HAL
open science

Population genetic structure in common ash : a focus on southeastern european genetic resources

Myriam Heuertz

To cite this version:

Myriam Heuertz. Population genetic structure in common ash : a focus on southeastern european genetic resources. Life Sciences [q-bio]. Université Libre de Bruxelles [Bruxelles], 2003. English. NNT : . tel-02831564

HAL Id: tel-02831564

https://hal.inrae.fr/tel-02831564v1

Submitted on 7 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dépôt Institutionnel de l'Université libre de Bruxelles /
Université libre de Bruxelles Institutional Repository
Thèse de doctorat/ PhD Thesis

Citation APA:

Heuertz, M. (2003). *Population genetic structure in common ash : a focus on southeastern european genetic resources* (Unpublished doctoral dissertation). Université libre de Bruxelles, Faculté des sciences, Bruxelles.

Disponible à / Available at permalink : https://dipot.ulb.ac.be/dspace/bitstream/2013/211332/1/b0ae6555-e30d-49c6-aebf-ddc9be80f92c.txt

(English version below)

Cette thèse de doctorat a été numérisée par l'Université libre de Bruxelles. L'auteur qui s'opposerait à sa mise en ligne dans DI-fusion est invité à prendre contact avec l'Université (di-fusion@ulb.ac.be).

Dans le cas où une version électronique native de la thèse existe, l'Université ne peut garantir que la présente version numérisée soit identique à la version électronique native, ni qu'elle soit la version officielle définitive de la thèse.

DI-fusion, le Dépôt Institutionnel de l'Université libre de Bruxelles, recueille la production scientifique de l'Université, mise à disposition en libre accès autant que possible. Les œuvres accessibles dans DI-fusion sont protégées par la législation belge relative aux droits d'auteur et aux droits voisins. Toute personne peut, sans avoir à demander l'autorisation de l'auteur ou de l'ayant-droit, à des fins d'usage privé ou à des fins d'illustration de l'enseignement ou de recherche scientifique, dans la mesure justifiée par le but non lucratif poursuivi, lire, télécharger ou reproduire sur papier ou sur tout autre support, les articles ou des fragments d'autres œuvres, disponibles dans DI-fusion, pour autant que :

- Le nom des auteurs, le titre et la référence bibliographique complète soient cités;
- L'identifiant unique attribué aux métadonnées dans DI-fusion (permalink) soit indiqué;
- Le contenu ne soit pas modifié.

L'œuvre ne peut être stockée dans une autre base de données dans le but d'y donner accès ; l'identifiant unique (permalink) indiqué ci-dessus doit toujours être utilisé pour donner accès à l'œuvre. Toute autre utilisation non mentionnée ci-dessus nécessite l'autorisation de l'auteur de l'œuvre ou de l'ayant droit.

---------- **English Version** ----------

This Ph.D. thesis has been digitized by Université libre de Bruxelles. The author who would disagree on its online availability in DI-fusion is invited to contact the University (di-fusion@ulb.ac.be).

If a native electronic version of the thesis exists, the University can guarantee neither that the present digitized version is identical to the native electronic version, nor that it is the definitive official version of the thesis.

DI-fusion is the Institutional Repository of Université libre de Bruxelles; it collects the research output of the University, available on open access as much as possible. The works included in DI-fusion are protected by the Belgian legislation relating to authors' rights and neighbouring rights. Any user may, without prior permission from the authors or copyright owners, for private usage or for educational or scientific research purposes, to the extent justified by the non-profit activity, read, download or reproduce on paper or on any other media, the articles or fragments of other works, available in DI-fusion, provided:

- The authors, title and full bibliographic details are credited in any copy;
- The unique identifier (permalink) for the original metadata page in DI-fusion is indicated;
- The content is not changed in any way.

It is not permitted to store the work in another database in order to provide access to it; the unique identifier (permalink) indicated above must always be used to provide access to the work. Any other use not mentioned above requires the authors' or copyright owners' permission.

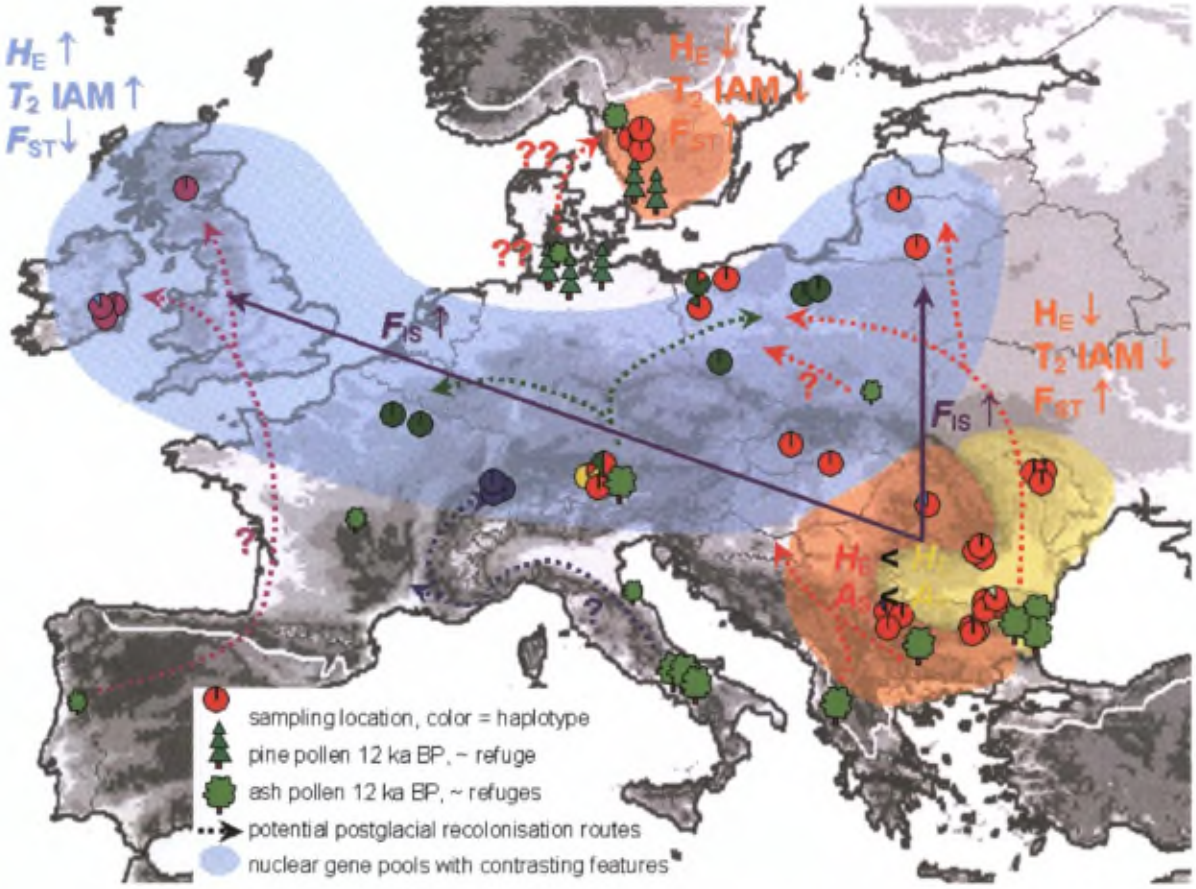
D 03166

BRE DE BRUXELLES
: DES SCIENCES
E BIOLOGIE VEGETALE
LABORATOIRE DE GENETIQUE ET ECOLOGIE VEGETALES

en collaboration avec le

CENTRE DE RECHERCHE PUBLIC-GABRIEL LIPPMANN
CELLULE DE RECHERCHE EN ENVIRONNEMENT ET BIOTECHNOLOGIES

POPULATION GENETIC STRUCTURE IN COMMON ASH: A FOCUS ON SOUTHEASTERN EUROPEAN GENETIC RESOURCES



Thèse présentée en vue de l'obtention du titre de
Docteur en Sciences,
orientation Biologie végétale

Universite Libre de Bruxelles
003190407

r

HEUERTZ

DIRECTEUR DE THESE: PROF. XAVIER VEKEMANS
ANNEE ACADEMIQUE 2002 / 2003

UNIVERSITE LIBRE DE BRUXELLES
FACULTE DES SCIENCES
DEPARTEMENT DE BIOLOGIE VEGETALE
LABORATOIRE DE GENETIQUE ET ECOLOGIE VEGETALES

en collaboration avec le

CENTRE DE RECHERCHE PUBLIC-GABRIEL LIPPMANN
CELLULE DE RECHERCHE EN ENVIRONNEMENT ET BIOTECHNOLOGIES

POPULATION GENETIC STRUCTURE IN COMMON ASH: A FOCUS ON SOUTHEASTERN EUROPEAN GENETIC RESOURCES

Thèse présentée en vue de l'obtention du titre de
Docteur en Sciences,
orientation Biologie végétale
par

MYRIAM HEUERTZ

DIRECTEUR DE THESE: PROF. XAVIER VEKEMANS
ANNEE ACADEMIQUE 2002 / 2003

Acknowledgements

In the very first place, I would like to thank my supervisor Xavier Vekemans, who did a great job coaching my thesis despite the 220 km that separated us most of the time. You are the person who introduced me to the fascinating field of population genetics and I enjoyed very much working with you over the past five years, master's dissertation included. I would like to thank you especially for encouraging me in the more difficult moments during the thesis and for making me feel that there was no doubt I could do it (although I had many doubts...). Thank you Xavier!

Many thanks also the Claude Lefèbvre, the father of population geneticists in the Brussels laboratory, who always has an open ear for all our problems!

Then, I would like to thank Lucien Hoffmann and Jean-François Hausman for giving me the opportunity to work on the GENFOR project on forest genetic resources in the CRP-GL in Luxembourg and for allowing me to pursue that work into a PhD thesis. Thank you Jean-François especially for your patience with me, desperately trying to arouse my diplomatic features (do they exist??). In the future, we will try to reconcile population genetics and adaptation studies; that is the future!

The Luxembourg team in the laboratory also deserves many thanks and a huuuge hug! Thanks for your company, for you advice with laboratory problems and for your constant support! Some special acknowledgements go to the technical staff for making the lab run, to Sandra for assistance with sequencing, to Jenny for some interesting chats sharing the office, and to Danièle for her understanding and tightening the team. And I don't forget the computer specialists who always have a solution when ordinary people capitulate.

Then I would like to thank Claude Muller from the Immunology Department of the National Health Laboratory in Luxembourg, for letting me use the automated sequencer. Thanks also to Sabrina and mainly Mick for flexibly scheduling the sequencer use and for their constant availability in case of problems. I was quite astonished that you remained that relax when I broke the precious glass plates, and moreover, when I nearly burned the machine because of such a (dam...) broken plate...

Many thanks to the project partners for their hospitality and their collaboration during the collection of ash samples in southeastern Europe, notably Magdalena Palada from ICAS Bucarest and her amazing group of "ash-hunters", Ivo Tsvetkov, Denitsa Pandeva, Vesselka Gyuleva, Emil Popov, the driver Bobi and their boss Alexander Alexandrov from the Forest Research Institute in Sofia, and Nathalia Verminciuk and Georghe and Dragos Postolache and their team from the Institute of Botany, Chisinau. At the same time, thanks to Jean-François Hausman, Agnès Labarthe and Thierry Frank for assistance with field work.

In the framework of this project, the International Plant Genetic Resources Institute played a key role. Thanks especially to Jozef Turok for coordinating and for stimulating discussions.

An important person for getting me started in this thesis was also Nathalie Frascaria. Thank you and your team, and particularly Marie-Elise Morand, for interesting discussions and sharing samples and unpublished results.

And then, there was Beppe Vendramin with his Italian charm (are you ready??) and his team with Maria, Roberta, 2 x Silvia and Gabriele. I will never forget the lunches at Strizzi's Bar where they played music from Luxembourg (wasn't it Beppe?) and where all was about calcio and AC Milan! But by the way, thanks also for giving me the opportunity to analyse cpDNA in your lab, and perform many extractions with the "supershaker".

Now, the moment for EUFORGEN has come! In the first place, many thanks to Frank Wolter from the Luxembourgish Forestry Administration, for allowing me to participate in the EUFORGEN Noble Hardwoods Network (and also a little in the Network for temperate oaks and beech). It was a great chance for me to present my results in the network meetings and receive feedback from forest geneticists and conservationists. Without these networks, sampling ash at the scale of Europe would not have been possible, and moreover, a met some good friends!

As we speak about sampling at the scale of Europe, I would like to acknowledge the collaboration of the following people and institutions: Lennart Ackzell from the Swedish Forestry Board; Sandor Bordasz from Hungary; Patrick Bonfils and Marcus Ulber from WSL, Switzerland; Ned Cundall from the Northern Research Station, Midlothian, Scotland; John Fennessy from the Irish Forestry Board; Jan Kowalczyk and Justyna Nowakowska from Poland, Ladislav Paule from the Technical University in Zvolen, Slovakia; Alfas Pliura from the Lithuanian Forest Research Institute and Randolph Schirmer from the Bavarian Institution for Forest Seeding and Planting, Germany.

We have travelled all over Europe, now we can come back home! Thanks to the staff of the laboratory in Brussels for your encouragements, mainly during the last six months! And especially, thanks to Olivier, the magician in population genetics, who produces ideas in population genetics just as magicians produce rabbits... Thank you for your help and supervising my work during the last few months!

Thanks also to the Groupe de Biologie et Génétique des Populations in Belgium. It was a really good idea to create such a group among Belgian university laboratories, where it is always possible to informally discuss results and perspectives.

During these four years, I recognised once more that there are friends! Without you, life would have been much less amusing, and work much harder. Thank you Fabienne (and hello to South America) for your friendship, sharing the flat with me in Luxembourg. Thanks also to Nadine, Anne, Insa, Malou and Christiane for encouraging me and spending good moments together. Not to forget boys, and quite astonishingly, many of them have excellent features in making maps, especially in urgent situations :-) I think especially of Michele, Christian, Frank and Claude. Thanks also to all karate friends (mawashi, geakuzuki!) and all others.

My "little" sisters Conny and Diane are really important people for me, and I enjoyed our common time in Brussels, where you constantly kept me motivated. Thanks an een décke Kuss! Finally, I am really happy that my parents supported me over these years, and helped printing! Thanks! Hello also to my brother Claude, who I see more rarely, but we are a good team!

I acknowledge financial support from the Public Research Centre-Gabriel Lippmann CRP-GL; the Luxembourgish Ministry of Culture, Higher Education and Research MCESR; the International Plant Genetic Resources Institue IPGRI; the Free University of Brussels ULB; the Luxembourgish National Research Fund FNR; the Belgian National Fund for Scientific Research FNRS; the Centre de Documentation et d'Information sur les Etudes Supérieures in Luxembourg CEDIES; the Rotary Club section Ettelbruck-Diekirch in Luxembourg and the European Science Foundation ESF.

Thanks to you all, and to all those I forgot! And finally a little hello and a message for those who work in the European Union-funded ash projects RAP and FRAXIGEN: There is still a lot more to do; go ahead!

THANKS TO ALL OF YOU! *Myriam*

Table of contents

Table of contents 1
Introduction 1
1. Conservation and sustainable use of forest genetic resources 1
1.1. The policy framework of forest genetic resources conservation in Europe 1
1.2. The challenge 2
1.3. Evolutionary processes 3
1.4. Factors causing endangerment 6
1.5. Utility of genetic markers for conservation 6
1.6. Conservation methods 8
2. Population genetic structure in European forest trees 11
2.1. Genetic diversity within populations of sessile and pedunculate oak analysed with nuclear DNA markers. 12
2.2. Chloroplast DNA patterns at the European scale 13
2.3. Structure of genetic diversity in Quercus petraea investigated with nuclear markers at the European scale. 18
2.4. Comparison of chloroplast and nuclear patterns of variation at genetic marker loci 19
2.5. Comparison of chloroplast patterns of variation with nuclear controlled phenotypic traits 19
3. Materials and Objectives 20
3.1. Forest resources in southeastern Europe 20
3.2. Common ash (*Fraxinus excelsior* L.) 22
3.3. Objectives 26
Chapter I: Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.) 29
Chapter II: Estimating seed versus pollen dispersal from spatial genetic structure in the common ash 39
Chapter III: Nuclear microsatellites reveal contrasting patterns of diversity in western and eastern European populations of the common ash (*Fraxinus excelsior* L.) 58
Chapter IV: Different evolutionary history and geographical distribution of even- and odd-sized alleles at a microsatellite locus in *Fraxinus excelsior* (L.) 79
Chapter V: Incongruence between chloroplast and nuclear marker data among European populations of common ash 95
References 121
Appendices 130

Introduction

This thesis is an attempt to understand the factors that shaped the present day spatial structure of genetic diversity in the common ash. In the following introduction, I will give some backgrounds on the conservation and sustainable use of forest genetic resources and highlight the importance of genetic knowledge for the setting up of conservation strategies. Using examples from the literature, I will present genetic diversity patterns observed at different spatial scales in European broadleaved forest tree species and show how they have been interpreted. Finally, I will introduce the context, the study species and the objectives of the present work.

1. Conservation and sustainable use of forest genetic resources

1.1. The policy framework of forest genetic resources conservation in Europe

In Europe, major threats affecting forests and forest biodiversity were recognised in the 1980's: acid rain, uncontrolled forest genetic erosion, forest fires, decline of mountain forests, lack of research and scientific knowledge on forest tree physiology and forest ecosystems (Arbez 2000). European countries expressed their will to commonly address the opportunities and threats related to forests and forestry and initiated a forum at the ministerial level in 1990 in order to develop common instruments and measures for the protection and sustainable management of forests. Forty European countries including the European Union participate in this initiative, which operates through the organisation of Ministerial Conferences, expert level meetings, round table meetings and working groups. To date, three ministerial conferences on the protection of forests in Europe were held; in Strasbourg 1990, in Helsinki 1993 and in Lisbon 1998. A key action in the ministerial process on forests was the adoption of Strasbourg resolution S2 on forest genetic resources. Signatory countries committed themselves to apply this resolution at the national level by taking measures for the conservation of forest genetic resources by whatever means they judged appropriate. At the international level, the **European Forest Genetics Resources Programme** (EUFORGEN) was launched in 1995 as an instrument of international cooperation for the implementation of resolution S2. It is financed by participating countries and is coordinated by the International Plant Genetic Resources Institute (IPGRI), in collaboration with the Forestry Department of the UN Food and Agriculture organisation (FAO). EUFORGEN operates through five networks for target groups of species (*Populus nigra*, noble hardwoods, conifers, mediterranean oaks and temperate oaks and beech) with the main objective of encouraging and strengthening national programmes. It is overseen by a steering committee of national

coordinators nominated by the participating countries. EUFORGEN in now in its second phase (2000-2004), and specific objectives for this period include (1) the exchange of information, (2) the development, coordination and promotion of long-term management and protection strategies, (3) the elaboration of technical guidelines for foresters for management and genetic conservation of different target species, (4) the development of descriptors and databases and (5) raising of public awareness (Turok 2002). These objectives are addressed in network meetings, held about every 18 months, which bring together forest geneticists and other forestry specialists of the participating countries. They further analyse common needs in research and forestry, implement work plans and monitor the progress in the different countries.

Beyond its original focuses, EUFORGEN has proved to be a useful platform implicated in the development of research projects, essentially financed by the European Union, and in the dissemination of research results and their incorporation in practical forestry. It is also involved in organising training workshops on forest genetic resources. Future proposed scopes are a stronger focus on gene conservation within the frame of sustainable forestry management and to facilitate the implementation of gene conservation networks at the European scale. Further, membership should be expanded to additional countries within Europe, and collaboration with other network programmes and linkage with international processes should be strengthened (Turok 2002).

At the ministerial level, cooperation with the ministerial process "Environment for Europe" was initiated and a joint "Work Programme on the Conservation and Enhancement of Biological and Landscape Diversity in Forest Ecosystems" (1997-2000) was implemented. The next ministerial conference on the protection of forests in Europe is to be held in Vienna in April 2003 and major themes are (1) forest biological diversity, (2) and economically viable forest management, (3) climate change and sustainable forest management in Europe and (4) cultural dimension of sustainable forest management.

1.2. The challenge

Effective conservation strategies demand a clear definition of objectives, adequate knowledge on the biological units to be conserved and appropriate conservation methods (Eriksson *et al.* 1993). The central objective for conservation in forestry is to **safeguard** the **evolutionary potential**, the **capacity to adapt** in an ever-changing environment (e.g. Eriksson *et al.* 1993, Kanowski & Boshier 1997, Namkoong 1998). This is evident as trees are often exposed during their long-lasting life cycles to a variety of ecological conditions, concerning both the physical environment and biotic interactions. The environment is dynamic and especially in the view of a threatening global climate change, it would be wrong to aim at preserving or freezing today's genetic composition of forests. The time-scale of concern for conservation

also matters (Frankel *et al.* 1995). Hence, in order to prepare forest genetic resources for an uncertain future, the objectives of long-term sustainable management must be conservation and even enhancement of biological diversity.

Biological diversity can be viewed in a hierarchical manner focusing from the diversity of communities or ecosystems to the genetic diversity within species (Primack 1993). The **species** certainly has a central place in conservation science as it contains a wealth of diversity from genes to populations and provides a measure for the diversity of communities (Frankel *et al.* 1995). It must nevertheless be stressed that meaningful species conservation in their natural environment (*in situ*) is impossible without conservation of their ecosystems (Frankel *et al.* 1995).

The evolutionary potential of forest tree species lies in their **within-species genetic diversity**. The level and structure of this diversity can be investigated by analysing patterns of neutral (i.e. non-adaptive) variation at genetic marker loci such as allozymes, random amplified polymorphic DNA (RAPDs), amplified fragment length polymorphisms (AFLPs) or microsatellites (e.g. Milligan *et al.* 1994). In particular, in a review of allozyme polymorphisms of 322 woody taxa, Hamrick *et al.* (1992) pointed out that **woody species maintain more variation within species and within populations** than other life forms but have **less variation among populations.** These diversity patterns can be explained partly by some common features shared by the majority of tree species, such as (i) mainly outbreeding mating systems and thus high levels of gene flow, (ii) individual longevity and fecundity and (iii) little or no history of domestication (with the exception of several fruit-bearing species, Loveless & Hamrick 1984, Hamrick *et al.* 1992).

1.3. Evolutionary processes

The evolutionary forces that govern genetic diversity act on allelic frequencies in **populations**. This is why population structure is of chief importance for conservation efforts at the level of individual species. The evolutionary forces are selection, mutation, gene flow and genetic drift. It is always essential to consider the balance of evolutionary forces in a particular situation.

1.3.1. Evolutionary forces

a. Selection

Selection is the only force that causes adaptive evolution by acting on the adaptive traits. It arises because different genotypes have different rates of reproduction and survival (reproductive fitness) and will transmit their genes accordingly.

The adaptive diversity of a species is mostly expressed trough quantitative variation among individuals in characters or traits. Some adaptive characters are monogenic (e.g. those involved in resistance to pathogens), but most traits are governed by many loci, which contribute small, often additive effects to total variance, and they are strongly affected by the environment (e.g. Frankham *et al.* 2002). The proportion of the variance of a particular trait in a population due to genetic differences among individuals is called heritability. The heritability of a trait determines its response to directional selection (selection favouring one extreme, Frankham *et al.* 2002) and therefore the evolutionary potential of the population with respect to that trait. The heritability is comprised between 0 in highly inbred populations without genetic variation and 1 in outbred populations where all genetic variance is additive and there is no environmental variance (Frankham *et al.* 2002). Most quantitative traits that have been studied show additive genetic variance (heritability > 0, e.g. Lynch & Lande 1993), which means that populations and species have the ability to respond to changing environmental conditions.

If the environment relevant to a particular trait is perceived as heterogeneous in the long term, selection will favour different genotypes in different areas and among-population differentiation will establish. In this case, **stabilising selection** occurs in populations within areas with homogeneous selection pressure regarding the particular trait, and **disruptive selection** is experienced among them, i.e. at the species level (Eriksson 1998). Hence, selection will maintain overall genetic diversity in heterogeneous environments.

b. Mutation

Mutations are a source of novel genetic, possibly adaptive variation. Rates of point mutations for individual nucleotide sites are low, approximately 10^{-10} to 10^{-9} per generation (Hancock 1999). Microsatellites have mutation rates on the order of 10^{-5} - 10^{-2} (Jarne & Lagoda 1996, Hancock 1999), which is two to three orders of magnitude higher than values known for allozymes (Jarne & Lagoda 1996). Pooled mutation frequency in loci influencing a quantitative trait may be somewhat higher, on the order of 10^{-3} - 10^{-2} (Lande & Barrowclough 1987). These rates are overall low, and the reintroduction of allelic diversity by mutations after for example a bottleneck takes time. Mutations are however of prime importance to maintain and increase adaptive potential of species at the long term.

c. Gene flow

In plants, gene flow refers to the movement of genes through the dispersal and successful establishment of gametes (pollen) or zygotes (seeds). Gene flow is a homogenising force, and acts thus against the forces that promote differentiation of local populations, namely selection, mutations and genetic drift. For instance, Slatkin (1985) has shown that one migrant per population per generation is enough to prevent differentiation due to genetic drift among populations (for populations approximating an island model, Wright 1931). Gene flow can

increase the genetic diversity of a population if it brings in migrants from populations with different alleles and allele frequencies.

d. Genetic drift

Genetic drift designates the random change in allele frequency that occurs because gametes transmitted from one generation to the next carry a biased sample of the alleles present in the parental population. In a random mating population under the effect of genetic drift, genetic variation is lost within populations at a rate $1/2N_e$ per generation where N_e is the effective population size (see below). As this loss affects genetic variation in a random fashion, differentiation inevitably increases among populations. The expected increase of inbreeding, or mating among relatives, is also equal to $1/2N_e$ per generation. As inbreeding increases levels of homozygosity, deleterious recessive alleles become expressed, which results in a decline in (reproductive) fitness in inbred individuals, known as inbreeding depression.

1.3.2. Population size

The **effective population size N_e** is a fundamental parameter for conservation biology as it determines the relative importance of directional (migration and selection) and stochastic (genetic drift) genetic factors. The effective population size is the size of an ideal population that would be subjected to the same amount of genetic drift as the actual population. An ideal population corresponds to a closed random mating population with discrete generations, constant population size, equal sex ratio and Poisson variation in family sizes (e.g. Frankham *et al.* 2002). Real populations rarely fulfil the assumptions of ideal populations. As a consequence, the effective size N_e is usually smaller than the census size N_c. Correct estimation of the ratio N_e/N_c is a difficult task, but many estimates fall within the range 0.1 - 0.2 (Nunney & Campbell 1993) or 0.1 - 0.4 (Waples 2002).

In **large populations** (populations with large effective size), stochastic genetic factors are negligible. Diversity patterns in large random mating populations are therefore essentially determined by **migration and selection**, mutations contributing only weakly. Migration brings in new variants that reflect allele types and frequencies from their source population(s) whereas selection affects beneficial and detrimental alleles, but not neutral ones (except under hitchhiking; i.e. when neutral genes are linked to selected ones). Under distinct environmental conditions, natural selection favours distinct alleles. This is why patterns of among-population differentiation can vary for different traits and particularly between traits and neutral (i.e. non-selected) genetic markers (e.g. Karhu *et al.* 1996).

In **small populations**, the effect of **stochastic genetic factors** is strong. Variation is lost at a rate $1/2N_e$, regardless whether it is adaptive or not. Hence, selection is less efficient in small than in large populations. This has two corollaries: small populations adapt less well to changing environmental conditions than large ones, and deleterious alleles are less likely to be

eliminated by selection in small populations and may even become fixed. A strong reduction in population size is called a population **bottleneck**. A bottleneck can occur for example during colonisation of a new area (founder effect) or if a population faces threats (section 1.4.). The magnitude and duration of a bottleneck strongly influence the population's genetic features (allelic richness *vs.* gene diversity, Nei 1975; inbreeding depression *vs.* genetic load, Kirkpatrick & Jarne 2000).

1.4. Factors causing endangerment

The threats to a population or a species can be divided into deterministic factors on the one hand, and stochastic factors on the other.

Deterministic factors are directional and relatively consistent in magnitude. They are related to human action. Deterministic factors threatening forest tree species include deforestation, exploitation, fragmentation, demographic and habitat alterations, environmental deterioration, translocation and domestication (Ledig 1992). Global climate change is a threat of eminent importance belonging to this category (e.g. Mátyás 1998, Eriksson 1999).

Stochastic factors are those that affect populations, in addition, when they have a small effective size. Shaffer (1981) has identified four forms of stochasticity or uncertainty in small populations: (i) demographic stochasticity arises from variation in the survival and reproductive success of individuals; (ii) environmental stochasticity refers to unpredictable fluctuations in rainfall, temperature, density of competitors, pathogens or herbivores that affect birth and death rates; (iii) natural catastrophes include fires, floods and extremes in climate hazards (e.g. drought, frost) and (iv) genetic stochasticity encompasses inbreeding depression, loss of genetic variation and accumulation of deleterious mutations.

As a response to deterministic threats, for instance global climate change, a population can either evolve or migrate; on the contrary in order to face detrimental effects of stochastic factors, only an increase in population size could be useful.

The threats caused by genetic uncertainty have not fully been recognized for a long time; for instance Lande's (1988) view was that demographic and environmental stochasticity or natural catastrophes would lead to population extinction before genetic factors could act. The role of genetic threats affecting species or population endangerment is however well established now (e.g. Hedrick 2001, Allendorf & Ryman 2002).

1.5. Utility of genetic markers for conservation

We saw that the central objective for conservation of forest genetic resources is to safeguard their adaptive diversity, their evolutionary potential. The question that promptly arises is: does

population genetic structure investigated with neutral molecular markers reflect patterns of adaptive diversity? Hedrick (2001) inspected this question and recommends caution in inferring biological differences from significant marker differences among groups. His main argument is that highly variable marker loci used in population biology (e.g. Sunnucks 2000) lead to high statistical power of differentiation tests (Goudet *et al.* 1996). As a result, the differences at marker loci are real, but they do not necessarily reflect biological differences. An example of the opposite situation is given in Karhu *et al.* (1996): non-significant differentiation at allozyme loci reveals the high levels of gene flow among populations in Scots pine, but diversifying selection for an adaptive trait, bud set, was strong enough to overcome the homogenising effects of gene flow and produce large adaptive genetic differences between populations.

Neutral molecular markers can nevertheless be used to track adaptive variation when they are employed for association mapping with adaptive traits. The linkage of a marker locus to a trait is used for example in marker-assisted selection in forest tree breeding and allows considerable shortening of generation turnover (O'Malley & Whetten 1997).

Molecular markers are used to address many other questions concerning species biology relevant to conservation biology (e.g. Hedrick 2001):

- Resolve taxonomic uncertainties (phylogenetics) and identify hybridisation
- Mark the boundaries of evolutionary significant units or management units and investigate the origin of individuals
- Measure effective population size
- Detect past bottlenecks
- Assess total gene flow at different scales and relative contributions of pollen and seed to gene flow
- Identify founder contributions
- Investigate historical gene flow and locate colonisation routes (phylogeography)
- Estimate levels of inbreeding
- Study mating systems

Furthermore, molecular markers permit non-intrusive sampling methods in animals, they provide help in choosing populations and suitable sites for reintrodution and they are used for forensics (e.g. find traces of protected species in goods for consumption, Frankham *et al.* 2002).

1.6. Conservation methods

1.6.1. In situ versus ex situ conservation

There are globally two conservation approaches: *In situ* conservation refers to conservation of a population or species in the wild, in its natural habitat; whereas *ex situ* conservation designates conservation away from the natural habitat, which for forest trees could be in provenance trials, in seed orchards or clonal archives, or in *in vitro*- or cryopreservation. In contrast to *ex situ* methods, ***in situ* conservation** is a **dynamic** means of conservation, genetic resources are continuously exposed to selection pressures. Another advantage is that *in situ* conservation mostly allows sustainable use of the material to be conserved. Both *in situ* and *ex situ* approaches can complement each other. In order to know which method to use or how to combine them for a particular species, its state of endangerment has to be evaluated, and associated objectives to long-term gene conservation, like for example improvement through breeding, have to be identified.

In situ and *circa situ* conservation are critically important in forest tree species: indeed, only few tree taxa, around 100, are effectively conserved *ex situ* and there are often technical and economic limitations to the establishment of *ex situ* programmes (Kanowski & Boshier 1997). Regarding *in situ* conservation for forest trees, it is however well known from provenance tests that the local material is not always the best adapted to the actual site conditions (e.g. Mátyás 1996). This has raised polemics about the utility of promoting natural regeneration and using autochthonous reproductive material for long-term conservation against non-autochthonous material for which the genetic adaptedness of the future stand has been evaluated to be higher (e.g. Geburek & Müller 2000).

1.6.2. Importance of population size and structure

The influences of many threats affecting genetic resources of a species reduce essentially to a decrease in the effective population size. There has been much debate on how large a population should be to maintain sufficient genetic variation in order to be able to adapt in the long term (Allendorf & Ryman 2002). Basically, an isolated population must be large enough for mutations to compensate the loss of variation caused by genetic drift. The suggestions for the effective population sizes needed range from 500-5000, which in many species cannot be accomplished unless the connectivity among populations is increased over a wide geographic area (Allendorf & Ryman 2002).

It appears thus that population size and among-population structure are central issues for conservation management. In the special case of forest tree species, Eriksson *et al.* (1993) described five basic categories of natural population structure ranging from a single large continuous random mating population over several populations interconnected to various degrees to a continuous population where the relatedness between individuals depends on

their spatial distance. Extinction and recolonisation dynamics may also influence the structure of populations (Wade & McCauley 1988).

1.6.3. Example of a strategy for conservation and management of forest trees: the multiple populations breeding system applied to noble hardwoods in Europe

Noble hardwoods represent a heterogeneous group of species, which includes, among others, maples, hornbeam, ashes, chestnut, walnut, lime trees, elms, wild fruit trees and *Sorbus* species. The common features of those species are scattered distribution patterns in European mixed forests, high ecological demands and timber of high quality (Turok *et al.* 1996). Noble hardwoods are rare and some even endangered (Rotach 1999). Many have been neglected until recently, which explains why information on crucial life-history characteristics of these species (e.g. mating system, pollen and seed dispersal) and on their patterns of genetic variation is still scarce. As biodiversity, sustainability and adaptability of ecosystems are being perceived as highly important, and not least because of a large demand of high-quality timber, noble hardwoods are being promoted now. Their effective conservation and sustainable utilisation are supported by the EUFORGEN Noble Hardwoods Network (Turok *et al.* 1996, Turok *et al.* 1998, Turok *et al.* 1999, Turok *et al.* 2002). Whenever possible, genetic resources of noble hardwoods should be managed and conserved *in situ*, in the frame of forestry, to allow for joint gene conservation and production of valuable timber (Rotach 1999).

The gene resource conservation strategy that the EUFORGEN Noble Hardwoods network currently recommends for European noble hardwoods is the multiple populations breeding system (MPBS) (Eriksson 2001 and references therein). The MPBS concept was initially proposed for combined breeding and gene conservation by Namkoong (1984). Eriksson *et al.* (1993) suggested that MPBS could be applied for various management strategies, from very elaborated breeding programs to simple *in situ* gene resource populations with no or limited silvicultural intervention. The main idea of MPBS is to identify gene resource (sub)-populations and to manage them *in situ*, or, jointly, *in situ* and *ex situ*, in order to increase among-population variance. Ideally, the subpopulations are chosen from heterogeneous environments (according to more than one variable) and/or in a way to maximise their initial differentiation among each other. *In situ* management aims at increasing the adaptation to local conditions while sites for establishment of populations *ex situ* are chosen with the objective of minimising among-population gene flow. The total gene resource population should contain approximately 20 subpopulations each of which should have $N_e \geq 50$. In these subpopulations, variability is lost at a rate of 0.01 and inbreeding accumulates at the same rate, which is a figure generally accepted for short-term conservation (Allendorf & Ryman 2002). The strength of MPBS however is the enhanced among-population differentiation, so that the total adaptive diversity is globally increased.

Eriksson *et al.* (1993) stressed that a proper sampling of gene resource populations is crucial for efficient gene conservation strategies, both *in situ* and combined *in situ* and *ex situ*. If available, genetic information on among-population differentiation for marker loci and/or traits should be accounted for when choosing gene resource populations. Otherwise, the steepness of ecological gradients and biological characteristics of the species could be helpful to produce educated guesses of among-population differentiation for traits and the strength of gene flow (Eriksson 1998).

According to their occurrence, use and susceptibility to devastating diseases, noble hardwoods are divided into four categories for management following MPBS (Eriksson 2001):

1. Commonly occurring species. This category addresses mainly ash and maple. Gene resource populations should be chosen *in situ* from heterogeneous environments, guided by genetic information when available, in order to sample a maximum of variation. In some populations, a larger area should be conserved in order to match the objective of conserving associated species.
2. Rarely occurring species. This category concerns most noble hardwoods. Gene resource populations should be chosen *in situ* for joint conservation of several species when possible. A commonly encountered problem is small effective population size in the wild. Increase of the natural population sizes is encouraged through thinning of competing species and through planting of material of autochthonous origin (e.g. Rotach 1999) conserved in seed orchards or clonal archives (*ex situ*).
3. Multipurpose species, e.g. used for timber and fruits. The long history of domestication of for example nut-bearing trees such as chestnut is likely to have strongly differentiated cultivated populations from those in the wild. Breeding activities in the cultivated gene pool should be continued and kept as isolated as possible from natural wild populations in order to minimise genetic erosion of the latter by the cultivated gene pool, which has reached a high domestic fitness at the cost of natural fitness.
4. Populations of elms (*Ulmus sp.*) have drastically declined due to Dutch elm disease (DED). For *Ulmus minor*, which reproduces vegetatively through root suckers, clones from the whole distribution area should be conserved *ex situ* in clonal archives and partly in cryopreservation. For *U. glabra* and *U. laevis*, MPBS should be applied *in situ*, except when the risk of DED is high. In that case, conservation is recommended in low clonal hedges, which protect elm plants from the DED vector, *Scolytus* insects.

Some of those recommendations are being set up, for instance in elms, which are clearly endangered (Eriksson 2001). EUFORGEN facilitates the dissemination of information and collaborative initiatives; political commitments are not within its scope.

Box 1 Statistics commonly used in population genetics

Nuclear genetic markers

Measures of polymorphism

A **or** K	Mean number of alleles; noted A_S when computed **within populations** and A_T when computed **in the total sample** (all populations considered together). Values are sometimes standardised to a particular population size.
H_O	Observed heterozygosity, or proportion of heterozygotes (only for codominant markers). Varies from 0 (homozygotes only) to 1 (all are heterozygous).
H_E	Gene diversity or heterozygosity expected under random mating; equivalent to the probability of drawing two different alleles when sampling at random (with replacement) in a population. Varies from 0 to ≈1. This statistic is noted H_S **within populations** and H_T **in the total sample**.

Measure of inbreeding

F_I	Inbreeding coefficient, or fixation index within individuals, computed as $(H_E-H_O)/H_E$; noted F_{IS} **within populations** and F_{IT} **in the total sample**. Varies from −1 to +1; there is inbreeding if $F_I > 0$, selfing species have $F_I \approx 1$.

Measures of among-population differentiation

F_{ST}	Fixation index among populations, computed as $(H_T-H_S)/H_T$, measures the proportion of diversity distributed among populations. Varies from 0 (all populations equal) to 1 (populations completely different). The statistic was initially developed for loci with 2 alleles.
R_{ST}	Similar to F_{ST}, this statistic was developed to measure differentiation at microsatellite loci, taking into account the differences of allele sizes (≈ number of mutations).

Chloroplast genetic markers

Measures of polymorphism

A **or** K	Mean number of alleles.
h_S **or** h_T	Haplotypic[1] diversity within populations or in the total sample.
v_S **or** v_T	Haplotypic diversity within populations or in the total sample considering the phylogenetic differences among haplotypes.

Measures of among-population differentiation

G_{ST}	Similar to F_{ST}, computed as $(h_T-h_S)/h_T$.
N_{ST}	Differentiation statistic that takes into account the phylogenetic differences among haplotypes, computed as $(v_T-v_S)/v_T$.

[1] In the chloroplast genome, all genes are fully linked on a single DNA molecule. Therefore, "alleles" or variants at individual loci cannot be considered isolated from each other; the combination of particular variants over all loci is called haplotype.

2. Population genetic structure in European forest trees

Today, there is a wealth of genetic markers available for the investigation of patterns of genetic variation (Parker 1998, Sunnucks 2000, Table 1). The patterns of genetic variation in plant species at a large geographical scale, like the European continent, have been found to result from the combined actions of a multitude of factors. Some of these are intrinsic to the species life history (breeding system, modes of seed and pollen dispersal, life form, gregariousness), while others are perturbations induced by natural processes (ice ages, climatic stochasticity) or human impact (habitat fragmentation, global change).

Tree species, as briefly mentioned before, present high genetic diversity at nuclear marker loci at the species and at the population level, but very little genetic differentiation among populations, compared to species with other life forms (Hamrick & Godt 1989, Hamrick *et al.* 1992, Box 1, Table 2). Furthermore, woody species with large geographical ranges, outcrossing breeding systems and wind or animal-ingested seed dispersal maintain more diversity within species and populations, but less variation among populations than woody species with other combinations of traits (Hamrick *et al.* 1992). Although life history and ecological traits explained 34% of the variation among species, the authors pointed to the large proportion of variation that remained unexplained and suggested the importance of the specific evolutionary history of each species in determining the level and pattern of genetic diversity.

The climatic changes of the Quaternary (the last 1.6 million years) have considerably influenced the evolutionary history of temperate tree species by triggering successive contractions and expansions of their distribution ranges, involving population extinctions, size reductions and long-term isolation. In many species, these historical imprints can still be identified today at a large geographical scale (see Comes and Kadereit 1998, Newton *et al.* 1999 and Hewitt 2000 for reviews). At the regional or local geographical scale, these patterns are usually blurred by the combined effects of selection, migration and genetic drift (e.g. Pigliucci *et al.* 1990, Leonardi and Menozzi 1996, Streiff *et al.* 1998, Lee *et al.* 2002).

Two main methodological approaches have been used to investigate the patterns of genetic variation in plant species at different geographical scales with the aim of making inference on their causal factors. Those methods employed either (i) organelle (mainly chloroplast) DNA markers or (ii) nuclear genetic markers (Table 1). In the following sections, I will present patterns of population genetic structure in European forest trees unravelled with those types of markers and give possible explanations for them according to authors of the respective studies. I chose to approach the issue by giving examples mainly of the species *Quercus petraea* (Matt.) Liebl., sessile oak, and more generally the white European oaks species

Table 1 Frequently used molecular markers in population genetics

	Markers that address variation in nuclear DNA				Chloroplast DNA markers	
	Isozymes / allozymes:	**Nuclear microsatellites / simple sequence repeats (SSRs)**	**Random amplified polymorphic DNA (RAPD)**	**Amplified fragment length polymorphism (AFLP)**	**Polymerase chain reaction – Restriction fragment length polymorphism (PCR-RFLP)**	**Chloroplast microsatellites**
Inheritance	Codominant	Codominant	Dominant	Dominant	Uniparental	Uniparental
Polymorphism	Modest	High	High	High	Modest	Moderate
Main use	Genetic diversity and structure	Genetic structure at local scale	Genetic diversity and structure	Genetic diversity and structure	Phylo-geography	Phylo-geography
Advantage	Many studies available for comparison[1]	Application in related species sometimes possible	Multilocus technique	Multilocus technique, high reproducibility		Higher polymorphism than RCR-RFLP
Disadvantage	Selection cannot be excluded[2]	Homoplasy[3], null alleles[4], scoring problems	Low reproducibility	Scoring is difficult		Homoplasy[3]
Cost	Moderate	High, especially the development	Low	High	Moderate	Moderate

[1]Hamrick & Godt (1989), Hamrick *et al.* (1992), Hamrick & Godt (1996).
[2]Sander *et al.* (2000).
[3]Alleles have the same size but are not identical by descent.
[4]Alleles that do not amplify in the polymerase chain reaction (PCR) because of mutations in the primer annealing sites.

Table 2 Levels of allozyme diversity within and among populations of species belonging to different life forms. H_E, genetic diversity within populations; G_{ST}, proportion of the total diversity among populations. Means followed by the same letter are not significantly different at the 5% probability level. From Hamrick *et al.* (1992).

	H_E	G_{ST}
All species	0.113 (0.004)	0.228 (0.010)
Annual	0.101^{b} (0.007)	0.355^{a} (0.021)
Short-lived perennial: herbaceous	0.098^{b} (0.006)	0.253^{b} (0.018)
Short-lived perennial: woody	0.096^{b} (0.015)	0.155^{bc} (0.038)
Long-lived perennial: herbaceous	0.082^{b} (0.009)	0.278^{ab} (0.033)
Long-lived perennial: woody	0.148^{a} (0.006)	0.084^{c} (0.008)

complex. These examples are meant to introduce the reader to population genetic structure of European forest trees and will serve for comparison in the discussion of the genetic structure patterns of common ash presented in this thesis.

Oaks are abundant and widely spread in Europe and play important roles from both economical and ecological points of view. Their morphology and molecular genetics have been thoroughly studied for practical reasons, for example in the view of developing methods for tracing the geographic origin of seed lots, identifying the source of plantations, checking the autochthonous status of oak woods or forests for conservation purposes or even tracing the origin of wood products (Petit *et al.* 2002a). Biologically, oaks represent a highly challenging study subject, as many of them undergo extensive introgressive hybridisation, forming a species complex. Sessile oak (*Quercus petraea*) is a commercially important widespread species covering most parts of Europe from Spain to Russia and Scotland to Turkey. It is sympatric with pedunculate oak (*Quercus robur*). Both species often occur in the same stands and occupy different but proximal ecological niches.

2.1. Genetic diversity within populations of sessile and pedunculate oak analysed with nuclear DNA markers.

Due to the sympatric occurrence of sessile (*Quercus petraea*) and pedunculate (*Quercus robur*) oaks, many studies have investigated diversity patterns in both species. Levels of within-population genetic diversity in both species are summarised in Table 3.

It appears that pedunculate and sessile oaks are highly polymorphic species, for instance when comparing their mean within-population genetic diversity for allozymes ($H_E \approx 0.25$) to that of other long-lived woody species (mean $H_E = 0.148$, Table 2). Microsatellites exhibit higher levels of diversity ($A \approx 20$, $H_E \approx 0.87$) than allozymes ($A \approx 3$, $H_E \approx 0.25$), whereas the dominant multilocus methods RAPD and AFLP show the lowest diversity among the markers assessed ($H_E \approx 0.21$).

Significant heterozygote deficiencies, as reflected by comparison of H_O *vs.* H_E or by positive values of the fixation index (inbreeding coefficient) F_{IS}, were found in several studies and the trend was stronger for allozymes than for microsatellites. Possible explanations for high F_{IS} values include inbreeding, especially due to non-random spatial genetic structure because of limited seed and/or pollen dispersal within stands (Bacilieri *et al.* 1994, Streiff *et al.* 1998); and a Wahlund effect (Degen *et al.* 1999) possibly due to phenological differences (Bacilieri *et al.* 1994). Degen *et al.* (1999) attributed markedly lower F_{is} values in microsatellites than in allozymes to scoring errors of microsatellites, because similar trends for all biparentally inherited markers should be expected if the mating system were the only process responsible for hetrozygote deficiency. The usual pattern, however, is to find higher F_{IS} values with

Table 3 Levels of genetic diversity within populations of *Quercus robur* and *Q. petraea*: A, mean number of alleles per population; H_O, observed heterozygosity; H_E, gene diversity; F_{IS}, fixation index.

Marker	Species	No. pops	No. trees/pop	No. loci	A	H_O	H_E	F_{IS}	Reference
allozymes	*Q. robur*	7	120	13	2.703	0.184	0.252		Zanetto *et al.* 1994
allozymes	*Q. robur*	1 mixed	217	7				0.244	Bacilieri *et al.* 1994
allozymes	*Q. robur*	2	156	7	3.065	0.298		0.072	Degen *et al.* 1999
allozymes	*Q. petraea*	7	120	13	2.725	0.222	0.245		Zanetto *et al.* 1994
allozymes	*Q. petraea*	1 mixed	190	7				0.304	Bacilieri *et al.* 1994
allozymes	*Q. petraea*	81	120	8	2.629	0.228	0.257		Zanetto, Kremer 1995
allozymes	*Q. robur & Q. petraea*	1 mixed	183 *Q. r.* 166 *Q. p.*	4	4.3	0.25	0.34	0.16	Streiff *et al.* 1998
SSR	*Q. robur*	2	156	7	17.88	0.896		-0.022	Degen *et al.* 1999
SSR	*Q. robur*	7 mixed	174	6	18.86	0.792	0.868	0.083	Mariette *et al.* 2002
SSR	*Q. petraea*	7 mixed	168	6	20.22	0.831	0.878	0.047	Mariette *et al.* 2002
SSR	*Q. robur & Q. petraea*	1 mixed	183 *Q. r.* 166 *Q. p.*	6	21.7	0.81	0.87	0.07	Streiff *et al.* 1998
RAPD	*Q. petraea*	21	23	31			0.233		Le Corre *et al.* 1997
AFLP	*Q. robur*	7 mixed	174	155			0.195		Mariette *et al.* 2002
AFLP	*Q. petraea*	7 mixed	168	155			0.202		Mariette *et al.* 2002

microsatellite markers than allozymes, because of the frequent occurrence of null alleles at microsatellite loci (e.g. Callen *et al.* 1993).

2.2. Chloroplast DNA patterns at the European scale

2.2.1. Background

In tree species, organelle (mainly chloroplast) DNA markers have mostly been used to address phylogeographic[1] issues, i.e. dealing with the spatial distribution of differentiated lineages of a taxon over a large geographic area (e.g. Demesure *et al.* 1996, King & Ferris 1998, Dumolin-Lapègue *et al.* 1997, Petit *et al.* 2002a,b). In Gymnosperms in particular, organelle markers have additionally served for a wider scope of applications, for instance to investigate patterns of gene flow within populations (e.g. Latta *et al.* 1998), population genetic structure (e.g. Echt *et al.* 1998, Hayashi *et al.* 2000, Sperisen *et al.* 2001), or hybridisation between species (e.g. Bucci *et al.* 1998).

Several attributes of chloroplast DNA make it especially suitable for phylogeographic studies. Chloroplast DNA is haploid, non-recombining and predominantly uniparentally inherited (Birky 1988, Clegg 1989). It therefore evolves under a twofold higher pressure of genetic drift than the nuclear genome in hermaphrodite diploid plants (Birky *et al.* 1983). This allows for a strong differentiation of isolated lineages, for example during climatic cold stages. Further, in most angiosperms, chloroplast DNA is maternally inherited, i.e. transmitted only through seeds (Ennos 1994), and has thus more restricted dispersal than nuclear markers, which are transmitted by both seeds and pollen. This feature and its low mutation rate (Clegg *et al.* 1994) allow chloroplast DNA to retain historical imprints in space over long time periods.

The climate of the late Quaternary, which corresponds to approximately the last 700.000 years, was marked by a series of cold and dry glacial periods, each lasting approximately 100.000 years, interrupted by shorter periods (10.000-20.000 years) of warmer and moister climate (Webb & Bartlein 1992). During the cold stages, glaciers covered vast areas, for instance the Scandinavian ice sheet covered Britain and most of northern Europe throughout the last glacial cycle (115.000 BP -10.000 BP, Peltier 1994). The distribution ranges of plants and animals were then contracted to more southern locations (glacial refuges) and expanded northwards again as the climate improved (Comes & Kadereit 1998, Newton *et al.* 1999). Many tree species have reached their present distribution ranges only recently, i.e. a few thousand years ago. Owing to their long life cycles, only 100 to 1000 generations (depending

[1] The term 'phylogeography' was initially coined by John C Avise in 1987 to describe his findings from earlier mitochondrial DNA surveys in animals, namely that there was concordance among related organisms of phylogenetic relationships and spatial distribution of genotypes.

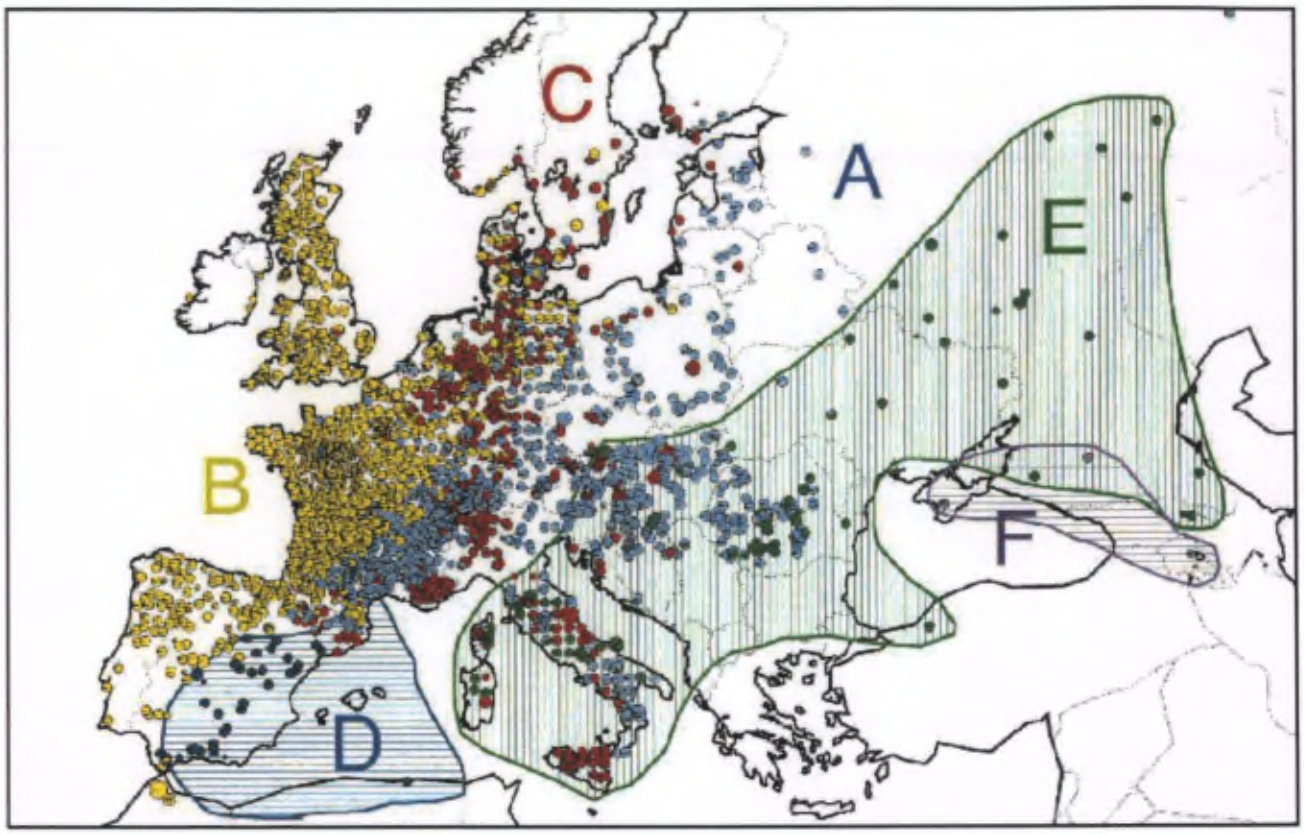


Fig. 1 Map of cpDNA lineages in 8 European oak species. The geographic distribution of the six cpDNA lineages identified (A-F) is provided. From Petit *et al.* 2002b.

on the species) have elapsed since the end of the last ice age and therefore their extant genetic structure often still contains historical imprints dating back to that period (Kremer 1994).

Chloroplast DNA surveys on a wide geographical scale allow to identify phylogenetic lineages within taxa and to retrace their postglacial recolonisation routes. They can suggest the location of glacial refuges by pinpointing high diversity areas, but cannot precisely locate them without evidence from fossil data, i.e. pollen or macrofossils (Taberlet *et al.* 1998, Comes & Kadereit 1999). Fossil material on the other hand can often not provide taxonomic resolution, e.g. palynological data didn't resolve oak species further as 'deciduous *Quercus*' (Brewer *et al.* 2002), but the necessary temporal resolution is provided through radiocarbon dating of fossil material. It appears that both methods are complementary, and molecular genetic markers were therefore jointly used with palynological tools to investigate the phylogeography of oak species (Petit *et al.* 2002a,b, Brewer *et al.* 2002).

2.2.2. Phylogeography of European white oaks: location of refuges and recolonisation routes

The studies of Petit *et al.* (2002a,b) describe phylogeographic patterns in European white oaks (eight species) based on chloroplast DNA (cpDNA) variation in over 2600 oak populations from all over Europe. Six haplotype lineages, noted from A to F were identified with the PCR-RFLP (polymerase chain reaction – restriction fragment length polymorphism) technique, the geographic distribution of which is shown in Fig 1. Chloroplast data have been interpreted in comparison with Brewer *et al.*' (2002) data on 'deciduous *Quercus*' (22 species).

The most important results are the following (Figs. 1 and 2):

- Existence of three main refuge areas in the three southern European peninsulas

Fossil pollen data from the end of the late glacial maximum 15000 (15 ka) BP identified three regions of Europe acting as full glacial refuges for oaks: (1) the southern Iberian Peninsula, (2) southern Italy and (3) the southern Balkan Peninsula with one site in Greece and one near the Bulgarian Black Sea border. These refuges were located in or near mountainous areas, supporting the idea that refuges for deciduous trees would have been situated at mid-altitude sites where the precipitation would have been higher than in the plains during the arid glacial period (see Brewer *et al.* 2002).

