Population Genetics – an Overview 2

Population geneticists spend most of their time doing one of two things: describing the genetic structure of populations or theorizing on the evolutionary forces acting on populations. On a good day, these two activities mesh and true insights emerge (Gillespie 1998).

2.1 Introduction

Population genetics provides the framework for an understanding of the dynamics of genetic structures or evolution. The following considerations are neither specific for nor confined to forest species. Most aspects mentioned in this chapter are as relevant for tropical forest species as for any other higher-plant or animal species. However, the population genetics approach taken throughout this book to discuss genetic processes in tropical forests requires a basic understanding of the most important fundamentals concerning genetic structures and their dynamics in time and space, and introduces the concept of genetic markers as well as the most important molecular and biochemical gene markers.

2.2 The Population

The genetic information or genotype of an organism does not change throughout its lifetime. It is necessary to observe and to compare genetic information of many plants or animals in order to describe genetic variation and its dynamics in time. The fundamental units for studies of genetic variation are collectives of individuals exchanging their genetic information among each other for the sexual production of the next generation. Such units of (sexual) reproduction are defined as populations.

A species rarely consists of one single population. It is usually composed of several more or less isolated populations. Isolation means the absence of mating

events between plants (or animals) from different populations in this context. However, an assessment of reproductive isolation usually requires a fairly detailed understanding of the reproduction system of a species. Furthermore, the definition of a population is operational only if an occasional (but rare) exchange of genetic information among populations is taken into consideration. Thus, interfertile, spatially clustered plants are often regarded as belonging to the same population, while spatially separated plants are regarded as parts of different populations.

The delineation of populations is particularly difficult for immobile organisms with poorly described spatial distribution patterns and largely unknown means of gene dispersal through seed and pollen. In most cases, it is far from obvious which trees in a species-rich tropical forest belong to the same population and how big tree populations are. This has far-reaching consequences for levels of genetic variation of tropical forest trees, as will be discussed in later chapters.

Analyses of spatial patterns of genotypes without a priori information on the delineation of populations or partially isolated subpopulations have recently been proposed as an alternative to the "traditional" approach of dividing the overall variation into components within and among demes or populations (Diniz-Filho and Telles 2002; Manel et al. 2003). Spatially explicit analyses of genetic variation patterns might be particularly informative for plant species occurring in low density in tropical forests; however, currently available information is mainly confined to comparatively small areas (Ng et al. 2004).

2.3 Variation at Gene Loci

2.3.1 The Molecular Basis of Genetic Variation

The genetic information of all organisms is stored as a sequence of the four bases or nucleotides adenine, thymine, cytosine, and guanine in their DNA. It is multiplied by the process of replication and translated to a sequence of amino acids by the processes of transcription (from DNA to messenger RNA, mRNA) and translation (from mRNA to polypeptides) (Griffiths et al. 2000). A cistron is a part of the DNA coding for a particular polypeptide, for example, an enzyme. From this point of view, the variation at a particular gene or gene locus can be assessed by observing differences in the DNA sequence within a cistron. The total number of "genes" is unknown for any tropical forest plant; however, it may be assumed to be at least on the order of more than 25,000 estimated for the first fully sequenced flowering plant, the herbaceous

Arabidopsis thaliana (The Arabidopsis Genome Initiative 2000). The number of genes of the first sequenced tree species (*Populus trichocarpa*) is estimated to be even higher (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html).

The observation of variation at the most basic level of DNA sequences is routine for tropical trees just as for other organisms, but it is time-consuming and costly if sequences of hundreds of organisms need to be compared. To date, only a few population studies have been conducted based on the observation of variation of DNA sequences within a single or a few closely related tropical forest species (but see, e.g., Ishiyama et al. 2003). Sequencing is more frequently applied to clarify taxonomic subdivisions and phylogenies (Chap. 8).

Numerous alternatives to sequencing are available. The choice of the most appropriate method depends on several factors, including available resources and the purpose of a study. A comprehensive overview of currently available molecular tools to assess patterns of genetic variation in plants has been given by Weising et al. (2005).

2.3.1.1 Molecular Markers

The most important methods to assess patterns of genetic variation within and among species rely on the amplification of short DNA fragments by means of polymerase chain reaction (PCR). PCR allows short fragments of the DNA of a target organism to be selectively amplified (multiplied) for further analyses. The sequence of two short oligonucleotides (primers), each usually between ten and 25 base pairs in length, is decisive for the amplified DNA region. The following marker types, all based on the PCR method, have gained particular importance in the study of genetic variation patterns of tropical forest trees.

(Partial) Gene Sequences and Single Nucleotide Polymorphisms

Selective primers can be designed to amplify the DNA coding for a particular gene (cistron). DNA regions translated into a polypeptide (exons) and non-translated regions (introns) of structural genes as well as regulatory genes can be analyzed. It is possible to sequence the amplified PCR products, often after a cloning step. For example, Kamiya et al. (2005) investigated the phylogeny of the species-rich genus *Shorea* (Dipterocarpaceae) on the basis of a (partial) sequence of the *PgiC* gene (Example 8.1).

An investigation of genetic variation is particularly rewarding at gene loci with known or putative function and a directly observable effect on phenotypic traits responsible for the adaptation to particular environmental conditions (candidate genes). Populations can be screened for single nucleotide polymorphisms (SNPs) without the need for repeated sequencing (Morin et al. 2004).

