GENETIC STRUCTURE AT DIFFERENT SPATIAL SCALES IN METAPOPULATIONS OF *SILENE TATARICA*

NIINA TERO

Faculty of Science,
Department of Biology,
University of Oulu

OULU 2005

NIINA TERO

GENETIC STRUCTURE AT DIFFERENT SPATIAL SCALES IN METAPOPULATIONS OF *SILENE TATARICA*

Academic Dissertation to be presented with the assent of the Faculty of Science, University of Oulu, for public discussion in Kuusamonsali (Auditorium YB210), Linnanmaa, on August 26th, 2005, at 12 noon

OULUN YLIOPISTO, OULU 2005

Supervised by
Docent Jouni Aspi
Docent Pirkko Siikamäki

Reviewed by
Docent Anni Harju
Docent Anna Westerberg

ISBN 951-42-7768-6 (nid.)
ISBN 951-42-7769-4 (PDF) http://herkules.oulu.fi/isbn9514277694/

ISSN 0355-3191 http://herkules.oulu.fi/issn03553191/

OULU UNIVERSITY PRESS
OULU 2005

Tero, Niina, Genetic structure at different spatial scales in metapopulations of *Silene tatarica*
Faculty of Science, Department of Biology, University of Oulu, P.O.Box 3000, FIN-90014 University of Oulu, Finland
2005
Oulu, Finland

Abstract

The genetic structure at different spatial scales and growing habitats was studied on *Silene tatarica*, using AFLP and microsatellite markers. *S. tatarica* is a rare perennial plant occurring along riverbanks and shores of two annually flooding rivers in Finland. Regional scale analysis based on AFLP fragment analysis showed that at Oulanka River population structure represented mostly classical metapopulation model. In general, colonization-extinction processes had an important role, dispersal between subpopulations was limited and genetic differentiation was independent of geographic location.

The same subpopulations were partly used to study spatial genetic structuring within subpopulations. Spatial autocorrelation revealed clear spatial genetic structure in each subpopulation. Paternity analysis in an isolated subpopulation showed small amounts of inbreeding, restricted seed dispersal and pollen flow through the subpopulation. Factors affecting the creation and maintenance of spatial genetic structure within subpopulation were most likely colonization events and restricted seed dispersal.

The impact of river regulation on the genetic structure of populations was studied by comparing results from Oulanka River to the results obtained from second main growing area, Kitinen River. Oulanka River is a natural river system, whereas Kitinen is a regulated river. The overall regional scale studies did not indicate major differences between river systems. There were some clear population genetic differences between rivers but there were no clear evidence that those would have been caused by river regulation. More likely differences were related to the marginal location of Kitinen population at the edge of the distribution range. Studies indicated that regardless of the species rarity in Finland, active management measures are not currently needed in either *S. tatarica* growing area.

Species specific microsatellite loci were isolated to complement AFLP studies. During the microsatellite isolation, an interesting amplification pattern was detected and studied further. It was suggested that there were repetitive areas within genome containing microsatellites resulting in unusual amplification. The most likely explanation for this phenomenon would be transposable elements containing proto-microsatellite areas. The microsatellites isolated could have evolved mostly from those proto-microsatellites.

Keywords: AFLP, microsatellite, repetitive DNA, riparian

Acknowledgements

This work was carried out at the Department of Biology, Division of Genetics, University of Oulu. I would like to thank my supervisors Jouni Aspi and Pirkko Siikamäki for providing interesting research topic and help in various aspects during my work. I am grateful for University of Oulu Department of Biology, especially Division of Genetics Professors Pekka Pamilo and Outi Savolainen, for providing an excellent working environment and who also have been very helpful and supportive.

I spent three months in Vienna at Prof. Christian Schlötterer's lab and I would like to thank people I met there: Claus, Gerhard, Viola, Ram, Hannah, Graham, Francesco, Carla and Jean-Michel for friendly assistance in the lab and Vienna habits. Especially I will thank Prof. Schlötterer for spending long hours helping with laboratory works and manuscripts.

From the technical assistance I thank Hannele Parkkinen, Minna Rinta-Pukkila and Anita Ala-Poikela and people working in the field for providing samples and ecological information for me: Anne Jäkäläniemi, Anna Kilpiä, Sirpa Marttila and the staff of the Oulanka Research Station.

From the second floor labs I would like to thank people in ants group: Anu, Hannaleena, Lumi, David, Sami, Marianne, Riitta and Christian B and from the next lab: Minna, Laura and Maria and their group members and others who have spend smaller periods of time in the lab. From our small plant group I thank Marjut and also Maare who shared the office with me many years. In generally, I would like to thank people with whom I have spend time at work and during the free time.

I would also like to thank my parents Mauri and Leena, brother Ismo, Reeta, friends and relatives. Luckily there is also other life than just work. Last but not least I thank Liisa for being very good sister and friend.

I am grateful to docents Anni Harju and Anna Westerberg for the valuable comments on my thesis and to Dr. David Hughes for revising the language. This work was financially supported by the University of Oulu grants to Niina Tero and Pirkko Siikamäki and Population Genetic Graduate School (Academy of Finland).

Oulu, May, 2005 Niina Tero

Abbreviations

AFLP	amplified fragment length polymorphism
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism
SMM	stepwise mutation model
PSM	proportional slippage model
SNP	single nucleotide polymorphism
SSCP	single strand conformational polymorphism
cpDNA	chloroplast DNA
LTR	long terminal repeat
SINE	short interspersed repetitive element
LINE	long interspersed repetitive element
MITE	miniature inverted-repeat transposable element
UTR	untranslated region

List of original papers

This thesis is based of the following papers, which are referred to in the text by their Roman numerals:

I Tero N, Aspi J, Siikamäki P, Jäkäläniemi A & Tuomi J (2003) Genetic structure and gene flow in a metapopulation of an endangered plant species, *Silene tatarica*. Mol Ecol 12: 2073-2085.

II Tero N, Aspi J, Siikamäki P & Jäkäläniemi A (2005) Local genetic population structure in an endangered plant species, *Silene tatarica* (Caryophyllaceae). Heredity 94: 478-487.

III Tero N & Schlötterer C (2005) Isolation and characterization microsatellite loci from *Silene tatarica*. Mol Ecol Notes, In press.

IV Tero N (2005) Genetic structuring of *Silene tatarica* along regulated Kitinen River in northern Finland (manuscript)

V Tero N, Neumeier H, Gudavalli R & Schlötterer C (2005) Microsatellites isolated from *Silene tatarica* are frequently located in repetitive DNA (submitted to J Mol Evol).

Contents

1 Introduction

1.1 Population genetic structure

Genetic variation within a species may be structured both by spatial factors and by genetic background. The neutral theory of molecular evolution provides a null model against which alternative hypotheses can be tested. However, selection and many other evolutionary processes can cause departures from neutral theory predictions (Harrison & Hastings 1996, Charlesworth *et al*. 2003). Many natural species are subdivided into local breeding units, a condition that can lead to the genetic differentiation of local demes. As a result of the spatial structure, the demography and genetics of populations will be a product not only of local environmental conditions but also processes operating on a regional scale (Husband & Barrett 1996). Much of the theory of population structure contrasts the effects of finite local population size and the rate of the gene flow among populations on genetic structure. Gene flow strongly influences the spatial scale over which genetic differentiation will be observed (Slatkin 1985). Population structure is critically affected by various dynamics at different scales. In understanding genetic variation the properties of new populations, like founding number, probability of common origin and kin structure, are also important (Whitlock & McCauley 1990, Hanski 1999).

1.1.1 Metapopulation approach

Metapopulation approach stems from the general notion of the hierarchical structure of nature and that spatially delimited local populations are connected by some degree of migration (Hanski 1999). Some possible models for linearly arranged subpopulations are presented in Fig. 1 and from those the last two resemble metapopulations. A metapopulation approach refers to research or management that in some way adopts the view that local populations are more or less discrete entities in space and that these local populations interact via migration and gene flow (Hanski & Gaggiotti 2004). Because

population geneticists have been studying metapopulations for decades before the word metapopulation was coined, the terminology can be confusing in the literature. In the original papers presented here 'population' or 'metapopulation' has been used to describe regional scale areas and 'subpopulation' to refer to geographically local populations within a metapopulation. Three fundamental processes form the core of metapopulation biology: local population extinction, (re)colonization and migration (Hanski 1999).

