3	2	1
40a	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	3	2	1
40b	IV	1	2	1	1	1	1	1	1	1	1	1	1	-	-	-	3	1	2	3	2	1
41	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	3	3	1
42	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	3	5	1
48	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	-	2	2	3	1	1
49	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	-	2	2	3	2	1
50	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	-	2	3	3	2	1
53	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	2	3	1	1
53a	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	3	1	1
53b	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	-	-	1	1
53c	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	2	3	1	-
53d	IV	1	2	1	1	1	1	1	1	1	1	1	-	-	-	1	1	2	3	1	1	2
53e	IV	1	2	1	1	1	1	1	1	1	1	1	-	-	-	1	3	1	2	3	1	1
56	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	3	1	-	1	2	3	1	1
57	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	2	3	1	1
58	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	2	3	-	-
59	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	3	2	3	1	2	3	-	-
60	IV	1	2	1	1	1	1	1	1	1	1	1	1	2	1	1	3	1	2	3	1	1
61	IV	1	2	1	1	1	1	1	1	1	1	1	1	2	3	1	-	1	2	3	1	1
62	IV	1	2	1	1	1	1	1	1	1	1	1	1	2	3	1	-	1	2	3	3	1
63	IV	1	2	1	1	1	1	1	1	1	1	1	1	2	3	1	-	1	2	3	3	1
64	IV	1	2	1	1	1	1	1	1	1	1	1	1	4	4	2	3	1	2	3	-	-
65	IV	1	2	1	1	1	1	1	1	1	4	1	1	1	1	1	-	1	2	3	1	1
66	IV	1	2	1	1	1	1	1	1	1	4	1	1	1	3	1	-	1	2	3	1	1
69	IV	1	2	1	1	1	1	1	1	1	4	4	1	1	1	1	-	1	2	3	3	1
70	IV	1	2	1	1	1	2	1	1	1	1	1	1	1	1	1	-	1	2	3	2	1
73	IV	1	2	1	3	2	1	1	1	1	1	1	1	1	1	1	-	1	2	3	3	1
74	IV	1	2	2	3	2	1	1	1	1	1	1	1	1	1	1	-	1	2	3	2	1
75	IV	1	2	2	3	2	1	1	1	1	1	1	1	1	1	1	-	1	2	3	3	1
76	IV	1	2	2	3	2	3	1	1	1	1	1	1	1	1	1	-	1	2	3	1	1

cpDNA marker development and application in black Poplar

43	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	6	2	1	2	
44	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	-	1	2	6	3	1	2	
67	IV	1	2	1	1	1	1	1	1	4	1	1	1	3	1	-	1	2	6	1	1	2	
71	IV	1	2	1	1	1	3	1	1	1	1	1	1	1	1	-	1	2	6	2	1	2	
72	IV	1	2	1	1	1	3	1	1	1	1	1	1	1	3	-	1	2	6	2	1	2	
45	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	-	1	3	7	3	1	1	
46	IV	1	2	1	1	1	1	1	1	1	1	1	1	1	1	-	1	3	7	3	1	2	
68	IV	1	2	1	1	1	1	1	1	4	1	1	1	3	1	-	1	3	7	1	1	2	
83	IV	3	2	1	1	1	1	1	1	5	1	1	1	3	1	-	1	3	7	2	1	2	
78	IV	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	0	1	1	
82	IV	2	2	3	4	3	1	1	1	1	1	1	1	1	1	-	1	2	3	2	1	2	
90	IV	4	3	1	3	2	1	1	1	1	1	1	1	1	1	3	1	2	3	-	-	-	
79	II	2	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	-	-	1	1	2	
80	II	2	2	1	1	1	2	2	2	-	1	2	2	2	1	1	1	-	-	1	1	2	
81	II	2	2	1	1	1	2	2	2	2	1	2	2	2	1	1	2	2	2	2	1	1	
84	I	4	3	1	1	1	1	1	1	1	1	1	4	4	4	4	0	4	5	0	1	3	
85	I	4	3	1	1	1	1	1	1	3	1	1	1	1	1	1	1	4	5	0	1	3	
86	I	4	3	1	1	1	1	1	1	3	1	1	1	1	4	4	0	4	5	0	1	3	
87	I	4	3	1	1	1	1	1	1	3	1	1	4	4	1	1	1	1	1	0	1	3	
88	I	4	3	1	1	1	1	1	1	3	1	1	4	4	4	4	0	2	3	0	1	3	
89	I	4	3	1	1	1	1	1	1	3	1	1	4	4	4	4	0	4	5	0	1	3	
91	I	4	3	1	3	2	4	1	1	3	3	1	4	4	4	0	4	0	2	2	4	2	3
92	I	4	3	1	3	2	4	1	1	3	3	1	4	4	4	-	0	4	2	6	1	3	
93	I	4	3	1	3	2	4	1	1	3	3	1	4	4	4	4	0	2	2	4	2	3	
94	I	4	3	1	3	2	4	1	1	3	3	1	4	4	4	4	0	4	5	4	2	3	

Missing values are indicated by a dot. Haplotypes are numbered (1, 2, . . . , 94) and are arranged according to membership of individual groups. A letter following the haplotype number indicates that there are missing values in the analysis but that all the fragments obtained are identical to those of the haplotype with the same identifier number but without the following letter.

2.4 The mitochondrial DNA – specific work

2.4.1 Introduction

The mitochondrial DNA in plants is double-stranded, circular or linear and sometimes forms sub-genomic circles, much smaller and simpler than the nuclear genome. It replicates independently. Mitochondrial genomes of higher plants are more than ten times as large as mitochondrial DNAs in animals. Mitochondrial DNAs have shown a wide variation in genome size and complexity between various plant species with average size of mt genomes ranging from 400–500 kb (Brennicke *et al.*, 1996; Marienfeld *et al.*, 1999). The size difference is explained by additional genes, by internal duplications, and by integrated nuclear and chloroplast sequences. An example for an average mt genome size is *Brassica* sp with about 200 kb (Palmer & Stein 1986a) whereas an exceptionally big mt genome can be found in muskmelon with 2400 kb. Currently (as of Oct. 26th 2009), 1946 full mitochondrial genome sequences of 450 different species (mostly animal) and only few plant mtDNA sequences like rice, sorghum or sugar beet (Kubo *et al.*, 2000) are available from (<http://www.ncbi.nlm.nih.gov/genomes/> Organelles), which have been generated in the course of various genome sequencing projects. Still the configuration of the plant mt genome remains unsolved as circular chromosomes as well as associated mt plasmids are being reported.

The organization of the mitochondrial genome of higher plants is complex. It has two striking features: a large size that can vary among plant species; and the ability to undergo homologous recombination that results in variation within species. Analysis has shown that the mitochondrial genome often is arranged into a single circular molecule, referred to as the master chromosome. This circular DNA molecule contains repeated sequences that can generate, via intramolecular recombination, either isomeric forms of the master chromosome or smaller subgenomic circular DNA molecules. For Maize it has been shown that the mt genome organization varies considerably among the different maize cytotypes (Fauron *et al.*, 1995).

In each somatic cell 10^2 - 10^5 mitochondria can be found. Depending on tissue and physiological state, each organelle contains up to 10 copies of the mtDNA molecule. In general, even though pollen as well as ovary cells contain mitochondria, only one type of mtDNA is found within an individual (one haplotype), meaning that the mtDNA from one parent is lost after fertilization of the ovary cell. Plant mtDNA is almost exclusively maternally inherited, unlike cpDNA, which generally is maternally inherited in angiosperms, but paternally in gymnosperms (Birky 2001; Vendramin *et al.*, 2000b; Vendramin & Ziegenhagen 1997). Rare cases of paternal transmission of mtDNA have been reported (McCauley *et al.*, 2007; Pearl *et al.*, 2009), particularly in interspecific crosses, but paternal mtDNA is rapidly degraded in the oocyte cytoplasm in intraspecific fertilisation. In general the mitochondrial genome is maternally transmitted as a single haploid supergene.

Nucleotide substitution occurring in the mitochondrial genome has been extensively exploited in animals due to typically high mutation rates (Wolfe *et al.*, 1987) and has been particularly useful for phylogenetic studies. In contrast, the plant mitochondrial genome, which is larger in size, undergoes extensive structural rearrangements. The average synonymous substitution rate for plant mitochondrial genomes is around $0.2\text{--}1.0 \times 10^{-9}$ nucleotides per site per year, 40–100 times lower than the corresponding rate in mammals ($20\text{--}55 \times 10^{-9}$ nucleotides per site per year; (Brennicke *et al.*, 1996; Wolfe *et al.*, 1987), but still 5–10 times faster than nuclear single copy number DNA, where coding genes are conserved. The approximately 100 protein-coding genes in plant mitochondria are transcribed from scattered promoters with unique structures. Nonlethal mutations in mitochondrial genes affect pollen maturation and lead to male sterile plants (Engelke *et al.*, 2004; Engelke & Tatlioglu 2000).

MtDNA currently is being used for DNA barcoding efforts for various species. The DNA barcoding initiatives rely on the premise that all organisms have a unique and identifiable molecular tag i.e a short region of mtDNA of the cytochrome oxidase subunit 1 (COI) that can be amplified using universal primers (Song *et al.*, 2008). Comparing the sequences of this region originating from various species allows the generation of phylogenetic relations between species.

The slower rate of mtDNA evolution in plants as compared to cpDNA, less background knowledge, its structural rearrangements and thus not conserved gene order makes marker development for mtDNA more difficult, therefore, limiting its use as tool for plant diversity studies (Bereiter-Hahn *et al.*, 2008; Cummings *et al.*, 2003; Fauron *et al.*, 1995; Galtier *et al.*, 2009). Due to these reasons mtDNA has been investigated only in some angiosperms where mtDNA is transmitted maternally whereas cpDNA migrates with the pollen (Scotti *et al.*, 2008; Sperisen *et al.*, 2001a). The advantage of cpDNA markers over mtDNA markers is the fact that the cp genome is conserved in structure and gene rearrangements are rare in cpDNA (Reboud & Zeyl 1994; Sugiura 1992; Sugiura 2003) allowing the transfer of markers between species, thus substantially easing access to this marker system.



Fig. 11: sucrose gradient with intact mitochondria extracted from *P.abies* callus cultures

2.4.2 Mitochondrial DNA marker development and application in Norway spruce

Norway spruce (*Picea abies*) was among the first tree species which re-colonised central Europe after the last glaciation, reaching the Alpine area approximately 12 000 years ago (Huntley & Birks 1983; Lang 1992; Tollefsrud *et al.*, 2008). In Austria, Norway spruce is a naturally occurring forest tree species that is of ecological and economical importance. In total, 1.87 million ha or 56.1 % of the productive forest area is covered by spruce. The Austrian spruce forests are found from lower to sub-alpine regions. Studies on genetic variation based on allozymes (Lagercrantz & Ryman 1990) or molecular markers (Gugerli *et al.*, 2001; Scotti *et al.*, 2006;) have shown that, over the whole range of *Picea abies*, genetic differentiation among populations is low. This may be regarded as a result of the last glacial period as well as as the effect of strong gene flow among populations especially as large-scale planting may have a drastic impact on the genetic uniformity of Norway spruce throughout Europe. The first, more extensive artificial sowing of Norway spruce was reported at the end of the 18th century in Upper Austria. Later, significant plantings were carried out throughout Austria. During the 19th century, extensive pure spruce forests were artificially established, especially in mountainous regions (Geburek *et al.*, 1998a; Perry *et al.*, 1999).

In order to assess existing genetic diversity in Norway spruce, molecular markers need to be developed and applied. Grivet (Grivet *et al.*, 1999) first demonstrated maternal inheritance of mtDNA in *Picea abies*. A PCR based technique based on mtDNA fragments generated by one pair of universal mitochondrial primers, followed by restriction enzyme analysis was established. Population genetics studies of Norway

spruce using maternally inherited mitochondrial tandem repeats (Bastien *et al.*, 2003; Gugerli *et al.*, 2001) and paternally inherited chloroplast microsatellites (Vendramin *et al.*, 2000a) reveal centres of diversity in the Carpathian mountains, the Baltic region, and the eastern Alps. The observed diversity appears to result from different phylogeographic processes, such as areas of glacial relict populations or of secondary contact.

Sperisen (Sperisen *et al.*, 2001b) investigated a tandem repeat region in the second intron of the mitochondrial NADH dehydrogenase subunit 1 (*nad 1*) gene of Norway spruce, which was found to be polymorphic and maternally inherited in intra-specific crosses of this species. Gugerli *et al.*, (2001) used the same region to show that the haplotypic diversity in the Alps is reduced in populations at high altitude as compared to populations at low elevation. Other than these regions, no molecular markers of the mt genome were available for Norway spruce.

Therefore an attempt has been made to characterize novel regions of the mt genome of Norway spruce which are variable and can be used as markers for population genetic analysis.

2.4.2.1 Development of mitochondrial markers for population genetics of Norway spruce [*Picea abies* (L.) Karst.]

F. Maghuly^{1), 4)}, K. Burg¹⁾, F. Nittinger²⁾, W. Pinsker²⁾, W. Praznik³⁾ and S. Fluch¹⁾

¹⁾ Austrian Research Center Seibersdorf, Division of Biogenetics and Natural Resources, 2444 Seibersdorf, Austria.

²⁾ Museum of Natural History Vienna, Burgring 7, 1010 Vienna, Austria.

³⁾ University of Natural Resources and Applied Life Sciences, Department of Chemistry, Muthgasse 18, 1190 Vienna, Austria.

⁴⁾ Present address and correspondence. University of Natural Resources and Applied Life Sciences, Institute of Applied Microbiology, Plant Biotechnology Unit, Nussdorfer Lände 11, 1190 Vienna, Austria.

Published: *Silvae Genetica*, 57(1) p. 41-44

Keywords: mtDNA, marker development, maternal inheritance, population genetics, length polymorphism, null alleles, *Picea abies*

Abstract

Norway spruce is an important commercial tree species in northern and central Europe. Pure mitochondrial DNA isolated from tissue culture materials grown in the dark were used to construct a partial mitochondrial library. 100 clones were randomly selected and 19 markers were isolated. Three of these markers proved to be polymorphic and two showed maternal inheritance in controlled crosses. These markers will be useful for population genetic studies in *P. abies*.

Introduction

Norway spruce is one of the commercially and ecologically most important coniferous species in Europe. Therefore, knowledge on the existing genetic variation and population differentiation is helpful for various aspects of management of its genetic resources. Plant mitochondrial DNA (mtDNA) is exclusively maternally inherited (NEALE and SEDEROFF, 1989; GRIVET *et al.*, 1999), unlike chloroplast DNA (cpDNA), which generally is maternally inherited in angiosperms, but paternally in gymnosperms (VENDRAMIN *et al.*, 1996). This differential mode of inheritance makes organellar genomes an ideal tool for studies of gene flow and population structures (MITTON *et al.*, 2000; GUGERLI *et al.*, 2001). From the geographic distribution of variation in the organellar DNA it is possible to draw conclusions about past gene flow through migration. In contrast to bi-parentally transmitted markers, maternally inherited alleles are expected to exhibit greater levels of population differentiation at equilibrium, due to the absence of meiotic recombination and to the fact that haploid

genomes are more sensitive to the effects of random drift (PETIT *et al.*, 1993; ENNOS, 1994). Furthermore, mtDNA markers are maternally inherited and thus indicative of seed dispersal, while cpDNA (and nuclear DNA) is paternally (or biparentally) inherited and thus indicative of gene flow by pollen and seeds, not only seeds (LATTA *et al.*, 1998).

The mitochondrial (mt) genome of plants is not conserved in size and structure between species (WOLFE *et al.*, 1987; PALMER, 1992; PALMER *et al.*, 2000). Therefore, markers developed for a particular species often cannot be applied for other taxa. Up to now, a few mitochondrial regions have been identified in Norway spruce which show only moderate degrees of diversity throughout Europe (SPERISEN *et al.*, 2001; GUGERLI *et al.*, 2001; JEANDROZ *et al.*, 2002; BASTIEN *et al.*, 2003; JARAMILLO-CORREA *et al.*, 2003).

In this study, we developed new primers for the mitochondrial genome of *P. abies*, which will be useful in population genetics and provide valuable information for forest management. Pure mtDNA of Norway spruce was isolated from cell cultures according to a modified method after MACKENZIE (1994). Out of 600 g of tissue (dark grown embryogenic callus cultures) 1-2 µg of mtDNA could be isolated. The purity of mtDNA was tested by PCR using published primer pairs specific for cpDNA, nuclear DNA and mtDNA. The DNA was considered as authentic if amplification was successfully achieved only with the mtDNA specific primers (data not shown). Due to the limited amount of mtDNA available, the Random Amplified Polymorphic DNA (RAPD) technique was used to generate clonable fragments for the construction of partial mt genome libraries. Each RAPD amplification reaction contained 1x buffer (Dynazyme), 2 mM MgCl₂, 10 pmol primer (OPERON: OPD-02, OPE-20, OPH-20, OPJ-20), 0.75 unit Dynazyme II DNA polymerase, 20-30 ng DNA template and PCR water up to 25 µl. PCR-cycling conditions consisted of an initial denaturation step of 94°C for 3 min, followed by 40 cycles of 94°C for 1 min, 36°C for 1 min, 72°C for 2 min, and a final extension step at 72 °C for 8 min. PCR products obtained with different RAPD primers were cloned using the TOPO TA cloning Kit for sequencing (Invitrogen). Purified plasmid DNA of the clones was sequenced (BigDye sequencing Kit, PE Applied Biosystem) on a capillary sequencer (ABI 3100 Genetic Analyser) and analysed using softwares Sequencher™ 4.0 and DNAsis 2.5. Out of 7200 identified positive clones a total of 100 were randomly selected and sequenced. Primers were designed (Oligo Primer Analysis Software; version 4.1) for 19 mtDNA-bearing clones. The clones were selected according to homology to known mt sequences or the length of the insert (> 1000 bp). Out of 19 potential loci, 14 were effectively amplified, but 11 of them proved to be monomorphic even with PCR-RFLP analysis. In total, three of the investigated loci (*Table 1*) were polymorphic: mt15-D02, mt23-D02, and mt1H01. Maternal inheritance was tested in a set of 11 controlled crosses (5-8 offsprings of each family) using different combinations of five male and six female parents (obtained from SkogForsk/Sweden and the Federal Research Centre for Forestry and Forest Products BFH Hamburg-Grosshansdorf/Germany).

For the analysis of polymorphism at the three loci, PCR amplifications were conducted in a total volume of 25 µl according to the HotStarTaq polymerase (QIAGEN) instructions. PCR-cycling consisted of an initial denaturation step of 95°C for 15 min, followed by 35 cycles of 1 min at 95°C, 1 min at annealing temperature (see *Table 1*), 2 min at 72°C, and a final extension step of 10 min at 72°C.

Maternal inheritance was proved for markers mt15-D02 and mt23-D02 in controlled crosses. For marker mt15-D02 two alleles corresponding to PCR fragments of 1107 bp (mitotype *a*) and 1249 bp (mitotype *b*) were detected. All female parents yielded mitotype *a*, while all male parents yielded mitotype *b*. Analysis of all 70 F₁ individuals from the 11 families possessed the maternal mitotype *a*, thus confirming maternal inheritance (*Figure 1a* and *Table 1*). With the primers for marker mt23-D02 fragments of either 1117 bp (mitotype *a*) or 1122 bp (mitotype *b*) were obtained, or no amplification product at all (mitotype *c*) (*Figure 1b* and *Table 1*). For the analysis of marker mt23-D02 a high-resolution method such as PCR-RFLP was necessary to visualize the 5 bp size difference between mitotypes *a* and *b* (*Figure 2*). The PCR products were digested with restriction enzyme *Hinf*I (5 U; 3h at 37°C) and separated through 8% non-denaturing polyacrylamide gels [Rotiphorese® Gel 30 (37,5:1)] at 14°C for 3 h using Dual Vertical Slab System DSG-250 (C.B.S. Scientific), followed by silver staining for band detection (DUMOLIN *et al.*, 1995; DEMESURE *et al.*, 1996). Of the 40 F₁ individuals from the various crosses all possessed the maternal mitotype (data not shown). The third marker (mt1H01) showed a monomorphic 720 bp fragment (*Figure 1c*) in all eleven parents (mitotype *a*). In a population study (MAGHULY *et al.*, 2007), however, some individuals occurred from which no amplification product could be obtained. The lack of a PCR fragment was interpreted as a null allele (mitotype *b*) that may be caused by putative mutations in the primer binding sites (*Table 1*).

In mt15-D02 sequence analysis of the two different mitotypes *a* and *b* (*Table 1*) revealed the presence of a 142 bp insertion/deletion (indel). In mt23-D02 mitotypes *a* and *b* differ by a 5 bp indel and mitotype *c* consists of a null allele (*Table 1*).

BLASTN searches revealed that marker mt1H01 is highly similar (91%) to a 729 bp section of the mt *nad1* gene of the evening primrose *Oenothera berteriana* [Score = 908 bits (458), E-value = 0.0]. For the other two markers similarity was found only to very short sections of known mt sequences, which is quite meaningless with respect to homology. For region mt23-D02 homology was found to 32 bp of the mt *nad1* genes of various *Picea* species [Score = 63.9 bits (32), E-value = 1e-06]. Region mt15-D02 showed homology to 25 bp of the mt rRNA gene (large subunit) of *Ceratophyllum submersum* [Score = 42.1 bits (21), E-value = 6.2].

All three markers for the mt genome were already successfully employed in the analysis of 37 populations (185 individuals) originating from various geographic regions of Austria (MAGHULY *et al.*, 2006, 2007) and thus provide a valuable tool for further population studies of *P. abies*. They will be also useful for forest management programs and for reconstructing the phylogeographic history and postglacial migration routes of *P. abies*.

Acknowledgement

This study was supported by the Austrian Exchange Service (Österreichischer Austauschdienst, ÖAD).

References

- JARAMILLO-CORREA, J. P., J. BOUSQUET, J. BEAULIEU, N. ISABEL, M. PERRON and M. BOUILLÉ (2003): Cross-species amplification of mitochondrial DNA sequence-tagged-site markers in conifers: the nature of polymorphism and variation within and among species in *Picea*. *Theoretical and Applied Genetics* **106**: 1353–1367.
- BASTIEN, D., J. M. FAVRE, A. M. COLLIGNON and C. SPERISEN (2003): Characterization of a mosaic minisatellite locus in the mitochondrial DNA of Norway spruce [*Picea abies* (L.) Karst.]. *Theoretical and Applied Genetics* **3**: 574–580.
- DUMOLIN, S., B. DEMESURE and PETIT R. J. (1995): Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics* **91**: 1253–1256.
- DEMESURE, B., B. COMPAS and R. J. PETIT (1996): Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica*) in Europe. *Evolution* **50**: 2515–2520.
- Ennos, R. A. (1994): Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**: 250–259.
- GRIVET, D., S. JEANDROZ and J. M. FAVRE (1999): *Nad1* b/c intron polymorphism reveals maternal inheritance of the mitochondrial genome in *Picea abies*. *Theoretical and Applied Genetics* **99**: 346–349.
- GUGERLI, F., C. SPERISEN, U. BÜCHLER, F. MAGNI, T. GEBUREK, S. JEANDROZ and J. SENN (2001): Haplotype variation in a mitochondrial tandem repeat of Norway spruce (*Picea abies*) populations suggests a serious founder effect during postglacial re-colonization of the western Alps. *Molecular Ecology* **10**: 1255–1263.
- Jeandroz, S., D. Bastien, A. Chandelier, P. Du Jardin and J. M. Favre (2002): A set of primers for amplification of mitochondrial DNA in *Picea abies* and other conifer species. *Molecular Ecology Notes* **2**: 389–392.
- Latta, R. G., Y. B. Linhart, D. Fleck and M. Elliot (1998): Direct and indirect estimates of seeds versus pollen movement within population of ponderosa pine. *Evolution* **44**: 38–53.
- MACKENZIE, S. (1994): Isolation of plant mitochondria and mitochondrial nucleic acids, pp. D3/1-D3/12 in *Plant Molecular Biology Manual*, edited by S. B. GELVIN, R. A. SCHILPEROORT and D. P. S. VERMA, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- MAGHULY, F., W. PINSKER, W. PRAZNIK and S. FLUCH (2006): Genetic diversity in managed subpopulations of Norway spruce [*Picea abies* (L.) Karst.]. *Forest Ecology and Management* **221**: 266–271.
- MAGHULY F., F. NITTINGER, W. PINSKER, W. PRAZNIK and S. FLUCH (2007) Differentiation among Austrian populations of Norway spruce [*Picea abies* (L.) Karst.] assayed by mitochondrial DNA markers. *Tree Genetics and Genomes*, in press, DOI 10.1007/s11295-006-0055-z.
- MITTON, J. B., B. R. KREISER and R. G. LATTA (2000): Glacial refugia of limber pine (*Pinus flexilis* James) inferred from the population structure of mitochondrial DNA. *Molecular Ecology* **9**: 91–7.
- NEALE, B. and R. R. SEDEROFF (1989): Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in Loblolly pine. *Theoretical and Applied Genetics* **77**: 212–216.
- PALMER J. D. (1992): Mitochondrial DNA in plant systematic: application and limitations, pp. 36–49 in *Molecular systems of plants*, edited by P.S. SOLTIS, D.E. Soltis and J. J. DOYLE, Chapman & Hall, New York.
- PALMER, J. D., K. L. ADAMS, Y. CHO, L. PARKINSON, Y.L. QIU and K. SONG (2000): Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proceeding of the National Academy of Sciences USA* **97**: 6960–6966
- PETIT, R. J., A. KREMER and D. B. WAGNER (1993): Finite island model for organelle and nuclear genes in plants. *Heredity* **71**: 630–641.

- SPERISEN, C., U. BÜCHLER, F. GUGERLI, G. MATYAS, T. GEBUREK and VENDRAMIN G. G. (2001): Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. *Molecular Ecology* **10**: 257–263.
- VENDRAMIN, G. G., L. LELLI, P. ROSSI and M. MORGANTE (1996): A set of primers for the amplification of 20 chloroplast microsatellites in Pinacea. *Molecular Ecology* **5**: 595–598.
- Wolfe, K.H., W.H. Li and P.M. Sharp (1987): Rates of Nucleotide Substitution Vary Greatly among Plant Mitochondrial, Chloroplast, and Nuclear DNAs. *Proceeding of the National Academy of Sciences USA* **84**: 9054–9058.

Table 1. - PCR primers for amplification of three mitochondrial markers

Name	PCR primer sequence 5'-3' fwd	PCR product (bp)	GenBank Acc. Nr	mitotype
mt15-D02	TATCTGACTTGCCTTATC	1107, 1249	AY897577	a, b
	ATCCGAATACATACACC			
mt23-D02	CACCCTTGGGTAGACTGG	1122, 1117	AY897578	a, b and c
	GGTTCACGCAGTGCTTCT	and n.a.		
mt1H01	AAGATGGATCGCCCTTACGC	702 and n.a.	AY897576	a, b
	GAGGAGGAGGCTTCGTCGTC			

Figure 2. - PCR-RFLP for mitochondrial markers mt23-DO2. PCR fragments were digested with restriction Enzyme Hinf I. Fragments were separated on a high resolution on acrylamide gel. Lane 1-5: F1, lane 6: male, lane 7: female, lane 8: 100 bp ladder (Boehringer).

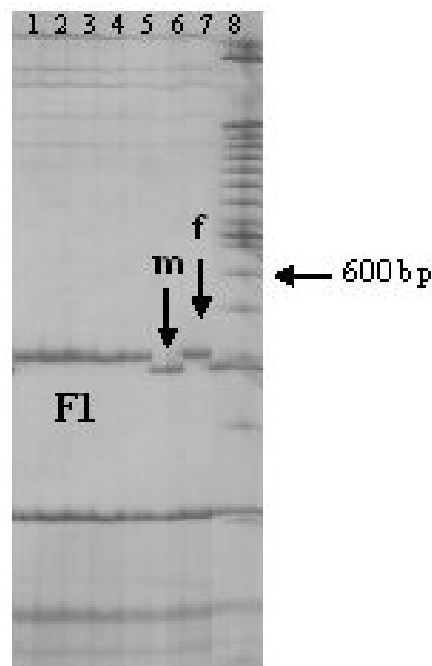
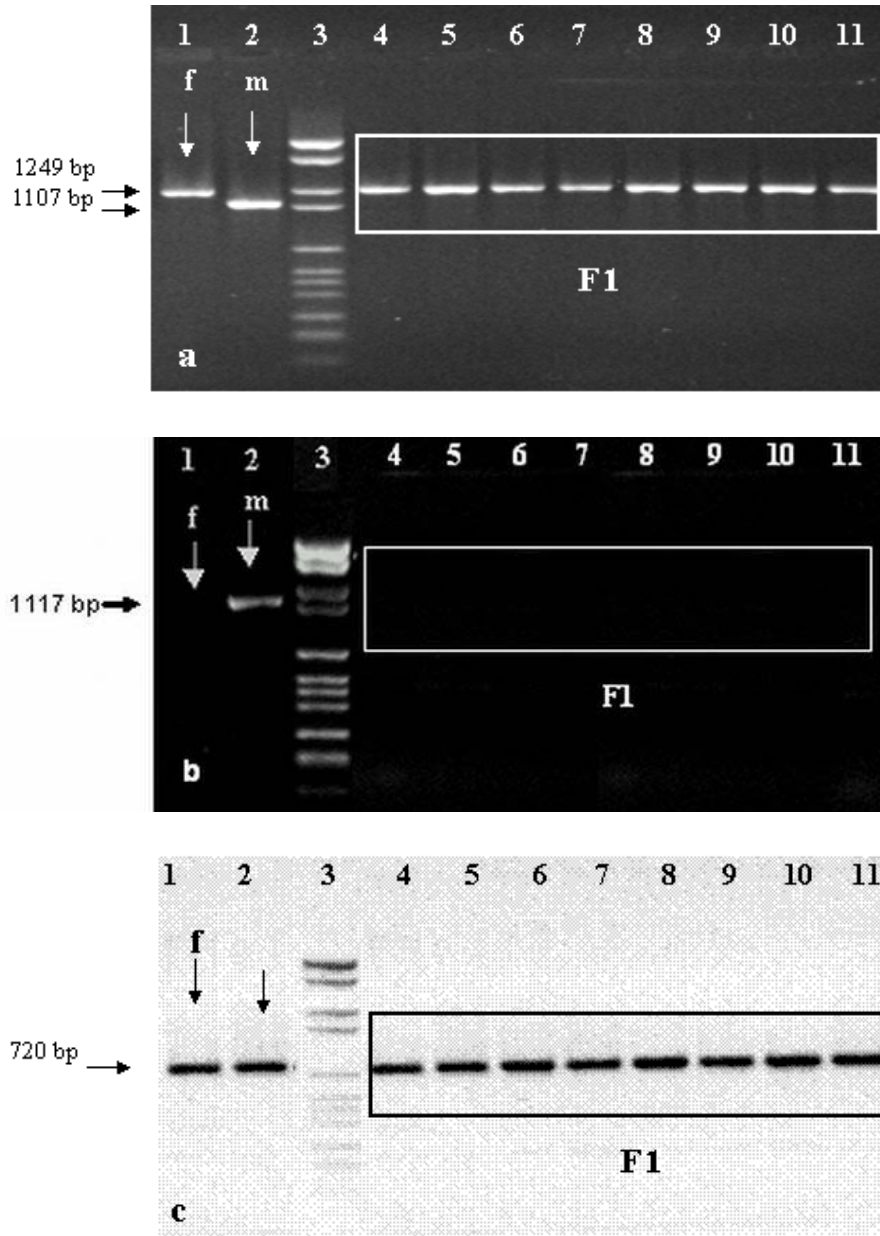


Figure 1. - Amplification products obtained with the three markers separated on agarose gels (1.2 %): a) mt15-02, b) mt23-DO2, and c) mt1H01. Parents and offspring of controlled crosses are shown. Lane 1: female parent, lane 2: male parent, lane 3: length marker VI, lanes 4-11: F₁ offspring. In mt15-02 the F₁ progeny shows the same fragment as the female parent. In mt23-DO2 the female parent and the F₁ offspring show the null allele whereas the male parent carries the 1117 bp allele. In mt1H01 the parents and offspring are monomorphic for the same allele.