- Proximity of the Italian and Balkan refuges and existence of landbridges

Italy and the Balkan Peninsula share haplotypes from lineages A, C and E (Fig. 1). This is mainly because the sea level was about 100 m lower during the last glacial period than at present (Peltier 1994), leaving the northern part of the bed of the Adriatic Sea exposed, which resulted in exchanges of haplotypes between the Italian and Balkan refuges (Fig. 2). As a consequence, phylogenetically divergent haplotypes have sometimes followed similar colonisation routes, limiting phylogeographic structure to some extent. The colonisation of

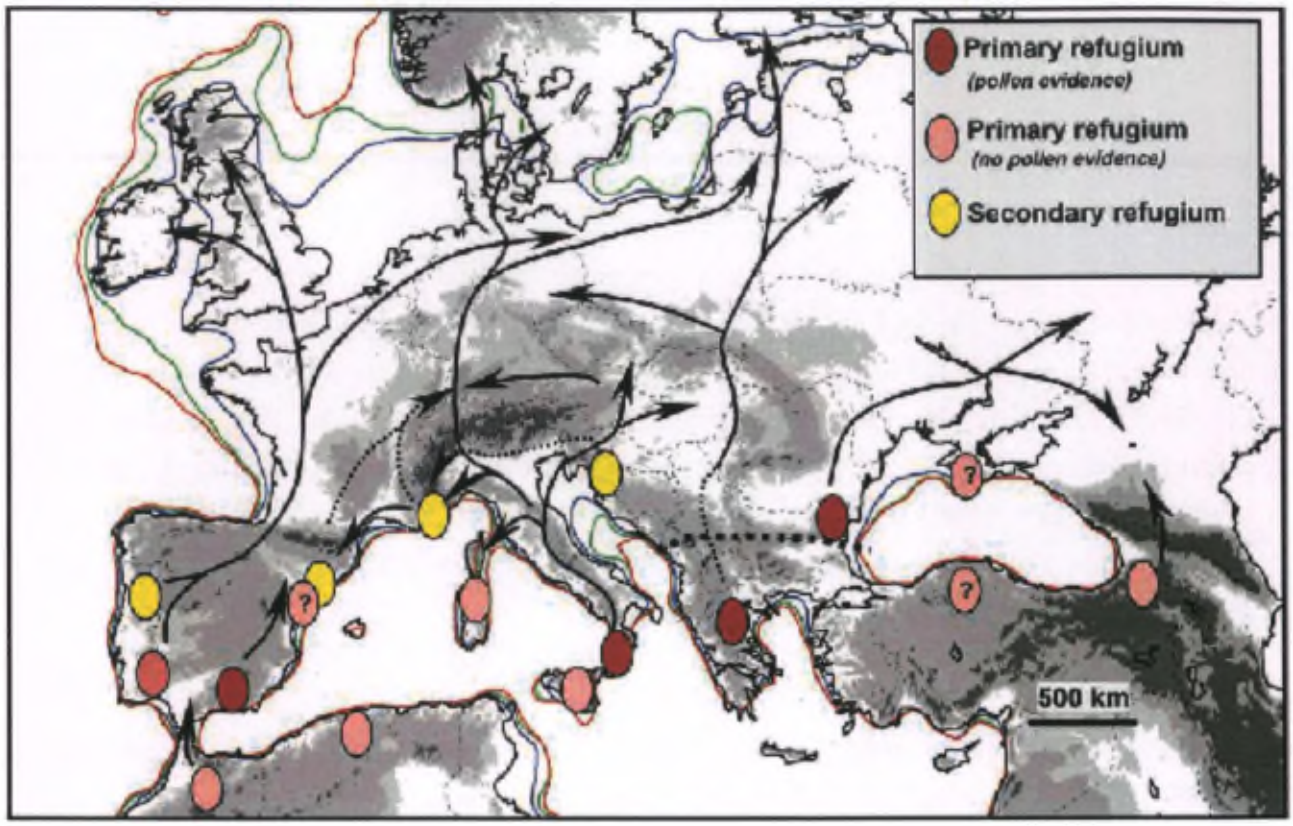


Fig. 2 Schematic distribution of primary and secondary oak refugia and post-glacial movements. From Petit *et al.* 2002a.

the British Isles through what is now the Channel is also an example for the existence of a landbridge.

- Existence of primary and secondary refuges

The climate change between the late glacial maximum and the onset of the Quaternary involved an initial warming to a thermal maximum between 13 and 12.5 ka BP. During this warming, oaks spread rapidly northwards from the full glacial or *primary* refuges. A short climatic deterioration was noted around 12 ka BP (the Older Dryas), followed by higher but declining temperatures until the severe cooling of the Younger Dryas (approximately 11 to 10 ka BP). Pollen records noted a decline of oak pollen deposition during this period. Oaks survived the Younger Dryas in *secondary* refuges located probably in mountainous areas south of a line running from the Pyrenees in the west to the Carpathians in the east (Fig. 2).

- Northward colonisation: the west to east orientation of mountain chains was sometimes a barrier

During the late glacial interstadial (i.e. before the Younger Dryas), haplotypes from Italy and the Balkans had reached the Po-plain, but could not spread farer north because the Alps represented a major obstacle. Instead migration continued westward, circumventing the Alps to reach secondary refuges. Today, oaks do not grow at altitudes higher than 1400 m in the Alps. However, during a climatic optimum at the onset of the Holocene with higher temperature than at present, oaks managed to cross the Alps via high altitude passes (more than 2000 m), and pollen records confirm their presence in the northern slopes by 9 ka BP. From there, Scandinavia was reached within 1000 years. On the other hand, haplotypes from lineage B from Iberia had circumvented the Pyrenees trough the Basque country (Fig. 2) and subsequently colonised their northern slopes (Fig. 1), central and western France, and areas further to the northeast up to Sweden. The south of Ireland and England was reached as early as between 9.5 and 9 ka BP. The present distribution range of oaks was reached around 6 ka BP.

- Discontinuities in the range of haplotyes and the influence of man

An example of range discontinuity is shown by haplotype 2 from lineage C, which occurs in Sicily and north of Rome, but not further south on the Italian peninsula. This suggests that, in this region, populations may have been overlooked, that they have gone extinct, or that the discovered haplotypes are distinct but could not be resolved with the techniques employed. Long-distance seed transport might also account for discontinuities in the range of haplotyes. Although acorns have constituted an important nutritional resource for man and his cattle, it is not likely that man was much implicated in the post-glacial recolonisation of oaks (König *et al.* 2001). More recent human influence is however confirmed by the identification populations carrying haplotypes of non-autochthonous origin (König *et al.* 2001).

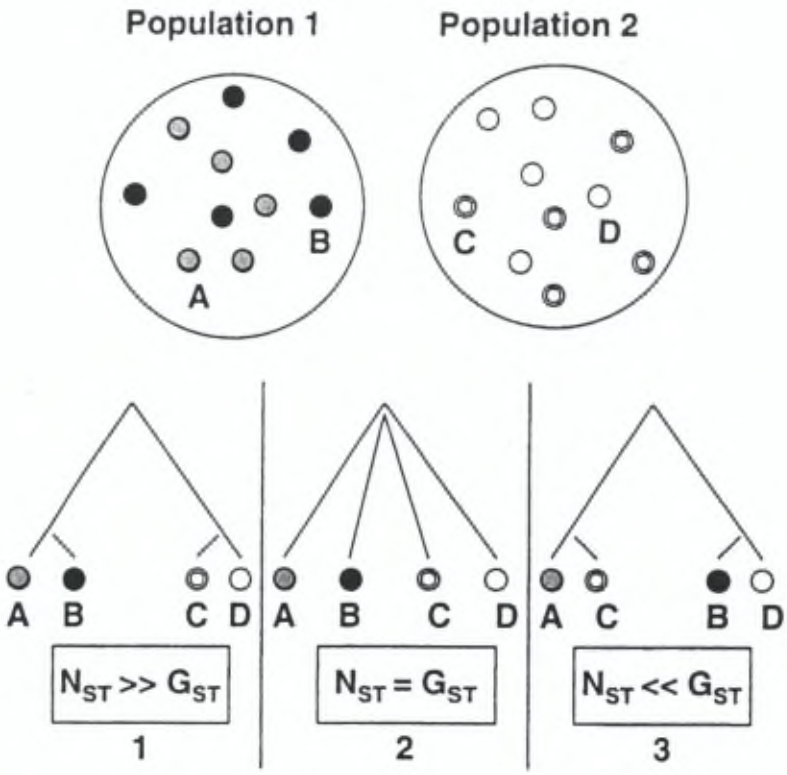


Fig. 3 Illustration of the correspondence between the phylogenies of the haplotypes and their geographic distribution. When there is correspondence (case 1), the differentiation measured by taking into account the similarities between haplotypes (N_{ST}) is larger than the differentiation based only on the frequency of the haplotypes (G_{ST}). When the haplotypes are equally related, $N_{ST} = G_{ST}$ (case 2). When the most strongly related haplotypes are never together but always found in different populations, $N_{ST} \leq G_{ST}$ (case 3). From Pons and Petit (1996).

Table 4 Levels of diversity and differentiation in *Quercus petraea*. From Petit *et al.* (2002b).

	No. Pops.	No. Haplo types	h_S	v_S		h_T	v_T		G_{ST}	N_{ST}	
Alpine region	121	6	0.062	0.075	ns	0.574	0.642	ns	0.892	0.884	ns
France	170	6	0.118	0.061	ns	0.712	0.539	ns	0.835	0.886	ns
Central Europe	169	9	0.154	0.174	ns	0.765	0.857	ns	0.799	0.797	ns
Great Britain	84	4	0.094	0.034	ns	0.605	0.239	ns	0.845	0.857	ns
North Balkans	97	11	0.139	0.070	**	0.792	0.593	ns	0.825	0.883	**
Northern Europe	27	8	0.107	0.068	ns	0.842	0.892	ns	0.873	0.924	*
Q. petraea	650	17	0.122	0.091	ns	0.847	0.835	ns	0.856	0.891	**

2.2.3. Phylogeography of European white oaks: patterns of diversity and differentiation

Genetic diversity of haplotype data can be evaluated by two methods: (1) by considering allele identity with the statistic h or (2) by taking into account the phylogenetic distance between alleles with the statistic v (Box 1, Pons and Petit 1996). The analysed samples display a phylogeographic structure when the among-population differentiation based on phylogenetic distance between haplotypes, N_{ST}, is significantly larger than the differentiation based on haplotype identity, G_{ST} (Pons & Petit 1996). This is the case when phylogenetically divergent haplotypes are preferentially distributed among populations rather than within populations (Fig. 3). However, a difference between N_{ST} and G_{ST} could also be caused by other factors, such as unequal effects of mutation rates or sampling artefacts (Pons and Petit 1996).

Table 4 summarises diversity and differentiation levels observed in *Quercus petraea*. It is immediately obvious that within-population diversity is low (mean h_S = 0.122, mean v_S = 0.091), but total diversity is high, resulting in strong differentiation among populations (typically $G_{ST} \geq 0.8$). Overall, phylogeographic structure is observed, as N_{ST} is significantly larger than G_{ST}. A refuge effect can be observed in the Northern Balkans. This region is characterised by a high number of haplotypes (11) and relatively high diversity h_S and h_T, but low phlyogenetic information content (v_S and v_T smaller than the corresponding h values). Central and northern Europe also display high numbers of haplotypes, but this time, v_T is higher than h_T (although not significantly so), indicating colonisation from different refuges. Great Britain, on the contrary, presents overall low diversity values reflecting colonisation from a single refuge.

At the level of the whole dataset, the tendencies outlined for *Q. petraea* become even clearer (Petit *et al.* 2002b). (1) Areas where refuges had been postulated such as the Iberian and Italian peninsulas generally display higher chloroplast diversity. (2) Superimposed on the 'refuge' effect, the authors found a 'mixing' effect: regions located at the junction of several colonisation routes, i.e. in central and northern Europe also displayed high diversity.

2.2.4. Mode of colonisation

It is currently believed that post-glacial recolonisation most likely occurred through 'leading edge dispersal', i.e. successive but rare events of long-distance dispersal, rather than by a compact wave of advance (Hewitt 1996, Dumolin-Lapègue *et al.* 1997, Petit *et al.* 1997, Comes & Kadereit 1998, Petit *et al.* 2002a). This mode of colonisation results in a patchy distribution of haplotypes at a local scale with low within-population diversity and high among-population differentiation, which is known both from observed (Petit *et al.* 1997) and from simulated data (Ibrahim *et al.* 1996, Le Corre & Kremer 1998). It logically involves successive population bottlenecks and a reduction of diversity at increasing distance from the refuge area. The first colonisers could establish an exponentially growing population, whereas

later migrants would reach the population already at carrying capacity, thus contributing comparatively little to the established diversity pattern.

The sharing of haplotypes among white oak species represents an interesting example for leading edge colonisation. Petit and colleagues (2002a) argue in favour of an early association of haplotypes to particular species (although not exclusively). Leading edge colonisation followed by pollen-mediated gene flow could account for the establishment of a second oak species in the newly colonised area through hybridisation, using the maternal type of the first colonisers and swamping out their nuclear genome through backcrossing (Petit *et al.* 1997, Petit *et al.* 2002a).

In contrast to this mode of colonisation, Austerlitz *et al.* (2000) have presented a model that proposes the arrival of a relatively high numbers of colonisers owing to the prolonged juvenile phase of tree species. This model can hardly explain the patchy distribution of maternal types observed in oaks, but allows for more realistic seed to pollen dispersal ratios (i.e. higher than 1:200, Ennos *et al.* 1994) by reducing substantially the founder effects.

2.2.5. Phylogeographic structure differs among species

In analogy to the example of European white oaks, most studies dealing with chloroplast genetic markers in temperate woody species revealed strong differentiation among populations and a relatively concordant geographical distribution of phylogenetically related haplotypes, resulting in a phylogeographic pattern. In Europe, the contribution of the different refuges is species-specific. In beech (Demesure *et al.* 1996) for instance, much of central Europe was colonised from a single refuge located in the Carpathians, and distinctive cpDNA haplotypes restricted to Italy suggest that beech was unable to cross the Alps during postglacial recolonisation. In hazel, on the other hand, most of Europe was colonized from a western refuge (Palmé and Vendramin 2002). Other European woody species where phylogeographic structure was detected include alder (King and Ferris 1998), ivy (Grivet and Petit 2002), Scots pine (Soranzo *et al.* 2000) and olive (Besnard *et al.* 2002). Some recent surveys of cpDNA over a wide geographic range were however unable to detect phylogeographic patterns. This was the case for sloe, *Prunus spinosa* (Mohanty *et al.* 2000), wild cherry, *Prunus avium* (Mohanty *et al.* 2001) and for wild service tree, *Sorbus torminalis* (Oddou-Muratorio *et al.* 2001). Low differentiation among populations was detected with cpDNA markers in these bird-dispersed early successional species, suggesting that mode of seed dispersal and successional behaviour have an impact on historical imprints of cpDNA.

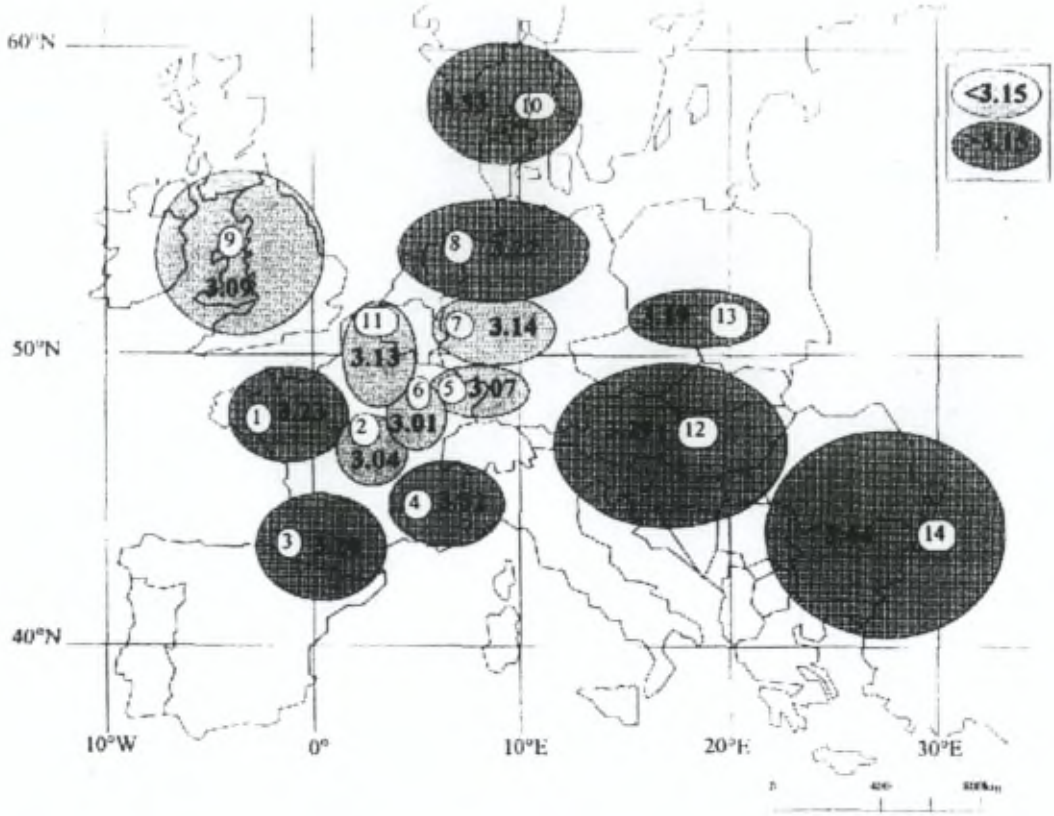


Fig. 4 Geographical distribution of mean number of alleles per population of *Quercus petraea*. Cut-off point was chosen *a priori* for illustration purposes. From Zanetto and Kremer (1995).

2.3. Structure of genetic diversity in Quercus petraea investigated with nuclear markers at the European scale.

Zanetto and Kremer (1995) investigated allozyme diversity in acorns from 81 populations of sessile oak from the whole distribution area. Diversity at the species level was high with an average number of alleles per locus of A = 4.77 and genetic diversity H = 0.265. Genetic differentiation among populations was very low, F_{ST} = 0.025, lower than the average for woody species (G_{ST}=0.084, Table 2). The comparison of F_{ST} for biparentally inherited markers to differentiation for maternally inherited chloroplast markers (G_{ST} = 0.865) illustrates the homogenising power of gene flow through pollen.

Despite the low level of differentiation identified by Zanetto and Kremer (1995), allelic richness and gene diversity varied among populations and variation was organised geographically at the European scale. Populations from the centre of the range (from the Loire to the Rhine river) displayed higher genetic diversity values than those from the edges of the distribution. For allelic richness, the trend was exactly opposite, with populations at the margins of the distribution comprising more alleles than those from the central part (Fig. 4). This pattern resulted in a positive correlation with longitude for allelic richness, and a negative one for genetic diversity.

The observed patterns of clinal variation in allele frequencies with longitude were concordant in seven out of eight loci, hence the authors concluded that they were not due to selection, but rather a consequence of postglacial recolonisation. To explain the geographic variation of allelic richness, the authors hypothesised loss of alleles in the northwestern part and/or gain of alleles on the edges of the distribution. They excluded the hypothesis of repeated bottlenecks bringing about a loss of alleles during recolonisation, because some northern populations exhibited high numbers of alleles. Selection against rare alleles on optimum sites (in northern France) was considered a possible explanation for the low allelic richness. A gain of alleles on the (mainly southern) edges of the distribution was attributed to an increase in the frequency of rare alleles above equilibrium values after recovery from bottlenecks. The antagonistic trends of variation in allelic richness and gene diversity could be explained if the loss of rare alleles was paralleled by an increase in frequency of an allele with intermediate frequency, and not the most common allele.

In their allozyme data set on 389 European populations of the common beech (*Fagus sylvatica*), Comps *et al.* (2001) observed a similar pattern of antagonistic trends in allelic richness and gene diversity over the distribution area as the authors of the sessile oak study. Comps *et al.* (2001) took up again the hypothesis of successive bottlenecks, producing a gradual loss of allelic richness with post-glacial recolonisation. They proposed that a concomitant increase in gene diversity might be due to combined effects of (i) admixture of

populations colonising from different refuges, (ii) selection against less viable, inbred individuals during population establishment and (iii) increased pollen flow during colonisation.

2.4. Comparison of chloroplast and nuclear patterns of variation at genetic marker loci

A recent study by Kremer *et al.* (2002) tested the hypothesis that common refuge origin may result in a correlation between cpDNA variation and nuclear variation. The authors performed Mantel tests to investigate the correlation between the matrix of chloroplast genetic distances and the matrix of standard genetic distances for either allozymes or RAPDs.

According to Kremer *et al.* (2002), cytonuclear disequilibria were established through genetic drift during the glacial period. Drift would have been strong enough to differentiate populations among refuges for both chloroplast and nuclear DNA, because populations were small and isolated for a long time. During recolonisation, cytonuclear disequilibria would have been erased to a large extent by at least 2 forces: (i) gene flow, homogenising nuclear genes more strongly than chloroplast genes, because of more extensive pollen flow, and (ii) selection, acting on specific gene products. Gene flow was likely to be the only force in action for allozymes and RAPDs, as they are supposedly neutral markers. Kremer *et al.* (2002) found significant cytonuclear disequilibria for both RAPDs and allozymes when not correcting for geographic distance, a result approximately equivalent to the detection of allele frequency variation according to geographical position found by Zanetto and Kremer (1995). After correction for geographical distance, however, the correlation with chloroplast genetic distances remained significant for RAPDs, but not for allozymes. Hence, Kremer *et al.* (2002) demonstrated that postglacial gene flow had been able to break up the linkages between chloroplast and nuclear genes in the case of allozymes, and they also revealed some influence of selection pressures on RAPD markers.

2.5. Comparison of chloroplast patterns of variation with nuclear controlled phenotypic traits

In oaks like in most plant species, phenotypic traits exhibit higher differentiation among populations than allozymes; for instance in sessile oak, differentiation coefficients for bud burst and height growth were generally higher than 20% (Kremer *et al.* 1997) compared to 2.5% for allozymes (Zanetto & Kremer 1995). Kremer *et al.* (2002) assessed the correlation with chloroplast lineage in four categories of traits: survival, growth, stem form and phenology. Only six out of 62 trait x provenance trial combinations were correlated with chloroplast genetic distances, and only two of them remained significant after correction for geographic distance. Vice versa, correlation with geographic distance was significant for 13

trait combinations, and all remained significant after correction for chloroplast genetic distance. This indicates that traits were more closely correlated to geographic distance than to chloroplast genetic distance. The 13 significant correlations all concerned phenology and none involved growth traits or form. This confirmed earlier results (references in Kremer *et al.* 2002) that phenological traits are under photoperiodic or heat sum control and exhibit geographic patterns at large spatial scales, whereas growth and form, which are influenced by local soil and competition conditions, show patterns at a microscale level.

According to Kremer *et al.* (2002), the observed patterns are due to local selection pressures acting on recolonised populations since their establishment and differentiating them increasingly over time for phenotypic traits. The extant differentiation among populations at phenotypic traits is therefore very different from the one of ice age refuges, and significant association with maternal lineages is no longer observed.

3. Materials and Objectives

This thesis was initiated in the framework of the cooperation project *GENFOR: Genetic Resources of Broadleaved Forest Tree Species in Southeastern Europe*, financed by the Luxembourgish Ministry of Finance and coordinated by the International Plant Genetic Resources Institute (IPGRI). The financial support to this project by the Luxembourgish Government can be seen as an important contribution to the implementation of Strasbourg resolution S2 on forest genetic resources. Partners in the first phase of the project were the Forest Research Institute in Sofia, Bulgaria; the Institute of Botany in Chisinau, the Republic of Moldavia; the ICAS Forestry Administration, Bucharest, Romania; and the Public Research Centre – Gabriel Lippmann, Luxembourg. The main objectives of the project were to carry out new forest inventories in southeastern European countries to investigate distribution and health status of forest tree species and to organise exchanges of researches for training and common research activities in the fields of *in vitro* propagation, adaptation to physiological stresses and population genetics of forest tree species. The present work was initiated from the latter subject and focuses on the geographic structure of genetic diversity in the common ash (*Fraxinus excelsior* L.), mainly in southeastern Europe, but also in the context of diversity patterns in other European regions.

3.1. Forest resources in southeastern Europe

Southeastern Europe is a rich area with respect to forest genetic resources because it contains a variety of ecological regions. The climate in the Balkan Peninsula is mainly continental, but mediterranean on the west coast, and the topography is highly variable with the mountain chains of the Carpathians reaching from Slovakia to Romania, the Balkans in central

Fig. 5 Topographic map of the Balkan Peninsula

Bulgaria, the Rhodopy mountains in southern Bulgaria, the Dinaric Alps in the area of former Yugoslavia and in Albania and the Pindos in Greece; interspersed with large plains in Hungary and the north of former Yugoslavia (rivers Danube, Save and Tisa), in southern Romania and northern Bulgaria (river Danube) and in the Republic of Moldavia (river Dniestr, Fig. 5). As a result, the species diversity of forests is high in these regions because they contain different forest ecological zones, comprising broadleaved deciduous forests, mixed ecotonal forests with conifers as well as evergreen mediterranean forests (Longauer 1999). These ecosystems, which are often less fragmented than in western Europe, provide habitats for a large variety of animal species some of which have gone extinct in their western range (i.e. bears in the Pyrenees or lynxes in Germany, Klimo *et al.* 1999).

Another reason for particularly high species diversity of forest trees in southeastern Europe is that the European distribution of broadleaved woody genera is sometimes characterised by one (or few) common species distributed over most of Europe, and additionally one (or few) species restricted to the southeastern part. This is for instance the case for beech with *Fagus sylvatica* commonly occurring all over Europe *vs. F. orientalis* restricted to southeastern Europe, for hornbeam with *Carpinus betulus vs. C. orientalis*, and for hazelnut with *Corylus avellana vs. C. colurna* and *C. maxima* (Tutin *et al.* 1964). This high species diversity in the Balkan Peninsula compared to the Italian and Iberian Peninsulas might partly be due to the proximity of Turkey and the near and middle East. These areas may have acted as a "buffer zone" during range contractions in the glacial periods of the Quaternary, offering areas of favourable climate for the retraction of species, while species retracting themselves to the other peninsulas may have gone extinct because of the presence of the Mediterranean Sea.

The Balkan Peninsula has held refuge areas for many tree species during the glaciations of the Quaternary (Huntley & Birks 1983) and is therefore expected to harbour also high within-species diversity (Hewitt 1996). Several southeastern European sites, mainly mountainous ones, are known to have acted as glacial refuges from fossil pollen (Huntley & Birks 1983, Brewer 2001), and they have been confirmed with molecular data, mostly chloroplast DNA surveys, e.g. the Carpathians for beech (*Fagus sylvativa*, Demesure *et al.* 1996) and black alder (*Alnus glutinosa*, King & Ferris 1998); Croatia and Bulgaria for silver fir (*Abies alba*, Konnert & Bergmann 1995, Vendramin *et al.* 1999); northern Greece for ivy (*Hedera* sp., Grivet & Petit 2002) and northern Greece and the Black Sea coast for oaks (*Quercus* sp., Petit *et al.* 2002a).

The countries of focus in our study are Bulgaria, Romania and the Republic of Moldavia. The most common forest type encountered there is broadleaved deciduous forest, making up more than 50 % of the forest area, followed by coniferous, and then mixed forest. Almost all European forest have been managed for hundreds of years, but a recent survey reports that in this region, and particularly in Bulgaria, the actual forest composition corresponds quite closely to what is expected to occur there naturally (Longauer 1999). Also, in contrast with

Fig. 6 (a) Flowering common ash trees in the Mamer river floodplain forest (Luxembourg); (b) imparipennate leaves and samaras; (c) flushing juvenile, note the black bud scales; (d) opening male inflorescences.

western Europe where several countries have faced severe deforestation at the end of the Middle Ages, forests have been continuously present in most eastern European countries, and large areas of semi-natural forests still predominate especially in the Carpathians and in the Balkans (Klimo *et al.* 1999).

In Bulgaria, Romania and the Republic of Moldavia, forests are mainly managed as even-aged communities according to shelterwood systems, i.e. part of the mature forest is left intact as a light canopy cover after felling of a large area, allowing for natural regeneration (Longauer 1999). Natural regeneration from seed is indeed the largely prevailing regeneration practice in Romania and Bulgaria (there are no data for the Republic of Moldavia) with over 50% of the total regeneration, distinguishing these countries clearly from the near Hungary and Slovakia, where most regeneration occurs by planting (Longauer 1999). It is noteworthy that in Bulgaria, more than 30% of the forest cover is coppice forest, whereas this management form makes up only a few percent in Romania (there are no data for the Republic of Moldavia, Longauer 1999). Further, large forest areas in our focal countries (more than 30%) have been designated for special functions such as water and soil protection or conservation and are therefore managed at low intensity (i.e. the semi-natural forests mentioned before, Klimo *et al.* 1999, Longauer & Krajmerova 1999).

The previously stated features of southeastern European forests suggest that they are likely to represent a more close-to-nature situation than most western European forests. They should therefore contain genetic resources that are suitable for the application of molecular markers in order to investigate population genetic structure in forest tree species and to infer of relevant biological information concerning for instance tree species life history characteristics (e.g. Milligan *et al.* 1994).

3.2. Common ash (*Fraxinus excelsior* L.)

Common ash (*Fraxinus excelsior*, Oleaceae) is the tallest of four native ash species in Europe (Tutin *et al.*, 1972). Just as *F. angustifolia* Vahl and *F. pallisiae* Wilmott, it is wind-pollinated and belongs to section Fraxinus of the genus, whereas the insect-pollinated *F. ornus* belongs to section Ornus (Wallander 2001). Common ash is a tree of 25-35 (40) m, with grey bark, black buds and imparipennate leaves with 7-13 leaflets (Fig. 6). Its mean longevity is 150-200 years.

3.2.1. Distribution

The natural distribution of common ash covers large parts of Europe (Fig. 7), ranging from northern Spain to southern Scandinavia and from the shores of the Atlantic in the west to the Volga river in the east. In southeastern Europe, the southern limit of common ash is in

Fig. 7 *Fraxinus excelsior* distribution map (from EUFORGEN Noble Hardwoods Network).

northern Greece and in northern Turkey; it is present around the Black Sea and as far east as the Caspian Sea and northern Iran.

3.2.2. Ecology

Common ash is a postpioneer species, exhibiting properties intermediate between typical pioneer woody species (birch, willow, poplar) and permanent forest components (oaks, beech, Giertych 1995). It readily colonises fertile, temporarily forest-absent habitats (forest clearings, roadside and streamside slopes and ravines) but its competition ability is high only when its ecological requirements are met. Common ash mainly occurs in floodplain forests, where it can represent a high percentage of the canopy, associated with black alder and/or elms (Thill 1970, Claessens *et al.* 1994). It also occurs in fertile mixed broadleaved forests associated with mainly oak, hornbeam and maple (Thill 1970, Claessens *et al.* 1994), but its distribution is discontinuous and scattered and it is not as common as the social species oak and beech (Giertych 1995, Pliura 1999). The occurrence of ash in habitats with contrasting features, such as wet lowland sites as opposed to dry chalk-rich sites suggested the existence of differentiated soil-ecotypes, which could however not be confirmed experimentally (Weiser 1995 and references therein). Common ash best grows in fertile, fresh and profound soils, with high organic matter content and good moisture availability.

3.2.3. Mating system

Common ash is a polygamous species; it shows continuous phenotypic sex expression from male via hermaphrodite to female individuals (Picard 1982, Lamb & Boshier 1994, Binggeli & Power 1999, Wallander 2001). Five basic types of flowers [(i), male with 2 stamens; (ii) male with 2 stamens and a rudimentary pistil; (iii) hermaphrodite; (iv) female with a pistil and rudimentary stamens; and (v) female with a pistil only; Binggeli & Power 1999] combine into four types of inflorescences (male; mixed male and hermaphrodite; hermaphrodite; and female; Wallander 2001). Their occurrence in individuals allows to define the three classes of phenotypic gender expression, although some variation was noted between individuals and years (Wallander 2001). Flowering occurs before leafing in spring, as in most wind-pollinated trees. Hermaphrodite flowers are self-compatible, but they may avoid self-pollination by being protogynous (Wallander 2001). For instance, Morand *et al.* (2002) stated self-fertilisation as one of the possible explanations for the high values of Wrights inbreeding coefficient (F_{IS} = 0.16 in seedlings and F_{IS} = 0.29 in adults) that they observed in a microsatellite study of 12 French populations. The fruiting pattern of common ash is irregular and synchrony among individuals indicates that it is probably triggered by some environmental component (for instance early leafing in the preceding year, Tapper 1992, 1996). The syncarpous ovary contains two ovules in each of two locules, but only one develops after fertilisation. The fruit is a one-seeded samara (Fig. 6 b), which is mainly wind-dispersed, although some authors argue in favour of additional bird-mediated dispersal to

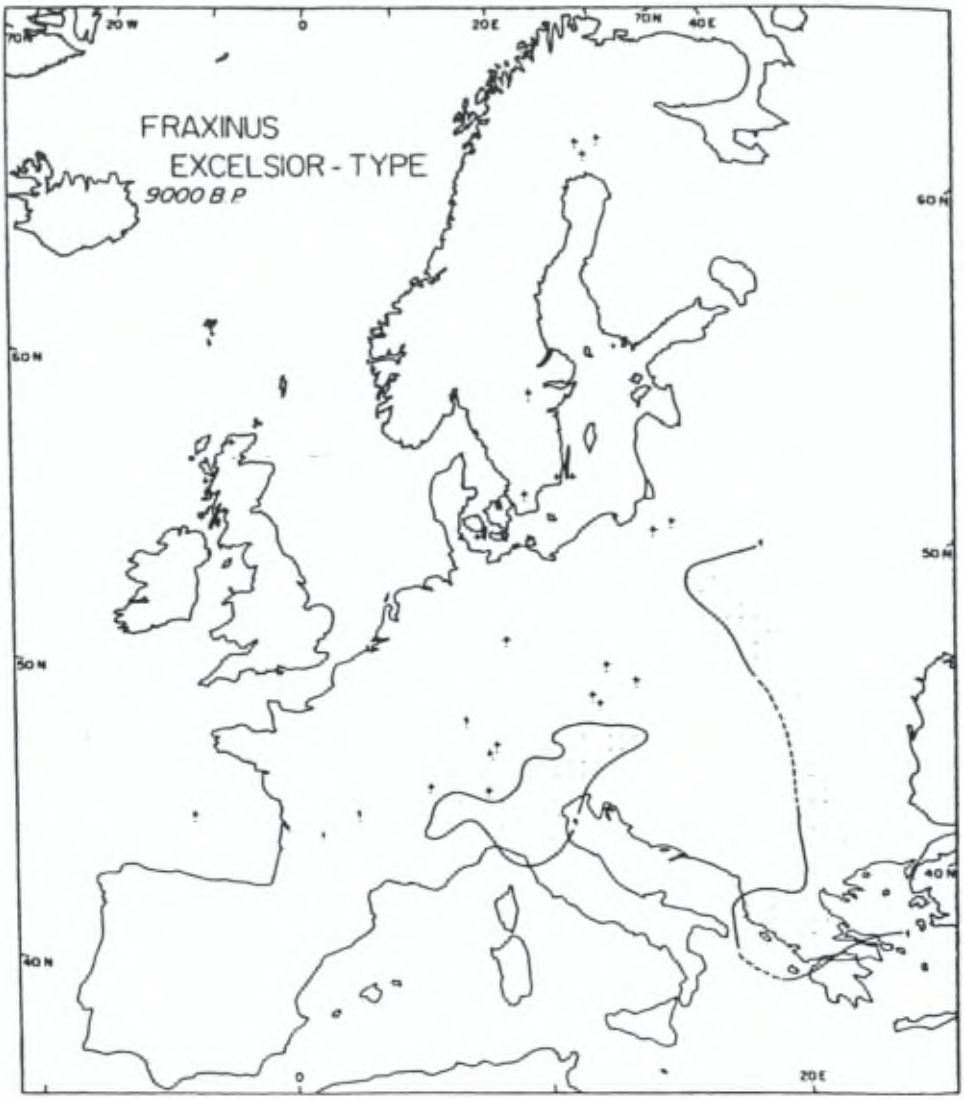


Fig. 8 Distribution of *Fraxinus excelsior* – type pollen deposits 9.000 radiocarbon years ago. From Huntley and Birks (1983).

explain features of postglacial recolonisation (Wilkinson 1997) and the strong colonising behaviour of ash (Giertych 1995).

3.2.4. Location of glacial refuges and postglacial recolonisation history inferred from fossil pollen evidence

Three species belong to the *Fraxinus excelsior*-type pollen taxon: *Fraxinus excelsior*, *F. angustifolia* and *F. pallisiae* (Huntley & Birks 1983). Two subspecies are recognised in *F. excelsior*, ssp. *excelsior*, which occurs throughout the range of the species and ssp. *coriariifolia*, which is distributed in southeastern Europe and Iran (Tutin *et al.* 1972). *Fraxinus angustifolia* occurs in southern and south-central Europe; it also contains two subspecies, with non-overlapping ranges, ssp. *angustifolia*, spread in the western mediterranean region and Portugal, and ssp. *oxycarpa*, which occurs in eastern-central Europe and in southern Europe, from northeastern Spain eastwards (Tutin *et al.* 1972). *Fraxinus pallisiae* is distributed in southeastern Europe, from Turkey to Moldavia.

As *F. excelsior* has a poor pollen representation, Huntley and Birks (1983) suggested that values above 1% in pollen cores would indicate local presence, whereas values above 5% would be associated with woodlands in which ash is dominant or co-dominant. These authors reported the presence of ash pollen during the last glacial cycle (115 ka BP – 10 ka BP) in Greece and Italy. In the early Holocene (10 ka BP), they documented expansions of ash from a southeastern European and an Alpine centre. Large amounts of pollen were recorded around 9 ka BP in eastern central Europe (Fig. 8) and interpreted either as a rapid expansion of the southern centre or, alternatively, as an expansion from a putative northeastern European refuge. These centres subsequently merged and expansion continued, reaching England, Denmark, the southern Baltic and western France by 6 ka BP. By this time, a possible fourth centre became apparent in the Pyrenees. Since 5 ka BP, ash became more and more abundant in western Europe, especially in the British Isles, while a retreat was observed in eastern Europe. The decline in the east was attributed to forest clearance and/or the immigration of taxa which expanded later (e.g. *Carpinus betulus*). Another possible reason was a reduction of the precipitation/evaporation ratio in that area, as *Fraxinus excelsior* favours moist soils (Huntley & Birks, 1983). According to these authors, the highest migration rates during recolonisation were observed in the period up to 6000 BP, with values comprised between 200 and 500 m per year.

A more recent survey based on a greater density of pollen records (Brewer 2001) identified sites with more than 5% ash pollen representation 12 ka BP at the Bulgarian Black Sea coast and in southern Italy near Naples (Fig. 9). These are probably full glacial refuges. Additional sites with ash pollen representation above 1% (i.e. local presence of ash *sensu* Huntley & Birks 1983) were located in central Portugal in the Serra da Estrela, in the French Massif Central, in central Italy in the Apennines, in the eastern Alps at the German-Austrian border,

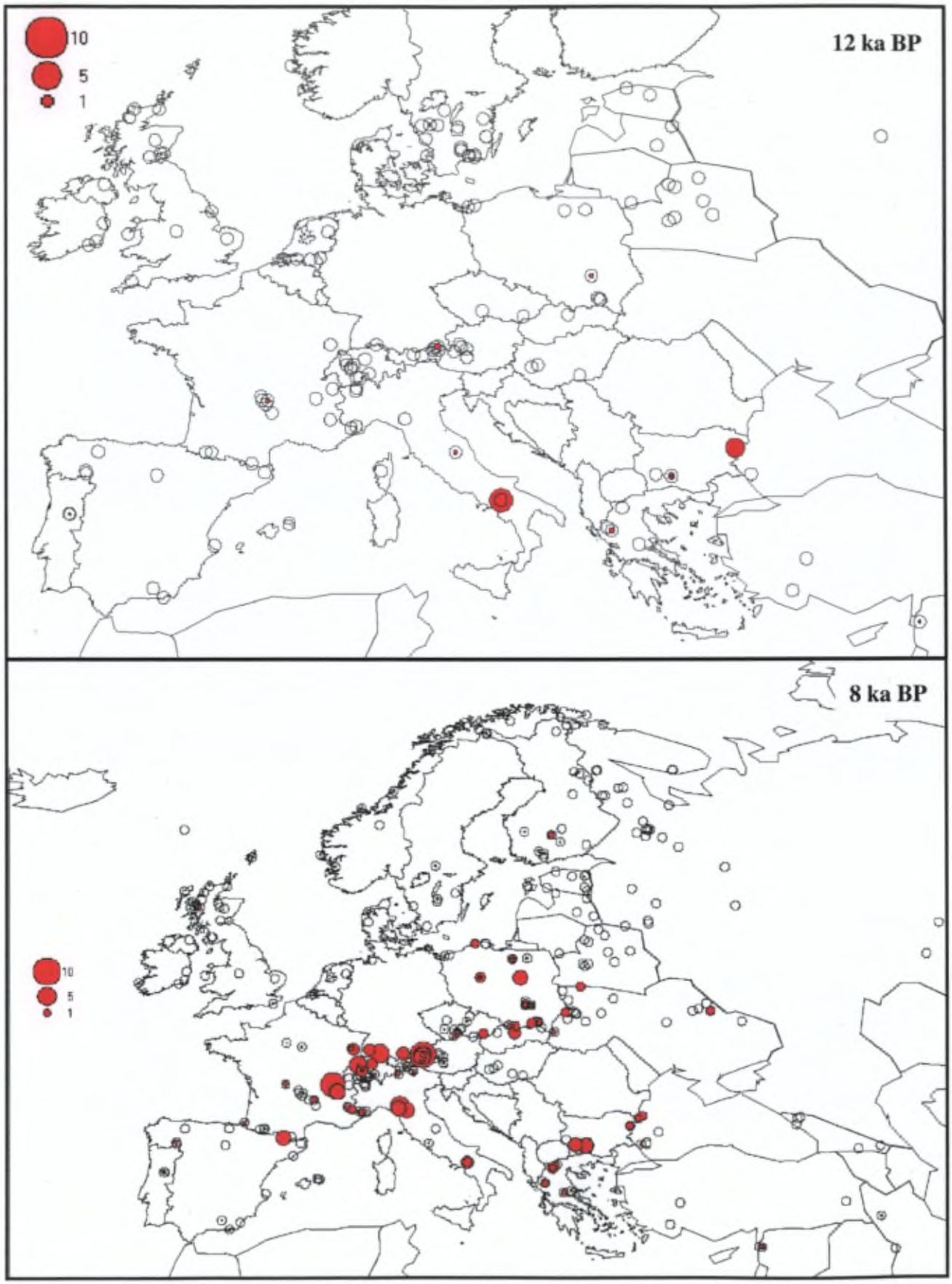


Fig. 9 Distribution of *Fraxinus excelsior* – type pollen deposits 12.000 and 8.000 radiocarbon years ago. From Brewer (2001).

in the Greek Pindos mountains, in the Bulgarian Rhodopy mountains, just north of the Carpathians in southeastern Poland, in Scotland at the northern coast of the Firth of Forth and in southern Sweden near Göteborg. Based on Peltier's (1994) maps on deglaciation history, the Scottish and Fennoscandian ice sheets must have freed the two latter areas only very recently before pollen deposition. Little change was observed in fossil pollen maps until after 10 ka BP, the end of the Younger Dryas cold period. Ash then spread north and west from the refuges and by 8 ka BP it had reached the Pyrenees, it was abundant at the northern slopes of the Alps and in Poland, and it was present, but to a lesser extent, in the Baltic States and in southern Scandinavia (Brewer 2001, Fig. 9). Ash had fully colonised the British Isles around 4,5 ka BP, after which a decline of ash became evident over large parts of the range, but mainly in the eastern part (Brewer 2001). According to Huntley and Birks' (1983) maps, a minimum of ash occurrence in Europe took place around 2000 BP; thereafter pollen values increased again in Denmark, southwest Sweden and western France. Today, ash pollen is recorded mainly in northwestern Europe, the highest values being reported in England. It is also found in small quantities in central Europe, mainly in upland (but not mountain) areas (Huntley & Birks 1983).

Another recent fossil pollen survey by Gliemeroth (1997) located the Iberian refuge for ash rather in the south of that Peninsula; it confirmed the refuges situated in the Alps and the Balkan Peninsula suggested by the previous authors; and it proposed additional refuges for ash in the Turkish mountains and at the northern Black Sea border.

Summarising the results from these three pollen surveys, it appears that there is strong support for full glacial refuges for ash in the Alps and the Balkan Peninsula, and weaker support for refuges located in Iberia, Turkey and at the northern Black Sea coast. Early occurrence of ash was reported in some other sites in France, Scotland, Sweden and Poland and their role as putative primary or secondary refuges deserves further attention. The colonisation of the European continent was first marked by an east-west migration, followed by a south-north trend, and the distribution limits were reached between 6000 and 4000 BP.

3.2.4. Geographic patterns of variation.

Not much information is available on geographic variation in common ash. One study assessed genetic diversity in French populations with nuclear microsatellite markers (Morand *et al.* 2002); and first surveys of phenotypic markers have been reported from two recently established provenance trials describing, respectively, patterns of variation in central Europe (Kleinschmit *et al.* 1996) and Scandinavia (Baliuckas *et al.* 2000). I will briefly review the results of these works here.

In their study of twelve common ash populations from northeastern France with five microsatellite loci, Morand *et al.* (2002) found high within-population gene diversity, H_E = 0.900 ± 0.047 (SD), as expected for a wind-pollinated tree species (compare with values for

oak, Table 4, p.16). A low proportion of the total diversity was distributed among populations, as expected, $F_{ST} = 0.043 \pm 0.017$ (SD), and no correlation was observed between geographic distance and genetic differentiation.

Kleinschmit *et al.* (1996) reported results on a total of 52 provenances, from Germany (43), Switzerland (3), Austria (4) and Romania (2), established in a total of 26 test sites in Germany (19), The Netherlands (2), Belgium (2), France (2) and Romania (1). Analysed traits included bud burst, bud set and damages due to late frost (recorded in the nursery); as well as losses, height at age 6 or 7 and shape (recorded on test sites). All characters displayed extensive ecotypic variation patterns at local geographic scale, and generally, differences within provenances were as large as among provenances. Some geographic trends of variation were identified, notably for bud burst and bud set (growth cessation), which are important adaptive characters. Dates of bud burst and bud set were however not correlated to each other. Forty percent of the variation in bud burst was explained by longitude, whereas latitude and longitude together explained 27% of the variation in bud set. Romanian provenances flushed early and set buds late, and therefore suffered from early and from late frosts in Germany. This also resulted in bad stem form because terminal buds were killed by frost. Height was strongly correlated to the date of bud set (i.e. length of the growth period), and decreased with increasing latitude.

Baliuckas *et al.* (2000) analysed 20 to 29 open-pollinated families from four Swedish provenances of common ash, sampled along a latitudinal gradient and established in two test sites. They assessed bud burst, growth cessation and height in saplings at ages 3, 4 and 5 years. They found a negative relationship between growth cessation and latitude, confirming the results of Kleinschmit *et al.* (1996). Large variation was identified among families within provenances, and the northernmost provenances exhibited as much variation as the others, which conflicts with the expectation that populations at the distribution edges should present reduced diversity because of the occurrence of bottlenecks and genetic drift (e.g. Hewitt 1996, Soltis *et al.* 1997).

3.3. Objectives

The central objective of this thesis is to investigate which factors shaped the spatial structure of genetic diversity in the common ash and how strong are their relative contributions. The region of main interest is the Balkan Peninsula, and the results are expected to lead to the formulation of recommendations in the context of management and conservation of common ash populations. The genetic diversity of common ash was analysed at different spatial scales with a series of molecular tools. The work was divided into five parts which are briefly introduced here and presented in the form of scientific manuscripts in the following chapters:

1. Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.)

   The main objectives of this chapter were to provide a first measure of the level of genetic marker diversity in the species using nuclear microsatellite markers and to investigate the partitioning of diversity at three geographic scales in Bulgaria, (i) within populations, (ii) among populations within natural regions of Bulgaria, and (iii) among regions. Biological characteristics of the species, such as the level of inbreeding and the variance of gene dispersal distance were inferred.

2. Estimating seed versus pollen dispersal from spatial genetic structure in the common ash

   This part deals with the estimation of the variance of gene dispersal distance within a large Romanian common ash population. The aim was to gather information on the relative contributions of seed and pollen to total gene dispersal by comparing the decrease of kinship between individuals against distance in the study population with data from simulations on an analogous population, where seed and pollen dispersal were under control.

3. Nuclear microsatellites reveal contrasting patterns of diversity in western and eastern European populations of the common ash (*Fraxinus excelsior* L.)

   In this chapter, patterns of genetic diversity were investigated within and among 36 European common ash populations with nuclear microsatellite markers. Data were analysed with classic statistical tools and with recently developed approaches which exploit the high levels of polymorphism characteristic for microsatellites. The objective was to explain the diversity patterns in the context of population demography and postglacial recolonisation history, with particular emphasis on the inference power of the new analysis methods.

4. Different evolutionary history and geographical distribution of even- and odd-sized alleles at a microsatellite locus in *Fraxinus excelsior* (L.)

   In the genetic diversity surveys in Bulgaria and in Europe, even- and odd allele sizes were observed at the microsatellite locus FEMSATL4 in common ash. The purpose of this chapter was to investigate the molecular diversity underlying this electrophoretic variation and to regard it in the context of the evolutionary history of the locus by sizing and sequencing amplification products from four species at increasing phylogenetic distance from common ash.

5. Incongruence between chloroplast and nuclear marker data among European populations of common ash

The aim of this part was to identify patterns of chloroplast genetic diversity at the European scale and compare them to the patterns previously obtained at nuclear markers in the context of the separation of populations in ice age refuges and subsequent postglacial recolonisation. The utility of using two differentially inherited types of genetic markers for the inference of population history is assessed.

Chapter I: Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.)

M. Heuertz, J.-F. Hausman, I. Tsvetkov, Nathalie Frascaria-Lacoste and X. Vekemans.

Abstract

We analysed genetic variation within and between populations of the common ash from Bulgaria in order to extract biological information useful in the context of conservation management of eastern European genetic resources of noble hardwood species. A total of 321 trees from three regions of Bulgaria were typed at six highly polymorphic microsatellite loci. Analysis of within-population inbreeding suggests an upper bound value of 2.7% for the selfing rate. Significant spatial genetic structure consistent with models of isolation by distance was detected within four out of ten populations as well as among populations. Estimates of neighbourhood size in the range 38 to 126 individual trees were obtained based on spatial genetic structure analyses at either the intrapopulation or interpopulation level. Differentiation among populations explained only about 8.7% of total genetic diversity. These results are discussed in comparison with data from social broadleaved species such as oak and beech.

Molecular Ecology (2001) 10, 1615–1623

Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.)

MYRIAM HEUERTZ,*† JEAN-FRANÇOIS HAUSMAN,* IVAYLO TSVETKOV,‡ NATHALIE FRASCARIA-LACOSTE§ and XAVIER VEKEMANS†

*Centre de Recherche Public-Gabriel Lippmann, CREBS Research Unit, 162a, avenue de la Faïencerie, 1511 Luxembourg, Luxembourg, †Université Libre de Bruxelles, Laboratoire de Génétique et Ecologie Végétales, 1850, chaussée de Wavre, 1160 Bruxelles, Belgium, ‡Forest Research Institute, 132, Kliment Ohridski Blvd., 1756 Sofia, Bulgaria, §ENGREF, Laboratoire Evolution et Systématique, UPRESA CNRS 8079, Bât 362, Université Paris-XI, F-91405 Orsay cedex, France

Abstract

We analysed genetic variation within and between populations of the common ash from Bulgaria in order to extract biological information useful in the context of conservation management of eastern European genetic resources of noble hardwood species. A total of 321 trees from three regions of Bulgaria were typed at six highly polymorphic microsatellite loci. Analysis of within-population inbreeding suggests an upper boundary value of 2.7% for the selfing rate. Significant spatial genetic structure consistent with models of isolation by distance was detected within four out of 10 populations as well as among populations. Estimates of neighbourhood size in the range 38–126 individual trees were obtained based on spatial genetic structure analyses at either the intrapopulation or interpopulation level. Differentiation among populations explained only about 8.7% of total genetic diversity. These results are discussed in comparison with data from social broad-leaved species such as oak and beech.

Keywords: *Fraxinus excelsior*, genetic structure, inbreeding, microsatellites, neighbourhood size, southeastern Europe

Received 19 October 2000; revision received 1 February 2001; accepted 1 February 2001

Introduction

Forest genetic resources are threatened by a broad array of factors such as climate change, environmental pollution, habitat destruction, competition and pest attacks. The capacity of forest tree species to survive these threats and to persist in spatially and temporarily heterogeneous environments is dependent on their adaptive potential, which is determined by the within-species genetic diversity. This is why conservation strategies for forest trees should be based on an evolutionary approach and focus on the maintenance of within-species genetic diversity (Geburek 1997; Eriksson 1999). Noble hardwood species (e.g. *Fraxinus, Acer, Tilia, Castanea, Juglans, Prunus, Alnus, Ulmus*) have a scattered distribution within mixed forests owing to their limited interspecific competition ability and to their rather narrow ecological plasticity. The conservation and promotion of these tree species is most promising *in situ*, within the frame of forestry management (Rotach 1999), with the aim of combining gene conservation efforts and the production of high quality timber. As conservation programmes of noble hardwood species are built up (Turok *et al.* 1996, 1998, 1999), it appears that information on crucial life-history characteristics of these species (e.g. mating system, pollen and seed dispersal) and on their patterns of genetic variation is still lacking. In this context, the application of molecular genetic techniques to the conservation of noble hardwood species is anticipated to bring valuable data that could be used to extract relevant biological information (Milligan *et al.* 1994), to document hotspots of genetic diversity, and to infer their phylogeography (Newton *et al.* 1999).

Forests from southeastern Europe constitute highly valuable reservoirs of genetic resources of noble hardwood

Correspondence: Myriam Heuertz. Centre de Recherche Public-Gabriel Lippmann, CREBS Research Unit, 162a, avenue de la Faïencerie, L-1511 Luxembourg, Luxembourg. Fax: + 352 470261 389; E-mail: heuertz@crpgl.lu

© 2001 Blackwell Science Ltd

species for two reasons. Firstly, many forest stands have been maintained in a seminatural status because of the low intensity of economic exploitation (European Commission Phare Programme 1999). Secondly, southeastern Europe has been identified as an important refuge for plant species during the Quaternary glacial periods (for reviews see Comes & Kadereit 1998 and Newton *et al.* 1999). For *Abies alba*, allozyme (Konnert & Bergmann 1995) and chloroplastic microsatellite (cpSSR) markers (Vendramin *et al.* 1999) revealed high-diversity potential refuges in the southern and southeastern Balkan peninsula and in the Croatian Balkans. Chloroplast DNA restriction fragment analysis revealed that central and western European populations of *Fagus sylvatica* (Demesure *et al.* 1996) and *Alnus glutinosa* (King & Ferris 1998) may have evolved exclusively from the expansion of their respective Carpathian refuges. Important glacial refuges in southeastern Europe were also documented in *Quercus* ssp. and *Picea abies* (Comes & Kadereit 1998). Huntley & Birks (1983) reported expansion of *Fraxinus excelsior* in the early Holocene from the northern Apennine peninsula and from the northern and northwestern Black Sea coasts based on fossil pollen evidence.

The common ash, *F. excelsior* L., is a postpioneer heliophilous tree species occurring in mixed deciduous forests all over Europe. It is a protandreous wind-pollinated species with a complex mating system, showing variation in sex expression from male to female with hermaphroditic intermediates (Wardle 1961; Picard 1982). Forestry management in Europe has shown increased interest in the common ash in the last 30–40 years, as a consequence of the recognition of its high economic value (Pliura 1999). However, because of technical difficulties with methods of allozyme electrophoresis, little information is so far available on the patterns of genetic variation in the common ash.

In this study, we assess the level of genetic differentiation in *F. excelsior* from southeastern Europe at three different geographical scales (among individuals within populations; among populations within natural regions of Bulgaria; and between regions) using six nuclear microsatellite loci, with the aim of inferring biological characteristics useful for conservation management such as the level of inbreeding and the variance of gene dispersal distances.

Materials and methods

Plant material

Samples of *Fraxinus excelsior* were collected in mixed deciduous forests originating from three distinct regions of Bulgaria. Buds or young leaves from about 30 trees were sampled in each of 10 populations (Fig. 1): three populations were located in the west, near Sofia; three in central Bulgaria; and the remaining four in the northeastern part of the country. Characteristics of the forest stands are given in Table 1. Elevations vary between 100 m and 1050 m. The origin of the common ash populations is putatively autochthonous. Two stands (1 and 2) result from coppice re-sprouting, the others are naturally regenerated from seeds. Within each stand, individual trees were sampled at regular intervals and their position was identified with respect to a reference point. In total, 321 *F. excelsior* trees were sampled. Samples of young leaves were dried and kept at room temperature prior to DNA extraction. Buds were shipped to the laboratory on their twigs wrapped in

Fig. 1 Locations of *Fraxinus excelsior* populations sampled in Bulgaria. Population numbers are the same as in Table 1.

Table 1 Characteristics of common ash populations sampled from Bulgaria

Population	Stand type	Elevation (m above sea level)	Location	Proportion (%)*	Basic substrate
1. Zli dol	coppice	100	43°50′ N, 27°03′ E	30	limestone
2. Kodga ormani	coppice	200	43°54′ N, 26°54′ E	40	loess
3. Iri hissar	mature light	150	43°51′ N, 26°46′ E	10	loess
4. Dulovo	mature dense	200	43°54′ N, 26°54′ E	30	loess
5. Bukatchov chukar	mature high	850	42°50′ N, 26°02′ E	20	marl
6. Golyamoto ravnishte	young dense	800	42°50′ N, 26°02′ E	100	marl
7. Elena	young dense	750	42°50′ N, 26°03′ E	30	marl
8. Kokalyane monastir	mature high	1000	42°33′ N, 23°26′ E	30	schist
9. Ljulin monastir	mature high	1050	42°39′ N, 23°11′ E	30	conglomerates
10. Vitosha	mature high	900	42°38′ N, 23°14′ E	30	andesite

*Proportion of common ash in the canopy.

Table 2 Allelic diversity of the microsatellite loci scored in the common ash

		Original publications		This study	
Locus	Repeat motif	No. of alleles	Range of sizes (bp)	No. of alleles	Range of sizes (bp)
M2–30*	$(TG)_{15}(AG)_{23}$	18	182–248	59	182–294
FEMSATL4†	$(CA)_2(AG)_{24}$	9	164–228	50	158–251
FEMSATL11†	$(GA)_{20}(TA)_4$	11	180–226	32	176–266
FEMSATL12†	$(GA)_6$ CA $(GA)_8$	9	180–262	18	181–264
FEMSATL16†	$(CA)_3$ CG $(CA)_{10}$	4	180–200	10	176–204
FEMSATL19†	$(CA)_6$ CGGC $(CA)_{13}$	12	174–214	33	142–229

*Brachet *et al.* 1999. †Lefort *et al.* 1999.

wet paper. They were then dissected out, separated from their scales, and frozen at −70 °C until DNA extraction.