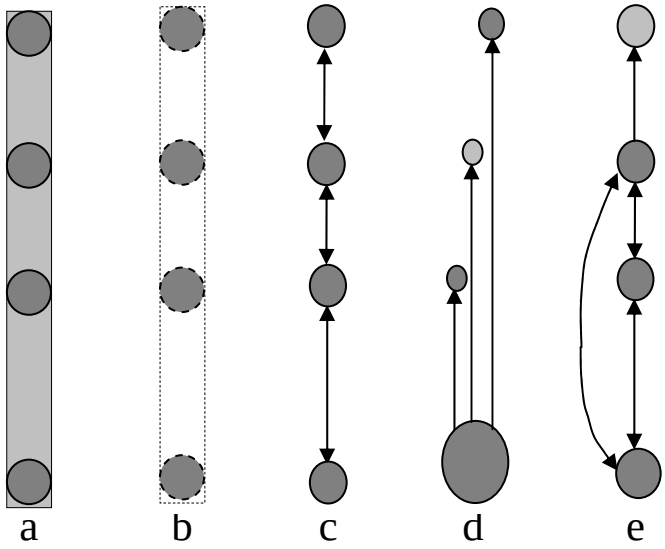

Fig. 1. Schematic representation of the possible population structure and migration models between linearly arranged subpopulations: a) genetically uniform population with a free gene flow across the population b) fragmentized population without recurrent gene flow between the subpopulations c) stepping-stone population d) source-sink population and e) 'classical' metapopulation.

Plants possess special features compared to animals such as seed dormancy, restricted dispersal and local adaptations presenting both challenges and opportunities for metapopulation biology (Husband & Barrett 1996). Several approaches to studies of dispersal can be found in the literature (see for example Levin *et al*. 2003, Manel *et al*. 2005). One is to use molecular markers and population genetic analysis. Two basic genetic methods can be followed in the study of plant dispersal. Indirect methods are based on the quantification of genetic divergence between populations or individuals and interpret this divergence in terms of the amount of past gene flow. Second, direct methods try to establish the parent-offspring relationship between individuals in space and give estimates of instantaneous dispersal. In other words, indirect methods give estimates from cumulative effects of gene flow whereas directs estimates apply only to the interval of time and space over which observations are made (Slatkin 1985, Neigel 1997, Ouborg & Eriksson 2004). Indirect methods were used on regional scale and both indirect and direct methods on local scale in the studies presented here. For plant populations, which are

subject to catastrophic forms of disturbance like *Silene tatarica* in its natural habitat (studied here), long-distance seed dispersal is essential for structure like metapopulation to persist (Cain *et al*. 2000). It was important to estimate the likelihood and amount of that kind of events.

In general, in a metapopulation with population turnover, average within-deme diversity, measures of total species diversity and F_{ST} may deviate strongly from those expected in an equivalent species with a stable demography (Pannell & Charlesworth 2000). Stable genetic or spatial structure may increase variability because lineages may remain associated with different environments for a long time. On the other hand, fluctuations in the structure of a population can greatly reduce variation, like bottlenecks which mainly decrease the amount of rare alleles (Luikart *et al*. 1998). Even in constant environments, there is fluctuation in subpopulation sizes caused by factors like reproductive success. The effect of structure depends on how long genes remain associated with particular genetic backgrounds and demes, relative to the timescales of coalescence and fluctuations in structure (Charlesworth *et al*. 2003).

In the past, single population approaches have dominated ecology and evolutionary biology. The metapopulation framework recognizes and provides a conceptual tool for dealing with the interactions of within and among subpopulation processes. Not all species whose populations have undergone fragmentation fit the definition of a metapopulation (Thrall *et al*. 2000). Different types of regional populations have been suggested in addition to metapopulations in a strict sense (see for example Freckleton & Watkinson 2003) but as suggested by Ouborg and Eriksson (2004, p 469) “it would be erroneous to force a diversity of regional dynamics into a narrow set of concepts and definition”. The metapopulation concept has been productive in stimulating studies of regional populations and broadening of the metapopulation concept may be more useful than developing new terms and concepts (Ouborg & Eriksson 2004). There are increasing number of studies demonstrating the importance of spatial structure in ecological and evolutionary questions and metapopulation approach provides tools for dealing with the interactions (Thrall *et al*. 2000).

1.1.2 Spatial genetic structure

One of the evolutionary consequences of localized dispersal is the development of fine-scale local adaptation in plant populations (Husband & Barrett 1996). According to the neutral theory, microgeographic isolation-by-distance structure may be generated in a population after a few generations as a consequence of fine-scale genetic processes. Four alternative scenarios are possible according to different seed and pollen dispersal and their variances. Either both pollen and seed dispersal are highly localized around the maternal plant and have similar variances and then inbreeding and genetic substructuring evolve as described by the isolation-by-distance model or both are random when neither inbreeding nor spatial genetic structure will develop. If pollen dispersal is localized but seed dispersal random then neither inbreeding nor genetic structure will develop without selfing. When pollen dispersal is random but seed dispersal is highly localized there will

not be inbreeding but significant fine-scale genetic structure will evolve (reviewed in Ennos 2001, Kalisz *et al*. 2001, Rousset 2001).

Spatial genetic structure may be generated in plant populations also as a consequence of sampling events that occur when a population is founded or regenerated. Spatial genetic structure generated in this fashion will be greatest where regeneration sites are colonized from a limited number of maternal/paternal seed parents. A key difference between this form of genetic structure and that generated purely by restricted gene flow is that it represents a non-equilibrium situation (Whitlock & McCauley 1990, Ennos 2001, Kalisz *et al*. 2001). Thus, if founder events are behind the spatial genetic structure, it should diminish or disappear with increasing stand age. Microspatial, as well as macrospatial, genetic structuring has been found from *Silene* species (McCauley 1997, Giles *et al*. 1998, Gehring & Delph 1999, Richards *et al*. 1999) and was studied here from *S. tatarica*.

Several statistical methods have been used to describe local spatial structure (see Escudero *et al*. 2003 and Manel *et al*. 2003). Spatial genetic structure has been usually described by means of spatial autocorrelation. Other often used measures to describe spatial structure within subpopulations are ‘patch width’ and ‘neighborhood size’ (Chambers 1995, Rousset 2001, Vekemans & Hardy 2004). Parentage analysis was also included in one *S. tatarica* subpopulation while studying microspatial structure.

1.2 Neutral molecular markers

Natural plant populations routinely and consistently show small-scale genetic differentiation (more than 100 species demonstrated) although important exceptions are known (Linhart & Grant 1996). Most molecular variation observed is presumed to be neutral or nearly neutral characters. Regardless of that, a minority of evolutionary biologists argue that genetic differentiation has been typically produced by natural selection in response to environmental heterogeneity. It has been shown that beneficial mutations are more common than assumed and methods have been developed to screen them (Linhart & Grant 1996, Schlötterer 2002).

Neutral theory of molecular evolution predicts that the majority of evolutionary changes and variability within species are caused by random genetic drift of alleles which are selectively neutral. Selection may operate but the intensity is weak compared to neutral processes (Hartl & Clark 1997, Graur & Li 2000). Most molecular markers are presumed to be neutral and these can be used to analyze problems in evolution. Different kind of markers can be useful for different purposes. Mutational processes resulting in differences in the data between markers should be taken into account when applying analyzing methods. Codominant markers allow distinguishing heterozygotes from homozygotes and can be more informative than dominant markers which show presence/absence variation. Some widely used markers are allozymes, RFLPs, AFLPs, SNPs and microsatellites. Many studies of the evolution of genes and genomes include analyzing variation within and among populations and species (Hartl & Clark 1997, Futuyma 1998, Silvertown & Charlesworth 2001). Molecular markers used in studies presented here are described later.

1.3 The genesis of microsatellites

The origin of microsatellites is poorly understood, even though microsatellites are very frequent in genomes (Ellegren 2004). Microsatellite mutations occur during replication by DNA slippage (Tautz 1989) and mechanisms with point mutation has been proposed for microsatellite genesis as well. In the first model, short tandem repeat sequences existing in the genome are starting point for microsatellite expansion. A second model suggests that point mutations first create tandemly duplicated motifs. In both cases, areas with few repeat units may start slowly expand through slippage-like mutation processes or mutate back to non-repeat structure (Schlötterer 2000, Ellegren 2004). Neutral origin hypothesis was supported in *Drosophila melanogaster* research where genomic regions with different GC content had significantly different microsatellite density and amount of $(TA)_n$ repeats (Bachtrog *et al*. 1999). Analysis from rice (*Oryza sativa* L.) also indicated that the occurrence of particular microsatellite motif was strongly associated with the GC content (Temnykh *et al*. 2001).

Other mechanisms of microsatellite evolution have been also suggested. Probably the most striking observation is the high frequency of microsatellites in the proximity of interspersed repetitive elements, such as short interspersed repeats (SINEs) and long interspersed elements (LINEs). It has been proposed that these microsatellites originated from the 3' poly (A) tails of reverse transcribed RNA that has been inserted into the genome (Nadir *et al*. 1996). Despite this overwhelming evidence, it is unlikely the majority of microsatellites originated through this mechanism. First, mainly A-rich microsatellites are associated with the 3' poly (A) tails (Nadir *et al*. 1996). Second, at least in plants, microsatellites are more abundant in non-repetitive DNA (Schlötterer 2000, Morgante *et al*. 2002).

