

Microsatellite markers as a tool for the detection of intra- and interpopulational genetic structure

Ivan Scotti*, Gianpaolo Paglia, Federica Magni, Michele Morgante

Dipartimento di Produzione Vegetale e Tecnologie Agrarie, Università degli Studi di Udine,
Via delle Scienze 208, 33100 Udine, Italy

*Corresponding author: Email: ivan.scotti@dpvta.uniud.it

Introduction

The microsatellites (or SSRs, Simple Sequence Repeats) have proven to be very useful for the purpose of unveiling genetic diversity in forest tree species. In this paper our experience in the application of microsatellites is described, with the aim of providing a guide for the transfer of this technique to the end-users.

The variability at microsatellite loci is due to the differences in the number of repeat units. It is therefore easily detectable as variation in length of a DNA fragment obtained through a PCR amplification from the whole genome.

Since their discovery and establishment as a molecular technique, the SSRs have been applied to several problems in which the fingerprinting of genetic diversity and differentiation was needed. The detection of within-population diversity and structure is made easy by the high polymorphism of this kind of markers, that allows to identify single trees even analysing few loci. For the same reasons, this technique can be applied to paternity analysis. The latter can also take advantage of the availability of uniparentally inherited SSR markers, especially in conifers, where the transmission of the chloroplast is strictly paternal (Vendramin *et al.*, 1996) and the mitochondrion is maternal, at least in Norway spruce (Sperisen *et al.*, 1999, this Compendium). The microsatellites have advantages in among-population differentiation studies as well. Forest tree species generally display a low level of among-population divergence. This is confirmed by analyses carried out using microsatellites. However, in this case the significance of the pairwise differentiation values is somehow enhanced. This represents an advantage compared to other marker classes such as isozymes. Data from SSR markers can be used either as single loci (if they are unlinked) or as haplotypes (if they are linked, as is the case with plastidial markers). Moreover, it must be highlighted that microsatellites convey an extra amount of information, compared to other classes of markers, thanks to the underlying mutational model (Stepwise Mutation Model), and that they often carry high numbers of alleles at very low frequencies or "private" alleles, that is, alleles present in only one or few populations. This greatly contributes to the assessment of the genetic relationships among populations and among individual trees. As an example, in our survey on natural populations of Norway spruce (*Picea abies* Karst), we found in two loci that no allele had a frequency higher than 0.10 over a sample of more than 1300 individuals.

Methods and results: Evaluation of the markers and of their cost-benefit balance

The SSR markers have characteristics that can be very useful for the end-user. Not only are they highly variable, as discussed above. They are generally recognised as neutral, so that selection and environmental pressure do not influence their expression directly. In Norway spruce, we have found that only 2 loci out of 45 departed from Mendelian segregation, when used for the analysis of the progeny of a controlled cross (unpublished data). Moreover, the SSRs generally are codominant markers, so that all alleles can be scored. Another advantage is that the gel runs can be multiplexed, that is, PCR products of different markers can be run on the same gel, thus saving time, labour and money. Up to four markers can be run in the same lane on a standard sequencing gel. Once the data have been scored, they can easily be compared to results obtained in other laboratories, provided that the method used for the gel run is the same, and that at least one sample is used as a standard shared among laboratories. This is because different detection techniques (e.g. acrylamide gels with radio-labelling, or automated sequencers of different classes) imply different estimates of the size of DNA fragments. Nevertheless, protocols are easy to exchange across groups since they are rather insensitive to changes in experimental conditions. The same pattern could be obtained in our lab over a range of magnesium chloride concentration from 1.5 to 2.5 mM and with DNA amounts ranging from 100 ng to 1.5 ng, in a reaction volume spanning from 10 to 50 µL. For the establishment of the technique in a new lab, no transfer of material is needed, since the primers for the PCR amplification can be synthesised starting from the sequence information.

Some drawbacks must also be noted in the interpretation of SSR amplification patterns. Stutter bands can appear along with the band corresponding to the expected fragment. This will cause (i) uncertainty in the estimation of the allele size and (ii) the possibility of mistaking a heterozygote for a homozygote, if the two bands are so close on the gel that the ladders produced by the two alleles overlap. The former problem can be easily overcome by running an internal standard (an allele of known size), plus a molecular size marker, along with the test samples. The standard will help to identify the “true” band among those amplified in the PCR. Generally speaking, the introduction of internal size standards is a very efficient strategy in microsatellite gel runs, because they can help in all cases in which the scoring of fragment sizes is uncertain.

One further problem that needs to be taken into account when using SSR loci is the presence of *null* alleles. These alleles are not amplified and therefore are not scored at all on the gels. They can therefore lead to underestimates of heterozygosity. This problem has no direct troubleshooting, since there is no direct way to turn these (partially) dominant loci into codominant, and can have consistent effects. In Norway spruce, we have computed, based on segregation data, that 20% of all the SSR alleles were *null* in a controlled cross (unpublished data). This high proportion of loci showing dominance as mode of gene action contradicts the commonly held notion that codominance is the usual mode of gene action at nuclear microsatellite loci.