2.4.2.2 Population differentiation of Austrian Norway spruce [*Picea abies* (L.) Karst.] populations assayed by mitochondrial markers

F. Maghuly¹, W. Praznik², W. Pinsker³ and S. Fluch¹

¹ARC Seibersdorf research GmbH; Biogenetics/Natural Resources; Bioresources; 2444 Seibersdorf, Austria

²University of Natural Resources and Applied Life Sciences, Department of Chemistry; Muthgasse 18, 1190 Vienna, Austria.

³Museum of Natural History Vienna, 1st Zoological Department, Molecular Systematics; Burgring 7, 1014 Vienna, Austria.

Published: *Tree Genetics & Genomes* (2007) 3: 199-206

Keywords Mitochondrial DNA . Population differentiation . *Picea abies* . Eastern Alps

Abstract

Thirty-seven populations of Norway spruce [*Picea abies* (L.) Karst.] across the Austrian Alps and Bohemian Massif were sampled to elucidate the geographical pattern of genetic differentiation. Three polymorphic mitochondrial DNA (mtDNA) loci were surveyed. Two or three alleles were detected at each locus, resulting in seven multilocus mtDNA haplotypes (A–F). Western populations proved to be monomorphic, whereas eastern and central Austrian populations were slightly to highly polymorphic, respectively. As revealed by spatial analysis of molecular variance and Monmonier's analysis, the two main haplotypes A and B are not randomly distributed. Haplotype A was restricted to central and eastern Austria, whereas haplotype B occurred in all Austrian populations but was the only haplotype identified in western populations. This pattern may be explained by different glacial refugia located in the Dinaric Alps and the Carpathian mountains.

Introduction

Norway spruce [*Picea abies* (L.) Karst.] is one of the commercially and ecologically most important coniferous species of Europe, and therefore, knowledge on the existing genetic variation and population differentiation is helpful for various aspects of genetic management. The distribution area of *P. abies* can be divided into a northeast boreal distribution and disjunct areas in the central European mountains. Previous studies of the genetic structure of *P. abies* showed pronounced differentiation between these two geographically separated regions (Vendramin *et al.*, 2000; Sperisen *et al.*, 2001; Collignon *et al.*, 2002; Acheré *et al.*, 2005), supporting the hypothesis of two distinct main glacial refugia postulated from pollen data (Huntley and Birks 1983; Lang 1994). Recent extended pollen analyses suggest a spread of *P. abies* from a putative refuge in Byelorussia and northern Russia to Fennoscandia (Giesecke and Bennett 2004). For the Alpine populations, putative

refuge areas (Huntley and Birks 1983; Lagercrantz and Ryman 1990) were located in the Carpathian mountains and the Balkans (Dinaric Alps). A genetic study based on tandem repeats in the second intron of the mitochondrial *nad1* gene (Gugerli *et al.*, 2001) confirmed that postglacial recolonization of the Alps most likely occurred from these two refugia. The western alpine populations proved to be completely monomorphic, whereas the eastern and southern alpine populations ranged from slightly to highly polymorphic.

This indicates a loss of variation owing to successive founder events after the expansion from the refugia starting about 10,000 years ago, as it is the case for many species (Hewitt 1996). However, many species in recolonized forests displayed high diversity, an apparent attenuation of founder events that can be explained through the admixture of haplotypes from merging postglacial recolonization routes (Petit *et al.*, 2003). The increase in genetic variation at mtDNA markers in secondary contact zones has been shown, for example, for *Quercus* in Switzerland (Mátyás and Sperisen 2001).

The limited sampling of Norway spruce from the eastern Alps (Bad Gastein, Dachstein and Gesäuse) in a previous study on mitochondrial tandem repeats (Gugerli *et al.*, 2001), however, did not allow more detailed inferences about the genetic structure of populations in the eastern Alps. In an earlier study on *P. abies* in Austria based on isozyme allele frequencies, Geburek (1999) found an increase in genetic variation towards the east. Hence, the question remains unanswered: How robust are the previously observed patterns of genetic structure in the eastern Alps at a finer sampling scale traced with mtDNA markers? In the present study, we report three new polymorphic markers from the mitochondrial genome of Norway spruce. We used these markers to assess the population structure in Austria at a finer scale by the extended sampling in the eastern Alps and adjacent areas of the Bohemian Massif. In doing so, we attempt to confirm the previously postulated east–west differentiation and the existence of a putative contact zone of different recolonization routes in the eastern Alps. Furthermore, we want to test the hypothesis of a decrease in genetic variation from east to west.

Materials and Methods

Plant material and sampling

Our study comprises all nine main growth districts defined by Kilian *et al.*, (1994) and characterizes the major forest climatic and geomorphological conditions in Austria: Region I=central Alps; Region II=northern alpine internal zone; Region III=south-eastern internal alpine zone; Region IV=north-western Alps; Region V=eastern Alps; Region VI=southern calcareous Alps; Region VII=northern Alp-foreland; Region VIII=south-eastern edge of the Austrian Alps and Pannonian region; Region IX=Mühl and Waldviertel comprising the Austrian part of the Bohemian Massif. Each sampling location, referred as population hereafter, consisted of five randomly selected seeds from more than 50 mother trees. Altogether, a total of 185

individuals from 37 populations were sampled (Fig. 1 and Table 1). Seed material was provided from certified seed stands used for forest plant production. Based on information from the local forest services, the seed stands studied may be regarded as autochthonous; however, ancient anthropogenic actions resulting in non-autochthonous plants cannot be excluded. Populations were consecutively numbered from the west to the east.

Molecular analysis

The collected seeds were germinated on wet filter paper in petri dishes, and approximately 5 days after germination, the seedlings (1 cm long) were used for DNA extraction using DNeasy Plant Mini Kit (QIAGEN). In a previous study, primers have been developed for 19 mitochondrial DNA (mtDNA) loci, which were screened for polymorphisms within *P. abies* from Austria (Maghuly *et al.*, unpublished). Of the primer pairs tested, polymorphism was observed only at three loci. The corresponding mtDNA regions were designated as mt15-D02 (GenBank accession number AY897577), mt23-D02 (AY897578) and mt1H01 (AY897576). Mt15-D02 exhibited a 142-bp insertion/deletion (indel) polymorphism. Mt23-D02 showed a polymorphism for a 5-bp indel, and an additional mitotype was detected consisting of a null allele, where the lack of amplification product may be caused by mutated primer binding sites. In region mt1H01, two alleles were found, one given rise to a 702-bp fragment and the other being a null allele. Maternal inheritance was proved for all three markers in controlled crosses (Maghuly *et al.*, unpublished). In a database search using BLASTN (Altschul *et al.*, 1990), the sequence mt1H01 showed high homology to the *rps13* gene and a section of the *nad1* gene of *Oenothera berteriana*. For region mt23-D02, some weak homology was found to a partial section of the *nad1* gene of various *Picea* species. Interestingly, this region showed high homology to chloroplast sequences, but this could be due to a chloroplast progenitor, as has been shown for *nad* genes in *Arabis* (Michalecka *et al.*, 2003). Region mt15-D02 showed no homology to any GenBank sequence at all. Nevertheless, the origin of all three markers from the mitochondrial genome was additionally tested by isolating pure mtDNA according to a method modified after Mackenzie (1994).

In the present study, these three mtDNA markers were used in combination to search for differentiation among spruce populations in Austria. Amplifications were conducted in a total volume of 25 μ l using 1 $\times$ PCR buffer (QIAGEN), 2 mM MgCl₂, 0.2 mM dNTPs, 4 pmol of each primer, 0.6 units HotStarTaq polymerase (QIAGEN HotStarTaq TM PCR) and 20–30 ng template DNA. PCR-cycling conditions consisted of an initial denaturation step of 95°C for 15 min, followed by 35 cycles of 1 min at 95°C, 1 min at annealing temperature (54°C for marker mt15-D02 [15F 5'-TATCTGACTTGCCTTATC-3'; 15R: 5'-ATCCGAATACATACACC-3'] and mt23-D02 [23F 5'-CA CCCTTGGGTAGACTGG-3'; 23R: 5'-GGTTCACGCAGTG CTTCT-3']; 52°C for marker mt1H01 [1H01F: 5'-AAGA TGGATCGCCCTTACGC-3'; 1H01R: 5'-GAGGAG GAGGCTTCGTCGTC-3']), 2-min extension at 72°C, and a final extension

step of 10 min at 72°C. To detect the presence of indels, PCR products were electrophoresed through 2% agarose gel. Five microliters of PCR product of region mt23-D02 was further digested with restriction enzyme HinfI (5 U; 3 h at 37°C) and separated through 8% non-denaturing polyacrylamide gels at 14°C for 3 h, followed by silver staining for band detection. This method allows visualizing by eye the 5-bp indel of marker mt23-D02. In markers mt23-D02 and mt1H01, the absence of the amplification product was treated as a character state (null allele). In this case, we carried out three independent PCR amplifications to confirm the obtained results.

Data analysis

For each spruce sample, the haplotype was defined by combining the polymorphisms detected with the three analysed marker. Four different analyses were then employed to identify possible structures in the Austrian Norway spruce populations. The first method, a cluster analysis using the unweighted pair-group method (UPGMA), is based on the analysis of a genetic distance matrix without consideration of the geographical location of populations. The other two methods, a spatial analysis of molecular variance (SAMOVA; Dupanloup *et al.*, 2002) and an analysis based on the Monmonier algorithm (Monmonier 1973), include also geographical information and aim to identify most differentiated groups of populations that are as spatially clustered as possible. For the UPGMA analysis, a minimum genetic distance (Nei 1972) implicated in the software programme TFPGA (Miller 1998) was used. Consistency of internal nodes was determined by running 1,000 bootstrap replicates.

SAMOVA was performed with the software SAMOVA 1 (Dupanloup *et al.*, 2002) to identify possible affinities among the populations of *P. abies*. This analysis defines groups of populations that are geographically homogeneous and maximally differentiated from each other. The method is based on a simulated annealing procedure that aims at maximizing the proportion of total genetic variance due to differences between groups of populations (largest FCT value). The Monmonier algorithm defines zones of maximum genetic change and genetic barriers within the network of 37 Norway spruce populations. This analysis was performed with the software BARRIER 2.2 (Manni *et al.*, 2004). Robustness of the barriers was determined by resampling the distance matrices 100 times. Moreover, we tested for the presence of correlation between geographic and genetic distance performing a Mantel's test using the procedure of Smouse *et al.*, (1986) implemented within the programme GenAlEx 6 (Peakall and Smouse 2005). Statistical significance of the values was obtained via 9,999 random permutations.

Results

Geographical distribution of mitochondrial variation

For each of the 185 individuals of *P. abies* originating from 37 different Austrian populations, the haplotype was defined by combining the polymorphisms detected

with the three markers. Two to three variants per locus that combined to seven different haplotypes (A–G; Table 2) were found. The number of distinct haplotypes per population ranged from 1 to 4 and the haplotype diversity per population ranged from 0.000 to 0.720 (Table 1).

Haplotypes A and B were the most abundant, and variation among populations and regions were evident (Fig. 3). Haplotype B, which is predominant in most of the western Austrian populations, occurred in almost all Austrian populations. Haplotype A prevailed in populations originating from eastern Austria. With the exception of populations 5 and 9, all populations from the Central Alps (region I) and the north-western Alps (IV) are fixed for haplotype B. Haplotype C was found in four populations (9, 12, 31 and 35) from regions I, VII and V. Haplotype D was present in populations 5, 28, 30, 35 and 27 from the Central Alps, the south-eastern internal alpine zone, the eastern Alps and the Mühl and Waldviertel (regions I, III, V and IX). Haplotype F was found in three geographically separated regions (regions VI, VII and VIII; populations 19, 33 and 37, respectively) in the southern calcareous Alps, the northern Alp-foreland and the Pannonian region. The less frequent haplotypes E and G were observed in eastern and south-eastern Austria. Haplotype E occurred twice in the eastern Alps (region V; populations 31 and 35). Haplotype G was found only once in population 37 from region VIII (Fig. 3).

Genetic distance using mtDNA data In a UPGMA dendrogram based on genetic distances, two main clades (clade 1 and clade 2; Fig. 2) can be observed. Both clades are subdivided into two groups. Group 1.1 is composed of populations from six regions (I, III, IV, V, VIII and IX) carrying haplotype A at high frequencies and haplotype G with the lowest frequencies. Population 19 from region VI forms group 1.2 and proved monomorphic for haplotype F. Within clade 2, group 2.1 comprises populations from regions I, II, II, IV, V, VII and IX, which have high frequencies of haplotype B. Group 2.2 includes populations 12 and 33, the only two populations from the northern Alp-foreland (region VII).

Geographical structure of mitochondrial variation In the SAMOVA, three groups could be distinguished (Table 3). Group 1 in the SAMOVA corresponds to clade 2 in the UPGMA analysis, with the exception of population 5. In the UPGMA tree, population 5 belongs to clade 1 because of the dominance of haplotype A, which is not seen in the other populations from region I. This may be explained by ancient anthropogenic influence resulting in non-autochthonous plants, as will be discussed below. Group 2 in the SAMOVA corresponds to group 1.2 of the UPGMA tree, which consists only of population 19 from the southern calcareous Alps (region VI). Group 3 corresponds to clade 2 (except population 5). A maximum FCT value of 0.766 was estimated between these three groups delineated by SAMOVA.

Monmonier's algorithm identified four genetic boundaries (Fig. 3). By definition, barriers correspond to zones of most abrupt genetic change in space. The first barrier (a in Fig. 3), which was also delineated by the SAMOVA, separated the southernmost population 19 of the southern calcareous Alps (Region VI) from all northern populations. The second barrier (b in Fig. 3) separated population 22 of the

Bohemian Massif (region IX) from all surrounding populations. The third barrier (c in Fig. 3) separated populations harbouring haplotype B (populations 24, 25, 28 and 29) of the south-eastern internal alpine area (region III) and population 34 of the south-eastern edge of the Austrian Alps (region VIII). The fourth barrier (d in Fig. 3) separated the western Austrian populations 1–18 of regions I, II, IV and VII from the eastern Austrian populations. All populations (except population 5) located westward of this boundary are dominated by haplotype B, whereas haplotype A prevailed in populations originating from the area east of boundaries c and d (line Salzburg–Graz). A Mantel test showed no significant relationship between F_{st} and geographic distance among the Norway spruce populations ($r^2=0.032$; $P=0.360$).

Discussion

Our study of three mitochondrial markers in *P. abies* revealed genetic differentiation between western and eastern populations within the eastern Alpine range in Austria. The variation in the selected mitochondrial markers also provided information about haplotype variation within different populations. Since mitochondria are maternally inherited, the seed material reflects the haplotype of mother trees.

The data analysis showed different haplotype diversity in the east and central Austrian populations as compared to the western populations. The Austrian populations located in the east were moderately to highly variable, whereas almost all western populations were monomorphic (except population 5) in the investigated sections. This result is in accordance with a previous study by Geburek (1999), revealing a slight east–west trend in the amount of genetic variation at isozyme loci in Austrian Norway spruce populations. This trend is also seen at a broader scale in the entire range of the Alps, where mitochondrial tandem repeat markers revealed monomorphic western Alpine populations but slightly to highly polymorphic eastern Alpine populations (Gugerli *et al.*, 2001).

A key result of the present study is the distribution of two most common haplotypes A and B in Austria, which appeared to be not random, as revealed by SAMOVA and Monmonier's analysis. Haplotype A is restricted to central and eastern Austria, whereas haplotype B occurs in all Austrian populations but is the only haplotype identified in western populations. It is likely that these two haplotypes became fixed in different glacial refugia. Similar patterns of haplotype distribution have been observed for various tree species in Europe and North America (*Picea*, Sperisen *et al.*, 2001; *Pinus*, Jaramillo-Correa *et al.*, 2004; *Fraxinus*, Heuertz *et al.*, 2004; *Pinus*, Godbout *et al.*, 2005). In *P. abies*, the occurrence of two main haplotypes (A and B) in our study furnishes evidence that *P. abies* populations in Austria are derived from two putative glacial refugia: the Dinaric Alps and the Carpathian mountains. This is consistent with previous results from *P. abies* in the Alps inferred from several mitochondrial and nuclear genetic markers (Collignon and Favre 2000; Gugerli *et al.*, 2001; Acheré *et al.*, 2005) but contradicts the results obtained with isozyme markers from Austrian Norway spruce populations (Geburek

1999), where no differentiation among the Austrian populations was found. Nevertheless, the present data do not reveal the origin of the two main haplotypes A and B with certainty since the adjacent areas of the Carpathian mountains and the Balkan peninsula have not been sampled yet.

However, according to findings of Collignon and Favre (2000) and Acheré *et al.*, (2005), the western Austrian populations (haplotype B in our study) belong to a Dinaric lineage. In contrast, the more eastern haplotype A might represent the genetic lineage that has migrated from a Carpathian refuge. Both lineages (A and B) met during their postglacial expansion in the easternmost Alpine range and exchanged genetic material resulting in the highly mixed and polymorphic populations seen in our study. Günzl (1979) reported that provenances of Norway spruce of the Bohemina Massif show similarities to provenances originating from the Carpathian mountains, which supports Kral's (1979) hypothesis that populations originating from a Carpathian refuge influenced this region. In our data, populations 20 and 27 are monomorphic for haplotype A. On the other hand, populations 22 and 33 are dominated by haplotype B. However, these latter populations might be non-autochthonous since a large portion of the low-elevation forests in Europe have been intensively used and managed (Schmidt-Vogt 1986). For the high-elevation forests in the Inner Alps, this seems less likely. The occurrence of the mixed population 5 (haplotypes A, B and D) from Tyrol, however, could also be best explained by human actions resulting in introduced haplotypes since all other populations from that region are either fixed for haplotype B or at least dominated. We could assume that the less frequent haplotypes C–G might have derived from the two most abundant haplotypes A and B, a way of interpretation suggested by Crandall and Templeton (1993). For instance, the frequent haplotype A might be ancestral to the less frequent haplotype E, which is only found in coexistence with haplotype A. However, we cannot rule out the possibility that less frequent haplotypes stem from a neighbouring gene pool beyond the borders of Austria, which may be dominated by these haplotypes. This could also be the case for the rare haplotype F, which is fixed in population 19 from the southern Alps. Haplotype F is also found in the easternmost population 37 from the Pannonian region and in population 33. This distinct distribution pattern makes interpretation difficult, and various factors, such as forest management, may account for blurring the phylogeographic pattern of genetic differentiation.

In conclusion, the mitochondrial markers analysed in this study proved to be useful for detecting genetic variation in Austrian Norway spruce populations, indicating the importance of genetic investigations on a small geographical scale. Easily scorable on agarose and polyacrylamide gels, these markers can be employed for further genetic diversity studies of *P. abies* in Central Europe and provide helpful implications for forest management.

Acknowledgements

We would like to thank Dr. I. Strohschneider (BFW, Federal Research and Training Centre for Forests, Natural Hazards and Landscape, Vienna) for providing

seed material from certified seed stands. We also wish to thank Dr. B. Heinze and two anonymous reviewers for their helpful comments and suggestions on the manuscript. This study was supported by the Austrian Exchange Service (Österreichischer Austauschdienst, ÖAD).

References

- Bergmann F, Ruetz W (1991) Isozyme genetic variation and heterozygosity in random tree samples and selected orchard clones from the same Norway spruce populations. *Forest Ecology and Management*, **46**, 39-48.
- Collignon AM, Favre JM (2000) Contribution to the postglacial history at the western margin of *Picea abies* natural area using RAPD markers. *Annals of Botany*, **85**, 713-722.
- Demesure B, Compas B, Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515-2520.
- Dumolin S, Demesure B, Petit RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics*, **91**, 1253-1256.
- Dumolin-Lap  gue S, Pemonge MH, Petit RJ (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology*, **6**, 393-397.
- Geburek T, Mottinger-Kroupa S, Burg K (1998) Genetic variation of Norway spruce (*Picea abies* [L.] Karst.) populations in Austria. II. Microspatial patterns derived from nuclear sequence tagged microsatellite sites. *Forest Genetics*, **5**, 231-237.
- Geburek T (1999) Genetic Variation of Norway Spruce (*Picea abies* [L.] Karst.) Populations in Austria. III. Macrospatial allozyme patterns of high elevation populations. *Forest Genetics*, **6**, 201-211.
- Grivet D, Jeandroz S, Favre JM (1999) Nad1 b/c intron polymorphism reveals maternal inheritance of the mitochondrial genome in *Picea abies*. *Theoretical and Applied Genetics*, **99**, 346-349.
- Gugerli F, Anzidei M, Madaghiele A, B  chler U, Sperisen C, Senn J, Vendramin GG (2001a) Chloroplast microsatellites and mitochondrial nad1 intron 2 sequences indicate congruent phylogenetic relationships among Swiss stone pine (*Pinus cembra*), Siberian stone pine (*P. sibirica*), and Siberian dwarf pine (*P. pumila*). *Molecular Ecology*, **10**, 1489-1497.
- Gugerli F, Sperisen C, B  chler U, Magni F, Geburek T, Jeandroz S, Senn J (2001b) Haplotype variation in a mitochondrial tandem repeat of Norway spruce (*Picea abies*) populations suggests a serious founder effect during postglacial re-colonization of the western Alps. *Molecular Ecology*, **10**, 1255-1263.
- Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNACys (GCA). *Nucleic Acids Research*, **28**, 2571-2576.
- Liedeker H (1988) Fichtensterben und spruce decline - a diagnostic comparison in Europe and North America. In: *Proceedings of the US/FRG research symposium: Effects of atmospheric pollutants on the spruce-fir forests of the eastern United States and the Federal Republic of Germany*; 1987 October 19-23; Burlington, VT. Gen. Tech. Rep. NE-120. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station, pp. 245-255.
- Mari  nfeld J, Unseld M, Brennicke A (1999) The mitochondrial genome of *Arabidopsis* is composed of both native and immigrant information. *Trends in Plant Science*, **4**, 495-502.
- McCauley DE (1994) Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: Implications for studies of gene flow in plants. *Proceedings of the National Academy of Sciences of the USA*, **91**, 8127-8131.
- Morgante M, Vendramin GG (1991) Genetic variation in Italian populations of *Picea abies* (L.) Karst. and *Pinus leucodermis* Ant. In: *Genetic Variation in European Populations of Forest Trees* (eds. M  ller-Starck G, Ziehe M), pp. 205-227. Sauerl  nder's, Frankfurt a.M.
- M  ller-Starck G (1995) Genetic variation in high elevated populations of Norway spruce (*Picea abies* (L.) Karst.) in Switzerland. *Silvae Genetica*, **44**, 356-362.
- Nei M (1972) Genetic distance between populations. *American Naturalist*, **106**, 283-292.

- Remy JP, Demesure B, Dumolin S (1998) CpDNA and mtDNA primers in plants. In: *Molecular Tools for Screening Biodiversity* (eds. Krap A, Isaac PG, Ingram DS), pp. 256-261. Chapman & Hall, London.
- Schmidt-Vogt H (1977) *Die Fichte*. Parey, Hamburg.
- Schuster W, Brennicke A (1994) The plant mitochondrial genome: Structure, information content, RNA editing and gene transfer. *Annual Review of Plant Physiology and Plant Molecular Biology*, 45, 61-78.
- Scotti I, Vendramin GG, Matteotti LS, Scarponi C, Sari-Gorla M, Binelli G (2000) Postglacial recolonization routes for *Picea abies* K. in Italy as suggested by the analysis of sequence-characterized amplified region (SCAR) markers. *Molecular Ecology*, 9, 699-708.
- Sperisen C, Büchler U, Gugerli F, Mátyás G, Geburek Th, Vendramin GG (2001) Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. *Molecular Ecology*, 10, 257-263.
- Vendramin GG, Anzidei M, Madaghiele A, Sperisen C, Bucci G (2000) Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). *Genome*, 43, 68-78.
- Wu J, Krutovskii KV, Strauss SH (1998) Abundant mitochondrial genome diversity, population differentiation, and convergent phenotypic evolution in pines. *Genetics*, 150, 1605-1614.
- Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany*, 129, 156-159.

Table 1 Geographical and genetic parameters of 37 *Picea abies* populations in Austria

	Population	Location	Altitude (m)	Latitude	Longitude	Region ^a	NH ^b	H ^c
1	Stürcher,	Vorarlberg	1,300	47°15'N	09°42'E	IV	5B	0.000
2	Winklertobel,	Vorarlberg	1,100	47°06'N	09°59'E	IV	5B	0.000
3	Unsinner,	Tyrol	1,150	47°20'N	10°32'E	II	5B	0.000
4	Oberletzen	Tyrol	1,300	47°30'N	10°42'E	IV	5B	0.000
5	Antonalm,	Tyrol	1,700	47°06'N	11°18'E	I	2A+2B+D	0.640
6	Öxlbach-Moltetal,	Tyrol	1,300	47°21'N	11°49'E	II	5B	0.000
7	Walchsee,	Tyrol	1,000	47°39'N	12°19'E	II	5B	0.000
8	Moaralm,	Carinthia	1,750	47°06'N	12°30'E	I	5B	0.000
9	Aschwinkel,	Tyrol	1,350	46°44'N	12°25'E	I	4B+C	0.320
10	Auzervillgraten,	Tyrol	1,500	46°47'N	12°25'E	I	5B	0.000
11	Rauris,	Salzburg	1,600	47°13'N	12°59'E	I	5B	0.000
12	Hörndl,	UpperAustria	450	48°15'N	13°02'E	VII	3A+B+C	0.560
13	Forstern,	UpperAustria	600	48°01'N	13°14'E	IV	5B	0.000
14	Kesselwand,	Carinthia	1,300	46°47'N	13°29'E	I	5B	0.000
15	Leobengraben,	Carinthia	1,200	46°55'N	13°37'E	I	5B	0.000
16	Bundschuh,	Salzburg	1,425	47°03'N	13°42'E	I	5B	0.000
17	Ramingstein,	Salzburg	1,545	47°04'N	13°50'E	I	5B	0.000
18	Bundschuh,	Salzburg	1,450	47°03'N	13°50'E	I	5B	0.000
19	Ossiacher Tauern	Carinthia	750	46°39'N	13°57'E	VI	5F	0.000
20	Mittelriedl-Röhre,	UpperAustria	560	48°33'N	13°57'E	IX	5A	0.000
21	Kleinreifling,	UpperAustria	1,050	47°50'N	14°40'E	IV	5A	0.000
22	Sternwald,	UpperAustria	850	48°34'N	14°14'E	IX	5B	0.000
23	Tragl,	UpperAustria	1,050	47°53'N	14°10'E	IV	4A+B	0.320
24	Kleinreifling,	UpperAustria	1,080	47°49'N	14°38'E	IV	5B	0.000
25	Kleinreifling,	UpperAustria	1,080	47°48'N	14°42'E	IV	A+4B	0.320
26	Sonntagberg,	LowerAustria	380	47°59'N	14°45'E	IV	4A+E	0.320
27	Zilleck,	LowerAustria	900	48°23'N	14°58'E	IX	4A+D	0.320
28	Trofaiach,	Styria	870	47°25'N	15°00'E	III	4B+D	0.320
29	Lobming,	Styria	1,100	47°18'N	14°59'E	III	5B	0.000
30	Bürgerwald,	Styria	950	47°22'N	15°05'E	III	5A	0.000
31	Gallmannsegg,	Styria	1,100	47°10'N	15°04'E	V	3A+B+C	0.560
32	Neubruck,	LowerAustria	500	47°58'N	15°11'E	IV	4B+D	0.320
33	Paltnühl,	LowerAustria	500	48°20'N	15°30'E	VII	4B+F	0.320
34	Schweinsbach,	Styria	300	46°46'N	15°32'E	VIII	5B	0.000
35	Kohleben,	Styria	940	47°37'N	15°39'E	V	2 ^a +B+C+D	0.720
36	Mariensee,	LowerAustria	875	47°32'N	15°59'E	V	4A+E	0.320
37	Lackenbach,	Burgenland	470	47°35'N	16°28'E	VIII	3A+F+G	0.560

Sample size was five seeds in all cases.

^a Geographic region following Kilian *et al.*, (1994): Region I=central Alps; Region II=northern alpine internal zone; Region III=south-eastern internal alpine zone; Region IV=north-western Alps; Region V=eastern Alps; Region VI=southern calcareous Alps; Region VII=northern Alp-foreland; Region VIII=south-eastern edge of the Austrian Alps and pannonic region; Region IX=Mühl and Waldviertel with Austrian part of Bohemian Massif.

^b Number and identity of haplotypes.