Finally, microsatellite genesis has been associated with proto-microsatellite sequences carried in transposable elements. This phenomenon was first described for the mini-me element, which is observed in a broad range of Dipteran species (Wilder & Hollocher 2001). The authors proposed that the proto-microsatellite was distributed through transposition, and in inactive elements the proto-microsatellite could expand to a full microsatellite. Thus, a large number of microsatellites could be generated from one sequence that has an intrinsic property to mutate to a microsatellite sequence. Consistent with the hypothesis, most microsatellites associated with repetitive elements were expansions of transposable elements internal sequences in barley (Ramsay *et al*. 1999). Furthermore, in rice $(AT)_n$ microsatellites were frequently associated with the Micropon family of miniature inverted-repeat transposable elements (MITEs) (Temnykh *et al*. 2001).

Retrotransposons are the most abundant and widespread class of eukaryotic transposable elements, consisting of the long terminal repeat (LTR) and the non-LTR retrotransposons (Kumar & Bennetzen 1999, Graur & Li 2000) which are present throughout the plant kingdom and can be found in high copy numbers. In general, plant genomes contain an exceptionally abundant and diverse set of retrotransposons compared to most other eukaryotic genomes, with the exception that functional retroviruses are either absent or present in very small numbers (Kumar & Bennetzen 1999). Even though, the retrotransposon activity has been very high in both the distant and recent past

(Wessler *et al*. 1995). It has been also suggested that some tandem repeats localized centromeric regions of plants may have originated from parts of retrotransposons (Cheng & Murata 2003).

1.4 Study species and molecular methods

1.4.1 Silene tatarica

Silene tatarica (L.) (Caryophyllaceae) is a rare perennial plant growing along periodically disturbed riverbanks and shores of two rivers in Finland. The main distribution area of *S. tatarica* is on the Russian steppes with disjunctive occurrences in Hungary, Germany, Lithuania and NW-Russia (Ulvinen 1997). North-western range of distribution area of the species is in the northern Finland, where it has invaded naturally the riverside habitats of the Oulanka and the Kitinen Rivers after last ice age about 10,000 years ago. These rivers are over 100 km apart and separated by a watershed (Fig. 2). According to the present distribution of the *S. tatarica*, the colonization route to Finland has been from east. There is no common phylogeographies found for colonization after ice age, and the postglacial colonization routes of *S. tatarica* have not studied in detail. General theory according to studies on other species suggest that northern part of the Europe was colonized primarily from the Iberic and the Balkan refugia but the importance of Eastern Europe and Fennoscandia may be underestimated because of lack of data (Taberlet *et al*. 1998).

The plants grow from open sand and gravel shores and erosion banks to more densely vegetated meadows shores and riverbanks. Subpopulations are widely scattered in more or less continuous areas of suitable long and narrow habitats along riversides. In addition to these natural populations, a few secondary populations are found from road- and railway sides indicating successful anthropochoric and/or zoochoric dispersal. Establishment of new subpopulations and the expansion of old ones occur solely by seed dispersal and the species does not spread vegetatively (Jäkäläniemi *et al*. 2005). The seeds presumably are dispersed by gravity at the slopes, wind, water and probably by animals and man as well (I). Flowers are pollinated both by bumble bees and night-flying moths and according to investigations with UV-dusts, the direct pollen dispersal in *S. tatarica* was on average only 7 meters (Siikamäki *et al*. unpublished result). The ecology of *S. tatarica* has been studied in many years field surveys concerning topics like survival strategies (Jäkäläniemi *et al*. 2004) and patch dynamics (Jäkäläniemi *et al*. 2005).

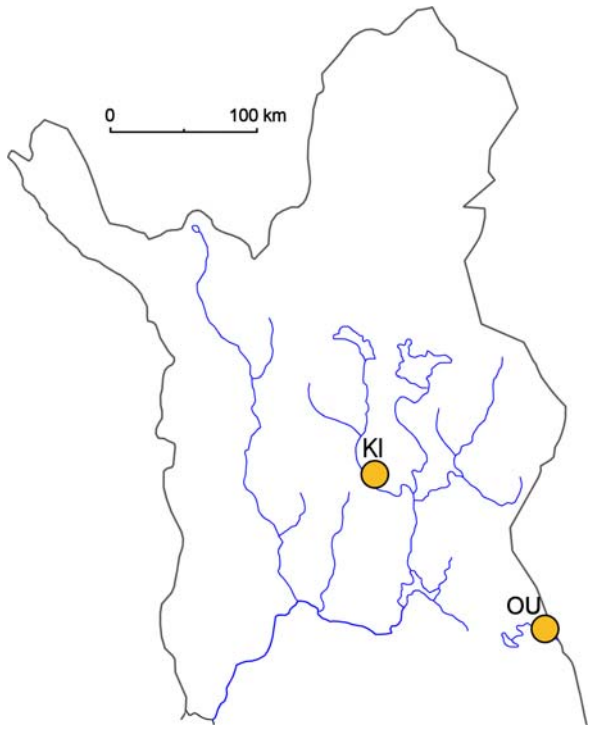

Fig. 2. Locations of Oulanka (OU) and Kitinen (KI) River sample sites in northern Finland.

1.4.2 AFLP fragment analysis

Amplified fragment length polymorphism (AFLP) marker system is a technique through which selected fragments from the digestion of total DNA are amplified by the PCR (Fig. 3). The marker shows a presence/absence variation among individuals throughout the genome (Vos *et al*. 1995). The AFLP method has many advantages: there is no need for prior sequence knowledge, the repeatability is generally good, quantity and quality of DNA requirements are small, and the resulting DNA fingerprints provide a large number of genetic markers. The drawback of this method is that markers are dominant so that heterozygotes cannot be distinguished from homozygotes and also factors like reproducibility has been questioned (Nybom 2004, Schlötterer 2004).

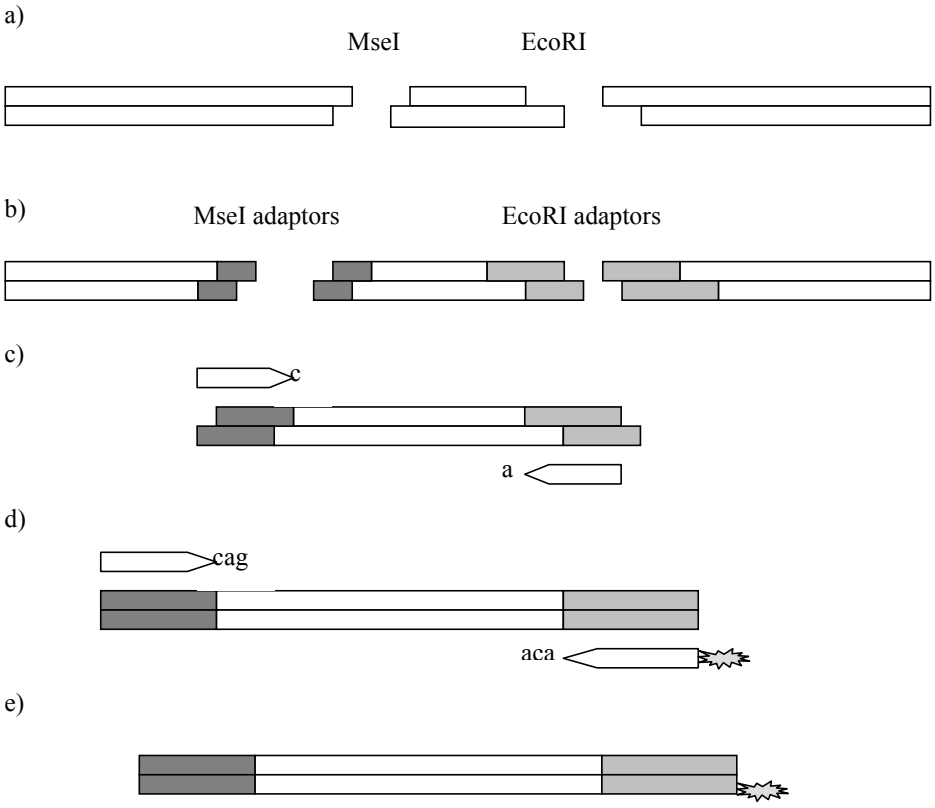

Fig. 3. Short introduction to AFLP protocol: a) genomic DNA digested with MseI and EcoRI enzymes b) MseI and EcoRI adaptors ligated c) preselective amplification (one selective nucleotide) d) selective amplification with fluorescently labeled primer (three selective nucleotides) and e) final selected and labeled AFLP product.

