determined by use of different host plants in *Drosophila mojavensis* and *Drosophila arizonae*. *J. Chem. Ecol.* 23, 2803–2824
57 Etges, W.J. and Ahrens, M.A. (2001) Premating isolation is determined by larval rearing substrates in cactophilic *Drosophila mojavensis*. V. Deep geographic variation in epicuticular hydrocarbons among isolated populations. *Am. Nat.* 158, 585–598
58 McClintock, W.J. and Uetz, G.W. (1996) Female choice and pre-existing bias: visual cues during courtship in two *Schizocosa* wolf spiders (Araneae: Lycosidae). *Anim. Behav.* 52, 167–181
59 Scheffer, S.J. *et al.* (1996) Sexual selection, male morphology, and the efficacy of courtship signalling in two wolf spiders (Araneae: Lycosidae). *Behav. Ecol. Sociobiol.* 38, 17–23
60 Hebets, E.A. and Uetz, G.W. (1999) Female responses to isolated signals from multimodal male courtship displays in the wolf spider genus *Schizocosa* (Araneae: Lycosidae). *Anim. Behav.* 57, 865–872

Biodiversity assessment using markers for ecologically important traits

Peter H. van Tienderen, Anita A. de Haan, C. Gerard van der Linden and Ben Vosman

Most studies of genetic variation within species to date are based on random markers. However, how well this correlates with quantitative variation is contentious. Yet, functional, or 'ecotypic' variation in quantitative traits determines the ecological niche of a species, its future evolutionary potential, and, for livestock, crops and their wild relatives, their usefulness as a genetic resource for breeding. But nowadays we can also assess genetic diversity using markers directly targeted at specific genes or gene families. Such gene-targeted, multilocus profiles of markers can contribute to *ex-situ* management of genetic resources, ecological studies of diversity, and conservation of endangered species.

Published online: 26 September 2002

Peter H. van Tienderen*
Institute for Biodiversity and Ecosystem Dynamics, PO Box 94062, 1090 GB, University of Amsterdam, Amsterdam, the Netherlands.
*e-mail: tienderen@science.uva.nl

Anita A. de Haan
Netherlands Institute of Ecology, NIOO-CTO, PO Box 40, 6666 ZG Heteren, the Netherlands.

C. Gerard van der Linden
Ben Vosman
Plant Research International, PO Box 16, 6700 AA Wageningen, The Netherlands.

Biodiversity within species concerns the amount, distribution and adaptive value of variation within and among populations in their natural environment [1]. Many biodiversity studies within species to date have focused on random molecular markers, such as microsatellites [2,3]. The breeding structure of populations, population bottlenecks and the biogeographical history of a species are expected to affect all markers in similar ways. By contrast, variation in functional regions (expressed or regulatory sequences) might reflect the past influences of selection, which can be different for each gene, superimposed on the pattern of variation as a result of history, migration and drift. The characteristics that enable a species or ecotype to occupy a certain geographical range or niche might depend on a limited set of genes, so that variation in such traits might not be detected by their correlation with random markers. Reed and Frankham [4] concluded that variation in molecular markers was not indicative of the adaptive evolutionary potential or differentiation of populations with respect to quantitative traits, and suggested that measuring quantitative genetic variation should be done directly. By contrast, Merilä and Crnokrak found a significant correlation between diversity in quantitative traits and molecular markers [5] but, at the same time, diversity in quantitative traits was consistently higher, indicative of a role for local adaptation and natural selection. Thus, studies of genetic diversity could benefit from targeting genes that exhibit ecologically relevant variation, rather than targeting random markers. Clearly, this is not a trivial exercise. One needs to assess which traits matter, identify the genes that potentially affect such traits, and develop markers within, or flanking the genes. For crop plants, the traits of interest are defined by the targets of the breeders. However, worldwide, 780 000 and 480 000 accessions of wheat and barley have been collected, respectively [6]. Genetic profiling of the accessions is essential, as to determine which have the most potential for use in breeding programmes. Furthermore, it is too expensive to maintain all accessions indefinitely *ex situ* in gene banks.

