Molecular markers in plant ecology. Tansley Review Nr. 63

Bachmann, K.

Published in:
New Phytologist

Citation for published version (APA):
Tansley Review No. 63
Molecular markers in plant ecology

By Konrad Bachmann
Hugo de Vries Laboratory, University of Amsterdam, Kruislaan 318, NL-1098 SM Amsterdam, The Netherlands
(Received 9 November 1993)

Contents

Summary 403
I. Introduction 404
II. Most ecological applications concern markers in the nuclear genome 404
III. DNA polymorphisms differ from allozyme polymorphisms 404
IV. Types of molecular markers 405
1. Literature on techniques 405
2. Marker bands arise by restriction cutting or by PCR amplification of DNA 405
3. PCR amplification and sequencing of the internal transcribed spacers of nuclear genes for ribosomal RNA 406
4. Other general PCR primers useful in ecological research 406
5. Random amplified polymorphic DNA 407
(a) RAPDs have limitations 407
(b) Impurities 407
V. The application of markers in ecology 409
1. The relationship between genotype and phenotype 409
2. Identifying clonal populations and detecting residual variation 410
3. Relict and inbreeding sexual populations: mating system and dispersal 411
4. Introgression 412
VI. From marker to character 413
VII. Outlook 414
References 414

Summary

Various methods from molecular biology reveal sequence polymorphisms in organelle and nuclear DNA that can be used as highly informative markers for the structure and dynamics of genomes at the level of populations and individuals. Molecular markers that can be determined without regard to the phenotype permit an unbiased comparison of the adaptation of organisms to their environment, its genetic basis and its effect on evolution. Several marker types used in ecological research and their uses and limitations are briefly discussed. PCR-based methods, especially arbitrary-primer-based RAPDs, are likely to be most widely used and receive most attention. The limited use of DNA markers for overall quantitative (phenetic) comparisons of ‘genetic variability’ and ‘genetic distance’ is stressed and their power as qualitative markers for any and all relevant regions in the genome is emphasized. Specific applications relevant to plant ecology are illustrated. These are: identification of organism and genotype even where morphology is of little help, as in mycorrhiza; identification of clones in asexually-reproducing species, even when they are widespread and intermingled; determining if genetic variation in clonal populations comes from mutations within clonal lines or from independent origins of clonal lines; reconstructing the genotype phylogeny and fruit dispersal of clonal (apomictic) and inbred selfing organisms, measuring the degree of outcrossing by offspring exclusion analysis; detecting and analyzing introgression and characterizing recombinant genotypes in hybrid zones relative to differential adaptive responses: tracing the phylogenetic origin and extent of ecologically differentiated races or species; characterizing the genetic basis, mapping and isolating the genes responsible for special adaptive responses. In a final outlook, I speculate about unconventional sources of genetic variation affecting the ecological characteristics of plants that will become accessible to experimental analysis with the new molecular methods.

Key words: Adaptation, breeding system, clonal plants, dispersal, DNA fingerprints, introgression, RAPD, RFLP, QTL, VNTR.
I. INTRODUCTION

Methods from molecular biology are invading all fields of biology. The recent publication of a book on Molecular Genetic Ecology (Hoelzel & Dover, 1991), the appearance of a journal Molecular Ecology in 1992 and the molecular emphasis of the Symposium ‘Genes in Ecology’ (Berry, Crawford & Hewitt, 1992) are clear signals that, even in ecology, molecular methods have arrived to stay. The term ‘ecology’ will evoke in most biologists a vision of complex interactions among plants, animals and microorganisms under the constraints of soil and climate. The application of methods dealing with the nucleotide sequence of DNA to such a system may either look like the surrender to an irrelevant fashion and likely to distract from the essential questions of the field, or it may be welcomed as the long-awaited simple and general tool to unravel ecological complexity from the inside out. In fact, it supplies a very powerful set of tools for the investigation of some specific ecological questions. Molecular research requires relatively specialized skills and equipment. Easier and less expensive methods for molecular analysis are constantly being developed, but, as in all research, the results from quick and easy approaches usually are preliminary and suggestive. The full potential of molecular methods can only be realized when a considerable investment has been made. It is therefore important to know which molecular methods are available and what they may contribute to ecological research. This is a rapidly developing field, and I shall try to present a balanced view of the present situation.

II. MOST ECOLOGICAL APPLICATIONS CONCERN MARKERS IN THE NUCLEAR GENOME

The direct genetic manipulation of adaptive characters of plants (e.g. Nelson et al., 1988; Misra & Gedamu, 1989; Tarczynski, Jensen & Bohnert, 1993) provides the greatest challenge for ecological research. At the moment, genetic manipulation of adaptive characters is mainly a topic in agricultural genetics. Ecologists tend to see the dangers of genetically modified plants for natural populations more clearly than the unique research opportunities offered by controlled changes in adaptive characters. Consequently, molecular methods are used in ecological research mainly to obtain markers for the detailed description and observation of natural populations with the least possible interference (Klinger, Arriola &Ellstrand, 1992; Williamson, 1992).

Molecular markers can label the mitochondria, chloroplast, or nuclear genomes. It may be noted that the plant mitochondrial genome differs significantly from that of animals. Plant mitochondrial DNA has a particularly slow rate of sequence evolution but plant mtDNA regularly undergoes sequence rearrangements through recombination among a limited number of repetitive elements (Grabau, 1985; Palmer & Herbon, 1988; Palmer et al., 1990). The plant mitochondrial genome is predominantly maternally inherited (Conde, Pring & Levings, 1979), even in conifers in which chloroplast DNA typically shows a paternal inheritance pattern. Due to their origin from a limited number of sequence rearrangements, similar mitochondrial marker patterns can arise frequently and they have played only a minor role in plant ecology (Paige, Capman & Jenetten, 1991; Barrett et al., 1993; Dong & Wagner, 1993; Strauss, Hong & Hipkins, 1993). Most of the ecological applications of molecular methods will deal with the nuclear genome and this will be implied in the following unless organellar DNA is specified.

