Application of selected molecular markers in studies on forest trees

Anna Tereba ✉, *Agata Konecka*, *Justyna A. Nowakowska*

Forest Research Institute, Laboratory of Molecular Biology, Sękocin Stary, Braci Leśnej 3, 05-090 Raszyn, Poland,
e-mail: A.Tereba@ibles.waw.pl

ABSTRACT

The paper describes a number of molecular methods used in the past and now to analyze forest tree species. Taking into account the economic importance of forest trees and in view of the timber economy, wood properties and characteristics are essential factors subjected to control, observation and research. Molecular techniques that support traditional selection methods allow for genetic diversity analyses considering a range of research aspects. The development of these techniques at the turn of the last two decades has enabled wide-ranging use of molecular data in studies on forest tree populations. On the example of pine (*Pinus* L.), the paper presents data based on molecular studies as well as a variety of possibilities to apply the obtained results.

KEY WORDS

INTRODUCTION

Numerous forest tree species, and especially those coniferous, are of great economic importance, and consequently, attempts are made to improve wood properties that are significant with regard to forest management, such as: wood quality, pest resistance, adaptability to climatic conditions, etc. However, the duration of the generational change, as well as difficulties in evaluating traits in the early stages of seedling development, complicate and slow down the progress in selection of beneficial breeding traits. In order to support traditional selection methods, molecular techniques based on DNA analysis have been used for the last 20 years to test genetic and phenotypic diversity. These involve, among others, analyzing microsatellite markers DNA (single

sequence repeat – SSR) in view of population variability or analyzing single nucleotide polymorphisms (SNPs) in the coding DNA fragments. Consequently, this research field progression seems to be very promising.

MOLECULAR TECHNIQUES – EARLIER AND TODAY

The advancement of molecular methods used in genetic research reflects well the use of ever more precise techniques. Restriction Fragment Length Polymorphism (RFLP), where the analysis is based on detecting diversity using a restriction enzyme that cuts a sequence of DNA of known order of about 4 to 6 base pairs, was used already in 1988 for cystic fibrosis diagnostics (Williams et al. 1988). Another easy to use method of DNA

polymorphism analysis is Random Amplified Polymorphic DNA (RAPD), which randomly amplifies fragments of the DNA template (Williams et al. 1990). One development of the RFLP method is Amplified Fragment Length Polymorphism (AFLP), described in 1995 by Vos et al., based on selectively amplified restriction fragments. Currently, one of the most popular tools in population-based molecular research is the analysis of SSR markers (Simple Sequence Repeats), which are short tandem repeats (2 to 6 nucleotides) characterized by fast mutation rates (Avise 2004).

The progressive development of molecular methods and an increasing availability of databases containing DNA sequences allow for analyses of polymorphisms in coding and non-coding DNA fragments. Comprehensive genomic library databases allow to apply more precise molecular techniques, for example the detection and analysis of single-nucleotide polymorphisms (SNPs), accessibility of which is better than, for example, tandem repeat SSR markers. In addition, databases constitute a rich information source for molecular and breeding applications as they can provide data to analyze genotype – phenotype relationships.

The development and validation of new SNP molecular markers for forest tree species is not only an indispensable element in continuing studies on molecular variability, but will also make possible future analyses on the direct relationship between molecular polymorphism and phenotypic traits in individual trees as well as tree stands.

FOREST MANAGEMENT ON THE EXAMPLE OF SCOTS PINE AS A DOMINANT SPECIES

Contemporary forest species composition is largely attributable to human activities. Scots pine (*Pinus silvestris*), due to good wood parameters, rapid growth (especially in the early development stages) and low soil requirements, is the dominant species in Poland's forests, and represents about 58.5% share in the country's forest stands (http://www.lasy.gov. Pl-2015). Apart from its wood value for the timber industry, pine essential oils are important ingredients used in the perfumery and cosmetic industries. Furthermore, pine trees constitute appreciated source of resins, although their quantity and quality varies a lot, depending on pine species.

Besides, pine seeds, e.g. stone pine (*Pinus pinea*) nuts, are utilized in the food industry. Some pine species, on account of their high resistance to extreme climatic conditions or high soil salinity, are much useful in biotechnical plantings, such as dune strengthening or land reclamation.