^c Haplotype diversity

Table 2 The seven mitochondrial haplotypes (A–G) and their respective variants in the three markers (mt15-D02, mt23-D02 and mt1H01)

mtDNA loci			Haplotype
mt15-D02 (bp)	mt23-D02 (bp)	mt1H01 (bp)	
1,249	1,117	702	A
1,107	1,122	702	B
1,249	1,122	702	C
1,107	1,117	702	D
1,249	n.A.	702	E
1,249	1117	n.A.	F
1,249	n.A.	n.A.	G

n.A.=no amplification

Table 3 Group structure estimated by SAMOVA

Group	Populations	Region	F_{CT}
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 22, 24, 25, 28, 29, 32, 33, 34	I, II, III, IV, V, VII, VIII, IX	0.766
2	19	VI	
3	20, 21, 23, 26, 27, 30, 31, 35, 36, 37	III, IV, V, VIII, IX	

Fig. 1 A total of 185 individuals from 37 sampling locations (populations) were sampled

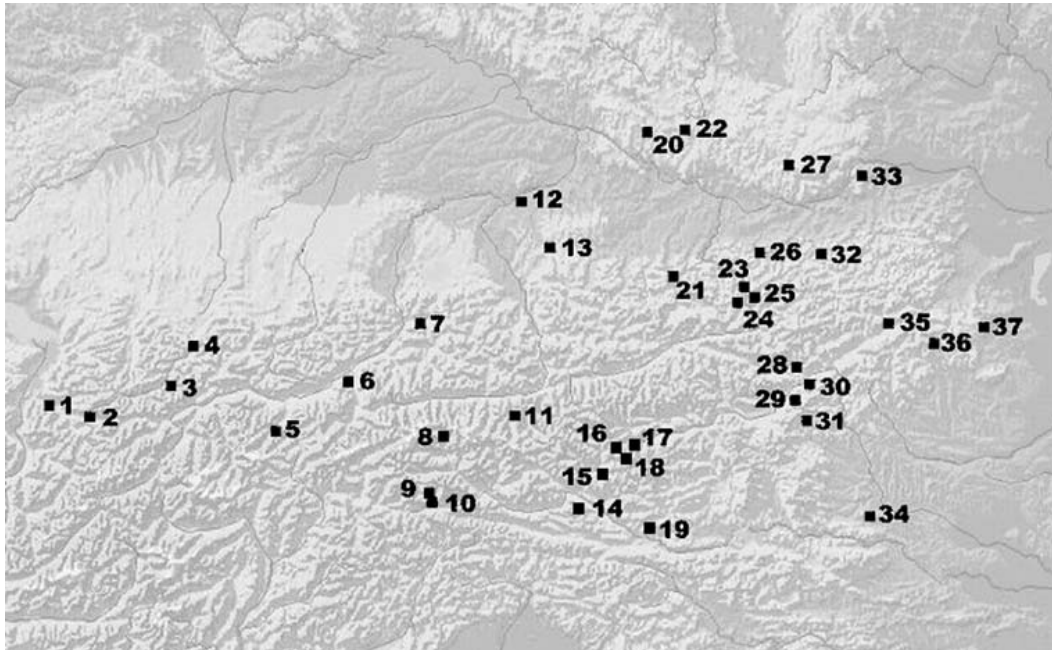


Fig. 2 UPGMA dendrogram based on Nei's (1972) genetic distances between 37 populations of *Picea abies*. Bootstrap support (>50) of internal nodes based on 1,000 replicates is indicated for relevant clades

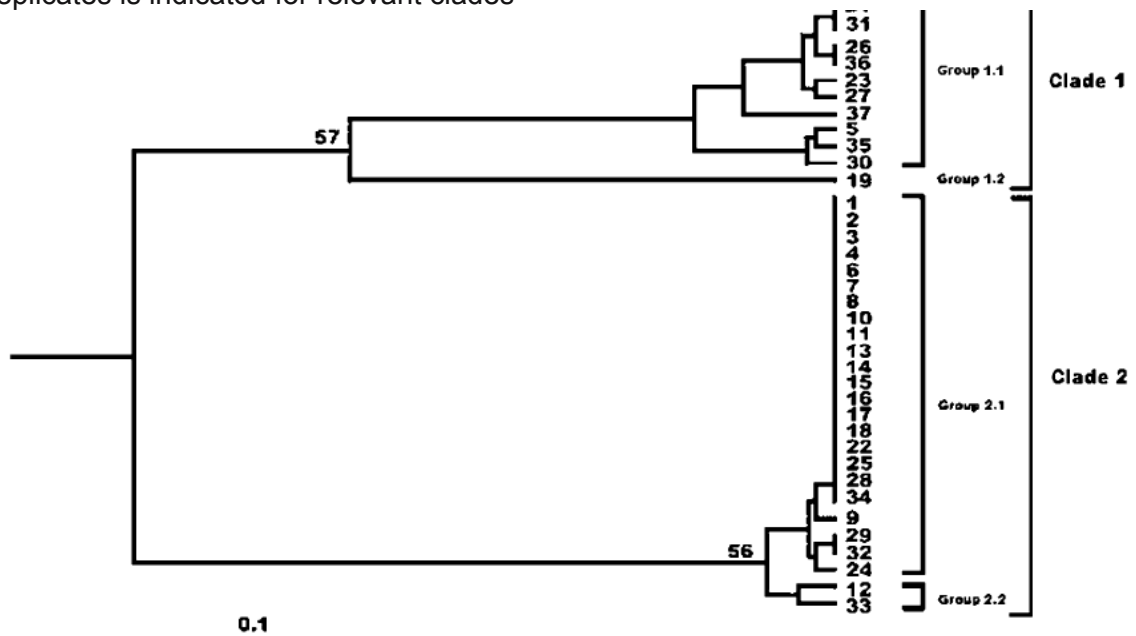
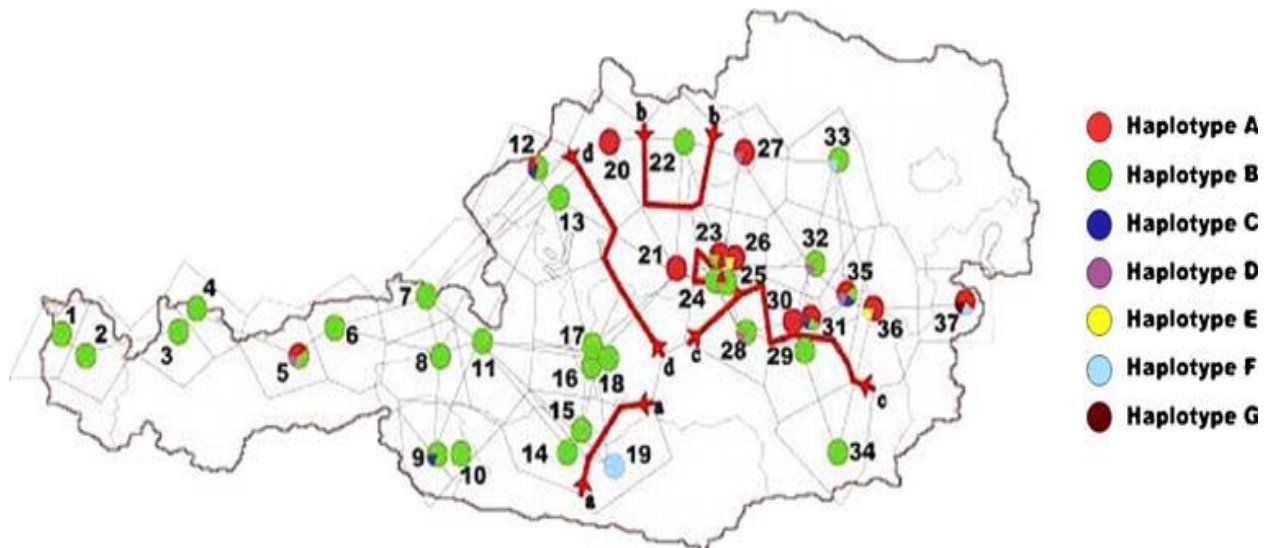


Fig. 3 Haplotype distribution of the seven haplotypes (A–G) found in 37 Austrian populations of Norway spruce. Genetic boundaries obtained with Monmonier's algorithm are indicated by red lines a, b, c and d. The black lines represent Delauney triangulation and Voronoi tessellation



3 Discussion & Conclusion

Biodiversity can be defined at genetic, species and community levels. Genetic diversity has an impact on the higher levels of biodiversity. Characterization of diversity has long been based on morphological traits mainly. However, morphological variability is not obvious at all stages of plant development and in addition it is affected by the environment. Environmental factors, genetic resources management as well as domestication and breeding has continuously narrowed genetic variation in plant communities due to continuous selection pressure for specific traits. This left these selected plants more vulnerable to disease and insect epidemics and in addition was reducing the potential for sustained genetic improvement over a long term. Thus it is important to study the genetic composition of the germplasm of existing cultivars in comparison to their putative ancestors and related species or to investigate the genetic make up of natural populations. This provides information on the phylogenetic relationship of species and populations, on evolution of species as well as post glacial migration and diversity hot spots or will help identifying useful genes for improved fitness.

There are a number of factors that affect genetic diversity of natural populations; probably the most important is the rate of gene flow between populations. One of the most significant factors is mobility, as greater mobility of an individual tends to give it greater migratory potential. Animals are more mobile than plants, although pollen and seeds may be carried long distances by animals or wind. Gene flow can also occur between species, either through hybridization (Burnier *et al.*, 2009) or gene transfer from bacteria or virus to new hosts. Barriers to gene flow need not always to be physical (Fraser *et al.*, 2007; Juan *et al.*, 2004; Kameyama *et al.*, 2002). Species can live in the same environment, yet show very limited gene flow due to limited hybridization as hybridization might yield 'unfit' hybrids (Eckert & Carstens 2008).

DNA mutations

Mutations are changes in the DNA sequence of a cell's genome and can be either caused by external factors (like chemicals, radiation, free radicals) sequence transposition (i.e. mobile DNA elements) as well as mistakes occurring during DNA replication. Changes in the DNA sequence of an organism form the basis for biodiversity, species development and genome evolution.

Mutations in the genome can either be single base pair mutations, indels or chromosome rearrangements, either having no visible effect, alter the product of a gene, or prevent the gene from functioning. Mutation rates vary significantly between coding and non coding regions (Nishant *et al.*, 2009) as well as between different genes (Yi *et al.*, 2006). Reports tell that a single mammalian cell accumulates 10,000 base lesions per day, most of which are being repaired by DNA repair systems to a frequency of 1 mutation every 10^9 nucleotides (McCulloch & Kunkel 2008).

Mutations can involve large sections of DNA becoming duplicated, usually through genetic recombination (Flagel & Wendel 2009). These duplications are a

major source for evolving genes, with tens to hundreds of genes duplicated in animal genomes every million years (Grant *et al.*, 2000). Most genes belong to larger families of genes of shared ancestry (Soltis *et al.*, 2007). Mutations in the genome can be subdivided into germ line mutations, which are passed on to the next generation, and somatic mutations, which occur in somatic cells and therefore are restricted to the individual carrying the mutation. These are not transmitted to the next generation. If an organism can reproduce asexually (via cuttings or budding like poplar or willow) somatic mutations also can be transferred to a new individual. These *de novo* mutations are influencing the estimates of effective mutation rates in different species.

Nonlethal mutations whose effects do not influence the fitness of an individual accumulate within the gene pool over time due to genetic drift and increase the amount of genetic variation (Eyre-Walker & Keightley 2007). The abundance of some genetic changes within the gene pool can be reduced by natural selection, while other, more favorable mutations, may accumulate and result in adaptive evolutionary changes. Natural selection acts on these mutations that generate advantageous new traits that multiply in 'fitter' offspring as well as disadvantageous traits, present in 'less fit' offspring that tends to die out.

In evolution, the most important role of mutations and chromosomal rearrangements may be to accelerate the divergence of a population into new species by making populations less likely to interbreed, and thereby accelerating speciation (Ayala & Coluzzi 2005). Shared evolutionary ancestry means that similar structures possibly evolved from some common ancestor. Homology of sequences is often concluded on the basis of sequence similarity. For example, if two or more genes have highly similar DNA sequences, it is likely that they are homologous. But sequence similarity may also arise without common ancestry: short sequences may be similar by chance, and sequences may be similar because both were selected to bind to a particular protein, still, they don't have common ancestry. There are two types of homologous sequences: orthologs and paralogs. Orthologous are sequences that were separated by a speciation event whereas sequences are called paralogous if they were separated by a gene duplication event (i.e. duplicated gene located in two different positions in the same genome). In a phylogenetic analysis like the DNA barcoding, orthologous genes are analyzed for taxonomic questions (Pennisi 2007) as well as genetic diversity measurements (Shneer 2009). The pattern of sequence divergence can be used to trace the relatedness of organisms. Comparative genomics is studying these genomic relationships between different species and utilizes synteny between genomes for marker development or the identification of orthologous genes of interest in closely related species (Ahn & Tanksley 1993; Conte *et al.*, 2008).

Nowadays, a variety of different genetic markers has been proposed to assess genetic variability as a complementary strategy to more traditional approaches in genetic resources management. Molecular tools provide valuable data on diversity through their ability to detect variation at the DNA level. For evaluation of species

diversity, it is essential that individuals can be classified accurately. The identification of taxonomic units and endangered species, being distinct from their more abundant relatives, is important in the development of appropriate conservation strategies. In population studies, molecular tools are being used to identify parents or to estimate the degree of relatedness among individuals.

DNA-based molecular markers have become the tool of choice for genetic diversity assessment (Ohsawa & Ide 2008; Vencovsky & Crossa 2003) and many DNA markers both specific as well as arbitrary have been used so far for DNA fingerprinting of various classes of germplasm (Vencovsky & Crossa 2003). For different questions in relation to studying the genetic diversity of a given sample set, criteria as defined by different authors (Agarwal *et al.*, 2008b; Gerber *et al.*, 2000; Mariette *et al.*, 2002) should be taken into account in particular considering the trade-off between using (i) numerous but less informative markers randomly distributed within the genome (AFLP, SSAP, ISSR, RAPD) or (ii) a few highly informative markers (microsatellites, DNA sequencing or SNPs). Genome analysis based on molecular markers has generated a vast amount of information, enhancing breeding as well as conservational biology and taxonomic studies, but with every class of markers, different restrictions have to be dealt with. Depending on the study, different molecular markers can be used, varying between very fast evolving microsatellites or some more slowly changing regions like intergenic spacers and introns.

Molecular markers, their applications and problems

Anonymous markers

The most widely used anonymous marker types are RAPD and AFLPs. Both of the techniques are generating multiband patterns targeting regions which are thought to be evenly distributed throughout the genome. The detected regions are hit either by arbitrary primers, or primers representing adaptor sequences which have been ligated to restriction fragments as in the case of AFLPs. The advantage of these marker types is that little developmental effort is needed in order to generate molecular data on any given species. Due to the fact that the methods rely on short sequence stretches (6–10 bp) which are present in all species and are more or less evenly distributed throughout the genome, no a priori sequence information is necessary to apply these techniques.

Anonymous markers gained popularity as tools for the study of genetic polymorphism especially in species where polymorphism is low. In wild barley, for example, AFLP or RAPD were used for studying variation with reference to salt or drought tolerance and associated ecogeography (Nazari & Pakniyat 2008; Pakniyat *et al.*, 1997). These studies have helped in the classification of existing biodiversity among plants, which can be further exploited in 'wild gene introgression programs'.

Still there are some issues that need attention. When applying these anonymous markers for diversity studies in natural populations, quality of DNA is an important

issue influencing results substantially. AFLP is quite sensitive to DNA quality and the presence of residual inhibitors in the DNA extract or degraded DNA samples as compared to standard PCR-based applications (Bensch & Akesson 2005). Degradation of the DNA sample will shift the range of amplified fragments towards shorter fragments, making comparison of fingerprints between individuals impossible whereas the presence of inhibitors results in incomplete digestion of the template DNA and thus altering the profile as well. Variation in DNA concentration seems less critical than variation in DNA quality.

Another problem are primer–dimers that can form in AFLP or RAPD reactions at lower stringency in the absence of a DNA template. These dimers subsequently produce three or four bands in blanks (i.e. ‘no template controls’). These bands might also appear in varying amount even when template DNA is present, causing problems when analysing closely related samples (i.e. paternity assignment) as the estimated error rate is between 1.9% and 2.5% (Busch *et al.*, 2000; Miller 2000). Another factor having potential of altering the fingerprint profile is DNA methylation. When using PstI as restriction enzyme for AFLP analysis, it has been observed that in mapping populations, due to parental PstI-site methylation up to 10% of the bands don’t fit the profile anymore (Isidore *et al.*, 2003).

A basic assumption of the AFLP technique is that the amplified DNA fragments are randomly distributed in the genome. A few studies have found fragments to cluster for example near centromeres (Lindner *et al.*, 2000). When using them for saturating linkage maps there have been observations that certain regions of the chromosomes lack or have underrepresented restriction sites leading to the fact, that in these regions marker saturation cannot be reached with the chosen marker type (Isidore *et al.*, 2003). In addition, when generating linkage maps, anonymous markers always have to be combined with sequence specific markers like SSRs or SNPs as positional ‘anchors’ along the genetic maps (Menz *et al.*, 2004; Yu & Wise 2000) where anonymous markers can be positioned in the space between the sequence characterized positions.

When using anonymous markers to discriminate resistant from susceptible plants by trying to identify one or a few bands distinguishing individuals, a subsequent step is the cloning of the band of interest, assuming that this DNA fragment is somehow related to the trait of interest. In RAPD as well as in AFLP one has to be aware, that besides the strongly visible bands, also a wide variety of DNA fragments is generated during PCR which might be equally big in size, but not visible on the gel used for analysis. Therefore turning anonymous markers over to STS (sequence tagged sites) needs careful evaluation of the isolated fragments (Béaszczyk *et al.*, 2005; Cao *et al.*, 2008; Karsai *et al.*, 2007).

Sequence specific markers

Markers targeting a specific known sequence in the genome are being increasingly used in all fields of plant genome analysis. From marker assisted selection to linkage map production, from taxonomy applications to diversity studies,

especially in the light of the novel ‘high throughput sequencing technologies’ which significantly ease the generation of, and thus access to, sequence data necessary to develop these sequence specific markers, sequence specific markers are getting increasingly important. Microsatellites (SSRs), transposon based marker systems as well as SNPs (Single Nucleotide Polymorphisms) are the most popular marker classes in this category.

Simple sequence repeats (SSRs)

Plant genomes contain large numbers of simple sequence repeats, or microsatellites, of < 6 bp which are tandemly repeated and widely scattered at many hundreds of loci throughout the genome. They can be found in the nuclear genome as well as in the organellar DNA. SSRs are robust, hypervariable and codominant molecular markers suitable for a wide range of applications. SSRs are routinely used in the study of population genetic structure and diversity, pedigree reconstruction and genetic linkage mapping. Their potential for cross-species amplification make microsatellites in addition interesting for applications in ‘orphan crops’ and less investigated species. A variety of methods for SSR isolation has been developed in recent years and the efforts required to obtain sufficient working SSR primer pairs have been comprehensively reviewed by (Zane *et al.*, 2002). With the advent of the 454 sequencing technology, the conventional strategies used to develop SSR markers which were usually labor-intensive, time-consuming, and expensive, have become obsolete. Nowadays thousands of SSR regions can be identified *in silico* simultaneously, using the sequencing data generated either from genomic shotgun sequencing or from sequencing the transcriptome of a given species (Jo *et al.*, 2009; Tangphatsornruang *et al.*, 2009; Varshney *et al.*, 2009). PCR amplification still needs to be optimized for the regions identified *in silico*. In genomes with large portions of repetitive elements like the conifers, using the transcribed part of the genome for marker development eases the generation of SSR markers targeting single copy regions of the genome (Scotti *et al.*, 2002). In addition, when using ESTs as basis for marker development, the detected variation is located in the transcribed region of the genome, possibly having functional impact on the gene concerned, thus allowing to measure potential functional diversity (Salem *et al.*, 2009; Sharma *et al.*, 2009).

Microsatellite variability is widely used to infer levels of genetic diversity in natural populations. However, a bias caused by selecting only the most polymorphic markers in the genome may lead to reduced sensitivity for identifying genome-wide levels of genetic diversity (Vajîli *et al.*, 2008). This potential limitation of microsatellite-based approaches has been tested to assess the degree of nucleotide diversity in noncoding regions of eight different carnivore populations. Estimates of nucleotide diversity varied in some cases with an order of magnitude despite very similar levels of microsatellite marker heterozygosity. These results imply that variability at microsatellite marker sets typically used in population studies may not accurately reflect the underlying genomic diversity. Therefore researchers might consider using

resequencing-based approaches for assessing genetic diversity when accurate inference is critical (Vařeli *et al.*, 2008).

Another issue that needs attention when using SSR for population genetic diversity measurements in natural populations are null alleles. For paternity testing in breeding populations these null alleles (alleles that fail to amplify from one allele due to mutations in the primer binding site and are thus undetected in heterozygotes), is not a large problem because of the small sample size (offspring and potential progenitors). However, when analyzing a natural population problems can arise from PCR-based microsatellite assays as null alleles would reduce heterozygosity measures in the investigated populations. With null allele frequencies ranging between 5% and 8% on average across loci, population genetic parameters such as genetic differentiation (F_{ST}) may be mostly unbiased. However, using these markers with such average prevalence of null alleles (up to 15% for some loci) can be seriously misleading in fine scale population studies and parentage analysis (Oddou-Muratorio *et al.*, 2009) in samples from uncontrolled crosses (i.e. collections in natural habitats).

Single nucleotide polymorphism (SNP)

Single nucleotide polymorphisms (SNPs) are abundant and evenly distributed throughout the genomes of most plant species. They have become an ideal marker system for genetic research in many crops such as maize (Ching *et al.*, 2002; Yan *et al.*, 2009) or rice (Feltus *et al.*, 2004; McNally *et al.*, 2009). There are virtually infinite possibilities to identify sites within the genome where a short stretch of DNA in a pair of homologous chromosomes of the same individual or between two individuals differs by a single base. SNPs arise as mutations involving base pair substitutions, from errors in DNA replication or from mutagenic agents. In terms of their location in the genome, SNPs may be found in coding or regulatory sequences of a gene as well as in noncoding regions. They comprise 90% of the genetic variation in any organism, with varying frequency in coding and noncoding regions. For example in soybean EST coding regions 1 SNP per 3.260 bp has been observed whereas in noncoding regions (UTRs) 1 SNP in 278 bp was found (Van *et al.*, 2004). In the COBL gene of poplar, on the other hand, one basepair mutation was found every 32,8 bp when investigating 40 individuals from a wide geographic range (Zhang *et al.*, 2009). For different *Eucalyptus* species, investigating 23 different genes in 1.764 individuals, similar results have been found for closely related species (one SNP every 33 bp for *E. nitens* and one in every 31 bp in *E. globulus* was detected), whereas for *E. camaldulensis* one SNP every 16 bp and one in 17 bp for *E. loxophleba* on average was found (Kuelheim *et al.*, 2009). This finding shows that two individuals of one species that are closely related are more likely to share identical SNPs than those that are more distant. Individuals belonging to the same family or the same population share more common SNPs because of their shared ancestry. Population and quantitative genetics theory is built on parameters that describe relatedness and estimation of these parameters from genetic markers enables progress in fields as

disparate as plant breeding, human disease gene mapping and forensic science (Weir *et al.*, 2006).

When SNPs occur in the coding sequences they may or may not affect the phenotype, depending on the base pair change concerned (same-sense, mis-sense or nonsense mutation). In such a situation the SNP will show association with the gene or trait concerned, and will be useful in association mapping and marker-assisted selection (MAS) (Gupta *et al.*, 2009; Kota *et al.*, 2008; Wang *et al.*, 2009a; Xing *et al.*, 2007).

In cases where, caused by a large genome size like in conifers, there is a rapid decay of linkage disequilibrium, often within gene limits, SNPs in coding regions can help to overcome this problem. In white spruce (*Picea glauca*) this method was used to identify genes linked to local adaptation. Genes potentially under natural selection were identified by estimating the differentiation in SNP frequencies among populations, resulting in a list of candidate gene variants for local adaptation specific to arid cold and humid environment (Namroud *et al.*, 2008; Wang *et al.*, 2009a). These investigations are made possible by the technological advances of the last years when reference genomes were generated and high throughput technologies were established. With these techniques – array based or bead based (Akhunov *et al.*, 2009; Gupta & Young 2009; Wang *et al.*, 2009a) – hundreds of genes and thousands of individuals can be investigated simultaneously at up to one million SNP loci at a time using automatized effective genotype analysis (Wang *et al.*, 2009a; Yan *et al.*, 2009). In combination with next generation sequencing technology, this allows map saturation in species where little genomic information was available before like in melon (Deleu *et al.*, 2009b) or for assembling genetic and physical maps (Kota *et al.*, 2008) as well as large scale association mapping or SNP discovery (Trick *et al.*, 2009). Their application in plant improvement programs (Gupta *et al.*, 2008; Wang *et al.*, 2009a) has become obvious and possible as in the case of rice, where introgression patterns of shared SNPs revealed the breeding history and relationships among 20 tested varieties; some introgressed regions are associated with agronomic traits that mark major milestones in rice improvement (McNally *et al.*, 2009). Especially for domesticated plants, the application of SNPs in mapping and breeding has some restrictions, as it has been observed that, when performing genome-wide comparison of 2 rice cultivars, 8% of the genome shows extremely low SNP rates indicating adaptive selection for domestication-related genes located in these genomic regions, suggesting that this low SNP frequency is strongly related to domestication (Wang *et al.*, 2009a). The availability of molecular markers for important agronomic traits combined with more efficient marker detection systems will help reach the full benefit of MAS in the breeding effort to reassemble potential genes and recapture critical genes among the breeding lines that were lost during domestication to help boost crop production worldwide (Utomo & Linscombe 2009).

Conservation and thus transferability of SNP marks between varieties and species was assayed by in a genome-wide analysis of SNP diversity in a collection of major cereal crops by in silico comparison of 3.5 million ESTs (Barker & Edwards

2009). The frequency of sequence polymorphism was lowest in diploid taxa (rice, barley and sorghum), intermediate in tetraploid maize and highest in allohexaploid wheat and octoploid sugarcane. SNPs generated due to sequencing errors which are inherently present in large datasets are getting less likely when detected between closely related varieties as they are unlikely to co-segregate with one another. Sequencing of the highly heterozygous *Vitis vinifera* L. cultivar Pinot Noir gave rise to several thousand SNPs which were used for a transferability survey in 69 grape accessions (Vezzulli *et al.*, 2008). Transferability of 31.5% of the 137 investigated SNPs using the PCR based SNPlex Genotyping System across unrelated cultivars of *V. vinifera*, of 18.8% across wild forms of *V. vinifera*, of 2.3% among non-*vinifera* *Vitis* species, and of 0% with *Muscadinia rotundifolia* was found. This limited applicability for non-*vinifera* *Vitis* species leads to the conclusion that a resequencing strategy should be applied instead of a PCR based approach for future applications such as mapping and diversity studies, accession identification or breeding when working with genetically less related species.

Nuclear/Chloroplast/Mitochondrial DNA

The marker types described above are present in the three genomes available in plants. The inheritance of the mitochondrial and chloroplast genome differs from that of the nuclear genome in showing uniparental inheritance, vegetative segregation and reduced recombination. Therefore the three genomes can have different modes of transmission and experience varying levels of gene flow. As reviewed by Mogensen, (Mogensen 1996) organelles are inherited in a nonMendelian fashion in all eukaryotic organisms investigated. Among the seed plants, plastids can be inherited strictly from the female parent, strictly from the male parent, or biparentally. Most flowering plants exhibit maternal plastid inheritance, but approximately one-third of the genera displays biparental plastid inheritance to some degree. Among the gymnosperms, paternal plastid inheritance is the rule in the conifers, whereas the other groups appear to have maternal plastid inheritance. Mitochondrial inheritance is predominantly maternal in the seed plants, except for a few coniferous families where it is predominantly paternal.

Chloroplast DNA (cpDNA) is mostly maternally inherited in the majority, but not all, of angiosperm species. The mode of inheritance of cpDNA is a critical determinant when using the genome for measuring molecular evolution and population genetic structure. Therefore careful investigations have to be made as to which extent cpDNA shows inheritance other than maternal i.e leakage. There are several examples in literature reporting on biparental or paternal via the pollen. Very often exceptions from the maternal mode of inheritance of cpDNA can be seen in inter- or intraspecific crosses as in the case of azalea (Kobayashi *et al.*, 2008), but also in natural populations as in the case of *Silene vulgaris*, where 1,5% of the offspring from natural populations did not match the mother's chloroplast type (McCauley *et al.*, 2007). When using cpDNA as phylogenetic marker like in barcoding

approaches (Kress & Erickson 2008), it has to be kept in mind, that hybridization is also a naturally occurring mechanism in speciation and evolution of genetic diversity.

Distorted ways of chloroplast inheritance have been observed in cases where nuclear-cytoplasmic incompatibility manifested as deficiency of chlorophyll pigmentation required proliferation of paternally inherited plastids for normal performance (Bogdanova 2007). MtDNA always seems to obey to the maternal way of inheritance, also in the case of *passiflora* hybrids, where mtDNA was inherited maternally, whereas cpDNA in some cases showed paternal inheritance (Muschner *et al.*, 2006). These results are of significance for the ongoing studies which are being performed on the molecular evolution of genera and the discussions on which marker region should be used for barcoding and constructing the 'tree of life'.

When deciding on the use of markers in different genomes also the mutation rate and thus the data drawn from investigations of these genomes have to be taken into account. MtDNA evolves about 5–10 times faster than nuclear single copy number DNA, where coding genes are conserved. Thus, different DNA signatures are left on different types of molecular markers. Both single-copy nuclear DNA and mtDNA contain historic genetic mutations and are suitable for resolving taxonomic questions; mtDNA markers and nuclear repetitive elements contain more information on recent genetic variation than single copy DNA. In recent studies, more than one genome was used to answer specific questions of hybridization, speciation or gene flow. Results are reviewed by (Petit *et al.*, 2005) on 183 species belonging to 103 genera and 52 families showing that the mode of inheritance appears to have a major effect on gene diversity (G_{ST}). Maternally inherited genomes experience considerably more subdivision (median value of 0.67) than paternally or biparentally inherited genomes (-0.10). Variation in cpDNA and mtDNA markers shows similar levels of diversity when both genomes are maternally inherited, whereas at paternally and biparentally inherited markers, variability is less similar. Variability at maternally inherited markers is largely independent of values based on nuclear markers. This also has been observed by Muir (Muir & Filatov 2007) where maternally inherited chloroplast genes as compared to paternally inherited Y-linked genes showed significant loss of polymorphism in chloroplast genes.

Not only the sample sizes but also the analysis tools and methods are currently changing significantly. A new generation of sequencing technologies, Illumina/Solexa, ABI/SOLiD, 454/Roche, and Helicos, provides new opportunities for high-throughput functional genomic research. These technologies are being applied in a variety of contexts, including whole-genome sequencing, targeted resequencing, and RNA expression profiling. The sudden availability of DNA sequencing technologies that rapidly produce vast amounts of sequence information transforming polymorphism discovery, analysis of genetic variation mutation mapping, DNA methylation, histone modifications, gene discovery, transcriptome sequencing, alternative splicing identification and small RNA profiling (Lister *et al.*, 2009; Morozova & Marra 2008). Thus, these deep sequencing technologies offer plant biologists unprecedented opportunities to increase the understanding of the functions and dynamics of plant

cells to whole populations. Although the methods have only been available for a short period of time, high-throughput sequencing techniques are strongly affecting life sciences and are opening out new potential avenues of research.