The characterization of expressed sequence tags, studies of molecular variation at the level of expressed genes, and analyses of variation at SNPs are expected to greatly improve our understanding of adaptive genetic variation and its dynamics in space and time. However, these recently developed marker types are not discussed in detail owing to the lack of published reports for tropical forest species (but see Sect. 7.2).

Microsatellites (Simple Sequence Repeats)

Microsatellites or simple sequence repeats (SSRs) are short DNA fragments of usually only two or three base pairs in length which are repeated several times in a particular location of the DNA (Fig. 2.1). SSRs can be studied by the development of primers in conserved DNA regions flanking the microsatellite. The development of primers is rather costly and time-consuming (Squirrell et al. 2003). Nuclear microsatellites developed for a particular species can only be transferred to closely related species, usually within the same genus.

Nuclear microsatellites are important genetic markers owing to their usually high variability and codominant mode of inheritance (Sect. 2.3.2, Chap. 3).

The variation of microsatellite fragment sizes is due to a different number of repeated motifs resulting in slightly different sizes of amplified fragments (Fig. 2.2). The availability of highly variable markers is particularly useful for analyses of the mating system and gene flow (Chaps. 5, 6). Microsatellites have been developed for a number of tropical forest tree species, including *Pithecellobium elegans* (Chase et al. 1996a), *Swietenia humilis* (White and Powell 1997), *Gliricidia sepium* (Dawson et al. 1997), *Shorea curtisii* (Ujino et al. 1998), *Caryocar brasiliense* (Collevatti et al. 1999), *Neobalanocarpus heimii* (Iwata et al. 2000), and *Prosopis* spp. (Mottura et al. 2005).

Microsatellite motifs are not only found in DNA of the nucleus, but also in DNA of chloroplasts (cpDNA; Sect. 2.3.2). Single nucleotide repeats are particularly common in cpDNA. Universal primers are available to investigate variation at cpSSRs in many different species of angiosperms (Weising and Gardner 1999) and gymnosperms (Vendramin et al. 1996).

AAGGATAAGTTAAA ACACACACACACACAC GTTGCCTCCATTT

TATGATGTATGAAT ACACACACACACACAC CCCACCTGGTTTT

TTGACGTCACACAG ACACACACACACACACAC TCTCTCATCCACA

Fig. 2.1. Three different DNA sequences of *Prosopis chilensis* containing the microsatellite repeat motif AC in nine (*lower* and *middle*) or ten (*upper*) copies (*underlined*). *A* adenine, *C* cytosine, *G* guanine, T thymine. (From Mottura, unpublished)

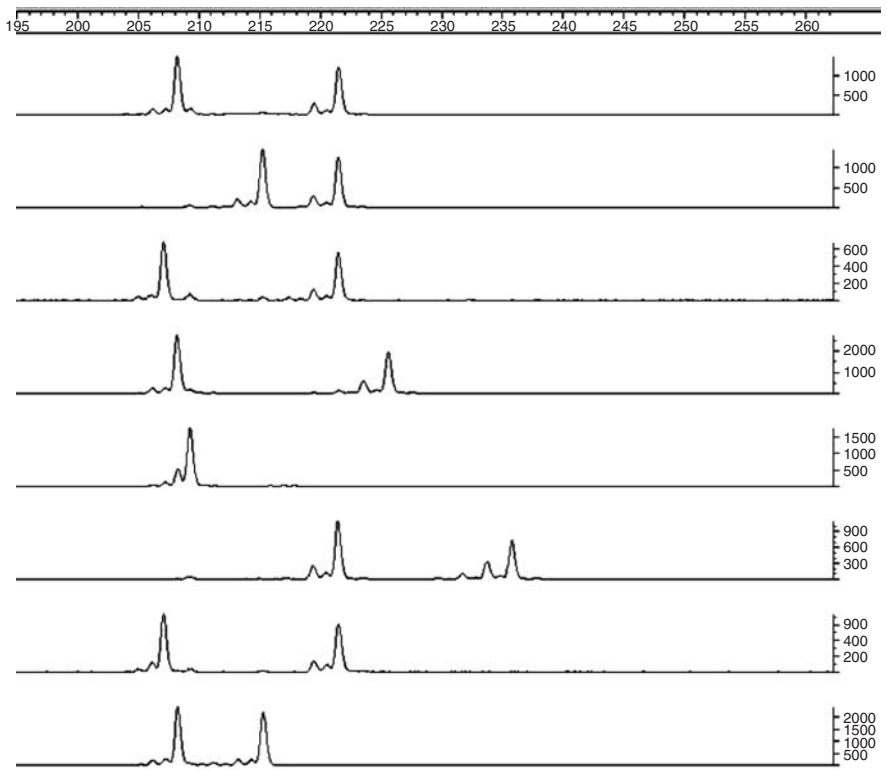

Fig. 2.2. Length variation for eight *Prosopis* spp. trees at a microsatellite gene locus (Mottura et al. 2005) separated in an automated sequencer (ABI 3100). Each line refers to one tree. Different sizes of amplified fragments are visualized as peaks at different positions in the chromatogram. Trees with two amplified fragments are heterozygous; tree 5 is homozygous with only one amplified fragment. (From Mottura, unpublished)