Equipment and costs

Among the positive characteristics of the SSR markers is the need for only basic molecular biology equipment. The facilities needed to perform these experiments are a PCR thermal cycler, an electrophoresis cell for vertical acrylamide gels and the equipment for film/gel development (including a dark room if the protocol used includes radioactive labelling; see Pfeiffer *et al.* (1997) for a protocol of this kind). These facilities can be purchased for approximately 12000 EURO.

The PCR reaction for the SSR amplification is a rather robust one and in general does not require highly purified DNA. Moreover, we have found that the PCR reaction can be scaled down, saving

money on the reagents and allowing to get a strong signal on the gel even with very small amounts of template genomic DNA (down to 1.5 ng).

We have computed that, from DNA extraction to gel scoring, it is possible to genotype one individual with one marker at the cost of no more than 1.5 EURO.

As for the time investment, generally available tools can help speed up all the procedures. The use of a multi-channel pipettor allows to set up a 96-sample PCR reaction in 20 minutes; if the gel is also loaded using the same tool, it allows to load 96 samples in less than 10 minutes. If the gel is used to run three markers at the same time, one gel is loaded in around 50 minutes (allowing for a 10 minute gel run between loadings). This means that one operator can easily process close to 300 samples a day, leaving time for data scoring, just with basic equipment.

Technical skills and background necessary to perform SSRs

Basic knowledge in molecular biology is required to set up an experiment involving the use of SSRs. The only techniques involved are PCR and gel run. Methods for performing PCRs and running acrylamide gels can be found in Maniatis *et al.* (1989) and in Newton and Graham (1997), and can be learnt quickly by any person having a background in biology with a training period of no more than three months. This is clearly proved by our experience with undergraduate students. They come to the lab only with theoretical background in biology and genetics and become able to run experiments with markers after few weeks of teaching. After a few months the students are able to manage the SSR characterisation on their own.

More effort is needed for the analysis of the data gathered from the experiments. For this step a sound background in statistics is required. Although several software packages are available for the analysis of data, it is nevertheless important that the experimenter knows what the software outputs mean and how they should be interpreted. This situation is common to all kinds of data analyses, of course. In the case of SSR markers, however, an improper use of the data can be particularly misleading, due to the very high information output that can be obtained from this class of loci.

Data analysis: availability of tools

Several software packages are available free of charge for the analysis of population microsatellite data. Most of these can be downloaded from web sites or via FTP. Examples of very useful tools in our experience are the following programmes: ARLEQUIN (Schneider *et al.*, 1997; download at <http://anthropologie.unige.ch/arlequin/>), RSTCALC (Goodman 1997; download at <http://helios.bto.ed.ac.uk/evolgen/>), GENEPOP (Raymond & Rousset, 1995; download by FTP at <ftp.cefe.cnrs-mop.fr>), GSED (Gillet, 1998; download at <http://www.uni-forst.gwdg.de/forst/fg/index.htm>). In general the programmes mentioned above can deal with huge datasets, although some of them have shown problems when tested with our data set of 1,300 individuals genotyped for 8 microsatellite loci (with a maximum of 62 alleles at a locus), summing up to 20,800 data points.

How to get microsatellites

To identify microsatellite markers from scratch is not as easy as using them. The isolation and characterisation of microsatellite regions, their sequencing and the testing of primers can be time consuming and expensive and requires more detailed knowledge of molecular biology and genetics. As far as the end-user is concerned, it is not economically advantageous to try to develop new markers, unless the required technology and expertise are already at hand. The most efficient

strategies to get new microsatellites are (i) to search the literature and the several web sites that carry information on the available markers, (ii) to design new primers on the sequences available in public databases, and (iii) to have the markers developed by a research lab with experience in this field. While the first option is quite straightforward, the second requires some experience in primer design and includes primer testing and optimisation. The third, on the other hand, usually completely relieves the end-user from the burden of marker development and offers the possibility of co-operation and exchange of information between parties (the end-user and the research lab) with different aims and interests.

Case studies

Several reports can be quoted as examples of the application of SSR markers to the analysis of within-population structure and between-population differentiation. High efficiency in the characterisation of diversity has been shown for nuclear microsatellites (*e.g.* La Scala *et al.*, 1999, this Compendium) and for chloroplast microsatellites (*e.g.* Anzidei *et al.*, 1999, this Compendium). Variability has also been reported for mitochondrial microsatellite loci (Soranzo *et al.*, 1999).