Forest tree populations are characterized by a substantial longevity as well as they exhibit, to a lesser or greater degree, adaptability to changing environmental conditions This concerns in particular pine species, due to their high tolerance for adverse environmental conditions. Local populations of *Pinus* species adapt to prevailing conditions, which translates into their large phenotypic diversity. The emergence of new populations (ecotypes) is associated with development of different phenotypic features (morphological, physiological and ecological) and affects great pine variability observed in the environment (Jaworski 2011).

In Europe, studies on pine phenotypic diversity have been conducted since the beginning of the 20th century (Giertych 1979). Research on extremely important issues, such as high adaptability, and also – natural phenotypic pine diversity which leads to the formation of local ecotypes, are still central to the concerns of many researchers. Pine populations described in subject literature are characterized by distinctive and specific morphological features, and thus, they are principally valuable not only from the practical point of view but also – for scientific knowledge. Modern research methods based on molecular biology techniques enable us to study variability of local populations at a DNA sequence level. Since biodiversity is one of the most important factors for ecosystem sustainability, having methods to monitor biodiversity is essential for all research and further conclusions.

SSR AND SNP MOLECULAR MARKERS – ADVANTAGES AND DISADVANTAGES

Microsatellite markers are widely used in the analyses of population genetic structure. Polymorphism of these markers is a result of differences in the length of tandem repeats located in the genomes. When compared to other plant species, relatively not many microsatellite markers have been characterized for the pine family, as the major obstacle are the size of the genome

[from 19 to 31 pg (Murray 1998)] and the fact that pine genomes are largely (75%- 86%) composed of duplicated fragments (Elsik and Williams 2000). There was also suggested non-homologous recombination (Nystedt et al. 2013). The long history of coniferous species divergence in combination with high genome complexity denote low efficiency of cross-species transfer of SSR markers occurring in the genome as single copies, which significantly limits the possibility of the use of the cross-species amplification technique. However the efficiency of this technics is low not only between different coniferous species, but also within pine species (Chagné et al. 2004). Problems with obtaining genotyping results for microsatellite genetic markers could lead to lack of PCR products or non-specific products, or else the lack of polymorphism. In addition, the results of analyses based on microsatellite markers should be considered carefully in view of the likelihood of homoplasia (appearance of identical alleles of different evolutionary origin). Reversed situation is also possible, when different alleles are observed for identical haplotypes, for example as a result of mutations in flanking regions (Rafalski 2002). The analysis of SSR markers may be affected by errors due to stutter bands seen in PCR amplification (Muncy et al. 2009; Ellegren 2004). Furthermore, null alleles (not detectable by genotyping; homozygote excess observed) cannot be ignored on account of their up to 35% frequency (Moriguchi et al. 2003). During the amplification reaction, there can also occur the phenomenon of the so-called slippage that leads to misinterpretation of alleles in further analyzes. This error can be minimized by the use of longer tandem repeats (for example 4 nucleotides). Unfortunately, for conifers only a few markers of this type have been so far identified (Guan et al. 2011). All the above described events can lead to erroneous genotype readings when SSR marker analysis is performed.

Single-nucleotide polymorphism (SNP) markers might constitute an alternative to microsatellite markers. SNP analysis requires no electrophoretic separation, can be automated, e.g. through using DNA-microarrays or sets for the detection of polymorphisms by sequencing or genotyping techniques. SNP markers are characterized by a low error rate of readout in laboratory procedures (Weller et al. 2006), are easy to map in the genome, as well as show high stability and easiness of standardization (Werner et al. 2004). On the other hand, a certain limitation of SNP markers is their bi-allelic nature (possible linked mutations of two different alleles in the same gene locus), which is associated with low expected heterozygosity of alleles in a given locus. However, when analyzing a large number of closely spaced SNP markers, a fully defined haplotype can be expected. Furthermore, e.g. in maize, there have been identified many SNP markers flanking the microsatellite regions, which allowed the direct use of these markers as substitutes for microsatellite markers (Edwards and Mogg 2001). Besides, with a large number of SNP markers it is possible to identified genotypes with a high degree of discrimination, even higher than for microsatellite loci. The polymorphism of SNP markers is currently considered a new trend in the analysis of genetic diversity in population studies. This is associated with the factors, such as: high frequency of SNPs in the genome, the better known and described mutation mechanism when compared to SSRs, low error rates and easiness of obtaining large numbers of markers in the analysis, and – additionally – a possibility to combine the results obtained with specific gene functionality, which is not possible in the analysis of random SSR or AFLP markers (Ryynänen et al. 2007).

