Chapter 2 in: **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**. Gillet, E.M. (ed.). 1999. URL http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm

DNA Markers - Concepts and characteristics

Elizabeth M. Gillet[1,2]

[1] Institut für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen, Büsgenweg 2, 37077 Göttingen, Germany

[2] Institut für Forstgenetik und Forstpflanzenzüchtung, Bundesforschungsanstaalt für Forst- und Holzwirtschaft, Sieker Landstrasse 2, 22927 Grosshansdorf, Germany
Email: egillet@gwdg.de

The objective of this compendium is to associate "Purposes" with "DNA markers". Before reading about associations between individual purposes and markers in the contributions, it can be helpful to survey these two topics separately. In this chapter, a systematic classification of the different characteristics of DNA markers that determine their usefulness is introduced.

1. Genetic markers and their characteristics

The concept of genetic marker and gene marker: Consider a given set of individuals of the same species and a set of characteristics of these individuals. The set of characteristics defines a **trait** in the set of individuals, if each individual possesses exactly one of these characteristics - its **trait state** or **phenotype**.

The genetic information possessed by each individual is termed its **genotype** and can refer to the entirety of its genetic information or a part of it. For each locus that is involved in the expression of

the phenotype, the individuals's genotype is the set of genes present at this locus. Since the following considerations apply both to coding and non-coding DNA, the term **gene** is used here in a wide sense to denote a defined segment of the DNA of an individual as a unit of transmission, and not only in the narrow sense of a "functional gene". A **locus** corresponds to a set of "transmission homologous" genes (Gillet 1996). Two genes at the same locus that differ in type are called **alleles**. A haploid locus has one gene, a diploid locus two genes or alleles, and a polyploid locus more than two. (In humans, all nuclear loci that are not located on the X- or Y-chromosomes are diploid. In most tree species, all nuclear loci are diploid. Many agricultural crop plants have been bred to be polyploid at all nuclear loci.)

The following definitions define particular types of relationships between phenotypes and genotypes (Gillet 1996; Bergmann *et al.* 1989). A trait is termed a **genetic trait**, if any two individuals possessing the same genotype also have the same phenotype, regardless of the environmental conditions in which they exist. In order to determine the relationship between phenotypes and genotypes, it is necesary to perform an **inheritance analysis**, *i.e.*, determination of the mode of inheritance of the trait states (see below). After successful inheritance analysis, a genetic trait is qualified as a **genetic marker**, if the relationship holds that each phenotype can be unambiguously assigned to a set of genotypes at one or more specified loci. Thus for a genetic marker it holds that, if one observes an individual's phenotype, then one knows that this individual possesses one of a defined set of genotypes. If each phenotype can be unambiguously assigned to exactly one genotype, then the genetic marker defines a **gene marker**. By this assignment, all involved genes become recognizable. Thus a trait can be called a gene marker if and only if there is a 1:1 relationship between phenotype and genotype, such that the alleles present at each of the involved loci are unambiguously specifiable for each phenotype. This hierarchy of trait types is summarized in Table 1.

Type of trait	**Definition**
Trait	Individual ===> Unique phenotype
Genetic trait	Genotype ===> Unique phenotype
Genetic marker	Phenotype ===> Unique set of genotypes
Gene marker	Phenotype <===> Unique genotype

Table 1: Hierarchy of trait types

Genetic markers preceding the development of DNA markers: Since the advent of recombinant DNA technology in population genetics in the mid-1980's, the repertoire of genetic markers available for population genetic studies in a number of tree species has increased enormously. In order to appreciate this fact, it is interesting to consider the history of genetic marker development in forest genetics. Until the beginning of the 1970's, the only genetic markers available in tree species were the rare morphological traits that could be shown to be controlled by alleles at a single gene locus, such as the *aurea* phenotype in Norway spruce (Langner 1953). Early attempts to interpret the relative or absolute quantities of the different monoterpenes in the resin of conifer trees (measured using gas chromatography) as genetic markers remained inconclusive due to difficulties in determining mode of inheritance and the probable dependence of their expression on environmental conditions (especially pathogen stress).

