Chapter 1: Introduction

1.1. The genetic system of species

Forests of the tropics harbour very high species richness, which was explained alternatively by different authors. FEDOROV (1966) regarded it as a consequence of small population sizes and genetic isolation but ASHTON (1969) assumed that population sizes even in species-rich forests were high due to the exchange of genes by efficient pollen and seed distributors in semi-isolated subpopulations. In fact, we cannot explain the evolution of the species richness in tropical forests without investigating the genetic system of the species (BAKER, 1970). Tropical forests harbour not only high biodiversity but also a great diversity of genetic systems (FINKELDEY and HATTEMER, in prep.).

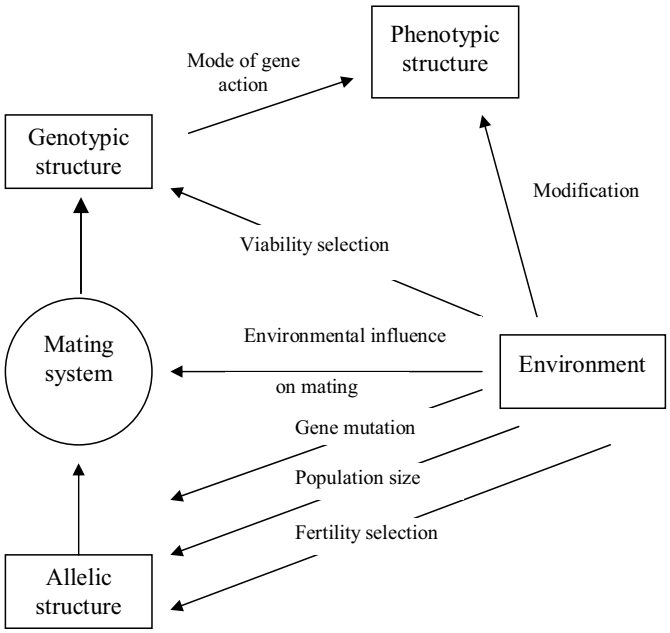

Figure 1. Evolutionary factors and their influence on the phenotypic structure of a population (according to HATTEMER and MÜLLER-STARCK, 1990).

The organization and transmission of genetic information and the kind and amount of genetic combinations produced by a population is determined by the genetic system (RIEGER *et al.*, 1976), which is composed of several subsystems, namely sexual, gene flow, mating, incompatibility, recombination, reproductive, selection and adaptation systems (GRANT,

1975). The subsystems interact to decide on the amount and distribution of genetic variation within and among populations of a species (Figure 1). Just as the other subsystems, the gene flow and mating system of a species are very closely related and influence the genetic structure of populations.

1.2. Genetic variation and genetic structure

Genetic variation is a common characteristic of most species (HATTEMER, 1991b). The maintenance of genetic diversity is a major focus of conservation biology. Genetic diversity is of utmost importance for the survival of populations because it helps populations to evolve and to adapt to environmental change. Thus, genetic variation is the basis for evolutionary processes.

Genetic structures are frequency distributions of genetic types such as alleles or genotypes within populations. Any change of the genetic structure of a population at at least one gene locus can be regarded as evolution. Processes that change the genetic structure of a population are called evolutionary factors, including mutation, gene flow and migration, genetic drift, selection and mating system. In other words, evolutionary factors act together to shape the genetic structures of all populations of biological organisms (FINKELDEY and HATTEMER, in prep.).

The dynamics of genetic structures cannot be observed in single organisms but only at the level of populations. Genetic information of an organism can be transferred to its offspring which may have different genotypes. Observation of genetic structures within a single population or several populations is required to investigate genetic systems and evolution.

The most important genetic structures are:

- The genotypic structure, i.e. the frequency distribution of genotypes in a population.
- The allelic structure, i.e. the frequency distribution of alleles in a population.

Both parameters are often seen as different basic measures to describe genetic variation of a population at a gene locus.

Activities in tree breeding and conservation of genetic resources should be also based on an understanding of patterns of genetic variation of a species. Investigation of the spatial distribution of genetic variation is clearly necessary to develop conservation strategies of species and for the management of forests and ecosystems. It can be done by genetic inventories, in which the genetic structure and variation of populations are inferred from data that are also used for the analysis of the mating system or other evolutionary factors.

1.3. The mating system

Outcrossing and selfing rates

The mating system determines the fusion of sexually different gametes into zygotes and is thus crucial for the genotypic structure of the next generation. Knowledge on the mating system and gene flow contributes to understand how genetic variation is distributed within and among individuals and populations and to explain the spatial and temporal patterns of genetic structure. Information on the mating system is also important for successful genetic management and breeding programs of a species.