Random Amplified Polymorphic DNA and Amplified Fragment Length Polymorphism

Random amplified polymorphic DNA (RAPD; Newbury and Ford-Lloyd 1993; Fig. 2.3) and amplified fragment length polymorphism (AFLP; Vos et al. 1995; Fig. 2.4) are genetic fingerprinting techniques resulting in more or less complex DNA patterns. Unlike most other techniques including microsatellites no previous sequence information is necessary for RAPD and AFLP studies. The presence (+ or 1) or absence (– or 0) of a DNA fragment of a particular size is scored for each sample plant investigated. This information is the basis for a matrix of size $n \times m$ (n is the number of plants investigated; m is the number of DNA fragments observed) with elements 1 (fragment present) or 0 (fragment absent). RAPDs and AFLPs are usually interpreted as dominant markers (see later). No

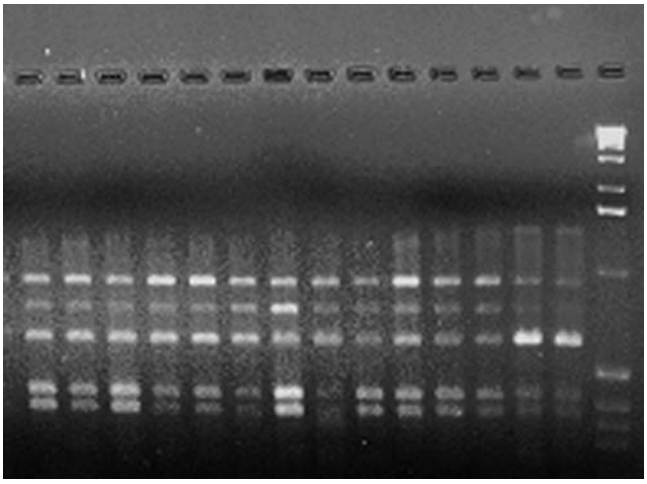

Fig. 2.3. Low variation at a random amplified polymorphic DNA marker in teak (*Tectona grandis*). (From Finkeldey, unpublished)

information on the function of a particular DNA fragment or its mode of inheritance is available; hence, RAPDs and AFLPs are "anonymous" markers.

The RAPD technique is a simple molecular tool to assess genetic variation within and among populations which has been widely used for tropical forest plants such as *Gliricidia* spp. (Chalmers et al. 1992), *Cedrela odorata* (Gillies et al. 1997), *Caesalpinia echinata* (Cardoso et al. 1998), *Prunus africana* (Dawson and Powell 1999), and many others. The reproducibility of the RAPD technique has been under dispute (Rabouam et al. 1999).

The AFLP method is more demanding with regard to laboratory equipment and experience, but allows more DNA fragments to be investigated from a single PCR reaction, and shows higher reproducibility in comparison with RAPDs. The potential of the AFLP technique to assess genetic variation patterns in tropical trees was recently reviewed by Kremer et al. (2005). AFLP studies to assess genetic variation within tropical tree species were conducted, for example, for *Moringa oleifera* (Muluvi et al. 1999), *Calycophyllum spruceanum* (Russel et al. 1999), and *Acer skutchii* (Lara-Gomez et al. 2005).

Restriction Fragment Length Polymorphism

A simple method to observe variation among DNA fragments from different plants is to digest the fragments into smaller pieces by restriction enzymes. Restriction enzymes cut DNA at a particular sequence. These recognition sequences are usually between four and six base pairs in length. The restriction fragment length polymorphism (RFLP) technique is based on the use of restriction enzymes to observe genetic variation. For example, nuclear RFLPs