Isoenzymes: It was not until the early 1970's, when Bartels (1971) and Bergmann (1971) developed enzyme electrophoresis for Norway spruce, that direct products of tree DNA were made

accessible to observation. This technique involved separation of functionally equivalent enzyme molecules according to their differing electrostatic charges, sizes, and molecular conformations, followed by their staining. Inheritance analysis of the resulting banding patterns enabled inference of their mode of inheritance (see below) and, consequently, allowed them to be used as genetic markers. Until recently, practically all progress in the experimental population genetics of forest tree species was achieved using multilocus isoenzyme analysis. **Isoenzymes**, the "electrophoretically separable variants of one enzyme system" (Bergmann *et al.* 1989), are coded by genes at one or often several loci. Variants that are coded by alleles at the same locus are called **allozymes**. Multilocus analysis considers the results for various loci belonging to one or, more commonly, a large number of enzyme systems. In fact, isoenzymes are still widely used as genetic markers for reasons that include the following: They are inexpensive compared to DNA markers, the laboratory protocols are well-established in numerous tree species, they are products of structural genes whose roles in metabolism are known in most cases, and, most importantly, their typical levels of variation makes them suitable markers for a number of purposes.

The usefulness of a marker completely depends on its characteristics. In order to appreciate what can be accomplished with DNA markers that couldn't already be done with isoenzymes, it will be helpful to begin by recalling the characteristics of isoenzymes. Enzyme molecules are direct products of genes, and thus of DNA, and play essential roles in the primary and secondary metabolism of their organism. Enzyme molecules are composed of chains of amino acids as determined by the DNA sequences of the coding genes. Differences in the total electrostatic charges of their amino acid sequences indicate the existence of differences in the DNA sequences. (Due to the redundancy of the genetic code that "assigns" amino acids to nucleotide triplets, the opposite is not necessarily true). Allozymes almost always differ due to single nucleotide substitutions at the locus that cause the substitution of single amino acids of different charges; as a rule, isoenzymes coded by different loci differ in size also (F. Bergmann, personal communication). Size differences result from insertions/deletions of nucleotides that lead to a longer/shorter amino acid sequence.

As summarized in Table 2 below, the typical mode of inheritance of isoenzymes is: Transmission by one or only a few nuclear gene loci; Codominance of gene action (with the exception of the null alleles typically found at some isoenzyme loci), which ensures the identifiability of both genes at a locus and thus of heterozygous individuals. An additional characteristic observed at a number of isoenzyme loci is the following: Prevalence of the same one, two, or even three alleles, accompanied by the same suite of rare alleles, in all studied populations even of related species, and a typically low level of differentiation among populations (Gregorius and Bergmann 1995). This has led to the widespread assumption that isoenzymes are selectively "neutral", *i.e.*, their frequency distributions are due to random effects (random mutation and drift). The authors of the above-cited study, however, consider the universal prevalence of the same two or three alleles at a locus to be evidence of "ontogenetic differentiation of enzyme function", while the universal rarity of the other alleles suggests "recurrent deleterious mutation", both of which are forms of selectivity.

2. Systematic classification of the characteristics of genetic markers

What new characteristics are possessed by DNA markers that earlier markers did not have? These new characteristics determine which previously intractable purposes can now be treated using genetic markers. In order to answer this question, the classification scheme in Table 2 for the description of the characteristics of genetic markers will be helpful:

Mode of inheritance - Mode of transmission

- Uniparental *vs.* biparental inheritance by cell compartment - Examples:
  - Biparental nuclear inheritance - All nuclear genes in plants
  - Maternal nuclear inheritance - Example: the expression of only the maternal allele in the haploid primary endosperm (megagametophyte) of conifer seeds
  - Maternal organellar inheritance - Examples: mitochondrial genome in all tree species; chloroplast genome in angiosperms; chloroplast DNA in the maternally inherited primary endosperm (megagametophyte) of some conifers (*Abies alba:* Vendramin and Ziegenhagen 1997)
  - Paternal organellar inheritance - Example: chloroplast genome in conifer species
- Degree of ploidy - Examples:
  - Diploidy or polyploidy for nuclear genes observed in diplophase material
  - Haploidy for nuclear genes observed in haplophase material
  - Pseudo-haploidy for organellar inheritance
  - Number of loci encoding the "phenotype"
  - Number of alleles per locus