The analysis of the mating system usually includes estimations of the rate of selfing and outcrossing. Selfing indicates the fusion of two sexually different gametes produced by the same monoecious or hermaphroditic plant. The selfing rate (s) is the relative proportion of ovules fertilized by pollen of the same individual. The fertilization of ovules by pollen from other individuals is called outcrossing. The outcrossing rate ($t = 1 - s$) is the relative proportion of ovules fertilized by "foreign" pollen. Measuring the outcrossing rate based on genetic variation at gene marker loci has advantages compared to other methods including those derived from the observation of morphological traits. The selfing/outcrossing rate can be estimated using two methods of practical importance, which are based on: (i) rare alleles and (ii) a mixed mating model. The second method is mainly used in this study.

The mixed mating model based on genetic variation at single gene loci was proposed by BROWN and ALLARD (1970) to estimate out-crossing rates assuming that each seed results from either selfing or random outcrossing. In addition to random mating, the model also assumes homogenous pollen allele frequencies in the effective pollen across all seed trees. In addition, genetic change due to mutation and selection after fertilization as well as assortative mating are assumed to be absent. Progenies of several seed trees are collected and analyzed.

RITLAND and JAIN (1981) suggested the mixed mating model based on multiple, simultaneously observed gene loci that is supposed to result in more precise estimates of the outcrossing rate and is less susceptible to violations of model assumptions.

Biparental inbreeding

Based on the mixed mating model, it is also possible to estimate the extent of biparental inbreeding and correlated mating. Inbreeding including biparental inbreeding (or mating between relatives) causes increased homozygosity relative to random mating (CROW and KIMURA, 1970; RITLAND, 2002). The difference between multilocus and single-locus estimates of outcrossing is used to characterize the degree of biparental inbreeding (RITLAND, 2002). This difference is often underestimated because it depends on the number of loci investigated. The precision of the estimate depends on the number of investigated loci and their variability.

Correlated mating

An aspect of interest in studies of mating systems is the correlated mating which estimates the degree to which siblings share the same father (male parent). This is an outcome of the pollination syndrome or population structure and has implications for the effectiveness of certain types of natural selection (RITLAND, 2002). It can be estimated based on the correlated-mating model proposed by RITLAND (1989).

1.4. Gene flow

The gene flow system describes how genetic information is transported within and among populations. Because plants and their female gametes are immobile, two basic mechanisms are important for the gene flow, namely transport of pollen and transport of seed (or vegetative propagules). Gene flow means the movement of pollen in its narrow sense and both the movements of pollen and seeds in its broad sense (FINKELDEY and HATTEMER, in prep.). It is a key factor for the spatial genetic structure in spatially distributed species (SMOUSE *et al*., 2001).

The change of forested landscapes by human influence is thought to interrupt the gene flow processes and might isolate populations. Populations isolated may loose their genetic

variation and fitness due to genetic drift and inbreeding (SMOUSE *et al*., 2001). Thus, it is necessary to assess the influence of human impacts on aspects of the gene flow system.

The gene flow between populations is often estimated indirectly by an analysis of the genetic structure (*F*- statistics; WRIGHT, 1969; WRIGHT, 1978) to infer the average effective number of migrants exchanged per generation. The approach is based on assumptions of evolutionary equilibrium and selective neutrality. Another approach, which is considered to be a direct measure, is based on parentage analysis or paternity-exclusion analysis of rare (foreign) alleles to estimate pollen gene movement within and among populations (DEVLIN and ELLSTRAND, 1990; ADAMS and BIRKES, 1991; SMOUSE *et al*., 1999). However, it needs to analyze a large number of local pollen donors and progeny per female. It is quite often observed that a rare allele comes to a population from outside and its origin is difficult to track.

The TwoGener approach

The TwoGener model was proposed by SMOUSE *et al*. (2001) to analyze the pollen flow across a landscape. This is a novel two-generation (parent-offspring) approach that allows to quantify heterogeneity among the male gamete pools gathered by maternal trees scattered across the landscape and to estimate mean pollination distance and effective neighborhood size. It is an indirect method of estimating pollen movement that is a hybrid of traditional paternity and genetic structure analysis (SMOUSE and SORK, 2004).

The effective number of mates (equal to $1/r_p$, which is comparable to the effective neighborhood size) can also be estimated using the correlated mating (r_p), which was proposed by RITLAND (1989; also see SAMPSON, 1998).

In this study, both approaches are used to infer pollen movement patterns within populations. The results are compared with those of the paternity exclusion analysis.

