Cover photos from upper left to lower right: Young shoot of *Fraxinus excelsior*, photograph Hartmut Weichelt; Receptive female strobilus of *Picea abies*, photograph Kjell Lännerholm; *Pseudotsuga menziesii* in an Oregon forest, photograph Gösta Eriksson; Aurea and strawberry *Picea abies* with assumed monohybrid inheritance, photograph Hartmut Weichelt.
An introduction to Forest Genetics

Second Edition

Gösta Eriksson
Inger Ekberg
David Clapham

ISBN 91-576-7190-7
This version of An Introduction to Forest Genetics is somewhat expanded compared to the book printed in 2001. We were encouraged to "publish" the revised version of the textbook on internet. Undergraduate students are the target group as well as graduate students with limited teaching in forest genetics. In this web version it is possible to use colors to a much larger extent than it is possible in a printed version. Colors can be used to strengthen the message. Thus, yellow and brown symbolize that a material originates from dry conditions or that the material is exposed to low-water treatment. Conversely green symbolizes wet origin or well-water treatment. Red and blue are used in an analogous way to symbolize high and low temperature, respectively. Most of the graphics and editing had to be done by ourselves, which explains that the finish of our product may not be totally perfect. Without the advice and help from Kjell Lännerholm, Björn Nicander, Johan Samuelsson and Hartmut Weichelt the editing would have been more troublesome. We express our sincere thanks to them.

In a separate folder power point presentations of the graphics of the different chapters will be published. You are free to use them. If you do this please keep the logotype on the graphics.

Uppsala, December 2006

Gösta Eriksson  Inger Ekberg  David Clapham
Chapter 8 Variation within populations.................111
Examples of variation among families for various traits........................................111
Heritabilities and coefficients of additive variation..............................................113
Why is there such a large within-population variation in *Picea abies* and *Pinus sylvestris* and many other tree species?.................................................................118
Summary.................................................................118
Further reading.......................................................118

Chapter 9 Forest tree breeding...............................119
What should be considered before the start of a breeding programme?......................119
Various types of tree breeding.........................121
   Species selection..............................................122
   History..........................................................123
   Long-term breeding.........................................124
   Population functions.......................................124
   Recurrent selection........................................125
   Multiple Population Breeding System..............126
   Sublining........................................................128
   Nucleus breeding..............................................128
   Short-term breeding........................................129
   One concrete example of a breeding strategy....129
   Selection of plus trees......................................130
   Seed orchards................................................130
   Clonal seed orchards.......................................130
   After effects....................................................132
   Vegetative propagation and clonal forestry.....134
   Progeny testing and mating design..................137
   Nested matings...............................................139
   Point of time for selection.............................139
   Early tests......................................................140
   Progress in breeding........................................141
   The sustainability of the gain........................144
   Summary........................................................145
   Further reading..............................................145

Chapter 10 Plant production....................................147
Summary ..........................................................149
Further reading................................................149

Chapter 11 Forest tree gene conservation.............151
The three cornerstones of gene conservation..................................................152
Objectives in gene conservation.................................................................152
   Prime objectives........................................152
   Other objectives.........................................153
   Genetic structure.........................................154
   In situ and ex situ gene conservation..............154
   Target species..............................................154
   Grouping of species in gene conservation.......155
   Ecological characteristics.............................155
   Involvement in breeding activities..................155
   Biological threats..........................................156
   Forest tree gene conservation methods...........156
   Safeguarding the potential for adaptation......158
   Methods for other objectives in gene conservation......................................164
   Miscellaneous...............................................166
   Species hybridisation and gene conservation...167
   Sustainable forestry........................................168
   Genetic pollution..........................................170
   Different levels of a conservation programme..170
   Summary......................................................171
   Further reading..............................................171

Chapter 12 Consequences of different breeding activities and silvicultural methods for the new generation of trees.................................173
The demand for genetic variation in the production population............................175
Summary..........................................................176
Further reading................................................176

Glossary............................................................179
Chromosome cytology

In this chapter we deal with chromosomes and basic concepts of the Mendelian genetics. We present further the two main types of nuclear divisions, the asexual nuclear division - mitosis - occurring in the somatic cells and the sexual nuclear division - meiosis - occurring in the gamete-forming tissues. Different types of chromosomal aberrations will be presented. Finally, the time of meiosis during the year and climatically induced injuries in meiosis will be highlighted.

Chromosome cytology deals with microscopic studies of chromosome number, size, morphology and chromosome behaviour during nuclear divisions.

Already in the middle of the 19th century, the chromosomes were discovered. They got the name chromosome because they became visible when stained with basic dyes, usually red. Cytology advanced rapidly during the latter part of the 19th century thanks to the improved light-microscope technique. More knowledge of the chromosomes could be acquired. The main features of fertilization at the cell level in animals and plants were revealed at that time. The divisions of the cell nucleus were described, both in the somatic cells - mitosis - and in the germ cells - meiosis.

Very soon after the rediscovery of the laws of Gregor Mendel, it became clear that the chromosomes were the potential carrier of the genes. The final proof was presented by the American geneticist Thomas Morgan and his co-workers Calvin Bridges and Alfred Sturtevant during the 1910s and onwards. Thus, in 1916 when studying the inheritance of the eye colour in the fruit fly (*Drosophila melanogaster*), Bridges observed that rare exceptions from the expected segregation in the progeny appeared when a large number of individuals was studied. He could verify that the exceptions were caused by the formation of abnormal egg cells with two X chromosomes (sex chromosomes, see below) instead of normal egg cells with one X chromosome. This provided an unequivocal evidence that genes are located on the chromosomes. Morgan’s group further observed that genes were not always inherited independently but sometimes behaved as if they were linked. Morgan’s group also revealed that exchanges of chromosome segments between two homologous chromosomes occur, a phenomenon called crossing-over (for the meaning of homologous chromosomes see Fig. 1-2).

Because of their size, the chromosomes have mainly been studied in the light microscope, using preparations of very thin sections from various tissues, or squash preparations where the cells are pressed (squashed) into a unicellular layer. Thanks to these light-microscope studies, chromosome number, size and morphology in a large number of plants and animals are now known. Studies in the 1960s and onwards using electron microscopy, have provided important information on the fine structure of the chromosomes.

Our understanding of the nature of gene action at the biochemical level is based on studies initiated already during the 1930s. In the beginning of 1940s, the American geneticists George Beadle and Edward Tatum launched the one gene - one enzyme hypothesis. Today we prefer to speak about one gene - one polypeptide since later studies have shown that many enzymes contain two or more polypeptides each being a product of a specific gene (see also Chapter 2). A polypeptide consists of amino acids.

The identification of the macromolecule, deoxyribonucleic acid, DNA, as the carrier of the hereditary information meant a great breakthrough for the genetic research in the middle of 1940s. The next great breakthrough came in 1953, when the double-helical structure of the DNA molecule was elucidated, see Chapter 2.

Karyotype

The karyotype of a species describes its chromosomes including chromosome number, size and morphology. In some instances, the karyotype can provide information on the relationship between species (Fig. 1-1).
In gymnosperms, nearly all species have two sets of chromosomes in their cells. The chromosomes appear in pairs. The chromosomes of such a pair are said to be homologous (Fig. 1-2). On every chromosome there is a centromere which can usually be seen as a constriction at a specific site on the chromosome. The centromere pulls the chromosome to one of the two spindle poles of the cell during the nuclear division. The centromere can be found anywhere on the chromosome and it divides the chromosome in two arms except when the centromere is located at the end of the chromosome. If the two arms are similar in length the chromosome is said to be metacentric, if the two arms are of unequal length the chromosome is acrocentric and if the centromere is located at a terminal position, at the telomere, the chromosome is telocentric. Some chromosomes have so-called secondary constrictions or nucleolar organizers associated with nucleoli (involved in ribosomal RNA synthesis) that in some species serve as useful landmarks for the identification of individual chromosomes.

![Figure 1-2: Two pairs of homologous chromosomes are shown. Genes with dominant alleles (capital letters) and genes with recessive alleles (small letters) located at loci in the same or different linkage groups are indicated.](image)

Typically, pine and spruce chromosomes are metacentric and their length very similar, and they often lack other landmarks that can be useful in cytogenetic studies. Because of this, the individual chromosomes are difficult to identify in common squash preparations. Recently, however, new techniques have been developed using *in situ* hybridization combined with fluorochrome staining that

<table>
<thead>
<tr>
<th>Chromosome 1</th>
<th>Chromosome 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>centromere</td>
<td>alleles</td>
</tr>
<tr>
<td>genes on the same chromosome belong to the same linkage group</td>
<td>alleles</td>
</tr>
<tr>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>d</td>
<td>e</td>
</tr>
<tr>
<td>Homologous chromosomes</td>
<td>Homologous chromosomes</td>
</tr>
</tbody>
</table>

Table 1-1. Chromosome number in some woody plants in Sweden. x = monoploid number, 2x = diploid, 3x = triploid, 4x = tetraploid

<table>
<thead>
<tr>
<th>Hardwoods</th>
<th>2x = 26</th>
<th>2x = 28</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Acer platanoides</em></td>
<td>3x = 39</td>
<td>2x = 24</td>
</tr>
<tr>
<td><em>Alnus glutinosa</em></td>
<td>2x = 28</td>
<td>2x = 28</td>
</tr>
<tr>
<td><em>Alnus incana</em></td>
<td>2x = 28</td>
<td>2x = 28</td>
</tr>
<tr>
<td><em>Betula nana</em></td>
<td>2x = 28</td>
<td>2x = 28</td>
</tr>
<tr>
<td><em>Betula pendula</em></td>
<td>4x = 56</td>
<td>4x = 56</td>
</tr>
<tr>
<td><em>Betula pubescens</em></td>
<td>2x = 24</td>
<td>2x = 24</td>
</tr>
<tr>
<td><em>Fagus sylvatica</em></td>
<td>2x = 24</td>
<td>2x = 24</td>
</tr>
<tr>
<td><em>Fraxinus excelsior</em></td>
<td>2x = 46</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conifers</th>
<th>2x = 22</th>
<th>2x = 24</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Juniperus communis</em></td>
<td>2x = 24</td>
<td>2x = 24</td>
</tr>
<tr>
<td><em>Larix decidua</em></td>
<td>2x = 24</td>
<td>2x = 24</td>
</tr>
<tr>
<td><em>Picea abies</em></td>
<td>2x = 24</td>
<td>2x = 24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conifers</th>
<th>2x = 24</th>
<th>2x = 24</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pinus sylvestris</em></td>
<td>2x = 24</td>
<td>2x = 24</td>
</tr>
<tr>
<td><em>Taxus baccata</em></td>
<td>2x = 24</td>
<td>2x = 24</td>
</tr>
<tr>
<td><em>Tilia cordata</em></td>
<td>2x = 24</td>
<td>2x = 24</td>
</tr>
<tr>
<td><em>Ulmus glabra</em></td>
<td>2x = 24</td>
<td>2x = 24</td>
</tr>
<tr>
<td><em>Populus tremula</em></td>
<td>2x = 38</td>
<td>2x = 38</td>
</tr>
<tr>
<td><em>Quercus robur</em></td>
<td>2x = 24</td>
<td>2x = 24</td>
</tr>
<tr>
<td><em>Salix sp</em></td>
<td>2x = 28</td>
<td></td>
</tr>
<tr>
<td><em>Sorbus aucuparia</em></td>
<td>2x = 34</td>
<td>2x = 34</td>
</tr>
<tr>
<td><em>Sorbus intermedia</em></td>
<td>4x = 68</td>
<td></td>
</tr>
<tr>
<td><em>Tilia cordata</em></td>
<td>2x = 82</td>
<td>2x = 82</td>
</tr>
<tr>
<td><em>Ulmus glabra</em></td>
<td>2x = 28</td>
<td>2x = 28</td>
</tr>
<tr>
<td><em>Betula pubescens</em></td>
<td>2x = 28</td>
<td>2x = 28</td>
</tr>
<tr>
<td><em>Alnus incana</em></td>
<td>4x = 76</td>
<td>4x = 76</td>
</tr>
<tr>
<td><em>Alnus glutinosa</em></td>
<td>2x = 28</td>
<td>2x = 28</td>
</tr>
<tr>
<td><em>Fraxinus excelsior</em></td>
<td>2x = 46</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1-2. Two pairs of homologous chromosomes are shown. Genes with dominant alleles (capital letters) and genes with recessive alleles (small letters) located at loci in the same or different linkage groups are indicated.
allow the identification of all 12 pairs of chromosomes (e.g. *Pinus elliottii*). Behind this newly-awakened interest in karyotype studies lies the need to assign linkage groups to physical chromosomes to be able to integrate physical and genetic maps.

In most conifers native to the northern hemisphere, the chromosome number is 24, but also 22 or 26 exist. The chromosomes are large and therefore easy to study in a light microscope. In contrast, the hardwoods often have very diminutive chromosomes. In Table 1-1, examples of chromosome numbers in some Swedish forest trees are given.

**Locus, genes, alleles, homozygosity, heterozygosity, dominant and recessive traits**

The hereditary units, the genes, are sited linearly on the chromosomes (cf Fig. 1-2). The site on a chromosome, where a certain gene is located, is called a **locus** (plural **loci**). Letters are used for symbolizing the genes in accordance with the suggestion of Gregor Mendel. Genes existing in more than one alternate form at the same locus are called **alleles**. **AA, Aa** or **aa** symbolize genotypes where A and a are alleles. An individual that carries both the **recessive** (a) and the **dominant** allele (A) is said to be **heterozygous**. Individuals that have allele A or a in a double set are said to be **homozygous**. Individuals with the genotypes **AA** and **Aa** have the same appearance or **phenotype** if A is completely dominant over a, while individuals with the genotype **aa** show another phenotype than **AA** or **Aa**. A diploid individual cannot have more than two different alleles at the same locus. But in a population of individuals more than two alleles belonging to the same locus can be found. This situation is termed a series of multiple alleles. We can also talk about polymorphism, i.e. when two or more alleles exist at the same locus in a population. With three alleles **a**, **a**, **a**, the following six genotypes will be formed, **a,a**, **a,a**, **a,a**, **a,a**, **a,a** and **a,a**.

If all these genotypes can be separated phenotypically, the alleles are said to be codominant. The alleles that determine the human ABO blood groups are examples of multiple alleles.

In higher animals and in those plants with male and female flowers on separate individuals, sex is determined by specific so-called sex chromosomes. In plants the sex chromosomes are not always discernible. In humans, there are one X chromosome and one Y chromosome. If an egg cell is fertilized by a sperm cell and both cells carry the X chromosome, this will give rise to a girl. Normally, a boy has the constitution XY and has received the Y chromosome from his father and the X chromosome from his mother. The Y chromosome is much smaller in size and lacks most of the genes located in the X chromosome. This explains why recessive defects determined by genes in the X chromosome, so-called X-linked genes, for example red-green colour blindness and hemophilia, are more common in males than in females. The recessive allele is expressed although the allele occurs in a single dose, because the Y chromosome lacks this locus and therefore there is no wildtype counterpart of the allele. This further explains why such defects omit one generation since a boy affected with red-green colour blindness has an allele, which originates from the X chromosome of his mother’s father.
Mitosis

Mitosis is the division of the cell nucleus, that ensures that the two daughter nuclei receive the same number and type of chromosomes as the parental nucleus. It is usually accompanied by the division of the cell, cytokinesis. Before entering mitosis, the chromosomes have duplicated and consist of two sister chromatids. In mitosis, the two chromatids separate and move to opposite spindle poles of the cell. The cell divides producing two daughter cells, each with an identical set of chromosomes. We use to divide mitosis into five stages called prophase, prometaphase, metaphase, anaphase, and telophase. The intervening stage between two mitoses is called interphase. The synthesis of DNA takes place during this stage.

Meiosis

Meiosis is the division of the cell nucleus, that ensures that the two daughter nuclei receive the same number and type of chromosomes as the parental nucleus. It is usually accompanied by the division of the cell, cytokinesis. Before entering mitosis, the chromosomes have duplicated and consist of two sister chromatids. In mitosis, the two chromatids separate and move to opposite spindle poles of the cell. The cell divides producing two daughter cells, each with an identical set of chromosomes. We use to divide mitosis into five stages called prophase, prometaphase, metaphase, anaphase, and telophase. The intervening stage between two mitoses is called interphase. The synthesis of DNA takes place during this stage.

Mitosis

Mitosis is the division of the cell nucleus, that ensures that the two daughter nuclei receive the same number and type of chromosomes as the parental nucleus. It is usually accompanied by the division of the cell, cytokinesis. Before entering mitosis, the chromosomes have duplicated and consist of two sister chromatids. In mitosis, the two chromatids separate and move to opposite spindle poles of the cell. The cell divides producing two daughter cells, each with an identical set of chromosomes. We use to divide mitosis into five stages called prophase, prometaphase, metaphase, anaphase, and telophase. The intervening stage between two mitoses is called interphase. The synthesis of DNA takes place during this stage.

Meiosis

It is obvious that the organisms must have a mechanism that prevents a doubling of the chromosome number after each generation. The formation of embryo sacs and pollen is preceded by a special process - meiosis - during which the nucleus divides twice but the chromosomes replicate only once, so that the chromosome number is halved (Picture 1-1). Meiosis occurs in specialized cells. The male cells of a tree are known as pollen (or microspore) mother cells while the female cells are called megaspore (or macrospore) mother cells.

It is in the first cell division, that the two homologous chromosomes separate, resulting in two daughter nuclei, each with only one of the two homologous chromosomes. For example, in Norway spruce, which consists of 12 pairs of homologous chromosomes, each daughter nucleus receives one chromosome 1, one chromosome 2 and so on up to one chromosome 12 (Table 1-1). This means that the chromosome number is halved but the DNA content is the same as in a diploid nucleus. This halving of the chromosome number is accomplished by the lateral pairing or synopsis of the homologous chromosomes during the early stages of the first cell division. An associated pair of homologous chromosomes is called a bivalent. The bivalents exist until the onset of anaphase I, when the two chromosomes in the bivalents separate and move to opposite poles. In addition to the random recombinations of chromosomes occurring in anaphase I (see Box 1-1), exchanges of segments between homologous chromosomes, so-called crossing-over, take place during pachytene - one of the early stages of meiosis - resulting in additional recombinations between genes.

The second cell division resembles a normal mitosis except that it is not preceded by chromosome duplication and the two separating sister chromatids have not an iden-
In summary, the function of meiosis is:

- To halve the chromosome number so that a pollen grain or an embryo sac will only contain half the chromosome number.
- To recombine genes from different chromosome pairs.
- To recombine genes from the same chromosome pair.

Recombination is an essential function of meiosis. Therefore, the mechanisms behind recombination will be discussed in more detail in the next section.

**Recombination.** Recombination means that genes from the male parent and the female parent are mixed in the gametes. The course of events, when genes in an individual with four chromosome pairs are mixed, is illustrated in Box 1-1. This individual, will show 16 different configurations of the maternal and paternal chromosomes in metaphase I of meiosis. In the box the number of possible recombinations in species with 12 chromosome pairs is given. Those who are interested can themselves calculate this number for species with a different number of chromosome pairs. The general formula for recombination, that is also valid for loci with multiple alleles, is given in Chapter 3, Box 3-1.

It seems as if *Tilia cordata* with its 41 chromosome pairs has tremendously more possibilities of recombination than *Picea abies* or *Pinus sylvestris*. However, in *Picea abies* and *Pinus sylvestris* many crossovers per chromosome pair can increase the number of possible recombinations and thus compensate for a lower chromosome number. Depending upon how many crossovers per chromosome pair that take place, species with a low chromosome number can attain the same level of recombinations as species with double or triple the number of chromosomes. The number of crossovers is estimated at 2-3 per chromosome pair in *Picea abies* and *Pinus sylvestris.*
Chromosome aberrations

Different types of chromosome aberrations can arise resulting in changes in the chromosome structure (Fig. 1-3). A loss of a chromosome segment is a deletion while a repetition of a segment is a duplication. A rearrangement of a segment in such a way that the segment order is turned 180 degrees is called an inversion. These three types of aberrations occur within homologous chromosomes (except for certain types of duplications). A fourth type implies an exchange of segments between nonhomologous chromosomes and is called a translocation, the reciprocal type being the most common (Fig. 1-3).

The prime cause of these aberrations is that at least two breaks have occurred in one or two chromosomes. The chromosomes can be restored by joining the broken ends. If the broken ends are joined in an incorrect way, aberrations are generated. One break in each of the homologous chromosomes is needed to generate a duplication, while two breaks in one of the homologous chromosomes are needed to induce a deletion or an inversion. A duplicated segment can also be found on a nonhomologous chromosome as well as at its original location. This means that the segment will be present in three copies, in the two original homologous chromosomes and in the nonhomologous chromosome. In addition to double-stranded breaks, duplications can also arise after errors in DNA replication or after unequal crossing-over owing to mistakes during the homologous pairing in meiosis. This results in duplications present in tandem arrangements. Duplications can also be induced by transposons. The evolutionary significance of duplications will be discussed further in Chapter 2. The exchanged segments in translocations can be of very unequal size. Furthermore, in some species all chromosomes can be involved in segmental exchanges as in the genus Oenothera.

All these aberrations can cause problems in meiosis. Individuals, heterozygous for an aberration, show a varying degree of sterility, because some of the gametes will be lethal. The larger the aberrations the greater the probability of producing lethal gametes. It can be of interest to mention that some of these aberrations are easily recognized cytologically, for instance in the meiotic divisions of the pollen mother cells.

There are very few reports of how common chromosomal aberrations are in forest trees. Minor inversions, however, have been observed in American pine species. Many plant species are polyploids, which means that they have more than two chromosome sets. An example of this is Sequoia sempervirens native to the coastal region of western North America, which has 6x=66 chromosomes. It is plausible that such species can tolerate deletions easier than their diploid relatives with only two sets of homologous chromosomes.

Development of egg cells and sperm cells

The final result of meiosis is the formation of four haploid daughter cells, a tetrad. In the male strobili each of these cells will give rise to a pollen grain. In the female strobili, on the contrary, only one of the four megaspores in a tetrad will continue to divide mitotically. This megaspore will give rise to an embryo sac and after further divisions of the nucleus, an egg cell is formed. The further development of the megaspore differs considerably between conifers and angiosperms (Fig. 1-4).

Conifers. The formation of the female gametophyte takes place within the remaining megaspore in the ovuliferous scale Figure 1-4. The megaspore grows to a large size and free nuclear divisions take place resulting in a large haploid megagametophyte or prothallium. The haploid gametophyte is sometimes incorrectly called endosperm. At the pole of the gametophyte a number of archegonia are formed that contains the large egg cell.

The formation of several archegonia and the possibility of fertilization of the egg cell in each of them result in a competition among the embryos formed. Embryos, which are less competitive may degenerate and disintegrate. This is one means to avoid formation of selfed seeds in conifers.
Self-sterility genes, which in one way or another prevent selfing, have not been found in conifers. The mechanism with fertilization in several archegonia has the same function as self-sterility genes. Especially in Finland analysis of number of archegonia and pollen production were carried out in *Picea abies* and *Pinus sylvestris*. As seen from Figure 1-5 three archegonia per megagametophyte was the most common number in *Picea abies*. Figure 1-6 demonstrates that the pollen production in the same stand varied between years and that the peak dispersal occurred at different dates. Initiation of generative organs, female and male strobili, and time for pollen dispersal and receptivity in female strobili are weather dependent.

At the time of pollination the pollen contains two prothallial cells, a generative cell and a tube cell (*Pinus* species). The pollen tube grows and the generative cell divides into a stalk cell and a spermatogenous cell. In *Picea abies* the mature pollen grain contains 5 cells: two prothallial cells, a tube cell, a spermatogenous cell, and a stalk cell. The spermatogenous cell divides under formation of two sperm nuclei.

After fertilization several free nuclei divisions takes place and a proembryo is formed at the distal pole of the former archegonium.

Angiosperms. The mononuclear embryo sac grows considerably in size and its nucleus starts to divide in 3 consecutive divisions. The result is 8 nuclei. Six of these 8 nuclei move to the poles of the embryosac and become enclosed in cell walls. The upper ones are referred to as egg apparatus, one of them tightly connected to the cell wall becomes the egg cell. The two others are coined synergids. The 3 cells at the bottom are called antipods. The remaining two so called pole nuclei move to the center of the embryo sac and unite to a diploid nucleus, called secondary embryo sac. Already in the pollen sacs of the anthers, the division in the pollen grain may start. One lens-formed nucleus is formed, which becomes the generative nucleus. The other nucleus is a vegetative nucleus. The pollination frequently takes place at this stage and when the pollen has reached the stigma the pollen tube starts to grow and another division of the generative nucleus starts and the two sperm nuclei are formed. The formed nuclei are always in the lower part of the pollen tube and the upper part is frequently degenerated as the pollen tube grows to the embryo sac. When the pollen tube reaches the egg apparatus the pollen tube releases its nuclei into the embryo sac but never directly into the egg cell but into one synergid cell. The vegetative nucleus of the pollen tube disintegrates. The sperm nuclei have spiral form and these nuclei probably have an own possibility to move. One of the nuclei enters the egg cell and

**Figure 1-4.** The development of the female organs after the formation of four megaspores in Gymnosperms, above, and Angiosperms, below. See the text for further explanation.
units with the egg nucleus under formation of a zygote. The other sperm nucleus unites with the diploid secondary embryo sac nucleus under the formation of a triploid endosperm nucleus.

Time of meiosis

A summary of the time of meiosis in a large number of conifer genera and in individual species in which a variation in this trait exists, is given in Table 1-2. As is evident from this table, there are three main types:
- start and completion in autumn
- start in the autumn and completion in spring
- start and completion in spring

Following the discovery that meiosis in pollen mother cells of different *Larix* species starts in autumn, the previous opinion that meiosis occurs either in autumn or in spring, had to be revised. Furthermore, detailed investigations of the situation in *Pseudotsuga menziesii, Thuja plicata* and *Tsuga heterophylla* showed that even these species exhibited the same type of timing of meiosis as *Larix* species. Both female and male meiosis take place in spring in the majority of species investigated. In most *Larix* species, the time of meiosis is dependent on the weather conditions. The maximum time span for meiosis in some conifers in Sweden is illustrated in Fig. 1-7.

In birch, hazel and alder, meiosis occurs in late summer. In elm, aspen and oak it takes place in spring. In those species in which meiosis starts in late summer or autumn the night length is probably the environmental factor that initiates meiosis. Thus, there is a continuous variation in time of initiation from northern Finland to southern Finland, with the earliest start in the northerly populations of *Betula pubescens*. In those species in which all or most of meiosis occurs in spring, the heat sum is the main factor that determines the timing of initiation of meiosis. Howe-
ver, there is a variation in heat-sum demand, so that the northern populations need smaller heat sum for initiating meiosis than the southern populations. But in spite of this, meiosis takes place later at the northern latitudes because of a much later spring.

Injuries and irregularities during meiosis

Certain stages of meiosis are known to be very susceptible to environmental factors during the lifetime of an individual. Especially, the effects of low temperature on meiosis have been studied in forest trees but also the effect of very high temperatures have been elucidated.

Long before climatically controlled cultivation facilities came into use in forest genetics, Enar Andersson was able to carry out very ingenious experiments for studying the effects of low temperatures on meiosis in pollen mother cells of Norway spruce. In 1948, in the end of April, he collected twigs with male strobili from trees growing at various levels along an alpine slope in Dalecarlia (Sälén), in central Sweden, and transferred them to different elevations. As the temperature decreases gradually during clear nights when approaching the bottom of the valley, it is possible to get information about how strongly the different temperatures affect meiosis. The results of such a transfer are illustrated in Fig. 1-8. The figure shows that the percentage damaged pollen mother cells was higher after a transfer to the level of 350 m than if the material was left at the 775 m level. Performing different transfers, Enar Andersson concluded that no injuries occurred at temperatures above -2°C, but at -11°C meiosis was so irregular that no pollen grains at all were produced. Large

Table 1-2. Time of meiosis in some conifer genera and species

<table>
<thead>
<tr>
<th>Meiosis starts and is completed during autumn</th>
<th>Meiosis starts during autumn and is completed during spring</th>
<th>Meiosis starts and is completed during spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cedrus</td>
<td>Meiosis of pollen mother cells in Larix, Pseudotsuga, Thuja, and Tsuga usually show this pattern</td>
<td>Abies</td>
</tr>
<tr>
<td>Cryptomeria</td>
<td></td>
<td>Athrotaxis</td>
</tr>
<tr>
<td>Juniperus chinensis</td>
<td></td>
<td>Cunninghamia</td>
</tr>
<tr>
<td>J. horizontalis</td>
<td></td>
<td>Juniperus communis</td>
</tr>
<tr>
<td>J. virginiana</td>
<td></td>
<td>J. rigida</td>
</tr>
<tr>
<td>Taxus</td>
<td></td>
<td>Keteleria</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Picea</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pinus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pseudolarix</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Megaspore mother cells in Larix and Taxus</td>
</tr>
</tbody>
</table>
temperature fluctuations between day and night occur occasionally in the Alps resulting in damage to meiosis in pollen mother cells in Norway spruce. From these and other results it was concluded that also high temperatures, above +20°C can induce disturbances of the susceptible stages of meiosis.

The most extensive investigations of temperature-induced injuries were performed in Swedish studies of the three *Larix species, Larix decidua, L. leptolepis* and *L. sibirica*. From these studies we have learnt that the most susceptible stages of meiosis are diakinesis - telophase I and prophase II - telophase II and that injuries will appear at temperatures below -2°C. Some years the pollen formation in *Larix sibirica* collapsed owing to frost injuries. Meiosis in the pollen mother cells of larch will also be discussed further in the section *Darwinian fitness* and *domestic fitness* in Chapter 7.

Figure 1-8. Transfers of male strobili of *Picea abies* to different elevations along an alpine slope, to determine at which temperature injuries of meiosis appear.

**Summary**

At the chromosome level, each species is characterized by its *karyotype*, the number, size and morphology of its *chromosomes*. The chromosomes are the carriers of the hereditary units. The genes are located linearly on the chromosomes at specific sites, *loci* (sing. *locus*) More than one alternative form of the genes - *alleles* - can exist at a locus. A gene can be *dominant* (*A*) or *recessive* (*a*). A *heterozygous* individual, *Aa*, carries both *A* and *a*, whereas a *homozygous* individual is either *AA* or *aa*. With complete dominance, individuals with the genotypes *AA* or *Aa* have the same performance or *phenotype*, which differs from individuals with the genotype *aa*. All genes on the same chromosome belong to the same *linkage*...
group. In a diploid organism, the same, homologous, chromosomes appear in pairs. In a diploid organism, there are two homologous chromosomes of each type. They contain the same loci in the same order.

When a somatic cell divides, the preceding division of the cell nucleus, mitosis, ensures that the two daughter cells receive the same number and type of chromosomes and thus the same genes as the parental cell. When a cell involved in gamete formation divides, two divisions of the cell nucleus, meiosis, result in halving of the chromosome number. Simultaneously recombination of genes between homologous as well as non-homologous chromosomes occur. Four haploid cells are formed, each with one chromosome set. During fertilization, the original chromosome number is restored.

In species such as *Picea abies* and *Pinus sylvestris* with 12 chromosomes pairs, an infinitely large number of possible recombinants can be produced.

Several types of chromosome aberrations occur such as deletion: loss of segment duplication: repetition of a segment inversions: a segment is inverted 180 degrees translocation: exchange of segments between non-homologous chromosomes

Chromosome aberrations usually cause irregularities in meiosis that are lethal to some of the gametes. The time of male meiosis in conifers during the annual cycle shows three types of pattern. It starts and is completed during autumn, e.g. *Juniperus* species. It starts in autumn and is completed in spring, e.g. *Larix sp*. Start and completion of meiosis during spring is the most common type, e.g. *Picea* and *Pinus sp*. Female meiosis takes place in spring only. In angiosperms such as birch, meiosis occurs in late summer, whereas in oak it takes place in spring.

Injuries and irregularities occurring during male meiosis are caused by sub-zero temperatures or very high temperatures.

Further reading


DNA, genes, molecular evolution, genetic engineering

It is obvious that the hereditary material must have great stability so that it can be transferred unchanged from one generation to the other in the overwhelming number of cases. On the other hand, the hereditary material must not be so stable that no changes whatsoever can take place. Heredity calls for a high degree of perpetuity but with opportunity for variation so that adaptation to new environmental conditions can take place. The molecule that fulfills these demands is the deoxyribonucleic acid - DNA. In addition to DNA, we will also deal with the structure of the genes, regulation of gene activity, the molecular clock, and genetic engineering including its applications.

DNA structure

As early as in the 1930s, it was known that DNA was a giant molecule much larger than a protein molecule. The four nucleotides of DNA were also known, each composed of one phosphate group, one sugar molecule and a purine- or a pyrimidin base. As the result of microbiological experiments in 1940s, it was shown that DNA was the molecular bearer of heredity. The great break-through, however, did not occur until the American James Watson and the Englishman Frances Crick in 1953 published their theory about the three-dimensional structure of DNA. At that time, Watson and Crick knew that the DNA molecule is composed of two long polynucleotides forming intertwined chains. The constant diameter of the DNA molecule was also known. But how these two chains were orientated relative to each other and kept together was unknown. By building three-dimensional models of DNA in such a way that the energetically most stable configurations were favoured, they soon came to the conclusion that the sugar-phosphate part forms the backbone on the outside of the DNA molecule and the purine- and pyrimidine bases are on the inside. The bases are oriented so that they can form hydrogen bonds, i.e. weak covalent bonds, between each other in the opposite chains. This is the way the two polynucleotide chains are kept together. When they built the model in such a way that a purine base always bound to a pyrimidine base, they also fulfilled the requirement that the diameter should be constant.

The DNA molecule is thus a double helix of two nucleotide chains running in opposite directions. The DNA molecule is like a helical ladder on which the two purine and pyrimidine bases are the rungs and the sugar-phosphate complex forms the backbone (Fig. 2-1). The purines consist of the two bases adenine (A) and guanine (G) with a double-ring structure including five or six atoms, respectively; two of the atoms in each ring are nitrogen, the others are carbon. The two pyrimidines, cytosine (C) and thymine (T), have a single ring with two nitrogen and four carbon atoms. To fulfil the rule of Chargaff, that the proportion of adenine in DNA equals that of thymine and the proportion of cytosine that of guanine, Watson and Crick assumed that adenine can pair only with thymine and cytosine only with guanine. This is called the base pairing of the DNA molecule which means that the two helices are complementary. If the sequence of the bases in one chain is known then the sequence in the other chain is known. The weak hydrogen bonds facilitate the split of the double helix. This in turn facilitates the replication of DNA.

DNA replication

Since the two strands of the double helix are fully complementary, they can serve as templates for generating two daughter double helices identical with the original double helix. Evidence for this model of replication was demonstrated by cultivating E. coli bacteria for several generations on a medium containing heavy carbon and nitrogen isotopes. In that way, the original double helix was labelled with these isotopes. The bacteria were then grown for one generation on a normal light medium. It was observed that the weight of all DNA molecules was intermediate between a heavy and a light double helix. The only possible interpretation of this result must be that the newly generated double helices consisted of one old, heavy strand (parent) and one new, light strand (Fig. 2-2). This is a semiconservative model of replication as opposed to a conservative model in which one of the two daughter helices is built only of two old strands while the other daughter helix consists only of new strands. The result of the experiment described above with E. coli bacteria demonstrates that the conservative model must be rejected. Furthermore, the structure of the DNA molecule proposed by Watson and Crick became quickly accepted by other researchers in the field. James Watson and Francis Crick were awarded the Nobel prize in physiology or medicine in 1962 together with the English physicist Maurice Wilkins.
Mutations - changes in DNA

Changes in the sequence of the bases in DNA can occur in many ways. During replication, a base pair can be deleted (deletion), or it can be added (addition), or it can be replaced by another base pair (substitution). Many different substances in the cell can interfere with the replication leading to such mutations and to incorrect base pairing.

Where to find DNA?

The largest amount of DNA is found in the nucleus of the cell. But DNA can also be found in the cytoplasmic organelles, the chloroplasts and the mitochondria. The least amount of DNA is located in the mitochondria. However, in leaf cells of some plant species, the amount of DNA in the cytoplasm can come up to 15%. One reason for this is that there are many chloroplasts and mitochondria in...
such a cell and that each DNA strand is present in many copies in each chloroplast and mitochondrion. There are usually 20-40 chloroplasts per cell, each with 100-150 copies of the chloroplast genome. Similarly, there are 100-3000 mitochondria per cell, each with 2-50 copies of the mitochondrial genome. This presumably reflects the need for high concentrations of photosynthetic and respiratory enzymes. As regards the number of genes in a woody plant, most of them are to be found in the nucleus, about 20 000 - 60 000. A chloroplast in higher plants consists of about 120 genes only and a mitochondrion of still fewer genes. Of importance for evolutionary studies is that chloroplast DNA in conifers shows paternal inheritance, i.e. DNA is transmitted with the pollen while it is inherited maternally in angiosperms. Mitochondria, however, exhibit maternal inheritance in both conifers and angiosperms.

Where is DNA located in the nucleus of the cell and how is DNA organized?

The next questions are where the DNA molecule is located in the nucleus and how it is organized. Already in the 1920s, the German chemist Robert Feulgen showed that DNA was located in the chromosomes. He developed a staining method, called the Feulgen method, where he used a DNA-specific purple dye. This method is still one of the most used for staining chromosomes. It is quite clear that the length of the DNA molecule is much larger than a chromosome in the metaphase stage of the mitosis. The problem to solve was how this very long DNA molecule is packed into chromosomes. A further question to be addressed was whether there are many DNA molecules or only one very long molecule. Only the last mentioned alternative is in accordance with the way the DNA replicates - the semiconservative model (see above: the semiconservative replication). In addition, all data from linkage studies point towards the fact that the genes on a chromosome/chromatid are located like strings of pearls. The conclusion drawn was that in a chromosome of higher organisms, DNA exists as only one continuous molecule. In humans, a cell contains about 1 meter of uncoiled DNA in a single, haploid chromosome set. In the largest chromosome, the length of the uncoiled DNA molecule is estimated to 8.5 cm. How can this very long DNA molecule be packed into the chromosome?

Fig. 2-2. The semiconservative replication of DNA as originally conceived by Watson and Crick. The current model is essentially the same but more complex. The identical daughter double helices consist of one old and one new strand. A, T, G and C denotes the four bases adenine, thymine, guanine and cytosine.
All the details in this very efficient packing 'system' are not known but the main features are illustrated in Fig. 2-3. Before looking into this figure in more detail, we need to get to know the nucleosome, discovered in the 1970s, because the nucleosome has been shown to be an important packing unit. In electron microscope pictures where the DNA molecule appears in its uncoiled state, the nucleosomes stand out like beads of a necklace. A nucleosome contains eight small chromosomal proteins called histones, that form an octomer. Two units of each of the histones H2A, H2B, H3 and H4 constitute together an octomer. The DNA is wound in a little less than 2 turns, 146 nucleotide pairs, around the nucleosome (see Fig. 2-3). The DNA molecule is then wound around the next nucleosome until the entire DNA molecule is wound around nucleosomes as is seen in Fig. 2-3. The region of the DNA molecule that links two nucleosomes is called linker DNA and to this region one molecule of a fifth type of histone, H1, is attached. This histone stabilizes the densely packed 30 nm structure during the further coiling into supercoils. In Fig. 2-3, additional stages in the coiling and folding of DNA is illustrated up to the final stage - a metaphase chromosome during mitosis. Conclusion: The DNA molecule exists in a densely packed condition in the chromosomes following progressive windings, coils and foldings.

What is a gene?

DNA consists of coding, genic DNA and non-coding regions, non-genic DNA. A clearly worded definition of a gene that completely covers the concept is difficult to achieve, but the definition below should be satisfactory:

A gene is a segment of DNA that is essential for a specific function.

Among some viruses, RNA (ribonucleic acid) constitutes the genetic material and in these cases, RNA replaces DNA in the definition above.

During 1940s, informative experiments were made with the fungus *Neurospora*. The two Americans George W. Beadle and Edward T. Tatum used mutants that were unable to synthesize specific amino acids. They could show that biochemical reactions in living cells occur in a series of discrete steps in which each reaction is catalyzed by a single enzyme. Furthermore, the experiments showed that there was a direct relationship between genes and enzymes: the one-gene-one-enzyme hypothesis could be approved. However, one gene often encodes more than one enzyme (protein) via so-called 'alternative splicing'. The concept of the indivisible gene and that the genes were located on the chromosomes like pearls on a necklace were firm convictions for a long time. The gene was the smallest unit of recombination. However, more recent research has shown that the gene, no more than the nucleus of the atom, is indivisible. Instead crossing-over can take place within the gene and the smallest unit consists of a base pair.

In Box 2-1, the steps from chromosomal DNA to the formation of a polypeptide is illustrated (see also the central dogma in the section The genetic code). This course of events is valid for those genes that are involved in protein synthesis, i.e. (1) genes encoding proteins. There is an additional type of genes that perform differently. This consists of (2) genes encoding functional RNA, for example ribosomal RNA (rRNA) or transfer RNA (tRNA). These genes are only transcribed, not translated.
Box 2-1. Transcription and translation

Transcription
The primary RNA transcript is formed as a complementary strand to the upper single-stranded DNA.

RNA processing
Non-coding regions, introns (shown as pink below) are spliced and removed before mRNA leaves the nucleus. The three coding exon regions are united in mRNA.

Translation
The polypeptide is synthesized as the ribosomes move along mRNA. A specific tRNA, transfer RNA, with its anticodon, specific for the amino acid it brings. The codon AUG on mRNA encodes the amino acid methionine. A = adeninine, U = uracil, G = guanine.
Eukaryotic DNA can also be classified as below:

* Unique, single and low-copy, functional genes, *i.e.* genes that exist in only one or a few copies per genome

* Repetitive DNA, which includes distinct subclasses:

  Moderately repetitive DNA with functional sequences, for example functional gene families and their closely related nonfunctional so-called *pseudogenes* which have completely or partly lost their protein-encoding function, and repetitive genes in tandem configuration encoding ribosomal RNA, transfer RNA and certain histones.

  Repetitive sequences with mainly unknown function

  Highly tandemly-repeated DNA located on both sides of the centromere, the *satellite DNA*, and located at the telomeres, the chromosome ends

  Variable number of tandem repeats, (VNTRs) known as *minisatellite DNA* used for *DNA fingerprinting*

  Simple tandem repeats (STRs) or simple sequence repeats, (SSRs) so-called *microsatellite DNA*, more often used for mapping purposes than minisatellites

  Transposed sequences, for example various types of *transposons*, so-called *‘jumping genes’*, found in *e.g.* maize, with the ability to transpose to new places within the genome without any intermediate, and retrotransposons that transpose via an RNA intermediate.

* Spacer DNA, which until now has not been classified in any of the categories above.

It is a paradox that there is no correlation between the total amount of DNA in plants and animals and their complexity. It is true that the amount of DNA increases from bacteria to human beings, but there are some insects, amphibians and plants with a much larger amount of DNA than humans. For example, *Fritillaria*, belonging to the lily genus, contains about 30 times more DNA than humans (Table 2-1). The explanation for this paradox is that in higher organisms most of the DNA does not code for amino acids in proteins, and the amount of non-coding DNA varies to a great extent among species. Moreover, in most higher organisms and especially in conifers, non-coding DNA comprises a very large part of the total DNA. More than 99% of DNA in Norway spruce and Scots pine is probably of minor or no importance for their environmental adaptation. This is in contrast to both *Eucalyptus* species and *Arabidopsis thaliana* (wall cress) which have a lower proportion of non-genic DNA (Table 2-2). One

<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome size (bp) haploid cell</th>
<th>Number of genes</th>
<th>Genic DNA %</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td>$4.7 \times 10^6$</td>
<td>4400</td>
<td>100</td>
</tr>
<tr>
<td>Yeast</td>
<td>$12 \times 10^6$</td>
<td>6000</td>
<td>50</td>
</tr>
<tr>
<td><em>Arabidopsis</em></td>
<td>$125 \times 10^6$</td>
<td>25500</td>
<td>50</td>
</tr>
<tr>
<td><em>Populus trichocarpa</em></td>
<td>$485 \times 10^6$</td>
<td>45000</td>
<td>25</td>
</tr>
<tr>
<td><em>Picea abies</em></td>
<td>$300000 \times 10^6$</td>
<td>25000 – 50000?</td>
<td>&lt;3</td>
</tr>
<tr>
<td><em>Fritillaria</em></td>
<td>$850000 \times 10^6$</td>
<td>25000?</td>
<td>0.02</td>
</tr>
<tr>
<td>Human</td>
<td>$3000 \times 10^6$</td>
<td>20000 - 35000</td>
<td>&lt;2</td>
</tr>
</tbody>
</table>

**Table 2-1. Genome size, number of genes, and per cent genic DNA in various organisms.**

<table>
<thead>
<tr>
<th>Trait</th>
<th>Pinus spp</th>
<th>Eucalyptus spp</th>
<th>Arabidopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size, pg per haploid cell</td>
<td>24</td>
<td>0.6</td>
<td>0.15</td>
</tr>
<tr>
<td>Haploid chromosome number</td>
<td>12</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Repetitive DNA, %</td>
<td>75</td>
<td>75</td>
<td>10</td>
</tr>
<tr>
<td>Single-copy DNA, %</td>
<td>25</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>Genic DNA, %</td>
<td>0.3</td>
<td>13.3</td>
<td>50</td>
</tr>
</tbody>
</table>

**Table 2-2. Comparison of genome traits between three parent species, pg = 10^{12}g.**
can ask why there exists such large amounts of seemingly redundant DNA, so-called junk DNA, i.e. DNA not encoding proteins or RNA, but may have other functions. Repetitive sequences such as transposons make up a large proportion of the genome. Introns also lack any obvious function. Transposons may be regarded as molecular parasites which do not have any specific function in their host. Frances Crick therefore called them 'selfish DNA' because they use their host for propagation only, apparently without being of any use for the host.

On the other hand, transposons have been of great significance for the origin of new genes by generating mutations during evolution:

1. They have induced rearrangements of DNA, especially gene duplications and deletions
2. They have given rise to spontaneous mutations by being inserted into a gene and thereby changing the gene product.

The transposons seem to be eliminated very slowly from the eukaryotic genome compared to their induction rate, leading to an accumulation during evolution. In many eukaryotes, the transposons therefore constitute a large part of the genome. It has been calculated that they constitute about 30% of the total human genome, while in maize they probably occupy more than 50%. The same type of transposons (retrotransposons) as in maize has also

---

Fig. 2-4. The genetic code. A, U, G and C stand for the four bases adenine, uracil (thymine is exchanged for uracil in mRNA), guanine and cytosine. The innermost cycle indicates position 1 of the codon, the intermediate cycle position 2 and the outer cycle position 3. At position 3 more than one alternative can exist, all resulting in the same amino acid. The code is said to be degenerate. For example, when the order of the bases is CAG or CAA both code for the amino acid glutamine, while AUG only codes for the amino acid methionine. Stop codons are UAA, UAG and UGA which means that the protein synthesis terminates at these codons.
been found in pine and spruce species. In *Pinus elliotii* these transposons have been shown to be distributed fairly evenly over all 12 chromosomes and they represent a significant part of the genome of this species.

Also the introns (see Box 2-1), DNA sequences located between the coding sequences of the genes expressed, belong to the so-called junk DNA. Their numbers vary largely among species, from a few as in *Arabidopsis* to a large number in humans and probably also in conifers.

This fragmentation of genes makes them more flexible. In humans, about 60% of the genes are subjected to alternative splicing, that is different protein molecules that are generated from the primary DNA transcript by changing the number and order of exons in the final mRNA after splicing out the introns. Alternative splicing seems to be less common in plants.

### Conservation of non-genic DNA

Why does the conifer genome contain so much more DNA, mainly non-coding, than the annual plant *Arabidopsis*? The explanations are probably as follows:

1. Extra DNA is a disadvantage if energy and material is needed for the formation of this DNA during each cell division. This disadvantage, however, seems to be negligible.

2. Of more importance is that strong correlations exist between a large genome and large cell nuclei, and between large cells and slow mitotic and meiotic cell divisions.

3. For most tree species, it does not matter whether the genome is large or not, because rapid cell divisions are not critical for survival. Some other factor limits the growth or a very rapid growth is not of ecological importance. Trees and herbs that grow at very northern latitudes, do have a fairly low DNA content, probably because they need to pass several developmental stages and go through the mitosis and meiosis during a short growing season.

4. *Arabidopsis* on the contrary has a very short life cycle of 2-3 weeks from seed to seed. Only very small genomes allow such rapid cell divisions as are needed for such a short cell cycle.

5. Furthermore, a positive selection for large genomes probably occurs in most conifers. With a few exceptions, these species do not have vascular vessels for water transport. Instead they have very long water-conducting cells, so-called tracheids. These tracheids need a large genome as there seems to be a strong correlation between the amount of nuclear DNA and the size of those cells in the cambium that generate tracheids. Thus, in a study of 18 North American pines species, it was found that those species adapted to dry areas had a larger genome than those species growing in more moist areas.

In most European plants the amount of non-genic DNA has increased during evolution, for instance, because the ‘nuclear’ parasites can easily reproduce without a decrease in the competitive capacity of the plants. For some species and in specific environments, natural selection has affected the DNA content in such a way that it is either conserved at a low level as in *Arabidopsis* or at an extremely high level as in most conifers.

### The genetic code

In the early 1950s, it was clear that a linear correlation exists between changes in the base pairs of DNA and changes in a protein. This means that when a change in DNA occurs, this is equivalent to a change at the corresponding site in the protein. This indicates that there is a strong correlation between a gene and a protein. However, the proteins contain 20 different amino acids, but DNA contains only four different bases. Therefore, it was assumed that groups of bases must constitute the code for the order of the individual amino acids in the protein. At that time, it was known that the major part of DNA is located in the chromosomes in the cell nucleus while the synthesis of a protein takes place outside the nucleus. It was also known that there is a large amount of RNA in those cells exhibiting a large protein synthesis. RNA is different from DNA, as the name indicates, as regards the sugar molecule - ribose - to which the bases are linked, and furthermore that uracil has replaced thymine. Based on these results, it was assumed that RNA could act as an intermediary of information from DNA in the nucleus to the site of the protein synthesis. The relationship between DNA, RNA and a protein is known as Crick’s central dogma of molecular genetics:

DNA → transcription → RNA → translation → protein

Crick emphasized that the flow of genetic information is from nucleic acid to protein. A reverse flow from protein to nucleic acid, which would enable the ‘inheritance of acquired characters’ envisaged by Lamarck in the 19th century, is apparently impossible. Some viruses possess an enzyme, reverse transcriptase, that synthesizes a single-stranded DNA molecule using RNA as a template; but this does not contradict the central dogma as formulated by Crick.

Later on it was shown that only one of the DNA strands serves as a template during the RNA synthesis. Thus, RNA consists of the complementary pattern to the strand of DNA that served as a template during RNA synthesis (see Box 2-1). The formation of the RNA strand is called transcription. In eukaryotes, the transcribed DNA contains both coding regions, exons, and non-coding regions, introns. But it is only the coding regions, the exons, that will transfer their information into a protein.
In early 1960, a now classical experiment was carried out. RNA molecule when located on the ribosomes. In the cytoplasm, the mRNA moves to the ribosomes on which the protein synthesis takes place using mRNA as a template, translation. When the ribosomes move along the mRNA molecule the amino acids are linked together forming a polypeptide in the order determined by mRNA. (For proteins consisting of more than one polypeptide, the individually synthesized polypeptides are later combined to form the complete protein.) Also another type of small RNA molecule, each binding an amino acid, was detected in the cell. These RNA molecules were named transfer RNA (tRNA), because they bring the amino acids to the RNA molecule when located on the ribosomes.

In early 1960, a now classical experiment was carried out with mRNA artificially produced and consisting of uracil only. This mRNA construct produced a polypeptide during the polypeptide synthesis that exclusively contained the amino acid phenylalanine, although all other amino acids were available for the polypeptide synthesis. Shortly after this discovery, it was shown that mRNA containing only adenine encoded a polypeptide consisting only of lysine. In the middle of 1960s, by synthesizing mRNA using different bases it could be stated that a sequence of three bases of the mRNA molecule codes for one specific amino acid. This triplet of three bases is called a codon. Thus, the genetic code consists of a series of mRNA codons, each specifying a particular amino acid.

Since the four bases can be combined in 64 different combinations (4^3), it is evident that more than one codon encodes a particular amino acid, i.e., the code is called degenerate. Out of the 64 possible combinations, 61 encode specific amino acids. The three combinations not encoding any amino acid, UAA, UAG and UGA, code for stop signals that terminate the synthesis of a protein. The codon AUG, which also codes for methionine, is the initiation codon for translation.