Microsatellite analysis

Total DNA was extracted from 60 to 90 mg dry leaves or 50–70 mg buds using the QIAGEN Plant mini kit (QIAGEN) with slight modifications, including a 15-min incubation at room temperature of the ground tissue with AP1 buffer and RNAse. Six primer pairs of microsatellite loci which had previously been shown to display easy to read band patterns and a high degree of polymorphism in *F. excelsior* were used for polymerase chain reaction (PCR) (Brachet *et al.* 1999; Lefort *et al.* 1999; see Table 2). The PCR reactions were performed in a mix containing 2.5 mM $MgCl_2$, 0.2 unit per reaction of *Taq* polymerase (Gibco BRL) in Gibco PCR buffer, 0.4 µM of each primer, 0.2 mM of each dNTP and approximately 30 pg of template DNA (1 µL of extract diluted 100 times) in a total reaction volume of 10 µL. After an initial denaturing step for 4 min at 95 °C, amplification comprised 35 cycles of 30 s at 94 °C, 45 s at either 52 °C (FEMSATL4, FEMSATL11, FEMSATL12, FEMSATL16, FEMSATL19) or 56 °C (M2–30) and 1 min at 72 °C. Final elongation was for 5 min at 72 °C. PCR reactions were performed on a Biometra® UNO II thermocycler or on a Perkin Elmer GeneAmp® PCR System 2400. The forward sequence of each primer pair was labelled with a fluorescent dye (PE Applied Biosystems) at its 5′ end: 6-FAM for FEMSATL4 and FEMSATL16, HEX for M2–30 and FEMSATL19, and NED for FEMSATL11 and FEMSATL12. This allowed the pooling of three PCR products together with an internal size standard (Genescan-350 with ROX dye) in each well after adjustment of their relative concentrations. As the size ranges of all loci are overlapping, it was not possible to load more than three PCR products in one well. Electrophoresis and detection of PCR products were carried out on denaturing polyacrylamide gels (4.4% acrylamide–bisacrylamide 19 : 1 from Biorad, 36 cm) using an ABI PRISM® 377 DNA sequencer from Perkin Elmer with filter set D. Gels were run for 2 h at 3000 V in TBE buffer. The resulting electrophoregrams were analysed with software Genotyper® 2.5 from Perkin Elmer.

Estimation of genetic diversity

The following statistics of genetic variation within populations were computed as averages over loci with the software GEN-SURVEY (Vekemans & Lefèbvre 1997): mean

number of alleles per locus, A; average observed heterozygosity, H_O; average gene diversity, H_E, computed according to Nei (1978); and Wright's inbreeding coefficient, F_{IS} corrected for small sample size (Kirby 1975). Deviation of genotypic frequencies from Hardy–Weinberg proportions was tested with the program GENEPOP version 3.1d (Raymond & Rousset 1995).

Within each population, a test of isolation by distance and an indirect estimation of the variance of gene dispersal distances were performed according to a spatial autocorrelation method (Hardy & Vekemans 1999) using the software AUTOCORG ver. 2.0 (ohardy@ulb.ac.be). Kinship coefficients were computed between all pairs of individuals using the statistic of Loiselle *et al.* (1995). The latter is defined for each allele k and each pair of individuals, i and j, as $F_{ij} = (p_i - p)(p_j - p)/(p(1-p)) + 1/(2n-1)$ where p_i, p_j are the allele frequencies of i and j, respectively (taking the following possible values: 0, 0.5, 1), p is the average allele frequency over the reference population, and n is the sample size used to estimate p. This formula produces kinship coefficients that are relative to the reference population.

Average multiallelic and multilocus estimators were obtained by weighting the $F_{i,j}$ values per allele by $p(1-p)$. The matrix of pairwise kinship coefficients was regressed on the matrix of the logarithm of pairwise geographical distances between individuals, giving the slope b_{log}. Mantel tests with 10 000 permutations were used to check the significance of the correlation between the two matrices. The extent of gene dispersal was estimated as $\hat{N}b = -(1 - F_{IS})/b_{log}$, where $\hat{N}b$ can be interpreted as an estimator of Wright's neighbourhood size defined as $Nb = 4\pi D\sigma^2$, with D equal to the population density, and σ^2 equal to the axial variance of gene dispersal distances. For graphical representation, average kinship coefficients were also computed for pairs of individuals separated by the following distance classes (0–10, 10–20, 20–40, 40–80, > 80 m).

The population genetic structure of the overall sample from Bulgaria was analysed with hierarchical F-statistics computed with the software ARLEQUIN (Schneider *et al.* 2000). The overall inbreeding coefficient (F_{IT}) was computed, and the proportion of genetic variance was determined for the following components: among populations in the overall sample (F_{ST}); among populations within regions (F_{SR}); and among regions (F_{RT}). The significance of each variance component was tested with permutation tests (Excoffier *et al.* 1992). Because at each locus many pairs of alleles were found that differ by less than one repeat unit, we did not compute statistics based on the stepwise mutation model such as R_{ST} statistics (Slatkin 1995).

Isolation by distance among populations was assessed by computing $F_{ST}/(1 - F_{ST})$ ratios for each population pair using GENEPOP version 3.1d (Raymond & Rousset 1995). A Mantel test on the matrix of pairwise $F_{ST}/(1 - F_{ST})$ ratios and that of the logarithm of geographical distances (natural logarithm-scale) was performed to test for isolation by distance. An indirect estimate of Nb was computed as $\hat{N}b = 1/B_{log}$ (Rousset 1997), where B_{log} is the slope from the regression of the pairwise $F_{ST}/(1 - F_{ST})$ values on the logarithm of the geographical distances.

Results

Allelic diversity of microsatellite loci

The total number of putative alleles at each locus and the size ranges of the PCR products corresponding to these alleles are given as observed in this study, and as reported in the original publications that first described these loci (Table 2). It appeared that all six microsatellite loci scored in this study were highly polymorphic, displaying a high number of alleles (from 10 to 59 alleles per locus) and a wide size range of PCR products. We recorded between two to five times more alleles per locus than in the original publications, but the original size ranges are only slightly expanded. The total number of alleles scored in 321 individuals over all loci was 202. For the computation of further statistics we omitted locus FEMSATL12 because it produced two singular features: (i) the single locus estimate of inbreeding coefficient ($F_{IS} = 0.697$) was much higher than for the other loci (F_{IS} ranging from −0.016 to 0.061); and (ii) in most populations a large proportion of individuals (varying from 5% to 35% of the sample size) did not give successful amplification after three separate trials. These features could be related to the presence of null alleles at high frequency or to unreliable amplification reactions. Analyses of offspring from controlled crossings are needed to resolve this issue.

Genetic variation within populations

Statistics of genetic diversity within populations are given in Table 3. High polymorphism was found within populations, as on average more than 12 alleles were observed per locus ($A = 12.36 \pm 3.22$) and the probability that two randomly sampled alleles in a given population were different was 73% ($H_E = 0.731$). The observed heterozygosity ($H_O = 0.720$) was slightly lower than the expected heterozygosity (H_E), causing a low but significantly positive mean inbreeding coefficient ($F_{IS} = 0.014$). Five out of 10 populations showed an overall significant departure from Hardy–Weinberg genotypic proportions (showing a deficit and an excess of heterozygotes in two and three populations, respectively), with a mean inbreeding coefficient comprised between $F_{IS} = -0.037$ and $F_{IS} = 0.089$. Assuming that the small positive value of the mean inbreeding coefficient was caused by partial selfing, we can compute an expected value of the selfing rate according to $s = 2F_{IS}/(1 + F_{IS})$ (Hartl & Clark 1989), which gives $\hat{s} = 0.027$.

Table 3 Statistics of genetic variation within populations at five microsatellite loci

Population	n	A	H_O	H_E	F_{IS}†
1. Zli dol	36	17.4	0.720	0.751	0.041*
2. Kodga ormani	36	14.6	0.756	0.738	−0.024
3. Iri hissar	20	12.4	0.740	0.716	−0.034
4. Dulovo	33	13.8	0.691	0.758	0.089***
5. Bukatchov chukar	30	11.2	0.641	0.687	0.066
6. Golyamoto ravnishte	37	15.4	0.762	0.807	0.055
7. Elena	30	12.8	0.780	0.779	−0.001
8. Kokalyane monastery	37	9.0	0.735	0.727	−0.012**
9. Ljulin monastery	30	6.4	0.661	0.638	−0.037*
10. Vitosha	32	10.6	0.709	0.713	−0.004***
Overall mean		12.36	0.720	0.731	0.014***
Standard deviation		3.22	0.045	0.048	0.045

n, number of trees sampled; A, average number of alleles per locus; H_O, average proportion of heterozygotes; H_E, average gene diversity; F_{IS}, average inbreeding coefficient.
†Exact test of departure from Hardy–Weinberg genotypic proportions: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; all other values, not significant.

In each population, the relationship between the matrix of pairwise kinship coefficients between individuals and that of logarithm of spatial distance was analysed with a Mantel test. Significant tests at the 5% level were obtained in populations 1, 2, 8 and 9 (Table 4). Values of the regression slopes (b_{log}) were negative, indicating that on average, individuals which were spatially close were more likely to be genetically related than individuals which were separated by a larger distance (Table 4). The multilocus estimates of neighbourhood size ($\hat{N}b$) based on the b_{log} values varied from 38 to 66 individuals (Table 4). This means that, on average, 86.5% of matings occur within local neighbourhoods comprising from 38 to 66 individuals. For graphical representation of isolation by distance in these populations, we computed for each of five classes of distance d between individuals the averages (F_d) of the multilocus pairwise kinship coefficients (F_{ij}). The average kinship coefficients (F_d) were then plotted against d (Fig. 2). Positive values of F_d were found at short distances, meaning that neighbour individuals had a higher genetic relatedness than random pairs of individuals, whereas negative values of F_d occurred

Table 4 Estimates of neighbourhood size (Nb) from the slope (b_{log}) of the regression of pairwise kinship coefficients on the logarithm of spatial distance for each population of common ash

	Pairs of individuals	b_{log}	P-value	Nb estimate
1. Zli dol	595	−0.024	0.0008	40.4
2. Kodga ormani	630	−0.015	0.0049	66.3
3. Iri hissar	190	−0.005	0.1940	NC*
4. Dulovo	528	−0.006	0.1799	NC
5. Bukatchov chukar	435	0.001	0.5044	NC
6. Golyamoto ravnishte	666	−0.005	0.1519	NC
7. Elena	435	−0.004	0.1452	NC
8. Kokalyane monastery	595	−0.022	0.0016	46.2
9. Ljulin monastery	435	−0.027	0.0002	37.9
10. Vitosha	496	−0.008	0.1578	NC

P-values correspond to the significance level of Mantel tests.
*NC, not computed because the Mantel test was not significant.

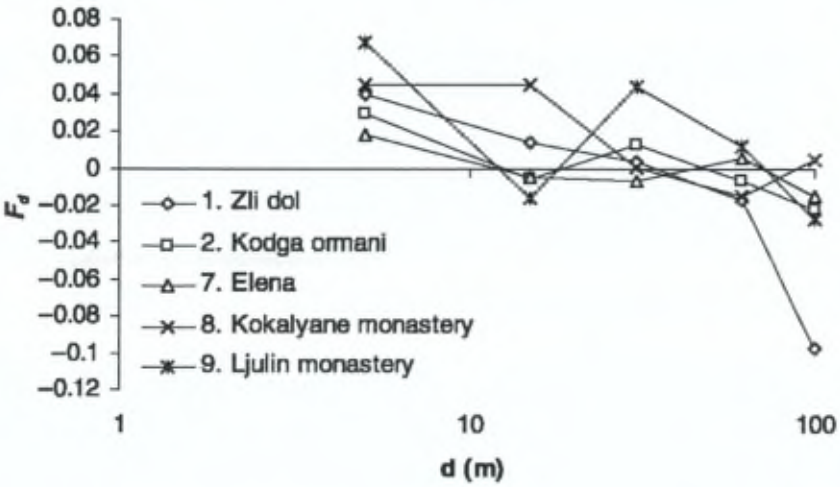


Fig. 2 Average kinship coefficients between pairs of individuals for each of five distance classes (F_d) within five populations of *Fraxinus excelsior* plotted against distance d between individuals (in metres). Positive and negative values of F_d indicate, respectively, higher or lower genetic relatedness between individuals separated by distance d as compared to random pairs of individuals sampled from the population.

Table 5 Hierarchical *F*-statistics at the five microsatellite loci

Loci	Overall inbreeding F_{IT}	Differentiation among: populations F_{ST}	populations within regions F_{SR}	regions F_{RT}
Femsatl4	0.081	0.086	0.065	0.023
Femsatl11	0.080	0.093	0.067	0.028
Femsatl16	0.124	0.063	0.075	−0.013
Femsatl19	0.133	0.099	0.082	0.019
M2–30	0.114	0.078	0.067	0.012
Multilocus estimates	0.104	0.087	0.070	0.018
Permutation test	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$

Significance of the multilocus estimates are computed by permutation tests.

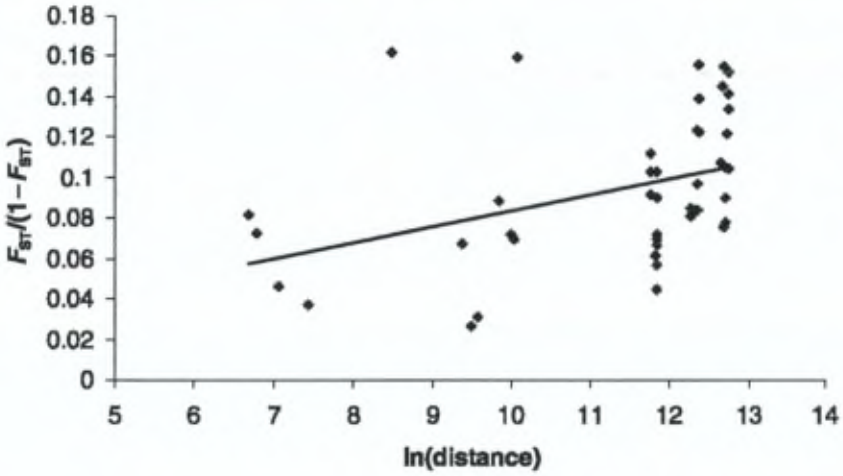


Fig. 3 Plot of $F_{ST}/(1-F_{ST})$ ratios against the logarithm of distances (in metres) among sample plots. The equation of the linear regression gives the slope B_{log} used to estimate the neighbourhood size. Under isolation by distance, values of the ratios are expected to increase linearly with the logarithm of distance (Rousset 1997).

at larger distances, a pattern indicating isolation by distance within a population (Hardy & Vekemans 1999). Population 7 was also included in Fig. 2 because the F_{ij} value at the shortest distance class was significantly higher than zero ($P < 0.05$) based on a permutation test.

Genetic differentiation among populations and regions

Total genetic diversity recorded in Bulgarian samples of *Fraxinus excelsior* was high ($H_T = 0.793 \pm 0.300$) and differentiation among populations was weak but significant: only 8.7% ($F_{ST} = 0.087$) of the total diversity was attributable to differentiation among populations (Table 5). Of the total differentiation among populations, 80% was explained by divergence among populations within regions ($F_{SR} = 0.070$), whereas only 20% accounted for divergence among regions ($F_{RT} = 0.018$). Substantial inbreeding in the overall sample was observed, as measured by the overall inbreeding coefficient ($F_{IT} = 0.104$).

The $F_{ST}/(1 - F_{ST})$ ratio for pairs of populations increased linearly with the natural logarithm of the geographical distance (Fig. 3; Mantel test: $P = 0.0005$, $R^2 = 0.152$), showing a typical pattern of isolation by distance (Rousset 1997), although explaining only 15% of the total variance. The reciprocal of the slope (B_{log}) of the regression of $F_{ST}/(1 - F_{ST})$ on the logarithm of geographical distance provides an estimate of neighbourhood size, $Nb = 126$ ($B_{log} = 0.00797$), which is higher but within the same order of magnitude as estimates obtained within populations.

Discussion

The diversity assessed with microsatellite markers in our sample of populations of *Fraxinus excelsior* from Bulgaria is high: the total gene diversity is $H_T = 0.793 \pm 0.300$ and the observed number of alleles per locus ranges from 10 to 59. The allelic diversity reported here is two to five times higher than in the original publications describing the screening of microsatellite loci in *F. excelsior* (Brachet *et al.* 1999; Lefort *et al.* 1999), while the size range of PCR products is slightly more expanded. The substantial difference in allelic richness is most probably due to the larger number of individuals scored in this study ($n = 321$) compared to the original studies (Brachet *et al.* 1999: $n = 50$, Lefort *et al.*

1999: $n = 16$). However, this observation could be an indication of a particularly high genetic diversity in Bulgarian populations of *F. excelsior*, in agreement with the hypothesis based on palynological data that this area of the Balkans acted as a glacial refuge for the species (Huntley & Birks 1983). Clearly, studies on microsatellite variation in common ash populations from other regions of Europe are needed to confirm this hypothesis. In studies of other deciduous temperate tree species using microsatellite markers, similarly high polymorphism has been reported: Isagi *et al.* (2000) found $H_E = 0.87 \pm 0.08$ and 7–40 alleles at eight microsatellite loci in *Magnolia obovata*; Streiff *et al.* (1998) reported mean gene diversity $H_E = 0.87$ and observed number of alleles comprised between 16 and 32 for six microsatellite loci in a mixed stand of *Quercus robur* and *Quercus petraea*; Degen *et al.* (1999) analysed four of the mentioned loci in two populations of *Q. robur* and found an average $H_E = 0.88$ and between 14 and 33 alleles per locus.

Overall, populations of *F. excelsior* from Bulgaria are slightly, but significantly inbred ($F_{IS} = 0.014$, $F_{IT} = 0.104$). For microsatellites, the most commonly reported cause of positive values of the inbreeding coefficient, indicating heterozygote deficiency, is the presence of null alleles (Bruford *et al.* 1998). This is a likely cause of the singular patterns found at locus FEMSATL12 that we omitted from the analyses. Although we cannot exclude this interpretation for additional loci, other possible explanations are self-fertilization and biparental inbreeding, i.e. inbreeding caused by matings between related individuals. Evidence for the occurrence of biparental inbreeding comes from the analysis of genetic structure within populations. Indeed, we detected an association between genetic relatedness and spatial position of individuals, in agreement with the concept of isolation by distance (Hardy & Vekemans 1999), in four out of 10 populations. Hence, the positive value of F_{IS} may be due in part to biparental inbreeding but knowledge of the dispersal curves of pollen and seeds would be necessary to quantify the contribution of this process to the observed value of F_{IS} (Fenster *et al.* in preparation). Ignoring the effects of biparental inbreeding we computed a maximum expected value of the selfing rate that gives $\hat{s} = 0.027$. Hence, our computation suggests that the selfing rate within Bulgarian populations of the common ash is lower than 3%.

We detected a pattern of genetic structure consistent with the model of isolation by distance at both intrapopulation and interpopulation levels. Estimates of the variance of gene dispersal distances obtained from intrapopulation analyses ranged from 38 to 66 individuals, when expressed as Wright's neighbourhood size, and a value of 126 was obtained from the interpopulation analysis. Similar congruence between estimates of gene dispersal obtained from analyses within and between populations has been reported in two other plant species (on oak, when comparing results from Le Corre *et al.* 1998 and Streiff *et al.* 1999; on *Chamaecrista fasciculata*, Fenster *et al.* in preparation). In another noble hardwood species, sugar maple, Perry & Knowles (1991) also found significant spatial genetic structure within three forest stands. From a reanalysis of their data sets with similar procedures as those used in this study, we obtained estimates of neighbourhood size close to those in the common ash, ranging from 29 to 59 individuals. In contrast, in oak, estimates of neighbourhood size ranging from 3000 to 4000 trees, and of neighbourhood area between 12 and 20 ha, have been obtained based on analyses of spatial distribution of allozyme frequencies (Le Corre *et al.* 1998) and on paternity analysis (Streiff *et al.* 1999). For comparison, the neighbourhood area in the common ash (computed as Nb/D where D is the population density) would vary between 0.63 and 0.81 ha based on within-population estimates of Nb when an average density of 200 mature trees per hectare in a mixed deciduous forest with a proportion of *F. excelsior* ranging from 10 to 100% (Table 1) is assumed. There is thus one order of magnitude difference between estimates of gene dispersal in oak and in the common ash. This could result from true differences in dispersal abilities of the two tree species. Indeed, very limited pollen dispersal has been reported in *F. excelsior* (50% and 95% of pollen dispersed at less than 10 m and 50 m, respectively, Altman & Dittmer 1964). Also seed dispersal mechanisms differ with heavy wind-dispersed seeds in the common ash and a combination of gravity and bird-mediated dispersal in oaks. Alternatively, the large difference in gene flow estimates could result from differences in sampling (much smaller sampling area in our study as compared to those for oak) and estimation procedures (indirect vs. direct estimation of variance in gene dispersal distances). In this study, an indirect method based on the effect of restricted gene flow on spatial genetic structure was used. Such methods are derived from particularly idealized models of populations from which many assumptions may be violated in nature (e.g. the equilibrium hypothesis, Whitlock & McCauley 1999). Moreover we did not find any indication of spatial genetic structure within five of 10 populations. This could be due to the particular history of these populations, i.e. recent origin or influence of forestry management, or it could result from the fact that the sampling area was shorter than the actual neighbourhood area in these forest stands (Rousset 2000). In the latter case, ignoring these populations would lead to a downward bias in the estimation of gene dispersal.

As commonly found in tree species, we observed a low overall differentiation among populations ($F_{ST} = 0.087$). In comparison, the average estimate of population differentiation among 322 woody taxa for allozyme loci is $G_{ST} = 0.084$ (Hamrick *et al.* 1992), whereas lower estimates were obtained for oak ($G_{ST} = 0.024$ and 0.032 in *Quercus petraea* and *Q. robur*, respectively, Zanetto *et al.* 1994) and

beech (in *Fagus sylvatica*, $G_{ST} = 0.054$, Comps *et al.* 1990; $F_{ST} = 0.046$, Leonardi & Menozzi 1995). These comparisons are in agreement with the suggestion that gene flow is more restricted in *F. excelsior* than in oak and beech, but could also be due to differences in processes of postglacial recolonization (Austerlitz *et al.* 2000). It must be recognized that comparisons of F_{ST} values between allozyme and microsatellite data may be misleading because of the very high mutation rates of microsatellite loci that can cause a downward bias in estimation of genetic differentiation (Nagylaki 1998). However, this should not hamper the conclusion that based on our study the differentiation among populations is higher in the common ash than in oak and beech. When compared to other noble hardwood species studied at a similar geographical scale, our results of population differentiation in the common ash fall within the reported range:, e.g. *Juglans regia*, F_{ST} within different European countries ranging from 0.018 to 0.071 (Fornari *et al.* 1999); *Castanea sativa* from Italy, $F_{ST} = 0.10$ (Pigliucci *et al.* 1990), from Turkey, $F_{ST} = 0.184$ (Villani *et al.* 1999); *Ulmus minor* from France, $F_{ST} = 0.178$ (Machon *et al.* 1997). Further investigations on pollen and seed dispersal with direct methods in noble hardwood species are needed to test whether differences in levels of genetic differentiation can be mainly explained by variation in gene dispersal.

Acknowledgements

This paper represents a portion of the doctoral research of Myriam Heuertz who received a scholarship (ref. BFR 99/046) from MCESR (Ministry of Culture, of Higher Education and of Research). This work was funded by IPGRI (International Plant Genetic Resources Institute) (project GENFOR). We are grateful to C. Muller from Laboratoire National de Santé, Department of Immunology, for letting us use the automated sequencer. We wish to thank M. E. Morand and O. Hardy for collaboration and stimulating discussions and to L. Hoffmann and two anonymous referees for helpful comments on an earlier draft of the manuscript.

References

Altman PL, Dittmer DS (1964) *Biology Data Book*, pp. 428–431. Federation of American Societies for Experimental Biology, Washington, DC. cited by: Richards AD (1997) *Plant Breeding Systems* 2nd edn, p. 183. Chapman & Hall. London.

Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics*, **154**, 1309–1321.

Brachet S, Jubier MF, Richard M, Jung-Muller B, Frascaria-Lacoste N (1999) Rapid identification of microsatellite loci using 5′ anchored PCR in the common ash *Fraxinus excelsior*. *Molecular Ecology*, **8**, 160–163.

Bruford MW, Ciofi C, Funk SM (1998) Characteristics of microsatelllites. In: *Molecular Tools for Screening Biodiversity* (eds Karp A, Isaac PG, Ingram DS), pp. 202–205. Chapman & Hall, London.

Comes HP, Kadereit JW (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432–438.

Comps B, Thiébaut B, Paule L, Merzeau D, Letouzey J (1990) Allozymic variability in beechwood (*Fagus sylvatica* L.) over Central Europe: spatial differentiation among and within populations. *Heredity*, **65**, 407–417.

Degen B, Streiff R, Ziegenhagen B (1999) Comparative study of genetic variation and differentiation of two pedunculate oak (*Quercus robur*) stands using microsatellite and allozyme loci. *Heredity*, **83**, 597–603.

Demesure B, Comps B, Petit R (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515–2520.

Eriksson G (1999) Global warming and gene conservation of Noble Hardwoods. In: *Noble Hardwoods Network. Report of the Third Meeting, 13–16 June 1998, Sagadi, Estonia* (compilers Turok J, Jensen J, Palmberg-Lerche C *et al.*), pp. 98–112. International Plant Genetic Resources Institute, Rome.

European Commission Phare Programme (1999) *Conservation and Sustainable Management of Forests in Central and Eastern European Countries*. Phare Environment Consortium, Brussels.

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

Fornari B, Cannata F, Spada M, Malvolti ME (1999) Allozyme analysis of genetic diversity and differentiation in European and Asiatic walnut (*Juglans regia* L.) populations. *Forest Genetics*, **6**, 115–127.

Geburek T (1997) Isozymes and DNA markers in gene conservation of forest trees. *Biodiversity and Conservation*, **6**, 1639–1654.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation and population genetics models. *Heredity*, **83**, 145–154.

Hartl DL, Clark AG (1989) *Principles of population genetics*, 2nd edn, p. 682. Sinauer Associates, Sunderland, MA.

Huntley B, Birks HJB (1983) *An Atlas of Past and Present Pollen Maps for Europe: 0–13 000 Years Ago*, pp. 209–223. Cambridge University Press, Cambridge.

Isagi Y, Kanazashi T, Suzuki W, Tanaka H, Abe T (2000) Microsatellite analysis of the regeneration process of *Magnolia obovata* Thunb. *Heredity*, **84**, 143–151.

King RA, Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology*, **7**, 1151–1161.

Kirby GC (1975) Heterozygote frequencies in small subpopulations. *Theoretical Population Biology*, **8**, 31–48.

Konnert M, Bergmann F (1995) The geographical distribution of genetic variation in silver fir (*Abies alba*, Pinaceae) in relation to its migration history. *Plant Systematics and Evolution*, **196**, 19–30.

Le Corre V, Roussel G, Zanetto A, Kremer A (1998) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. III. Patterns of variation identified by geostatistical analyses. *Heredity*, **80**, 464–473.

Lefort F, Brachet S, Frascaria-Lacoste N, Edwards KJ, Douglas GC (1999) Identification and characterisation of microsatellite loci in ash (*Fraxinus excelsior* L.) and their conservation in the olive family (Oleaceae). *Molecular Ecology*, **8**, 1088–1091.

Leonardi S, Menozzi P (1995) Genetic variability of beech (*Fagus sylvatica* L.) in Italy: the role of postglacial recolonization. *Heredity*, **75**, 35–44.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understorey shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Machon N, Lefranc M, Bilger I, Mazer SJ, Sarr A (1997) Allozyme variation in *Ulmus* species from France: analysis of differentiation. *Heredity*, **78**, 12–20.

Milligan BG, Leebens-Mack J, Strand AE (1994) Conservation genetics: beyond the maintenance of marker diversity. *Molecular Ecology*, **3**, 423–435.

Nagylaki T (1998) Fixation Indices in Subdivided Populations. *Genetics*, **148**, 1325–1332.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology and Evolution*, **14**, 140–145.

Perry DJ, Knowles P (1991) Spatial genetic structure within three sugar maple (*Acer saccharum* Marsh.) stands. *Heredity*, **66**, 137–142.

Picard JF (1982) Contribution à l'étude de la biologie florale et de la fructification du Frêne commun (*Fraxinus excelsior* L). *Revue Forestière Française*, **XXXIV**, 97–107.

Pigliucci M, Benedettelli S, Villani F (1990) Spatial patterns of genetic variability in Italian chestnut (*Castanea sativa*). *Canadian Journal of Botany*, **68**, 1962–1967.

Pliura A (1999) European long-term gene conservation strategies Ash (*Fraxinus* ssp). In: *Noble Hardwoods Network. Report of the Third Meeting, 13–16 June 1998, Sagadi, Estonia* (compilers Turok J, Jensen J, Palmberg-Lerche C *et al.*), pp. 8–20. International Plant Genetic Resources Institute, Rome.

Raymond M, Rousset F (1995) GENEPOP (Version 12) Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rotach P (1999) In situ conservation and promotion of Noble Hardwoods: silvicultural management strategies. In: *Noble Hardwoods Network. Report of the Third Meeting, 13–16 June 1998, Sagadi, Estonia* (compilers Turok J, Jensen J, Palmberg-Lerche C *et al.*), pp. 39–50. International Plant Genetic Resources Institute, Rome.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.

Schneider S, Roessli D, Excoffier L (2000) *Arlequin, Version 2000: A software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glössl J, Kremer A (1998) Within-population genetic structure in *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology*, **7**, 317–328.

Streiff R, Ducousso A, Lexer C, Steinkellner H, Gloessl J, Kremer A (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. *Molecular Ecology*, **8**, 831–841.

Turok J, Eriksson G, Kleinschmidt J, Canger S, compilers (1996) *Noble Hardwoods Network. Report of the First Meeting, 24–27 March 1996, Escherode, Germany*. International Plant Genetic Resources Institute, Rome.

Turok J, Collin E, Demesure B *et al.* compilers (1998) *Noble Hardwoods Network. Report of the Second Meeting, 22–25 March 1997, Lourizán, Spain*. International Plant Genetic Resources Institute, Rome.

Turok J, Jensen J, Palmberg-Lerche C *et al.* compilers (1999) *Noble Hardwoods Network. Report of the Third Meeting, 13–16 June 1998, Sagadi, Estonia*. International Plant Genetic Resources Institute, Rome.

Vekemans X, Lefèbvre C (1997) On the evolution of heavy-metal tolerant populations in *Armeria maritima*: evidence from allozyme variation and reproductive barriers. *Journal of Evolutionary Biology*, **10**, 175–191.

Vendramin GG, Degen B, Petit RJ, Anzidei M, Madaghiele A, Ziegenhagen B (1999) High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Molecular Ecology*, **8**, 1117–1126.

Villani F, Sansotta A, Cherubini M, Cesaroni D, Sbordoni V (1999) Genetic structure of natural populations of *Castanea sativa* in Turkey: evidence of a hybrid zone. *Journal of Evolutionary Biology*, **12**, 233–244.

Wardle P (1961) Biological Flora of the British Isles. *Fraxinus excelsior* L. *Journal of Ecology*, **49**, 739–751.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$. *Heredity*, **82**, 117–125.

Zanetto A, Roussel G, Kremer A (1994) Geographic variation of inter-specific differentiation between *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. *Forest Genetics*, **1**, 111–123.

Myriam Heuertz is a PhD student at CRP-Gabriel Lippmann and ULB working under the supervision of Jean-François Hausman and Xavier Vekemans on the conservation of genetic resources in the common ash from Eastern Europe. Jean-François Hausman is senior researcher at CRP-Gabriel Lippmann, working on the physiological response of woody species to environmental constraints. Nathalie Frascaria-Lacoste is a lecturer in forest genetics at the ENGREF Institution (PARIS). Ivaylo Tsvetkov is senior researcher at Forest Research Institue, Sofia, woking on micropropagation and regeneration of tree species. Xavier Vekemans is senior researcher at ULB working in plant population genetics.

Chapter II: Estimating seed versus pollen dispersal from spatial genetic structure in the common ash

M. Heuertz, X. Vekemans, J.-F. Hausman, M. Palada and O. J. Hardy.

Abstract

Spatial genetic structure was analysed with five highly polymorphic microsatellite loci in a Romanian population of common ash (*Fraxinus excelsior* L.), a wind-pollinated and wind-dispersed tree species occurring in mixed deciduous forests nearly all over Europe. Contributions of seed and pollen dispersal to total gene flow were investigated by analysing the pattern of decrease in kinship coefficients among pairs of individuals with geographic distance and comparing it with simulation results. Plots of kinship against the logarithm of distance were decomposed into a slope and a shape component. Simulations showed that the slope informed on the global level of gene flow, in agreement with theoretical expectations, whereas the shape component was correlated to the relative importance of seed *vs.* pollen dispersal. Hence, our results indicate that insights into the relative contributions of seed and pollen dispersal to overall gene flow can be gained from details of the pattern of spatial genetic structure at biparentally inherited loci. In common ash, the slope provided an estimate of total gene dispersal in terms of Wright's neighbourhood size of Nb = 519 individuals. No precise estimate of seed versus pollen flow could be obtained from the shape because of the stochasticity inherent to the data, but the parameter combinations that best fitted the data indicated restricted seed flow, $\sigma_s \leq 14$ m, and moderate pollen flow, 70 m $\leq \sigma_p \leq 140$ m.

Introduction

In natural plant populations, spatial genetic structure of sexually reproducing individuals is determined by the combined effects of gene flow, local genetic drift, various forms of natural selection and the spatial arrangement of individuals (Wright 1943; Epperson 1993; Doligez *et al.* 1998). Gene flow through pollen and seed dispersal is a key determinant in the establishment of genetic structure: when overall gene flow is restricted, a decrease in genetic relatedness among pairs of individuals is expected with increasing geographic distance between them, in agreement with models of isolation by distance (Wright 1943; Malécot 1950).

Many studies on forest tree species have tested for local spatial genetic structure using genetic markers and the tools of spatial autocorrelation analysis. In most cases, evidence for statistically significant spatial genetic structure was found (e.g. Schnabel *et al.* 1991; Sork *et al.* 1993; Berg & Hamrick 1995; Streiff *et al.* 1998; Chung & Epperson 2000; Takahashi *et al.* 2000; Ueno *et al.* 2000; Dutech *et al.* 2002). In some species, however, spatial distribution of genotypes was close to random (Epperson & Allard 1989; Knowles 1991; Xie & Knowles 1991; Chung *et al.* 2000). The latter situation was observed in tree species where both pollen and seed dispersal are wide-ranging, as for example in wind-pollinated and wind-dispersed conifers (*Pinus contorta* ssp. *latifolia*, Epperson & Allard 1989; *Pinus banksiana*, Xie & Knowles 1991; *Picea mariana*, Knowles 1991); or alternatively in insect-pollinated species but with very efficient seed dispersal through frugivorous birds (*Neolitsea sericea*, Chung *et al.* 2000). In contrast, in those species where significant spatial genetic structure was detected, either seed dispersal, or both pollen and seed dispersal are spatially restricted. For example spatial genetic structure in different oak species (*Quercus rubra*, Sork *et al.* 1993; *Q. laevis*, Berg & Hamrick 1995, *Q. robur* and *Q. petraea*, Streiff *et al.* 1998) was observed at short spatial scale. This pattern was attributed to localised, gravity-mediated seed dispersal, as pollen dispersal by wind is extensive in those species (Sork *et al.* 1993; Berg & Hamrick 1995, Streiff *et al.* 1998). Strong spatial genetic structure was found in tree species featuring substantial restrictions to both pollen and seed dispersal, i.e. pollination by small insects and seed dispersal by gravity (*Gleditsia triacanthos*, Schnabel *et al.* 1991; *Eurya emarginata*, Chung & Epperson 2000).

Relative pollen and seed contributions to gene flow can readily be estimated from genetic structure when it is analysed concomitantly at markers with different modes of inheritance (Latta *et al.* 1998, Latta *et al.* 2001). However, it remains unclear whether the relative contributions of pollen and seed dispersal could be deduced from the observed patterns of spatial genetic structure at nuclear loci alone. Methods for estimating the extent of local gene flow from the observed spatial genetic structure within populations have recently been proposed (Hardy & Vekemans 1999; Rousset 2000) and applied to several plant species

(Heuertz *et al.* 2001; Dutech *et al.* 2002; Fenster *et al.* in preparation). These methods are based on the rate of decrease of genetic relatedness between individuals with geographic distance (Hardy & Vekemans 1999; Rousset 2000). Theoretical models of isolation by distance in a two-dimensional space predict that relatedness decreases linearly with the logarithm of the distance at a rate inversely proportional to the product of the effective population density, D, and half the mean squared gene dispersal distance, σ_g^2, within an adequate distance range (*ca.* σ_g to $20\sigma_g$, Rousset 1997, 2000). Gene dispersal is contributed by pollen and seed dispersal ($\sigma_g^2 = \sigma_s^2 + \sigma_p^2/2$, where σ_s^2, σ_p^2 are half the mean squared dispersal distances of seed and pollen, respectively; Crawford 1984), but the relative contribution of each component to total gene dispersal cannot be inferred using this approach. However, since this method focuses on the rate of decrease of relatedness within a limited distance range only, it does not exploit the whole information available from the plot of relatedness against distance. In this study, we attempted to gather information on the relative contributions of seed and pollen dispersal by inspecting in detail how relatedness between individuals, as assessed by biparentally inherited markers, decreases over an unrestricted range of spatial distances.

Common ash (*Fraxinus excelsior* L., Oleaceae) is a temperate tree species which occurs in mixed deciduous forests nearly all over Europe. It is wind-pollinated and its single-seeded fruits, the winged samaras, are wind-dispersed. Its mating system is polygamous: there is a continuum from pure male to pure female individuals with hermaphroditic intermediates (Picard 1982; A Lamb and D Boshier unpublished; Wallander 2001). In a previous study, we have identified strong genetic structure at short geographic scale (less than 1 ha) in four out of 10 common ash populations from Bulgaria (Heuertz *et al.* 2001). This pattern was tentatively attributed to restricted seed dispersal (heavy wind-dispersed seed) and possibly restricted pollen dispersal. Our objective in the present study was to investigate whether a pattern of isolation by distance could be observed at larger geographical scale in a common ash population from southeastern Romania and whether some inference could be drawn on the relative contributions of pollen and seed to total gene flow from the observed spatial genetic structure. For this purpose we used intensive computer simulations of a theoretical population that closely matched features of the study population. We simulated different combinations of pollen and seed dispersal parameters and examined the resulting plots of relatedness against distance in order to (1) investigate how the relative contributions of seed *vs.* pollen dispersal affect the patterns of decrease of relatedness with distance and (2) infer levels of pollen and seed dispersal in the study population by comparing observed and simulated data.

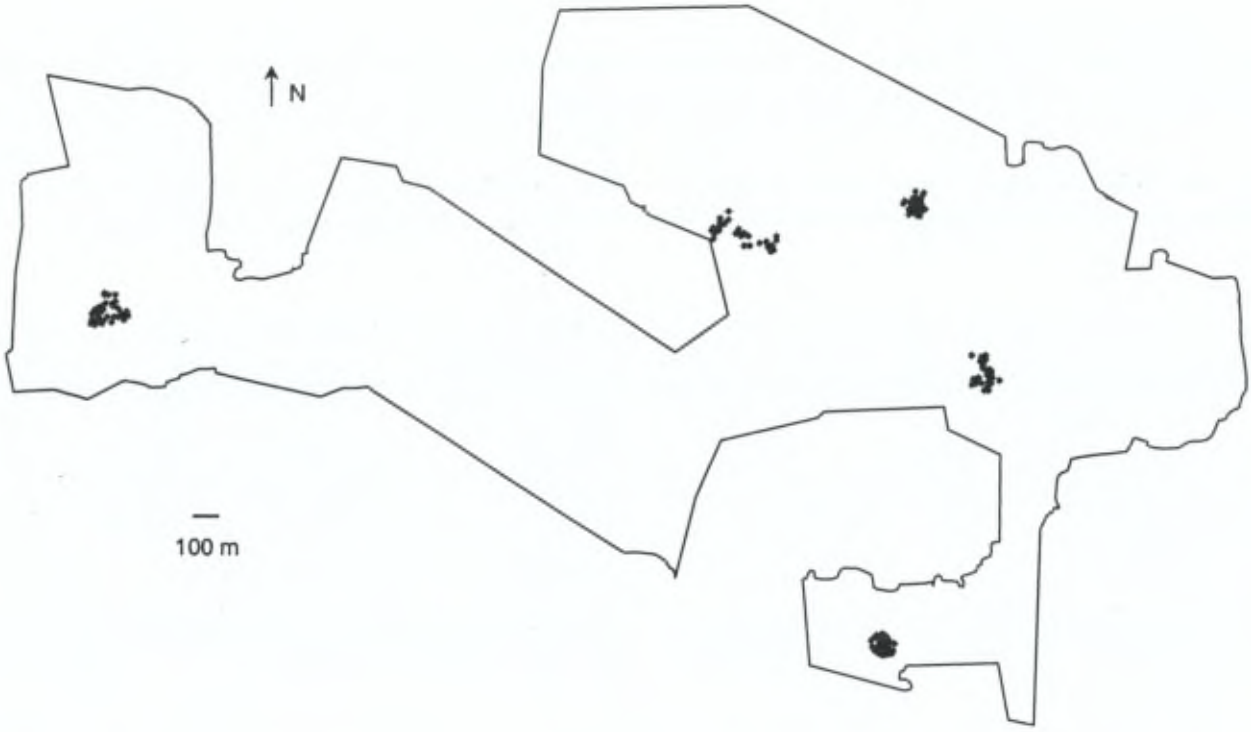


Fig. 1 Sampling locations of common ash trees in the Ploesti forest.

Materials and Methods

Plant material

The study site is a putatively autochthonous continuous mixed deciduous forest (approximately 710 ha) located in southeastern Romania (44°50'N, 26°04'E) near the city of Ploesti (Fig. 1). *Fraxinus excelsior* occurs there at high density, on average 200 mature trees (trees with diameter at breast height, d.b.h. > 15 cm) per ha, in mixture with *Quercus robur*, *Acer platanoides*, *A. campestre*, *A. tataricum*, *Ulmus minor* and other broadleaved tree species. An average of 30 non-adjacent common ash trees were sampled and mapped in each of five sampling locations, called subpopulations hereafter, covering an area of 0.71 to 3.69 ha (average 1.68 ha). Distances between subpopulations ranged from 514 to 3043 m. In total, 152 trees were sampled. Samples consisted of buds, which were shipped to the laboratory on their twigs wrapped in wet paper. They were then dissected out, separated from their scales and kept at –70°C until DNA extraction.

Microsatellite analysis

Total DNA was extracted with the CTAB procedure of the NucleoSpin Plant kit (Macherey Nagel) from 50 to 70 mg of buds ground in an automatic grinding mill (Retsch MM200). Microsatellite analysis was performed as described in Heuertz *et al.* (2001). Five highly polymorphic microsatellite loci were amplified with polymerase chain reaction (PCR) (Table1). The quantity of template DNA in the PCR could be increased to 3 to 10 ng compared with the previously published protocol, probably because metabolites susceptible to interfere with PCR were better eliminated with the DNA extraction kit used here. Fluorescent labeling of the forward primers allowed detection of amplification products on an automated DNA sequencer (ABI PRISM® 377 DNA sequencer). Sizing of fragments was performed with the software programs Genescan® 3.1 and Genotyper® 2.5 from Applied Biosystems by comparison with an internal sizing standard (Genescan-350 Rox). Another modification brought to the previously published protocol was that polyacrylamide gels were produced with 10% LongRanger® gel solution (Sanver Tech).

Microsatellite data analysis

The following statistics of genetic variation were computed for each locus and as averages over loci for each of the five subpopulations and for the entire population with the program GEN-SURVEY (Vekemans and Lefèbvre 1997): (1) the average number of observed alleles *A*, (2) the average expected heterozygosity or gene diversity *H* corrected for small sample size (Nei 1978) and (3) Wright's inbreeding coefficient *F* corrected for small sample size (Kirby

1975). Deviation of genotypic frequencies from Hardy-Weinberg proportions was tested with exact tests with the program GENEPOP version 3.3 d (Raymond and Rousset 1995).

Spatial genetic structure was analysed using kinship coefficients. The *a priori* kinship coefficient Φ_{ij} is the probability that a gene drawn at random from individual i and a gene drawn at random from individual j are identical by descent (IBD). From a sample of genotyped individuals, only relative kinship coefficients, $F_{ij} \equiv \left(\Phi_{ij} - \overline{\Phi}\right)/\left(1 - \overline{\Phi}\right)$, can actually be estimated, where $\overline{\Phi}$ is the mean Φ_{ij} over all pairs of sampled individuals. Kinship coefficients were computed for all pairs of individuals in each subpopulation and in the entire population using the statistic of Loiselle *et al.* (1995), which is defined for each allele k and each pair of individuals, i and j, as $F_{ij} = (p_i - \overline{p_k})(p_j - \overline{p_k})/(\overline{p_k}(1-\overline{p_k})) + 1/(2n-1)$ where p_i, p_j are the frequencies of allele k in individuals i and j (taking the values 0, 0.5 or 1) and p_k is the average allele frequency of allele k in the reference population with sample size n. Average multiallelic (multilocus) estimates were computed by weighting the F_{ij} for each allele k by its polymorphism index $\overline{p_k}(1-\overline{p_k})$.

In order to test for isolation by distance, the multilocus kinship coefficient for each pair of individuals was plotted against the logarithm of the geographic distance separating them (kinship-distance plot). The significance of the linear regression line was tested by Mantel tests with 10000 permutations. The extent of gene dispersal was estimated from the observed regression slope b_{ro} as

$$\hat{N}b = -(1-F_0)/b_{ro} \qquad (1)$$

where F_0 is the kinship coefficient between adjacent individuals, i.e. in distance class zero. $\hat{N}b$ can be interpreted as an estimator of Wright's neighbourhood size defined as $Nb = 4\pi D_e \sigma_g^2$, where D_e is the effective population density, and σ_g^2 the second moment of gene dispersal distance between parents and offspring (Hardy & Vekemans 1999, Rousset 2000). A linear decrease of pairwise kinship coefficients with the logarithm of distance is expected within a distance range comprised between approximately σ_g and $20\sigma_g$, and consequently the inference of Nb is most reliable in this range (Rousset 1997, 2000). All computations were performed with the program SPAGeDi (Hardy & Vekemans 2002, http://www.ulb.ac.be/sciences/lagev). For graphical representation of kinship in the total population, average kinship coefficients were computed for 17 distance classes, equivalent to those used for the analysis of simulated data (see below).

Computer simulation procedure

In order to determine whether inference on the relative contributions of pollen and seeds to total gene flow could be drawn from spatial genetic structure, we compared the kinship-

distance plot of the common ash population to kinship-distance plots from simulated data sets produced with various combinations of pollen and seed dispersal parameters. Computer simulations were carried out on a theoretical population that closely matched features of the real population. In the real population, the density was on average 200 mature trees/ha and the maximum distances between samples were 3.43 km and 1.75 km on the abscissa and the ordinate of an orthogonal grid. Hence, the population could be represented by 490 x 250 trees positioned on a rectangular grid where adjacent trees were separated by 7 meters. To minimise edge effects, a larger theoretical population of 600 x 300 regularly spaced mature hermaphrodite diploid individuals was simulated considering overlapping generations. Simulations proceeded by the following steps: (1) an initial population is created by defining individual genotypes, choosing alleles stochastically at five loci according to allele numbers and frequencies observed in the common ash population; (2) each individual on the grid is given a 80% chance of surviving (assuming one simulation cycle corresponds to 1/5 of average lifetime); (3) each empty cell (death event) is replaced by an offspring whose maternal parent is chosen stochastically around the empty cell according to a pre-defined seed dispersal curve, and the paternal parent is chosen stochastically around the maternal parent according to a pre-defined pollen dispersal curve (individuals present at the previous step were considered when selecting parents); (4) the genotype of each new offspring is defined by choosing, at each locus, one allele of each parent at random; (5) once all individuals have been considered, a new population is defined by keeping previously surviving individuals and replacing dead ones by the offspring. Stages 2 to 5 were repeated 1000 times. This corresponds to 200 generations, which is a realistic figure for postglacial recolonised populations (Kremer 1994) and which allowed to neglect mutation. Seed and pollen dispersal followed an isotropic bivariate normal distribution characterized by parameters σ_s and σ_p, respectively. We ran 100 independent simulation replicates in each of 56 combinations of dispersal parameters, where σ_s = 0.75, 1, 2, 3, 4, 5, 7, 10 and σ_p = 3, 5, 7, 10, 20, 50, 100 (in lattice units; these values are to be multiplied by seven to obtain σ in meters). After 1000 simulation cycles, equilibrium spatial genetic structure was reached for most parameter combinations, except when both seed and pollen dispersal were very narrow. In each replicate, 152 individuals were then sampled at the grid positions corresponding to those closest to the position of sampled trees in the reference population if the latter had been overlaid by a 7 x 7 m grid. Average multilocus kinship coefficients per distance intervals, F_k, were computed for the following 17 distance classes (upper bound distance in grid units; times 7 to obtain meters): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 40, 200, 300, 600. We used short intervals for the first distance classes to get a detailed picture of how kinship varies with distance at a small spatial scale (i.e. within subpopulations, ≤40 lattice units). At larger spatial scales, wider intervals were used because kinship is then expected to vary little and estimates suffer less stochastic variance.

Analysis of simulated data sets

In order to identify trends of variation among the kinship-distance plots, kinship coefficients F_k in each distance class k were averaged over the 100 simulation replicates, yielding $\overline{F_k}$ for each class k in each parameter combination.

First, to verify the expected linear decrease of kinship-distance plots, we plotted $\overline{F_k}$ against the logarithm of distance (measured in m), and computed the slope b_r of the linear regression (i) over all distances and (ii) over distances ranging from σ_g to 20 σ_g, with $\sigma_g^2 = \sigma_s^2 + \sigma_p^2/2$, where σ_s, σ_p are the standard deviations of seed and pollen dispersal distances, respectively, entered as parameters. To compute b_r, $\overline{F_k}$ values were weighted by the number of pairs of individuals belonging to class k, n_k:

$$b_r = \frac{\sum_k n_k (x_k - \bar{x})(\overline{F_k} - \overline{F})}{\sum_k n_k (x_k - \bar{x})^2} \quad , \qquad (2)$$

where x_k is the average of ln(distance between individuals), with distance in m, for distance class k; $\bar{x}$ is the weighted average of x_k, $\bar{x} = \sum_k n_k x_k \Big/ \sum_k n_k$; and $\overline{F}$ is the weighted average kinship coefficient $\overline{F} = \sum_k n_k \overline{F_k} \Big/ \sum_k n_k$. Equation (2) ensures that b_r is essentially independent of distance class designations because it is nearly equivalent to regressing all pairwise F_{ij} values on the logarithm of distance. The fit of the linear regression for each parameter combination was evaluated with the coefficient of determination R^2, which expresses the proportion of variation in $\overline{F_k}$ values explained by the linear regression model.

The extent of local gene flow, σ_g, was estimated from the regression slope b_r using the equation $\hat{\sigma}_g = \sqrt{\hat{N}b/(4\pi D_e)}$, with the estimated neighbourhood size $\hat{N}b = -(1-F_{k=1})/b_r$. Assuming that density was constant over time, the effective density D_e was computed from $D_e = D[4/(2+V)][1/(1+F_I)]$ (Crawford 1984), where D corresponds to the density of individuals, V is the variance of the lifetime reproductive success among individuals, and F_I is Wright's inbreeding coefficient. V and F_I were recorded in the course of the simulations and we found $V \approx 5$ and $F_I \approx 0.01$ for all parameter combinations, resulting in $D_e \approx 0.56D$. Similarly, as seed and pollen dispersal events occurring in the simulations were recorded, we also computed the realised neighbourhood size and pollen and seed dispersal standard deviations, Nb_r, σ_{sr} and σ_{pr}, respectively ($\sigma_r = \frac{1}{2}\sqrt{\sum r_i^2 / N}$, where r_i is the distance crossed by the ith dispersal event, N being the total number of dispersal events recorded;

$Nb_r = 4\pi(\sigma_{sr}^2 + \frac{1}{2}\sigma_{pr}^2)\frac{4}{(2+V)(1+F_I)}$). Realised values were slightly smaller than their corresponding parameter values because of edge effects and the use of a lattice rather than a continuous space. The estimates $\hat{\sigma}_g$ were compared to their expected values computed as $\sigma_{g\exp} = \sqrt{\sigma_{sr}^2 + 0.5\sigma_{pr}^2}$.

The shape of dispersal curves was further investigated by examining plots of residuals

$$d_k = \overline{F}_k - F_{ke} \tag{3}$$

with F_{ke} being the kinship coefficient for class k expected from linear regression over the entire distance range. Residuals d_k were divided by

$$s\sqrt{1-h_k}\,, \tag{4}$$

to obtain plots in standard deviation scale, where $s = \sqrt{\sum_k n_k d_k^2 \big/ (n-2)}$ is the unexplained standard deviation (i.e. not due to linear regression) and $h_k = 1\big/K + n(x_k - \overline{x})^2 \big/ \sum_k n_k (x_k - \overline{x})^2$ the leverage coefficient of distance class k, as recommended by Sokal and Rohlf (1995, p. 531), weighted by the number of pairs of individuals in class k. The total number of classes K was 17. The visual appearance of plots of standardised residuals against distance suggested that they might be explained by polynomial regression. We fitted polynomial functions up to the third degree of the logarithm of distance to the observed standardised residuals plots following Sokal and Rohlf (1995, p.616) and weighting each distance class by its respective number of pairs. Coefficients of these functions were compared to the ratio of seed *vs.* pollen contributions to total dispersal.

Fitting the observed spatial genetic structure to simulation results

The direct comparison of kinship values observed in the common ash population with those from simulated datasets required equivalent definitions of distance classes in the real and theoretical populations. Upper bounds of classes in the real population were determined by (i) multiplying those of the theoretical population by 7 m and (ii) adding or subtracting a maximum of 1.5 m to correct for rounding errors from positioning of the sample points in the theoretical population. The resulting 17 distance classes were (upper bound distances in m): 9.5, 14, 22, 27.5, 36.5, 42.5, 49, 56, 63, 71, 84, 105, 140, 280, 1400, 2100, 4200. Subsequently, the fit of the observed kinship plot was investigated (i) using a X^2 statistic describing the global departure between observed and simulated kinship coefficients over all distance classes, (ii) using two statistics based on the regression of kinship coefficients on the

logarithm of distances, one being the regression slope, the other being a shape index based on the residuals. These statistics are defined as follows.

Chi-square-like test. For each independent simulation replicate in each parameter combination, we computed

$$X^2 = \sum_k \left[\frac{(F_k - \overline{F}_k)^2}{\mathrm{var}(F_k)} \right] \tag{5}$$

where F_k is the average multilocus kinship coefficient in distance class k, and $\overline{F}_k$ and var(F_k) are, respectively, the average and variance of F_k over the 100 independent simulation replicates of the investigated parameter combination. The numerator of equation (4) measures the relative squared error of a single replicate from the average over replicates, and the divisor weights for sample size in the corresponding distance class (large numbers of observations, i.e. pairs of individuals, produce a small variance, and reciprocally). For each parameter combination, the observed X_o^2 was obtained by replacing F_k values in equation (5) by their respective values observed in the real population. The smaller X_o^2, the better the simulations fitted the observation. Parameter combinations for which X_o^2 was larger than the 95% lowest X^2 values from independent replicates were considered incompatible with the common ash data, and therefore rejected (one-sided test).

Linear regression. For each independent simulation replicate in each parameter combination and for the real population, kinship coefficients F_k in each distance class k were plotted against the logarithm of the distance (measured in m). The slope b_r of the linear regression was computed following equation (2), with F_k instead of $\overline{F}_k$ (the whole distance range was considered). The slope b_{ro} of the observed kinship plot was compared to the 100 ranked b_r values from independent simulation replicates for each parameter combination. A parameter combination was considered incompatible with the observation when b_{ro} was smaller than the third or larger than the 98th ranked b_r value; i.e. lying outside of a 96% confidence interval (two-sided test).

Polynomial fit of residuals. Residuals from the linear regression described above were computed for each independent simulation replicate in each parameter combination and for the real population as in equation (3) with F_k replacing $\overline{F}_k$. After standardising the scale following equation (4), a polynomial function involving the third power of the logarithm of distance, $y=a+bx+cx^2+dx^3$, was fitted to the plot of residuals against distance, weighting each distance class by its number of observations. As we observed that the relative contribution of seed *vs.* pollen dispersal was closely related to the coefficient d of the polynomial (see results), the coefficient d_o of the term in x^3 in the real population was compared to the 100 ranked values of d of independent replicates in each parameter combination. A parameter

Table 1 Statistics of genetic diversity of microsatellite loci in the overall sample.

Locus	*K*	*H*	F_I[a]
M2-30	42	0.918	0.043
FEMSATL4	37	0.880	0.007
FEMSATL11	32	0.918	0.003
FEMSATL16	10	0.663	0.165
FEMSATL19	27	0.568	-0.066

K, total number of alleles; *H*, gene diversity; F_I, Wright's inbreeding coefficient.
[a] Exact tests for departure from Hardy-Weinberg genotypic proportions were non-significant for all loci.

Table 2 Genetic diversity statistics within subpopulations and the total population.

Population	n	*A*	*H*	F_I[a]
Subpopulation 1	30	13.2	0.799	-0.011
Subpopulation 2	30	14.8	0.765	0.044
Subpopulation 3	30	13.8	0.773	0.045
Subpopulation 4	32	16.6	0.807	0.060
Subpopulation 5	30	14.8	0.794	-0.022
Mean subpopulations	30.4	14.6	0.788	0.023
Total population	152	29.6	0.792	0.029

n, sample size; *A*, average number of alleles per locus; *H*, gene diversity; F_I, Wright's inbreeding coefficient.
[a] Exact tests for departure from Hardy-Weinberg genotypic proportions were non-significant for all subpopulations and for the total population.

Table 3 Spatial genetic structure and estimates of gene dispersal.