‘Markers’ in the nuclear DNA are polymorphisms in the nucleotide sequence at homologous (allelic) sites. Polymorphic markers can define a multilocus genotype characteristic for an individual or a clone; selected markers can be diagnostic for a population or a species. The recombination of markers can be followed in genetic crosses and the position of markers in the genome can be mapped. In all of this, DNA markers behave like single-gene morphological markers or like allozymes, with the added advantage that their expression is not influenced by development or by environmental factors. To a certain extent, they can complement or replace allozyme markers in the same kinds of investigations. However, there are subtle practical and theoretical differences in the origin of the various kinds of DNA markers. Much inappropriate use can be prevented if we emphasize these differences and do not try to force DNA markers into the established routine of isozyme data-gathering and interpretation.

III. DNA POLYMORPHISMS DIFFER FROM ALLOZYME POLYMORPHISMS

Allozyme polymorphisms deal with a limited group of functional enzymes and mainly concern variations in the coding sequence of the genes. DNA polymorphisms can occur anywhere in the genome including coding and non-coding, single-copy or repetitive DNA. Genome sizes differ considerably even among closely related species (Bennett, 1985; Price, 1988a) and a large proportion of the DNA of one species may have no equivalent in another one (Zentgraf, King & Hemleben, 1992). More and more cases of intraspecific genome size variations are found (Bennett, Smith & Heslop-Harrison, 1982; Price, 1988b; Bennett & Bennett, 1992) which may contribute considerably to differences in the levels of polymorphism. For these reasons, we can expect to
get very different figures for such global measures of genetic variability as the percentage of polymorphic loci and the degree of heterozygosity depending on the kind of marker used to determine them (Zhang et al. 1993).

Allozyme polymorphisms have been employed as markers or as characters, and often the distinction has been blurred. Molecular markers label sites on the genome and in some cases may provide information on the adaptive functions coded at the marked region of the genome. The mutation that is scored as a marker may even be responsible for the different contributions of the marked alleles to fitness, but that has to be determined in each case. Markers, even 'marker phenotypes' (banding patterns compared directly, without genetic interpretation), describe the genotype, while characters describe the phenotype of the organism. In ecology, markers are research tools, characters are research objects. The treatment of allozymes as markers and as characters, and the assumption that markers are selectively neutral and characters are adaptive (and various variations of this), are responsible for the fact that, on balance, isozyme research has probably diverted attention from basic inquiries in the nature of adaptation more than it has furthered our understanding. In plants, especially, the crucial role of phenotypic plasticity in adaptation pointed out by Bradshaw (1965) had essentially to be rediscovered (Jain, 1979; Morisset & Boutin, 1984; Grime, Crick & Rincon, 1986; Schlichting, 1986; Sultan, 1987) and still awaits a definitive treatment. It is not surprising that Scharloo (1989) in the context of adaptation and plasticity speaks of the 'dark age of electrophoresis'. The fascination with bands on gels at the expense of research on the adaptive interactions of the living plant will not be revived by the introduction of DNA polymorphisms if they are consistently viewed as tools that provide access to relevant characters rather than as characters themselves.

One difference between allozyme and DNA markers concerns their number. Theoretically, DNA markers covering the entire genome can be found, and the number and density of available markers depends essentially on the amount of work, the experimental skill and the money invested in finding them. Appropriate markers can be found to suit the specific requirements of any project.

Once a DNA polymorphism has been detected, the method to find it can be used to isolate the sequence around it and study its genetics, evolution and function in detail. The most exciting property of DNA polymorphisms is that they provide direct experimental access to the region marked by them and its function. The marker need not stand for an inaccessible character but can be used to investigate the character itself. This property of DNA markers makes them invaluable tools for an urgent experiment-based reconsideration of the relationship between genetics and whole-organism physiology and development.

IV. TYPES OF MOLECULAR MARKERS

1. Literature on techniques

The basic techniques for obtaining molecular markers have frequently been described. Besides the book by Hoelzel & Dover (1991), there are several very useful books on molecular methods in evolution and systematics (Hillis & Moritz, 1990; Crawford, 1990; Hewitt, Johnston & Young, 1991; Li & Graur, 1991; Soltis, Soltis & Doyle, 1992). The books by Crawford (1990), Soltis et al. (1992) and the volume edited by Brown et al. (1990) also contain much valuable information relevant to the use of molecular methods in plant ecology (Clegg, 1990). Some more recent relevant literature references will be given below. Details of technical protocols and the most suitable enzymes and apparatus develop rapidly and will not be dealt with here. I shall briefly review the principal types of markers and comment on their uses and limitations as they are seen at the moment. Then I shall survey some of the ecological questions in which these markers are or will be useful.

2. Marker bands arise by restriction cutting or by PCR amplification of DNA

The bands on gels that typically serve as molecular markers (Rafalski & Tingey, 1993) arise either from cutting DNA at specific sites with restriction enzymes to detect restriction fragment length polymorphisms (RFLPs) or by in vitro synthesis of a stretch of the target DNA between specific sites to which short single-stranded primers attach and serve as starting (and end) points for a polymerase chain reaction (PCR).

A detectably high concentration of restriction fragments with identical length and nucleotide sequence accumulates after digestion with a restriction enzyme, because copies of homologous sequences are cut to pieces of the same length. In the polymerase chain reaction, repeated synthesis from the same primer sites selectively and exponentially amplifies the sequence between the primer sites and produces a high concentration of identical amplification products. PCR methods (Erlich, 1989) require much smaller samples of DNA than RFLP analyses and are therefore generally preferred. Of course, the methods can be combined. Sequence polymorphisms in PCR amplification products can be detected with restriction digestion (Arnold, Buckner & Robinson, 1991; Rieseberg, Hanson & Philbrick, 1992; Harada et al., 1993), the design of a specific PCR test may well begin with a restriction analysis, and PCR methods for the efficient isolation of cloned inserts from the cloning vectors for RFLP analysis have
been developed (Güüssow & Clackson, 1989; Xu, Magill & Hart, 1993).