In most cases, as is illustrated in Fig. 2-4, it is the two first bases of the triplet that specify a particular amino acid while the third base has no influence on which amino acid will be formed. For example, when the base cytosine is found in the positions 1 and 2 in the triplet, the amino acid proline is always formed irrespective of which base is located in position 3. When a base is replaced by another base without changing the amino acid encoded, the change is referred to as a synonymous substitution.

Each transcribable gene has a region upstream of the start of transcription that regulates the synthesis of mRNA. This region is called the promoter (Fig. 2-5). The synthesis of RNA is carried out with help of an enzyme, called RNA polymerase as it contributes to the formation of a RNA polymer. RNA polymerase binds to the site of transcriptional initiation of the promoter. Eukaryotes have three types of RNA polymerase, each with a different function. The first type mainly transcribes genes for ribosomal RNA, while the second type transcribes protein-coding genes leading to the synthesis of mRNA, and the third type transcribes tRNA genes and some other small nuclear RNA types. As discussed above, stop signals are needed to terminate the synthesis of a protein. Three codons have this function (see Fig. 2-5).

**Fig. 2-5. The promoter region in higher eukaryotes. The TATA box and the two upstream elements the CCAAT box and GC-rich box are shown. Elements acting on the promoter at great distances, enhancers and silencers, are not shown.**

**Regulation of gene activity**

Already before the genetic code was deciphered, it was clear that there must be ways of controlling the number and type of proteins formed in the cells. For example in a forest tree, all genes are not active - turned on - at the same time during the whole day and night, or during the annual cycle or during ontogenetic ageing from the juvenile to the adult, flowering-competent phase. A plant of Norway spruce would not be able to survive if genes for active height growth were active even in winter time. Referring to Box 2-1 and the central dogma of molecular genetics, it is easy to see that this control can be exerted at two levels at least: (1) at the time when DNA is transcribed to mRNA, transcriptional control, and (2) when mRNA is translated to a protein, translational control.

In eukaryotes there are many genes that are always constitutively expressed. These so-called house-keeping genes code for essential functions common to all or most cells. Other genes must be turned on or off at appropriate times. The major control of gene expression takes place at the transcriptional level. The essential factor for this
control is the promoter. The promoters are classified in three main groups: (1) constitutive (always active more or less), (2) developmentally regulated or tissue specific (active during specific developmental phases or in specific tissues), (3) inducible (activated by various physical and chemical factors). The promoter, often named the core promoter, is the region between the site of transcriptional initiation and the TATA box, a AT-rich sequence. The main controlling elements in most genes are the TATA box and promoter-proximal elements such as the CAAT box and the GC-rich box located upstream of the promoter (see Fig. 2-5). In addition there are elements that can act on the promoter at great distances. These latter elements can either increase the rate of transcription, so-called enhancers, or decrease the rate of transcription, so-called silencers. The enhancers act in such a way that the genes are only transcribed when the proper transcriptional activators are present and bind to the enhancers. The activators, also referred to as transcription factors, are needed for RNA polymerase to initiate transcription to mRNA. In a corresponding way, repressors bind to the silencers to slow transcription.

One example of translational control is provided by so-called masked mRNA. Unfertilized sea urchin eggs contain large amounts of mRNA which are translationally inactive until a few minutes after fertilization when the translations starts. Another example is that translational control can be regulated through factors that increase the lifetime of mRNA molecules during which the mRNA molecules are translated repeatedly. This means that fewer copies of a gene are required to produce a given amount of a particular translational product.

Number of functional genes in plants

Is it possible to estimate the number of genes in the genome of a conifer? What information is available about the number of genes in the genome of a conifer? From comprehensive DNA sequencing, the number of genes is known approximately for Arabidopsis, rice, and very recently, Populus trichocarpa, as 25 000, 50 000, and 45 000, respectively. Populus trichocarpa is thought to be a newly formed tetraploid. It is likely that conifers, also perhaps ancient polyploids, also have around 50 000 functional genes. Examples of genome size and number of genes in various organisms are given in Table 2-1.

Similar gene and gene order over wide taxonomic families

If one compares the genome size of the two cereal species rice and barley, rice has a very small genome (0.45 picogram in a haploid cell) while the barley genome is about 12 times larger (5.5 picogram). Both are diploid species. If we assume that rice and barley have about the same number of genes, we should expect that barley contains much more repetitive DNA. In both these species and in other cereal species, a large number of genes have been located on genetic maps and the order of the genes has been compared. These studies showed that not only did the same genes exist in both species, but more interestingly, the genes appeared in the same order, i.e. the gene order seems to be highly conserved. The picture of a plant chromosome that gradually appears, shows that repetitive DNA is found around the centromere, at the chromosome ends, and in regions of various size between the genes (Fig. 2-6). Also the genes themselves appear in clusters often located near the centromere region or near the chromosome ends. What is remarkable is that the genes occupy an insignificant part of the chromosomes as compared to the regions of various types of repetitive DNA. Also in so widely unrelated species as human and mouse partial conservation of gene order, called synteny, exists. What about conifers? This was intensively studied, notably in pine species. Synteny was found between P. taeda and P. pinaster as well as between P. taeda and P. sylvestris. This information can be used for mapping purposes. When the gene order in one species is known, this knowledge will facilitate the location of genes in a second species.
The molecular clock hypothesis is based on the following assumptions:

1. The substitution of nucleotides in DNA, or amino acids in proteins, occurs at a constant rate.

2. The degree of difference in DNA (or amino acid) sequence between similar genes (or the corresponding proteins), in two species, gives information about the time when the two species diverged from their common ancestor.

3. Most spontaneous mutations that persist in the population are neutral, which means that they are insignificant for natural selection; other mutations are often deleterious and will therefore be selected against if natural selection is allowed to act.

4. A plant can tolerate changes in some proteins but not in others.

With these assumptions we expect that each gene or DNA sequence changes at its own characteristic rate, and that this rate changes little over millions of years, i.e. each gene has its own molecular clock that ticks at a near constant rate. Knowledge of changes in DNA sequences is therefore used to calculate relationships between species, genera and higher taxonomic orders. This would correspond to the dating of geological times by measuring the decay of radioactive elements.

At the same time we have to pay attention to the fact that the molecular clocks are different in different species and vary even within a single plant. The difference between species may be due to differences in generation time; species with a long generation time probably accumulate spontaneous mutations due to DNA replication errors at a slower rate than species with a short generation time. This may be one reason why humans have a slower clock than rodents.

The variation in molecular clocks within a single individual might be due to the fact that nucleotide changes occur at a slower rate in such codons where a change in the third position causes an amino acid change and synthesis of a different protein. The molecular clock also runs much slower for those genes encoding e.g. histones. We have already seen that histones are significant proteins of the nucleosomes in the chromosomes and interact with DNA. Each amino acid along the histone protein is needed at its correct site for the correct formation of a nucleosome. When the histones in humans and in rodents are compared, the histones appear to be identical although the two species have separated about 80 million years ago (Table 2-3). Does this mean that the corresponding DNA also is identical? The answer is definitely no. As we discussed earlier, the genetic code is degenerate, i.e. different codons (triplets) can code for the same amino acid. As is seen from Table 2-3, in the histone 3 gene, synonymous substitutions have occurred at a fast rate while no nonsynonymous ones have occurred. Probably, nonsynonymous substitutions induce amino acid exchanges in the histones which are selected against by natural selection, because they are lethal or nearly so. Thus, the histone proteins have been strongly conserved during evolution. It is a general phenomenon that synonymous substitutions occur at a much faster rate than nonsynonymous ones.

At the time when the hypothesis of the molecular clock was introduced in 1965, there was great interest among researchers to use this tool for evolutionary studies. But it also caused much controversy. Notably, the original assumption that changes in DNA and proteins occur at a constant rate has been questioned. Today (in 2006) the constant molecular clock is accepted by most geneticists as an approximation. The molecular clocks have become very valuable tools for calculating the dates of speciation events and for constructing so-called phylogenetic trees. There is often a good correspondence with the expectations from the conventional tree constructions.

### Table 2-3. Variation in rates of synonymous and nonsynonymous substitution in various mammalian protein-coding genes. Comparisons between human and rodent genes. The time from divergence: 80 million years.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Substitution rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-synonymous (x10⁹)</td>
</tr>
<tr>
<td>Histones:</td>
<td></td>
</tr>
<tr>
<td>Histone 3</td>
<td>0.00</td>
</tr>
<tr>
<td>Histone 4</td>
<td>0.00</td>
</tr>
<tr>
<td>Contractile system proteins:</td>
<td></td>
</tr>
<tr>
<td>Actin α</td>
<td>0.01</td>
</tr>
<tr>
<td>Actin β</td>
<td>0.03</td>
</tr>
<tr>
<td>Insulin</td>
<td>0.13</td>
</tr>
<tr>
<td>Growth hormone</td>
<td>1.23</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>0.55</td>
</tr>
<tr>
<td>Interferon</td>
<td>2.79</td>
</tr>
</tbody>
</table>

Rate = substitutions per site per 10⁹ years
Chloroplasts and mitochondria have their own genetic systems resembling those of bacteria

Although only a few proteins are coded by organelle DNA, chloroplasts and mitochondria perform their own replication of DNA, transcription and translation (protein synthesis). Remarkably enough, these processes are more similar to those occurring in bacteria than to those occurring in the cytoplasm of eukaryotic organisms, including plants.

For example:
(1) Chloroplast ribosomes resemble ribosomes in *Escherichia coli* both structurally and functionally. For instance, protein synthesis is hampered by the same antibiotic substance. Nucleotide sequences in rRNA in chloroplasts and in *E. coli* are very similar. Chloroplast ribosomes can use bacterial tRNA for synthesis of proteins. In these characteristics, the chloroplast ribosomes differ from other ribosomes in the same plant cell.

(2) The protein synthesis in the chloroplasts starts with N-formylmethionine, as in bacteria, instead of starting with methionine as in the other ribosomes of the plant cell.

(3) In contrast to nuclear DNA, chloroplast DNA can be transcribed by RNA polymerase from *E. coli*, and the mRNA thus formed, can be translated by the bacterial machinery to proteins.

The genetic machinery of mitochondria also resembles that of bacteria, but to a lesser extent.

The endosymbiotic hypothesis explains the origin of organelles

How can it be that in some molecular details, chloroplasts and mitochondria are more similar to bacteria than to corresponding characteristics in other organelles of the same cell?

According to the *endosymbiotic hypothesis*, the eukaryotic cells started as anaerobic, free-living bacteria without any chloroplasts or mitochondria. After that, they established a symbiotic relationship with bacteria whose oxidative metabolism they took over and modified for their own use. The result was the progenitor of the mitochondrion enclosed in an early eukaryotic cell. This event occurred about $1.5 \times 10^9$ years ago when oxygen had entered the atmosphere, but before animals and plants were separated evolutionarily. The chloroplasts are assumed to have evolved from a bacterium with photosynthetic capacity. The chloroplasts are strikingly like the modern cyanobacteria, earlier named blue-green algae.

Interplay between the cell nucleus and the organelles

The endosymbiotic theory does not explain satisfactorily why the genomes of chloroplasts and mitochondria are so small, *i.e.* why there are so few coding genes in these organelles. In eukaryotes of today, nuclear DNA encodes about 90% of the proteins existing in chloroplasts and mitochondria. This means that we have to assume that a large number of genes has been transferred from the original endosymbiotic bacteria to the nucleus. Evidence for such a gene transfer is that nuclear genes coding for mitochondrial proteins are more similar to bacterial genes than to genes coding for corresponding proteins in other organelles. Furthermore, a large number of DNA segments from plant mitochondria can be traced in nuclear DNA, but they contain genes that in most cases have lost their functions. DNA transfer from organelle to nucleus, and even between organelles still occurs. After partial sequencing of chloroplast genomes, it was observed that the chloroplast genome of some plants contained DNA segments that were copies of segments of the mitochondrial genome. We may ask why this gene transfer has not been completed a long time ago. It must be "expensive" for the cell to preserve a separate device for the protein synthesis in the organelles, and perhaps unneeded, since there exist other organelles that have no DNA of their own. One possibility is that the transfer process was slowed down because the genetic code of proteins was changed in the nucleus while earlier versions of the code were conserved in the chloroplasts and mitochondria. In organisms living today, three or four codons have different meanings for the protein synthesis in the nucleus compared to the organelle protein synthesis.

Genetic linkage maps

Genetic linkage maps are employed for studying genes for quantitative traits (QTL, see Chapter 5) and for comparing genome structure and gene order of different species. If the gene order is conserved among conifer species or hardwoods, only one or a few species have to be investigated more closely. These maps can then be used for predicting localisation of genes in other species. This knowledge will also be significant for our understanding of the evolution of the conifer and hardwood genomes.

Genetic linkage maps with molecular markers evenly distributed along the chromosomes can significantly contribute to making the analysis of genetic diversity more efficient and to the characterization of gene resources of importance for gene conservation.

Genetic maps may also form the basis of gene cloning and the generation of transgenic trees (see below).

Genetic engineering

Genetic engineering or recombinant DNA technology implies that individual, interesting genes are transferred from one organism to another, often from one species to another species. Genetic engineering relies on molecular
Figure 2-7. Examples of restriction enzymes are given. Their recognition sites, where they will cut DNA and give rise to sticky or blunt ends, are indicated. A, T, C, and G stands for the four nucleic acid bases. Pu stands for any of the two purin bases and Py for any of the two pyrimidine bases.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Origin</th>
<th>Recognition site</th>
<th>Sticky or blunt ends</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI</td>
<td>E. coli</td>
<td>$5'$ GAAAAATTC</td>
<td>$5'$ GCTTAAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3'$ CTTAAC</td>
<td>$3'$ CTATT</td>
</tr>
<tr>
<td>HindII</td>
<td>Hemophilus influenzae</td>
<td>$5'$ GTPyPuAC</td>
<td>$5'$ GTPyPuPyTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3'$ CAPyPyPyTG</td>
<td>$3'$ CAPuPyPyTG</td>
</tr>
<tr>
<td>HindIII</td>
<td>H. influenzae</td>
<td>$5'$ AAGCTT</td>
<td>$5'$ TTCCGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3'$ TCCGAA</td>
<td>$3'$ TTCCGA</td>
</tr>
<tr>
<td>HpaII</td>
<td>H. para-influenzae</td>
<td>$5'$ CCGG</td>
<td>$5'$ CCGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3'$ GCCG</td>
<td>$3'$ GCCG</td>
</tr>
</tbody>
</table>

genetic methods developed since 1970s:

1. Methods of creating recombinant DNA molecules that include a sequence of DNA in which two non-homologous DNA segments have been combined often from quite different species.

2. Methods of determining the order of the nucleotide base pairs in a DNA segment, so-called DNA sequencing.

3. Methods of producing a large amount of specific DNA sequences by the Polymerase Chain Reactions, PCR.

4. Methods of producing synthetic genes or part of genes.

5. Methods of revealing gene function

Briefly, what do these methods mean?

1. Creation of a recombinant DNA molecule is the first step in producing many copies of a gene or part of a gene. This method is called gene cloning. First a DNA fragment including the gene to be cloned is "cut out" and then inserted into a so-called vector by ligation to produce the recombinant DNA molecule.

The vector is usually a plasmid of bacterial origin and thus often originates from a completely different species than the DNA fragment. A large number of copies of the DNA fragment can be produced in a bacterial cell through the use of such a vector. The plasmids are small, circular DNA molecules which exist in bacterial cells together with their ordinary chromosome. The plasmids contain a few genes only, among them one or more genes for antibiotic resistance. The plasmid can also be equipped with histochemical markers with specific staining properties or nutritional markers which enable cells containing this plasmid to survive on a medium lacking an essential nutrient. The plasmids replicate independently of the chromosome of the host cell and it is this property that is used for multiplication of an introduced DNA fragment. When the host cell, e.g. a bacterial cell, is cultivated on a suitable medium it starts to divide and copies of the recombinant DNA molecule are transferred to its daughter cells and a further amplification of the DNA fragment takes place. As the daughter cells originate from one and the same original cell, the colony produced via the cell divisions constitutes a clone (gene cloning).

Using the gene cloning technique, a DNA library can be established including a collection of clones with a nearly complete set of fragments that contains most of the genomic DNA or cDNA of e.g. a plant. cDNA is made from mRNA and thus reflects the expressed genes. Hence, the library can be either a genomic library or a cDNA library depending on the purpose of your investigation.

The discovery of a very specific type of enzyme, the so-called restriction enzymes, makes it possible to "cut" the DNA molecule at defined nucleotide sequences. These enzymes act as scissors that cut only at defined nucleotide sequences, recognition sequences, unique for each restriction enzyme. Furthermore, the enzymes produce two main types of DNA ends, blunt or sticky ends. In Fig. 2-7, examples of restriction enzymes and their recognition sequences are shown. The recognition sequences consist of between 4 and 8 base pairs. Fig. 2-7 also illustrates that blunt ends appear when the restriction enzyme HindIII is used, because it makes a simple, straight cut, while the other enzymes make a zigzag-like cut that produces sticky ends.
Knowledge of restriction enzymes is essential for the creation of recombinant DNA. In Fig. 2-8, the procedure for making a recombinant DNA molecule is illustrated. In this example, a plasmid isolated from the intestinal bacterium *Eschericia coli* (*E. coli*) is used as vector. This plasmid has one recognition sequence where the restriction enzyme *Eco*RI can cut. When the plasmid is treated with this enzyme, the result is a linear DNA molecule with single-stranded sticky ends. When the same treatment with *Eco*RI is applied to DNA from a plant or an animal cell, we get several DNA fragments all with the same type of single-stranded sticky ends as the vector. In the next step, the DNA fragments and the linear vector molecules are ligated together to form recombinant DNA molecules. The formation of a recombinant DNA molecule is due to the fact that the DNA fragment and the vector contain the same complementary DNA sequences. To ligate the DNA fragment and the linear vector molecule, another type of enzyme, DNA ligase, is used. These enzymes are naturally occurring and have the capacity to repair single-stranded breaks in a DNA molecule. The vector including the DNA fragments is then multiplied using the gene cloning technique discussed above. When a gene in the fragment is cloned, new possibilities open up for studying the gene structure, its function and how it is expressed. For example, the order of the nucleotides can be determined, so-called DNA sequencing (see under 2). A further application is that the gene can be introduced into host cells belonging even to different species, to pro-

**Fig. 2-8.** Creation of a recombinant DNA molecule and cloning (amplification) of genes. In addition to the important gene, the bacterial plasmid (vector) is carrying a gene for antibiotic resistance which is used for selection so that only those bacterial cells can proliferate that contain the recombinant DNA molecule.
duce a transgenic organism containing a foreign gene. An example of an industrial application of the recombinant DNA technique, is the former pharmaceutical company Kabi which initiated commercial production of a growth hormone that inhibits dwarf growth in humans, using genetic engineering. Earlier Kabi produced this hormone from deceased persons. Because of the risk of contamination with slow-acting infectious organisms (which could not be completely eliminated) such as prions that cause the Creutzfeld - Jacob’s disease (mad cow disease), it was important to be able to introduce an alternative production method of the hormone.

(2) In the mid-1970s, two methods of DNA sequencing were published. One method was based on chemical degradation of the DNA molecules, while the other method used enzymes. This latter method is now most commonly used for genome sequencing, mainly because it has been easier to adapt to automated DNA-sequencing machines. Such machines are now available in every molecular biology laboratory. The work to make the DNA sequencing more efficient is speeded up by the extensive project of mapping the human genome, the publicly funded HGP/HUGO project. When this project started, the goal was that the human genome should be completely sequenced in 2005. However, a preliminary draft sequence of the human genome was announced already in June 2000 by the International Human Genome Sequencing consortium. The eukaryotic genome first to be entirely sequenced was the genome of budding yeast Saccharomyces cerevisiae (in 1996). The plant genomes that have been sequenced are those of Arabidopsis, rice, and, in 2005, the first tree species, Populus trichocarpa. This species was completely mapped in 2006. The number of genes was estimated at 45 000, a surprisingly high number. There is a steady increase of partially or completely sequenced genomes in prokaryote and eukaryote genomes. The information is stored in international or private databases.

(3) By means of the polymerase chain reaction (PCR), a gene or a DNA sequence can be amplified in a more efficient way than by gene cloning. The method was very quickly accepted after the first reports in mid 1980s. The arrival of this method is a great breakthrough in molecular biology. Automated PCR machines can now be found in every well-equipped molecular laboratory. A DNA sequence can easily be amplified 1 million times. This technique provides a number of applications. As only a very small amount of DNA is required initially, this technique can assist very efficiently in for example, criminal cases in which the potential perpetrator only has left minor traces. Examples of application in human medicine, are within cancer therapy to reveal cancer cells among a large population of normal cells; fetus diagnosis to determine the sex or for diagnosis of heritable diseases. PCR has also opened up possibilities for studies of evolution at the molecular level. Fossil DNA, often occurring in minute amounts, can be amplified and becomes amenable for analysis. Furthermore, for gene mapping and studies of the structure and function of genes, PCR technique is an indispensable tool.

(4) In the search for more efficient methods of gene sequencing, techniques were developed for producing synthetic DNA, so-called oligonucleotides. Nowadays, also this technique is automated, thanks to the development of programmable machines. The oligonucleotides are mainly used as probes, or for production of synthetic genes. A probe with a known DNA sequence can be employed for investigating whether a gene of interest is expressed in your material or whether an unknown DNA sequence contains the same sequences as the probe. If so the unknown DNA sequence or gene can usually be identified independently of its origin. Examples of genes that were among the first to be synthesised are the interferons.

(5) The next great challenge will be to reveal the function of the located and sequenced genes. An unexpected result from completely sequenced organisms was that the number of genes is much higher than conventional genetic analysis had indicated. For instance, in yeast, S. cerevisiae, it was found that only 30 % of genes had been previously identified. There are three major methods most often used for determining the function of unknown genes. (1) Database searching with the purpose of finding homologous genes in other organisms for which the function is known. The premise for this approach is that there is a good correspondence between genes with similar DNA sequences and their functions even in distantly related organisms. (2) Inactivation of a gene and then search for loss of function. (3) In transgenic plants (or plant tissues) the gene can be overexpressed and the effect on the phenotype assessed. (4) One promising device for the functional analysis of genes consists of tiny droplets each containing a cloned and sequenced gene and placed in a microarray, on a microscope slide. This can be used to monitor the expression of thousands of genes simultaneously in a particular tissue, by hybridization with mRNA extracted from the tissue (more precisely, with a fluorescently labelled reverse-transcribed copy of the mRNA). A UV microscope is used to detect the fluorescing spots in the microarray representing genes in the microarray that have hybridized to the labelled cDNA and were therefore active in the tissue. This is a very powerful technique that among other things can potentially answer many of the questions raised by traditional quantitative genetics, by precisely defining how the individual members of a population vary genetically.

Here we should mention ‘real-time PCR.’ A real-time PCR machine follows the amplification of the DNA sequence in real time, using fluorescent markers. The method, ‘real-time reverse transcriptase PCR’, allows accurate quantification of the activity (mRNA abundance) of a specific gene relative to that of a reference gene, or group of reference genes. It is sensitive enough to need only a
small sample of tissue, such as a few needles. This enables proper replication and statistical analysis, so that the measurements of gene activity bear scrutiny by traditional geneticists. It is probably feasible to detect significant differences in gene activity as small as 20% between samples. Knowledge obtained from measurements of this kind can refine the results from microarray analysis and should be relatively easy to integrate into traditional forestry.

How can genetic engineering be applied to forest trees?

One major advantage of genetic transformation via genetic engineering is that only one important gene or a group of genes will be transferred at a time into a variety improved by conventional breeding methods and the remaining set of genes will be preserved more or less unaffected. For our forest trees and particularly those with a long breeding cycle, it is impossible to transfer one single gene by crossing experiments because the first cross must be followed by repeated backcrosses (See Fig. 9-3 for explanation) to recover most of the original set of genes. Therefore, in the future, genetic engineering can play an important role when integrated into conventional breeding programmes. In the near future, practical applications of transgenic plants will probably first appear in hardwood species, such as *Populus* and *Salix*, with rapid growth and short rotation cycles.

In basic research, genetic transformation has been intensively studied in those species where application of this technique is possible. In particular, it is a powerful tool for studying gene function and regulation of gene activity in forest trees. For example, it makes it possible to overexpress a gene if a strong promoter is added or to downregulate the expression of a naturally occurring gene. This technique has been commercially exploited, for instance in tomato; the expression of the gene coding for ethylene production has been reduced resulting in delayed maturation and improved storage capacity.

Genetic transformation means the transfer of recombinant gene constructs into plant cells, selection of transgenic cells and regeneration of these cells into transgenic plants. For achieving this, research is required within three major areas: (1) isolation and identification of genes of importance for tree breeding, whether of major or minor effects, and finding promoters that enable a suitable degree of gene expression in the appropriate cell type; (2) development of reliable means of gene transfer; (3) development of an efficient regeneration system for production and propagation of transgenic plants.

(1) Methods of isolation and identification of genes have been touched upon in previous sections. The promoters are derived from genes that are highly expressed such as the widely used 35 S promoter originating from cauliflower mosaic virus. The most common selectable marker gene confers antibiotic or herbicide resistance to the cells. When grown on selective medium containing an antibiotic or a herbicide only the transformed cell will survive. For example, in *Picea abies*, a reporter gene (GUS, to demonstrate transgene expression) and a gene conferring herbicide resistance (Basta) were fused to a ubiquitin promoter from maize. This promoter construct has successfully been involved in the production of hundreds of transgenic plantlets of *Picea abies* (2006).
(2) To date, the two main methods used for gene transfer are:

(a) Biological vectors, most often gained from bacteria that belong to the genus *Agrobacterium*

(b) Direct gene transfer using biolistic methods

The first method can exclusively be used in plants as animals cannot be infected by *Agrobacterium*. The second method has proved to be successful both in plants and in animals.

(a) Biological vectors. The most widely used vectors in plants are those isolated from the two bacterial species *Agrobacterium tumefaciens* and *A. rhizogenes*. Both are common soil bacteria. *Agrobacterium* infects injured plants and causes gall formation on the stem (*A. tumefaciens*, crown gall disease) or hairy roots (*A. rhizogenes*). When a bacterial cell infects a plant cell, a plasmid, the Ti plasmid, is transferred into the plant cell (Box 2-1). A part of the Ti plasmid, the T-DNA region, is integrated into the chromosomes of the plant cell. The plant cell starts to produce hormones in excess which induces uncontrolled cell divisions and a gall (tumor) develops (*A. tumefaciens*). *A. tumefaciens* mainly infects dicotyledons including woody plants like *Populus* and *Salix* while monocotyledons are usually resistant in field conditions. Both *Agrobacterium* species can accomplish this type of gene transfer also in conifers. In natural conditions, conifers are seldom infected by *Agrobacterium*, but infection can be induced, for instance, if the bacterial cells are inoculated in the stem. The genes that regulate the synthesis of hormones can be excluded from the bacterial plasmid and replaced by genes e.g. important for tree breeding. When this disarmed Ti plasmid is allowed to infect conifer cells growing in artificial culture, they no longer produce an excess of hormones, and can therefore be stimulated to proliferate, mature and finally grow on to a new plant. A premise for this vector method and also for the direct transfer method is access to an efficient regeneration system in vitro for production and propagation of transgenic plants.

*Agrobacterium* is routinely used for the production of transgenic plants in several herbaceous species including monocots nowadays. This is also the case for *Populus* and *Betula*. Among conifers, European larch (*Larix decidua*) was the first species in which transgenic plants were produced using *A. rhizogenes*. In *Picea abies*, as a contrast, to date this bacterium can only develop transgenic roots but no transgenic plants. However, *A. tumefaciens* can transform embryogenic suspension cultures of *Picea abies* at high efficiency resulting in many transgenic sublines.
(b) Direct gene transfer using biolistic methods. Successful transformation methods of this type are: electro- and chemical poration and microprojectile bombardement. The latter method makes use of a particle accelerator in which the acceleration of high-velocity microprojectiles for entering into plant cells (Fig. 2-9). The microprojectiles, usually gold or tungsten particles, have been coated with e.g. plasmid DNA. Successful production of transgenic plants has been achieved in *Picea glauca*. Also in other *Picea* species, including *Picea abies*, hundreds of transgenic cell lines and plants have been produced.

(3) An efficient regeneration system for proliferation of transgenic cells up to propagation of transgenic plants is a bottleneck in many forest tree species, in particular conifers. However, such generation systems are available in a number of woody species, including hybrid aspen, poplars, *Picea glauca*, *Picea abies*, *P. mariana*, *Pinus radiata*, *Pinus elliottii* and *Pseudotsuga menziesii*.

**Which traits are most amenable to genetic engineering?**

The economic benefits of transgenic tree crops can be great both for society and for forest industry. But also the environmental benefits can be substantial, reducing the use of herbicides and pesticides through the introduction of transgenic plants with herbicide tolerance or resistance to insects and pathogens. Furthermore, an increased wood fiber production via genetic engineering will reduce the need to harvest native forests. Of crucial importance is
of course the public and legal acceptance of transgenic plants.

There seems to be a general consensus about the major categories of traits amenable to genetic engineering including:

- Herbicide tolerance
- Resistance to insects and pathogens
- Reproductive capacity
- Lignin modification

In the near future also additional traits such as those affecting wood formation and fiber quality will be included.

Historically, herbicide tolerance (glyphosate) was the first trait introduced in *Populus* via genetic transformation. Herbicide tolerance has also been introduced in transgenic crop plants such as maize and soybean.

Plants have evolved efficient strategies for resistance to various insects and pathogens. For instance, plants can activate a biochemical defence when exposed to stressful conditions. Therefore, one breeding goal can be to increase this defence reaction (defence enzymes such as protease inhibitors) via genetic engineering. Another option is the transfer of genes coding for insect toxins obtained from the bacterium *Bacillus thuringiensis* (Bt). This bacterium contains a large number of genes coding for delta-endotoxins that punch holes in the guts of insect larvae. Transgenic *Populus* carrying a *Bt* toxic gene controlled by the constitutive 35S promoter showed endotoxin activity against insect larvae. As regards crop plants, large areas of maize, cotton and potatoes carrying *Bt* toxic genes are currently under cultivation.

Reproductive capacity can be changed in two ways: (1) accelerated flowering and (2) induction of male or female sterility. A desirable breeding goal would be to reduce the extended juvenile phase and breeding cycles in forest trees and in this way increase the genetic gain per time unit. Potential genes for this approach are floral meristem identity genes isolated from *Arabidopsis*. For example, early flowering was induced in transgenic hybrid aspen (*Populus tremula* x *P. alba*), by the transfer of one of these floral meristem identity genes. *Populus* also, as well as *Pinus radiata*, is being engineered for male sterility. A gene required for the development of pollen, or of the entire flower or male strobilus, can be blocked by *e.g.* the antisense RNA technique, or a gene encoding a toxic product can be fused to a promoter conferring expression only in male tissue. The benefits of male sterility in conifers are (i) to avoid unwanted spread of genes via pollen to native populations; (ii) to facilitate full-sib matings in indoor seed orchards, where no isolation of the female flowers will be needed; (iii) an opportunity to increase the biomass production via reduction of abundant male flowering which demands carbon resources.

The present state of research in flowering may be taken as a case study of the potentials and problems of the application of molecular genetics to forestry. The control of flowering is a prioritized area of traditional forest research. As discussed elsewhere in this book, shortening the time to flowering from the normal 20-30 years in conifers to nearer 10 years implies a prospect of two to three times faster progress in genetic improvement by traditional selective breeding.

Mainly because of its central importance for agricultural crops such as cereals, flowering is being intensively studied in *Arabidopsis*. This ‘model’ plant species has many advantages for modern genetic research in particular, a short generation time of about three weeks, a fully sequenced genome, easy genetic transformation, and a set of lines in each of which the expression of a known gene has been ‘knocked out’ or ‘knocked down’ by mutation or transformation so that its function can be studied in detail. Many genes from *Arabidopsis* have been shown to regulate critical aspects of flowering. This knowledge has led to the identification of genes of more or less closely related nucleotide sequence in other species. For woody plants, work of this kind has proceeded furthest with *Populus*, the first forest tree for which the genome has been sequenced. It is important to remember, however, that a gene of closely similar sequence to a well characterized flowering gene in *Arabidopsis* may have a different function in *Populus* or other species. One talks of a ‘candidate gene’ for the regulation of flowering in the other, less well characterized, species. A candidate gene is believed to influence the character of interest that requires further study.

An interesting example is the *FT* gene (*FLOWERING LOCUS T*). Mutants of *Arabidopsis* defective in the *FT* gene flower late in long days, leading to its identification as a gene regulating flowering. The *FT* gene product, probably the protein rather than the mRNA, is likely to be a main component of ‘florigen’. This is the substance or mixture of substances, hypothesized since the 1930s, that is induced in a leaf exposed to a floral inductive treatment, such as long days in *Arabidopsis*. The florigen is transported in the phloem to the site of action, a bud that is induced to flower. Useful practical applications of this knowledge are likely, as genes similar to *FT* have been identified in *Populus* and conifers. *FT* has recently been shown to regulate the photoperiodic reaction of budset as well as flowering in *Populus*, and budset in Norway spruce.

Another interesting gene apparently related to flowering is *DAL1*, isolated from Norway spruce. Its sequence is partly similar to a class of flowering genes in *Arabidopsis*. Plants of *Arabidopsis* transformed with *DAL1* from Norway spruce flower early. Such experiments are, however, difficult to interpret because of often occurring chan-
molecules of function of genes during evolution, as mentioned above. What makes DALI interesting from a practical forestry point of view is that it is unexpressed until Norway spruce trees are about 4-6 years old. Then it is expressed increasingly in the flanks of the meristem, bud-scale primordia, and vascular strands of needles and stems need-les of post-dormant trees, reaching maximum expression after 15-20 years. DALI expression is therefore believed to be a marker of the tree’s progress through the juvenile phase and young adult phase after which it is competent to flower. Later the gene is expressed in entire female and male cones at early stages of their development.

Because lignin removal during pulp and paper production is costly, modification of the lignin content and composition is a breeding goal amenable to genetic engineering. The lignin composition differs between angiosperms and gymnosperms. The lignin in angiosperms is relatively easier to extract by chemicals than lignin from gymnosperms. Therefore, besides reducing the total amount of lignin, a change of the lignin composition so that it will be more like the angiosperm lignin would be a desirable goal of genetic engineering in conifers.

The risks associated with transgenic trees are (1) an unwanted spread of transgenes to native populations, as indicated above, and (2) instability of gene expression. To mitigate the impact of these risks, two main options exist, to obtain reproductive sterility and to screen for stable gene expression. Both options pose great scientific challenges to genetic engineering in forest trees. However, a great incentive to take up these challenges is that forest biotechnology has the potential to offer significant economic and environmental benefits in the future.

It should be added that fears have been expressed that the diversity of livelihoods in the Third World will be eroded if the local varieties are driven out of the market and superseded by a few genetically engineered products. However, this applies equally to traditional breeding, if only a few commercial varieties are grown. In this context the function of different populations ought to be considered. This is discussed further in Chapters 9 and 11.

Summary

DNA is the molecule that carries the genetic information in most organisms. The three-dimensional structure of DNA was proposed by Watson and Crick in 1953. DNA, deoxyribonucleic acid, is a double-stranded polymer consisting of polynucleotides twisted around one another to form a double helix. A nucleotide is composed of a base, purine or pyrimidine, a deoxyribose sugar and a phosphate group. The ‘backbone’ of the polynucleotide consists of alternating sugars and phosphates. The bases adenine (A) and guanine (G) are purines, whereas cytosine (C) and thymine (T) are pyrimidines. Equal proportions of purines and pyrimidines are found in DNA because the bases are paired, adenine with thymine (A-T) and cytosine with guanine (C-G). The base pairing holds the two complementary polynucleotide chains together. This structure of the DNA molecule also indicates that the genetic information lies in the sequence of the bases, unique for each gene.

The replication of DNA follows the semiconservative model, in which each parental strand serves as a template for the synthesis of a new strand and thus the two daughter helices will consist of one old parental and one new strand. Although the DNA molecule is characterized by great stability, mutations can occur during replication that change the sequence of the bases.

DNA is located in the chromosomes, and exists as only one continuous molecule in each chromosome. It is densely packed following progressive windings, coils, and foldings.

DNA consists of coding, genic DNA and non-coding regions, non-genic DNA. Eukaryotic DNA can exist as unique, single and low-copy, functional genes, i.e. genes that are found in only one or few copies per haploid genome, and as different types of repetitive DNA, or as spacer DNA.

In most higher organisms and especially in conifers, non-coding DNA, so called junk DNA, comprises a very large part of the total DNA. More than 99% of DNA in Norway spruce and Scots pine is probably of minor or no importance for their environmental adaptation.

The central dogma of molecular genetics says that DNA transcribes its information to a RNA molecule called messenger RNA (mRNA) that moves out of the cell and to the ribosomes on which the information is translated into proteins. RNA, ribonucleic acid, is a single-stranded polynucleotide in which deoxyribose is replaced by ribose and the base thymine with uracil. Small RNA molecules, the transfer RNAs (tRNA), bring the amino acids to the mRNA on the ribosomes in the sequence determined by the order of the nucleic-acid bases on mRNA. Thus, a sequence of three bases of the mRNA, a codon, codes for a specific amino acid. The genetic code is the set of rules specifying the correspondence between the codons in DNA or RNA and the amino acids in the proteins. Special codons serve as start and stop signals for protein synthesis. Each transcribable gene has a region at which the transcription of the gene is regulated, i.e. that regulates the synthesis of mRNA. This region is called the promoter. The synthesis of RNA is carried out with help of an enzyme called RNA polymerase that binds to the site of transcriptional initiation of the promoter. Before the RNA molecule is released from the nucleus, it is processed in several steps. The main feature of the processing is to excise the so-called introns, short sequences of DNA that interrupt the coding regions, the exons. The
gene expression is controlled both at the transcriptional and translational level.

The gene order seems to be highly conserved both within genera, e.g. among pine species, and over wide taxonomic families, independently of genome size. This facilitates gene mapping.

The molecular clock hypothesis assumes that each gene has its molecular clock that ticks at an approximately constant rate. The molecular clocks have become very valuable for calculating the dates of speciation events and for constructing so-called phylogenetic (evolutionary) trees.

The largest amount of DNA is found in the nucleus of the cell. Additional DNA can be found in the cytoplasmic organelles, the chloroplasts and the mitochondria. The DNA in organelles performs its own replication, as well as transcription and translation. These processes are very similar to those occurring in bacteria, which according to the endosymbiotic hypothesis, indicates that these organelles are of bacterial origin.

Genetic linkage maps were developed for a number of tree species. The main purposes are to identify genes for quantitative traits (QTL) for use in tree breeding, and to form the basis of gene cloning and production of genetically engineered trees.

Another application of molecular genetic methods is DNA sequencing to determine the order of bases. To date (2006), nearly completely sequenced genomes of three higher plants were published: Arabidopsis thaliana, rice, and Populus trichocarpa.

The next great challenge will be to reveal the function of the sequenced genes. One promising method is the DNA microarray technique, which reveals the genes that are active in a particular tissue at a particular moment.

By means of the polymerase chain reaction (PCR), a gene or a DNA sequence can easily be amplified one million times in a very efficient way. The arrival of this method in mid 1980s was a great breakthrough in molecular genetics and and is now a routine in every well-equipped molecular laboratory for gene mapping and studies of the structure and function of genes. This technique provides a number of additional applications, e.g. in forensic medicine, as only a small amount of DNA is required initially.

Genetic engineering or recombinant DNA technology are synonyms for transferring genes between organisms, often from one species to another species using molecular genetic methods. The first step, after isolation of a DNA

*Figure 2-1. Steps involved in production of transgenic Picea abies plants. Identification of transformed cells was done by the blue-staining GUS-gene. Photograph Hartmut Weichelt.*
fragment including a desired gene, is to insert the DNA fragment into a so-called vector that can be a plasmid, a small circular chromosome from a bacterium. This technique relies heavily on the discovery of the restriction enzymes which cleave the DNA at defined nucleotide sequences, characteristic of each enzyme. The plasmid, now a recombinant DNA molecule, is then introduced into a host cell in which it can replicate and produce many copies. Also these host cells can proliferate if cultivated on a selective medium allowing only those cells with recombinant DNA molecule to survive and propagate. This multiplication of a gene is an example of gene cloning.

The next step is to introduce the gene into a plant cell or an animal cell. For this purpose, two main methods exist: gene transfer via biological vectors and direct gene transfer via biolistic methods. Finally, to produce a transgenic plant, an efficient regeneration system is needed. This step is a bottleneck in many forest tree species, in particular conifers. But the number of species in which transgenic plants can be produced is increasing.

A general consensus exists about traits amenable to genetic engineering in forest trees: herbicide tolerance, resistance to insects and pathogens, reproductive capacity and lignin modification. In the near future also traits such as those affecting wood formation and fiber quality will probably be included.

However, there are risks associated with transgenic trees such as an unwanted spread of transgenes to native populations, and instability of gene expression. The mitigation of these risks poses great scientific challenges to genetic engineering in forest trees.

Further reading


3 Qualitative inheritance

In this chapter we shall first discuss how to distinguish between genetic and nongenetic variation. After that we shall explain the different types of inheritance, and how they depend on the number of genes involved. Discrepancies from these basic types of inheritance are also discussed.

Even before the Czech monk Gregor Mendel, in 1865, demonstrated the laws of heredity for certain characteristics in plants, it was commonly known that characteristics are transmitted from one generation to the next. Children resemble their parents, and sisters and brothers resemble each other. Before we consider the details of heredity, we shall first discuss the difference between variation due to genetic effects and variation without any genetic influence.

Genetic variation and non-genetic variation

All variation in nature need not be due to heredity, instead part of the variation can be due to environmental effects. Plants, in particular, have the ability to modify their growth and morphology so that two plants may become identical even if they carry different genes. On the contrary, morphological variation may exist without any genetic differences. Physiological differences can thus be of great significance.

In the 1910s, the Danish botanist Wilhelm Johannsen made his classical selection experiments with beans from pure lines. A pure line means progenies from individual homozygous mother plants and are usually produced by repeated self pollination. When he made selections within a pure line, he could demonstrate that irrespective of whether he selected for big beans or selected for small beans he got the same result with respect to the mean bean weight: neither any increase in weight after selection for big beans nor any decrease after selection for small beans was obtained. His results showed that he got no response to selection if the material was genetically homogeneous and consequently the observed variation within the material must be due to environmental influences. This is a good proof of the fact that variation must be due to genetic differences for selection to cause a change. This experiment also shows that it is important not to take for granted that differences observed in nature are genetically determined.

Mendelian inheritance

Mendel made a series of crosses between pea plants from different pure lines with contrasting phenotypic characteristics. The progeny, which in genetic crossing experiments is conventionally designated \( F_1 \) (the first filial generation), was often identical with one of the parents. Then, individual plants from the \( F_1 \) progeny were crossed and Mendel observed that there was a segregation of the different parental traits in this second filial generation, \( F_2 \). This was a fundamental discovery, since the result could not be explained as being caused by a mixture of parental traits, as was commonly believed in the middle of the 19th century when Mendel was active. Mendel observed a certain conformity of segregation in the \( F_2 \) generation. In most cases he observed that 3/4 of the plants carried one of the traits while 1/4 carried the other trait. The trait observed in highest frequency in the progeny was called dominant while the other was called recessive. Following studies on an extensive crossing material, Mendel could state that the segregation ratio 3:1 was generally found. Based on these results, he could provide a theoretical explanation of the inheritance of certain traits. Mendel denoted genes by letters and he let capital letters stand for dominant traits and small letters for recessive traits. Even this must be considered a stroke of genius. If we have two traits and let the dominant be \( A \) and the recessive be \( a \), and make a cross between two individuals both with the constitution \( Aa \), the female and the male parent will produce the same number of sex cells with \( a \) as with \( A \). The sex cells are often called gametes in genetic contexts. The constitutions \( AA, Aa \) and \( aa \) are called genotypes as was discussed in Chapter 1. A Punnett square, or checkerboard, is often used by geneticists for studying the segregation in which we assume that the gametes are formed in equal frequency in both the female and the male.

![Punnett square](image_url)

The result from such a square shows that an egg cell with \( A \) is just as likely to be combined with \( A \) pollen as with \( a \) pollen. Similarly, it is just as likely that an \( a \) egg cell combines with \( A \) pollen as with \( a \) pollen. Therefore, the ratio of the genotypes in the progeny will be 1 \( AA \): 2\( Aa \): 1\( aa \). Because \( A \) is dominant the phenotypic ratio will be 3 dominant: 1 recessive. It is not possible to determine which plants are \( AA \) and which are \( Aa \). To identify which plants are \( AA \) or \( Aa \), the plants are self-pollinated, if the species allows self-pollination. The homozygotes \( AA \) produce only \( AA \) progeny, while the heterozygotes \( Aa \) segregate.

---

43
In a ratio of 3:1. Another alternative for identifying the AA and Aa, is by crossing the AA and Aa with plants that are homozygous recessive aa. This is called a test cross. If the recessive aa is crossed with the homozygote AA, the result is that all plants in the F₁ progeny will be Aa and show the phenotype of the dominant A allele. A segregation in the F₁ progeny will only occur if the recessive aa is crossed with the heterozygote Aa and the ratio will be 1:1. This can easily be verified by a Punnett square as described above. In this case the four squares are replaced by two squares as the homozygotes aa only produce one type of gamete, the a gamete.

In his book "Genetics: Basic and Applied", the Swedish geneticist Arne Münzing has made an excellent summary of the great significance of Gregor Mendel's discovery as follows:

Mendel realized that his results could only be explained by the assumption that hereditary differences between the intercrossed parents depended on individual, constant units of heredity – later on called genes, which in an unchanged condition are transmitted by the sex cells from one generation to the next.

This concept of constant units of heredity was something quite new for people at that time and it would take some additional decades until the discoveries of Mendel were more widely spread in biological research. This occurred in 1900, when three researchers independently rediscovered Mendel’s results.

In genetics, the genotype of the female is, by convention, written first in a cross. Thus, in the cross aa x Aa the female is homozygous recessive aa and the male is heterozygous Aa.

Mendel also studied what happened when more than one pair of traits appeared in the parents. For example, he crossed pea plants with yellow and round seeds with other pea plants with green and wrinkled seeds. All plants in the F₁ progeny were yellow and round. Then Mendel self-fertilized the F₁ progeny. In the F₂ progeny he observed four seed phenotypes and their segregation was in good accordance with

9 yellow and round
3 yellow and wrinkled (brown frame in the Punnett square in next column)
3 green and round
1 green and wrinkled (brown frame in the Punnett square in next column)

We immediately observe that two of the trait combinations were not present in the parents at all, i.e. yellow and wrinkled as well as green and round. How to explain this segregation? Because the F₁ progeny showed yellow and round seeds we must assume that these two traits are dominant and we let Y stand for yellow seeds and R for round seeds. Green and wrinkled seeds must be recessive traits and their genes are designated as y and r, respectively. If we assume that the segregation of the two seed traits, seed colour and seed shape, is dependent on two pairs of alleles in two loci and that these alleles undergo independent assortment, the F₂ progeny will produce four types of gamete in equal frequency, YR, Yr, yR and yr. This is illustrated in the 4 x 4 Punnett square above.

In this grid, 9 of the 16 squares include both Y and R, which means that these squares contain genotypes that will show yellow and round seeds. It should be observed that there are four different genotypes behind this phenotypic trait combination. Behind each of the two new trait combinations there are two genotypes. Finally, the recessive trait combination green and wrinkled seeds is represented by one genotype, the double homozygote yyrr. Crosses between parents differing at two loci display a dihybrid segregation while crosses between parents differing at one locus display monohybrid segregation.

If we have three pairs of traits, a plant heterozygous at all three loci will produce eight different gametes. If the alleles at the third locus is designated T and t, each of the four gametes produced in the dihybrid cross will combine with T or t, and eight different gametes will appear. The Punnett square for this segregation will thus contain 64 squares. The segregation ratio of phenotypes will be 27: 9: 9: 3: 3: 3: 1.

The phenotypes have different colors in the large Punnett square on top of next page. From this trihybrid square it is clear that plants homozygous at all three loci are very rare. The more loci involved in segregation the lower the frequency of plants being homozygous at all loci.

Based on these Punnett squares, a couple of general formulae can be framed concerning number of different gamete types and genotypes produced in the progeny when the parents are heterozygous at all loci.

<table>
<thead>
<tr>
<th>Gametes</th>
<th>Genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2ⁿ</td>
<td>3ⁿ</td>
</tr>
<tr>
<td>Monozygotic</td>
<td>2</td>
</tr>
<tr>
<td>Dihybrid</td>
<td>4</td>
</tr>
<tr>
<td>Trihybrid</td>
<td>8</td>
</tr>
</tbody>
</table>
The gametes are produced in a frequency of $2^n$, where $n$ stands for the number of heterozygous loci. The genotype frequencies are still larger = $3^n$. For Norway spruce and Scots pine with 12 chromosome pairs, heterozygosity at one locus on each chromosome pair for the two parents crossed, should theoretically generate $3^{12}$ different genotypes in their progeny. This is a number somewhat larger than 500000. For a lime tree with its 41 chromosome pairs, heterozygosity at one locus on each chromosome pair will result in $3^{41}$ different genotypes in the progeny. The number is unbelievably large, $= 3.6 \times 10^{19}$, probably greater than the total number of lime trees in Sweden. In those situations when a locus has more than two alleles, multiple alleles. With this knowledge about the immense possibilities of variation in mind, it is easy to see that all living individuals in cross-fertilized organisms are unique genetically, except for those individuals regenerated via cleavage of a fertilized egg, as is the case for identical twins.

If you are uncertain of how different traits are inherited you can carry out the same type of crosses as Mendel did. What is important to remember is that the number of plants in the progeny should be large enough so that the segregation can be verified with statistical significance. By chance, we will seldom or never expect to get exactly the segregation ratios 3:1 or 9:3:3:1 and so forth. The statistical method used to estimate the probability of one type or the other type of segregation is called the $\chi^2$ method. One example will illustrate the procedure for estimating the goodness-of-fit between the observed numbers and and the expected numbers. In an F$_2$ progeny obtained from the original cross between two parents, one with yellow and round seeds and the other with green and wrinkled seeds, the segregation in the progeny was as follows:

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>yellow and round</td>
<td>100</td>
</tr>
<tr>
<td>yellow and wrinkled</td>
<td>27</td>
</tr>
<tr>
<td>green and round</td>
<td>25</td>
</tr>
<tr>
<td>green and wrinkled</td>
<td>8</td>
</tr>
</tbody>
</table>

Based on these figures, we can suspect that the segregation is 9:3:3:1, which is the expected frequency of phenotypes. This expected, ideal frequency should be compared with the observed frequency. If we start with our 160 plants, we will get the following expected numbers:

- yellow and round   90
- yellow and wrinkled 30
- green and round 30
- green and wrinkled 10

To estimate the goodness-of-fit, we must first calculate the difference between the observed and the expected values, thus, 100-90; 27-30; 25-30 and 8-10. $\chi^2$ value is the sum of the squared deviations divided by the expected values, respectively. In general, this is written $\chi^2 = \sum(d^2/m)$, where $d$ is the deviation and $m$ the expected number. In our case we get

$$\chi^2 = \frac{(100-90)^2}{90} + \frac{(27-30)^2}{30} + \frac{(25-30)^2}{30} + \frac{(8-10)^2}{10} = 2.63$$

### Box 3-1

Number of genotypes — $N$ — formed during recombination of heterozygous loci on the assumption of no linkage between loci:

$$N = \left(\frac{r(r+1)}{2}\right)^n$$

where $r$ = number of alleles in each locus, $n$ = number of segregating loci, $N$ = number of genotypes resulting from recombinations.

Theoretically, *Picea abies* and *Pinus sylvestris*, both having 12 pairs of chromosomes, can produce 531441 different genotypes if only one locus per chromosome pair is heterozygous. This means that in reality the resulting number of genotypes will be infinitely large.
Using tables in textbooks of statistics, we can find the probability for the observed segregation being caused by chance under the hypothesis put forward for the expected frequencies. By this probability, called the p value, we can infer whether the observed segregation is caused by chance or is caused by the fact that the observed segregation did not agree with our hypothesis. In our material we have four classes and therefore three degrees of freedom. In a $\chi^2$ table for three degrees of freedom we find that the probability value is between 0.3 and 0.5. This means that it is a fairly high probability that we are dealing with a 9:3:3:1 segregation. The larger the $\chi^2$ value the less is the probability that the observed segregation ratio agrees with the expected segregation ratio. Mendel reported an extremely good agreement with the expected ratios. Therefore, he has afterwards been under suspicion of having to some extent changed the value in his experiments, since such a good agreement with the expected segregation ratios is very unlikely in repeated experiments.