Here, we review the potential of a gene-targeting approach for biodiversity studies within species. Marker systems for functional genes are now being developed, and existing sequence information is being used to develop markers that tag variation within the gene or in a flanking region. Although gene targeting appears to be technically feasible, more work is needed to increase our knowledge of candidate genes. Finally, we compare the merits of gene targeting with alternative approaches using random markers, gene expression profiling, and direct measurements of functional variation.

SSAPs and SNPs for diversity assessment

In the European Union biotechnology programme 'molecular tools for screening biodiversity' [7], different approaches are being evaluated for the development of markers within and flanking genes in plants and animals. The markers do not necessarily carry the mutations that cause the phenotypic effect. They are putative tags for functional variation at a nearby position within the targeted genes. Two types of strategy are being developed: (1) the use of conserved sequence motifs as anchors for sequence-specific amplification polymorphisms (SSAP: see Glossary) [8]; and (2) the selection of genes involved in key processes and sequencing of several genotypes to detect single nucleotide polymorphisms (SNPs).

 0169-5347/02/$ – see front matter PII: S0169-5347(02)02624-1

Box 1. Domain-directed profiling: a gene-targeted approach for functional markers

The gene-targeted approach makes use of conserved domains in gene families to target the PCR towards amplification of members of these families. Detected polymorphisms can be length polymorphisms as well as the absence and/or presence of bands caused by point mutations or larger deletions and/or insertions in or near to genes of interest.

In the first step, genomic DNA of several varieties is cut with a frequently cutting restriction enzyme (e.g. Rsa 1). This creates a pool of DNA fragments (two arbitrary genotypes are depicted in Fig. Ia). An adapter is then ligated to the ends of these fragments. In the following PCR reaction, a degenerate primer is used that selectively binds to domain-containing fragments, in combination with an adapter primer (Fig. Ib).

Most fragments that are amplified in this step originate from genes harbouring the targeted domain; therefore, polymorphisms in the banding pattern (Fig. Ic) are most likely to be associated with the function of the conserved motif. The DNA fragments are labelled and separated by electrophoresis on polyacrylamide gels, resulting in a banding pattern in which polymorphisms are easily detected. Figure Id shows such a profiling pattern, targeted to the nucleotide-binding site that is present in many resistance genes in plants. Varieties 1–4 (Fig. Id) were examined in duplicate (duplicates are run in neighbouring lanes). Over 75% of the polymorphic fragments (indicated by arrows) show significant homology with known resistance genes or resistance gene analogues.

