TISCIA 33, 9-30

MOLECULAR MARKERS IN ECOLOGY

Zs. Pénzes, Gy. Csanádi, G. M. Kovács and Zs. Beer

Pénzes, Zs., Csanádi, Gy., Kovács, M.G. and Beer. Zs. (2002): Molecular markers in ecology. — *Tiscia 33, 9-30.*

Abstract. The aim of this paper is to encourage reluctant ecologists thinking more on the use of genetic markers. No training in molecular methods is supposed. The resistance may be explained by the confusion of the high number of methods, including laboratory techniques (named by up to four characters) and evaluation. Inevitably, field work needs some extra step to collect DNA source, but sampling strategy is the same or even less restrictive owing to the new, powerful statistical methods. Laboratory techniques develop very fast, many phases can be done automatically and/or many customers provide services on reasonable price.
Despite the financial and technical requirements, genetic markers provide high quality information that can be obtained hardly otherwise, or simply impossible. We try to overview the most important genetic markers and technology used recently. Some examples are given by studies of parentage, population structure (migration, fragmentation) or population history.

Keywords: DNA techniques, DNA markers, molecular markers

Zs. Pénzes, Department of Ecology, University of Szeged, H-6701 Szeged, P.O. Box 51, Hungary, **Gy. Csanádi**, Department of Biotechnology, University of Szeged, H-6726 Szeged, Temesvári krt. 62., **G. M. Kovács**, Veterinary Medical Research Institute, Hungarian Academy of Sciences, H-1581 Budapest, P.O. Box 18, Hungary, **Zs. Beer**, Department of Forensic Medicine, University of Szeged

Introduction

Until the mid-1960-s, population genetics was mainly a theoretical science. Now, this view is changed and the impact of population genetics increased considerably in many fields of biology (Sunnucks 2000). Suddenly, the mainly theoretical quantities of population genetics (percent polymorph loci, inbreeding coefficient, population structure etc.) are inevitably incorporated into the terminology of other fields of biology, including conservation biology (Smith and Waine 1996, Hedrick 2001) and ecology (Snow and Parker 1998, Baker 2000a). Population genetics studies the underlying processes of genetic variation, defined in samples of individuals from different populations and species. Nowadays, the genetic variation is not only a criterion for natural selection to work. Genetic variation is a huge source of information about the biology of individuals carrying a given variant. It also keeps track the history and spatial relationships of populations composed by these individuals. This information can be extracted using genetic markers.

The use of genetic markers is not limited to the age of modern genetics. They were essential tools also of the classical breeders who followed the inheritance of selected traits of their plants and/or animals. This aspect sheds light on the two main characteristics of a genetic marker: (1) it must be distinguishable and (2) must be inherited genetically. The genetic basis of having markers in an individual or in a population is the presence of different alleles on a given genetic locus, which results in a genetic variability or genetic diversity or genetic polymorphisms. These expressions are synonyms describing the same thing: there are genetic differences between individuals which can be used for making genetic analysis on them for different purposes like individual identification (genetic "fingerprint"), genetic mapping, breeding or to describe their genetic relatedness.

For centuries only visible phenotypic differences could be used as markers such as colours, spots, heights and weights and other morphological characteristics, which caused a serious limitation in genetic analysis, especially in plants. Despite of the

fact that Gregor Mendel produced his historical results in a plant system, these organisms have unfortunately low amount of characteristics, which can be used for genetic analysis. In addition, many of the visible and important characteristics are not represented by a single genetic locus but coded by multiple genetic loci (e.g. height, weight), which makes their use as genetic markers almost impossible. Therefore the appearance of molecular genetic markers was a revolutionary breakthrough in determining and using genetic polymorphisms for different purposes. The first such molecular markers were the protein polymorphisms which were based on the isoenzyme variations of the same polypeptide. Using this methodology the number of detectable polymorphisms increased significantly indicating that most of the available genetic differences remain invisible at the morphological level (Hubby and Lewontin 1966, Lewontin and Hubby 1966).

The discovery of DNA markers had really enormous effect on the field of genetic analysis unveiling the waste majority of hidden allelic variations. Application of solution-based DNA-DNA hybridisation technique is an important step in this process (Bledsoe and Sheldon 2000). The revision of bird phylogeny by Sibley and Ahlquist (1990) stimulated biologists to study DNA variation from a different point of view: the use of DNA markers. The methods are changed, but molecular phylogenetics is one of the biggest "end-user" of genetic markers today (e.g. Hillis *et al.* 1996, Page and Holmes 1998, Nei and Kumar 2000). The growing number of inferred phylogenetic trees provides the basis of the comparative methods, which is one of the most important tools in the evolutionary biology (Harvey and Pagel 1991).

Another important step was the introduction of the Southern-hybridisation based RFLP (Restriction Fragment Length Polymorphism) technology to eukaryotic genetics in the early 1980s which made possible for the first time to investigate the genetic differences directly at the genotype level in higher taxonomic level. This was followed by the elaboration of the polymerase chain reaction (PCR) technique at the early 1990s, which became an essential tool for modern geneticists because of its amazing simplicity, effectiveness and versatility. With appropriately selected piece of DNA sequence (the primers), small amount of DNA can be amplified to usable quantity. But the technical refinements are not yet finished. Highest variability can be obtained on the level of DNA sequence. Every single base accessed by automated methods of DNA sequencing could have high importance in an ecological analysis. On this ground, the number of markers useful for a given topic grows continuously. Recent success of highly variable microsatellites (SSRs, single sequence repeats) provides an excellent example. Molecular methods generated new data, discoveries and controversies that stimulated new theories and development of powerful statistical methods, available almost at once in user-friendly computer programs (e.g. Luikart and England 1999). These facilitated further the study of variation in DNA related to fundamental questions of biology. These key innovations together reveal otherwise unobtainable information at all level of biotic hierarchy (e.g. Avise 1994, Ferraris and Palumbi 1996, Smith and Waine 1996, Burke *et al.* 1998, Goldstein and Schlötterer 1999, Baker 2000a). Many efforts are made to standardise methodology and getting comparable results (e.g. EU Molecular Screening Tools project, see Karp *et al.* 1988).

To give a rigorous definition, genetic markers are simply heritable characters with multiple states at each character (Sunnucks 2000), where character means a given genetic locus and character state is defined by the alleles. Furthermore, individuals have many loci and separate loci can provide independent characters. Applying the same logic to lower organisation level, even nucleotides in any position of the DNA sequence may be interpreted as character with four states. This means that with sufficient variability, genetic markers may provide never seen discriminatory power.

The information can be obtained about loci defines two types of genetic variation. We speak about genotypic variation where the genotype is of interest. Genotypes of multiple single loci compose a genotypic array. Sometimes allelic correlations are also needed (linkage disequilibrium). In other words, the gametic phase or haplotype pair is identified in a given individual. Genic variation means that alleles or haplotypes are of interest, but not their combination into genotypes. On a pooled analysis carried out on population level, we may have frequency data from alleles of a given locus, even if we have no imagination about the genotype of individuals.

Every individual organism has unique combination alleles and DNA sequence. This gives unique set of markers on the lowest level of biological hierarchy and provides extreme sensitivity of the genetic markers. As Waser and Strobeck said (1998): "very few birds have bands but all have genotypes". But, according to the nature of genetic differentiation, the rate of change of the different kind of markers may vary. This rate is influenced by the processes of recombination, mutation and selective constraint. Of course, models of population genetics can be applied to DNA markers, too. In this

way, markers can be used for answering questions on different scale in space and time: different markers can be selected for different scales. On the other hand, genetic variation itself is organised hierarchically (irrespectively of species concept): alleles within individual, individuals within (sub)population, or even metapopulations. Besides the scale of study, appropriate marker choice depends also on the level of hierarchy.

The same conclusion can be drawn on a different way. The fate of a given individual both in space and time depends on its biology and its environmental interactions. These may be reflected in the DNA sequence (often referred as the "natural history of DNA") resulting in polymorphism. Fundamentally, this polymorphism is shaped by the known processes of mutation, recombination or selection and their connection, as exemplified by population genetics for at least eighty years. In this way, the theory is simple: by measuring the genetic variation and translating it to useful quantities with the models of population genetics, we can make inferences about the history and/or relationship of populations or the biology of organisms. Information can be obtained about the biology or many population level processes in space and time by examining genetic markers with appropriate rates of change. Just to mention some of them: mating structure, relatedness, population structure (e.g. migration, heterogeneity) or population history (e.g. bottleneck events). Of course, this logic can be extended to the scale of evolutionary processes as it is demonstrated by the enormous literature of molecular phylogenetics.

There are many factors should be weight for selecting DNA marker for a given problem. The classification of markers on the basis of these factors is detailed below. In less rigorous way, three levels of molecular change are generally separated (e.g. Sunnucks 2000). Levels are related to different aspect of population biology, claiming for different markers.

The most sensitive markers are the genotypic arrays (e.g. multiple microsatellite loci scored in individuals). The relevant time scale is the generation and the rate of change depends mostly on the frequency of recombination. Genotypic arrays are used in the shortest and finest scale studies. Some examples of usage are individual identification, parentage (paternity) and relatedness.

Genic analysis (e.g. microsatellite, mitochondrial DNA (mtDNA)): individual genes are considered and allele frequencies, geographic distributions are recorded. These properties change on larger scale. Typical examples are the studies of gene flow (migration) or reconstruction of population history.

Relationship of alleles (creation of new alleles by mutation) provides information on larger, evolutionary scale. It is used for studying long-term processes of phylogeography, speciation or phylogenetic reconstruction. This level is beyond the scope of this paper.

Our aim is to give a short overview, classification of the fundamental markers and techniques proven to be useful in ecology. This topic is covered by excellent books (e.g. Baker 2000a) or reviews published in journals of ecology (e.g. Snow and Parker 1998) written for non-molecular biologists. We try to highlight information from these on the basic level or sometimes update them by some new important results. Reviews are referred whenever possible as they are available more generally and further details and references can be found in them. We focus on how to collect and store sample or the reason why a given marker should be used for solving a special question. The advent of statistical methods and software is outside of our topic, mentioned shortly in the discussion. Furthermore, this paper is based primarily on our experience in using these methods even in ecological laboratory. That is why some of the important class of markers is not covered here (e.g. MHC) or mentioned shortly (e.g. mtDNA). Other markers are simply ignored as they have (more) interest in taxonomy or evolutionary biology (e.g. interspersed class of repetitive sequences, like SINEs and LINEs) but not in ecology. Abbreviations are given at the end of the paper. Finally, the importance of this field is proven also by statistical indices of its bibliography. The leader journal in this topic is the Molecular Ecology being among the most cited primary ecological journals (source: ISI, cited in Sunnucks 2000).

DNA sample

The success of molecular analyses depends upon collecting the most suitable samples and storing them in a proper way as to minimise damage. Properly handled DNA sample can be used for many different analyses. Molecular techniques require very small amounts of DNA (typically less then 5 μg) permitting non-destructive sampling of large number of individuals. In theory, one molecule of DNA template is enough for a PCR amplification. Not only the quantity but the quality of the DNA is important, depending on the method (see there). Properly prepared and stored samples are needed not only to carry out the present study but for the unpredictable

future possibilities. As new techniques are being developed, old samples can be reanalysed again with more efficient methods. This may happened many times in literature and it worth to keep in mind.