A premise of attaining segregation ratios such as 3:1 and 9:3:3:1, is the independent assortment of alleles i.e. that genes in different pairs of alleles are inherited independently of each other. In some cases, it has been observed that the traits of the parents appear in a higher frequency than expected. The reason for this is probably that loci for these traits are sited on the same chromosome. The loci belong to the same linkage group. If the two loci are located very near each other on the same chromosome, perhaps only the parental combinations will be found in the $F_2$ progeny. The only way of breaking up the parental combinations is if crossing-over between the two loci takes place during meiosis. Let us assume that we make the following type of cross, $AAbb \times aabb$, where the $a$ locus is linked to the $b$ locus, and that the $F_1$ progeny is selfed. We further assume that one out of 10 gametes is a crossover gamete. This will give us the following gamete ratio: 9 $Ab$: 9 $ab$: 1 $AB$: 1 $ab$.

The two last gametes are results of crossing-over between the $a$ and $b$ loci. To derive the genotypes formed in the next generation we have to consider that the gametes do not occur in equal frequencies. In population genetics this is usually done by introducing the fractions of the gametes in the Punnett square. These fractions we do obtain by dividing each figure (9, 9, 1, 1) with 20, which results in two gamete frequencies 0.45 ($Ab$ and $aB$) and 0.05 ($AB$ and $ab$). However, to simplify the calculations in the Punnett square we use the whole numbers 9, 9, 1, 1.

In the Punnett square we introduce these for the gamete frequencies:

Phenotypically, we get the following segregation:

| 201 | $AB$ |
| 99  | $Ab$ |
| 99  | $aB$ |
| 1   | $ab$ |

Using the $\chi^2$ method this segregation can be compared with the expected segregation after independent assortment: 225:75:75:25. This will result in a very large $\chi^2$ value, indicating that the observed deviations from expected are not caused by chance but have other causes, i.e. the $a$ locus and the $b$ locus are closely linked.

Also other deviations from the expected 9:3:3:1 ratio can be found in the progeny from crosses between parents that are heterozygous at two loci. One deviation is caused by the $B$ allele. If this allele is only expressed in the presence of the $A$ allele, the phenotypic segregation ratio will be 9:3:4. In other cases, $A$ and $B$ can be mutually dependent on each other, which results in a 9:7 ratio that can be difficult to separate from a 1:1 ratio. A large progeny is needed to be able to statistically separate these two ratios. If a trait is expressed only when certain alleles at other loci are present, as in the two cases discussed above, this is called gene interaction or epistasis.

**Gene effects at the biochemical level**

As the relationship between genes and proteins has been demonstrated, we can more easily understand such phenomena as dominance and recessiveness. If the enzyme $E$, that the gene $A$ is encoding, is needed for a precursor $D$ to be transformed to substance $F$, we can easily realize that this transformation can take place in the homozygote $A_1A_1$ as well as in the heterozygote $A_1A_2$. In contrast, the homozygote $A_2A_2$ results in a plant/tree with the phenotype $D$. The picture will be more complicated if we assume that the phenotype $F$ can only be produced when the enzyme exists above a certain threshold value. If we further assume that $A_1$ produces more enzyme $E$ than $A_2$, the expression of the phenotype of $A_1A_2$ depends on whether the amount of enzyme $E$ is above or below the threshold value. If it is above the threshold, $A_1A_1$ and $A_1A_2$ will again show the phenotype $F$, while if it is below the threshold, the heterozygote $A_1A_2$ will have the $D$ phenotype. In some instances, the phenotype of the heterozygote will be completely intermediate between the two homozygotes. In such cases, we must assume that development of the different phenotypes is limited by the enzyme produced by the genes, so that two genes of $A_1$ produce double the amount of enzyme compared to the production of single genes. In Norway spruce and Scots...
pine, there are several chlorophyll mutants with decreased amounts of chlorophyll. In a heterozygote of such a mutant, Björn Walles observed that the amount of chlorophyll was half that in the homozygous dominant plants.

**Summary**

Gregor Mendel elucidated the hereditary transmission of the so-called qualitative traits. He realized that the hereditary determinants, later called genes, were transmitted unchanged from one generation to the other and that no unspecific mixture of the contribution of the two parents occurred. He also realized that some genes are dominant, others are recessive. Depending on how many traits are involved in the experiment with crosses between two heterozygote parents we shall get:

for **one gene pair**, a segregation ratio of 3:1 for dominant : recessive

for **two gene pairs**, a ratio of 9:3:3:1, dominance for both traits : dominance for one trait : dominance for the other trait : recessive for both traits.

In general, the number of gametes produced in a plant heterozygous at n loci = 2^n. The number of different genotypes produced after a cross between two parents both being heterozygous at the same n loci = 3^n. Linkage exists if loci that control two traits are located close to each other on the same chromosome arm. If this is the case, the two traits do not segregate independently. The closer the two loci are located, the larger are the deviations from the expected segregation when the two loci are located on different chromosomes.

To determine whether the observed segregation ratio is in accordance with a certain expected segregation ratio, a $\chi^2$ test is required.

**Further reading**


Population genetics - Hardy-Weinberg law

In this chapter we focus on the Hardy-Weinberg law. The concept of effective population size is presented. Estimates of population differentiation and inbreeding are introduced. A brief introduction to \( F \) statistics is also given. Issues on population genetics are also presented in Chapter 6.

Population genetics deals with studies of allele frequencies in populations and their changes. Such changes may be caused by mutations, genetic drift, gene flow, and natural selection. Therefore, population genetics is of great importance for evolutionary issues, which are treated in more detail in Chapter 6. The meaning of mutations has been outlined earlier. Genetic drift is a random process that is of greatest significance in small populations. Various types of migration among populations are called gene flow. Natural selection has occurred when certain individuals in a population have been more successful in passing their alleles to the progeny generation than other individuals of the same population. This ability must be attributed to a better vitality of the successful individuals under the environmental conditions where the population grows.

In the simple Punnett squares used for deriving mono-hybrid, dihybrid, and trihybrid segregations in diploid organisms, all alleles occur at frequencies of 50%. Mostly we express the frequency in fractions of 1 and in this case the frequency is written as 0.5. Populations in nature can have all kinds of allele frequencies between 0 and 1. In populations the allele frequencies of two alleles in one locus are usually designated as \( p \) and \( q \), where \( p + q = 1 \). When there are multiple alleles in one locus the alleles are designated as \( p, q, r \), etc. Note that the sum \( p + q + r \ldots \) also in this case is 1.

To analyse the changes in allele frequency of a population from one generation to the next we benefit from the Hardy-Weinberg law. The name emanates from the two men who independently of each other presented their findings. The law is most simply described by an example. In a very large population the genotype frequencies are assumed to be the following:

\[
\begin{align*}
\text{Male} & & \text{Female} \\
0.7 & a_1 & 0.3 & a_2 \\
0.3 & a_2 & 0.7 & a_1 \\
\end{align*}
\]

To enable us to derive the genotypic composition after complete random mating we have to assume that the genotypes contribute gametes in the frequency with which they occur. Thus \( a_1a_1 \) contributes 60% of the gametes while \( a_1a_2 \) and \( a_2a_2 \), each contributes 20% of the gametes. For the allele \( a_1 \) the genotype \( a_1a_1 \) will give rise only to \( a_1 \) gametes while half of the gametes from the heterozygote \( a_1a_2 \) will carry \( a_1 \) alleles:

\[
0.6 + 1/2 \times 0.2 = 0.7.
\]

In an analogous way we shall find the frequency of \( a_2 \) alleles to be:

\[
0.2 + 1/2 \times 0.2 = 0.3.
\]

The probability that an \( a_1 \) allele will participate in the fertilization in this large population is 0.7 while the corresponding probability for an \( a_2 \) allele is 0.3. These frequencies are valid if the matings are random, there are no new mutations in this \( a \) locus, that there is no gene flow, that genetic drift or natural selection does not occur. Since the frequencies of the two alleles differ we have to introduce these frequencies in the Punnett square above to obtain the frequencies of the three genotypes in the progeny. The probability that an \( a_1 \) pollen will fertilize an \( a_1 \) egg cell is in our case 0.7 x 0.7 = 0.49.

The genotype frequencies are equal to their probabilities and by summarising the probabilities for the heterozygotes in the Punnett square we obtain the following genotype frequencies:

\[
\begin{align*}
\text{Male} & & \text{Female} \\
0.7 & a_1 & 0.3 & a_2 \\
0.3 & a_2 & 0.7 & a_1 \\
\end{align*}
\]

\[
\begin{align*}
\text{Male} & & \text{Female} \\
0.7 & a_1 & 0.3 & a_2 \\
0.3 & a_2 & 0.7 & a_1 \\
\end{align*}
\]

Which gamete frequencies do we get from this population? For the \( a_1 \) allele we get the following frequency: 0.49 + 1/2 x 0.42 = 0.70; in an analogous way we get for \( a_2 \): 0.09 + 1/2 x 0.42 = 0.30. In other words the allele frequencies remain unchanged and we can use the same Punnett square as above to derive the genotypic frequencies in the second generation progeny. Hardy-Weinberg law says that the allele and genotype frequencies remain constant from generation to generation if none of the factors listed above (mutations, genetic drift, gene flow, natural selection) exert any influence on the population. The
Another characteristic of the Hardy-Weinberg law is that the equilibrium in one locus is obtained immediately after random mating. If we consider two or more loci the equilibrium is reached somewhat more slowly. This law also shows that genetic variation remains from generation to generation under the conditions given above. Deviations from the expected genotypic frequencies according to Hardy-Weinberg law suggest that random mating does not exist, that there is gene flow into the population or that natural selection is in operation.

Other important information from Hardy-Weinberg law is that rare alleles mainly occur in heterozygotes. An example will shed some light on this. If the rare allele occurs at a frequency of 0.01 we shall have the following genotypic frequencies:

\[
\begin{align*}
    a_1a_1 &= 0.0001 \\
    a_1a_2 &= 0.0198 \\
    a_2a_2 &= 0.9801
\end{align*}
\]

Generally the rarer an allele, the wider the gap between the frequencies of the homozygous and heterozygous carriers of the rare allele. This means that it is hardly possible to clean the population from a rare recessive vitality-reducing allele since most of the recessive alleles occur in the heterozygotes which cannot be distinguished from the dominant homozygote.

Hardy-Weinberg law helps to explain why the frequency of homozygotes decreases when isolates are broken, which has been called the Wahlund principle (see Box 4-1). Breaking of isolates is of positive significance for recessively conditioned human diseases such as cystic fibrosis and sickle-cell anemia.

One reason for deviations from random mating may be that all individuals do not participate in flowering or fruiting in a population. The sum of those individuals that contribute are referred to as the effective population size and are designated \( Ne \). It should be noted that \( Ne \) is usually estimated in a more complex way than presented here. It is a general biological phenomenon that all individuals in a population do not contribute to the production of a progeny. This is of great significance for decisions about how many trees should be included in a gene resource population.

### Box 4-1 Wahlund’s principle

One simple example makes the meaning of this principle clear. Two populations both large enough for random mating according to the Hardy-Weinberg law, have the following genotype and gamete frequencies

<table>
<thead>
<tr>
<th>Genotype frequency</th>
<th>( a_1a_1 )</th>
<th>( a_2a_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.64</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>0.32</td>
<td>0.48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene frequency</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.80</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.60</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
    \text{Total frequency of homozygotes} & = \frac{0.64 + 0.04 + 0.16 + 0.36}{2} = 0.6 \\
    \text{After the fusion of these two populations to a single population with random mating, according to the Hardy-Weinberg law, the following gamete and genotype frequencies will be obtained:}
\end{align*}
\]

\[
\begin{align*}
    a_1 &= \frac{0.8 + 0.4}{2} = 0.6 \\
    a_2 &= \frac{0.2 + 0.6}{2} = 0.4
\end{align*}
\]

\[
\begin{align*}
    a_1a_1 &= 0.36 \\
    a_1a_2 &= 0.48 \\
    a_2a_2 &= 0.16
\end{align*}
\]

The frequency of homozygotes has decreased from 0.60 to 0.52, which is what Wahlund’s principle says. By breaking an isolate the frequency of homozygotes will be lower if the merging populations have different gene frequencies.

In nature the effective population size may vary from one generation to the next. The effect of a strong reduction of \( Ne \) means that it becomes much less than the arithmetic mean over generations. To estimate the \( Ne \) the following equation is used:

\[
1/Ne = 1/t \times \Sigma 1/ni
\]

in which \( t \) stands for the number of generations, \( ni \) stands for \( Ne \) in a certain generation. If \( Ne \) over five generations is 20, 80, 100, 125, and 175, respectively, we obtain from the above equation \( Ne = 58 \), which is considerably less than the arithmetic mean of 100.
Before describing the concepts in Fig. 4-1. All three parameters, population differences, and inbreeding coefficient will be introduced in next chapter. $F_{ST}$ estimates the reduction in heterozygosity in a subpopulation due to genetic drift and thus is a measure of the relative differentiation in allele frequencies between subpopulations. Therefore, estimates of $F_{ST}$ are frequently presented in reports on population differentiation studied by aid of isozymes. The inbreeding within subpopulations is estimated by $F_{IS}$, which is a measure of the reduction of heterozygosity of individuals within subpopulations. $F_{IT}$ is an estimate of the reduction in heterozygosity of an individual in relation to the total population. Expressed in another way, $F_{IT}$ is the total inbreeding in all subpopulations. It is thus a combined effect of non-random mating within subpopulations ($F_{IS}$) and the effect of population subdivision ($F_{ST}$).

It should be noted that it is not always straight-forward to compare $F_{ST}$s from different studies since these estimates depend on the loci analysed and whether non-polymorphic loci are included in the estimate or not. The selection of populations influence the estimates of $F_{ST}$ too.

$G_{ST}$ is another parameter frequently used for estimation of population differentiation by markers. $F_{ST}$ and $G_{ST}$ are identical if there are only two alleles at a locus.

For species with different genders, the number of females and males play a role for $N_c$. For such a situation the following equation is valid:

$$N_c = 4 \frac{N_m N_f}{(N_m + N_f)}$$

in which $N_f$ stands for the number of females and $N_m$ for the number of males. If the number of females is 50 and the number of males is 200, $N_c$ becomes 160, which is considerably less than the total number of individuals.

**F statistics**

Before describing $F$ statistics it should be noted that it is beyond the scope of this book to carry out derivations of the concepts introduced in this section.

As will be discussed in more detail in chapter 11 it is of interest to encompass existing genetic variation when sampling gene resource populations. It is also of interest to avoid a high degree of inbreeding in the gene resource population. $F$ statistics are useful means to get information on population differentiation and amount of inbreeding. An attempt to visualise $F$ statistics parameters is done in Fig. 4-1. All three parameters, $F_{IS}$, $F_{IT}$, and $F_{ST}$, are a kind of inbreeding coefficients. The concept of inbreeding coefficient will be introduced in next chapter.

![Figure 4-1. A schematic illustration of the concepts $F_{ST}$, $F_{IS}$, and $F_{IT}$ and their relationship](image)

$F_{IS}$ stands for the number of females and $F_{IT}$ for the number of males. If the number of females is 50 and the number of males is 200, $F_{IS}$ becomes 160, which is considerably less than the total number of individuals.

$F_{IS}$ = reduction of heterozygosity of an individual due to non-random mating within subpopulations

$F_{IT}$ = reduction of heterozygosity of an individual in relation to the total population

$F_{ST}$ = fixation index = reduction in heterozygosity of a subpopulation due to genetic drift

$F_{IS} = \frac{H_s - H_i}{H_S}$

$H_i$ = The heterozygosity of an individual in a subpopulation

$F_{IT} = \frac{H_T - H_i}{H_T}$

$H_T = \bar{H}_S$ = The expected heterozygosity of an individual in an equivalent subpopulation averaged over all subpopulations

$F_{ST} = \frac{H_T - \bar{H}_S}{H_T}$

$\bar{H}_S = \bar{H}_T = $ Expected heterozygosity of an individual in an equivalent random mating total population

$\mathbf{I}$ = individual  $\mathbf{S}$ = subpopulation  $\mathbf{T}$ = total population
FST estimates have been used to derive the number of migrants per generation (cf Box 4.2). In this context it is worth stressing two things:

• Migration strongly prevents population differentiation,
• FST is inversely related to the number of migrants.

The latter is intuitively understood from the definition of gene flow, which is migration to a recipient population from another population with a different allele frequency. The stronger the migration into a population the smaller the difference between these two populations. Gene flow into a small population has a stronger impact than into a large population.

Most alleles involved in the regulation of a quantitative trait cannot be identified. To enable a comparison of population differentiation between marker traits and quantitative traits, estimates of the latter, designated QST, were derived based on different variance components, Vp, Vps, and Ve. (Variance is a statistical estimate of the variation in a specific trait in a population.) Vp is the population variance component, Vps is the population x block interaction variance component, and Ve is the variance of individuals (phenotypic value of a tree within a population). Assuming Hardy-Weinberg equilibrium the following formula is used for estimates of QST:

$$Q_{ST} = \frac{V_p}{V_p + 2h^2(V_{ps} + V_e)}.$$  

The denominator contains h^2, which is the heritability of the trait. Heritability is presented in chapter 5, Quantitative genetics. The heritability for a trait is an estimate of the resemblance between related individuals for that trait. As can be seen from the equation QST decreases by increasing heritability.

In the early part of this chapter it was stated that population genetics is of great importance for the understanding of evolution. Before evolution is discussed (Chapter 6) it is necessary to introduce different quantitative genetics concepts, which is done in the next chapter. Observed population differences for markers, FST and quantitative traits, QST, are presented in Chapter 7.

Summary.

Hardy-Weinberg law says that one generation of random mating causes equilibrium of the gene frequencies at one locus. This equilibrium is kept as long as the mating is random. This requirement for random mating is hardly ever fulfilled owing to mutations, natural selection, genetic drift, and gene flow. F statistics, with its parameters FST, FIS, and FIT, is frequently used in population genetics research. These three parameters are a kind of inbreeding coefficients. Inbreeding will be treated in next chapter. In many studies on population differentiation FST estimates based on isozyme variation are reported. The formula for estimates of population differentiation of quantitative traits is also presented.

Further reading

Quantitative genetics

The characteristics of quantitative traits are presented. The molecular genetics technique for detection and localization of quantitative trait loci (QTL) is outlined. Important concepts in quantitative genetics such as heritability, breeding value, combining ability, genotype x environment interaction, inbreeding, selection differential, selection intensity, genetic gain and genetic correlation are presented.

Characteristics of quantitative traits

In many plants, traits of value for the adaptation to certain environmental conditions such as frequency of flowering, seed production, growth rhythm and tolerance against diseases show continuous variation and are said to be quantitatively varying or quantitative traits. This means that there is no possibility to distinguish a distinct segregation in the progeny in contrast to the traits studied by Gregor Mendel.

After random mating from the crosses $Aa \times Aa$, $AaBb \times AaBb$, and $AaBbCc \times AaBbCc$, we obtain the genotypes shown in Chapter 3. If we prefer to study the segregation of genotypes only we shall get the following frequencies for the three simplest types of inheritance:

- monohybrid
  - No capital letter: 1
  - 1 capital letter: 4
  - 2 capital letters: 6
  - 3 capital letters: 15
  - 4 capital letters: 20
  - 5 capital letters: 15
  - 6 capital letters: 6

A mathematically skilled person sees that these figures are equal to the coefficients after expanding the binominal expression $(a + b)^n$, in which $n = 2, 4, 6$ respectively. This exercise in figures is meaningful for an understanding of the inheritance of quantitative traits, which may be affected by alleles at many more loci than the three discussed. For simplicity let us assume that that there are alleles at four loci that influence height growth in such a way that the homozygote $aabbccdd$ has a height of 20 meters at an age of 100 years. Let us also assume that each allele with a capital letter gives an additional height of 0.1 meter. If we know the coefficients for the binominal expression with $n = 8$ the frequency of the phenotypes are easily derived for this purely hypothetical case. The possibilities for us to distinguish the different classes 20.0, 20.1, 20.2,...,20.8 are evidently small owing to the slight differences among the different genotypes with different numbers of capital letters. In addition, the environmental conditions might blur the picture to make the distribution continuous rather than stepwise. Figure 5-1 illustrates that the distribution is close to a normal distribution. In most cases we assume that quantitative traits have a normal distribution.

The quantitative traits do not give distinct segregation in the progeny in contrast to what Mendel obtained in his crosses between yellow and green peas or wrinkled and smooth peas. The absence of distinct classes is a characteristic of quantitative traits. Frequently quantitative traits are affected by alleles at a large number of loci and the influence of each allele on the trait is minor. Modern molecular genetics has revealed that alleles at a certain segment of a chromosome exert a much larger influence than others. The technique is not detailed enough to say whether it is one locus that is involved or whether several linked loci are in action.

![Segregation from the cross AaBbCcDd x AaBbCcDd](image)

**Figure 5-1.** The distribution of tree height among different classes in the progeny from the cross $AaBbCcDd$ x $AaBbCcDd$ on the assumption that tree height is 20 metres for the recessive homozygote at all four loci and that each capital letter contributes 0.1 metre to the tree height.
Since the alleles at a locus do not always influence a trait in such a simplistic way as assumed for the tree height in the example above, the scale in Figure 5-2 is used to describe the allelic effect.

The use of capital letters and small letters to designate alleles might be misleading for quantitative traits. Therefore, the alleles are given indices to separate different alleles from each other. In the upper part of Figure 5-2 the genotypic value of the heterozygote, \(a_1a_2\), is closer to the homozygote \(a_1a_1\) than the other homozygote. The value \(d\) may be positive or negative. The heterozygote may even have its genotypic value outside the scale shown in Figure 5-2. At complete dominance, as was the case for the traits studies by Mendel, \(d\) is equal to 0. The heterozygote \(a_1a_2\) is therefore phenotypically equal to \(a_1a_1\).

The phenotype of a quantitative trait is the joint action of alleles at many loci as well as the effect of the environment at the site where the plant or tree is growing. Also for such traits as survival, for which there are only two classes, alive or dead, there is an underlying quantitative genetic variation for them. When the pooled genetic and environmental effects are below a certain value the plant will survive while it dies above this value. This value is usually called the threshold value. Healthy or diseased is another example of a trait that has only two classes but with an underlying genetic variation.

When the effect of the alleles is simply added to each other, as in the above hypothetical example of tree height, the gene action is completely additive. This was also the case in the first example of quantitative gene action described for kernel colour in wheat by the Swedish wheat breeder, Herman Nilsson-Ehle. His experiment revealed that there were three loci involved in the kernel colour of wheat, which is a hexaploid species. This is not surprising since there is one locus affecting the kernel colour in each of the three genomes of wheat.

### Figure 5.2. The genotypic values for three genotypes at one locus, in which the mean values of the homozygotes is 0. The value of the heterozygote is usually denoted as dominance deviation, \(d\), and may take any value on the scale, \(-a\) - \(+a\), as well as outside this scale. When the heterozygote is intermediate to the two homozygotes, below, \(d\) is zero and the gene action is completely additive.

QTL stands for Quantitative Trait Locus, i.e., those loci on the chromosomes that contain genes for quantitative traits. By making use of a large number of DNA segments it is sometimes possible to find associations with loci regulating a quantitative trait (QTLs). To date, traditional biometrical models have not been able to study individual genes for quantitative traits. Furthermore, in these models many simplified assumptions had to be made, for example, that a trait is affected by a very large number of genes, each with a small effect on the trait, that genes act additively and that they segregate independently. However, the molecular marker techniques developed for studying genes may significantly increase our knowledge about genes for quantitative traits. These techniques can provide information about the number of genes affecting the phenotypic variation of a trait, whether the variation is due to a few genes with large effects or a large number of genes each with small effects or whether a combination of both alternatives prevails. Information about interactions between genes at different loci and between genes and environment can also be gained. Construction of genetic linkage maps for QTL and gene markers will inform about the chromosomal positions of the genes for quantitative traits. Its application in breeding will be briefly presented in Chapter 9.

A prerequisite to make progress with QTL is that some of the quantitative alleles have a relatively strong influence on the development of the trait. Another prerequisite is that the parents have pairs of loci in linkage disequilibrium. This means that the alleles \(a_i\) and \(b_j\) always occur in the gametes of one parent and that \(a_i\) and \(b_j\) always occur together in the gametes of the other parent. The alternative linkage disequilibrium with \(a_i + b_j\) and \(a_i + b_j\), respectively, is equally useful. Constructing genetic linkage maps for QTL demands that a large number of plants/trees per family is analysed. Moreover, it is assumed that linkage disequilibrium is rare in wind pollinated tree species with a large and continuous distribution since there is a large gene flow among populations. Therefore, it is essential that the population designed for QTL detection is segregating (heterozygous) for as many of the QTLs as possible.

### Methods for constructing genetic linkage maps for QTL

It should be observed that QTL is a segment of a chromosome that may contain not only one but in some cases more than one locus affecting the quantitative trait of interest. The mapping of QTLs means that they are localized to their sites, respectively, on the chromosomes. At the same time, their number and the proportion of the total phenotypic variation that they can explain are estimated.
We assume that the budset timing is a completely additive trait. After selfing of F₁, an F₂ is obtained. In case of complete linkage between marker and QTL locus (no crossovers), F₂ will consist of three types of plants, early M₁/M₁, intermediate M₁/M₂, and late M₂/M₂. If significant differences in budset timing are found between the three classes of markers, linkage is established and a QTL locus has been detected and mapped.

In the lower part of the figure, an example is given showing how to proceed in order to determine whether the plants are homo- or heterozygous for the marker gene. The example given holds for marker genes of RFLP type, for which both alleles in the RFLP locus can be identified, here denoted M₁ and M₂. We assume that this locus has a recognition site for the enzyme EcoRI (see below) but that a change has occurred in allele M₂ so that the enzyme cannot cut. This causes unequal RFLP-fragments during the separation of the fragments on a gel using electrophoresis. The plant with late budset has a long fragment (indicated by the blue brace) that contains QTL₂ and M₂, and can therefore only move a short distance on the gel, while the plant with early budset has a short fragment (indicated by the green brace) with QTL₁ and M₁ and therefore moves a longer distance. The plant showing intermediate budset is heterozygous for both QTL and M loci and can be identified because both fragment sizes occur on the gel.

If the marker locus and the QTL locus are at long distance from each other, or on different chromosomes (in different linkage groups), all three marker gene categories will be intermediate in timing of budset.

### Phenotype

<table>
<thead>
<tr>
<th>Ratio at complete linkage</th>
<th>Early budset M₁/M₁</th>
<th>Intermediate budset M₁/M₂</th>
<th>Late budset M₂/M₂</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 : 2 : 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Recognition site where Eco R₁ can cut DNA**

Using electrophoresis, the fragments of unequal size are separated on a gel, on which DNA from the 3 F₂ plants has been added. Rows 1, 2, and 3 represent plant 1, 2, and 3, respectively.

---

**Figure 5-3.** Linkage determination between a marker gene locus with two alleles M₁ and M₂ and a locus for a qualitative trait, the timing of budset, with two alleles QTL₁ and QTL₂. Parent 1 shows an early budset and is homozygous for both alleles M₁ and QTL₁ while parent 2 shows a late budset and is homozygous for both alleles M₂ and QTL₂. After mating of the two parents an F₁ progeny is obtained that is intermediate in timing of budset, heterozygous for QTL₁/QTL₂, and heterozygous for marker loci M₁/M₂. We assume that the budset timing is a completely additive trait.

The radioactively labelled DNA probe visualizes the 2 bands using autoradiography and reveals whether the plants are homozygous or heterozygous for the marker gene. This information is needed for QTL detection.

<table>
<thead>
<tr>
<th>QTL₂ M₂</th>
<th>Longer fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>QTL₁ M₁</td>
<td>Shorter fragment</td>
</tr>
</tbody>
</table>

**Eco R₁ cannot cut** because the recognition site is changed. The plant with late budset has a long fragment (indicated by the blue brace) that contains QTL₂ and M₂, and can therefore only move a short distance on the gel, while the plant with early budset has a short fragment (indicated by the green brace) with QTL₁ and M₁ and therefore moves a longer distance. The plant showing intermediate budset is heterozygous for both QTL and M loci and can be identified because both fragment sizes occur on the gel.
Construction of genetic maps involves several steps:

1. Generation of maps based on genetic markers where the markers should cover the whole genome, the denser the sites of the markers the better is the chances to detect linkage between marker loci and QTLs; the genetic markers should show a high degree of polymorphism, which means that it is highly probable that two individuals carry different alleles at each locus; the markers should also be neutral so that they do not affect the trait of interest or affect the regeneration capacity.

2. Establish linkage between marker locus and QTL. This presupposes that there exists a sufficiently large phenotypic variation in the quantitative trait in populations used for mapping purposes and that the QTL segregates in the population (Fig. 5-3). Therefore, the selection of suitable mapping populations is very crucial. The mapping populations employed to map QTLs in forest trees consist of various full-sib or half-sib mating designs. In addition, the adequate marker, whether dominant or codominant, has to match the type of mapping population used.

3. Advanced statistical methods of analysis have to be elaborated to detect significant associations between markers and QTL. An array of methods is now available for different mapping populations. However, the methods have their limitations and need to be developed further.

Results from detection and mapping of QTLs

In forest trees, detection of QTLs has been published for both conifers and hardwoods. Interspecific mating schemes have been used mainly in hardwoods. The quantitative traits involved are economically important traits such as growth, wood quality, adaptive traits to abiotic and biotic stress, and reproduction capacity.

Examples from studies of growth traits in a few species are given in Fig. 5-4, in which the range of phenotypic variance explained by individual QTL is given. Thus, in Pinus taeda, three QTLs for diameter and four QTLs for height were detected. The range of phenotypic variance explained by individual QTL for these growth traits was 15 percentage units. A few examples from other traits are given in next paragraph.

Five QTLs were individually responsible for at most 5% of the phenotypic variance of wood density in Pinus taeda. For the same trait, the five QTLs detected in an interspecific hybrid, Eucalyptus grandis x E. urophylla, individually explained up to 10% of the phenotypic variance. Similar magnitude of explained phenotypic variance was obtained for frost tolerance in Eucalyptus nitens, but only two QTLs were detected. For the interspecific hybrid Populus trichocarpa x P. deltoides, significantly larger phenotypic variances were explained. For example for the bud phenology trait, leaf flush, five QTLs were detected and one of these were responsible for 52% of the phenotypic variance. In an interspecific cross between the male hybrid clone Salix viminalis and a female Salix viminalis clone, six QTLs were detected for the timing of budburst. Individual QTL explained 12% to 24% of the phenotypic variance.

According to the expectation QTLs in species hybrids usually explain larger part of the variation than in pure species, mainly due to linkage disequilibrium. Additional information gained from QTL mapping, is that QTLs for different traits seem to appear in clusters, suggesting tightly linked QTLs or pleiotropy, i.e. that a single QTL affects more than one trait. The latter is what you should expect if there is high genetic correlation between two traits. This was the case in a study of Salix hybrids. Sometimes QTLs are co-localized to the same chromosome arm.

As regards the results presented, one has to be somewhat cautious, because many of these estimations probably underestimate the number of QTLs involved in each trait and overestimate the percentage phenotypic variance explained by each trait. The main reason for this is that the size of the family used for mapping was too small. This was indicated by simulation experiments in which it was shown that at a family size of less than 250 a few QTLs were erroneously identified, each with an exaggerated effect on that trait.

Genetic marker maps were developed not only for markers like RFLP and RAPD but also for AFLPs, microsatellites (SSRs), ESTs (For explanation see Chapter 7), and SNPs.

Following mapping of QTLs, the next task will be to identify the genes associated with QTLs. A procedure

![Figure 5-4. The range of phenotypic variance explained by single QTL in Pinus pinaster (6), Pinus taeda (3-4) Castanea sativa (28), Eucalyptus nitens (2-3), and the species hybrids Populus trichocarpa x P. deltoides (1-2) and Salix dasyclados x S. viminalis (1-8). Figures in brackets give the number of QTL for each species or species hybrid.](image-url)
suggested for this purpose is the so-called "candidate gene" analysis. The assumption behind this analysis is that candidate genes, sequenced genes probably affecting the trait expression, occupy a large portion of the QTLs affecting the trait. One way to go is to search in available genetic databases to obtain their sequences when the genes are isolated.

An additional task will be to study whether the QTL expression is stable over years, various environments and genetic backgrounds. Even if the family size is satisfactory large in a QTL study it ought to be stressed that the QTLs identified are valid for that particular family. Therefore, it is urgent to test more than one family to identify QTLs that are of general importance and not limited to just one family. Such knowledge will be of importance for their potential use in marker assisted selection in breeding programmes (See Chapter 9).

**Heritability**

The heritability is a concept of great significance in breeding and evolution. The meaning of heritability is visualised in Fig. 5-5. The phenotypes of the progenies are plotted against the phenotypes of their parents. The upper part of Fig. 5-5 shows a fairly good agreement between the parental and progeny phenotypes. This is a case of high heritability. If there is a high heritability for a trait there are possibilities to improve this trait in breeding since a tree with a good phenotype will give rise to a progeny with good phenotypes, too. Such a trait has the potential to become changed by natural selection. In the lower part of Fig. 5-5 a case is illustrated in which there is poor agreement between parents and offspring, which means that the heritability is low. For traits with low heritability it is impossible to identify the good genotypes via their phenotypes. The only way to reveal the good parents is to test their progeny.

To estimate the heritability, statistical methods are applied to analyze data from progeny trials. Mathematically the heritability = the additive variance \((\sigma_a^2)\) divided by the phenotypic variance \((\sigma^2_p)\). Somewhat later the meaning of additive variance will be presented. In formal terminology heritability of a trait = an estimate of the degree of resemblance between relatives for this trait.

Heritability is a relative concept that depends on the individuals tested as well as the environmental conditions during the test. An example will illustrate this. If frost hardness is tested a few hundred meters above the timber line, the probability is high that all plants will die. We shall not be able to reveal any genetic variation in frost hardness, which is a condition for obtaining a value of the heritability departing from zero. If the progeny trial is located a few hundred meters below the timber line, the probability is high that we shall be able to reveal genetic variation in the survival of the local population. In consequence a heritability differing from zero may be estimated. In certain experiments with Scots pine in northern Sweden the open pollinated progenies from individual trees in a population had an amplitude of 50 percentage units or more (Fig 5-6). As will be discussed in Chapter 7 phenological traits, such as budburst and growth cessation are of great significance for survival and good performance. The size of heritabilities for this kind of traits is much dependent on the point of time for the assessment. If the assessment is too early or too late during the process of development, limited variation will be revealed

![Figure 5-5. The relationships between the performance in parents and offspring, in which the mean values are 100 for parents and offspring. At a good agreement between parents and offspring the heritability is high.](image1)

![Figure 5-6. The survival of open pollinated progenies from one population in a 20 year old field trial of Pinus sylvestris in northern Sweden.](image2)
and a misjudgement of the heritability will occur. The ideal date of assessment for such traits is when the grand mean is close to when 50% of the development has taken place. The unfortunate situation is that we do not know this beforehand.

The relationship between parental and progeny phenotypes is just one of several options for estimations of the heritability for a trait. In destructive tests, as often with freezing tests or after inoculation with pathogens, we are forced to estimate the heritability with the aid of siblings.

With well designed experiments we can estimate both the additive variance and the phenotypic variance, the ratio of which is the heritability. A prerequisite is that there are replications in the experiment. The meaning of additive variance is explained in Box 5-1. Ten Scots pine trees are mated in all possible combinations with all other trees. The seed harvested from these crosses is used for establishment of well designed progeny trials. The mean tree height for all progenies in which tree 1 is one of the parents is shown in the center part of Box 5-1. In a similar way all mean values of the other nine trees are calculated and illustrated. In order to facilitate the understanding it should be noted that the differences are exaggerated compared to a situation in nature. Tree number 5 has the highest progeny and its deviation (D) from the mean value of all trees is illustrated. The breeding value is defined as 2D and the variance of the breeding value is the additive variance which we are interested in. Of great interest for breeders is to estimate the coefficient of additive variance (frequently abbreviated CVₐ) which is the square root of the additive variance divided by the phenotypic mean of the trait under consideration. The CVₐ gives us a possibility to evaluate the potentials for improvement by breeding for this trait. Theoretically we can find all 4 combinations of high and low heritability with high and low CVₐ. The possibilities for genetic improvement may be as good for a trait combining low heritability with high CVₐ as for a trait with a high heritability and low CVₐ.

The reason for defining the breeding value as 2D is that only half of the genetic material comes from one of the parents, the other half coming from other parents. As is evident from Box 5-1 the breeding value is dependent on the other parents tested as well as the environmental conditions under which the testing took place. This means that breeding value, additive variance, and heritability are relative estimates. They are valid for the population under
Additive variance

![Additive variance graph]

Figure 5-7. The relationship between gene frequency and additive variance with completely additive gene action; $a$ is the value as illustrated in Figure 5-2.

Table 5-1. Hypothetical values for all families after crosses between all parents.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>$\bar{X}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>32</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>21</td>
<td>25.6</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>-</td>
<td>31</td>
<td>31</td>
<td>29</td>
<td>30</td>
<td>30.8</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>30</td>
<td>-</td>
<td>25</td>
<td>23</td>
<td>23</td>
<td>24.8</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>32</td>
<td>24</td>
<td>-</td>
<td>21</td>
<td>27</td>
<td>25.8</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>28</td>
<td>24</td>
<td>22</td>
<td>-</td>
<td>34</td>
<td>26.4</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>31</td>
<td>22</td>
<td>26</td>
<td>32</td>
<td>-</td>
<td>26.2</td>
</tr>
<tr>
<td>$\bar{X}$</td>
<td>25.0</td>
<td>30.6</td>
<td>25.0</td>
<td>25.8</td>
<td>26.2</td>
<td>27.0</td>
<td>26.6</td>
</tr>
</tbody>
</table>

Table 5-2. Deviations from the mean value for all families in Table 5-1, 26.6, and the values for the General Combining Abilities, GCA.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>$\bar{X}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>5.4</td>
<td>-2.6</td>
<td>-1.6</td>
<td>-0.6</td>
<td>-5.6</td>
<td>-1.0</td>
</tr>
<tr>
<td>2</td>
<td>6.4</td>
<td>-</td>
<td>4.4</td>
<td>4.4</td>
<td>2.4</td>
<td>3.4</td>
<td>4.2</td>
</tr>
<tr>
<td>3</td>
<td>-3.6</td>
<td>3.4</td>
<td>-</td>
<td>-1.6</td>
<td>-3.6</td>
<td>-3.6</td>
<td>-1.8</td>
</tr>
<tr>
<td>4</td>
<td>-1.6</td>
<td>5.4</td>
<td>-2.6</td>
<td>-</td>
<td>-5.6</td>
<td>0.4</td>
<td>-0.8</td>
</tr>
<tr>
<td>5</td>
<td>-2.6</td>
<td>1.4</td>
<td>-2.6</td>
<td>-4.6</td>
<td>-</td>
<td>7.4</td>
<td>-0.2</td>
</tr>
<tr>
<td>6</td>
<td>-6.6</td>
<td>4.4</td>
<td>-4.6</td>
<td>-0.6</td>
<td>5.4</td>
<td>-</td>
<td>-0.4</td>
</tr>
<tr>
<td>$\bar{X}$</td>
<td>-1.6</td>
<td>4.0</td>
<td>-1.6</td>
<td>-0.8</td>
<td>-0.4</td>
<td>0.4</td>
<td>26.6</td>
</tr>
</tbody>
</table>

To link the statistically derived heritability with the quantitative genetics the heritability might be expressed with genotypic values as was done in Fig 5-2 in the following way:

$$h^2 = \frac{\sum 2pq[a + d(q - p)]^2}{\sigma_h^2}$$  \(1\)

in which q and p are allele frequencies and a and d are the values explained in Fig. 5-2. The effects of all the alleles at all loci that affect the trait must be summed. In the equation the factor $[a + d(q - p)]$ is an expression for the effect on the genotypic value of an exchange of $a_1$ for $a_2$ or $a_2$ for $a_1$. It might seem surprising that such changes of one or the other allele and vice versa do not have the same effect. This is a consequence of the exchange of the allele frequencies p and q as is evident from the equation. In the equation $a$ stands for the proportion of the genotypic effects that are added to each other and that are easiest to exploit in breeding.

It should be noted that the allele frequencies p and q may vary from locus to locus and that the heritability has a maximum at $p = q = 0.5$ since the product pq which is part of the numerator has its maximum at 0.25. This is reflected in Fig. 5-7 which illustrates the relationship between allele frequency and additive variance for complete additive gene action. In cases with dominance the curve takes another shape but the alleles at low frequencies do not contribute much to the additive variance in that case either. Alleles at very low or very high frequencies do not contribute much to the additive variance. A consequence of this is that alleles at such frequencies are hard to change by breeding. Similarly such alleles are hardly changed by natural selection.

The effects of $a$ and $d$ can vary from locus to locus. All these conditions make it impossible to distinguish the effects of the alleles at a particular locus. We have to be satisfied with the knowledge of the joint effect of alleles at several loci that affect the trait. Equation 1 shows that it is possible to connect the statistically estimated heritability with known genetic concepts such as allele frequencies and the effects of exchanges of alleles $a_1$ for $a_2$ and vice versa.

There is a close relationship between heritability and another concept, the General Combining Ability (the abbreviation GCA is frequently used to denote combinations of alleles at several loci that affect the trait).
used in texts). As with heritability, it can be estimated in progeny trials having material raised from systematic matings. When data from such trials are analysed, one often finds that one parent independently of mating partner gives rise to well performing progenies. This is an example of a parent with good GCA. More precisely expressed, the average deviation of this parent’s progeny from the grand mean of the trial is an estimate of the GCA of that parent. To illustrate this, hypothetical values are given in Table 5-1 for a trial in which all possible matings between 6 parents are involved except for selfing. When parents serve both as females and males, the mating design is designated as diallel. (A more detailed description of mating designs is carried out in Chapter 9 in connection with tree breeding since mating designs are of great importance in breeding.) From Table 5-1 it is evident that parent No 2 has high values in its progenies. Parent No 2 is thus an example of a parent with good general combining ability. The progenies 5 x 6 and 6 x 5 have values that deviate in a conspicuous way from the mean values of these two parents, which are close to the grand mean of this trial. Such a deviation is designated as Specific Combining Ability (SCA).

The grand mean for all crosses in Table 5-1 is 26.6. The family deviations and the parental deviations from the grand mean are given in Table 5-2. These latter deviations are estimates of the parental GCA on the assumption that the experimental error is $= 0$. A comparison of the information in Box 5-1 and Table 5-2 reveals that the GCA of a parent = half the breeding value of this parent. Under the same assumption it is possible to estimate the specific combing abilities by the aid of the following general equation:

$$y_{ij} = m + GCA_i + GCA_j + SCA_{ij}$$

where

- $y_{ij}$ is the value for the cross $i \times j$
- $m$ is the overall mean value
- $GCA_i$ is the general combining ability of parent $i$
- $GCA_j$ is the general combining ability of the parent $j$
- $SCA_{ij}$ is the specific combining ability of the cross $i \times j$

For the cross 5 x 6 we can approximately estimate the SCA in the following way:

$$y_{5 \times 6} = 33 = 26.6 + (-0.3) + 0 + SCA_{5 \times 6}$$

$$SCA_{5 \times 6} = 33 - 26.3 = 6.7$$

Table 5-3. Effects that can be distinguished in an experimental series planted at more than one test site (or alternatively exposed to more than one treatment) and containing more than one population with full-sib families or open-pollinated families.

<table>
<thead>
<tr>
<th>Effects</th>
<th>Full-sib progeny trial</th>
<th>Open-pollination progeny trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>grand mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>population x site/treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (population)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male (population)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>female x male (population)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>female x site/treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male x site/treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(female x male) x site/treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>residual</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5-8. Below. The curves show the phenotypic performance of different genetic entries (provenance, population, family, clone) along an environmental gradient. In the part below there is a great change in ranking among the entries. This means that a genotype x environment interaction exists. This is not the case in the upper part of the figure.

Phenotype

Environmental gradient
The estimations of the two combining abilities under real conditions takes place by using a more complex statistical model in which mating design and experimental design are important components. This enables an estimation of the significances of the two combining abilities. One example of the effects that can be distinguished in an experimental series planted at more than one test site and containing more than one population is given in Table 5-3. As seen from this table experiments with full-sibs increases our possibilities to identify different effects compared to the situation for open-pollinated families. All experimental trials containing the same crosses are designated as an experimental series.

The estimation of GCA is one of the main objectives in forest tree breeding and enables an identification of the genetically most valuable trees. It should be noted that the general combining ability of a tree is a relative estimate and depends on which parents are tested and the environment of the testing.

**Genotype x environment interaction**

Another objective of progeny testing is to estimate how stable the performance of the progenies is when tested under different environmental conditions. In Fig. 5-8 two situations are illustrated. Above is shown that the ranking is totally stable over the environmental gradient tested. In the part below there are several ranking changes. Such changes in ranking are called *genotype x environmental interactions*. To verify such an interaction we need at least two experimental plantations which differ with respect to the environmental conditions. A study of the genotype x environment interaction at two similar sites is of no value for estimates of genotype x environment interaction.

Knowledge of genotype x environment interaction is of value both for breeding and evolution. Forest genetic progeny trials belonging to one experimental series are therefore frequently located to shifting site conditions. According to which objective is of greatest importance we can calculate the heritability on data from all trials orheritabilities from individual trials. With a large genotype x environment interaction the heritability based on data from all experiments will be low. To evaluate the importance of the genotype x environment interaction for breeding, forest geneticists relate the variance component for the interaction to the parental variance component. As a rule of thumb, with a value above 1.0 there is a need for delineation of different breeding zones with separate breeding in each zone.

**Inbreeding and heterosis**

It is well known that different types of inbreeding in cross fertilising organisms cause a decrease of the vigour of the affected individuals. This is called inbreeding depression. Thanks to Nils Sylvén’s pioneering effort, the oldest progeny trial with selfed Norway spruce was established in 1916 (Picture 5-1). This trial was established before statistics were considered and it has no replications. In spite of that, the results are spectacular with a stem volume of the selfed trees amounting to less than 50% of the stem volume of the outbred trees. Still poorer performances of selfed Douglas fir, noble fir, ponderosa pine, and Scots pine have been observed in experiments with replications. Somewhat lower inbreeding was noted in one experiment with *Pinus radiata* (Fig. 5-9). The good site
The derivation of inbreeding coefficients will be done by one example. We have a mating between a homozygous *a*₁*a*₁ with two non-related males to obtain two individuals (F₁) in the progeny, which in turn are mated and give rise to the individual designated F₂, which is the individual we should derive the inbreeding coefficient for. In order to enable this, the two *a*₁ alleles are given in red and blue, respectively. The probability that the red *a*₁ allele is transferred from the *a*₁*a*₁ female to its daughter is = ½. The probability for transfer of the *a*₁ allele from F₁ to F₂ is also ½. The probability of transfer from *a*₁*a*₁ to F₂ is obtained by multiplying these two probabilities, which gives ¼. Similarly the transfer on the right-hand path from *a*₁*a*₁ to F₂ is also ¼. These two probabilities have to be multiplied, ¼ x ¼, which gives the probability = 1/16 for obtaining a homozygous F₂ individual. Another way to estimate this probability is to start from the F₂ individual and follow the red lines back to the starting point and for each step that is taken multiply with ½. The latter procedure may be simpler in more complex pedigrees than in the box. The other homozygote that can have copies of the same allele is *a*₁*a*₁. The probability that this homozygote will arise is equally large as the probability for *a*₁*a*₁. To calculate the total probability for identity by descent the two probabilities have to be summed. After summation we obtain 1/8, which is the inbreeding coefficient for half-sib mating. Note that the males in the top line cannot give rise to identity by descent. Shown in an analogous way the inbreeding coefficient for full-sib mating is ¼.

Figure 5-9. The percentage inbreeding depression at different ages, 2-8 years, of Pinus radiata pine studied in New Zealand. The field test was carried out at a locality with good site conditions.
and the degree of inbreeding has been derived. To enable an understanding of this equation the concept, inbreeding coefficient, must be clarified. In trees with both female and male flowers a high degree of inbreeding can be obtained via repeated selfings. A prerequisite is of course that there are no prevention of fertilization with the pollen of the same tree. Fullsib, halfsib, and first cousin matings are other types of inbreeding with decreasing degree of relatedness in that order. In quantitative genetics, the inbreeding coefficient, \( F \), is an estimate of identity by descent of alleles. Identity by descent means that copies of one and the same allele at an ancestor have been brought together in an offspring. It is important to observe that it is not enough with homozygosity but the alleles at a homozygote must originate from one common allele at an ancestor. This is further explained in Box 5-2. It is certainly true that the degree of homozygosity also increases following inbreeding. The inbreeding coefficients for various types of inbreeding are evident from Fig. 5-10 and Picture 5-2. The relationship between the magnitude of the inbreeding depression and the inbreeding coefficient is evident from the formula:

\[
m_F = m_0 - 2F \sum dpq
\]

in which \( m_F \) designates the value of a trait such as tree height or stem volume in a population with the inbreeding coefficient \( F \) while \( m_0 \) is the value for the trait studied before any inbreeding took place, \( p \) and \( q \) represent the average gene frequencies in loci affecting the trait, \( d \) is the dominance deviation (cf Fig. 5-2). The sign \( \sum \) stands for summation of the effects from all loci involved.

From equation 3 we can extract the following information:

1. The equation shows that there is a linear relationship between the size of the inbreeding depression and the inbreeding coefficient, \( F \).
2. If the dominance deviation for all loci is equal to zero, there will be no inbreeding depression. When \( d = 0 \) the gene action is totally additive. This was the assumption we had in the example with tree heights to derive the quantitative inheritance. Since inbreeding depression occurs in most cross breeding organisms one can conclude that \( d \) is different from zero and mostly on the plus side according to Fig. 5-2.
3. The allele frequencies have great impact on the size of the inbreeding depression in agreement with the situation for heritability.

This interpretation is correct if all alleles involved operate in an additive way. For Norway spruce and Scots pine there are data suggesting that this may be the case. In spite of the additive gene action we have, as mentioned before, a large inbreeding depression in these species, which is contrary to the predictions according to equation 3. One possible explanation is that the inbreeding depression depends on vitality-decreasing alleles at very low frequencies. Homozygotes should mainly arise in such cases after crosses among related individuals.

Heterosis is the opposite to inbreeding depression and is thus an increase of vitality after mating between inbred individuals. The best known example is hybrid breeding in maize, in which species several generations of inbreeding were carried out before mating took place between individuals from different inbred lines. Through this method one has achieved very spectacular results but the value of the technique has been challenged in recent decades. Also in this case there is a quantitative genetic equation that describes the heterosis, \( H_{F1} \), that we expect in the first filial generation after mating between two inbred lines:

\[
H_{F1} = \sum d(p_1 - p_2)^2
\]

in which \( p_1 \) is the frequency of one allele at one of the inbred lines and \( p_2 \) is the frequency of the same allele at another inbred line. Summation of the effects in loci affecting the trait must take place in this case too. An analysis of the equation reveals the larger the difference in gene frequencies between lines, the larger \( H_{F1} \) will be. The largest effect is obtained when the allele frequency is 0 in one line and 1 in the other line, i.e. one line is homozygous \( a_1a_1 \) and the other homozygous \( a_2a_2 \). Also in this case \( d \) is involved and in analogy with the inbreeding depression there will be no heterosis if \( d \) at all loci involved is 0. Another condition for heterosis is that \( d \) at most loci is positive.

Matings between individuals from different inbred lines immediately restore the vitality lost by inbreeding, which is important for conservation genetics. Parenthetically it might be mentioned that this equation had a large impact on the early breeding of Norway spruce in Norway and Sweden.
Selection differential, selection intensity, and genetic gain

In this section we shall discuss the effects of different strength of artificial selection while the effects of natural selection will be discussed in next chapter.

In Fig. 5-11 the meaning of selection differential (S) is illustrated. The selection differential is equal to the difference between the mean of the selected part of the population and the mean of the total population. As may be seen from this figure the selection differential depends on the distribution of the trait. If the same proportion of individuals is selected, the selection differential is larger if the distribution is larger. To enable a comparison of different cases of selection the selection intensity (i) has been introduced. The selection intensity is obtained by dividing the selection differential by the standard deviation. The selection intensity is non-linearly related to the proportion selected. To increase the selection intensity from 2 to 3 requires a much larger population than the increase it from 1 to 2.