1.4.3 Microsatellite fragment analysis

Microsatellites are DNA sequences in which a short motif of 1-6 bases are tandemly repeated. Their inherent instability and wide distribution within most genomes makes microsatellite loci particulary useful for evolutionary and population genetic studies (Goldstein & Schlötterer 1999). Traditional stepwise mutation model (SMM) for microsatellite mutation assumes that the length of a microsatellite varies at a fixed rate independent of repeat length and with same probability of expansion and contraction. The model is not necessarily realistic for various reasons and new models have been developed and tested (Kruglyak *et al.* 1998, 2000, Whittaker *et al.* 2003, Calabrese & Durrett 2003, Sainudiin *et al.* 2004).

The length variation of microsatellite alleles has been created by process called replication slippage (Tautz 1989). One new mutation model suggested is proportional slippage (PS) model in which balance between slippage events and point mutations exists and point mutations might prevent expansion by breaking long microsatellites (Kruglyak *et al*. 1998, 2000). The simplest models assume that length, measured by number of repeat units, changes by only one unit per mutation without bias in expansion or contraction rates. The single unit changes appear to be the most common type but also larger mutations have been detected, especially in downward direction. Moreover, short alleles are biased toward expansion or unbiased whereas mutations in longer alleles favor contraction and alleles with a greater number of repeats appear to be more mutable violating assumption length independency (see for example Xu *et al*. 2000, Harr & Schlötterer 2000, Brohede *et al*. 2002, Vigouroux *et al*. 2002). Even though proved to be very useful in many kinds of studies, the major drawbacks with microsatellites are the time and costs needed to develop them, scoring problems and statistical model violations (Nybom 2004, Schlötterer 2004).

1.4.4 DNA sequence variation

Basic processes in the evolution of DNA sequences are the substitution of one nucleotide for another during evolutionary time and length variation through insertions and deletions. Basically, all variation in genetic markers systems is sequence polymorphic (e.g., Schlötterer 2004) but in this chapter I am discussing sequence variation assayed through DNA sequencing. Changes in nucleotide sequences are used in molecular evolutionary studies and cannot be dealt with by direct observation because the process of nucleotide substitutions is usually extremely slow. To detect evolutionary changes one needs to compare two sequences that have descended from a common ancestral sequences and such comparisons require statistical analysis (Graur & Li 2000). A sequencing approach provides the most fine-grained genetic information. Although method was initially time consuming and expensive, recent technological developments permit large scale analysis nowadays.

1.4.5 Chloroplast DNA analysis

Gene flow within and among plant populations is usually mediated by dispersal of pollen and/or seeds. A complete gene flow description in plants should then include the relative importance of pollen and seeds dispersal (Ennos 1994). Organelle DNA usually shows uniparental inheritance, little or no crossing over and different evolutionary rate from nuclear DNA. Chloroplast DNA (cpDNA), usually used in plant studies, are generally maternally inherited in angiosperms and can not disperse by pollen whereas nuclear genomes migrate by both pollen and seeds (Ennos 2001). Accordingly, the relative influences of seeds and pollen dispersals on total gene flow can be resolved by comparing nuclear and maternally (in some cases paternally) inherited markers (Ennos 1994, McCauley 1995). Techniques like RFLP have been used successfully to detect

polymorphism from chloroplast in several species whereas few or no polymorphisms have been found from others. The mutation rates associated with the chloroplast are low but new markers, as chloroplast microsatellites, may increase the variation found (Provan *et al*. 2001).

I tried to detect variation from chloroplast DNA, starting from simple methods and moving toward more fine-grained and faster evolving markers. The first method used was PCR-RFLP. Chloroplast areas trnT-L, trnL-F and trn intron were amplified from Oulanka River population samples using primers described Taberlet and co-authors. (1991) and products were digested using multiple restriction enzymes. Since there was no variation between samples, the same loci were run using more accurate SSCP (Single Strand Conformation Polymorphism) methods which can detect one nucleotide changes (Orita *et al*. 1989). The sample size was about ten individuals from four subpopulations. In the next step, the same chloroplast areas were sequenced directly from their PCR products. From 16 to 23 individuals were analyzed from each locus by sequencing one or both directions. Finally, four chloroplast microsatellite loci (ccmp1, ccmp4, ccmp6 and ccmp10) isolated by Weising and Gardner (1999) were used to analyze about 240 individuals from 10 subpopulations with fluorescently labeled primers and subsequent gel electrophoresis.

I could not detect any variation between individuals or subpopulations using any of the methods. Accordingly, chloroplast results were not used in original papers and more detailed methods or results are not presented here. Most likely it would be possible to detect chloroplast variation from *S. tatarica* either by increasing the number of loci or by using wider spatial scale. However, because it would be difficult to estimate how much effort was required to find polymorphism, it was decided to concentrate on nuclear variation for my thesis.

1.5 Goals of this work

The purpose of this study was to get basic population genetic information from the endangered riparian species *S. tatarica* to estimate the impact and importance of river dynamics to the plant species in these habitats. We estimated genetic parameters from subpopulations to determine the most probable population model. Our hypothesis was that natural river dynamics creates extinction and colonization of subpopulations, detected in ecological studies, which can create metapopulation structure. Amount of gene flow strongly affects the population genetic structure and related to that, the relative rates of pollen and seed dispersal may have impact on plant populations. Even though, it was not possible to determine these relative rates without cpDNA markers, many other methods were applied to detect the pattern of gene flow. The mating system of the species was not known and was one of the research topics. The subpopulations are scattered in the riverside and the combination of low migration between subpopulations and inbreeding could be serious threat to species survival. To estimate the lack of natural river dynamics, samples from Kitinen riverside were collected, analyzed and compared to results from free-flowing rivers. The properties of recently founded subpopulations may also have strong impact to whole population structure when the colonization is frequent

like in the case of *S. tatarica*. From the collected data we could compare old and new subpopulations and the age of subpopulations could be estimated relatively reliably.

My thesis consists of studies from geographical structuring at different scales which were studied mainly using AFLP markers. At the regional scale, indirect methods were used to estimate historical migration events between subpopulations and many methods were used to estimate other genetical parameters (papers I and IV). Direct and indirect methods were used to detect contemporary gene flow and spatial autocorrelation at microspatial scale (II). *S. tatarica* is endangered species so the river regulation could have impacted upon population viability and conservation management. Different environments may lead to alteration in genetic pattern of the species and this was studied by comparing subpopulations from different kinds of riverside habitats (IV). Collecting data from different marker types gives a better overall picture and the results are more reliable when two independent systems are used. We ended up isolating microsatellite markers for *S. tatarica* with reasonable success (III) but because that project was started rather recently microsatellite loci were used only for genetic research in Kitinen River population (IV).

The last article (V) concentrates on the evolution of microsatellites and the reasons leading to difficulties in the development of microsatellite markers for *S. tatarica*. Microsatellite isolation was rather demanding and we wanted to look in more detail at what caused problems in the procedure. Sequences isolated from enriched library were compared to each other and nucleotide diversity values were estimated between most similar ones to find out the degree of diversity between isolated loci.

2 Materials and methods

The materials and methods are given here only briefly. More detailed descriptions are found in the original papers (I-V).

2.1 Plant material and populations studied

We collected plant samples from c. 30 individuals at seven different sites along Oulanka River (Fig. 4; see also paper I, Table 2) in the summer 1999. Five of the sites were in Finland and two in Russia. Three of those were also used in microspatial scale studies (II).

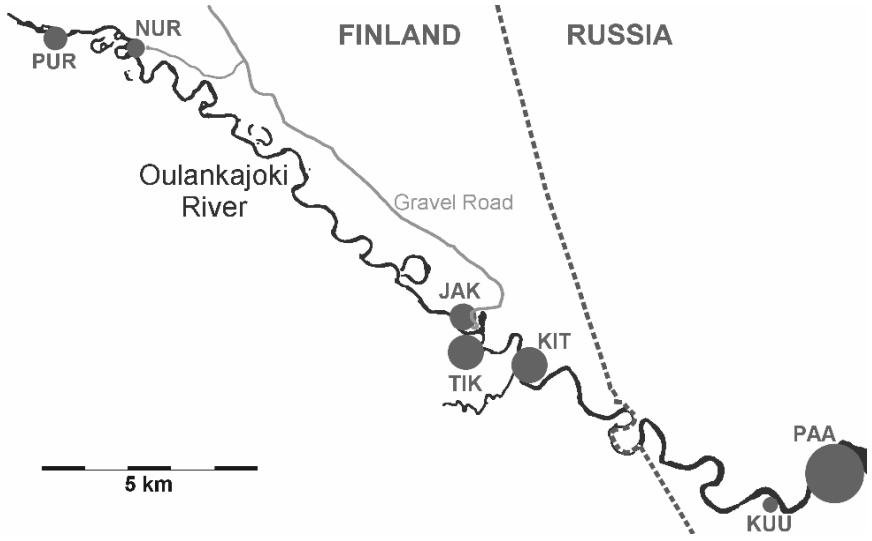

Fig. 4. Map of the sample sites at Oulanka River. The subpopulation name abbreviations refer to the Table 2 in paper I.