SNP's diversity of *Pinus* sp.

Identification of new molecular markers with high accuracy while maintaining a low error value is a key goal in the development of molecular methods. Analysis of selected SNP markers allows the investigation of genetic diversity and verification of introgression levels within pine species: lodgepole pine *Pinus contorta* and Jack pine *Pinus banksiana* (Cullingham et al. 2013). Study of the genetic structure and demographic processes in forest tree populations are carried out based on SNP markers (Wachowiak et al. 2011; Wachowiak 2015) as alternatives to other types of genetic markers, e.g. SSR, cpSSR (chloroplast simple sequence repeat) or allozyme (Suharyanto et al. 2012).

Another extremely important subject matter concerning the use of molecular data is an assessment of a degree of kinship. Many works based on SSR markers refer to progeny genetic variability, an extent of self-pollination, or the percentage of pollen from areas situated outside a given plantation (Slavov et al.

2005; Hansen and Kjaer 2006). The use of single nucleotide polymorphisms (SNPs) in studies on kinship is an additional possibility. For example, just 18 SNP polymorphisms are enough for verification of parental specimens in *Pinus thunbergii* (Suharyanto et al. 2012). For such a number of markers, the probability of occurrence of the identical genotype in unrelated individuals is 3.40×10^{-8}, and the cumulative power of discrimination is $1 - 3.40 \times 10^{-8}$. Taking into consideration such a high level of discriminatory power, Suharyanto et al. (2012) point out the usefulness of the developed panel of SNP markers in the creation of the genetic base for reproductive material.

The use of SNP markers has not been merely limited to traditional molecular techniques. Next-generation sequencing (NGS) technique in fact allows for analyzing an unlimited number of single nucleotide polymorphisms. By use of the Illumina platform, it was possible to create a genotyping chip containing 7252 SNP, the largest of the so far prepared ones for *Pinus pinaster* (Plomion et al. 2016). Analyses with the use of third-generation sequencing technology have been successfully performed also for other coniferous species (Pavy et al. 2008; Eckert et al. 2009; Chancerel et al. 2013; Liu et al. 2014). Data obtained by this method have wide-ranging applications: mapping genomes, identifying seeds with their origins, comparing natural and bred populations, investigating genetic diversity with respect to phenotypic diversity, concluding on historical processes, demographic events as well as selection and adaptation. Notwithstanding such impressive data size, in confrontation with the size of coniferous genome – 24.5G in maritime pine *Pinus pinaster* (Chagne et al. 2002) – it should be noted that the obtained SNP number (Plomion et al. 2016) only coincides with a small fraction of potential genetic diversity. In spite of the need for attaining more and more data, it has been found, however, that a few thousand of appropriately selected SNPs is as much as necessary to draw a reliable conclusion about polygenic patterns of adaptation process (Berg and Coop 2014).

As described above, SNP genetic markers are much comprehensive. They can be analyzed using a variety of molecular biology methods, from single mutation analysis to sequencing of genome fragments. The scope of use of the obtained data is exceptionally broad, and combine far and wide various genetic aspects.

CONCLUSION

Undoubtedly, single nucleotide polymorphism (SNP) is a fundamental and inexhaustible source of molecular variability. Rapidly supplemented libraries of gene sequences and whole genomes provide a rich base for typing polymorphic markers for many species, including forest trees. Additionally, the development of technologies for molecular analyses provides the prospect for the flourishing and automation molecular analysis techniques at increasingly higher levels of precision. This allows to accelerate significantly the development of a new molecular analysis course (Feder and Mitchell-Olds 2003), which is the analysis of SNP polymorphisms. In the near future, sequencing methods will probably be developed to allow direct reading of DNA sequences, eliminating many methodological problems. The studies based on sequence capture and direct sequencing seem to be the inevitable future of molecular analyses, creating the opportunity to generate vast amounts of data and learn about numerous issues related to the functioning of trees in forest ecosystems.

REFERENCES

Avise J.C. 1994. Molecular Markers, Natural History, and Evolution. Chapman & Hall, New York.