To make it possible to calculate the result of a certain selection for a particular trait it is necessary to know the genetic proportion of variation in this trait. If we aim at a mass selection, i.e. to select several individuals as parents for a new generation, the genetic effects that are added to each other are of importance. Thus, it is the heritability that is of interest and the improvement is equal to the heritability multiplied by the selection differential. The result of this product is usually referred to as genetic gain, \( \Delta G \). As equation we get:

\[
\Delta G = h^2 \times S \tag{5}
\]

or if we use selection intensity instead of selection differential:

\[
\Delta G = h^2 \times i \times \sigma_{ph} \tag{6}
\]

In order to visualise the great impact of the additive variance for making progress by artificial or natural selection the heritability can be expressed as the ratio \( \sigma_a^2 / \sigma_{ph}^2 \). If this ratio is included in equation 6 we get the genetic gain in the following way:

\[
\Delta G = i \times \sigma_a^2 / \sigma_{ph} \tag{7}
\]
in which \( i \) expresses the selection intensity for the selected trees, \( \sigma_i^2 \) is the additive variance, \( \sigma_{ph}^2 \) is the standard deviation in the entire population.

Under assumption of a heritability of 0.4, the genetic gain that is obtained after selection of 1 % or 50 % of the best individuals is illustrated in Box 5-3. The effects of the selection as illustrated in this box are a confirmation of what we said before that the larger the selection differential the larger the gain by selection.

Besides the selection of parents there are other types of selection. We can select all individuals in a progeny, which geneticists frequently refer to as family selection. We can select the best individual in the best family. When destructive tests are used such as at freeze testing of whole plants or inoculation with pathogens we can select siblings to the plants tested, which is designated as sibling selection.

### Genetic correlation

From classical genetics several cases are known in which the same allele influences two different traits. Therefore, it is quite logical that loci affecting quantitative traits may also affect more than one quantitative trait. For tree breeding it is particularly important to be able to disclose how other traits are affected when selecting for one specific trait. The understanding of evolution is also simplified if we know the genetic relationship between different traits.

It would not be surprising if two consecutive stages during bud burst in Norway spruce are affected by the same alleles. On the other hand it is less certain that the point of time for bud burst during spring and growth cessation during late summer or autumn are affected by the same alleles. To disclose whether this is the case genetic correlations are calculated. The genetic correlation is in most introductory texts referred to as a correlation of breeding values of two traits. In advanced texts it is disclosed that genetic correlations are not that simple but it is beyond the scope of this book to go into further detail. In agreement with estimates of breeding values, the genetic correlations are valid for the population tested and the conditions under which it is tested. In the equation for the genetic correlation, the covariance between traits \( x \) and \( y \) is one part; this covariance estimates the covariation between the two traits. The genetic correlation is frequently designated as \( r_s \) and it is equal to:

\[
 r_s = \frac{\text{cov}_{xy}}{(\text{var}_x \times \text{var}_y)^{1/2}}
\]

To enable high precision in the estimates of genetic correlations it is required that the experiments contain progenies of numerous parents, at least 100 being desirable.

Owing to the original design of many forest tree breeding programmes there are frequently no more than 40 parents in each experimental series. This means that the precision in the estimates is not as good as desired.

### Summary

Quantitative traits are affected by a large number of alleles, each with a small effect on the trait. This means that we cannot observe any discrete segregation in the progeny population, rather we frequently note a normal distribution of quantitative traits. The environment influences the traits as well. Finally, the gene action is rarely fully dominant or recessive; instead we have a certain degree of dominance.

Heritability, general combining ability and genotype \( \times \) environment interaction are parameters that are estimated by statistical methods in well designed experiments. All three are important to enable predictions of the effect selection has on a certain trait. Inbreeding depression and heterosis are explained by quantitative genetics equations. The inbreeding depression after selfing is substantial in Norway spruce and Scots pine as well as in many other cross-fertilizing species. With decreasing degree of relatedness of the parents inbreeding depression becomes less pronounced.

The selection differential is the difference between the mean values of the selected part of the population and the whole population. By division of the selection differential by the standard deviation the selection intensity is obtained, which is independent of the distribution of the trait. The genetic gain is equal to the heritability multiplied by the selection differential for the trait under study. The genetic correlation is an estimate of the strength of the relationship between the breeding values of two traits.

### Further reading


Evolution

In this chapter we mainly focus on the principles of evolution and we will present a limited number of empirical data. Population differentiation observed for many traits are shown in the next chapter. First we give a short presentation of natural selection, random genetic drift, mutations, and gene flow. The phenotypic plasticity of a trait and its role in evolution is also discussed. Later on we present in more detail the evolutionary factors. We raise the question whether or not any perfect form could be reached in nature.

Evolution is a continuously ongoing process, in which species arise, flourish, and become extinct. From the fossil record, some scientists have come to the conclusion that 99.9% of all species that once existed have become extinct. The most probable prediction we can reach for the biological world is that all presently existing species will become extinct in the long-run. Certain times during the millions of years in the past were characterized by mass extinction. Besides the extinction that is unavoidable in a world of evolution, man is causing even more extinction by various activities. Above all, the activities that cause unnecessary erosion of species have to be identified. There is no question that the human population explosion is the greatest threat against the majority of the vulnerable species.

Box 6-1 Definitions of adaptation, adaptedness, adaptability, fitness

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptation</td>
<td>= the process that leads to a better adaptedness in a specific environment</td>
</tr>
<tr>
<td>Adaptedness</td>
<td>= is the degree to which an organism is able to live and reproduce in a given set of environments</td>
</tr>
</tbody>
</table>
| Closely related to adaptedness according to the above definition is the fitness concept. Fitness is an expression for an individual’s contribution to the next generation in relation to other individuals in the same population. This type of fitness is sometimes referred to as Darwinian fitness. This latter term is used to distinguish processes in nature from cultivation (see section Darwinian and Domestic fitness in Chapter 7).
| Adaptability| = the ability of a population to respond genetically or phenotypically to changed environmental conditions. The amplitude of a trait of a genotype studied in at least two different environments is called phenotypic plasticity. The term reaction norm is used to describe the trait value change of a genotype studied along an environmental gradient. |
Terminology

The terms related to adaptation are sometimes used with some differences in meaning depending on the author. It is important to define terms that are frequently used in the literature. The definitions used in our text are those defined in Box 6-1.

Factors influencing evolution

Differentiation among populations is a major issue in evolution. We have tried to visualize the factors promoting and constraining differentiation among populations in Fig. 6-1. Natural selection, genetic drift, and mutations promote differentiation among populations. They raise the horizontal line to a higher level if they are in operation, which is equal to a larger differentiation. If gene flow is in operation the horizontal line is pushed downwards. Via natural selection certain individuals contribute more to the next generation than others and in this way cause a change in gene frequencies. Natural selection is regarded by many as weak. However, a careful scrutiny of many scientific papers on natural selection by an American scientist during the mid eighties indicated that natural selection might take any place on a scale from weak to as strong as in plant or animal breeding.

Random genetic drift is a random process that inevitably

Box 6-2 Definition of evolutionary factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutation</td>
<td>= alteration in a gene or a chromosome</td>
</tr>
<tr>
<td>Natural selection</td>
<td>= differential transfer of alleles to next generation resulting in increased fitness</td>
</tr>
<tr>
<td>Random genetic drift</td>
<td>= random loss of alleles in small populations</td>
</tr>
<tr>
<td>Gene flow</td>
<td>= migration to a recipient population from another population with a different allele frequency</td>
</tr>
<tr>
<td>Phenotypic plasticity</td>
<td>= the amplitude of a trait of a genotype studied in at least two different environments</td>
</tr>
</tbody>
</table>

Figure 6-1. Schematic illustration of evolutionary factors promoting population differentiation. The arrows pointing upwards increase the differentiation while gene flow reduces the differentiation between populations. Mutations have a limited impact on differentiation.

Figure 6-2. Remaining fraction of additive variance after 10 generations as a consequence of genetic drift. It is assumed that the effective population size was constant during these 10 generations.

Figure 6-3. The curve describes the change in phenotypic value of a genotype along an environmental gradient. The curve is the reaction norm within this range of environment of the genotype. The difference between the highest and lowest phenotypic value is the phenotypic plasticity of the genotype.
causes loss of alleles in small populations. This takes place whether or not the adaptedness of the small population is increased. Genetic drift is important in populations of a size less than 20 individuals (Fig. 6-2).

The mutation rate at individual loci is generally low, mainly within the range of one per ten thousand to one per million. For that reason the probability that the same mutation will arise in two populations is low, which explains why mutations are supposed to contribute very slightly to population differentiation. The mutation rate seems to be higher in conifers than in angiosperms.

Matings among individuals from different populations are part of the gene flow which is a strong obstacle to population differentiation. Transport of seeds, fruits, nuts, and acorns is another component of the gene flow which in many cases is believed to play a minor role compared to the gene flow via pollen. All factors treated above cause changes in gene frequencies.

Plants have a great capability to change their exterior shape depending on the growth conditions. A genotype that is tested in two or more environments may have different heights, crown form, density etc in the different environments. The amplitude of such a variation in a trait is a measure of the genotype's phenotypic plasticity (see Figure 6-3). Many textbooks in genetics do not at all treat phenotypic plasticity. Its role in evolution is somewhat ambiguous: On the one hand the phenotypic plasticity can be regarded as a disguise of the genotype which means that natural selection will not be as efficient as it would be without this disguise. On the other hand phenotypic plasticity may contribute to the fitness of a genotype, especially if it is a long-lived species with a wide distribution encompassing many different site conditions. If this is the case natural selection will increase the frequency of genotypes with a large phenotypic plasticity.

So far we have discussed differentiation among populations but the same factors operate within a population as well (see Figure 6-4). From this figure it is seen that mutations and gene flow increase the genetic variation within populations while natural selection, genetic drift and inbreeding reduce the within-population genetic variation. The matings that were realized are designated as the mating pattern. Gene flow, genetic drift, and inbreeding can thus be regarded as components of the mating pattern. A comparison of Figures 6-1 and 6-4 reveals that the same evolutionary factor influences differentiation among populations and genetic variation within populations in different ways. The difference between natural selection at the species level and within an individual population will be discussed when the three types of natural selection are presented below. Observe that recombination does not cause any change of gene frequencies but it creates new genetic combinations in the gametes.

In conclusion it should be emphasized that natural selection is not the only evolutionary factor. As outlined in the previous paragraph the genetic raw material that can be changed by natural selection depends on the matings that took place. In nature, the five factors discussed above interact in a complex way, and we cannot expect that the adaptedness will be perfect. In most cases evolution is a gradual change of the genetic composition of a population.

**Natural selection**

Ever since Charles Darwin presented his theories about evolution, natural selection has attracted great biological interest. It has sometimes been misinterpreted. Some have seen it as a dark force while others have seen it as a creative force driving evolution to a greater perfection. Sometimes it has been regarded as an ethical principle that man should not intervene with. None of these opinions is correct. Natural selection is a process that can be expressed by a statistical measure of the differences among individuals in their capacity to transfer genes to the filial generation. Natural selection is not caused by differences in transferring genes to the next generation, it is differences in transfers of genes to the coming generation. Changes in gene frequencies depend on the existing conditions and have nothing to do with future conditions. Therefore, there is no goal or any predetermined direction of natural selection. It cannot give populations such car-
characteristics that the probability for survival is increased in the future. The great evolutionist Ernst Mayr expressed the essence of natural selection in his book on evolution from 1988 in an excellent way: Selection is not a forward-looking process but simply a name for the survival of those few individuals that have successfully outlasted the "struggle for existence".

The target for selection has attracted interest among evolutionists. In most cases it is the fitness of the entire phenotype that is selected or rejected in natural selection. This means that there may be progress in one fitness-contributing trait and regression in another if these two traits are negatively correlated. The concept survival of the fittest has misguided many laymen to believe that natural selection requires that the phenotypic variation is fitness-dependent. This has strongly been criticised by R. Lewontin, one of the leading geneticist during the second part of the 20th century.

Since fitness is the trait that is selected for there will be no natural selection without genetic differences in fitness. One frequently used example of natural selection is the observed variation in critical night length for budset in Norway spruce seedlings (see Fig. 7-14). It must be assumed that natural selection has favoured genotypes with a date of budset that matches the date for autumn frost appearance at various localities. To give one example in which natural selection is not working we shall present a hypothetical case. Let us assume that we have a large forest population of one single clone growing under very heterogeneous site conditions. As a consequence of this heterogeneity, only some of the clonal members will produce offspring. This is a form of selection but it is not natural selection since there will be no change in allele frequency from the parent to the progeny generation, which is the case for natural selection. In summary, natural selection requires that the phenotypic variation is genetically regulated and that the variation leads to differences in fitness.

We may ask why certain individuals in nature have bet-
Figure 6-6. In many populations in nature there is stabilising selection. When a series of populations are growing along an environmental gradient, their phenotypic values will differ. This will be perceived as disruptive selection between populations.

index 2-alleles such as 14:12, 12:14, 15:11, 11:15 etc. Any scientific proof that the selection gives rise to such a situation does not exist. The above described form of increased homozygosity offers an explanation as to why additive variance remains even under stabilizing selection.

Under the harsh conditions that prevailed after the withdrawal of the ice after the last glaciation it is easy to perceive that plants in the "harsh tail" of the normal distribution had the highest fitness. This will favour a directional selection. Directional selection is probably of great significance in populations migrating along an ecological gradient. If it was the tail with the harshest individuals that was favoured after the glaciation perhaps the other tail has the highest fitness today when we are probably in a period of temperature increase owing to increase of greenhouse gases in the atmosphere.

We can look upon directional selection in an analogous way to what we did for stabilizing selection with alleles designated with index 1 and 2. We may assume that individuals in one of the tails mainly have index 1-alleles while mainly index 2-alleles dominate in the other tail of the normal distribution. Depending on the direction of selection the progeny will face, there will be an increase in the frequency of one of the alleles. If alleles with index 1 contribute to fitness under harsh conditions, an increased frequency of index 2-alleles is expected under global warming.

Figure 6-7. The change of allele frequency at a locus in a large population under the assumption that the difference in fitness between the two homozygotes amounts to 3% and that the initial allele frequency is 0.05.

If we assume that the individuals in the two tails have the highest fitness we shall observe disruptive selection. In northern Scandinavia with much snow during the long winters there may be an advantage for a tree either to have a narrow crown so that the snow glides down from the tree, or to have extremely strong branches. The narrow crown might be the result of natural selection. When branches are broken by heavy snow the tree crown will be reduced, which probably leads to reduced photosynthesis and reduced growth. This in turn means fewer flowers and a lower possibility that the genes of such a tree are transferred to the next generation. On the other hand if the branches are very strong they may carry the large amount of snow coming during the winter. This means that trees in the opposite end of the distribution are also equipped with high fitness and we have a situation that might provoke disruptive selection.

It is worth emphasising that stabilizing and directional natural selection cause a reduction of within population genetic variation while the variation among populations becomes larger. It is thus important to distinguish between the selection within an individual population from the selection that takes place at the species level. This is evident from Fig. 6-6 in which we have stabilizing selection in the four populations that are growing along an ecological gradient. When we introduce the phenotypic values we observe that the stabilizing selection within populations becomes disruptive among populations.

To understand the speed of change caused by directional selection we have to introduce an expression for how
frequency of one third of the

tion. The maximum speed of change occurs at an allele

sites for changes of allele frequencies via natural selec-

gives an interesting piece of information on the prerequi-

This formula, which is valid under complete dominance,
gives an interesting piece of information on the prerequi-
sites for changes of allele frequencies via natural selec-
tion. The maximum speed of change occurs at an allele
frequency of one third of the a allele. The equation also
shows that the speed of change is largest when both al-

\[ \Delta q = -spq^2/(1 - sq^2) \]

This formula, which is valid under complete dominance,
gives an interesting piece of information on the prerequi-
sites for changes of allele frequencies via natural selec-

In Figure 6-7 the effects of selection on a dominant and a
recessive allele are graphically shown. The curves are va-

Moreover, we have assumed that the difference between
the homozygotes, AA and aa, are 3 %, i.e. s is 0.03. As
expected, the largest change is observed for the dominant
allele during the first generations while the increase of the
recessive allele is very slow in the first hundreds of
generations. This is understandable since to begin with
the a allele is found only in heterozygotes which are con-

In Figure 6-7 the effects of selection on a dominant and a
recessive allele are graphically shown. The curves are va-

It might seem as if there is no possibility for a new mu-
tant to become established in a population. If the number
of individuals in the population is low there is a higher
chance for the new allele to become fixed (= homozy-
gous) than is the case in Figure 6-7. Thus, fragmentation
per se is no obstacle for evolution rather it might speed
up evolution. Some evolutionary geneticists have the op-
inion that limited population size is the explanation for
much of the speciation that occurs. The positive effects
of the integration of fitness-promoting alleles in a popula-
tion must be weighed against the disadvantages of small
populations. As will be shown below, genetic drift may
give rise to loss of fitness-promoting alleles as well as
to increased inbreeding with accompanying inbreeding

Natural selection under severe stress conditions

Studies of plants under severe stress have shown that na-
tural selection can change allele frequencies dramatically.
Investigations of heavy metal tolerance in grass species
growing on mining wastes show that natural selection
must play a major role for the frequency of tolerant plants
in adjacent populations growing on non-contaminated
soil (Figure 6-8). Since these species are wind pollinated
there is probably a strong gene flow between the two
populations. In spite of this the great difference between
the two populations remains.

Another example of rapid change concerns pesticide
resistance in insects that was built up after introduction
of pesticides on a large scale. In certain cases it did not
take more than 5 generations to achieve resistance which
might seem extremely few based on the curves presented
in Figure 6-7. This rapid building up of resistance must
be attributed to an extremely strong selection for the al-

It is probable that there is a high cost for keeping various
types of strong stress tolerance. This means that indivi-
duals equipped with strong stress tolerance are not very
competitive under more normal conditions (cf Figure
6-8); they have lower fitness under ordinary conditions.
Therefore, it is unlikely to find genotypes that have a high
fitness over the broad span of environmental conditions
that a species occupies. For gene conservation it is im-
portant that we capture genes of importance for growth
under stress conditions in the gene resource population.
As presented in chapter 11 this is most easily done by
Random genetic drift

In populations with a small effective population size ($N_e$), genetic drift is of more importance than the other evolutionary factors. By small $N_e$ is meant that there are few trees contributing to the progeny in the next generation (cf Chapter 4). A population may contain many trees but for various reasons only a few of them flower, which makes the population small as regards progeny production.

Numerous simulations have shown that genetic drift leads to allele fixation if a population remains small over a large number of generations. This means that the population either becomes homozygous $a_1a_1$ or $a_2a_2$. Such an allele fixation takes place independently of any evolutionary advantage of the homozygote. It is worth mentioning that allele fixation also takes place for the allele of the lower frequency of the two. If the allele frequencies are 0.9 for $a_1$ and 0.1 for $a_2$, at a locus and the population consists of a large number of small populations, after a large number of generations there will be 90% of the populations homozygous $a_1a_1$ and 10% of the populations homozygous $a_2a_2$. It is important to note that genetic drift takes place without an increase of fitness. Genetic drift does not rule out natural selection but by decreasing population size the impact of natural selection drops. In gene conservation and breeding we are interested in limiting the impact of genetic drift. It is therefore important to keep the population size large enough to prevent any major role of genetic drift.

As is illustrated in Figure 6-2 the remaining additive variance is dramatically reduced at effective population sizes lower than 20. The loss of additive variance per generation is $1/2N_e$. This means that the loss of additive variance with an effective population size of 50 amounts to 1% while an effective population size of 10 causes a loss of 5% per generation.

For wind-pollinated trees with wide and continuous distributions it is expected that genetic drift plays a minor role. The only exception may be populations at the margin of their distribution area. Strong gene flow might compensate to some extent for a low number of flowering trees. Perhaps genetic drift has played a role for speciation in the tropics where many tree species are represented by one or two adult trees per hectare.

Mutations

It has long been assumed that mutations are randomly distributed over the genome. This assumption has been challenged. In some grass species growing adjacent to mining wastes, heavy metal tolerant mutants have been found in some species but not in others. Thus, it seems as there are restrictions in the genome such that heavy metal tolerance mutants can only be induced in some species.

Estimates of the mutation rate per locus and per generation in higher organisms are mainly in the range of one per ten thousand to one per million. Even if mutations are prerequisites for evolution it is important that the hereditary material is resistant to change. A highly conservative characteristic of DNA is important for avoiding chaotic conditions. As regards mutations at loci regulating quantitative traits the knowledge is for obvious reasons scanty. There are estimates of the pooled mutation rate at all loci involved in the regulation of a quantitative trait. Such estimates are in the range of one per one hundred to one per thousand per generation. Since there are many or even numerous loci involved in the regulation of such a trait it is reasonable that the pooled mutation rate is higher than in individual loci but that it differs as much as 10-100 times is somewhat surprising.

As regards the influence on population differentiation mutations play a minor role owing to the low mutation rate at individual loci. Thus the effectiveness of mutations in promoting population differentiation is in most cases several times weaker than natural selection in large populations. Similarly it is several times weaker than genetic drift in small populations.

Gene flow

The meaning of this term is that individuals from one population participate in the procreation of a new generation in the recipient population and that the donor and recipient populations have different allele frequencies. For plants, which are mostly stationary there are gene flows via pollen, seed or fruit dispersal. Different species vary considerably with respect to distance of dispersal. Studies have shown that pollen grains of the wind pollinated tree species such as spruces and pines may spread their pollen hundreds of kilometers. Such figures do not tell us how important long-distance transport of pollen is for fertilization in a population. They only tell us that there is a potential for long-distance transfers of alleles which might be lacking in other species with more stationary pollen vectors.

The effect of gene flow in a recipient population is visualised in Figure 6-9. As is evident from this illustration, large differences between donor and recipient populations cause a large change in the recipient population. Slighter differences, as well as a low fraction of immigrants, also
After immigration

$M = 0.2$

$q_1 = q_0 - M(q_0 - q_i)$

$q_1 = \text{allele frequency after immigration} \quad q_0 = \text{allele frequency in recipient population} \quad M = \text{fraction of immigrants} \quad q_i = \text{allele frequency in the contributing population}$

$\text{Increase of q} = 0.06$

**Figure 6-9. Illustration on how gene flow (=immigration) from one population influences the allele frequency in the progeny of a recipient population. The increase of the q allele frequency is indicated.**

lead to deviations from what is expected according to the Hardy-Weinberg law. It is worth mentioning that gene flow in most cases has a greater impact on the population structure than mutations have. Exchange of one single gamete between two populations prevents a fixation of neutral alleles in the recipient population. It is of great interest to estimate the magnitude of the gene flow among populations.

In a study of 66 populations belonging to 3 subspecies of lodgepole pine, it was shown that the number of migrants between populations was larger than 1. This means that the possibilities for differentiation among these populations are small. Another way of describing this phenomenon is to estimate the effective population size. For *Picea abies* and *Pinus sylvestris* the estimates vary between 8,000 and 16,000. These figures show that these species are efficient dispersers of their genes over large areas. Detailed studies of the pollination pattern were carried out for *Quercus petraea*, *Q. robur*, *Tilia cordata* and *Castanea sativa*. In the two oak species the gene flow into the studied French populations was considerable, above 60% while the selfing did not exceed 2% (Fig. 6-10). Such a large gene flow is a serious constraint to adaptation to the

**Figure 6-10. The percentage of matings within French populations of Quercus petraea and Q. robur. The percentage of pollination by pollen from other populations is also indicated.**
prevailing site condition at the growth localities. Contrary to the situation for oak, the selfing was high in *Tilia cordata*, amounting to 25% in the stand studied and to 65% in four isolated trees outside this stand (Fig. 6-11). The average pollination distance was estimated at 150 meters while the corresponding estimate for maximum pollination distance was 1666 meters. These figures for *T. cordata*, which is an insect-pollinated species, were slightly lower than the results from *Castanea sativa*, which is a wind-pollinated species. Another observation in the *Tilia* study was that the pollinating insects flew preferentially to large trees and stayed longer in them. In *Castanea sativa* the number of migrants exceeded 1 in naturalized and coppice populations from Greece and Italy while the fruit orchard populations had low estimates (Fig. 6-12).

**Figure 6-11.** The percentage of selfings in a German population of *Tilia cordata* and in four isolated trees outside the main population.

Data on outcrossing, defined as matings between unrelated individuals, do not give us information about gene flow but such data inform us about the occurrence of matings among related individuals or selfing in populations. Data on outcrossing can be related to life-history traits of various species and therefore give us an understanding of mating pattern in species with different characteristics. In Figure 6-13 data for several temperate forest tree species are presented based on a compilation by David Boshier. As seen from this figure outcrossings occur to more than 70% in all species in this figure. In conclusion outcrossing data suggest that most males are unrelated and grow in the vicinity of the seed tree studied, but long-distance pollinations occur. The number of pollen donors varies

**Figure 6-12.** Number of migrants per generation in three types of Greek and Italian *Castanea sativa* populations, naturalised, orchard, and coppice, and in natural populations of *Acer platanoides* and *Betula pendula*.

**Figure 6-13.** Percentage outcrossing rates estimated by several isozymes for *Abies lasiocarpa*, *Juglans nigra*, *Liriodendron tulipifera*, *Picea Engelmannii*, *Pinus attenuata*, *Pinus banksiana*, *Pinus contorta*, *Pinus ponderosa*, *Pinus sylvestris*, *Pseudotsuga menziesii*, and *Quercus robur*.
considerably. Obviously successful mating requires that there is overlap between female receptivity and pollen dispersal.

Especially in southern Sweden a large number of new stands of Norway spruce have been established with eastern European *Picea abies*. Some people have regarded the pollination of the domestic *Picea abies* from the eastern sources as genetic pollution. For traits of high adaptive value the progeny will with high probability be intermediate to the exotic and domestic *Picea abies*. This in turn must be attributed to changes in allele frequencies. The issue will be elaborated somewhat more under Genetic pollution in Chapter 11.

The evolutionary importance of gene flow and phenotypic plasticity is further discussed in the section Ecotypes and ecoclines later on in this chapter.

## Phenotypic plasticity

Especially plants have a great ability to develop in different ways depending on the environmental conditions. It has become more and more clear that this plasticity in traits such as plant size, flower number etc can be genetically regulated. The morphology of flowers seems to have less plasticity. If such traits were very plastic it would threaten the reproduction of the species. Different species probably have different phenotypic plasticity according to their ecological characteristics.

Our knowledge about variation in phenotypic plasticity within populations is still limited. From theoretical points of view we expect that it is large in species such as Norway spruce, Scots pine and lodgepole pine. It may be less in insect pollinated species with a scattered distribution and limited gene flow among populations (See also the section Ecotype and ecocline below).

### Will the adaptedness ever be perfect?

Many laymen seem to believe that the adaptedness is perfect at a particular site if the population has grown there for many generations so that natural selection chisels out something perfect for just this site condition. Even among geneticists adaptation lag is discussed meaning that if natural selection was allowed to proceed over enough number of generations we should eventually observe a perfect adaptedness. Besides, one tries to find adaptive advantage in all traits an individual carries. If a perfect adaptedness should exist it is required that the following prequisites are fulfilled:

* the environment is constant
* all traits in the population are totally independent of each other

That the environment is not constant over time is so evident that there is no need to elaborate much on this issue. Let us just remember that we who live in the far north have different seasons during a year and that conditions during two summers are never the same. Someone might believe that the year to year variations are so slight that they would not make any difference to natural selection. However, we do not know whether subtle differences would be evolutionarily very important. In Sweden the winters during the last two decades of the 20th century are a good illustration of the annual weather variability. The last cold winter in southern Sweden occurred in 1987. After that, most winters have been milder than normal. These conditions have probably had consequences for survival and growth of Norway spruce plants. The mild winters have resulted in melting of the snow and plants were sometimes exposed to high day temperatures while the ground was still frozen. This leads to frost desiccation with death or retarded growth as a consequence. As regards variable weather we must assume that natural selection can change gene frequencies in different directions under different ambient conditions.

The relationship between two traits may be positive, negative, or absent. If it is negative it means that progress in one trait leads to regression in the second trait as is illustrated in Figure 6-14. Since natural selection operates on the individual as an entity it means that there may be regression in a trait of adaptive value if it is negatively correlated with another trait of still higher adaptive value. It is the phenotype of an individual as an entity that is the selection target in most cases. Therefore, traits cannot be disentangled from each other and as corollary of this; natural selection may lead to good adaptedness in certain traits but with low adaptedness in others. Already Darwin stated that ”Natural selection will not produce absolute perfection.”
Besides the two prerequisites raised above there are other conditions that make it unlikely to find perfect adaptedness. In small populations there is a chance-conditioned loss of alleles owing to genetic drift. Once more it is important to stress that it is the number of the trees that contribute to the formation of the next generation that counts. This number can be considerably less than the trees of a species growing in a forest.

Wind pollination, which is important for avoidance of genetic drift, may slow down the adaptation. This is the case if the pollen emanates from other populations with a certain degree of adaptedness to other site conditions than in the recipient population. The impact of gene flow on among- and within-population variation is further discussed in connection with Fig. 6-15. Similarly the role of gene flow for conferring fitness to phenotypic plasticity is discussed with starting point in this figure.

As mentioned in the chapter on quantitative traits it is expected that several different genotypes can give rise to one phenotype. A simple example might be used to exemplify that. Let us assume that alleles with index 1 contribute equally to the adaptedness and differently from alleles with index 2, which in turn contribute equally much. Under these conditions the genotypes \( a_1a_1b_1b_1c_1c_1 \) and \( a_2a_2b_2b_2c_2c_2 \) would have the same adaptedness, which is also valid for all other genotypes with 4 index-1 and 2 index-2 alleles. The phenotype with highest fitness may differ genetically and therefore natural selection will not favour just one genotype. The population which is growing in nature as a consequence of natural selection, with all its limitations must therefore be regarded as one solution of a great number of possible solutions.

It might be a little trivial to remark that most of the newly arisen mutations reduce the vitality and thereby the adaptedness of its carrier.

The conclusion from the discussion above is that we can never regard the present genetic constitution of a population as perfect or sacred, rather it must be regarded as transient and one of several possible. Therefore, the present genetic constitution should not be targeted in dynamic gene conservation (cf. Chapter 11).

Ecotype and ecocline

Göte Turesson was probably the first to discuss genetic adaptation to different site conditions in the early 1920s. He introduced the term ecotype for this type of adaptation. He mainly studied perennials growing under quite distinct site conditions, such as rocky localities as contrasted to beach meadows. After cultivation at other site conditions the "rocky" and the "meadow" ecotypes kept their morphology, proving that their characteristics were genetically conditioned.

As will be shown in Chapter 7, growth rhythm, budburst during spring and inwintering at the end of the growth period show a continuous variation in Norway spruce and Scots pine. Such a variation is designated as clinal and instead of ecotypes we have ecoclines in Norway spruce and Scots pine.

For most tree species from the northern hemisphere that have been studied, the night length is the primary trigger for onset of growth cessation. If the ambient conditions are harsh e.g. owing to drought this might also induce growth cessation for obvious reasons. The regulation by the night length of growth cessation means that a population transferred northwards gets a longer growth period than at its original site. Similarly, transfers southwards reduce the duration of the growth period since the critical night length for growth cessation occurs earlier than at the original site.

Are there any proven cases of specific adaptation to edaphic conditions within tree species? There are a few reports suggesting this. However, later reports describing the same materials have disclosed that there were no longer any indications of specific adaptation to edaphic conditions. This does not exclude that individual genotypes differ in their ability to take up or utilize nutrients.

Detailed studies of Scots pine seedlings cultivated at different availability of nitrogen, which is a limiting nutrient element for pine growth in Sweden, resulted in some genotype x nitrogen treatment interaction but the interaction was not larger than the variation among families in the experiment. Is it possible to understand such a result evolutionarily? In southern Sweden Scots pine grows at various site conditions, which ought to give rise to specific adaptedness. However, the sites occur in mosaics and there is a large gene flow between trees growing at the different site conditions. As might be remembered from the section on gene flow, it is a strong factor tending to eliminate population differentiation. Thus, to allow a specific adaptation to take place there must not be any gene flow among the different types of site. It might even be an evolutionary advantage to develop genotypes that give rise to progeny that grow well over a broad span of site conditions. This means that phenotypic plasticity will contribute to fitness.
The above example illustrates well that the mating pattern is of utmost importance for the genetic structure. The American philosopher Robert Brandon, who has devoted much of his research to adaptation, has introduced the concept of selective environmental neighborhoods (SEN). Within such an area there is no genotype x environment interaction as regards fitness which means that there is a large environmental homogeneity within an SEN. In Figures 6-15 and 6-16 two contrasting situations as regards gene flow among different SENs are illustrated. In the first there is a gene flow among all SENs, in the second there is no gene flow between the two SENs. When there is no gene flow between the two SENs there are good opportunities for specific adaptation to the site conditions in each SEN. The examples in figures 6-15 and 6-16 were selected consciously to illustrate a situation that is typical for northern conifers and broadleaved trees, respectively.

The broadleaved tree species consists of isolated populations and is pollinated by insects which are flying over short distances only, the schematic picture becomes close to reality. This type of tree has higher probability for specific adaptation than tree species which do not share these characteristics.

We have tried to illustrate schematically a situation that is typical for a species with wide and continuous distribution in Figure 6-17. In such a case the environment changes gradually, e.g. there is frequently a gradual change from south to north with respect to climate. There will be no sharp boundary between SENs and pollination and seed transfer may take place between adjacent SENs. If the environmental conditions are fairly stable, natural selection will improve the adaptedness along this gradient but gene flow will slow down this adaptation. However, such a gene flow may be useful under rapid global change as will be discussed in the next section of this chapter.

Many tropical tree species are represented by one or a few trees per hectare. Huge areas may constitute one selective environmental neighborhood in wet tropical forests. In such a case the zone for shared pollination may be much smaller than a SEN (Figure 6-18). The situation for many tropical tree species constitutes a great contrast to the situation for a species with continuous distribution as depicted in Figure 6-17.

In ecological texts the niche concept is frequently used to describe site conditions. The advantage with the selective environmental neighborhood concept is that it is not bound to one specific geographic area but it may vary dependent on which trait is under consideration. Thus, for a strictly neutral trait in a species there is just one SEN.
whereas there may be many SENs for adaptive traits.

It has been hypothesized that life history-trait such as type of distribution, pollen and seed dispersal, and stage in ecosystem may influence the variation within and between populations. In Fig. 6-19 we have summarized the life-history traits that promote the ratio (among-population differentiation)/(within-population variation), and its inverse. In a wind-pollinated species with a wide and continuous distribution gene flow may be considerable, which means a leveling of allele frequencies between populations. This may be strengthened if the species is one of the climax species in the ecosystem under consideration. Contrary to this, a species with scattered distribution and with limited dispersal of pollen and seed there is room for a larger population differentiation than in tree species with the life-history trait combinations shown to the left in Fig. 6-19. Some studies give support to this but there are data from the latest decade (1995-2005), which indicate that species sharing the life-history traits to the right in Fig. 6-19 have ecoclinal rather than ecotypic variation. It must be assumed that these species have passed the threshold, which causes ecoclinal variation, even if their gene flow is lower than in wind-pollinated species. Pioneer species such as weed species, which occupy various types of bare ground may benefit from great uniformity to effectively utilize the open land. Asexual propagation, such as in Taraxacum vulgare, may be advantageous and this species consists of clones. Generally there is an inverse relationship between adaptedness and adaptability. Adaptedness may reach a high level by eliminating what is referred to as genetic load resulting in high genetic uniformity. Such a uniformity means that the additive variance and thereby the adaptability goes down. Thus, high adaptedness may be very useful under constant environmental conditions but may be disastrous under rapid change of the environment.

Some proponents of the ecotype concept have claimed that what we observe as continuous variation is actually a stepwise variation which should be designated as ecotypic. The prerequisite for us to detect stepwise variation is that there is no gene flow among populations which are growing under different site conditions. The pattern of pollination and seed dispersal are decisive as to whether there will be an ecotypic or an ecoclinal variation along an environmental gradient. It is highly unlikely to find ecotypes in wind pollinated species which have a broad and continuous distribution. If plants are exposed to extreme stress, like the grass species growing on mining wastes, there can be ecotypic variation as was earlier shown in this chapter. In summary, the lengthy and animated controversy among scientists whether forest trees really show continuous variation or not, is best resolved by trying to identify the evolutionary factors of significance for each species separately.

Figure 6-18. Schematic illustration of a common situation for many tree species from the wet tropical forests with one or a few tree species per hectare. It is assumed that the environmental conditions are fairly uniform over a huge area, which thus constitutes one selective environmental neighborhood.

Figure 6-19. Schematic illustration of combinations of life-history traits promoting population differentiation and within-population variation, respectively.
Evolution and global warming

As components of ecosystems, trees and plants are continuously exposed to environmental changes. Under global warming the speed of change might be faster than before. The changes connected with a greenhouse effect are more a question of degree than of new types of genetic processes differing from those occurring under “normal” changes in the environment.

It is evident that long-lived tree species under global change will be exposed to a gradual change of weather conditions during their life times. To endure such a change trees have to be equipped with large phenotypic plasticity, which thus is of great importance for trees in the forests today. However, the phenotypic plasticity is simultaneously a constraint if it allows continuous existence of an already existing population, which prevents establishment of a new population with better adaptedness to the changed conditions (cf Fig 6-20). For long-term success trees must rely on two other options. The first is dispersal ability and the second is the ability to respond genetically, i.e. that there is ample additive variance for traits of adaptive significance. If either of them is large enough to cope with the changes in the environment the species will survive. A continuously distributed species with long-distance pollen transfer may benefit from this gene flow in contrast to a species with scattered distribution and no or limited gene flow (cf Fig 6-21). However, for the population from the warmest location there is no pollen donor to benefit from.

Phenology of growth and flowering are critical for survival and good growth of many forest tree species from the temperate part of the world. We will discuss growth phenology for Picea abies, which is one of the most important forest tree in Scandinavia but first we will give some comments on flowering phenology.

Flowering intensity and flowering time are of great significance in any sexually reproducing species. It has been proven that flowering initiation is dependent on high temperatures in several tree species, at least at high latitudes. It is likely that flowering will take place earlier during the season in case of global warming since flowering like many other phenology traits is triggered by the heat sum. The expectation following a temperature increase can be phrased in the following way: With a prediction of more temperature extremes, this early flowering may lead to exposure to low and damaging temperatures. In consequence severe frost damage may occur owing to early flowering followed by a frost spell.

Most forest tree geneticists agree that growth cessation is triggered by night length and that budburst is dependent on the heat sum. The northern populations require short-
Table 6-1. A summary of possible effects of global warming on phenology in Picea abies.

<table>
<thead>
<tr>
<th>Phenological event</th>
<th>Occurs earlier</th>
<th>Occurs later</th>
</tr>
</thead>
</table>
| Triggering of growth cessation         | No difference since night length, which is the triggering factor, is not influenced by climate change.  
Sustained drought may provoke earlier growth cessation. |                                                                              |
| Reaching of dormancy                   | If the low temperatures are present and the high temperatures speeds up the development. | If the low temperatures occur later than under present conditions.          |
| Breaking of dormancy and start of growth activities | If temperatures low enough occur and high temperatures occur during winter. | If dormancy is built up later and low temperatures are less frequent than under present conditions. |

Phenological gardens, *i.e.* plots located in different climatic zones with the same genetic material, are useful for estimations of effects of global warming on phenology traits. Based on data from phenological gardens distributed over Europe it was estimated that budburst in *Prunus avium* would take place 5 days earlier per degree of temperature increase. The corresponding figures for *Tilia cordata* and *Sorbus aucuparia* were estimated at 2-3 days. The prediction that leaf fall will not be changed by increased temperature was confirmed for *Tilia cordata*. Strong temperature dependence for budburst in some *Fagus sylvatica* populations was also reported. This means that budburst will take place earlier in case of global warming. Many models have been put forward to predict effects of global warming on phenology traits. In one paper possible outcomes of models that try to predict effects of global warming were formulated in the following way: *Both models and experiments show that the response of phenology to climate change, and in particular to global warming, will depend on the species, the latitude at which the populations are observed and the intensity of changes.* It seems, as the effects on phenology will be more pronounced at higher latitudes after a temperature increase than in the Mediterranean region. In analogy with this the effect in the latter region will be largest at high elevation.

One possible consequence of global warming is fragmentation of a continuously distributed species. This could lead to lower effective population sizes with increased importance of genetic drift in the scattered and sometimes small populations. This means that the mating pattern may be changed. Mating pattern is defined as the matings that are realized, *i.e.* the zygotes formed in a population. Fragmentation of a previously continuously distributed species may have consequences for its mating pattern. One leading scientist has stated that fragmentation might be of importance only if the fragmentation results in populations with effective population sizes less than 100. However, any predictions are hard to put forward owing to lack of empirical data. Generally, the effect is dependent on the gene flow before fragmentation, the pattern of migration between separated populations after fragmentation, as well as local recolonisation and extinction.
This will be discussed by the aid of Fig. 6-22. In the centre of this figure the gene flow between populations is illustrated by arrows of different thickness, the thicker the arrow the larger the gene flow. There is no direct gene flow between the two most distant populations. However, there is a possibility for a stepwise gene flow between these populations via the central populations. If the two central populations become extinct there may be no gene flow between the most distant populations. Intuitively, the change in mating pattern depicted to the right in Fig. 6-22 is expected to result in an increased differentiation between populations thanks to adaptation in absence of gene flow by pollen from other populations. If this break of gene flow is associated with reduced population size, genetic drift may also contribute to increased population differentiation. However, there are results pointing at a reduced differentiation between distant populations (Fig. 6-22, left part). If fragmentation occurs in an insect pollinated species, loss of central populations may force insects to fly further than before to find food, which will result in gene flow between earlier isolated populations with reduced population differentiation as a result. Therefore, it is unlikely that a general prediction for the outcome of fragmentation can be put forward.

Climate change may also have consequences at the species level. Related allopatric (see below) species may after climate change migrate in such a way that they will occupy the same habitat. If they have no means of isolation except for the previous geographic isolation, interspecific hybridisation may occur. There are supposed examples for this in the *Abies* and *Pinus* genera.

From the study of fossil records many ecologists have come to the conclusion that most species will not be able to migrate fast enough to cope with the speed of change caused by global warming. If this conclusion is true, species have to rely on the genetic ability to respond to the changes caused by global warming.

In conclusion for long-term survival of a species under global climatic change one of the two following conditions must be fulfilled:

* the dispersal ability is greater than the speed of environmental change
* the genetic response is greater than the speed of environmental change

It should be noted that these conditions apply irrespective of the duration of the environmental change. However, it must be remembered that a tree species with a generation time of 25 years needs a much larger amount of additive variance than an annual species that can respond 25 times during this period. Species with exclusively asexual reproduction have to rely on dispersal ability to cope with global change.

**Coevolution**

Coevolution is usually defined as: Mutual evolutionary changes in two interacting species as a response to changes in these species. The typical example is that a host plant builds up a defense mechanism against one of its herbivorous species. This is then followed by development of a mechanism in the herbivore to overcome the defense system of the host plant. In turn the host plant develops a new defense mechanism which again is overcome by the herbivore. Coevolution has been regarded as an arms race. In the breeding of agricultural crops such a kind of arms race is quite common. The resistance against a harmful organism is followed by a change in the pest or disease organism to overcome the defense. This has lead to a constant search for new resistance alleles since one crop variety after the other has lost its resistance. There has been a constant struggle to be ahead of the pest or disease organism. Does such an arms race occur in nature?
Many investigations of herbivores and their host plant have been carried out. Some of the results are presented below.

Coevolution seems to be an exception among host plants and herbivorous insects. Many plant species have secondary metabolites which are sometimes toxic and may slow down the digestion in herbivores. Especially among herbivores of plants belonging to families Cruciferae and Umbelliferae it is common that the toxic substances are signals for recognition such that the insects are enticed to visit these plants. One such example is given in Box 6-3 for small ermine moths belonging to genus Yponomeuta. Within this genus there are three closely related species which feed on different tree species. As is evident from the crosses between Yponomeuta cagnagellus and Yponomeuta malinellus it seems as if one single dominant allele can break the defence of a host tree. A species lacking this allele cannot break the defence.

As a rule it does not seem as if the host plants have developed their defence mechanisms against their own herbivore. Both among host plants and herbivores it seems that the defense mechanisms and the means to overcome the defense are more general than a strict coevolution requires. Thus it is likely that vascular plants early during their evolution produced secondary metabolites that raised the fitness of its carriers. Long-lived tree species, such as the conifers, rarely have specific toxic substances but they have secondary metabolites which slow down the digestion in the herbivores.

Specialization of the herbivorous insect on certain host plants might suggest that coevolution would be beneficial. However, there are other reasons why a specialization might be advantageous. Certain of the toxic metabolites, which the herbivore gets from feeding from plants with toxins, might protect it from its own parasites or from other animals which have the herbivore as a prey animal.
Speciation

Speciation must be regarded as a logical continuation of population differentiation (See Box 6-4). When the differentiation between populations has gone so far that there is no gene flow between the two populations the main condition for speciation is fulfilled. If no gene flow occurs the populations will become reproductively isolated after some time. In Box 6-4 it is illustrated how one homogeneous population (below) after some time converts into two populations that are differentiated genetically although some gene flow occurs (horizontal arrows). When there is no longer any gene flow between the two populations, two new species have arisen. Each of these two species becomes internally differentiated over time. A little later the population to the furthest right becomes isolated from the rest of its species and a new species appears on the arena. After still further time a new speciation takes place in the right part of the box while the speciation in the left part of the box does not occur until the most recent time. The end result of this hypothetical example is that we now have 5 species from the once homogeneous population. In the left part it is also illustrated that differentiation between populations might cease. A prerequisite is that the gene flow becomes strong enough between the two populations so that the matings can be regarded as totally random.

The major difference between speciation and population differentiation is that the former implies reproductive isolation from other species. The isolation mechanisms can be of various kinds. The requirement for reproductive isolation is easy to understand if we remember what was stated above, viz that it is a strong constraint to differentiation among populations.

Usually three types of speciation are distinguished, allopatric, sympatric, and parapatric. Allopatric means that the speciation takes place in populations which exist in different regions. Sympatric is the opposite, i.e. speciation in a common area. Parapatric means that two different alleles in adjacent populations are favoured. This type of speciation will not be further commented on.

Allopatric and sympatric speciation.

Indications of the occurrence of allopatric speciation are numerous and are built on information about geographic variation. Distant populations often have sterility barriers or differ more in ethological behaviour than adjacent populations. A good example of that is the example of the seagulls referred to in the beginning of this chapter. Often biological differences covary with geographic barriers. This is the case for many fresh water fish species inhabiting different mountain lake systems. One of the most prominent evolutionists, Ernst Mayr, has very stron-
Sewall Wright introduced the concept of adaptive landscapes, which consists of peaks and valleys. The higher the position on a slope a population has reached the better the adaptedness. A population, in which natural selection is the dominating evolutionary factor will figuratively climb higher and higher on the slope.

A condition for the population on the X slope to reach peak Z is that it passes the valley bottom Y. This cannot take place via natural selection, which only results in improved adaptedness.

If the effective population size of the population is reduced such that genetic drift outweighs natural selection the population may theoretically reach valley bottom Y.

If the effective population size of the population increases such that natural selection can operate the population may theoretically climb the Z slope and in this way reach a higher peak.

When the differences have been accumulated between two populations they have also obtained differences in alleles which may not be of immediate importance for survival. Such changes can lead to inferior vitality of hybrids between the two populations. This is referred to as outbreeding depression. Such a hybrid inferiority means that the parental populations are more or less reproductively isolated from each other. Another means to prevent gene flow among populations is that their flowering times do not overlap. At the species level this seems to be the case for European and Siberian larch growing in Sweden. It should be noted that neither of the two species is native in Sweden.

Sympatric speciation is debated. One condition for sympatric speciation is that there is limited or no gene flow between groups of individuals in a population inhabiting a certain area. Such a situation may occur if the two groups of individuals have totally different flowering times. There is one unquestionably important means of sympatric speciation, hybridisation between two species followed by polyploidisation. This kind of speciation is treated in a separate section below.

Speciation per se is seldom adaptive, rather it is a byproduct of adaptation to different site conditions. Speciation, though, is a good starting point for future evolution. Some scientists claim that speciation can be due to single alleles. Above all, this might be the case for alleles influencing floral structure, which in consequence must lead...
to reproductive isolation between individuals with the original floral shape and the mutant form individuals. Speciation via a difference at just one locus is probably rare.

**Speciation by polyploidy.**

Already early in the 20th century it was detected that species belonging to the same genus had multiples of the somatic chromosome number. In one of the first studies the Swede Otto Rosenberg found that Drosera rotundifolia had 20 chromosomes while D. longifolia had 40 chromosomes. Later on it was found that many cultivated plants such as wheat, oats, cotton, banana, sugarcane, coffee, potato and tobacco are polyploid. Already during the second decade of the 20th century the hypothesis on speciation via doubling of the chromosomal number after species hybridisation was presented. The first example of a replication of spontaneous speciation via polyploidy in nature was the creation of an artificial Galeopsis tetrahit after crosses between two other Galeopsis species G. pubescens and G. speciosa by Arne Müntzing during the thirties.

Meiosis is frequently disturbed in species hybrids since there are no homologous chromosomes available for pairing. Sometimes so called restitution nuclei are formed, in which all chromosomes from the two crossing partners are included. These nuclei have twice as many chromosomes as the normal gametes. Even if only the egg cell is diploid, a polyploid (triploid) embryo will be formed. A plant which is developed from such an embryo usually differs much from the parental species. Progeny derived from spontaneous back crosses with the parental species are rare if they occur at all since the hybrid is highly sterile owing to the problems with bivalent formation during meiosis. Therefore, a tetraploid plant that has arisen after a doubling of the chromosome number in a species hybrid will be reproductively isolated from the parental species. This is, as pointed out above a prerequisite for speciation. Species created in this way are usually referred to as allotetraploids. To simplify the description of polyploids we use one letter to designate an entire genome of a polyploid species. An allotetraploid is thus designated as AABB.

Spontaneous doubling of the chromosome number may occur without a preceding species hybridisation. This is designated as autotetraploidy and is written as AAAA. The meiotic division in such a polyploid is disturbed since four chromosomes try to find their homologue for pairing. This leads to variable chromosome numbers in egg and pollen grains.

**The speed of speciation**

The speed of speciation seems to vary much among different groups. The European and American species of Platanus have been separated for numerous generations.

**Summary**

If we look upon adaptation from an analytical perspective we can distinguish two steps. During the first step genetic variation is created and recombination of alleles takes place. This is mainly a random process. Natural selection constitutes the second step, during which the allele frequencies of populations are changed.

It is stressed that natural selection is one of several factors that influence genetic variation within and among populations. Natural selection is a change of gene frequencies and it reduces the genetic variation within populations. There are three types of natural selection. Stabilizing selection means that phenotypes close to the population mean are favoured. In directional selection individuals in one tail of the distribution are favoured. Finally, disruptive selection favours individuals in both tails of the distribution. Stabilizing selection is common within stationary populations. A stabilizing selection within a series of populations growing along an environmental gradient will be experienced as disruptive selection among populations. Natural selection improves the adaptedness but other evolutionary factors participate in the evolution. Therefore, perfect adaptedness will never be observed in nature.

Genetic drift is a random process that leads to allele fixation independent of the fitness contribution of the fixed allele; this reduces the within-population genetic variation. By chance different alleles will be fixed in different populations, contributing to among-population variation. The effect of genetic drift increases exponentially with decreasing effective population size.

Mutations occur at a low frequency and increase the genetic variation within populations. Since the mutation rate per locus and generation is so low, the probability for the same mutation to arise in two populations is infinitesimal. Therefore, mutations will give rise to a small difference
among populations.

**Gene flow** is a strong constraint to among-population differentiation. At the population level it is a strong contributor to increased within-population variation. Data on gene flow and outcrossing suggest that a large gene flow is not restricted to wind-pollinated species with a wide and continuous distribution but also occur in scattered and insect pollinated tree species.

The role of **phenotypic plasticity** is ambiguous. On the one hand it can confer fitness to its carrier and thus is favoured by natural selection. On the other hand it may be regarded as a disguise of the genotype. In this way natural selection becomes less efficient in the presence of pronounced phenotypic plasticity.

The relationship between evolution, natural selection, and genetic drift is illustrated in Figure 6-23. All three require that there is genetic variation available. Differences in fitness are not a prerequisite for evolution but it is facilitated if it exists. Difference in fitness is what characterises natural selection and separates natural selection from the other two. Evolution means that genetic change has taken place.