For more detailed microspatial analysis recently colonized, isolated island Valtikkasaari (VAL) subpopulation was sampled. From VAL we collected leaf material, capsules from all fertile individuals and mapped all individuals. The seeds from each capsule were sown in a greenhouse in a turf-sand mixture and seedlings were counted and sampled after three months.

For the comparison between river systems, about 30 individuals from six subpopulations were sampled from Kitinen riverside during summers 2001-2002 (Fig. 5; see also paper IV, Table 1).

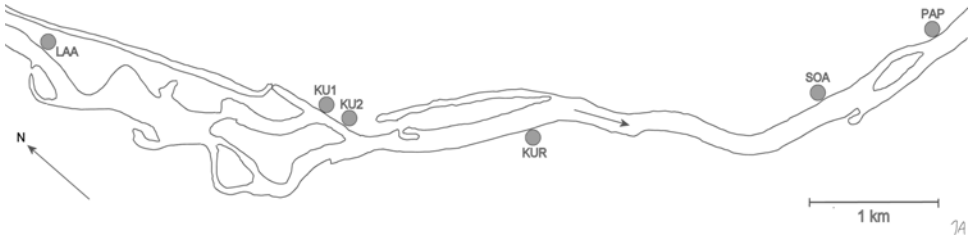

Fig. 5. Map of the sample sites at Kitinen River. The subpopulation name abbreviations refer to the Table 1 in paper II.

2.2 DNA isolation

Leaf samples were collected from natural habitats or from greenhouses and transported on ice to the laboratory. Fresh material was stored at -20ºC until DNA was extracted. Genomic DNA was isolated from 0.1 g frozen leaves of individual plants using the slightly modified CTAB method (Rogers & Bendich 1985). Total DNA concentration was measured at UV absorbance A_{260} using BioPhotometer (Eppendorf).

2.3 Molecular methods

2.3.1 AFLP fragment analysis

Amplified fragment length polymorphism reactions were conducted using the AFLP™ Plant Mapping kit supplied by Perkin-Elmer, Applied Biosystems (USA). A kit optimized for a genome size of 500 to 6000 Mb was used according to the manufacturer's instructions with slight modifications. The denatured samples were analyzed on an automated DNA sequencer (ABI Prism 377™) with ROX-500 internal size standard (Applied Biosystems). Electropherograms were analyzed by GeneScan and Genotyper softwares (Applied Biosystems). More detailed protocol has been described in paper I.

2.3.2 Microsatellite isolation and fragment analysis

Microsatellite isolation procedure has been described in papers III and V. Genomic DNA (10 µg) was digested with *MboI* (Fermentas), separated on agarose gel and DNA fragments sized 500-800 bp were recovered from the gel (Schlötterer 1998). SauLA and SauLB adaptors were annealed and ligated to approximately 200 ng of the total digested, size-selected DNA. 3'-end biotinylated $(GT)_{22}$ microsatellite probe was used to enrich the library in microsatellite sequences (Hammond *et al.* 1998). 300 µg Streptavidin beads (Dynabeads M-280, Dynal Biotech) was incubated with genomic DNA/probe solution. The washings were done according Hammond and co-authors (1998). Genomic DNA was eluted from Dynabeads, amplified with PCR and products were cloned into M13mp19 vector and sequenced (see below).

Microsatellite polymorphism was analyzed by PCR both with radioactive and fluorescent labeled primers and subsequent gel electrophoresis. Radioactive labeling and runs were done according Schlötterer (1998). PCR products labeled with γ^{32}P-ATP were separated on 7% denaturing polyacrylamide gel and visualized by autoradiography. For fluorescent labeled runs (papers III and V), microsatellite polymorphism was analyzed by PCR with one fluorescently labeled primer and subsequent gel electrophoresis on a ABI Prism 377™ sequencer with TAMRA-350 internal size standard. Electropherograms were analyzed by GeneScan and Genotyper softwares (Applied Biosystems).

2.3.3 Cloning and DNA sequence analysis

Enriched fragment products were cloned into M13mp19 vector (III and V). To investigate further what caused multiple banding, PCR products from one randomly chosen microsatellite was ligated pCR®2.1-TOPO vector (Invitrogen™) and transformed to chemical competent TOP10F' cells (V). Plasmid DNA was isolated from clones and inserts were sequenced directly using plasmids or through standard PCR reaction. Sequencing was done using BigDye 1.1 sequencing chemistry (Applied Biosystems) and products were separated on a MegaBase sequencer.

2.4 Statistical methods

2.4.1 Data analysis on macrospatial scale studies

Overall population genetic parameters were estimated by multiple methods and programs (papers I, II and IV). For AFLP's the number and percent of polymorphic loci and Nei's gene diversity was calculated using POPGENE program (Yeh *et al.* 1995). Linkage equilibrium analysis was conducted using chi-square test (I) following Miyashita and co-authors (1999) or LIAN (ver 3.1) software (II and IV) which tests for independent assortment (Haubold & Hudson 2000). The presence of genetic structure in the total

population was analyzed using Bayesian method by program Hickory (Holsinger & Lewis 2002).

The program MICRO-CHECKER was used to detect genotyping errors from microsatellite markers (Oosterhout *et al*. 2004). To estimate from microsatellites allele frequencies, F_{ST} values for whole sample and pairwise comparison between subpopulation, average gene flow (*Nm*) and linkage disequilibrium between loci, GENEPOP software web version was used (http://wbiomed.curtin.edu.au/genepop/).

For both markers the presence of genetic structure on a regional and subpopulation level was tested by an analysis of variance framework using Analysis of Molecular Variance (AMOVA) by program Arlequin ver. 2.00 (Schneider *et al*. 2000). Software Structure (ver 2.0) was used to infer population structure and assign individuals into subpopulations (Pritchard *et al*. 2000, Falush *et al*. 2003).

2.4.2 Data analysis on microspatial scale studies

Microspatial scale was analyzed in paper II. Some population genetic parameters were estimated as described in the previous chapter for the new Valtikkasaari (VAL) subpopulation. The spatial autocorrelation analysis was conducted in each subpopulation using Hardy´s kinship coefficient between individuals versus distance in logarithmic scale using SPAGeDi (Hardy & Vekemans 2002). To characterize the spatial genetic pattern of subpopulations, we estimated the indirect estimate of neighborhood size (*Nb*) and *Sp* statistics for each subpopulation on the basis of spatial autocorrelation (Hardy 2003, Vekemans & Hardy 2004). For the parentage analysis, Famoz software was used (Gerber *et al*. 2003) with categorial allocation approach (Jones & Arden 2003).

2.4.3 Data analysis on nucleotide variation

Sequences included to families found by BLAST (Altschul *et al*. 1990) were aligned first using ClustalW online program (http://www.ebi.ac.uk/clustalw/index.html) and the results were checked and corrected manually using BioEdit ver. 5.0.9 software (Hall 1999). Nucleotide diversities (π) were estimated using MEGA ver. 2.1 (http://www.megasoftware.net/). All values estimated are presented in paper V.

3 Results and Discussion

3.1 Metapopulation structure at Oulanka river

The first paper (I) investigated the distribution of genetic variation within and between seven subpopulations at Oulanka River. In *S. tatarica* mean Nei's estimate for gene diversity of the subpopulations was 0.127. This seems to be at intermediate level of genetic diversity similar to other endangered or vulnerable species obtained with AFLP markers (Travis *et al*. 1996, Gaudeul *et al*. 2000, Keiper & McConchie 2000). We could not detect any correlation between the subpopulation census size and amount of genetic variation (paper I, Fig. 3).

The overall pattern of genetic variation within and between subpopulations suggested a classical metapopulation structure of the species. AFLP marker data contained seven clusters where each original sample site seems to contain a distinct subpopulation. AMOVA analysis and UPGMA-dendrogram did not suggest hierarchical regional structuring among the subpopulations. Most genotypic diversity was found within subpopulations (63%) but even though only about 37% was between subpopulations, the Bayesian estimate of F_{ST} was relatively high (0.287). Respective studies in related species have suggested less differentiation among subpopulations (see Giles & Goudet 1997, McCauley 1997, Gehring & Delph 1999). There was no correlation between geographic and genetic distances among the subpopulations indicating lack of isolation-by-distance.