3. **PCR amplification and sequencing of the internal transcribed spacers of nuclear genes for ribosomal RNA**

Identical nucleotide sequences of 20–30 nucleotides are so unlikely to arise independently by chance that PCR primers of that length guarantee the homology of the amplified DNA. Such primers are designed individually for regions (genes) of which the sequence is known and are useful only as far as the primer attachment sites are invariant. Ecological applications usually require highly variable sequences with polymorphisms within species or between closely related species. Evolutionary stable primer sites bordering highly variable sequences are found in the nuclear genes for ribosomal RNA (‘rDNA’), where the highly conserved sequences for ribosomal 18S, 5.8S, and 25S RNA are interrupted by the ‘internal transcribed spacers’ ITS1 (between 18S and 5.8S) and ITS2 (between 5.8S and 25S). Both of these spacers are about 200 basepairs in length (longer in conifers, see Karvonen, Karjalainen & Savolainen, 1993) and vary among closely related species (Torres, Galal & Hemleben, 1990; Baldwin, 1992; Wojciechowski et al., 1993; Baldwin, 1993a: Van Houten, Scarlett & Bachmann, 1993). Genes for ribosomal RNA occur in the mitochondrial, chloroplast and nuclear genomes. The ribosomal DNA of the three genomes has evolved independently (Grabau, 1985; Jorgensen & Cluster, 1988; Hamby & Zimmer, 1992) and differs sufficiently so that specific primers amplify nuclear rDNA sequences only.

Transcription units specifying the sequence of the 18S, 5.8S and 25S rRNAs and the transcribed spacers occur in clusters of hundreds of copies in tandem (one after the other), separated by non-transcribed spacers, on one or a few chromosomes in the genome (Long & Dawid, 1980). All of these copies usually have the same sequence, which means that a mutation in one of the repeat units gets eliminated or incorporated in all of them (‘concerted evolution’ of tandem repeated copies: Smith, 1976; Dover, 1982, 1986; Dvorák, 1990). The rDNA sequence is therefore treated as if it were a single-copy gene. This may not work in some cases interesting to ecologists. Considerable polymorphism in the ITS sequences within the genome of a few species of *Amelanchier* (Rosaceae) may be related with their apomictic reproduction which may prevent concerted evolution (Campbell et al., 1993). On the other hand, there are indications that concerted evolution homogenizes ITS sequences in hybrids so that allopolyploid species contain the sequence of only one of the parents (Hahn, Karol & Sýtsma, 1993) or possibly a recombinant sequence (Van Houten et al., 1993).

The ribosomal 5S RNA is also encoded in tandem repetitive units which are different from the 18S-5.8S-25S clusters and usually not linked to them (Long & Dawid, 1980; Dvorák, 1990). The 5S coding sequence is about 120 nucleotides long. The length of the intervening spacers varies. They are about 480 nucleotides long in Brassicaceae (Capesius, 1993). They contain regulatory sequences that are likely to be conserved (Hemleben & Werts, 1988) but overall are considerably more variable than the coding sequences. Primers amplifying the entire Brassicacean repeat unit from conservative parts of the coding region have been designed (Capesius, 1993). From a recent investigation of the 5S sequences from the 35 species of the Triticeae (Poaceae), it appears that there is a considerable amount of sequence variation among the repeat copies within one plant (Kellogg & Appels, 1993). The use of these sequences for population studies remains to be investigated.

4. **Other general PCR primers useful in ecological research**

The PCR amplification of diagnostic DNA markers from small amounts of DNA in mixed samples can help to identify species involved in close ecological associations. One such association is mycorrhiza, in which the identification of the fungus and sometimes even that of the host plant is difficult. Cullings (1992) has developed a plant-specific primer, 28KJ, in the 28S (= ‘25S’ of plants) ribosomal RNA gene. The primer represents a sequence of 25 nucleotides found in all angiosperms and conifers. Six of these nucleotides consistently differ from those found in fungi. Together with the ‘universal’ primer 28C, a part of the 28S gene can be amplified from angiosperm and conifer DNA, while fungal DNA is not amplified.

There is, of course, no guarantee that a particular sequence should be constant in all members of one taxonomic group, and the stability of taxon-specific PCR primers is an empirical property depending on the sample that has been tested.

A taxon-specific primer for the ITS region of fungi and one specific for basidiomycetes have been developed by Gardes & Bruns (1993). The basidiomycete primer (at the 3’ end) combined with a universal or a fungus-specific primer at the 5’ end amplified basidiomycete DNAs and discriminated against ascomycete DNAs. It also produced some amplification product with some plant species, but preferentially amplified fungal DNA when plant and fungal DNA were present together. The fact that PCR primers can attach to imperfectly-matching primer sites and that during the repeated PCR
amplification cycles there is a selection of the most efficiently-amplified sequences needs to be remembered for the interpretation of PCR results, especially the anonymous and statistical amplification products produced by arbitrary short primers.

5. Random amplified polymorphic DNA (RAPD)

PCR with single arbitrary short primers relies on the statistical chance that the complementary primer sites occur somewhere in the genome as inverted repeats enclosing a relatively short stretch of DNA (up to a few thousand nucleotide pairs). The DNA between the two opposite primer sites can be amplified (Arbitrarily Primed PCR, AP-PCR: Welsh & McClelland, 1990; Random Amplified Polymorphic DNAs, RAPDs: Williams et al., 1990, 1993; DNA Amplification Fingerprinting, DAF: Caetano-Anollés, Bassam & Gresshoff, 1991). Hadrys, Balick & Schierwater (1992) discuss the application of the technique to molecular ecology. The RAPD method of Williams et al. (1990) using 10 base-pair (decamer) primers is probably the most convenient for many applications. Since a virtually unlimited number of arbitrary decamer primers can be synthesized, finding a genetic polymorphism even in the least variable population is a question of persistence. Decamer primers are commercially available from various sources. For easy comparison, the standard primer sets from Operon Technologies Inc. (Almeda, Calif.) are frequently used. As a first approximation, amplification products of the same length obtained with one primer pair are likely to be homologous. Homology can be ascertained by sequencing the amplification product. The convenient length and the known terminal (i.e. primer) sequence of RAPD bands facilitates sequencing. This makes RAPDs a method of choice for many molecular approaches: a relatively easy and not too expensive method that requires no previous sequence knowledge.