Berg J.J., Coop G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genetics*, 10 (8), e1004412.

Chagne D., Lalanne C., Madur D., Kumar S., Frigérioa J.-M., Kriera C., Decroocqa S., Savouréc A., Bou-Dagher-Kharratc M., Bertocchia E., Bracha J., Plomion C. 2002. A high density linkage map of *Pinus pinaster* based on AFLPs. *Annals of Forest Science*, 59, 627–636.

Chagné D., Chaumeil P., Ramboer A., Collada C., Guevara A., Cervera M.T., Vendramin G.G., Garcia V., Frigerio J.-M., Echt C., Richardson T., Plomion C. 2004. Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theoretical and Applied Genetics*, 109 (6), 1204–1214.

Chancerel E., Lamy J.B., Lesur I., Noirot C., Klopp C., Ehrenmann F., Boury C., Le Provost G., Label P., Lalanne C., Léger V., Salin F., Gion J.-M., Plomion C. 2013. High density linkage mapping in a pine

tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. *BMC Biology*, 11, 50.

Cullingham C.I., Cooke J.E.K., Dang S., Coltman D.W. 2013. A species-diagnostic SNP panel for discriminating lodgepole pine, jack pine, and their interspecific hybrids. *Tree Genetics and Genomes*, 9 (4), 1119–1127.

Eckert A.J., Pande B., Ersoz E.S., Wright M.H., Rashbrook V.K., Nicolet C.M, Neale D.B. 2009. High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). *Tree Genetics and Genomes*, 5, 225–234.

Edwards K.J., Mogg R. 2001. Plant genotyping by analysis of single nucleotide polymorphisms. Plant Genotyping: The DNA Fingerprinting of Plants. Henry RJ. Wallingford, CABI, 1–13.

Ellegren H. 2004. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5, 435–445. DOI: 10.1038/nrg1348.

Elsik C.G., Williams C.G. 2000. Retroelements contribute to the excess low-copy-number DNA in pine. *Molecular and General Genetics*, 264 (1/2), 47–55.

Feder M.E., Mitchell-Olds T. 2003. Evolutionary and ecological functional genomics. *Nature Reviews Genetics*, 4, 651–657.

Giertych M. 1979. Summary of results on Scots pine (*Pinus sylvestris* L.) height growth in IUFRO provenance experiment. *Silvae Genetica*, 28 (4), 136–152.

González-Martínez S.C., Wheeler N.C., Ersoz E., Nelson C.D., Neale D.B. 2007. Association Genetics in *Pinus taeda* L. I. Wood Property Traits. *Genetics*, 175 (1), 399–409.

Guan L., Suharyanto, Shiraishi S. 2011. Isolation and characterization of tetranucleotide microsatellite loci in *Pinus massoniana* (Pinaceae). *American Journal of Botany*, 98, e216–e217. DOI: 10.3732/ajb.1100076.

Hansen O.K., Kjaer E.D. 2006. Paternity analysis with microsatellites in a Danish *Abies nordmanniana* clonal seed orchard reveals dysfunctions. *Canadian Journal of Forest Research*, 36, 1054–1058.

Jaworski A. 2011. Hodowla lasu. Charakterystyka hodowlana drzew i krzewów leśnych. PWRiL, Warszawa.

Liu J.-J., Sniezko R.A., Sturrock R.N., Chen H. 2014. Western white pine SNP discovery and high-throughput genotyping for breeding and conservation applications. *BMC Plant Biology*, 14, 1586.

Moriguchi Y., Yamazaki Y., Taira H., Tsumura Y. 2010. Mating patterns in an indoor miniature *Cryptomeria japonica* seed orchard as revealed by microsatellite markers. *New Forest*, 39, 261–273. DOI: 10.1007/s11056-009-9169-0.

Munyard K.A., Ledger J.M., Lee C.Y., Babra C., Groth D.M. 2009. Characterization and multiplex genotyping of alpaca tetranucleotide microsatellite markers. *Small Ruminant Research*, 85 (2/3), 153–156. DOI: 10.1016/j.smallrumres.2009.07.012.

Murray B.G. 1998. Nuclear DNA Amounts in Gymnosperms. *Annals of Botany*, 82 (Supplement A), 3–15.