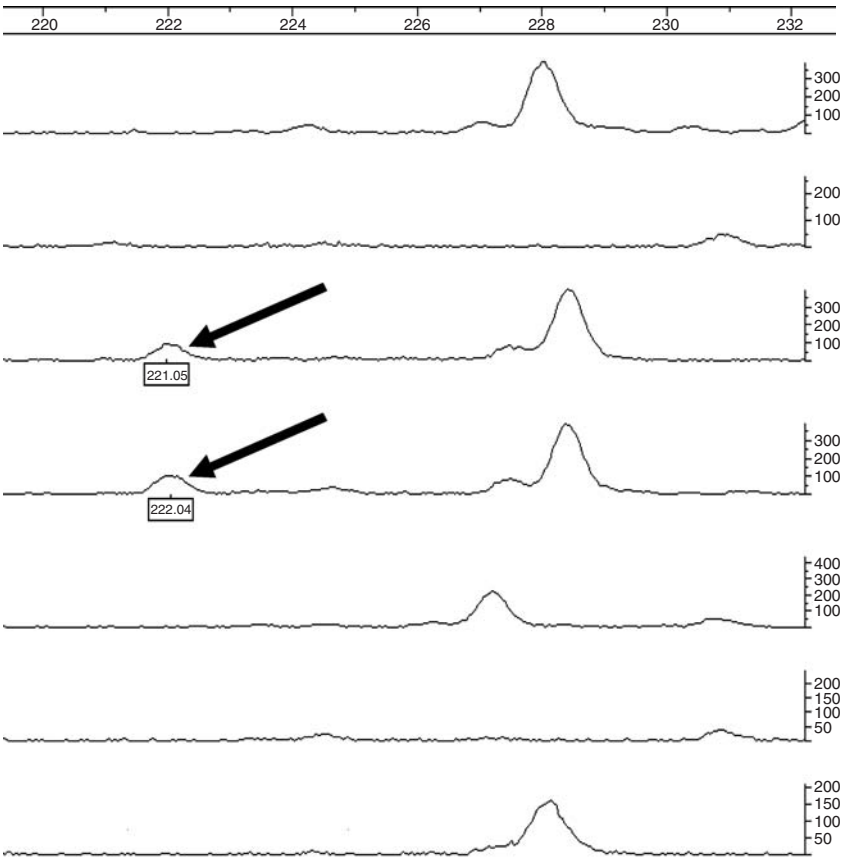

Fig. 2.4. Variation at amplified fragment length polymorphisms in seven *Shorea parvifolia* trees after separation in an automated sequencer (ABI 3100). Only fragments in the range 220–232 bp are shown. The fragment with a size of 220 bp (*arrows*) is "diagnostic" for the population Berau (Borneo), since it was not observed in six other populations. (From Cao, unpublished)

were developed for *Acacia mangium* to assess genetic diversity and differentiation among populations (Butcher et al. 1998) and with the objective to incorporate marker-based approaches in breeding programs (Butcher et al. 2000; Butcher 2004).

More frequently, the RFLP techniques is combined with a PCR step and is used to assess variation of cpDNA. For example, Tsumura et al. (1996) and Indrioko et al. (2006) (Fig. 2.5) analyzed phylogenetic relationships among Southeast Asian dipterocarps using the PCR-RFLP technique (Example 8.1), and Cavers et al. (2003) observed strong genetic differentiation among five cpDNA haplotypes of *Cedrela odorata* in Central America (Example 3.6).

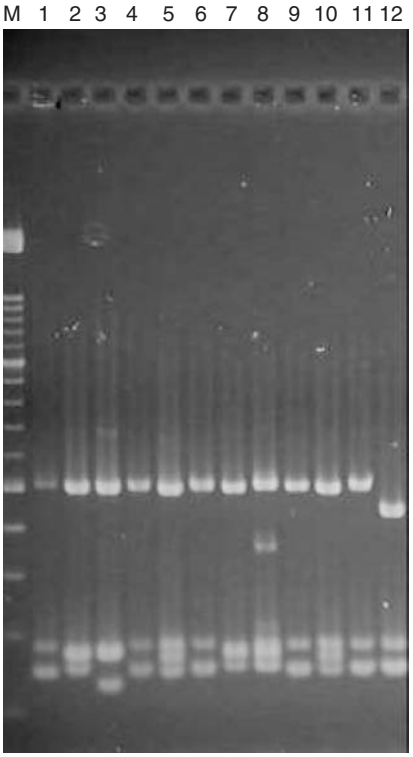

Fig. 2.5. Polymerase chain reaction restriction fragment length polymorphisms of Indonesian Dipterocarpaceae after amplification with the primer pair *TrnL–TrnF* and restriction with *TaqI*. *M* marker (size standard), *1 Hopea celebica*, *2 Vatica rassak*, *3 V. pauciflora*, *4 H. bancana*, *5 Dipterocarpus oblongifolius*, *6 S. javanica*, *7 V. granulata*, *8 Upuna borneensis*, *9 S. mecistopteryx*, *10 D. rigidus*, *11 S. mecistopteryx*, 12 *H. grifithii*. (From Indrioko, unpublished)

2.3.1.2 Biochemical Markers – Isozymes

The sequence of nucleotides of the DNA is converted to a sequence of amino acids of polypeptides by the processes of transcription and translation. Enzymes are the most important group of polypeptides catalyzing all kinds of biochemical reactions in the metabolism; thus, there is a close relation between enzymes and controlling genetic information according to the "one gene – one polypeptide hypothesis." Isozymes are enzymes with similar or even identical functions. Electrophoresis of isozymes followed by biochemical staining is a simple way to observe differences with regard to their electric load and/or spatial structure (Rothe 1994; Fig. 2.6). It is often possible to directly relate the variation observed by isozyme electrophoresis to genetic variation (Bergmann and Hattemer 1998). A formal inheritance study (Sect. 2.3.2) is strongly recommended prior to the use of isozymes as genetic markers.

Isozyme gene loci became the first widely used markers to assess patterns of genetic variation of tropical forest species at single gene loci. The ease of isozyme

Fig. 2.6. Zymogram showing variation of 16 *Dalbergia sissoo* trees from Nepal at the enzyme system alcohol dehydrogenase. The variation is controlled by a gene locus *ADH-A*. The inferred genotypes of the respective trees are indicated. (From Pandey, unpublished)

inventories and the comparatively low costs have made them useful tools for population genetics studies, although they were often substituted by various types of molecular DNA markers during the last decade.

2.3.2 The Gene As a Unit of Heredity

Genetics as the science of heredity dates back to the experiments of the Austrian monk Gregor Mendel (1866), if not earlier. The definition of a gene as a unit of heredity was introduced in the early twentieth century, long before the role of DNA as the material basis of genetic information was recognized. A gene in this sense is identified by the observation of segregation within progenies of particular parents.