The logic followed in this chapter based primarily on Carter's (2000a) work. It provides also a review of sampling, preservation considerations and fundamental techniques not explained here. Good and continuously updated protocols are given in the cited reviews. We suggest also Karp *et al.* (1998) as it contains information about equipment and safety considerations. Shortly, the basic steps of molecular analyses are the followings:

1. Sample collection
2. Sample storage
3. Extraction of DNA
4. Preliminary modifications (e.g. digestion)
5. PCR or hybridisation
6. Visualisation (electrophoresis, sequencing)
7. Interpretation of the results.

In this section, sample handling (1-3.) and general techniques are overviewed in some details. Marker technologies with their special requirements are considered in the next section (4-6). Step 7 is outside of our scope, but some notes are given in the discussion.

Sample source

Almost any cells contain a copy of the genome and can be used to extract DNA. Owing to the advances of PCR, even the catching of individuals can be avoided: feathers of birds or hairs and faeces of mammals can be potential sources of DNA. These non-invasive techniques are of great interest (Taberlet *et al.* 1999), especially in conservation biology. For example, mtDNA and microsatellites markers can be assayed from faeces (for the topic of molecular scatology, see e.g. Kohn and Wayne 1997) or urine (Valiere and Taberlet 2000).

Nevertheless, not all cells and tissues are equally suitable sources of DNA. Secondary metabolites may interfere with the enzymes used to manipulate DNA: high concentration of melanin inhibits *Taq* polymerase used in PCR or polyphenols in plants inhibits many DNA modification enzymes. Bones or woody plant tissues are very hard and difficult to work with. In these cases some extra steps are needed to extract DNA.

Blood is one of the most convenient samples for DNA isolation in non-mammalian vertebrates, as they have nucleated red cells. As little as 1-2 µl of blood is sufficient, but usually much more is taken by venipuncture. For mammalians, 1-2 ml would be needed to work with easily. In some cases, tissues may be preferred, like skin biopsies. For insects, the whole individual or muscles may be used. Pigmented parts should be avoided. For plants, young leaves (or needles) are the most obvious sources (more cells per weight and less polyphenols and polysaccharides). "Ancient DNA", even thousands of years old may be isolated from bones and teeth and used for PCR amplification and sequencing (e.g. Hagelberg and Clegg 1991). Preserves used in museum collections (like formalin and Carnoy's solution) are not good for DNA preservation: usually large genomic DNA can not be isolated but PCR-based techniques can be used.

Sample preservation

DNA is very stable and robust, if it is handled properly. It is prone to damage by nucleases, chemical degradation, extremes of pH, mechanical shearing, excessive heat and strong light. It is a good idea to extract the DNA as soon as possible and store it –80 oC. Freshly collected samples are the best for DNA isolation but even these must be preserved until the extraction. There are many buffers suggested in the literature but they vary in efficiency. Generally true that the ratio of the preservative and sample must be kept.

The two basic methods of preservation are the chelation and dehydration. Agents such as EDTA are able to chelate the free magnesium ions (essential to activity of nucleases) and prevent the nucleases from degrading DNA. Detergents (SDS) are also included in cell lysing buffers. Disadvantages of these methods may be the some extra steps to remove proteins. Owing to the viscosity of the suspension, it can be difficult. Nucleases can also be inactivated by dehydration. One of the most generally applied ways is using several volumes of absolute ethanol (for example, suspending one drop of blood in 1 ml ethanol in an Eppendorf tube). Ethanol can be used also for preserving whole insects. Samples handled in this way can be stored for years even at room temperature without significant degradation, although alcohol contaminants may cause degradation.

There are many other possibilities for sample storage, like sodium chloride with 10 % DMSO used for DNA fingerprinting and mtDNA analysis from large tissues. Dried blood is also frequently used or one can also keep samples alive, like blood cells in Alseviers solution (for weeks only).

It worth to mention that samples should be stored away from strong light because of the photochemical degradation of the DNA. It is also a good idea to keep it in a cool place. Samples should be subdivided to avoid repeated cycles of freezing and thawing,

which causes loss in quality and quantity. In summary, we have plenty of possibilities. Appropriate preservation system should be chosen carefully and should be tested before the starting of the field season.

DNA extraction

The aim of the procedure is to produce sufficiently pure and adequate quantity and quality of total (nuclear and organelle) DNA for analysing markers. Protocols are straightforward: lyse the cells (e.g. SDS), remove the proteins and other cellular components (e.g. Proteinase K, phenol-chloroform extraction or salting) and recover DNA in the intact form. The concentration of the DNA (even as PCR template) may be important (e.g. Linacero *et al.* 1998). It should also be mentioned, that there are protocols for direct PCR from tissues without extracting the DNA directly. We had success using the "single fly" method for insects (Gloor *et al.* 1993).

For samples that are known to yield high quantities of DNA (e.g. blood and tissues) the standard phenol-chloroform-ethanol precipitation method may be used (Sambrook *et al.* 1989). The nuclear DNA could be easily extracted also from plant or fungal tissues. One of the most common is the CTAB method, which is widely used by botanists and mycologists (e.g. Štorchová *et al.* 2000). The simultaneous extraction of fungal and herbal DNA could make problems for example when natural samples are used with non-specific methods (e.g. RAPD) but these could be excluded with specific primers. Isolation of DNA from cellular organelles (chloroplasts and mitochondria) may need extra steps, to digest nuclear DNA. However, this can be avoided using PCR with specific primers.

Some techniques are developed to be fast, others to be cheap or to give high quality output. There are protocols to avoid phenol extraction, as special safety equipment is needed. The most time-consuming step is the extraction of DNA from hundreds of individuals. However, variety of commercial kits can decrease this time, used widespread in ecology (e.g. produced by Qiagen and Sigma). They are efficient and have reasonable cost, especially for unit time. They can be used for different sources and produces high quality and high yields of DNA. Based on our experience, kits (we used Sigma genomic DNA kit for insects) should be preferred. Phenol-chloroform extraction may yield higher amount, but kits are much faster, more robust and produce about the same quantity of DNA (quantification is not needed). Safety equipment can also be avoided in this way. Chelex with its simple protocol is also proven to be very useful.

Finally, protocols even may depend on the species. Genomic size and complexity may differ, so preliminary experiments should be performed to minimise data loss in the further studies.

Further steps, basic techniques

If we have the DNA in hand, further steps vary between marker technologies. Sometimes (minisatellite fingerprinting, generally for hybridisation based methods or direct sequencing) DNA must be digested by restriction endonucleases. For PCR, this step is not needed. Before going to details of markers, some words about the basic techniques are given. Any marker analysis results finally in a set of DNA fragment. These fragments may differ in size that must be separated and visualised. As already explained, even if fragment sizes are the same, or differences are not detectable by the method selected, DNA sequencing may provide the last solution.

Southern hybridisation

The basis of the method is the restriction enzyme digestion of the total genomic DNA isolated from the individuals. For this one has to use a laborious procedure because only the pure DNA can be digested properly. This digestion generates a huge amount of restriction fragments with different sizes. Approximately 10 μg of digested DNA is loaded on an agarose gel and separated according to their sizes. This extremely high amount of different fragments contains many overlapping sized fragments thus resulting in a continuous smear on the agarose gel. The high complexity of fragments can not tell us anything about genetic variance by itself. To be able to see the differences in the fragment sizes we have to use a radioactive- or fluorescent-labelled probe which is (Southern-) hybridised to this separated DNA. Because the probe DNA hybridises only to its complementary sequence(s) it can detect the fragment size of its complementers in the fragment population. If the probe hybridises to a single locus, it results a single-locus pattern, if it has many complementary sequences in the fragments we will see a multi-locus pattern after visualisation. For visualisation we detect the radioactive or fluorescent radiation of the probe by using X-ray film or digital imaging technology. Reliable and sensitive commercial kits may provide help us in this process (e.g. Digoxygenin DNA labelling and hybridisation kit, Boehringer-Mannheim).

Polymerase Chain Reaction (PCR)

The introduction of PCR technique (Mullis and Faloona 1987, Saiki *et al.* 1988) revolutionalised not only marker methodology but sampling strategy. Using PCR-based methods we need much less amount of isolated DNA (usually few ng is enough) and the detection procedure is also less laborious. The general PCR-procedure is the following: the genomic and/or organelle DNA is called as template DNA for the amplification. There are two short oligonucleotide sequences called primers that are specifically designed to be complementary to the two borders of the sequence on the template DNA which we want to amplify. These are for priming the DNA-strand elongation following the denaturation of the two strands of template DNA and the hybridisation (annealing) of the primer sequences to their complementary sequences on the template DNA. The repetition of the denaturation/ annealing /elongation steps results in exponential increase of the DNA amount bordered by the two primer sequences. It means that after thirty cycles the few fg DNA sequence can be increased to several μg (even 1000 0000 0000x increase!). The length differences between the amplified fragments of different individuals can be visualised on agarose or acrylamide gels after adequate staining or the amplified fragments can be sequenced.

There are many variations of methods applying this simple logic. Some of them are explained below, related to a given marker. See e.g. Birt and Baker (2000) for a recent overview, including variants and optimisation of PCR reaction.

DNA Sequencing

In supraindividal studies often resolution given by differences in fragment length is enough. However, sometimes we need the nucleotide sequence of a certain DNA region even for the analysis or to design adequate markers for further studies. The base of the sequencing used generally nowadays is the polymerase chain reaction with dideoxy nucleotides with which the products of the reaction are stopped at certain bases along of the whole target sequence so with appropriate separation of these products the sequence of the target could be determined. The productivity of the automatic sequencing — particularly with the capillary technique — has improved to such a level which is nearly incomparable with the early methods, however manual sequencing is still used — although not often — for short target regions (e.g. Milot *et al.* 2001). As the technique has developed the cost of the sequencing decreased and — although the equipment is expensive — the sequencing conducted in the frame of co-operations or even as a customer of the sequencing services of plenty of biotechnological companies is relative easy to pay.

Although high fidelity enzymes are available, sequencing of several quasi-independent samples from one region is particularly important when the template is a PCR product. The cloning of the PCR products and sequencing different clones could give a solution for this problem. Although this is more time and equipment consuming method, the samples could be kept as reference material and can be used for further studies and on the other hand using of cloned material could be one of the cheapest ways of sequencing. See e.g. Palumbi (1996) for technical notes related to different questions.

Fragment separation and visualisation

Nearly all the DNA methods with which data can be obtained to answer ecological questions after all based or connected somehow with the size-based separation and detection of the nucleotide fragments. One of the easiest methods — with which however very important and usable data could be obtained (see e.g. RAPD) is the agarose-based electrophoresis of the DNA visualised by ethidium-bromide. After the samples were run on the gel with appropriate agarose concentration using the electric charge of the DNA molecules and staining the gel in ethidium-bromide solution, the fragments can be visualised by ultra-violet light and documented by photo or by special gel documentary systems. Although this is a lightly manageable and relative cheap method and a lab could be easily supplied with the equipment necessary for collecting analysable data, one of the biggest limitations of the technique is its relative poor resolution. Small differences (typically less than 10 nucleotides) — even with using high concentration agarose gel — between fragments are hardly detectable.