Depending on the ecological characteristics of a species they show **ecoclinal** or **ecotypic** differentiation. Ecotypic differentiation occurs if gene flow among populations is much restricted. Ecolinal variation occurs in widespread species with a large gene flow among populations. Ecolinal variation means a continuous variation along environmental gradients while ecotypic variation occurs stepwise.

From an evolutionary point of view differentiation of populations and **speciation** are related. The difference is that reproductive isolation exist at the species level whereas some gene flow might occur among populations within a species. Small populations with no or restricted gene flow are a good basis for rapid speciation. Speciation is facilitated by geographical isolation. Doubling of the chromosomal number in species hybrids has been shown to have occurred frequently in plant speciation and is an outstanding example of speciation without geographical isolation.

Evolution in the past has created various patterns of population structure in different tree species. Some of the types of population structure are illustrated schematically in next chapter, Box 7-1. Examples of species having the various patterns are also given. These patterns are discussed in the next chapter while variation within populations is discussed in Chapter 8.

**Further reading**


Genetic variation and provenance research

In Chapter 2 we have presented the origin of genetic variation. In this chapter the main emphasis is on the genetic variation within a species and particularly provenance differences. The important distinction between Darwinian and domestic fitness is also outlined.

Genetic structure and how it is estimated

By genetic structure we mean how alleles and genotypes are distributed among and within populations. The previously described evolutionary processes have contributed in different ways to the present genetic constitution in nature. The history of a population or a species is therefore important for the genetic variation we can observe today.

Some geneticists have the opinion that we should only talk about genetic structure when we have identified the alleles which affect different traits. In this narrow sense the quantitative traits are excluded which are affected by alleles that we cannot identify with certainty. Since most of the traits with high adaptive value are quantitative it would be unfortunate if they were not included in estimates of the genetic structure. Below genetic structure is used to describe genetic variation in both qualitative and quantitative traits.

In Box 7-1 we present a number of possible population structures as well as examples from tree species representative of the various structures. In nature the structure is often not as distinct as illustrated; rather we can observe

<table>
<thead>
<tr>
<th>Box 7-1 Potential population structures and gene flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>One large contiguous population; Example: Pinus resinosa from northeastern North America</td>
</tr>
<tr>
<td>Continent-island, gene flow occurs mainly from the large population to the small island populations; Example: Picea omorica</td>
</tr>
<tr>
<td>Small disjunct populations without any clear gene flow between them; Example: Pinus radiata in Californien</td>
</tr>
<tr>
<td>Stepping stone structure, where gene flow occurs between adjacent populations; Example: Abies fraseri in the Appalachian mountains in eastern USA</td>
</tr>
<tr>
<td>A large continuous population, where geographic distance affects the similarity between populations; Examples: Picea abies and Pinus sylvestris in Europe</td>
</tr>
</tbody>
</table>

In gene conservation it is important to include the entire genetic variation in a species. To enable this we must know the population structure of the species. In reality the population structure is rarely as clear as might be assumed from the population structures visualised.
Restriction Fragment Length Polymorphism (RFLP) is a DNA technique that produced markers. The name derives from the use of so-called restriction enzymes for cleaving DNA (see Chapter 2). This technique is rather laborious and it does not limit the DNA analysis to coding regions only. For Scots pine and Norway spruce it is estimated that the coding part of DNA is only 0.5% of the nuclear DNA. It is likely that the ratio between coding and non-coding regions in other conifers is of a similar magnitude. Theoretically, a large number of fragments might be identified with this technique, as with the RAPD technique. RFLP is a codominant marker, which means that both alleles at a locus can be detected.

Unlike the RFLP technique the RAPD technique (Random Amplified Polymorphic DNA) is faster and it does not require work with radioactive labelling. It is not possible to separate coding from non-coding regions of DNA. The PCR technique is used for amplification of the DNA segments. This technique has come into frequent use in forensic applications. A disadvantage with RAPD is that the segments amplified are dominant. Therefore, it is not possible to discern if there is any difference between two homologous chromosomes as regards a particular segment.

AFLP (Amplified Fragment Length Polymorphism) is a more recent (1995) type of DNA marker where certain DNA segments are amplified by the PCR technique. AFLPs are dominant and identification of coding segments cannot be done. A larger number of polymorphic fragments can be obtained than with RAPD. This means that genetic linkage maps obtained from AFLP are of higher quality than those obtained from RAPD. This is attributed to close location of the AFLP markers which gives a so-called high density map.

Microsatellites (SSRs simple sequence repeats) are regions of DNA containing short segments (2-6/8 base pairs) replicated after each other a variable number of times. Such replications are called tandem repeats. They occur all over the genome, mainly in non-coding regions of DNA. A very large number of variants occur which makes them useful for identification of single individuals. Therefore, they are also very useful for studies of gene flow among populations.

The methods for cleaving DNA from mitochondria and chloroplasts do not differ from those for nuclear DNA. Unlike nuclear DNA, mtDNA and cpDNA are not very polymorphic.

EST (Expressed Sequence Tag) is a partial cDNA sequence, i.e. a sequence within the coding region of a gene. ESTs are used for recognizing active genes in a tissue and may also be used for constructing comparative genetic maps of conifers. They can, for example, be labelled and used as probes for RFLP.
Recently a large number of single nucleotide polymorphisms (SNPs) distributed throughout the human genome have been mapped. They will be used e.g. in studies of human population genetics. Their role in forest genetics is under investigation.

In Table 7-1 we present our opinions about the usefulness of different traits for estimates of among-population differentiation and for identification of individuals.

Metric traits are superior when there is an interest in revealing differentiation as a result of adaptation to various environmental conditions. This is particularly the case if natural selection played the major role for the present population structure. The overwhelming majority of markers are neutral, which means that they are not affected by natural selection. The possibilities of detecting differences are limited if few markers are available but increase with higher numbers of markers. If there is linkage between a marker and traits of value for adaptedness it is possible to detect differentiation for markers too. The higher the number of marker loci analysed, the greater the probability that some marker loci are linked to loci affecting adaptive traits. Neutral markers may therefore reflect previous adaptation.

The assumptions given in Table 7-1 as regards such metric traits as growth rhythm, survival, and tree growth or isozymes may be analysed using available data for Norway spruce and Scots pine. Growth rhythm is the point of time for onset of growth during spring and cessation of growth during autumn. These points of time are important for avoidance of exposure to late spring frosts or early autumn frosts. All isozyme studies show a much smaller differentiation than for the adaptive traits mentioned above. It is worth mentioning that the statistical technique used for markers is less precise than for metric traits. This means that the differences might be underestimated. In spite of this it is evident that studies of isozyme variation and variation in metric traits give different types of variation pattern. The variation we observe for neutral markers may be attributed to linkage as mentioned above or to the fact that it takes some generations to reach equal allele frequencies in different populations via gene flow.

Even if metric traits well reflect past adaptation in populations they are more or less useless for genetic identification of individuals. This is because many alleles at many loci affect a quantitative trait, each allele contributes little.

As with identification of a human father using markers, the genetic identification of a tree is facilitated if numerous markers are available. Microsatellites with their hyper-variable DNA seem to be the best choice for such identification.

Since mitochondrial and chloroplast DNA provide only a few markers, they are not particularly well suited for genetic identification of individuals. They have, however

<table>
<thead>
<tr>
<th>Type of trait</th>
<th>Differentiation of populations</th>
<th>Identification of single genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metric</td>
<td>significant for traits of adaptive value</td>
<td>non-existent</td>
</tr>
<tr>
<td>morphological</td>
<td>insignificant</td>
<td>insignificant</td>
</tr>
<tr>
<td>single isozyme locus</td>
<td>insignificant</td>
<td>insignificant</td>
</tr>
<tr>
<td>simultaneous analysis of many isozyme loci</td>
<td>limited</td>
<td>the more loci the better</td>
</tr>
<tr>
<td>Nuclear DNA:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td>limited</td>
<td>significant</td>
</tr>
<tr>
<td>RAPD</td>
<td>limited</td>
<td>significant</td>
</tr>
<tr>
<td>AFLP</td>
<td>limited</td>
<td>significant</td>
</tr>
<tr>
<td>EST</td>
<td>limited</td>
<td>insignificant</td>
</tr>
<tr>
<td>Microsatellites*</td>
<td>limited</td>
<td>highly significant</td>
</tr>
<tr>
<td>Chloroplast DNA</td>
<td>limited – when inherited paternally significant - when inherited maternally</td>
<td>the more loci the better</td>
</tr>
<tr>
<td>Mitochondrial DNA</td>
<td>significant</td>
<td>the more loci the better</td>
</tr>
</tbody>
</table>

* If several microsatellites are identified for cpDNA or mtDNA they have the same characteristics as nuclear microsatellites

Table 7-1. Schematic summary of the possibilities to identify population differences and single genotypes using different traits.
The development of the so-called neutral theory was one consequence of the results of isozyme research. According to this theory most of the molecular changes in DNA are selectively neutral and their future existence in a population is dependent on genetic drift. The neutral theory is not accepted by all geneticists. The probability for loss of a molecular change in the genetic code per generation is much higher \[\left(\frac{2N_e}{N}\right)\ln(2N)\] than for fixation of the change \[\left(\frac{4N_e}{N}\right)\]. \(N\) is the total number of trees while \(N_e\) is the effective population size. An example will be used to illustrate this. If \(N = 100\) and \(N_e = 80\), the probability for loss is only \(2 \times 0.8 \times 5.08 \approx 8\) generations while the time to fixation is \((4 \times 80) = 320\) generations.

The conclusion of this discussion is that neutral changes in the genetic code and changes of amino acids (isozymes) are suitable for phylogenetic determination while they are much less suitable for determination of adaptive variation. To investigate adaptation, studies of traits that influence fitness are required. In next section we will present results obtained for quantitative traits and markers studied in the same populations.

Comparison of markers and quantitative traits.

In 1984 the American geneticist Richard Lewontin carried out an analysis of the discrimination power of markers, such as isozymes, and quantitative, i.e. metric, traits. To have the same discrimination power as the metric trait, the latter trait must not be regulated by more genes than given by the ratio: \(1/h^2\). If the heritability of a quantitative trait is \(0.2\) the markers will have the same discrimination power as the quantitative trait if the latter is regulated by no more than five loci. Since it is expected that more than five loci regulate most quantitative traits, it is anticipated that isozymes and many other markers show lower differentiation among populations than quantitative traits. A few examples of estimates of population differentiation for markers and quantitative traits in the same populations are shown in Figs. 7.1-7.4.

The first example concerns a Canadian study of \(Pinus contorta\), in which 2 growth traits and four quality traits were compared with isozymes. Except for branch angle all quantitative traits had much larger \(Q_{ST}\) estimates than the \(F_{ST}\) value for isozymes. The interpretation of this is that isozymes and branch angle seem to be neutral traits, which are not changed by natural selection whereas the rest of the traits are strongly affected by natural selection.

Growth rhythm, such as budburst and budset, is extremely important for northern tree species. In a Finnish investigation with \(Pinus sylvestris\) populations, originating from entire Finland, 34\% of the variation in budset of \(Pinus sylvestris\) was attributed to population differences while markers such as isozymes, RFLPs, and microsatellites showed limited population differentiation in agreement with expectation that they are neutral (Fig. 7-2). Ribosomal DNA (rDNA) took an intermediate position. However, no geographic differentiation was noted for rDNA.
In France oak species have played a great role in forest genetics studies. In one case population differentiation by isozymes was compared to the $Q_{st}$ for two quantitative traits, budburst and height in *Quercus petraea*. As outlined in chapter 4, $Q_{st}$ is dependent on the heritability of the trait under study. Heritability is a term in the denominator of the equation used to calculate $Q_{st}$. $Q_{st}$ estimates for the range of heritabilities noted for budburst and height are illustrated in Fig 7-3. It is evident that the population differentiation estimated by isozymes is several times lower than for the two quantitative traits even at the highest heritabilities. It should be remarked that populations included in the isozymes study originated from a wider range than the populations included in the study of budburst and height growth. This means that the differences between the two types of traits, markers and quantitative traits, were probably underestimated.

The shrub species *Salix viminalis* has attracted much interest in Sweden as a source for energy production. Even in this species, which differs considerably from the long-lived tree species, it is evident that $Q_{st}$ estimates are higher than the $F_{st}$ estimate for isozymes (Fig 7-4). It should be noted that the drop in $Q_{st}$ for number of shoots at ages 3 and 5 may be explained by the increased competition in the field trial and its accompanying increase in heritability.

One example of a study with higher population differentiation of markers than for quantitative traits will be given. *Cedrela odorata* is a Central American tree species growing from Mexico to Panama. Population differentiation in this species was studied by aid of chloroplast DNA, nuclear AFLP, and 17 quantitative traits, both growth and morphology traits. Most populations, 26 of 29, were monomorphic for cpDNA. A consequence of such a high frequency of monomorphic populations is that high $G_{st}$ estimates are expected. In agreement with this expectation, $G_{st}$ was estimated at 0.96 while the $Q_{st}$ estimate was much lower, 0.34. The growth traits are with high probability of adaptive significance, whether this is true also for the leaf shape traits studied is uncertain. The $Q_{st}$ may be somewhat higher if only truly adaptive traits were included in the derivation of this parameter. AFLP was analysed for Costa Rican populations only and the differentiation was estimated at approximately 83%. This data suggest that many neutral substitutions have taken place in DNA without a corresponding change in adaptive and morphological traits.
In conclusion, population differentiation estimated by isozymes is several times lower than for traits that are of adaptive significance. Along an environmental gradient clines for adaptive traits are expected to be steeper than clines for isozymes (Fig 7.5). This agrees with the assumption that isozymes are neutral markers.

Variation among populations in metric traits

Most of the information about among-population variation derives from provenance research which has played a great role in forest research. One definition of provenance is a population or group of individuals of the same species occurring within or originating from one more or less rigorously defined geographic area. The important thing is that seeds were harvested from a geographically identified area. It should be noted that the term provenance is not always identical with the term origin. Thus the seed of Pinus contorta harvested in the province of Lapland in Sweden is provenance Lapland although this species originates from North-America. Therefore, provenance experiments contain genetic entries whose seeds were collected in geographically different localities and should represent a much larger area than an individual stand. Even in those cases where the seed collection is limited to one stand within a provenance the experiments are usually referred to as provenance experiments. Population would be a more accurate designation when seeds are collected in individual stands.

Provenance trials generally comprise a large number of provenances (populations) from geographically widely separated areas. Mostly such experiments are located at a series of test sites. Thus, most provenance experiments are a part of a series of experiments.

During an international conference in 1965, provenance researchers agreed on the requirements that should be fulfilled by provenance tests. Each provenance should be represented by progenies from at least 20 trees but preferably from 50 trees. The following objectives were also agreed upon:

1. The primary objective of provenance research is applied, concerned with identifying the provenances giving the highest value production within a certain area.
2. There is also a scientific objective of provenance research to trace the adaptation that has taken place as well as the environmental factors that have influenced the adaptation.

As soon as the identification of the best provenance(s) has been carried out, the best stands within the provenance area should be selected for seed harvesting. This is a complicated task since we frequently neither have access to the history of the stand nor to the silviculture applied within a seed tree stand under scrutiny. For approval the stand should be of such an age that an evaluation of tree quality could be carried out. Moreover, the stand should have such a size that selfing is unlikely. Generally, seed harvesting is only carried out in stands fulfilling certain phenotypic standards. Stands, in which segregation of phenotypically inferior trees occurs, are excluded since this suggests that vitality reducing alleles occur in such a stand. In many countries a federal organisation approves stands for seed harvests. Sometimes this approval is also given to stands in other countries, from which imports can then take place.

Pinus sylvestris and Picea abies provenance research

Already during the early part of the 1900s it was clear to Central European researchers that there was a large variation among populations of Scots pine and Norway spruce. The Austrian forest researcher Cieslar concluded that the physiological varieties were hereditarily adapted to the length of the vegetation periods in their respective native habitats. Based on their experiments, both Cieslar and his contemporary colleague from Switzerland, Engler, were aware that there was a continuous variation of Norway spruce and Scots pine from north to south and from valley bottoms to high elevations.
A pioneer achievement was that of Olof Langlet who presented his thesis *Studies on the physiological variability in Scots pine and its relationship with the climate* (translated from German) in 1936, a publication that has attracted much international attention. Langlet demonstrated that the dry matter at a certain point of time during the autumn varied continuously from south to north in Sweden (Figure 7-6). The dry matter content reflects the degree of hardiness obtained in a certain material. Thereby, the frost tolerance attained is indirectly revealed. Langlet was probably the first to introduce replications in provenance trials.

The Swedish forest researcher Gunnar Schotte, who worked in the early 20th century, was probably the first who established real provenance trials in northern Europe, starting in 1904. It took a few decades before forest researchers were aware of the need for establishing experimental plantations with replications. In spite of this, his results give us some guidance about survival and yield of different provenances. His pioneering work was followed by others and during the 1930s it was evident to provenance researchers that the local Scots pine in northern interior Sweden did not have satisfactory survival. Some researchers even observed that there was a large variation within a provenance as well. It was not until the 1960s that foresters in Sweden realised that Scots pine seed transfers to the south must take place in the northerly harsh areas of Sweden to get satisfactory regeneration. The credit for this must be given the Vilhelms Eiche who during the late 1940s carried out a country-wide collection of seeds in approximately 100 stands for establishment of a country-wide experimental series of provenances. The series contains a few non-Swedish populations as well. This series of provenance trials differs from conventional ones by including different provenances in different test plantations. Eiche’s intention was to evaluate the effect of transfer on the provenance performance. He included different transfers in latitudinal (north - south) and altitudinal (up - down cf Fig 7-7) direction. Evidently trials close to the timber line or to sea level could not have all possible transfers. Each provenance in this series is represented by open-pollinated progenies from 20 trees per stand which makes it unique. Thanks to Vilhelms Eiche we have good knowledge not only about effects of transfer but also about variation within each population for a large number of traits.

For the northern part of Sweden the results as regards survival agree extremely well within the provenance series established by Eiche. Transfers to a northern test locality reduce survival while an opposite transfer improves survival (Figure 7-8 and 7-9). Figure 7-8 reveals that the

**Figure 7-6. The relationship between dry matter content and original latitudes of Pinus sylvestris populations at a certain date during inwintering.**

**Figure 7-7. The principle for testing of transfer effects – north-south, upwards-downwards, in the Eiche series of Pinus sylvestris provenance trials.**
mortality in the best populations was above 40% at this harsh site. Data from the series established by Eiche suggest that one degree of latitudinal transfer causes a change in survival of approximately 10 percentage units while a change of 100 meters in elevation gives a change of approximately 3 percentage units.

Figure 7-10 is an illustration of the impact of transfer on yield per hectare from the largest experimental plantation in the series established by Eiche at 400 masl close to latitude 64°. As seen from this figure a long transfer southwards seems to give the best yield in this plantation.

The results from the provenance trials can also be used to map biologically the harshness of individual test plantations. In other words we can use the results to map Sweden biologically with respect to Scots pine hardness to provide a severity index (Figure 7-11). Severity index is the expected plant mortality in per cent of the local popu-

Figure 7-8. The percentage mortality of transferred Pinus sylvestris provenances at one of the harshest trials located at latitude 66°16’, 440 masl.

Figure 7-9. The percentage tree mortality in a Pinus sylvestris provenance trial at latitude 64°19’, 400 masl. The position of the bars indicates the transfer in latitudinal and elevational direction. The provenances above the horizontal 0-line were transferred to the south and provenances to the right of the diagonal 0-line were transferred upwards.
Figure 7-10. The relationship between volume per hectare and latitudinal transfer of Pinus sylvestris provenances in a trial at latitude 64°19', 400 masl.

we know that transfers to southern test localities increase the number of high quality trunks. The effects of transfer are most pronounced at southern hilly plantations and northern low-level plantations.

Also for Picea abies there are provenance trials in Sweden over half a century old, which have given us useful information about provenance variation as regards growth and yield. In Figure 7-12 the results are summarised from one of the test plantations in the largest Norway spruce provenance series in Europe.

Neither in this case does the local provenance give the best result with respect to survival or growth. In Sweden Norway spruce should be transferred to the north to utilise the growth potential of this species to the full extent. The evolutionary explanation for the inferiority of the local provenance will be treated in the section Darwinian fitness and domestic fitness below.

For Norway spruce the timing of budburst in spring is extremely important for the adaptedness to the weather conditions at the reforestation site. For Norway spruce and many other tree species at high latitudes it is important that they do not start their growth too early during spring to avoid late spring frosts. For timing of budburst there is a large variation among provenances (Figure 7-13). Budburst time is mainly regulated by temperature. Northern populations require a
lower heat sum than southern populations for bud burst. It is also important to attain hardiness before the early autumn frosts appear. Apical bud-set is a trait fairly well correlated with hardiness and can be used to get an estimate of the degree of hardiness in a material. The onset of inwintering and thus building up of hardiness is mainly triggered by the night length. For this trait there are large differences among provenances. Northern populations require a shorter night length for onset of this process than southern populations (Figure 7-14). Pronounced altitudinal clinal variation was observed for populations from the Alps (Figure 7-15).

Both for budburst and inwintering it is evident that adaptation to the climatic conditions at the sites of origin has played a major role for the observed differences. For northern and high elevation populations it might be advantageous to respond rapidly to warm weather during spring to make use of the short summer conditions prevailing at high latitudes. To avoid early autumn frost exposure it is important that northern populations respond to short night lengths for building up of hardiness. Both bud burst and the critical night length for budset display clinal variation from north to south in Scandinavia.

Provenance research in some other conifers

Most North American conifers show clinal variation in agreement with the observations for Scots pine and Norway spruce in northern Europe. *Pseudotsuga menziesii* (Mirb.) Franco, Douglas fir, and *Pinus contorta*, lodgepole pine, are of importance both in their native countries and Europe. Therefore, we will focus on these two species.

*Pseudotsuga menziesii* is native to western North America and ranges from scattered populations at 19°N in Mexico
Two geographic subspecies are distinguished, the coastal, green variety, *Pseudotsuga menziesii* var. *menziesii* and the interior, blue one, var. *glauca*. The coastal, var. *menziesii* grows in the Pacific Northwest along the coastal range down to latitude 34°N in California west of the summit of the Cascade Range and Sierra Nevada. Near the Pacific Ocean, the winters are mild and wet and the summers cool and relatively dry with long frost-free periods. In the mountainous zone of the Cascades and Sierra Nevada the climate tends to be harsher. The interior, var. *glauca* grows along the Rocky Mountains into central Mexico. The climate is most continental in the central region with long and harsh winters and hot and dry summers compared to the northern and southern regions. The rainfall varies within the three regions depending upon the aspects of the mountains.

Douglas-fir is well-known as being one of the most important timber trees in the world, often marketed as ‘Oregon pine’. It was introduced to Europe and extensively planted in the middle of the 19th century. Since then, a large number of provenance trials have been established both in the United States, Canada and Europe. At the turn of the 21st century, Douglas-fir is the most extensively planted introduced species in Western and Central Europe. Plantations exist in all countries in the European Union, comprising an afforested area of altogether 630,000 ha. About half of this area is in France. In Sweden, Douglas-fir has been of no importance for Swedish forestry. Plantations exist on about 100 ha, mainly in the southern and central parts. Provenance trials were established from the IUFRO seed collection made in 1966/1970 in Northwestern America. The survival was very low in most of these trials owing to the use of too southern provenances for Swedish climatic conditions.

The extremely broad distribution range of Douglas-fir is reflected in the large genetic variation. The main limit-
in the inland, dry sites of the coastal Douglas-fir than near the Pacific Ocean whereas a second flush is not so common in provenances from interior, northern latitudes and high altitudes.

The results from Douglas fir reveal a more complex relationship with geographic variables than the simple relationship between latitude and growth rhythm traits in Scandinavia. In Scandinavia the climate mainly varies with latitude whereas the mountain chains in western North America weaken the relationship between latitude and climate.

*Pinus contorta* has played a prominent role in Swedish forestry during the last half of the 20th century. For this reason many provenance trials were established in Sweden north of latitude 60°. The results from these trials indicate large differences among populations, which are expected considering the wide distribution of lodgepole pine in North-America. One way to estimate the usefulness of a provenance is to multiply the percentage survival by a growth trait such as height or stem volume. This is particularly useful for areas in which survival of many provenances is unsatisfactory. In Figure 7-17 the relationship between the product, % survival x mean stem volume, and the latitudinal origin of the provenances is shown. In the lower part which deals with data from a southern test locality there is no problem with survival. The southern populations, which respond to a longer night for growth cessation than northern populations, give rise to taller trees. As we approach harsher conditions, the % survival becomes the most important component of the product and northern provenances are superior to southern since the latter do not have a satisfactory survival. The results from lodgepole pine shown in Figure 7-17 illustrate well that we have to weigh the survival against growth to reach an optimum yield per hectare. Moreover, the relative importance of these two changes from mild to harsh climatic conditions.

Sometimes freeze tests are used to assess the hardness attained in a material during the process of inwintering. Plants are first grown under growth promoting conditions. After that a continuous night prolongation is applied. Freeze tests are then applied at certain intervals. If the freeze testing is carried out too early during the inwintering most plants will be severely damaged and if it is carried out too late, most plants have attained full hardiness. It is therefore of importance to carry out the freeze testing such that approximately 50 % of the plants are severely damaged to obtain the best resolution with respect to hardiness in the material. Often the results are related to some environmental variable such as latitude or elevation. One example for *Pinus contorta* is shown in Figure 7-18, in which the percentage of severely damaged plants is plotted against the latitudinal origin of the provenances. Similar relationships between tree mortality in four field trials and latitudinal origin of provenances are also shown in this graph. The provenances included in the field trials and the freeze testing are not identical but they originate from the same latitudinal range in Canada. The agreement between the slopes of the five curves is good. This suggests that freeze testing well reflects the field mortality.

Another important issue is the susceptibility to pests and diseases of an introduced species such as lodgepole pine. Introduced species and populations are sometimes referred to as exotics. At the end of the 1980s there were certain weather conditions which caused severe attacks by the *Gremeniella abietina* fungus on lodgepole pine plantations in northern Sweden. Weather-conditioned damage has in many cases been the gateway for fungal attacks. To avoid attacks as much as possible it is important to use provenances with good hardiness.
Pinus ponderosa is an important tree species from western North America. In Fig 7-19 the results from a study in two contrasting nurseries with a range-wide collection of populations are summarised. Thanks to the large number of populations it is possible to identify in an accurate way the importance of the origin-population effect and of the interaction-population x site effect. As seen from Fig 7-19 the population effect was stronger than the interaction for elongation and growth cessation in spite of the contrasting growth conditions in the two nurseries. The interpretation of these results is that there was a strong natural selection for elongation and growth cessation during the past evolution of this species. Contrary to this, the growth initiation seems to have been less affected by natural selection in the past. In this respect *P. ponderosa* resembles *P. sylvestris*.

In Figure 7-20 results on growth at 285m asl of *Pinus ponderosa* populations from an elevational transect in California are illustrated. As seen from this figure the same trend as in Scandinavia is observed with poor performance of high elevation populations at a low test site.

There are two north American species, *Pinus monticola* and *Pinus resinosa*, which differ from the general pattern of large provenance differences although they have a wide distribution area.

In *P. monticola* there is a sharp difference between populations south and north of central Oregon in northwestern USA. Except for this sharp border there is almost no genetic variation in growth and phenology either in the northern part or in the southern part of the distribution of this species. A very large phenotypic plasticity might be one explanation but the experts on this species have ruled out phenotypic plasticity as an explanation. One possible explanation is that the species occupies a specific habitat and is evidently outcompeted by other spe-

![Figure 7-19. The population and population x test site variance components for growth initiation, growth cessation, and plant elongation in Pinus ponderosa populations grown in two nurseries with differing water availability.](image)

![Figure 7-20. Tree height at age 29 of Pinus ponderosa populations from a Californian elevation transect grown at a low elevation, 285 meters above sea level.](image)
cies in other habitats. In terms of selective environmental neighborhoods it would occupy two SENs, one south of and one north of central Oregon. A less probable explanation is that *P. monticola* experiences the environment as very variable over time. In consequence, natural selection has figuratively operated in varying directions during the evolution of the species. Evidently most other conifers with the same distribution area did not experience the environment as so variable as *P. monticola* did. Therefore, this explanation is less likely.

*Pinus resinosa* grows mainly in xeric habitats in a 700 kilometers wide band from eastern Manitoba and Minnesota to the Atlantic coast in the east. There are several climatic zones in this huge distribution area, which ought to have caused a population differentiation, but there is almost no differentiation. One hypothesis is that the species after the last glaciation has passed through several bottlenecks, *i.e.* the effective population size was low at several occasions. This might have eroded the genetic variation of the species. The low inbreeding depression observed in the species lends some support to this hypothesis. Another explanation could be that it only occupies xeric conditions and that the species for this reason experiences the environment as fairly homogeneous. The larger genetic variation in *Pinus strobus*, which has a similar distribution area, might be attributed to its occupation of a wider range of site conditions than *P. resinosa*.

*Figure 7-21. Right. The relationship between volume per hectare (extrapolations from smaller plots were done) at age 22 of *Betula pendula* and provenance transfers at latitude 60° 21' in Finland. The curve, explaining 81% of the variation, is based on unpublished data kindly provided by A. Vihreä-Aarnio, and P. Velling, Finland.*

*Picture 7-3. Leaf colouring and leaf fall in *Betula pubescens*. Trees from latitude 67° N (far left) are defoliated while the trees in front of the path from latitude 60° N still have green leaves. Photograph Gösta Eriksson*
they do not build up hardiness in due time. Therefore, they become frost damaged most years.

As stated before, oak species have played a great role in forest genetics studies in France. In Fig. 7-3 an example of estimated population differentiation in *Quercus petraea* is given for some important traits. The range of $Q_{st}$ estimates are given for the span of heritabilities observed for the different traits. The generally high minimum estimates for the growth rhythm traits and height indicate large population differentiation for these traits.

Good growth is a question of optimisation, since both too late an onset of growth during spring and too early growth cessation during the autumn will give rise to small plants and trees which will not set good seed. The latter means that they have a low fitness. One example of this problem of optimisation is taken from a Finnish experiment with silver birch. As may be seen from Figure 7-23 leaf colouring and growth are mirror images of each other. The plants with the advanced autumn colouring are smallest. The highest fitness at a certain locality will those trees have which show the best balance between onset of growth and growth cessation at this particular site. Since fitness is a relative concept it is valid for the trees in one population growing in one environment. The ranking with respect to fitness of the same trees might be totally different at another growth locality.

The general trend of earlier budburst and growth cessation in northern than in southern populations has been demonstrated for several tree species. Examples on growth cessation from studies in growth chambers and nursery are illustrated in Fig. 7-24 and 7-25.

The reason for the clinal variation from south to north in Scandinavia and from low elevation to high elevation is that the climate varies in a similar way. What we observe as provenance differences is a confirmation of what was illustrated in Fig. 6-6 that there is disruptive selection among provenances. In countries outside Scandinavia in which the climate is not so much influenced by latitude there might be other relationships with geographic variables. Thus, in Spanish populations of *Castanea sativa* the southern populations had an earlier budburst than the
northern ones (Fig 7-26). The southern populations originate from localities with severe summer drought. Under such conditions it may be an advantage to have an early budburst to capitalise on the favourable growth conditions during the early part of the summer before the ambient conditions become too limiting for growth. Plants with a late budburst will have a shorter spring – early summer growth period, which in turn means that such plants will be shorter and less competitive than the early flushing plants. In some cases it might be the precipitation that is the most decisive environmental factor. In other cases climate changes with the distance from the coast.

Adaptation to edaphic conditions

The examples presented above clearly indicate that there is a pronounced population differentiation for growth and growth rhythm traits that must be attributed to climatic differences over the range of distribution of species. There is limited information about the impact of edaphic conditions on population differentiation. At the species level there is a clear difference on preferences with respect to edaphic conditions. One clear example comes from the Pirin valley in Bulgaria with its two pine species, *Pinus peuce* and *P. Heldreichi*. Each species occupies just one of the two slopes. The two slopes differ with respect to soil conditions, *P. peuce* prefers silicate and *P. Heldreichi* prefers limestone soil. Therefore, it might be speculated that such differences may be extended to populations within a species. There are few data available on well designed experiments to study adaptation to edaphic conditions. There is one example for *Fraxinus excelsior* studied in Germany. Two populations from dry and wet sites were included in an experimental series planted at three sites, one wet, one dry, and one intermediate locality. The tree height at age 10 is illustrated in Figure 7-27, which

![Figure 7-25. Bud set in five Ulmus laevis populations studied in a nursery in Uppsala, Sweden. The blue circle sector indicates the percentage budset a certain date.](image)

**Figure 7-25. Bud set in five Ulmus laevis populations studied in a nursery in Uppsala, Sweden. The blue circle sector indicates the percentage budset a certain date.**

![Figure 7-26. Bud flushing stage in Spanish Castanea sativa populations studied in a Spanish nursery. The water availability during summer varied considerably at the site of origin in the two groups of populations.](image)

**Figure 7-26. Bud flushing stage in Spanish Castanea sativa populations studied in a Spanish nursery. The water availability during summer varied considerably at the site of origin in the two groups of populations.**

![Figure 7-27. Tree height at age 10 of Fraxinus excelsior populations from two types of origin, wet and dry sites, tested at three sites with varying water availability. Green bars refer to two populations from wet sites, brown bars refer to populations from dry sites.](image)

**Figure 7-27. Tree height at age 10 of Fraxinus excelsior populations from two types of origin, wet and dry sites, tested at three sites with varying water availability. Green bars refer to two populations from wet sites, brown bars refer to populations from dry sites.**
shows that the populations from the wet origin outgrew the two populations from the dry sites at the wet test site. At the two other test sites there was no significant difference between populations from the two origins. Data from tree height at age 33 at the dry test site confirm the absence of population differences at age 10 for the dry test plantation.

Utilization of provenance results

How can the results from provenance research be utilized in applied forestry? Continuous variation along ecological gradients has been demonstrated repeatedly. The question is: Over how large an area could a provenance be used without losing in production as we move from the optimum of this provenance? To elaborate on this, a hypothetical situation is shown in Figure 7-28. In this figure the relative production of four provenances is shown graphically. If we accept a drop of production to 90% of the maximum, these four provenances cover precisely the range in the figure. If we do not accept a greater loss than 5% then we do not have any suitable provenance for areas I and II in this ecological gradient. The reason why we do not find a provenance that satisfies our requirements might be that the provenances tested so far have their origins too far apart from each other. In other words another provenance test with a denser net of provenances might give us the proper provenances for the entire ecological gradient under the requirement of no more than 5% drop in production from the optimum.

In many countries seed transfer rules are based on results from provenance research. In Sweden the most recent forestry act from 1994 says that local provenances should be used, which is in conflict with most provenance research. Optimum production will not be obtained if this recommendation is followed.

Markers

Most forest geneticists agree that the majority of isozymes are neutral markers and as such they do not contribute to fitness. However, it is evident that different ambient conditions influence the efficiency of certain isozymes. As a corollary of this it has also been assumed that different isozymes vary with respect to adaptedness. One way to estimate this is to compare the genetic distances of different loci in genetic entries varying widely in their response to the ambient conditions. One example of such a comparison is the characterisation of isozymes in German populations of *Fagus sylvatica* tolerant and susceptible to air pollutants (Fig 7-29). As seen from this figure the distances for loci PGM-A and LAP-A are much larger than for 6-PGDH-A and PGI-B loci. Therefore, it may be speculated that the 2 former loci may have been changed by natural selection while the latter are less affected by natural selection.

Another indication for non-neutrality of isozymes is the clinal variation in isozymes allele frequencies along environmental gradients. However, it should be noted that it takes several generations to level allele frequencies in different populations after a new mutation has arisen. This means that there is always a time lag until allele frequencies are levelled; the further apart the larger the difference in allele frequency between populations. Clinal variation
of isozymes alleles may therefore be attributed to the time lag for levelling of allele frequencies. It should be noted that many population studies did not show clinal variation of the isozymes. One example of this is presented in the next paragraph.

In Fig 7-30 the genetic distances estimated by isozymes between Swedish *Pinus sylvestris* populations are shown to the left. It should be noted that there is a limited genetic differentiation among these populations. The geographic origin of the populations is also indicated and it is clear that there is no geographic trend. The absence of geographic trends in studies involving isozymes is an indication that the isozymes studied are neutral. In this study it can be remarked that the E1 population was closest related to the B1 population and less related to its neighbour population E2. Both E populations would have no or limited survival at the B locality four degrees further north.

In * Castanea sativa * a conspicuous transition in allele frequencies were noted in Bithynia in western Turkey (Fig. 7-31). The transition zone was estimated at 324 km. The allele frequencies in the eastern and western Turkey populations were relatively uniform but different from each other. The estimated number of migrants per generation was higher within each of the three regions, western, Bithynian, and eastern, than between regions. One possible explanation is that the western and eastern populations developed isolated from each other over several generations and recently came into contact with each other. The development in isolation should in this case have resulted in different allele frequencies in the two regions. Another alternative explanation is that the variation in allele frequency from east to west is a reflection of past natural selection. A closer look at the climatic conditions in the three regions shows that the climate does not vary much within the Bithynian region while it varied considerably in the eastern region. These facts speak against the selection interpretation of the results. Palynological data give some support to the first hypothesis of recent contacts of two previously isolated populations.

The population differentiation estimates (F<sub>ST</sub> or G<sub>ST</sub>) based on isozymes in most widespread conifers rarely exceed 0.05. This means that there is limited population
differentiation with respect to isozymes markers and the most likely reason for this is strong gene flow among populations. In Table 7-2 a compilation of estimates of population differentiation is given for some broad-leaved tree species. It should be remarked that the estimates are dependent on how the selection of population was done and on the number of isozymes markers that was analyzed. With these limitations in mind it is anyhow a tendency that widespread and wind-pollinated species such as *Betula pendula*, *Castanea sativa*, *Quercus petraea*, and *Quercus robur* show lower estimates than species with non-continuous populations such as *Acer platanoides*, *Alnus glutinosa*, and *Sorbus torminalis*. The highest estimate was noted for small, scattered, and marginal populations of *Ulmus laevis* at the northern margin of distribution in Finland. It is likely that genetic drift has played a great role during preceding generations of these small populations. *Pinus cembra* has relatively small and scattered populations in central European mountains. In spite of its scattered distribution, population differentiation was limited, $F_{ST} = 0.047$. The most likely explanation for this is that the isolation of the populations occurred relatively recently. In this context recently means that the isolation occurred a few generations ago.

In Fig 7-32 one example of a study including chloroplast DNA (cpDNA) markers in three oak species is illustrated. Such markers are frequently referred to as haplotypes. As seen from this figure there is a large population differentiation in all three species. The reason for this is that there are few markers and that several populations have just one marker. When several populations are monomorphic with respect to one marker and several other populations are monomorphic with respect to another marker a large differentiation is obtained. In such a case there is no discrimination between the monomorphic populations sharing the same cpDNA marker.

In Fig 7-33 population differentiation based on isozymes and mitochondrial DNA (mtDNA) in some western American pine species *P. attenuata*, *P. muricata*, and *P. radiata* is given. The figures in the bars refer to number of trees and populations included in the study.

<table>
<thead>
<tr>
<th>Species</th>
<th>$F_{ST}$ or $G_{ST}$ estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Acer platanoides</em></td>
<td>0.10</td>
</tr>
<tr>
<td><em>Alnus glutinosa</em></td>
<td>0.20</td>
</tr>
<tr>
<td><em>Betula pendula</em></td>
<td>0.03</td>
</tr>
<tr>
<td><em>Castanea sativa</em></td>
<td>0.11</td>
</tr>
<tr>
<td><em>Quercus petraea</em></td>
<td>0.02</td>
</tr>
<tr>
<td><em>Quercus robur</em></td>
<td>0.05</td>
</tr>
<tr>
<td><em>Sorbus aucuparia</em></td>
<td>0.06</td>
</tr>
<tr>
<td><em>Sorbus torminalis</em></td>
<td>0.15</td>
</tr>
<tr>
<td><em>Ulmus laevis</em></td>
<td></td>
</tr>
<tr>
<td>marginal populations</td>
<td>0.33</td>
</tr>
<tr>
<td><em>Ulmus minor</em></td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 7-2. Compilation of observed population differentiation in European broad-leaved tree species by aid of isozymes.

Figure 7-32. Genetic distance, $G_{ST}$ based on chloroplast DNA in *Quercus pubescens*, *Q. robur* and *Q. petraea* populations.

Figure 7-33. Genetic distances based on isozymes and mitochondrial DNA (mtDNA) in populations of the North American pine species *Pinus attenuata*, *P. muricata*, and *P. radiata*. The figures in the bars refer to number of trees and populations included in the study.
The clinal variation can be an effect of a time-lag in le-

The data for these shortcomings general trends were noted. In spite of how the studied populations were sampled, whether monomorphic loci were included in the analysis or not, as well as the number of trees and loci included. In spite of these shortcomings general trends were noted.

The data for *Pinus cembra* presented above do not conform to the expectation of relatively large inter-population difference for a species with a limited range of distribution.

In conclusion, isozymes usually show limited population differentiation that sometimes is clinal. The latter does not necessarily mean that isozymes contribute to fitness. The clinal variation can be an effect of a time-lag in velling of allele frequencies between populations. Other markers frequently show large population differentiation owing to many monomorphic populations. In spite of the large estimates of differentiation there is a weak discrimina-

### Darwinian and domestic fitness

Geneticists sometimes distinguish between Darwinian fitness, *i.e.* the adaptedness in nature and domestic fitness, which is the ability of a genetic entry to produce biomass, high quality timber, shelter, or any other utility for us as human beings. These two types of fitness might coincide but usually this is not the case. In nature extremely good growth is in vain if this ability is not transferred to a progeny, which normally takes place via the seeds formed after sexual mating. Seed production might be an unnecessary and energy-demanding process for production of human utilities. This is especially pronounced for crops

| Table 7-3. Some general trends in population differentiation estimated by isozymes |
|---------------------------------|---------------------------------------------------------------|
| **Species characteristics**     | **General pattern**                                           |
| Continuous and wide distribution| Little genetic differentiation among populations within regions, most variation occurs within populations |
| Large geographic range but with separation into subspecies | Large differentiation between subspecies but little genetic differentiation within subspecies |
| Small and disjunct ranges       | Great inter-population differentiation and moderate within-population variation |
| Extremely small geographic range| Relatively large inter-population differentiation |

During the early nineties one attempt to summarise the results on isozymes’ $F_{st}$ and $G_{st}$ for conifers was done (Table 7-3). Before an analysis of the table is done it is important to point out that the results are dependent on how the studied populations were sampled, whether monomorphic loci were included in the analysis or not, as well as the number of trees and loci included. In spite of general trends were noted.

In conclusion, isozymes usually show limited population differentiation that sometimes is clinal. The latter does not necessarily mean that isozymes contribute to fitness. The clinal variation can be an effect of a time-lag in velling of allele frequencies between populations. Other markers frequently show large population differentiation owing to many monomorphic populations. In spite of the large estimates of differentiation there is a weak discrimina-

### Box 7-2 Why is the local population of *Picea abies* not the best choice for reforestation of a clearcut area in southern Sweden?

| **Comparison of the type of regeneration** |
|---------------------------------|---------------------------------------------------------------|
| **Planting at a clearcut area** | **Natural regeneration**                                      |
| Day temperature                | higher                                                         |
| Night temperature              | lower                                                          |
| Budburst                       | earlier                                                        |
| Type of fitness required       | domestic                                                       |
|                                | higher                                                         |
|                                | later                                                          |
|                                | Darwinian                                                      |

The high day temperatures at a clearcut area induces an earlier budburst than in an opening in the forest. This means that the budburst in a clearcut area takes place at the time when the probability for frost exposure is higher than when budburst occurs in the opening in a stand. Since the local population has a lower heat demand for budburst than some exotic populations, *e.g.* Byelorussian populations, the latter are better than the local population.

At natural regeneration in small openings in the forest an early budburst might be advantageous since such genotypes outcompete other late budbursting genotypes and are also competitive against other plant species.
where vegetative propagation is applied such as for potatoes. The difference might be best understood when we realise how important regeneration is in nature. As regards cultivated plants man has taken over the responsibility for propagation and this ability is no longer decisive for the continued use of a species. Many ornamental plants would be outcompeted if there was no human intervention. As human beings we cultivate them for their beauty and propagation is taken over by nurserymen, which means that their Darwinian fitness is obsolete. Below a few examples of domestic fitness are given.

Byelorussian Norway spruce outgrows the domestic Norway spruce in southern Sweden. Most people have interpreted this as a consequence of the migration of Norway spruce into Sweden after the last glaciation. From the refuge in Russia Norway spruce migrated in a western direction and one branch migrated northwards in Finland and entered Sweden after passing to the north of the Gulf of Botnia. Finally it migrated southwards in Sweden. The passing of latitude 66° is assumed to have caused an enrichment of hardiness alleles at the cost of growth promoting alleles. This explanation is not valid since Norway spruce passed latitude 66° when the climate was much warmer than today. The reason for the superiority of the Byelorussian Norway spruce is probably the type of reforestation used, ie planting after clear cutting. Norway spruce is not well adapted for regeneration on clear-cut areas since it has evolved under regeneration in small openings in forests. The reason for the superiority of the Byelorussian Norway spruce is further outlined in Box 7-2.

The later budburst is accompanied by later growth cessation, which might be harmful for building up of frost tolerance during the autumn. However, there are no signs of any problems of Byelorussian Norway spruce up to latitude 60° in Sweden. Therefore, it seems as if both the domestic and the Byelorussian Norway spruce start their inwintering too early and do not utilise the growth potential that the southern Swedish climate offers. If a population avoids frost damage it means that not only the survival is increased but also that stem defects are avoided and that the duration of the phase of establishment is reduced. The latter increases the productivity per year.

Why do we need to transfer Scots pine seeds from north to south in northern Sweden to get satisfactory survival (see Figures 7.8 - 7.10)? Once more it is a question of comparing self regeneration with planting. At planting the majority of the plants must survive to get a closed stand while it might suffice with one plant per mille under natural regeneration, since Scots pine has a profuse seed production. A Finnish scientist has estimated that one pine tree during its life time might produce one million seeds. To keep the range of distribution unchanged, it is theoretically sufficient that each pine tree gives rise to one new tree. One might question the need for such a waste of energy. This is further discussed in next chapter.
Meiosis in pollen mother cells of European, Japanese, and Siberian larch starts during late autumn. Once the diplotene stage is reached, which is one of the earliest stages during meiosis, a rest, usually called dormancy, is initiated. To break the rest a certain amount of chilling is required. This is a general phenomenon in woody plant species from the temperate and boreal zones. The chilling requirement varies among the three larch species. Increased chilling is required in the order: Siberian - Japanese - European larch. In European larch cultivated in Sweden the dormancy is normally broken in February -March when continuation of meiosis takes place. These stages of the meiotic division are probably the most frost sensitive during the life cycle of an individual.

In southern and central Sweden, with its maritime climate, there are continual changes between cold and mild periods during the winter. During certain years dormancy was broken in the pollen mother cells of Siberian larch already during November - December. The limited heat during the following mild period was enough to induce a continuation of the meiotic divisions. If the period of mild weather is short and then followed by a frost period before the completion of the meiotic divisions, severe frost damage might be induced. In certain years there is a total collapse of the meiotic divisions with temperature induced pollen sterility as a consequence. Since pollen mother cells of European larch have a larger demand for chilling to break the dormancy, the pollen formation in this species did not show as much damage as the Siberian larch. Japanese larch takes an intermediate position between the other two species. How can these differences between the three larch species be evolutionarily explained? One plausible explanation is that European larch grows under less continental climatic conditions than the Siberian larch. This means that changes between mild and cold periods occur frequently. European larch genotypes that have a large chilling demand will have a higher fitness than those with a low chilling demand. In contrast in Siberian larch there has not been any need for an increased chilling demand owing to the more stable cold winters in its distribution area. A large chilling demand probably did not contribute fitness to Siberian larch.

**Summary**

The genetic structure of a species is reflected in different ways by different traits. There are minor differences among populations as regards biochemical markers while metric traits frequently reveal major population differentiation. This is the case for the majority of tree species from the temperate and boreal zones which have been studied. Climatic conditions have played a major role for tree populations’ adaptation to the ambient conditions. At high latitudes early inwintering is important. Similarly, the higher the elevation the earlier the inwintering. At southern latitudes where the photoperiodic conditions are less dramatic other climatic elements such as drought have played a role in previous adaptation. Darwinian fitness is the ability of a genotype or a population to transfer its alleles to the progeny generation. Domestic fitness is the ability of a genotype or a population to produce some kind of utility for man. The latter is of great significance in all kinds of plant cultivation whether it is for biomass production or beautification.

**Further reading**


Variation within populations

In this chapter we present observed variation within populations and genetic parameters derived from experiments with full-sib or half-sib progenies. Heritabilities and coefficients of additive genetic variation are presented for growth, growth rhythm traits, and disease tolerance. First results for conifers are shown and after that data for broad-leaved tree species are presented.

More or less all early studies of variation within populations were connected to breeding programs. For this reason there is more information from tree species with high economic value such as Picea abies, Pinus elliottii, Pinus sylvestris, Pinus taeda, and Pseudotsuga menziesii than from other tree species. First we will present some observations of variation within populations and after that turn to estimates of genetic parameters.

Examples of variation among families for various traits

Large variation in timing of bud burst of open-pollinated progenies of Picea abies in different populations from Slovakia and Poland was noted in a south Swedish nursery (Figure 8-1). In some of the populations the range in time is approximately two weeks. The variation in phenology has also consequences for the growth; a short growth period usually means a poor growth. Variation in juvenile growth in a Norwegian nursery test of several

Norwegian and a few exotic populations of Picea abies is demonstrated in Fig 8-2. The variation of open-pollinated family means for age 3 heights is considerable. In a study of Picea sitchensis there was a large variation both within populations and between populations in tree height at age 10 (Fig. 8-3). In spite of the conspicuous difference between the families the heritability was as low as 0.07. This must be attributed to a large variation of the seedlings in a family both within and between replications. Figure 5-6 is another example of a large within-population variation of such an important trait as survival in Pinus sylvestris

Figure 8-1. The variation in timing of budburst in open-pollinated progenies from individual parents in Polish and Slovak populations of Picea abies studied in a nursery at 59°30’.

Figure 8-2. The range of family means of plant height at age 3 of Norwegian open-pollinated families of Picea abies and a few exotic sources. CZ = Czech, DE = German, PL = Polish, and SF = Finnish populations, respectively.
in northern Sweden. During the late part of the previous century increasing knowledge of within-population variation in broad-leaved tree species has accumulated. An example from *Quercus robur* illustrates (Figure 8-4) this for budburst. It should be stressed that too early or too late assessment will underestimate the genetic variation.

Disease resistance or rather disease tolerance is of utmost importance in several conifers. Fusiform rust, *Cronartium quercuum*, causes great losses for forest owners in South Eastern US owing to attacks on the important pine species, *Pinuselliottii* and *P. taeda*. There is a large variation in susceptibility as is illustrated for one experiment with 16 *Pinus taeda* open-pollinated families (Fig. 8-5, Picture 8-1). Blister rust is another important disease affecting *Pinus strobus*, *P. lambertiana*, and *P. monticola*. In four experiments with over 200 open-pollinated families in each of the latter two species the survival at age 5 after artificial inoculations with *Cronartium ribicola* varied between 1.6 and 13.1%. In spite of this low field survival the family means varied from 0 to 54.8%.

**Figure 8-3.** Range of family means for tree height at age ten of *Picea sitchensis* open-pollinated families from seven populations.

**Figure 8-4.** The variation in timing of budburst in open-pollinated progenies from individual parents in five Swedish *Quercus robur* populations studied in a nursery at latitude 58°38'. The latitudinal origin of the populations is given. The coefficients for additive variation for budburst of individual populations are given.

**Picture 8-1.** A *Pinus taeda* tree severely infected with fusiform rust. Photograph Gösta Eriksson.

**Figure 8-5.** Variation in fusiform rust tolerance in *Pinus taeda* families.
Dutch elm disease caused by *Ophiostoma novo-ulmi* and chestnut blight caused by *Cryphonectria parasitica* seem to be the most spectacular diseases in European and North American deciduous tree species. Great efforts have been devoted to identify tolerant material and a few elm cultivars have been released. The success in obtaining disease tolerance in *Castanea dentata* is meagre. The limited success in elms and American chestnut suggests that there is limited variation in disease tolerance in these two species.