The standard RAPD procedure uses one decamer primer, separates the amplification products on agarose gels and stains them with ethidium bromide. More sensitive variations have been designed to detect rare polymorphisms among similar genomes, e.g. somatic mutations among clonal derivatives (Collins & Symons, 1993). These methods include using two different arbitrary decamer primers instead of one so that sequences bordered by primer 1 at both ends, those bordered by primer 2 on both ends and those bordered on one end by primer 1 and on the other by primer 2 are amplified. This can produce quite a few amplification products, and these are more sensitively separated on denaturing acrylamide gels and revealed by silver staining (Bassam, Caetano-Anollés & Gresshof, 1991). The banding patterns obtained when amplification products are run on acrylamide gels, even under non-denaturing conditions, i.e. as double-stranded DNA fragments, differ from the patterns obtained when the same products are run on agarose. This is due to the separation of bands comigrating in agarose and the more sensitive staining procedure, and it is not practical to switch between the two separation methods. Since all uncertainties in the interpretation of anonymous banding patterns increase with the number of bands in the pattern, a highly sensitive RAPD method is recommended only when there are very few polymorphisms. Using one primer at a time, separating the amplification products on agarose, concentrating on unequivocal and consistent polymorphisms and treating these as individual qualitative markers is probably the simplest and safest approach for most investigations.

6. Comments on the RAPD method

(a) RAPDs have limitations. The banding patterns obtained with the RAPD method provide relevant information for a wide range of applications, even if the bands are not further characterized by formal genetic analysis, restriction pattern, hybridization or sequencing. When conclusions are drawn directly from banding patterns, some important limitations should be kept in mind, especially for quantitative interpretations.

(b) Impurities. Amplification products obtained with the RAPD method are anonymous pieces of DNA, and since decamer primers amplify DNA from virtually all sources this may include impurities, infections and parasites in the material from which the DNA has been isolated. This may not create problems when clean plants have been raised under controlled conditions and any unspecific DNA is much more dilute than the DNA under study. However, it should always be kept in mind when conclusions are based on anonymous bands.

(c) Statistical nature of random amplification. Amplification from arbitrary decamer primers is essentially a statistical procedure and depends on the criteria used. Primers can bind (with reduced efficiency) to targets containing one or two base mismatches. Also, the distance between primer sites that still allows efficient amplification depends on amplification conditions. There will be imperfect primer attachment sites and site distances such that under the conditions of the experiment sometimes amplification is successful, sometimes not (Ruano, Brash & Kidd, 1991). Standardizing the conditions for a certain project should reduce the number of unreliable amplification products. They cannot be completely excluded and it is safe to disregard weak and unreliable bands. Of course this means also that
quantitative estimates of genetic variability based on RAPD bands are more influenced than the number of isoenzyme loci and alleles or the number of RFLP bands by details of the method.

Several different factors may contribute to differences in the RAPD banding patterns between genotypes and the comparison becomes ever less informative with an increasing phylogenetic distance between the genomes. A precisely repetitive DNA with primer sites at regular intervals will produce a single band, while degenerate repetitive DNAs may produce many bands. Large genomes may statistically produce more bands, and the number of bands may not increase linearly with the number of potential sites due to competition or interaction between amplification products during the chain reaction. Additivity of the bands can always be checked by comparing the amplification products from two samples run separately with those obtained after pooling them (Hadrys et al., 1992). Patterns based on allelic differences in a segregating offspring family are usually additive. Additivity is less certain for the comparison of genomes from different species, which may be important for the analysis of distant hybrids and allopolyploids.

(d) Homology of RAPD bands. Even when banding patterns are reproducible, there is no assurance of the homology of comigrating amplification products. DNA can be isolated from a band, reamplified and labelled and hybridized to a Southern blot of the gel (Southern, 1975). It should label all comigrating bands, and frequently does (Wilkie, Isaak & Slater, 1993), but it seems to be just as easy to find exceptions (Williams et al., 1993). If the hybridization signal is stronger with the band from which the labelled probe has been amplified, it points to internal sequence variation among the comigrating bands (Jessup, 1993). If other bands besides the comigrating ones are labelled, it may point to length alleles at homologous sites or to the existence of multiple homologous RAPD sites in a genome, i.e. paralogous loci (various members of a gene family) or homoeologs in a polyploid (Jessup, 1993). Williams et al. (1990) report that five of 11 rehybridized amplification products from soybean labelled many bands of different lengths that must have come from middle or highly repetitive DNA fractions. Kazan, Manners & Cameron (1993) found that four of five rehybridized amplification products came from repetitive DNA.

(e) Dominant/codominant segregation in genetic crosses. Typically RAPDs are absence/presence polymorphisms due to mutations in primer sites and segregate as dominants (3 : 1) in F2 families (Williams et al., 1990; Carlson et al., 1991; Martin, Williams & Tänksley, 1991; Kazan et al., 1993). However, RAPDs due to short insertions/deletions between non-mutated primer sites have been found. These produce two alternative bands that are both labelled when one is used as the probe and segregate 1 : 2 : 1 in hybrid F2s. Rieseberg et al. (1993) found that 11 among 161 RAPD loci in Helianthus anomalus segregated as codominants. We have sequenced a pair of allelic amplification products from Microseris bigelovii and detected a 30 bp insertion in the longer one (A. Hofman, M. van der Bijl & K. Bachmann, unpublished).

Sequence variation between constant primer sites in allelic amplification products can be used to generate codominant markers. For this, the amplified DNA is digested with a restriction enzyme that cuts frequently (Harada et al., 1993) and then separated on the gel. Allelic bands of the same length differing by a restriction site mutation will be converted from non-polymorphic comigrating bands into length variants.

7. Restriction length polymorphisms in nuclear and chloroplast DNA

Digestion of a nuclear genome with a restriction enzyme produces so many fragments that individual homologous bands will have to be identified by hybridization with a labelled probe in a Southern blot of the gel (Apuya et al., 1988). The method therefore depends on the availability of suitable probes. Cloned segments of the conservative parts of ribosomal genes hybridize to homologous regions from many species and have been used extensively to demonstrate restriction site variation in the rDNA within and among populations (Schaal & Lear, 1988). Otherwise, probes have been specifically cloned from the genome that is to be analyzed ('homologous probes'; Keim et al., 1989; Hughes et al., 1990). Many of these will hybridize to homologous segments in related species ('heterologous probes'; Van Houten et al., 1993).

Ecologically relevant results obtained with RFLP analysis of nuclear DNA include the distinction between the weedy and serpentine races of the wild sunflower, Helianthus bolanderi (Rieseberg, Soltis & Palmer, 1988) with RFLPs in the rDNA and the study of herbivore-plant interaction in a cottonwood hybrid zone (Paige & Capman, 1993), which I shall describe below.