The acrylamide-based gel electrophoresis with silver staining has much better resolution – one nucleotide differences of fragments could be detected — and although the method is more time and practice consuming it is commonly used in molecular works in ecology (see below) and the necessary accessories are relative not expensive. Isotope-based techniques are still also used for visualisation and documentation of DNA fragments but the work with them needs special safety equipment and permissions beside the special experience.

Very large fragments (e.g. chromosome sized) are separated by different methods, such as pulsed-

field gel electrophoresis (PFGE). Actually, efficient electrophoresis methods are developed to detect minimal differences as small as one nucleotide (e.g. SSCP).

Markers

As it was mentioned before there are two main types of molecular markers such as protein and DNA ones. However, they include several “subgroups” which can differ essentially from each other. Before going into details of their characteristics and grouping we have to consider some important characteristics of the different marker methodologies.

Genetic background

The basic background of genetic polymorphisms in organisms is the differences occurring in the DNA sequences due to mutations accumulated in their genomes. It is important to know what kind of mutations can occur because the type of a certain mutation significantly determines the selection of the most useful marker methodology for its detection. The main types of mutations are: (1) insertion or deletion of a certain DNA-fragment, (2) gene duplication or multiplication, (3) variable number of repetitive sequences such as mini- or microsatellites, all of them resulting in altered fragment lengths (4) or exchange of multiple or single nucleotides, the latter one called single-nucleotide polymorphisms (SNPs), without alteration of fragment length.

The mutations occurring in coding genes can influence the expression of them causing differences in visible phenotypes. Others are non-visible but result differences at protein level. All the others are so-called "neutral" mutations either in coding genes without affecting them or in non-coding regions such as repetitive or heterochromatic regions (e.g. “satellites”). These alterations are not represented even in the protein pattern therefore their existence can be demonstrated only at DNA level. This allows us to draw several important conclusions to be considered when selecting a right tool for analysis of genetic variability. First, the expected number of variations is the highest when using DNA markers and lowest when following visible phenotypic characters. This is because only very small portion of the total genomic DNA is coding for genes, which have phenotypic effect, therefore most of the genetic variations are not represented in the phenotype. Second, the less influenced therefore the most reliable markers are also the DNA markers because they are not influenced by the so-called "penetration-effect": i.e. some genes are not expressed, or suppressed by other gene's products, therefore they are not represented at protein or morphological level. Third, the use of visible markers is the most misleading because they can be influenced also by environmental factors, which effect is completely eliminated at DNA level.

It must also be mentioned here that because there may be a strong evolutionary pressure on the protein coding genes which doesn't allow them to mutate very frequently because sometimes they are essential for the survival of an organism. Therefore most of the mutations are concentrated mainly on the non-coding regions of the genome. According to this there are so-called "mutation hot-spots" which are hypervariable regions such as short repetitive sequences or microsatellites. It means that if we want to detect high level of polymorphisms we have to search them in the hypervariable regions of the genome.

Two classes of tandem repeat sequences are of great interest in the hypervariable region. These are called VNTRs (Variable Number of Tandem Repeats) including mini- and microsatellites (called as SSRs and sometimes STRs). Nowadays, VNTR tends to be maintained for minisatellites only. The function and evolutionary significance of these sequences is unknown. They are repeated many times in the genome and extremely variable in length. SSRs consist of the so-called core sequences from 30-150 bases, where the length of the repeat units is 1-8 (10) bases (that is 10-50 copies), most often 2-5. Minisatellites has more than 10 bases length unit and core sequence size much bigger. Minisatellites also have lower number of copies in the genome. The underlying genetics of high mutation rate are more or less known: unequal crossing over or strand slippage during replication may cause changes in the number of repeat units (Jeffreys *et al.* 1998) and consequently the size of the fragment we detect.

It must also be mentioned that if we investigate different regions of the genome we can not find the same mutation rate which can be misleading by the evaluation of results. Higher mutation rate of a genomic region results in higher genetic variability, which can be interpreted as higher genetic distance. Nevertheless, the regions where the selection works in the direction of keeping polymorphism at high level (called diversifying selection) can also be used as a powerful tool in population studies (e.g. MHC).

Mitochondrial (and chloroplast) DNA has different characteristics than the nuclear one. Most important ones are that they are haploid and can be considered as haplotypes. Furthermore, they are transmitted intact through generations owing to the (assumed) lack of recombination and inherited

usually maternally. Nevertheless, the lack of recombination is questioned today (e.g. Lunt and Hyman 1997, Hagelberg *et al.* 1999, Wallis 1999). These characteristics provide special importance for mtDNA: "The genome of mitochondria has been the workhorse of this molecular revolution." (Randi 2000).

Classification of marker methods on different basis

The grouping of the different marker approaches can be performed based on different aspects. The most widely applied classifications are based on the methodology (A) and genetic information content (B) but the other ones can serve as important source of information, too (C-E). Of course, they can be classified on their use in ecology, generality (connectibility) in taxonomic scale or variability (sensitivity). These are not highlighted here but explained in the description of methodologies. (Full names of the markers are given at the end of the paper.)

A. Classification on technical basis:

1. Protein markers:
a. isoenzyme
b. total protein (SDS gels)
2. DNA markers:
a. Southern-hybridization based:
RFLPs of single or low copy number loci
mini- (VNTRs) and microsatellites (SSRs)
multilocus fingerprint hybridisation (e.g. multicopy genes, loci or transposons)
b. PCR-based:
(i) random primed methods (no need for sequence information):
RAPDs, DAFs,
AFLPs
(ii) specific primer-methods (need for sequence information):
specific gene amplifications (mtDNA, scnDNA)
VNTRs (rarely)
SSRs

Notes: referring mtDNA as marker means usually its sequence, although mtDNA RFLP is also used. scnDNA actually a pool for single locus nuclear markers (excluding repeated sequences and sometimes introns) but it can be evaluated many different way including solution DNA-DNA hybridisation. See below.

B. Classification on genetic information content

1. markers detecting single or low copy number loci (codominant inheritance):
a. isoenzymes
b. RFLPs
c. SSRs
d. scnDNA
2. markers detecting multiple loci (dominant/recessive inheritance):
a. total protein analysis
b. Southern hybridisation fingerprints using repeated sequences
c. RAPDs, DAFs
d. AFLPs

Note: mitochondrial and chloroplast DNA are haploid, so (single locus) mtDNA sequence and RFLPs are codominant by definition.

C. Classification on effectiveness and reproducibility/reliability

1. isoenzymes: needs low amount of protein, laborious, highly reproducible and reliable
2. total protein: needs low amount of protein, less laborious, middle reproducible and reliable
3. RFLPs: needs high amount of DNA, very laborious, highly reproducible and reliable
4. AFLPs: needs low amount of DNA, laborious optimisation and use, good reproducibility, high reliability
5. SSRs: needs low amount of DNA, very laborious to develop, but middle-laborious use, very high reproducibility and reliability
6. RAPDs, DAFs: need low amount of DNA, less laborious, low reproducibility and reliability
7. scnDNA: needs low amount of DNA, laborious to develop, but middle-laborious use, very high reproducibility and reliability
8. mtDNA: needs low amount of DNA, but middle-laborious use, very high reproducibility and reliability

Note: developing markers is related to the possibility of the transfer to new taxa. It varies. Furthermore, type of assay including separation (classification D below) may depend on the scale of question/variability.

D. Classification on separation and visualisation methods:

1. protein markers: acrylamide gel electrophoresis & protein staining
2. RFLPs: agarose gel electrophoresis & radioactivity/fluorescence
3. AFLPs: acrylamide gel electrophoresis & silver staining, radioactivity/fluorescence
4. SSRs: acrylamide gel electrophoresis & silver staining, radioactivity/fluorescence
5. RAPDs: agarose gel electrophoresis & EtBr staining
6. DAFs: acrylamide gel electrophoresis & silver staining

For scnDNA and mtDNA, see Table E.

E. Specific methods for detecting SNPs when using PCR amplification based methods

1. DGGE: Denaturing Gradient Gel Electrophoresis
2. SSCP: Single Strand Conformation Polymorphisms Both methods use acrylamide gel separation under denaturing conditions thus detecting single nucleotide alterations
3. Sequencing (mtDNA)

There are strong correlations between different classifications as it is explained below. On the other hand, the classification may suggest strict boundaries, which is not true. Methods can be combined on many different ways. For example, RAPD or AFLP fragments can be converted to single locus markers (isolating from the gel, sequencing and designing a primer pair), in which case they can be used as scnDNA. Another example can be the method for searching for microsatellites using AFLP (Hakki and Akkaya 2000).

Description of the different marker techniques

Protein markers

The isoenzyme-variation analysis (Wedel and Weeden 1989) is based on the knowledge that the same enzyme can have different subunit-composition. These subunit differences can be visualised after running of protein samples on acrylamide gels and staining them with a definite substrate resulting in a colour product thus indicating the presence of a certain protein allele. This method has the serious limitation by finding the right substrate and conditions for enzymatic reactions. The main advantage is its high reproducibility and reliability. Despite of the wide-range use of DNA markers there are examples for using it as efficient tool of genetic relatedness (Rouamba *et al.* 2001) and studying population structure (e.g. Meglécz *et al.* 1999).

The other way of using proteins as genetic markers is the analysis of total protein content of a given tissue of an individual. In this case the total protein content is visualised after acrylamide gel electrophoresis of the proteins under denaturing conditions at which protein subunits are separated by Coomassie or silver staining method. Using this technique we can follow the presence or absence of a certain subunit of different anonymous proteins and these differences can be characteristic for the individuals or groups (Vollmann *et al.* 2000).

Application. They are not detailed here. Isosymes are single locus markers and may show sufficient variability. The methodology is well founded and cheaper than the DNA methods. Nevertheless, studying protein polymorphism requires high quality samples and gives limited genealogical information comparing to the relevant DNA methods. Collection and storage of DNA sample is usually much easier. See e.g. Baker (2000b) for more technical details and usage.

DNA markers: non-PCR-based methods

The most widely used methods are the DNA-based genotyping techniques. According to the classification used before we should see first the Southern-hybridisation based methods like RFLP and hybridisation-based fingerprints.

The RFLP method was developed and first used in human genetic mapping (Botstein *et al.* 1980, Lander and Green 1987, Lander and Botstein 1989). We call **RFLP** in narrow sense if we follow the pattern of low number of loci at one hybridisation, and it is called hybridisation fingerprint if we use highly repetitive sequences as probes for hybridisation. The use of fingerprint methodology became highly significant by the exploration of the variable number of tandem repeats (**VNTR**, Horn *et al.* 1989) or minisatellites which are highly repetitive and polymorphic DNA-sequences capable to detect a very high number of different loci at the same time (Jeffreys 1985, 1990). Using this approach we can distinguish up to 10-30 loci parallel in one experiment.

If we isolate DNA from many different individuals and use them for Southern-hybridisation we can expect that there will be differences in the

length of restriction fragments between these individuals. We can detect fragment length differences if there was an insertion/deletion event in the region we can visualise or the recognition site of the enzyme had been changed between the individuals. The limitations of this technique are coming from two facts. First, only the genetic differences generated by the selected restriction enzyme can be detected. This means that if we have a polymorphism e.g. in the *Eco*RI site of the individuals and we use *Bam*HI enzyme we will not see the differences. The second is, if we use a given enzyme but our probe hybridises not to a polymorphic but to a non-polymorphic fragment. In this case the genetic differences remain also hidden. That is, selecting the right enzyme with a probe needs some optimisation.