Identification and tracking of genetic variation is becoming efficient and precise so that thousands of variants can be tracked within large populations, which for the first time allows wide application of markers analyzing differences at the nucleotide level. New applications and approaches are just being developed, such as allele-specific expression assays to identify cis, trans, and cis-by-trans regulatory variation (Main *et al.*, 2009) or new methods for genotyping of thousands of samples in parallel using only one sequencing run (Erlich *et al.*, 2009).

Important areas of application will be large-scale development of molecular markers for linkage mapping, association mapping, transcript profiling, population genetics and de novo genome/organellar genome assembly as well as the detection of epigenetic modifications.

Epigenetics

Apart from variations on the DNA level, there are also epigenetic factors influencing gene expression – and to some extent also marker applications. The term epigenetics refers to heritable changes not encoded by DNA. As epigenetic factors, DNA methylation, histone protein methylation, heterochromatin structure and microRNAs are influencing the phenotype of an individual (Djupedal & Ekwall 2009). Large parts of eukaryotic genomes consist of constitutively highly condensed heterochromatin, important for maintaining genome integrity but also for silencing of genes within these regions.

DNA methylation occurring on the 5 position of the pyrimidine ring of cytosines in the context of the dinucleotide sequence CpG forms one of the mechanisms of epigenetic control and modulation of gene expression. DNA methylation plays an important role in many aspects of biology, including development and disease. Methylation can be detected using bisulfite conversion, methylation-sensitive restriction enzymes, methyl-binding proteins and anti-methylcytosine antibodies. Combining these techniques with DNA microarrays and high-throughput sequencing has made the mapping of DNA methylation feasible on a genome-wide scale (Zilberman & Henikoff 2007), using the results as biomarkers of diseases (Tost 2009) or stress adaptation in plants (Chinnusamy & Zhu 2009). When using traditional DNA markers, one has to be aware that DNA methylation can strongly influence the result of restriction enzyme or PCR based marker applications due to the fact that methylation at CpG islands could either prevent the binding of a PCR primer to the target region or hinder a methylation sensitive restriction enzyme from cutting at a specific sequence. Especially in AFLP approaches, where PstI is frequently used as restriction enzyme, differences in banding patterns between individuals might as well originate from differences in the methylation pattern at the PstI site rather than from sequence mutations. These methylation differences can as well be exploited as molecular markers when detecting them either by differences in

restriction sites using methylation sensitive enzymes, or by bisulfite sequencing (Hayatsu *et al.*, 2008; Reed *et al.*, 2010). Bisulfite deaminates cytosine in nucleic acids, while 5-methylcytosine resists this bisulfite action. For this reason, bisulfite treatment can be used for detecting 5-methylcytosine in DNA (Hayatsu *et al.*, 2008).

The use of methylation sensitive enzymes has led to methylation filtration techniques that were used directly as markers such as methylation-sensitive amplified polymorphism (MSAP) (Rabinowicz *et al.*, 2005; Tan 2010). The hypomethylated fraction of plant genomes is usually enriched in genes and has been used to selectively clone gene space related sequences from the maize (Emberton *et al.*, 2005) or rice (Berenyi *et al.*, 2009) genome. Methylation differences can not only be found between individuals but also between the different tissues of one individual (Lu *et al.*, 2008).

Richards (Bossdorf *et al.*, 2008; Richards 2008) reports on the first attempt to investigate epigenetic differences at the population level. Population epigenetics is emerging as an addition to molecular genetics, genomics, and population biology, and addresses questions concerning the prevalence and importance of epigenetic variation in the natural populations. Heritable variation in ecologically relevant traits can be generated through a suite of epigenetic mechanisms, even in the absence of genetic variation. Moreover, recent studies indicate that environmentally induced epigenetic changes may be inherited to future generations (Bossdorf *et al.*, 2008). Comparative studies on stress-responsive epigenomes and transcriptomes are just starting and will enhance our understanding of stress adaptation of plants (Elling & Deng 2009; Wang *et al.*, 2009b).

The development of a diversity of DNA-based markers in the past twenty years has revolutionized biological, forestry and agricultural science. Molecular markers have demonstrated usefulness for plant genotyping and population genetic analysis. Genome characterization by DNA-fingerprinting has generated valuable information on population structure and molecular genetic diversity. Large sets of molecular markers were developed for hundreds of species, genetic and physical maps of high density were constructed for agricultural breeding material as well as for trees or ornamentals. With the advent of the new generation of molecular tools, screening large natural populations or segregating progenies has become possible.

With the work presented here, we contributed to the marker development in different areas of plant genomics, and proved the applicability of the different marker types targeting the three different genomes (nuclear, mitochondrial, chloroplast) available in plants.

In order to investigate the genetic variation present in Alpine *Rhododendron* populations, nuclear SSRs were developed for *Rhododendron ferrugineum* and *Rhododendron hirsutum* by generating genomic libraries enriched for dinucleotide repeats. These newly generated markers subsequently were used to investigate relict populations of the alpine rose in Italy and Austria. The same method was applied to isolate SSR regions for the identification of *Paeonia suffruticosa* varieties available in a germ plasm collection in Schönbrunn. In the latter case, also an in silico approach

was used in order to identify SSRs present in publicly available ESTs (sequence information on Expressed Sequence Tags). By this approach, 5 additional SSR markers for *Paeonia suffruticosa* as well as 24 EST-SSRs for *Picea abies* could be generated in relatively short time. These EST-SSRs have added value as they are part of expressed sequences and thus can be used for measuring the functional diversity in Norway spruce as well as *Paeonia*.

For *Quercus robur* L. and *Quercus petraea* L. SSRs as well as VNTRs were isolated by a standard wet lab approaches and used for diversity measurements in Austrian populations of oak. In addition, these markers, together with RAPDs, RFLPs and SCARs were mapped on a genetic map of a French controlled oak cross. Due to their hypervariable nature and sensitivity to mutations, nuclear microsatellites for oak were used to test for somaclonal variation occurring in tissue culture material over time. With this approach we could prove the usefulness of the SSRs for testing the genome stability under stress.

Maternally inherited markers were developed for *Populus nigra* L. and *Picea abies*. In order to assess the genetic diversity in *Populus nigra* L. sampled throughout Europe, a chloroplast DNA based marker system was developed and applied. For Norway spruce on the other hand, the maternally inherited mitochondrial genome was targeted for marker development. Using these new mt markers for Norway spruce it could be shown that there is moderate genetic variation in Austrian Norway spruce strands, not detected before.

The newly developed nuclear as well as organellar markers are valuable tools for the assessment of genetic diversity in natural populations as well as plant selection programs. Using this DNA based molecular information it is possible to analyze the genetic structure of populations, to estimate gene flow between populations and, as in the case of poplar, to identify naturally occurring clonal material. With these new markers, specific population genetic as well as phylogeographical questions can be answered, and, as in the case of rhododendron, decisions on conservational measures taken.

751 TCCCTTTTAAGAAACAGATATGTCAAATAGTAGATGAATAATACAACCTT
AGGGAAAAATCTTTGTCTATACAGTTTATCATCTACTTATTATGTTGGAA

801 AATGCACAAGAAGCATTATATATTGAGTTTTTTTTTTTTTTTTTTTTTTTTT
TTACGTGTTCTTCGTAATATATAACTCAAAAAAAAAAAAAAAAAAAAAA

851 TTACGAGTGTACTTTTCATTCCCTGAATTTAAGAAATTGAAGCGCAACATT
AATGCTCACATGAAAAAGTAAGGACTTAAATTCTTTAACTTCGCGTTGTAA

901 CAGTAGCTCATGACCTAATAAGATAGTCATAAA**AGAGAGAGGACTAAGGA**
GTCATCGAGTACTGGATTATTCTATCAGTATTT**TCTCTCTCCTGATTCC**T

951 AAAATCAAGATTAAATTATAAATTACACTC**AAGCTTTAGGGGATTATTTT**
TTTTAGTTCTAATTAAATATTTAATGTGAG**TTCGAAATCCCCTAATAAAA**

F2Core repeat (RepMot No.2)

1001 TGACTTTTACTCCCTAAAGTTTCAAAATTTCAATATGCTCTCATAACATA
ACTGAAAAATGAGGGATTTCAAAGTTTAAAGTTATACGAGAGTATTGTAT

1051 AAGTCTCATGAGTA**GTC****CAATCCCCTTCCAACTTTGTG**TTGTGCCTAGCTGT
TTCAGAGTACTCAT**CAGTTAGGGAAGGTTGAAACAC**AAACACGGATCGACA

1101 CTTAAGCGTATAAGATTG**GGCC**GCAGTGCTG**CCCTCTATGTTTGTA**CTT
GAATTCGCATATTCTAAC**CCGG**CGTCACGAC**GGGAGATACAAACATTGA**A

Mn1 repeat1 (RepMot No.1)

1151 ACA**TCTATGTTTGAAAT**TGTATTGATT**TGAACCTCTATGTTTGGA**ACTG
TGT**AGATACAAACTTTAA**CATAACTAACTT**GGGAGATACAAACCTTG**AC

Mn1 repeat4

1201 TCATACA**CCCTCTATCTTTGAAA**TTGTTTTTAATTGATTTCGAACTTAAC
AGTATGT**GGGAGATAGAACTTTA**CAAAAAATTAATAAGCTTGAATTG

1251 CAGCTAACCAAAGT**CATTACAAGCCCTCTCCTCAG**GGAAGGATTTTTTTT
GTCGATTGGTTTC**AGTAATGTTTCGGGAGAGGAGTC**CCCTTCCATAAAAAA

1301 TATAAGGCTTCACACATTGTATGGGAAGCAAATCCAAATGATGATACCTG
ATATTCCGAAGTGTGTAACATACCCCTTCGTTTAGGTTTACTACTATGGAC

1351 AAGTTAGGAGAGCCAATC**AAAAATATTGAACTTTAGGGGAGCAAAATTAA**
TTCAATCCTCTCGGTTAG**TTTTATAACTTTGAAATCCCCTCGTTTTAATT**

1401 **AATTACCCCAAAGCTT**TATATATAT**TATAGATGGGATTGACGAGAA**CATGT
TTAATGGGGTTTCGAAATATATAT**TATCTACCCTAACTGCTCTT**GTACA

1451 TATAGTGACGAGGTAGCGTTATGACGAGGTTATCCCATGACGACAAACAA
ATATCACTGCTCCATCGCAATACTGCTCCAATAGGGTACTGCTGTTTGT

1501 ACCATCAA**CAACGAGGAAAAGGAAGGTCATTCA**ATGGTGCCAGCAAGTAAC
TGGTAGTT**GTTGCTCCTTTCCCTTCCAGTAAGT**TACCACGGTCGTTTCATTG

BLAST result for the 2.3kb fragment:
Total number of bases is: 2331.
DNA sequence composition: 747 A; 495 C; 440 G; 649 T;
Sequence name: B001

>  ref|XM_002283496.1|  PREDICTED: Vitis vinifera similar to ATP binding, transcript variant 2 (LOC100254725), mRNA
Length=2038

GENE ID: 100254725 GSVIVT00018481001 | similar to HSC70-1 (heat shock cognate 70 kDa protein 1); ATP binding [Vitis vinifera] (10 or fewer PubMed links)

Score = 361 bits (400), Expect = 7e-96
Identities = 471/634 (74%), Gaps = 22/634 (3%)
Strand=Plus/Minus

```

Query  131  AGAATCCTCTTTGCT-TCTCACAAGCAGTTCTTAACCTCCTAAGAGCT-TGGAATTCCCA 188
          ||| | ||||| ||||| | || | || | ||||| | || | || |
Sbjct  957  AGAGTTCTCTTTGCCCTCTCACACGATGTCCTCAATCTTCTAAGAGATCTGGCGTTTCCA 898

Query  189  CTAATGTCCACCTTGTATTGCCTCTTAAATATTTCAACAAAGTGATTCACTATTCGGTTA 248
          ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
Sbjct  897  GTAATGTCCTTCTTGTGTTTCTCTTAAACTCCTGAACAAAGTGTTGACCATTCTATTA 838

Query  249  TCAAAGTCCTCACCTCCAAGGTAAGTGCTCCAAC---AGCCTTCACCTCAAAGATACCC 305
          ||||| ||||| ||||| | || ||||| | ||||| ||||| ||||| ||
Sbjct  837  TCAAAGTCCTCTCCTCCAAGATGGGTATCTCCAGCTGTTGCCCTTCACCTCAAAGATACCC 778

Query  306  TTGTCAATAGTAAGTAGTGAGACAT-GTAGTTGCCACCACTCAAATCAAAGTCAACACA 364
          | ||||| || || || || || || || || || || || || || || ||
Sbjct  777  TCTTCAATAGTGAGAAGAGACACATCAAAGTGCCGCCACCCAGATCGAAAATCAGCACG 718

Query  365  TTCTCATTGCCAATGCTATT--GTAACCTGTCTAGACCGTAAGCAAGGGCTGAAGCAGT- 421
          |||| | |||| | || | ||||| ||||| ||||| ||||| ||||| ||
Sbjct  717  TTCTTCTCACCAACACTGGTCGCCTTCTTGTCAAGACCATAAGCAATAGCGGCAGCTGTT 658

Query  422  ---TCATTAATTATACGTAGGACATTTAGACCTGCAAT-GCTTTTGCATCCTTTATAGCC 477
          |||| | |||| | |||| | ||||| || ||||| || ||||| || ||
Sbjct  657  GGCTCATTGATGATACGCATCACATTGAGGCCTGCAATAACTCCCGCATCCTTGGTGGCT 598

Query  478  CTACGCTGGGAGTCATTGGAGTAGCGAGG-ACAGTAACAACCGGTTTCTTCGACAGTTTT 536
          ||||| ||||| ||||| ||||| || || ||||| || ||||| ||||| ||
Sbjct  597  TGACGCTGAGAGTCATTGAAGTAGGCAGGCACTGTAACAACAGCGTTCTTC-ACAGTTGA 539

Query  537  GCCTAGGTATGCCTCAGCTATTTACGCATCTTTATTAG-ACCATATATGAGATTTCTC 595
          || ||||| ||||| || || ||||| || || ||||| || || ||||| ||
Sbjct  538  ACCGAGGTATGCCTCGGCAATCTCACGCATCTTAATCAGAACCATGGACGAGATTTCTC 479

Query  596  AGCAGC-AAGTGCTTCTCTTCGCCTTTATAGTTGACCACAATCATAAGCTTATCACCA-G 653
          || | || ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||
Sbjct  478  GGC GGCAAAC T GCTTATCTTCA C C C T TATAGGTGACTGTGATCATGGGCTTGTCA C CAGG 419

Query  654  ACCTTCAA-TACCTTAAATG-CCAAAGGTTGAAATCACTCTGAACCAAAGGATCACTGAA 711
          ||| || ||||| ||||| || || ||||| ||||| || ||||| ||||| ||
Sbjct  418  ACCAGGAACGACCTTGAATGACCAGTGCTTGATATCACTCTGAACCGAGGCATCAGAGAA 359

Query  712  TCTCCTACCGATCAACCGCTTTGCATCTAGAAAA 745
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Sbjct  358  TCTTCTACCAATCAAACGCTTGGCATC--GAAAA 327

```

5 Appendix II

A genetic linkage map of *Quercus robur* L (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers.

Barreneche T, Bodenes C, Lexer C, Trontin J F, Fluch S, Streiff R, Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glössl J, Kremer A.

T. Barreneche, C. Bodenes, R. Streiff, C. Plomion, G. Roussel, A. Kremer, Laboratoire de Genetique et Amelioration des Arbres Forestiers, INRA, B.P. 45, Pierroton, F-33610 Cestas Cedex, France Fax: (33) 5 57 97 90 88 E-mail: kremer@pierroton.inra.fr
C. Lexer, J. Glössl, H. Steinkellner, Zentrum für Angewandte Genetik, Universität für Bodenkultur Wien, Muthgasse 18, A-1190 Vienna, Austria
J.-F. Trontin, J.-M. Favre, Laboratoire de Biologie Forestie`re, associe« INRA, Universite« Henri Poincare«, Nancy, B.P. 239, 54506 Vandoeuvre-le`s-Nancy, France,
S. Fluch, K. Burg Forschungszentrum Seibersdorf, A-2444 Seibersdorf, Austria

Theor Appl Genet (1998) 97: 1090-1103

Abstract

A genetic map of Pedunculate oak (*Quercus robur* L.) was constructed based on one 5S rDNA, 271 RAPD, ten SCAR, 18 microsatellite, one minisatellite, and six isozyme markers. A total of 94 individuals from a full-sib family was genotyped. Two maps, including 307 markers, were constructed according to the 'tway pseudo-testcross' mapping strategy. Testcross markers segregating in the 1 : 1 ratio were first used to establish separate maternal (893.2 cM, 12 linkage groups) and paternal (921.7 cM, 12 linkage groups) maps. Both maps provided 85–90% genome coverage. Homologies between the male and female linkage groups were then identified based on 74 intercross markers segregating in the 3 : 1, 1 : 2 : 1 and 1 : 1 : 1 : 1 ratios (RAPDs, SCARs, SSRs, 5S rDNA and isozymes) in the hybrid progeny. In each map, approximately 18% of the studied markers showed segregation distortion. More than 60% of the skewed markers were due to an excess of heterozygote genotypes. This map will be used for: (1) studying the molecular organisation of genomic regions involved in inter- and intraspecific differentiation in oaks and (2) identification of QTLs for adaptive traits.

1 Introduction

Depending on the authors the number of existing oak species varies between 300 and 600. They are spread over the northern hemisphere (Europe, Africa, North America and Asia) from boreal zones to semi-arid areas. Pedunculate oak (*Quercus robur* L.) is one of the most important and widespread oak species in Europe. It belongs to the white oak section (*epidobalanus*) and is distributed throughout Europe from central Spain to the Urals. According to Camus (1954) the *epidobalanus* section is the richest with respect to numbers and diversity and comprises more than

150 species spread over all four continents. *Quercus robur* L. has a broad ecological range, from very acid to chalky and from moist to mesophilic soils, but it prefers to grow on neutral and moist soils (Becker and Levy 1990). Genetic studies in pedunculate oak have mostly concentrated on population genetics, provenance research and the inheritance of economically important traits (Kremer *et al.*, 1993, for a review). Cytogenetic investigations have been restricted to karyotypic analysis. *Quercus robur* L. is a diploid species ($2n=2x=24$), although in some cases supernumerary chromosomes (B type) have been reported (Ohri and Ahuja 1990; Besendorfer *et al.*, 1996). Natural triploid individuals have been described (Johnsson 1946; Butorina 1993) that usually are larger than diploid trees. Compared to data available on other woody angiosperms (e.g. *Populus*, *Eucalyptus*, *Acacia*, *Pyrus*, *Fraxinus*) the physical genome size of *Q. robur* is larger (1.88 pg/2C) (Zoldos *et al.*, 1998), although its size is among the lower values found in plants (Favre and Arnould 1996; Favre and Brown 1996). Segregation studies with gene markers have been conducted in *Quercus robur* L. with isozymes (Zanetto *et al.*, 1996), SCARs (sequence characterised amplified regions, Bode«ne`s *et al.*, 1996) and microsatellites (Steinkellner *et al.*, 1997a). However, due to the small number of loci analysed in each of these studies, no linkage groups could be constructed except with isozymes when two linkage groups were identified (Zanetto *et al.*, 1996). Our objective was to construct a genetic map with all earlier developed markers (isozymes, SCARs, microsatellites, minisatellites, 5S rDNA). In order to locate these markers in the genome, we added randomly distributed markers using amplification with arbitrarily designed primers [random amplified polymorphic DNA (RAPD) Williams *et al.*, 1990]. The pedigree used for the map construction is a full-sib F1 progeny of *Quercus robur* L. Linkage groups were assembled according to the 'two-way pseudo-testcross' method described by Grattapaglia and Sederoff (1994). To our knowledge, this is the first reported linkage map in the genus *Quercus*, and in the *Fagaceae* family as well. Some of the markers used in the construction of the map have been reported to be applicable to other oak species: isozymes in *Q. petraea* and *Q. pubescens* (Müller-Starck *et al.*, 1996), microsatellites in other sections of the genus *Quercus* (Steinkellner *et al.*, 1997b). Because oaks of a given botanical section usually exhibit high genetic similarities (Kremer and Petit 1993), we expect that the present map will provide a useful tool for various applications in population and quantitative genetics in different white oak species.

2 Materials and methods

Plant material

Segregation data from a two-generation full-sib pedigree including 94 individuals were used to construct two genetic maps, one for each parental tree. The male parent originated from near Arcachon (southwest of France), and the female parent was located on the Forestry Research Station in Pierroton (southwest of France). The controlled cross was made in the spring of 1992 according to the techniques

described in Zanetto *et al.*, (1996). About 400 acorns were collected in pollination bags, of which 94 were used for map construction. Preliminary segregation analysis of the 94 offspring with a few isozymic loci did not detect any acorn resulting from fertilisation with foreign pollen (Zanetto *et al.*, 1996). Marker analysis Genomic DNA was extracted from leaves according to Saghai- Maroof *et al.*, (1984) with some minor modifications described by Moreau *et al.*, (1994). Procedures used to study different markers are summarized in Table 1. Specific primer sequences and annealing temperatures for the SCAR, microsatellite, minisatellite and 5S rDNA markers are indicated in Table 2.

Data analysis

After segregation was scored for each marker, Mendelian segregation was tested in the offspring using Chi-square goodness-of-fit tests. Linkage analysis of the markers was performed using a combination of MAPMAKER (Lander *et al.*, 1987) and JOINMAP (Stam 1993) software (see Results section). The Kosambi (Kosambi 1944) mapping function was used to convert recombination frequencies to map distances in centiMorgans. Genome length, $E(G)$, of both parental maps was estimated under the assumption of random marker distribution according to method of Chakravarti *et al.*, (1991): G^2MX/K , where $M^n(n!1)/2$ is the number of informative meioses (n is the number of loci analysed), X is the maximum observed map distance among the locus pairs above a threshold LOD, Z , and K is the number of locus pairs having LOD values at or above Z . The values tested for Z were 3, 4 and 5. The values of X and K were obtained using the function of MAPMAKER with the maximum recombination fraction set to 0.5 and minimum LOD threshold set to Z . Values for X were obtained using the Kosambi map function. These estimations were calculated by taking into consideration only pairwise comparisons between framework markers.

3 Results

Segregation analysis

Markers were subdivided in four different groups on the basis of the segregation patterns in the progeny: A) Loci that were in the testcross configuration (Grattapaglia and Sederoff 1994) between the parents (heterozygous in one parent and homozygous null in the other, or vice-versa). A 1 : 1 segregation ratio was observed in the F1 family for most RAPDs, seven SCARs (six dominant and one codominant), two isozymes, 1 microsatellite (ssrQpZAG104) and one minisatellite (Omin3). B) Loci that were heterozygous in both parents and that segregated in a 3 : 1 ratio in the offsprings. Fifty RAPDs followed this pattern. C) Loci that were heterozygous in both parents and that segregated in a 1 : 2 : 1 ratio. This was the case for 2 microsatellites and three isozymes. D) Loci that were heterozygous in both parents and that segregated in a 1 : 1 : 1 : 1 ratio. Three SCARs, 15 microsatellites, one isozyme and one 5S rDNA locus followed this pattern. Markers segregating in the 1 : 1 : 1 : 1 ratio

were partitioned according to the alleles contributed by the relevant parent (Jacobs *et al.*, 1995), resulting in a marker segregating 1 : 1 in both the female and the male parent, as for category A markers.

Approximately 80% of the markers analysed in the progeny fit the expected Mendelian segregation ratios (1 : 1, 3 : 1, 1 : 2 : 1 or 1 : 1 : 1 : 1). A total of 61 markers showed segregation distortion at $\alpha=0.05$.

Map construction

The map construction was done in two steps. First, two separate parental maps (paternal and maternal) were constructed using markers segregating in the 1 : 1 ratio (category A and D markers) with a confidence level of $P=0.01$. These 'testcross' markers were grouped with the help of MAPMAKER (Lander *et al.*, 1987) using the following criteria:

- a minimum LOD score of 3.0 for statistical acceptance of linkage.
- a maximum recombination fraction h of 0.40.

In the initial step of map construction, the order of non-distorted markers was approximated using the 'FIRST ORDER' command. From this order, a subset of markers that could be locally ordered with a likelihood ratio greater than LOD^2 was used to establish a framework map using the 'RIPPLE' command. Markers that could not meet this ordering criteria were dropped from the framework and placed as accessory markers. Accessory markers were then located beside their closest framework marker. Distorted markers that were excluded from the framework map were also placed in the established framework map as accessory markers using the 'NEAR' command. These markers are indicated with a boldface asterisk in Fig. 1.

In a second step, intercross markers segregating in both parents (category B, C) in addition to category D markers already mapped in the first step were used to identify homologous counterparts between parental maps. Markers of category D were mapped using MAPMAKER. However, because MAPMAKER does not accept mixed segregation data, intercross dominant and codominant markers of category B and C were placed to an approximate location on both maps with the JOINMAP programme (Stam 1993). Finally, a total of 307 Loci (271 RAPDs, ten SCARs, 18 microsatellites, one minisatellite, six isozymes and one 5S rDNA marker) were located on both oak linkage maps (Fig. 1). Among the 50 intercross RAPD loci (category B) only the 24 markers showing strong linkage ($\text{LOD} \geq 5.5$) with framework markers in both maps were retained.

Female map (Fig. 1)

A total of 136 markers were used to establish the female map. Approximately 58% of them (79 loci) were located on the framework map. Twelve linkage groups were set up for the female map covering 893.2 cM, with an average distance

between 2 framework loci of 11.30 cM. The size of the linkage groups ranged from 33 cM to 124.8 cM, with an average of 74.40\$ 16.90 cM.

Male map (Fig. 1)

The male map had 136 markers with 83 loci (60%) establishing the framework and covering 921.7 cM. The average distance between 2 framework loci was 11.10 cM. The framework loci were mapped into 12 linkage groups. The length of the groups ranged between 9.1 cM and 191.8 cM, with an average of 70.90\$27.80 cM. Fifty-two markers in the female map and 44 markers in the male map (testcross loci), including RAPDs, SSRs, isozymes and SCARs, were used as accessory markers. The number of linkage groups corresponded to the number of haploid chromosomes of *Quercus*: $n=12$, in both maps.

Homology between parental maps

Markers segregating in both parents including intercross markers (50 RAPDs, three isozymes and 2 microsatellites) and loci segregating 1 : 1 : 1 : 1 (three SCARs, 15 microsatellites, one isozyme and one 5S rDNA marker) were used to identify homology between parental maps. These 'bridge' markers are connected by bold lines in Fig. 1. Hence, ten homologous linkage groups were identified. Because of low information content between markerpairs corresponding to category A, B and A, C (Ritter *et al.*, 1990), the use of codominant markers segregating 1 : 1 : 1 : 1 and therefore segregating 1 : 1 from both parents is a prerequisite for constructing an accurate combined map. Merging the male and female maps was not our primary goal but this would be possible if more microsatellites were available.

Segregation distortion

At $\alpha=0.05$, segregation distortions from the Mendelian expectation of testcross markers (category A and D markers) were detected at 25 marker loci in the female map and 24 marker loci in the male map. The distortion level was quite similar in both parents (18%). At $\alpha=0.01$, the number of distorted markers decreased to 13 in the female map and 11 in the male map. Sixty percent of category A-distorted RAPD markers in the male map (66% in the female) displayed an excess heterozygote genotypes. Fifteen intercross RAPD markers showed strong deviations from Mendelian expectations, but only 2 of these were associated with a LOD score 54.5 to a framework marker in the map. Finally, a total of 61 markers (20%) showed segregation distortions at $\alpha=0.05$. This distortion level is higher than expected by chance alone (15 distorted markers at $\alpha=0.05$). Markers exhibiting segregation distortion were not randomly distributed in the genome. They tended to be preferentially grouped in a few clusters as shown on Fig. 1.

Genome length and map coverage

The estimated total map length (average of three values) for the female parent amounted to $E(G)=1192$ cM and for the male parent to $E(G)=1235$ cM. These values

are the first to be reported for oak. The different estimates, obtained from a range of Z values, were very close (data not shown) and were higher than the observed genome length of 893.2 cM for the female and 921.7 cM for the male map. This indicated that the mapped loci did not provide full coverage of the oak genome. Although the number of linkage groups corresponded to the number of haploid chromosomes ($n=12$), 2 small linkage groups (LG12) were identified for both parental maps. In addition, linkage group 1 of the female map was homologous to 2 unlinked linkage groups (1a and 1b) of the male map. This gap could be filled in by additional markers. Following Lange and Boehnke (1982), the minimal number, N, of randomly distributed markers required to cover a proportion of $P=95\%$ of a genome of size of $L=1200$ cM at a maximum distance between markers of $d=20$ cM is: $N = \ln(1/p) / \ln(1 - d/L) = 178$. This expectation was higher than the number of markers actually used in the initial step of map construction (136 markers for both the male and female map). According to this formula, 136 markers would provide 90% coverage for a 20-cM map, a percentage coverage that was very close to that obtained following Bishop *et al.*, (1983) (85% for both maps). The assumption of a random distribution of markers in the genome seemed to fit our experimental data mainly based on RAPDs. The observed and expected distributions of markers were compared for 20-cM intervals. A chi-square test for departure from a Poisson distribution was computed for six classes or groups of classes containing at least five observations. The mean of the Poisson distribution was set to the mean number of markers per 20-cM interval length. No significant departure ($\alpha=0.01$) from the Poisson expectation was observed in this goodness-of-fit test. A random distribution of RAPD markers was also observed in maritime pine (Plomion *et al.*, 1995).