There is a voluminous literature on restriction analysis of chloroplast DNA in plant systematics. The method is best suited for the analysis of subgenera and genera and therefore of limited use in ecological studies. Of course there is no reason why cpDNA should not vary within species, and some intraspecific variation is regularly found in sufficiently extensive surveys (Soltis, Soltis & Milligan, 1992a). The relative rarity of this variation together with the (usual) maternal inheritance of cpDNA in angiosperms can make individual variants highly
informative about diploid and polyploid hybrid taxa and introgression (Rieseberg & Brunsfeld, 1992), the origin of allopolyploids (Soltis, Doyle & Soltis, 1992b), intraspecific biogeography (Soltis et al., 1992c) and the recognition of ecotypes. The 'weedy' and 'serpentine' races of the wild sunflower, Helianthus bolanderi (Rieseberg et al., 1988) mentioned above have also been distinguished by cpDNA markers.

8. Minisatellites, variable number tandem repeats (VNTRs)

A variation of the RFLP method uses various suitable probes to detect short tandem repeated sequences in which the number of repeats between flanking restriction sites is highly variable. This causes variations in the length of restriction fragments containing the repeats. 'Minisatellites' are detected with probes isolated from human DNA (Jeffreys, Wilson & Thein, 1985) or from the phage M13. The sequence of a minisatellite from the wheat genome has been determined by Martienssen & Baulcombe (1989). The method has been applied since 1988 to detect polymorphisms in plants (Dallas, 1988; Rogstad, Patton & Schaal, 1988a; Rykov et al., 1988) and various early uses have been summarized by Bachmann (1992). Very simple synthetic oligonucleotides, e.g. (GATA), regularly detect VNTRs ('microsatellites') (Tautz & Renz, 1984; Weising et al., 1989, 1991; Klinklicht & Tautz, 1992). Typically, one probe detects VNTRs at many highly variable loci in the genome and produces a many-banded DNA-'fingerprint'. Since the repeat sequences are short and ubiquitous, hybridization of the blotted DNA with a common probe in this case is no proof of homology of the labelled bands. As in RAPDs, bands of the same length are likely to be homologous and the similarity of patterns can be calculated via a band-sharing index (Lynch, 1988, 1990; Piper & Rabenold, 1992). Since VNTRs are among the most variable polymorphisms (Van Houten et al., 1991) and consequently have many alleles at each locus (Nybom et al., 1992), they are useful for the detection of close relatives, including clonal derivatives (Rogstad, Nybom & Schaal, 1991). Generally, the method is considered more difficult and expensive than 'fingerprinting' with RAPDs (Weatherhead & Montgomery, 1991). It needs more starting material, requires a labelled probe and produces anonymous bands. To exploit the high variability of the VNTRs, individual VNTR loci can be cloned and sequenced (Condit & Hubble, 1991), and specific PCR primers for the unique locus-specific sequences flanking the VNTR can be designed. These can be used to detect length alleles of individual VNTR loci via PCR. This is one of several ways to use site-specific sequence information to mark a locus (to create a 'sequence-tagged site'). PCR of specific microsatellite loci has been used intensively for mapping polymorphisms in the human and rodent genomes (Weissenbach et al., 1992; Serikawa et al., 1992). For most applications in plant ecology, the preparatory work needed to obtain specific primers is probably too difficult and expensive. However, when specific and highly polymorphic markers are needed for larger routine surveys, the isolation of VNTR loci should be considered. One ongoing application is a survey of the United States National Brassica and Malus genetic resources collections at the Plant Genetic Resources Unit of the USDA-ARS in Geneva, New York (Lamboy et al., 1993).

Much technical information and a complete list of the literature on DNA fingerprinting is contained in the five volumes of 'Fingerprint News' (1989–1993) published by Drs W. Amos and J. Pemberton of the Department of Genetics in Cambridge (UK).

V. THE APPLICATION OF MARKERS IN ECOLOGY

1. The relationship between genotype and phenotype

DNA markers can identify the organism and its taxonomic association (Kurtzman, 1987), even from fragmentary remains (Bruns, Fogel & Taylor, 1990; Hoss et al., 1992) and even where morphology cannot distinguish strains (Schäfer & Wöstemeyer, 1992). Marked genetic loci can be followed through recombination and gene flow. In plants, this can be used to investigate breeding systems (Brown, 1990) and dispersal (Barrett & Husband, 1990). Differences and similarities between organisms can be related to geographic and ecological distribution. Where organisms with different ecological adaptations can be crossed and produce hybrid offspring, molecular markers will greatly facilitate the genetic analysis of adaptive character differences and provide access to the relevant genes at the level of DNA sequences (Rafalski & Tingey, 1993).

There seems to be general agreement that natural selection creates organisms that are specifically and heritably adapted to the biotic and abiotic factors in their environment. At the same time, there seems to be little consensus on the speed and flexibility of heritable adaptive change: is each character state of each organism an optimal genetically-determined response to the present conditions? How closely do (genes affecting) adaptive characters track environmental changes in space and time? How physiologically or genetically independent are changes in various characters? How can the genetic variability needed for further adaptation be maintained without reducing the present adaptiveness? In the following, I shall point out how the detailed monitoring of the genotype possible with molecular markers can provide information on the degree to which genotypes of organisms are shaped by the environment.
without having to infer genotypes from phenotypes. As I have suggested above, this advance in methods should help to resolve conceptions about the relationship between genotype and phenotype and the nature of phenotypic adaptation which are more a concession to earlier limitations in methods than a result of the objective analysis of the ecologically relevant facts.

2. Identifying clonal populations and detecting residual variation

Each plant in a population may have a unique genotype through recombination in a variable gene pool, a local population of plants may consist of the genetically identical clonal offspring of a single ancestor, or the genetic variation of a population may have a more complex structure depending on mating system, gene flow, dispersal and selection. Isozymes have become routine indicators of the genetic structure of natural populations. DNA polymorphisms frequently allow a more detailed and statistically-reliable analysis, especially when genetic variation is low as it frequently is within clonal and inbreeding populations.

Nyborg & Schaal (1990) have collected 20 second-year shoots each of the two raspberry species, Rubus occidentalis and R. pensylvanicus, along a 600 m stretch of a dirt road near St Louis, Missouri. Multilocus VNTR fingerprinting with the phage M13 probe differentiated 15 genotypes among the 20 shoots of the predominantly sexual R. occidentalis, but only five genotypes in the presumably apomictic R. pensylvanicus. Rogstad et al. (1991) have used the M13 probe to analyze the clonal structure in quaking aspen (Populus tremuloides) and found that it can detect morphologically cryptic aspen clones interdigitating with others. This single-probe test seems to be more sensitive than the isozyme analyses of Mitton & Grant (1980).