Application. RFLP is expensive and time-consuming. Even it is true for its PCR analogue, not detailed here (see e.g. Snow and Parker 1998, but see ascnDNA below). Although RFLPs can be useful for studying population structure, they are rarely used today. It is replaced by more efficient PCR based single locus methods, such as mtDNA sequence analysis or microsatellites.

In contrast, the hybridisation based fingerprinting using repetitive sequences, such as multilocus and single locus minisatellites are still widely used. Owing to their variability, they are used especially for individual identification ("fingerprint") and testing parentage (paternity exclusion) but can be used to estimate diversity on population level. Multilocus fingerprint gives one of the most easily accessible (but not the cheapest) way for individual identification in higher vertebrate taxa, as it can be done by one hybridisation using universal probes for wide range of species (Burke and Bruford 1987). Nevertheless, it is very time-consuming and suffers of drawbacks of multilocus techniques (see below).

Single locus minisatellites would be an ideal tool for individual identification, owing to the advantages of one locus techniques and the use of agarose gel. However, development of probes is very time-consuming and expensive (involves library construction) and probes usually can not be used for other taxa. Single locus minisatellites could also be assayed by PCR. However, owing to the large fragment size, it is rarely used today and has some disadvantages comparing to microsatellites.

Microsatellites can be assayed with hybridisation on a multilocus manner and they have the same usage as minisatellites. The more common PCR based assay is detailed below. Some more notes of fingerprinting are given in the application section.

DNA markers: PCR-based methods

As explained before, for most of the PCR-based methods we need less amount of isolated DNA and the detection procedure is also less laborious. PCR-based assay can be used both of single and multilocus manner. For single locus methods, specific primer pair is needed similarly to the specific probes in Southern-hybridisation. The length of these primers is typically 20 bases at least.

Multilocus techniques

One of the most widely used PCR-technique in genetic variance analysis is the **RAPD** analysis (Random Amplified Polimorphic DNA; Williams *et al.* 1990, Welsh and McClelland 1990). In this case the primer is relatively short (10 base oligonucleotide) and only one sequence is used. The sequence is randomly selected, thus there is no need for prior sequence knowledge. These 10mer oligonucleotides will hybridise to their complementary sequences, which are dispersed randomly in the genome. Therefore many different fragments will be amplified at the same time due to the random distribution of the primers on both strands of the template DNA. The location and the sequence of these fragments are unknown but they can represent fragment length differences between different individuals, if any deletion/insertion happened in the regions they amplify. It can also be supposed that the primer-binding site was altered by mutations in some individuals resulting also differences in the amplification pattern. The amplified fragments are loaded on a concentrated agarose gel and underlain to electrophoretic separation, which is followed by an EtBr-stained visualisation on UV-lamp. The results are usually documented by photography. By this method usually 5-10 loci can be detected in one experiment.

Comparing the protocols of specific primed PCR and RAPD, the length of the oligonucleotide primers is significantly different and the annealing temperature of RAPD is much lower. These two characters are correlated: the shorter the sequence is the lower the hybridization (annealing) temperature must be (because there are much less amount of hydrogen bonds). The GC-nucleotide content has also very important effects on amplification. The higher the GC-content is the higher the annealing temperature can be (because there are 3 H-bonds between G and C nucleotides in contrast to A and T ones). Specificity of primer bound also depends on the temperature (because of the number of hydrogen bounds).

A slightly modified version of the random PCR-amplification methods is the **DAFs** (DNA Amplification Fingerprint, Caetano-Anolles *et al.* 1991). This is also based on the random primed PCR methodology but uses 8mer primers instead of 10mers and uses acrylamide gel electrophoresis for fragment separation instead of agarose. For fragment visualisation the silver staining technique is used instead of EtBr. This is because the shorter primer sequence allows lower specificity in annealing and therefore much higher number of possible homologue sequences in the genome. As a result, much more fragments will be amplified by this method, which can not be separated on simple agarose gels only on acrylamide. The silver staining method is a more sensitive way of fragment visualisation as ethidium bromide, so at the end much higher number of amplified fragments can be seen on the gel compared to the RAPD technique. Using this method it is possible to detect more than 20 fragments in a gel. DAFs can be used on the same manner as RAPD, but not detailed here.

The most recent PCR-based fingerprinting technology is the so-called **AFLPs** (Amplification Fragment Length Polymorphism, Vos *et al.* 1995, Mueller and Wolfenbarger 1999). The name resembles the RFLPs which is not accidental. Namely this method combines the advantages of both RFLPs and PCR methods. The theoretical/technical basis of the methodology is the following. The first step is the template DNA isolation which is then followed by a restriction enzyme digestion like in RFLPs. But the fragment length differences will not be visualized by the use of DNA-DNA hybridisation but will be amplified by PCR primers. For this specifically designed primers are used which are complementary to the oligonucleotides ligated to the ends of the digested DNA fragments. It is easy to ligate the adapters to the fragments because the restriction enzyme we used for digestion will result so-called "sticky ends" which can be used for ligation of adapters. So the sequence of the protocol is: restriction digestion of template DNA, ligation of specific adapters to the ends of the fragments containing the complementary oligonucleotide sequences of the primers, then PCR amplification using primers complementary to the ligated adapters. We can imagine if all the restriction fragments will be amplified, the gel separation of them will result in a very complex, useless pattern. Therefore specific "selective nucleotides" are planned to the ends of the primers allowing the amplification only a smaller portion of the digested fragments. The fragments will then be separated on polyacrylamide gels and visualised either by silver staining or radioactive/fluorescent reactions. As a result we will get a serious number of amplified fragments with different lengths and can check them for genetic polymorphisms. The genetic background of polymorphisms detected by this method is the change of nucleotides at restriction enzyme sites (like RFLPs), or the insertion/deletion events resulting in altered fragment sizes between individuals. By this method the number of distinguishable fragments is over 50 in some cases.

Application (RAPD, AFLP). At this time, both RAPD and AFLP are applied for plants more intensively (see e.g. Ritland and Ritland 2000). AFLP is relatively new method, but at least in plants, it replaces RAPD almost everywhere. Although, highly variable microsatellites are available for plants, even in the chloroplast genome. Contrary, in animals, where we have a choice, single locus microsatellites are preferred. If no primers are available, RAPD is the first to try. The difference in preferred markers between animals and plants can be explained rather by tradition and earlier experience than efficiency. It is probably result of the great use of RAPDs in plant genome mapping, as controlled crosses with large number of progenies are much easier. Of course, genomic composition also differs between animals and plants, giving rise to a preference of a given type of marker (see discussion).

Nonetheless, the explanation of wide use of RAPD is clear, irrespectively its theoretical drawbacks: it is cheap, simple and can be used without any preliminary sequence knowledge after some optimisation for any species. Furthermore, level of polymorphism may be tuned using different primers. Nevertheless, the price of simplicity is paid at the evaluation. It is a multilocus marker with dominant inheritance (such as AFLP) even if theoretical results make possible to use models of population genetics in a restrictive manner (Lynch and Milligan 1994). Furthermore, high quality DNA is needed for reproducible result (but see notes below).

RAPD and AFLP can be used in a wide range of studies owing to the high range of variability it can show. Sometimes it used for studying relatedness (like the fingerprinting) or mating structure, but much common use is the study of population structure. Linkage mapping and quantitative trait locus (QTL) mapping must also be mentioned. RAPD is also used in taxonomy because plenty of characters can be obtained easily to compare species. See also the application section for further details.

Single locus techniques (Sequenced-tagged-sites, STS)

By definition, STS markers are those that reveal codominant polymorphisms in specifically targeted sequences, including scnDNA, introns, microsatellites and mtDNA sequence. The main difference between the previous methods and these ones is based on the genetic information obtained. Alleles of loci can be identified (codominance) providing a huge advantage in many situations. It means that the full power of theoretical tools piled up by population genetics in 80 years can be used together with the advantages of specific-priming PCR technique provided in sampling.

Specific, conserved primers have also a disadvantage: the target sequences must be obtained usually from genomic library, which needs a lot of work and high laboratory requirements. Furthermore, flanking regions used for priming are usually species specific. In same cases, where conservative sequences — kept unchanged in evolutionary time scale — can be found, universal primers can be constructed that works for a large number of taxa (see e.g. bird sexing, below). But generally, moving even to a related species needs new primers. Fortunately, owing to some recent studies and technical refinements, these disadvantages seem to decrease at least for microsatellites (see below).

scnDNA (single copy nuclear DNA), as suggested by its name, means that alleles of an unique nuclear locus is amplified and visualised. It is the PCR analogue of RFLP. It has two types. When target locus in the nuclear genome is actually unknown but polymorphic, we may have a useful marker termed as anonym scnDNA (ascnDNA). Anonymous scnDNA has the advantages of technical convenience and may yield variable region. Its obvious advantage is the random priming in the genome (see RAPD) but on a (multiple) single locus manner. The disadvantage of ascnDNA is the lack of generality, it can be used only in closely related taxa. However, when the flanking region of a variable target locus is evolutionarily conserved, we can design specific PCR primer pairs for that. This marker is termed as specific scnDNA. It may be used in many taxa, depending on how much the flanking region (i.e. the primers we designed) is conserved. Commonly, scnDNA is sequenced but all methods of detecting SNPs can be used. These are recent methods, their full advantages are not known (Palumbi and Baker 1996, Karl 1996).

Nuclear **introns** are predicted to be of greater importance in the future (Friesen 2000). Introns are widely used target regions of sequencing based molecular ecological studies. It can also be evaluated using electrophoretic methods of detecting SNPs (SSCP, TGGE), so it can be screened routinely. Introns are transcribed but not translated, so the mutations at these regions are presumed as neutral for selection. It means also that it has higher mutation rate than other scnDNA markers. Furthermore, it is more representative for the genome than mtDNA which of special inheritance. The functional regions (exons) framing them make possible to design PCR primers for introns. This step is based usually on date from sequence databases. Sequences submitted to databases may contain introns that can be recognized easily. This fact itself suggests future success of intron-based methods. Conservative exon regions also provide wider range of species where primers can be used as explained before.

Application. The ribosomal genes are widely used in molecular phylogenetics and taxonomy, like their intron regions or the internal transcribed spacer (ITS) between genes coding for the small and large subunit of the ribosome (e.g. Kovács *et al.* 2001). Another use is the gender identification (Griffith 2000). As an example, two sets of universal primers are available for sexing non-ratitae birds (Griffith *et al.* 1998, Fridolfsson and Ellegren 1999). It is based on the highly conservative, sex-linked gene of the chromo-domain-helicase DNA binding (CHD) protein. Intron length of the CHD gene copies in the avian W and Z chromosomes differ that can be visualised on agarose gel with EtBr staining. Females — the heterogametic sex in birds — produce two bands (W and Z) while males do one (Z). This method is used routinely today starting from a drop of blood, even for sexing offsprings or individuals of species without sexual dimorphism (but see Dawson et al (2001) for a critique). Nevertheless, recently introns are used mainly in phylogenetics and evolutionary biology. There is not enough data to evalute its advantages in the mainstream of ecology.