3 Discussion

Pseudo-test-cross mapping in Q. robur

The *Q. robur* maps presented here were constructed following the two-way pseudo-testcross ' strategy. Two maps of similar size (893 cM in the female parent, 922 cM in the male parent) were obtained for each parent, each one represented by 12 linkage groups corresponding to the 12 haploid chromosomes identified in oaks (Ohri and Ahuja 1990). The two maps are furthermore equally balanced in number of framework markers, mean size of linkage groups and mean distance between framework markers. As oaks exhibit high levels of heterozygosity in natural populations (Kremer and Petit 1993; Zanetto *et al.*, 1994), 18% of the RAPD markers were found to be heterozygous in both parents (these are referred to as intercross loci segregating in the 3 : 1 ratio) compared to the 82% that segregated in a testcross configuration. Reported corresponding values in other species using the 'two-way pseudo-testcross ' are quite different: the number of RAPDs segregating in 3 : 1 proportions varied between 2% and 5% in *Eucalyptus* studies (Grattapaglia and Sederoff 1994; Verhaegen and Plomion 1996) and amounted to 5% in pines (Kubisiak *et al.*, 1995). These differences can likely be imputed to the differences in

levels of heterozygosity between these species. The higher proportion of heterozygote RAPD loci in both parents also allowed the assignment of homologous linkage groups of the two parents as suggested by Hemmat *et al.*, (1994). However, homologies between groups became more evident with codominant markers: microsatellites, minisatellite, SCARs, isozymes and the 5S rDNA marker. Among the 12 linkage groups identified in the two parents, 10 pairs could clearly be associated because of the presence of several common markers. A complete merging of the two maps into one single consensus map could be accomplished only if the ordering of markers were more precisely depended and if cosegregation analysis were conducted in larger families. Within the intraspecific F1 cross of *Q. robur*, a large number of markers (18% in both parental maps) exhibited segregation distortion regardless of the type of marker. In other examples of 'pseudo-testcross' mapping, distorted markers are present in lower proportions (9% in *Eucalyptus urophylla*, 8% in *Eucalyptus grandis* Verhaegen and Plomion 1996, 15% in *Pinus palustris* and 14% in *Pinus elliottii* Kubisiak *et al.*, 1995). As in pines, eucalypts and pedunculate oak, there is a general trend towards a clustering of distorted markers on the linkage groups (linkage groups 2 and 4 in the female map, and linkage groups 4 and 9 in the male map, Fig. 1).

The large number of distorted markers in *Q. robur* may be related to the high genetic load observed in this species. Interestingly, the variation of diversity among eucalypts, pines and oaks as illustrated by the number of loci being heterozygous in the two parents follows the same pattern as the variation in the number of distorted loci. Although quantitative data on inbreeding depression are not available in oaks, observations of seedlings originating from control selfing crosses reveal that they experience several detrimental characters during their early development (poor germination, chlorophyll deficiency, reduced growth) (Kleinschmit and Kleinschmit 1996; and personal observations.).

Genetic map of *Q. robur* and other woody angiosperms and annual plants

To our knowledge this is the first reported genetic map of an oak species and of any member of the *Fagaceae* family. Other linkage studies based on a reduced number of isozyme loci and on several full-sib families have indicated that 4 loci (*Aap*, *ap*, *Idh*, *Acp*) are located on the same linkage group in *Q. robur* (Zanetto *et al.*, 1996) and *Q. petraea* (Müller-Starck *et al.*, 1996). Interestingly these 4 loci are distributed within linkage group 2 of our genetic map (Fig. 1). Genetic maps in other genera of broadleaved forest trees are currently only available in *Eucalyptus* (Grattapaglia and Sederoff 1994; Verhaegen and Plomion 1996; Byrne *et al.*, 1995) and in *Populus* (Bradshaw *et al.*, 1994). The comparison with woody angiosperms is therefore extended to fruit trees having related species in natural forests and to a few annual plants for documentary purposes (Table 3). With a few exceptions (*Eucalyptus* and *Quercus*) most existing maps of woody angiosperms cover less than 80% of the

genome. Based on the data of Table 3, the physical genome size of woody angiosperms varies between 0.27 pg DNA/C (*Prunus persica*) and 0.9 pg DNA/C (*Quercus robur* L.), lying in the range of variation between *Arabidopsis* (0.15 pg DNA/C) and *Lycopersicon esculentum* (0.98 pg DNA/C) (Arumuganathan and Earle 1991). Most woody angiosperms have lower physical and genetic genome sizes than annual plants. *Quercus robur* L. has the largest physical genome among the angiosperm woody species so far investigated, but its size is not greater than the genome of tomato. Curiously, the number of chromosomes and the genetic length of the genome in tomato and oak are also of a similar magnitude. Given that 1 pg represents 0.915 $\times$ 10⁹ bp (Bennet and Smith 1976), the physical distance per unit of genetic distance in *Q. robur* amounts to 0.77 $\times$ 10⁶ bp/cM, which represents only 3.21 times that of *Arabidopsis thaliana* (Table 3).

Microsatellites and consensus maps in oaks and Fagaceae

As illustrated in our example where two maps were constructed (Fig. 1), homologies between linkage groups became obvious as codominant markers segregating in 1 : 2 : 1 or 1 : 1 : 1 : 1 or dominant markers segregating in 3 : 1 could be mapped in both parents. Because the standard error of recombination fraction between markers segregating 1 : 1 : 1 : 1 (recoded 1 : 1) and markers segregating 1 : 1 is lower than for the others (Ritter *et al.*, 1990), their position on a linkage group is more precise. Hence, they are preferential candidates as anchor markers and for comparisons across maps.

Microsatellite primers are transportable across oak species as shown by the experiment conducted by Steinkellner *et al.*, (1997b). Among the 18 microsatellites mapped in *Q. robur*, which were originally developed from *Q. petraea*, polymerase chain reaction (PCR) amplification and polymorphism were tested for 14 loci in other oak species and other genera of the *Fagaceae*. All amplified products for the 14 loci were polymorphic in oak species belonging to the same botanical section (*epidobalanus*.) The number of successful loci was 11/14 reduced to 64% in oak species of section *Cerris* and to 47% in oak species of section *Erythrobalanus*. These results are reinforced by our own experience with microsatellite loci developed from *Quercus macrocarpa* (Dow *et al.*, 1995) and *Quercus salicina* (Kawahara T, unpublished data) and that were applied for mapping purposes in the present contribution. However, successful amplification of oak microsatellites in related genera of the family *Fagaceae* was reduced to 24% in *Fagus* and to 47% in *Castanea*. Even though the proportion of amplified fragments and the level of polymorphism decreased with increasing evolutionary distance, microsatellites can be regarded as preferential anchor and consensus markers across different oak species in the *Quercus* genus.

In the future these maps will be saturated using additional markers like amplified fragment length polymorphism (AFLP, Vos *et al.*, 1995) and further integrated into a single consensus map when more microsatellites can be localised. This consensus

map will be of great value in identifying quantitative trait loci of interest and particularly of those involved in adaptive traits of oaks.

Acknowledgements

The authors gratefully acknowledge Dr. T. Kawahara for sharing primers of the MicJ-AG22 locus, Anne Zanetto and Alexis Ducousso for their contribution in the controlled cross and Bernard Montousse for field technical assistance. The research was supported by an EU research project (B104 CT96 0706) 'Development, optimization and validation of molecular tools for assessments of biodiversity in forest trees'.

References

- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 35 : 1439-1445
- Becker M, Levy G (1990) Le point sur l'écologie comparée du chêne sessile et du chêne pédonculé. *Rev For Fr* 42 : 148-154
- Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* 19 : 137-144
- Bennet MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. *Philos Trans R Soc London* 216 : 179-199
- Besendorfer V, Zoldos V, Peskan T, Krnik-Rasol M, Littway T, Papes D (1996) Identification of potential cytogenetical and biochemical markers in bioindication of common oak forest. *Phyton* (Horn, Austria), Special issue 36 : 139-146
- Bishop DT, Cannings C, Skolnick M, Williamson JA (1983) The number of polymorphic DNA clones required to map the human genome. In: Wier BS (ed) *Statistical analysis of DNA sequence data*. Marcel Dekker, New York, pp 181-200
- Bodenes C, Laigret F, Kremer A (1996) Inheritance and molecular variations of PCR-SSCP fragments in pedunculate oak (*Quercus robur* L.). *Theor Appl Genet* 93 : 348-354
- Bodenes C, Joandet S, Laigret F, Kremer A (1997) Detection of genomic regions differentiating two closely related oak species *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Heredity* 78 : 433-444
- Bradshaw HD Jr, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. *Theor Appl Genet* 89 : 167-179
- Butorina AK (1993) Cytogenetic study of diploid and spontaneous triploid oaks, *Quercus robur* L. *Ann Sci For* 50 : 144s-150s
- Byrne M, Murrell JC, Allen B, Moran GF (1995) An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. *Theor Appl Genet* 91 : 869-876
- Cai Q, Guy CL, Moore GA (1994) Extension of the linkage map in *Citrus* using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. *Theor Appl Genet* 89 : 606-614
- Camus A (1954) *Les Chênes - Monographie du genre Quercus*. Editions Le Chevalier, Paris
- Causse M, Santoni S, Damerval C, Maurice A, Charcosset A, Deatrick J, de Vienne D (1996) A composite map of expressed sequences in maize. *Genome* 39 : 418-432
- Chakravarti A, Lasher LA, Reier JE (1991) A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* 128 : 175-182
- Chaparro JX, Werner DJ, O'Malley D, Sederstrom SS (1994) Targeted mapping and linkage analysis of morphological, isozyme, and RAPD markers in peach. *Theor Appl Genet* 87 : 805-815
- Crouzillat D, Lerceteau E, Petiard V, Morera J, Rodriguez H, Walker D, Phillips W, Ronning C, Schnell R, Osei J, Fritz P (1996) *Theobroma cacao* L.: a genetic linkage map and quantitative trait loci analysis. *Theor Appl Genet* 93 : 205-214

- Dirlewanger E, Proner V, Parvery C, Rothan C, Guye A, Monet R (1998) Genetic linkage map of peach (*Prunus persica* (L.) Batsch) using morphological, RFLP, isoenzyme, RAPD and AFLP markers. TAG (in press)
- Dow BD, Ashley MV (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Mol Ecol 5 : 615-627
- Dow BD, Ashley MV, Howe HF (1995) Characterisation of highly variable (GA/CT)/microsatellites in the bur oak, *Quercus macrocarpa*. Theor Appl Genet 91 : 137-141
- Durham RE, Liou PC, Gmitter FG, Moore GA (1994) Linkage of restriction length polymorphisms and isozymes in *Citrus*. Theor Appl Genet 84 : 39-48
- Favre JM, Arnould MF (1996) Nuclear DNA content and interspecific ribosomal variation between *Q. robur* and *Q. petraea*. In: Kremer A, Muhs H (eds) Inter- and intra-specific variation in European oaks: evolutionary implications and practical consequences. European Commission, EUR 16717EN, Luxembourg, pp 41-51
- Favre JM, Brown S (1996) A flow cytometric evaluation of the nuclear DNA content and GC percent in genomes of European oak species. Ann Sci For 53 : 915-917
- Foolad MR, Aruselkar S, Becerra V, Bliss FA (1995) A genetic map of *Prunus* based on an interspecific cross between peach and almond. Theor Appl Genet 91 : 262-269
- Grattapaglia D, Sederberg R (1994) Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137 : 1121-1137
- Hemmat M, Weeden NF, Manganaris AG, Lawson DM (1994) Molecular marker linkage map for apple. J Hered 84 : 4-11
- Jacobs JME, Van Eck HJ, Arens P, Verkerck-Bakker B, te Lintel Hehhert B, Bastiaanssen HJM, El-Kharbotly A, Pereira A, Jacobsen E, Stiekema WJ (1995) A genetic map of potato (*Solanum tuberosum*) integrating molecular markers, including transposons and classical markers. Theor Appl Genet 91 : 289-300
- Johnsson H (1946) Chromosome numbers of twin plants of *Quercus robur* L. and *Fagus sylvatica*. Hereditas 32 : 469-472
- Kleinschmit JRG, Kleinschmit J (1996) Artificial hybridization between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. In: Kremer A, Muhs H (eds) Inter- and intra-specific variation in European oaks: evolutionary implications and practical consequences. European Commission, EUR 16717EN, Luxembourg, pp 69-87
- Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of *Citrus*. Theor Appl Genet 94 : 701-706
- Kleinhofs A, Kilian A, Shagai-Maroo MA, Biyashev RM, Hayes P, Chen FQ, Lapitan N, Fenwick A, Blake TK, Kanazin V, Ananiev E, Dahleen L, Kudrna D, Bollinger J, Knapp SJ, Liu B, Sorrells M, Heun M, Franckowiack JD, Homan D, Skadsen R, Steenson BJ (1993) A molecular, isozyme and morphological map of barley (*Hordeum vulgare*) genome. Theor Appl Genet 86 : 705-712
- Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12 : 172-175
- Kremer A, Petit RJ (1993) Gene diversity in natural populations of oak species. Ann Sci For 50 : 186s-202s
- Kremer A, Savill PS, Steiner KC (1993) Genetics of oaks. Ann Sci For 50 : 1s-472s
- Kubisiak TL, Nelson CD, Nance WL, Stine M (1995) RAPD linkage mapping in a longleaf pine slash pine F1 family. Theor Appl Genet 90 : 1119-1127
- Lanaud C, Risterucci AM, NÖGoran AKJ, Clement D, Flament MH, Laurent V, Falque M (1995) A genetic linkage map of *Theobroma cacao* L. Theor Appl Genet 91 : 987-993
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1 : 174-181
- Lange K, Boehnke M (1982) How many polymorphic genes will it take to span the human genome? Am J Hum Genet 34 : 842-845
- Lodhi MA, Daly MJ, Ye G-N, Weeden NF, Reisch BI (1995) A molecular marker based linkage map of vitis. Genome 38 : 786-794
- Luro F, Lorieux M, Laigret F, Bove JM, Ollitrault P (1995) Cartographie du genome des agrumes a l'aide de marqueurs moleculaires et distortions de segregation. In: Berville A, Tersac M (eds) Techniques et utilisation des marqueurs moleculaires. INRA Editions, Paris, pp 69-82

- Moreau F, Kleinschmit J, Kremer A (1994) Molecular differentiation between *Q. petraea* and *Q. robur* assessed by random amplified DNA fragments. *For Genet* 1 : 51-64.
- Müller-Starck G, Zanetto A, Kremer A, Herzog S (1996) Inheritance of isoenzymes in sessile oak (*Quercus petraea* (Matt.) Liebl.) and offspring from interspecific crosses. *For Genet* 3 : 1-12
- Ohri D, Ahuja MR (1990) Giemsa C - banded karyotype in *Quercus* L. (Oak). *Silvae Genet* 39 : 216-219
- Plomion C, O'Malley DM, Durel C-E (1995) Genomic analysis in maritime pine (*Pinus pinaster*). Comparison of two RAPDs maps using selfed and open-pollinated seeds of the same individual. *Theor Appl Genet* 90 : 1028-1034
- Rajapakse S, Belthoff LE, He G, Estager AE, Scorza R, Verde I, Ballard RE, Baird WV, Callahan A, Monet R, Abbott AG (1995) Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. *Theor Appl Genet* 90 : 503-510
- Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV, Scolnik PA (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. *Proc Natl Acad Sci USA* 89 : 1477-1481
- Ritter E, Gebhardt C, Salamini F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics* 125 : 645-654
- Saghai-Marouf MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. *Proc Natl Acad Sci USA* 81 : 8014-8018
- Sharon D, Cregan PB, Mhammed S, Kusharska M, Hillel J, Lahav E, Lavi U (1997) An integrated genetic linkage map of avocado. *Theor Appl Genet* 95 : 911-922
- Specht T, Szymanski M, Barciszewska MZ, Barciszewski J, Erdmann VA (1997) Compilation of 5S rRNA and 5S rRNA gene sequences. *Nucleic Acids Research* 25 : 96-97
- Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. *Plant J* 3 : 739-744
- Steinkellner H, Fluch S, Turetschek E, Lexer C, Streið R, Kremer A, Burg K, Glössl J (1997a) Identification and characterization of (GA/CT)_n - microsatellite loci from *Quercus petraea*. *Plant Mol Biol* 33 : 1093-1096
- Steinkellner H, Lexer C, Turetschek E, Glössl J (1997b) Conservation of (GA)_n microsatellite loci between *Quercus* species. *Mol Ecol* 6 : 1189-1194
- Streið R, Labbe T, Bacilieri R, Steinkellner H, Glössl J, Kremer A (1998) Within population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Mol Ecol* 7 : 317-328
- Tanksley SD, Ganai MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Miller JC, Miller L, Paterson AH, Pineda O, Röder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132 : 1141-1160
- Verhaegen D, Plomion C (1996) Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. *Genome* 39 : 1051-1061
- Viruel MA, Messegueur R, de Vicente MC, Garcia-Mas J, Puigdomenech P, Vargas F, Arus P (1995) A linkage map with RFLP and isozymes markers for almond. *Theor Appl Genet* 91 : 964-971
- Vos P, Hogers R, Bleeker M, Reijmans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23 : 4407-4414
- Williams JGK, Kubelik AR, Livak KJ, Rafalski A, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18 : 6531-6535
- Zanetto A, Roussel G, Kremer A (1994) Geographic variation of inter-specific differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) For Genet 1 : 111-123
- Zanetto A, Kremer A, Müller-Starck G, Hattermer HH (1996) Inheritance of isozymes in Pedunculate oak (*Quercus robur* L.) *J Hered* 87 : 364-370
- Zoldos V, Papes D, Brown S, Panaud O, Sijjak-Yakovlev S (1998) Genome size and base composition of seven *Quercus* species: inter- and intra-population variation. *Genome* (in press)

Table 1 Protocols used for the molecular analysis of the different markers

Marker type	Number of markers	PCR		protocol		Electrophoresis	Staining conditions	References
		Primer type		Amplification conditions				
RAPD	271	Ten-base primers Operon Technologies kits A-Z	1 step	94°C	4 min	1.8% agarose gel	Ethidium bromide	Williams et al. 1990
			35 cycles	92°C 40°C 72°C	45 s 45 s 1 min 45 s			
SCAR	10	*	1 step	94°C 94°C T°C 72°C	3 min 45 s 45 s 1 min 30 s	1% agarose gel or 8% non-denaturing polyacrylamide gel	Ethidium bromide Silver staining	Bodénès et al. 1997
			30 cycles					
Microsatellite	18	*	1 step	95°C T°C 72°C 92°C 72°C	6 min 1 min 30 s 1 min 8 min	6% denaturing polyacrylamide gel	Fluorescence (ALF express)	Steinkellner et al. 1997b
			30 cycles				Silver staining	Streiff et al. 1998
			1 step					
Minisatellite	1	*	1 step	94°C 94°C 50°C 72°C	4 min 45 s 45 s 30 s	2% agarose gel	Ethidium bromide	Fluch and Burg unpublished data
			30 cycles					
5S rDNA	1	*	1 step	96°C 94°C 55°C 72°C	3 min 40 s 1 min 40 s	8% non-denaturing polyacrylamide gel	Silver staining	Specht et al. 1997
			30 cycles					
Isozymes	6	—	—	—	—	11%–12% starch gel or IEF	^b	Zanetto et al. 1996

^a See Table 2 for specific primer sequences and annealing temperatures^b See Zanetto et al. (1996) for staining conditions of various enzyme systems**Table 2** Continued

Locus name	Primer sequence	Annealing temperature (°C)	References
MSQ13	F: TGG CTG CAC CTA TGG CTC TTA G B: ACA CTC AGA CCC ACC ATT TTT CC	50	Dow et al 1995
MSQ16	F: GGA ACA ACT AGA GAG AAC CAA GTC AGG B: TTG CCT ATC CTG CC CCG TAT CAC	50	Dow and Ashley 1996
MicJ-AG22	F: CGA TTT TGA ACA CCT TCT TC B: CGA TGT GGG ATT TTG ATT TT	46	T. Kawahara (personal communication)
Omin3	F: AGA ATC GAC TTC AAT GCA AG B: CCA TTC ACA GAG GTT GTC GG	50	Fluch et al. (personal communication)
5S rDNA	F: TGG GAA GTC CTC GTG TTG CA B: ATT AGT GCT GGT ATG ATC GC	55	Specht et al. 1997

Fig. 1 Framework linkage maps of *Quercus robur* L. Female map on the *left* and male map on the *right*. Markers were mapped with a minimal LOD score of 3.0. Accessory markers (testcross and intercross types) were located beside their closest framework marker. Intercross markers (category B and C, see Results section) and some recoded testcross markers (category D) were used to identify homologous linkage groups. These markers are interconnected with *bold lines*. Intercross RAPD markers are *boxed*, and microsatellite,

minisatellite, SCAR and 5S rDNA markers are *circled*. Markers displaying distorted segregation ($\alpha^2 0.05$) are denoted by **boldface asterisk**

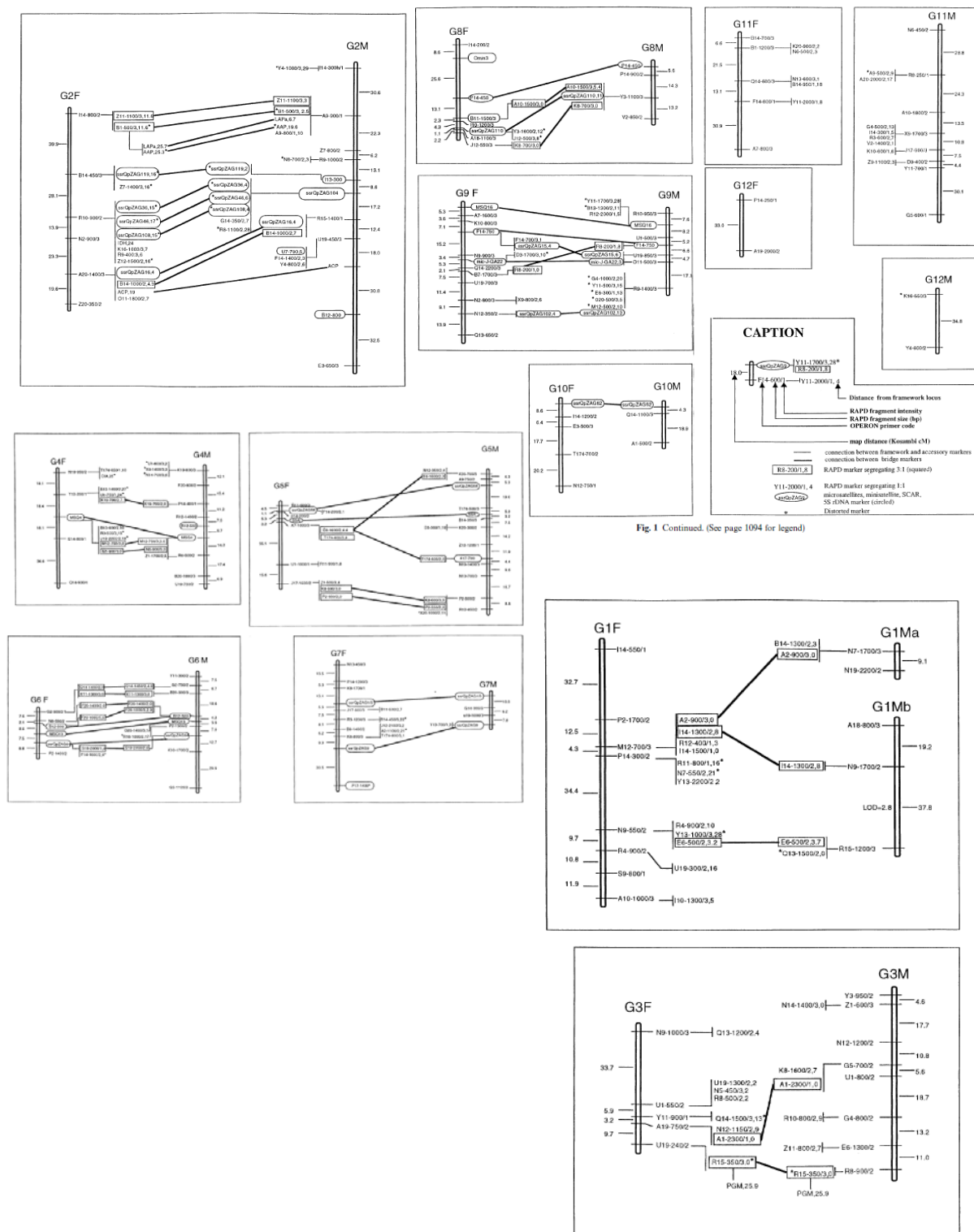


Table 3 Continued

Species (chromosome number) (pg/C)	Population type	Number of maps	Marker type	Number of markers	Observed map length	Map length coverage	Estimated map length ^b	References
<i>Theobroma cacao</i> L. (2n = 20) ((0.4 pg/C))	Intraspecific F ₁	1	RAPD, RFLP, isozymes	193	759 cM	—	1000–1500 cM	Lanaud et al. 1995 (this study)
	Intraspecific BC ₁	1	RAPD, RFLP	138	1068 cM	—		Crouzillat et al. 1996
<i>Vitis</i> (2n = 38) ((0.5 pg/C))	Interspecific F ₁	2	RAPD, RFLP, isozymes	523	1196 cM (female)	—	1900–2850 cM	Lodhi et al. 1995
					1477 cM (male)	—		
<i>Arabidopsis thaliana</i> (2n = 10) ((0.15 pg/C))	RIL	1	RAPD, microsatellites	302	675 cM	~ 100%	500–750 cM	Reiter et al. 1992
				30				Bell et al. 1994 ^c
<i>Hordeum vulgare</i> (2n = 14) ((5.50 pg/C))	DH	1	RAPD, RFLP, isozymes	295	1250 cM	~ 100%	700–1050 cM	Kleinhofs et al. 1993
<i>Lycopersicon esculentum</i> (2n = 24) ((0.98 pg/C))	F ₂	1	RFLP, isozymes, morphological	1030	1276 cM	~ 100%	1200–1800 cM	Tanksley et al. 1992
<i>Zea mays</i> (2n = 20) ((2.5 pg/C))	RIL, F ₂	1	RFLP, isozymes, molecular markers	275	1765 cM	~ 100%	1000–1500 cM	Causse et al. 1996

^aF₁, Two-generation outbred pedigree; F₂, three-generation inbred pedigree; G₂, three-generation outbred pedigree; RIL, recombinant inbred lines; DH, doubled-haploid lines; BC₁, first-generation backcross

^b100–150 cM per chromosome according to Rajapakse et al. (1995)

^cSee <http://cbil.humgen.upenn.edu/Natgc/SSLP.info/SSLP.map.html>

Table 3 Continued

Species (chromosome number) (pg/C)	Population type	Number of maps	Marker type	Number of markers	Observed map length	Map length coverage	Estimated map length ^b	References
<i>Theobroma cacao</i> L. (2n = 20) ((0.4 pg/C))	Intraspecific F ₁	1	RAPD, RFLP, isozymes	193	759 cM	—	1000–1500 cM	Lanaud et al. 1995 (this study)
	Intraspecific BC ₁	1	RAPD, RFLP	138	1068 cM	—		Crouzillat et al. 1996
<i>Vitis</i> (2n = 38) ((0.5 pg/C))	Interspecific F ₁	2	RAPD, RFLP, isozymes	523	1196 cM (female)	—	1900–2850 cM	Lodhi et al. 1995
					1477 cM (male)	—		
<i>Arabidopsis thaliana</i> (2n = 10) ((0.15 pg/C))	RIL	1	RAPD, microsatellites	302	675 cM	~ 100%	500–750 cM	Reiter et al. 1992
				30				Bell et al. 1994 ^c
<i>Hordeum vulgare</i> (2n = 14) ((5.50 pg/C))	DH	1	RAPD, RFLP, isozymes	295	1250 cM	~ 100%	700–1050 cM	Kleinhofs et al. 1993
<i>Lycopersicon esculentum</i> (2n = 24) ((0.98 pg/C))	F ₂	1	RFLP, isozymes, morphological	1030	1276 cM	~ 100%	1200–1800 cM	Tanksley et al. 1992
<i>Zea mays</i> (2n = 20) ((2.5 pg/C))	RIL, F ₂	1	RFLP, isozymes, molecular markers	275	1765 cM	~ 100%	1000–1500 cM	Causse et al. 1996

^aF₁, Two-generation outbred pedigree; F₂, three-generation inbred pedigree; G₂, three-generation outbred pedigree; RIL, recombinant inbred lines; DH, doubled-haploid lines; BC₁, first-generation backcross

^b100–150 cM per chromosome according to Rajapakse et al. (1995)

^cSee <http://cbil.humgen.upenn.edu/Natgc/SSLP.info/SSLP.map.html>

6 References

- Achere, V., Favre, J.M., Besnard, G., and Jeandroz, S. 2005. Genomic organization of molecular differentiation in Norway spruce (*Picea abies*). *Mol. Ecol.* **14**: 3191-3201.
- Adams, S.P., Leitch, I.J., Bennett, M.D., and Leitch, A.R. 2000. Aloe L. - A second plant family without (TTTAGGG)(n) telomeres. *Chromosoma* **109**: 201-205.
- Agarwal, M., Shrivastava, N., and Padh, H. 2008b. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* **27**: 617-631.
- Ahn, S. and Tanksley, S.D. 1993. Comparative linkage maps of the rice and maize genomes. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 7980-7984.
- Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, H., Hayden, M.J., Howes, N., Sharp, P., Vaughan, P., Rathmell, B., Huttner, E., and Kilian, A. 2006. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theoretical and Applied Genetics* **113**: 1409-1420.
- Akhunov, E., Nicolet, C., and Dvorak, J. 2009. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theoretical and Applied Genetics* **119**: 507-517.
- Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X., Richmond, T.A., Middle, C.M., Rodesch, M.J., Packard, C.J., Weinstock, G.M., and Gibbs, R.A. 2007. Direct selection of human genomic loci by microarray hybridization. *Nature Methods* **4**: 903-905.
- Allaby, R.G. and Brown, T.A. 2003. AFLP data and the origins of domesticated crops. *Genome* **46**: 448-453.
- Allan, G., Williams, A., Rabinowicz, P.D., Chan, A.P., Ravel, J., and Keim, P. 2008. Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLPs and SSRs. *Genetic Resources and Crop Evolution* **55**: 365-378.
- Alwine, J.C., Kemp, D.J., and Stark, G.R. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5350-5354.
- Angaji, S.A. 2009. Single nucleotide polymorphism genotyping and its application on mapping and marker-assisted plant breeding. *African Journal of Biotechnology* **8**: 908-914.
- Ayala, F.J. and Coluzzi, M. 2005. Chromosome speciation: Humans, *Drosophila*, and mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 6535-6542.
- Béaszczyk, L., Tyrka, M., and Che+ékowski, J. 2005. PstIAFLP based markers for leaf rust resistance genes in common wheat. *Journal of Applied Genetics* **46**: 357-364.
- Bacilieri, R., Ducouso, A., and Kremer, A. 1995. Genetic, Morphological, Ecological and Phenological Differentiation Between *Quercus petraea* (Matt) Liebl and *Quercus robur* L. in A Mixed Stand of Northwest of France. *Silvae Genetica* **44**: 1-10.
- Bacilieri, R., Ducouso, A., and Kremer, A. 1996a. Comparison of morphological characters and molecular markers for the analysis of hybridization in sessile and pedunculate oak 41. *Annales des Sciences Forestieres* **53**: 79-91.
- Bacilieri, R., Ducouso, A., Petit, R.J., and Kremer, A. 1996b. Mating system and asymmetric hybridization in a mixed stand of European oaks 39. *Evolution* **50**: 900-908.
- Bacilieri, R., Ducouso, A., Petit, R.J., and Kremer, A. 1996c. Mating system and asymmetric hybridization in a mixed stand of European oaks 39. *Evolution* **50**: 900-908.
- Bancroft, I. 2001. Duplicate and diverge: The evolution of plant genome microstructure. *Trends in Genetics* **17**: 89-93.
- Bansal, V., Bashir, A., and Bafna, V. 2007. Evidence for large inversion polymorphisms in the human genome from HapMap data. *Genome Research* **17**: 219-230.
- Barakat, A., Dilorieto, D.S., Zhang, Y., Smith, C., Baier, K., Powell, W.A., Wheeler, N., and Sederoff, R. 2009. Comparison of the transcriptomes of American chestnut (*Castanea*

- dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *Bmc Plant Biology* **9**.
- Barbosa,A.M.M., Geraldi,I.O., Benchimol,L.L., Garcia,A.A.F., Souza,J., and Souza,A.P. 2003. Relationship of intra- and interpopulation tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. *Euphytica* **130**: 87-99.
- Barker,G.L.A. and Edwards,K.J. 2009. A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. *Plant Biotechnol. J.* **7**: 318-325.
- Barre,P., Moreau,L., Mi,F., Turner,L., Gastal,F., Julier,B., and Ghesquire,M. 2009. Quantitative trait loci for leaf length in perennial ryegrass (*Lolium perenne* L.). *Grass Forage Sci.* **64**: 310-321.
- Barreneche,T., Bodenes,C., Lexer,C., Trontin,J.F., Fluch,S., Streiff,R., Plomion,C., Roussel,G., Steinkellner,H., Burg,K., Favre,J.M., Glossl,J., and Kremer,A. 1998a. A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers 27. *Theoretical and Applied Genetics* **97**: 1090-1103.
- Barreneche,T., Casasoli,M., Russell,K., Akkak,A., Meddour,H., Plomion,C., Villani,F., and Kremer,A. 2004b. Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs) 11. *Theoretical and Applied Genetics* **108**: 558-566.
- Bassam,B.J., Caetano-Anolles,G., and Gresshoff,P.M. 1992. DNA amplification fingerprinting of bacteria. *Applied Microbiology and Biotechnology* **38**: 70-76.
- Bastianel,M., Cristofani-Yaly,M., de Oliveira,A.C., Freitas-Asta,J., Garcia,A.A.F., de Resende,M.D.V., Rodrigues,V., and Machado,M.A. Quantitative trait loci analysis of citrus leprosis resistance in an interspecific backcross family of (*Citrus reticulata* Blancoüüá C. *sinensis* L. Osbeck) C. *sinensis* L. Osb. *Euphytica* . 2009.
- Bastien,D., Favre,J.M., Collignon,A.M., Sperisen,C., and Jeandroz,S. 2003. Characterization of a mosaic minisatellite locus in the mitochondrial DNA of Norway spruce [*Picea abies* (L.) Karst.]. *Theoretical and Applied Genetics* **107**: 574-580.
- Bausher,M.G., Singh,N.D., Lee,S.B., Jansen,R.K., and Daniell,H. 2006. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': Organization and phylogenetic relationships to other angiosperms. *Bmc Plant Biology* **6**.
- Beavis,W.D. and Grant,D. 1991. A linkage map based on information from four F2 populations of maize (*Zea mays* L.). *Theoretical and Applied Genetics* **82**: 636-644.
- Becker,J. and Heun,M. 1995. Barley microsatellites: Allele variation and mapping. *Plant. Mol. Biol.* **27**: 835-845.
- Bender,J. 2002. Plant epigenetics. *Current biology* : CB **12**: R412-R414.
- Bennett,M.D. and Leitch,I.J. 2005. Plant genome size research: A field in focus. *Ann Bot* **95**: 1-6.
- Bennetzen,J.L. 2002. Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**: 29-36.
- Bensch,S. and Akesson,M. 2005. Ten years of AFLP in ecology and evolution: Why so few animals? *Mol. Ecol.* **14**: 2899-2914.
- Bereiter-Hahn,J., Vath,M., Mai,S., and Jendrach,M. 2008. Structural implications of mitochondrial dynamics. *Biotechnology Journal* **3**: 765-780.
- Berenyi,M., Gichuki,S.T., Schmidt,J., and Burg,K. 2002. Ty1-copia retrotransposon-based S-SAP (sequence-specific amplified polymorphism) for genetic analysis of sweetpotato. *Theoretical and Applied Genetics* **105**: 862-869.
- Berenyi,M., Mauleon,R.P., Kopecky,D., Wandl,S., Friedl,R., Fluch,S., Boonruangrod,R., Muge,E., and Burg,K. 2009. Isolation of Plant Gene space-Related sequence elements by high C+G patch (HCGP) filtration: Model Study on rice. *Plant Mol. Biol. Report.* **27**: 79-85.
- Bergmann,F. and Ruetz,W. 1991. Isozyme genetic variation and heterozygosity in random tree samples and selected orchard clones from the same Norway spruce populations. *Forest Ecology and Management* **46**: 39-47.
- Birky,C.W. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. 35, 125-148. 2001. Ref Type: Serial (Book,Monograph)
- Blas,A.L., Yu,Q., Chen,C., Veatch,O., Moore,P.H., Paull,R.E., and Ming,R. 2009. Enrichment of a papaya high-density genetic map with AFLP markers. *Genome* **52**: 716-725.
- Bodenes,C., Joandet,S., Laigret,F., and Kremer,A. 1997. Detection of genomic regions differentiating two closely related oak species *Quercus petraea* (Matt) Liebl and *Quercus robur* L 35. *Heredity* **78**: 433-444.

- Bogdanova,V.S. 2007. Inheritance of organelle DNA markers in a pea cross associated with nuclear-cytoplasmic incompatibility. *Theoretical and Applied Genetics* **114**: 333-339.
- Boonruangrod,R., Fluch,S., and Burg,K. Elucidation of origin of the present day hybrid banana cultivars using the 5'ETS rDNA sequence information. *Molecular Breeding* . 2009.
- Bossdorf,O., Richards,C.L., and Pigliucci,M. 2008. Epigenetics for ecologists. *Ecology Letters* **11**: 106-115.
- Botstein,D., White,R.L., Skolnick,M., and Davis,R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**: 314-331.
- Brennicke,A., Klein,M., Binder,S., Knoop,V., Grohmann,L., Malek,O., Marchfelder,A., Marienfeld,J., and Unseld,M. 1996. Molecular biology of plant mitochondria. *Naturwissenschaften* **83**: 339-346.
- Brewer,M.T., Moyseenko,J.B., Monforte,A.J., and Van Der Knaap,E. 2007. Morphological variation in tomato: A comprehensive study of quantitative trait loci controlling fruit shape and development. *Journal of Experimental Botany* **58**: 1339-1349.
- Budahn,H., Peterka,H., Mousa,M.A.A., Ding,Y., Zhang,S., and Li,J. 2009. Molecular mapping in oil radish (*Raphanus sativus* L.) and QTL analysis of resistance against beet cyst nematode (*Heterodera schachtii*). *Theoretical and Applied Genetics* **118**: 775-782.
- Burnier,J., Buerki,S., Arrigo,N., Kupfer,P., and Alvarez,N. 2009. Genetic structure and evolution of Alpine polyploid complexes: *Ranunculus kuepferi* (Ranunculaceae) as a case study. *Mol. Ecol.* **18**: 3730-3744.
- Busch,J.D., Miller,M.P., Paxton,E.H., Sogge,M.K., and Keim,P. 2000. Genetic variation in the endangered Southwestern Willow Flycatcher. *Auk* **117**: 586-595.
- Caetano-Anolles,G. and Bassam,B.J. 1993. DNA amplification fingerprinting using arbitrary oligonucleotide primers. *Applied Biochemistry and Biotechnology* **42**: 189-200.
- Caetano-Anolles,G., Callahan,L.M., and Gresshoff,P.M. 1997. The origin of bermudagrass (*Cynodon*) off-types inferred by DNA amplification fingerprinting. *Crop Science* **37**: 81-87.
- Cao,Z., Deng,Z., Wang,M., Wang,X., Jing,J., Zhang,X., Shang,H., and Li,Z. 2008. Inheritance and molecular mapping of an alien stripe-rust resistance gene from a wheat-*Psathyrostachys huashanica* translocation line. *Plant Science* **174**: 544-549.
- Cato,S.A. and Richardson,T.E. 1996. Inter- and intraspecific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. *TAG Theoretical and Applied Genetics* **93**: 587-592.
- Causse,M., Duffe,P., Gomez,M.C., Buret,M., Damidaux,R., Zamir,D., Gur,A., Chevalier,C., Lemaire-Chamley,M., and Rothan,C. 2004. A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. *Journal of Experimental Botany* **55**: 1671-1685.
- Causse,M.A., Fulton,T.M., Yong,G.C., Sang,N.A., Chunwongse,J., Wu,K., Xiao,J., Yu,Z., Ronald,P.C., Harrington,S.E., Second,G., McCouch,S.R., and Tanksley,S.D. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* **138**: 1251-1274.
- Celton,J.M., Tustin,D., Chagn,D., and Gardiner,S. 2009. Construction of a dense genetic linkage map for apple rootstocks using SSRs developed from *Malus* ESTs and *Pyrus* genomic sequences. *Tree Genetics & Genomes* **5**: 93-107.
- Cervera,M.T., Ruiz-Garcia,L., and Martinez-Zapater,J.M. 2002. Analysis of DNA methylation in *Arabidopsis thaliana* based on methylation-sensitive AFLP markers. *Molecular Genetics and Genomics* **268**: 543-552.
- Chaim,A.B., Borovsky,Y., Rao,G.U., Tanyolac,B., and Paran,I. 2003. fs3.1: A major fruit shape QTL conserved in *Capsicum*. *Genome* **46**: 1-9.
- Chan,K.F. and Sun,M. 1997. Genetic diversity and relationships detected by isozyme and RAPD analysis of crop and wild species of *Amaranthus*. *Theoretical and Applied Genetics* **95**: 865-873.
- Chappell,M., Robacker,C., and Jenkins,T.M. 2008. Genetic diversity of seven deciduous azalea species (*Rhododendron* spp. section *Pentanthera*) native to the Eastern United States. *J. Am. Soc. Hortic. Sci.* **133**: 374-382.
- Charlesworth,B., Sniegowski,P., and Stephan,W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**: 215-220.
- Ching,A., Caldwell,K.S., Jung,M., Dolan,M., Smith,O.S.H., Tingey,S., Morgante,M., and Rafalski,A.J. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC genetics* **3**: 19 doi:10.1186/1471-2156-3-19

- Chinnusamy,V. and Zhu,J.K. 2009. Epigenetic regulation of stress responses in plants. *Current Opinion in Plant Biology* **12**: 133-139.
- Cho,Y.G., Ishii,T., Temnykh,S., Chen,X., Lipovich,L., McCouch,S.R., Park,W.D., Ayres,N., and Cartinhour,S. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **100**: 713-722.
- Chokchaichamnankit,P., Anamthawat-Jonsson,K., and Chulalaksananukul,W. 2008. Chromosomal mapping of 18S-25S and 5S ribosomal genes on 15 species of Fagaceae from northern Thailand. *Silvae Genetica* **57**: 5-13.
- Conte,M.G., Gaillard,S., Lanau,N., Rouard,M., and Perin,C. 2008. GreenPhylDB: a database for plant comparative genomics. *NUCLEIC ACIDS RES.* **36**: D991-D998.
- Cui,L., Veeraraghavan,N., Richter,A., Wall,K., Jansen,R.K., Leebens-Mack,J., Makalowska,I., and dePamphilis,C.W. 2006. ChloroplastDB: the Chloroplast Genome Database. *Nucleic acids research.* **34**: D692-D696.
- Cullis,C.A., Vorster,B.J., Van Der Vyver,C., and Kunert,K.J. 2009a. Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants? *Ann Bot* **103**: 625-633.
- Cullis,C.A., Vorster,B.J., Van Der Vyver,C., and Kunert,K.J. 2009b. Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants? *Ann Bot* **103**: 625-633.
- Cummings,M.P., Nugent,J.M., Olmstead,R.G., and Palmer,J.D. 2003. Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcl* to the mitochondrial genome in angiosperms. *Current Genetics* **43**: 131-138.
- Daniell,H., Cohill,P.R., Kumar,S., and Dfourmantel,N. 2004. Chloroplast genetic engineering. *Molecular Biology and Biotechnology of Plant Organelles* 443-490.
- Darrigues,A., Yang,W., and Francis,D.M. DNA-microarray detection of molecular markers for improving color and nutritional quality in tomato. 789, 241-248. 2008.
- Ref Type: Serial (Book,Monograph)
- Deleu,W., Esteras,C., Roig,C., Gonz+ílez-To,M., Fern+índez-Silva,I., Gonzalez-Ibeas,D., Blanca,J., Aranda,M.A., Ars,P., Nuez,F., Monforte,A.J., Pic,M.B., and Garcia-Mas,J. 2009a. A set of EST-SNPs for map saturation and cultivar identification in melon. *Bmc Plant Biology* **9**.
- Deleu,W., Esteras,C., Roig,C., Gonz+ílez-To,M., Fern+índez-Silva,I., Gonzalez-Ibeas,D., Blanca,J., Aranda,M.A., Ars,P., Nuez,F., Monforte,A.J., Pic,M.B., and Garcia-Mas,J. 2009b. A set of EST-SNPs for map saturation and cultivar identification in melon. *Bmc Plant Biology* **9**.
- Deluc,L.G., Quilici,D.R., Decendit,A., Grimplet,J., Wheatley,M.D., Schlauch,K.A., Morillon,J.M., Cushman,J.C., and Cramer,G.R. 2009. Water deficit alters differentially metabolic pathways affecting important flavor and quality traits in grape berries of Cabernet Sauvignon and Chardonnay. *Bmc Genomics* **10**.
- Devos,K.M., Moore,G., and Gale,M.D. 1995. Conservation of marker synteny during evolution. *Euphytica* **85**: 367-372.
- Dimitri,P., Corradini,N., Rossi,F., and Verni|Ç,F. 2005. The paradox of functional heterochromatin. *BioEssays* **27**: 29-41.
- Djupedal,I. and Ekwall,K. 2009. Epigenetics: Heterochromatin meets RNAi. *Cell Research* **19**: 282-295.
- Dumolin-Lapegue,S., Kremer,A., and Petit,R.J. 1999a. Are chloroplast and mitochondrial DNA variation species independent in oaks? *Evolution* **53**: 1406-1413.
- Dwivedi,S.L., Crouch,J.H., Mackill,D.J., Xu,Y., Blair,M.W., Ragot,M., Upadhyaya,H.D., and Ortiz,R. The Molecularization of Public Sector Crop Breeding: Progress, Problems, and Prospects. 95, 163-318. 2007.
- E.Southern 1975. Detection of specific sequences among DNA fragments separated by gel-electrophoresis. *J Mol Biol* **98**: 503.
- Echt,C.S., May-Marquardt,P., Hseih,M., and Zahorchak,R. 1996. Characterization of microsatellite markers in eastern white pine. *Genome* **39**: 1102-1108.
- Eckert,A.J. and Carstens,B.C. 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Mol. Phylogenet. Evol.* **49**: 832-842.
- Edwards,K.J., Barker,J.H.A., Daly,A., Jones,C., and Karp,A. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. *BioTechniques* **20**: 758-760.

- Ellegren, H. 2008a. Comparative genomics and the study of evolution by natural selection. *Mol. Ecol.* **17**: 4586-4596.
- Ellegren, H. 2008b. Sequencing goes 454 and takes large-scale genomics into the wild. *Mol. Ecol.* **17**: 1629-1631.
- Elling, A.A. and Deng, X.W. 2009. Next-generation sequencing reveals complex relationships between the epigenome and transcriptome in maize. *Plant Signaling and Behavior* **4**: 760-762.
- Ellis, T.H.N., Poyser, S.J., Knox, M.R., Vershinin, A.V., and Ambrose, M.J. 1998. Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. *Molecular and General Genetics* **260**: 9-19.
- Ellstrand, N.C. 1992. Gene flow among seed plant populations. *New Forests* **6**: 241-256.
- Ellstrand, N.C. and Schierenbeck, K.A. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences of the United States of America* **97**: 7043-7050.
- Emberton, J., Ma, J., Yuan, Y., Sanmiguel, P., and Bennetzen, J.L. 2005. Gene enrichment in maize with hypomethylated partial restriction (HMPR) libraries. *Genome Research* **15**: 1441-1446.
- Engelke, T., Agbicodo, E., and Tatlioglu, T. 2004. Mitochondrial genome variation in *Allium ampeloprasum* and its wild relatives. *Euphytica* **137**: 181-191.
- Engelke, T. and Tatlioglu, T. 2000. Mitochondrial genome diversity in connection with male sterility in *Allium schoenoprasum* L. *Theoretical and Applied Genetics* **100**: 942-948.
- Erich, Y., Chang, K., Gordon, A., Ronen, R., Navon, O., Rooks, M., and Hannon, G.J. 2009. DNA Sudoku - Harnessing high-throughput sequencing for multiplexed specimen analysis. *Genome Research* **19**: 1243-1253.
- Escaravage, N., Pornon, A., Doche, B., and Till-Bottraud, I. 1997. Breeding system in an alpine species: *Rhododendron ferrugineum* L. (Ericaceae) in the French northern Alps. *Canadian Journal of Botany* **75**: 736-743.
- Escaravage, N., Questiau, S., Pornon, A., Doche, B., and Taberlet, P. 1998. Clonal diversity in a *Rhododendron ferrugineum* L. (Ericaceae) population inferred from AFLP markers. *Mol. Ecol.* **7**: 975-982.
- Escaravage, N. and Wagner, J. 2004. Pollination effectiveness and pollen dispersal in a *Rhododendron ferrugineum* (Ericaceae) population. *Plant Biol.* **6**: 606-615.
- Eyre-Walker, A. and Keightley, P.D. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet* **8**: 610-618.
- Faccioli, P., Pecchioni, N., Stanca, A.M., and Terzi, V. 1999. Amplified fragment length polymorphism (AFLP) markers for Barley malt fingerprinting. *Journal of Cereal Science* **29**: 257-260.
- Fauron, C.M.R., Casper, M., Gao, Y., and Moore, H. 1995. The maize mitochondrial genome: Dynamic, yet functional. *Trends in Genetics* **11**: 228-235.
- Fazekas, A.J., Kesanakurti, P.R., Burgess, K.S., Percy, D.M., Graham, S.W., Barrett, S.C.H., Newmaster, S.G., Hajibabaei, M., and Husband, B.C. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources* **9**: 130-139.
- Feleke, Y., Pasquet, R., and Gepts, P. 2006. Development of PCR-based chloroplast DNA markers that characterize domesticated cowpea (*Vigna unguiculata* ssp. *unguiculata* var. *unguiculata*) and highlight its crop-weed complex. *Plant Systematics and Evolution* **262**: 75-87.
- Feltus, F.A., Wan, J., Schulze, S.R., Estill, J.C., Jiang, N., and Paterson, A.H. 2004. An SNP resource for rice genetics and breeding based on subspecies Indica and Japonica genome alignments. *Genome Research* **14**: 1812-1819.
- Ferris, C., Oliver, R.P., Davy, A.J., and Hewitt, G.M. 1995a. Using chloroplast DNA to trace postglacial migration routes of oaks into Britain. *Molecular Ecology* **4**: 731-738.
- Flagel, L.E. and Wendel, J.F. 2009. Gene duplication and evolutionary novelty in plants. *New Phytol.* **183**: 557-564.
- Fluch, S., Olmo, C.C., Tauber, S., Stierschneider, M., Kopecky, D., Reichenauer, T.G., and Matusíková, I. 2008. Transcriptomic changes in wind-exposed poplar leaves are dependent on developmental stage. *Planta* **228**: 757-764.
- Forneck, A., Wegscheider, E., and Benjak, A. 2009. Clonal variation in pinot noir revealed by S-SAP involving universal retrotransposon-based sequences. *American Journal of Enology and Viticulture* **60**: 104-109.

- Fraser,D.J., Hansen,M.M., Ostergaard,S., Tessier,N., Legault,M., and Bernatchez,L. 2007. Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. *Mol. Ecol.* **16**: 3866-3889.
- Gale,M.D. and Devos,K.M. 1998. Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 1971-1974.
- Galindo,C.L., McIver,L.J., McCormick,J.F., Skinner,M.A., Xie,Y., Gelhausen,R.A., Ng,K., Kumar,N.M., and Garner,H.R. 2009. Global microsatellite content distinguishes humans, primates, animals, and plants. *Mol Biol Evol* **26**: 2809-2819.
- Galtier,N., Jobson,R.W., Nabholz,B., Glamin,S., and Blier,P.U. 2009. Mitochondrial whims: Metabolic rate, longevity and the rate of molecular evolution. *Biology Letters* **5**: 413-416.
- Garcia-Mas,J., Oliver,M., Gomez-Paniagua,H., and De Vicente,M.C. 2000. Comparing AFLP, RAPD and RFLP markers for measuring genetic diversity in melon. *Theoretical and Applied Genetics* **101**: 860-864.
- Geburek,T., Mottinger-Kroupa,S., Morgante,M., and Burg,K. 1998b. Genetic variation of Norway spruce (*Picea abies* (L.) Karst.) populations in Austria. II. Microspatial patterns derived from nuclear sequence tagged microsatellite sites. *For Genet* **5**: 231-237.
- Gedil,M., Kolade,F., Raji,A., Ingelbrecht,I., and Dixon,A. 2009. Development of molecular genomic tools for verification of intergeneric hybrids between castor bean (*Ricinus communis* L.) and cassava (*Manihot esculenta* Crantz). *Journal of Food, Agriculture and Environment* **7**: 534-539.
- Gerber,S., Mariette,S., Streiff,R., Bodenes,C., and Kremer,A. 2000. Comparison of microsatellites and amplified fragment markers for parentage analysis. *Mol. Ecol.* **9**: 1037-1048.
- Grant,D., Cregan,P., and Shoemaker,R.C. 2000. Genome organization in dicots: Genome duplication in Arabidopsis and synteny between soybean and Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 4168-4173.
- Gregory,T.R., Nicol,J.A., Tamm,H., Kullman,B., Kullman,K., Leitch,I.J., Murray,B.G., Kapraun,D.F., Greilhuber,J., and Bennett,M.D. 2007. Eukaryotic genome size databases. *NUCLEIC ACIDS RES.* **35**: D332-D338.
- Grivet,D., Jeandroz,S., and Favre,J.M. 1999. Nad1 b/c intron polymorphism reveals maternal inheritance of the mitochondrial genome in *Picea abies*. *Theoretical and Applied Genetics* **99**: 346-349.
- Gugerli,F., Sperisen,C., Büchler,U., Magni,F., Geburek,T., Jeandroz,S., and Senn,J. 2001a. Haplotype variation in a mitochondrial tandem repeat of Norway spruce (*Picea abies*) populations suggests a serious founder effect during postglacial re-colonization of the western Alps. *Mol. Ecol.* **10**: 1255-1263.
- Gunderson,K.L., Kuhn,K.M., Steemers,F.J., Ng,P., Murray,S.S., and Shen,R. 2006. Whole-genome genotyping of haplotype tag single nucleotide polymorphisms. *Pharmacogenomics* **7**: 641-648.
- Guo,X., Zhang,Z., Gerstein,M.B., and Zheng,D. 2009. Small RNAs originated from pseudogenes: cis- or trans-acting? *PLoS Computational Biology* **5**.
- Gupta,M. and Young,B.D. 2009. Application of SNP genotype arrays to determine somatic changes in cancer. *Methods in molecular biology (Clifton, N. J.)* **538**: 179-206.
- Gupta,P.K., Langridge,P., and Mir,R.R. Marker-assisted wheat breeding: present status and future possibilities. *Molecular Breeding* . 11. Dec. 2009. 10.1007/s11032-009-9359-7
- Gupta,P.K., Rustgi,S., and Mir,R.R. 2008. Array-based high-throughput DNA markers for crop improvement. *Heredity* **101**: 5-18.
- Hall,N. 2007. Advanced sequencing technologies and their wider impact in microbiology. *Journal of Experimental Biology* **210**: 1518-1525.
- Hamilton,M.B., Pincus,E.L., Di Fiore,A., and Fleischer,R.C. 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques* **27**: 500-507.
- Hanson,M.R. and Folkerts,O. 1992. Structure and function of the higher plant mitochondrial genome. *International Review of Cytology* **141**: 129-172.
- Hauge,B.M., Hanley,S.M., Cartinhour,S., Michael Cherry,J., Goodman,H.M., Koornneef,M., Stam,P., Chang,C., Kempin,S., Medrano,L., and Meyerowitz,E.M. 1993. An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. *Plant J.* **3**: 745-754.
- Hayatsu,H., Shiraishi,M., and Negishi,K. 2008. Bisulfite modification for analysis of DNA methylation. *Current Protocols in Nucleic Acid Chemistry*. June 2008; Unit 6.10; DOI: 10.1002/0471142700.nc0610s33; <http://www.currentprotocols.com/protocol/nc0610>

- Heinze, B. 2007. A database of PCR primers for the chloroplast genomes of higher plants. *Plant Methods* **27**:3:4
- Helmersson, A., Von Arnold, S., Burg, K., and Bozhkov, P.V. 2004. High stability of nuclear microsatellite loci during the early stages of somatic embryogenesis in Norway spruce. *Tree Physiology* **24**: 1181-1186.
- Hemmat, M., Weeden, N.F., Manganaris, A.G., and Lawson, D.M. 1994. Molecular marker linkage map for apple. *Journal of Heredity* **85**: 4-11.
- Herrera, P., Alvarado, A., Cortina, G., Combes, M.C., Romero, G., and Lashermes, P. 2009. Genetic analysis of partial resistance to coffee leaf rust (*Hemileia vastatrix* Berk & Br.) introgressed into the cultivated *Coffea arabica* L. from the diploid *C. canephora* species. *Euphytica* **167**: 57-67.
- Heslop-Harrison, J.S., Brandes, A., Taketa, S., Schmidt, T., Vershinin, A.V., Alkhimova, E.G., Kamm, A., Doudrick, R.L., Schwarzacher, T., Katsiotis, A., Kubis, S., Kumar, A., Pearce, S.R., Flavell, A.J., and Harrison, G.E. 1997. The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* **100**: 197-204.
- Heslop-Harrison, J.S. and Schwarzacher, T. 2007. Domestication, Genomics and the Future for Banana. *Ann Bot mcm191*.
- Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A., and Cox, D.R. 2005. Whole-genome patterns of common DNA variation in three human populations. *Science* **307**: 1072-1079.
- Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., Van Der Bank, M., Chase, M.W., Cowan, R.S., Erickson, D.L., Fazekas, A.J., Graham, S.W., James, K.E., Kim, K.J., Kress, W.J., Schneider, H., Van AlphenStahl, J., Barrett, S.C.H., Van Den Berg, C., Bogarin, D., Burgess, K.S., Cameron, K.M., Carine, M., Chacun, J., Clark, A., Clarkson, J.J., Conrad, F., Devey, D.S., Ford, C.S., Hedderson, T.A.J., Hollingsworth, M.L., Husband, B.C., Kelly, L.J., Kesanakurti, P.R., Jung, S.K., Kim, Y.D., Lahaye, R., Lee, H.L., Long, D.G., Madrianín, S., Maurin, O., Meusnier, I., Newmaster, S.G., Park, C.W., Percy, D.M., Petersen, G., Richardson, J.E., Salazar, G.A., Savolainen, V., Seberg, O., Wilkinson, M.J., Yi, D.K., and Little, D.P. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 12794-12797.
- Huang, C.Y., Ayliffe, M.A., and Timmis, J.N. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* **422**: 72-76.
- Huntley, B. and Birks, H.J.B. 1983. *An Atlas of past and Present Pollen Maps for Europe: 0-13,000 Years Ago*. Cambridge: University Press.
- Ikeda, H., Senni, K., Fujii, N., and Setoguchi, H. 2008. Post-glacial range fragmentation is responsible for the current distribution of *Potentilla matsumurae* Th. Wolf (Rosaceae) in the Japanese archipelago. *Journal of Biogeography* **35**: 791-800.
- Inoue, K. and Lupski, J.R. Molecular mechanisms for genomic disorders. *3*, 199-242. 2002.
- Isidore, E., Van Os, H., Andrzejewski, S., Bakker, J., Barrena, I., Bryan, G.J., Caromel, B., Van Eck, H., Ghareeb, B., De Jong, W., Van Koert, P., Lefebvre, V., Milbourne, D., Ritter, E., Van Der Voort, J.R., Rousselle-Bourgeois, F., Van Vliet, J., and Waugh, R. 2003. Toward a Marker-Dense Meiotic Map of the Potato Genome: Lessons from Linkage Group I. *Genetics* **165**: 2107-2116.
- Isobe, S., Kalliker, R., Hisano, H., Sasamoto, S., Wada, T., Klimenko, I., Okumura, K., and Tabata, S. 2009. Construction of a consensus linkage map for red clover (*Trifolium pratense* L.). *Bmc Plant Biology* **9**.
- Jacobs, J.M.E., Van Eck, H.J., Arens, P., Verkerk-Bakker, B., te Lintel Hekkert, B., Bastiaanssen, H.J.M., El Kharbotly, A., Pereira, A., Jacobsen, E., and Stiekema, W.J. 1995. A genetic map of potato (*Solanum tuberosum*) integrating molecular markers, including transposons, and classical markers. *Theoretical and Applied Genetics* **91**: 289-300.
- Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985. Hypervariable 'minisatellite' regions in human DNA. *Biotechnology (Reading, Mass.)* **24**: 467-472.
- Jo, B.H., Suh, D.S., Cho, E.M., Kim, J., Ryu, G.H., and Chung, K.W. 2009. Characterization of polymorphic microsatellite loci in cultivated and wild *Panax ginseng*. *Genes and Genomics* **31**: 119-127.
- John, S., Shephard, N., Liu, G., Zeggini, E., Cao, M., Chen, W., Vasavda, N., Mills, T., Barton, A., Hinks, A., Eyre, S., Jones, K.W., Ollier, W., Silman, A., Gibson, N., Worthington, J., and Kennedy, G.C. 2004. Whole-genome scan, in a complex disease, using 11,245 single-

- nucleotide polymorphisms: Comparison with microsatellites. *American Journal of Human Genetics* **75**: 54-64.
- Johnk, N. and Siegmund, H.R. 1997. Population structure and post-glacial migration routes of *Quercus robur* and *Quercus petraea* in Denmark, based on chloroplast DNA analysis. *Scandinavian Journal of Forest Research* **12**: 130-137.
- Juan, A., Crespo, M.B., Cowan, R.S., Lexer, C., and Fay, M.F. 2004. Patterns of variability and gene flow in *Medicago citrina*, an endangered endemic of islands in the western Mediterranean, as revealed by amplified fragment length polymorphism (AFLP). *Mol. Ecol.* **13**: 2679-2690.
- Kuelheim, C., Hui Yeoh, S., Maintz, J., Foley, W.J., and Moran, G.F. 2009. Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *Bmc Genomics* **10**: 452.
- Kameyama, Y., Isagi, Y., and Nakagoshi, N. 2002. Relatedness structure in *Rhododendron metternichii* var. *hondoense* revealed by microsatellite analysis. *Mol. Ecol.* **11**: 519-527.
- Kane, N.C. and Rieseberg, L.H. 2008. Genetics and evolution of weedy *Helianthus annuus* populations: Adaptation of an agricultural weed. *Mol. Ecol.* **17**: 384-394.
- Karp, A., Edwards, K.J., Bruford, M., Funk, S., Vosman, B., Morgante, M., Seberg, O., Kremer, A., Boursot, P., Arctander, P., Tautz, D., and Hewitt, G.M. 1997. Molecular technologies for biodiversity evaluation: Opportunities and challenges. *Nature Biotechnology* **15**: 625-628.
- Karro, J.E., Yan, Y., Zheng, D., Zhang, Z., Carriero, N., Cayting, P., Harrison, P., and Gerstein, M. 2007. Pseudogene.org: A comprehensive database and comparison platform for pseudogene annotation. *NUCLEIC ACIDS RES.* **35**: D55-D60.
- Karsai, I., Szucs, P., Maszáros, K., Puskás, K., Bedo, Z., and Veisz, O. 2007. Barley (*Hordeum vulgare* L.) marker linkage map: A case study of various marker types and of mapping population structure. *Cereal Research Communications* **35**: 1551-1562.
- Keeling, P.J. 2004. Diversity and evolutionary history of plastids and their hosts. *American Journal of Botany* **91**: 1481-1493.
- Kim, G.H., Yun, H.K., Choi, C.S., Park, J.H., Jung, Y.J., Park, K.S., Dane, F., and Kang, K.K. 2008. Identification of AFLP and RAPD markers linked to anthracnose resistance in grapes and their conversion to SCAR markers. *Plant Breeding* **127**: 418-423.
- Kim, S., Lee, Y.P., Lim, H., Ahn, Y., and Sung, S.K. 2009. Identification of highly variable chloroplast sequences and development of cpDNA-based molecular markers that distinguish four cytoplasm types in radish (*Raphanus sativus* L.). *TAG Theoretical and Applied Genetics* **119**: 189-198.
- Kirst, M., Myburg, A., and Sederoff, R. 2004. Genetic mapping in forest trees: markers, linkage analysis and genomics. *Genetic engineering* **26**: 105-141.
- Kobayashi, N., Mizuta, D., Nakatsuka, A., and Akabane, M. 2008. Attaining inter-subgeneric hybrids in fragrant azalea breeding and the inheritance of organelle DNA. *Euphytica* **159**: 67-72.
- Kofler, R., Schlotterer, C., and Lelley, T. 2007. SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics* **23**: 1683-1685.
- Koreth, J., O'Leary, J.J., and McGee, J.O. 1996. Microsatellites and PCR genomic analysis. *Journal of Pathology* **178**: 239-248.
- Kostia, S., Varvio, S.L., Vakkari, P., and Pulkkinen, P. 1995. Microsatellite sequences in a conifer, *Pinus sylvestris*. *Genome* **38**: 1244-1248.
- Kota, R., Varshney, R.K., Prasad, M., Zhang, H., Stein, N., and Graner, A. 2008. EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. *Functional and Integrative Genomics* **8**: 223-233.
- Kremer, A., Dupouey, J.L., Deans, J.D., Cottrell, J., Csai, U., Finkeldey, R., Espinel, S., Jensen, J., Kleinschmit, J., Van Dam, B., Ducousso, A., Forrest, I., de Heredia, U.L., Lowe, A.J., Tutkova, M., Munro, R.C., Steinhoff, S., and Badaeu, V. 2002a. Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science* **59**: 777-787.
- Kress, W.J. and Erickson, D.L. 2008. DNA barcodes: Genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 2761-2762.
- Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A., and Mikami, T. 2000. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). *NUCLEIC ACIDS RES.* **28**: 2571-2576.