Population	Pairs of individuals	b_{ro}	*P*-value	F_0	$\hat{N}b$	sampling area (ha)
Subpopulation 1	435	-0.0177	0.0201	-0.012	57.3	0.83
Subpopulation 2	435	-0.0165	0.0217	0.054	57.3	1.38
Subpopulation 3	435	-0.0039	0.2475	0.047	nc	3.69
Subpopulation 4	496	-0.0170	0.0241	0.064	55.1	1.76
Subpopulation 5	435	0.0005	0.5001	-0.022	nc	0.71
Mean subpopulations	2236	-0.0106	0.0019	0.027	92.2	1.68
Total population	11476	-0.0019	0.0038	0.031	518.5	~710

b_{ro}, slope of the regression of pairwise kinship coefficients on the logarithm of geographic distance; *P*-value of the one-sided Mantel test with H0: observed $b_{ro} \geq 0$; F_0, average kinship coefficient between adjacent individuals; $\hat{N}b$, estimate of neighbourhood size; nc, not computed because the Mantel test was not significant

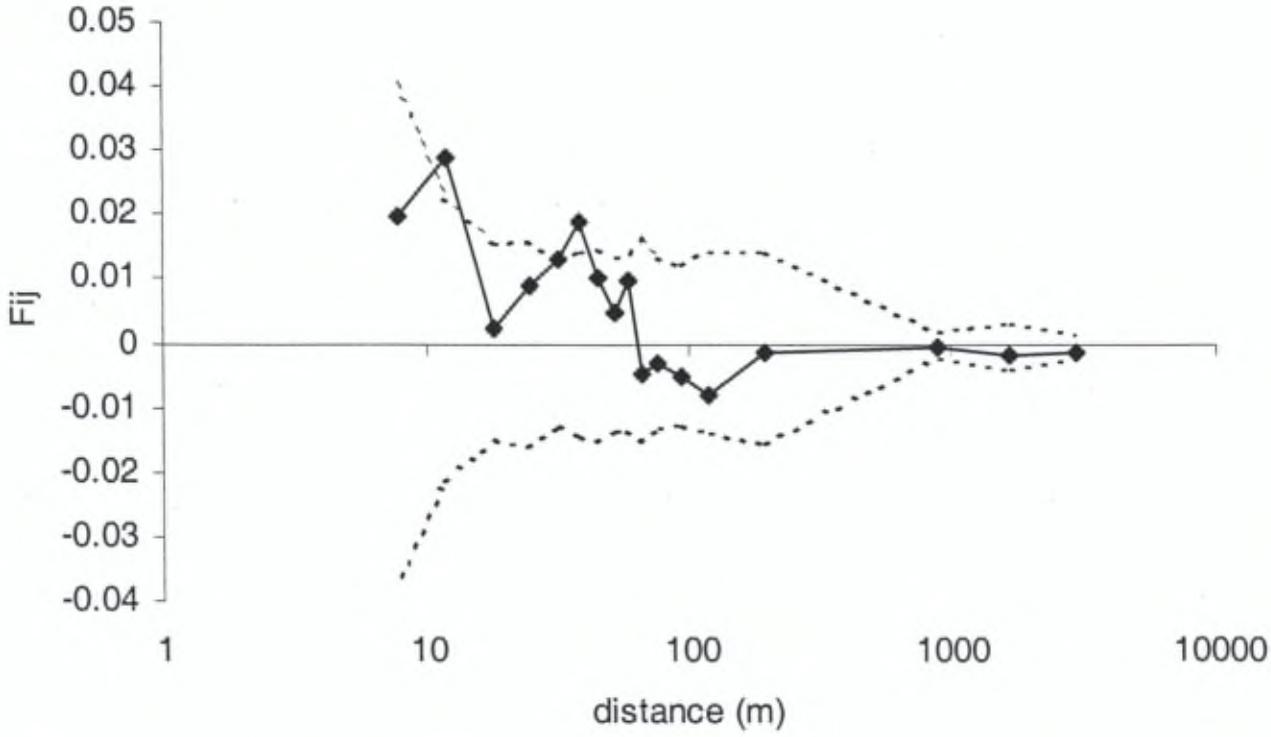


Fig. 2 Average kinship coefficients F_{ij} between pairs of individuals plotted against the logarithm of geographical distance in the whole population. Dashed lines represent 95% confidence intervals for F_{ij} under the null hypothesis that genotypes are randomly distributed.

combination was considered incompatible with the observation when d was smaller than the third or larger than the 98th ranked d value (two-sided test).

To represent graphically the quality of the fit to the simulation results according to each of the three statistics described above, we attributed P-values for each parameter combination according to the ranking of the statistic obtained for the common ash data with those of the simulation replicates. P-values were plotted on a grid representing all 56 parameter combinations and contours of $P = 0.05$, $P = 0.2$ and $P = 0.5$ were interpolated automatically with Microsoft Excel. Parameter combinations that had not been rejected in any of the three tests defined the space of parameters compatible with the common ash data.

Results

Microsatellite analysis of the common ash population

Allelic diversity of microsatellite loci. The total number of alleles per microsatellite locus in the total sample of 152 individuals was comprised between 10 and 42, with an overall total of 148 alleles scored over the five loci (Table 1). Gene diversity H ranged from 0.568 to 0.918 per locus, and Wright's inbreeding coefficient F_I from –0.066 to 0.165 (Table 1). The exact test for departure from Hardy-Weinberg genotypic proportions was non-significant for all loci.

Genetic diversity within subpopulations and the total population. Genetic diversity statistics for subpopulations and the total population are given in Table 2. Average number of alleles per locus A and gene diversity H were similar in all subpopulations; they ranged respectively from 13.2 to 16.6 and from 0.765 to 0.807. For the total population, $A = 29.6$ and $H = 0.792$. Wright's inbreeding coefficient F_I was comprised between –0.022 and 0.06 within subpopulations and $F_I = 0.029$ for the total population. No departure from Hardy-Weinberg genotypic proportions was detected.

Spatial genetic structure. In agreement with models of isolation by distance, a significant linear decrease of pairwise kinship coefficients F_{ij} with the logarithm of increasing geographic distance was detected in subpopulations 1, 2, and 4 ($P < 0.05$) and in the total population ($P < 0.01$) (Table 3), although only few average kinship values per distance class lay outside the 95% confidence interval obtained from 10000 random permutations of locations among individuals (Fig. 2). The neighbourhood size $\hat{N}b$ could be estimated from equation (1) considering the regression slopes observed for the different subpopulations and the whole population. In the three subpopulations with a significant isolation by distance pattern, $\hat{N}b$ ranged from 55.1 to 57.3. The average regression slope observed within all five subpopulations resulted in $\hat{N}b = 92.2$. The regression applied on the whole population gave

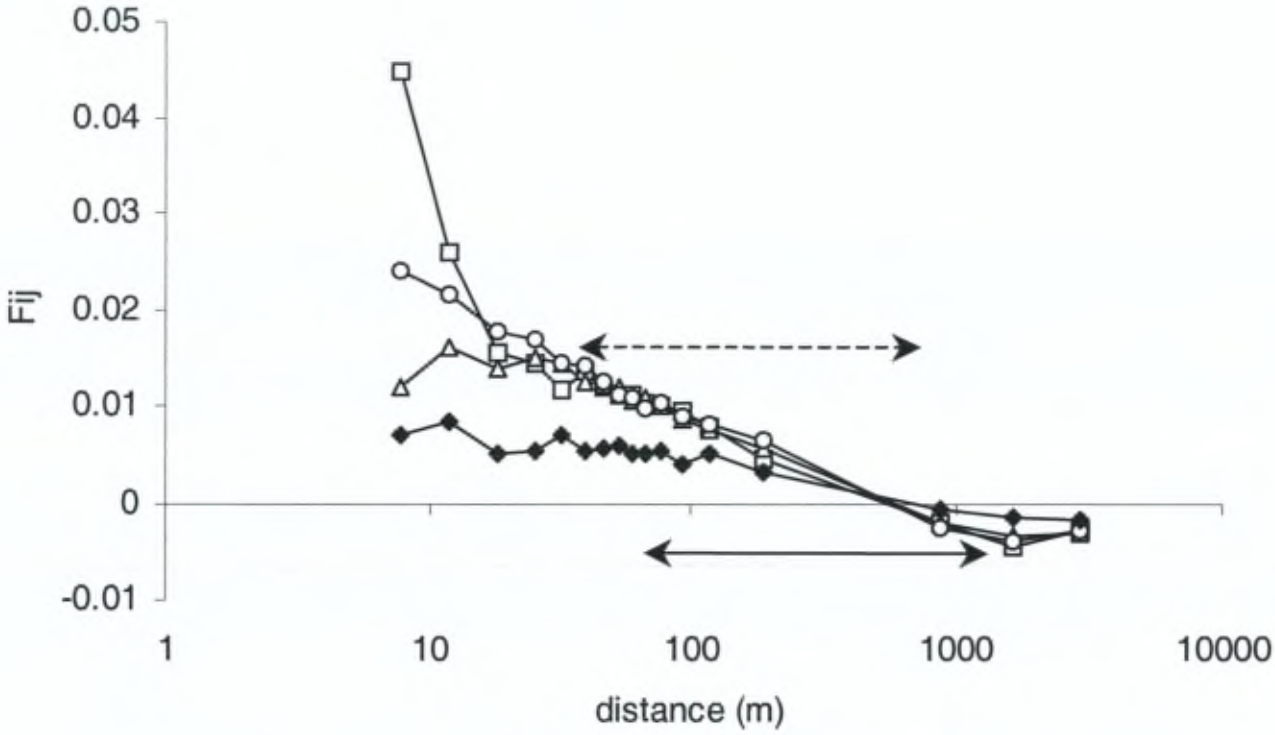


Fig. 3 Average kinship coefficients F_{ij} over 100 independent replicates of four simulated parameter combinations, plotted against the logarithm of distance. For squares, σ_s=5 m, σ_p=49 m; for circles, σ_s=14 m, σ_p=49 m; for triangles, σ_s=28 m, σ_p=35 m; for diamonds, σ_s=49 m, σ_p=49 m. Open symbols stand for parameter combinations with $\sigma_g \approx$ 35-37m, whereas filled symbols feature $\sigma_g \approx$ 60m. Arrows represent the distance range between σ and 20σ, where a linear decrease of F_{ij} is expected, the dashed arrow corresponding to σ_g=37 m, the solid arrow to σ_g=60 m.

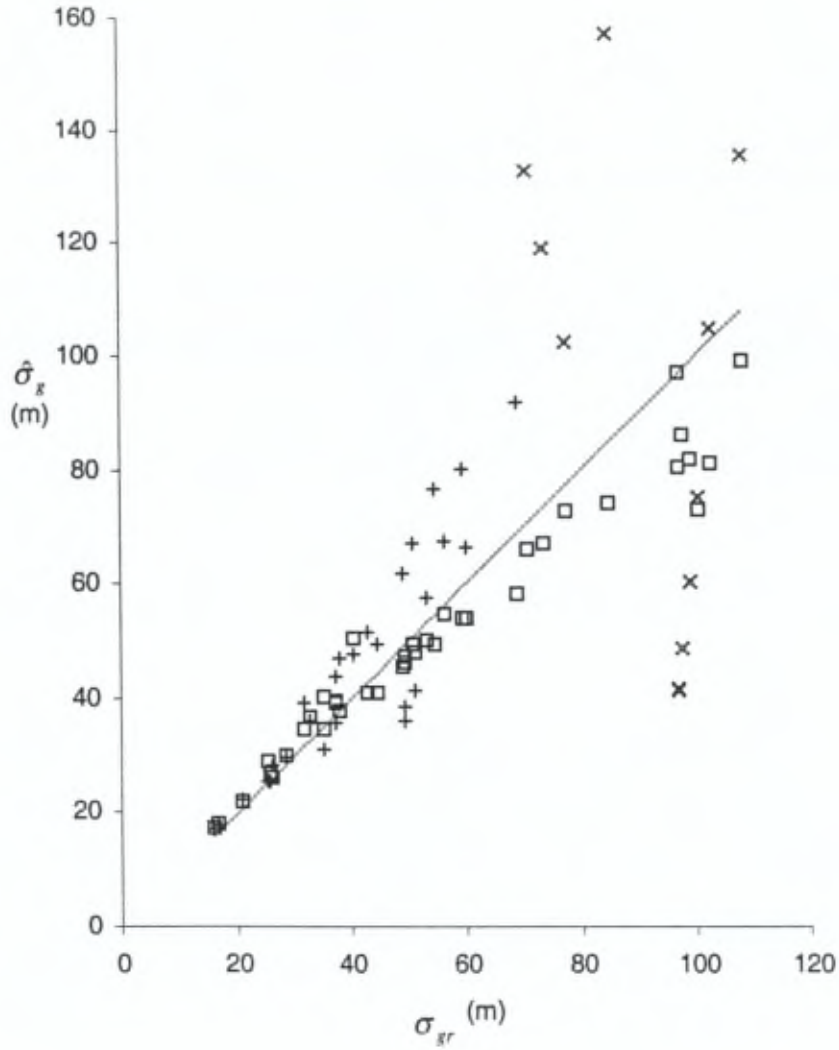


Fig. 4 Plot of the estimates of gene dispersal $\hat{\sigma}_g$ computed from the regression slopes b_r of kinship-distance plots from simulated datasets against the realised σ_{gr}. The line shows the expected relationship for an unbiased $\hat{\sigma}_g$ estimate. When regression is performed over an adequate range (σ_g to 20 σ_g, Rousset 1997, 2000), $\hat{\sigma}_g$ estimates are little biased (squares). When regression is done within subpopulations, $\hat{\sigma}_g$ estimates are little biased under narrow dispersal (σ_g<70m, + symbols), but can become substantially biased under wide-ranging dispersal (σ_g>70m, x symbols).

$\hat{N}b$ = 518.5, an estimate five times larger than at short spatial scale (Table 3). Inferred neighbourhood sizes within subpopulations are in agreement with those previously obtained at a similar spatial scale in Bulgaria ($\hat{N}b$ = 37.9 - 66.3, Heuertz *et al.* 2001). Note that regressions were applied over the whole available distance ranges, which did not necessarily correspond to the adequate distance range for *Nb* inference based on equation (1).

Computer simulations

Analysis of simulated data sets. Within the distance range comprised between σ_g and $20\sigma_g$, average kinship coefficients decreased linearly with the logarithm of geographic distance (Fig. 3), as verified by the average value of the coefficient of determination, $\overline{R}^2 = 0.968 \pm 0.044$ (SD), for parameter combinations with $\sigma_p \leq 140$m. For combinations with larger σ_p, the regression could not be computed because the upper bound, $20\sigma_g$, lay outside the simulated distance range. Wide-ranging gene dispersal ($\sigma_g \approx 60$m, filled symbols in Fig. 3) produced shallower slopes than narrower dispersal ($\sigma_g \approx 37$m, open symbols). Linear regression over the total distance range resulted in $\overline{R}^2 = 0.737 \pm 0.272$ (SD) for all parameter combinations and $\overline{R}^2 = 0.875 \pm 0.100$ (SD) for those with $\sigma_p \leq 140$m, showing that deviation from linearity is substantially more pronounced than within the σ_g to $20\sigma_g$ distance range. Interestingly, deviations from the linear relationship at a distance shorter than σ_g depended much on the relative contributions of pollen and seed dispersal, as illustrated by three parameter combinations with similar total gene dispersal ($\sigma_g \approx 37$ m; Fig. 3): narrow dispersal of seeds as compared to pollen produced an upward concave initial curving (squares in Fig. 3); and, as the relative seed dispersal contribution increased, the shape of the curve flattened (circles) before becoming downward concave (triangles). It is noteworthy that, in agreement with theoretical expectations, the three parameter combinations with contrasting σ_s and σ_p but very similar σ_g values show the same pattern of decrease of kinship coefficients within the σ_g – 20 σ_g distance range (Fig. 3).

Estimates of standard deviations of gene dispersal distances $\hat{\sigma}_g$ using slopes b_r computed over the adequate distance range (σ_g to 20 σ_g, Rousset 1997, 2000) were close to the expected values (Fig. 4). When the regression was applied on a too short distance range, such as within common ash subpopulations, substantial errors appeared on the estimation of σ_g (Fig. 4): dispersal was underestimated when the seed *vs.* pollen contribution was small (i.e. steep negative slope at short distances, Fig. 3); and it was overestimated when the relative seed contribution was important (shallower slope at short than at large distance, Fig. 3).

After subtraction of the linear regression slope considering the whole distance range, polynomial functions of the third degree of the logarithm of distance were fitted to the

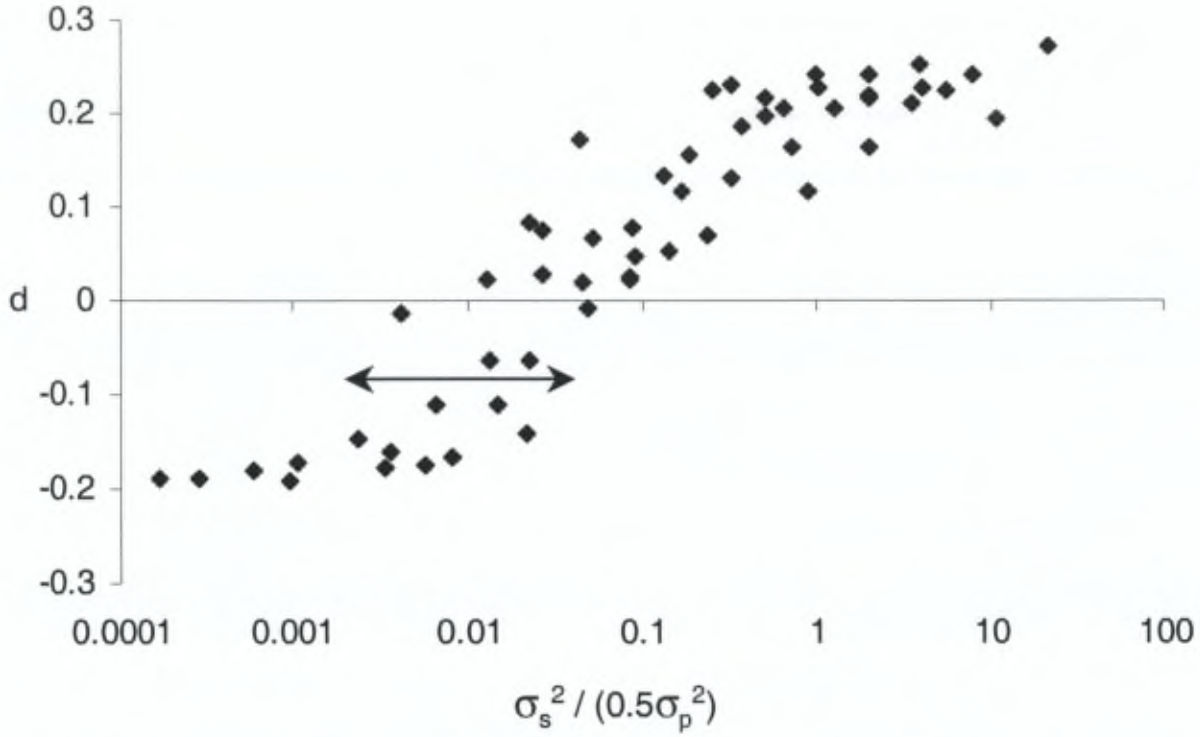


Fig. 5 Shape parameter *d* of the kinship-distance plots according to the ratio of seed *vs.* pollen contributions to total gene dispersal, $\sigma_s^2/(0.5\sigma_p^2)$. The quantity *d* corresponds to the coefficient of the term of third power of the polynomial functions ($y=a+bx+cx^2+dx^3$) that were fitted to the standardized residuals of kinship coefficients (see text for details). The arrow stands for the observed *d* value in the common ash population and the corresponding range of estimates of $\sigma_s^2/(0.5\sigma_p^2)$.

standardised residuals, giving a mean $\overline{R}^2$ over all parameter combinations equal to 0.700 (±0.201, SD). Cubic regression functions are indicators of the shape of plots of kinship against distance, independently of their slope. The form of these equations was $y=a+bx+cx^2+dx^3$, and all coefficients, *a*, *b*, *c*, and *d*, were correlated to the ratio of contributions of seed and pollen to total gene dispersal, $\sigma_s^2/(0.5\sigma_p^2)$. The narrowest relationship was observed for *d*, which corresponds to 1/6 times the third derivative of these functions (Fig. 5). Linear regression of *d* against $\sigma_s^2/(0.5\sigma_p^2)$ produced $R^2 = 0.846$ (Fig. 5), suggesting that *d* might be a good indicator of the relative contributions of seed and pollen to gene dispersal.

Analysis of the fit of the observed spatial genetic structure to that obtained from simulations

Chi-square-like test. A large proportion of simulated parameter combinations (36 out of 56 combinations) were compatible with the observed data (Fig. 6a, shaded area); the fit was rejected (α = 0.05) only when both pollen and seed dispersal were either narrow (approximately $\sigma_s \leq 21$ m and $\sigma_p \leq 35$ m), or wide-ranging (approximately $\sigma_s \geq 35$ m and $\sigma_p \geq$ 350 m); or, alternatively, when seed dispersal was narrow and pollen dispersal wide-ranging (Fig. 6a, white area). Relatively good fit ($P > 0.2$) was obtained for $\sigma_s \leq 35$ m and $\sigma_p = 70$ m on the one hand, and $\sigma_s = 49$ m and $\sigma_p \leq 49$ m on the other. These conditions correspond to a realised neighbourhood size approximately comprised between 200 and 1000.

Slope b_r. Wide-ranging pollen dispersal and a combination of narrow pollen and narrow seed dispersal were incompatible with the observed data (Fig. 6b). Overall, 31 of the 56 simulated parameter combinations remained compatible at $\alpha = 0.04$. A zone of relatively good fit (here: $P > 0.5$) was observed for roughly the same parameter combinations as in the Chi-square-like test (compare Figs. 6a and 6b).

Shape d. Generally, narrow pollen dispersal ($\sigma_p \leq 49$m) produced shapes of kinship plots differing from the observed data: although only 13 parameter combinations with $\sigma_p \leq 49$ m could be rejected at $\alpha = 0.04$ (Fig. 6c), several others were at the verge of significance. Good fit ($P > 0.5$) was observed for narrow seed dispersal ($\sigma_s \leq 14$ (21) m) and moderate to wide-ranging pollen dispersal ($\sigma_p \geq$ (70) 140m).

The direct comparison of the shape parameter in the common ash population, $d = -0.084$, with the relationship of *d* to seed *vs.* pollen dispersal contributions (Fig. 5) produced an estimate of $\sigma_s^2/(0.5\sigma_p^2)$ comprised between approximately 0.002 and 0.04. The resulting ratio of seed *vs.* pollen dispersal standard deviations was $7.1 \leq \sigma_p/\sigma_s \leq 31.6$.

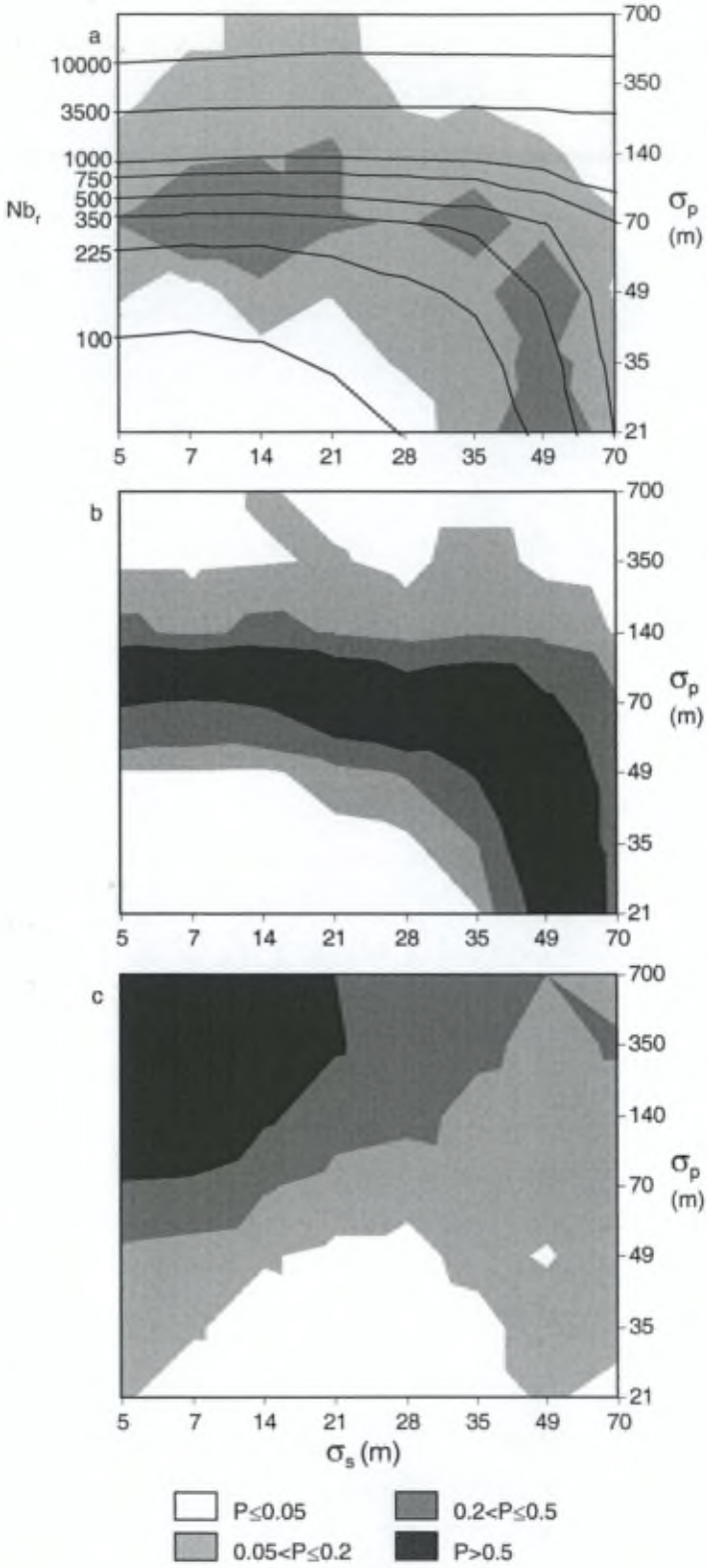


Fig. 6 Fit of kinship-distance plots from simulated data sets to the observed kinship-distance plot for three statistics: a, Chi-square; b, slope, b_r statistic; c, shape, d statistic. The axes represent the seed and pollen dispersal standard deviations used for simulations (scales not linear). Parameter combinations in white areas were rejected; darker shades represent an increasingly better fit of simulations with the observed data (see legend). The contours of shaded areas were interpolated. Lines of equal realised neighbourhood size Nb_r were also plotted in (a).

The comparison of the results from the three tests revealed that 24 simulated parameter combinations out of 56 remained compatible with the observed data. Regarding pollen, very narrow as well as very wide-ranging dispersal conditions were effectively rejected, intermediate values of roughly 70m $\leq \sigma_p \leq$ 140m being in good agreement with the observed data. The picture was less clear for seed dispersal; no type of dispersal could be efficiently rejected. However, good fit for both slope b_r and shape d was observed for intermediate pollen and narrow seed dispersal (Figs. 6b and 6c), delimiting an area of parameter combinations (70 m $\leq \sigma_p \leq$ 140 m and $\sigma_s \leq$ 14 m) that were best compatible with the observed data.

Discussion

In the Ploesti forest in southeastern Romania, we detected spatial genetic structure in common ash in three of the five subpopulations and over the area of the total population. In general, kinship coefficients among pairs of individuals decreased with increasing logarithm of geographic distance between individuals. However, this decrease was not uniformly linear over the whole range of distances, as reflected by steep regression slopes at short distance, and a shallower slope at large distance. Resulting estimates of the neighbourhood size were more than five times larger at the scale of the total population than within subpopulations. In order to find explanations for this observed pattern, possibly in terms of seed *vs.* pollen dispersal, we thoroughly examined the shape of kinship to distance plots from simulated datasets.

Estimating gene dispersal from spatial genetic structure by the regression approach

Simulations have shown that linearity of the decrease of kinship coefficients against the logarithm of distance (Rousset 1997, 2000) is verified when regression is performed over an adequate distance range (Fig. 3). The resulting estimates of gene dispersal are close to their expectations (Fig. 4), confirming the models of Hardy and Vekemans (1999) and Rousset (2000).

However, substantial errors occurred on the estimation of gene dispersal when regression was performed over a too short distance range (Fig. 4). Using the proper interval for regression is thus essential for a reliable estimate of *Nb*. As σ_g is usually unknown, the following approach can be attempted to determine the right interval: (i) estimating σ_g from the first *Nb* estimate, knowing *De*; (ii) estimating *Nb* again by performing linear regression over an interval of $\hat{\sigma}_g$ to $20\hat{\sigma}_g$; and (iii) repeating the procedure iteratively until the estimates $\hat{\sigma}_g$ and $\hat{N}b$ stabilise (e.g. Fenster *et al.* in preparation). When this method was initialised on the entire dataset

using within-subpopulation dispersal estimates, successive $\hat{\sigma}_g$ shifted repeatedly to larger values, suggesting that the within-subpopulation scale is too narrow to estimate *Nb*. However, even at the level of the whole population, $\hat{\sigma}_g$ did not stabilise because the *a posteriori* regression slope became positive for $\hat{\sigma}_g \geq 53$m or $\hat{Nb} \geq 395$ (using $D_e = 0.56D$). This was attributed to stochastic variation of the data (see below). Hence, the suggested approach to determine the relevant interval for estimating *Nb* could not be applied to the current data set. Application to other data sets will be needed to test the usefulness of the approach.

Impact of seed vs. pollen dispersal on the observed spatial genetic structure at nuclear loci

The kinship-distance plot in the entire common ash population was characterised by a markedly sharper decrease of kinship at short than at long distances. One possible explanation is restricted seed *vs.* pollen dispersal as revealed by our simulations (Fig. 3), where we assumed normal dispersal distributions. More generally, simulation results showed that a steep decrease of kinship at short distance occurs when gene dispersal follows a highly leptokurtic distribution (O. J. Hardy, unpublished data). Highly leptokurtic gene dispersal distributions result, firstly, from situations where seeds are substantially less dispersed than pollen; i.e. half the genes move over short distances (the maternal ones), and the other half can move over large distances (the paternal ones), producing a leptokurtic composite distribution. If in that case, seed and pollen dispersal distributions are not normal as we assumed, but for example exponential, the adequacy of our method to estimate the ratio of seed *vs.* pollen dispersal still needs to be tested. Secondly, highly leptokurtic gene dispersal distributions also occur when both seed and pollen dispersal distributions are (i) equal and (ii) highly leptokurtic themselves. In that case, our method would produce a wrong estimate of seed *vs.* pollen dispersal, as it assumes normality. Wind-dispersed pollen and seed, which are both characteristic for the common ash, may follow leptokurtic dispersal functions, as suggested in the literature. For instance, pollen deposition patterns in anemophilous plants were explained through leptokurtic dispersal functions (e.g. Levin & Kerster 1974, Ellstrand 1992, Richards 1997), and paternity analyses have often left high proportions of offspring with no potential father within the study population, demonstrating a strong long-distance component in wind-mediated pollen dispersal (Burczyk & Prat 1997, Streiff *et al.* 1999, Vassiliadis *et al.* 2002). Indications for leptokurtic dispersal of wind-borne winged seeds are weaker, referring mainly to an important deposition close to the source detected in direct studies (e.g. Levin & Kerster 1974, Johnson 1988). Wind-dispersed seed of tree species are not expected to move as far as pollen, because of their generally much greater mass. More restricted dispersal of wind-borne seed compared to pollen was also confirmed by a stronger spatial genetic structure at maternally than paternally inherited markers in conifer species

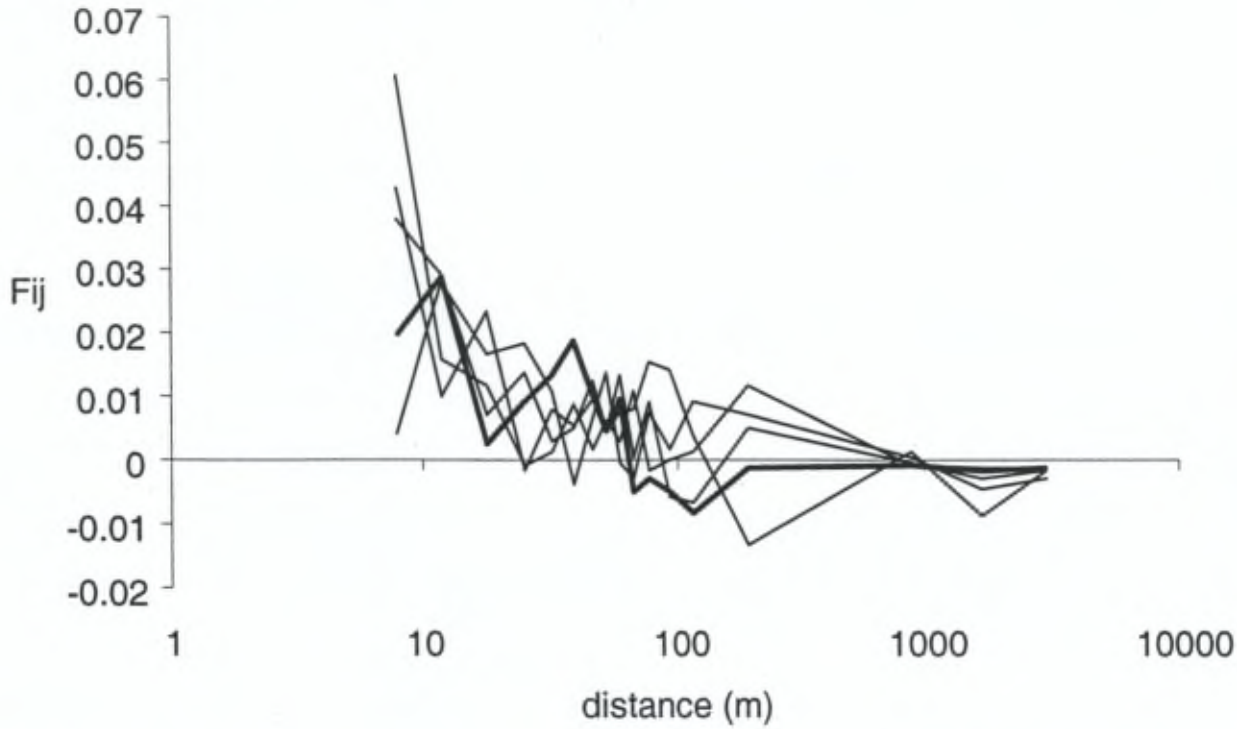


Fig. 7 Stochasticity in plots of average kinship coefficients F_{ij} against the logarithm of distance: observed values in the common ash population (bold line) and four independent simulation replicates for σ_s=7m and σ_p=70m (plain lines)

(Latta *et al.* 1998, Liepelt *et al.* 2002). Hence, although it is likely that dispersal functions of pollen and seed in common ash may contain a leptokurtic component potentially biasing our estimate of relative dispersal levels, we think that most of the leptokurtosis of gene dispersal is due to the relative difference between mean seed and pollen dispersal distances.

Implications of the shape of the kinship-distance plot on the estimate of dispersal

In the case of the common ash data, the steep negative initial slope of the kinship-distance plot and the demonstration that the within-subpopulation scale is inadequate for the regression approach indicate that we most certainly underestimated *Nb* at short distance.

In general, departures from linearity of kinship-distance plots, possibly originating from unequal seed *vs.* pollen dispersal, result in a non-random distribution of residuals from linear regression. The latter can be approximated by cubic regression functions whose coefficients, and notably the coefficient d of the term of third power of the logarithm of distance, are strongly correlated to the relative contributions of seed *vs.* pollen to total dispersal. Hence, in the context of a particular data set, d can provide an estimate of the relative magnitudes of seed and pollen dispersal, independently of the slope of the kinship-distance plot. The effect of seed and pollen on d is especially verified for seed contributions 1000 times smaller up to equalling pollen contributions (Fig. 5). Larger seed than pollen contributions probably do not reflect a biologically realistic situation in wind-pollinated and wind-dispersed species like common ash, and seed contributions more than 1000 times smaller than pollen contributions most likely represent rather isolated cases of long-distance pollination.

Power of methods and stochastic variation

The analysis of the fit of the observed spatial genetic structure to that obtained from simulations revealed that the decomposition of the dispersal function into a slope and a shape component allowed to withdraw more information than considering simply the overall departure using a Chi-square statistic; i.e. the combined tests on the slope b_r and the shape d were able to reject 31 parameter combinations at $\alpha = 0.04$, compared to the Chi-square-like test, which rejected only 20 at $\alpha = 0.05$. Notwithstanding, a large range of parameter combinations remained compatible with the observation. Although more simulation replicates (200 or 500 instead of 100) might have reduced the number of compatible combinations by rejecting those at the limit of significance, we argue that the main cause for weak rejection power is the stochasticity inherent to the data. Large fluctuations of kinship values were indeed found in the observed and the simulated data, despite the analysis of 5 highly polymorphic microsatellites in 152 individuals (Fig. 7). Such fluctuations could be related to the sampling scheme, however, simulations with other sampling locations but respecting the

same distance classes showed similar variation. The precision that can be reached on an *Nb* estimate based on spatial genetic structure is actually much dependent on the real *Nb* value itself: the stronger the spatial structure (i.e. the smaller the *Nb*), the smaller the coefficient of variation of the regression slope, hence the more precise the *Nb* estimate. The present study suggests that a larger sample size would have been necessary to infer gene dispersal with sufficient precision in common ash.

Estimates of dispersal in common ash

For the common ash, results from our investigations suggest a neighbourhood size $\hat{N}b = 519$, relative contributions of pollen *vs.* seed dispersal of $7.1 \leq \sigma_p/\sigma_s \leq 31.6$, and restricted seed and moderate pollen dispersal, with most probable parameters of, respectively, $\sigma_s \leq 14$m and 70m $\leq \sigma_p \leq 140$m. The latter would translate into a neighbourhood size estimate comprised between 363 and 1468 individuals when computed as $Nb = 4\pi D_e\left(\sigma_s^2 + 0.5\sigma_p^2\right)$.

These estimates of *Nb* are much higher than our previous results from Bulgaria ($\hat{N}b = 37.9$-66.3, Heuertz *et al.* 2001); but the latter estimates are likely biased downward because the samples were taken at too narrow distance. The Ploesti common ash population is an unusually dense one, as this species mostly occurs scattered in European mixed deciduous forests. The very restricted seed dispersal suggested here probably reflects the high vertical terminal velocities of ash seeds (1.2-1.7 m/s, Johnson 1988, Greene and Johnson 1995) dispersing under a relatively closed forest canopy where wind velocity and turbulence are reduced (Levin and Kerster 1974). Our estimate is smaller than that of Morand-Prieur *et al.* (in preparation) who found a mean seed dispersal distance of 125m in a parentage study in a French common ash population (which corresponds to $\hat{\sigma}_s$ = 99.7 m in case of a normal dispersal function, using distance $\hat{d} = \hat{\sigma}\sqrt{\pi/2}$). Various factors might contribute to explain this discrepancy, for instance a lower density and an elongated shape following a brook, with a preferential direction of seed dispersal in the latter population. Our own estimates of σ_s and σ_p might also be underestimated if the effective density in the real population was much lower than we assumed (i.e. $D_e/D = 0.5$). Our 7 to 32-fold farer pollen than seed dispersal estimate based on standard deviations compares favourably with the estimated ratio of 13 based on mean dispersal distances by Morand-Prieur *et al.* (in preparation). Estimates of pollen dispersal from the present study are slightly lower, but of the same order of magnitude than in the widely occurring European white oaks, where paternity analysis revealed σ_p = 140 m for *Q. petraea* and σ_p = 176 m for *Q. robur* (Streiff *et al.* 1999); and pollen sedimentation velocities in oaks and in ashes are similar (~ 3 cm/s, Jackson & Lyford 1999).

Concluding remarks

We have shown that our method permitted to obtain information on the relative contributions of seed and pollen to gene flow from the spatial genetic structure at nuclear loci alone. However, a large sample size and/or a large number of highly variable loci are required to overcome the stochasticity of the genetic structure and gain precision in the estimates, particularly when gene dispersal is extensive (low level of genetic structure). Further, the robustness of the method against other types of dispersal, e.g. leptokurtic, should be tested.

References

Berg EE, Hamrick JL (1995) Fine-scale genetic structure of a turkey oak forest. *Evolution*, **49**, 110-120.

Burczyk J, Prat D (1997) Male reproductive success in *Pseudotsuga menziesii* (Mirb.) Franco: the effects of spatial structure and flowering characteristics. *Heredity*, **79**, 638-647.

Chung MG, Chung MY, Oh GS, Epperson BK (2000) Spatial genetic structure in a *Neolitsea sericea* population (Lauraceae). *Heredity*, **85**, 490-497.

Chung MG, Epperson BK (2000) Clonal and spatial genetic structure in *Eurya emarginata* (Theaceae). *Heredity*, **84**, 170-177.

Crawford TJ (1984) The estimation of neighbourhood parameters for plant populations. *Heredity*, **53**, 273-283.

Doligez A, Baril C, Joly HI (1998) Fine-scale spatial genetic structure with non-uniform distribution of individuals. *Genetics*, **148**, 905-919.

Dong J, Wagner DB (1994) Paternally inherited chloroplast polymorphism in *Pinus*: estimation of diversity and population subdivision and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics*, **136**, 1187-1194.

Dutech C, Seiter J, Petronelli P, Joly HI, Jarne P (2002) Evidence of low gene flow in a neotropical clustered tree species in two rainforest stands of French Guiana. *Molecular Ecology*, **11**, 725-738.

Ellstrand NC (1992) Gene flow by pollen: implications for plant conservation genetics. *Oikos*, **63**, 77-86.

Epperson BK (1993) Recent advances in correlation analysis of spatial patterns of genetic variation. *Evolutionary Biology*, **27**, 95-155.

Epperson BK, Allard RW (1989) Spatial autocorrelation analysis of the distribution of genotypes within populations of lodge-pole pine. *Genetics*, **121**, 369-377.

Greene DF, Johnson EA (1995) Long-distance wind dispersal of tree seeds. *Canadian Journal of Botany*, **73**, 1036-1045.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation and population genetics models. *Heredity*, **83**, 145-154.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile compute program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618.

Heuertz M, Hausman JF, Tsvetkov I, Frascaria-Lacoste N, Vekemans X (2001) Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.). *Molecular Ecology*, **10**, 1615-1623.

Jackson ST, Lyford ME (1999) Pollen dispersal models in quaternary plant ecology: assumptions, parameters and prescriptions. *The Botanical Review*, **65**, 39-75.

Johnson WC (1988) Estimating dispersibility of Acer, Fraxinus and Tilia in fragmented landscapes from patterns of seedling establishment. *Landscape Ecology*, **1**, 175-187.

Kirby GC (1975) Heterozygote frequencies in small subpopulations. *Theoretical Population Biology*, **8**, 31-48.

Knowles P (1991) Spatial genetic structure within two natural stands of black spruce (*Picea mariana* (Mill.) B.S.P.). *Silvae Genetica*, **40**, 13-19.

Kremer A 1994 Diversité génétique et variabilité des caractères phénotypiques chez les arbres forestiers. *Genetics Selection Evolution*, **26** (Suppl.1), 105-123.

Latta RG, Linhart YB, Fleck D, Elliot M (1998) Direct and indirect estimates of seed versus pollen movement within a population of ponderosa pine. *Evolution*, **52**, 61-67.

Latta RG, Linhart YB, Mitton JB (2001) Cytonuclear disequilibrium and genetic drift in a natural population of ponderosa pine. *Genetics*, **158**, 843-850.

Levin DA, Kerster HW (1974) Gene flow in seed plants. *Evolutionary Biology*, **7**, 139-220.

Liepelt S, Bialozyt R, Ziegenhagen B (2002) Wind-dispersed pollen mediates postglacial geneflow among refugia. *Proceedings of the National Academy of Sciences of the USA*, **99**, 14590-14594.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understorey shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420-1425.

Malécot G (1950) Quelques schémas probabilistes sur la variabilité des populations naturelles. *Annales de l'Université de Lyon A*, **13**, 37-60.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics,* **89**, 583-590.

Picard JF (1982) Contribution à l'étude de la biologie florale et de la fructification du Frêne commun (*Fraxinus excelsior* L). *Revue Forestière Française,* **XXXIV**, 97-107.

Raymond M, Rousset F (1995) GENEPOP (version 12) Population genetics software for exact tests and ecumenicism. *Journal of Heredity,* **86**, 248-249.

Richards AD (1997) *Plant Breeding Systems* Second edition. Chapman & Hall, London.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219-1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58-62.

Schnabel A, Laushman RH, Hamrick JL (1991) Comparative analysis of population genetic structure of two co-occurring tree species, *Maclura pomifera* (Moraceae) and *Gleditsia triacanthos* (Leguminosae). *Heredity*, **67**, 357-364.

Sokal RR, Rohlf FJ (1995) Biometry: the principles and practice of statistics in biological research, 3rd ed. WH Freeman & Co, New York.

Sork VL, Huang S, Wiener E (1993) Macrogeographic and fine-scale genetic structure in a North American oak species, *Quercus rubra* L. *Annales des Sciences Forestières* **50(S)**, 261s-270s.

Streiff 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. *Molecular Ecology,* **7**, 317-328.

Streiff R, Ducousso A, Lexer C, Steinkellner H, Gloessl J & Kremer A (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Molecular Ecology,* **8**, 831-841.

Takahashi M, Mukouda M, Koono K (2000) Differences in genetic structure between two Japanese beech (*Fagus crenata* Blume) stands. *Heredity*, **84**, 103-115.

Ueno S, Tomaru N, Yoshimaru H, Manabe T, Yamamoto S (2000) Genetic structure of *Camellia japonica* L. in an old-growth evergreen forest, Tsushima, Japan. *Molecular Ecology*, **9**, 647-656.

Vassiliadis C, Saumitou-Laprade P, Lepart J, Viard F (2002) High male reproductive success of hermaphrodites in the androdioecious *Phillyrea angustifolia*. *Evolution*, **56**, 1262-1373.

Vekemans X, Lefèbvre C (1997) On the evolution of heavy-metal tolerant populations in *Armeria maritima*: evidence from allozyme variation and reproductive barriers *Journal of Evolutionary Biology*, **10**, 175-191.

Wallander E (2001) Evolution of wind-pollination in *Fraxinus* (Oleaceae) – an ecophylogenetic approach. PhD thesis, Göteborg University.

Wright S (1943) Isolation by distance. *Genetics*, **28**,114-138.

Xie CY, Knowles P (1991) Spatial genetic substructure within natural populations of jack pine (*Pinus banksiana*). *Canadian Journal of Botany*, **69**, 547-551.

Chapter III: Nuclear microsatellites reveal contrasting patterns of diversity in western and eastern European populations of the common ash (*Fraxinus excelsior* L.)

M. Heuertz, J.-F. Hausman, O. J. Hardy, G. G. Vendramin, N. Frascaria-Lacoste and X. Vekemans.

Abstract

Thirty-six common ash populations from Europe were genotyped at five nuclear microsatellite loci. Surprisingly, estimates of allelic richness and genetic diversity were lowest in regions close to putative ice age refuges, whereas most recently recolonised areas featured high values for both diversity statistics. Values of Wright's inbreeding coefficient and Cornuet and Luikart's T_2 statistic exhibited a similar geographic pattern; namely populations from recently recolonised regions showed comparatively higher inbreeding and a higher tendency for excess of heterozygosity relative to equilibrium heterozygosity expected from the observed number of alleles. Differentiation among populations was low within western and central European regions (F_{ST} = 0.016 and F_{ST} = 0.015, respectively), but high in the northeast and the southeast of the continent (F_{ST} = 0.090 and F_{ST} = 0.088, respectively). The application of a model-based clustering method and of a test for the information content of microsatellite allele sizes notably contributed to reveal details of the diversity patterns. An evolutionary scenario comprising processes of divergence among populations under stable demography and restricted gene flow in southeastern Europe and of admixture of previously differentiated gene pools during post-glacial recolonisation of western and central Europe was suggested.

Introduction

The patterns of genetic variation within plant species over a wide distribution range, like the western European continent, have been found to result from a combination of factors, some of which are intrinsic to the species life history (breeding system, modes of seed and pollen dispersal, life form, gregariousness), while others are perturbations induced by natural processes (ice ages, climatic stochasticity) or human impact (habitat fragmentation, global change).

For temperate tree species, a common observation is that they present very high genetic diversity at nuclear loci, but very little genetic differentiation among populations (Hamrick & Godt 1989, Hamrick *et al.* 1992). Many biological characteristics of temperate tree species have been put forward to explain this general pattern: long generation time, woody life form, high fecundity, outcrossing mating system, predominant wind pollination (Hamrick & Holden 1979, Loveless & Hamrick 1984). Indeed, some of these characteristics are expected to influence genetic variation and population genetic structure through their effects on the mutation-selection balance (Charlesworth *et al.* 1993), effective population size (Pollak 1987, Schoen & Brown 1991), and gene flow (Levin 1981).

Details of historical processes, such as the most recent post-glacial recolonisation events, have also been shown to imprint the genetic structure of several temperate tree species (Comes & Kadereit 1998, Newton *et al.* 1999, Hewitt 2000). In many species successive founder effects during recolonisation lead to a gradual decrease in genetic diversity from ancient glacial refuges towards recently recolonised areas (Tomaru *et al.* 1997, Ally *et al.* 2000, Ledig 2000). Populations founded by recolonisation from distinct refuges are often highly divergent (Lagercrantz & Ryman 1990, Konnert & Bergmann 1995), and regions where previously separated lineages merge often feature high diversity (Lagercrantz & Ryman 1990, Zanetto & Kremer 1995, Comps *et al.* 2001). Interactions between these historical events and biological characteristics of temperate tree species, such as potential for long-distance dispersal (Ibrahim *et al.* 1996, Le Corre & Kremer 1998) and an extended juvenile phase (Austerlitz *et al.* 2000), have probably contributed to a large extent to shaping their current diversity patterns. Human impact, through past and present management, i.e. multiplication of interesting phenotypes (in chestnut, Fineschi *et al.* 2000), harvesting within stands (i.e. Rajora *et al.* 2000) and forest fragmentation (i.e. Aldrich *et al.* 1998), has also strongly influenced the patterns of genetic diversity of tree species.

Large-scale patterns of nuclear genetic diversity in tree species have generally been investigated by analyzing the variation of population genetics statistics such as allelic richness, gene diversity and Wright's F_{ST} as a function of geographical variables like latitude, longitude and altitude. Such analyses revealed either similar (mountain hemlock, Ally *et al.* 2000) or contrasting (sessile oak, Zanetto & Kremer 1995; beech, Comps *et al.* 2001) trends

of variation in allelic richness and gene diversity over the investigated distribution area. Another approach is to study founder events associated to the recolonisation process by comparing actual gene diversity and gene diversity estimated from the observed number of alleles assuming mutation-drift equilibrium (Cornuet & Luikart 1996), knowing that bottlenecks produce transiently a stronger decrease in allelic richness than gene diversity (Nei *et al.* 1975). Results with this method reach from the detection of significant bottlenecks, i.e. in 19% of European beech populations, all located away from refuge areas (Comps *et al.* 2001); to the opposite pattern, compatible with recent population expansion rather than size reduction, i.e. in mountain hemlock (Ally *et al.* 2000).

Recently, a new approach has been proposed to identify patterns of genetic diversity: a model-based clustering method that uses multilocus genotypes to infer population structure and assign individuals to populations, allowing for population admixture (Pritchard *et al.* 2000). The method has successfully identified population structure and population admixture in several mammal species including humans (Rosenberg *et al.* 2001, Wilson *et al.* 2001) and cats (Randi *et al.* 2001) but has not yet been applied to large-scale investigations of population structure in tree species. This is possibly because hitherto most large-scale genetic surveys in tree species have been conducted with allozyme markers, which have limited resolution in identifying multilocus genotypes. Another limitation of allozymes is their low phylogeographic information content, making it difficult to distinguish alternative causes for diversity patterns, i.e. recent gene flow or recolonisation history (Zanetto & Kremer 1995). In this respect, nuclear haplotype data are becoming a promising new tool for evolutionary inference (Hare 2001). Alternatively, when using microsatellite markers, phylogenetic information may be contained in the distribution of allelic sizes. A useful test for detecting such information has been developed (Hardy *et al.* 2002).

The general objective of our study is to evaluate whether the use of microsatellite markers may substantially improve the identification of the factors responsible for the current genetic patterns in tree species. Microsatellites allow better estimates of population genetic statistics because their high polymorphism ensures low standard errors (Goudet *et al.* 1996) and they enable the application of new approaches of data analysis such as the model-based clustering method and a test of phylogenetic information content of allele sizes. Our model plant is the common ash, *Fraxinus excelsior*, a temperate tree species occurring in mixed deciduous forests from nearly all Europe, with the exclusion of the most southern and most northern parts. Common ash exhibits intermediate properties between a pioneer species and a permanent forest component; it features strong colonization capacity, but competitive ability is high only when ecological requirements are met. Flowers are wind-pollinated and the single-seeded fruits, the samaras, are wind-dispersed. The mating system of common ash is complex and variable: flowers are male, hermaphroditic or female and there is a continuum from pure male to pure female individuals with hermaphroditic intermediates (Wardle 1961,

Table 1 Sampling locations and sample sizes for common ash.

	Population	Acronym	Latitude	Longitude	n	Region[a]	Zone[a]
1	Camolin	IreCa	52°36′N	6°28′W	30	Ireland	West
2	Avova	IreAv	52°54′N	6°11′W	29	Ireland	West
3	Kilmacurra	IreKi	53°00′N	6°10′W	30	Ireland	West
4	Loch Tay	ScotTa	56°35′N	4°03′W	30	Scotland	West
5	Saint Gobain	FraGo	49°35′N	3°22′E	20	France	West
6	La Romagne	FraLr	49°40′N	4°19′E	20	France	West
7	Bremgarten	SwiB	47°20′N	8°18′E	30	Switzerland	Center
8	Ehrendingen	SwiG	47°29′N	8°21′E	30	Switzerland	Center
9	Eglisau	SwiE	47°35′N	8°31′E	30	Switzerland	Center
10	Tiroler Ache	GerTi	47°50′N	12°31′E	30	Germany	Center
11	Chiemsee	GerCh	47°48′N	12°31′E	29	Germany	Center
12	Freilassing	GerFr	47°50′N	12°59′E	30	Germany	Center
13	Rödjan	SweRoe	57°19′N	13°59′E	30	Sweden	Northeast
14	Ehd	SweEhd	57°14′N	13°59′E	30	Sweden	Northeast
15	Åkerås	SweAk	57°17′N	14°04′E	29	Sweden	Northeast
16	Zeimelis	LitZei	56°16′N	24°03′E	30	Lithuania	Northeast
17	Kaisiadorys	LitKai	54°53′N	24°22′E	30	Lithuania	Northeast
18	Slovakia	SloPa	48°34′N	19°08′E	26	Slovakia & Hungary	Southeast
19	Lillafüred Lusta-Völgy	HunLue	48°08′N	20°40′E	28	Slovakia & Hungary	Southeast
20	Ljulin monastir	BulLj	42°39′N	23°11′E	29	Western Bulgaria	Southeast
21	Vitosha	BulVi	42°38′N	23°14′E	32	Western Bulgaria	Southeast
22	Kokalyane monastir	BulMo	42°33′N	23°26′E	37	Western Bulgaria	Southeast
23	Tutuleac	RomTu	46°13′N	24°48′E	31	Romania	Southeast
24	Balota 1	RomBaa	44°50′N	26°04′E	30	Romania	Southeast
25	Balota 2	RomBab	44°50′N	26°04′E	30	Romania	Southeast
26	Balota 3	RomBac	44°50′N	26°03′E	30	Romania	Southeast
27	Golyamoto ravnishte	BulGr	42°50′N	26°03′E	36	Central Bulgaria	Southeast
28	Bukatchov chukar	BulBc	42°50′N	26°03′E	30	Central Bulgaria	Southeast
29	Elena	BulEl	42°50′N	26°04′E	30	Central Bulgaria	Southeast
30	Iri hissar	BulIh	43°51′N	26°46′E	20	Eastern Bulgaria	Southeast
31	Dulovo	BulDu	43°54′N	26°54′E	32	Eastern Bulgaria	Southeast
32	Kodga ormani	BulKo	43°54′N	26°54′E	36	Eastern Bulgaria	Southeast
33	Zli dol	BulZd	43°50′N	27°03′E	35	Eastern Bulgaria	Southeast
34	Hrjauca 1	MolHr	47°18′N	28°12′E	31	Republic of Moldova	Southeast
35	Hrjauca 2	MolHi	47°17′N	28°15′E	27	Republic of Moldova	Southeast
36	Hrjauca 3	MolHa	47°20′N	28°17′E	32	Republic of Moldova	Southeast

n, sample size per population.

[a] For the definition of regions and zones, see Material and Methods.

Picard 1982, Binggeli & Power 1999). Chloroplast haplotype data (G. G. Vendramin, in preparation) and fossil pollen records (Huntley & Birks 1983, Gliemeroth 1997, Brewer 2001) suggest glacial refuges for ash in the Balkans and Italy and/or the Alps, while refuges in Iberia and north of the Black Sea are less strongly supported. Postglacial recolonisation of continental Europe would have occurred from those refuges.

We sampled thirty-six populations of common ash in Europe and analysed them at five nuclear microsatellite loci with the objectives of (1) identifying geographical patterns of within- and among-population diversity, (2) clarifying genetic and phylogenetic relationships between regions using distance or model-based clustering methods and analyses of allelic size distributions, and (3) attempting to interpret the observed patterns in terms of originating evolutionary events with particular emphasis on discussing whether the properties of microsatellite markers added additional power in the evolutionary inference.

Materials and Methods

Plant material

Samples of Fraxinus excelsior were collected in 36 putatively autochthonous forests in Europe (Table 1, Figure 3). Samples consisting of buds or leaves were taken from an average of 30 non-adjacent trees within each population. Buds were shipped to the laboratory on their twigs, wrapped in wet paper. They were dissected out, separated from their scales and conserved at –70° prior to DNA extraction. Leaves were dried between paper sheets or in plastic bags in the presence of silica gel. They were kept at room temperature until DNA extraction.

DNA extraction

Total DNA was extracted from 60 to 90 mg of dry leaves or from 50 to 70 mg fresh weight of buds. The plant material was ground by hand or in the automatic grinding mill MM200 (Retsch) for extraction with the DNeasy Plant mini kit (Qiagen) or the CTAB procedure of the NucleoSpin Plant kit (Macherey Nagel). Alternatively, high throughput DNA extraction was performed by grinding simultaneously 192 samples of about 20 mg of dry leaves in the mill MM300 (Retsch) and subsequently extracting the DNA with the DNeasy 96 Plant Kit (Qiagen). Overall, the DNA extracts of ash performed with the NucleoSpin plant kit had to be less diluted for polymerase chain reaction (PCR) than the ones performed with the DNeasy kits, probably because of a better elimination with the former kit of metabolites susceptible to interfere with PCR.

Microsatellite analysis

Microsatellite analysis was performed as previously described (Heuertz *et al.* 2001): Five primer pairs of highly polymorphic microsatellite loci (table 2) were chosen on the basis of the quality of their banding pattern and used for PCR. Fluorescent labeling of the forward primers allowed detection of amplification products on an automated DNA sequencer (ABI PRISM® 377 DNA sequencer). Sizing of fragments was performed with the software programs Genescan® 3.1 and Genotyper® 2.5 from Applied Biosystems by comparison with an internal sizing standard (Genescan-350 Rox). The following modifications were applied to the previously published protocol: (1) when DNA was extracted with the NucleoSpin Plant kit (Macherey-Nagel), the quantity of template DNA in the PCR was increased to 3-10 ng, (2) samples which produced bad amplifications were reamplified with LA Taq polymerase (TaKaRa) and (3) polyacrylamide gels were produced with 10% LongRanger® gel solution (Sanver Tech).