As explained before, **microsatellites** (SSRs, sometimes called STRs, short tandem repeats) belongs to the class of tandem repeat sequences, with usually less than five bases length of repeat units. Owing to the supposed mechanism of mutation already mentioned, alleles differ in size of the repeat unit: larger unit size means easier differentiation. For microsatellites, smaller differences must be recognised comparing to minisatellites. That is why acrylamide gel and silver staining are needed, but often fragments are simply sequenced owing to the automated methods. Actually, expected polymorphism depends on the characteristics of the loci (like length), it is not detailed here.

Microsatellite loci are assayed on the standard way: target region is amplified by specific primer-pair. Different primers are needed for different loci, which means that PCR condition may also differ. Resolution is increased by analysing many loci consecutively from the same individual. Using specific primers has the disadvantage of the limited applicability to a new species. However, for microsatellites this problem tends to decrease.

Nowadays, according to the high interest for microsatellites, screening for useful markers in the genome is less and less laborious. There are techniques available to find microsatellite-enriched sequences efficiently. AFLP can also be used without cloning and screening, as mentioned before. But the primary source of information is the literature and the sequence databases, such as NCBI-GENBANK (http://www.ncbi.nlm.nhi.gov) and EMBL (http://www.embl-heidelberg.de). Further useful sources are the species specific databases or microsatellites can be queried in many different places (Scribner and Pearce 2000).

Specificity is much better as it was supposed earlier, at least for birds. For example, microsatellite primers developed for a species of swallow detected polymorphic microsatellite markers for 32 of 39 other species within the same order and 6 of 19 bird species within different order (Primmer *et al.* 1996). We are also using primers developed for house sparrow (*Passer domesticus*) in tree sparrow studies (*Passer montanus*), some of them seems to be polymorphic (Pénzes *et al.*, unpublished data). It means that if no primers are available for a given species, we have a chance that primers developed for a related species may work.

The efficiency of SSRs resolution is also demonstrated in plant samples e.g. in soybeans in searching for genetic polymorphism and using them for genetic mapping studies (Csanádi *et al.* 2001). Finally, technical advances can be shown best by a human example. First, owing to the human genome project, screening for any sequence means actually database search (which is much faster and cheaper, of course). So target sequences can be selected much easily, keeping in mind of their independence (e.g. lack of linkage: different chromosomes) and different size range (can be scored on one gel). Commercial kits are available for microsatellites (used for paternity test in courts, they are very expensive, e.g. Genetic Analyzer, Applied Biosystems). Nine independent microsatellite loci (plus one locus for sex identification) can be analysed in the same time — 10 pairs of different primers are used in one PCR reaction. Furthermore, fragments are separated by acrylamide based capillary electrophoresis, where samples of individuals are put in different capillary. Alleles of a given locus are recognised by its size range and/or labelling. As an example, we used these microsatellite loci for a population structure study of humans (Beer *et al*. unpublished data).

There are numerous recent reviews of microsatellites according to its importance. See e.g. Scribner and Pearce (2000) for a general overview with many useful information and we suggest Ritland and Ritland (2000) for applying them in plants. Mitochondria and chloroplasts contain also microsatellites, not detailed here.

Application. Microsatellites are extremely useful markers for studies on whole scale of ecology interested in, from individual level to populations. First, it is assayed by PCR, using specific primer with the full set of advantages on sampling. Microsatellites may be highly variable, in a single locus as many as 50 alleles can be examined. More loci are used generally together composing a multilocus pattern where both alleles of every locus can be identified (codominant inheritance). These provide sufficient statistical power for individual identification and parentage determination or infer relatedness. By extending it to more generation, pedigree or population level patterns can be constructed. We can have insight into the mating structure or estimate effective population size. Numerous studies apply microsatellites to determine the magnitude of differences between populations or understanding population subdivision and gene flow. Taxonomy and evolutionary biology also uses microsatellites (speciation, hybridisation). Results of different population studies can also be combined using the models of molecular evolution, opening the way for metaanalyses. Finally, besides its practical importance, there are many opened questions about microsatellites not detailed here: it is “an active arena for theoretical and empirical work” (Scribner and Pearce 2000).

Mitochondrial DNA has importance in many fields (Randi 2000). As it was mentioned earlier, it can be considered as a haplotype and inherited mainly maternally. It evolves faster than nuclear genes (but not the hypervariable sequences), at least on average — owing to the less efficient DNA repair mechanism, resulting in higher level of variation. Furthermore, its special transmission provides unique tool for reconstructing genealogy in a wide timespan, from populations to phylogenetics. The conservative protein coding regions can be used to trace phylogenetical relationships back to million years. Besides this, the main noncoding sequence, the control region (called D-loop in vertebrates and AT-rich region in invertebrates) is much more variable (it

regulates the replication and transcription of the whole mitochondrial genome) targeted in many studies. mtDNA is assayed by RFLP or most frequently PCR amplification followed by gelelectrophoresis (in any form) or DNA sequencing. For vertebrates, there are universal primers for PCR-based sequencing the genome (Sorensen, 1999). Some examples of application are the studying population variability and gene flow, hybridisation or phylogenetics and conservation biology. We emphasise its use in intraspecific phylogeography or historical biogeography, which has incredible importance in recent understanding in evolutionary (or simply general) biology (Avise 1994, Burke 1998).

Advantages and disadvantages of different marker techniques: which marker to use

Numerous reviews have been born to compare the efficiency and usage of different marker technologies (e.g. Lu *et al.* 1996, Powell *et al.* 1996, Jones *et al.* 1997, Milbourne *et al.* 1997, Russell *et al.* 1997, Sunnucks 2000 and others). To compare these techniques we can take into account the aspects of classifications explained above. On this basis we can compare them on the basis of technology, genetic and efficiency aspects.

The first decision should be about the **sensitivity** of the marker needed. Lots of data accumulated for today to help us in this decision, but pilot studies may make the picture cleaner. Less sensitive marker does not give polymorphism on the level of our interest (like mtDNA or ITS for studying paternity). Or, using sensitive marker, like microsatellite, for a question of large scale, e.g. taxonomy, would give a random pattern. Some methods can be scaled to give the desired level of polymorphism (e.g. selecting RAPD primer, different mtDNA regions). Among the markers with suitable resolution, choices can be made on more practical bases.

On **genetic basis** we can take into account two main aspects: the **information content** and the **robustness** of the technology. The information content can be described by the dominant/recessive or codominant heredity of the alleles on the locus. All the marker types inherit codominantly if all the alleles of the individuals can be distinguished. In this case in a diploid individual if one allele gives longer fragment and the other a shorter one, the heterozygote will have double fragments. This is true for RFLPs and SSRs. The other methods distinguish only the "presence" or "absence" of a certain fragment (such as RAPDs, DAFs and AFLPs). In this case if the fragment is not present we can say it is a homozygous individual for the allele but if it is present, we cannot decide if it is homozygous for the presence of both fragments or heterozygous because only one allele is present. This basic difference can cause information loss in the case of dominant markers compared to the codominants, because the genotypization is not so exact as by them. This suggests using rather codominant than dominant marker types.

On the other hand RAPDs, DAFs and AFLPs produce much higher number of visible fragments at the same time, so they are the most robust ones even if they are dominantly inherited. The higher amount of fragments can compensate for the loss of information content by the dominant inheritance.

With other words the information content of a given marker type can be numerically characterised by the expected heterogeneity (H_{av}, number of polymorphisms detected) and by the effective multiplex ratio (E, the number of effective bands) (Powell *et al.*, 1996). The distinctive capacity (characterised by marker indices, MI) of the marker system is the product of E and H_{av}. This means that the efficiency of a given marker system can be enhanced by increasing the value of E and/or H_{av}. The higher the Marker Index is the better the marker technique for a given species. According to the recent data in the literature the AFLPs proved to be the most and RAPDs the less efficient technology regarding both heterogeneity and multiplex ratio (Powell *et al.* 1996, Lu *et al.* 1996, Bohn *et al.* 1999).

The **reproducibility and reliability** of the different techniques is also widely investigated. In a very detailed analysis, the reproducibility of the different marker technologies was tested in several different European labs (Jones *et al.* 1997). It was found, that RAPDs are poorly reproducible in contrast to other marker techniques and besides AFLPs, SSRs proved to be the best in this comparison. This is very important in the evaluation and interpretation of the results obtained. Taking into account that AFLPs are the most robust technique between all genotyping methods one can conclude to use this methodology for checking genetic variance. Why most of the people choose RAPDs most preferably? To answer this question we have to check a different aspect of the molecular genetic methods. This is the cost- and labor-efficiency of a marker methodology.

The **efficiency** of the marker methodologies is depending on several different parameters. The best way to measure it when we count the polymorphic bands/ invested time and energy for each case. At the first glance RAPDs seems to be the less laborious,

moderately robust and not very expensive method. Comparing the cost of RAPDs and RFLPs with their efficiency, Ragot and Hoisington (1993) found that RAPDs are most cost-efficient when using small sample sizes in contrast to RFLPs. But basically it cannot be concluded which one is generally better in this aspect. When checking the efficiency we have to take in account several aspects of the different methods:

1. Price of the equipment
2. Price of the chemicals used
3. The number of working hours
4. The amount of polymorphic bands expected

1. Price of equipment:

RFLPs: hybridisation oven, tubes and sometimes documentation tools

RAPDs: usually require a PCR-machine, an electrophoresis system

and some documentation tools

SSRs: PCR machine, electrophoresis system and documentation tools

AFLPs: PCR machine, sequencing acrylamide gel system and documentation tools

From these the most expensive investment is related to the AFLPs. This is one of the reason why it is not so widely used as RAPDs

2. Chemicals:

RFLPs: hybridisation membranes, polymerase for probe labelling, radioactive or non-radioactive chemicals for detection

RAPDs: primers, nucleotides, polymerase, agarose and ethidium-bromide

SSRs: primers, nucleotides, polymerase, acrylamide and silver staining chemicals or radioactive / non-radioactive chemicals for detection

AFLPs: chemicals in kits: restriction enzymes, primers, adapters, polymerase,

nucleotides, radioactive or non-radioactive chemicals for detection

It can be clearly seen from this list, that most expensive investment is related to AFLPs again. The less expensive is the RAPDs.

3. Number of working hours

In this aspect to standardise one method can take for different times because sometimes it is very hard job to optimise even the simplest RAPD reaction. In general the more steps the method has the more time is necessary for optimisation and working. For a routine analysis one RAPD or SSR reaction can be performed in 4-5 hours, while RFLPs take for several days similarly to AFLPs. This would suggest working rather with the first two methods.

4. Number of polymorphic loci expected in one experiment

RFLPs: up to 10 loci

RAPDs: up to 10 loci

SSRs: up to 5 loci

AFLPs: up to 50 loci

In this case it would be optimal to work with AFLPs.

We can draw some conclusions from the above aspects. In general if we do not want to analyse a huge amount of samples or we do not have enough money for molecular analysis we can start with RAPDs. It is the less laborious and money-consuming method when using low number of samples, but it is the less reliable and reproducible, too. If we decide to analyse a huge amount of individuals which are relatively close to each other but we do not have much money for investment it is reliable to start with SSRs, or if they are not developed for the species — with RFLP fingerprints using random probes. We can also try to use DAFs after a certain optimisation procedure. If we have enough money for investment and want to generate a huge amount of polymorphic samples we should invest into AFLPs technology because of its highest productivity, reliability and reproducibility.