- Kuhn,D.N., Motamayor,J.C., Meerow,A.W., Borrone,J.W., and Schnell,R.J. 2008. SSCP markers provide a useful alternative to microsatellites in genotyping and estimating genetic diversity in populations and germplasm collections of plant specialty crops. *Electrophoresis* **29**: 4096-4108.
- Kumar,A. and Bennetzen,J.L. Plant retrotransposons. 33, 479-532. 1999.
- Labra,M., Imazio,S., Grassi,F., Rossoni,M., and Sala,F. 2004. Vine-1 retrotransposon-based sequence-specific amplified polymorphism for *Vitis vinifera* L. genotyping. *Plant Breeding* **123**: 180-185.
- Lagercrantz,U. and Ryman,N. 1990. Genetic structure of Norway spruce (*Picea abies*): Concordance of morphological and allozymic variation. *Evolution* **44**: 38-53.
- Lalli,D.A., Abbott,A.G., Zhebentyayeva,T.N., Badenes,M.L., Damsteegt,V., Polík,J., Kríka,B., and Salava,J. 2008. A genetic linkage map for an apricot (*Prunus armeniaca* L.) BC1 population mapping plum pox virus resistance. *Tree Genet. Genomes* **4**: 481-493.
- Lang,G. 1992. Some aspects of European late- and post-glacial flora history. *Acta Botanica Fennica* **144**: 1-17.
- Latta,R.G. 2004. Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytol.* **161**: 51-58.
- Latta,R.G., Linhart,Y.B., Fleck,D., and Elliot,M. 1998. Direct and indirect estimates of seed versus pollen movement within a population of ponderosa pine. *Evolution* **52**: 61-67.
- Latta,R.G. and Mitton,J.B. 1997. A comparison of population differentiation across four classes of gene marker in limber pine (*Pinus flexilis* James). *Genetics* **146**: 1153-1163.
- Lecomte,L., Duffa,P., Buret,M., Servin,B., Hospital,F., and Causse,M. 2004. Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. *Theoretical and Applied Genetics* **109**: 658-668.
- Leitch,I.J. and Heslop-Harrison,J.S. 1993. Physical mapping of four sites of 5S rDNA sequences and one site of the amylase-2 gene in barley (*Hordeum vulgare*). *Genome* **36**: 517-523.
- Leitch,I.J., Kahandawala,I., Suda,J., Hanson,L., Ingrouille,M.J., Chase,M.W., and Fay,M.F. 2009. Genome size diversity in orchids: Consequences and evolution. *Ann Bot* **104**: 469-481.
- Leitch,I.J., Soltis,D.E., Soltis,P.S., and Bennett,M.D. 2005. Evolution of DNA amounts across land plants (Embryophyta). *Ann Bot* **95**: 207-217.
- Lexer,C., Heinze,B., Alia,R., and Rieseberg,L.H. 2004. Hybrid zones as a tool for identifying adaptive genetic variation in outbreeding forest trees: lessons from wild annual sunflowers (*Helianthus* spp.). *Forest Ecology and Management* **197**: 49-64.
- Lexer,C., Heinze,B., Gerber,S., Macalka-Kampfer,S., Steinkellner,H., Kremer,A., and Glossl,J. 2000a. Microsatellite analysis of maternal half-sib families of *Quercus robur*, pedunculate oak: II. inferring the number of pollen donors from the offspring
24. *Theoretical and Applied Genetics* **100**: 858-865.
- Lexer,C., Kremer,A., and Petit,R.J. 2006a. Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology* **15**: 2007-2012.
- Li,R., Li,Y., Fang,X., Yang,H., Wang,J., Kristiansen,K., and Wang,J. 2009. SNP detection for massively parallel whole-genome resequencing. *Genome Research* **19**: 1124-1132.
- Liewlaksaneeyanawin,C., Zhuang,J., Tang,M., Farzaneh,N., Lueng,G., Cullis,C., Findlay,S., Ritland,C.E., Bohlmann,J., and Ritland,K. 2009. Identification of COS markers in the *Pinaceae*. *Tree Genet. Genomes* **5**: 247-255.
- Lindner,K.R., Seeb,J.E., Habicht,C., Knudsen,K.L., Kretschmer,E., Reedy,D.J., Spruell,P., and Allendorf,F.W. 2000. Gene-centromere mapping of 312 loci in pink salmon by half-tetrad analysis. *Genome* **43**: 538-549.
- Lister,R., Gregory,B.D., and Ecker,J.R. 2009. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Current Opinion in Plant Biology* **12**: 107-118.
- Litt,M. and Luty,J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* **44**: 397-401.
- Lu,Y., Rong,T., and Cao,M. 2008. Analysis of DNA methylation in different maize tissues. *Journal of Genetics and Genomics* **35**: 41-48.
- Lupski,J.R. 2007. Genomic rearrangements and sporadic disease. *Nature Genetics* **39**: S43-S47.

- Lynch, M., Koskella, B., and Schaack, S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science* **311**: 1727-1730.
- Lyon, E., Rong Mao, and Swensen, J. 2009. Mutation Detection by Real-Time PCR in Real-Time PCR. In Logan, J., Edwards, K., and Saunders, N., eds. Caister Academic Press.
- Müller-Starck, G., Baradat, P., and Bergmann, F. 1992. Genetic variation within European tree species. *New Forests* **6**: 23-47.
- Ma, C., Kimura, Y., Fujimoto, H., Sakai, T., Imamura, J., and Fu, T. 2000. Genetic diversity of Chinese and Japanese rapeseed (*Brassica napus* L.) varieties detected by RAPD markers. *Breeding Science* **50**: 257-265.
- Ma, J., Wing, R.A., Bennetzen, J.L., and Jackson, S.A. 2007. Plant centromere organization: a dynamic structure with conserved functions. *Trends in Genetics* **23**: 134-139.
- Main, B.J., Bickel, R.D., McIntyre, L.M., Graze, R.M., Calabrese, P.P., and Nuzhdin, S.V. 2009. Allele-specific expression assays using Solexa. *Bmc Genomics* **10**: 422.
- Manos, P.S., Doyle, J.J., and Nixon, K.C. 1999. Phylogeny, Biogeography, and Processes of Molecular Differentiation in *Quercus* Subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution* **12**: 333-349.
- Mantovani, P., Maccaferri, M., Sanguineti, M.C., Tuberosa, R., Catizone, I., Wenzl, P., Thomson, B., Carling, J., Huttner, E., DeAmbrogio, E., and Kilian, A. 2008. An integrated DArT-SSR linkage map of durum wheat. *Mol. Breed.* **22**: 629-648.
- Marchelli, P., Gallo, L., Scholz, F., and Ziegenhagen, B. 1998. Chloroplast DNA markers reveal a geographical divide across Argentinean southern beech *Nothofagus nervosa* (Phil.) Dim. et Mil. distribution area. *Theoretical and Applied Genetics* **97**: 642-646.
- Marefat, A., Ophel-Keller, K., Scott, E.S., and Sedgley, M. 2006. The use of ARMS PCR in detection and identification of *Xanthomonads* associated with pistachio dieback in Australia. *European Journal of Plant Pathology* **116**: 57-68.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., and Rothberg, J.M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.
- Marienfeld, J., Unseld, M., and Brennicke, A. 1999. The mitochondrial genome of *Arabidopsis* is composed of both native and immigrant information. *Trends Plant Sci.* **4**: 495-502.
- Mariette, S., Cottrell, J., Csaikl, U.M., Goikoechea, P., Kořenig, A., Lowe, A.J., Van Dam, B.C., Barreneche, T., BodeüeneCs, C., Streiff, R., Burg, K., Groppe, K., Munro, R.C., Tabbener, H., and Kremer, A. 2002. Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (MATT.) LIEBL. and *Q. robur* L. stands. *Silvae Genetica* **51**: 72-79.
- Marques, C.M., Brondani, R.P.V., Grattapaglia, D., and Sederoff, R. 2002. Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *Theoretical and Applied Genetics* **105**: 474-478.
- McCauley, D.E. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology and Evolution* **10**: 198-202.
- McCauley, D.E., Sundby, A.K., Bailey, M.F., and Welch, M.E. 2007. Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris* (Caryophyllaceae): Evidence from experimental crosses and natural populations. *Am. J. Bot.* **94**: 1333-1337.
- McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **36**, 344-355. 1950.
- Ref Type: Abstract
- McCulloch, S.D. and Kunkel, T.A. 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Research* **18**: 148-161.
- McDonald, M.W., Brooker, M.I.H., and Butcher, P.A. 2009. A taxonomic revision of *Eucalyptus camaldulensis* (Myrtaceae). *Australian Systematic Botany* **22**: 257-285.
- McGuigan, F.E.A. and Ralston, S.H. 2002. Single nucleotide polymorphism detection: Allelic discrimination using TaqMan. *Psychiatric Genetics* **12**: 133-136.

- McNally,K.L., Childs,K.L., Bohnert,R., Davidson,R.M., Zhao,K., Ulat,V.J., Zeller,G., Clark,R.M., Hoen,D.R., Bureau,T.E., Stokowski,R., Ballinger,D.G., Frazer,K.A., Cox,D.R., Padhukasahasram,B., Bustamante,C.D., Weigel,D., Mackill,D.J., Bruskiewich,R.M., Reftsch,G., Buell,C.R., Leung,H., and Leach,J.E. 2009. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 12273-12278.
- Menz,M.A., Klein,R.R., Unruh,N.C., Rooney,W.L., Klein,P.E., and Mullet,J.E. 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. *Crop Science* **44**: 1236-1244.
- Metzgar,D., Bytof,J., and Wills,C. 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research* **10**: 72-80.
- Meyers,B.C., Tingey,S.V., and Morgante,M. 2001. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Research* **11**: 1660-1676.
- Milbourne,D., Meyer,R., Bradshaw,J.E., Baird,E., Bonar,N., Provan,J., Powell,W., and Waugh,R. 1997. Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. *Mol. Breed.* **3**: 127-136.
- Milbourne,D., Meyer,R.C., Collins,A.J., Ramsay,L.D., Gebhardt,C., and Waugh,R. 1998. Isolation, characterisation and mapping of simple sequence repeat loci in potato. *Molecular and General Genetics* **259**: 233-245.
- Millen,R.S., Olmstead,R.G., Adams,K.L., Palmer,J.D., Lao,N.T., Heggie,L., Kavanagh,T.A., Hibberd,J.M., Gray,J.C., Morden,C.W., Calie,P.J., Jermin,L.S., and Wolfe,K.H. 2001. Many parallel losses of infa from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **13**: 645-658.
- Miller,L.M. 2000. Classifying genealogical origins in hybrid populations using dominant markers. *Journal of Heredity* **91**: 46-49.
- Mochida,K., Saisho,D., Yoshida,T., Sakurai,T., and Shinozaki,K. 2008. TriMEDB: A database to integrate transcribed markers and facilitate genetic studies of the tribe Triticeae. *BMC Plant Biology*. Jun 30; 8:72.
- Mogensen,H.L. 1996. The hows and whys of cytoplasmic inheritance in seed plants. *Am. J. Bot.* **83**: 383-404.
- Monna,L., Miyao,A., Inoue,T., Fukuoka,S., Yamazaki,M., Sun Zhong,H., Sasaki,T., and Minobe,Y. 1994. Determination of RAPD markers in rice and their conversion into sequence tagged sites (STSs) and STS-specific primers. *DNA Research* **1**: 139-148.
- Morgante,M., Hanafey,M., and Powell,W. 2002b. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* **30**: 194-200.
- Morgante,M., Hanafey,M., and Powell,W. 2002a. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* **30**: 194-200.
- Morozova,O. and Marra,M.A. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**: 255-264.
- Moustafa,A., Beszteri,B., Maier,U., Bowler,C., Valentin,K., and Bhattacharya,D. 2009. Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms. *Science* **324**: 1724-1726.
- Moyzis,R.K., Buckingham,J.M., Cram,L.S., Dani,M., Deaven,L.L., Jones,M.D., Meyne,J., Ratliff,R.L., and Wu,J.R. 1988. A highly conserved repetitive DNA sequence, (TTAGGG)(n), present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 6622-6626.
- Muir,G. and Filatov,D. 2007. A selective sweep in the chloroplast DNA of dioecious silene (Section *Elisanthe*). *Genetics* **177**: 1239-1247.
- Muir,G., Fleming,C.C., and Schlötterer,C. 2000. Species status of hybridizing oaks. *Nature* **405**: 1016.
- Muir,G., Lowe,A.J., Fleming,C.C., and Vogl,C. 2004. High nuclear genetic diversity, high levels of outcrossing and low differentiation among remnant populations of *Quercus petraea* at the margin of its range in Ireland. *Annals of Botany* **93**: 691-697.
- Mullis,K.B. and Faloona,F.A. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology* **155**: 335-350.
- Murayama,S., Habuchi,T., Yamagishi,H., and Terachi,T. 2003. Identification of a sequence-tagged site (STS) marker linked to a restorer gene for Ogura cytoplasmic male sterility in radish (*Raphanus sativus* L.) by non-radioactive AFLP analysis. *Euphytica* **129**: 61-68.

- Muschner, V.C., Lorenz-Lemke, A.P., Vecchia, M., Bonatto, S.L., Salzano, F.M., and Freitas, L.B. 2006. Differential organellar inheritance in *Passiflora*'s (*Passifloraceae*) subgenera. *Genetica* **128**: 449-453.
- Naito, K., Isagi, Y., and Nakagoshi, N. 1998. Isolation and characterization of microsatellites of *Rhododendron metternichii* Sieb. et Zucc. var. *hondoense* Nakai. *Molecular Ecology* **7**: 927-928.
- Nakamura, Y. 2009. DNA variations in human and medical genetics: 25 Years of my experience. *Journal of Human Genetics* **54**: 1-8.
- Namroud, M.C., Beaulieu, J., Juge, N., Laroche, J., and Bousquet, J. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol. Ecol.* **17**: 3599-3613.
- Nason, J.D., Ellstrand, N.C., and Arnold, M.L. 1992. Patterns of Hybridization and Introgression in Populations of Oaks, Manzanitas, and Irises. *American Journal of Botany* **79**: 101-111.
- Nazari, L. and Pakniyat, H. 2008. Genetic diversity of wild and cultivated barley genotypes under drought stress using RAPD markers. *Biotechnology* **7**: 745-750.
- Neale, D.B., Saghai-Maroo, M.A., Allard, R.W., Zhang, Q., and Jorgensen, R.A. 1988. Chloroplast DNA diversity in populations of wild and cultivated barley. *Genetics* **120**: 1105-1110.
- Nishant, K.T., Singh, N.D., and Alani, E. 2009. Genomic mutation rates: What high-throughput methods can tell us. *BioEssays* **31**: 912-920.
- Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* **13**: 1143-1155.
- Oddou-Muratorio, S., Vendramin, G.G., Buiteveld, J., and Fady, B. 2009. Population estimators or progeny tests: What is the best method to assess null allele frequencies at SSR loci? *Conservation Genetics* **10**: 1343-1347.
- Ohsawa, T. and Ide, Y. 2008. Global patterns of genetic variation in plant species along vertical and horizontal gradients on mountains. *Global Ecology and Biogeography* **17**: 152-163.
- Olmstead, R.G. and Palmer, J.D. 1994. Chloroplast DNA systematics: A review of methods and data analysis. *Am. J. Bot.* **81**: 1205-1224.
- Olson, M., Hood, L., Cantor, C., and Botstein, D. 1989. A common language for physical mapping of the human genome. *Science* **245**: 1434-1435.
- Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 2766-2770.
- Ortiz, M., Tremetsberger, K., Stuessy, T.F., Terrab, A., Garcia-Castano, J.L., and Talavera, S. 2009. Phylogeographic patterns in *Hypochaeris* section *Hypochaeris* (Asteraceae, Lactuceae) of the western Mediterranean. *Journal of Biogeography* **36**: 1384-1397.
- Pakniyat, H., Powell, W., Baird, E., Handley, L.L., Robinson, D., Scrimgeour, C.M., Nevo, E., Hackett, C.A., Caligari, P.D.S., and Forster, B.P. 1997. AFLP variation in wild barley (*Hordeum spontaneum* c. Koch) with reference to salt tolerance and associated ecogeography. *Genome* **40**: 332-341.
- Pakull, B., Groppe, K., Meyer, M., Markussen, T., and Fladung, M. 2009. Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). *Tree Genet. Genomes* **5**: 505-515.
- Palmer, J.D. 1990. Contrasting modes and tempos of genome evolution in land plant organelles. *Trends in Genetics* **6**: 115-120.
- Palmer, J.D., Singh, G.P., and Pillay, D.T.N. 1983. Structure and sequence evolution of three legume chloroplast DNAs. *MGG Molecular & General Genetics* **190**: 13-19.
- Palmer, J.D. and Stein, D.B. 1986a. Conservation of chloroplast genome structure among vascular plants. *Current Genetics* **10**: 823-833.
- Palomeque, T. and Lorite, P. 2008. Satellite DNA in insects: A review. *Heredity* **100**: 564-573.
- Parasnis, A.S., Ramakrishna, W., Chowdari, K.V., Gupta, V.S., and Ranjekar, P.K. 1999. Microsatellite (GATA)(n) reveals sex-specific differences in papaya. *Theoretical and Applied Genetics* **99**: 1047-1052.
- Parfrey, L.W., Lahr, D.J.G., and Katz, L.A. 2008. The dynamic nature of eukaryotic genomes. *Mol Biol Evol* **25**: 787-794.
- Parh, D.K., Jordan, D.R., Aitken, E.A.B., Mace, E.S., Jun-Ai, P., McIntyre, C.L., and Godwin, I.D. 2008. QTL analysis of ergot resistance in sorghum. *Theoretical and Applied Genetics* **117**: 369-382.

- Park, S.O., Hwang, H.Y., and Crosby, K.M. 2009. A genetic linkage map including loci for male sterility, sugars, and ascorbic acid in melon. *Journal of the American Society for Horticultural Science* **134**: 67-76.
- Patil, N., Bero, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., Hacker, C.R., Kautzer, C.R., Lee, D.H., Marjoribanks, C., McDonough, D.P., Nguyen, B.T.N., Norris, M.C., Sheehan, J.B., Shen, N., Stern, D., Stokowski, R.P., Thomas, D.J., Trulson, M.O., Vyas, K.R., Frazer, K.A., Fodor, S.P.A., and Cox, D.R. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**: 1719-1723.
- Paun, O., Fay, M.F., Soltis, D.E., and Chase, M.W. 2007. Genetic and epigenetic alterations after hybridization and genome doubling. *Taxon* **56**: 649-656.
- Pearl, S.A., Welch, M.E., and McCauley, D.E. 2009. Mitochondrial heteroplasmy and paternal leakage in natural populations of *Silene vulgaris*, a gynodioecious plant. *Mol Biol Evol* **26**: 537-545.
- Peleg, Z., Saranga, Y., Suprunova, T., Ronin, Y., Röder, M.S., Kilian, A., Korol, A.B., and Fahima, T. 2008. High-density genetic map of durum wheat x wild emmer wheat based on SSR and DArT markers. *Theoretical and Applied Genetics* **117**: 103-115.
- Pennisi, E. 2007. Taxonomy. Wanted: A barcode for plants. *Science* **318**: 190-191.
- Perry, D.J., Isabel, N., and Bousquet, J. 1999. Sequence-tagged-site (STS) markers of arbitrary genes: The amount and nature of variation revealed in Norway spruce. *Heredity* **83**: 239-248.
- Pestsova, E., Meinhard, J., Menze, A., Fischer, U., Windhavel, A., and Westhoff, P. 2008. Transcript profiles uncover temporal and stress-induced changes of metabolic pathways in germinating sugar beet seeds. *BMC Plant Biology* **8**: 122.
- Petit, R.J., Bodenes, C., Ducousso, A., Roussel, G., and Kremer, A. 2004a. Hybridization as a mechanism of invasion in oaks. *New Phytol.* **161**: 151-164.
- Petit, R.J., Brewer, S., Bordacs, S., Burg, K., Cheddadi, R., Coart, E., Cottrell, J., Csakl, U.M., Van Dam, B., Deans, J.D., Espinel, S., Fineschi, S., Finkeldey, R., Glaz, I., Goicoechea, P.G., Jensen, J.S., König, A.O., Lowe, A.J., Madsen, S.F., Matyas, G., Munro, R.C., Popescu, F., Slade, D., Tabbener, H., de Vries, S.G.M., Ziegenhagen, B., de Beaulieu, J.L., and Kremer, A. 2002. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence; *Forest Ecology and Management* **156**: 49-74.
- Petit, R.J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., and Vendramin, G.G. 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol. Ecol.* **14**: 689-701.
- Petit, R.J., Kremer, A., and Wagner, D.B. 1993. Geographic Structure of Chloroplast Dna Polymorphisms in European Oaks; *Theoretical and Applied Genetics* **87**: 122-128.
- Petit, R.J., Pineau, E., Demesure, B., Bacilieri, R., Ducousso, A., and Kremer, A. 1997a. Chloroplast DNA footprints of postglacial recolonization by oaks; *Proceedings of the National Academy of Sciences of the United States of America* **94**: 9996-10001.
- Pfeiffer, A., Olivieri, A.M., and Morgante, M. 1997. Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome* **40**: 411-419.
- Pich, U., Fuchs, J., and Schubert, I. 1996. How do *Alliaceae* stabilize their chromosome ends in the absence of TTTAGGG sequences? *Chromosome Research* **4**: 207-213.
- Podlaha, O. and Zhang, J. 2009. Processed pseudogenes: the 'fossilized footprints' of past gene expression. *Trends in Genetics* **25**: 429-434.
- Pomou, A. and Escaravage, N. 1999. Genotypic structure in clonal *Rhododendron ferrugineum* L. (Ericaceae) populations: Origin and maintenance. *Plant Ecology* **141**: 145-150.
- Powell, W., Morgante, M., McDevitt, R., Vendramin, G.G., and Rafalski, J.A. 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 7759-7763.
- Prasad, M., Kumar, N., Kulwal, P.L., R+Âder, M.S., Balyan, H.S., Dhaliwal, H.S., and Gupta, P.K. 2003. QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. *Theoretical and Applied Genetics* **106**: 659-667.
- Price, Z., Schulman, A.H., and Mayes, S. 2003. Development of new marker methods - An example from oil palm. *Plant Genetic Resources: Characterisation and Utilisation* **1**: 103-113.

- Qin,D., Wu,H., Peng,H., Yao,Y., Ni,Z., Li,Z., Zhou,C., and Sun,Q. 2008. Heat stress-responsive transcriptome analysis in heat susceptible and tolerant wheat (*Triticum aestivum* L.) by using Wheat Genome Array. *Bmc Genomics* **9**.
- Rabinowicz,P.D., Citek,R., Budiman,M.A., Nunberg,A., Bedell,J.A., Lakey,N., O'Shaughnessy,A.L., Nascimento,L.U., McCombie,W.R., and Martienssen,R.A. 2005. Differential methylation of genes and repeats in land plants. *Genome Research* **15**: 1431-1440.
- Rae,S.J., Macaulay,M., Ramsay,L., Leigh,F., Matthews,D., O'Sullivan,D.M., Donini,P., Morris,P.C., Powell,W., Marshall,D.F., Waugh,R., and Thomas,W.T.B. 2007. Molecular barley breeding. *Euphytica* **158**: 295-303.
- Rahimmalek,M., Sayed Tabatabaei,B.E., Arzani,A., and Etemadi,N. 2009. Assessment of genetic diversity among and within *Achillea* species using amplified fragment length polymorphism (AFLP). *Biochemical Systematics and Ecology* **37**: 354-361.
- Rahman,M.H. and Rajora,O.P. 2001. Microsatellite DNA somaclonal variation in micropropagated trembling aspen (*Populus tremuloides*). *Plant Cell Reports* **20**: 531-536.
- Rajora,O.P. and Rahman,M.H. 2003. Microsatellite DNA and RAPD fingerprinting, identification and genetic relationships of hybrid poplar (*Populus x canadensis*) cultivars. *Theoretical and Applied Genetics* **106**: 470-477.
- Ramos,A.M., Crooijmans,R.P.M.A., Affara,N.A., Amaral,A.J., Archibald,A.L., Beever,J.E., Bendixen,C., Churcher,C., Clark,R., Dehais,P., Hansen,M.S., Hedegaard,J., Hu,Z.L., Kerstens,H.H., Law,A.S., Megens,H.J., Milan,D., Nonneman,D.J., Rohrer,G.A., Rothschild,M.F., Smith,T.P.L., Schnabel,R.D., Van Tassell,C.P., Taylor,J.F., Wiedmann,R.T., Schook,L.B., and Groenen,M.A.M. 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS ONE* **4**.
- Ramsay,L., Macaulay,M., Degli Ivanisovich,S., MacLean,K., Cardle,L., Fuller,J., Edwards,K.J., Turesson,S., Morgante,M., Massari,A., Maestri,E., Marmiroli,N., Sjakste,T., Ganai,M., Powell,W., and Waugh,R. 2000. A simple sequence repeat-based linkage map of Barley. *Genetics* **156**: 1997-2005.
- Reboud,X. and Zeyl,C. 1994. Organelle inheritance in plants. *Heredity* **72**: 132-140.
- Reed,K., Poulin,M.L., Yan,L., and Parissenti,A.M. 2010. Comparison of bisulfite sequencing PCR with pyrosequencing for measuring differences in DNA methylation. *Anal. Biochem.* **397**: 96-106.
- Ricca,M. and Shaw,A.J. Allopolyploidy and homoploid hybridization in the *Sphagnum subsecundum* complex (Sphagnaceae: Bryophyta). *Biological Journal of the Linnean Society* 99[1], 135-151. 2010.
- Richards,E.J. 2008. Population epigenetics. *CURR. OPIN. GENET. DEV.* **18**: 221-226.
- Richards,E.J. 2009. Quantitative epigenetics: DNA sequence variation need not apply. *Genes and Development* **23**: 1601-1605.
- Rinaldi,C., Kohler,A., Frey,P., Duchaussoy,F., Ningre,N., Couloux,A., Wincker,P., Le Thiec,D., Fluch,S., Martin,F., and Duplessis,S. 2007. Transcript profiling of poplar leaves upon infection with compatible and incompatible strains of the foliar rust *Melampsora larici-populina*. *Plant Physiology* **144**: 347-366.
- Ronaghi,M., Karamohamed,S., Pettersson,B., Uhle  n,M., and Nyre  n,P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**: 84-89.
- Rostoks,N., Mudie,S., Cardle,L., Russell,J., Ramsay,L., Booth,A., Svensson,J.T., Wanamaker,S.I., Walia,H., Rodriguez,E.M., Hedley,P.E., Liu,H., Morris,J., Close,T.J., Marshall,D.F., and Waugh,R. 2005. Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Molecular Genetics and Genomics* **274**: 515-527.
- Roy,A.M., Carroll,M.L., Kass,D.H., Nguyen,S.V., Salem,A.H., Batzer,M.A., and Deininger,P.L. 1999. Recently integrated human Alu repeats: Finding needles in the haystack. *Genetica* **107**: 149-161.
- Ruas,C.F., Weiss-Schneeweiss,H., Stuessy,T.F., Samuel,M.R., Pedrosa-Harand,A., Tremetsberger,K., Ruas,P.M., Schl  ter,P.M., Ortiz Herrera,M.A., K  nig,C., and Matzenbacher,N.I. 2008. Characterization, genomic organization and chromosomal distribution of Ty1-copia retrotransposons in species of *Hypochaeris* (Asteraceae). *Gene* **412**: 39-49.