Data analysis

Estimation of genetic variation within populations. The within-population statistics are presented as average values over populations grouped into twelve regions defined on the basis of geographic proximity and on an additional criterion of inclusion of between two to four populations per region (Table 1: Ireland, France, Switzerland, Germany, Sweden, Slovakia and Hungary, Lithuania, Romania, the Republic of Moldova, and eastern, western and central Bulgaria), and as averages over populations grouped into four main geographic zones (Table 1: west, comprising Ireland, Scotland and France; center, comprising Switzerland and Germany; northeast, comprising Sweden and Lithuania; and southeast, comprising Hungary, Slovakia, Romania, Bulgaria and the Republic of Moldova). The limits of the main geographic zones were chosen arbitrarily based on geographic proximity and on the additional criterion of inclusion of a minimum of five populations per zone.

The following statistics of genetic variation within populations were computed as averages over loci: (1) The allelic richness A quantifies allelic diversity with the property of being independent of sample size variation. It estimates the expected number of alleles A_n in a sub-sample of n genes, given that N genes have been sampled ($N \geq n$), according to the rarefaction method (Hurlbert 1971) adapted by El Mousadik and Petit (1996). The allelic richness within populations was computed with the software FSTAT version 2.9.3 (J. Goudet 2001, available from http://www.unil.ch/izea/softwares/fstat.html) as A_S_36 for a sub-sample of n = 36 gene copies or 18 diploid individual trees, which corresponds to the smallest sample size with a complete genotype at all five loci in a population (in population 30, BulIh). (2) The average expected heterozygosity or gene diversity H_E corrected for small sample size (Nei 1978) and (3) Wright's inbreeding coefficient F corrected for small sample size (Kirby 1975) were

computed with the program GEN-SURVEY (Vekemans & Lefèbvre 1997). Deviation of genotypic frequencies from Hardy-Weinberg proportions was tested with Fisher's exact test with the program GENEPOP version 3.3 (Raymond & Rousset 1995). Heterogeneity among local regions or zones in mean values of these statistics was tested with a permutation procedure with FSTAT version 2.9.3 (J. Goudet 2001). (4) The "bottleneck" statistic T_2 from Cornuet and Luikart (1996) measures the deviation in observed gene diversity (H_E) from the expected equilibrium value computed from the observed number of alleles (H_A), assuming mutation-drift equilibrium. The principle of the statistic is derived from the realization that a population suffering from a recent bottleneck experiences a larger reduction of its number of alleles at selectively neutral loci than of its gene diversity (Nei *et al.* 1975). Hence, positive values of T_2 indicate an excess in observed heterozygosity, consistent with a recent population bottleneck, whereas negative values are consistent with recent population expansion without immigration. T_2 was computed using the software BOTTLENECK version 1.2.02 (Cornuet & Luikart 1996) according to two mutation models, the Infinite Alleles Model (T_2 IAM), and the Stepwise Mutation Model (T_2 SMM), which represent two extreme models as our loci do not follow a strict SMM (i.e. we observe many alleles differing by one base pair, which is less than one repeat unit). We also report populations with significant deviation from equilibrium heterozygosity H_A as determined with the Wilcoxon signed rank test, which is the most appropriate test when only few polymorphic loci are analysed (Piry *et al.* 1999).

Analysis of linkage disequilibrium. Linkage disequilibrium was analysed for all pairs of loci in each population with GENEPOP version 3.3.

Estimation of total diversity and population genetic structure. Statistics of total diversity are given for the overall sample and for each of the twelve previously defined regions and the four zones. Total allelic richness A_T was computed standardised to a sample size of 36 gene copies (A_T_36), and relative differentiation among populations on the basis of allelic richness was determined as $A_{ST} = 1-(\overline{A}_S\_36-1)/(A_T\_36-1)$ as suggested by Comps et al. (2001). Total gene diversity H_T corrected for small sample size (Nei & Chesser 1983) and average absolute gene diversity between populations $D_{ST} = H_T - \overline{H}_E$ were computed with GEN-SURVEY. Relative differentiation values among populations, F_{ST} based on allele identity and R_{ST} based on allele size, were computed following an ANOVA approach (Weir & Cockerham 1984, for F_{ST}; Michalakis & Excoffier 1996, for R_{ST}) with the program SPAGeDi (Hardy & Vekemans 2002). This program also permitted to test for significance of the information content of allele sizes, i.e. whether R_{ST} was significantly different from F_{ST}, by means of a permutation procedure where allele sizes were permuted among alleles within populations (Hardy *et al.* 2002)

Correlation analysis between geographical variables and diversity statistics. The Pearson moment correlation coefficient was computed between the population latitude and longitude

Table 2 Allelic diversity of the nuclear microsatellite loci scored in the common ash.

Locus	Repeat motif	Range of sizes (bp)	K	H_T	F	F_{ST}	R_{ST}
M2-30	$(TG)_{15}$ $(AG)_{23}$ [a]	176-294	83	0.979	0.018	0.059	0.150
FEMSATL4	$(CA)_2$ $(AG)_{24}$ [b]	155-298	71	0.898	0.062	0.066	0.066
FEMSATL11	$(GA)_{20}$ $(TA)_4$ [b]	176-249	48	0.904	-0.018	0.067	0.034
FEMSATL16	$(CA)_3$ CG $(CA)_{10}$ [b]	170-210	16	0.433	0.203	0.093	0.079
FEMSATL19	$(CA)_6$ CGGC $(CA)_{13}$ [b]	142-230	57	0.949	-0.001	0.106	0.350

K, total number of alleles; H_T, total gene diversity; F, Wright's inbreeding coefficient; F_{ST}, relative differentiation based on allele identity; R_{ST}, relative differentiation based on allele size.
[a] from Brachet *et al.* 1999
[b] from Lefort *et al.* 1999

and the genetic diversity statistics presented above, as well as among diversity statistics. Significance levels are given after sequential Bonferroni correction (Rice 1989).

Patterns of differentiation among populations. Patterns of among-population differentiation were investigated with distance-based methods (1) and with a model-based clustering method (2). (1) Bootstrapped distance matrices involving pairwise population comparisons for Cavalli-Sforza and Edwards' (1967) chord distance and for Goldstein *et al.*'s (1995) $(\delta\mu)^2$ distance were computed with the software MICROSAT.C version 1.5e (E. Minch, A. Ruiz-Linares, D. Goldstein, M. Feldman & L. L. Cavalli-Sforza, 1995, 1996 available from http:// www.lotka.stanford.edu/microsat.html). Midpoint-rooted neighbor-joining trees were produced from those distance matrices with the software package PHYLIP version 3.5c (J. Felsenstein 1993 available from http://evolution.genetics.washington.edu/phylip.html) and visualised with TREEVIEW version 1.6.6 (R. D. M. Page 2001 available from http://taxonomy/zoology.gla.ac.uk/rod/rod.html). (2) Clusters of genetically similar diploid individuals were identified from their multilocus genotypes without prior knowledge of their population affinities with the computer program STRUCTURE (Pritchard *et al.* 2000). We ran the program for 100000 iterations after a burn-in period of 100000 with a number of clusters varying from one to eight, allowing for population admixture. For each individual, the fraction of ancestry from each of the clusters was estimated. The partition of the samples into clusters was computed as the averages over individuals belonging to a sample of the fractions of ancestry from each cluster.

Isolation by distance. Isolation by distance was investigated by computing the slope *Blog* of the linear regression of the $F_{ST}/(1- F_{ST})$ and $R_{ST}/(1- R_{ST})$ ratios among pairs of populations on the natural logarithms of their geographic distances (Rousset 1997) with SPAGeDi version 0.0. The significance of the isolation by distance pattern was tested by comparing the observed slope with values obtained after 1000 random permutations of population locations. The analysis was performed for the whole data set as well as for a subset of populations found to have a similar composition of clusters as determined with STRUCTURE, comprising the populations from Ireland, Scotland, France, Switzerland, Germany, Hungary, Slovakia and Lithuania.

Results

Allelic variation at microsatellite loci

The total number of putative alleles per locus in the overall sample of 1069 individuals ranged from 16 to 83, with an overall total of 275 alleles scored over the five loci (Table 2). The size ranges of the PCR products corresponding to these alleles are given in Table 2. Overall gene diversities (H_T) were very similar for each locus with the exception of FEMSATL16 that

Table 3 Genetic variation within populations of the common ash analyzed with nuclear microsatellite loci

Group[a]	N_p	$\bar{n}$	$\bar{A}_S$_36	H_E	F		T_2 IAM	T_2 SMM
Ireland	3	29.7	11.92 (4.66)	0.831 (0.144)	0.087 (0.068)	***	0.814	-3.261
France	2	20.0	12.21 (5.40)	0.840 (0.147)	0.121 (0.076)	***	1.167	-1.350
Switzerland	3	30.0	12.60 (5.76)	0.766 (0.311)	0.060 (0.020)	**	0.340	-3.300
Germany	3	29.7	11.51 (5.15)	0.795 (0.207)	0.014 (0.038)	**	0.858	-2.686
Sweden	3	29.7	8.65 (3.47)	0.685 (0.178)	-0.001(0.040)	ns	-1.397	-8.553
Slovakia & Hungary	2	27.0	11.80 (4.45)	0.825 (0.166)	0.043 (0.048)	***	0.701	-3.015
Western Bulgaria	3	32.7	7.34 (2.76)	0.684 (0.256)	-0.044 (0.025)	***	0.623	-3.172
Lithuania	2	30.0	14.19 (6.08)	0.850 (0.135)	0.090 (0.064)	***	-0.088	-4.576
Romania	4	30.3	10.97 (3.87)	0.753 (0.179)	0.015 (0.024)	ns	-1.332	-8.892
Central Bulgaria	3	32.0	10.40 (4.00)	0.747 (0.262)	0.056 (0.035)	*	-0.138	-5.256
Eastern Bulgaria	4	30.8	11.36 (4.63)	0.735 (0.312)	0.010 (0.012)	ns	-1.227	-7.068
Republic of Moldova	3	30.0	13.89 (5.64)	0.800 (0.267)	0.012 (0.027)	ns	-0.414	-4.657
West	6	26.5	12.23 (5.08)	0.832 (0.153)	0.090 (0.067)	***	0.925	-2.624
Center	6	29.8	12.06 (5.43)	0.780 (0.255)	0.037 (0.025)	***	0.599	-2.993
Northeast	5	29.8	10.87 (4.45)	0.751(0.160)	0.035 (0.050)	***	-0.873	-6.962
Southeast	19	30.6	10.94 (4.06)	0.752 (0.236)	0.014 (0.014)	***	-0.454	-5.743
Overall mean		29.7	11.33 (4.48)	0.770 (0.213)	0.033 (0.024)	***	-0.107	-4.934
S.D. (populations)		3.74	2.02	0.062	0.052		1.181	2.937

Diversity statistics were computed as multilocus statistics averaged over populations from the same region or zone. N_p, number of populations included in each group; $\bar{n}$, average sample size per population; $\bar{A}_S$_36, allelic richness for standardized samples of 36 gene copies (standard deviation over loci); H_E, gene diversity (standard deviation over loci); F, Wright's inbreeding coefficient (standard deviation over loci) and exact test of departure from Hardy-Weinberg genotypic proportions: *, P<0.05; **, P<0.01; ***, P<0.001; ns, not significant; T_2, bottleneck statistic from Cornuet and Luikart (1996) computed on the basis of the infinite alleles model (T_2 IAM) and the stepwise mutation model (SMM); S.D. (populations), standard deviation over populations.

[a] For the definition of groups, see Material and Methods

showed less then half the polymorphism of other loci. Little variation in F_{ST} (CV = 25.9%) but strong variation in F (CV = 168.8%) and R_{ST} (CV = 93.5%) were observed among loci (Table 2). In a previous publication concerning only 321 individuals from ten populations of Bulgaria (Heuertz *et al.* 2001), we observed an overall total of 184 alleles for the same set of loci. Total allelic richness standardised to the size of the Bulgarian data set is higher for the present study (A_T = 226), indicating that the higher allelic diversity detected is not only due to a larger sample size, but also to the larger geographic area investigated. In a study of 240 ash seedlings from 12 French populations, Morand *et al.* (2002) found 169 alleles for four of the loci that we used (all except FEMSATL16); omitting this locus, our present data set featured 259, and the Bulgarian set 174 distinct alleles.

Genetic variation within populations

Statistics of genetic variation are given as population averages for the twelve local regions and the four zones (Table 3). Correlation coefficients between geographical variables and within-population statistics are given in Table 4.

The mean allelic richness (A_S_36) within individual populations varied from 6.14 to 14.90 (standardised samples of 36 gene copies) with an overall mean of 11.33. Nei's gene diversity (H_E) varied from 0.650 to 0.853, with a mean value of 0.770. An overall significant departure from Hardy-Weinberg genotypic proportions (Fisher's exact test, $P < 0.05$) was observed in 15 out of 36 populations, two of which showed an excess and 13 a deficit of heterozygotes. Statistically significant values of Wright's inbreeding coefficient (F) were comprised between -0.028 and 0.157, and the overall mean value was 0.033, indicating a low overall deficit in heterozygotes within populations. Averages of the three statistics varied significantly among local regions (A_S, $P < 0.001$; H_E, $P < 0.01$ and F, $P < 0.001$), based on 1000 permutations of populations among regions. Among zones, variation was significant for H_E ($P < 0.05$) and marginally significant for F ($P = 0.051$). Gene diversity was significantly positively correlated to both A_S and F (Table 4). No correlation was found between latitude and any of the within-population statistics. In contrast, we observed a negative association between longitude and values of H_E (r = -0.345) and F (r = -0.425) but the correlation was non-significant after application of the sequential Bonferroni correction. Accordingly, high gene diversity was found in populations from Ireland and France, whereas several regions from eastern Europe showed lower values of H_E (Bulgaria and Romania). Values of Wright's inbreeding coefficient F were also found to be the highest in western populations (Ireland and France), whereas the lowest values occurred in southeastern Europe. Values of allelic richness did not show the same pattern as the highest values occurred in populations from Lithuania and the Republic of Moldova. However, average values for all three statistics (A_S, H_E and F) at the zonal level showed a similar overall pattern of a decrease from western to eastern Europe

Table 4 Correlation matrix between within-population diversity statistics and geographical variables.

	Lat	**Long**	**A_S_36**	**H_E**	**F**	**T_2 IAM**
Long	-0.569 **					
A_S_36	0.204	-0.122				
H_E	0.268	-0.345	0.823 ***			
F	0.290	-0.425	0.431	0.545 *		
T_2 IAM	0.052	-0.455	0.164	0.523 *	0.346	
T_2 SMM	0.015	-0.404	0.234	0.503 *	0.309	0.929 ***

Lat, latitude; Long, longitude; A_S_36, allelic richness for standardised samples of 36 gene copies; H_E, gene diversity; F, Wright's inbreeding coefficient; T_2 IAM, bottleneck statistic based on the infinite alleles model; T_2 SMM, bottleneck statistic based on the stepwise mutation model. Significance levels of correlation coefficients are given after sequential Bonferroni correction as *, P<0.05; **, P<0.01; ***, P<0.001.

Table 5 Statistics of regional population genetic structure.

Group	Populations	A_T_36	H_T	A_{ST}	D_{ST}	F_{ST}	R_{ST}	P(2-sided)
Ireland	3	12.79 (5.65)	0.841 (0.140)	0.074	0.009	0.015	0.039	ns
France	2	12.84 (6.22)	0.845 (0.151)	0.054	0.002	0.004	-0.015	ns
Switzerland	3	13.09 (6.92)	0.771 (0.312)	0.040	0.004	0.008	0.003	ns
Germany	3	12.37 (6.33)	0.802 (0.207)	0.076	0.007	0.013	0.018	ns
Sweden	3	10.45 (5.37)	0.766 (0.182)	0.191	0.081	0.153	0.077	ns
Slovakia & Hungary	2	12.84 (5.21)	0.840 (0.165)	0.088	0.014	0.033	0.059	ns
Western Bulgaria	3	10.94 (4.60)	0.752 (0.282)	0.362	0.068	0.126	0.138	ns
Lithuania	2	14.19 (6.60)	0.856 (0.132)	0.001	0.004	0.010	-0.003	ns
Romania	4	12.97 (4.95)	0.808 (0.160)	0.167	0.055	0.089	0.328	**
Central Bulgaria	3	12.18 (5.64)	0.782 (0.275)	0.159	0.034	0.064	0.097	ns
Eastern Bulgaria	4	13.81 (6.63)	0.765 (0.328)	0.191	0.030	0.048	0.060	ns
Republic of Moldova	3	14.48 (6.39)	0.801 (0.268)	0.044	0.002	0.003	0.000	ns
West	6	13.63 (6.40)	0.844 (0.154)	0.111	0.011	0.016	0.021	ns
Center	6	13.18 (6.87)	0.790 (0.258)	0.092	0.010	0.015	0.030	ns
Northeast	5	12.53 (6.05)	0.811 (0.162)	0.144	0.059	0.090	0.045	ns
Southeast	19	15.94 (7.17)	0.822 (0.255)	0.335	0.070	0.088	0.216	***
Overall	36	16.13 (7.66)	0.832 (0.266)	0.318	0.062	0.076	0.173	***

A_T_36, total allelic richness for standardized samples of 36 gene copies (standard deviation over loci); H_T, total gene diversity (standard deviation over loci); A_{ST}, differentiation computed from allelic richness; D_{ST}, absolute differentiation; F_{ST}, relative differentiation based on allele identity; R_{ST}, relative differentiation based on allele size; P(2-sided), test of information content of allele sizes from Hardy *et al.* (2002): ns, not significant; **, P<0.01; ***, P<0.001.

Table 6 Correlation matrix between total diversity and genetic structure statistics and geographical variables and within-population diversity statistics

	A_T_36	H_T	A_{ST}	D_{ST}	F_{ST}
Lat	-0.059	0.462	-0.466	-0.076	-0.041
Long	0.208	-0.329	0.319	0.296	0.278
A_S_36	0.904 **	0.650	-0.918 ***	-0.876 **	-0.887 **
H_E	0.662	0.923 ***	-0.847 *	-0.872 **	-0.862 **
F	0.417	0.753	-0.748	-0.677	-0.669
T_2 IAM	0.025	0.449	-0.285	-0.553	-0.514
T_2 SMM	0.111	0.382	-0.326	-0.652	-0.608

Differentiation statistics were computed among populations from twelve local regions (see Materials and Methods), and diversity statistics as averages over populations from each of those regions. A_T_36, total allelic richness for standardized samples of 36 gene copies; H_T, total gene diversity; A_{ST}, differentiation computed from allelic richness; D_{ST}, absolute differentiation; F_{ST}, relative differentiation based on allele identity; Lat, latitude; Long, longitude; A_S_36, allelic richness for standardized samples of 36 gene copies; H_E, gene diversity; F, Wright's inbreeding coefficient; T_2 IAM, bottleneck statistic based on the infinite alleles model; T_2 SMM; bottleneck statistic based on the stepwise mutation model. Significance levels of correlation coefficients are given after sequential Bonferroni correction as *, P<0.05; **, P<0.01; ***, P<0.001.

(Table 3). This pattern hides some local peculiarities, for instance very high genetic variation and high inbreeding coefficients were found in populations from Lithuania.

Linkage disequilibrium was detected ($\alpha = 0.05$) for one pair of loci in seven populations, two pairs in five populations, three pairs in three populations, four pairs in two populations, five pairs in four populations and all ten pairs in one population (12, GerFr from Germany). No correlation between the number of pairs of loci in linkage disequilibrium and within-population diversity statistics or geographical location was identified. For a given pair of loci, the number of populations with significant linkage disequilibrium was comprised between two and ten.

The two bottleneck statistics T_2 IAM and T_2 SMM were strongly positively correlated ($r = 0.935$, $P < 0.001$, Table 4) but the values of T_2 SMM were consistently lower than those of T_2 IAM. The lowest values for both statistics were observed in Romania (mean T_2 IAM = -1.332, mean T_2 SMM = -8.892) and Sweden (mean T_2 IAM = -1.397, mean T_2 SMM = -8.553), whereas the highest were recorded in France (mean T_2 IAM = 1.167, mean T_2 SMM = -1.350) and Germany (mean T_2 IAM = 0.858, mean T_2 SMM = -2.686)(Table 3). Similarly to the other intrapopulation statistics, T_2 decreased from western to eastern Europe. For T_2 IAM, a negative association with longitude was observed ($r = -0.455$, $P < 0.08$): the western zone featured a mean T_2 IAM ≈ 0.9, the eastern zones' mean T_2 IAM was ≈ -0.9 in the northeast and ≈ -0.5 in the southeast. Hence, western populations show an excess of heterozygosity relative to that expected from the observed number of alleles, which is consistent with a recent bottleneck. Conversely eastern populations show a deficit of heterozygosity explicable by population expansion. Wilcoxon's signed rank test ($\alpha = 0.05$) applied to T_2 IAM identified significant signatures of bottlenecks in seven populations (IreAv, 2 and IreCa, 3 from Ireland, FraGo, 5 and FraLr, 6 from France, GerTi, 10 from Germany and BulLj, 20 and BulMo, 22 from western Bulgaria) and of expansion in two populations (RomTu, 23 from Romania and BulVi, 21 from western Bulgaria). For T_2 SMM, the negative association with longitude was less marked ($r = -0.404$). All populations but one showed negative T_2 SMM values and consequently Wilcoxon's signed rank test identified no significant signal of bottlenecks, but 28 populations showed a significant signal of demographic expansion. Summarizing these results, it can be said that populations from western and central Europe showed a stronger tendency towards an excess heterozygosity than populations from eastern Europe.

Total diversity and population genetic structure

Statistics of total diversity and differentiation among populations were computed for each of the twelve local regions, for each of the four zones and at the European scale (Table 5). Correlation coefficients between total diversity and population structure statistics and both geographical variables and within-population statistics are presented in Table 6.

Total diversity in the overall data set was very high with an average of 16 different alleles per locus (A_T_36 = 16.13), when considering a standardised overall sample of 36 gene copies, and total gene diversity (H_T) equal to 0.832. At the regional level, total allelic richness (A_T) and total gene diversity (H_T) showed a geographical pattern similar to that described for the corresponding within-population statistics (A_S and H_E, respectively) to which they were significantly positively correlated. Accordingly, the highest values for A_T were observed in the Republic of Moldova and Lithuania, whereas the lowest values were found in Sweden and western Bulgaria. The latter two regions were also characterised by the lowest values for total gene diversity H_T, while the highest values of H_T occurred in western European populations and in Lithuania. At the zonal level, however, the patterns of variation of A_T and H_T were distinct from the longitudinal gradient described for within-population statistics, mainly because the southeastern zone exhibited low within-population diversity but very high total diversity.

The overall population differentiation measured as F_{ST} was 0.076, indicating that more than 92% of the genetic variation at nuclear microsatellite loci occurs within populations, whereas the overall allelic differentiation statistic, A_{ST}, showed a four-fold higher value (A_{ST} = 0.318). The population differentiation statistics A_{ST}, D_{ST} and F_{ST} varied widely at the regional level, and showed a pattern of variation very different from either the within-population or the total diversity statistics. Indeed, each of these differentiation statistics was significantly negatively correlated to both A_S and H_E. The observation that D_{ST} showed a very similar pattern to F_{ST} indicates that the negative correlation between F_{ST} and the within-population statistics was not an artifact due to the computation of F_{ST} as a ratio statistic. Regions with low genetic variation, i.e. Sweden and western Bulgaria, thus exhibited strong population genetic structure ($F_{ST} > 0.1$). In contrast, regions with high genetic variation like the Republic of Moldova, Lithuania and France showed low differentiation among populations ($F_{ST} < 0.01$). At the zonal level, all differentiation statistics displayed high values for the two eastern European zones, whereas populations from western and central Europe showed very little population genetic structure.

The differentiation statistic based on the distributions of allele size (R_{ST} = 0.173) was about two times higher than that based on allele identities (F_{ST} = 0.076) on the overall data set (Table 5). The difference between the two statistics was highly significant ($P < 0.001$), when permutations of allele sizes among alleles were performed. At the regional level, R_{ST} was found to be significantly positively correlated to F_{ST} ($r = 0.615$, $P < 0.05$), with values higher than F_{ST} in seven out of the twelve regions, but a significant difference between F_{ST} and R_{ST} occurred only in one region (Romania, $P < 0.01$). In Romania, the R_{ST} statistic was more than three-fold higher than the F_{ST}. In contrast, in Sweden, the R_{ST} statistic was half the value of F_{ST}. At the zonal level, values of the R_{ST} were higher than the F_{ST} in three out of four cases,

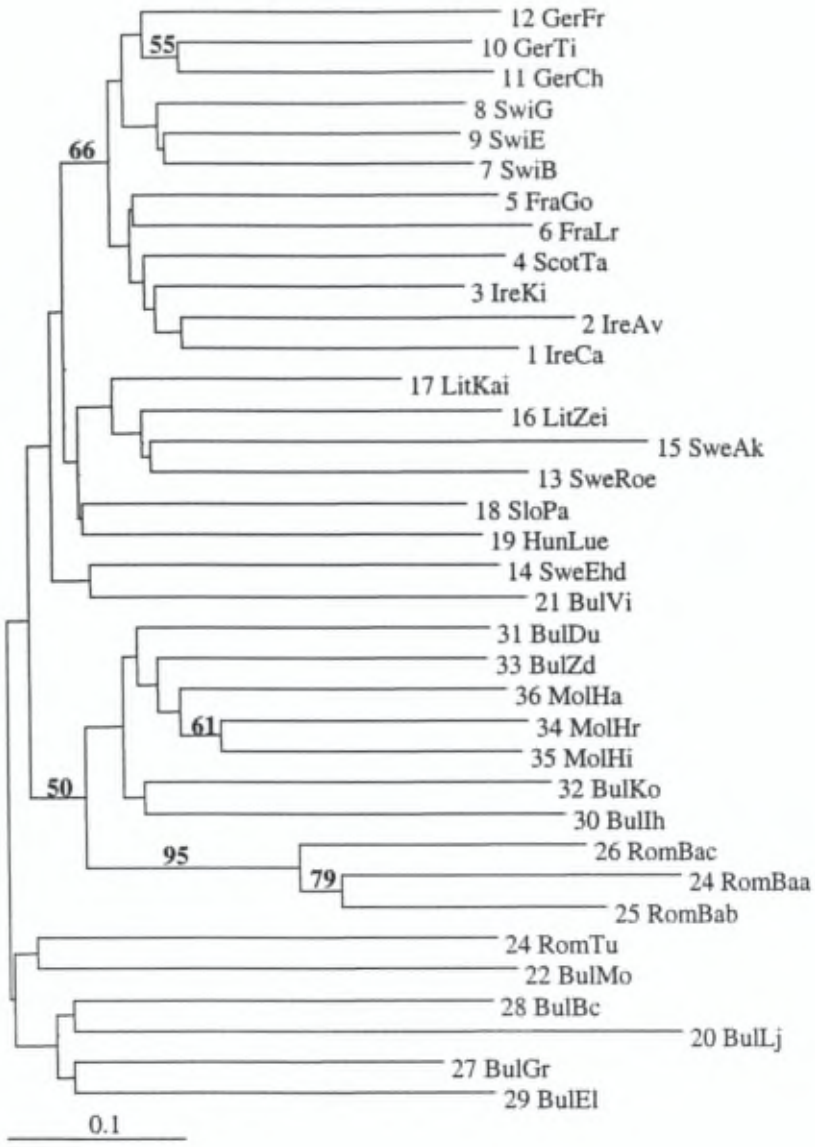


Fig 1 – Midpoint-rooted neighbour joining tree computed from Cavalli-Sforza and Edwards' (1967) chord distance among pairs of populations. The bootstrap support of the branches is given in bold numbers as the percentage out of 126 possible bootstraps.

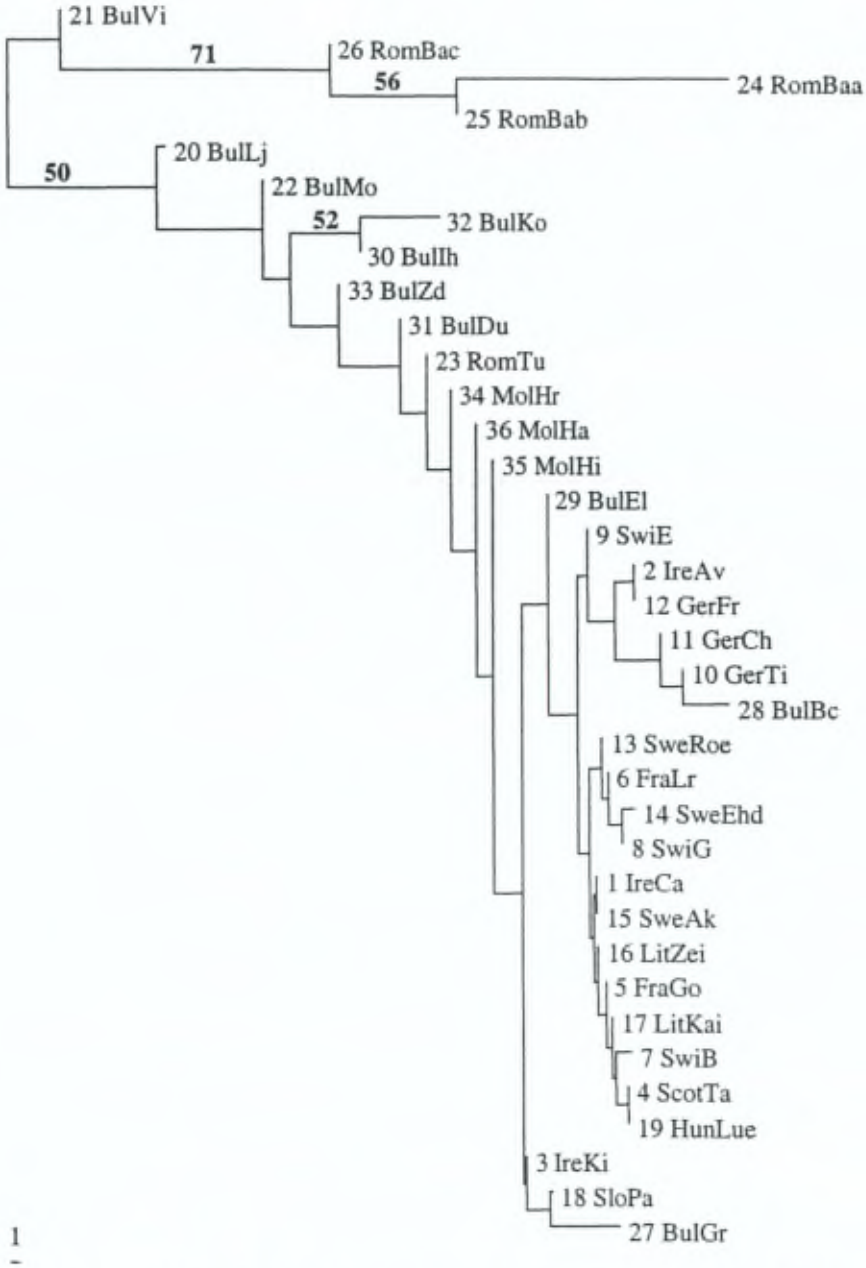


Fig 2 – Midpoint-rooted neighbour joining tree computed from Goldstein *et al.*'s (1995) $(\delta\mu)^2$ distance among pairs of populations. The bootstrap support of the branches is given in bold numbers as the percentage out of 126 possible bootstraps.

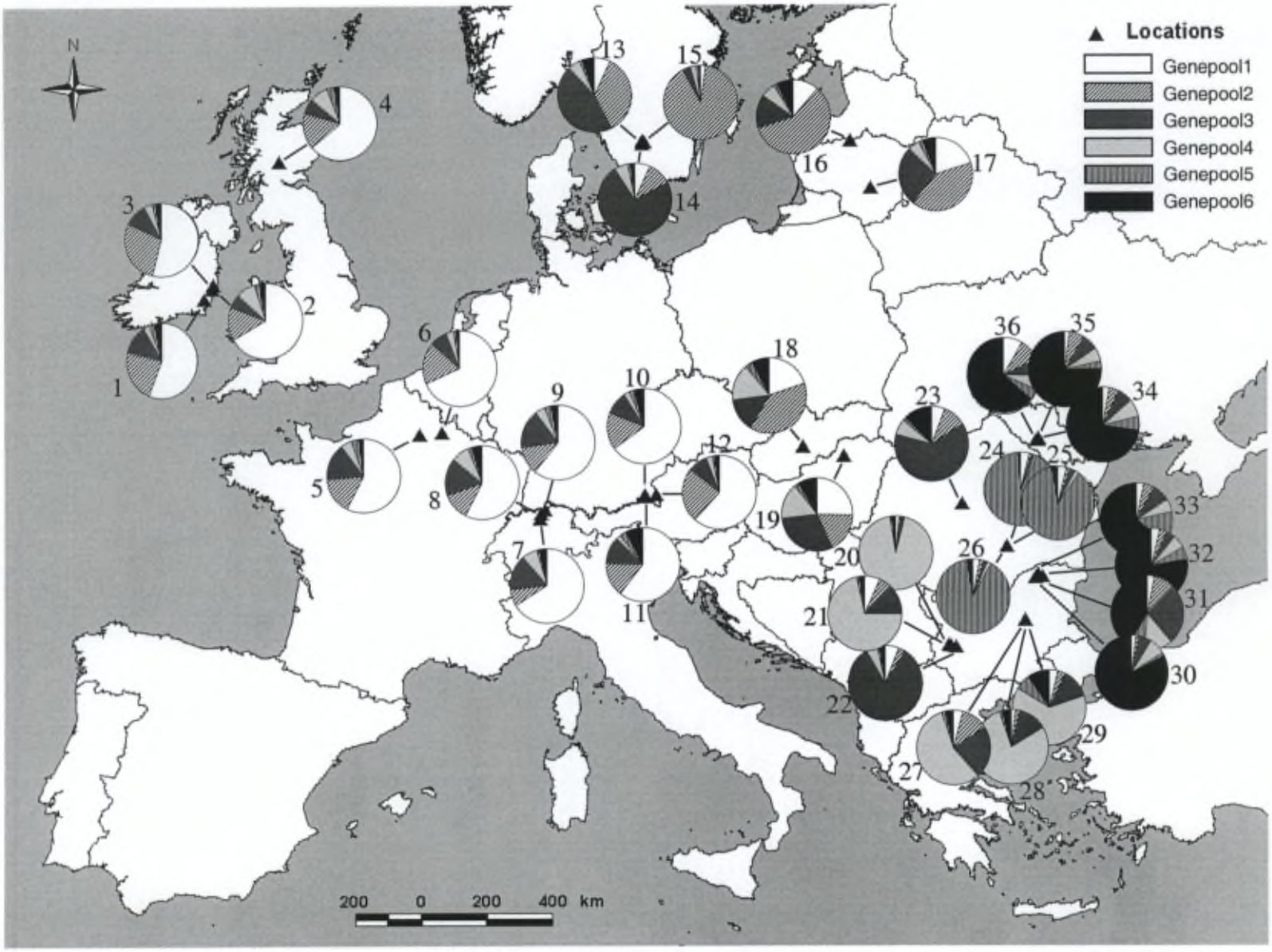


Fig 3 – Proportions of ancestry of each population in each of 6 gene pools defined with the model-based clustering method from Pritchard *et al.* (2000).

but a significant difference between F_{ST} and R_{ST} (P<0.001) occurred only in the southeastern group.

Patterns of differentiation among populations

Strong geographic differentiation was apparent from the midpoint-rooted neighbor joining tree based on Cavalli-Sforza and Edwards' (1967) chord distance (Figure 1). One major cluster contained all populations from the western and the central zones. It was close to populations from Sweden, Lithuania, Slovakia and Hungary and was inserted among clusters with populations from eastern Europe. Among eastern populations, those from the Danube and Dniestr plains, i.e. eastern Bulgaria, the Republic of Moldova and eastern Romania, were grouped together. A strikingly long internal branch separated the three populations from eastern Romania (24, RomBaa; 25, RomBab and 26, RomBac) from the rest of the tree. Because of the very limited number of loci, only 126 bootstraps could be performed. Very low support for internal branches was observed, with the separation of the western and central European group being given a score of 66% of bootstraps. The tree based on the $(\delta\mu)^2$ distance also isolated the three populations of eastern Romania trough a long internal branch, supported by 71% of the bootstraps (Figure 2). No clear geographical structure was found among the other populations.

Inference of population structure was obtained from an analysis of multilocus genotypes using the model-based clustering method of Pritchard *et al.* (2000) allowing for population admixture. The highest likelihood of the observed genotypic data X conditional to a given number K of clusters or gene pools was obtained for K=6, $\ln[\Pr(X|K=6)] = -27583$, and the posterior probability $\Pr(K=6|X)$ was very close to one. These six identified gene pools showed a strong geographic structure. A first gene pool was predominant in all populations from the western and central zones where 54 to 68% of the genes from each population were assigned to it (Figure 3). It was also present in Slovakia, Hungary and Lithuania. A second gene pool was mainly present in populations from Slovakia, Lithuania and two Swedish populations (SweRoe, 13 and SweAk, 15), whereas a third gene pool occurred in populations from Hungary, central Romania, population BulMo (22) from western Bulgaria and populations SweRoe (13) and SweEhd (14) from Sweden. A fourth gene pool was primarily found in all populations from western and central Bulgaria except in population BulMo (22). Populations from the Danube and Dniestr plains were assigned to two separate gene pools, the fifth gene pool comprising populations from eastern Romania and the sixth gene pool prevailing in those from the Republic of Moldova and eastern Bulgaria. Estimates of population admixture varied greatly among populations. On average the proportion of genes assigned to the best scoring gene pool within a population was $P_{best} = 0.66 \pm 0.15$ (S.D.). Lower than average population admixture was found within some geographic groups in

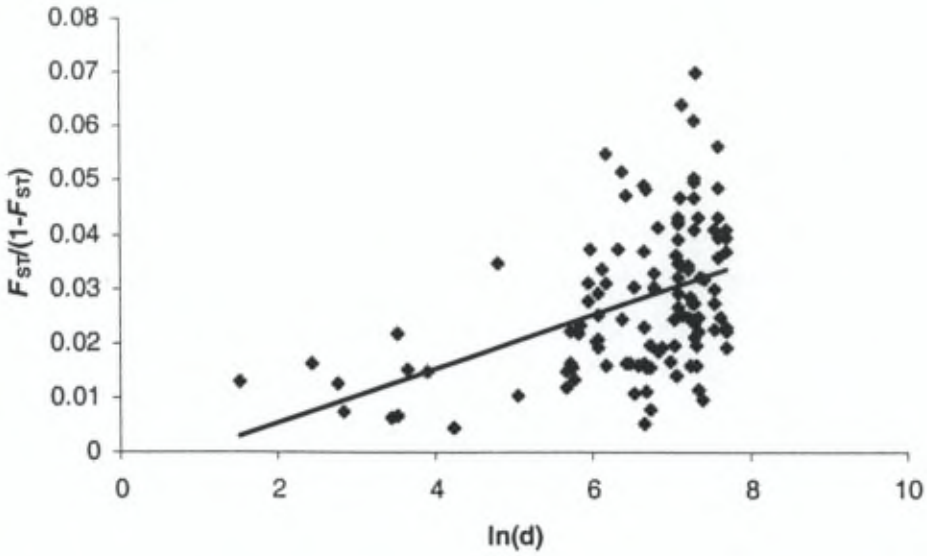


Fig 4 – Plot of pairwise $F_{ST}/(1- F_{ST})$ ratios against the logarithm of distances (in km) between populations. Under isolation by distance, values of the ratios are expected to increase linearly with the logarithm of distance (ROUSSET 1997).

southeastern Europe: P_{best} = 0.89 in eastern Romania; P_{best} = 0.74 in western and central Bulgaria; P_{best} = 0.70 in the Republic of Moldova. Four different gene pools (gene pools 3 to 6) appeared to be the most common gene pool in distinct populations from those regions. The highest population admixture was found in populations from Lithuania, Slovakia and Hungary (P_{best} = 0.42) with on average 19%, 39% and 21% of genes assigned to the first, second and third gene pools, respectively. High population admixture was also found in all populations from the western and central zones (P_{best} = 0.62), which showed a strikingly homogeneous pattern with the first gene pool predominating, and, on average 17% of genes assigned to the second gene pool and 11% of genes assigned to the third gene pool. The situation in Sweden was singular as two populations show little evidence for population admixture, but with distinct predominant gene pools that were found in admixture in the third Swedish population.

Isolation by distance tested at the European level by a permutation procedure was rejected both for F_{ST} and R_{ST} computed as multilocus statistics. However, a significant pattern of isolation by distance (P<0.001) was detected with the $F_{ST}/(1-F_{ST})$ statistic but not with $R_{ST}/(1-R_{ST})$ in a subset of the data, comprising populations from Scotland, Ireland, France, Germany, Slovakia, Hungary and Lithuania (i.e. from the regions in which admixture of gene pools 1 to 3 was found from the model-based cluster analysis). In this subset, the $F_{ST}/(1-F_{ST})$ ratio increased with the natural logarithm of geographical distance (*Blog* = 0.00491, R^2 = 0.190; Figure 4).

Discussion

The overall pattern of genetic variation at microsatellite loci that we observed for the common ash is typical for a long-lived outcrossing species with high genic diversity and a low level of inbreeding within populations (H_E = 0.77 and F = 0.03), and low differentiation among populations (F_{ST} = 0.08). Nevertheless, over the distribution range, we observed two contrasting patterns in terms of level of within and among-population diversity: (1) regions with high allelic and genic diversity within populations, but very low differentiation among populations; and (2) regions with relatively low allelic and genic diversity within, but relatively strong genetic divergence among populations. The former pattern occurs mainly in the western and central European zones, although it also occurs in Lithuania, whereas the latter is found only in Sweden and in the southeastern zone. These patterns were also associated to striking differences in levels of the inbreeding coefficient, that was positively correlated to the within population genetic variation, and the T_2 statistics of Cornuet and Luikart (1996), which indicated a tendency to excess of heterozygosity with respect to that expected from the observed number of alleles in regions with high within-population genetic variation, and the opposite pattern in regions of high genetic differentiation among populations. Finally, the southeastern region, which has high genetic differentiation in terms

of F_{ST}, also shows a large and significant excess of R_{ST} over the F_{ST}. Although our sampling scheme is too limited to make full inference of historical processes, we will discuss below possible causes of such clear contrasting patterns.

Patterns of genetic diversity and postglacial recolonisation processes

Palynological data since the last ice age suggest the occurrence of glacial refuges for ash in the eastern Alps and in the southern Balkans, as well as, but with lower support, northeast of the Black Sea and in the Pyrenees (Huntley & Birks 1983). From interpretation of Brewer's palynological maps (Brewer 2001), additional refuges were located within the Italian peninsula and at the Bulgarian Black Sea coast; the western refuge is weakly supported and may rather be situated within the Iberian peninsula, and the refuge north of the Black sea is not confirmed. The analysis of pollen diagrams by Gliemeroth (1997) supported the occurrence of refuges in the Iberian peninsula and at the northern Black Sea coast, whereas refuges in the three peninsulas and in the Alps are supported by chloroplast haplotype data (G. G. Vendramin, in preparation). Recolonisation from the refuges started northward about 12000 years BP and was followed by a westward migration, which began roughly 9000 years BP (from Brewer 2001). In several tree species from North America (*Tsuga mertensiana*, Ally *et al.* 2000; *Pinus coulteri*, Ledig 2000) and Japan (*Fagus crenata*, Tomaru *et al.* 1997) a decrease of both allelic richness and gene diversity within populations has been reported from southern to northern areas of the distribution, and interpreted as a consequence of founder effects associated to postglacial recolonisation. In the common ash, allelic richness and gene diversity also varied concomitantly with geography but genetic variation was highest in regions that have been colonized more recently, i.e. France and the British Isles. Thus in the common ash, we found no evidence for genetic impoverishment due to founder effects during the recolonisation process. The relationship between allelic richness and gene diversity is condensed in Cornuet and Luikart's (1996) bottleneck statistic T_2 for which we recorded low values close to refuges, i.e. a relative excess of allelic richness, but high values in recently recolonised areas, i.e. a relative excess of gene diversity. Similar variation of the bottleneck statistic with geography was observed by Comps *et al.* (2001) in European *Fagus sylvatica*. However, these authors, as well as Zanetto and Kremer (1995) in European *Quercus petraea*, found a negative correlation between regional averages of allelic richness and gene diversity, with higher allelic richness in regions corresponding to the glacial refuges but higher gene diversity in recently colonized regions. Comps *et al.* (2001) interpreted these observations as a consequence of post-glacial recolonisation from multiple refuges simultaneously, in analogy with the "migrant pool" model of metapopulation dynamics (Wade & McCauley 1988). Hence, the drift effect associated to founder events could have been counteracted by admixture of previously differentiated gene pools and they presented simple calculations illustrating that this process may act differentially on allelic richness and gene diversity

(Comps *et al.* 2001). We propose that admixture of differentiated gene pools originating from distinct refuges may also explain the diversity patterns found in the common ash. This hypothesis is reinforced by the results from the model-based cluster analysis, which identified higher levels of population genetic admixture in recently colonized areas (i.e., western and central zones) than in regions closest to glacial refuges (southeastern zone). The topography of Europe allowing recolonising lineages to merge, and the evidence of a westward colonization of ash after migration from the refuges (from S. Brewer, 2001) also support this proposition. The lack of a decrease in allelic richness from refuges to recolonised areas in the common ash may also be attributable to the genetic markers employed in this study: microsatellites have indeed a higher mutation rate than allozymes (Jarne & Lagoda, 1996) and may thus have recovered alleles more quickly after founder events.

Geographic variation in within-population level of inbreeding

We observed a low positive but significant average inbreeding coefficient within European populations of the common ash ($F = 0.033$) but among-population variability for F was high (CV = 158%). Moreover a longitudinal gradient of decreasing F from west to east was observed. This could be explained by variation across Europe either in the mating system, in the level of biparental inbreeding (mating among relatives), or in the frequency of null alleles. The latter explanation is a general feature of microsatellite loci, the higher the frequency of null alleles, the higher the value of the observed inbreeding coefficient (Bruford 1998). Although we cannot rule out this hypothesis, we did not observe unreliable amplification reactions at any analysed locus. In a previous paper on populations from Bulgaria, we studied the association between the spatial position of individual trees within stands and their genetic relatedness (Heuertz *et al.* 2001). Our results indicated the presence of spatially-determined biparental inbreeding in several populations. A comparison of the average values of the inbreeding coefficient ($F = 0.014$) with that of the kinship coefficient between neighbor plants ($F_{\text{short distance class}} = 0.021$) indicated that most of the observed inbreeding seemed to be caused by biparental inbreeding. In contrast, within a large French population, M.-E. Morand (in preparation) observed a much higher value of the inbreeding coefficient ($F = 0.122$) than kinship coefficients between neighbour plants ($F_{\text{short distance class}} = 0.026$). Their results suggest that in this French population most inbreeding is due either to self-fertilization or to non-spatially determined biparental inbreeding, as for instance assortative mating. We thus suggest that in the common ash a gradient of increasing selfing may occur from eastern to western Europe, with an inferred average selfing rate in the western zone as high as 17% [computed as $s=2F/(1+F)$]. Alternatively, following the hypothesis of admixture of differentiated gene pools at the time of recolonisation in western Europe, biparental inbreeding could have been enforced in the descendant populations through the occurrence of some reproductive barriers between individuals from the co-occurring gene pools. Data on

variation in morphology, phenology and sexual phenotype within populations, as well as direct assessments of outcrossing rates would be necessary to distinguish between these two hypotheses. For instance in *Fraxinus lanuginosa*, an androdioecious ash species from Japan, increasing levels of inbreeding attributed to self-fertilization were observed in seeds as the frequency of male individuals decreased within populations (Ishida & Hiura 2002).

Patterns of divergence among populations

Results from the distance-based and the model-based cluster analyses clearly showed that populations from the western and central zones, as well as populations from Slovakia, Hungary and Lithuania but to a lower extent, have a very similar genetic composition ($F_{ST} < 0.02$, $R_{ST} < 0.03$). This suggests either a common history of recolonisation, probably occurring mainly westwards in agreement with fossil pollen maps, or several events of recolonisation followed by homogenisation through extensive gene flow. A higher than average level of population admixture in these populations suggests that intermixing of genotypes derived from several refuges did occur, and the observed pattern of isolation by distance supports some common migration history. Chloroplast haplotype data are necessary to clarify the proposed hypotheses. In contrast, in the southeastern zone, we found evidence for the co-occurrence of several well differentiated gene pools ($F_{ST} = 0.09$, $R_{ST} = 0.22$) at relatively short geographic distance. This observation would agree with restricted inter-regional gene flow, and accordingly, in a previous investigation on spatial patterns of genetic differentiation within and among populations from Bulgaria, we found indirect evidence for relatively limited levels of pollen and seed effective dispersal in the common ash (Heuertz *et al.* 2001). Hence, the observed strong genetic differentiation in the southeastern zone could be explained by long-term coexistence of several gene pools under a stable demographic history and restricted gene flow. This pattern is not at odds with the observed low levels of differentiation among populations in western and central Europe: it has been shown that during recolonisation similar levels of dispersal of seed or pollen into low-density stands could result in higher effective gene flow than under stable demographic conditions (Austerlitz et al. 2000). Finally, another striking pattern was the very large divergence between populations 24 to 26 from Romania and all other samples. This pattern was even more pronounced when the variances in microsatellite allelic sizes were taken into account ($R_{ST} >> F_{ST}$; see also Figure 2). Although this pattern could be partly explained by evolution from separate glacial refuges, we suggest that additional processes such as interspecific hybridisation could be responsible for it, as other species of Fraxinus, i.e. *F. angustifolia*, *F. pallisiae* and *F. ornus* are common in this region (Huntley & Birks 1983; M. Heuertz, personal observation).

Population admixture

Our results suggested that areas most recently colonized showed evidence for higher population admixture than those from areas closer to the glacial refuges. It is striking that most genes from the populations with higher than average admixture were assigned with different proportions to a combination of mainly three gene pools using the model-based clustering approach: gene pools 1 to 3, which were predominant in a very large geographic area ranging from the British Isles to eastern Romania and eastern Bulgaria, including northern populations from Sweden and Lithuania. Hence these gene pools could originate from several identified potential refuges, the Iberian Peninsula, Italy and/or the Alps, or the Balkans. These results should be taken with caution, however, because the model-based clustering method has not yet been applied to large datasets on plant species, and evolutionary processes such as pollen gene flow, isolation by distance, founder effects may have confounding effects on the observed patterns. When the method was applied to populations from the western and central zones alone, the signal of population admixture disappeared. This suggests either that the detection of admixture in the overall analysis is an artifact, or that the events leading to population admixture occurred a long time ago and most linkage disequilibrium within populations has already decayed. The geographic coherence of the distribution of the putative gene pools suggests however that the signal of admixture has a historical foundation, with for instance, the occurrence of a suture zone between the western group and the southeastern group in Slovakia and Hungary where the highest levels of admixture were recorded.

A putative northern cryptic refuge

In Sweden, the patterns of genetic variation were very similar to those observed in the southeastern zone, although Swedish populations are thought to have a history of postglacial recolonisation: low allelic richness and gene diversity; low inbreeding coefficient; low values of the T_2 statistics; lower than average population admixture. A careful investigation of fossil pollen maps (from Brewer *et al.* 2001) confirms that pollen records are known from southern Sweden as early as 12500 years BP. We thus suggest that our data might confirm a putative cryptic northern glacial refuge for common ash in southern Sweden. Such cryptic northern refuges have recently been identified in other plant and animal species (Stewart & Lister 2001). Moreover, they could have played a role in the postglacial recolonisation process, because one of the six identified gene pools (number 2) is predominant in one Swedish population but is found in admixture across the western and central zones as well as in Lithuania.

Microsatellite versus allozyme patterns of variation

Up to now, most large-scale studies of patterns of genetic variation in tree species were performed with allozyme markers. Although no data on allozymes are available in *Fraxinus excelsior*, our results on relative statistics such as F and F_{ST} are comparable to those obtained on species with similar life histories but with allozymes (Hamrick & Godt 1989, Hamrick *et al.* 1992). Like these other studies, we also observed geographic patterns of variation in statistics describing within- and among-population genetic variation. However, data from five highly variable microsatellite loci allowed us to identify several additional patterns that may have been missed with allozyme markers. First we observed a subtle pattern of geographic variation in F and in the statistic T_2, that was detectable thanks to the comparatively lower variance associated to microsatellite markers (Goudet *et al.* 1996). Second, a comparison between F_{ST} and R_{ST} values allowed us to identify a region with potential long-term genetic differentiation and/or interspecific hybridisation. Third, the application of the model-based cluster analysis of Pritchard *et al.* (2000) on multilocus genotypes allowed to suggest the occurrence of population admixture associated to recolonisation of northwestern Europe.

Concluding remarks

We suggest that the refuge areas analysed, i.e. southeastern Europe, contain differentiated genetic resources of the common ash, which may have been coexisting for a long time without substantial genetic exchanges. These areas thus constitute important source populations for the conservation of genetic resources in the common ash. In contrast, central and western European populations harbor much more homogeneous resources which would result from the processes of recolonisation and admixture of previously differentiated gene pools.

References

Aldrich PR, Hamrick JL, Chavarriaga P and Kochert G (1998) Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Molecular Ecology*, **7**, 933-944.

Ally D, El-Kassaby YA and Ritland K (2000) Genetic diversity, differentiation and mating system in mountain hemlock (*Tsuga mertensiana*) across British Columbia. *Forest Genetics*, **7**, 97-108.

Austerlitz F, Mariette S, Machon N, Gouyon PH and Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics*, **154**, 1309-1321.

Binggeli P, Power JJ (1999) *Gender variation in ash.* Available from http://members.tripod.co.uk/WoodyPlantEcology/species/ash.htm

Brachet S, Jubier MF, Richard M, Jung-Muller B, Frascaria-Lacoste N (1999) Rapid identification of microsatellite loci using 5' anchored PCR in the common ash *Fraxinus excelsior*. *Molecular Ecology*, **8**, 160-163.

Bruford MW, Ciofi C. and Funk SM (1998) Characteristics of microsatellites. In: *Molecular Tools for Screening Biodiversity* (eds Karp A, Isaac PG, Ingram DS), pp. 202-205. Chapman & Hall, London.

Brewer S (2001) *Fossil pollen maps*. Available form http://www.pierroton.inra.fr/cytofor

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics*, **19**, 233-257.

Charlesworth B, Morgan MT and Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289-1303.

Comes HP, and Kadereit JW (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432-438.

Comps B, Gömöry D, Letouzey J, Thiébaut B, Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics*, **157**, 389-397.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001-2014.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* L. Skeels) endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832-839.

Felsenstein J (1995) PHYLIP (Phylogeny Inference Package) version 3.57, available from http://evolution.genetics.washington.edu/phylip.html.

Fineschi S, Taurchini D, Villani F, Vendramin GG (2000) Chloroplast DNA reveals little geographical structure in *Castanea sativa* Mill. Fagaceae throughout southern European countries. *Molecular Ecology*, **9**, 1495-1503.

Gliemeroth AK (1997) Paläoökologische Aspekte der Einwanderungsgeschichte einiger Baumgattungen während des Holozäns nach Europa. *Angewandte Botanik*, **71**, 54-61.

Goldstein DB, Linares AR, Feldman MW, Cavalli-Sforza LL (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463-471.

Goudet J, Raymond M, de Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933-1940.

Goudet J (1995) FSTAT (vers.1.2), a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In *Plant population genetics, breeding, and genetic resources* (eds. Brown ADH, Clegg MT, Kahler AL, Weir BS), pp. 43-63. Sinauer Associates, Inc., Sunderland.

Hamrick JL, Holden LR (1979) Influence of microhabitat heterogeneity on gene frequency distribution and gametic phase disequilibrium in *Avena barbata*. *Evolution*, **33**, 521-533.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95-124.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile compute program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618.

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution*, **16**, 700-706.

Heuertz M, Hausman JF, Tsvetkov I, Frascaria-Lacoste N, Vekemans X (2001) Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.). *Molecular Ecology*, **10**, 615-623.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Huntley B, Birks HJB (1983) *An atlas of past and present pollen maps for Europe: 0-13000 Years Ago*, pp. 209-223. Cambridge University Press, Cambridge, UK.

Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, **52**, 577-586.

Ibrahim KL, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, **77**, 282-291.

Ishida K, Hiura T (2002) Mating system and population genetic structure of an androdioecious tree, *Fraxinus lanuginosa* Koidz. (Oleaceae) in northern Japan. *Heredity*, **88**, 296-301.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424-429.

Kirby GC (1975) Heterozygote frequencies in small subpopulations. *Theoretical Population Biology*, **8**, 31-48.

Konnert M, Bergmann F (1995) The geographical distribution of genetic variation in silver fir (*Abies alba*, Pinaceae) in relation to its migration history. *Plant Systematics and Evolution*, **196**, 19-30.

Lagercrantz U, Ryman N (1990) Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution*, **44**, 38-53.

Le Corre V, Kremer A (1998) Cumulative effects of founding events during colonization on genetic diversity and differentiation in an island and stepping-stone model. *Journal of Evolutionary Biology*, **11**, 495-512.

Ledig FT (2000) Founder effects and the genetic structure of Coulter pine. *Journal of Heredity*, **91**, 307-315.

Lefort F, Brachet S, Frascaria-Lacoste N, Edwards KJ and Douglas GC (1999) Identification and characterisation of microsatellite loci in ash (*Fraxinus excelsior* L.) and their conservation in the olive family (*Oleaceae*). *Molecular Ecology*, **8**, 1088-1091.

Levin DA (1981) Dispersal versus gene flow in plants. *Annals of the Missouri Botanical Garden*, **68**, 233-253.

Loveless MD, Hamrick JL (1984) Ecological determinants of the genetic structure in plant populations. *Annual Review of Ecology and Systematics*, **15**, 65-95.

Michalakis Y, Excoffier L (1996a) Genetic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, **142**, 1061-1064.

Morand ME, Brachet S, Rossignol P, Dufour J and Frascaria-Lacoste N (2002) A generalized heterozygote deficiency assessed with microsatellites in French common ash populations. *Molecular Ecology*, **11**, 377-385.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583-590.

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, **47**, 253-259.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1-10.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology and Evolution*, **14**, 140-145.

Picard JF (1982) Contribution à l'étude de la biologie florale et de la fructification du frêne commun (*Fraxinus excelsior* L). *Revue Forestière Française*, **XXXIV**, 97-107.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **86**, 502-503.

Pollak E (1987) On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics*, **117**, 353-360.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Rajora OP, Rahman MH, Buchert GP and Dancik BP (2000) Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. *Molecular Ecology*, **9**, 339-348.