Mode of inheritance - Mode of gene action

- Codominance *vs.* dominance at a nuclear gene locus - Definitions:
  - **Codominance** means that both alleles present at a (diploid or polyploid) locus are always scorable, that is, heterozygotes are always recognizable as such. A locus showing codominance of gene action defines a **gene marker**.
  - **Dominance** means that one allele, a so-called dominant allele, masks the presence of the another, the recessive, allele, so that heterozygotes of genotype Dominant/Recessive are not distinguishable from homozygotes of genotype Dominant/Dominant. A locus showing dominance of gene action defines a **genetic marker**, but not a gene marker.
- Epistasis - Epistasis is defined between two loci and means that the expression of an allele at one locus masks the expression of an allele at another locus. A genetic trait showing epistasis defines a **genetic marker**, but not a gene marker.

Level of genetic variability

- Variation within stands or subpopulations - Assessment using genetic diversity measures. Extreme cases are:
  - Monomorphy, *i.e.*, all individuals have the same "phenotype";
  - Hypervariability, *i.e.*, each individual has a unique "phenotype".
- Variation between stands or subpopulations - Either qualitative (different types) or quantitative (differences in the frequency distributions of the same types). Assessment using a genetic differentiation measure with the following properties (*e.g.* the measure of subpopulation differentiation *delta* (Gregorius and Roberds 1986); (Wright's F_{ST}) = (Nei's G_{ST}) do not fulfill the second requirement (Gregorius 1988):
  - Populations are completely undifferentiated (minimum differentiation value, *e.g.* = 0) if and only if the frequency distributions of types are equal in all populations;
  - Populations are completely differentiated (maximum differentiation value, *e.g.* = 1) if and only if the types present in each population are completely distinct from the types in all other populations.

Function
Functional equivalence *vs.* functional difference of the alleles of a locus - Functional differences may or may not imply selective differences. Alleles at non-coding loci are considered to be functionally and thus selectively equivalent, unless argued otherwise.

Table 2: The characteristics of genetic markers

3. The characteristics of the DNA markers developed within this research project

It will be left to the single contributions to explain how the respective DNA markers are observed in the laboratory. It will suffice here to list their characteristics (some of which were first described during the project) according to the above classification, since it is these characteristics that decide on the suitability of a marker for a given purpose:

Marker type	***Mode of inheritance***		***Level of genetic variability***	***Function***
	Mode of transmission	***Mode of gene action***		
AFLP® fingerprint	biparental nuclear, many loci, unknown no. alleles per locus	dominance at some loci, codominance at others	hypervariable, *i.e.* each individual has unique banding pattern	unknown
Nuclear microsatellites	biparental nuclear, few loci, many alleles per locus	codominance, with exception of null alleles at some loci	large variation within populations, low differentiation between populations	non-coding, may contribute to genome stability
Chloroplast microsatellites	uniparental (maternal in angiosperms, paternal in conifers), pseudo-haploid, single locus, many alleles per locus	each cytotype is expressed	low variation within populations, large differentiation between populations	non-coding
Mitochondrial intron marker	uniparental (maternal), pseudo-haploid, single locus, many alleles per locus	each cytotype is expressed	low variation within populations, large differentiation between regions	non-coding
ITS of ribosomal DNA	biparental nuclear, several loci, several alleles per locus	codominance	high variability, even within a single individual	non-coding
cDNA markers	biparental nuclear, one to a few loci, few alleles per locus	codominance	low variation within populations, low differentiation between populations	functional differences possible between alleles of a locus
Isoenzymes (for comparison)	biparental nuclear, 1-5 loci, 1-7 alleles per locus	codominance, with exception of null alleles at some loci	low to medium variation within populations, low differentiation between populations	functional differences possible between alleles of a locus

Table 3: Characteristics of the DNA markers developed in the project

4. Links to the contributions treating the different types of DNA marker

---

AFLP® fingerprint

Chapter 4: Comparison of microsatellites and AFLP markers for parentage analysis - S. Gerber, S. Mariette, R. Streiff, C. Bodénès, A. Kremer