The basic idea for the identification of a gene as a unit of heredity has remained unchanged since the experiments Mendel (1866). He investigated segregation ratios in progenies after controlled pollination at simple morphological traits such as the color of petals in peas.

If a plant is heterozygous at a controlling gene locus, it will transmit only one of the two different alleles to a particular progeny (the terms heterozygosity and allele are described later). Thus, segregation is expected in progenies if at least one parent was heterozygous at a controlling gene locus. It is possible to compare the observed segregation in a sample of progenies after controlled pollination with an expectation based on the assumption of a simple "Mendelian" inheritance of the trait. The controlling gene or the gene locus is identified, if the

differences between the observed and expected values are not significant (Hattemer 1991). Suitable statistical tests are, for example, described by Sokal and Rohlf (1998, p. 686 ff.; χ^2 test or goodness-of-fit test, G test).

The formal identification of a genetic marker locus by an observation of segregation is possible only for environmentally stable, genetic traits. It is also advisable for the identification of gene loci based on the observation of biochemical or molecular markers. For example, Moran and Bell (1983) observed segregation at the isozyme system *GOT* in progenies of *Eucalyptus regnans* after controlled pollination and found no evidence to reject the hypothesis of the genetic control of the variation by a single gene locus.

Forest trees, in particular those of the tropics, are extraordinarily difficult to cross owing to their large size, short viabilities of pollen, unknown reproductive biology, and other obstacles. Thus, the identification of a gene locus by inheritance studies is preferentially conducted on the basis of alternative approaches such as the observation of segregation in the haploid megagametophyte (the "endosperm") of gymnosperms (Bergmann 1974) or the observation of segregation in the progenies of putatively heterozygous seed trees after open pollination (Gillet and Hattemer 1989). The first method has been used to identify enzyme gene loci in *Pinus merkusii* from Thailand (Changtragoon and Finkeldey 1995b); the latter approach was used to clarify the inheritance of isozymes in *Pterocarpus indicus* from the Philippines (Finkeldey et al. 1998).

2.3.3 The Mode of Inheritance

The analysis of variation patterns at environmentally stable, genetic traits (Chap. 3) requires an understanding of the transmission of the trait from parents to progenies and from one generation to the next. Two considerations are of particular importance in this context: the transmission of genetic information from both parents or from a single parent only, and the impact of a single or two alleles at a gene locus on the observed trait expression (the phenotype).

Most of the DNA in a plant cell is located in the nucleus. Nuclear DNA (nDNA) is inherited by both parents. For diploid organisms, one set of chromosomes is transmitted by each of the two parents to their common progenies; thus, nDNA is typically **biparentally** inherited. In addition, DNA of plants is located in two types of plastids: mitochondria (mitochrondrial DNA, mtDNA) and chloroplasts (cpDNA). DNA in plastids, both mitochondria and chloroplasts, is usually transmitted by a single parent to its progenies (uniparental inheritance). mtDNA is mainly inherited from the female or seed parent only

(**maternal** inheritance). The same holds for cpDNA in angiosperms. However, cpDNA of gymnosperms is typically **paternally** inherited, i.e., only the genetic information of the pollen parent is transferred to a progeny. cpDNA was investigated in most studies on genetic variation of tropical forest plants based on uniparentally inherited markers (Example 3.6). Maternally inherited mtDNA was studied in several animal species such as Central American frogs (Crawford 2003) and the giant panda in China (Lu et al. 2001).

The different types of genetic information at any biparentally inherited gene locus are called **alleles**. Since two sets of chromosomes are inherited (one from the seed parent, the other from the pollen parent), each nuclear gene is represented in two copies in each progeny of a diploid species. Polyploids are not discussed in this introduction. A tree is **homozygous** (e.g., A_1A_1) at a particular locus A, if two identical alleles (both A_1) were inherited from its parents. A **heterozygous** tree (e.g., A_1A_2) received two different alleles (A_1 from the seed parent and A_2 from the pollen parent or vice versa) from its parents.

An allele is **dominant**, if its possession in a single copy is sufficient to express a particular phenotype. If A_1 is dominant, the phenotypic expressions of trees with the genotypes A_1A_1, A_1A_2, A_1A_3, etc. are identical. The other alleles (A_2 and A_3 in the example) are **recessive**. If it is possible to distinguish all heterozygotes (e.g., A_1A_2) from all homozygotes (A_1A_1 and A_2A_2), the alleles are defined as **codominant**. A simple observation of genetic structures (Sect. 2.4) is only possible at codominant marker loci.

2.3.4 Definition and Classification of Gene Markers

On the basis of the previous considerations, it is possible to define a gene marker in different contexts:

- A **gene marker** (synonyms gene locus and marker locus) **in a wide sense** is an environmentally stable trait; thus, the variation is determined by genetic factors only.
- A **gene marker in a narrow sense** is an environmentally stable, biparentally inherited, codominant trait.

Although many different traits may be regarded as gene markers in a wide sense, only the molecular and biochemical markers briefly described in Sect. 2.3.1 have considerable practical importance to assess genetic variation patterns of tropical forest species. A rough classification of important, currently available marker types with regard to their mode of inheritance and their variability is presented in Table 2.1.