Heritabilities and coefficients of additive variation

In breeding programs estimates of heritability, $h^2$, have taken a prominent role and more recently estimates of the coefficient of additive variance, $CVA$, have been published. It should be noted that in most breeding programs estimates of $h^2$ and $CVA$ were based on a phenotypically limited part of the entire populations. Therefore, the estimates may be lower than they would be if there had been a representative selection of parents in the tested populations. In other cases the estimates are based on plus trees from different populations and if they are of a wide origin, the estimates may be inflated with a strong population effect and thus exaggerate the true estimate for a single population.

In a review article from the early 1990s Cornelius summarized published data on $h^2$ and $CVA$ for tree species (Fig 8-6). From this figure it is seen that the growth trait mean value for heritability was approximately 0.20 while it was about twice as high for wood density. It should once more be stressed that heritability is valid for the population under study as well as the ambient conditions prevailing at the test sites. This figure also reveals that the growth traits have at least twice as large $CVA$ estimates as density. This means that the prospects for genetic gain in growth generally is higher than for gain in density even if the heritability is twice as large for density as for growth.

Finland is a country with a large number of *Pinus sylvestris* progeny trials. In a thesis from 2002 results for tree height measurements at ages 12-18 from several series of Finnish progeny trials were summarized (Figure 8-7). The heritability estimates varied between 0.033 and 0.21. The variation could to some extent be attributed to the relative impact of G x E interaction. Thus the lowest heritability estimate was noted for the series with the highest ratio family x site interaction variance/family variance and conversely the highest heritability estimate was noted for the series with one of the lowest values for this ratio. Within the same series of trials, heritability estimates were obtained at various ages, only the heritabilities for the latest assessments are shown in Fig. 8-7. There was no clear age trend in the heritability estimates. In some cases the heritability increased with age in others it decreased.
In Norway Jon Dietrichson initiated a study of variation in three domestic populations of *Picea abies* by carrying out all possible crosses between 10 trees in each population. In Fig. 8-8 one example of results for the percentage of plant elongation a certain date during the 7th growth period is illustrated. As seen from this figure the differences in breeding values were not extremely large, but still statistically significant. An early growth cessation means that the growth period is not fully utilized. One of the conclusions from this Norwegian study was that the variation within individual populations for growth rhythm traits was larger than the variation among populations.

Several studies in the border area genetics-physiology were carried out in growth chambers in Sweden. In growth chambers plants can be exposed to various photoperiodic (day/night length) temperatures, nutrient and watering regimes. One example of an experiment with two nutrient regimes, free access and restricted access, are illustrated in Fig. 8-9. All *Pinus contorta* populations originated from approximately the same latitude and the same elevation. There was a strong response to free access of nutrients in all populations. One reason for the comparatively poor growth of the population from longi-
Picture 8-3. Plant growth after 5 weeks in a study of variation in uptake of nitrogen in Norway spruce seedlings. Treatments from left to right, free access of nitrogen, free access of nitrogen + mycorrhiza Laccaria bicolor, strongly restricted access of nitrogen + mycorrhiza Laccaria bicolor, strongly restricted access of nitrogen. Photograph Per Lindén.

High N   High N+M   Low N +M   Low N

Picture 8-4. Two-year old seedlings of Norway spruce exposed to free access, F.a., of a balanced nutrient solution and restricted access of nutrients, R.a., pipetted daily during the growth period. Photograph Per Lindén.

in controlled environments that lead to a low phenotypic variance. Since this variance is the denominator of the heritability it explains the high estimates of heritability under these uniform conditions.

Plant and tree growth is a complex trait that may be decomposed into several components (cf Fig. 8-10). At high latitudes there are two major components, growth rate and duration of the growth period. Growth rate may be split into nutrient efficiency, water use efficiency, and photosynthetic efficiency. The former may be further sub-

![Diagram](image)

**Figure 8-10. A compilation of heritabilities for water use efficiency, uptake of nutrients, and nutrient utilization based on studies in the Uppsala phytotron with various tree species.**
divided into uptake of nutrients, utilization of nutrients once the nutrients are inside the plant, and reallocation of nutrients. Separate estimates for each nutrient element may be obtained for each component of nutrient efficiency. Several of these components were studied in the Uppsala phytotron and the range of heritability estimates for these components were obtained for several tree species. In many cases high estimates of heritability (Fig. 8-10) and $CVA_\alpha$ were obtained. The Norway spruce families that had the poorest growth at low nitrogen level in the treatment without mycorrhiza benefitted most from mycorrhiza association. This explains the lower heritability in the treatment with mycorrhiza. It should be noted that only juvenile plants can be studied in growth chambers. If these components are regulated by different sets of genes it may be possible in breeding to combine several of these components in one progeny. Even combinations that never have existed can be obtained in breeding if the sets of genes regulating different growth components are identified. In agreement with results for *Pinus contorta* the heritability estimates reached in many cases much higher levels than ever reported for growth traits from field experiments. It should be noted that the estimates in Fig. 8-10 were not inflated by any population effect.

In a Lithuanian series of progeny trials with *Quercus robur* juvenile growth, budburst, and autumn leaf coloring were studied. As seen from Fig. 8-11 the family variance component for budburst was several times larger than

Figure 8-11. Variance components for family and family x site effects for budburst, autumn leaf colouring, and height growth in Lithuanian *Quercus robur* populations.

Figure 8-12. Family and family x site interaction for budset, budburst, and plant height in juvenile material of *Castanea sativa*.

Figure 8-13. Coefficients of additive variation for budburst, budset, and plant height of deciduous tree species with varying combinations of life history traits. Based on the thesis by Virgilijus Baliuckas.
Naturally regenerated material of Scots pine in northern interior Sweden has a large genetic variation. In this part of Sweden the plants are exposed to extreme strains during late winter when they have reached a size of approximately one meter. Once the trees have emerged from the snow cover during late winter the large amplitude in temperature between day and night may be harmful to plants, which respond rapidly upon the high day temperatures. Depending on the ambient weather conditions the bottleneck will take different positions along the vertical environmental scale. During these critical years the genetic variation will be narrowed considerably. When the plants have developed into trees the environmental conditions no longer constitute a strain to them. Plants which were culled during the phase of establishment would if they had survived be able to grow and even outcompete some of the trees that passed the bottleneck unharmed.

At the next occasion for regeneration the segregation of genotypes is different and so is the position of the bottleneck. Thanks to the broad segregation there are plants able to pass the new bottleneck.

What would happen if alleles at just one locus had been responsible for the survival, i.e. survival showed qualitative inheritance? To illustrate this we have assumed that the three genotypes in one locus take different positions on the environmental scale. In order to be able to pass the second bottleneck it is required that either \(a_1a_2\) or \(a_2a_2\) pass the first bottleneck to have the required \(a_2a_2\) genotype for the second bottleneck. Since neither \(a_1a_2\) nor \(a_2a_2\) passed the first bottleneck the population would not give rise to a second generation and thus become extinct. If the inheritance is quantitative there is a large segregation and some individuals would guarantee the continued survival of the population.
A CV$_A$ of 20 must be regarded as promising for breeders to change the trait by selection. Similarly, such a value is beneficial for future adaptation in nature if changes will occur in the environment. In many of the species the highest CV$_A$ values were noted for budburst while CV$_A$ for height never reached 20. There is no clear tendency that the hypothesis outlined above is true. It ought to be remarked that a comparison of the species is not totally straightforward since the populations studied were represented with different number of trees and the distribution was different. Thus, *Fagus sylvatica* is limited to the mildest climate in Sweden while *Alnus glutinosa* has a much wider distribution, which means that *F. sylvatica* had to be sampled from much smaller climatic range than *A. glutinosa*.

**Why is there such a large within-population variation in *Picea abies* and *Pinus sylvestris* and many other tree species?**

The large within-population variation described above seems to reflect poor adaptedness of the populations studied. Unique for many tree species is the long generation time. This means that a tree during its lifetime will experience large annual fluctuations in weather conditions and even climatic changes. For these reasons it might be an advantage to have a large variation around a mean value such that there are always some genotypes well-adapted to the conditions prevailing at the time of regeneration. Expressed in another way there is a trade off between high adaptedness in the short-time perspective and the potential for response to changes in a long-time perspective. A prerequisite for a large segregation is that the traits of adaptive value are quantitative. Quantitative inheritance per se might be of adaptive significance in long-lived tree species. Natural selection changes allele frequencies in different directions depending on the ambient conditions, which promotes large within-population variation (See Box 8-1 and the discussion on stabilizing selection in Chapter 6).

**Summary**

Large genetic variation for many traits of adaptive significance occurs in most tree species. The estimates of heritability and coefficient of additive variance, CV$_A$, for growth traits vary from low to moderate in field trials. The heritability for wood density was in many cases high but the CV$_A$ was low. The heritability in growth chamber studies was usually a few times higher than in field trials, which must be attributed to the uniform conditions in growth chambers. For most traits of adaptive significance there are good prospects for genetic change via natural selection or breeding. The American chestnut, *Castanea dentata*, does not seem to have any tolerance against chestnut blight, *Cryphonectria parasitica*. Most *Ulmus glabra* families are highly susceptible to Dutch elm disease, *Ophiostoma novo-ulmi*, which is a great constraint to improvement of tolerance against this disease.

**Further reading**


Forest tree breeding

General questions related to forest tree breeding are first presented. Then selection of species and the principles of long-term forest tree breeding are discussed. Finally, operative aspects of selection of plus trees, seed orchards, mating design and observed gains in tree breeding are presented.

What should be considered before the start of a breeding programme?

Several aspects both genetic and non-genetic, must be considered before a breeding programme is established. First of all the objective(s) of the tree breeding programme must be identified. In many countries it might seem self-evident that the objective is to produce the raw material for saw mills, pulp and paper industry. Less evident is that breeding might be focused on production of material for amenity forests. In Iceland, which was once covered with much larger forests than today, there is a great interest in extending the forests to non-forest land. The use of forests for prevention of erosion is another breeding objective. Related to this is the use of forests as lee plantations. Christmas tree cultivation and street tree improvement are of economic importance in several countries (Pictures 9-1 and 9-2). As a consequence of the varying objectives of forest plantations different selection criteria must be used to build up breeding populations that will meet the different objectives.

Of greatest importance is of course the economic value of the products obtained from the tree species included in tree breeding activities. This value must be weighed against the investment in staff and materials that are required. If it is assumed that the species even on a long-term basis will have a considerable economic value it is motivated to plan for long-term breeding. All around the world there are many long-term breeding programmes.

In Figure 9-1 different intensities in the improvement of ornamental trees and shrubs in Sweden are presented. At the lowest intensity only identification of good seed stands takes place. For Norway maple, which is of great importance for various urban plantations and landscaping, there is an economic incentive for the establishment of seed orchards.

Of primary interest in any breeding programme is to decide which traits should be improved. The more traits that are included in the improvement programme, the harder the breeding activity. If we assume that one tree per 100 is carrying a trait and the traits are uncorrelated, one million trees (1/100 x 1/100 x 1/100) are required to find one tree with the desired combination of all three traits. If the traits are positively correlated the tree with the desired


traits might be found among a lower number of trees.

When the traits for improvement have been identified, it is important to estimate the genetic variation in these traits and the mode of inheritance of each trait. Estimates of the additive variance or the additive coefficient of variation are important, since the additive variance can be exploited in mass selection. When the additive variance is known, we can calculate possible genetic gains. If the proportion of non-additive genetic variance is considerable, the breeding becomes more complex.

Knowledge of flowering biology is a prerequisite for a successful breeding. Without flowering no breeding can be carried out, and it is important to know the conditions that promote flowering. This is probably easier to trace in the boreal and temperate zones with their changes of seasons than in the tropics.

Flowering phenology, *i.e.* the occurrence of different phases of flower development over time, should also be determined in order to be able to predict the probability for matings within seed orchards or in other plantations aimed for seed production. Pollen dispersal is a factor of great significance for predictions of matings with pollen from unbred forests in the surroundings of the seed orchard. Contamination with unbred pollen generally reduces the genetic gain in the seed produced in proportion to the amount of contamination (Figure 9-2). In certain cases it is more serious as will be discussed later on in this chapter.

Norway spruce and Scots pine and many other conifers carry both female and male strobili on the same tree. Other species such as ash and aspen are monoecious and usually carry one sex only. Theoretically, selfing may occur in many tree species. As is evident from Chapter 5 selfing is mostly accompanied by a pronounced inbreeding depression. The North-American red pine and yellow cedar are exceptions to this. Certain species such as the birches have self-sterility alleles which prevent selfing. A tree with the self-sterility alleles s1 and s2 does not form any seeds if the pollen grain contains either of the alleles s1 and s2. It does not matter whether the pollen originates from the same tree or another tree; the female tissue prevents fertilization with pollen containing these alleles. Conifers do not seem to have self-sterility alleles. Instead they have varying numbers of lethal alleles. Besides, polyembryonic embryos are frequently formed in conifers. This means that there is frequently competition between embryos such that only one forms a viable embryo in each seed.

If a decision is taken that the breeding should be of long-term character it is important that there is a stable tree breeding organisation that lasts for decades. Without such stability there is a high risk that short-term problems are given priority at the expense of long-term and perhaps less glamorous tasks.
Various types of tree breeding

Forest tree breeding may be structured in many ways, one of them is shown below.

Selection
- species level
- provenance level
- population - stand level
- individual tree level

Breeding to combine desired traits

Polyploidy breeding

Breeding using mutations, molecular markers, and genetic engineering

Generally, breeding aims at combining useful traits from different parents via matings among them. This is followed by selection of the best performing trees in the progeny. Selection at either levels without crossing can hardly be regarded as breeding in a strict sense. In spite of this we treat introduction of exotic species in this chapter. Provenance research was extensively treated in the Chapter 7. As was stressed in that chapter, forest tree breeders hardly distinguish between populations and provenances when seed is collected in natural forests. A major focus in the rest of the chapter is on selection of individuals or plus trees, i.e. trees with desirable phenotypic characteristics (Picture 9-3), and how matings among plus trees should be done to improve breeding. Before coming into species and plus tree selection and their breeding we will briefly comment on polyploidy breeding and mutation breeding.

Polyploidy occurs frequently in higher plants and has played an important role in agricultural plant breeding. In several cases the polyploids in a genus are larger than the diploid species of that genus. Polyploidy also had a leading role for establishment of the Swedish forest tree breeding. In 1935 a famous wheat breeder, Herman Nilsson-Ehle, detected a giant aspen tree in a forest in southern Sweden. It turned out that this tree was a triploid. Nilsson-Ehle envisaged polyploidy breeding of Norway spruce and Scots pine in Sweden to obtain giant trees of these species. He convinced influential foresters that Sweden ought to have an organised tree breeding, and there has been such an organization since 1936. One of its first tasks was to produce polyploids of Norway spruce and Scots pine. Triploid trees of these two species did not grow into giants but rather they were dwarfs. Different genera have different ploidy optima. Certain grass species have their optima as hexaploids while the optimum for Norway spruce and Scots pine is evidently at the diploid level.

Mutation breeding raised great expectations during the 1950s and 1960s. These expectations were mainly linked to the hope that certain chemicals would bring about mutations at particular loci. Mutation breeding has the best prospects in highly bred crops, in which breeders might be interested in a change at one locus. If this could be achieved the breeders do not need to use the labour demanding back crossing over 7 - 8 generations to transfer one specific allele into the crop. Back crossing means that the original parent is used as one mating partner over several generations and selection for the desired trait takes place in each generation (Fig. 9-3). Owing to the long generation time of forest trees the back crossing technique is hardly possible. Mutation breeding is of little or no value for most forest trees, though it has had some importance in changing flower colours in ornamental plants. In many respects mutation breeding and allele transfer via gene technology are similar. One difference is that a modern molecular geneticist knows which allele he/she transfers to a recipient whereas the induction of mutations is brought about blindly.

*Picture 9-3. An excellent plus tree of Eucalyptus grandis growing in Australia. Photograph Gösta Eriksson*
Species selection

In Scandinavia the flora is poor owing to the short time since the last glaciation. This is pronounced for forest trees. In consequence we may not have the tree species which would give the best yield. It is motivated to compare the performance of domestic trees with the performance of exotic tree species. At the start of tree cultivation in developing countries it is useful to evaluate which species should be included in breeding programmes. For this information, species trials are required.

When establishing species trials it is urgent to carry out a careful selection of the provenances that should be included. An idle selection of provenances may cause misleading results as illustrated in Figure 9-4. In this graph the true production of different provenances of species A, B, and C is given over an environmental gradient. In Fig. 9-4 the provenances tested are shown as filled symbols. Among the tested provenances, the one coming from test locality 4 belonging to species B gives the best test result. Since we know the true production we also know that the best production can be brought about by species C. However, the proper provenance of species C was not tested. If there had been only the three provenances marked with arrows in the species trial, the result would have been still further away from the truth.

The conclusion that might be drawn from Figure 9-4 is that species trials must have several provenances of each species to be meaningful. If a species has not shown a maximum such as is the case for species C in the graph it may be questioned whether we have complete information on the ranking of the species. From the previous chapter it is evident that Norway spruce and Scots pine show pronounced clines, which is why we expect that introduced species originating from climatic conditions similar to the Scandinavian also show clinal variation. If the knowledge about provenance differences in a domestic species is as good as it is for Norway spruce and Scots pine in Scandinavia it is easy to select the provenances of a domestic species for species trials. In such situations one or two provenances might be sufficient. A larger number of provenances of the exotic species that should be tested ought to be selected. They should be selected from areas with similar climate and edaphic conditions to those of the test area.

In summary, species trials require large test plantations at more than one locality. The experimenter requires great intuition and skill to select the proper test localities and provenances to be included in the experiment. Only then we can expect to get accurate information about the potential of different species.

Figure 9-3. The theoretical contribution of the A parent to the offspring over several generations in a back crossing programme.

Figure 9-4. Illustration of the importance of having several provenances of each tree species in species trials. For further information see the text.
Large test plantations mostly mean that it will be hard to find sufficiently uniform localities. One way to overcome this problem is to employ a two-step strategy. In the first step growth rhythm studies are carried out in greenhouses or nurseries. After evaluation of data from the first step there ought to be information on the provenances that have potential for a certain test locality. Such tests are of particular value when frost damage significantly interferes with growth. In Scandinavia frost damage occurs frequently. Large savings may be achieved by running a first step species trial in a nursery before the costly field trials are established. Thanks to this the test plantations do not need to be as large as they would have been without this pretesting. The probability of finding small homogeneous test plantations is much larger than of finding large ones. Thus, there are several advantages with this two-step strategy.

### History

The history of the Swedish forest tree breeding will be used to illustrate some of the thinking in many breeding organisations during the early stages of tree breeding. Even if an organised tree breeding was established in 1936 it took until the mid 1940s before large scale selections of plus trees took place. Scions were collected from the plus trees and grafting was done. The grafts were later planted in seed orchards for commercial production of seed. For further improvements, crosses were carried out between the plus trees in the seed orchards. The progenies were raised and planted in progeny trials. Normally such a trial contains several progenies. Progenies are frequently referred to as full-sib or half-sib families depending on the type of cross used to create the progeny. One major objective of the progeny trials is to estimate the genetic value of the parental trees. Thus, the parental tree genetic quality is revealed by its offspring in well designed experiments. Parents are selected for new seed orchards based on the evaluation of the progeny trials. Such a selection cannot be carried out more than once or twice since we soon reach a situation where there are no more parents to select among and the number of trees in the breeding population would not reach a satisfactorily large $N_e$ (Figure 9-5). Gradually it became evident that the best trees in the best families had to be selected. Scions are then collected from these trees for establishment of the second generation seed orchards. When progeny-tested parents are used for establishment of new seed orchards, American forest tree breeders call them one and a half generation seed orchards.

Another way of mitigating the reduction in number of trees in the breeding population is to select plus trees in unbred native populations. However, this means that the gain achieved from earlier selection will be lost to a certain extent (see Figure 9-6). The loss is proportional to the portion of unbred material which is inserted into the breeding population. If only 50% of the trees have passed previous selection and breeding, the gain will drop to half of what is possible if the most advanced bred material is used. Insertions from the wild become less attractive the higher the degree of breeding. The same can be said about pollen contamination from the wild.

### Figure 9-5

*Figure 9-5. Recurrent selection of a certain fraction of parents makes additional selections after a couple of generations impossible. To have satisfactory size of the population, selection must be carried out in the progeny.*

### Figure 9-6

*Figure 9-6. Inclusion of unbred material in the breeding population after a few generations of breeding causes a drastic reduction of the genetic gain.*
Long-term breeding

Long-term breeding might be envisaged as a cyclic course of events, in which crossings, establishment and evaluation of progeny trials, selection of trees/plants for the next generation of the breeding population based on the evaluation, and planting of grafts of the selected trees are the components of this circle (see the left part of Figure 9-7). For each completed cycle the material for cultivation has been improved. Theoretically we have three options for exploitation of the improvement in the breeding population. We may establish seed orchards for seed production, establish clonal archives for production of cuttings, or produce plants via tissue culture.

Population functions

To enable an understanding of breeding and its consequences for genetic erosion it is important to clarify the function of different populations which might be distinguished in Figure 9-7 (See Box 9-1). The core of a breeding activity is the breeding population which is to be found in the cyclic part of Figure 9-7. Seeds from seed orchards or vegetatively propagated plants from clonal collections constitute the production population, i.e. they are the starting material for wood producing forests, if the production of wood is the breeding objective. Generally the production population is the population that should produce human utilities whether it is biomass or beautification. The starting material for the production population is the propagule population. Seed orchards, clonal archives, and plant material for tissue culture propagation are all components of the propagule population. It should be noted that one and the same seed orchard simultaneously might function as a breeding population and a propagule population. In the former case the seed orchard is used for crossings, the resulting seeds giving rise to seedlings, which are established in progeny trials. The role of seed orchards as propagule populations is fulfilled when seeds are produced for sowing in nurseries or for direct seeding in forests.

To guarantee a sustainable gain in the breeding work a high additive variance is required in the breeding population. In Box 9-2 it is illustrated schematically why in a long-term perspective it might be a disadvantage to have few trees in the breeding population. We can have a lower genetic variation in the production population without loss of cultivation security.

Box 9-1 Functional types of populations

Breeding population: the collection of trees that will carry the advancement of breeding into future generations

Gene resource population: the seeds, acorns, nuts, plants, or trees that are included in the gene conservation

Production population: the trees that shall produce human utilities

Propagule population: the trees or plants utilized in sexual or vegetative propagation

Figure 9-7. The principle of forest tree breeding with estimation of breeding values, selection, matings and progeny trials in a cycle as well as generation of material in three different ways for the production forests. (Somewhat modified from an idea by Öje Danell.)
In multiple generation breeding of crop plants three different types of recurrent selections have been applied: simple recurrent selection, recurrent selection for general combining ability, and reciprocal recurrent selection. Recurrent selection is used when something is repeated over and over again in a cyclic way as is illustrated in Figure 9-8. The principles of simple recurrent selection to the left and recurrent selection for general combining ability to the right. OP = open pollination.

9-7. For the most complex recurrent selection it is difficult to illustrate the different components in a cyclic way. Therefore, we prefer to show all three types of recurrent selection as linear flow charts to facilitate comparisons among them.

Simple recurrent selection is not any intensive type of breeding (Figure 9-8). The seed from seed orchards is used to raise seedlings in nurseries or for direct seeding in forests to establish a production population without any pedigree. The best trees in the production population are selected for establishment of a new generation of seed orchards and the process is repeated again. When funding for breeding is limited this is one option that can be used.

In recurrent selection for GCA, matings are carried out for establishment of progeny trials. Open-pollinated seed is used to establish a production population. The progeny trials are evaluated and the best trees in the best families are selected and crosses among them are carried out. The
Reciprocal recurrent selection is the most complex form of recurrent selection. Since it is mostly used in species hybridization we have illustrated it for such a case. Based on the evaluation of progeny trials of the two species, trees are selected for interspecific matings. The hybrids obtained are used for establishment of interspecific progeny trials. The data from this type of trials are used for selection of parents which give good interspecific hybrids. Thus selected trees are used for establishment of the seed orchard that should produce the seed for the production population. The selected parents are also used for interspecific matings to generate material for interspecific progeny trials. Selection of parents are then carried out in these interspecific progeny trials. These parents are used for interspecific mating and the process is repeated again. As may be seen from Figure 9-9 in this type of recurrent selection it takes two generations to obtain the seed for the production population. For this reason it is not much used in forestry. In South Korea two north American pines, *Pinus rigida* and *Pinus taeda*, were introduced for hybridization. The former species is hardy but has a bad stem form. The latter species does not have a satisfactory hardiness for this part of the world but has an acceptable stem form. Therefore, efforts are taken to combine hardiness and growth form in the interspecific hybrids. The breeding programme for this interspecific hybridization is carried out according to reciprocal recurrent selection. This is a typical case of species hybridization used to combine two good traits from each of the parental species.

It should be stressed that Figures 9-8 and 9-9 show the principles of the three types of recurrent selection. In practice modifications of them occur.

### Multiple Population Breeding System

One of the major problems in breeding is that the high priority breeding objectives of today may be of limited value when it is time to harvest the gains from tree breeding. Another factor of great uncertainty is that the environmental conditions may change dramatically over a rotation time of 50 - 150 years. Changes of the reforestation and silvicultural methods will take place with high probability over such a period. To this must be added the environmental change, which to some extent is beyond human control. Today when climatic change is probably a fact, the forest tree breeder faces great problems. Unlike the cultivation of cereals there is no possibility to change cultivars every or every second year. An effective forest tree breeding programme ought to be designed such that it matches the future changes in breeding objectives and environmental change. The American forest geneticist, Gene Namkoong, developed a breeding concept that essentially fulfils these
requirements. His concept means that the breeding population is subdivided into smaller subpopulations instead of being kept as one big breeding population. His concept of breeding is called the Multiple Population Breeding System (MPBS). The subpopulations are preferentially planted over a broad span of site conditions (cf Fig. 9-10). The target trait might be the same in all subpopulations or a stronger emphasis might be given to stem quality rather than biomass production in some of the subpopulations. The MPBS means that disruptive selection takes place among subpopulations. The MPBS concept is adopted in the Swedish breeding programmes for silver birch, Lodgepole pine, Norway spruce, and Scots pine. As is seen from Figure 9-11 different subpopulations will be distributed to various combinations of temperature and photoperiodic conditions. A world-wide inventory during 1999 showed that the MPBS concept is adopted in many breeding programmes.

Each subpopulation should have 50 trees, which may seem a low number. If the entire breeding population has 20 subpopulations the \( N_e \) becomes much larger than 50. At such an effective population size there are few alleles lost for random reasons unless they are extremely rare. In Figure 9-12 the minimum number of trees required to save one rare allele per locus is illustrated for three different cases; one rare allele in each of 10, 50, or 100 loci. As is seen from this figure the allele frequency plays a greater role for the minimum number of individuals that ought to be saved than the number of loci with rare alleles. To be sure that alleles at a frequency of 0.01 and higher will be saved only a few hundred trees are required.

The inbreeding that may take place at \( N_e = 50 \) trees amounts to 1 % (\( F = 1/2N_e \)) and will not cause any in-breeding depression of importance. In various breeding programmes with other organisms than forest trees sustainable gains over 50 generations have been obtained at population sizes lower than 50. If the selection of plus trees considers the adaptation that might have taken place under different site conditions, the probability of including rare alleles increases. An allele might be rare at species level but be more frequent in a subpopulation of the species thanks to its contribution to fitness in this subpopulation.
With subdivision of the breeding population it is no longer the case that only one population passes through the circle in Figure 9-7, rather each subpopulation passes through the circle. The speed with which the subpopulations pass through the circle will probably vary depending on the site conditions or the breeding goals of the individual subpopulations. In all subpopulations the larger part of their additive variance will be kept while the additive variance among subpopulations will increase. This is an ingenious system which in its simplicity guarantees an increased additive variance and which simultaneously offers possibilities for changes of breeding goals. Finally, recurrent selection for general combining ability is mostly used within each subpopulation.

In summary the main advantage of the MPBS is that it combines the capture of the total existing genetic variation with a satisfactory variation within each subpopulation and that it allows the target populations to adapt to the prevailing environmental conditions. Another advantage is that the speed of evolution might be faster in a population of 50 trees than in a large population containing thousands of trees.

Sublining

The above described MPBS should not be confused with sublining which is also a subdivision of the breeding population but in this case it is targeted for one breeding goal. The purpose of sublining is to avoid inbreeding in the production population. This is accomplished by selection of one clone from each subline for establishment of seed orchards for production of commercial seed. Inbreeding is in this case permitted in each subpopulation. The reason for launching this concept was that it was feared that it would not be possible to avoid inbreeding in the breeding population in a long-term perspective. This fear is probably exaggerated, at least in breeding populations with several hundred trees. It should be noted that sublining does not aim at an increase of the among-population additive variance which is in contrast to the MPBS concept.

Dag Lindgren has developed the concept of status number, which can be interpreted as the size of a population comprised of unrelated trees. Status number has given breeders a possibility to estimate the unrelatedness in the breeding population.

Nucleus breeding

Another system for long-term breeding which has been applied in certain programmes is to split the breeding population into subpopulations of unequal size. The smaller nucleus contains 30 - 50 trees while the larger part keeps 300 - 400 trees. The most intensive breeding occurs in the nucleus, which has given the name nucleus breeding to this system (Figure 9-13). The objective is that gene conservation and long-term gain will be guaranteed in the larger subpopulation while the breeder profits from the larger gain that may be obtained in the smaller subpopulation. As is evident from Figure 9-13, the difference between the two subpopulations will increase over the generations and it will be tempting to concentrate all breeding efforts to the nucleus only. In some programmes which apply this system a transfer of material from the larger subpopulation to the nucleus is envisaged. The fear for
inbreeding is also in this case the reason for the latter suggestion. However, it should be remembered that genetic gain is lost when material is taken from a lower level of breeding to a higher (Figure 9-6). As is the case for the MPBS method recurrent selection for general combining ability is mostly used within the two populations.

**Short-term breeding**

Whichever type of breeding that is selected, it may be complemented with intensive breeding under a few generations to identify clones for elite tree seed orchards. Figuratively it can be seen as a means to skim the cream off the milk at the cost of narrowing down the additive variance. In principle it differs from long-term breeding in the number of trees included in the breeding operation and in that there is no long-term intention in this operation. The latter is a contrast to the nucleus breeding in which the elite part of the population is aimed for long-term breeding.

In reality breeding programmes may be more complex than the pure forms indicated in the above sections. In the breeding cooperative in New Zealand a new breeding strategy for *Pinus radiata* was launched in the late 1990s. It has elements of MPBS, nucleus breeding and sublining. It consists of two sublines, which are referred to as superfamilies. Each superfamily consists of a nucleus with a large main population and 7 subpopulations with separate breeding goals in accordance with the MPBS concept. The breeding goal of the main population is general improvement of growth and quality traits. The breeding goals in 6 of the 7 subpopulations are:

- High wood density
- Structural timber, *i.e.* strong stiff, stable timber with small knots and low spiral-grain angle
- Clear cuttings, *i.e.* “clear wood from unpruned trees”
- Long internodes, good growth rate
- Good growth rate, low *Dothistroma* infection
- Excellent growth and form

The 7th subpopulation consists of introduced material from the Guadeloupe island. The breeding goal of this subpopulation comprises most of the goals in the other 6 subpopulations.

As seen from the list above, some subpopulations have breeding goals mostly related to timber quality, while others are related to good growth or resistance to *Dothistroma*, which is a serious disease in certain parts of the country.

To make the story still more complex, breeding in each subpopulation might be regarded as short-term breeding.

The main population is seen as a “genetic insurance”. In addition to this there are specific gene resource plantings in New Zealand of the five existing provenances as well as a large number of clones in archives.
Selection of plus trees

In most intensive breeding programmes the scrutiny of plus tree candidates was rigorous during the original selection of plus trees. One problem was that the selection mostly took place in stands originating from self regeneration. In such a stand the quality development is different from the development in a planted stand in which the bred material will grow. Planted stands usually have a much wider spacing than in naturally regenerated forests. Another problem with selection of plus trees in stands close to the end of the rotation time is that imperfections in the most valuable part of the stem might be hidden inside the trunk. In connection with a new selection of plus trees in Sweden during the 1980s it was decided that the new selections should preferably take place in planted stands which had reached a third of the rotation time (Picture 9-4). Thereby it was possible to carry out stem quality selection based on the economically most important part of the stem.

During the first plus tree selection, the growth of the plus trees was compared with the growth of the tallest trees in the same stand. Wood cores were taken to enable a correction for different ages of these comparison trees and the age of the plus tree. In spite of this it is difficult to carry out an unbiased selection in uneven-aged stands.

Seed orchards

Seed orchards might be classified in different ways. One way is to distinguish between seedling seed orchards and clonal seed orchards. Another classification takes into account the type of material included in the seed orchard. There may be clones from two species, interspecific seed orchards, from two provenances, interprovenance seed orchards, or finally the clones may originate from one provenance, intraprovenance seed orchards.

Seedling seed orchard

This type of seed orchard is usually established as a progeny trial with seedlings raised from open pollination or controlled crosses. The aim is to use the best trees in the best families as seed producers. One disadvantage with this type of seed orchard is that progeny trials rarely stimulate an abundant flowering. In many cases seed orchards are located in warmer climates to stimulate flowering. If the progeny trials are located in another climate this may result in a selection of wrong clones owing to genotype x environment interaction. If this interaction occurs the best trees in the best families in the progeny trial are not identical with the best trees in the best families in the climatic zone in which the seed should be used.

Another disadvantage with seedling seed orchards is that flowering usually starts later in this type of seed orchard than in grafts in clonal seed orchards. Therefore, this type of seed orchard is most suitable for tree species with an early flowering. For species in which there are problems with union of the scion and the root stock it may be necessary to use seedling seed orchards. This kind of problem is designated as grafting incompatibility.

For sanitary reasons, seedling seed orchards of Pinus contorta were established since imports of scions of this
Table 9-1. Different types of clonal seed orchards and their application in Scandinavia

<table>
<thead>
<tr>
<th>Type of seed orchard</th>
<th>Application in Scandinavia</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interspecific</td>
<td>For production of hybrid larch <em>Larix decidua x L. leptolepis L. decidua x L. sibirica</em></td>
<td>Species frequently have non-overlapping receptivity and pollen dispersal</td>
</tr>
<tr>
<td>Interverprovenance</td>
<td>Most old <em>Picea abies</em> seed orchards were of this type with varying number of Scandinavian and continental clones</td>
<td>The objective is to obtain provenance hybrids. Under the most favourable conditions 50% hybrids may be obtained, the others are the result of matings among clones within each of the two provenances</td>
</tr>
<tr>
<td>Intranprovenance</td>
<td>This is the dominating type of seed orchard for <em>Picea abies</em> and <em>Pinus sylvestris</em></td>
<td>Only the general combining ability can be exploited</td>
</tr>
<tr>
<td>Biclonal</td>
<td>Exists only for research purpose</td>
<td>Strong prerequisite on isolation to avoid contamination from surrounding stands</td>
</tr>
<tr>
<td>Monoclonal*</td>
<td>Does not exist</td>
<td>Successful mass pollination without preceding isolation of female strobili is a prerequisite</td>
</tr>
</tbody>
</table>

* When such a seed orchard is pruned to a maximum height of 3 metres it is known as a hedge seed orchard

species to Sweden is not permitted. In order to have an approximately even spacing after thinning, progenies from one female are planted in groups with a denser spacing within groups than between groups. The intention is to save one tree per group based on phenotypic examination. To stimulate flowering, the *Pinus contorta* seed orchards are located south of the climatic zone in which the seed should be used. Parallel to the establishment of seed orchards, progeny trials were established in the zone in which the seed should be used. This guarantees that the best families are selected while the selection within family has to be carried out in the seedling seed orchard outside the zone of cultivation.

Clonal seed orchards

In Table 9-1 clonal seed orchards are classified and characterized. Whether or not they are applied in Scandinavia is also indicated in the table.

Biclonal or monoclonal seed orchards are only of interest for progeny-tested clones. If the disadvantages mentioned in Table 9-1 can be avoided, the largest gains may be obtained from these two types of seed orchard. At the turn of the century only a few biclonal seed orchards existed in Scandinavia.

The early breeders were aware of the problem that not only interspecific or interprovenance hybrids were obtained in interspecific and interprovenance seed orchards, respectively. In the early days of tree breeding the labour cost was several times lower than presently and the breeders counted on some manual seedling classification in nurseries with culling of all non-hybrid seedlings. This does not seem to be possible either from a biological or economic point of view. In some instances one species or one population was represented by one single clone. Such a clone is used as a female parent and cones are harvested from this clone only.

Intraprovenance seed orchards are the most suitable type of seed orchard for newly selected plus trees. At planting of the grafts the breeders aim at a maximum distance between grafts of the same clone. This is done to reduce the probability for selfing. Another condition is that the best possibilities for random mating in the orchard should exist. Biclonal and monoclonal seed orchards are interesting alternatives for future, intensive breeding with artificial mass pollination.

Scots pine seed orchards have generally been successful with respect to their role as propagule population, i.e. to produce seed for production populations. Many conventional seed orchards have limitations, which means that the gains that ought to be obtained from a theoretical point of view are not obtained. Many Norway spruce seed orchards in Scandinavia were not properly located, resulting in reduced flowering. To obtain good flowering in Norway spruce high temperatures are required at the time of bud initiation, which takes place one year before flowering. To have the Norway spruce seed orchards as far away as possible from Norway spruce stands, farmland was preferred for location of seed orchards in Sweden. The reason for this was to avoid pollen contamination from stands as much as possible. However, the occurrence of cool winds makes the local climate unsuitable for flower induction.

The female strobili of Norway spruce appear in the apical part of a twig which prevents further vegetative develop-
Based on a detailed analysis in a large number of seed orchards in Scandinavia, the ideal location of a seed orchard was developed (Box 9-3). Before locating a seed orchard it is urgent to clarify the local climate of the candidate locality in order to get flowering at all.

After effects

During the early 1980s worrying reports on the poor hardiness of seed orchard progenies were published in Norway. The seed material from clones growing a few degrees of latitude south of their origin had a longer growth period and reduced hardiness compared to the progenies from the same clones in the original stands. This phenomenon was called after effects. An earlier term for this phenomenon is preconditioning. One example of after effects is illustrated in Figure 9-15, in which the development of budset in three materials is given. The first material is the local Norwegian material from latitude 63°50', the second has seedlings from the exotic Harz population growing at the same locality, finally there are seedlings of Harz origin coming directly from Germany. It is clear that there is a dramatic difference between the two Harz materials. The performance of the Harz seedlings coming from the harvest in Norway behave in an almost identical way to the domestic Norwegian seedlings harvested at the same latitude. Natural selection must be ruled out as an explanation since the hybrids between Harz and Norwe-

![Female strobili of a Norway spruce graft. Note the apical location of the strobili. Photograph Kjell Lännerholm.](image)

*Figure 9-6. Female strobili of a Norway spruce graft. Note the apical location of the strobili. Photograph Kjell Lännerholm.*

ment of the apical part of a strobilus-carrying twig (Picture 9-6). This means that there are no possibilities to have abundant flowering in the same Norway spruce tree in two consecutive years. The flowering in Norway spruce seems to be cyclic (Figure 9-14) with a good flowering year followed by one or a few poor flowering years.

![Figure 9-14. The number of days with a temperature > 20°C during flower initiation and mean flowering next year in a clonal trial of Norway spruce at latitude 59°30'.](image)

*Figure 9-14. The number of days with a temperature > 20°C during flower initiation and mean flowering next year in a clonal trial of Norway spruce at latitude 59°30'.*
gian parents should have a mean value half-way between their two parents, which they evidently did not have.

Great efforts have been devoted to this phenomenon of after effects in Norwegian forest genetics research. A systematic testing of temperature and photoperiodic conditions has indicated that it is the temperature conditions from the proembry stage to the mature seeds that is critical for the change of the growth rhythm. The explanation may be that signals from the environment give an imprint on the female genome such that certain genes are expressed. A signal at a southern locality would thus cause a southern behaviour of the progenies produced at a southern locality and conversely a northern locality would cause a northern behaviour. One hypothesis under testing (2006) is that methylation of DNA is the mechanism behind appearance of after effects.

After effects may be due to a purely physiological effect. Such an explanation is based on the fact that growth in trees and shrubs is dependent on the current conditions as well as conditions during previous years. One example of this is illustrated in Figure 9-16. As is evident from this illustration the plant to the right is smallest owing to the longer nights it was exposed to during its first growth period. The right plant continued to grow less than the sister plant that had the shorter night during the first growth period. It should be noted that the plants had the same photoperiodic conditions during growth periods 2-4. The influence from the first growth period remained even during growth period 6. The plants seemingly had a memory mechanism. An example from Norway spruce of such a memory is presented in Fig. 9-17. The different combinations of photoperiod, (continuous light and 8-10 hours of night, respectively) and temperature (heat sum in degree days, dd, 1000 and 2000, respectively) during seed maturation resulted in variation in growth rhythm. As seen from Fig. 9-17 the combinations high temperature + long night (upper left) and low temperature + continuous light (lower right) had later growth cessation than the other combinations. The late growth cessation was accompanied with the largest frost damage.

One explanation for the "memory" in trees is that much of the growth for the next season is programmed in the bud. The growth that we observe is actually an elongation of already formed stem units. In the light of this the lower hardiness of the southern progenies might be explained by the longer time for formation of stem units at the southern locality. In consequence their elongation takes a longer time.
than is the case for the northern material. As a corollary of this, budset and hardiness take place later during the season in material matured at southern than at northern localities.

Of greatest significance for understanding of the phenomenon of after effects is to study the progenies to the material with changed behaviour, *i.e.* northern behaviour of southern origin and vice versa. A final proof will not be obtained until progenies are raised from the trees that had unexpected behaviour as young seedlings. If their progeny keep the character it will be a proof.

To determine which explanation is true is not of purely academic interest but of great practical significance. Especially in Scandinavia, seed orchards were located south of the area in which their progenies will grow. This was done to stimulate flowering and secure good seed development as touched upon previously. If after effects are of genetic nature the seed from seed orchards located far outside the zone of cultivation cannot be used as intended. If after effects are of physiological nature it is fairly simple in modern nurseries to programme the cultivation conditions such that the problem with long growth period and late hardening is overcome.

![Picture 9-6. Above. Clonal rows of cuttings in a nursery in Escherode in Germany. Photograph Gösta Eriksson.](image)

**Vegetative propagation and clonal forestry**

All cuttings produced from one donor plant (= ortet) as well as all cuttings produced from them belong to a clone. All plants/trees of a clone are genetically identical. If a tree breeder has identified a super tree it is tempting to multiply it vegetatively on a large scale and market it. Many ornamental plants, berries, and fruit trees are vegetatively propagated and are marketed as individual clones. From a genetic perspective a vegetative propagation means that not only the additive variance is exploited but also the non-additive variance (Figure 9-18). The great majority of results from conifers suggest that there is not much non-additive variance to exploit in traits of interest to improve. However, exceptions do occur.

Another reason for breeders to use vegetative propagation is for mass propagation of valuable families. In species such as Norway spruce with its irregular flowering it might be useful to propagate the plants of families obtained from crosses between parents with high breeding values. Artificial crosses are also motivated owing to the high degree of pollen contamination (see page 143) in seed orchards. Contaminations reduce the gain considerably in conventional seed orchard seed.

Still another reason for vegetative propagation is to use the material in the evaluation of parents (Box 9-4). One great advantage in this case is that one genotype can be tested under several different environmental conditions. For species which are easy to propagate vegetatively, such tests are in operation in some breeding programmes. Simulations have shown that the gain might be increased considerably by clonal testing compared to ordinary progeny testing with sexually propagated material.
Following mating, the mature or immature embryos are induced by the hormones cytokinin and auxin to produce somatic embryogenic cultures. From each genotype, some of the embryogenic cultures are cryostored in liquid nitrogen. Others are further treated with the hormone auxin in order to produce mature embryos. From these embryos, plants are regenerated for testing in the nursery and field. After evaluation and selection of the best genotypes, new plants are regenerated from the cryostored embryogenic cultures for establishing production populations in the forest. Plants are also regenerated for use in a new breeding cycle.
There is a general public fear that clonal forestry is risky since clones might be attacked by pests or diseases. Several theoretical analyses have been carried out. They all show that 30 - 40 clones give the same or better cultivation security than much larger numbers of clones.

Even if there are no attacks from pests or diseases we might expect that a clone that grows very well under certain site conditions might perform poorly under other site conditions. The reverse might be the case for another clone. Therefore, clones or clonal mixtures should require more rigorous testing to avoid losses in commercial plantations than is required for ordinary seed lots from stands. The latter are assumed to be buffered by their broader genetic variation. The results from young field trials are somewhat contradictory. One series of trials indicated significant clone x site interaction while another did not. Calculation of ecovalence values is a statistical method to estimate the percentage contribution of a clone to the clone x site interaction. The larger the ecovalence of a clone the more it contributes to the interaction. In Fig 9-19 the distribution of ecovalence values for one of the series of clone trials mentioned above are presented. This series has 11 field trials in Denmark and southern Sweden and it has 96 clones from four provenances. Although there was a strong clone x site interaction none of the clones contributed more than 3 % to the interaction (Fig. 9-19).

Vegetative propagation of Cryptomeria japonica has taken place in Japan for centuries and extensive reforestation with this species occurs (Picture 9-8). Many poplars and willows are easy to propagate vegetatively and are used in the production population. The so-called energy forestry with willows in Sweden relies on vegetative propagation of outstanding clones. Some of the most productive forests in the world consist of Eucalyptus clones. At the turn of the century approximately half of all planted Eucalyptus forests consisted of clonal plantations. There are thus several examples from all over the world in which vegetatively propagated material is utilised in the production population.
Progeny testing and mating design

Progeny testing plays a major role in forest tree breeding, above all to identify parents with good general combining ability. Selection of parents based on data from progeny tests is usually designated as selection backward. Estimates of variances is another objective of progeny testing. Such estimates are used for future breeding and for prediction of possible gains from tree breeding. Finally the progeny trials are sources for selection of trees for a new generation of the breeding population. Such a selection is designated selection forward, i.e. the best trees in the best families are selected.

There are three main types of mating design:

- **Diallel matings**
- **Factorial matings**
- **Nested matings**

In addition, **polycross** and **open pollination** may be used. The meaning of the different types of mating design is given in connection with the presentation of their advantages and disadvantages. All artificial mating work is labour demanding and thus expensive. It is important to clarify the objective of the mating work before it is decided which mating design should be used.

The meaning of diallel mating is that the parents serve both as female and male (Figure 9-20). If we want to have total information about the genetic quality of a set of trees the best thing to do is to carry out all possible crosses among all parents, i.e. 1x2, 2x1, 1x3, 3x1 etc. With this mating design we shall theoretically obtain the best estimates of additive and non-additive effects as well as selfing and reciprocal effects. The progeny plantation from such a mating design is also the best for selection of the best trees in the best families. This is the only mating design in which all families are present. The major disadvantage with the complete diallel is that it becomes cumbersome when the number of parents is high. If we assume that 50 trees should be progeny tested, a complete diallel mating requires 50 x 49 = 2,450 crosses of the trees. Selfings are not included in this figure. Both mating work and field trials will be too large to make this mating design realistic in applied breeding. Another thing that is frequently overlooked in connection with choice of mating design is that mating designs with large number of families require a large homogeneous area of forest land. Mostly it is hard to find forest land larger than 2-3 hectares with a satisfactory homogeneity. A complete diallel mating with 50 parents at a spacing of 2 x 2 meters would only allow 3 plants per family in a field trial of 3 hectares. To avoid the requirement for large progeny trials it is necessary to reduce the number of families. Such reductions are called **partial diallel matings**.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>4</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>5</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>6</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>7</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>8</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Figure 9-20. A complete diallel mating design without selfing.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>4</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>5</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>6</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>7</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>8</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Figure 9-21. Mating design described as half-diallel.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>4</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>5</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>6</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>7</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>8</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Figure 9-22. Partial diallel mating design according to Kempthorne and Curnow.
The largest with respect to remaining families after reduction of the complete diallel mating design is the **half diallel** (Figure 9-21). As the name says half of all possible matings are carried out. Mostly this is done by excluding the reciprocals. It is assumed that maternal effects can be neglected. A partial diallel that has frequently been used is the one in Figure 9-22, which also excludes selfings. This type of mating is a good compromise between different objectives in progeny testing, such as identification of parents with good general combining ability, estimates of variance components and possibilities for forward selection.

**Factorial mating** means that a parent either serves as female or as male. When a factorial mating design has a few male clones and numerous females it is designated as **common tester** mating (Figure 9-23). The major advantage of common tester design is that the estimates of female GCAs are fairly accurate. The number of unrelated families is low and does not exceed the number of males. This mating is unbalanced with respect to the number of females and males. The estimates of the GCA of each male is very precise. Since the males are few this is a waste of resources. Historically this was the first systematic mating design used world-wide in tree breeding programmes. Earlier, seemingly haphazard matings were carried out. In the early days of tree breeding it was important to compare the performance of plus tree progenies with ordinary seed lots. Since flowering in the young seed orchards was erratic, systematic matings were almost impossible. The early tree breeders had to rely on data from unsystematic matings. The greatest efficiency of factorial matings is obtained at equal numbers of females and males. Also for factorial matings there are possibilities to reduce the number of matings to enable a simplified handling of the progeny testing.

---

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>6</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>7</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>8</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>9</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>10</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>11</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>12</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

*Figure 9-23. A factorial mating design that is frequently called matings with common testers.*
Disconnected half-diallels (Figure 9-24) are groups of half-diallels that have no clones in common. This mating design became popular worldwide around 1980 and substituted the common tester mating design in many tree breeding programmes. The major advantage with this mating design is that small half-diallels are easy to complete. The parents are selected according to flowering a certain year. Clones which are not flowering one year may perhaps flower the next year so that another half-diallel can be accomplished that year. Flowering has been a great obstacle in certain species for completion of mating designs using several clones. The greatest disadvantage with all mating designs without connections between groups of progenies is that a comparison of breeding values of parents from different groups is not totally unequivocal.

Single-pair mating means that each parent is mated just to one other parent. This mating design might be good to mate parents with good breeding values for generating families for selection forward. The possibilities to estimate genetic variance components are more or less nonexistent.

Polycross and open pollination are two satisfactory alternatives for estimates of breeding values. Non-additive estimates cannot be obtained in these two cases. Polycross means that each parent is pollinated with a pollen mix, usually with a large number of males. Open pollination means that seeds are harvested from trees without any artificial pollination. At the first selection of plus trees a simultaneous collection of seeds enables an early establishment of progeny trials, which gives a gain in time in the breeding work. Since each parent is represented by one progeny only, the trial area is much less than for other mating designs, even if the number of trees per family should be larger in progeny testing using polycross or open pollination than in other mating designs.

Nested matings

Nested means that the parents are grouped into a series of nests, preferably no less than 20 in each nest. In its most complete form each female is mated with pollen mixes from each nest. This may result in some selfing but it is judged as negligible since selfed seedling will be outcompeted by the outcrossed seedlings. The estimates of parental GCA are good if the pollen mix is composed of 20-30 parents. There are possibilities to modify the complete nested design with less labour demanding designs. Since they do not seem to have been used in forest genetics research or breeding we will not discuss them.

Point of time for selection

For trees with long rotation times amounting to several decades it is impossible to postpone the evaluation of the progeny trials until harvest. Some breeders claim that one third of the rotation time is enough for a ranking of the parents with respect to growth. Even one third of the rotation time means many years for high latitude progeny trials. A ranking of the parents for growth at an age of 15 - 20 years will probably not result in significant mistakes. The long-term growth potential is probably best obtained from the growth increment during the last five or ten years. The breeders are generally careful about selection of the locality for the progeny trial to get as homogeneous ground as possible. However, the phase of establishment is a very sensitive part of the development of a progeny trial. Planting shocks might be random so that some plants are hit severely while others are less affected. The competition with weeds is another matter which might affect juvenile plants in a random way. The effect of such environmental effects will diminish with time and more of the genetic quality will determine later growth.

As trees in a progeny trial grow, they will face an increased competition for resources such as water, nutrients, and light from the other trees in the progeny trial. If the competition is allowed to be very strong this will lead to a stronger differentiation among the families. This will facilitate the selection of the best parents. However, if we are interested in an estimation of future gains via the heritability which is derived from the results in the progeny trial we shall probably exaggerate the potential gain from the material under strong competition.