Besides these aspects there are also some other points to be considered. First ones are the source of DNA and its quality. For RFLPs, the most critical point is the isolation and purification of proper amount of source DNA. The isolation can be automatised by using modern extracting methods but it will increase our costs. Because of the restriction digestion step, the low amounts of source DNA must have high quality when using the AFLP technique: incomplete digestion can generate false positive results for genetic analysis. In addition, all PCR-based methods have the risk to be contaminated with foreign DNA, which is completely excluded in RFLPs analysis. In summary, high quality, non-degraded DNA is generally needed for multilocus methods. Again, this is less important for using short, single (multiple) locus markers assayable by PCR.

Second one is the automation possibility of a certain technique. The automation of RFLPs is possible and solved but only at industrial level: it is so expensive. The use of radioactive and non-radioactive stains for SSRs and AFLPs makes their detection easier but a bit more expensive. This pays out in longer times if we can detect significantly higher number of polymorphisms as by other methods.

It is worth to summarise the trade-off between multilocus (with dominance) and single locus (with codominance) methods. Most of the time, single

locus markers assayed by PCR are preferred. Besides the advantages in sampling, it gives genealogy and comparable results (called "connectivity", Sunnucks 2000). Thus, data from many different studies are comparable directly in meta-analysis. For studies including many species the universality of primers may be also important. In this, single locus methods vary considerably. Multilocus methods are usually more convenient but limited connectivity. Sometimes multilocus methods considered being more economical but it can be questioned. Inevitably, recent advances based on genotypic arrays are provided by single locus methods of codominant markers (Sunnucks 2000).

Rate of evolution in plant and animal DNA is often very different claiming for different markers for the same kind of question. Chloroplast in plants is uniparentally inherited like mitochondria and its DNA (cpDNA) evolves slow rate. It is useful for deeper studies in phylogenetics, like ribosomal RNA genes in nuclear DNA. New advances of chloroplast microsatellites may provide useful tools for population level studies in plants (Provan *et al.* 2001). Differences between nuclear and organelle DNA must also be kept in mind. For example, it is a bad practice to use mtDNA because of the availability of universal primers. mtDNA is inherited mainly maternally and this must be considered in conclusion. mtDNA has also a lower effective population size resulting in more sensitivity for keeping past events.

In summary we can suggest when selecting an ideal method for the analysis the first choice must be made on the sensitivity but we have to take into account the efficiency, reliability and productivity of a certain method and this must be related to the aim of our genetic work.

Applications

The most important general use of different markers is mentioned in the previous section. Now, a different classification is given: the topic of interest. The aim of this section is contrasting markers in some highlighted topics of ecology without rigorous overview. Molecular markers in ecology are used both on individual and population level. One of the main interests is the estimation of demographic parameters that would be difficult to obtain otherwise: reproductive success of individuals, dispersion patterns in space and time or population growth and fluctuation of effective population size. Furthermore, these may include the knowledge of gender, genetic relatedness of individuals or individuals must be assigned to a given group (population).

Five main areas can be outlined: (1) individual identification, (2) identification of sex, (3) testing parentage including pedigree or mating structure reconstruction, (4) population structure (dispersion) and (5) population history. These categories overlap both in theory and practice. Individual identification and testing parentage needs highly variable markers, discussed as DNA fingerprinting. Pedigree construction means subdividing populations into families, at least in time, giving rise to population structure. It also provides data for determining mating structure. Population history reconstruction is mentioned briefly here, although it is one of the most important fields as it provides completely new information for us.

Individual identification and testing parentage: DNA fingerprinting and profiling

This field needs the most sensitive markers. DNA fingerprinting was developed to detect individual-specific patterns for humans (Jeffreys *et al.* 1985). In its original form, it was a multilocus method, minisatellites were detected by Southern blot hybridisation. Using a 33 bp repeated sequence from a human intron, Jeffreys and co-workers isolated two probes, named 33.6 and 33.15, from a human genomic library. These are the most widely used probes today, not only in humans. They have extreme discriminatory power even in birds (Burke and Bruford 1987). Fragments are generated by restriction enzymes with 4 bp recognition site resulting in a set of bands with extreme allelic length variation. Important feature of the bands that they are generally inherited in Mendelian manner (half of the bands are inherited from each parent). In this way, close familiar relationships can be analysed, especially parentage: decision is based on the band sharing, calculated between offspring and supposed parents.

Multilocus fingerprinting detects many loci simultaneously using universal probes, usually by Southern blot hybridisation at low stringency (i.e. less specificity). However, allelism between bands is not known without complex segregation analysis, although sometimes codominant inheritance is suggested. It is a multilocus technique, without the knowledge of the number of loci. High quality and high molecular weight DNA is essential for successful analysis in standard concentration (it should be assayed). Even different gels can be compared carefully, but for this internal size markers are needed. It means that all potential parents must be analysed simultaneously. In summary, universal

probes have advantage, but technical requirements are high.

Single locus fingerprinting is the detection of alleles at a given minisatellite locus using a single locus probe, usually by Southern blot hybridisation at high stringency (high specificity). Offspring can be examined for the presence of non-parental alleles, as evidence for multiple paternity or maternity. Offspring exclusion can be done even if all supposed parents are not known. Alleles can be organised into a database and identification of individuals can be done routinely (e.g. Wetton *et al.* 1995). Analysis of different seasons and years can be performed more readily using single locus method but not with multilocus one. Its main drawback is the lack of general probes. Probes are often synthetic oligonucleotides as it has been realised that simple tandem repeats (microsatellites) detect fingerprint-like patterns. Good commercial kits are available for hybridisation for non-radioactive labelling (e.g. Boehringer's digoxygenin based kit). DNA profiling term is coined for single locus fingerprinting because single locus pattern is not individual specific: it has less discriminatory power. This can be solved when different loci are hybridised consecutively. In both type of fingerprinting, fragments are separated on agarose gels. Usually maxi gel (20x25 cm) is needed. In summary, as a single locus technique, it has many advantages, however, probes are not available generally.

Minisatellites are rarely used in PCR. Even if used, usually not as many locus are available as in microsatellite and primers seems to be more specific resulting in more restricted applicability for different taxa (Sunnucks 2000). Use of microsatellites for testing parentage is equivalent to minisatellites in theory: offspring must carry parental alleles in each locus. Non-parental alleles suggest e.g. extra pair paternity or maternity. For microsatellites, acrylamide gel separation is necessary.

To mention one of the numerous examples of their usage (e.g. Carter 2000b, Scribner and Pearce 2000), studies of extra pair paternity of birds is analysed in this way (e.g. Griffith *et al.* 2002). Using single locus method for analysing pedigree provides not only information about the reproductive success of individuals but the mating structure in the population or effective population size. The classical example is given by Craighead and co-workers (1995) studying 30 family groups of grizzly bears.

Population structure

Population structure, that is difference between populations or subdivision can be studied using many different markers: proteins, RFLPs, RAPDs, mtDNA sequences or microsatellites (minisatellites are rarely). It is related to relatedness studies, at least in methodology. But instead of single individuals (as in relatedness) our main interest is on the group of individuals, populations. Traditionally, Wright's F-statistics are calculated from allele frequencies (or its modern analogues for different markers) and/or variance components are estimated using a tuned analysis of variance (AMOVA, Analysis of Molecular Variance, Excoffier *et al.* 1992). Again, the advantages of single locus methods are clear as they provide allele frequencies. RAPD and AFLP may also be used although it has less power (see Lynch and Milligan (1994) for theoretical background). Methods used to analyse data are very similar to statistics familiar to ecologists (analysis of variance, spatial statistical models and multivariate methods).

Population structure study starts with partitioning variability into different levels, like within and between population components using e.g. phenotypic similarity indices of ecology (e.g. Nei and Li, 1979). This can be used in cluster analysis or PCA. The following approach is more reliable. Bands may provide allele data directly (single locus) or scored for presence (multilocus pattern) that can be analysed further using AMOVA. Correlation between genetic (e.g. Nei's) and geographical distances can also be tested using Mantel test. This is a common list of steps in studying genetic structure in fragmented landscape or gene flow. Of course, different markers may be useful for different scales, as explained before.

Population history, recent advances

Most of the codominant markers (mtDNA sequence, mtDNA RFLP, microsatellites and scnDNA) are extremely useful as they can provide information about gene genealogies. This property adds the time dimension to allele frequency distributions (Sunnucks 2000). Nowadays, clear tests can be carried out about history and spatial patterns. Some examples are: relative timing of past events, like range expansion and differences in gene flow, bottleneck events can be inferred. Genealogies may provide information from population processes to phylogeographic events and have of great importance today. Theoretical advances must also be mentioned, like coalescence approach (Kingman 1982, see e.g. Nordborg 2001, Posada and Crandall 2001, Stephens 2001) and nested clade analysis (Templeton 1998) related to history. Although comparative phylogeography has incredible contribution to ecology and

conservation, we can not detail this field here (see e.g. Burke *et al.* 1998).

Discussion

Molecular tools are inevitable incorporated into the methodology of ecology. To access molecular markers is more complex than the evaluation of the morphological ones. Using DNA techniques need special equipment. Field methods might be changed. However, we believe, this investment will be recovered soon. Students of biology at most of the universities learn these DNA techniques and may have some practice using them. Therefore, the taste of uniqueness will disappear soon. Some of the DNA work can be done even in ecological laboratories (e.g. DNA isolation using kits and PCR with specific primers as in molecular sexing of birds). The most laborious steps (e.g. sequencing) are provided by many companies for reasonable price. Collaboration between ecological and molecular laboratories is also a potential solution, this is how this paper has been born: both sides has same interests. See also the notes of Snow and Parker (1998). It must also be mentioned that various local, national or international laws may regulate the collection and use of biological samples.

New approaches, including markers and statistical methods are also useful in studying nonequilibrium situations. These are in the heart of conservation biology or useful in studies of invasions, population foundations (Davies *et al.* 1999, Waser and Strobeck 1998). The importance of molecular methods in conservation biology increases continuously (e.g. Smith and Wayne 1996, Karp *et al.* 1998), not only on the level of the researches but also on the level of arrangements and acts. It is not accidental that for example a non-profit high-tech was recently established within (till 2003) the Max Planck Institute (Laboratory for Conservation Genetics, LCG, Leipzig, www.raredna.com) to serve the technical and experimental possibilities for applied conservation.

The subject of the unit of the conservation provides an example for the meeting of the results of the molecular methods and conservation biology. The term of the adaptive evolutionary conservation was born from the operativisation of the definitions of the evolutionarily significant unit (ESU) (Fraser and Bernatchez 2001). This term has important legal part in e.g. the USA Endangered Species Act (ESA) (Waples 1991, 1995), or in the Australian Endangered Species Protection Act (Moritz 1994).

The connection and continuity between population genetics and systematics is self-evident today shown by many textbooks of evolution. The same molecular markers with sufficient sensitivity can be used on both fields, as mentioned in methodologies. The parallelism is clear: individuals of different species can be identified — e.g. like the different haplotypes; the phylogenetical relations can be determined — e.g. like the structure of the populations; lineage can be inferred – e.g. like the population histories.