Single clones can be enormous in size and age (Cook, 1983). A clone of quaking aspen has been reported to cover an area of 43 ha with about 47000 stems (Barnes, 1975) and a genetically uniform mycelium of the fungus Armillaria bulbosa has covered 15 ha. That this was a clone rather than a group of inbreed siblings, was confirmed by demonstrating uniform heterozygosity at five RFLP and 11 RAPD loci (Smith, Bruhn & Anderson, 1992).

While individual clones in clonal plants usually seem to be restricted to one or a few populations (Ellstrand & Roose, 1987; Parks & Werth, 1993), single clones can sometimes be dispersed over a large geographic area. DNA markers have only recently been applied to determine clonal diversity and distribution in vegetatively propagating aquatic weeds (Furnier & Olfelt, 1993) which may form clones that are both locally very large and also widely distributed (Cook, 1987). Many of these are of considerable ecological interest due to their invasiveness.

The dandelion genus Taraxacum (Asteraceae) contains sexual diploid and many apomictic polypl oid species which form seed of the maternal genotype from parthenogenetically developing unreduced eggs (Richards, 1973). Each of hundreds of named ‘microspecies’ presumably is the clonal offspring of one original apomict hybrid. Many clones of apparently independent origin can coexist in natural dandelion populations, and clone identification on the basis of morphology alone is problematical. Menken & Morita (1989) have shown with isozyme electrophoresis that the Japanese pentaploid obligate agamosperm, Taraxacum albidum, seems to consist of a single clone across its entire distribution area. Van Heusden, Roupppe van der Voort & Bachmann (1991) have used multilocus microsatellite fingerprinting to show that plants of the European microspecies T. hollandicum collected in Czechoslovakia, France and The Netherlands seem to be members of a single clone differing by a few mutations. In fact, plants collected as representatives of three different named microspecies in The Netherlands had identical fingerprint patterns. The very wide distribution of single clones together with the coexistence of many clones at one site raises questions about adaptation that can be addressed once clones can be unequivocally identified.

Asexual clones will accumulate mutations. King & Schaal (1990) have found non-parental restriction variants in the ribosomal 18S-25S and the alcohol dehydrogenase-1 genes in the apomictic offspring of two dandelions, and a few polymorphic fingerprint bands were found in the asexual microspecies by Van Heusden et al. (1991). The statistical treatment of band-sharing in fingerprint patterns of clonal populations accumulating mutations in various lineages has been examined by Brookfield (1992).

Polymorphisms in clonal reproduction evolve in a strictly cladistic manner even at the level of the individual (‘ramet’) so that the phylogeny and geographic dispersal of clones can be reconstructed. King (1993) has examined North American dandelions, which have originally been introduced from Europe, and European ones. Some European genotypes were found also in North America and some mutations in the rDNA and chloroplast DNA have apparently arisen in North America. However, more of the observed genetic variation is apparently due to multiple origins of similar clonal lines (from sexual dandelions in Europe) than to evolution within a clonal line. Battjes, Menken & Den Nijs (1992) have shown with isozymes that both uniclonal and highly variable ‘microspecies’ of Taraxacum coexist in Czechoslovakia. Not all dandelions are invasive weeds. Some microspecies have very narrow ecological requirements (Sterk, Groenhart & Mooren, 1983). With the availability of many nuclear and
cytoplasmic markers, the possibilities for ecological genetic investigations of *Taraxacum* and other agamic complexes have greatly increased.

3. Relict and inbreeding sexual populations: mating system and dispersal

The genetic diversity in narrowly endemic or relict populations with sexual reproduction and in populations of predominant selfers can be as narrow as that in clonal populations. Crawford *et al.* (1991) have found no variation in 22 allozyme loci in 83 plants from 12 populations of *Lactoris fernandeziana* (Lactoridaceae), an endemic polygamo-dioecious shrub of the island of Masatierra in the Juan Fernandez Archipelago. A total of 27 plants from 15 populations were later examined for RFLPs in the 18S–25S rDNA and for RAPDs. Different rDNA repeat lengths and restriction site mutations were detected within individuals and within and among populations. Of 106 RAPD bands per plant produced with 16 primers, 26 were polymorphic. Fifteen of the 27 plants had at least one variant band (Brauner, Crawford & Stuessy, 1992). RAPDs are obviously a key to finding residual genetic variation. However, no variation has been discovered at 23 isozyme loci, and none with RAPDs using 69 primers in the highly self-fertile red pine, *Pinus resinosa*, in Newfoundland (Mosseler, Egger & Innes, 1993). If this is the effect of a post-glacial population bottleneck, the species has had about 12000 yr to recover.

We have to keep in mind that RAPD bands can arise from all kinds of DNA fractions including species-specific repetitive DNA and that the kinetics of amplification select among potential amplification sites. RAPD banding patterns derived from different species are considerably less likely than isozyme patterns to represent the genetic variation at a comparable sample of loci in the various genomes, and bands of the same strength and position on the gel are less likely to mark identical alleles of homologous loci. In addition, the dominant inheritance of most RAPD bands adds a factor of uncertainty to allele counts and estimates of heterozygosity. The routine calculations developed for isozyme alleles should be applied to RAPD data with critical care, especially since the high numerical precision that can be achieved with a large data set may disguise the uncertainty about the factors contributing to the measured diversity. The possibility that we get very precise data on effects that are effectively evolutionary noise is as high with anonymous RAPD markers as it is with quantitative phenotypic variation without a knowledge of the genetic basis.

This consideration certainly is valid for the quantitative comparison of levels of genetic variation and for the interpretation of genetic similarities or differences among populations or species. Since we (Van Heusden & Bachmann, 1992c, table 5) have published one of the first interspecies comparisons of genetic variation based on RAPDs, I can use this to point out that we used different numbers of samples per species and different numbers of primers, both of which can create enough noise to hide any real difference. However, we were sure that the genomes were of about the same size, ploidy and chromosome number and we made quite clear that ours was a rough estimate. Since we found a decrease in genetic variability where we used more primers, the direction of the effect is likely to be biologically significant. Experience with generating RAPD maps in these species has confirmed this. It is hoped that no quantitative comparisons of genetic variabilities or genetic distances based on RAPD data using different (numbers of) primers in different sized samples and applying different criteria for band recognition in unrelated species with different genome sizes and ploidies will be attempted.