Chapter 5: DIG-labelled AFLPs in oaks - A DNA marker for reconstruction of full- or half-sib family relationships? - B. Ziegenhagen, V. Kuhlenkamp, R. Brettschneider, F. Scholz, B.R. Stephan, B. Degen

Chapter 12: Amplified Fragment Length Polymorphisms and Microsatellites: A phylogenetic perspective - J.P. Robinson, S.A. Harris

Chapter 14: Application of the AFLP® technique in marker assisted breeding - J. Peleman

Nuclear microsatellites

Chapter 4: Comparison of microsatellites and AFLP markers for parentage analysis - S. Gerber, S. Mariette, R. Streiff, C. Bodénès, A. Kremer

Chapter 6: Microsatellite analysis of anonymous seedlot samples from oak: a promising approach to monitor the number of different seed parents and pollen donors - C. Lexer, B. Heinze, S. Gerber, H. Steinkellner, B. Ziegenhagen, A. Kremer, J. Glössl

Chapter 7: Nuclear microsatellites as a tool in the genetic characterization of forest reproductive material. A case study in sessile oak (*Quercus petraea* Matt., Liebl.) - S. La Scala, R. Schubert, G. Müller-Starck, K. Liepe

Chapter 8: Microsatellite markers as a tool for the detection of intra- and interpopulation genetic structure - I. Scotti, G. Paglia, F. Magni, M. Morgante

Chapter 12: Amplified Fragment Length Polymorphisms and Microsatellites: A phylogenetic perspective - J.P. Robinson, S.A. Harris

Chloroplast microsatellites

Chapter 9: Chloroplast microsatellites for analysis of the geographic distribution of diversity in conifer species - M. Anzidei, A. Madaghiele, C. Sperisen, B. Ziegenhagen, G.G. Vendramin

Mitochondrial intron marker

Chapter 10: Mitochondrial DNA variation provides a tool for identifying introduced provenances: A case study in Norway spruce - C. Sperisen, U. Büchler, G. Mátyás, L. Ackzell

ITS of ribosomal DNA

Chapter 11: Limitations to the phylogenetic use of ITS sequences in closely related species and populations - a case study in *Quercus petraea* (Matt.) Liebl. - G. Muir, C. Schlötterer

cDNA markers

Chapter 13: Isolation and sequence analysis of oak and spruce cDNA clones - M. Berenyi, S. Fluch, K. Hohl, K. Burg, R. Schubert, R. Riegel, G. Müller-Starck

---

References

Bartels H (1971) Genetic control of multiple esterases from needles and macrogametophytes of *Picea abies. Planta (Berl)* 99:283-289.

Bergmann F, Gregorius H-R, Scholz F (1989) Isoenzymes, indicators of environmental impacts on plants or environmentally stable gene markers? In: Scholz F, Gregorius H-R, Rudin D (eds.). *Genetic Effects of Air Pollutants in Forest Tree Populations.* Springer-Verlag Heidelberg, New York, Tokyo, pp. 17-28.

Bergmann F (1971) Genetische Untersuchungen bei *Picea abies* mit Hilfe der Isoenzym-Identifizierung. I. Möglichkeiten für genetische Zertifizierung von Forstsaatgut. *Allgemeine Forst- und Jagdzeitung* 142: 278-280.

Gregorius H-R, Bergmann F (1995) Analysis of isoenzyme genetic profiles observed in tree populations. In: Baradat P, Adams WT, Müller-Starck G (eds.). *Population Genetics and Genetic Conservation of Forest Trees*. Amsterdam: SPB Academic Publishing, pp. 79-96.

Gillet EM (1996) Qualitative inheritance analysis of isoenzymes in haploid gametophytes: principles and a computerized method. *Silvae Genetica* 45: 8-16.

Gregorius H-R, Roberds J (1986) Measurement of genetical differentiation among subpopulations. *Theoretical and Applied Genetics* 71: 826-834.

Gregorius H-R (1988) The meaning of genetic variation within and between subpopulations. *Theoretical and Applied Genetics* 76: 947-951.

Langner W (1953) Eine Mendelspaltung bei Aurea=Formen von *Picea abies* (L.) Karst. als Mittel zur Klärung der Befruchtungsverhältnisse im Walde. *Z. Forstgenetik u. Forstpflanzenzüchtung* 2: 49-51.

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