1.5. Genetic markers

Genetic markers are traits controlled by one or a few loci. Different types of genetic markers have been used in genetic studies: morphological polymorphisms, colour traits, secondary

products of metabolic pathways, isoenzymes (or isozymes) and DNA markers (for example: Restriction fragment length polymorphisms – RFLPs, Random amplified polymorphic DNA – RAPDs, Amplified fragment length polymorphisms – AFLPs, Microsatellites or Simple sequence repeats – SSRs).

Variation of morphological polymorphisms like rare leaf forms and colour traits such as the purpurea form of some trees in the temperate zone are not common in natural populations (HATTEMER, 1991a). They used to be main markers applied in the early time of genetic research. Quantitative traits were mostly used to study patterns of genetic variation until the 1970s. However, such studies were laborious and slow. It is usually impossible to assess the number of loci involved in the observed variation of a quantitative trait as well as the genetic structures at these loci (FINKELDEY and HATTEMER, in prep.). The use of the content of certain secondary products, in particular terpenes, as a genetic marker resulted in many difficulties such as large experimental effort, difficulties in performing inheritance studies, low polymorphism and dominant alleles. Since the beginning of the 1970s, the application of new biochemical and molecular markers has enabled genetic researchers to work with genetic markers of better resolution and with more loci. Recently, many methods have been developed to investigate DNA polymorphisms allowing to choose genetic markers of a resolution suitable for particular study objectives. Two genetic markers, isozymes and AFLPs, were used in the present study. The mating system was studied using isozymes while the structure and genetic variation within and among populations was investigated with both marker types.

1.4.1. Isozymes

Analysis of isozymes has been widely used over the past several decades as a powerful technique to investigate many aspects of the genetic system of a large number of species, from fruit flies to wild herbs and to humans (LIENGSIRI *et al*., 1990). They have been the most important tool to study the genetic system of tropical forest trees in the past (FINKELDEY and HATTEMER, in prep.).

Biochemically, isozymes are enzymes that catalyze identical biochemical reactions in the metabolism but may differ with regard to their intracellular location. In other words, different forms of functionally identical enzymes are called isozymes (LIENGSIRI *et al*., 1990). They are composed of amino acids. Amino acids are organic molecules that have a basic amino (NH_2)

group and an acidic carboxyl (COOH) group. They can bind with each other by peptide bonds to form polypeptide chains. Polypeptides are primary gene products. Thus, their variation is closely related to the genetic information of their coding regions in the DNA. Because the nucleotide sequence in the DNA codes for the corresponding sequence of amino acids, a change in the DNA sequence possibly results in a change in the respective sequence of amino acids. If such a mutation occurs, the sequence of amino acids will be slightly different from the sequence before the mutation. This might lead to a slight change of the enzyme structure, but does normally not affect the enzyme function. If this alters the net charge or the shape of the enzyme, the two (original and mutated) forms of the enzyme will differ in their electrophoretic mobility. This difference can be detected by suitable electrophoresis techniques. Thus, electrophoresis of enzymes can reflect the genetic difference resulting from mutations. While different forms of a gene that produce structurally altered proteins are termed alleles, structurally different enzymes produced from different alleles of one and the same gene locus are called allozymes (LIENGSIRI *et al*., 1990).

In spite of new markers with higher resolution, isozymes are still important genetic markers until today because of the following properties:

- Many isozymes are environmentally constant.
- They can be observed in many different ontogenetic stages.
- A small amount of tissue is needed for investigations.
- They are codominant.
- A comparatively large amount of variation is observed in most tree species.
- Many isozymes and samples can be analyzed simultaneously.
- Laboratory techniques are comparatively simple. Equipment and chemicals are relatively cheap.
- It is possible to investigate the variation of allozymes simultaneously at many loci.
- Inheritance studies have been successfully carried out for many tree species.

The main disadvantage of isozymes is that they are less variable as compared to certain other marker types. But hypervariable codominant microsatellite markers such as microsatellites (simple sequence repeats; SSRs) are expensive and difficult to develop. It is often not possible to apply already developed SSR-primers for a particular species to species of different genera. Thus, isozymes are suitable gene markers that are nowadays still widely applied to investigate

the mating and gene flow systems of plant populations. As genetic markers, they have often been used for studying the structure and genetic variation of populations. The use of isozymes for systematic studies was reviewed by BUTH (1984) and BUTH and MURPHY (1999). Many enzyme systems and large amounts of genetic variation at enzyme gene loci have been described for many trees including some dipterocarps.