Nystedt B., Street N.R., Wetterbom A., Zuccolo A., Lin Y.-C., Scofield D.G., Vezzi F., Delhomme N., Giacomello S., Alexeyenko A., Vicedomini R., Sahlin K., Sherwood E., Elfstrand M., Gramzow L., Holmberg K., Hällman J., Keech O., Klasson L., Koriabine M., Kucukoglu M., Käller M., Luthman J., Lysholm F., Niittylä T., Olson Å., Rilakovic N., Ritland C., Rosselló J.A., Sena J., Svensson T., Talavera-López C., Theissen G., Tuominen H., Vanneste K., Wu Z.-Q., Zhang B., Zerbe P., Arvestad L., Bhalerao R., Bohlmann J., Bousquet J., Garcia Gil R., Hvidsten T.R., de Jong P., MacKay J., Morgante M., Ritland K., Sundberg B., Lee Thompson S., Van de Peer Y., Andersson B., Nilsson O., Ingvarsson P.K., Lundeberg J. Jansson S. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497, 579–584.

Pavy N., Pelgas B., Beauseigle S., Blais S., Gagnon F., Gosselin I., Lamothe M., Isabel N., Bousquet J. 2008. Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics*, 9, 21.

Plomion C., Bartholomé J., Lesur I., Boury C., Rodríguez-Quilón I., Lagraulet H., Ehrenmann F., Bouffier L., Gion J.M., Grivet D., de Miguel M., de María N., Cervera M.T., Bagnoli F., Isik F., Vendramin G.G., González-Martínez S.C. 2016. High-density SNP assay development for genetic analysis

in maritime pine (*Pinus pinaster*). *Molecular Ecology Resources*, 16 (2), 574–587.

Rafalski A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, 5 (2), 94–100.

Ryynänen H.J., Tonteri A., Vasemägi A., Primmer C.R. 2007. A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). *Journal of Heredity*, 98 (7), 692–704.

Slavov G.T., Howe G.T., Adams W.T. 2005. Pollen contamination and mating patterns in a Douglas-fir seed orchard as measured by simple sequence repeat markers. *Canadian Journal of Forest Research*, 35, 1592–1603. DOI: 10.1139/x05-082.

Suharyanto, Nose M., Shiraishi S. 2012. Development and application of a multiplex SNP system to evaluate the mating dynamics of *Pinus thunbergii* clonal seed orchards. *Molecular Breeding*, 30, 1465. DOI: 10.1007/s11032-012-9733-8.

Wachowiak W., Salmela M.J., Ennos R.A., Iason G., Cavers S. 2011. High genetic diversity at the extreme range edge: nucleotide variation at nuclear loci in Scots pine (*Pinus sylvestris* L.) in Scotland. *Heredity*, 106 (5), 775–87. DOI: 10.1038/hdy.2010.118.

Wachowiak W. 2015. Genetic relationships between Polish and reference populations of Scots pine (*Pinus sylvestris* L.) in Europe based on nucleotide polymorphism study at nuclear loci. *Sylwan*, 159 (1), 53–61.

Weller J.I., Seroussi E., Ron M. 2006. Estimation of the number of genetic markers required for individual animal identification accounting for genotyping errors. *Animal Genetics*, 37, 387–389. DOI: 10.1111/j.1365-2052.2006.01455.x.

Werner F.A., Durstewitz G., Habermann F.A., Thaller G., Kramer W., Kollers S., Buitkamp J., Georges M., Brem G., Mosner J., Fries R. 2004. Detection and characterization of SNPs useful for ID control and parentage testing in major European dairy breeds. *Animal Genetics*, 35, 44–49. DOI: 10.1046/j.1365-2052.2003.01071.x.

Williams C., Williamson R., Coutelle C., Loeffler F., Smith J., Ivinson A. 1988. Same-day, first-trimester antenatal diagnosis for cystic fibrosis by gene amplification. *Lancet*, 2 (8602), 102–103.

Williams J.G.K., Kubelik A.R., Livak K.J., Rafalski J.A., Tingey S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18 (22), 6531–6535. DOI: 10.1093/nar/18.22.6531.

Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Friters A., Pot J., Paleman J., Kuiper M., Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23 (21), 4407–4414. DOI: 10.1093/nar/23.21.4407.