One option to determine the point of time for selection is to estimate age-age correlations, i.e. correlations estimated on the same tree individual between the same trait at different ages. for example, in a Swedish progeny trial of Pinus sylvestris, high age-age correlations were estimated for tracheid length between ages 11 and 31, and for wood density between ages 8-11, and ages 28-33. The results also showed that the genetic gain per year for these
traits was two to three times larger when selection was carried out at age 11 rather at age 31 or 33. This indicates that the optimum selection age might be even lower than 11. Moreover, early tests for these traits should increase the efficiency of the Pinus sylvestris tree breeding program.

Early tests

Great hopes have been invested in possibilities of predicting future growth performance on seedlings or even seeds. The advantage with early tests is that the circle in Figure 9-7 can be completed much faster than is possible with long-time field testing. One problem with early testing is to identify the trait or the combination of traits in the juvenile material that gives a strong correlation with the valuable adult traits. Up to the end of the 20th century the early tests for growth have not given any consistent results. Strong juvenile - mature (J-M) correlations were obtained in a few cases while no correlations were found in other cases. It is of interest to analyse the reasons for weak juvenile-mature genetic correlations.

1. Human failure may have resulted in mislabelling. The scions, the grafts, and the seed lots might have been mixed with wrong identity as a consequence. Pollen contamination may have occurred since it is extremely hard to avoid pollen contamination in wind pollinated species. The experimental plan may not reflect the true identity.

2. The additive variance may be low either at the juvenile or the mature stage.

3. Different sets of alleles regulate the trait at the juvenile and mature stages. One probable case might be the presence of free growth in spruce at the juvenile stage which disappears at a certain age.

4. As discussed in Chapter 5 the same phenotype might be created by several different combinations of alleles. A fast growing juvenile plant might be caused by a genotype that differs from that of a fast growing mature tree.

5. The environmental conditions are mostly different in growth chambers, greenhouses, or nurseries and in the field. This may result in a genotype x environment interaction.

6. As discussed above, non-genetic effects may dominate during the phase of establishment. This means that strong J-M genetic correlations cannot be expected until the genetic effects are dominating in the field trials.

7. The results in the field trials may not reflect the genetic capacity fully owing to imperfections in experimental design or to other causes leading to imprecision of the estimates.

8. Since growth is a complex trait individual components of growth may not give strong J-M correlations. Weighting of the components in an index may be one way to overcome this.

The simplest way to develop early tests is to utilise the results from existing field trials. This is possible when the parents of such field trials are present in seed orchards or clonal archives. Crossings can be repeated or seeds might be obtained after open pollination and young siblings to the more mature trees in field trials can be studied at the juvenile stage. Such an early test is called retrospective.

The same Norway spruce material was tested with respect to nutrient efficiency and water availability to test if explanation 8 above would be true. However, there was no indication that this explanation is the correct one for the poor juvenile-mature relationships. The most likely reason for poor juvenile-mature genetic correlations is that there are different sets of genes active during the juvenile
It was relatively easy to develop early tests for frost tolerance (Picture 9-12 and 9-13). This trait is of greatest significance during the phase of establishment. At this time the plants are close to the ground and the temperatures during clear nights with cool air is much lower than the temperatures recorded by weather stations. Normally the temperature is recorded at 1.3 meter above ground.

In Sweden, freeze testing of individual progenies or bulked seed lots from an orchard is routinely carried out for Scots pine for the interior part of northerly Sweden. The principle of the procedure followed for this kind of freeze testing is illustrated in Box 9-5. With the help of such freeze tests the frost hardiness of a material can be obtained already half a year after seed harvest. Under field conditions the critical period for survival occurs when the young trees are above the snow cover during late winter. However, the critical weather conditions do not appear regularly. It can take some years before the trees are exposed to the critical temperatures. Twenty years is usually the time period required for reliable results as regards frost tolerance of Scots pine in northerly Sweden.

Around 1990 a molecular genetics method for early testing was developed. This method requires the identification of quantitative trait loci, QTLs, (see Chapter 5). It is expected that the results from molecular marker techniques will enable early selection of individuals with a desirable phenotype and thus increase genetic gain per time unit. However, detection of QTLs is most efficient for traits that have high heritabilities, which often means that these traits are affected by genes with large effects. But for these traits phenotypic selection is often more efficient. This dilemma can be solved by using marker-assisted selection (MAS) only when a phenotypic selection for a trait with high heritability is more expensive or takes longer time. Traits with low heritabilities can be subdivided into components, each with a relatively higher heritability, for example height growth can be subdivided into time for growth initiation, length of growing period and time for growth cessation.

**Progress in breeding**

With the help of equations such as number 6 in the section about genetic gain in Chapter 5 we can theoretically estimate the genetic gain that can be obtained from different methods of breeding.

If a seed orchard is established and we designate the proportion of selected plus trees as \( i \), the theoretical gain becomes identical to the gain in equation 7 in Chapter 5, i.e.

\[
\Delta G = i \frac{\sigma^2}{\mu} + \frac{1}{2i} \frac{\sigma^2}{\mu}
\]

Contrary to this, the seeds from a seedling seed orchard will only have half of that gain, \( \frac{1}{2i} \frac{\sigma^2}{\mu} \), if the seed is collected after open pollination. The reason is that there was no selection among the pollen producing parents.

The progeny trials are established in order to guide roguing in existing seed orchards or to guide which crosses should be carried out to obtain a filial generation in which the best trees in the best families should be selected. From Box 9-6 it may be seen that one can obtain different partial gains. It is beyond the scope of this book to give all equations for the different partial gains that might be obtained.

One major objective of seedling seed orchards is to identify the best families and the best trees in those families. Once this information is available, culling of inferior individuals and families can take place (4 in Box 9-6).

Except for biclonal seed orchards, only the additive vari-
Volume gain % over unimproved material

**Picea abies and Pinus sylvestris, Sweden**

<table>
<thead>
<tr>
<th>Method</th>
<th>Partial gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Seed from clonal seed orchard without roguing</td>
<td>One gain related to the plus tree selection</td>
</tr>
<tr>
<td>2. Seed from seedling seed orchard without roguing</td>
<td>The gain is A) equal to 1 if the seeds were obtained from crosses among the selected trees or B) is half of that gain if the seeds were obtained from open pollination</td>
</tr>
<tr>
<td>3. Selection backward</td>
<td>One partial gain related to the original plus tree selection</td>
</tr>
<tr>
<td>4. Selection forward</td>
<td>One partial gain related to the original plus tree selection</td>
</tr>
<tr>
<td>5. Biclonal seed orchard based on progeny testing of full-sib families obtained from controlled crosses in a clonal seed orchard</td>
<td>One partial gain related to the original plus tree selection</td>
</tr>
</tbody>
</table>

In most countries seed orchards were not established until the end of the 1940s. This means that crosses to raise progenies for estimation of realised gains could not be started until 1960. All progeny trials are young and predictions of gains at full rotation cannot be given. However, there is a large number of progeny trials with fairly uniform data that suggest that considerable gains could be obtained. In 2001 The Forest Research Institute of Sweden summarized the results from approximately 40 progeny trials of *Pinus sylvestris* and *Picea abies*. The average improvement for these seed orchards with untested clones, *i.e.* trees selected in stands, amounted to 10% (Fig. 9-25). An additional gain of 2% can be obtained by roguing in this type of seed orchard. The real improvement, 25%, can be achieved by establishment of a second generation of seed orchards with the best parental clones from the first generation of seed orchards. For the northern harsh parts of Sweden where survival is a serious problem in *Pinus sylvestris* plantations, improved survival increases the gain further. The most recent data from Scots pine progeny trials in northern Sweden evaluated at age 27 indicate that the gain in stem volume was 18.9%, which is substantially higher than shown in Fig.9-25.

The improvement in the breeding programme of *Pinus taeda* in south-eastern USA is fairly similar to the Swe-
Pollen contamination is a serious problem in most conifer seed orchards. With the aid of biochemical markers it was estimated that the contamination amounts to such a high figure as 50 % on average. Cases with up to 70 % contamination are not uncommon. This is particularly important for the Scots pine seed orchards for the northerly part of Sweden. These seed orchards are located in a milder area than the region in which the seed should be used. The reason for this was to obtain regular seed crops from the seed orchards. In the northern interior part of Sweden good seed crops occur at intervals of 15-20 years. The seed from such an orchard will be a mixture of crosses among the clones in the seed orchard and crosses between seed orchard clones and trees in stands. Since there is no method available that identifies seedlings that are the result of pollen contamination the seed crop is hardly of any use at all. This is particularly pronounced when the contamination is 50 %, which will result in a distribution with two peaks. For regions in which the hardiness problem does not exist such as for Scots pine in southern Sweden, contamination will "only" result in a reduced gain in growth in proportion to the contamination (Fig. 9-2).

Some of the most advanced breeding programmes in the world occur in two cooperatives in south-eastern US. In this region the breeders are facing a large problem as regards pollen contamination. Owing to the forest ownership there are small land owners who do not utilise any bred material. Therefore, there are fairly large areas with native forests which can spread its pollen to the third generation of seed orchards. In the case of 100 % pollen contamination the gain is reduced to half the potential gain. Thus if the gain can be 30 %, a total pollen contamination would reduce the gain to 15 %. In summary it is most urgent to avoid pollen contamination after completion of several cycles in the breeding population. The difference between what is theoretically possible and what is obtained is unfortunately maximised under such a situation.

Besides the problems with pollen contamination there are others which contribute to deviations from the ideal composition of the seed crop after random mating in a seed orchard. Large differences in the number of female (Fig. 9-27) and male strobili per clone occur frequently, especially in young seed orchards. Similarly the point of time of receptivity and pollen dispersal vary among clones in a seed orchard (Fig. 9-28). These factors contribute to a skewed distribution of the contribution to the filial generation (Fig. 9-29).

To overcome the problems with pollen contamination and unequal contribution of the parents to the filial generation breeders are working actively to develop alternatives to the conventional seed orchards. There is a large potential to improve the gain as indicated in Figs 9-25 and 9-26.
The sustainability of the gain

In model studies in maize and *Drosophila* as well as in some breeding programmes there has been a response to selection for quantitative traits even after 100 generations of directional selection. These results suggest that there is no substantial decrease in the additive variance even if the population sizes were as low as 20-40 individuals. In such small populations exposed to a strong selection, the existing additive variance at the start of the selection would be eroded after 10-20 generations. Since there has been a steady response to selection new genetic variation must have arisen, or alternatively, previously neutral alleles have contributed to the regulation of the trait in the new genetic environment. Since mutations per locus arise at a rate of one per hundred thousand per generation there must be a large number of loci involved in the regulation of the trait if the hypothesis on mutations is correct. Earlier we have mentioned that the pooled mutation rate for one quantitative trait is considerably higher and might reach one per thousand or even higher. The true explanation for the steady response to selection remains to determine. Evidently, heterozygosity per se is not the explanation.

The knowledge that the variance remains in spite of an intensive selection is fundamental, since absence of variance would cause stagnation instead of progress. This is true for traits which the breeders want to improve. What is the situation for traits not included in any breeding programme? Under the assumption that these traits are not linked to traits included in the breeding, the relationship for loss of additive variance at random selection is valid \((1 - 1/2N_e)\). From this formula it is evident that selection
of one individual means that 50% of the additive variance remains (see Figure 9-30). Even a selection of such a low number as 10 individuals means that 95% of the additive variance remains. As is evident from Fig 9-30 a randomly selected population with 500 trees has almost the same additive variance as one with 1000 trees. Therefore, we do not gain much by increasing the population size above 500 trees.

Summary

Forest tree breeding is a cyclic process, in which the gains are obtained by selecting the best trees in the breeding population for seed production or for vegetative propagation. Different populations have different functions. The breeding population should safeguard long-term gain in the breeding. Seed orchards serve as propagule population, breeding population, and gene resource population. It is recommended to split the breeding population into 20 subpopulations with slightly varying selection criteria. This will cause an increased variance among the subpopulations, which facilitates sustainable gains in the breeding. A full diallel cross is the best mating design with respect to estimation of additive and non-additive variance. Owing to the large number of crosses that must be carried out it is not feasible when large number of clones should be tested. Mating designs have been developed in which a reduced number of crosses are carried out.

Before the establishment of new seed orchards it is important to find sites with a warm local climate to stimulate flowering. Seed from existing seed orchards contain a considerable genetic gain. The great weakness of the conventional seed orchards all around the world is that the theoretically possible gains are not reached owing to pollen contamination, and the variable pollen and seed production of the seed orchard clones. Differences in the points of time for receptivity and pollen dispersal also contribute to deviations from theoretical expectations. Pollen contamination is very harmful for Scots pine seed orchards for northerly Sweden since they are located far south of the area in which the seed should be used.

Further reading


Plant production

In this chapter knowledge of the varying requirements for growth cessation and building up of frost hardiness is illustrated from an applied perspective. Genetic aspects of container cultivation as compared to open air cultivation are presented. Finally the possibilities for vegetative propagation are touched upon.

In Scandinavia many companies raise their seedlings indoors in polythene houses. Early on there were some failures since the sowing took place in late February or early March with immediate budset after the development of the cotyledons. During this part of the year the long nights induce immediate budset after germination in northerly materials. Already during the 1960s it was shown that growth cessation is mainly regulated by the night length. The clinal variation in critical night length for budset in Norway spruce was presented in Chapter 7. Once it was realised that the nights were too long, artificial light was used to prolong the day and the plants continued to grow.

The reason for the early sowing was that nurserymen wanted to complete two growth periods during one season. The plants were therefore exposed to 4-6 weeks of 16-hour nights in May and June, after which the plants were moved outdoors where they began their second growth period. This cultivation technique has led to development of robust plants for reforestation.

Since materials of varying origin have different critical night lengths it is important that the nurserymen know approximately the critical night length of each material to avoid too early growth cessation. During the autumn before cold storage or planting it is essential that the plants are frost hardy. In many nurseries artificial night prolongation to 16 hours is carried out to complete the process of attaining stable hardiness. Most materials need a continuous period with such long nights to achieve this. Extremely northern populations have set terminal buds after exposure to one single 16-hour night.

During the 1970s and 1980s many Swedish nurseries stopped their open-air production of seedlings and started to use various types of container for plant production. This means that 1- or 2-year old seedlings instead of 4-year old plants are planted in forests. This in turn means that a material with a much longer growth period is planted in forests since the duration of the growth period declines strongly during the first years (Figure 10-1). This also means that small seedlings are planted in the forests and that the seedlings are closer to the ground, at which the temperature is lowest during clear cool nights. Both these conditions mean that the change of production system leads to reforestation with a plant material at higher risk for frost exposure and frost damage.

**Figure 10-1.** The figure illustrates that the duration of the growth period declines with increasing age.
Container raising of seedlings also means that each seed should germinate and give rise to one seedling that is planned in the forest. This means that the entire distribution with both poor and good genotypes is represented in the material that is planted in the forest. In open-air cultivation it is assumed that the poor genotypes are outcompeted and never reach the forest. The hypothesis is that good cultivation conditions in the nurseries with adequate water and nutrients lead to minor differences between poor and good genotypes. The performance of selfed Norway spruce in nursery and in field revealed a difference in 30 percentage units (Fig. 10-2) and thus support the hypothesis. If these data are confirmed in other investigations such a cultivation system would lead to higher mortality in the afforestation with a reduced yield in future.

From a theoretical point of view the selection differential is probably not very large between the two types of cultivation. A large proportion of the difference in number of seedlings germinated per kilogram seed must probably be attributed to differences other than genetic. To avoid potential genetic risks connected with container raising of seedlings it is important to have material with high genetic quality.

Some of the problems referred to above can be overcome by regulating the cultivation regimes in nurseries. Thus, there are possibilities to influence the duration of the growth period during one season via the cultivation regime used during the preceding growth period. Cultivation of Norway spruce seedlings for 24 weeks without any night leads to a shortening of the second growth period. Since the probability for exposure to frost decreases with time during spring and increases with time during autumn, the probability for frost exposure is reduced during the second growth period in plants cultivated for 24 weeks without any night during the first growth period. The treatment has caused a more aged performance, i.e. shorter growth period, during the second growth period (Fig. 10-3), which might last even during the third growth period. Norway spruce seedlings which from their start grow under continuous light also grow continuously with steady cell divisions. Many scientists believe that the number of cell divisions determines whether a plant will be juvenile or adult. This would explain the aged performance of the young Norway spruce seedlings described above.

During the 1970s there was a growing interest in the production of cuttings from valuable genotypes. The main interest was to mass propagate trees which had shown promising results in field trials. It was soon realised that there were great difficulties connected with vegetative propagation. Thus, the percentage rooted Norway spruce cuttings from old trees was low and those which had roots frequently showed the branch type growth rather than the orthotrophic growth characteristic of young seedlings. Frequently the root formation was abnormal with one unbranched root growing perpendicular to the stem. The branchlike growth habit and the abnormal root growth is attributed to ontogenetic aging. Cuttings do not perform like young plants but rather as parts of a mature tree. This type of aging appears already at an age of 5 years in Norway spruce. Cuttings can successfully be produced from 1-4 years old seedlings. Among pines Pinus radiata is one example of a species in which vegetative propagation is commercially possible.

If scions are taken from different parts of the crown of an adult Norway spruce tree, those taken at the bottom of the

![Figure 10-2. The inbreeding depression of genetically identical material in field and in greenhouse.](image)

![Figure 10-3. Duration of the second growth period, GP2, in Norway spruce seedlings cultivated under continuous light for 8 and 24 weeks during the first growth period, GP1, respectively.](image)
crown give the highest percentage of rooting. The scions taken from the apical part of the crown show most symptoms of ontogenetic aging which might seem remarkable. The scions from the apex are evidently chronologically the youngest but they are formed from a meristem that has passed through hundreds of cell divisions and definitely more than the meristem at the bottom of the tree crown. Once again, this is a sign that the number of cell divisions is critical for the ontogenetic age.

Since the 1970s much research has focused on tissue culture techniques, by which it is possible to obtain plants from ordinary somatic cells. Such techniques are common in horticulture. Great hopes were raised when successful plant regeneration via somatic embryos was reported. By means of certain hormone treatments it was possible to obtain embryos from somatic cells. In 2000 it is possible to obtain somatic embryo plants on a large scale in Norway spruce and Scots pine. Such plants are sometimes referred to as emblings.

The Forest Research Institute of Sweden has developed a computer programme for selection of plant material for reforestation in northern Sweden. Via the Institute’s home page anyone has free access to this programme, which gives several options for reforestation material to be used at a particular clear-felled area.

Summary

Genetic knowledge of importance for plant cultivation is mainly related to genetic variation in growth rhythm of different materials. There have been fears that container raising of seedlings where even poor seedlings are planted out in the forests should lead to inclusion of poor genotypes in coming production populations in contrast to the case with open air cultivation in nurseries. It is assumed that inferior seedlings are outcompeted in the latter type of cultivation, but there is no definite proof for this. There is great interest among breeders in the vegetative propagation of outstanding trees, and this has been done successfully for a small number of tree species.

Further reading

Forest tree gene conservation

The difference between gene and genotype conservation is first explained. After that the three key components of gene conservation - objectives, genetic structure & dynamics, and methods are presented. Identification of target species is discussed. The discussion of methods occupies the largest space in this chapter.

The prime aim of gene conservation is to save enough genetic variation for the targeted species to enable it to cope with changes in the environment. Expressed in another way, the potential for adaptation should be focused in forest tree gene conservation. Similarly, broad genetic variation is also required for long-term success in breeding. These matters will be elaborated somewhat more under objectives below. In other cases the reason for conservation may be more of affectionate nature, such as the case for Scots pine, once an important commercial tree species in Great Britain, now with a strongly reduced number (Picture 11-1).

The future demand for wood or conversion of forestland for crop cultivation is of significance when we design gene conservation. Without these considerations the most sophisticated gene conservation program may be futile in a long-term perspective. One problem in designing gene conservation programs is that we, in the overwhelming number of cases, do not know the relationship between gene diversity and production of utilities. If this relationship follows curve A in Fig. 11-1 there is no conflict between the two goals whereas curve E describes a situation harder to solve. Curve E suggests that production of utilities and gene conservation should be carried out separately. Research priority should be given to studies of the relationship between diversity and production.

A fundamental question is whether we should conserve genes or genotypes. So far there is no technique available for conservation of genes like books in a library, and for the foreseeable future gene conservation will take place by conservation of certain genotypes. This means that commonly occurring genes will be conserved, but only in a limited number of genotypes. Is this satisfactory? The question is justified since genes may interact with each other and there can be difficulties in recreating valuable genotypes. In Chapter 3 it was shown that heterozygosity even at a rather limited number of loci gives rise to a large number of genotypes. In cross-fertilizing species, all individuals except for identical twins have a unique genotype. For this reason conservation of each genotype is impossible. Conservation of all genotypes would require a gigantic global museum of all existing individuals. As in many other cases a compromise must be developed, which means that a representative sample of existing genotypes is conserved in such a way that the largest possible number of genes will be included in the gene conservation programme.

Before we discuss gene conservation methods it is useful to repeat some of the knowledge from Chapters 5 and 6, which is of great significance for forest tree gene conservation. Therefore, Figures 11-2 – 11-4 are included once more. The gene frequency constituting the lower limit for conservation has been debated. Some forest geneticists claim that gene conservation should be designed such that rare alleles are included in gene resource populations. Others claim that they are of limited or no importance for gene conservation. From Fig. 11-2 it is evident that alleles in very low frequencies do not contribute to the

![Figure 11-1. Potential relationships between genetic diversity and production of other utilities.](image-url)

**Picture 11-1.** A Scots pine population in Scotland. Photograph Gösta Eriksson.
additive variance; they do not safeguard the potential for adaptation and are of limited value in gene conservation. Another way to explain this is that rare, recessive alleles mainly occur in heterozygotes and they cannot be increased in frequency by natural selection. Therefore, it is expected that natural populations will have a large number of recessive alleles in low frequencies, which cause a reduction of the vitality of the recessive homozygotes. This was observed in Douglas fir for which a very large experimental material was analysed. The most common reason for a low frequency of rare alleles is that they do not contribute to fitness of the individuals having these alleles. Other rare alleles are probably the result of recent mutations.

Since loss of additive variance is considerable in populations with an effective population size of 20 or lower (Fig. 11-3) gene resource populations should be large enough to avoid serious losses of additive variance. From Fig 11-4 it is seen that allele frequency is more important than number of low-frequency alleles for the possibility to include this type of alleles in gene resource populations.

The three cornerstones of gene conservation

Gene conservation has three cornerstones— the objectives, the genetic structure & dynamics, and the methods. Genetic structure or population structure was discussed in Chapter 7. When the objective is clearly identified and the genetic structure is known the most adequate method for gene conservation that matches the objective should be developed.

Objectives in gene conservation

Prime objective

In Chapter 6 we stressed that for species in nature, several evolutionary factors may be in operation simultaneously. There is a steady change of gene frequency in a population according to the ambient conditions and/or the genetic qualities (above all the Ne) of the population. Therefore, dynamic forest tree gene conservation has been argued for in most instances. The Swedish parliament took a decision in 1991 according to which, in a word for word translation, The biological diversity and the genetic variation should be safeguarded...and...naturally occurring plant and animal species should be given conditions for continued existence under natural conditions and in vigorous populations. This is in agreement with decisions in many other countries that signed the Rio declaration, and with what one of the forefront persons in gene conservation, Michael Soulé, has stated: Conservation genetics exists for one reason only: To promote the fitness of tar-

Figure 11-2. The relationship between gene frequency and additive variance with completely additive gene action; a is the value a illustrated in Figure 5-2.

Figure 11-3. Remaining fraction of additive variance after 10 generations as a consequence of genetic drift. It is assumed that the effective population size was constant during these 10 generations. Loss of additive variance is considerable at an effective population size of 20 or lower.

Figure 11-4. The minimum number of individuals required to save one rare allele at each of 10, 50, or 100 loci.
geted populations. A related formulation is to safeguard the potential for adaptation of a species, which has been used by the EUFORGEN network on noble hardwoods. If the potential for adaptation is guaranteed, the species has a greater chance to cope with the changes continuously occurring in the environment.

For the tree species included in breeding programmes it is important to analyse whether it is possible to include breeding within dynamic gene conservation. Simultaneous gene conservation and breeding might be one objective.

Other objectives

Preservation of the present genetic constitution is another objective in gene conservation. Behind this objective we may distinguish four different causes:

1. The need to have a reference material for comparisons in future research
2. The belief that natural selection has chiselled out individuals that are perfectly adapted to the site condition where they live.
3. The material has highly desired characteristics that may be lost if the present constitution is not preserved
4. Wild tree species may hybridise with highly bred cultivars

In breeding of agricultural crops it has turned out that point 1 has been useful and therefore may be useful in forestry as well.

In Chapter 6 it was clarified that point 2 is incorrect. The present genetic structure is one out of many possible and it is transient. However, it is a good starting material for dynamic gene conservation.

Point 3 is frequently referred to gene conservation although it might be better to classify it as a breeding or production objective since it is the human utility of a tree species that is the motif for preservation.

Wild fruit trees such as apples and pears are rarely occurring and as such exposed to great threats of hybridization with cultivated varieties. Preservation of the wild status is one objective.

Some scientists argue for conservation of unknown genetic variation. The reason for this is the expectation that useful substances for mankind might be detected.

Another objective is to save populations that are endangered directly or indirectly by human activities. Tree species that today have their distribution restricted to a low number of mountain peaks face a serious threat if there is a climatic change with higher temperatures as a result (Fig. 11-5). At such a change there is no possibility for the species to migrate to higher elevation. Abies fraseri in the Appalachian mountains in south-eastern USA and Abies pinsapo in southern Spain and North-Africa are examples of species that may suffer if there is considerable global warming. The cost for saving the gene resources of such species will be considerable. Especially riparian species in the industrialised world have been exposed to urbanisation, since the banks of many rivers have been stabilised, preventing the natural dynamics of rivers and thereby the natural regeneration sites of such species. Fragmentation of populations is one result of such activities. Biological threats will be discussed below under Grouping of Species.

The objectives discussed so far have concerned a single species, i.e. the species targeted in the gene conservation. For most forest tree species, many other species are dependent on them for their existence. In this case the tree species is designated as an ecological keystone species. By associated species is meant a species dependent on other species for its existence. A totally objective separation between keystone species and non-keystone spe-
cies does not exist. There are many transition cases since species are always dependent on other species to some extent. It is obvious that forest trees thanks to their age and size are of great significance for many other species. A final objective for gene conservation is to include the conservation of species associated with a target species.

**Genetic structure**

In Chapter 7, several possible population structures were presented. The strategy for gene conservation depends on the genetic structure of the species in question. Knowledge of the genetic structure is indispensable for genetically satisfactory gene conservation. There will never be funding enough for studies of the genetic structure of all species, so such studies will be restricted to a few commercially important species. Measures to be taken in absence of knowledge about genetic structure will be discussed under “Gene conservation methods”.

**In situ and ex situ gene conservation**

Traditionally two main methods have been distinguished in gene conservation, *in situ* and *ex situ*. *In situ* means “on the spot” and is understood by most forest geneticists as conservation of naturally regenerated forests. *Ex situ* means that gene conservation is carried out by seed or pollen banks or that the gene resource population occurs in some kind of plantation. A certain confusion of concepts exists since many gene conservationists of agricultural crops interpret the *in situ* concept in another way. They regard a growing crop as *in situ* while *ex situ* for them is limited to banks of seed, pollen, or tissue culture.

Owing to this confusion of concepts it would be better to classify methods according to the function of the different gene resource populations. However, the *in situ* and *ex situ* terms are so commonly used that a new but unequivocal terminology would not be able to outcompete these terms.

Before an analysis of the different methods is carried out it is important to emphasize that gene conservation will always have too limited funding. Therefore, it is important to develop methods that unite as many objectives as possible. It is also important to give priority to certain species. Species given priority are referred to as target species. Once target species are identified a grouping of species with respect to gene conservation methods has to be carried out.

**Target species**

There are several options to select a species as a target species:
- scientific reason
- threat
- charisma
- economic reasons

For scientific reasons there is a desire to choose species in order to combine different ecological characteristics such as distribution, pollen vector, seed dispersal, and stage in ecosystem. We shall return to these characteristics in the next section.

If species are selected according to their ecological characteristics and designated as ecological keystone species, they will probably be maintained as target species. Whether a species has any close relatives is another scientific reason to base the selection of target species on. Some scientists suggest that species without any close relative should not be selected as target species, since they appear to belong to an evolutionary dead end. If target species are selected among species which have shown recent speciation the funding is spent on vigorous species with high potential for continued evolution. Other scientists claim the opposite, that species without any close relatives are probably genetically unique and deserve to be target species. This type of species is frequently endemic, *i.e.* they occur within one restricted area only. They are sometimes endangered, and therefore they have often been conserved. For practical reasons it is natural that this has been the case. However, an analysis of the potential for adaptation of such a species ought to be carried out before decisions on large investments are taken.

The selection of charismatic species might seem to be an incorrect way to utilize limited funding for a few species only. The spotted owl in North Western USA is an example of a charismatic species, which has taken large resources. Also such an investment might be justified since it draws public attention to conservation issues. As a consequence of this, it might be easier to raise funding for gene conservation.

We are facing an explosion of the human population never experienced before. In this perspective the demand for fibres for various purposes will probably increase dramatically. The human population increase also contributes to an aggravation of the effects of pollutants and climatic change. Improvement in the living conditions in developing countries is one way to come closer to eco-
Economic equality among countries. One means to achieve this goal is to utilize the renewable forest resources. The demand for wood will probably increase dramatically up to the middle of the 21st century.

In countries with hundreds of tree species a scoring system has been used to identify target species. Scientists, farmers, local peasants, and business people scored the species with respect to utility, ecological importance, and threat.

Probably a weighting of the economic and the ecological reasons is a good starting point for selection of target species. The scoring mentioned above is very close to such a weighting. The various reasons have to be judged for each individual case. It is also of great significance that the methods for gene conservation that we suggest will stand future pressure of demand for wood. If not, the gene resource populations may be cut, like most of the park trees in Sarajevo during the civil war in former Yugoslavia.

Grouping of species in gene conservation

Ecological characteristics

Tree species differ in their characteristics and may require different methods for their gene conservation. Of interest is to identify whether a grouping of species with respect to gene conservation can be done by the aid of species’ characteristics such as:

- Distribution, wide range - limited; continuous - scattered; large - small populations
- Mating system, wind pollinated - animal pollinated
- Stage in ecosystem, climax - intermediate - pioneer species

Population size covers everything from the extreme situation of large random mating populations to widely scattered single trees. The effective population size is of great significance for random mating. A species consisting of many small and scattered populations without any gene flow among the populations will give rise to large but non-adaptive genetic variation among populations in contrast to the adaptive variation that may exist in random mating populations. The species are therefore first grouped according to rarity, i.e. whether they are rarely or commonly occurring.

Some of the rare tree species are intermediate species and are therefore cut during thinning to promote the growth of the climax species. Owing to limited taxonomic knowledge about rare species among foresters, thinning might have included rarely occurring tree species without any intention to exterminate them.

Wind pollination and pollination by insects, birds, or bats are two major types of mating system. Studies of the mating system by aid of isozymes have revealed that mating system influences the genetic structure of a species. Wind pollinated species have generally a higher within-population/among-population variation ratio than insect pollinated species (Chapter 6). This is attributed to gene flow over larger distances of wind pollinated species than in insect pollinated species. This does not mean that every wind-pollinated species has wider pollen dispersal than every insect pollinated species. Generally, an insect pollinated species will need a larger number of gene resource populations than a wind-pollinated species since the differentiation in a given area is assumed to be larger for the insect pollinated species than for the wind-pollinated species. However, the difference between the two types of mating system is probably of another magnitude than differences between rarely and commonly occurring species.

By stage in ecosystem we mean whether a species is a pioneer species, a climax species or takes a position between these two extremes. Typically, pioneers invade open areas with fairly homogeneous growth conditions that do not call for a large genetic variation. Rather once a genotype with good adaptedness to the prevailing conditions at the open area arises, it would, teleologically speaking, be an advantage for the species in a short-term perspective to rely on that genotype. Therefore, asexual propagation of highly adapted genotypes like in *Taraxacum vulgare* would be advantageous. Contrary to this, climax species experience heterogeneity both in space and time during their lifetime, and genetic variation within populations must be assumed to be advantageous. The climax – pioneer difference in genetic variation is analogous to the contrast wind pollination – insect pollination and is probably of a lower magnitude than the classification: rarely occurring – commonly occurring.

Involvement in breeding activities

Long-term breeding efforts require breeding populations with satisfactory additive variance, which is also a major prerequisite for gene conservation. If this requirement for a large additive variance is fulfilled, gene conservation is well taken care of in breeding. Whether or not a species is included in breeding might therefore be used in grouping of species. Moreover, the intensity of breeding might vary and as a corollary the amount of additive variance in the breeding population might vary. The objectives in breeding might include one or several traits. When breeding objectives in a species comprise such disparate traits as high-quality timber and nuts it is not self-evident that improvement can be achieved in one common breeding population. This is especially pronounced for chestnut and walnut, in which selection for nut quality and yield has gone on for millennia while timber improvement has not taken place to any large extent. So gene conservation in multipurpose species needs special treatment.
Biological threats

Besides the threats caused by human activities, biological threats in the form of diseases or pests play a prominent role. Well-known cases are the serious fungal diseases in elms in Europe and America and the American chestnut \((\textit{Castanea dentata})\). Only the most northerly populations of wych elm \((\textit{Ulmus glabra})\) in Europe are not affected by the Dutch elm disease since the \textit{Scolytus} insects, which transmit the disease from tree to tree do not survive under the harsh northern conditions.

Forest tree gene conservation methods

We shall first once more return to the multiple population breeding system, MPBS, in its role for combined breeding and gene conservation. The advantage of splitting the combined gene resource and breeding population into subpopulations is visualised with the help of Figure 11-6. Each point in the cube is assumed to be one subpopulation, which gives a total number of 64 subpopulations. It needs to be emphasized that the growth conditions are not as simple as indicated in this graph. It may very well be that certain of these 64 combinations do not exist in reality. The merit of this subdivision is that each subpopulation might be enriched with alleles promoting fitness under this particular combination of environmental conditions. This means that rare alleles that are valuable under extreme environmental conditions might increase in frequency. For random reasons such alleles would be lost if there was just one large gene resource population. To strengthen this still more, some of the gene resource subpopulations might be planted outside the present range of the species. This will lead to a broader genetic variation than if we have one large gene resource population. The MPBS method of gene conservation and breeding is gaining terrain which means that tree species included in intensive breeding programs do not need a separate gene conservation activity.

To reiterate the summary of the merits of MPBS: The main advantage of the MPBS is that it combines the capture of the total existing genetic variation with a satisfactory variation within each subpopulation and that it allows the target populations to adapt to the prevailing environmental conditions. Another advantage is that the speed of evolution might be faster in a population of 50 trees than in a large population containing thousands of trees.

Besides MPBS another system, coined HOPE, Hierarchical Open Ended, was developed for simultaneous gene conservation and sustainable breeding of agricultural crops for one environmental condition (see Box 11-1) and is of limited relevance for forest trees which will grow under variable site conditions. Another disadvantage is that backcrosses cannot be carried out as easily for forest trees as for crops like barley or maize.

In Tables 11-3 and 11-4 we have summarised the methods that should be used to match different objectives in forest tree gene conservation. The greatest emphasis is given to the prime objective of gene conservation, to safeguard the potential for adaptation of the target species. Before coming to specifics about methods it is of importance to discuss how to select gene resource populations in absence...
Box 11-1 Comparison of the two systems aiming at a combined gene conservation and tree breeding.

HOPE stands for Hierarchial OPen Ended system, which implies that new material may continuously be incorporated into the system. MPBS stands for Multiple Population Breeding System, implying that the system has many equivalent subpopulations.

The level of improvement is shown on the Y-axis. On the X-axis there are many generations of breeding over time. In HOPE there is a transfer of genetic material from the unimproved level and less improved gene resource subpopulations to the green elite population by aid of backcrosses. The level of improvement remains more or less constant in all subpopulations except for the elite population. The elite population has a narrow genetic base. The cultivars for commercial production are generated from the elite population. For each breeding generation the gaps between the elite population and the subpopulations are broadened. HOPE is of limited importance for long-generation species such as is the case for most forest trees.

Over the time the gaps between all the 19 subpopulations in the MPBS system are broadened while keeping satisfactory genetic variation (= additive variance) within each subpopulation. This means that the total additive variance in this case increases over time. Selection of material for seed orchard establishment or clonal propagation takes place in some of the subpopulations according to the demand for reforestation material. Thus, if there is a demand for reforestation for fibre farming, clones are selected from the uppermost subpopulation. When breeding of a species is carried out according to the MPBS system it conserves the genetic variation of that species in a good way.
of genetic knowledge. In this case we have to select the subpopulations based on knowledge about the life history traits and the genetic structure these traits might have given rise to. This means that we may benefit from what we have learned from Chapter 6. For species with random mating populations disruptive natural selection and gene flow are the two dominating evolutionary factors.

In Table 11-1 educated guesses about within – and among-population variation are given for contrasting combinations of these two evolutionary factors. Studies have shown that species with pollen vectors flying over short distances will probably be more differentiated than species that are wind pollinated. This is probably also the case for species with scattered distributions with no or limited gene flow among the scattered populations. In case of non-random mating populations we also have to consider genetic drift and its impact on within- and among-population variation (Table 11-2). The strongest population differentiation is projected for the combination: genetic drift + strong disruptive selection + limited gene flow. The contrasting combination, weak disruptive selection and large gene flow will have the lowest differentiation of populations. In cases with weak disruptive selection and limited gene flow, genetic drift will be the dominating evolutionary factor. Since genetic drift leads to fixation of alleles the variation within populations is expected to be low for this combination. It ought to be stressed that these projections are theoretical and should be used with care and only when information on adaptive differentiation is lacking.

**Safeguarding the potential for adaptation**

As stated above, species included in serious long-term breeding programmes do not need separate gene conservation programmes. The main method for this type of species will be the *ex situ* MPBS (Table 3). This means that there will be a series of progeny trials in which the best phenotypes are crossed to obtain a new generation in the combined breeding and gene resource population. This process is repeated in a recurrent way. Possibly the breeding/gene resource subpopulations should be complemented with additional populations when the MPBS subpopulations do not well cover the entire genetic variation of the species.

Within the CASCADE project, “Securing gene conservation, adaptive and breeding potential of a multipurpose tree species (*Castanea sativa*) in a changing environment”, conservation values were developed under the leadership of Gabriele Bucci for adaptive traits (ATCV), pathogen tolerance (PTCV), and marker traits (MTCV). The number of populations in this study varied from 6 for adaptive traits to 78 for markers. The full use of the ATCVs and the PTCVs cannot be done here since the number of studied populations was too limited. However,
the calculations used in our project can be applied in future for sweet chestnut or any other tree species studied in detail.

The additive trait conservation value (ATCV) can be based on the evolutionary potential or population divergence. The latter is based on how much a specific population differs from the other populations studied. As seen from Fig. 11-7, the Greek population Paiko was the only population that showed low evolutionary potential. The Spanish population Coruna and the Greek population Hortiatis showed the largest population divergence. The former showed good juvenile growth and the latter showed poor growth.

**Figure 11-7.** Adaptive trait conservation value, ATCV, of six Castanea sativa populations. ATCV considers the potential of the population to evolve and its unique genetic constitution.

**Figure 11-8.** Pathogen trait conservation value, PTCV, of Castanea sativa populations. The PTCV combines high tolerance to Phytophthora cambivora and high potential for improvement of tolerance against this pathogen. The value is given for naturalised, coppice and orchard populations separately.

**Figure 11-9.** Marker-based conservation value, MBCV, of Castanea sativa populations from 9 regions in Europe. The MBCV is mainly attributed to richness of genetic variability in individual populations. The value is given for naturalised, coppice and orchard populations separately.
The pathogen tolerance conservation value, PTCV, was based on the inoculations of the material with one strain of Phytophthora cambivora and it was calculated separately for three domestication levels, naturalised, coppice, and orchard populations. The PTCV was calculated in such a way that a high PTCV value means good tolerance against P. cambivora as well as large evolutionary potential for improvement of tolerance. The two coppice populations from Greece showed high PTCVs as well as the naturalised Greek population from Hortiatis (Fig. 11-8). The French populations and the Spanish orchard populations had low PTCVs.

Three estimates were used for derivation of the marker-based conservation value MBCV, expected heterozygosity $H_e$, $F_{ST}$ and $N_e$. For the markers it turned out that $H_e$ had the greatest influence on MBCV. Generally the orchard populations showed the lowest $H_e$ as expected for grafted material. The southern Greek populations had a genetic constitution differing from most other populations and for that reason the Greek populations have a special value for the network of gene resource populations (Fig. 11-9). Noteworthy is the high $H_e$ in both English populations, (Glouchestershire and Suffolk), which both are coppice forests.

In cases where there is low-intensity breeding or no breeding but we know the genetic differentiation of the species, the less intensive in situ MPBS method is recommended (Table 11-3). This method is also recommended for those cases for which we lack the desired genetic knowledge. Cork oak (picture 11-2), Quercus suber, is one example of a species, for which knowledge about genetic differentiation of adaptive traits is just emerging (year 2006). Based on existing knowledge from juvenile experiments the subpopulations suggested in Fig. 11-10 may be one solution for gene conservation of cork oak. Another example about selection

<table>
<thead>
<tr>
<th>Commonly occurring: Species included in intensive breeding</th>
<th>Ex situ MPBS + complementation with populations in the wild when needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commonly occurring: No breeding or low-intensity breeding</td>
<td>In situ MPBS selected according to genetic knowledge</td>
</tr>
<tr>
<td>Commonly occurring: No breeding or low-intensity breeding without any genetic knowledge</td>
<td>In situ MPBS selected on ecogeographic principles</td>
</tr>
<tr>
<td>Commonly occurring: Multipurpose breeding: wood and nuts</td>
<td>MPBS for wood in naturalised forests + clone archives for nuts</td>
</tr>
<tr>
<td>Commonly occurring: Endangered by Dutch elm disease</td>
<td>Low clone hedges + in situ MPBS whenever possible</td>
</tr>
<tr>
<td>Rarely occurring: With possibilities for investment</td>
<td>Clone archives + progeny plantations for each ecogeographic zone</td>
</tr>
<tr>
<td>Rarely occurring: low-cost alternative</td>
<td>promotion of the growth conditions + delivery of seedlings free of charge to forest land owners</td>
</tr>
</tbody>
</table>

Table 11-3. Gene conservation methods to meet the objective of safeguarding the potential for adaptation in various groups of species.


from Hortiatis (Fig. 11-8). The French populations and the Spanish orchard populations had low PTCVs.
of gene resource subpopulations for gene conservation in absence of genetic knowledge is illustrated in Fig. 11-11. In southern Sweden there are thousands of so called habitat protected areas. In free translation from Swedish such an area is defined in the following way: A habitat protected area is a forest area, evaluated from its present structure, species composition, history, and physical environment being of great significance for the forestry flora and fauna. It contains or is expected to contain red-listed species. Since our knowledge about noble hardwood genetics is rather limited, we have suggested a selection of gene resource subpopulations in habitat-protected areas. A climatically representative sample from southern Sweden is aimed at. The suggested subpopulations will be surveyed with respect to numbers of trees of individual noble hardwood species and will be approved if their effective population size is judged to be satisfactory large. If not, neighbouring habitat protected areas will be surveyed.
In the Mediterranean region breeding for nut yield and quality in chestnut (Castanea sativa, Picture 11-3) and walnut (Juglans regia) has gone on for millennia. Grafting of superior nut producing genotypes has taken place, which means that fruit orchards usually have few cultivars/genotypes. Some of them are even male sterile. In high forests (Picture 11-4), adaptation to the prevailing conditions has probably taken place whereas the Darwinian fitness in most cultivars for nut production may be low. According to the concepts introduced in chapter 7, the nut cultivars might have a high degree of domestic fitness while populations in the wild have some degree of Darwinian fitness. This might cause a problem for gene conservation if there is a large gene flow from cultivars to gene resource populations since there is a fear that such a gene flow would drastically reduce the Darwinian fitness of the populations in the wild. A similar problem is due to the introduction of the Asian chestnut species Castanea crenata and C. mollisima for hybridisation with C. sativa to obtain hybrids tolerant to diseases caused by Cryphonectria and Phytophthora species. The species hybrids are less drought tolerant and have another growth rhythm. The latter means that they cannot be used in areas in which late spring frosts are a constraint for chestnut growth. To solve the problems mentioned, the ex situ MPBS method is suggested for breeding with the objective of improvement of wood yield and quality. If funding for that is not available the in situ MPBS is recommended for the naturalised forest populations. The in situ subpopulations should be selected such that gene flow from cultivars is minimised (Fig. 11-12). Probably the nut breeding has suffered from low effective population sizes. Therefore, a series of clonal archives in different ecogeographic regions is suggested. This will permit more efficient breeding when there are more genotypes represented. For areas with no summer drought and limited spring frost problems, clones of the two Asian species might be included in the clonal archives.

Figure 11-12. Suggested principle for gene conservation of Castanea sativa. Separate conservation for wood production and fruit production are suggested. For both purposes the multiple population breeding system concept will be applied.
Dutch elm disease affects wych elm most seriously but in cycles. The long-term gene conservation of this species must rely on low hedges (Picture 11-5). Such hedges do not constitute a breeding ground for the insect vectors (Scolytus insects). Whenever possible the in situ MPBS method is suggested. To overcome the problem with Dutch elm disease and other serious diseases in a long term perspective, breeding for disease tolerance should be carried out. A restoration programme of the severely affected American chestnut is going on by hybridisation of this species with disease tolerant Asian species.

One focal point for gene conservation of rarely occurring tree species is to increase the effective population size to avoid random genetic drift. The most costly way of doing this is to collect scions of trees and produce grafts for clone archives or seed orchards. The seeds obtained from them are used to raise seedlings, which will be planted in forests. The seedlings and trees will be exposed to natural selection. Clones from different ecogeographic regions should not be mixed. Therefore, this intensive gene conservation will be carried out according to the MPBS method. In most cases such a high-cost method will not be possible to carry out; instead it is recommended that seedlings are raised and offered to forest landowners free of charge in order to raise the population size. To overcome the problem of unintentional cutting of rare trees during thinnings, taxonomic training of all kinds of foresters might be a remedy. Many of the rare tree species in temperate forests are intermediate species that might be outcompeted by climax species. Once foresters are aware of the existence of rare species they might even be promoted in silvicultural operations by thinning of competing tree species.

For all types of in situ gene resource populations it is important that their regeneration is guaranteed. Natural regeneration of some gene resource populations may in some cases fail owing to severe competition from other species such that the regeneration takes place with another species than the target species. It is evident that this is a dead end of that gene resource population. If this is the case, active measures must be taken to support the regeneration of the target species. This has to be done even if the gene resource happens to grow in a protected area with “hands off” management regime.
Methods for other objectives in gene conservation

Preservation of the existing genetic constitution means freezing the current genetic structure. The need for reference material (Reason No 1 for preservation presented under Objectives above) is most simply satisfied for many tree species by storage of seeds or other propagules. Seeds, acorn or nuts of several tree species cannot be kept in long-term storage. For these species clone archives are the remedy and such archives are used for conservation of wild relatives to fruit trees such as apple, pear and cherry. Pollen storage is also an alternative for static gene conservation. To match reason No 3 for preservation, a large buffer zone (Fig. 11-13) around the population/forest that should be preserved, is suggested. However, it should be noted that a total preservation is not possible unless the site conditions remain constant, which is highly unlikely.

Table 11-4. Gene conservation methods to meet some gene conservation objectives other than safeguarding the potential for adaptation.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preservation: as reference for future experimentation; avoidance of contamination from cultivars</td>
<td>Clone archives, seed orchards, seed banks (pollen banks)</td>
</tr>
<tr>
<td>Unknown variation</td>
<td>Encompass as much variability as possible; is mostly obtained if the MPBS method is applied</td>
</tr>
<tr>
<td>Threat: from pollution or other anthropogenic causes, urbanization</td>
<td>Clone archives, seed orchards, seed banks, complementary plantations</td>
</tr>
<tr>
<td>Gene conservation of associated species:</td>
<td>Large MPBS, 200 – 300 ha including management of some subpopulations while a few others are nature reserves</td>
</tr>
</tbody>
</table>

To safeguard unknown genetic variation the only solution is to try to include as much of the existing genetic variability as possible. This is probably most efficiently done by the MPBS method.

Threat. In Germany great efforts have been devoted to saving *Picea abies* populations exposed to air pollution. This was done by collecting scions for grafting and establishment of clone archives outside the polluted area. The unique Guadeloup population of *Pinus radiata* has no natural regeneration owing to the goat population of this island. During the 1990s several seed collections were carried out. Seedlings raised from these seed collections were established in California, Australia, and New Zealand. For riparian species, which have lost part of their habitats, the *in situ* MPBS is suggested if breeding is not under consideration. If the fragmentation has gone far, with a strong reduction of the effective population size, complementary plantations with seedlings or cuttings (black poplars, *Populus nigra*) are recommended. Clearings along natural riverbanks may be needed to obtain a satisfactory regeneration.

Figure 11-13. The principle design for preservation of the core gene resource population from undesired gene flow by having a buffer zone.
To satisfy the objective of conservation of associated species we suggest that some of the subpopulations in the MPBS are extended to 200-300 hectares. Many of the rare or endangered species depend on specific habitats for their survival. In all ecosystems, species appear during different stages of the succession, which thus constitute different habitats. Therefore, it is important to design the gene conservation so that all stages from the juvenile to the over-mature phase with dead trees are represented (Fig. 11-14). Controlled forest fire might also be necessary. The subpopulations should be selected so that they cover the site conditions occupied by the target species. Also within each of these large subpopulations as broad cover of site conditions as possible should be aimed at.

Pictures 11-6 and 11-7, which were taken a few hundred metres from each other, reveal that the flora is strongly dependent on the management regime. In the upper picture limited management takes place while the lower population is managed like a park landscape. These two pictures reveal that the management is of great significance for the floral composition of the associated species. To match the demands of different associated species it is important to include different management regimes in the subpopulations of the size 200-300 hectares. If this is done gene

![Picture 11-6 and 11-7. The two pictures were taken of the same Quercus robur population growing approximately 200 metres apart. The forest in the above picture does not have much human intervention while the lower is managed to get a park landscape character. Photograph Gösta Eriksson.](image1)

![Picture 11-8. A mixed Abies alba and Fagus sylvatica climax forest in Slovenia. Note that regeneration consists of Fagus sylvatica seedlings only. Photograph Gösta Eriksson.](image2)

![Figure 11-14. Schematic illustration how a large gene resource population, 200-300 hectares, might be subdivided into 12 plots, 10 of which have different age classes, I- X, and two plots being nature reserves, NR. Different site indices are also indicated with different colours. Plots I-X may be managed for production of utilities.](image3)
conservation of both target and associated species is taken care of. It ought to be stressed that several species are dependent on human activities; many species would become extinct if no human activities are allowed in gene resource populations. Other species depend on virgin forests without any human intervention. It is important that both managed and untouched forests are included in the gene conservation of associated species. This is particularly well demonstrated for the picture from a *Fagus sylvatica* – *Abies alba* climax forest in Slovenia (11-8). In this nature reserve no management is permitted, which has resulted in regeneration of *Fagus sylvatica* only since all *Abies alba* seedlings were eaten by deer animals. If no management is allowed this mixed climax forest will turn into one-species climax forest over time.

**Miscellaneous**

Clone archives, provenance and progeny trials constitute a form of gene resource population. Their major merit is that they can be utilized for crossings in gene conservation and breeding. The genetic structure of the seed in a provenance trial is hard to predict. In most cases there will be a mixture of within-provenance crosses, crosses among the provenances in the trial, and hybrids between the provenances in the trial and the surrounding stands. Especially for wind-pollinated species with wide pollen dispersal, the latter will be significant. In many progeny trials the parents originate from one population and the seed crop will be more homogeneous than seeds from a provenance trial. Depending on the composition of the clone archives, the seed crop from them will either have a similar structure as in a provenance trial or in a progeny trial.

Botanical gardens mostly contain one or a few trees of each species and it might be questioned whether they should be regarded as gene resources. Some botanical gardens carry out an active gene conservation on annual and perennial herbs.

*Coppice* populations occur in some species like the situation for *Castanea sativa* (Picture 11-9). Mostly, this type of silviculture prevents a natural regeneration since any seedlings occurring are outcompeted by the vigorously growing stems in the coppice population. This means that no adaptation takes place in coppice populations and they are incompatible with the safeguarding of the potential for adaptation. Picture 11-10 was taken in a *Taxus baccata* forest in Sardinia. In this population there were no seedlings on the ground, and thus no regeneration. The shoots on the trunks were the only juvenile material. This presents an analogous situation to coppice forests of *Castanea sativa*.