Table 2.1. Classification of genetic markers in a broad sense commonly used in studies on genetic variation of tropical forest species with regard to their mode of inheritance and their variability

Variation	Uniparental	Anonymous[a]	Codominant[b]
High	Chloroplast simple sequence repeats; sequences of chloroplast DNA	Amplified fragment length polymorphisms	Microsatellites (simple sequence repeats)
Moderate to low	Polymerase chain reaction restriction fragment length polymorphisms of chloroplast DNA and mitochondrial DNA	Random amplified polymorphic DNAs	Isozymes

[a]Mode of inheritance usually unknown; interpretation as dominant markers
[b]Gene markers in a narrow sense

2.4 Genetic Structures Within Populations

The final result from a laboratory study to assess the genetic constitution of a plant at a particular biochemical or molecular gene marker system is a specific pattern, which can be used:

- To infer a genotype in case of gene markers in a narrow sense (e.g., isozymes, SSRs)
- To identify a haplotype in the case of uniparentally inherited markers (e.g., PCR-RFLPs of cpDNA)
- To assign a genetic fingerprint usually consisting of many anonymous marker loci to the plant (e.g., AFLPs)

Genetic structures are frequency distributions of such genetic types in populations. The most important frequency distributions to assess levels of genetic variation within and among populations are allelic and genotypic structures.

2.4.1 Allelic and Genotypic Structures

Genetic variation is assessed by an investigation of all plants or a (random) sample of plants from a population. At a gene marker in a narrow sense, the genotypes of these plants are known, and it is possible to calculate the (relative) frequency of a particular genotype A_iA_j as

$$P_{ij} = N_{ij}/N, \tag{2.1}$$

where N_{ij} is the (absolute) frequency of genotype A_iA_j in the population (or sample) and N is the population size (or sample size).

The frequency vector of all (relative) frequencies of genotypes is defined as the **genotypic structure** of the population at the respective marker gene locus. It obviously holds that $\sum_i \sum_j P_{ij} = 1$.

Each genotype of a diploid species consists of two alleles; thus, the total number of alleles in a population (or a sample) is twice the number of plants investigated, and the frequency of a particular allele A_i can be calculated as

$$p_i = N_i / 2N, \tag{2.2}$$

where N_i is (absolute) frequency of allele A_i in the population (or sample) and N is the population size (or sample size). The respective allele A_i is counted twice in case of homozygous plants A_iA_i.

The frequency vector of all (relative) frequencies of alleles is defined as the **allelic structure** of the population at the respective marker gene locus. Again, it holds that $\sum p_i = 1$.

If the genotypic structure of a population is known, it is easily possible to compute the allelic structure at the respective gene locus. The frequency of allele A_i (p_i) can be computed by calculating the sum of the frequency of the corresponding homozygote (A_iA_i) and half of the frequencies of all heterozygotes where the respective allele occurs (A_iA_j, A_iA_k, ...). However, the calculation of genotypic structures based on allele frequencies requires additional information or assumptions (e.g., Sect. 6.1.2).

Example 2.1: Genetic Structures at an Isozyme Gene Locus in *Dalbergia sissoo*

The observed variation at the isozyme gene locus *ADH-A* is illustrated for 16 *Dalbergia sissoo* trees in Fig. 2.6. Three different genotypes were observed. Six trees (nos. 1, 3, 4, 7, 15, and 16) show the homozygote genotype A_2A_2, tree 10 is the only tree with genotype A_1A_1, and the remaining nine trees exhibit the heterozygous genotype A_1A_2. Thus, the **genotypic structure** of the sample is as follows:

P_{11}=1/16=0.0625,
P_{12}=9/16=0.5625,
P_{22}=6/16=0.3750.

Eleven of the 32 observed alleles (each tree has two alleles) are A_1, the remaining 21 alleles are of type A_2. The **allelic structure** may also be computed as follows:

$$p_1 = P_{11} + 0.5P_{12} = 0.0625 + 0.5625/2 = 11/32 = 0.34375,$$
$$p_2 = P_{22} + 0.5P_{12} = 0.375 + 0.5625/2 = 21/32 = 0.65625.$$

2.4.2 Variation at Uniparentally Inherited Markers

The uniparental inheritance of cpDNA and mtDNA implies the absence of recombination among genes. Since cpDNA and mtDNA are usually transmitted without any changes in maternal or, in the case of cpDNA of gymnosperms, paternal lineages, it is possible to interpret variation patterns observed at different loci together and to assign particular cpDNA or mtDNA **haplotypes** to different plants. The interpretation of genetic diversity within and among populations (Chap. 3) is conventionally based on these haplotypes rather than on variation at particular loci.

Example 2.2: Diversity of cpDNA Haplotypes in *Dalbergia sissoo*

Six to ten seeds from randomly selected, different seed trees were investigated in each of ten populations of *D. sissoo* in Nepal. The inventory was conducted in five regions, each represented by one natural population and a neighboring plantation. Variation was detected by the PCR-RFLP technique (Sect. 2.3.1) by restriction of the amplified fragment *Trn K1/K2* with the enzymes *AluI* and *RsaI*, and by amplification of the chloroplast microsatellites (cpSSRs) *ccmp6* and *ccmp7* (Weising and Gardner 1999). A total of eight haplotypes with two or three haplotypes in each population were observed (Table 2.2). Haplotypes 1–3 were only observed in natural populations, while haplotypes 4–8 occurred mainly in plantations (Pandey et al. 2004).