- Ruf, S., Karcher, D., and Bock, R. 2007. Determining the transgene containment level provided by chloroplast transformation. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 6998-7002.
- Rungis, D., Barub, Y., Zhang, J., Ralph, S., Ritland, C.E., Ellis, B.E., Douglas, C., Bohlmann, J., and Ritland, K. 2004. Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics* **109**: 1283-1294.
- Sabatti, M., D'ovidio, R., Tanzarella, O.A., and Mugnozza, G.E.S. 2001. Assessment of geographic variation by RAPD markers among Italian open-pollinated progenies of *Populus alba* L. *Genetic Resources and Crop Evolution* **48**: 423-428.
- Saiki, R.K., Scharf, S., and Faloona, F. 1985. Enzymatic amplification of globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350-1354.
- Salem, K.F.M., Varshney, R.K., Rader, M.S., and Barner, A. 2010. EST-SSR based estimates on functional genetic variation in a barley (*Hordeum vulgare* L.) collection from Egypt. *Genetic Resources and Crop Evolution* **57**, 4: 515-521.
- Samuel, R. 1999. Identification of hybrids between *Quercus petraea* and *Q. robur* (Fagaceae): Results obtained with RAPD markers confirm allozyme studies based on the Got-2 locus. *Plant Systematics and Evolution* **217**: 137-146.
- Sanger, F. and Coulson, A.R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* **94**: 441-448.
- Sanger, F., Nicklen, S., and Coulson, A.R. 1992. DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology (Reading, Mass.)* **24**: 104-108.
- Sato, Y. and Nishio, T. 2003. Mutation detection in rice waxy mutants by PCR-RF-SSCP. *Theoretical and Applied Genetics* **107**: 560-567.
- Schlotterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**: 365-371.
- Schmalenbach, I. and Pillen, K. 2009. Detection and verification of malting quality QTLs using wild barley introgression lines. *Theoretical and Applied Genetics* **118**: 1411-1427.
- Schondelmaier, J., Schmidt, T., Heslop-Harrison, J.S., and Jung, C. 1997. Genetic and chromosomal localization of the 5S rDNA locus in sugar beet (*Beta vulgaris* L.). *Genome* **40**: 171-175.
- Scotti, I., Gugerli, F., Pastorelli, R., Sebastiani, F., and Vendramin, G.G. 2008a. Maternally and paternally inherited molecular markers elucidate population patterns and inferred dispersal processes on a small scale within a subalpine stand of Norway spruce (*Picea abies* [L.] Karst.). *Forest Ecology and Management* **255**: 3806-3812.
- Scotti, I., Magni, F., Paglia, G.P., and Morgante, M. 2002. Trinucleotide microsatellites in Norway spruce (*Picea abies*): Their features and the development of molecular markers. *Theoretical and Applied Genetics* **106**: 40-50.
- Scotti, I., Paglia, G., Magni, F., and Morgante, M. 2006a. Population genetics of Norway spruce (*Picea abies* Karst.) at regional scale: Sensitivity of different microsatellite motif classes in detecting differentiation. *Annals of Forest Science* **63**: 485-491.
- Scotti, I., Troggio, M., Soranzo, N., Vendramin, G.G., and Bucci, G. 1998. A new set of PCR-based, locus-specific markers for *Picea abies* (L.) Karst. *Mol. Ecol.* **7**: 789-792.
- Shahmuradov, I.A., Akbarova, Y.Y., Solovyev, V.V., and Aliyev, J.A. 2003. Abundance of plastid DNA insertions in nuclear genomes of rice and *Arabidopsis*. *Plant. Mol. Biol.* **52**: 923-934.
- Sharma, R.K., Bhardwaj, P., Negi, R., Mohapatra, T., and Ahuja, P.S. 2009. Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). *Bmc Plant Biology* **9**.
- Shendure, J., Mitra, R.D., Varma, C., and Church, G.M. 2004. Advanced sequencing technologies: Methods and goals. *Nature Reviews Genetics* **5**: 335-344.
- Sherman, J.D. and Talbert, L.E. 2002. Vernalization-induced changes of the DNA methylation pattern in winter wheat. *Genome* **45**: 253-260.
- Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. 2001. DbSNP: The NCBI database of genetic variation. *NUCLEIC ACIDS RES.* **29**: 308-311.
- Shneer, V.S. 2009. DNA barcoding of animal and plant species as an approach for their molecular identification and describing of diversity. *Zhurnal obshcheĖ biologii* **70**: 296-315.

- Sánchez, C., Martínez, M.T., Vidal, N., San Jose, M.C., Valladares, S., and Vieitez, A.M. 2008. Preservation of *Quercus robur* germplasm by cryostorage of embryogenic cultures derived from mature trees and RAPD analysis of genetic stability. *Cryo-Letters* **29**: 493-504.
- Sirivatanauksorn, V., Sirivatanauksorn, Y., McKie, A.B., and Lemoine, N.R. 2002. Use of DNA fingerprinting to detect genetic rearrangements in human cancer. *Methods in molecular medicine* **68**: d-114.
- Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* **236**: 787-792.
- Smith, D.N. and Devey, M.E. 1994. Occurrence and inheritance of microsatellites in *Pinus radiata*. *Genome* **37**: 977-983.
- Soltis, D.E., Gitzendanner, M.A., Strenge, D.D., and Soltis, P.S. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution* **206**: 353-373.
- Soltis, D.E., Ma, H., Frohlich, M.W., Soltis, P.S., Albert, V.A., Oppenheimer, D.G., Altman, N.S., DePamphilis, C., and Leebens-Mack, J. 2007. The floral genome: an evolutionary history of gene duplication and shifting patterns of gene expression. *Trends Plant Sci.* **12**: 358-367.
- Soltis, D.E., Soltis, P.S., Bennett, M.D., and Leitch, I.J. 2003. Evolution of genome size in the angiosperms. *Am. J. Bot.* **90**: 1596-1603.
- Soltis, E.D. and Soltis, P.S. 2000. Contributions of plant molecular systematics to studies of molecular evolution. *Plant. Mol. Biol.* **42**: 45-75.
- Song, H., Buhay, J.E., Whiting, M.F., and Crandall, K.A. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 13486-13491.
- Soranzo, N., Provan, J., and Powell, W. 1998. Characterization of microsatellite loci in *Pinus sylvestris* L. *Mol. Ecol.* **7**: 1260-1261.
- Sperisen, C., Büchler, U., Gugerli, F., Matyas, G., Geburek, T., and Vendramin, G.G. 2001a. Tandem repeats in plant mitochondrial genomes: Application to the analysis of population differentiation in the conifer Norway spruce. *Mol. Ecol.* **10**: 257-263.
- Sperisen, C., Büchler, U., Gugerli, F., Matyas, G., Geburek, T., and Vendramin, G.G. 2001b. Tandem repeats in plant mitochondrial genomes: Application to the analysis of population differentiation in the conifer Norway spruce. *Mol. Ecol.* **10**: 257-263.
- Stallings, R.L., Torney, D.C., Hildebrand, C.E., Longmire, J.L., Deaven, L.L., Jett, J.H., Doggett, N.A., and Moyzis, R.K. 1990. Physical mapping of human chromosomes by repetitive sequence fingerprinting. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 6218-6222.
- Steinkellner, H., Fluch, S., Turetschek, E., Lexer, C., Streiff, R., Kremer, A., Burg, K., and Glossl, J. 1997a. Identification and characterization of (GA/CT)(n)-microsatellite loci from *Quercus petraea*. *Plant Molecular Biology* **33**: 1093-1096.
- Steinkellner, H., Lexer, C., Turetschek, E., and Glossl, J. 1997b. Conservation of (GA)(n) microsatellite loci between *Quercus* species. *Molecular Ecology* **6**: 1189-1194.
- Streiff, R., Labbe, T., Bacilieri, R., Steinkellner, H., Glossl, J., and Kremer, A. 1998a. Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology* **7**: 317-328.
- Sugiura, M. 1992. The chloroplast genome. *Plant. Mol. Biol.* **19**: 149-168.
- Sugiura, M. 2003. History of chloroplast genomics. *Photosynthesis Research* **76**: 371-377.
- Sunnucks, P., Wilson, A.C.C., Beheregaray, L.B., Zenger, K., French, J., and Taylor, A.C. 2000. SSCP is not so difficult: The application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Mol. Ecol.* **9**: 1699-1710.
- Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant. Mol. Biol.* **17**: 1105-1109.
- Tamaki, K. and Jeffreys, A.J. 2005. Human tandem repeat sequences in forensic DNA typing. *Legal Medicine* **7**: 244-250.
- Tan, M. 2010. Analysis of DNA methylation of maize in response to osmotic and salt stress based on methylation-sensitive amplified polymorphism. *Plant Physiology and Biochemistry* **48**: 21-26.
- Tangphatsornruang, S., Somta, P., Uthapaisanwong, P., Chanprasert, J., Sangsakru, D., Seehalak, W., Sommanas, W., Tragoonrung, S., and Srinives, P. 2009. Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). *BMC Plant Biology* **9**: 137

- Tanksley, S.D., Ganai, M.W., Prince, J.P., De Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messeguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Roder, M.S., Wing, R.A., Wu, W., and Young, N.D. 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**: 1141-1160.
- Teixeira, A.P.M., Barreto, F.A., and Camargo, L.E.A. 2008. An AFLP marker linked to the Pm-1 gene that confers resistance to *Podosphaera xanthii* race 1 in *Cucumis melo*. *Genetics and Molecular Biology* **31**: 547-550.
- Terrab, A., Hampe, A., Lepais, O., Talavera, S., Vela, E., and Stuessy, T.F. 2008. Phylogeography of North African Atlas cedar (*Cedrus atlantica*, Pinaceae): Combined molecular and fossil data reveal a complex Quaternary history. *American Journal of Botany* **95**: 1262-1269.
- Terrab, A., Ortiz, M.A., Talavera, M., Ariza, M.J., Moriana, M., Garcia-Castano, J.L., Tremetsberger, K., Stuessy, T.F., Baeza, C.M., Urtubey, E., Ruas, C., Casimiro-Soriguer, R., Balao, F., Gibbs, P.E., and Talavera, S. 2009. AFLP and breeding system studies indicate vicariance origin for scattered populations and enigmatic low fecundity in the Moroccan endemic *Hypochaeris angustifolia* (Asteraceae), sister taxon to all of the South American *Hypochaeris* species. *Molecular Phylogenetics and Evolution* **53**: 13-22.
- Tollefsrud, M.M., Kissling, R., Gugerli, F., Johnsen, O., Skroppa, T., Cheddadi, R., Van Der Knaap, W.O., Latalowa, M., Terhürne-Berson, R., Litt, T., Geburek, T., Brochmann, C., and Sperisen, C. 2008. Genetic consequences of glacial survival and postglacial colonization in Norway spruce: Combined analysis of mitochondrial DNA and fossil pollen. *Mol. Ecol.* **17**: 4134-4150.
- Torress, A.M., Weeden, N.F., and Martin, A. 1993. Linkage among isozyme, RFLP, and RAPD markers. *Plant Physiol* **101**: 394-452.
- Tost, J. 2009. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Methods in molecular biology* (Clifton, N. J.) **507**: 3-20.
- Travis, S.E., Proffitt, C.E., and Ritland, K. 2004. Population structure and inbreeding vary with successional stage in created *Spartina alterniflora* marshes. *Ecological Applications* **14**: 1189-1202.
- Travis, S.E., Ritland, K., Whitham, T.G., and Keim, P. 1998. A genetic linkage map of pinyon pine (*Pinus edulis*) based on amplified fragment length polymorphisms. *Theoretical and Applied Genetics* **97**: 871-880.
- Trick, M., Long, Y., Meng, J., and Bancroft, I. 2009. Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnol. J.* **7**: 334-346.
- Troupin, D., Nathan, R., and Vendramin, G.G. 2006. Analysis of spatial genetic structure in an expanding *Pinus halepensis* population reveals development of fine-scale genetic clustering over time. *Mol. Ecol.* **15**: 3617-3630.
- Tullu, A., Tar'an, B., Warkentin, T., and Vandenberg, A. 2008. Construction of an intraspecific linkage map and QTL analysis for earliness and plant height in lentil. *Crop Science* **48**: 2254-2264.
- Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, M., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroove, S., D+òjardin, A., DePamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J.C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van De Peer, Y., and Rokhsar, D. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596-1604.

- Utomo, H.S. and Linscombe, S.D. 2009. Current patents and future development underlying marker-assisted breeding in major grain crops. *Recent Patents on DNA and Gene Sequences* **3**: 53-62.
- Van Der Knaap, E. and Tanksley, S.D. 2003. The making of a bell pepper-shaped tomato fruit: Identification of loci controlling fruit morphology in Yellow Stuffer tomato. *Theoretical and Applied Genetics* **107**: 139-147.
- Van, K., Hwang, E.Y., Kim, M.Y., Kim, Y.H., Cho, Y.I., Cregan, P.B., and Lee, S.H. 2004. Discovery of single nucleotide polymorphisms in soybean using primers designed from ESTs. *Euphytica* **139**: 147-157.
- Varshney, R.K., Penmetsa, R.V., Dutta, S., Kulwal, P.L., Saxena, R.K., Datta, S., Sharma, T.R., Rosen, B., Carrasquilla-Garcia, N., Farmer, A.D., Dubey, A., Saxena, K.B., Gao, J., Fakrudin, B., Singh, M.N., Singh, B.P., Wanjari, K.B., Yuan, M., Srivastava, R.K., Kilian, A., Upadhyaya, H.D., Mallikarjuna, N., Town, C.D., Bruening, G.E., He, G., May, G.D., McCombie, R., Jackson, S.A., Singh, N.K., and Cook, D.R. Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). *Molecular Breeding*. 2009. DOI 10.1007/s11032-009-9327-2
- Varshney, R.K., Sigmund, R., Borner, A., Korzun, V., Stein, N., Sorrells, M.E., Langridge, P., and Graner, A. 2005. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Science* **168**: 195-202.
- Varshney, R.K., Thiel, T., Stein, N., Langridge, P., and Graner, A. 2002. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cellular & Molecular Biology Letters* **7**: 537-546.
- Vaughn, C.P. and Elenitoba-Johnson, K.S.J. 2004. High-resolution melting analysis for detection of internal tandem duplications. *Journal of Molecular Diagnostics* **6**: 211-216.
- Vaëli, U., Einarsson, A., Waits, L., and Ellegren, H. 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Mol. Ecol.* **17**: 3808-3817.
- Vencovsky, R. and Crossa, J. 2003. Measurements of Representativeness Used in Genetic Resources Conservation and Plant Breeding. *Crop Science* **43**: 1912-1921.
- Vendramin, G.G., Anzidei, M., Madaghiele, A., Sperisen, C., and Bucci, G. 2000a. Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). *Genome* **43**: 68-78.
- Vendramin, G.G., Anzidei, M., Madaghiele, A., Sperisen, C., and Bucci, G. 2000b. Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). *Genome* **43**: 68-78.
- Vendramin, G.G. and Ziegenhagen, B. 1997. Characterisation and inheritance of polymorphic plastid microsatellites in *Abies*. *Genome* **40**: 857-864.
- Vezzulli, S., Micheletti, D., Riaz, S., Pindo, M., Viola, R., This, P., Walker, M.A., Troggio, M., and Velasco, R. 2008. A SNP transferability survey within the genus *Vitis*. *BMC Plant Biology* **8**: 128 doi:10.1186/1471-2229-8-128
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: A new technique for DNA fingerprinting. *NUCLEIC ACIDS RES.* **23**: 4407-4414.
- Wall, P.K., Leebens-Mack, J., Chanderbali, A.S., Barakat, A., Wolcott, E., Liang, H., Landherr, L., Tomsho, L.P., Hu, Y., Carlson, J.E., Ma, H., Schuster, S.C., Soltis, D.E., Soltis, P.S., Altman, N., and de Pamphilis, C.W. 2009. Comparison of next generation sequencing technologies for transcriptome characterization. *Bmc Genomics* **10**.
- Wang, L., Hao, L., Li, X., Hu, S., Ge, S., and Yu, J. 2009a. SNP deserts of Asian cultivated rice: Genomic regions under domestication. *Journal of Evolutionary Biology* **22**: 751-761.
- Wang, X., Elling, A.A., Li, X., Li, N., Peng, Z., He, G., Sun, H., Qi, Y., Liu, X.S., and Denga, X.W. 2009b. Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in Maize. *Plant Cell* **21**: 1053-1069.
- Waugh, R., McLean, K., Flavell, A.J., Pearce, S.R., Kumar, A., Thomas, B.B.T., and Powell, W. 1997a. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Molecular and General Genetics* **253**: 687-694.
- Waugh, R., McLean, K., Flavell, A.J., Pearce, S.R., Kumar, A., Thomas, B.B.T., and Powell, W. 1997b. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Molecular and General Genetics* **253**: 687-694.

- Weber, J.L. and May, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**: 388-396.
- Weir, B.S., Anderson, A.D., and Hepler, A.B. 2006. Genetic relatedness analysis: Modern data and new challenges. *Nature Reviews Genetics* **7**: 771-780.
- Welsh, J. and McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *NUCLEIC ACIDS RES.* **18**: 7213-7218.
- Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Klein, A., and Kilian, A. 2004. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 9915-9920.
- Wilhelm, E., Hristoforoglu, K., Fluch, S., and Burg, K. 2005. Detection of microsatellite instability during somatic embryogenesis of oak (*Quercus robur* L.). *Plant Cell Reports* **23**: 790-795.
- Williams, C.G., Goodman, M.M., and Stuber, C.W. 1995. Comparative recombination distances among *Zea mays* L. inbreds, wide crosses and interspecific hybrids. *Genetics* **141**: 1573-1581.
- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. 1990a. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *NUCLEIC ACIDS RES.* **18**: 6531-6535.
- Wilson, A.S.G., Van Der Kamp, B.J., and Ritland, C. 2005. Spatial genetic and clonal structure in *Maianthemum dilatatum* as defined by AFLP markers. *Canadian Journal of Botany* **83**: 1126-1132.
- Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteegen, J.G., and Pryor, R.J. 2003. High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical Chemistry* **49**: 853-860.
- Wolf, P.G., Doche, B., Gielly, L., and Taberlet, P. 2004. Genetic structure of *Rhododendron ferrugineum* at a wide range of spatial scales. *J. Hered.* **95**: 301-308.
- Wolfe, K.H., Li, W.H., and Sharp, P.M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America* **84**: 9054-9058.
- Wu, K.S., Jones, R., Danneberger, L., and Scolnik, P.A. 1994. Detection of microsatellite polymorphisms without cloning. *NUCLEIC ACIDS RES.* **22**: 3257-3258.
- Xing, Y., Frei, U., Schejbel, B., Asp, T., and Lübberstedt, T. 2007. Nucleotide diversity and linkage disequilibrium in 11 expressed resistance candidate genes in *Lolium perenne*. *BMC Plant Biology* **7**: 43 doi: 10.1186/1471-2229-7-43.
- Xu, X., Peng, M., Fang, Z., and Xu, X. 2000. The direction of microsatellite mutations is dependent upon allele length. *Nature Genetics* **24**: 396-399.
- Yan, J., Yang, X., Shah, T., Sanchez-Villeda, H., Li, J., Warburton, M., Zhou, Y., Crouch, J.H., and Xu, Y. 2010. High-throughput SNP genotyping with the GoldenGate assay in maize. *Molecular Breeding* **25**: 441-451.
- Yang, R.W., Zhou, Y.H., Zhang, Y., Zheng, Y.L., and Ding, C.B. 2006. The genetic diversity among *Leymus* species based on random amplified microsatellite polymorphism (RAMP). *Genetic Resources and Crop Evolution* **53**: 139-144.
- Yi, G.B., Lee, J.M., Lee, S., Choi, D., and Kim, B.D. 2006. Exploitation of pepper EST-SSRs and an SSR-based linkage map. *Theoretical and Applied Genetics* **114**: 113-130.
- Yu, G.X. and Wise, R.P. 2000. An anchored AFLP- and retrotransposon-based map of diploid *Avena*. *Genome* **43**: 736-749.
- Zane, L., Bargelloni, L., and Patarnello, T. 2002. Strategies for microsatellite isolation: A review. *Mol. Ecol.* **11**: 1-16.
- Zanetto, A., Kremer, A., Müller-Starck, G., and Hattemer, H.H. 1996b. Inheritance of isozymes in pedunculate oak (*Quercus robur* L.). *Journal of Heredity* **87**: 364-370.
- Zeller, G., Clark, R.M., Schneeberger, K., Bohlen, A., Weigel, D., and Reñtsch, G. 2008. Detecting polymorphic regions in *Arabidopsis thaliana* with resequencing microarrays. *Genome Research* **18**: 918-929.
- Zhang, D., Yang, X., Zhang, Z., and Li, B. 2010. Expression and nucleotide diversity of the poplar COBL gene. *Tree Genetics and Genomes* **6** (2): 331-344.
- Zhang, W., Sun, X., Yuan, H., Araki, H., Wang, J., and Tian, D. 2008. The pattern of insertion/deletion polymorphism in *Arabidopsis thaliana*. *Molecular Genetics and Genomics* **280**: 351-361.

- Zhou,W.C., Kolb,F.L., Bai,G.H., Domier,L.L., Boze,L.K., and Smith,N.J. 2003. Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant Breeding* **122**: 40-46.
- Zietkiewicz,E., Rafalski,A., and Labuda,D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20**: 176-183.
- Zilberman,D. and Henikoff,S. 2007. Genome-wide analysis of DNA methylation patterns. *Development* **134**: 3959-3965.
- Zombori,Z., Jancso,M., Zvara,A., Pauk,J., and Györgyey,J. 2008. Investigation of the effect of drought stress on the rice transcriptome. *Acta Biol. Szegediensis* **52**: 143-145.

7 Curriculum Vitae

Mag. Silvia Fluch, MSc

Geboren: 21.11.1966, Graz, Österreich



Beruflicher Werdegang:

- ab 1996: Anstellung als wissenschaftliche Mitarbeiterin am AIT-Austrian Institute of Technology GmbH (früher Austrian Research Centers GmbH - ARC, Dept. Biogenetics/Natural Resources)
- ab 1998: Projektleiterin in nationalen und internationalen Forschungsprojekten mit den Aufgabengebieten: Projektakquisition sowie finanzielles und wissenschaftliches Projektmanagement

Forschungsaktivitäten:

- Funktionale Pflanzenforschung an Eiche, Fichte, Banane, Süßkartoffel in kooperativen nationalen und internationalen Projekten (EU, CGIAR)
- Etablierung einer Microarray-Plattform
- Markerentwicklung und Anwendung für Biodiversitätsstudien bei Forstgehölzen und Kulturpflanzen
- seit 2003: Leitung der Arbeitsgruppe PICME und Aufbau des internationalen Ressourcenzentrums für Pflanzengene (www.picme.at);
- seit 2009: Stellvertretende Abteilungsleiterin Bioresources

Ausbildung

- 2009–2010 Dissertation am AIT/Uni Wien
- 1992–1996 Arbeiten am Dissertationsthema: 'Identification of polymorphic regions in the oak genome (*Quercus* spp)'; Die Arbeiten wurden im Rahmen von 2 EU Projekten: BIOTEC3 und BIOTEC4 (Development of molecular tools for screening diversity) an den ARC in Kooperation mit der BOKU, ZAG durchgeführt.
- 1987–1991 2. Studienabschnitt im Bereich Biologie Genetik an der Universität Wien; Diplomarbeit am Institute für physikalische Chemie (Doz. Schmetterer): Investigation of the glucose transport in fructose resistant mutants of *Synechocystis* PCC 6803
- 1985–1987 1. Studienabschnitt Biologie an der Karl-Franzens-Universität Graz
- 1985 Matura am BG Stift Rein (Neusprachlich)

Zusatzqualifikationen:

- 2004–2006 Masterlehrgang Organisationsentwicklung, Uni Klagenfurt (iff); (MSc)

Weitere Tätigkeiten:

2004–2006: Chair der COST Action E28 GENOSILVA; European forest genomics Network (<http://www.genosilva.org>)

Awards:

2002: Gewinnerin des Austrian Business-plan Wettbewerbs *i2B* (Ideas to Business)

Publikationsliste:

- Steinkellner H., Fluch S., Turetschek E., Lexer C., Streiff R., Kremer A., Burg K., Glössl J.; Identification and characterisation of (GA/CT)_n-microsatellite loci from *Q. robur* and *Q. petraea*; (1997) Plant Mol. Biol.:33:1093-1096;
- Barreneche T, Bodenes C, Lexer C, Trontin J F, Fluch S, Streiff R, Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glössl J, Kremer A. (1998) A genetic linkage map of *Quercus robur* L (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. - Theor Appl Genet 97: 1090-1103
- Storme V., A. Vanden Broeck, B. Ivens, D. Halfmaerten, J. Van Slycken, S. Castiglione, F. Grassi, T. Fossati, J. E. Cottrell, H. E. Tabbener, F. Lefèvre, C. Saintagne, S. Fluch, V. Krystufek, K. Burg, S. Bordács, A. Borovics, K. Gebhardt, B. Vornam, A. Pohl, N. Alba, D. Agúndez, C. Maestro, E. Notivol, J. Bovenschen, B. C. van Dam, J. van der Schoot, B. Vosman, W. Boerjan and M. J. M. Smulders; Ex-situ conservation of Black poplar in Europe: genetic diversity in nine gene bank collections and their value for nature development TAG Theoretical and Applied Genetics ; Issue: Volume 108, Number 6 Date: April 2004 Pages: 969 - 981
- Wilhelm E., K Hristoforoglu, S. Fluch, K. Burg (2004) Detection of microsatellite instability during somatic embryogenesis of oak (*Quercus robur* L.) Plant Cell Rep. Oct.19.
- Bajla, I. Holländer, S. Fluch, K. Burg, M. Kollar (2005) An Alternative Method for Electrophoretic Gel Image Analysis in the GelMaster Software; Computer Methods and Programs in Biomedicine; Volume 77, Issue 3, Pages 209-231
- Cottrell J.E., V. Krystufek, H.E. Tabbener, A.D. Milner, T. Connolly, L. Sing, S. Fluch, K. Burg, F. Lefevre, P. Achard, S. Bordacs, K. Gebhardt, B. Vornam, M.J.M. Smulders, A.H. Vanden Broeck, J. Van Slycken, V. Storme, W. Boerjan, S. Castiglione, T. Fossati, N. Alba, D. Agundez, C. Maestro, E. Notivol, J. Bovenschen, V. van Dam, (2005). Postglacial migration of *Populus nigra* L.: lessons learnt from chloroplast DNA; Forest Ecology and Management 206:
- Maghuly F., Nittinger F., Pinsker W., Praznik W., Fluch S.; Differentiation among Austrian Populations of Norway Spruce [*Picea abies* (L.) Karst.] assayed by mitochondrial DNA markers; Tree Genetics & Genomes (2007) 3: 199-206
- Soler M, Serra O, Molinas M, Huguet G, Fluch S, Figueras M.; A Genomic Approach to Suberin Biosynthesis and Cork Differentiation. Plant Physiol. 2007 May; 144(1):419-31;
- Rinaldi C., A. Kohler, P. Frey, F. Duchaussoy, N. Ningre, A. Couloux, P. Wincker, D. Le Thiec, S. Fluch, F. Martin and S. Duplessis; (2007) Transcript Profiling of Poplar Leaves upon Infection with Compatible and Incompatible Strains of the Foliar Rust *Melampsora larici-populina*, Plant Physiology, 2007 May;144(1):347-66
- Fluch S., Olmo, C. Ch., Tauber, S., Stierschneider, M., Kopecky, D., Reichenauer, T., Matušíková, I. (2008) Transcriptomic changes in wind-exposed poplar leaves are dependent on developmental stage. Planta 228(5):757-64, DOI: 10.1007/s00425-008-0777-2,
- Maghuly F., Nittinger F., Burg K., Pinsker W., Praznik W., Fluch S.; Development of mitochondrial markers for population genetics of Norway spruce [*Picea abies* (L.) Karst.] Silvae Genetica, 57(1) p. 41-44
- Smulders M.J.M., J.E. Cottrell, F. Lefèvre, J. van der Schoot, P. Arens, B. Vosman, H.E. Tabbener, F. Grassi, T. Fossati, S. Castiglione, V. Krystufek, S. Fluch, K. Burg, B. Vornam, A. Pohl, K. Gebhardt, N. Alba, D. Agúndez, C. Maestro, E. Notivol, R. Volosyanchuk, M. Pospíšková, S. Bordács, J. Bovenschen, B.C. van Dam, H.P. Koelewijn, D. Halfmaerten, B. Ivens, J. van Slycken, A. Vanden Broeck, V. Storme, W. Boerjan Structure of the genetic diversity in Black poplar (*Populus nigra* L.) populations

- across European river systems: consequences for conservation and restoration. *Forest Ecology and Management*, Vol. 255, No. 4-5. (April 2008), pp. 1388-1399
- Berenyi, M., Mauleon, R.P., Kopecky, D., Wandl, S., Friedl, R., Fluch, S., Boonruangrod, R., Muge, E. and K. Burg* Isolation of plant gene space related sequence elements by High C+G patch (HCGP) filtration; model study on rice. *Plant Mol Biol Rep* (2009) 27:79–85
- Boonruangrod R, Desai D, Fluch S, Berenyi M and K. Burg* Identification of cytoplasmic ancestor gene-pools of *Musa acuminata* Colla and *Musa balbisiana* Colla and their hybrids by chloroplast and mitochondrial haplotyping. *TAG*, 118(1):43-55
- Boonruangrod R., S. Fluch, K. Burg*; Elucidation of origin of the present day hybrid banana cultivars using the 5'ETS rDNA sequence information; *Molecular Breeding*, (2009) Volume 24, Number 1 p: 77-91