Protected areas play a significant role in gene conservation in developing countries. There is an increased understanding that the previous philosophy of “hands off”, which means that any human intervention is banned from protected areas, has to be revised to reach goals in gene conservation. There is also an increased understanding that sustainable use of the natural resources in protected areas by local people ought to be integrated with conservation. Gene conservation in forests that are included in protected areas in future will easier reach its conservation goal if a management plan is developed for the protected area. There is a need that governmental organisations include non-governmental groups, indigenous people, community groups, and the private sector in establishment and management of protected areas. It is fundamental that an appropriate management of the protected area is carried out, such that forest gene resources are not lost or degraded inadvertently. If local people are involved in establishment of protected areas the risk for degradation or loss of such gene resources is probably much reduced.
It deserves to state once more that silvicultural treatments to guarantee regeneration of gene resource populations are of greatest significance in management of protected areas. Especially for tropical forests the regeneration issue is most relevant. Logging of very valuable tree species that occur at low densities, 1-2 mature trees per hectare, poses a risk for the continued survival of such species. In these cases measures to promote regeneration should be taken. Common sense gives ideas about how to promote regeneration:

- Logging of trees is permitted only if the tree has reached a minimum breast height diameter. This will allow younger trees to reach flowering age.
- Logging at optimum time, e.g. after maturation of seeds or fruits and not during main flowering.
- Opening of forests; the size of the openings is dependent on the stage in the ecosystem of the tree species. Pioneer tree species are light demanding while climax species usually are shade tolerant.

Finally logging for commercial reasons means that there may be some revenue of forests. This in turn may convey a message to local people that forests have an economic value. For that reason slash and burn for shifting agriculture hopefully will become less attractive.

**Species hybridisation and gene conservation**

Hybridization is common in some genera and must be regarded as a part of evolution in nature. The role of hybridization is somewhat controversial. Some regard hybridization as something that must be totally avoided while others regard hybridization as a means to increase the additive variance and thereby save genetic material that would be lost without hybridization. In Box 11-2 some possible outcomes of species hybridization are given.

In nature *Trochetiopsis erythroxylon* became extinct during the 1950s. Seeds were collected from the tree be-
fore extinction but they were affected by large inbreeding depression. During the 1980s two trees of the related species *Trochetiopsis ebenus* were detected. Crosses were carried out between these two species, which resulted in a vigorous offspring. This project did not lead to saving of the two species but genes from these two extremely rare species were preserved.

There are efforts going on to save the American chestnut, *Castanea dentata*, which just survives as root suckers owing to the devastating chestnut blight disease caused by *Cryphonectria parasitica*. Crosses are carried out with East Asian *Castanea* species to get resistant hybrids. Via back crossing it is hoped that the American chestnut with resistance against chestnut blight will be restored. Efforts are also made to identify and transfer resistance genes by modern molecular genetics methods.

A possible loss of Darwinian fitness in *Castanea sativa* populations in nature owing to hybridisation with East Asian *Castanea* species was discussed in connection with conservation of *Castanea sativa*.

Species hybridization may be a threat to rarely occurring species. If the hybrids between the rare species (red circle in Fig 11-15) and the common species have higher fitness than the rare species, the rare species may eventually be eliminated. Since the census number is much larger in the common species the gene flow will mainly be unilateral from the common to the rare species. If the hybrid between the two species has a higher fitness than the rare species, hybrids will increase in number at the cost of the rare species. Over time the hybrids will become more and more similar to the common species.

Another situation is illustrated in Fig.11-16. In this case the hybrid has an inferior fitness compared to both parental species. This is usually referred to as *outbreeding depression*. The rare species (red circle) may in such a situation waste its gametes in crosses with the commonly occurring species or the hybrid, leading to reduction in number of the rare species.

**Sustainable forestry**

Sustainable forestry, when it is being most environmentally conscious, can be regarded as a form of gene conservation. If all forest land has such a silvicultural regime, the gene conservation will be dispersed to all forests. As is evident from earlier chapters of this book there are no population genetics needs for such a gene conservation of a target species. Associated species might be dependent on larger number than required for the target species, especially if it is an ecological keystone species. The requirement is rarely so large that all forestland is needed for gene conservation. All projections of future demand for wood suggest a steady increase of the demand, partly owing to the dramatic increase in the human population. In this perspective many scientists advocate gene conservation on a landscape basis rather than on a stand basis. This means that certain forests constitute nature reserves, others have a total focus on production of wood, still others take an intermediate position to these two extremes. Such a landscape approach is anticipated to satisfy different objectives in a better way than one silvicultural regime over the entire area. A schematic sketch of how this can be achieved is given in Fig. 11-17. This illustration applies to the tropics in the first place but might in its purely forestry parts be applied outside the tropics as well. In New Zealand most of the wood is supplied from plantations of *Pinus radiata*. This is an example of a landscape approach and it has probably been of great importance for keeping the unique domestic forests untouched (Picture 11-11).

Figure 11-18 illustrates the decline in self-sufficiency of forest products in California from the Second World War until 1990s. The main reason for this is that forestland, to
an increasing extent, has been set aside as various kinds of nature reserves. In the early years, the demand for wood in California was satisfied by “imports” from the neighbouring states, Oregon and Washington. When larger areas in the latter states also were converted to nature reserves and the demand for wood increased, imports came also from British Columbia. During the last years the demand had to be satisfied by wood from tropical Asia. Thus, the strong protection and gene conservation of plants and animals in California have increased the pressure on forests in other countries. This pressure is most serious for the endangered forests in South-East Asia. The lesson to learn from this example is that we ought to have a global and landscape perspective on forest tree gene conservation.

We have tried to summarise this global perspective in Fig. 11-19. As stated several times the human population will increase dramatically. Human aspirations for a better life will also increase. Expansion of the human population will also result in increased air pollution. These conditions will lead to increased demand for wood and as corollary of this, forest decline and loss of biodiversity. Rapid global change means that large adaptability is required. All these conditions mean that we should not treat gene conservation isolated from production of human utilities. Such a production is much dependent on tree breeding. Thus, tree breeding, production of utilities, and gene conservation ought to be done in conjunction.
Some scientists have expressed a great fear that the adaptation that has caused an increased adaptedness of a population in nature will be destroyed if there is gene flow from surrounding, introduced populations. Such a gene flow is designated as introgression. Some have even called such gene flow pollution, which is an emotionally strong word. The use of the word pollution probably emanates from the belief that the adaptedness is perfect and that any gene flow will reduce the adaptedness of the recipient population and the belief that any gene flow will break a finely tuned genetic set-up of the recipient population leading to drastic reduction in fitness.

The latter belief requires that a specific adaptation to particular site conditions has taken place such that the activity of many genes depends on the presence of many other genes. This implies the evolution of what is often referred to as coadapted gene complexes. Once this specific combination of genes is broken up by crosses with alien pollen the adaptedness would be drastically reduced. Such situations probably exist but it is unlikely for the majority of wind pollinated tree species such as Norway spruce, Scots pine, loblolly pine, slash pine, Douglas fir, sessile oak, cork oak, sweet chestnut and many other wind-pollinated species with a wide and continuous distribution.

In southern Sweden many of the stands originate from eastern and south-eastern Europe. To evaluate if introgression from these introduced provenances to the Swedish Norway spruce populations is of any significance we need to know whether such an introgression causes:

1. a change but with a possibility to recreate the genetic constitution of the domestic population
2. a change that is irreversible.

To get an apprehension of which alternative might be true for Norway spruce we benefit from the knowledge that there is an additive gene action for important traits. As an example of additive gene action, data for budburst in different crosses are illustrated in Fig. 11-20. We have selected budburst since this trait is of great significance for survival of Norway spruce plants; mainly it is a question of avoidance of low temperature exposure of the frost sensitive stages just after budburst. In each part of the graph the mean values for the four hybrids northern x southern and the intra-provenance crosses are shown. As is seen from the graph the mean values of the hybrids are close to the means between the northern x northern and the southern x southern crosses. In one case the mean is somewhat closer to the northern cross, in the other the opposite situation prevails. From Norway spruce and Scots pine we have several such examples suggesting an additive gene action. This means that gene flow to an autochthonous south Swedish Norway spruce population will lead to a progeny that will be intermediate to the two origins. Especially if such hybridisations occur between widely differing populations there will be an increased additive variance. Via backcrosses it is possible to recreate the domestic population, certainly a very cumbersome task, but possible. It should be noted that we do not know how the situation is for the trees pollinated by insects, which fly over short distances and thus may give rise to specific adaptation (cf Fig. 6-16). For such tree species gene flow might be more serious.

For traits of adaptive significance, gene flow causes a change in allele frequency in the recipient population. Difference in gene frequencies among populations is one result of natural selection. Therefore, there is principally no difference between gene flow and natural selection; in both cases a change in gene frequency takes place. Pollution is too strong a word to use for gene flow, though it must be admitted that gene flow will in most cases reduce the Darwinian fitness of the recipient population.

**Different levels of a conservation programme**

For practical reasons one has frequently distinguished between conservation at the ecosystem, the species, and the gene levels. Even if there are practical reasons for doing so there is no biological reason for that (see Chapter 6). It was stressed that genetic differentiation between populations or species is the same type of process in a dynamic evolution, already envisaged by Darwin. To limit the conservation to the species level only, is a neglect of the fact that speciation is merely a part of the continuous evolution that takes place. Moreover, ecosystems are not stable or static since ecosystems are composed of species, which in turn are composed of populations, both of which are participants in a dynamic evolutionary process. Fossil data give support to this since they show that climatic change caused different migrations of the different components of an ecosystem. Thus, ecosystems did not migrate as ecosystems but rather each constituent species migrated independently of each other.
Summary

A gene conservation programme consists of three main components: objectives, genetic knowledge, and methods. Many objectives might be identified, of which the most important is to safeguard the future potential for adaptation of the species. Other objectives are to preserve the present genetic constitution and to have as a means for comparisons in the future. Preservation of the unique qualities of some populations used for production of highly valued human utilities is another objective in gene conservation. Simultaneous gene conservation and breeding, and conservation of associated species, are other objectives. Certain populations might be threatened and they deserve to be conserved for this reason. The methods in gene conservation should ensure that objectives in gene conservation are fulfilled while taking the structure and dynamics of the gene resource into account. For the majority of target species we lack knowledge about the genetic structure. From the ecological characteristics of the species educated guesses about the genetic structure have to be made in these cases. Gene conservation according to the Multiple Population Breeding System is the best way to meet the prime objective of gene conservation. This method cannot always be applied in its most sophisticated form, i.e. as ex situ plantations of the subpopulations. Less intensive variants of this method can be used by simply selecting the subpopulations in existing forests with the intention of safeguarding as much as possible of existing adaptedness. For the gene conservation of associated species an enlargement of some of the subpopulations to a few hundred hectares is suggested. Preferably, all stages of succession in the ecosystem should be represented. Since genotypes, populations, and species constitute components of an ecosystem it is biologically artificial to separate conservation methods for these three levels.

The essence of dynamic and static gene conservation is illustrated in Figure 11-21. It needs to be emphasized that regeneration of gene resource populations is crucial for long-term success of gene conservation. Since the demand for forestland is expected to increase in future it is important to minimise the risk that the gene resource population will be exploited for other purposes. Especially, for the developing countries it is important that conservation is tightly linked to local communities to guarantee their support of the conservation measures.

Finally, a landscape perspective should be applied in forest tree gene conservation.

Further reading


Forest genetic resources conservation and management, 3 volumes. 2001 - 2004. FAO, Danida Forest Seed Centre, and IPGRI.


![Figure 11-21. A synthesised summary of dynamic and static gene conservation.](image-url)
Consequences of different breeding activities and silvicultural methods for the new generation of trees

First the impact of breeding and silviculture on the progeny generation is presented. Since there are limited results on these topics the presentation has to be largely based on theoretical considerations. Finally the demands for genetic variation in the breeding and the production populations are outlined.

The first question to raise is: To what extent do various breeding or silvicultural activities lead to drastic genetic changes in the filial generation? The knowledge about the genetic consequences of different breeding or silvicultural activities is limited even if there was an increasing number of studies related to this issues during the late 1990s. Most of these studies were carried out with isozyme markers and they might therefore not well reflect what the consequences have been for traits of adaptive significance. However, if they do not reveal a loss of variability compared to the situation in natural stands it is unlikely that there have been losses in adaptive trait variability.

There are several occasions when there is a potential for genetic change during the course - breeding - raising of seedlings - silvicultural practice. In Fig 12-1 we have visualised a chronological step of events, which may have genetic consequences.

Seed orchards are almost always composed of trees from different stands. This means that relatedness, which might have existed in individual stands, will be broken when clones are brought together in a seed orchard. Therefore, the genetic variability in seed orchards is frequently higher than in natural stands. This has been confirmed by isozyme analysis.

As regards the progenies from seed orchards we know that changes have occurred both with respect to growth and stem quality even if we do not know the alleles that have caused this effect. Moreover, genetic change is what is aimed at in breeding. Breeding may influence other traits by being correlated with the selected trait either through pleitropy or close linkage. The knowledge of this matter is limited. Since the heritability is mostly low and the selection differential is not strong, we do not expect any great changes in traits not involved in breeding. The results available do not suggest any major changes in other traits.

Genetic drift and increased inbreeding might occur as a consequence of differences in female and male flowering frequencies among clones as well as asynchrony in receptivity and pollen dispersal among clones (cf Chapter 9). For tree species which are wind-pollinated and widely spread there is low probability that genetic drift or inbreeding will be of any significance in the progeny generation if the number of clones in the seed orchard is not extremely low. As pointed out in chapter 6 the loss of additive variance per generation is equal to $1/2N_e$. Most available data, which mainly originate from conifers, indicate that there is no genetic drift or inbreeding. The situation may be different in species which are not wind-pollinated and where the distribution is scattered.

Figure 12-1. A chronological illustration of factors in breeding and silviculture, which may influence the genetic composition.
Some studies showed that seed storage, seed germinability, and raising of seedlings had changed the genetic composition from that at harvest. The probability for deviations from the ideal composition increases with decreasing number of clones.

As regards clonal forestry the genetic variation might be narrowed down considerably. If the characteristics of the clones are well known we may also design clonal mixtures to have a specific genetic variation. Since self regeneration is not an objective of clonal forestry the occurrence of genetic drift or inbreeding is not relevant for clonal forestry. On the other hand this will be of interest if the clonal forest is used as a seed tree stand in the future. The amount of inbreeding will depend on the number of clones in the clonal plantation and gene flow from surrounding non-clonal forest plantations. In many countries the regulations for the number of clones required for clonal forestry do not give rise to any concern for considerable negative consequences of clonal forestry. Only for the cases when large areas have few clones and self regeneration is permitted there are potential risks that the following generation will suffer from inbreeding depression.

Studies on the impact of different regeneration methods on the genetic variability in the progeny populations have not revealed any large differences whether the studies were carried out in North-America, Australia, or tropical Asia. There was a tendency for the lowest genetic variability to occur in the offspring from unmanaged stands. In the tropical forests with low occurrence of many species it is useful to set limits with respect to the size of the trees that may be logged.

In one Malaysian lowland mixed dipterocarp forest, the effect of logging was studied approximately 40 years after logging. An adjacent unlogged stand was used as reference. Of the six species studied only one had a lower frequency of observed isozyme heterozygotes in the logged stands than in the unlogged control. The average increase of observed heterozygosity in the six species in the logged stand was slightly above 15 %.

A Canadian study of the mating pattern in forests with four different regeneration regimes, including self regeneration, showed high levels of inbreeding in all regimes. If this holds for other forests as well it means that self regeneration can lead to some inbreeding depression in the regenerated material.

The number of alleles present after different types of silvicultural activities is frequently reported. In some instances there are losses of alleles but it mostly concerns rare alleles, which probably are of no or very limited significance for additive variance. The effect of thinning on the expected heterozygosity was reported in a Canadian study of two stands. Thinning was carried out to promote the growth of Douglas fir. In Fig. 12-2 the change in species composition in the two stands is illustrated. This study is of significance since we will see the impact of a large change in species composition on genetic constitution. In the first stand the Douglas fir percentage increased from 72.4% to 84.6% while the corresponding figures for the

Figure 12-2. A The species composition before and after thinning of two Douglas fir stands. B The expected heterozygosity before and after thinning estimated by isozymes is given. The data for the two stands are given separately when data were obtained from both stands.
other stand were 51.4 and 75.9%. Two species, western redcedar and pacific silver fir, were lost in the thinning, which was not unexpected since they did not pass two percent in the stand before thinning. In spite of the large change in species frequency the expected heterozygosity did not change much in any of the remaining species. Even in western hemlock, which lost 7 rare alleles, the expected heterozygosity did not change.

Some geneticists have stressed that repeated cutting of the best trees over many generations may lead to genetic erosion, which sometimes is called dyogenic selection. The probability for this must be regarded as low. In spite of this, the stem form of the two oak species Quercus petrea and Q. robur suggests that such an erosion may have taken place in Denmark owing to repeated selective cuttings of the trees with the best stem form. When decisions are taken about which trees should be left in a seed tree stand it is not possible to select the phenotypically superior trees only, since the spacing after thinning must be fairly even in a seed tree stand. This means that the selection differential is smaller than if we consider only the proportion of trees remaining in the seed tree stand. As a corollary of this we do not expect any large genetic changes from thinnings.

Some caution as regards the relevance of these studies for adaptive traits is justified since all these studies were carried out with different types of markers.

Laymen frequently claim that the genetic variation in the production population must be as large as the variation in the breeding population. In the chapter about tree breeding we have learnt that a large additive variance is crucial for success in breeding. We should therefore focus on the additive variance when we are discussing the demand for genetic variation in a parental population. When the purpose of a population is not to produce a new generation we may simply talk about demand for genetic variation in that population instead of demand for additive variance. In Figure 12-3 two situations are illustrated schematically. To the left we assume that the production population will be used in self regeneration. To avoid genetic drift and inbreeding in this case, the demand for additive variance is as large as in the breeding population. If we need to keep the additive variance it means that we cannot use the same selection intensity as when we obtain the production population from seed orchards or vegetatively from clonal hedges (to the right in Fig. 12-3). In the latter case we might figuratively skim the cream off the milk. This causes a narrowing of the additive variance. This does not matter as long as we have genetic variation enough in the production population to cope with the environmental conditions during its life time. At each occasion when reforestation is to be carried out the plant material is obtained from the genetically most advanced

---

**Figure 12-3.** Schematic illustration of the requirements for additive variance and genetic variation in the production population at self regeneration of this population (left) or at artificial regeneration with the best material obtained from the breeding population at each occasion (right).
seed orchards or clonal hedges. Genetic progress may be much larger in this case than when the production population is to be used for regeneration.

Even if we never intend to use the production population for self regeneration the demand for genetic variation in the production population might vary. This has mainly been discussed in connection with clonal forestry. Figure 12-4 is based on a scientific paper from a group of scientists who have discussed the demand for genetic variation in relation to clonal forestry. The figure must not be perceived as exact differences in demand for genetic variation in any of the contrasting pairs. Above all, the distances between the two areas do not constitute any precise estimate of the demand for genetic variation.

The demand for genetic variation is least when the rotation time is short and when the cultivation takes place under uniform conditions. Heterogeneous environment and long generation times require broad genetic variation to mitigate the variations in time and space that might occur. To reach an understanding of what conditions require the largest genetic variation in the production population we might argue in an analogous way concerning the other contrasting pairs in this figure. In this context it is important to remember once again that both growth under heterogeneous environments and long rotation times probably promote evolution of large phenotypic plasticity in a species. If this is the case the demand is slightly reduced.

**Summary**

There is limited information about the consequences of different silvicultural methods and breeding activities for genetic changes in future production populations. All kinds of breeding aim at genetic change of the target traits. There are no signs that breeding has caused any dramatic changes in non-targeted traits. A large additive variance is required if self regeneration of the production population is envisaged. When the production population is replaced by material from the breeding population, the genetic variation can be less as long as the production population has enough variation to cope with the environmental conditions during its rotation time. The lower the demand for genetic variation the stronger the selection and the higher the gain.

**Further reading**


Adaptability The ability to respond genetically or phenotypically to changed environmental conditions.

Adaptation The process of genetic change of a population, owing to natural selection, resulting in a better adaptedness in a specific environment.

Adaptedness The degree to which an organism is able to live and reproduce in a given set of environments.

Additive gene action When alleles at two or more loci combine additively, the gene action is described as additive; this means that the value of a genotype with respect to alleles at several loci is the sum of the values attributable to the alleles at the separate loci. If alleles at the same or different loci interact the gene action is non-additive. See also Dominance and Epistasis

Additive genetic variance The part of the total genetic variance due to additive gene effects; the variance of breeding values. Additive variance can be exploited in mass selection.

Adenine A purin basis, one of the four nucleotide bases of DNA and RNA. Adenine is paired with thymine in the double helix.

After-effects If the performance of a progeny depends on the conditions during seed maturation or conditions during preceding growth period(s), there are said to be after-effects. This phenomenon has also been called pre-conditioning.

Allele One of two or several alternative forms of a gene that can exist at a single locus; if the number of alleles is larger than 2 the alleles form a system of multiple alleles; if the number of alleles in the same population is two or more and relatively common, the alleles are said to be polymorphic. Each individual chromosome has just one allele at each locus.

Allele fixation Allele fixation at a locus has taken place when there is only one type of allele at that locus in a population.

Allopolyploidy see Polyploidy

Allozyme see Isozyme

Alternative splicing different protein molecules are generated from the primary DNA transcript by changing the number and order of exons in the final mRNA after splicing out the introns.

Amplified Fragment Length Polymorphism (AFLP) A DNA marker that is based on the polymerase chain reaction (PCR) technique for amplification of restriction fragments and is exploited in the construction of genetic maps of chromosomes; the technique allows a high number of polymorphic loci to be detected. AFLP markers usually show dominance (the heterozygote cannot be distinguished from the homozygote).

Antipods The three nuclei located in the pole opposite to the egg apparatus in the embryo sac.

Archeogonium Female, multicellular, sex organ of most gymnosperms in which a single egg cell is produced.

Artificial selection Usually selection of superior phenotypes by man. See also Natural selection

Autopolyploidy see Polyploidy

Back crossing Repeated crosses with one of the original parents in each generation. Back-crossing is usually made to incorporate a single desirable trait from a species or a variety.

Base pair and base pairing The pairing between the nucleotide base adenine-thymine (adenine-uracil in RNA) or cytosine-guanine leads to formation of base pairs; commonly abbreviated bp.

Biclonal seed orchard see Seed orchard

Biochemical markers Qualitatively inherited genetic traits that are revealed by biochemical methods.

Bivalent A pair of homologous chromosomes, each consisting of two chromatids, appear during the first meiotic division; the number of bivalents is equal to half the chromosome number.

Bottle neck The occurrence of reduced effective population size during one or more generations.

Breeding population see Population

Breeding value The genotypic value of an individual judged by the mean value of its progeny. If an individual is crossed with a large number of randomly selected individuals in a population, its breeding value = the double deviation of its mean from the grand mean of this population. The breeding value is 2 x the general combining ability.

cDNA see Complementary DNA

Central dogma The flow of genetic information is from nucleic acid to protein, never in the reverse direction. More popularly, DNA makes RNA, and RNA makes protein.

Centromere The region of a chromosome that is essential for chromosome movements during cell divisions.

Chi-square test (χ² test) A statistical test to assess the goodness of fit between an observed and an expected segregation.

Chloroplast The cell organelle in plants in which photosynthesis takes place; chloroplasts have several circular DNA molecules containing approximately 120 genes. Chloroplasts are strikingly similar to cyanobacteria (previously called blue-green algae).

Chromatid One of the two subunits of a duplicated chromosome.

Chromosomal optimum The degree of polyploidy that gives rise to the most vigorous growth.

Chromosome In eucaryotes, a DNA molecule that contains genes in linear order to which numerous proteins are bound and that has a telomere at each end and a centromere. Chromosomes are dark-staining with basic dyes.
and microscopically observable in the cell during mitosis.

Cline A continuous change of population means along an ecological gradient attributed to changes in allele frequency, see also Ecocline

Clonal seed orchard see Seed orchard

Clone genetically identical individuals multiplied by vegetative propagation: grafts, cuttings, root suckers, somatic embryos.

Codominance Both alleles at a heterozygous locus can be identified phenotypically.

Codon A triplet of nucleotides in an mRNA molecule that codes for a particular amino acid or a stop signal in protein synthesis.

Coefficient of additive genetic variation. (CVA) The ratio between additive genetic standard deviation and the mean value of a trait expressed as percentage; CVA = 100σ²/x.

Coevolution Mutual evolutionary changes in two interacting species as a response to changes in these species, e.g. host-parasite interactions.

Combining ability Two types exist:

General Combining Ability (GCA) the value of an individual judged by the mean value of its progeny. If an individual is mated to a large number of randomly selected individuals in a population, GCA = the deviation of the mean of its progeny from the overall mean of the entire population. High GCA usually implies the presence of genes with additive effects.

Specific Combining Ability (SCA) specific pairs of parents after a cross give a progeny that strongly deviates from what is expected based on their general combining ability. High SCA usually implies the presence of dominance or epistasis.

Common tester see Mating design

Complementary DNA, (cDNA) A DNA molecule that is synthesised from an mRNA molecule by the enzyme reverse transcriptase; cDNA has no introns.

Controlled pollination Female flowers are isolated before they are receptive to prevent pollination with unknown male pollen. At maximum receptivity female flowers are dusted with pollen from one specific male or from a mixture of males.

Critical night length for budset The night length at which 50% of the plants belonging to a genetic entry are induced to form an apical bud.

Crossing-over Reciprocal exchange of non-sister chromatid segments in a pair of homologous chromosomes resulting in the recombination of genes within a linkage group.

Cytology Chromosome cytology deals with the microscopic studies of chromosome number, size, morphology, and behaviour during nuclear divisions.

Cytosine A pyrimidine base, one of the four nucleotide bases of DNA and RNA. Cytosine is paired with guanine in the DNA double helix.

Cytosol The remaining compartment of the cytoplasm in which the organelles have been excluded.

Darwinian fitness see Fitness

Degenerate code means that more than one codon encodes one particular amino acid.

Deletion Loss of a chromosomal segment.

Diallel crosses see Mating design

Dihybrid cross A cross between two individuals heterozygous at two different loci.

Dihybrid segregation see Segregation

Diploid An individual with two sets of homologous chromosomes (denoted 2x).

Disconnected half-diallel see Mating design

DNA, (deoxyribonucleic acid) The carrier of the hereditary material in most organisms; DNA is a double helix consisting of 4 nucleotide bases, (adenine, cytosine, guanine, thymin), one deoxyribose residue, and one group phosphate.

DNA fingerprinting A method to generate a pattern of DNA restriction fragments that is unique to an individual.

DNA library A collection of transformed cells each of which contains DNA fragments that represent the total genome of a species (genomic library) or contains cDNA fragments (cDNA library).

DNA replication Synthesis of DNA leading to the duplication of chromosomes.

DNA sequencing The technique for determining the base (nucleotide) sequence of a DNA molecule.

DNA vector A DNA molecule, that can replicate in a cell, into which a gene or a DNA segment has been inserted by recombinant DNA techniques; can serve as a vehicle to transfer a gene or a DNA segment to a host cell; bacterial plasmids are frequently used as vectors.

Dominance The interaction of alleles at homologous loci; the degree of deviation of the heterozygote from the mean value of the two homozygotes at the locus.

Dominant allele The allele (A) that is phenotypically expressed in a heterozygous (Aa) individual as well as in a homozygous individual (AA); at complete dominance both Aa and AA have the same phenotype.

Duplication A chromosome aberration in which more than one copy of a chromosome segment is present in the haploid genome. In a tandem duplication the two segments are adjacent.

Early test Such a test aims at identification of good predictors for adult performance in juvenile material.

Ecocline Sometimes used to distinguish clinal variation from ecotypic variation in a species, see Cline.

Ecotype Group of individuals in a species with a certain adaptedness to the conditions at a specific site.

Effective population size, (Nₑ) In a simplified version it is the number of individuals contributing to the filial generation.

Egg apparatus The egg cell and the two synergids in the embry sac in angiosperms.

Emasculatio Removal of male organs - anthers, male strobili - prior to pollination.

Embryo sac The female haplophase in higher plants usually developed from one of the macrospores.
Endemic  A species is endemic if it occurs naturally only within one area.
Endosperm  The triploid tissue in seeds of angiosperms. This term is sometimes erroneously used for the haploid female gametophyte in conifers.
Endosymbiont  An organism that lives in symbiosis in cells or tissues of another organism.
Endosymbiotic hypothesis  The proposal that microchondria and chloroplasts were originally free living organisms that entered into a symbiotic relationship with nucleated cells.
Epistasis  Interaction between alleles at different loci, i.e. denotes the non-additive effects between loci.
EST  see  Expressed Sequence Tags
Eukaryote  An organism in which the cells have a nucleus and other membrane-bound organelles, in contrast to a prokaryote like bacteria which lack these features. Fungi, algae, protozoa, higher plants, and animals are all eukaryotes.
Evolution  Cumulative change in the genetic composition of a population through time.
Expressed Sequence Tags, EST  A partial cDNA sequence, i.e. a sequence within the coding region of a gene.
Exon  see  Gene
F1  generation  Offspring from a cross between parents, progenies from crosses between F1 individuals are called F2, and so on.
F statistics  F statistics are useful means to get information on population differentiation and amount of inbreeding.
Factorial mating  see  Mating design
Family  The progeny from a controlled cross or from open pollination of one individual.
full-sibs  a progeny with both parents in common
half-sibs  a progeny with one parent, usually the female, in common.
Fitness  An expression for the average contribution of one allele or one genotype to the progeny of an individual in relation to the contribution of other individuals in the same population.
Darwinian fitness  The adaptedness in nature, which means the ability of an individual within a population to transfer its genes to the next generation, usually relative to that of other individuals within the same population, in contrast to domestic fitness.
Domestic fitness  The ability of a genetic entry to produce utilities for man.
Fns  an estimate of population differentiation for marker genes.
Gamete  A mature reproductive cell that is haploid.
Gene  A unit that transfers information from one generation to the next; is a segment of DNA of a chromosome (or RNA in certain viruses) with similar biochemical function; most genes in eukaryotes have (1) coding sequences (exons) that are transcribed to mRNA that in turn are transformed to proteins, (2) inserted non-coding sequences (introns), (3) a promoter, a regulating part that enables transcription, and (4) terminal (stop) sequences.
Gene bank  Collection of genotypes; seed bank, tissue culture bank, clone archives, genetic tests, the main objective of which is preservation of genetic material.
Gene cloning  Insertion of a DNA fragment, carrying a gene, into a vector molecule, such as a plasmid, capable of replication in the same or a different organism.
Gene conservation  
ex situ  in forestry it generally stands for storage or cultivation of a gene resource population.
in situ  in forestry it generally stands for a naturally regenerated gene resource population.
Gene fixation  see  Allele fixation
Gene flow  Migration to a recipient population from another population with a different allele frequency. For wind-pollinated species gene flow is mainly the result of pollen dispersal.
Gene frequency  The frequency of a gene in a population; gene frequencies are usually expressed as fractions of 1.
Gene map  The genes or small chromosomal segments that have been located to their respective chromosomes are arranged linearly in the map and the distances between the genes on the chromosome are usually known.
Gene resource population  see  Population
Geneecology  To study of adaptation to varying environmental conditions.
General combining ability (GCA)  see  Combining ability
Generative cell  In many gymnosperms, the cell of the male gametophyte that divides to form the stalk and spermatogenous cells; in angiosperms, the generative nucleus of the male gametophyte that divides to form two sperm nuclei.
Genetic code  The series of 64 triplets of bases, mRNA codons, each of which specifies one of the 20 amino acids in proteins and the signals for initiation and termination of polypeptide synthesis.
Genetic correlation  Correlation of breeding values, an estimate of the degree to which certain genes influence two different quantitative traits.
Genetic drift  Random fixation of alleles in small populations.
Genetic engineering, or recombinant DNA technology  The use of molecular genetics techniques to produce DNA molecules containing new genes or new combination of genes for the purpose of generating organisms with new desired characteristics.
Genetic entry  Stands for clone, clonal mixture, half-sib family, full-sib family, population or provenance.
Genetic gain  The mean progress of the progeny compared to the original population.
Genetic roguing  Culling of genetically inferior individuals (in seed orchards).
Genetic structure  The distribution of the genetic variation within and among populations.
Genome  One set of chromosomes; the gametes of diploid organisms have one genome, the gametes of polyploid species have two genomes or more.
Genotype (1) the sum of genes, the genetic constitution; (2) the alleles at one or more loci.

Genotype x environment interaction In a somewhat simplified way, a rank change of genetic entries from one environment to another.

G_{ST}, an estimate of population differentiation for marker genes.

Guanine A purin base, one of the four nucleotides of DNA and RNA; guanine is paired with cytosine in the DNA double helix.

Haploid chromosome number (n) The number of chromosomes in a haploid cell; gametes are haploid; the megagametophyte in conifers is haploid.

Hardy-Weinberg law The allele frequencies and genotype frequencies are constant from generation to generation in a random mating population with no selection, mutation or migration.

Heritability (h^2) The ratio of additive variance to phenotypic variance. The heritability of a certain trait is an estimate of the resemblance between individuals for that trait and it takes values between 0 and 1.

Heterosis Occurrence of increased size or vitality in hybrids compared with the parents or the parental generation.

Heterozygote An individual that forms more than one kind of gamete since it carries dissimilar alleles of one or more genes or dissimilar gene arrangement such as inversion and translocation heterozygotes.

Homologous chromosome Chromosomes that are identical with respect to size, form, and type of genes but the genes at a locus may differ. Diploid organisms have pairs of homologous chromosomes.

Homozygote An individual that carries the same alleles of one or more genes.

HOPE, Hierarchical OPen Ended A breeding system in which genes continuously and stepwise can be transferred via crosses to an elite population; the degree of improvement increases with each step.

House-keeping genes code for essential functions common to all or most cells in an organism.

Hybrid Progeny produced by mating of genetically different parents.

Inbreeding Selfing or mating between related individuals.

Inbreeding coefficient, F An estimate of identity by descent of alleles; identity by descent means that copies of one and the same allele in an ancestor have been brought together in an offspring.

Inbreeding depression Reduction of vitality after inbreeding.

Incompatibility Prevention of selfing or of mating between different individuals, usually caused by genes for self-incompatibility. The term is also used for the hindrance of good union of graft and root stock.

Intron see Gene

Inversion The reversal of the linear sequence of the genes in a segment of a chromosome owing to erroneous reunion of two breaks in the same chromosome.

Isozyme or allozyme Enzymes existing in different molecular forms but with function similar in character.

Jumping genes see transposones

Junk DNA DNA not encoding proteins or RNA, but may have other functions not yet identified.

Juvenile - mature correlation Correlation between the expression of a trait in the juvenile stage and in the mature stage.

Karyotype Description of the chromosomes of a species including chromosome number, size, and morphology; in some instances, the karyotype can provide information on the relationship between species.

Linkage The genes are not inherited independently of each other but rather as if they were linked to each other since they are located on the same chromosome. The larger the distance between two genes the weaker the linkage. Genes located far apart on the same chromosome usually appear unlinked, because at least one crossing-over will take place in the region between the two genes.

Linkage disequilibrium means that the alleles a_1 and b_1 always occur in the gametes of one parent and that a_2 and b_2 always occur together in the gametes of the other parent.

Linkage group All genes present in the same chromosome.

Locus (plural loci) Fixed position on a chromosome at which a gene is located.

Macrospore mother cell A cell that gives rise to the female gamete.

Marginal population A population close to the limit of distribution of a species.

Maternal effect Influence of the mother on the progeny that is not of genetic nature.

Mating design Systematic crosses of varying character:

- Common tester A special case of factorial crosses, usually a lower number of males than females are used.
- Disconnected half-diallel In a half-diallel only half of the possible crosses are carried out, disconnected means that the parents are split into groups, in each group a half-diallel cross is carried out.
- Factorial A mating design in which one group of parents are used as females and another group is used as males.
- Full-diallel mating A mating in which all parents are crossed with all other parents including reciprocal crosses.
- Nested One female may be mated to one series of males while another female is mated to another series of males.
- Partial diallel A limited number of the theoretically possible matings according to a full-diallel are carried out.
- Polycross Artificial pollination with a mixture of pollen from several individuals.

Mating pattern The matings that are realized, i.e. the zygotes formed in a population.
Mating system There are two major types: wind pollination and animal pollination; the latter type can be pollination by insects, birds, and bats.

Megagametophyte = prothallium The result of the free nuclei formation in the embryo sac. It is haploid.

Meiosis The process of nuclear division that leads to the formation of haploid gametes; the nucleus of the pollen mother cell or megaspore mother cell divides twice: in the first division the homologous chromosomes separate, in the second division the chromosomes divide.

Mendelian inheritance The rules of hereditary transmission from one generation to the next; two alleles of a gene segregate from each other in meiosis and pass to different gametes; alleles belonging to different loci segregate independently, and combine randomly in the progeny, except for loci near each other on the same chromosomes, so-called linked genes; this is valid for qualitative and quantitative traits but are usually impossible to detect for quantitative traits.

Messenger RNA (mRNA) The information stored in DNA is transcribed to mRNA, which in turn translates it into proteins.

Microarray A technique, which reveals the genes that are active in a particular tissue at a particular moment.

Microsatellite Highly repetitive, polymorphic short tandemly repeated sequences of DNA; 2-6/8 base pair repeat units; also called tandem repeat (STR) or simple sequence repeat, (SSR). Microsatellites can be used for DNA fingerprinting.

Mitochondrion An organell, 1-3 µm x 1 µm, occurring in each eucaryotic cell. Mitochondria are the most important energy sources in cells and contain enzymes involved in the final steps of oxidation of organic material to carbon dioxide and water.

Mitosis The division of the nucleus in somatic cells leading to the formation of two daughter nuclei, which are enclosed in two separate cells after the division is completed.

Molecular clock Based upon the hypothesis that mutations in a gene occur at equal rate during the course of evolution as long as the function of the gene is unchanged.

Monohybrid segregation see Segregation

MPBS see Multiple Population Breeding System

Multiple alleles see Alleles

Multiple Population Breeding System (MPBS) Split of the breeding or gene resource population into approximately 20 subpopulations that are cultivated under different environmental conditions or are exposed to different selection criteria.

Mutation A chemical change in DNA or a structural change of DNA; mutations are neutral if they do not change the fitness of the organism.

Natural selection Improvement of adaptedness via differential transfer of alleles to the next generation. It requires that there is a genetically conditioned phenotypic variation causing variable fitness.

Directional natural selection individuals with extreme phenotypes in one tail of the distribution contribute more to the progeny generation than others.

Disruptive natural selection individuals with extreme phenotypes in both tails of the distribution contribute more to the progeny generation than others.

Stabilizing selection individuals close to the mean of the distribution contribute more to the progeny generation than others. See also Artificial selection.

Nested mating design see Mating design

Night length The duration of the dark period.

Nonsense DNA see Junk DNA

Norm of reaction The phenotypic expression of a genetic entry along an environmental gradient.

Nucleolar organizer A region of the chromosome containing the genes for ribosomal RNA, also called secondary constriction.

Nucleosome Structural component in eukaryotic chromosomes. It consists of 8 histones (2 of each of H2A, H2B, H3, H4) which are proteins binding to DNA. This structure is called octomer. The DNA-molecule is wound twice around the nucleosome.

Nucleotide The building block of nucleic acids; it is composed of a sugar molecule (deoxyribose in DNA, ribose in RNA), a phosphate group, and an organic nitrogen base (purine or pyrimidine).

Nucleus breeding The breeding population is split into one small nucleus population (usually 50-70 trees) and one large subpopulation (= 250 trees); the selection intensity is largest in the nucleus subpopulation; over the generations the gap in progress between the two subpopulations will be broadened.

Nutrient efficiency a plant’s ability to produce biomass in relation to available nutrients, whether it can be attributed to uptake of nutrients from a substrate or to utilization of nutrients.
Nutrient uptake The content of nitrogen measured in the whole plant or in parts of it. Oligonucleotides a linear sequence of about 10-20 linked nucleotides, natural or synthetic.

Open pollination There is no human influence on the seed formed, selfing may occur; seeds from open-pollinated families have been collected from individual mother trees following wind or animal pollination. Ortet Original plant, the plant or the tree which is the founder for vegetative propagation.

Outbreeding The opposite to mating between related individuals. Partial diallel mating see Mating design PCR see Polymerase Chain Reaction Pedigree A record of ancestry, often shown as pedigree diagrams representing the familiar relationships among relatives, such as full-sib and half-sib families. Phenology Timing of periodic phenomena such as bud burst, budset, flowering, especially related to seasonal changes in temperature and photoperiod. Phenotype The observable properties of an individual; the sum of the characteristics of a certain genotype at a certain occasion. The phenotype is determined by the genotype and the environment: phenotype = genotype + environment. Phenotypic plasticity the amplitude for a trait of a genotype studied in at least two different environmental conditions. Phenotypic variance The variance of the assessed values of a trait; phenotypic variance = genotypic variance + environmental variance. Photoperiod The length of the period with daylight. Photoperiodic response A type of change initiated by changes in the relation between day length and night length. Phylogeny the evolutionary history of a group of organisms or genes. Phytotron A series of growth chambers in which several environmental factors such as temperature, photoperiod, and air humidity can be regulated. Plasmid Usually a self-replicating segment of DNA and independent of the host cell; plasmids occur mainly in bacteria but also in eukaryotic cells. Plasticity see Phenotypic plasticity Plot The smallest research unit e.g. in a field or a nursery trial consisting of for example a single provenance. Plots are assembled into blocks. Plus tree Selected tree with superior phenotype. Pollen contamination Pollination with alien pollen either in stands or seed orchards. Pollen mother cell A cell that gives rise to the male gamete. Polycross see Mating System Polymembryony Occurrence of more than one embryo in each seed. Polygenic inheritance see Quantitative inheritance

Polymerase Chain Reaction (PCR) A technique that results in exponential amplification of a specific region of double-stranded DNA. Polymorphism The occurrence in a population of two or more alleles at the same locus; the most common allele has a frequency of less than 0.95. Polypeptide A chain of amino acids linked together by peptide bonds; a protein consists of one or more polypeptide chains. Polyploidy Occurrence of more than two complete sets of chromosomes. Allopolyploidy Polyploidy as result of species hybridization. Autopolyploidy Polyploidy that has arisen after chromosome doubling within a species. Population Usually a collection of individuals from a limited area that have a certain degree of adaptedness to that area. Breeding population The collection of trees that will carry the advancement of breeding into future generations. Gene resource population the seeds, acorns, nuts, plants, or trees that are included in the gene conservation. Production population A population intended to produce human utilities. Propagule population The plants or trees utilized in sexual or vegetative propagation. Population genetics Studies of gene frequencies in populations and their changes. Preconditioning See After-effects. Production population see Population Progeny trial A trial in which different families are tested. Promoter A sequence of double-stranded DNA upstream of the start of transcription at which RNA polymerase binds and initiates transcription of the structural gene. Propagule population see Population Prothallial cell The sterile cell or cells found in the male gametophytes of gymnosperms but not in angiosperms; believed to be remnants of the vegetative tissue of the male gametophyte Prothallium see Megagametophyte Provenance One definition of provenance is a population or group of individuals of the same species occurring within or originating from one more or less rigorously defined geographic area. Provenance hybrid seed orchard see Seed orchard Pseudogene A non-functional gene with sequence homology to a functional gene elsewhere in the genome. Purine A nucleotide base with two carbon-nitrogen rings; adenine and cytosine are purin bases. Pyrimidine A nucleotide base with one carbon-nitrogen ring; guanine, thymine and uracil are pyrimidine bases. Qform an estimate of population differentiation of quantitative traits.
Quantitative trait locus (QTL) The genes in such loci participate in the regulation of quantitative traits.

Qualitative inheritance One gene strongly influences the phenotype.

Quantitative inheritance Genes in many loci influence a trait, the influence of each gene is usually small.

Ramet An individual obtained from vegetative propagation; a member of a clone.

Random genetic drift see Genetic drift

Random Amplified Polymorphic DNA (RAPD) A DNA marker that is based on the polymerase chain reaction (PCR) technique for amplifying specific DNA fragments by using arbitrary 10-base oligonucleotides as primers; RAPDs are usually dominant i.e. the heterozygote cannot be distinguished from the homozygote.

Real-time PCR A real-time PCR machine follows the amplification of the DNA sequence in real time, using fluorescent markers. This allows accurate quantification of e.g. gene expression.

Receptivity The stage of female flower or strobilus at which success of pollination is expected.

Recessive An allele that is only expressed when homozygous.

Reciprocal crosses Two crosses in which each parent serves as female in one of the crosses and as male in the other; female A x male B and the reciprocal cross Female B x male A.

Recombinant DNA DNA created by bringing together DNA segments often from different species.

Recombination The creation of new combinations of genes in F$_2$ through segregation of chromosomes and crossing-over at meiosis e.g. a$_1$a$_2$b$_1$b$_2$ and a$_1$a$_2$b$_2$b$_1$ may be obtained in F$_2$ following the original cross a$_1$a$_2$b$_1$b$_2$ x a$_1$a$_2$b$_2$b$_1$.

Recurrent selection Selection repeated over several generations to obtain progressive change.

Regulator gene A gene that regulates the expression of another gene. Many regulator genes encode transcription factors, which are proteins binding to the promoter region and influence how actively the gene is transcribed.

Repetitive DNA Certain sequences are repeated many times in the haploid genome, even up to one million times. It comprises 70-80% of total DNA in conifers.

Replication see DNA replication

Rest Budrest or bud dormancy is 'the temporary suspension of visible growth of any plant system containing a meristem'; a meristem is a tissue where new cells are formed by cell division; budrest is built up in the buds soon after budset and prevents an untimely budburst; budrest is broken by temperatures a few degrees above zero or by long nights.

Restriction enzymes restriction endonucleases are site-specific enzymes each recognizing specific DNA sequences and cleaving DNA at these sites producing DNA fragments.

Restriction Fragment Length Polymorphism (RFLP) A DNA marker in which the size of the fragments varies within a genetic entry, such as population; RFLPs are codominant (both alleles are expressed) and selectively neutral.

Retrospective early test Studies of young siblings of the families studied in field trials.

Ribosome A cellular organelle on which the translation of mRNA into amino acids in protein synthesis occurs.

Ribosomal RNA (rRNA) RNA molecules that constitute part of the structure of ribosomes.

RNA, ribonucleic acid It consists of a chain of nucleotides linked through the phosphate groups. Each nucleotide contains the sugar ribose, and one of the four bases adenine, cytosine, guanine, and uracil. RNA is typically single-stranded unlike DNA.

Satellite DNA Highly tandemly repeated DNA sequences located on both sides of the centromere or at the chromosome ends.

Secondary embryo sac The merged two nuclei in the centre part of the embryo sac.

Specific Combining Ability, SCA see Combining ability

Seed orchard An establishment for production of genetically superior seeds.

Clonal seed orchard Grafts or cuttings are used.

Seedling seed orchard Seedlings of full-sib or half-sib families are used.

Seed orchards can further be classified as:

Biclonal seed orchard Two clones are used, its main implementation is for clones with high specific combining ability.

Monoclonal seed orchard One clone is used mainly aimed for seed production after artificial mass pollination.

Interprovenance seed orchard The genetic entries originate from two or more provenances.

Interspecific seed orchard The genetic entries originate from two species.

Intraprovenance seed orchard The genetic entries originate from one provenance.

Seed tree stand A stand within the best provenances; in many countries approved for seed harvests by a federal organisation.

Seedling seed orchard see Seed orchard

Segregation Separation of the two alleles of a gene into different gametes at meiosis.

Selection see Artificial selection and Natural selection

Selection backward Selection of parents using data from progeny tests.

Selection differential, S The difference between the mean of the selected part of the population and the overall mean of that population.
Selection forward Selection of trees in progeny trials for generating a new breeding population.

Selection intensity, i the selection intensity is obtained by dividing the selection differential by the phenotypic standard deviation, i.e. the standardized selection differential.

Selective Environmental Neighbourhood, SEN An area within which there is no genotype x environment interaction as regards fitness which means that there is a large homogeneity within an SEN.

Self-fertility Ability to form viable offspring by fusion of female and male gametes from the same individual.

Self-sterility Inability to form viable offspring by fusion of female and male gametes from the same individual.

Self-sterility alleles Alleles that prevent selfing which means that a tree with the self-sterility alleles s1 or s2 does not form any seeds if the pollen grains contain s1 or s2. It does not matter whether the pollen originates from the same tree or another tree; the female tissue prevents fertilization with pollen containing these alleles. Conifers do not seem to have self-sterility alleles.

Selfing Fusion of female and male gametes from the same individual.

Selfish DNA Sequences use their host for propagation only, apparently without being of any use for the host.

Semi-conservative replication after replication of DNA the newly generated double helices consist of one old strand and one new strand

SEN see Selective environmental neighbourhood

Severity index The expected plant mortality in per cent of the local population 20 years after establishment of the test plantation. The reason for using such a high age as 20 years for establishment is that the results have shown that it may take 20 years before knowledge about hardiness is complete.

Single Nucleotide Polymorphism, SNP It is caused by the change of a single nucleotide. Most genetic variation between individual humans is believed to be due to SNPs.

Speciation The differentiation between two populations has gone so far that they have become reproductively isolated from each other and therefore gene flow between them is prevented; two main types of speciation are distinguished:

Allopatric The speciation takes place in geographically separated populations.

Symbiotic The speciation takes place in geographically common area.

Specific combining ability see Combining ability

Spermatogenous cell The cell of the male gametophyte of gymnosperms, which divides mitotically to form two sperm nuclei.

Stalk cell One of the two cells produced by the division of the generative cell in developing pollen grains of gymnosperms; it eventually degenerates.

Status number An estimate of the size of a population comprised of unrelated trees. A breeding population of 50 trees may have a much lower status number than 50 owing to various degrees of relatedness among the 50 trees.

Strobilus (plural strobili) Reproductive structure in Pinaceae; the pollen cone consists of microsporophylls with microsporangia containing pollen grains; the seed cone consists of ovule-bearing scales, the ovules contain egg cells.

Sublining The breeding population is divided in smaller populations, sublines, so that inbreeding is avoided in the production population, but permitted in each subline; from each subline one clone is selected for establishment of seed orchards for production of commercial seed.

Synergids Usually the two cells located adjacent to the egg cell in the embryo sac.

Synteny Partial conservation of gene order among species.

Target species A species given priority in gene conservation for scientific reason, threat, charisma, or economic reason.

Telomere The DNA sequence at the end of a chromosome that provides stability to the chromosome.

Terminator region Includes the stop codons for termination of the polypeptide (protein) synthesis.

Tetraploid Species or individuals with four chromosome sets (denoted 4x).

Thymine A pyrimidine base, one of the four nucleotide bases of DNA; thymine is paired with adenine in the DNA double helix.

Transcription The synthesis of an RNA transcript on a DNA template.

Transcription factor A protein that activates the initiation of eukaryotic transcription, either at all loci (general transcription factor) or at specific loci (specific transcription factor).

Transformation, stable The incorporation of a new gene(s) into the host cells’ genome using genetic engineering.

Transgenic plant Plants into which genes have been transferred using genetic engineering.

Translation The synthesis of a polypeptide whose amino acid sequence is determined by the codon sequence of an mRNA molecule.

Translocation Interchange of chromosomal segments between non-homologous chromosomes.
Transposon A piece of DNA that can move spontaneously from one position to another within the same chromosome or between chromosomes; also called jumping gene.

Triplet The three nucleotide pairs that constitute a codon.

Triploid Species or individuals with three chromosome sets (denoted 3x).

Transfer RNA, tRNA A small RNA molecule that serves in protein synthesis; it binds an amino acid, specified by tRNA anticodon which pairs with a codon on the mRNA, and tRNA delivers its amino acid to the growing polypeptide during translation of mRNA.

Tube cell In male gametophytes, the cell that develops into the pollen tube.

Uracil One of the two pyrimidine bases found in RNA, it is replaced by thymine in DNA.

Vector, cloning A DNA molecule capable of replication in a host cell, into which a gene or DNA segment is inserted by recombinant DNA techniques and can serve as a vehicle for transfer of DNA to a host cell.

Wahlund’s principle The frequency of homozygotes decreases in the progeny after matings among individuals of two previously isolated populations.

Water use efficiency The ratio of carbon gain to water losses.
The two orchids in the Siberian taiga serve as examples that conservationists should design forest tree gene conservation such that associated species are included in the conservation efforts. Photograph Gösta Eriksson