1.4.2. AFLPs (Amplified Fragment Length Polymorphisms)

Amplified Fragment Length Polymorphisms are molecular markers generated by the use of selective PCR (polymerase chain reaction) amplification of restriction fragments from a total digest of genomic DNA (VOS *et al*., 1995). The technique basically includes three steps: (i) restriction of the DNA and ligation of oligonucleotide adapters, (ii) preselective and selective PCR amplification of sets of restriction fragments, and (iii) analysis of the amplified fragments by gel electrophoresis or in an automated sequencer. AFLPs have the following advantages:

- No need to store fresh material.
- The technique can be applied for DNA of any origin and complexity.
- Fingerprints are produced without prior sequence knowledge using a limited set of generic primers.
- A large number of fragments are generated by a single PCR reaction.
- Only a small amount of tissue or DNA is needed for analysis.
- Markers are usually highly polymorphic.

Although the technique has been applied primarily for mapping studies, the high polymorphism revealed by AFLPs has also encouraged their use in other fields. In population genetics, AFLPs are used mainly to study genetic variation. AFLPs proved to be useful to detect genetic differentiation among individuals, populations and independently involving lineages, such as species (MUELLER and WOLFENBARGER, 1999). However, the use of AFLPs for studies of the mating system is limited because they are dominant markers. To date, AFLPs are not reported to have been used to study population genetics of dipterocarps.

1.5. The Dipterocarpaceae and their genetic variation

1.5.1. *Dipterocarpaceae*

Dipterocarps are trees with alternate entire leaves and pentamerous flowers. The family type genus is the Asian *Dipterocarpus* Gaertn.f.. The family *Dipterocarpaceae sensu stricto* is restricted to Asian plants while the *Dipterocarpaceae sensu lato* include three subfamilies: *Dipterocarpoideae* in Asia, *Pakaraimoideae* in South America, and *Monotoideae* in Africa and South America (Figure 2). The position of the new African and South-American taxa relatively to the Asian group is still in debate. Consequently, the family contains either 15, 16 or 19 genera and 470 to 580 or more species depending on authors. However, the numbers have reduced with the increase in collections and systematic expertise (MAURY-LECHON and CURTET, 1998). In Vietnam, there are 6 genera and 46 species of the *Dipterocarpaceae* (PHAM-HOANG, 1991-1993).

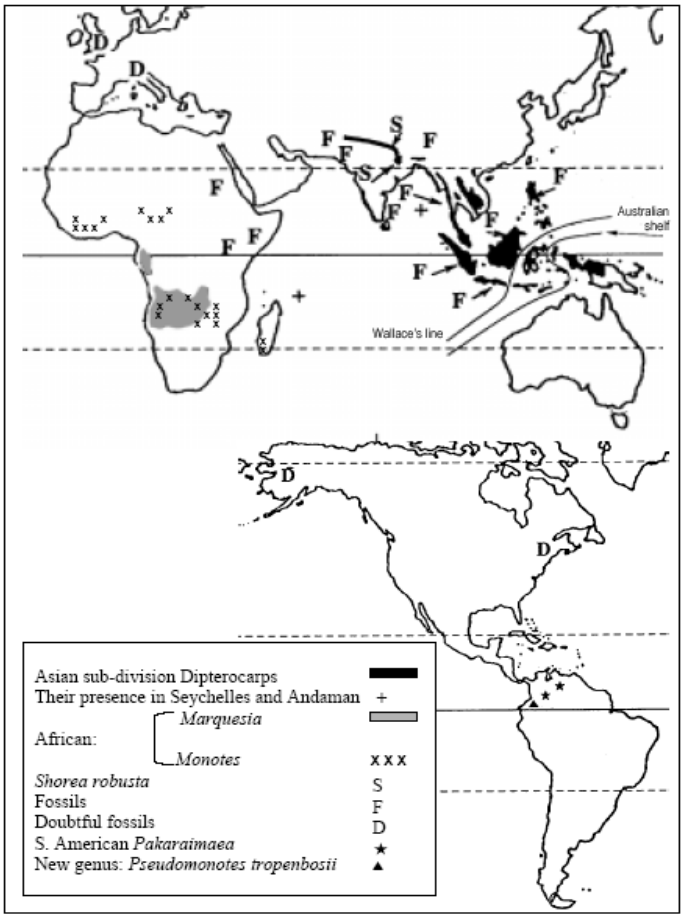

Figure 2. Distribution of dipterocarps in the world (from MAURY-LECHON and CURTET, 1998)

Species of the Asian subfamily occur from the Seychelles and Sri Lanka through southern and eastern India to Hainan, southeast to Sundaland where the species diversity reaches its zenith and east to New Guinea and the Louisiades. They are mostly abundant at low altitude and can reach an altitude of 2000 meters. Normally, they are dominant in the forest canopy (ASHTON, 1989).