Our comparison was a byproduct of a project primarily concerned with the structure of the intraspecific variation in three annual species of *Microseris* (Asteraceae). All three species, *M. elegans* and *M. bigelovii* of North America and the closely related *M. pygmaea* of Chile, have very patchy distributions in isolated local populations, and selfing is the predominant mode of reproduction. The populations frequently consist of one or very few morphologically distinct 'local biotypes' which breed essentially true and can be indefinitely maintained as inbred strains. It appears that local populations arise from founder events, possibly founder individuals, and that only occasionally plants from different origins colonize the same site and then may hybridize to create a temporary 'hybrid swarm' and eventually a new recombinant local biotype (Chambers, 1955). Except for occasional bursts of recombination, evolution in such inbreeding populations resembles that of clonal populations, and it should be possible to reconstruct an intraspecific cladistic tree of stepwise dispersals and founder events. RAPD markers are ideal for this because of their great number and the fact that we compare essentially homozygous biotypes. The results were clear and showed subtle differences among the three species. *M. elegans* conformed most closely to the expected pattern. Populations containing closely related biotypes are interspersed with populations containing genetically very different plants, and a pattern of very rare intraspecific achene (fruit) dispersal to unoccupied suitable sites can be inferred (Van Heusden & Bachmann, 1992a). The Chilean *M. pygmaea* is the result of long-distance dispersal from North America and has spread from the point of establishment in two genetically isolated series of populations, one at the coast and one inland (Van Heusden & Bachmann, 1992c). *M. bigelovii* has a very extended, nearly linear, distribution along the
Pacific coast of North America from Southern California to mid-Oregon with a few disjunct populations near Victoria, British Columbia. A cladistic analysis of RAPD bands of biotypes from along the distribution range clearly clustered multiple isolates from the same or neighbouring populations and confirmed that the Canadian disjunct populations are the result of a single colonization event. Beyond this, the RAPD markers were essentially randomized among populations so that the cladogram collapsed into a star-shaped polytomy (Van Heusden & Bachmann, 1992b). This means that gene flow by achene dispersal is rare enough to allow local populations to evolve characteristic biotypes by inbreeding and selection, but still sufficient to randomize allele distributions throughout the range.

Besides the indication that dispersal events too rare for direct observation suffice to randomize alleles throughout the range of a species, our results with Microseris raise an ecological question. Successful long-distance seed dispersal requires the dispersed genotype to be sufficiently adapted to the target site to become established. It is likely that the chances for establishment are orders of magnitude lower than those of reaching a site, and they may involve a very strong selective effect. However, there is little evidence for an adaptive basis of the character distribution in the annual Microseris. An exception is a north–south gradient in the genetically determined flowering times in M. bigelovii (Bachmann et al., 1987). Most of the visible characters and RAPD marked sites that seem to vary randomly among the populations may either contribute to local adaptations that we are unable to recognize or they are irrelevant for adaptation. We are now producing genetically-marked hybrid offspring in which the characters of the parental biotypes appear in random combinations that can be used for selection experiments under natural circumstances.

Founder events can be inferred even in the presence of much recombination when a disjunct population contains markers that are a subset of the central marker pool. In many cases, isozymes are sufficient for this demonstration (e.g. Watson, Kornkven & Miller, 1993, in Eriocaulon kornickianum). Recently, RAPDs have confirmed isozyme data for the derivation of disjunct populations of flowering dogwood, Cornus nuttallii (Brunsfeld & Rieseberg, 1993).

Fritsch & Rieseberg (1992) illustrate in an exemplary manner the determination of outcrossing rates in partially self-fertilizing populations. Their example is Datisca glomerata (Datiscaeae), the only plant known to maintain androdioecy, i.e. male and hermaphrodite individuals in a breeding population (Liston, Rieseberg & Elias, 1990). A cladistic treatment of chloroplast DNA restriction site data has shown that, contrary to the expected scenario (Bawa, 1980), androdioecy in Datisca is derived from dioecy and not an intermediate step from hermaphroditism to dioecy (Rieseberg et al., 1992).

Androdioecy should only be stable under rather stringent circumstances: the fertility of males should be at least twice the male fertility of the hermaphrodites, and even greater in partially self-fertilizing populations (Charlesworth & Charlesworth, 1978; Charlesworth, 1984). A previous study of outcrossing in Datisca using isozymes had revealed an inbreeding coefficient (F) of 0.671, which suggested high rates of self-fertilization (Liston et al., 1990). F, however, is influenced by any type of non-random mating and selection. Outcrossing estimates can be generated by examining allelic variation over many loci in progeny arrays from plants allowed to pollinate naturally (open pollination) and using a maximum likelihood model to exclude progeny resulting from self-fertilization (Ritland, 1990; Lewis & Snow, 1992). Isozyme methods did not detect enough loci for progeny exclusion in Datisca. From a survey of 340 decamer primers 12 polymorphic RAPD loci were selected to determine multilocus outcrossing rates. These were considerably higher than the original estimates based on isozymes.

4. Introggression

The genetics of ecologically-relevant characters can be quite different from that of the visible markers. Even many randomly-distributed markers are a biased representation of the genome, because they have been selected for variability. Many crucial autecological characters and their genetic basis may be uniform throughout a species or a higher taxon, and population genetic parameters may be irrelevant for quite a few ecological problems. However, when intraspecific hybridization and introgression cause genetic recombination, ecological consequences seem to be the rule (Anderson, 1949). The contribution of molecular methods to our understanding of introgression is already substantial. Rieseberg & Brunsfeld (1992) have reviewed the topic and shown that morphological evidence suggesting introgression has been misleading in some cases while quite a few cases of introgression are not recognizable in the morphological variation. Especially, more and more cases of cytoplasmic introgression without (detectable) nuclear introgression are being found when chloroplast and nuclear phylogenies are compared.