Estimate of *Nm* (0.333) suggested a low level of gene flow between the subpopulations, and the assignment tests proposed a few long-distance dispersal events between the subpopulations. Long-distance dispersal was most likely mediated by zoochoric or hydrochoric dispersal of seeds. Relative high amount of linkage disequilibrium at subpopulation level indicated recent population bottlenecks or local substructuring (McVean 2002). Extinction-colonization dynamics of the subpopulations has been characterized during a five years field study in the Oulanka River region (Jäkäläniemi *et al*. 2005) supporting metapopulation model with frequent population turnover.

The metapopulation structure of *S. tatarica* was somewhat expected on the basis of preliminary ecological studies and studies from related species. Different genetical

parameters built up a general view from the subpopulations studied and gave valuable information for the species. The AFLP method was applied successfully for *S. tatarica* and proved to be effective method regardless of statistical and other problems mentioned in the Introduction.

3.2 Microspatial genetic analysis

Existence and mechanism behind the microspatial genetic structure of *S. tatarica* in four subpopulations was investigated in paper II. Codominant markers are usually the prime choice for this type of studies but as shown here and in other articles (see for example Krauss & Peakell 1998, Mariette *et al*. 2002, Hardy 2003), also dominant markers like AFLPs may be used.

Our results suggested that the local spatial genetic structure in *S. tatarica* was attributed merely to the isolation-by-distance process rather than founder effect, and despite free pollen movement across population, restricted seed dispersal maintains local genetic structure in this species. Linkage disequilibrium values tended to be higher in older than in younger subpopulations probably because the continuous isolation-by-distance process could have enhanced linkage disequilibrium in the older ones. This was further supported by the *Sp* statistics values which increased with stand age whereas in the case of founder effect opposite is expected (Vekemans & Hardy 2004). Restricted seed dispersal seems to be typical mechanism creating structuring for perennial herb species (see for example Westerbergh & Saura 1994, Miyazaki & Isagi 2000, Kalisz *et al*. 2001).

The number of seedlings produced was very uneven between mother plants (from zero to over hundred) and among the greenhouse grown offspring the genetic polymorphism appeared to be higher than within parental population. *S. tatarica* commonly produce large amount of seeds but seedling survival is low (Marttila 2004). Parentage analysis in an isolated island subpopulation indicated that species is predominantly outcrossing with a minimal amount (0.47%) of selfing. This result does not exclude the possibility of biparental inbreeding, especially in the small subpopulations. Mean estimated pollen dispersal distance (24.10 m; SD = 10.5) was significantly longer whereas the mean seed dispersal distance (9.07 m; SD = 9.23) was considerably shorter than the mean distance between the individuals (19.20 m; SD = 13.80). Distributions are shown in Fig. 3 in paper II.

The indirect genetic estimates of neighborhood sizes varied from 15.8 to 35.8 and direct estimate of VAL subpopulation was 37.6, respectively. In Vekemans and Hardy's (2004) review the mean neighborhood size for herbaceous species was 21.8. Our estimates were close to that. Neighborhood sizes were much smaller (from 1/10 to 1/5) than the surveyed subpopulation sizes, except in the VAL subpopulation where it was 85%.

The first large scale study at Oulanka River raised many questions, as level of structuring and inbreeding, which were answered in the paper II. Local scale study increased our knowledge from the species and showed that there is no immediate genetic

threat for species existence at Oulanka River. Likewise, we did not detect inbreeding regardless of spatial genetic structuring.

3.3 Development of microsatellite markers

Five new microsatellite loci were isolated from *S. tatarica*, published in paper III. The markers presented are the first microsatellites reported for this species. We used an enriched genomic library method which was optimized during the work and the same protocol has not previously used to our knowledge.

We characterized *S. tatarica* individuals originating from two riverbank populations in northern Finland and observed between four and nine alleles per locus (paper III, Table 1). We had far less individuals from Oulanka (44) than from Kitinen riparian subpopulations (178 individuals). Even though, we found two alleles from both riversides using three loci (ST5, ST21 and ST28) and only one allele less from Oulanka samples when ST12 locus was analyzed. Moreover, one locus was monomorphic at Kitinen samples but there were four alleles at Oulanka samples. Results indicate that there are more alleles at Oulanka River population. Because most alleles were different between rivers, the systems are most likely somewhat genetically divergent from each others, also supported by AFLP markers (Tero *et al*. unpublished result).

Observed heterozygosity was consistently lower than gene diversity (mean $Ho = 0.125$ and mean $He = 0.455$). This deficiency of heterozygous genotypes was observed in most locus/population combinations, which may be caused by spatial genetic structuring within subpopulations detected in other study (II). Wahlund effect i.e., that there is genetic structuring within subpopulation, is the most likely explanation for excess homozygosity in outcrossing populations (Silvertown & Charlesworth 2001).

There was significant linkage disequilibrium between loci within subpopulations in most comparisons as in other studies and will not be discussed further here. From each locus pairs across all populations seven from ten comparisons were significant after conservative Bonferroni correction ($P < 0.005$).

Some species have been proven to be difficult to obtain microsatellites from and *S. tatarica* seems to belong to these (see also 3.5). During the work we optimized microsatellite enrichment procedure which can be used if more loci are needed and is most likely suitable for other species as well. The loci were used successfully to estimate population genetic parameters and conformed AFLP marker results (IV).

3.4 Comparison between river systems

Large scale genetic structuring from six *S. tatarica* subpopulations growing along regulated Kitinen River and comparison of genetic estimates to free-flowing river system is presented in paper IV. Kitinen River represents marginal distribution area of this species. This may have many genetic consequences in addition to regulation like increase in the effect of drift because of fluctuation in offspring number and survival over generations. Harsh environmental conditions are also predicted to decrease migration and

increase differentiation among subpopulations (Sun & Ritland 1998, Pamilo & Savolainen 1999, Silvertown & Charlesworth 2001).

Microsatellites were used for the first time in addition to AFLPs and in general the results were consistent between markers. Subpopulations were genetically differentiated but the genetic distances between subpopulations were not dependent on geographic distances. The $\varnothing_{ST}$ values overall subpopulations were rather high, 0.402 for AFLPs and 0.507 for microsatellites, and assignment analysis indicated few migration events between subpopulations supported by *Nm* estimate (0.138). Spatial and temporal Analysis of Molecular Variance (AMOVA) models between subpopulations were not significant. Most of the variation was within subpopulations using AFLP markers (59.78%) and rest among (40.22%) differing slightly from microsatellite results where the ratio was about half in both.

Genetic differentiation, defined as F_{ST} and related estimates, were larger between young than between old subpopulations. Estimated $\varnothing$ value was 0.21 and suggested mixed colonization mostly explained by the migrant pool model e.g., colonization from multiple sources. Differentiation between young subpopulations could then be mainly caused by low numbers of the initial colonist. Result was supported by F_{ST} values which indicated mixed colonization mostly caused by migrant pool model.

Generally, genetic estimates were similar compared to free-flowing river system suggesting that impoundment has not created major changes to population genetic parameters and metapopulation model can be applied also at Kitinen River. Largest differences, percentage of polymorphic loci, number of private alleles and number of migrants may have been caused by multiple factors. We found two private alleles using dominant markers from Kitinen River subpopulations whereas the number was 68 at Oulanka River. Mean percentage of polymorphic loci and number of migrants in Kitinen populations was about half of those observed in the Oulanka population. Population subdivision has been also found from related *S. dioica* where comparison between habitats showed importance of the surrounding vegetation to the gene flow pattern (Westerbergh & Saura 1994). Their results indicated that denser vegetation limited pollinator movement. However, the *S. tatarica* subpopulations studied here were growing at quite open riverside habitats where the pollinator behavior is most likely much the same as among the rivers.

Differences caused by marginal location of the population could be applied to the results and explained differences better than river regulation. Enhanced drift may have removed alleles whereas expected loss of genetic diversity was small as expected if the period enhanced drift is short. I detected also decrease in migration but the genetic distances among subpopulations were not larger. Differences between common and marginal populations have been found from other plant species and the genetic consequences vary between articles (Sun & Ritland 1998, Gao *et al*. 2000, Butcher *et al*. 2002). Even though, it is not possible to rule out other factors affecting genetic differences between rivers like impoundment or postglazial colonization according to the analysis presented in paper IV. At this moment this species does not need active management measures in its main growing areas in Finland. Creating open habitats and restoring connectivity could be considered if the need arises in the future.

3.5 Microsatellite evolution studies

The sequences isolated from enrichment library were studied with details in paper V. Contrary to expectations of unique microsatellite loci, 54% of all clones carrying a microsatellite repeat were found to be associated with other repetitive DNA. Within the limitation of a potential bias in the enrichment procedure, this implies that a very large proportion of the $(GT)_n$ microsatellites in *S. tatarica* are associated with, and most likely originate from, repetitive elements. We observed a rather high sequence divergence among the sequences grouping into a given family (paper V, Table 1). Within the limitations of a moderate number of family members, we could not identify a genomic region that is conserved either within or between families and database searches identified only few hits to known transposable elements.