Randi E, Pierpoli M, Beaumont M, Ragni B, Sforzi A (2001) Genetic identification of wild and domestic cats (*Felis sylvestris*) and their hybrids using Bayesian clustering methods. *Molecular Biology and Evolution*, **18**, 1679-1693.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2) Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223-225.

Rosenberg AN, Woolf E, Pritchard JK, Schoof T, Gefel D, Shpirer I, Lavi U, Bonné-Tamir B, Lillel J, Feldman MW (2001) Distinctive genetic signatures in the Libyan Jews. *Proceedings of the National Academy of Sciences USA*, **98**, 858-863.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219-1228.

Schoen DJ, Brown ADH (1991) Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences of the USA*, **88**, 4494-4497.

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution*, **16**, 608-613.

Tomaru N, Mitsutsuji T, Takahashi M, Tsumura Y, Uchida K, Ohba K (1997) Genetic diversity in *Fagus crenata* (Japanese beech): influence of the distributional shift during the late-Quaternary. *Heredity*, **78**, 241-251.

Vekemans X, Lefèbvre C (1997) On the evolution of heavy-metal tolerant populations in *Armeria maritima*: evidence from allozyme variation and reproductive barriers. *Journal of Evolutionary Biology*, **10**, 175-191.

Wade MJ, McCauley DE (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution*, **42**, 995-1005.

Wardle P (1961) Biological Flora of the British Isles. *Fraxinus excelsior* L. *Journal of Ecology*, **49**, 739-751.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB (2001) Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proceedings of the National Academy of Sciences of the USA*, **98**, 5078-5083.

Zanetto A, Kremer A (1995) Geographical structure of gene diversity in *Quercus petraea* Matt. Liebl. I. Monolocus patternsof variation. *Heredity*, **75**, 506-517.

Chapter IV: Different evolutionary history and geographical distribution of even- and odd-sized alleles at a microsatellite locus in *Fraxinus excelsior* (L.)

Abstract

The evolutionary history of the molecular determinism of even- and odd-sized alleles in the compound dinucleotide microsatellite locus FEMSATL4 in *Fraxinus excelsior* was investigated by sequencing a total of 38 alleles in five species of the olive family (Oleaceae). In *F. excelsior*, the even-sized allele type was mainly observed in southeastern Europe and differed from the more common odd-sized allele type by the absence of the stretch $AAA(CA)_n$. The odd-sized allele type was also found in the closely related *F. ornus*, suggesting that it corresponds to the ancestral allele type in *F.* excelsior, but its origin could not be traced further back into the past. Two evolutionary scenarios are proposed to account for the current phylogeographic pattern in *F. excelsior*, involving (i) a deletion of the motif within *F. excelsior* during the last glacial period, or (ii) hybridisation with a closely related ash species occurring in southeastern Europe, probably *F. angustifolia*, which would carry the odd-sized allele type at high frequency. Further, our results suggest that two hexanucleotide repeats have evolved independently in *F. ornus* and *Olea europaea,* probably trough a substitution within a perfect $(GA)_n$ repeat, followed by slippage affecting six nucleotides. This finding challenges the view that base substitutions occurring within microsatellites usually stabilise the locus by preventing further slippage events.

Introduction

Microsatellites or simple sequence repeats (SSRs) are tandemly repeated DNA motifs, one to six nucleotides long (Queller *et al.* 1993). They are highly polymorphic mendelian markers, extensively used for mapping and for population genetic and demographic issues (e.g. Jarne & Lagoda 1996, Estoup & Angers 1998).

Mutation rates at microsatellite loci are generally high; on the order of 10^{-5} to 10^{-2} per locus per generation for animals (Jarne & Lagoda 1996, Estoup & Angers 1998). In plants, estimates ranged from 5.1×10^{-5} to 1.1×10^{-3} in maize (Vigouroux *et al.* 2002) and from 3.9×10^{-3} to 10^{-2} in chickpea (Udupa & Baum 2001). The favoured mutation mechanism for microsatellites is slip-strand mispairing, i.e. relative slippage of the nascent and the template strands during DNA replication (Levinson & Gutman 1997, Primmer *et al.* 1996a), increasing or decreasing the allele size by one repeat motif. More complicated mutation processes may occur, involving multi-repeat mutations (e.g. Karhu *et al.* 2000, Vigouroux *et al.* 2002), single nucleotide substitutions or indels (e.g. Primmer & Ellegren 1998, van Oppen *et al.* 2000, Barrier *et al.* 2000) or large indels (van Oppen *et al.* 2000, Colson & Goldstein 1999). Mutations in flanking regions are also common (Brohede & Ellegren 1999, Colson & Goldstein 1999), and sometimes linked to certain allele types (Blankenship *et al.* 2002). Frequent and variable mutations lead to size homoplasy, i.e. alleles have the same electrophoretic size but they are not identical by descent (reviewed in Estoup *et al.* 2002). Homoplasious alleles can have the same sequence or not. In the latter case, size homoplasy is referred to as "molecularly accessible" (Estoup *et al.* 2002), and it was observed among alleles from the same species (e.g. Culver *et al.* 2001), or from different species (Estoup *et al.* 1995, Primmer & Ellegren 1998, Markova *et al.* 2000, van Oppen *et al.* 2000).

Although there is substantial interlocus variation of mutation rates, empirical studies have revealed associations between mutations and features of the microsatellite (revieved in Estoup & Angers 1998, and Eisen 1999): (i) mutation rates are on average inversely proportional to the size of the repeat unit; (ii) the likelihood for slip-strand mispairing, and *a fortiori* the mutation rate, increase with repeat count; (iii) tracts with multiple repeats are more subjected to large, multi-repeat deletions than short tracts; and (iv) the presence of interruptions or variant repeats stabilises the locus by reducing the opportunity for slip-strand mispairing.

For adequate inference of population genetics parameters from microsatellite data, statistics have been developed based on models which take into account the peculiarities of microsatellite mutations (Slatkin 1995, Goldstein *et al.* 1995). The simplest model of microsatellite evolution, the stepwise mutation model (SMM, Ohta & Kimura 1973) assumes evolution by addition or deletion of one repeat motif. Recent models additionally allow for more complex mutation events, involving several repeat motifs (the two-phase model, TPM, Di Rienzo *et al.* 1994) or point mutations (e.g. Kruglyak *et al.* 1998). However, statistics

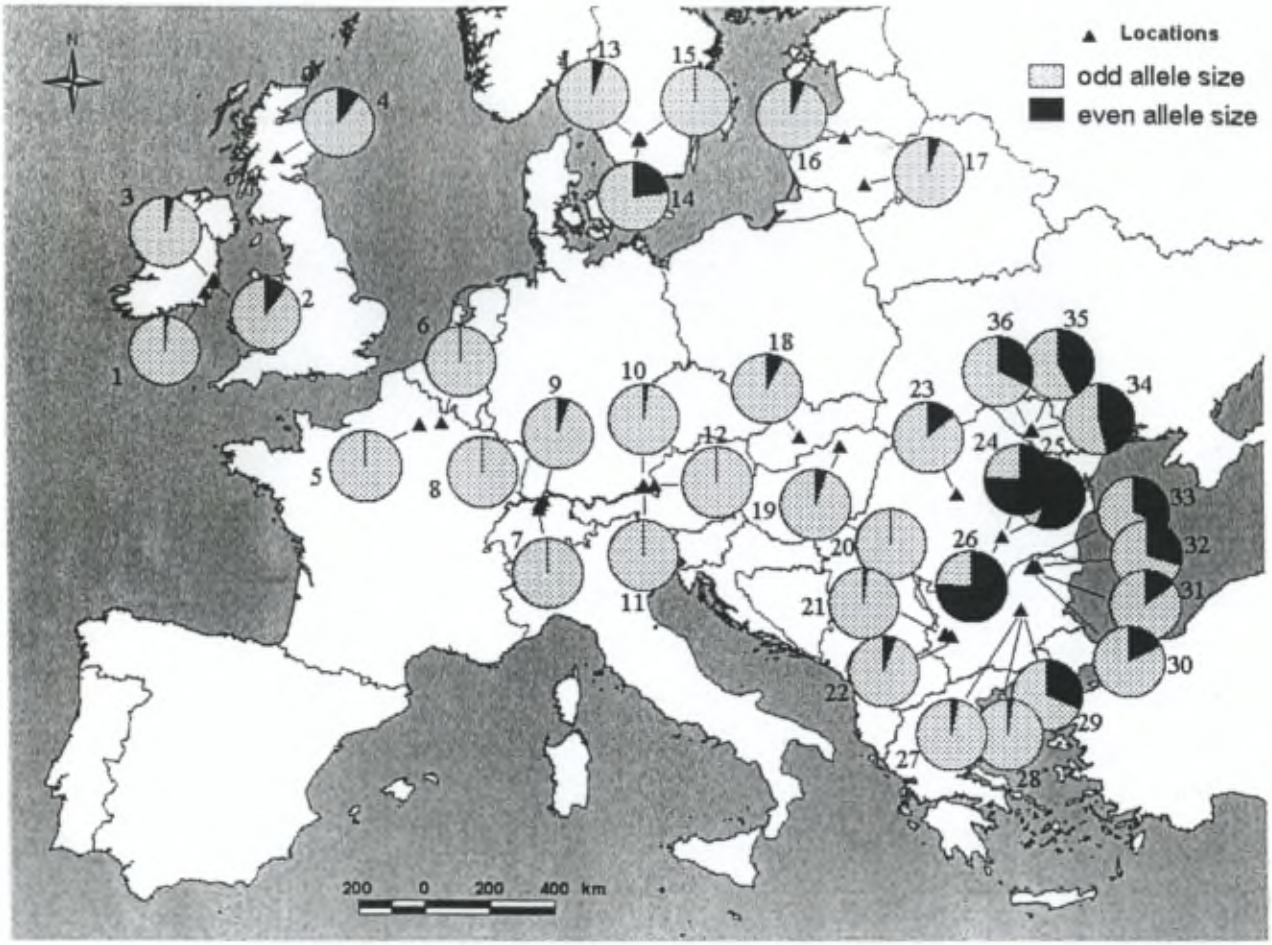


Fig. 1 Geographical distribution of odd- and even-sized alleles at microsatellite locus FEMSATL4 in 36 European *Fraxinus excelsior* populations.

based on allele size variation typically suffer higher sampling variances than their counterparts based on allele identity information, as demonstrated for instance for the estimators of among-population differentiation R_{ST} and F_{ST} (Slatkin 1995, Balloux & Goudet 2002). In this context, a useful test was recently developed which assesses whether step-wise like mutations significantly contributed to population differentiation (Hardy *et al.* 2002).

The mutation process at microsatellite loci can be investigated directly from observation of mutations in the germline, for example in humans (Amos *et al.* 1996) or barn swallows (Primmer *et al.* 1996a). Alternatively, sequence variation accumulated over evolutionary time can be studied by using the same primers in related taxa (Primmer & Ellegren 1998, van Oppen *et al.* 2000, Karhu *et al.* 2000). This is the approach we used for the compound dinucleotide repeat locus FEMSATL 4, originally isolated as $(CA)_2(AG)_{24}$ in common ash, *Fraxinus excelsior*, a European forest tree species (Lefort *et al.* 1999). Priming sites of this locus have been conserved in many ash species and in three additional genera (*Ligustrum*, *Olea* and *Phillyrea*, Lefort *et al.* 1999) from the tribe Oleeae, in the relatively young family Oleaceae (most recent common ancestor ~ 37 MYBP, E. Wallander personal communication). Our study was motivated by the finding of alleles of even and odd electrophoretic sizes with different geographic distribution in a population genetic study of common ash at the European scale (Heuertz *et al.* unpublished). Our objectives were (i) to investigate the molecular diversity underlying the electrophoretic variation in *Fraxinus excelsior* and (ii) to regard it in the context of the evolutionary history of the locus by analysing amplification products from four species with increasing phylogenetic distance to *F. excelsior*: *F. ornus*; (*Olea europaea* and *Phillyrea angustifolia*), and *Ligustrum vulgare*.

Materials and methods

Materials

The studied species include common ash (*Fraxinus excelsior*), a large European forest tree; flowering ash (*F. ornus*), a small south European tree; *Ligustrum vulgare*, a European bush, sometimes planted in hedges; olive tree (*Olea europaea*), a largely cultivated tree species with a Mediterranean distribution; and *Phillyrea angustifolia*, a Mediterranean bush. Samples were collected in order to represent a large geographical range: leaf or bud samples from a total of 1050 *F. excelsior* trees from 36 European populations (Heuertz *et al.* unpublished, Fig. 1); leaf or DNA samples from a total of 21 *F. ornus* trees from two Bulgarian and one Italian population; leaf samples from a total of 36 *Ligustrum vulgare* bushes from one Dutch, one Belgian, two French and two Moldavian populations; DNA samples from a total of 69 olive trees belonging to 26 cultivars from different south-European, north African and near Eastern countries or to five natural populations from south Italy, leaf samples from three ornamental

Table 1 Statistics of genetic diversity of the microsatellite locus FEMSATL4 in five Oleaceae species.

Species	n	*K*	*A*	*H*	F_I[a]		allele range	mean allele size
Fraxinus excelsior	1050	71	19.2	0.897	0.123	**	155-298	175.7 (17.4)[c]
Fraxinus ornus	21	11	11.0	0.885	0.085	ns	179-213	189.8 (10.7)[b]
Ligustrum vulgare	37	13	10.2	0.786	0.002	ns	152-210	169.8 (9.3)[d]
Olea europaea	72	10	7.5	0.764	0.091	***	159-201	170.7 (15.9)[d]
Phillyrea angustifolia[e]	4	8					140-198	170.0 (20.5)[cd]

n, sample size; *K*, total number of alleles; *A*, allelic richness standardised to a sample size of 21 diploid individuals; *H*, total gene diversity; F_I, Wright's inbreeding coefficient; allele range, range of allele sizes in nucleotides; mean allele size, mean size of alleles in nucleotides and standard deviation in parentheses.

[a] Exact test of departure from Hardy-Weinberg genotypic proportions: ns, non significant; **, P<0.01; ***, P<0.001.

[bcd] Values followed by the same letter are not significantly different at the 5% probability level (T-tests).

[e] Diversity statistics were not computed because the sample size was too small.

olive trees from Portugal; and DNA samples from four *Phillyrea* bushes from southern France, the Balearic Islands and Sicily.

Screening for allelic diversity

Total DNA was extracted from 60 to 90 mg of dry leaves or from 50 to 70 mg fresh weight of buds. The plant material was ground by hand or in the automatic grinding mill MM200 (Retsch) and extracted with the DNeasy Plant mini kit (Qiagen) or the CTAB procedure of the NucleoSpin Plant kit (Macherey Nagel). Polymerase chain reaction (PCR) was performed with FEMSATL4 primers (Lefort et al 1999) as described in Heuertz *et al.* (2001) and Heuertz *et al.* (unpublished). Difficult templates were PCR amplified with LATaq polymerase (TaKaRa). Fluorescent labelling of the forward primer allowed detection of amplification products on an ABI PRISM® 377 DNA sequencer. Sizing of fragments was performed with the software programs Genescan® 3.1 and Genotyper® 2.5 from Applied Biosystems by comparison with an internal sizing standard (Genescan-350 Rox), and alleles were called by eye. Differences in mean allele sizes among pairs of species were tested with T-tests. For the two *Faxinus* species, for *Olea* and for *Ligustrum*, the following genetic diversity statistics were computed for the overall species samples with the program FSTAT v. 2.9.3.2 (Goudet 2001): the allelic richness *A*, which quantifies allelic diversity with the property of being independent of sample size variation, for a standardised sample size of n=21 diploid individuals (corresponding to the smallest sample size, observed in *F. ornus*) according to El Mousadik and Petit (1996); the total gene diversity *H*; and the overall inbreeding coefficient F_I. Deviation from Hardy-Weinberg genotypic proportions was tested with exact tests with the program GENEPOP v. 3.3 (Raymond & Rousset 1995).

Cloning and sequencing of PCR products

For cloning, we selected individuals carrying alleles from throughout the observed size range (Table 1). In order to minimize the number of cloning reactions, heterozygous individuals were preferred. FEMSATL4 was amplified with non-fluorescent primers in a total volume of 25 µl; 4 µl were run on 1% agarose gels to check the quality of PCR products, the rest was purified with the QIAQuick PCR purification kit (QIAGEN). Purified DNA was cloned using the TOPO TA® cloning kit (Invitrogen). For each cloning reaction, ten bacterial colonies were screened for the presence of an insert of suitable size as follows. The colonies were suspended in 25 µl sterile water, boiled for 5 min and spun for 1 min at 12000 G. Three µl of the supernatant were used as a PCR template for M13 forward and reverse primers annealing to the cloning vector. The PCR reactions were performed in a mix (25µl) containing 2.5 mm $MgCl_2$, 0.5 U per reaction of Taq polymerase (APBiotech) in APBiotech PCR buffer, 0.5 µM

of each primer and 0.1 μM of each dNTP. After an initial denaturation step of 5 min at 95°C, PCR consisted of 25 cycles of 1 min 94°C, 1 min at 55°C and 1 min at 72°C, followed by a final elongation step of 5 min at 72 °C. Size of products was verified running 4μl on 1% agarose gels. Per cloning reaction, a mean of eight of these PCR products were cleaned with the QIAQuick PCR purification kit (QIAGEN) and cycle sequenced with the ABI Prism BigDye™ v.2 (*F. excelsior* samples) or v.3 cycle sequencing kit according to manufacturer's protocol. Sequence products were run on an ABI PRISM® 377 DNA sequencer (gel electrophoresis, *F. excelsior* samples) or on an ABI PRISM® 310 DNA sequencer (capillary electrophoresis).

Data analysis

Sequences were checked and completed against their electopherograms. They were aligned by eye using the program Genedoc (Nichols & Nichols 1997). Flanking regions were screened for variation and repeat sequences were formalised. For each species, sequences were assigned to the following categories: (i) alleles; (ii) stutter bands shorter than the actual allele size, resulting from Taq polymerase slippage; (iii) stutter bands longer than the actual allele size; (iv) unexpected sequences containing a microsatellite; and (v) sequences without a microsatellite or unreadable sequences. Phylogenetic relationships among species were analysed based on sequence variation in the flanking regions. As only one nucleotide site was phylogenetically informative (see results), the direct sequence information could not be used. Instead, we used a distance matrix based on the number of nucleotide differences between all pairs of flanking region haplotypes to compute a neighbour joining cladogram with the software package PHYLIP v. 3.57c (Felsenstein 1993). In *F. excelsior*, alleles of even and odd size corresponded to two different types of microsatellite repeat sequences (see results). Their respective frequencies and average sizes were investigated based on allele sizing data (Heuertz *et al.* unpublished) from 36 European populations. The correlation of within-population allele type frequency with geographic coordinates was tested with the Pearson moment correlation coefficient. Geographical trends in the other loci of that study (Heuertz *et al.* unpublished) were investigated by testing the correlation of mean allele size with geographical position. Genetic differentiation was analysed at locus FEMSATL4 among the 36 populations and among 9 southeastern European populations containing more than 25% of even-sized alleles (i) based on all alleles, and (ii) according to allelic type. Estimators based on allele identity (F_{ST}) or allele size (R_{ST}) were computed and the contribution of step-wise mutations to population differentiation was tested (Hardy *et al.* 2002). Results were compared to single locus estimates for four other loci in the overall dataset, and to multilocus estimates for five microsatellite loci in both datasets (data from Heuertz *et al.* unpublished). All computations were performed with the program SPAGeDi (Hardy & Vekemans, 2002). When R_{ST} was significantly larger than F_{ST}, a neighbour joining tree based on Goldstein *et al.*'s

Table 2 Sample sizes and efficiency of cloning and sequencing reactions.

Sample sizes	*Fraxinus excelsior*	*Fraxinus ornus*	*Ligustrum vulgare*	*Olea europaea*	*Phillyrea angustifolia*
Cloning reactions	14	5	9	8	4
Alleles chosen for cloning	26	9	9	8	8
Alleles recovered from sequencing reactions	14	6	6	8	4
Sequencing products (number)	>100	40	84	58	27
of which alleles (%)	25	50	35	83	26
shorter stutter bands (%)	23	13	19	5	4
longer stutter bands (%)	6	0	9	0	0
unexpected sequences with microsatellite (%)	2	13	7	7	22
no microsatellites or bad sequences (%)	50	25	30	5	48

(1995) $(\delta\mu)^2$ distance was constructed with PHYLIP v. 3.57c (Felsenstein 1993) in order to clarify geographic patterns of variation.

Results

Screening for allelic diversity

Statistics of genetic diversity at the microsatellite locus FEMSATL4 in five Oleaceae species are reported in Table 1. After standardising for equal sample sizes, allelic richness was highest in *F. excelsior* ($A = 19.2$) and lowest in *O. europaea* ($A = 7.5$). In *P. angustifolia*, a standardised allelic richness could not be computed, because we analysed only four individuals, but all of them were heterozygous and all detected alleles were different ($K = 8$); for comparison, in *O. europaea*, 72 samples were analysed and only two more alleles were identified ($K = 10$). Size ranges of observed alleles were overlapping in all species, but comparatively large alleles were found in *F. excelsior*. Mean allele size was largest in *F. ornus* (189.8 ± 10.7 nucleotides), followed by *F. excelsior* (175.7 ± 17.4 nucleotides). Gene diversity was high in all species, ranging from $H = 0.764$ in *O. europaea* to $H = 0.897$ in *F. excelsior*. Significant inbreeding in the overall species samples was found in *F. excelsior* and *O. europaea*.

Efficiency of cloning and sequencing reactions

Table 2 summarises the number of cloning reactions performed in each species, the number of alleles involved, and the assignment of sequencing products to different categories. It appears that for *F. excelsior*, only 14 of the 26 alleles implicated in cloning were recovered in sequencing products. Similarly, only four out of eight alleles were recovered in *P. angustifolia*. The experiment performed slightly better in *F. ornus* and *L. vulgare*, and it was truly successful in *O. europaea*, where all eight cloned alleles were recovered in sequencing. More precisely, in *O. europaea*, as much as 83% of the recovered sequences corresponded to alleles; and the species also featured a very low percentage of stutter bands (5%). In all other species the percentage recovery of fragments of expected size was worse, notably in *P. angustifolia* and in *F. excelsior* where only a quarter of the sequences corresponded to alleles and half of them were unreadable or contained no microsatellite. The bad result for *F. excelsior* could however be linked to experimental conditions (see discussion).

Sequence variation in the microsatellite repeat region and in flanking regions

Microsatellite repeat region. Sequence variation at the microsatellite locus FEMSATL4 is summarized in Fig. 2. All species carried a $(GA)_n$ motif in their repeat region (box in Fig. 2),

```
                                  10        20
F. excelsior odd 1    : AGAACGTTCTCTGCTAGTG-TAGGACACG(CA)3    T GAAAA(CA)2-6(GA)9-40
F. excelsior odd 2    : ..........................(CA)3    T GAAAA(CA)2-3(GA)24-39
F. excelsior even 1   : ...................-.........(CA)3    T(GA)12-21
F. excelsior even 2   : ...................-.........(CA)3    T(GA)26
                      :
F. ornus 1            : ....T......A.......-.........(CA)3    T GAAAA(CA)3-8        (GAGACA)0-6(GA)4-19
F. ornus 2            : ....T......A...G...-.........  CACCCA T GAAAA(CA)6          (GAGACA)5   GA
F. ornus 3            : ....T......A.......-.........(CA)3    T GAAAA(CA)3  TA(CA)4(GAGACA)3   (GA)4
F. ornus 4            : ....T......A.......-.........(CA)2 CCT GAAAA(CA)17         (GAGACA)2   (GA)7
                      :
L. vulgare 1          : .T..T..............G......T..TGCAC    T(GA)7-38
L. vulgare 2          : .T..T..............G....G.T..TGCAC    T(GA)16-19
                      :
O. europaea 1         : ....T.......T......-C....(CA)8-9              (GAGTGA)2   (GA)3
O. europaea 2         : ....T.......T......-.....(CA)10-16    (GA)0-1(GAGTGA)2-8(GA)3
O. europaea 3         : ....T.......T......-.....(CA)2      AA(CA)8-9(GAGTGA)2   (GA)3
O. europaea 4         : ....T.......T......-.....(CA)16               (GAGTGA)2   (GA)3
                      :
P. angustifolia 1     : ....T.......T....A.-G....       (GA)14-17
P. angustifolia 2     : ....T.......T......-G....(CA)14(GA)15

                          30        40        50        60        70        80
F. excelsior odd 1    : CCTCATTAAATGAATGACACAAAGCCTCTTCACCCAGAGATGCATATATGGTCACAACAGT
F. excelsior odd 2    : ....T.........................................................
F. excelsior even 1   : ................................................................
F. excelsior even 2   : ....T...........................................................
                      :
F. ornus 1            : ................................................................
F. ornus 2            : ................................................................
F. ornus 3            : ................................................................
F. ornus 4            : ................................................................
                      :
L. vulgare 1          : ...................................T....................G.....
L. vulgare 2          : ...................................T....................G.....
                      :
O. europaea 1         : ...........................................................G..
O. europaea 2         : ...........................................................G..
O. europaea 3         : ................G...........................................G..
O. europaea 4         : ...........................................................G..
                      :
P. angustifolia 1     : ...................T...........................................
P. angustifolia 2     : ...................T...........................................
```

Fig. 2 Alignment of nucleotide sequences between five Oleaceae species for the microsatellite locus of the study. Primers are excluded from the sequences. Dots (.) show identical nucleotides. Gaps (-) have been placed to increase the similarity. The box contains what is defined as the microsatellite repeat region. Numbers next to aligned sequences indicate nucleotide positions when the repeat region is excluded. Species names in bold correspond to haplotypes defined by variation in the flanking regions, and the corresponding microsatellite repeat sequences. Plain sequence names indicate sequences with additional variation in flanking or repeat regions, which were occasionally observed (at least two observations for each sequence).

but in addition to that, there was substantial variation, especially among species. In *F. excelsior*, two different patterns of variation were identified in the repeat region and unambiguously assigned to odd and even-sized alleles respectively. In odd-sized alleles, the $(GA)_n$ repeat was preceded by the stretch $AAA(CA)_n$ (sequence types *F. excelsior* odd 1 and 2 in Fig. 2; detected in alleles of length 163, 165, 167, 169, 171, 175, 183, 187, 195, 219 and 227 nucleotides). This motif was missing in even-sized alleles (sequence types *F. excelsior* even 1 and 2 in Fig. 2; verified in alleles of length 160, 166 and 188, and in stutter bands of length 162, 170, 172 and 178 nucleotides). In *F. ornus*, the motif characteristic for odd allele size in *F. excelsior* appeared in all sequences and it was mostly followed by a hexanucleotide repeat $(GAGACA)_n$. Some more variation in the repeat region (sequence types *F. ornus* 2-4 in Fig. 2) was observed in few sequences (at least 2 sequences for each type), but not unambiguously confirmed (e.g. observed in sequences cloned from only one individual and therefore possibly due to a replication error of the Taq polymerase, and/or in sequences with inferior resolution). The repeat region in *L. vulgare* was a perfect dinucleotide repeat $(GA)_n$, but the first nucleotide positions immediately flanking the repeat showed little homology to the *Fraxinus* species. The variable sequence in *O. europaea* was a compound dinucleotide / hexanucleotide repeat $(CA)_n(GAGTGA)_n$, sometimes interrupted by a single GA motif; additional variation in the repeat region (sequence type *O. europaea* 3) was not unequivocally confirmed. Three of the four detected *P. angustifolia* alleles carried a perfect $(GA)_n$ repeat, in the fourth allele, it was preceded by a $(CA)_n$ motif.

Flanking regions. The overall beginning of the repeat region was defined at position 26 (box in Fig. 2), where it is located in *O. europaea* and *P. angustifolia*. According to this definition, eleven sites in the sequenced flanking regions were polymorphic among species, comprising one insertion/deletion and ten substitutions (compare sequence types in bold in Fig. 2). The variation at these sites defined six haplotypes (sequence types in bold, Fig. 2); one in each species except in *O. europaea*, where there were two due to a single polymorphic site (position 21). In the other species, no variation in flanking regions was confirmed, but there were five candidate substitutions (in non-bold sequence types in Fig. 2, e.g. *F. ornus* 2, position 16: G) and one deletion (in *O. europaea* 4). Cladistic methods could not be applied to sequence variation in flanking regions among haplotypes because there was only one phylogenetically informative site (position 13), which allowed to group *O. europaea* and *P. angustifolia* against the other species. The number of mutations between pairs of species (with the indel coded as a separate character) was comprised between two (*F. excelsior – F. ornus*) and eight (*L. vulgare – P. angustifolia*), and phylogenetic relationships between species based on the number of mutations in flanking regions of locus FEMSATL4 were in agreement with the phylogeny of the tribe Oleae established by Wallander and Albert (2000) from sequence data at two non-coding chloroplast loci (Fig. 3, compare a and b).

Fig. 3 Agreement of phylogenetic relationships among five species of the tribe Oleae, based on (a) number of mutations in flanking regions of microsatellite FEMSATL4 (rectangular cladogram of midpoint-rooted neighbour joining tree) and (b) on sequence variation at two non-coding chloroplast loci, the *rps16* intron and the *trnL-F* region (the species of interest are indicated with arrows; part of Fig.1 extracted from Wallander and Albert, 2000). The original figure in this reference shows a strict consensus tree of the most parsimonious trees from the combined data set at the two loci, with indels coded as separate characters, for 76 Oleaceae species. Jackknife support values over 50% are shown above the branches.

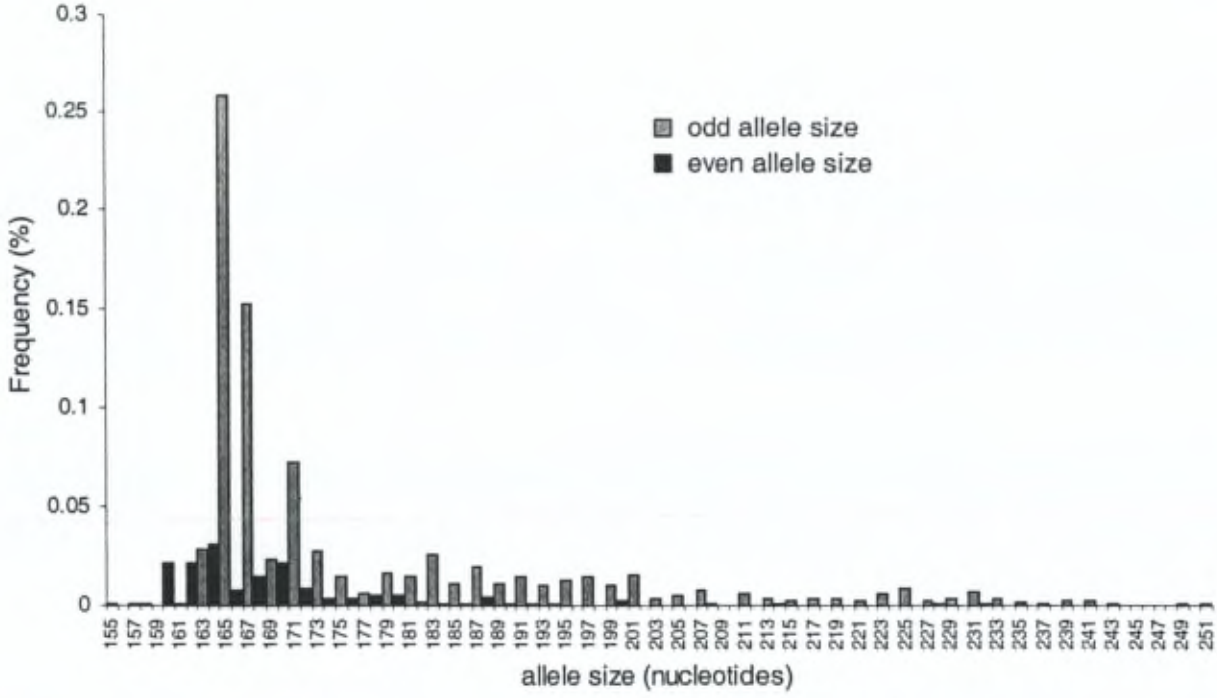


Fig. 4 Allele frequency distribution at locus FEMSATL4 in 1050 *Fraxinus excelsior* trees from 36 European populations (Fig. 1). Odd and even allele sizes are represented by white and grey bars, respectively. Two alleles larger than 251 nucleotides were found in low frequency and are not shown.

Table 3 Structure of genetic diversity among populations based on allele identity (F_{ST}) or allele sizes (R_{ST}).

Populations	Locus	F_{ST}[a]	R_{ST}[a]	P (1-sided)
36 European[b]	5 loci	0.076***	0.173***	***
36 European[b]	M2-30	0.059 ***	0.150***	**
36 European[b]	FEMSATL 11	0.067***	0.034***	ns
36 European[b]	FEMSATL 16	0.093***	0.073***	ns
36 European[b]	FEMSATL 19	0.106***	0.350***	***
36 European	FEMSATL 4 all alleles	0.066***	0.066***	ns
36 European	FEMSATL 4 odd alleles	0.071***	0.064***	ns
9 SE European[bc]	5 loci	0.066***	0.177***	**
9 SE European	FEMSATL 4 all alleles	0.042***	0.009 ns	ns
9 SE European	FEMSATL 4 odd alleles	0.053***	0.023 ns	ns
9 SE European	FEMSATL 4 even alleles	0.053***	-0.025 ns	ns

P (1-sided), test of information content of allele sizes (Hardy et al 2002): ns, non significant; **, P<0.01; ***, P<0.001.
[a] When five loci were analysed, multilocus statistics were computed. Values are followed by the significance level for the tests with H0: $F_{ST} = 0$, and H0: $R_{ST} = 0$, respectively, assessed with 1000 permutations: ns, non significant; ***, P<0.001.
[b] Data from Heuertz *et al*. (unpublished).
[c] Nine south-eastern European populations with more than 25 % even-sized alleles at locus FEMSATL4.

Discussion

Allelic diversity

Among the five Oleaceae species screened at microsatellite locus FEMSATL4, allelic diversity was highest in *Fraxinus excelsior*. This is in agreement with Primmer *et al.*'s (1996b) finding that microsatellite loci tend to be more polymorphic in the species in which they were initially isolated than in related species. Additionally, cross-species priming may produce a higher amount of null alleles in related species, because of mutations in the priming sites (e.g. Estoup & Angers 1998). Our data set bears no evidence for this, since inbreeding coefficients were not higher in related species than in *F. excelsior*. It has been argued that microsatellites were longer in the focal species, due to preferential selection of loci with a large number of repeats during the isolation process (ascertainment bias, e.g. Ellegren *et al.* 1997, but see Crawford *et al.* 1998). We found that *F. excelsior* alleles were indeed longer than those found in *L. vulgare* or in *O. europaea*, but shorter than those identified in *F. ornus*.

Efficiency of cloning and sequencing reactions

The number of cloned alleles recovered in sequencing reactions was generally low in our study; we re-identified on average only one allele from each cloning reaction, despite choosing as much as eight colonies for sequencing. In some species, we found many stutter bands, 2 to 4 nucleotides shorter, or sometimes longer, than the allele length determined in the sizing experiment. One possible cause for the multiplicity of stutter bands may lie in the experimental protocol: instead of using DNA from a miniprep for sequencing, we used a purified PCR product, generated by amplification with M13 primers, hence adding one step of putative polymerase slippage. The use of a proofreading Taq polymerase instead of a conventional one might also have improved the situation. In *F. excelsior*, the low recovery of alleles is probably due to the fact that we failed to clone and sequence several long alleles in this species (repeat region of up to 190 bases). The cloning and sequencing of long alleles is complicated in two respects: (i) if they occur in heterozygous individuals together with a short allele, they amplify more weakly because of competition in the PCR reaction (short allele dominance, Wattier et al 1998); hence there is a lower concentration of the longer allele in the PCR product for cloning; and (ii) they display more stutter bands than short alleles as there is more opportunity for slippage in a long repeat region than in a short one (in agreement with higher mutation rates in long microsatellites, e.g. Estoup & Angers 1998). In heterozygous individuals where allele lengths differ by more than 20 nucleotides, we were indeed unable to recover the longer allele. Furthermore, we found sequences carrying very long AG repeats (>60 motifs), but the part the sequence downstream of the microstallite was degenerated, probably because of slippage that had occurred in the pre-sequencing PCR and during cycle sequencing. In *O. europaea* on the contrary, very few stutter bands were observed, and most

alleles differed by 6 bases. Together with the relatively low polymorphism in *O. europaea*, this indicates that there might be less slippage, i.e. less potential for mutation, at hexanucleotide than at dinucleotide repeats, in agreement with Chakraborty *et al.*'s (1997) result that mutation rates are on average inversely proportional to the size of the repeat unit. In *L. vulgare*, *F. ornus* and *P. angustifolia*, we sequenced numerous copies of a product of 157 bases that did not contain a microsatellite and showed 100% similarity among these species. A gene duplication can be excluded because the fragment shows no homology with the flanking regions of the microsatellite. A BLAST search to determine its origin did not reveal any conclusive results. Hence, it is probably a locus with priming sites similar to those of FEMSATL4, which does not amplify in *F. excelsior*, because primers are more specific in the species they have been designed for; although it cannot be excluded that this fragment corresponds to a contamination. In our category "unexpected sequences with microsatellite", we grouped two types of products: (i) products where repeat and flanking regions were those of the analysed species, but the product was of unexpected length, and (ii) products with repeat and/or flanking regions atypical for the species. The first type was mainly found in *F. excelsior* and *O. europaea*, and the second in the other species. In *P. angustifolia*, we identified sequence products with flanking and repeat regions typical for *O. europaea* in three out of four cloned individuals. In *F. ornus*, similar *O. europaea* products were found in 3 sequences from one cloned individual; and in *L. vulgare*, two sequences corresponded to the odd-sized allele type of *F. excelsior*. The simplest explanation for the detection of these sequences is contamination, which is perhaps more likely to occur in related species where primer annealing may not be optimal. However, one *F. ornus* sequence carried a hexanucleotide repeat typical for *O. europaea*, but corresponded entirely to the *F. ornus* allele type otherwise. Hence, the hypothesis of a gene duplication and independent evolution to that type of hexanucleotide repeat might deserve further attention in *F. ornus*. A larger sample size should help resolve whether we deal with an artefact here, or not.

Molecular diversity at locus FEMSATL4 in Fraxinus excelsior: origin and geographical distribution

In *F. excelsior*, the ancestral state of the microsatellite locus probably corresponds to the odd-sized allele type. It is indeed more abundant, more widespread geographically and notably, it can be identified in the phylogenetically close *F. ornus*, which is not the case for the even-sized allele type. The narrower size range of even-sized alleles also points to a shorter evolutionary history of this type. We were however unable to determine the origin of the odd-sized allele, because too few Oleaceae species were studied. The even-sized allele type has probably evolved from the odd-sized type since the *F. excelsior* – *F. ornus* split, by the deletion of the motif $AAA(CA)_n$. Its geographical distribution suggests two possible evolutionary scenarios, involving hybridisation with another *Fraxinus* species, or not.

Firstly, the deletion would have taken place in *F. excelsior* during the last glacial cold stage (115.000 – 10.000 radiocarbon years BP) in the Balkan area, and possibly near the Black Sea coast, where fossil pollen data provide evidence for a glacial refuge for ash (Brewer 2001). During post-glacial recolonisation (since 15.000 BP), the new allele type has colonised areas towards the north, but it could not spread all over Europe because of competition with northward migrating ashes from other glacial refuges. Therefore, its occurrence remained skewed towards southeastern Europe. This scenario can only be explained if the locus exhibits a very high mutation rate, for example on the order of 10^{-2} per generation, because 23 new alleles have been evolved since the deletion. The expounded mutation rate is possibly a realistic figure, regarding the tremendous polymorphism of the locus. A mutation rate at least one of magnitude lower would place the deletion back to at least 700.000 BP, approximately, the time before the European continent experienced repeated glacial periods (Webb and Bartlein 1992). During the latter, recurrent range contractions and expansions would have swept the phylogeographical pattern that is now observed among odd- and even-sized alleles.

Secondly, an alternative explanation for the observed phylogeographic pattern would be regular hybridisation of *F. excelsior* with another *Fraxinus* species, possibly *F. angustifolia*, which would carry the even-sized allele type at high frequency. Four ash species occur sympatrically in the Balkan area: *F. excelsior*, *F. angustifolia* ssp. *oxycarpa*, *F. ornus* and *F. pallisiae* (Tutin *et al.* 1972, and personal observation). *Fraxinus excelsior* and *F. angustifolia* ssp. *oxycarpa* have been reported to hybridise naturally in southeastern France (Rameau *et al.* 1989), and morphologically intermediate individuals are encountered in Romania (personal observation). According to Wallander (2001), *Fraxinus pallisiae* is a synonym for *F. angustifolia* ssp. *oxycarpa*; these species should hybridise more easily to *F. excelsior* than *F. ornus*, because the former three belong to section Fraxinus of the genus, whereas *F. ornus* belongs to section Ornus. Under the hybridisation scenario, the motif $AAA(CA)_n$ would have been lost in *F. angustifolia* s.l. after the *F. excelsior – F. angustifolia* divide. In that case, the presence of even-sized alleles in *F. excelsior* would be an indication for hybridisation with *F. angustifolia* s.l. The relative restriction of even-sized alleles to the Balkan peninsula can be explained by (i) the distribution of *F. angustifolia*, which is restricted to southern and eastern central Europe (Tutin *et al.* 1972), (ii) by the fact that we did not analyse *F. excelsior* populations from another region where both species co-occur (such as Italy or southern France), and (iii) by a stable demographic situation in the Balkan area, with little gene flow among regions (see Heuertz *et al.* unpublished). Under this scenario, the deletion would be much more ancient than under the first scenario and a lower mutation rate ($<10^{-3}$ per generation) for locus FEMSATL4 is therefore suggested. The hybridisation scenario is supported by the fact that a Romanian floodplain population of uncertain taxonomic status (*F. excelsior – F. angustifolia*) carried 60% of even-sized alleles.

Further, we found that step-wise like mutations did not significantly contribute to population differentiation at locus FEMSATL4, in neither allele-size type. This could be due to a low mutation rate and a relatively strong impact of the gene flow / genetic drift equilibrium to generate population differentiation. This result seems to weaken our first evolutionary scenario, where the even-sized allele type evolved in *F. excelsior* without hybridisation.

Evolution of microsatellite locus FEMSATL4 in the Oleaceae family

Our results reveal that the compound dinucleotide repeat locus originally isolated from *F. excelsior* bears a more complicated repeat structure when related species are analysed. The only elements common to all species are a $(CA)_n$ and a $(GA)_n$ repeat region, and might therefore have been contained in the repeat type ancestral to these five species. The phylogenetic proximity of *F. excelsior* and *F. ornus* on the one hand, and of *O. europaea* and *P. angustifolia* on the other (Fig. 3b) can be recognized by comparing their respective repeat regions (Fig. 2). Both species pairs differ from each other by a 16 nucleotides insertion/deletion of the stretch $CACG(CA)_3TGA_4$. After the split of *F. excelsior* and *F. ornus*, a transversion G $\rightarrow$ C has probably occurred in the *F. ornus* lineage, transforming a perfect (GA) into an interrupted sequence, which then continued to evolve by slippage as a hexanucleotide repeat (GAGACA). In *O. europaea*, the hexanucleotide repeat (GAGTGA) has probably evolved similarly, through an A $\rightarrow$ T transversion. It is recognized that microsatellites evolve by a combination of slippage events and base pair substitutions, however, the mutated nucleotide mostly seems to have the function of a repeat interruption, stabilising the microsatellite by reducing the possibility for further slip-strand mispairing (e.g. Angers & Bernatchez 1997, Kruglyak *et al.* 1998, Eisen 1999, Estoup & Cornuet 1999). Our results suggest that base substitutions in the microsatellite region produced an elongated repeat motif, but did not inhibit further evolution by slippage. The study of Primmer and Ellegren (1998) also describes length variation of the repeat unit in a microsatellite locus (in birds), but that situation is more complicated than ours, and it has been interpreted by a combination of substitutions and indels.

Mutations in regions immediately flanking the microsatellites repeat have been reported to be more common than in sequences further away (Brohede & Ellegren 1999). Our results agree with this, notably because the sequence immediately adjacent to the left side of the $(GA)_n$ repeat in *L. vulgare* shows much less homology to the corresponding sequence in both *Fraxinus* species than the 25 sequenced bases located upstream. More generally, the left side of our microsatellite locus seems to be more variable overall than the right side, showing more substitutions in the flanking regions, an indel of the sequence flanking immediately the repeat, and point mutations in this sequence when it has been conserved.

Concluding remarks

The results of this study unravelled the molecular diversity accounting for the appearance of odd- and even-sized allele types at locus FEMSATL4 in *Fraxinus excelsior* and suggested two evolutionary scenarios to explain the current geographic distribution of these allele types. Further, support for the evolution of a hexanucleotide repeat through a base substitution and subsequent slippage was given on two occasions. In contrast, our results were unable to determine (i) the origin of the stretch $AAA(CA)_n$, which distinguishes both *F. excelsior* allele types, and (ii) the time of its deletion, i.e. discriminate the two evolutionary scenarios. The sequencing of the microsatellite locus in other closely related genera (Cormoranthus, Schrebera and from the subtribe Oleinae, see Wallander & Albert 2000) and in other *Fraxinus* species, especially in *F. angustifolia*, is expected to resolve these issues.

References

Amos W, Sawcwe SJ, Feakes RW, Rubinsztein (1996) Microsatellites show mutational bias and heterozygote instability. *Nature Genetics*, **13**, 390-391.

Angers B, Bernatchez L (1997) Complex Evolution of a Salmonid Microsatellite locus and its consequences in inferring allelic divergence from size information. *Molecular Biology and Evolution*, **14**, 130-138.

Balloux F, Goudet J (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology*, **11**, 771-783.

Barrier M, Friar E Robichaux R, Purugganan M (2000) Interspecific evolution in plant microsatellite structure. *Gene*, **241**, 101-105.

Blankenship SM, May B, Hedgecock D (2002) Evolution of a perfect simple sequence repeat locus in the context of its flanking sequence. *Molecular Biology and Evolution*, **19**, 1943-1951.

Brewer S (2001) *Fossil pollen maps*. Available form http://www.pierroton.inra.fr/cytofor

Brohede J, Ellegren H (1999) Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. *Proceedings of the Royal Society of London B*, **266**, 825-833.

Cakraborty R, Kimmel M, Stivers DN, Davidson J, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences of the USA*, **94**, 1041-1046.

Colson I, Goldstein DB (1999) Evidence for complex mutations at microsatellite loci in Drosophila. *Genetics*, **152**, 617-627.

Crawford AM, Kappes SM, Paterson KA, de Gotari MJ, Dodds KG, Freking BA, Stone RT, Beattie CW (1998) Testing the ascertainment hypothesis. *Journal of Molecular Evolution*, **46**, 256-260.

Culver M, Menotti-Raymond MA, O'Brien SJ (2001) Patterns of size homoplasy at 10 microsatellite loci in pumas (*Puma concolor*). *Molecular Biology and Evolution*, **18**, 1151-1156.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational Processes of simple sequence repeats in human populations. *Proceedings of the National Academy of Sciences of the USA*, **91**, 3166-3170.

Eisen JA (1999) Mechanistic basis for microsatellite instability. In: *Microsatellites – Evolution and applications* (Goldstein DB, Schlötterer C, eds.), pp. 34-48. Oxford University Press, Oxford, UK.

Ellegren H, Moore S, Robinson N, Byrne K, Ward W, Sheldon BC (1997) Microsatellite evolution: a reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Molecular Biology and Evolution*, **14**, 864-860.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* L. Skeels) endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832-839.

Estoup A, Angers B (1998) Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In: *Advances in molecular ecology* (ed. Carvalho GR), pp 55-86. IOS Press, Amsterdam, The Netherlands.

Estoup A, Cornuet J-M (1999) Microsatellite evolution: inferences from population data. In: *Microsatellites – Evolution and applications* (Goldstein DB, Schlötterer C, eds.), pp. 49-65, Oxford University Press, Oxford, UK.

Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591-1604.

Estoup A, Tailliez C, Cornuet J-M, Solignac M (1995) Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Molecular Biology and Evolution*, **12**, 1074-1084.

Felsenstein J (1995) PHYLIP (Phylogeny Inference Package) version 3.57, available from http://evolution.genetics.washington.edu/phylip.html.

Goldstein DB, Linares AR, Feldman MW and Cavalli-Sforza LL (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463-471.

Goudet J (1995) FSTAT (vers.1.2), a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile compute program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618.

Hardy OJ, Charbonnel N, Freville H, Heuertz M (2002) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* (in press).

Heuertz M, Hausman JF, Hardy OJ, Vendramin GG, Frascaria-Lacoste N, Vekemans X. Nuclear microsatellites reveal contrasting patterns of diversity in western and eastern European populations of the common ash (*Fraxinus excelsior* L.) (unpublished).

Heuertz M, Hausman JF, Tsvetkov I, Frascaria-Lacoste N, Vekemans X, 2001. Assessment of genetic structure within and among Bulgarian populations of the common ash (*Fraxinus excelsior* L.). *Molecular Ecology* 10, 1615-1623.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424-429.

Karhu A, Dieterich J-H, Savolainen O (2000) Rapid extension of microsatellite sequences in pines. *Molecular Biology and Evolution*, **17**, 259-265.

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences of the USA*, **95**, 10774-10778.

Lefort F, Brachet S, Frascaria-Lacoste N, Edwards KJ, Douglas GC (1999) Identification and characterisation of microsatellite loci in ash (*Fraxinus excelsior* L.) and their conservation in the olive family (*Oleaceae*). *Molecular Ecology*, **8**, 1088-1091.

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, **4**, 203-221.

Markova KD, Nekrutenko A, Baker JB (2000) Evolution of microsatellite alleles in four species of mice (genus Apodemus). *Journal of Molecular Evolution*, **51**, 166-172.

Nicholas KB, Nicholas HB (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Available from http://www.psc.edu/biomed/genedoc.

Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*, **22**, 201-204.

Primmer CR, Ellegren H, Saino N, Møller AP (1996a) Directional evolution in germline microsatellite mutations. *Nature Genetics*, **13**, 391-393.

Primmer CR, Møller AP, Ellegren H (1996b) A wide-range survey of cross-species microsatellite amplification in birds. *Molecular Ecology*, **5**, 365-378.

Primmer CR, Ellegren H (1998) Patterns of evolution in avian microsatellites. *Molecular Biology and Evolution*, **15**, 997-1008.

Queller DC, Strassmann JE, Hughes CR (1993) Microsatellites and kinship. *Trends in Ecology and Evolution*, **8**, 285-288.

Rameau JC, Mansion D, Dumé G (1989) Flore forestière française, guide écologique illustré. 1. Plaines et collines. Institut pour le développement forestier, Paris, France.

Raymond M, Rousset F (1995) GENEPOP version 1.2 Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics,* **139**, 457-462.

Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA, eds (1972) *Flora Europaea*, volume 3. Cambridge University Press, UK.

Udupa SM, Baum M (2001) High mutation rate and mutational bias at (TAA)(n) microsatellite loci in chickpea (*Cicer arietinum* L.). *Molecular Genetics and Genomics*, **265**, 1097-1103.

Van Oppen MJH, Rico C, Turner GF, Hewitt GM (2000) Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in lake Malawi Cichlids. *Molecular Biology and Evolution*, **17**, 489-498.

Vigouroux Y, Jaqueth J, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J (2002) Rateand pattern of mutation at microstellite loci in maize. *Molecular Biology and Evolution*, **19**, 1251-1260.

Wallander E (2001) Evolution of wind-pollination in *Fraxinus* (Oleaceae) – an ecophylogenetic appraoch. PhD thesis. Botanical Institute, Göteborg University, Sweden.

Wallander E, Albert VA (2000) Phylogeny and classification of Oeaceae based on *rps16* and *trnL-F* sequence data. *American Journal of Botany*, **87**, 1827-1841.

Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology*, **7**,1569-1573.

Webb T and Bartlein PJ (1992) Global changes during the last 3 million years: climatic controls and biotic response. *Annual Review of Ecology and Systematics*, **23**, 141-173.

Chapter V: Incongruence between chloroplast and nuclear marker data among European populations of common ash

Abstract

Forty-two common ash populations from Europe were analysed at six chloroplast microsatellite loci. Overall, little polymorphism was detected and allelic variants at four polymorphic loci combined into eight haplotypes. The clear-cut geographic organisation of chloroplast haplotype variation suggested that recolonisation of western and central Europe occurred from refuges in the Apennine and the Balkan Peninsulas, and potentially the Iberian Peninsula, in agreement with fossil pollen data. The comparison with previously obtained nuclear microsatellite data revealed contrasting patterns. Most interestingly, strong divergence among populations from southeastern Europe and Sweden was observed at nuclear markers, whereas chloroplast markers detected only very little variation. In southeastern Europe, distinct diversity patterns were identified east and west of the Carpathian and Balkan mountains. Diversity patterns at chloroplast and nuclear markers are jointly examined in different regions of Europe and discussed in the context of population demography and postglacial recolonisation in the current glacial interstadial. The particular value of southeastern Europe and Sweden for the conservation of genetic resources in common ash is emphasised. Our results suggest that efficient management and conservation strategies should be based on data from differentially inherited genetic markers.

Introduction

Two main methodological approaches have been used to characterise patterns of genetic variation over a wide distribution range, with the aim of making inference on the main causal factors: (1) phylogeographic studies based on organelle DNA (mainly chloroplast) patterns of variation; (2) population genetic structure studies based on analyses of nuclear genetic marker diversity.

In flowering plants, where the chloroplast genome is usually maternally transmitted (Birky 1988), much stronger population genetic structure is generally found for chloroplast than nuclear markers (Ennos 1994, Le Corre *et al.* 1997, Raspé *et al.* 2000). This genetic structure often presents a clear phylogeographic pattern, which means that there is concordance of phylogenetic relationships and spatial distribution of genotypes (e.g. King & Ferris 1998, Petit *et al.* 2002a, Palmé & Vendramin 2002; see Taberlet *et al.* 1998, Comes & Kadereit 1998, Newton *et al.* 1999 for reviews). Several factors may account for these observations. In angiosperms, the chloroplast genome is generally subjected to a lower level of gene flow than nuclear genes because it is dispersed only in the seed, whereas nuclear genes are transmitted by both seed and pollen. Further, as the chloroplast genome is haploid and uniparentally inherited (Birky 1988, Clegg 1989), it evolves in hermaphrodite diploid plants under a two-fold higher pressure of genetic drift than the nuclear genome (Birky *et al.* 1983). Therefore, when populations are relatively small and isolated, substantial genetic impoverishment may occur within, and large genetic divergence between them at chloroplast DNA markers. In many species, this was for instance the case for populations surviving in separate refuges during the glaciations of the Quaternary (Hewitt 2000). The comparison of phylogenetic relationships among organelle haplotypes with their current geographical occurrence allows to make inference on postglacial recolonisation routes (Ferris *et al.* 1998, Petit *et al.* 2002b), to propose high diversity areas as putative glacial refuges (King & Ferris 1998, Huang *et al.* 2002, Petit *et al.* 2002b) and to corroborate the location of refuges suggested on the basis of palynological data (Demesure *et al.* 1996, Petit *et al.* 2002b). Uniparentally inherited organelle markers are also used to study the directionality of gene flow in species complexes (Bucci *et al.* 1998, Dumolin-Lapègue *et al.* 1999) and to identify populations of non-local origin (Demesure *et al.* 1996, Sperisen *et al.* 1999, Vendramin *et al.* in preparation). However, some important evolutionary processes such as recent demographic changes, or interspecific hybridisation events may remain undetected with organelle markers because they often display low polymorphism and have uniparental transmission. Moreover, each organelle genome represents only one occurrence of a genealogical process, which obviously does not keep trace of all major evolutionary events in the species history (Hare 2001).

Nuclear markers are highly suited to study large-scale geographic variation in levels of within-population diversity, and by this means to identify imprints of founder effects resulting

from long-distance dispersal during postglacial recolonisation (Lagercrantz & Ryman 1990, Zanetto & Kremer 1995, Comps *et al.* 2001). Geographic variation in patterns of genetic structure can also be studied with nuclear markers at high resolution (Machon *et al.* 1997, Murillo & Finkeldey 2000), and hybridisation events between closely related species or between lineages at the intraspecific level can be identified (Marchelli & Gallo 2001, Breitenbach-Dorfer *et al.* 1997). However, a limitation with nuclear markers in tree species is that most variation is found within populations so that their phylogeographic information content is low. As a consequence, alternative causes of the observed patterns of variation, i.e. recent gene flow or recolonisation history, cannot be easily distinguished (Zanetto & Kremer 1995). In this respect, nuclear haplotype data are becoming a promising new tool for evolutionary inference (Hare 2001). Another strategy is to assess simultaneously organelle and nuclear DNA markers on the same set of populations. The combination of both markers may lead to a better understanding of the relative roles of different evolutionary factors, i.e. gene flow, postglacial history, and demographic events (Latta & Mitton 1997, Le Corre *et al.* 1997, Terry *et al.* 2000). Moreover, when no phylogeographic structure occurs, this allows to determine the relative contributions of seed and pollen to gene flow (Ennos 1994).

The general objective of this work was to combine both approaches on the same population samples to better characterise diversity patterns and to test whether the combination provides a clearer identification of the factors shaping these patterns. Our study plant is the common ash, *Fraxinus excelsior*, a temperate tree species occurring in floodplain forests and in mixed deciduous forests nearly all over Europe. Common ash exhibits intermediate properties between a pioneer species and a permanent forest component; it features strong colonisation capacity, but competitive ability is high only when ecological requirements are met. Flowers are wind-pollinated and single-seeded fruits, the samaras, are wind-dispersed. The mating system of common ash is polygamous and highly variable: flowers are male, hermaphroditic or female and there is a continuum from pure male to pure female individuals with hermaphroditic intermediates (Picard 1982, A Lamb & D Boshier unpublishe,; Bingelli & Power 1999, Wallander 2001). Palynological data strongly support full glacial refuges for ash in the eastern Alps and in the Balkan Peninsula (Huntley & Birks, 1983; Gliemeroth 1997; Brewer 2001), with up to three sites in the latter (Brewer 2001). Fossil pollen evidence is weaker for putative refuges in the Iberian Peninsula, in the Italian Peninsula, in Turkey and at the northern Black Sea coast (Huntley & Birks, 1983, Gliemeroth 1997, Brewer 2001). Further, the most detailed survey (Brewer 2001) identified sites compatible with early occurrence of ash in the French Central Massif, in southeastern Poland, in Scotland at the northern coast of the Firth of Forth and in southern Sweden near Göteborg. The role of these sites in postglacial history of ash still needs to be clarified. Postglacial colonisation was marked by an east-west migration in the early Holocene (after 10.000 BP), followed by a south-north trend.