Even where morphology is sufficient to demonstrate introgression, molecular markers reveal additional significant details. Arnold et al. (1991) have compared RAPD and chloroplast DNA variation in Louisiana irises and deduced a wealth of new details on a classical case of introgression. Molecular details also become important for the ecologically
VI. FROM MARKER TO CHARACTER

In most of the ecological applications cited above, molecular polymorphisms were used as anonymous markers for the structure and dynamics of populations. Implicit in some application of population genetic parameters is the assumption that the genes influencing ecologically relevant characters are as variable and share the dynamics of the markers. I have suggested above that this may not necessarily be the case. It is certainly advisable to complement the general population genetic information with an approach that begins with the identification of physiological and morphological characters relevant for the ecological status of the plants under investigation, and to determine the population genetic parameters for genes influencing these characters.

These genes can be found, mapped, isolated, analyzed and modified using molecular markers. The most general method will consist in treating the ecological character as a quantitative trait and using cosegregation between the trait and nuclear DNA markers in a segregating hybrid progeny to find, map and isolate the relevant genes as ‘quantitative trait loci’ (QTLs) (Paterson et al., 1988, Doebley et al., 1990). Among ecologically important traits that have been mapped are: time to first flower in soybean (Keim et al., 1990); winter hardiness in barley and other plants (Hayes et al., 1993; see also Thomas, 1990); resistance to low-phosphorus stress in maize (Reiter et al., 1991); and resistance to fungal infections in barley (Heun, 1992). Any heritable variation in a hybrid can be analyzed this way.

Eventually, this may be the most profound contribution of molecular methods to ecology. At the moment, most such applications are in agriculture, but their application in the ecology of wild populations, via wild relatives of domesticated plants, is just a question of time.

Each step in this analysis is aided by molecular markers. The character is marked ever more closely until molecular polymorphisms are found that are in the relevant genes, and these genes can be isolated using the marker sequence as a starting point.

The initial step consists in relating an ecologically significant character distribution with that of appropriate markers at a taxonomic level where hybridization for an eventual genetic analysis is possible. Some taxa with specific ecological tolerances have already been found in molecular taxonomic studies. ITS sequences from the ribosomal RNA genes indicate that the serpentine-endemic species in the n = 9 lineage of *Madia* (Asteraceae) form a sister group to the volcanic ash endemic species (Baldwin, 1993b). I have mentioned above the differentiation between the weedy and serpentine races of the wild sunflower, *Helianthus bolanderi* with RFLPs in the cpDNA and nuclear rDNA (Rieseberg et al., 1988). The genus *Oonopsis* (Asteraceae) is related to *Haplopappus* and contains six taxa that are all herbaceous selenium-accumulating perennials found on the highplains of eastern Colorado and central and eastern Wyoming. Chloroplast DNA restriction site data (Morgan & Simpson, 1992) support the monophyly of the plants which share that adaptation.

Of course, there are genetic and molecular analyses of ecologically adaptive syndromes that precede the days of molecular markers, and these indicate the kinds of genetic changes we might expect to find. One of these is the evolution of plants with *C*₄ photosynthesis from *C*₃ ancestors. This has happened several times independently. From a comparison of the two types of plants, the essential step could be a complex alteration in the control of enzyme genes present in the *C*₃ ancestor: an alteration of the kinetic and regulatory properties, an increase in the expression levels of least an order of magnitude, and the confinement of the expression to mesophyll or bundle-sheath cells of the leaf (Westhoff et al., 1993). The changes can be studied in detail by comparing species in a genus such as *Flaveria* (Asteraceae), which contains *C*₃, *C*₄ and *C*₃–*C*₄ intermediate plants (Ku et al., 1991).

Another classical case of an ecological adaptation that has been studied from the ecological to the molecular level is cyanogenesis in plants such as white clover, *Trifolium repens* (Hughes, 1991; Kakes, 1990). Breeding genetics had shown that cyanogenesis in white clover depends on active alleles at two independently segregating loci. Recently, a specific molecular probe for one of these, the gene for the hydrolytic enzyme linamarase has been
isolated (Hughes et al. 1990). Ecologically, cyanogenesis seems to be primarily a defence against herbivory.

Plant–animal interactions are a fertile field for the applications of molecular markers. One recent example concerns the hybrid swarm of cottonwoods in the Weber River drainage which I have mentioned above. In each spring, these trees are attacked by colonizing stem mothers of the leaf galling aphid, *Pemphigus betae* (Whitham, 1989). Successful colonization results in the formation of a leaf gall, wherein up to 300 progeny are produced parthenogenetically (Whitham, 1978). Colonization success varies from 0% to 80% among these trees and the resistance of the trees has a genetic basis (Whitham, 1989). Pure *Populus fremontii* is resistant, and less than 3% of the aphids survived on the F1 hybrid and first backcross plants. Second to fourth backcross trees and pure *P. angustifolia* ranged from highly resistant to highly susceptible with no significant differences among these genotype classes (Paige & Capman, 1993). Leaf morphology differs considerably between the parental species, but a leaf shape score for the hybrids is not a sufficiently accurate predictor of the genotype and the resistance relationships based on leaf shape had to be reassessed on view of the molecular data (Paige et al., 1990; Paige & Capman, 1993).

More such cases are going to accumulate rapidly and the further refinement of the genotype analysis up to finding the responsible genes presents no insurmountable difficulties, even though the intensive molecular work necessary may require cooperation between ecologists and molecular biologists.

**VII. OUTLOOK**

My assessment of the potential role of molecular markers in plant ecology emphasizes more accurate, more detailed or easier ways of following established leads. Some of these refinements will effectively constitute qualitatively novel approaches. There are indications that, in the course of this work, some concepts that until now have been marginal and suggestive may become important and central. Among these will be unusual ways in which the genome can influence the phenotype, and the possibility that some environmental effects may act directly on the genome in a more specific manner than mutagens. Some quantitative characters such as leaf shape seem than mutagens. In our experience with *Microseris*, the possibility that some environmental effects may act directly on the genome in a more specific manner than mutagens. Some of these refinements will effectively constitute qualitatively novel approaches. There are indications that, in the course of this work, some concepts that until now have been marginal and suggestive may become important and central. Among these will be unusual ways in which the genome can influence the phenotype, and the possibility that some environmental effects may act directly on the genome in a more specific manner than mutagens. Some quantitative characters such as leaf shape seem

**REFERENCES**


Molecular markers in plant ecology


This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.