Table 2.2. Variation of chloroplast DNA haplotypes in five natural populations and five plantations of *Dalbergia sissoo* in Nepal. (Adapted from Pandey et al. 2004)

Region	Population	*N*	Chloroplast DNA haplotypes							
			1	2	3	4	5	6	7	8
Natural populations										
A	Hetauda N	10	9	1						
B	Shivapur	10	8		2					
C	Hattisar	6	5	1						
D	Godawari	10	4		6					
E	Pipariya	10			9	1				
Plantations										
A	Hetauda P	10					4	6		
B	Surai	10				1			7	2
C	Thakurdwara	10				4			3	3
D	Attaria	10				6			4	
E	Shuklaphanta	10				5			5	

N sample size

2.5 Evolution and Evolutionary Factors

Today's biodiversity is the result of the biological evolution which started at least three and a half billion years ago on earth. The number of existing species is still unknown, since a large proportion of the species living on our planet have not yet been described. The majority of "unknown" species are insects living in the canopy of tropical forests (Erwin 1988). Current estimates of the global species diversity vary from approximately two million to 30 or even 50 million species (Erwin 1997).

The long history of evolution on earth does not preclude the study of its effects in short periods. Evolution is permanently ongoing and it is possible to observe its effects even in long-living organisms such as forest trees. Furthermore, the majority of evolutionary change is within existing species and does not immediately result in the creation of new taxa.

The main principle of evolution was recognized and described by Charles Darwin (1859): Biological evolution is based on heritable variation within species or populations. Darwin recognized the key role of natural selection in this context. Fundamental observations, which led to the development of the selection theory, were made both by Darwin and Alfred Russel Wallace, who independently developed a similar theory, after extended studies in tropical areas (Lefèvre 1984; Sect. 8.2).

Our advanced knowledge about the process of inheritance allow us to define evolution more precisely today that during Darwin's era. **Evolution is a change of the genetic** (allelic or genotypic) **structure of a population** at one or several gene loci. Only changes improving the adaptedness of a population to particular environmental conditions were regarded as evolution by Dobzhansky et al. (1977); however, Kimura (1983) proposed including all changes of genetic structures including those at selectively neutral loci into the concept.

Evolutionary factors are the causes of changes of genetic structures (Hedrick 2000). The significance of evolutionary factors for the dynamics of genetic structures of tropical forest plants is the main topic in later chapters of this book; thus, only a brief description is given here:

- **Mutation** is the prerequisite for any genetic variation. Mutations are random changes of genetic information of an organism which might affect a single nucleotide (a spontaneous change of a single nucleotide resulting in a SNP; Sect. 2.3.1), the chromosomal structure (e.g., fissions, translocations, inversions), or the number of genes (e.g., duplications, polyploidy). Most mutations are detrimental or neutral for an organism. Only few advantageous mutations increase the fitness, i.e., the capacity to produce offspring.

Mutations are rare events; mutation rates are estimated to be in the range between 10^{-6} and 10^{-8} for most genes. However, the frequency of mutations is not uniform. For example, considerably higher estimates of mutation rates were reported for microsatellites (Goldstein and Pollock 1997).

- **Gene flow** and **migration** increase the genetic multiplicity of a population, if genes which were previously restricted to certain populations migrate to other populations. Transport of pollen (gene flow in a narrow sense) and seeds (migration) are the main processes involved in the dispersal of genes for most plant species (Chap. 5). By definition, gene flow and migration among populations are rare events.
- The **mating system** is decisive for the fusion of male and female gametes. Gene flow through pollen is an obvious prerequisite for mating between two seed plants. Thus, gene flow and the mating system are closely connected and are the two most important aspects of the reproduction system of a plant species. The mating system decides on the combination of alleles to genotypes. The genotypic structure of the progeny generation is influenced by the mating system, which primarily does not cause changes of allelic structures. Inbreeding due to selfing or mating among relatives is a particularly important aspect of the mating system (Chap. 6).
- Random fluctuations of genetic structures are described as **genetic drift**. Genetic drift is a consequence of limited population size and becomes stronger in small populations. The strength of genetic drift can be described as a function of the population size, although its final outcome is unpredictable. Alleles, in particular rare alleles, are likely to completely disappear from populations as a result of genetic drift, and previously polymorphic (variable) gene loci may become monomorphic (fixed). A sudden reduction of the population size for one or several generations is often described as a genetic "bottleneck." Bottlenecks result in reduced genetic variation even after later population expansion owing to the effects of genetic drift while populations are small. Genetic differentiation among partially or completely isolated populations is expected to increase as a result of genetic drift.
- **Selection** is a consequence of the differing abilities of organisms to produce offspring (fitness). Selection does not increase multiplicity, but often contributes to the maintenance of genetic diversity by stabilizing polymorphisms, for example, owing to a selective advantage of heterozygous genotypes. In general, selection results in an improved adaptedness (Chap. 7). The basis of selection is the different survival of genotypes (viability selection) and the different contributions of genotypes to the formation of a progeny generation (fertility selection).

Population geneticists are concerned with the effect of these evolutionary factors on patterns of genetic variation within and among populations

(Gillespie 1998). Genetic structures result from complicated interactions between the different evolutionary factors in natural and managed populations.