Interestingly, we detected different repeat types in the same family and a similar phenomenon has been previously described in *Drosophila* (Harr *et al*. 2000). One example is family 3 in which two clones were detected where the repeat motif $(CTGTGT)_5$ was embedded in a $(GT)_n$ microsatellite. Moreover, the clone obtained from the enrichment procedure carried a $(GT)_n$ microsatellite and so did two of the cloned PCR products. In the remaining eight cloned PCR products, the $(GT)_n$ microsatellite was replaced by a highly variable $(AT)_n$ microsatellite (paper V, Fig. 4). Additional evidence for a turnover of microsatellite repeat types at homologous positions comes from a study in pines (Karhu *et al*. 2000). Between primate species has been detected microsatellite motif replacements in UTR (untranslated region) and intron areas (Riley & Krieger 2004, 2005).

For certain taxonomic groups, the isolation of microsatellites has been proven to be a major problem. Well documented examples include many plant species, several invertebrate groups, some dipterans and gastropods, and a number of avian groups (Goldstein & Schlötterer 1999). In this study, PCR primers for 23 loci amplified a product in the expected size range but only five loci amplified a single polymorphic locus. Four were designed from sequences, which did not show a significant match to other clones obtained from the enrichment procedure. One locus (ST8) however, was designed from a clone showing significant similarity to other clones. This observation indicates that due to the high sequence divergence among the repeated sequences flanking regions, even microsatellites linked to repetitive DNA could be converted into a unique locus marker.

4 Concluding remarks

The population genetic analyzes of *S. tatarica* species revealed clear overall pattern in investigated riparian habitats. Within and between subpopulation structures were affected with multiple factors operating in different spatial scales. Populations at both rivers mainly represented classical metapopulation, even though all model predictions were not fulfilled. I agree with Ouborg and Eriksson (2004) that forcing a diversity of regional dynamics into a narrow set of concepts and definition may not be the right way. Broadening of the metapopulation concept may be more useful than developing new terms and concepts.

There have been long term ecological studies in main distribution areas of *S. tatarica*. By combining both ecological and genetical data it is possible to get better understanding of the processes affecting species viability. The species survival is not threatened in Finland and the impact of human activity is already small at Oulanka national park. At Kitinen River it seems like impoundment has had little effect to the genetic parameters but also the evolutionary timescale is very small.

During the work I tested different types of markers and estimated how suitable those are for this kind of research. For example, I tried to find polymorphism from cpDNA without detecting any. Even though chloroplast markers would have been very useful for separating pollen and seed dispersal as shown in related species (e.g., McCayley 1997), for *S. tatarica* and at the studied scale, it was not possible to find variation from cpDNA and this marker type was not suitable. The microsatellite isolation proved to be difficult for *S. tatarica* and the interpretation concerning the genetic structure of the population in the regional and local scale was performed mainly by applying dominant AFLP markers. However, the microsatellite markers developed during the project may be used in the further genetic studies and it would be possible to optimize more microsatellite loci on the basis of pilot work done during my thesis.

In the future, the microspatial scale studies could be applied to Kitinen River subpopulations. It would be interesting to know if there are differences also at the local scale. Other additions to a population genetic research topic could be secondary habitats in Finland and main growing areas in Russia. It would be possible to gather more detailed information from microsatellite loci in *S. tatarica* increasing data set and including other repeat types during enrichment. This type of studies has not been common especially at

natural plant species. Since microsatellites are missing from many species and many isolation projects are surely ongoing currently, this type of project could have wider application.

References

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.

Bachtrog D, Weiss S, Zangerl B, Brem G & Schlötterer C (1999) Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. Mol Biol Evol 16: 602-610.

Brohede J, Primmer CR, Møller A & Ellegren H (2002) Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. Nucleic Acids Res 30: 1997-2003.

Butcher PA, Otero A, McDonald MW & Moran GF (2002) Nuclear RFLP variation in *Eucalyptus camaldulensis* Dehnh. from northern Australia. Heredity 88: 402-412.

Cain ML, Milligan BG & Strand AE (2000) Long-distance seed dispersal in plant populations. Am J Bot 87: 1217-1227.

Calabrese P & Durrett R (2003) Dinucleatide repeats in the Drosophila and human genomes have complex, length-dependent mutation processes. Mol Biol Evol 20: 715-725.

Chambers SM (1995) Spatial structure, genetic variation, and the neighborhood adjustment to effective population size. Conserv Biol 9: 1312-1315.

Charlesworth B, Charlesworth D & Barton NH (2003) The effects of genetic and geographic structure on neutral variation. Annu Rev Ecol Evol Syst 34: 99-125.

Cheng Z-J & Murata M (2003) A centromeric tandem repeat family originated from a part of Ty3/gypsy-retroelement in wheat and its relatives. Genetics 164: 665-672.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5: 435-445.

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

Ennos RA (2001) Inferences about spatial processes in plant populations from the analysis of molecular markers. In: Silvertown J & Antonovics J (eds) Integrating ecology and evolution in a spatial context. British ecological society Blackwell science, Oxford, p 45-71.

Escudero A, Iriondo JM & Torres ME (2003) Spatial analysis of genetic diversity as a tool for plant conservation. Biol Conserv 113: 351-365.

Falush D, Stephens M & Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Freckleton RP & Watkinson AR (2003) Are all plant populations metapopulations? J Ecol 91: 321-324.

Futuyma DJ (1998) Evolutionary biology, 3th edition. Sinauer Associates, Inc, USA.

Gao L, Chen W, Jiang W, Ge S, Hong D & Wang X (2000) Genetic erosion in northern marginal population of the common wild rice *Oryza rufipogon* Griff. and its conservation, revealed by the change of population genetic structure. Hereditas 133: 47-53.

Gaudeul M, Taberlet P & Till-Bottraud I (2000) Genetic diversity in an endangered alpine plant, *Eryngium alpinum*, L. (Apiacae), inferred from amplified fragment length polymorphism markers. Mol Ecol 9: 1625-1637.

Gehring JL & Delph LJ (1999) Fine scale genetic structure and clinal variation in *Silene acaulis* despite high gene flow. Heredity 82: 628-637.

Gerber S, Chabrier P & Kremer A (2003) FaMoz: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. Mol Ecol Notes 3: 479-481.

Giles BE & Goudet J (1997) Genetic differentiation in *Silene dioica* metapopulations: estimation of spatiotemporal effects in a successional plant species. Am Nat 149: 507-526.

Giles BE, Lundqvist E & Goudet J (1998) Restricted gene flow and subpopulation differentiation in *Silene dioica*. Heredity 80: 715-723.

Goldstein DB & Schlötterer C (1999) Microsatellites evolution and applications. Oxford University Press, Inc, New York.

Graur & Li (2000) Fundamentals of molecular evolution. Sinauer Associates, Inc, USA.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95-98.

Hammond RL, Saccheri IJ, Ciofi C, Coote T, Funk SM, McMillan WO, Bayes MK, Taylor E & Bruford MW (1998) Isolation of microsatellite markers in animals In: Karp A, Isaag PG & Ingram DS (eds) Molecular tools for screening biodiversity plant and animals. Chapman & Hall, London, p 279-285.

Hanski I (1999) Metapopulation ecology. Oxford University Press, Oxford.

Hanski I & Gaggiotti OE (2004) Metapopulation biology: past present, and future In: Hanski I & Gaggiotti OE (eds) Ecology, genetics and evolution of metapopulations. Elsevier Academic Press, USA, p 3-22.

Hardy OJ (2003) Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. Mol Ecol 12: 1577-1588.

Hardy OJ & Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population level. Mol Ecol Notes 2: 618-620.

Harr B & Schlötterer C (2000) Long microsatellites alleles in Drosophila melanogaster have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. Genetics 155: 1213-1220.

Harr B, Zangerl B & Schlötterer C (2000) Removal of microsatellite interruptions by DNA replication slippage: phylogenetic evidence from Drosophila. Mol Biol Evol 17: 1001-1009.

Harrison S & Hastings A (1996) Genetic and evolutionary consequences of metapopulation structure. Trends Ecol Evol 11: 180-183.

Hartl DL & Clark AG (1997) Principles of population genetics, 3th edition. Sinauer Associates, Inc, USA.

Haubold B & Hudson RR (2000) Lian, version 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics 16: 847-848.

Holsinger KE & Lewis PO (2002) Hickory: A package for analysis of population genetic data v1.0. http://darwin.eeb.uconn.edu/hickory/documentation/Hickory.pdf

Husband BC & Barrett SCH (1996) A metapopulation perspective in plant population biology. J Ecol 84: 461-469.