The speed of the development of the powerful statistical methods can be applied to DNA data is comparable to that of the markers. Clearly, technical and analytical methods have facilitated each other. We could not detail the field of evaluation here, but some notes must be given. See Luikart and England (1999) for an overview and software, including relatedness/parentage and dispersal. Assignment (and related) test already mentioned (Paetkeau *et al.* 1998, Davies *et al.* 1999, Luikart and England 1999) as powerful tool studying migration. Exact tests, computer based algorithms are used frequently claiming for update our views of statistics (Rousset and Raymond 1997). On the basis of underlying theory of coalescence, maximum likelihood-based estimators are constructed to estimate population parameters using Markov chain Monte Carlo sampling. Microarray data confronted presently have raised also important issues for statistical testing (Nadon and Shoemaker 2002). The list is endless and grows further including statistical issues of database search and Bayesian approach.

Software for carrying out analyses is essential. We have a choice from plenty of possibilities. Citations and web sites are given usually in the reviews of the different methods. We mention GENEPOP, ARLEQUIN and GDA as they are used frequently. Usually, special questions need special software, and they can be often downloaded from the author's web site. We mention R (Ihaka and Getleman 1996, http://cran.r-project.org) a general software package with efficient data handling, excellent statistical and graphical capabilities, as it is free and develops very fast. Many packages are available for R, including tools for manipulate DNA data (J. Lindsey's DNA package). There are many free software available also for Linux operating system (which is very stable, fast and can be obtained also freely), including NCBI tools, Phylip, TreeView, Arlequin, R and many more (see biology software in http://www.debian.org).

The application of DNA markers should be planned carefully. Common mistake is to use inappropriate markers for a given problem (see e.g. Sunnucks 2000, Baker 2000a). Decision is often made on the availability of universal primers for

amplification irrespectively of the scale of question. This can lead to false conclusion. Sequence databases (like GenBank) are good sources of primers for the studied (or closely related) species even for species specific primers.

Finally, some cautions should be mentioned. Financial needs and technical details often override the importance of conclusions. These are methods to study problems in many fields of biology. It may be expensive relative to the classical field methods but provides high quality information. The money-information quality trade-off can be tracked in the DNA methods, too (e.g. RAPD compared to microsatellites). But the same sampling considerations must be followed (random, sufficiently large samples, etc), it is imposed by the methods of evaluation. Nevertheless, differences between populations detected by microsatellites can be resulted by plenty of processes (e.g. drift, selection, migration). Common mistake to assign differences to one of them, to migration for example (see Bossart and Prowell 1998). These are indirect methods. To decide between the alternative hypotheses, more specialised tests or direct methods are needed. Migration can be detected by field studies. Methods are good but must be used carefully.

Future prospect of applying DNA markers seems to be clear. Probably, new classes of genetic markers will be developed owing to the large scale genomic studies. The fast speed of development both in technology and related statistical methods can be illustrated by microsatellites (Jarne and Lagoda 1996, Luikart and England 1999) or the phylogeography (Burke *et al.* 1998). Interspersed nuclear elements (SINE) may provide a recent example: they are efficient markers for phylogenetic studies, as they can be handled as derived characters (e.g. Takahasi *et al.* 1998). Development of markers and screening also supposed to be simplified. Probably the best example is provided by the recent advances of microchip based technology (which is very expensive today, it is used in medicine and agriculture) for screening SNPs.

Abbreviations

AFLP: Amplification Fragment Length Polymorphism
CHD: chromo-domain-helicase DNA binding protein (gene)
cpDNA: chloroplast DNA
CTAB: cationic hexadecyltrimethyl ammonium bromide
DAF: DNA Amplification Fingerprint
DGGE: Denaturing Gradient Gel Electrophoresis
DMSO: dimethylsulphoxide
EDTA: ethylenediamine-tetraacetate
EtBr: Etidium bromide
ITS: Internal Transcribed Spacer
LINE: Long INterspersed Repeat
mtDNA: mitochondrial DNA
MHC: Major Histocompatibility Complex
PCR: Polymerase Chain Reaction
PFGE: Pulsed-Field Gel Electrophoresis
QTL: Quantitative Trait Locus
RAPD: Random(ly) Amplified Polymorphic DNA
RFLP: Restriction Fragment Length Polymorphism
scnDNA: Single Copy Nuclear DNA, it can be specific or anonymous (ascnDNA)
SDS: Sodium dodecyl sulphate
SINE: Short INterspersed Repeat
SNP: single-nucleotide polymorphisms
SSCP: Single Strand Conformation Polymorphism
SSR: Simple Sequence Repeat (microsatellite)
STR: Short Tandem Repeat, synonym of SSR (microsatellite)
STS: Sequenced-tagged-sites (including SSR, scnDNA)
TGGE: Temperature Gradient Gel Electrophoresis
VNTR: Variable Number of Tandem Repeat (minisatellite)

Acknowledgement

Our research was supported by Hungarian Scientific Research Fund (OTKA F 30577, Zs. Pénzes), Biotechnology 2001 (Biotechologia 2000, BIO-00126, Zs. Pénzes and Gy. Csanádi), research grant of Hungarian Ministry of Education (FKFP 0375/1999) and “Békésy György “ grant of Hungarian Ministry of Education (Zs. Pénzes and Gy. Csanádi).

References

Avise, J. C. (1994): Molecular markers, Natural history and Evolution – Chapman & Hall
Baker, A. J. ed. (2000a): Molecular methods in ecology – Blackwell Sci. , Oxford
Baker, A. J. (2000b): Protein electrophoresis. In: Baker, A. J. (ed): Molecular methods in ecology – Blackwell Sci., Oxford, 65-88 pp.
Birt, T. P. and Baker, A. J. (2000): Polymerase chain reaction. In: Baker, A. J. (ed): Molecular methods in ecology – Blackwell Sci., Oxford, 50-64 pp.

Bledshoe, A. H. and Sheldon, F. H. (2000): Solution DNA-DNA hybridization. In: Baker, A. J. (ed): Molecular methods in ecology – Blackwell Sci., Oxford, 89-112 pp.

Bohn, M., Utz, F., and Melchinger, A. E. (1999): Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs and SSrs and their use for predicting progeny variance. — Crop Sci. 39,228-237.

Bossart, J. L. and Prowell, D. P. (1998): Genetic estimates of population structure and gene flow: limitations, lessons and new directions. – Trends Ecol. Evol. 13, 202-206.

Botstein, D., White, R. L., Skolnick, M. H. and Davis, R. W. (1980): Construction of a genetic map in man using restriction fragment length polymorphisms. — Am. J. Hum. Genet. 32, 314-331.

Burke, T. and Bruford, N. W. (1987): DNA fingerprinting in birds. – Nature 327, 149-152.

Burke, T., Davies, N. B., Bruford, N. W. and Hatchwell, B. J. (1989): Parental care and mating behaviour in polyandrous dunnocks *Prunella modularis* related to paternity by DNA fingerprinting. – Nature 338, 249-251.

Burke, T. *et al.* eds. (1998): Special issue – Phylogeography. – Mol. Ecol. 7, 367-545.

Caetano-Anolles, G., Bassam, B. J. and Gresshoff, P. M. (1991): DNA amplification fingerprinting: A strategy for genome analysis. — Plant Mol. Biol. Rep. 9, 294-307.

Carter, R. E. (2000a): General molecular biology. In: Baker, A. J. (ed): Molecular methods in ecology – Blackwell Sci., Oxford, 7-49 pp.

Carter, R. E. (2000b): DNA fingerprinting using minisatellite probes. In: Baker, A. J. (ed): Molecular methods in ecology – Blackwell Sci., Oxford, 113-135 pp.

Craighead, L., Paetkau, D, Reynolds, H. V., Vyse, E. R. and Strobeck, C. (1995): Microsatellite analysis of paternity and reproduction in Artic grizzly bears. – J. of Heredity 86, 255-261.

Csanádi, Gy., Vollmann, J., Stift, G. and Lelley, T. (2001): Seed quality QTL identified in a molecular map of early maturing soybean. – Theor. Appl. Genet. 103, 912-919.

Davies, N., Villablanca, F. X. and Roderick, G. K. (1999): Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. – Trends Ecol. Evol. 14, 17-21.

Dawson, D. A., Darby, S., Hunter, F. M., Krupa, A. P., Jones, I. L. and Burke, T. (2001): A critique of avian CHD-based molecular sexing protocols illustrated by a Z-chromosome polymorphism detected in auklets. – Mol. Ecol. Notes 1, 201-204.

Excoffier, L., Smouse, P. E. and Quattro, J. M. (1992): Analysis of molecular variance inferred from metric distances among DNA haplotypes; application to human mitochondrial DNA restriction data. Genetics 131, 479-491.

Ferraris, J. D. and Palumbi, S. R. eds. (1996): Molecular zoology. Advances, strategies and protocols. – Wiley, New York

Fraser, D. J., Bernatchez, L. 2001: Adaptive evolutionary conservation: towards a unified concept for defining conservation units. — Mol. Ecol. 10, 2741-2752.

Fridolfsson, A.-K. and Ellegren, H. (1999): A simple and universal method for molecular sexing of non-ratite birds. – J. Avian Biol. 30, 116-121.

Friesen, V. (2000): Introns. In: Baker, A. J. (ed): Molecular methods in ecology – Blackwell Sci., Oxford, 274-294 pp.

Gloor, G. B., Preston, C. R., Johnson-Schlitz, D. M., Nassif, N. A., Phyllis, R. W., Benz, W. K., Robertson, H. M. and Engels, W. E. (1993): Type I repressor of P element mobility. – Genetics 135, 81-95.

Goldstein, D. B. and Schlötterer, C. eds. (1999): Microsatellites: Evolution and Applications – Oxford Univ. Press, Oxford

Griffith, S. C., Owens, I. P. F. and Tuman, K. A. (2002): Extra pair paternity in birds: a review of interspecific variation and adaptive function. – Mol. Ecol. 11, 2195-2212.

Griffiths, R., Double, M. C., Orr, K. and Dawson, J. G. (1998): A DNA test to sex most birds. – Mol. Ecol. 7, 1071-1076.

Griffiths, R. (2000): Sex identification using DNA markers. In: Baker, A. J. (ed): Molecular methods in ecology – Blackwell Sci., Oxford, 295-321 pp.

Hakki, E. E. and Akkaya, M. S. (2000): Microsatellite isolation using amplified fragment length polymorphism markers: no cloning, no screening. – Mol. Ecol. 9, 2149-2154.

Hagelberg, E. and Clegg, J. B. (1991): Isolation and characterization of DNA from archeological bone. – Proc. Roy. Soc., London B. 244, 45-50.

Hagelberg E., Goldman N. and Lio, P. (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. — Proceedings of the Royal Society of London B, 266, 485-492.

Harvey, P. H. and Pagel, M. D. (1991): The comparative methods in evolutionary biology. — Oxford Univ. Press, Oxford

Hedrick, P. W. (2001): Conservation genetics: where are we now? – Trends Ecol. Evol. 16, 629-636.

Hillis, D. M., Moritz, C and Mable, B. K. eds. (1996): Molecular systematics (2nd ed). Sinauer Ass. Inc., Sunderland USA

Horn, G. T., Richards, B. and Klinger, K. W. (1989): Amplification of highly polymorphic VNTR segment by the polymerase chain reacton. — Nucl. Acids. Res. 17, 6463-6472

Hubby, J. L. and Lewontin, R. C. (1966): A molecular approach to the study of genic heterozygosity in natural populations. I. – Genetics 54, 577-594.

Ihaka, R. and Getleman, R. (1996): R: A language for data analysis and graphics. – J. Comput. Graph. Stat. 5, 299-314.