The role of different evolutionary factors for the creation and maintenance of the extraordinary species richness of tropical forests has been widely discussed in the past. A high frequency of selfing and restricted gene flow and migration resulting in small populations, which become differentiated from each other mainly owing to the effects of genetic drift, have been proposed as the main forces for the evolution of the enormous species richness in tropical forests by Fedorov (1966). This view was challenged by Ashton (1969), who argued in favor of more efficient means of gene flow through seed and pollen than previously thought. The evolution of species diversity is promoted by selective differentiation in partially isolated subpopulations which eventually develop reproductive barriers. The latter proposition is greatly supported by maker-based assessments on genetic variation patterns in tropical forest species (Chap. 3) and on the reproduction system of tropical forest plants (Chaps. 5, 6).

2.6 Phenotypic Variation

The vast majority of directly observable traits of trees are an outcome of the interaction between the genetic constitution of a plant at a large number of gene loci (genotype, G) and the environmental conditions (E). Thus, it holds for the phenotype (P) that $P = G \times E$, where $\times$ symbolizes the interaction between environmental and genotypic effects, which often differs from simple additivity of both components.

Phenotypic traits deserve particular interest if they are important for the adaptation of plants to their environment (adaptive traits) or for the value of a tree (economic traits). Adaptive traits such as the tolerance against biotic (pests) or abiotic (e.g., draught) stress and economic traits such as volume growth are controlled by both genetic and environmental factors for most plant species. The analysis of phenotypic traits in a breeding context will be discussed in Chap. 12 in more detail. It is also described for forest trees in other textbooks (Wright 1976; Zobel and Talbert 1984; Williams et al. 2002).

The majority of the gene markers described in Sect. 2.3.1 are unlikely to have strong, recognizable effects on adaptive traits or traits of economic importance. They are regarded as "neutral" or "nearly neutral" markers with regard to an adaptation to the environment; however, this does not preclude their application to monitor adaptive processes (Ziehe et al. 1999; Sect. 7.2).

In line with the definition of genetic structures, the frequency distribution of particular phenotypes is defined as the phenotypic structure of a population. Direct and indirect effects of the environment on the phenotypic structure are illustrated in Fig. 2.7. Apart from the direct impact of the environment

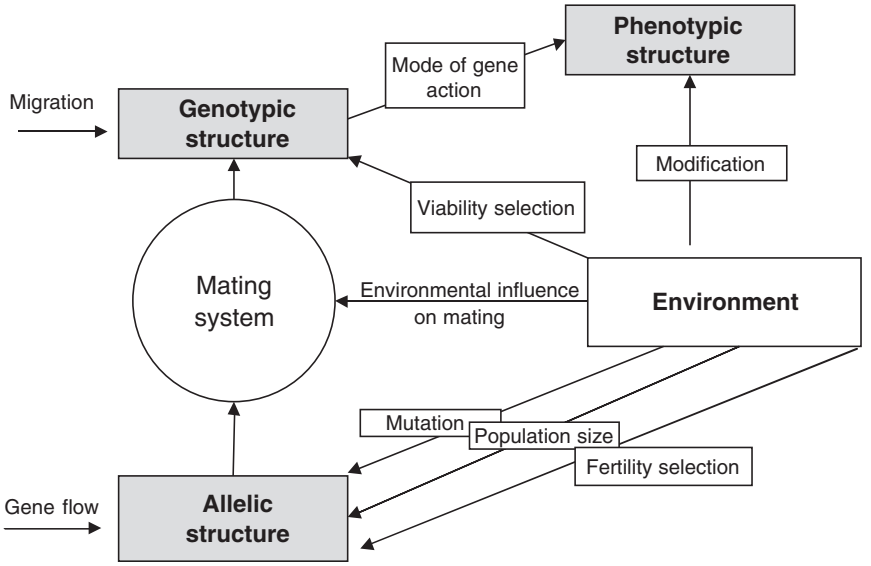

Fig. 2.7. The environmental impact on phenotypic structures. (Adapted from Hattemer and Müller-Starck 1990)

on the phenotype ($P = G \times E$), there are numerous impacts of the environment on genotypic structures. The environmental conditions influence mutation rates, are crucial for population sizes and, hence, the importance of genetic drift, cause changes of allelic structures owing to fertility selection and genotypic structures owing to viability selection (Chap. 7), and have an impact on the mating system (Chap. 6). Thus, manifold human alterations of environmental conditions do not only directly effect phenotypes of forest plants, but also change the genotypic structure of populations.

2.7 Recommended Literature

The molecular basics of genetics is covered in many recently published textbooks, such as the one written by Griffiths et al. (2000). A useful and detailed description of genetic markers and their application to study plant genetics is given by Weising et al. (2005). Altukhov and Salmenkova (2002) provide a comprehensive survey on currently available markers to study genetic variation of plants. The need for inheritance studies in order to identify gene markers was emphasized by Hattemer (1991). The basics of population genetics are covered by Hedrick (2000). The book of Altukhov (2006) represents another introduction to population genetics putting genetic variation in an ecological context.

http://www.springer.com/978-3-540-37396-4

Tropical Forest Genetics
Finkeldey, R.; Hattemer, H.H.
2007, XII, 316 p. 44 illus., Hardcover
ISBN: 978-3-540-37396-4