Jones AG & Arden WR (2003) Methods of parentage analysis in natural populations. Mol Ecol 12: 2511-2523.

Jäkäläniemi A, Kauppi A, Pramila A & Vähätaini K (2004) Survival strategies of *Silene tatarica* (Caryophyllaceae) in riparian and ruderal habitats. Can J Bot 82: 491-502.

Jäkäläniemi A, Tuomi J, Siikamäki P & Kilpiä A (2005) Colonization-extinction and patch dynamics of the perennial riparian plant, *Silene tatarica*. J Ecol, in press.

Kalisz S, Nason JD, Hanzawa FM & Tonsor SJ (2001) Spatial population genetic structure in *Trillium grandiflorum*: the roles of dispersal, mating, history and selection. Evolution 55: 1560-1568.

Karhu A, Dieterich JH & Savolainen O (2000) Rapid expansion of microsatellite sequences in pines. Mol Biol Evol 17: 259-265.

Keiper FJ & McConchie R (2000) An analysis of genetic variation in natural populations of *Sticherus flabellatus* [R. Br. (St John)] using amplified fragment length polymorphism (AFLP) markers. Mol Ecol 9: 572-581.

Krauss SL & Peakall R (1998) An Evaluation of the AFLP Fingerprinting Technique for the Analysis of Paternity in Natural Populations of *Persoonia mollis* (Proteaceae). Aust J Bot 46: 533-546.

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. Proc Natl Acad Sci U S A. 95: 10774-10778.

Kruglyak S, Durrett RT, Schug MD & Aquadro CF (2000) Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. Mol Biol Evol 17: 1210-1219.

Kumar A & Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33: 479-532.

Levin SA, Muller-Landau HC, Nathan R & Chave J (2003) The ecology and evolution of seed dispersal: a theoretical perspective. Annu Rev Ecol Evol Syst 34: 575-604.

Linhard YB & Grant MC (1996) Evolutinary significance of local genetic differentiation in plants. Ann Rev Ecol Syst 27: 237-277.

Luikart G, Allendorf FW, Cornuet J-M & Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered 89: 238-247.

Manel S, Schwartz MK, Luikart G & Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18: 189-196.

Manel S, Gaggiotti OE & Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. Trends Ecol Evol 20: 136-142

Mariette S, Le Corre V, Austerlitz F & Kremer A (2002) Sampling within the genome for measuring within-population diversity: trade-offs between markers. Mol Ecol 11: 1145-1156.

Marttila S (2004) Tataarikohokin (*Silene tatarica*) taimettumis- ja kolonisaatiodynamiikka häiriönalaisilla jokirannoilla. Masters thesis, Department of biology, University of Oulu. [In Finnish]

McCauley DE (1995) The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends Ecol Evol 10: 198-202.

McCauley DE (1997) The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. J Hered 88: 257-263.

McVean GAT (2002) A genealogical interpretation of linkage disequilibrium. Genetics 162: 987-991.

Miyashita NT, Kawabe A & Innan H (1999) DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment level polymorphism analysis. Genetics 152: 1723-1731.

Miyazaki Y & Isagi Y (2000) Pollen flow and the intrapopulation genetic structure of *Heloniopsis orientalis* on the forest floor as determined using mikrosatellite markers. Theor Appl Genet 101: 718-723.

Morgante M, Hanafey M & Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30: 194-200.

Nadir E, Margalit H, Gallily T & Ben-Sasson SA (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proc Natl Acad Sci U S A 93: 6470-6475.

Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. Annu Rev Ecol Syst 28: 105-128.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 13: 1143-1155.

van Oosterhout C, Hutchinson WF, Wills DPM & Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4: 535-538.

Ouborg NJ & Eriksson (2004) Toward a metapopulation concept for plants. In: Hanski I & Gaggiotti OE (eds) Ecology, genetics and evolution of metapopulations. Elsevier Academic Press, USA, p 447-469.

Orita M, Owahana H, Kanazawa H Hayashi K & Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci U S A 86: 2766-2770.

Pamilo P & Savolainen O (1999) Post-glacial colonization, drift, local selection and conservation value of populations: a northern perspective. Hereditas 130: 229-238.

Pannell JR & Charlesworth B (2000) Effects of metapopulation processes on measures of genetic diversity. Philos Trans R Soc Lond B Biol Sci 355: 1851-1864.

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

Provan J, Powell W & Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16: 142-147.

Ramsay L, Macaulay M, Cardle L, Morgante M, Ivanissevich S, Maestri E, Powell W & Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J 17: 415-425.

Richards CM, Church S & McCauley DE (1999) The influence of population size and isolation on gene flow by pollen in *Silene alba*. Evolution 53: 63-73.

Riley DE & Krieger JN (2004) Simple repeat replacements support similar functions of distinct repeat in inter-species mRNA homologs. Gene 328: 17-24.

Riley DE & Krieger JN (2005) Short tandem repeat (STR) replacements in UTRs and introns suggest an important role for certain STRs in gene expression and disease. Gene 344: 203-211.

Rogers SO & Bendich AJ (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Mol Biol 5: 69-76.

Rousset F (2001) Inferences from spatial population genetics In: Balding DJ, Bishop M & Canning C (eds) Handbook of statistical genetics. John Wiley & Sons. Ltd, England, p 239-269.

Sainudiin R, Durrett RT, Aquadro CF & Nielsen R (2004) Microsatellite mutation models: insight from a comparison of humans and chimpanzees. Genetics 168: 383-395.

Schlötterer C (1998) Microsatellites In: Hoelzel AR (ed) Molecular genetic analysis of populations, a practical approach. Oxford University Press Inc, New York, p 237-261.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365-371.

Schlötterer C (2002) Towards a molecular characterization of adaptation in local populations. Curr Opin Genet Dev 12: 683-687.

Schlötterer C (2004) The evolution of molecular markers – just a matter of fashion? Nat Rev Genet 5: 63-69.

Schneider S, Roessli D & Excoffier L (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Silvertown J & Charlesworth D (2001) Introduction to plant population biology, 4th edition. Blackwell Science, USA.

Sun M & Ritland K (1998) Mating system of yellow starthistle (*Centaurea solstitialis*), a successful colonizer in North America. Heredity 80: 225-232.

Slatkin M (1985) Gene flow in natural populations. Annu Rev Ecol Syst 16: 393-430.

Taberlet P, Gielly L Pautou G & Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17: 1105-1109.

Taberlet P, Fumagalli L, Wust-Saucy A-G & Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol 7: 453-464.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17: 6463-6471.

Temnykh S, DeClerck G, Lukashove A, Lipovich L, Cartenhour S & McCouch S (2001) Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11: 1441-1452.

Thrall PH, Burdon JJ & Murray BR (2000) The metapopulation paradigm: a fragmented view of conservation biology. In: Young AG & Clarke GM (eds) Genetics, demography and viability of fragmented populations. Cambridge University Press, Cambridge, UK, p 75-95.

Travis SE, Manchinski J & Keim P (1996) An analysis of genetic variation in *Astragalus cremnophylax* var. *cremnophylax*, a critically endangered plant, using AFLP markers. Mol Ecol 5: 735-745.

Ulvinen T (1997) Tataarikohokki – ryssglim. In: T. Ryttäri & T. Kettunen (eds) Uhanalaiset kasvimme. Suomen Ympäristökeskus and Kirjayhtymä Oy, Helsinki, p 246.

Vekemans X & Hardy OJ (2004) New insights from fine-scale spatial genetic structure analysis in plant populations. Mol Ecol 13: 921-935.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith SC & Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol 19: 1251-1260.

Vos P, Hogers R, Bleeker M, Reijans SM, Vandelee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M & Zabeau M (1995) AFLP: A new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407-4414.

Weising K & Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42: 9-19.

Wessler SR, Bureau TE & White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5: 814-821.

Westerbergh A & Saura A (1994) Gene flow and pollinator behaviour in *Silene dioica* populations. Oikos 71: 215-224.

Whitlock MC & McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. Evolution 44: 1717-1724.

Whittaker JC, Harbord RM, Boxall N, Mackay I Dawson G & Sibly RM (2003) Likelihood-based estimation of microsatellite mutation rates. Genetics 164: 781-787.

Wilder J & Hollocher H (2001) Mobile elements and the genesis of microsatellites in Dipterans. Mol Biol Evol 18: 384-392.

Xu X, Peng M, Fang, Z & Xu X (2000) The direction of microsatellite mutations is dependent upon allele length. Nat Genet 24: 396-399.

Yeh F, Yang RC & Boyle T (1995) POPGENE, the User-Friendly Shareware for Population Genetic Analysis. Molecular Biology and Biotechnology Center, Universty of Alberta, Edmonton.