Jarne, P. and Lagoda, P. J. L. (1996): Microsatellites, from molecules to populations and back. – Trends Ecol. Evol. 11, 424-429.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985): Individual-specific "fingerprints" of human DNA. — Nature 316, 76-79

Jeffreys, A. J., Neumann, R. and Wilson, V. (1990): Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. — Cell 60, 473-485

Jones, C.J., Edwards, K.J., Castaglione, S, Winfield, M.O., Sala, F., van de Viel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N., Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez and Karp, A. (1997): Reproducibility testing of RAPD, AFLP and SSR markers in plants by network of European laboratoires. — Molecular Breeding 3:381-390

Karl, S. A. (1996): Application of anonymous nuclear loci to conservation biology. In: Smith, T. B. and Wayne, R. K. (eds): Molecular genetic approaches in conservation – Oxford Univ. Press, Oxford, 38-53 pp.

Karp, A., Isaac, P. G. and Ingram, D. S. eds. (1998): Molecular tools for screening biodiversity. Plants and animals. – Chapman & Hall, London

Kingman, J. F. C. (1982): The coalescent. — Stoch. Proc. Appl. 13, 235-248.

Kohn, M. H. and Wayne, R. K. (1997): Facts from faces revisited. – Trends Ecol. Evol. 12, 223-227.

Kovács, G. M., Rudnóy, Sz., Vágvölgyi, Cs., Lásztity, D., Rácz, I., and Bratek, Z. (2001): Intraspecific invariability of the internal transcribed spacer region of rDNA of the truffle *Terfezia terfezioides* in Europe. — Folia Microbiologica 46, 423-426.

Lander, E. S. and Botstein, D. (1989): Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. – Genetics 121, 185–199.

Lander, E. S. and Green, P. (1987): Construction of multilocus genetic linkage maps in humans. — Proc Natl Acad Sci USA 84, 2363–2367.

Lewontin, R. C. and Hubby, J. L. (1966): A molecular approach to the study of genic heterozygosity in natural populations. II. – Genetics 54, 595-609.

Linacero, R., Rueda, J. and Vázquez, A. M. (1998): Quantification of DNA. In: Karp, A., Isaac, P. G. and Ingram, D. S. (eds.): Molecular tools for screening biodiversity. Plants and animals. – Chapman & Hall, London, , 18-21 pp.

Lu, J., M.R., Knox, M.J., Ambrose, J.K.M. Brown and T.H.N. Ellis. (1996): Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. — Theor.Appl.Genet. 93,1103-1111.

Luikart, G. and England, P. E. (1999): Statistical analysis of microsatellite data. — Trends Ecol. Evol. 14, 253-256.

Lunt, D. H. and Hyman, B. C. (1997): Animal mitochondrial DNA recombination. Nature, 387, 247

Lynch, M. and Milligan, B. G. (1994): Analysis of population genetic structure with RAPD markers. – Mol. Ecol. 3, 91-99.

Meglécz, E., Néve, G., Pecsenye, K. and Varga, Z. (1999): Genetic variations in space and time in Parnassius mnemosyne (L.) (Lepidoptera) populations in north-east Hungary: implications for conservation. – Biol. Conservation 89, 251-259.

Milbourne, D., Meyer, R., Bradshaw, J.E., Baird, E., Bonar, N., Provan, J., Powell, W. and Waugh, R. (1997): Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. — Molecular Breeding. 3, 127-136.

Milot, E., H. Gibbs, L. and Hobson, K. A. (2001): Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). — Mol. Ecol. 9, 667-681.

Mueller, U. G. and Wolfenbarger, L. L. (1999): AFLP genotyping and fingerprinting. – Trends Ecol. Evol. 14, 389-394.

Mullis, K. B. and Faloona, F. A. (1987): Specific synthesis of DNA in vitro via polymerase catalysed reacton. — Meth. Enzymol. 255, 335-350

Nadon, R. and Shoemaker, J. (2002): Statistical issues with microarrays: processing and analysis. – Trends Genet. 18, 265-271.

Nei, M. and Kumar, S (2000): Molecular evolution and phylogenetics. — Oxford Univ. Press, Oxford

Nei, M. and Li, W. H. (1979): Mathematical model for studying genetic variation in terms of restriction endonucleases. – Proc. Natl. Acad. Sci. of USA 76, 5269-5273.

Nordborg, M. (2001): Coalescent theory. In: Balding D. J., Bishop M. Cannings C. (eds): Handbook of statistical genetics — John Wiley, Chichester, UK. 179-212 pp.

Moritz, C. (1994): Defining `evolutionary significant units` for conservation. – Trends Ecol. Evol. 9, 373-375.

Paetkeau, D., Shields, G. F. and Strobeck, C (1998): Gene flow between insular, coastal and interior populations of brown bears in Alaska. – Mol. Ecol 7, 1283-1292.

Page, R. D. and Holmes, E. C. (1998): Molecular evolution: a phylogenetic approach. — Blackwell Sci., Oxford

Palumbi, S. R. (1996): Macrospatial genetic structure and speciation in marine taxa with high dispersal abilities. In: Ferraris, J. D. and Palumbi, S. R. eds.: Molecular zoology – Wiley, New York, 101-118 pp.

Palumbi, S. R. and Baker, C. S. (1996): Nuclear genetic analysis of population structure and genetic variation using intron primers. In: Smith, T. B. and Wayne, R. K. (eds): Molecular genetic approaches in conservation – Oxford Univ. Press, Oxford, 25-37 pp.

Posada, D. and Crandall, K. A. (2001): Intraspecific gene genealogies: trees grafting into networks. – Trends. Ecol. Evol. 16, 37-45.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. and Rafalski, A. (1996): The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. — Molecular Breeding. 2, 225-238.

Primmer, C. R., Moller, A. P. and Ellegren, H. (1996): A wide range survey of cross-species microsatellite amplification in birds. – Mol. Ecol. 5, 365-378.

Provan, J., Powell, W. and Holingsworth, P. M. (2001): Chloroplast microsatellites: new tools for studies in plant ecology and evolution. – Trends Ecol. Evol. 16, 142-147.

Ragot, M and Hoisington, D.A. (1993): Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. -Theor. Appl. Genet. 975-983

Randi, E. (2000): Mitochondral DNA. In: Baker, A. J. (ed): Molecular methods in ecology – Blackwell Sci., Oxford, 136-167 pp.

Ritland, C. and Ritland, K. (2000): DNA-fragment markers in plants. In: Baker, A. J. (ed): Molecular methods in ecology – Blackwell Sci., Oxford, 208-234 pp.

Rouamba, A., Sandmeier, M., Sarr, A. and Ricroch, A. (2001): Allozyme variation within and among populations of onion (Allium cepa L.) from West Africa. — Theor Appl Genet 103:855–861

Rousset, F. and Raymound, M. (1997): Statistical analyses of population gnetic data: new tools, old concepts. – Trends Ecol. Evol. 12, 313-317.

Russell, J.R., Fuller, J.D., Macaulay, M., Hatz, B.G., Jahoor, A., Powell, W. and Waugh, R. (1997): Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. — Theor.Appl.Genet. 95:714-722

Saiki, R. K., Gelfand, D. H. , Stoffel, S., Scharf, S. J. , Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988): Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. – Science, 239, 487–491

Sambrook, J., Fritsch, E. F. and Maniatus, T. (1989): Molecular cloning: a laboratory manual. 2nd ed. – Cold Spring Harbor Laboratory Press, New York

Scribner, K. T. and Pearce, J. M. (2000): Microsatellites: evolutionary and methodological background and empirical applications at individual, population and phylogenetic levels. In: Baker, A. J. (ed): Molecular methods in ecology – Blackwell Sci., Oxford, 235-273 pp.

Sibley, C. G. and Ahlquist, J. E. (1990): Phylogeny and classification of birds. Yale Univ. Press, New Haven

Smith, T. B. and Wayne, R. K. eds. (1996): Molecular genetic approaches in conservation – Oxford Univ. Press, Oxford

Snow , A. A. and Parker, P. G. eds. (1998): Special feature – Molecular technics in ecology. – Ecology 79, 359-425.
Sorensen, M. D., Ast, J. C., Dimcheff, D. E., Yuri, T. and Mindell, D. P. (1999): Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. – Mol. Phylogenet. Evol. 12, 105-114.
Štorchová, H., Hrdlicková, R., Chrtek Jr., J. , Tetera, M., Fitze, D. & Fehrer, J. (2000): An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. — Taxon 49, 79.
Sunnucks, P. (2000): Efficient genetic markers for population biology. — Trends Ecol. Evol. 15, 199-203.
Stephens M. (2001): Inferences under the coalescent. In: Balding D. J., Bishop M. and Cannings C. (eds): Handbook of statistical genetics — John Wiley, Chichester, UK. 213-238 pp.
Taberlet, P., Waits, L. P. and Luikart, G. (1999): Non-invasive genetic sampling: look before you leap. – Trends Ecol. Evol. 14, 323-327.
Takahasi, K., Terai, Y., Nishida, M. and Okada, N. (1998): A novel family of short interspersed repetitive elements (SINEs) from cichlids. – Molecular Phylogeny and Evolution 15, 391-407.
Templeton, A. R. (1998): Nested clade analysis of phylogenetic data: testing hypothesis about gene flow and population history. – Mol. Ecol. 7, 381-397.
Valiere, N. and Taberlet, P. (2000): Urine collected in the field as a source of DNA for species and individual identification. – Mol. Ecol. 9, 2149-2154.
Vollmann, J., Fritz, C. N., Wagentristl, H. and Ruckenbauer, P. (2000): Environmental and genetic variation of soybean seed protein content under Central European growing conditions. – J. Sci. Food Agric. 80, 1300-1306.
Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995): AFLP: a new technique for DNA fingerprinting. — Nucleic Acids Res 23, 4407–4414.
Wallis, G. P. (1999): Do animal mitochondrial genomes recombine? – Trends Ecol. Evol. 14, 209-210.
Waples R.S. (1991) Pacific Salmon, *Oncorhynchus* spp. & the definition of `species` under the endangered species act. — Marine Fisheries Reviews, 53, 11-22.
Waples R.S. (1995) Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. In: Evolution and the Aquatic Ecosystem: Defining Unique Units in Population Conservation (eds Nielsen, J.L. and Powers, G.A), 8-27 pp. Symposium 17. American Fisheries Society, Bethesda, Maryland. pp. 8-27.
Waser, P. M. and Strobeck, C. (1998): Genetic signatures of interpopulation dispersal. – Trends Ecol. Evol. 13, 43-44.
Welsh, J. and McClelland, M. (1990): Fingerprinting genomes using PCR with arbitrary primers. — Nucl. Acids. Res. 18, 7213-7218.
Wendel, J. F. and Weeden, N. F. (1989): Visualization and interpretation of plant isozymes. In: Soltis, D. E. and Soltis, P. S. (eds): Isozymes in plant biology. — Dioscorides Press, Portland, Oreg, 5–45 pp
Wetton, J. H., Burke, T., Parkin, D. T. and Cairns, E. (1995):Single locus DNA fingerprinting reveals that male reproductive success increases with age through extra-pair paternity in the house sparrow (*Passer domesticus*). – Proc. Roy. Soc. London B. 260, 91-98.
Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. (1990): DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. — Nucleic Acids Res 18,6531–6535.