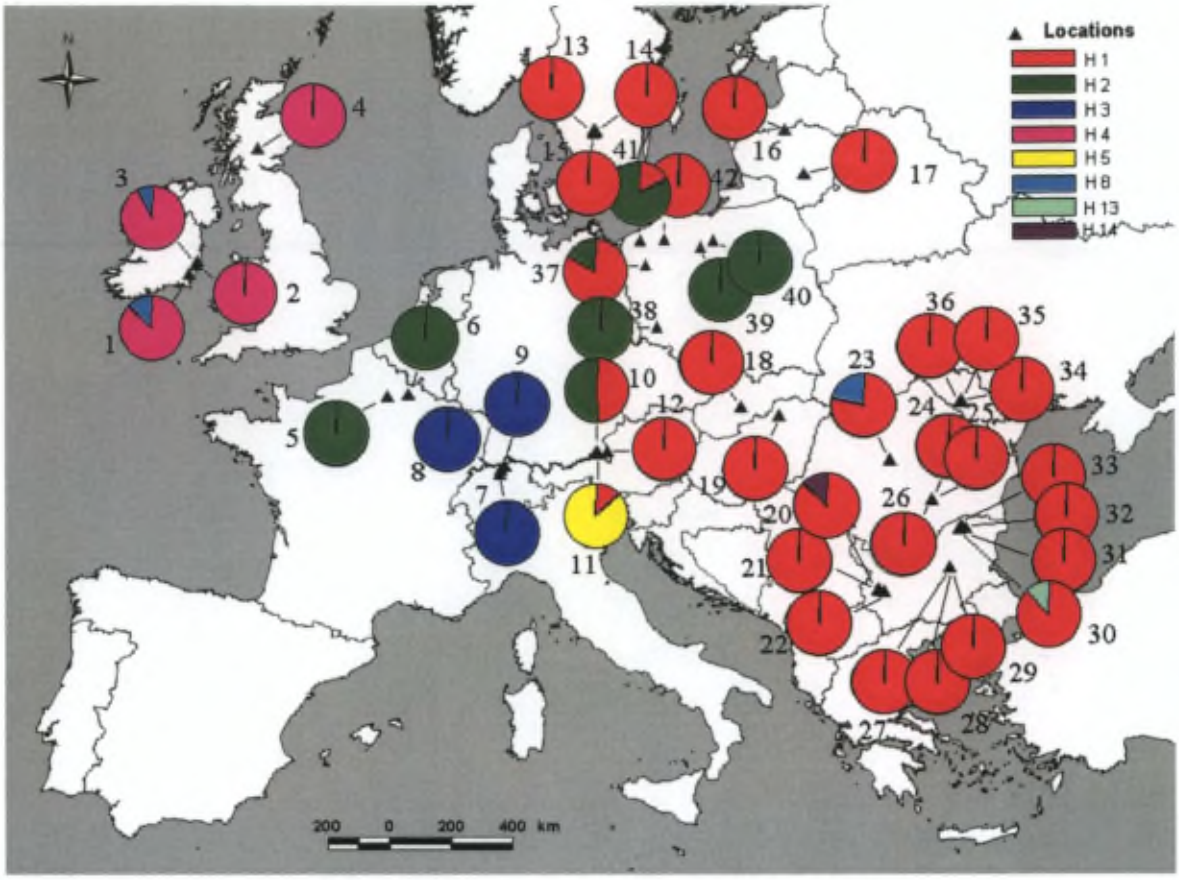


Fig. 1 Geographic distribution of the haplotypes identified with chloroplast microsatellites.

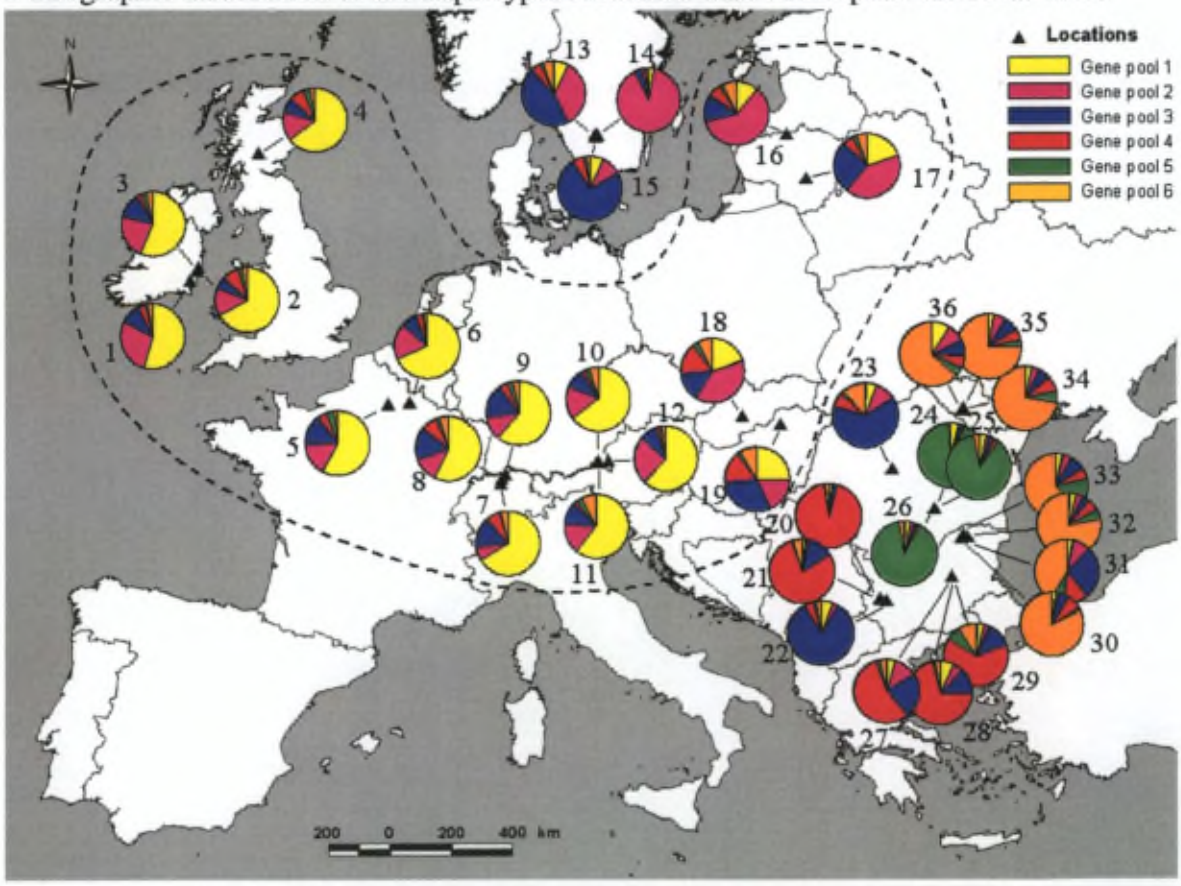


Fig. 2 Geographic distribution of the gene pools identified from nuclear microsatellites according to the model-based clustering method by Pritchard *et al.* (2000) (from Heuertz *et al.* unpublished, chapter 3). The "western European group" is delimited with dashes (see text).

Nuclear microsatellites revealed strongly differentiated gene pools for common ash in southeastern Europe, whereas the partitioning of diversity in central and western Europe is much more uniform (Heuertz *et al.* unpublished, chapter 3). Also, we observed an increase of genetic diversity from populations close to refuges towards recently recolonised areas, which we attributed to admixture events of distinct genepools during recolonisation (Heuertz *et al.* unpublished, chapter 3).

In this study we assessed genetic diversity at six chloroplast microsatellite loci in 42 European populations of common ash, including the 36 populations previously analysed with nuclear microsatellite markers. Our objectives were (i) to identify geographical patterns of within- and among-population diversity at chloroplast markers, (ii) to compare them to the structure of genetic variation at nuclear markers, and (iii) to explain the current diversity patterns at both types of markers in terms of population demography and postglacial recolonisation dynamics during the current glacial interstadial.

Materials and Methods

Material

Samples of *Fraxinus excelsior* were collected in 42 putatively autochthonous forests in Europe, comprising the 36 populations previously analysed at nuclear microsatellite loci (Heuertz *et al.* unpublished, chapter 3; Figs. 1 and 2) and six additional populations from Poland (number 37, Choszno, 53°10'40"N, 15°18'09"E; 38, Jawor, 50°58'58"N, 16°12'23"E; 39, Jamy-Chelmo, 53°20'59"N, 18°23'52"E; 40, Jamy-Bialochowo, 53°32'59"N, 15°18'09"E; 41, Nowogard, 53°38'27"N, 15°05'27"E; 42, Szczecinek, 53°44'35"N, 16°42'28"E; Fig. 1). Samples consisted of buds or leaves. Buds were brought to the laboratory on their twigs, wrapped in wet paper. They were dissected out, separated from their scales and conserved at –70°C prior to DNA extraction. Leaves were dried between paper sheets or in plastic bags in the presence of silica gel. They were kept at room temperature until DNA extraction. Total DNA was extracted from 60 to 90 mg of dry leaves or from 50-70 mg fresh weight of buds with the DNeasy plant mini kit (Qiagen) or the CTAB procedure of the NucleoSpin Plant kit (Macherey Nagel).

Chloroplast microsatellite analysis

A total of 323 trees were analysed at chloroplast microsatellite (cpSSR) loci, which corresponds to an average of 7.7 ± 3.2 (SD) trees per population. Chloroplast microsatellites corresponding to poly(A) or poly (T) repeats were polymerase chain reaction (PCR) amplified with six universal primer pairs for angiosperms (ccmp2, ccmp3, ccpm4, ccmp6, ccmp7, and ccmp10 from Weising & Gardner 1999) in a total reaction volume of 25 µl with the following

profile: 5 min 96°C, 25 cycles of 1 min 94°C, 1 min 55°C, 1 min 72°C with a final extension step of 7 min at 72°C. The reaction mix contained four dNTPs, each 0.2 mM, 2.5 mM of $MgCl_2$, 0.2 μM of each primer, approximately 20 ng of template DNA and 1 U of Taq polymerase (Pharmacia) in Pharmacia PCR buffer. Amplification products were multiplexed by size (ccmp2, ccmp3 and ccmp10 on the one hand, and ccmp4, ccmp6 and ccmp7 on the other) and loaded onto Reprogel Long Read acrylamide gels (Pharmacia). Electrophoresis was run for 70 min at 1500V on an automated ALF express DNA sequencer (Pharmacia) in TBE buffer. Fragment sizes were determined by comparison with internal and external size standards with the software Fragment Manager 1.2 (Pharmacia).

Data analysis

A chloroplast haplotype network was constructed based on the minimum number of length differences of PCR-amplified fragments separating the haplotypes, using TCS 1.13 (Clement *et al.* 2000). The distance matrix among haplotypes was computed defining the distance between two haplotypes as the sum over the 6 loci of the absolute number of nucleotides distinguishing the haplotypes, assuming that a one-nucleotide difference corresponded to a single mutation, i.e. a slippage event at a microsatellite locus involving a single nucleotide.

Within populations, we determined the number of alleles K and the haplotypic diversity based on haplotype identity (h_S) or on the minimum number of microsatellite length mutations separating haplotypes (v_S) following Pons and Petit (1996). In order to allow for straightforward comparison of within-population haplotypic diversity statistics, the weights for v-type statistics were divided by a correction factor computed from the overall matrix of differences among haplotypes (the matrix that we also used to build the haplotype network, see above) according to Petit *et al.* (2002a). Geographic variation of within-population diversity was investigated by computing the Pearson product moment correlation between diversity statistics (number of alleles, h_S, v_S) and geographical coordinates.

In the overall sample, total haplotypic diversity statistics based on ordered or unordered alleles (h_T and v_T, respectively) were calculated following Pons and Petit (1996). Differentiation among populations was computed from ordered and from unordered alleles (N_{ST} and G_{ST}, respectively) and a permutation procedure was carried out to test for the presence of phylogeographic structure (i.e. whether $N_{ST} > G_{ST}$) following Burban *et al.* (1999), using the program PERMUT 2 (Pons & Petit 1996). An analysis of isolation by distance was performed using the matrix of F_{ST} values among pairs of populations computed with ARLEQUIN 2.0 (Schneider *et al.* 2000) and the matrix of the logarithms of geographic distances among pairs of populations computed with SPAGeDi 1.0 (Hardy & Vekemans 2002). Values of F_{ST} were regressed on the logarithm of distance and a Mantel test between both matrixes was computed with the program CADM (Legendre 2001).

A previous analysis of 36 common ash populations with five nuclear microsatellite (nSSR) loci has revealed distinct patterns of genetic diversity among eastern and western European populations (Heuertz *et al.* unpublished, chapter 3). More precisely, our data have suggested congruent patterns of variation in western and central Europe and in Lithuania on the one hand (hereafter referred to as "western group"), and in southeastern Europe and Sweden on the other (hereafter referred to as "eastern group"). A synthetic visualisation of the patterns was obtained with the model-based clustering method by Pritchard *et al.* (2000) (Fig. 2). In this study, we tested the discrepancy between the two suggested groups of populations at the European scale, as well as between populations located east or west of the Carpathian and Balkan mountain range at the scale of southeastern Europe. We performed permutation tests with FSTAT 2.3.5 (Goudet 2001) for the following statistics in each group: the within-population allelic richness, A_S, standardised to a sample size of 18 diploid individuals (which corresponds to the smallest sample size among the 36 populations) according to El Mousadik and Petit (1996), the within-population genetic diversity (H_E), the within-population inbreeding coefficient (F_I), and the statistic of among-population differentiation, F_{ST}. We also tested with an analysis of variance whether there were differences among groups in the mean value of the bottleneck statistic T_2 (Cornuet & Luikart 1996). This statistic is based on the principle that a drastic size reduction in a population, i.e. a bottleneck, transiently produces a stronger decrease of the number of alleles than of gene diversity H_E (Nei *et al.* 1995). The T_2 statistic measures an average over loci of the difference between the actual gene diversity H_E and the gene diversity expected from the number of alleles in the population, assuming drift-mutation equilibrium. Hence positive values of T_2 are expected in case of a recent bottleneck whereas negative values indicate population expansion without immigration. We provide T_2 values under the infinite alleles model (T_2 IAM) and under the step-wise mutation model (T_2 SMM), which represent two extreme mutation models, as our loci do not follow a strict step-wise mutation model (we observed many alleles differing by one nucleotide, which is less than the repeat unit).

Chloroplast and nuclear patters of variation were compared by means of the Pearson product moment correlation coefficient between within-population diversity statistics at both types of markers, including (i) the number of alleles (K) and haplotypic diversity based on ordered (v_S) or unordered (h_S) alleles for chloroplast microsatellites; and (ii) allelic richness (A_S), gene diversity (H_E) and the inbreeding coefficient (F_I) for nuclear microsatellites (Heuertz *et al.* unpublished, chapter 3). As we tested multiple correlations, a sequential Bonferroni correction (Rice 1989) was applied to significance levels in order to exclude correlations that might have appeared by chance. We also compared pairwise population differentiation (F_{ST}) matrixes for both types of markers with a Mantel test using CADM (Legendre 2001). These comparisons were performed over all 36 populations and within each of the two groups of populations for which nuclear markers had suggested contrasting diversity patterns.

Table 1 Characteristics of the cpDNA haplotypes of the common ash identified from six microsatellite loci. Codes of loci are as in Weising and Gardner (1999) and codes of haplotypes as in Vendramin *et al.* (in preparation).

Haplotypes	number of individuals	size of amplified fragment (nucleotides)					
		ccmp2	ccmp3	ccmp4	ccmp6	ccmp7	ccmp10
H1	207	194	97	140	97	118	103
H2	39	194	97	140	99	117	104
H3	16	194	97	140	99	117	103
H4	47	194	97	140	98	118	104
H5	7	194	97	140	98	117	103
H8	5	194	97	140	97	118	104
H13	1	194	97	140	97	118	106
H14	1	194	96	140	97	118	103

Table 2 Chloroplast marker diversity statistics from Pons and Petit (1996) from 42 common ash populations in Europe. Standard errors are given between brackets.

Unordered alleles		Ordered alleles	
h_S	0.066 (0.0212)	v_S	0.060 (0.0243)
h_T	0.556 (0.0762)	v_T	0.591 (0.0786)
G_{ST}	0.881 (0.0389)	N_{ST}	0.898 (0.0410)

Table 3 Statistics of chloroplast genetic diversity in polymorphic populations of the common ash: *n*, number of individual trees analysed; *K*, number of haplotypes identified; h_S and v_S, genetic diversity based on unordered and ordered alleles, respectively. All 33 remaining populations were monomorphic.

Population	Country	n	*K*	h_S	v_S
1 Camolin	Ireland	8	2	0.250	0.087
2 Kilmacurra	Ireland	12	2	0.167	0.058
10 Tiroler Ache	Germany	8	2	0.571	0.792
11 Chiemsee	Germany	8	2	0.250	0.170
20 Ljulin Monastery	Bulgaria	8	2	0.250	0.087
23 Tutuleac	Romania	14	2	0.363	0.126
30 Iri Hissar	Bulgaria	8	2	0.250	0.250
37 Choszno	Poland	6	2	0.333	0.462
41 Nowogard	Poland	6	2	0.333	0.462

Results

Variation at chloroplast microsatellites

Two of the six cpSSR loci analysed were monomorphic (ccmp2 and ccmp4, Table 1). The other loci showed low levels of polymorphism. Two distinct size variants separated by one nucleotide were observed at both ccmp3 and ccmp7 (Table 1). Ccmp6 and ccmp10 displayed three size variants each; with amplification fragment sizes of 96, 97 and 98 nucleotides, and 103, 104 and 106 nucleotides, respectively (Table 1). These size variants combined into a total of 8 haplotypes (Table 1). In the overall sample of 323 individuals, two haplotypes (H13 and H14) were encountered in only one individual; the frequencies of the others ranged from 5 individuals (1.5 %, H8, only observed in population 11 from Germany) to 207 individuals (64.1 %, H1, Table 1). The minimum numbers of evolutionary events separating the haplotypes is visualised in the haplotype network in Fig. 3. Most haplotypes are related to 1 to 3 others by one single mutational step. Three putative haplotypes corresponding to intermediate evolutionary steps were not detected in our dataset (circles in Fig. 3).

Geographical distribution of cpSSR polymorphism

Within-population variation at chloroplast microsatellites was low, with an average number of K=1.21±0.41 (standard deviation) haplotypes per population and average gene diversity values of h_S=0.066 and v_S=0.060 based on unordered or ordered alleles, respectively (Table 2). Accordingly, most populations carried a single haplotype, and only nine population samples displayed two distinct haplotypes (Fig. 1). The diversity measure based on ordered alleles (v_S) was higher than that based on allele identity (h_S) in polymorphic populations from Germany and Poland; in the other polymorphic populations, v_S was smaller than h_S (Table 3). No correlation was observed between within-population diversity statistics and geographical coordinates (Table 4); hence our dataset bears no evidence for gradual variation of diversity according to latitude or longitude.

Haplotypic diversity in the overall sample was much higher than within populations, h_T = 0.556, v_T = 0.591, and consequently, differentiation among populations was important: G_{ST} = 0.881 and N_{ST} = 0.898 (Table 2). The permutation procedure to test for phylogeographic structure revealed that N_{ST} was not significantly larger than G_{ST}, indicating that on average, phylogenetically close haplotypes were not found together in the same population more often than different haplotypes which would have been chosen randomly. A geographical organisation of genetic diversity is nevertheless evident from Fig. 1, with haplotype H1 being predominant in eastern and northern Europe, haplotypes H2, H3 and H5 occurring in central and western Europe, and haplotype H4 being found only in the British Isles. Also, the test of

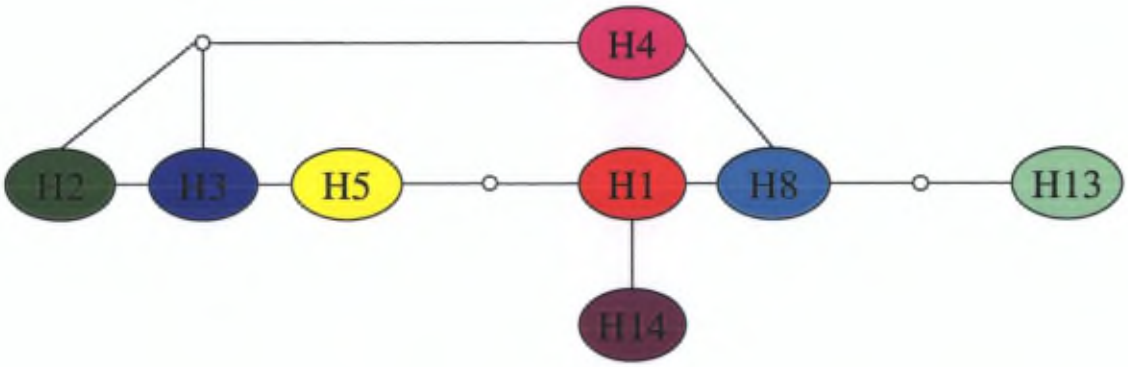


Fig. 3 Haplotype network constructed from variation at the chloroplast microsatellite loci. The circles represent putative haplotypes that were not detected in our dataset.

Table 4 Pearson moment correlation coefficients among within-population diversity statistics computed from chloroplast microsatellite data and geographical coordinates or within-population diversity statistics computed from nuclear microsatellite data. All correlations were non significant.

	Diversity at chloroplast microsatellites		
	K	h_S	v_S
Latitude	0.011	-0.000	0.050
Longitude	-0.209	-0.140	-0.084
Diversity at nuclear microsatellites			
A_S	-0.239	-0.158	0.017
H_E	-0.167	-0.109	0.035
F_I	-0.118	-0.217	-0.244

Table 5 Statistics of genetic diversity and differentiation at nuclear microsatellite markers between groups of common ash populations. Compare population numbers with Fig. 2. A_S, allelic richness for population samples of standardised size 18 diploid individuals; H_E, gene diversity; T_2 IAM and T_2 SMM, bottleneck statistics based on the infinite alleles model and the step-wise mutation model, respectively; F_I, Wright's inbreeding coefficient; F_{ST}, relative among-population differentiation based on allele identity; Test, significance level (P) of tests with 10000 permutations for A_S, H_E, F_I and F_{ST}, and results of the analysis of variance for the bottleneck statistics T_2 IAM and T_2 SMM, where F is the ratio of mean squares to compare with the F-distribution, and P the significance level obtained.

European scale			
Populations from western and central Europe and Lithuania (numbers 1-12, 16-19)	Populations from Sweden and southeastern Europe (numbers 13-15, 20-36)	Test	
$A_S = 12.355$	$A_S = 10.508$		$P = 0.0048$
$H_E = 0.814$	$H_E = 0.729$		$P = 0.0001$
T_2 IAM = 0.648	T_2 IAM = -0.711	$F = 17.22$	$P = 0.0002$
T_2 SMM = -3.055	T_2 SMM = -6.438	$F = 17.27$	$P = 0.0002$
$F_I = 0.064$	$F_I = 0.010$		$P = 0.0008$
$F_{ST} = 0.027$	$F_{ST} = 0.105$		$P = 0.0001$
Southeastern European scale			
Populations from within or west of the Carpathians and Balkans (numbers 20-23, 27-29)	Populations from east of the Carpathians and Balkans (numbers 24-26, 30-36)	Test	
$A_S = 8.941$	$A_S = 12.162$		$P = 0.0013$
$H_E = 0.714$	$H_E = 0.769$		$P = 0.0261$
T_2 IAM =-0.045	T_2 IAM =-0.970	$F = 2.58$	$P = 0.1292$
T_2 SMM =-4.942	T_2 SMM =-6.850	$F = 1.68$	$P = 0.2143$
$F_I = 0.004$	$F_I = 0.017$		$P = 0.5709$
$F_{ST} = 0.101$	$F_{ST} = 0.067$		$P = 0.2310$

isolation by distance showed that among-population differentiation increased significantly with the logarithm of geographical distance (Mantel test: $P < 0.001$, linear regression: $R^2 = 0.222$).

Geographical distribution of nSSR polymorphism

The contrasting patterns in the geographical distribution of genetic diversity at nuclear microsatellite markers at the European scale suggested in our previous study (Heuertz *et al.* unpublished, chapter 3) were confirmed (Table 5): western and central European, as well as Lithuanian populations (i.e. the western group) displayed highly significantly larger allelic richness, gene diversity and inbreeding coefficients, but lower levels of among-population differentiation than populations from southeastern Europe and Sweden (i.e. the eastern group). The difference in gene diversity was more marked ($P = 0.0001$) than that in allelic richness ($P = 0.0048$). As a consequence, we observed a gene diversity excess with respect to the one expected from the number of alleles in the western group (T_2 IAM = 0.648), and a gene diversity deficiency in the eastern group (T_2 IAM = -0.711). Assuming a stepwise mutation model, all values for the bottleneck statistic T_2 SMM were negative, but still higher in the western than in the eastern group. Hence, western European populations carry stronger imprints from recent population bottlenecks than eastern European populations.

Within the Balkan Peninsula, we were able to identify two groups of populations separated by the physical barrier of the Carpathian and Balkan mountains. Populations at the foothills of the mountains or west of them harboured significantly lower allelic richness and gene diversity than populations from the Danube and Dniestr plains in the east (Table 5). The model-based clustering method by Pritchard *et al.* (2000) also assigned different gene pools to these groups of populations, gene pools 3 and 4 in the west, and gene pools 5 and 6 in the east (Fig. 2).

Comparison of chloroplast and nuclear patterns of variation

The comparison between chloroplast and nuclear within-population diversity statistics over all 36 populations revealed low and non-significant correlation coefficients (Table 4). When populations were split into a western and an eastern group, as suggested by patterns of nuclear microsatellite diversity, a weak negative association was observed in the western group between the phylogenetic diversity of haplotypes, v_S, and the inbreeding coefficient F_I (Pearson correlation coefficient $r = -0.517$), but it was non-significant after application of a sequential Bonferroni correction. The Mantel test identified a negative correlation between the matrixes of pairwise F_{ST}-values among populations at chloroplast and nuclear markers (Spearman rank correlation coefficient $r = -0.263$, $P = 0.993$ for the one-tailed Mantel test

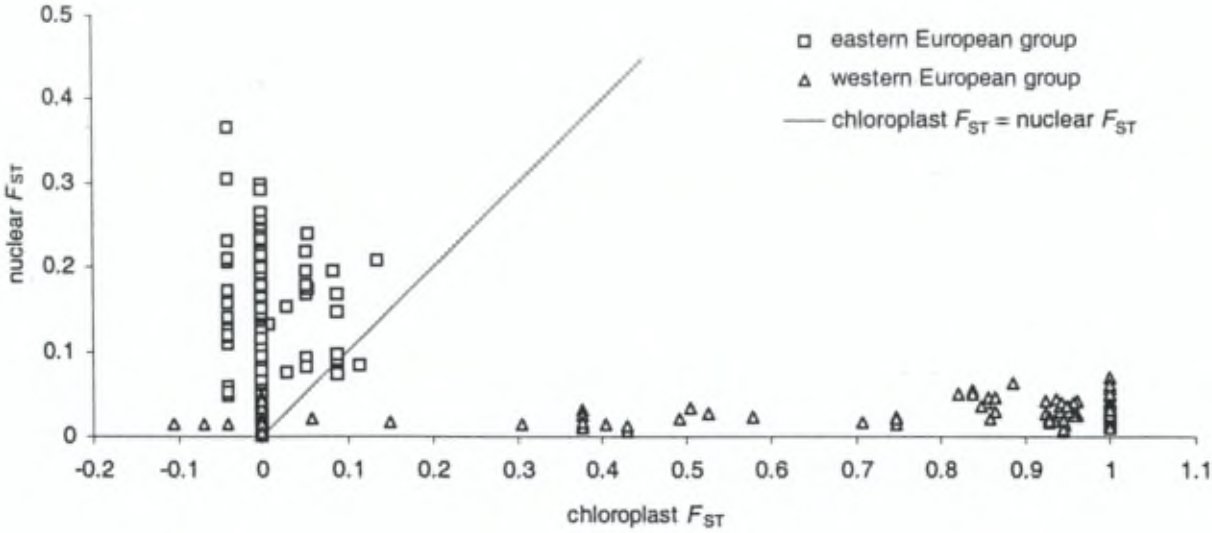


Fig. 4 Patterns of differentiation among populations. F_{ST} statistics were computed between pairs of populations and values at nuclear loci were plotted against those at chloroplast loci. The plot represents values computed among populations within each of two groups defined on the basis of polymorphism patterns at nuclear loci (see materials and methods). Squares represent differentiation values among eastern European populations; triangles stand for differentiation among western European populations. Between-group comparisons are not shown.

with H1: $r > 0$, $P = 0.000$ for the two-tailed Spearman rank correlation test), indicating that high differentiation at nSSRs occurred concomitantly with low differentiation at cpSSRs, and reciprocally. It appeared that this negative correlation was essentially due to the comparison of pairs of populations within each of the groups defined from diversity patterns at nuclear loci. The western European group showed low differentiation among populations at nuclear markers, but high divergence among populations at chloroplast markers (Fig. 4). Conversely, the eastern European group displayed relatively high among-population differentiation at nuclear microsatellites, but very weak differentiation at chloroplast markers, as haplotype H1 largely prevails in that region (Fig. 4, compare also Figs. 1 and 2). The comparison of differentiation between the two groups did not show a specific pattern (results not shown); differentiation at chloroplast markers varied from slightly negative values to one, whereas differentiation at nuclear markers took low to intermediate values.

Discussion

Properties of chloroplast microsatellites

The haploid, non-recombinant and predominantly uniparentally inherited nature of chloroplast DNA makes it a useful tool for evolutionary studies (e.g. Olmstead & Palmer 1994, Provan *et al.* 2001). Some limitations may however occur, especially in phylogenetic applications, when using chloroplast microsatellites as genetic markers. Like their nuclear counterparts, chloroplast microsatellites are prone to homoplasy, which means that variants at cpSSR loci can be identical in state without being identical by descent (Provan *et al.* 2001, Goldstein & Pollock 1997). Consequently, phylogenetic distances between chloroplast haplotypes should be interpreted with caution, especially when using samples that are likely to be highly divergent, such as in broad-scale phylogeographic studies. In this latter case, Provan *et al.* (2001) recommended to preferentially use PCR-restriction fragment length polymorphism (PCR-RFLP) markers, which are based on nucleotide substitution polymorphisms, if these display enough polymorphism. In our study organism, PCR-RFLP markers show very little polymorphism (G.G. Vendramin & D. Salvini, personal communication), which is why we opted for cpSSRs. We can assume that the mutation rate in our loci is not very high, as we detected only little polymorphism (8 haplotypes in 323 individuals) and because these markers are highly conserved across angiosperm species (Weising & Gardner 1999). Nevertheless, homoplasy is likely, and it would be useful to test the relationships we found among cpSSR haplotypes by comparing them with PCR-RFLP markers, even if the latter provided only a low resolution. Because of the little overall number of mutations separating haplotypes (Fig. 3), we were unable to identify strongly divergent haplotype lineages. For this reason, we chose not to present our data according to groups of populations carrying similar haplotypes.

Geographic structure of chloroplast microsatellite variation

Chloroplast haplotype variation in common ash was highly structured in Europe ($G_{ST} = 0.88$). Similar levels of among-population differentiation were observed for instance in black alder, *Alnus glutinosa* ($G_{ST} = 0.87$, King & Ferris 1998), another mid-successional colonising species with wind-dispersed pollen and wind-disseminated seed, and in the more widely occurring European white oaks (*Quercus robur*, $G_{ST} = 0.78$; *Quercus petraea*, $G_{ST} = 0.86$, Petit *et al.* 2002a). In those species, a phylogeographic pattern of chloroplast DNA polymorphisms was observed, which means that phylogenetically close haplotypes were found together in the same population more often than randomly chosen different haplotypes. In common ash however, we did not observe phylogenetic structure (N_{ST} was not different from G_{ST}) and we attribute this to the overall low haplotypic variation, and especially to the low number of mutational steps separating haplotypes (Fig. 3). Other species that lack phylogenetic structure mostly display high within-population chloroplast haplotype diversity and low differentiation among populations, owing to very efficient bird-dispersal of fleshy fruits (e.g. $G_{ST} = 0.29$ in *Prunus avium*, Mohanty *et al.* 2001; $G_{ST} = 0.34$ in *Sorbus torminalis*, Oddou-Muratorio *et al.* 2001). Such a situation is clearly incompatible with common ash regarding its dispersal features and the low level of chloroplast DNA polymorphism within populations (Table 2).

A geographic organisation of chloroplast haplotype variation in common ash was observed (Fig. 1) and confirmed by a significant pattern of isolation by distance among populations. Isolation by distance at maternally inherited markers at a large geographical scale can be due to imprints of postglacial recolonisation with haplotypes from distinct refuges colonising distinct areas (e.g. Palmé & Vendramin 2002) potentially accompanied by loss of diversity due to recurrent founder events (e.g. Hewitt 2000), and/or to clinal variation in haplotype frequencies established for instance by postglacial gene flow and genetic drift (e.g. Liepelt *et al.* 2002). As in our study, haplotypes occurred in clearly delimited geographic areas, in good agreement with palynological data (Huntley & Birks 1983, Gliemeroth 1997, Brewer 2001), we suggest that postglacial recolonisation was the main cause of the isolation by distance pattern. Correlation analyses of genetic diversity measures with geographical coordinates were non-significant in all cases; hence, we were unable to identify losses of diversity due to founder events during postglacial recolonisation, which was most likely due to the low polymorphism of the chloroplast markers we employed. From the joint interpretation of palynological and genetic data, we propose that colonisation of continental Europe by common ash occurred from at least two distinct glacial refuge areas (*sensu lato*). Regarding the eastern European distribution of haplotype H1, it is most likely that it has survived the last glacial period in the Balkan Peninsula and thereafter colonised eastern central Europe northwards to Lithuania and Sweden. The central and western European distribution of

haplotypes H2, H3 and H5 suggests that they originate from one or, more probably, a few refuges in the Alps and/or Italy. In line with the now broadly accepted view that refuge populations were able to accumulate high diversity and/or unique haplotypes owing to their persistence and relative stability over glacial cycles (Hewitt 1996, Newton *et al.* 1999, Tzedakis *et al.* 2002), the occurrence of private haplotypes in populations from the Alps and the Balkan Peninsula in our dataset (H5 in population 11 in the Alps, and H13 and H14 in populations 20 and 30 from Bulgaria, respectively) provides additional support for the location of refuges in these areas. A suture zone between haplotypes from distinct refuges (H1 and H2, Fig. 1) occurs in Germany and Poland, explaining the relatively high values of phylogenetic diversity within populations, v_S, in this area (Table 3). The fact that we identified haplotype H4 exclusively in the British Isles evokes two scenarios: either a colonisation from another distinct refuge, possibly in the Iberian Peninsula, or, but less likely, fixation of a mutant haplotype during colonisation from, potentially, the Balkan Peninsula (H8 or H4, differing from the Eastern European haplotype H1 by only one or two steps, respectively, Fig. 3) and extinction of geographically intermediate populations. The occurrence of H8 in population 23 from Romania as well as in populations 1 and 3 from Ireland weakly supports the latter scenario. Yet, in Vendramin *et al.*'s (unpublished) chloroplast microsatellite survey at the European scale in common ash, haplotype H4 was also found in northern Spain, which confirms the existence of an Iberian refuge, as suggested from palynological data (Gliemeroth 1997, Brewer 2001).

Joint interpretation of chloroplast and nuclear patterns of genetic diversity

From the previous paragraph it appears that the four major chloroplast haplotypes detected in European populations of the common ash present a clear geographical structure (Fig. 1), which was most likely shaped during postglacial recolonisation. The geographic structure of genetic diversity at nuclear markers showed a completely different pattern. In western and central Europe, differentiation for cpSSRs was high, as all four major chloroplast haplotypes occurred there, but very little differentiation among populations was observed at nSSRs. Conversely, in eastern Europe, differentiation for chloroplast markers was very weak because of the predominance of haplotype H1, but most surprisingly, strong differentiation among populations was identified at nuclear markers (compare Figs. 1 and 2, Fig. 4).

Western and central Europe

In western and central Europe, we found a very homogeneous distribution of nuclear marker diversity with $F_{ST} = 0.027$ and a most likely number of gene pools of one, using the method of Pritchard *et al.* (2000), i.e. when the method was applied to populations from the western group only, the admixture of gene pools 1, 2 and 3 (Fig. 2) basically reduced to a single gene

pool (Heuertz *et al.* unpublished, chapter 3). Compared to southeastern Europe and Sweden, this region was also characterised by comparatively high gene diversity and high values of the inbreeding coefficient, and it showed slight imprints from recent population bottlenecks.

Little structure among populations at nuclear loci, despite potentially high divergence for chloroplast markers, as well as relatively high gene diversity, seem to be general observations in late-successional broadleaved tree species in western and central Europe. Sessile oak, which has recolonised continental Europe from the three southern peninsulas (Dumolin-Lapègue *et al.* 1997, Petit *et al.* 2002b) similarly to common ash, displayed low overall differentiation at nuclear allozyme loci among 85 mainly western European populations (G_{ST} = 0.025, Zanetto & Kremer 1995), and the highest gene diversity values were observed in northwestern Europe. Beech colonised most of Europe from one Carpathian refuge (Demesure *et al.* 1996) and differentiation for nuclear allozymes in western and central Europe was also low (F_{ST} = 0.04, Comps *et al.* 2001). Both of these studies showed a decrease of within-population allelic richness and a concomitant increase of gene diversity following the direction of postglacial recolonisation, which can be understood as a consequence of population bottlenecks during recolonisation. Hence there seem to be some congruent patterns in the recolonisation of western and central Europe by ash, oak and beech.

In the light of our results, we propose the following scenario for postglacial recolonisation of western and central Europe in common ash. (1) During glacial times, populations from different refuges were differentiated at chloroplast and at nuclear markers (e.g. Kremer *et al.* 2002). (2) As climatic conditions improved and suitable areas became available in the beginning of the Holocene (10000 BP), differentiated maternal lineages from different refuges started colonising northward. Pollen data suggest that colonisation from Spain started about 1000 years later than from refuges in Italy, the Alps and the Balkan area, which resulted in an east to west trend of migration (Huntley & Birks 1983, Gliemeroth 1997, Brewer 2001) (3) Pollen flow established contact between the colonising lineages, homogenising progressively the nuclear gene pool of migrating populations. The resulting nuclear gene pool featured high diversity, because it resulted from the admixture of differentiated gene pools, which had evolved in separate glacial refuges. As seed flow was much more limited in general, only little mixing of chloroplast haplotypes occurred. (4) Maternal lineages expanded and continued colonising northward and westward, where new land became available. Colonisation occurred through leading edge dispersal, i.e. successive but rare events of long-distance dispersal of seed followed by a more compact wave of advance (e.g. Hewitt 1996, Petit *et al.* 1997). Long-distance dispersal resulted in founder events (or population bottlenecks), of which we still detect some imprints in the nuclear genome today. Multiple founders carrying different alleles helped to maintain overall diversity in the colonising gene pool (Ibrahim *et al.* 1996, Le Corre & Kremer 1998). (5) Extensive gene flow through pollen between the core and the newly colonised populations moderated the consequences of founder

events in the nuclear gene pool and also contributed to maintain diversity by assuring a high effective population size.

Even though our previous results suggested restrictions to pollen flow (Heuertz *et al.* unpublished, chapter 2), we argue that wide-ranging gene flow through pollen was possible during colonisation. Colonising populations typically feature low densities, and therefore there is little competition against immigrating pollen, allowing for high levels of effective gene flow. Theoretical results of Austerlitz *et al.* (2000) also predicted stronger gene flow under colonisation than under stable demographic situations. Additional support for such a scenario comes from silver fir, for which Liepelt *et al.* (2002) revealed a range-wide cline for a paternally inherited chloroplast marker, demonstrating extensive postglacial pollen flow in a species with much larger pollen grains (Jackson & Lyford 1999) and a longer lifespan than ash. In oaks, the lack of an association between chloroplast and nuclear marker diversity was also mainly attributed to high levels of pollen flow during colonisation (Kremer *et al.* 2002).

Eastern Europe

In eastern Europe, the strong genetic structure we observed at nuclear loci in the common ash was rather unexpected, especially because chloroplast patterns of diversity did not allow to discern different glacial origins of populations in this area. The nuclear allozyme study in beech (Comps *et al.* 2001) showed some similarities with our results: among-population differentiation in the postulated refuge area in the Carpathians was higher than average, and the bottleneck statistic was overall low in southeastern Europe, as in our study. In oaks, for which at least two refuges were suggested in the Balkan Peninsula (Brewer *et al.* 2001), sampling efforts for genetic data were too limited in this area to permit any comparisons.

In common ash, fossil pollen data have identified one glacial refuge at the Bulgarian Black Sea coast, one in the Greek Pindus mountains and one in the Bulgarian Rhodopy mountains. It is likely that, during glacial times, populations in these refuges were differentiated at chloroplast markers and at nuclear markers (e.g. Kremer *et al.* 2002). In this context, the large prevalence of haplotype H1 in the Balkan Peninsula suggests that either (i) refuge populations were differentiated at chloroplast markers and postglacial recolonisation occurred from a single refuge, or (ii) refuge populations were not differentiated and haplotype H1 had been fixed in the Balkan area already before the last ice age (before 115000 BP).

The examination of diversity patterns at nuclear microsatellites within the Balkan Peninsula revealed high divergence among populations and some congruence according to geography, with higher levels of diversity in plain populations east of the Carpathians and Balkans than in populations within these mountains or west of them (Table 5), and a different gene pool composition of both groups of populations (Fig. 2). Three non exclusive hypotheses can be proposed to explain the observed patterns. (1) Gene pools might reflect differentiated lineages

of common ash, for example ecotypes adapted to mountainous or floodplain habitats, and each lineage would display characteristic diversity patterns. (2) Some gene pools might reflect hybridisation with other ash species, like *F. angustifolia* or *F. pallisiae*, occurring in the Balkan Peninsula (Tutin *et al.* 1972). As *F. excelsior* and *F. angustifolia* have been reported to hybridise in southeastern France (Rameau *et al.* 1989), hybridisation might account for some of the differentiation between *F. excelsior* populations, for instance between hybridising and non-hybridising ones. (3) Gene pools might reflect glacial origin. The strong differentiation and the very heterogeneous gene pool composition among populations in the Balkan Peninsula suggest that postglacial recolonisation most probably occurred from multiple refuges, in agreement with fossil pollen data (Huntley & Birks 1983, Gliemeroth 1997, Brewer 2001). Under this hypothesis, haplotype H1 must have been fixed in the Balkan area before the last glacial period. A rapid expansion in the early Holocene from several refuges in the southern Balkan Peninsula as suggested by Huntley and Birks (1983) is very compatible with our results and evokes the following steps. (1) As the climate improved, multiple populations differentiated for their nuclear gene pools expanded mostly at their edges. (2) Rare long-distance colonisation events would not have been very successful because of the proximity of the refuges and the rapidly achieved high density of ash in the area. (3) Similarly, because of high density, high competition would have taken place against incoming pollen, and therefore, among-population pollen flow would have rarely resulted in effective gene flow. (4) Additionally, the complex topography might have made gene flow even more difficult, keeping the gene pools differentiated. A scenario of expansion of multiple differentiated gene pools with little immigration is also in agreement with low values of the bottleneck statistic in the Balkan Peninsula.

Contrasting demographic features of refuge populations could contribute to explain the higher levels of within-population diversity in the Danube and Dniestr plains than in mountainous areas. For instance, floodplain refuge populations near the Black Sea, which probably provided the colonisers to the Danube and Dniestr plains, could have remained relatively large throughout cold periods, conserving high levels of diversity, whereas mountainous populations could have been smaller and more scattered, and therefore subjected to stronger genetic drift. Moisture availability was a key factor for the persistence of refuge populations in the arid glacial periods, and for this reason, many refuges were located in mid-altitude sites (Brewer *et al.* 2002 and references therein). In sheltered floodplains, like in the Kamcija plain near Varna at the Bulgaria Black Sea coast, we suggest that sufficient moisture was available to sustain relatively large refuge populations. A larger pollen representation for ash near the Black Sea than in the mountainous area 12000 years ago (Brewer 2001) provides additional support to this matter. The slightly larger differentiation values in the western part of the Balkan Peninsula could also reflect a stronger fragmentation of populations in mountain refuges. Further, hybridisation is probably a stronger issue for floodplain than for mountain

populations, as *F. angustifolia* and *F. pallisiae* preferentially occur in floodplain habitats (Tutin *et al.* 1972). Therefore, hybridisation might also contribute to explain the higher levels of diversity found in *F. excelsior* populations from the Danube and Dniestr floodplains compared to those from the mountains, and/or the strong divergence of populations 23-25 from populations 30-36 in that area.

Eastern European contact zone

A zone of admixture of haplotype H1 of eastern European origin and haplotype H2 from the Alps was observed in Germany and Poland. This confirms an early northeastward colonisation form an Alpine refuge, joining the westward expansion observed in Poland as suggested from fossil pollen data (around 8000 BP, Huntley & Birks 1983). Whether haplotype H1 in the contact zone originates from the southern Balkan Peninsula, or possibly from a northern Carpathian refuge in southeastern Poland (Huntley & Birks 1983, Brewer 2001) is yet unclear. In case such a refuge contributed to postglacial recolonisation, it is most likely that it contained haplotype H1, providing another argument for an ancient fixation of this haplotype in eastern Europe. At the level of the nuclear gene pool, the strong admixture revealed in populations 18 and 19 from Slovakia and Hungary by the method of Pritchard *et al.* (2000) (Fig. 2) suggests contact of central and eastern European gene pools. The advance of a central European maternal lineage up to northern Poland can also partially explain the high levels of admixture discovered in Lithuania (Fig. 2).

A particular diversity pattern in Sweden

Populations from Sweden displayed diversity patterns at nuclear loci that were strikingly different from those we observed in other populations from western and central Europe. In Sweden, within-population gene diversity and allelic richness were low, the bottleneck statistic was negative, indicating population expansion without immigration, and populations were in Hardy-Weinberg equilibrium and strongly differentiated from each other (Heuertz *et al.* unpublished, chapter 3). These features are very similar to those we observed in southeastern European populations. Interestingly, in oak, populations from Denmark and Norway also showed higher allelic richness and lower gene diversity compared to western European populations (Zanetto & Kremer 1995); and in beech, populations from southern Sweden were more differentiated (F_{ST}) and displayed lower values of the inbreeding coefficient (F_I) than populations from Germany or France (Comps *et al.* 2001). Comps *et al.* (2001) attributed this to quicker dynamics of the fixation indices compared to gene diversity measures in recently colonised populations. However, these populations also showed negative bottleneck statistics, indicating population expansion, and higher than average total allelic richness (Comps 2001). Populations from the British Isles could be interesting to confirm

such diversity patterns as they should also have been recolonised most recently, but such populations were not analysed in beech. In oaks, diversity patterns in the British Isles were similar to those in western France, and did not show the particularities of Danish and Norwegian populations.

These unusual diversity patterns in southern Scandinavian populations of ash, but also of oak and beech brings about the question whether glacial refuges for ash may have existed in northern central Europe. Palynological data indeed identified proportions of ash pollen above one percent in southern Sweden as early as 12000 BP (Brewer 2001), which is an indication for local presence of ash according to Huntley and Birks (1983). The inference of glacial refuges on the basis of pollen data alone is however increasingly questioned (e.g. Brewer *et al.* 2002), as pollen could have been transported over long distances. In this context, the importance of macrofossils is more and more stressed, as they represent the only method to provide unequivocal evidence for the presence of a taxon in a particular site (e.g. Birks & Birks 2000). Today, increasing evidence in many thermophilous plant and animal species suggests that southern European glacial refuges were supplemented by small "cryptic" refuges in northern Europe, usually located in areas of sheltered topography (Stewart & Lister 2001). In the particular case of Sweden, macrofossils of boreal tree species like Norway spruce, Scots pine or birch discovered in the Scandes mountains of central Sweden have been dated to the late glacial period (the Younger Dryas, 11000 – 10000 BP), suggesting colonisation of early deglaciated mountain peaks (nunataks) before complete deglaciation of the massif by 9200 BP (Kullman 2001). It was suggested that these species immigrated into the Scandes during the Younger Dryas cold period from a coastal refuge in the west, where they had persisted as particularly cold-adapted ecotypes (krummholz, i.e. vegetative shrub-like growth forms) (Kullmann 2001). For ash, colonisation of Sweden from the west is not very likely not only because krummholz variants are not known, but also as our populations carried the eastern haplotype H1. Early colonisation from the east is also improbable because the area was glaciated (Kullman 2001). In the Swedish Scandes, macrofossils indicate that thermophilous trees such as oaks, elms, black alder, hazel and lime tree had appeared as early as between 8500 and 8000 BP, suggesting that colonisation probably occurred from small stands close to the ice sheets rather than from the southern refuges (Stewart & Lister 2001). During the late glacial (around 12000 BP), the northern limit of pine forest was in northern Germany and southern Sweden, as inferred from palynological data (Stewart & Lister 2001). Further, Willis *et al.* (2000) provided evidence that thermophilous trees survived in microenvironmentally favourable pockets within refuges for more cold-adapted species in Hungary. Hence, a possible explanation for an early occurrence of ash in southern Sweden could be the existence of cryptic glacial refuges for ash in this pine forest, close to the Fennoscandian ice sheet. This hypothesis is not at odds with the Swedish populations carrying

haplotype H1, if the latter had been fixed before glacial times, regarding that H1 is found all over eastern Europe.

In ash, the nuclear gene pool 2 inferred with the method of Pritchard (2000) (Fig. 2) is dominant in a population from southern Sweden only, and it is present in admixture in all western and central European populations. Considering this together with the particular diversity patterns identified in oak and beech, the existence of cryptic refuges for thermophilous tree species in northern central Europe and their contribution to colonisation of western and central Europe might deserve more attention in the future.

Concluding remarks

Our study showed that the combined analysis of chloroplast and nuclear markers provided interesting insights into postglacial history of common ash. Nuclear microsatellites unravelled diversity patterns at high resolution, largely owing to an efficient model-based clustering method, suggesting that they might constitute more valuable tools for the investigation of population history than previously thought. High differentiation and specific gene pool compositions were detected among populations from southeastern Europe and from Sweden with nuclear markers, whereas chloroplast markers revealed only very little divergence. These results identified southeastern Europe and Sweden as important centres for the conservation of genetic resources in ash. This study demonstrated that the investigation of population history with only one genetic marker approach may lead to overlook essential diversity patterns in a taxon; for example nuclear markers did not discriminate the different maternal lineages in western and central Europe and chloroplast markers ignored strong differentiation for nuclear markers in the Balkan Peninsula. Therefore, we suggest that strategies for the management and conservation of genetic resources should be based on genetic markers with different modes of inheritance, especially when maternal markers display low polymorphism.

References

Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics*, **154**, 1309-1321.

Bingelli P, Power J (1999) http://members.tripod.co.uk/WoodyPlantEcology/species/ash.htm

Birky CW (1988) Evolution and variation in plant chloroplast and mitochondrial genomes. In: *Plant evolutionary biology* (eds. Gottlieb LD, Jain SK), p. 23-55 Chapman and Hall, London.

Birky CW Jr., Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, **103**, 513-527.

Birks HH, Birks HJB (2000) Future uses of pollen analysis must include plant macrofossils. *Journal of Biogeography*, **27**, 31-35.

Breitenbach-Dorfer, Konnert M, Pinsker W, Starlinger F, Geburek T (1997) The contact zone between two migration routes of silver fir, *Abies alba* (*Pinaceae*), revealed by allozyme studies. *Plant Systematics and Evolution*, **206**, 259-272.

Brewer S (2001) Fossil pollen maps. Available form http://www.pierroton.inra.fr/cytofor

Brewer S, Cheddadi R, de Beaulieu JL *et al.* (2002) The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecology and Management*, **156**, 27-48.

Bucci G, Anzidei M, Madaghiele A, Vendramin GG (1998) Detection of haplotypic variation and natural hybridisation in *halepensis*-complex pine species using chloroplast simple sequence repeat (SSR) markers. *Molecular Ecology*, **7**, 1633-1643.

Burban C, Petit RJ, Carcreff E, Jactel H (1999) Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. *Molecular Ecology*, **8**, 1593-1602.

Clegg MT (1989) Molecular diversity in plant populations. In: *Plant population genetics, breeding and genetic resources* (eds. Brown ADH, Clegg MT, Kahler AL, Weir BS), p. 98-115. Sinauer Inc. Sunderland, MA.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657-1659.

Comes HP, Kadereit JW (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432-438.

Comps B, Gömöry D, Letouzey J, Thiébaut B, Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics*, **157**, 389-397.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001-2014.

Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515-2520.

Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475-1487.

Dumolin-Lapègue S, Kremer A, Petit R (1999) Are chloroplast and mitochondrial DNA variation species independent in oaks? *Evolution*, **53**, 1406-1413.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* (L.) Skeels) endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832-839.

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250-259.

Ferris C, King RA, Väinölä R, Hewitt GM (1998) Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. *Heredity*, **80**, 584-593.

Gliemeroth AK (1997) Paläoökologische Aspekte der Einwanderungsgeschichte einiger Baumgattungen während des Holozäns nach Europa. *Angewandte Botanik*, **71**, 54-61.

Goldstein DG, Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity*, **88**, 335-342.

Goudet J (1995) FSTAT (vers.1.2), a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile compute program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618.

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution*, **16**, 700-706.

Hewitt G (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Huntley B, Birks HJB (1983) An atlas of past and present pollen maps for Europe: 0-13000 Years Ago, pp 209-223. Cambridge University Press, Cambridge, UK.

Huang SSF, Hwang SY, Lin TP (2002) Spatial patterns of chloroplast DNA variation of *Cyclobalanopsis glauca* in Taiwan and East Asia. *Molecular Ecology*, **11**, 2349-2358.

Ibrahim KL, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, **77**, 282-291.

Jackson ST, Lyford ME (1999) Pollen dispersal models in quaternary plant ecology: assumptions, parameters and prescriptions. *The Botanical Review*, **65**, 39-75.

King RA, Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology*, **7**, 1151-1161.

Kremer A, Kleinschmit J, Cottrell J *et al.* (2002) Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? *Forest Ecology and Management*, **156**, 75-85.

Kullman L (2001) A new approach to postglacial forest history of northern Scandinavia. Review of megafossil and macrofossil evidence. *Recent Research Development Ecology*, **1**, 1-19.

Lagercrantz U, Ryman N (1990) Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution*, **44**, 38-53.

Latta RG, Mitton JB (1997) A comparison of population differentiation across four classes of gene markers in Limber pine (*Pinus flexilis* James). *Genetics*, **146**, 1153-1163.

Le Corre V, Kremer A (1998) Cumulative effects of founding events during colonization on genetic diversity and differentiation in an island and stepping-stone model. *Journal of Evolutionary Biology*, **11**, 495-512.

Le Corre V, Dumolin-Lapègue S, Kremer A (1997) Genetic variation at allozyme and RAPD loci in sessile oak *Quercus petraea* (Matt.) Liebl.: the role of history and geography. *Molecular Ecology*, **6**, 519-529.

Legendre P (2001) CADM, available from http://www.fas.umontreal.ca/biol/casgrain/en/labo/cadm.html

Liepelt S, Bialozyt R, Ziegenhagen B (2002) Wind-dispersed pollen mediates postglacial geneflow among refugia. *Proceedings of the National Academy of Sciences of the USA*, **99**, 14590-14594.

Machon N, Lefranc M, Bilger I, Mazer SJ, Sarr A (1997) Allozyme variation in *Ulmus* species from France: analysis of differentiation. *Heredity*, **78**, 12-20.

Marchelli P, Gallo LA (2001) Genetic diversity and differentiation in a southern beech subjected to introgressive hybridization. *Heredity*, **87**, 284-293.

Mohanty A, Martín JP, Aguinagalde I (2001) A population genetic analysis of chloroplast DNA in wild populations of *Prunus avium* L. in Europe. *Heredity*, **87**, 421-427.

Murillo O, Finkeldey R (2000) Genetic diversity in natural populations of *Alnus acuminata* ssp. *Arguta* (Schlechtendal) Furlow in Costa Rica and Panama. *Forest Genetics*, **7**, 121-131.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1-10.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology and Evolution*, **14**, 140-145.

Oddou-Muratorio S, Petit RJ, Le Guerroue B, Guesnet D, Demesure B (2001) Pollen- versus seed-mediated gene flow in a scattered forest tree species. *Evolution*, **55**, 1123-1135.

Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany*, **81**, 1205-1224.

Palmé AE, Vendramin GG (2002) Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, *Corylus avellana*. *Molecular Ecology*, **11**, 1769-1780.

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997) Chloroplast footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences of the USA*, **94**, 9996-10001.

Petit RJ, Csaikl UM, Bordács S *et al.* (2002a) Chloroplast DNA variation in European white oaks: Phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156**, 5-26.

Petit RJ, Brewer S, Bordács S *et al.* (2002b) 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.

Picard JF (1982) Contribution à l'étude de la biologie florale et de la fructification du Frêne commun (*Fraxinus excelsior* L). *Revue Forestière Française*, XXXIV, 97-107.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144**, 1237-1245.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Provan J, Powll W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology and Evolution*, **16**, 142-147.

Rameau JC, Mansion D, Dumé G (1989) Flore forestière française, guide écologique illustré. 1. Plaines et collines. Institut pour le développement forestier, Paris, France.

Raspé O, Saumitou-Laprade P, Cuguen J, Jacquemart AL (2000) Chloroplast DNA haplotype variation and population differentiation in *Sorbus aucuparia* L. (Rosaceae: Maloideae). *Molecular Ecology*, **9**, 1113-1122.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223-225.

Schneider S, Roessli D, Excoffier L (2000) Arlequin v. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sperisen C, Büchler U, Mátyás G, Ackzell L (1999) Mitochondrial DNA variation provides a tool for identifying introduced provenances: a case study in Norway spruce. In: *Which marker for which purpose?* European Union DGXII Biotechnology FW IV Research programme *Molecular tools for biodiversity* (ed. Gillet EM) http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution*, **16**, 608-613.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453-464.

Terry RG, Nowak RS, Tausch RJ (2000) Genetic variation in chloroplast and nuclear ribosomal DNA in Utah juniper (*Juniperus osteosperma*, Cupressaceae): evidence for interspecific gene flow. *American Journal of Botany*, **87**, 250-258.

Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA, eds (1972) *Flora Europaea*, volume 3. Cambridge University Press, UK.

Tzedakis PC, Lawson IT, Frogley MR, Hewitt GM, Preece RC (2002) Buffered tree population changes in a Quaternary refugium: evolutionary implication. *Science*, **297**, 2044-2047.