Both conifer and angiosperm tree species have been investigated using microsatellite markers. Among the broadleaf trees, a clear example of the usefulness of SSRs for the analysis of natural diversity is given in Streiff *et al.* (1998) as well as in White and Powell (1998). Paternity analysis is demonstrated by Lexer *et al.* (1997). In conifers, the applications of chloroplast microsatellites has been fully developed, as is shown in Echt *et al.* (1998), Vendramin and Ziegenhagen (1997), and Anzidei *et al.* (1999, this Compendium). A first set of nuclear SSR markers was developed by Pfeiffer *et al.* (1997), followed by an extended set developed within the frame of this project both from cDNA clones (Scotti *et al.*, in press) and from genomic libraries enriched for dinucleotide and trinucleotide repeats (Scotti I, Magni F, Paglia GP, Zuccolo A, Felice N, Morgante M, in prep.). The performances of nuclear dinucleotide, nuclear trinucleotide and chloroplast SSRs have also been compared in a test of among-population differentiation (Scotti I, Magni F, Paglia GP, Morgante M, in prep.).

Also gene flow studies have been carried out successfully using microsatellites. This is shown in Dow and Ashley (1996) and in Streiff *et al.* (1999).

Conclusion

In this paper we have shown that the SSR technique is well established in the analysis of diversity and differentiation in forest tree species and that several methods and tools are available for this purpose. We want to highlight that to set up an experiment for the analysis of microsatellite loci only requires minimal efforts and limited technological investment, and that a large set of data can quickly be gathered using SSR markers. All these characteristics make the SSRs the markers of choice for all those that are interested in individual identification, paternity analysis, and in the assessment of differences among populations, either natural or artificial. A positive exchange of information and requests between end-users and research institutions can quickly lead to the development of SSR markers for specific purposes and of general interest for the research in genetic diversity.

References

Anzidei M, Madaghiale A, Sperisen C, Ziegenhagen B, Vendramin GG (1999) Chloroplast microsatellites for the analysis of diversity in forest tree species. Chapter 10 in this Compendium: Gillet EM (ed.). *Which DNA Marker for Which Purpose?* Final Compendium of the Research

Project *Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees* in the European Union DGXII Biotechnology FW IV Research Programme *Molecular Tools for Biodiversity*. URL

<http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>

Dow BD, Ashley MV (1996). Microsatellite analysis of seed dispersal and parentage of saplings in bur oak *Quercus macrocarpa*. *Molecular Ecology* 5: 615-627.

Echt CS, DeVerno LL, Anzidei M, Vendramin GG (1998) Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Molecular Ecology* 7(3): 307-316.

Gillet, E. (1998) GSED - Genetic Structures from Electrophoresis Data. Institut für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen.

Goodman SJ (1997) RST Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* 6(9): 881-885.

La Scala S, Schubert R, Müller-Starck G, Liepe K (1999) Nuclear microsatellites as a tool in the genetic characterization of forest reproductive material. A case study in sessile oak (*Quercus petraea* Matt., Liebl.). Chapter 7 in this Compendium: Gillet EM (ed.). *Which DNA Marker for Which Purpose?* Final Compendium of the Research Project *Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees* in the European Union DGXII Biotechnology FW IV Research Programme *Molecular Tools for Biodiversity*. URL <http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>

Lexer C, Streiff R, Steinkellner H, Glössl J (1997) Paternity tests for trees with micro-satellites. [German] *Oesterreichische Forstzeitung* 108(6): 43-44.

Maniatis T, Sambrook J, Fritsch EF (1989) *Molecular Cloning : A Laboratory Manual*. Cold Spring Harbor Laboratory Press.

Newton CR, Graham A (1997) *PCR (Introduction to Biotechniques Series)*. Springer Verlag, New York.

Pfeiffer A, Olivieri AM, Morgante M (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome* 40(4): 411-419.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86: 248-249.

Schneider S, Kueffer J-M, Roessli D, Excoffier L (1997) Arlequin ver. 1.1: A software for population genetic data analysis. Genetics and Biometry Laboratory. University of Geneva, Switzerland.

Scotti I, Magni F, Fink R, Powell W, Binelli G, Hedley PE (1999) Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. *Genome*, in press.

Soranzo N, Provan J, Powell W (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. *Genome* 42(1): 158-161.

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. Chapter 9 in this Compendium: Gillet EM (ed.). *Which DNA Marker for Which Purpose?* Final Compendium of the Research Project *Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees* in the European Union DGXII Biotechnology FW IV Research Programme *Molecular Tools for Biodiversity*. URL <http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>

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

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(3): 317-328.

Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Molecular Ecology* 5(4): 595-598.

Vendramin GG, Ziegenhagen B (1997) Characterisation and inheritance of polymorphic plastid microsatellites in *Abies*. *Genome* 40(6): 857-864.

White G, Powell W (1997) Cross-species amplification of SSR loci in the *Meliaceae* family. *Molecular Ecology* 6(12): 1195-1197.