Wallander E (2001) Evolution of wind-pollination in *Fraxinus* (Oleaceae) – an ecophylogenetic appraoch. PhD thesis. Botanical Institute, Göteborg University, Sweden.

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chlorolplast genomes of dicotyledonous angiosperms. *Genome*, **42**, 9-19.

Zanetto A, Kremer A (1995) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. I. Monolocus patterns of variation. *Heredity*, **75**, 506-517.

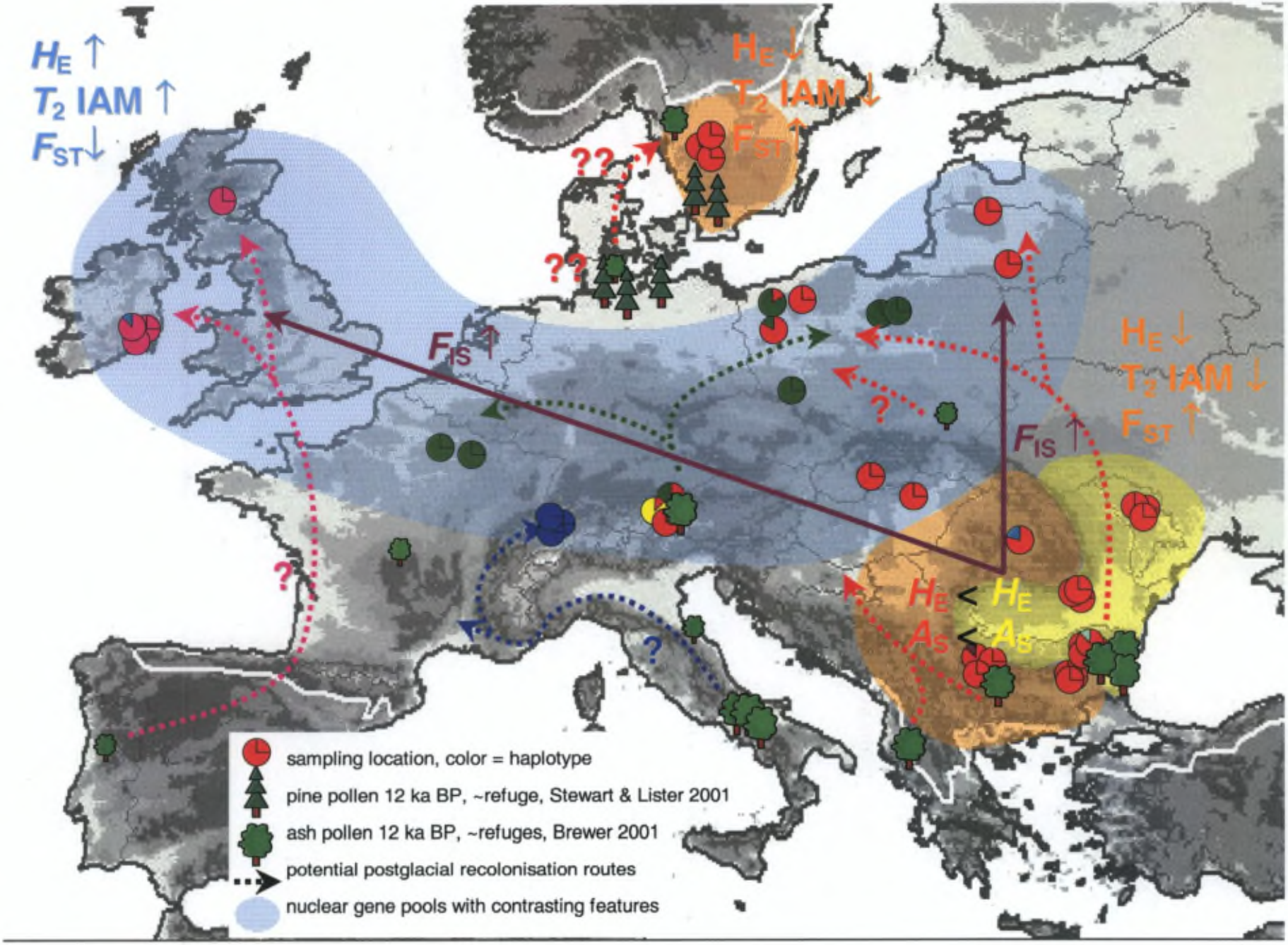


Fig. 1 Synthetic map of the genetic diversity patterns of common ash analysed in this thesis. Within population diversity statistics at nuclear microsatellite loci: A_S, allelic richness; H_E, genetic diversity; T_2 IAM, bottleneck statistic; F_{IS}, inbreeding coefficient. Statistic of among-population differentiation at nuclear microsatellites: F_{ST}. The distribution limits of common ash are indicated as white lines.

General discussion

A synthetic view of the results of this work is given in Fig. 1. Chloroplast microsatellite analysis allowed us to corroborate glacial refuges for ash situated in the Alps and and the Balkan Peninsula, as suggested from fossil pollen data (Huntley & Birks 1983, Gliemeroth 1997, Brewer 2001), and our results are in agreement with other proposed refuges in Italy and the Iberian Peninsula. Putative postglacial recolonisation routes are indicated. Nuclear microsatellite analysis revealed distinct diversity patterns, and notably contrasting situations in western and central Europe on the one hand, and in southeastern Europe and Sweden on the other. In western and central Europe, slight imprints from bottlenecks, high genetic diversity and a homogeneous gene pool over the entire area were interpreted as the result of extensive pollen flow, homogenising the nuclear gene pools of colonisers from differentiated refuges during the process of postglacial recolonisation. In southeastern Europe, high among-population differentiation as well as relatively low levels of genetic diversity and no signs from bottlenecks were attributed to the expansion from different glacial refuges with little among-population gene flow, hypothesising that a demographically stable condition with high density of ash was rapidly reached in the area. In Sweden, diversity patterns were similar to southeastern Europe, which lead us to suspect the existence of cryptic full glacial refuges in northern central Europe following Stewart and Lister (2001).

Recommendations for management and conservation of genetic resources in common ash

The central aim in forest genetic resources conservation is to safeguard a species' adaptive potential (Eriksson *et al.*. 1992, Namkoong 1998). As the real adaptive potential of populations is generally not known, one possibility is to concentrate on existing patterns of adaptation. This is the scope of the multiple populations breeding system (MPBS) recommended as a conservation strategy for noble hardwoods (Eriksson 2001) by the EUFORGEN Noble Hardwoods network. The concept of MPBS is to enhance among-population differentiation by promoting natural selection within populations with the aim to increase the overall adaptive diversity. There is a broad agreement among conservation geneticists that, together with adaptive divergence, historical isolation is an important concept for the delimitation of significant units for conservation (reviewed by Fraser & Bernatchez 2001). This is because historically isolated populations have most certainly accumulated different mutations that prepare them differentially for future challenges, even if these

populations are not readily distinguishable by obvious adaptive differences under present conditions.

In order to capture actual patterns of adaptive variation and enhance them, the EUFORGEN technical guidelines (unpublished) for common ash (see also Pliura 1999) foresee that in each country, eco-geographic or provenance regions should be delimited according to climatic variation, topography, soil and vegetation. Then, two complementary networks of gene conservation populations should be established: (i) a network of 20-30 *in situ* populations selected throughout provenance regions, and (ii) a network of *ex situ* populations (progeny trials, provenance trials, collections). When possible, *in situ* conservation activities should be jointly performed for other noble hardwoods.

If common ash occurs in large populations within a country, *in situ* conservation is sufficient according to EUFORGEN technical guidelines (unpublished), and 1-3 stands of sizes 5-15 ha (or at least 100 flowering trees) should be selected in each region of provenance. In these stands, promoting natural regeneration of the target species, monitoring health condition and protection against damage are necessary. Management should aim at creating multi-age structure and rapid generation turnover. For populations that are marginal, isolated, endangered, growing under special ecological conditions or carrying rare features, *in situ* conservation should be complemented by *ex situ* activities, which is most effective in the form of progeny trials, as it permits joint gene conservation and breeding.

Some additional recommendations for conservation of common ash genetic resources can be formulated on the basis of the molecular marker analyses in this thesis.

The high differentiation among populations detected with nuclear markers at the level of southeastern European countries most probably has a historical origin, regarding the proximity of different glacial refuges inferred from fossil pollen data. Hence, we suggest that southeastern European common ash populations contain differentiated genetic resources, which prepare them well for future uncertainties. In order to value the distinctiveness of genetic resources for common ash in this region, we suggest that a higher density of *in situ* gene conservation populations should be established in southeastern Europe, especially in Bulgaria and Romania, which were probably colonised from different glacial refuges.

Similarly, in Sweden, genetic resources in ash are highly differentiatied, although the historical origin of this differentiation still has to be verified. Additionally, the adaptation to photoperiod was experimentally confirmed (Baliuckas 2000) and signs for population bottlenecks were not identified neither in that study, nor in the present one. We therefore also recommend particular conservation efforts in northern central Europe.

In western and central Europe, as well as Lithuania, our study revealed an important homogeneity of the nuclear gene pool and no particular patterns of divergence despite the presence of different maternal lineages. Hence in these regions, it is probably better to look at

present patterns of adaptation according to ecological zones, and to identify particular features. For instance, an important dieback of young ash populations due to a yet unidentified pest is actually reported in Poland and Lithuania (J. Nowakowska and A. Pliura, personal communication), and resistant populations should benefit from complementary *ex situ* conservation. Despite little differentiation at nuclear markers in our study, we identified gradients of increasing levels of inbreeding from southeastern Europe towards the west and the north, with the exception of Sweden. If this pattern originated from increased selfing ability (which we stated as one possible explanation, Heuertz *et al.* unpublished, chapter 3), it could be meaningful in the context of adaptation, at least during colonisation, where long-distance colonisers with a potential for self-fertilisation would have had a selective advantage over colonisers lacking such a mechanism. In order to capture such large-scale trends of potentially adaptive variation, we propose that in large geographical regions with similar ecological features, several geographically distant populations should be chosen for *in situ* gene conservation.

Further, it is also important to conserve populations close to glacial refuges or even the refuge populations themselves because of their significant role in the evolutionary history of the species in the past and potentially also in the future (e.g. Tzedakis *et al.* 2002); for instance we identified private chloroplast DNA haplotypes in populations close to refuges indicating peculiarities in their genetic composition.

Perspectives

Several perspectives for future research can be identified in the light of the results of this thesis.

A major subject with potentially serious financial implications is certainly the hybridisation between common ash and narrow-leaved ash. Narrow-leaved ash (*Fraxinus angustifolia*) is a termophilous tree species mainly occurring in southern Europe (Tutin *et al.* 1972). It is a smaller tree (up to 25 m high) than *F. excelsior* (up to 40 m), and its wood is of lower quality (Picard 1983). Common ash can be distinguished from narrow-leaved ash by its black buds and its leaves which show more serrations than lateral veins; in narrow-leaved ash, buds are brown and the leaves show as many serrations as lateral veins (Tutin *et al.* 1972). These two ash species have been reported to commonly hybridise in southeastern France (Rameau *et al.* 1989) and hybrid individuals were also found in nurseries among seedlings of *F. excelsior* (Picard 1983). We observed many ash populations featuring a continuum from light brown to black buds in Romania, in the floodplain of the river Mures, west of the Carpathians, and close to the rivers Olt and Jiu east of the Carpathians. Because we were not sure of the taxonomic status, these samples were not included in our survey. The development of

molecular markers for the identification of hybrids is strongly needed in forestry, and some studies have already focused on this theme using RAPD (Jeandroz *et al.* 1996) or chloroplast microsatellite markers (Morand-Prieur *et al.* 2002). Our results (chapter 4) suggest that progress in this field could potentially be made through the sequencing of nuclear microsatellite loci.

In our nuclear microsatellite diversity survey of common ash at the European scale (chapters 3 and 5), we identified southeastern Europe as a region harbouring populations with high conservation value. In oaks, the location of glacial refuges suggests similar diversity patterns in southeastern Europe, which still remain to be investigated with nuclear markers. Further, we found peculiar diversity patterns in Sweden, such as no traces from population bottlenecks and high among-differentiation, which we tentatively explained by the putative existence of cryptic full glacial refuges in northern central Europe. The comparison of our results with those from other broadleaved species, like oak (Zanetto & Kremer 1995) and beech (Comps *et al.* 2001), revealed similar patterns of genetic diversity in the same region. Hence, we suggest that future research should aim at clarifying why these diversity patterns are dissimilar from those in much of western and central Europe, and whether this is indeed a general observation for broadleaved European forest trees, mainly in the context of the conservation of forest genetic resources.

Further, our study did not thoroughly reveal the different sources that contributed to the diversity of the homogeneous nuclear gene pool, which is now widely spread over western and central Europe. Some candidate regions that would potentially bring more clarity to this matter were not included in this survey. These regions include northern Spain, Italy and southern France. It would also be interesting to look at nuclear diversity patterns within Italy, especially because two putative glacial refuges were identified within the Apennine Peninsula and to understand whether the Alpine refuge contributed to (southward) postglacial recolonisation of Italy.

A technical problem we encountered during this work was the difficulty of correctly scoring the hypervariable microsatellite loci, especially when incorporating many populations over a large geographic area like Europe (chapter 3). Our observed diversity patterns are concordant, which is why we think that we made few scoring mistakes, and that we revealed the true organisation of neutral genetic variation at nuclear markers. If nuclear genetic diversity patterns of ash have to be studied at a large scale within a research project involving many laboratories, the reproducibility of scoring among laboratories could however be a serious issue (as it is with microsatellite markers in oaks, Mariette *et al.* 2002). Therefore, it would be interesting to develop a set of less polymorphic nuclear markers. This would also allow the verification of the diversity patterns obtained in our study.

At the within-population level, we obtained some estimates for the variance of pollen and seed dispersal distances from spatial genetic structure (chapter 2). Our estimates suggested a much lower seed flow in our Romanian population than preliminary results in a French population where a parentage study is in progress (Morand-Prieur *et al.*, in preparation). It would be interesting to verify our results by a direct method, similarly to Morand-Prieur *et al.*, in order to corroborate or invalidate our estimates from spatial genetic structure. Several local-scale studies in selected populations from all over Europe might also clarify the issue of increasing values of the inbreeding coefficient from southeastern Europe towards the west and towards the north (with the exception of Sweden).

References

Allendorf FW, Ryman N (2002) The role of genetics in population viability analysis. In: *Population viability analysis* (eds. Beissinger SR, McCullough DR) p. 50-85. The University of Chicago Press, Chicago, USA.

Arbez M (2000) Implementation of the Strasbourg Resolution S2 on the conservation of forest genetic resources in Europe. In: *International collaboration on forest genetic resources: the role of Europe. Proceedings of the Second EUFORGEN Steering Commitee meeting, 26-29 November 1998, Vienna, Austria.* (eds. Turok J, Geburek T), pp. 9-13. International Plant Genetic Resources Institute, Rome, Italy.

Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B (2000) Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics*, **154**, 1309-1321.

Avise JC, Arnold J, Ball RM Jr, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489-522.

Bacilieri R, Labbé T, Kremer A (1994) Intraspecific genetic structure in a mixed population of *Quercus petraea* (Matt.) Liebl. and *Q. robur* L. *Heredity*, **73**, 130-141.

Baliuckas V, Lagerström T, Eriksson G (2000) Within-population variation in juvenile growth rythm and growth in *Fraxinus excelsior* L. and *Prunus avium* L. *Forest Genetics*, **7**, 183-192.

Besnard G, Khadari B, Baradat P, Bervillé A (2002) *Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism. *Theoretical and Applied Genetics*, **104**, 1353-1361.

Binggeli P, Power JJ (1999) *Gender variation in ash.* Available from http://members.tripod.co.uk/WoodyPlantEcology/species/ash.htm

Birky CW Jr., Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, **103**, 513-527.

Birky CW (1988) Evolution and variation in plant chloroplast and mitochondrial genomes. In: *Plant evolutionary biology* (eds. Gottlieb LD, Jain SK), pp. 23-55. Chapman and Hall, London, UK.

Brewer S (2001) *Fossil pollen maps*. Available form http://www.pierroton.inra.fr/cytofor

Brewer S, Cheddadi R, de Beaulieu JL *et al.* (2002) The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecology and Management*, **156**, 27-48.

Bucci G, Anzidei M, Madaghiele A, Vendramin GG (1998) Detection of haplotypiv variation and natural hybridization in *halepensis*-complex pine species using chloroplast simple sequence repeat (SSR) markers. *Molecular Ecology*, **7**, 1633-1643.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR (1993) Incidence and origin of 'null' alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics*, **52**, 922-927.

Claessens H, Thibaut A, Lecomte H, Delecourt R, Rondeux J, Thill A (1994) *Le frêne en Condroz – stations et productivité potentielles.* Faculté des Sciences Agronomiques de Gembloux, Institut pour l'Encouragement de la Recherche Scientifique dans l'Industrie et l'Agriculture, Bruxelles, Belgium.

Clegg MT (1989) Molecular diversity in plant populations. In: *Plant population genetics, breeding and genetic resources* (eds. Brown ADH, Clegg MT, Kahler AL, Weir BS), pp. 98-115. Sinauer Inc. Sunderland, MA, USA.

Clegg MT, Gaut BS, Learn GH Jr., Morton BR (1994) Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences of the USA*, **91**, 6795-1801.

Comes HP, Kadereit JW (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432-438.

Comps B, Gömöry D, Letouzey J, Thiébaut B, Petit RJ (2001). Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics*, **157**, 389-397.

Degen B, Streiff R, Ziegenhagen B (1999) Comparative study of genetic variation and differentiation of two pedunculate oak (*Quercus robur*) stands using microsatellite and allozyme loci. *Heredity*, **83**, 597-603.

Demesure B, Comps B, Petit R (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515-2520.

Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475-1487.

Echt CS, DeVerno LL, Anzidei M, Vendramin GG (1998) Chloroplast microsatelites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Molecular Ecology*, **7**, 307-316.

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250-259.

Eriksson G, Namkoong G, Roberds JH (1993) Dynamic gene conservation for uncertain futures. *Forest Ecology and Management*, **62**, 15-37.

Eriksson G (1998) Sampling for genetic resources populations in the absence of genetic knowledge. In: *Noble Hardwoods Network. Report of the second meeting, 22-25 March 1997, Lourizán, Spain* (compilers Turok J, Collin E, Demesure B, Eriksson G, Kleinschmit, Rusanen M, Stephan R), pp. 61-75. International Plant Genetic Resources Institute, Rome, Italy.

Eriksson G (1999) Global warming and gene conservation of noble hardwoods. In: *Noble Hardwoods Network. Report of the third meeting, 13-16 June 1998, Sagadi, Estonia* (compilers Turok J, Jensen J, Palmberg-Lerche C, Rusanen M, Russell K, de Vries S, Lipman E), pp. 98-112. International Plant Genetic Resources Institute, Rome, Italy.

Eriksson G (2001) Conservation of noble hardwoods in Europe. *Canadian Journal of Forest Research* **31**, 577-587.

Frankel OH, Brown ADH, Burdon JJ (1995) *The conservation of plant biodiversity.* Cambridge University Press, Cambridge, UK.

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to conservation genetics.* Cambridge University Press, Cambridge, UK.

Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, **10**, 2741-2752.

Geburek T, Müller F (2000) Genetic aspects of the Helsinki and Lisbon resolutions – Second and third ministerial conferences on the protection of forests in Europe. In: International collaboration on forest genetic resources: the role of Europe. Proceedings of the Second EUFORGEN Steering Commitee meeting, 26-29 November 1998, Vienna, Austria (eds. Turok J and Geburek T), pp. 14-24. International Plant Genetic Resources Institute, Rome, Italy.

Giertych M (1995) Outline of Ecology – Summary. In*: Nasze Drzewa Lesne: Monografie Popularnonaukowe. Tom 17: Jesion wyniosly – Fraxinus excelsior L. [Our forest trees: popular monograph. Volume 17: European ash – Fraxinus excelsior L.]* (ed. Bugala W), pp. 302-304. Institute of Dendrology, Poznan, Poland.

Gliemeroth AK (1997) Paläoökologische Aspekte der Einwanderungsgeschichte einiger Baumgattungen während des Holozäns nach Europa. *Angewandte Botanik*, **71**, 54-61.

Goudet J, Raymond M, de Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933-1940.

Grivet D, Petit RJ (2002) Phylogeography of the common ivy (Hedera sp.) in Europe: genetic differentiation through space and time. *Molecular Ecology*, **11**, 1351-1362.

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: *Plant population genetics, breeding, and genetic ressources* (eds. Brown ADH, Clegg MT, Kahler AL, Weir BS), pp. 43-63. Sinauer Associates, Inc., Sunderland, MA, USA.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95-124.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London B*, **351**, 1291-1298.

Hancock JM (1999) Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: *Microsatellites: Evolution and Applications* (eds. Goldstein DB, Schlötterer C), pp. 1-9. Oxford University Press, Oxford, UK.

Hayashi E, Ubuka M, Iizuka K and Itahana N (2000) Genetic differentiation of organelle DNA polymorphisms in Saghalin fir from Hokkaido. *Forest Genetics*, **7**, 31-38.

Hedrick PW (2001) Conservation genetics: where are we now? *Trends in Ecology and Evolution*, **16**, 629-636.

Hewitt G (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Huntley B, Birks HJB (1983) An atlas of past and present pollen maps for Europe: 0-13000 Years Ago. Cambridge University Press, Cambridge, UK.

Ibrahim KL, Nochols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, **77**, 282-291.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424-429.

Jeandroz S, Frascaria-Lacoste N, Bousquet J (1996) Molecular recognition of the closely related *F. excelsior* and *F. oxyphylla* (Oleaceae) by RAPD markers. *Forest Genetics*, 3, 237-242.

Karhu A, Hurme P, Karjalainen M, Karvonen P, Kärkkäinen K, Neale D, Savolainen O (1996) Do molecular markers reflect patterns of differentiation of adaptive traits in conifers? *Theoretical and Applied Genetics*, **93**, 215-221.

Kanowski P, Boshier D (1997) Conserving the genetic resources of trees *in situ*. In: *Plant Genetic Conservation* (eds. Maxted N, Ford-Lloyd BV, Hawkes JG), pp. 207-219. Chapman & Hall, London, UK.

Kirkpatrick M, Jarne P (2000) The effects of a bottleneck on inbreeding depression and the genetic load. *The American Naturalist*, **155**, 154-167.

King RA, Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology*, **7**, 1151-1161.

Kleinschmit J, Svolba J, Enescu V, Franke A, Rau HM, Ruetz W (1996) Erste Ergebnisse des Eschen-Herkunftsversuches von 1982. *Fortarchiv*, **67**, 114-122.

Klimo E, Paule L, Longauer R (1999) Forest biodiversity and ecological networks. In: *Conservation and sustainable management of forests in central and eastern European countries*, pp. 35-48. Phare Environment Consortium, European Commission Phare Programme, Brussels, Belgium

Konnert M, Bergmann F (1995) The geographical distribution of genetic variation in silver fir (*Abies alba*, Pinaceae) in relation to its migration history. *Plant Systematics and Evolution*, **196**, 19-30.

Kremer A (1994) Diversité génétique et variabilité des caractères phénotypiques chez les arbres forestiers. *Genetics Selection Evolution*, **26** (Suppl.1), 105-123.

Kremer A, Zanetto A, Ducousso A (1997) Multilocus and multitrait measures of differentiation for gene markers and phenotypic traits. *Genetics*, **145**, 1229-1241.

Kremer A, Kleinschmit J, Cottrell J *et al.* (2002) Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? *Forest Ecology and Manegement*, **156**, 75-85.

König AO, Ziegenhagen B, van Dam BC *et al.* (2001) Chloroplast DNA variation of oaks in western Central Europe and genetic consequences of human influences. *Forest Ecology and Management*, **156**, 147-166.

Lamb A, Boshier D (1994) *The reproductive biology of common ash (Fraxinus excelsior): a literature review*. Oxford Forestry Institute, University of Oxford, UK.

Lande R (1988) Genetics and demography in biological conservation. *Science*, **241**, 1455-1460.

Lande R, Barrowclough GF (1987) Effective population size, genetic variation and their use in population management. In: *Viable populations for conservation* (ed. Soulé ME), pp. 87-123. Cambridge University Press, Cambridge, UK.

Latta RG, Linhart YB, Fleck D, Elliot M (1998) Direct and indirect estimates of seed versus pollen movement within a population of ponderosa pine. *Evolution*, **52**, 61-67.

Le Corre V, Dumolin-Lapègue S, Kremer A (1997) Genetic variation at allozyme and RAPD loci in sessile oak *Quercus petraea* (Matt.) Liebl.: role of history and geography. *Molecular Ecology* **6**, 519-529.

Le Corre V, Kremer A (1998) Cumulative effects of founding events during colonization on genetic diversity and differentiation in an island and stepping-stone model. *Journal of Evolutionary Biology*, **11,** 495-512.

Ledig FT (1992) Human impacts on genetic diversity in forest ecosystems. *Oikos*, **63**, 87-108.

Lee SW, Ledig FT, Johnson DR (2002) Genetic variation at allozyme and RAPD markers in *Pinus longaeva* (Pinaceae) of the White Mountains, California. *American Journal of Botany*, **89**, 566-577.

Leonardi S, Menozzi P (1996) Spatial structure of genetic variability in natural stands of *Fagus sylvatica* L. (beech) in Italy. *Heredity*, **77**, 359-368.

Longauer R (1999) Forest resources of the PHARE countries. In: *Conservation and sustainable management of forests in central and eastern European countries*, pp. 7-20. Phare Environment Consortium, European Commission Phare Programme, Brussels, Belgium.

Longauer R, Krajmerova D (1999) Protective, conservation and special forest functions. In: *Conservation and sustainable management of forests in central and eastern European countries*, pp. 21-34. Phare Environment Consortium, European Commission Phare Programme, Brussels, Belgium.

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. *Annual Reviews in Ecology and Systematics*, **15**, 65-95.

Lynch M, Lande R (1993) Evolution and extinction in response to environmental change. In: *Biotic interactions and global change* (eds. Kareiva PM, Kingsolver JG, Huey RB), pp. 234-250. Sinauer Associates Inc., Sunderland, MA, USA.

Mariette S, Cottrell J, Csaikl U *et al.* (2002) Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Quercus petraea* (Matt.) Liebl. and *Q. robur* stands. *Silvae Genetica* (in press).

Mátyás C (1996) Climatic adaptation of forest trees: rediscovering provenance test data. *Euphytica*, **92**, 45-54.

Mátyás C (1998) Setting priorities in conservation of genetic diversity – with special reference to widely distributed conifer species. In: *Conservation of forest genetic resources in Europe. Proceedings of the European Forest Genetics Workshop, 21 November 1995, Sopron, Hungary* (eds. Turok J, Palmberg-Lerche C, Skrøppa T, Ouédraogo AS), pp. 11-18. International Plant Genetic Resources Institute, Rome, Italy.

Milligan BG, Leebens-Mack J, Strand AE (1994) Conservation genetics: beyond the maintenance of marker diversity. *Molecular Ecology*, **3**, 423-435.

Mohanty A, Martín JP, Aguinagalde I (2000) Chloroplast DNA diversity within and among populations of the allotetraploid *Prunus spinosa* L. *Theoretical and Applied Genetics*, **100**, 1304-1310.

Mohanty A, Martín JP, Aguinagalde I (2001) A population genetic analysis of chloroplast DNA in wild populations of *Prunus avium* L. in Europe. *Heredity*, **87**, 421-427.

Morand ME, Brachet S, Rossignol P, Dufour J, Frascaria-Lacoste N (2002) A generalized heterozygote deficiency assessed with microsatellites in French common ash populations. *Molecular Ecology*, **11**, 377-385.

Moran-Prieur ME, Vedel F, Raquin C, Brachet S, Sihachakr D, Frascaria-Lacoste N (2002) Maternal inheritance of a chloroplast microsatellite marker in controlled hybrids between *Fraxinus excelsior* and *Fraxinus angustifolia*. *Molecular Ecology*, **11**, 613-617.

Namkoong G (1984) Strategies for gene conservation in forest tree breeding. In: *Plant genetic resources: a conservation imperative, AAAS Selected Symposium 87* (eds. Yeatman CW, Krafton D, Wilkes G), pp.93-109. Westview Press, Boulder, CO, USA.

Namkoong G (1998) Forest genetics and conservation in Europe. In: *Conservation of forest genetic resources in Europe. Proceedings of the European Forest Genetics Workshop, 21 November 1995, Sopron, Hungary* (eds. Turok J, Palmberg.Lerche C, Skrøppa T, Ouédraogo AS), pp. 3-10. International Plant Genetic Resources Institute, Rome, Italy.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1-10.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology and Evolution*, **14**, 140-145.

Nunney L, Campbell KA (1993) Assessing minimum viable population size: demography meets population genetics. *Trends in Ecology and Evolution*, **8**, 234-239.

Oddou-Muratorio S, Guesnet D, Ozdemir E, Petit RJ, Demesure B (2001) Patterns of seed dispersal in a scattered forest tree species (*Sorbus torminalis*) based on multi-scale investigation of population genetic structure for chloroplast DNA. *Genetic response of forest systems to changing environmental conditions*, **70**, 271-280.

O'Malley DM, Whetten R (1997) Molecular markers and forest trees. In: *DNA markers: Protocols, Applications and Overviews* (eds. Caetano-Anollés G, Gresshoff PM), pp. 237-257. John Wiley & Sons, New York, USA.

Palmé AE, Vendramin GG (2002) Chloroplast DNA variation, postglacial recolonisation and hybridization in hazel, *Corylus avellana*. *Molecular Ecology*, **11**, 1769-1780.

Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA (1998) What molecules can tell us about populations: choosing and using a molecular marker. *Ecology*, **79**, 361-382.

Peltier WR (1994) Ice Age Paleotopography. *Science*, **265**, 195-201.

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997) Chloroplast footprints of postglacial recolonisation by oaks. *Proceedings of the National Academy of Sciences of the USA*, **94**, 9996-10001.

Petit RJ, Brewer S, Bordács S *et al.* (2002a) 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 RJ, Csaikl UM, Bordacs S *et al.* (2002b) Chloroplast DNA variation in European white oaks: Phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156**, 5-26.

Picard JF (1982) Contribution à l'étude de la biologie florale et de la fructification du Frêne commun (*Fraxinus excelsior* L.). *Revue Forestière Française*, **XXXIV**, 97-107.

Picard JF (1983) A propos du frêne oxyphylle *Fraxinus angustifolia. Forêt-entreprise*, **83**, 9.

Pigliucci M, Benedettelli S, Villani F (1990) Spatial patterns of genetic variability in Italian chestnut (*Castanea sativa*). *Canadian Journal of Botany*, **68**, 1962-1967.

Pliura A (1999) European long-term gene conservation strategies – Ash (*Fraxinus* ssp.) In: *Noble Hardwoods Network. Report of the third meeting, 13-16 June 1998, Sagadi, Estonia* (compilers Turok J, Jensen J, Palmberg-Lerche C *et al.*) pp 8-20. International Plant Genetic Resources Institute, Rome, Italy.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144**, 1237-1245.

Primack RB (1993) *Essentials of conservation biology.* Sinauer Associates, Inc., Sunderland, MA, USA.

Rameau JC, Mansion D, Dumé G (1989) Flore forestière française, guide écologique illustré. 1. Plaines et collines. Institut pour le développement forestier, Paris, France.

Rotach P (1999) *In situ* conservation and promotion of noble hardwoods: silvicultural management strategies. In*: Noble Hardwoods Network. Report of the third meeting, 13-16 June 1998, Sagadi, Estonia* (compilers Turok J, Jensen J, Palmberg-Lerche C, Rusanen M, Russell K, de Vries S, Lipman E), pp.39-50. International Plant Genetic Resources Institute, Rome, Italy.

Sander T, König S, Rithe GM, Janβen A, Weisgerber H (2000) Genetic variation of European beech (*Fagus sylvatica*) along an altitudinal transect at mount Vogelsberg in Hesse, Germany. *Molecular Ecology*, **9**, 1349-1361.

Shaffer ML (1981) Minimum population sizes for species conservation. *Bioscience*, **31**, 131-134.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics,* **16**, 393-430.

Soltis DE, Glitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, **206**, 353-373.

Soranzo N, Alia R, Provan J, Powell W (2000) Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. *Molecular Ecology*, **9**, 1205-1211.

Sperisen C, Büchler U, Gugerli F, Mátyás G, Geburek T, 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.

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution*, **16**, 608-613.

Streiff 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. *Molecular Ecology,* **7**, 317-328.

Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology and Evolution*, **15**, 199-203.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453-464.

Tapper PG (1992) Irregular fruiting in *Fraxinus excelsior*. *Journal of Vegetation Science*, **3**, 41-46.

Tapper PG (1996) Long-term patterns of mast fruiting in *Fraxinus excelsior*. *Ecology*, **77**, 2567-2572.

Thill A (1970) *Le frêne et sa culture*. Presses Agronomiques de Gembloux, Editions J. Duculot, S.A., Gembloux, Belgium.

Turok J (2002) Collaborative networking on forest genetic resources in Europe. In: *Proceedings of the DYGEN Conference: Dynamics and conservation of genetic diversity in forest ecosystems*, p.20. European Parliament, Strasbourg, France.

Turok J, Eriksson G, Kleinschmit, Canger S, compilers (1996) Noble Hardwoods Network. Report of the first meeting, 24-27 March 1996, Escherode, Germany. International Plant Genetic Resources Institute, Rome, Italy.

Turok J, Collin E, Demesure B, Eriksson G, Kleinschmit, Rusanen M, Stephan R, compilers (1998) Noble Hardwoods Network. Report of the second meeting, 22-25 March 1997, Lourizán, Spain. International Plant Genetic Resources Institute, Rome, Italy.

Turok J, Jensen J, Palmberg-Lerche C, Rusanen M, Russell K, de Vries S, Lipman E, compilers (1999) Noble Hardwoods Network. Report of the third meeting, 13-16 June 1998, Sagadi, Estonia. International Plant Genetic Resources Institute, Rome, Italy.

Turok J, Eriksson G, Russel K, Borelli S, compilers (2002) Noble Hardwoods Network. Report of the 4th meeting, 4-6 September 1999, Gmunden, Austria and the 5th meeting 17-19 May 2001, Blessington, Ireland. International Plant Genetic Resources Institute, Rome, Italy.

Tutin TG, Heywood VH, Burges NA, Valentine DH, Walters SM, Webb DA, eds. (1964) *Flora Europaea*, volume 1. Cambridge University Press, UK.

Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA, eds. (1972) *Flora Europaea*, volume 3. Cambridge University Press, UK.

Tzedakis PC, Lawson IT, Frogley MR, Hewitt GM, Preece RC (2002) Buffered tree population changes in a Quaternary refugium: evolutionary implication. *Science*, **297**, 2044-2047.

Vendramin GG, Degen B, Petit RJ, Anzidei M, Madaghiele A, Ziegenhagen B (1999) High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Molecular Ecology*, **8**, 1117-1126.

Wade MJ, McCauley DE (1998) Extinction and recolonisation: their effects on the genetic differentiation of local populations. *Evolution*, **42**, 995-1005.

Wallander E (2001) *Evolution of wind-pollination in Fraxinus (Oleaceae) – an ecophylogenetic appraoch*. PhD thesis. Botanical Institute, Göteborg University, Sweden.

Waples RS (2002) Definition and estimation of effective population size in the conservation of endangered species. In: *Population viability analysis* (eds. Beissinger SR, McCullough DR), pp.147-168. The University of Chicago Press, Chicago, USA.

Webb T, Bartlein PJ (1992) Global changes during the last 3 million years: climatic controls and biotic response. *Annual Review of Ecology and Systematics*, **23**, 141-173.

Weiser F (1995) Beitrag zur Existenz von Ökotypen bei Gemeiner Esche (*Fraxinus excelsior* L.). *Forstarchiv*, **66**, 251-257.

Wilkinson DM (1997) Plant colonization: are wind-dispersed seeds really dispersed by birds at larger spatial and temporal scales? *Journal of Biogeography*, **24**, 61-65.

Willis KJ, Rudner E, Sümegi P (2000) The full glacial forests of central and southeastern Europe. *Quaternary Research*, **53**, 203-213.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97-159.

Zanetto A, Roussel G, Kremer A (1994) Geographic variation of inter-specific differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Forest Genetics*, **1**,111-123.

Zanetto A, Kremer A (1995) Geographical structure of gene diversity in *Quercus petraea* Matt. Liebl. I. Monolocus patterns of variation. *Heredity*, **75**, 506-517.

Appendices

Appendice I: Sizing of microsatellite loci with Genescan and Genotyper

FEMSATL 4:

Dye: 6-FAM

Nice banding pattern with 1 major peak, 2 to 3 smaller stutter bands. Small +A peak.

Even allele sizes: $(AG)_n$ mostly in south-eastern Europe

Odd allele sizes: AAA $(AC)_n$ $(AG)_n$ all over Europe

Scored size		Assigned size	
lower limit	upper limit	lower limit	upper limit
155	172	155	172
172.5	188.5	173	189
190	229	191	230

FEMSATL 11:

Dye: NED

Nice banding pattern, nearly always double peaks, nearly no stutters. I chose the peak of the largest allele size (the right-most peak) and rounded it down. Most alleles have even size and differ by 2 bp or multiples of 2 bp. Exceptions occur in southeastern Europe (Moldaovia).

FEMSATL 16:

Dye: 6-FAM

Very easy to read banding pattern, small number of alleles. Alleles differed by 2 bp or multiples of 2 bp. The overall most common allele was sized to 184 bp, an allele of 183 bp occured in Romania.

FEMSATL 19:

Dye: HEX

Double peaks were common. Even and uneven allele sizes occured. There seem to be different repeat patterns in alleles of similar size (homoplasy), as differences between allele sizes were sometimes quite continuous (especially in south-eastern Europe). A criterion had to be fixed how to round up or down the sizes. In each population, I checked if alleles of similar size were put into the same category.

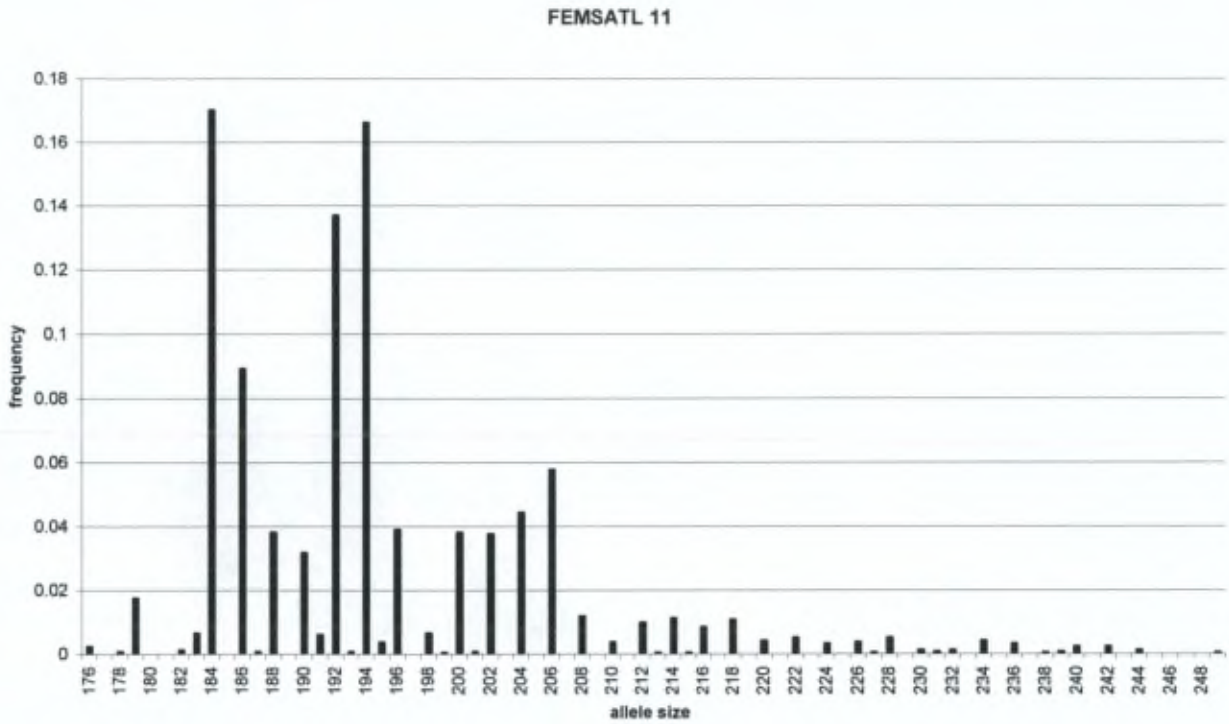
Scored size		Assigned size
lower limit	upper limit	
142	158	leftmost peak was rouded down
170	196	rightmost peak was rounded down
197	210	rightmost peak was rounded up
211	230	size OK

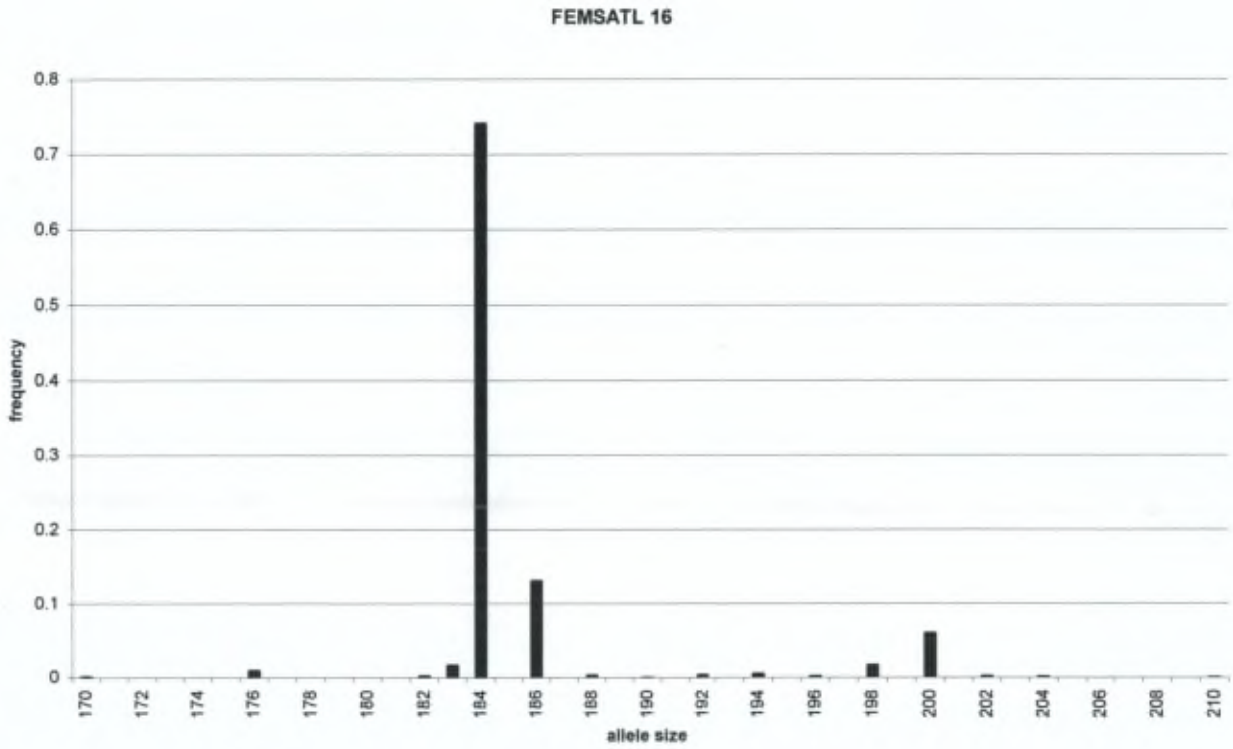
M230:

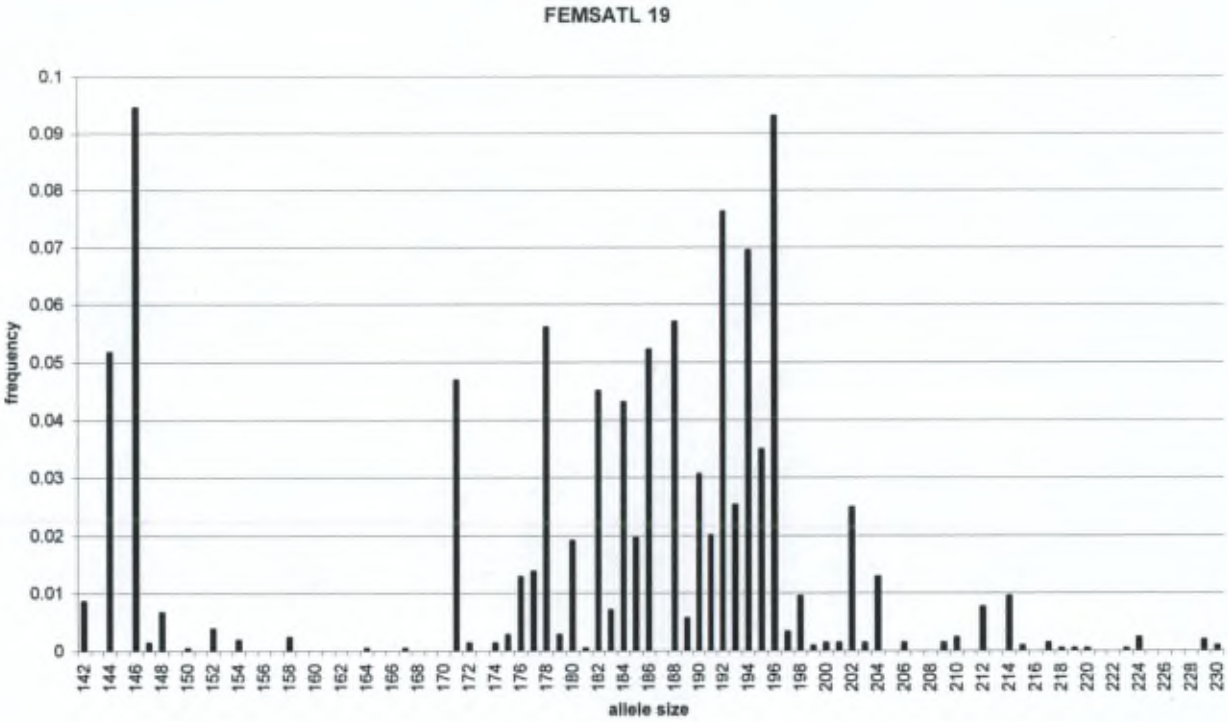
Dye: HEX

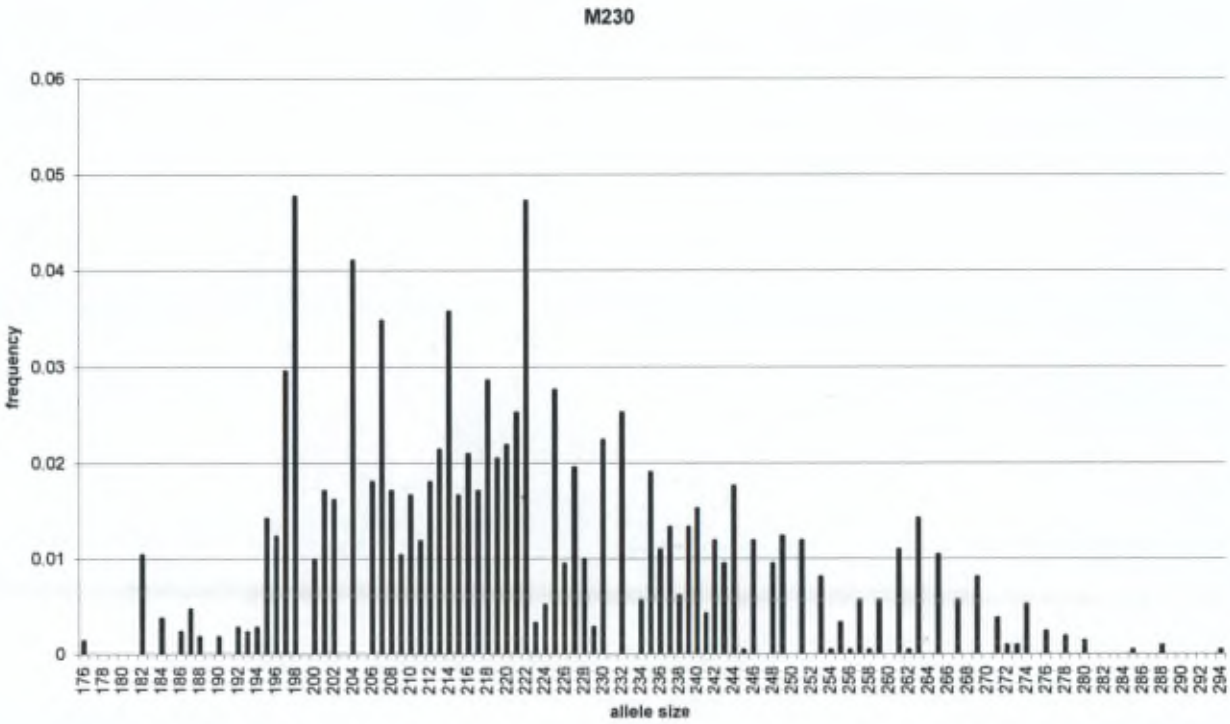
Many stutters! Small alleles were quite easy to score, large alleles exhibited a "hedgehog" pattern. I compared many large allele patterns and decided what peak should be the allele (not always the tallest peak!). Regarding sizes all possible decimals occured, the repeat pattern is probably highly variable. I produced a frequency distribution of allele sizes and made classes of alleles with different widths (e.g. from 0.7 to 1.5 bp). Cut-offs between classes were made where the number of observations was small. In each population, I checked if alleles of similar size were put into the same class. Approximately one percent of alleles had to be re-classified with this method, so it seems to be quite robust.

Appendice II: Allele frequency distributions in four microsatelite loci over a mean of 1069 common ash trees






FEMSATL 19
frequency
allele size


M230
frequency
allele size

Appendice III: Statistics of genetic diversity in 36 European populations of common ash

	Population	A_S_36	H_E	F_{IS}		T_2_IAM	T_2_SMM
1	IreCa	11.475	0.8368	0.0916	***	1.467	-1.698
2	IreAv	12.684	0.8465	0.0574	ns	1.149	-2.585
3	IreKi	11.5984	0.8108	0.1109	***	-0.174	-5.501
4	ScotTa	13.214	0.8191	0.0389	***	0.776	-3.262
5	FraGo	12.9946	0.8423	0.1566	**	1.289	-1.235
6	FraLr	11.4212	0.8376	0.0845	**	1.045	-1.465
7	SwiB	12.7026	0.7644	0.0478	*	0.334	-3.079
8	SwiG	12.502	0.7429	0.0396	ns	-0.361	-4.885
9	SwiE	12.5984	0.7899	0.0922	*	1.048	-1.936
10	GerTi	12.7518	0.8291	-0.0235	ns	1.215	-2.144
11	GerCh	10.3068	0.7365	0.0378	ns	0.397	-3.440
12	GerFr	11.4668	0.8193	0.0268	**	0.963	-2.473
13	SweRoe	10.0816	0.7434	0.0191	ns	-0.427	-6.046
14	SweEhd	8.5356	0.6609	-0.0393	ns	-1.96	-9.199
15	SweAk	7.3316	0.6505	0.0159	ns	-1.805	-10.415
16	LitZei	13.4694	0.8475	0.103	***	0.24	-4.857
17	LitKai	14.9014	0.8528	0.0773	***	-0.415	-4.295
18	SloPa	11.8542	0.8305	0.0844	**	0.847	-2.331
19	HunLue	11.742	0.8195	0.0015	ns	0.555	-3.699
20	BulLj	6.1466	0.6499	-0.0217	*	1.974	0.008
21	BulVi	7.962	0.6793	-0.1085	ns	-1.769	-8.778
22	BulMo	7.906	0.7239	-0.0018	***	1.663	-0.747
23	RomTu	9.3636	0.6761	-0.0197	ns	-1.771	-9.311
24	RomBaa	11.1366	0.7994	-0.0111	ns	-0.377	-6.797
25	RomBab	11.803	0.7648	0.0438	ns	-2.317	-12.011
26	RomBac	11.57	0.7731	0.0449	ns	-0.861	-7.449
27	BulGr	11.334	0.7938	0.0192	ns	0.418	-4.295
28	BulBc	8.4638	0.6661	0.093	ns	-1.288	-7.583
29	BulEl	11.4122	0.7823	0.0551	ns	0.457	-3.889
30	BulIh	10.8918	0.7076	-0.029	ns	-1.854	-6.317
31	BulDu	11.1984	0.7435	0.065	**	-1.019	-8.325
32	BulKo	11.4784	0.7416	-0.0164	ns	-0.131	-5.386
33	BulZd	11.8866	0.7456	0.0221	ns	-1.902	-8.242
34	MolHr	14.3858	0.8019	-0.0052	ns	-0.657	-5.092
35	MolHi	13.6782	0.7832	0.0161	ns	-0.24	-3.743
36	MolHa	13.5956	0.8138	0.0255	**	-0.345	-5.135

A_S_36, allelic richness for standardised samples of 36 gene copies; H_E, gene diversity; F, Wright's inbreeding coefficient and exact test of departure from Hardy-Weinberg genotypic proportions: *, P<0.05; **, P<0.01; ***, P<0.001; ns, not significant; T_2, bottleneck statistic from Cornuet and Luikart (1996) computed on the basis of the infinite alleles model (T_2 IAM) and the stepwise mutation model (SMM); S.D. (populations), standard deviation over populations.

Samples from Bulgaria were rescored between chapters I and III from the thesis, which explains the slight differences between the values.

Résumé: Cette thèse s'intéresse aux facteurs responsables de la structure de la diversité génétique chez le frêne commun. L'accent est mis sur les patrons de diversité génétique au sud-est de l'Europe, qui sont comparés à la structure génétique dans d'autres régions d'Europe dans le contexte de la démographie des populations et de la recolonisation postglaciaire du frêne.
L'analyse de microsatellites chloroplastiques nous a permis de confirmer la localisation de refuges glaciaires pour le frêne dans les Alpes et dans la Péninsule Balkanique, suggérée par des données de pollen fossile. Nos résultats sont aussi en accord avec d'autres refuges proposés en Italie et dans la Péninsule Ibérique. Les microsatellites nucléaires ont révélé des patrons de diversité différents, avec notamment des situations contrastées à l'ouest et au centre de l'Europe d'un côté, et au sud-est de l'Europe et en Suède de l'autre. A l'ouest et au centre de l'Europe, une diversité élevée, de légères empreintes de goulots d'étranglement et un pool génique homogène sur toute la région ont été interprétés comme résultant d'un flux de pollen considérable au cours de la recolonisation postglaciaire, homogénéisant les pools géniques de colonisateurs issus de refuges différenciés. Au sud-est de l'Europe, une différenciation importante ainsi qu'une diversité relativement faible et l'absence de signes de goulots d'étranglement ont été attribuées à une expansion depuis plusieurs refuges avec peu de flux de gènes entre populations, sous l'hypothèse qu'une situation démographique stable avec une densité élevée en frêne aurait été rapidement atteinte dans la région. Une diversité plus élevée à l'est qu'à l'ouest de la Péninsule Balkanique pourrait être due à de l'hybridation avec d'autres espèces de frênes sympatriques dans des populations proches des grandes rivières à l'est. Alternativement, ce patron pourrait aussi être expliqué par des caractéristiques contrastées des populations de refuges glaciaires, avec par exemple des populations fragmentées dans les montagnes du sud-est de l'Europe qui auraient subi une dérivé génétique plus importante qu'une large population dans un refuge de plaine alluviale près de la Mer Noire.
Nous avons estimé un flux de graines très restreint et un flux de pollen modéré dans une population roumaine de frêne commun en comparant la structure de la diversité génétique dans la population avec des résultats de simulations où nous avions contrôlé la dispersion de graines et de pollen.
Nous nous sommes également intéressés à l'histoire évolutive d'un locus microsatellite nucléaire et nous avons identifié un déterminisme moléculaire distinct et des distributions géographiques distinctes pour des allèles de taille paire ou impaire. Deux scénarios évolutifs ont été proposés pour expliquer ces patrons, impliquant ou non l'hybridation avec une autre espèce de frêne, peut-être le frêne à feuilles étroites. Au sein de la famille des oléacées, le locus microsatellite montre une variabilité importante de la région répétée.
Dans l'ensemble, ce travail a identifié le sud-est de l'Europe et le nord de l'Europe centrale comme des centres importants pour la conservation des ressources génétiques chez le frêne.

Summary: This thesis is an attempt to understand the factors that shaped the present day spatial structure of genetic diversity in the common ash. It focuses on the structure of genetic diversity in southeastern Europe and compares it to diversity patterns in other European regions in the context of population demography and postglacial recolonisation dynamics during the current glacial interstadial.
Chloroplast microsatellite analysis allowed us to corroborate glacial refuges for ash situated in the Alps and in the Balkan Peninsula, as suggested from fossil pollen data, and our results are in agreement with other proposed refuges in Italy and the Iberian Peninsula. Nuclear microsatellite analysis revealed distinct diversity patterns, and notably contrasting situations in western and central Europe on the one hand, and in southeastern Europe and Sweden on the other. In western and central Europe, slight imprints from bottlenecks, high genetic diversity and a homogeneous gene pool over the entire area were interpreted as the result of extensive pollen flow, homogenising the nuclear gene pools of colonisers from differentiated refuges during the process of postglacial recolonisation. In southeastern Europe, high among-population differentiation as well as relatively low levels of genetic diversity and no signs from bottlenecks were attributed to the expansion from different glacial refuges with little among-population gene flow, hypothesising that a demographically stable condition with high density of ash was rapidly reached in the area. Higher diversity in the eastern than in the western part of the Balkan Peninsula was tentatively explained by hybridisation with other sympatric ash species in floodplain populations in the east. Alternatively, this pattern could also be due to contrasting demographic features of glacial refuge populations, with for instance scattered and fragmented refuge populations in the southeastern European mountains that would have been subjected to stronger genetic drift than a large floodplain refuge population near the Black Sea. In Sweden, diversity patterns were similar to southeastern Europe, which lead us to suspect the existence of cryptic full glacial refuges in northern central Europe.
Further, very narrow seed dispersal and moderate pollen dispersal were estimated in a Romanian common ash population by comparing the spatial genetic structure in the population with results from computer simulations where seed and pollen dispersal were under control.
We also investigated the evolutionary history of a nuclear microsatellite locus and identified a different molecular determinism and different geographical distributions of odd- and even-sized alleles in common ash. Two evolutionary scenarios were proposed to explain the observed patterns, involving or not the hybridisation with another ash species, possibly narrow-leaved ash. Within the olive family, the microsatellite locus shows a high variability of the repeat region.
Overall, this work identified southeastern Europe and northern central Europe as important centres for the conservation of genetic resources in common ash.