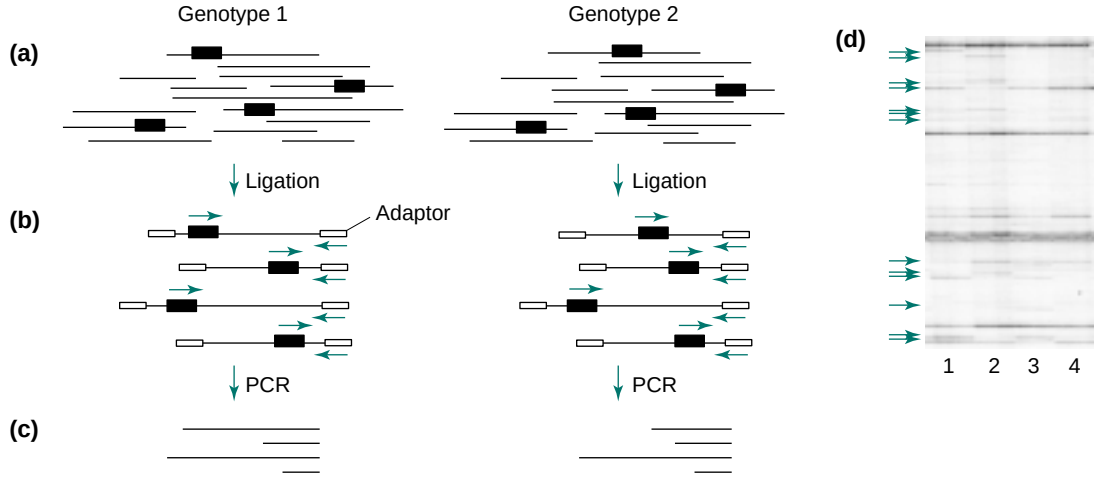

Fig. I

Sequence-specific amplification polymorphism

In the SSAP procedure [8], genomic DNA is digested with a restriction enzyme and ADAPTERS are ligated to the fragments obtained, in a similar manner to the production of amplification fragment length polymorphisms (AFLPs)® [9]. A PCR reaction is carried out using a primer that is based on the sequence of the adapter and a specific primer that is based on a CONSERVED MOTIF. Use of conserved motifs will thus direct the PCR reactions towards amplifying fragments comprising the conserved sequence together with flanking DNA. The resulting fragments are (radioactively) labelled and separated by gel electrophoresis, resulting in a multilocus DNA fingerprint (Box 1). This detects variation in fragment presence and length caused by the presence and/or absence of a restriction site near the target sequence. In the original protocol, the conserved sequence was derived from a RETROELEMENT [8], and the procedure has also been shown to work with microsatellite motifs [10] (SAMPL). Functional variation can be detected by using motifs known to affect specific traits; for instance, they could comprise coding sequences for a conserved protein domain.

An advantage of the SSAP procedure is that the DNA can be analyzed for specific functional regions in a relatively short time, without previous knowledge about specific loci and alleles. The amount of work involved is similar to standard AFLP. The crucial step is the design of the conserved (gene-targeting) primer, which determines the specificity of the DNA profile. This primer is usually based on the consensus of motif sequences of gene family members present in the nucleotide data base. Depending on the level of conservation, this primer will be partly degenerate, with either INOSINES or mixed nucleotides at less conserved positions, and should be optimal for annealing to as many target gene family members as possible, whilst still being specific and repeatable. The system delivers dominant markers, and it is usually unknown whether different fragments are allelic or come from different loci.

Single nucleotide polymorphisms

An SNP is a polymorphic site for which the variants differ by a single nucleotide substitution. SNPs can be found by comparing the sequences of target fragments from a (limited) set of genotypes (Box 2). SNPs are abundant in many organisms and genomic regions, although their frequency decreases with the relatedness of the genotypes compared. In *Arabidopsis thaliana*, for instance, 25 274 SNPs were identified between the Landsberg and Columbia strains [11]. The abundance compensates for the fact that SNPs only have two alleles (the probability that a third nucleotide is present at exactly the same position is negligible).

Box 2. Detecting and transporting SNPs in phytochromes

Single nucleotide polymorphisms (SNPs) can be found by sequencing homologous fragments (often a specific gene) obtained from a limited number of genotypes or accessions. Sequences are aligned so that positions can be identified with a variable nucleotide present, and the resulting data can be submitted to a GENBANK data base (http://www.ncbi.nlm.nih.gov/SNP/). Once detected, SNPs can be transformed into an efficient scoring system for genetic variation, for example, by incorporating the SNP in a primer of a PCR, or high-throughput screening using microarrays.

Phytochrome SNPs were detected by Kuittinen *et al.* [a]. Phytochromes are photopigments comprising a chromophore attached to an apoprotein at a conserved Cys residue. In *Arabidopsis thaliana*, five genes encoding the apoprotein parts of PHYA–PHYE, respectively, have been identified [b]. Unique primer combinations were designed for each phytochrome gene using the published sequences of *A. thaliana* [e.g. the PHYE sequence (GENBANK acc: X76610) was used to design primers that amplified a 712-bp fragment, including the largest intron, in *A. thaliana* (Fig. Ia: schematic representation of the PHYE gene, containing three introns, of *A. thaliana*, and the primers used to amplify fragments from the *A. thaliana* genotype as well as related species)]. Amplification was tested in *A. thaliana* (Lanes 1–3, Fig. Ib), *A. lyrata* (Lanes 4,6, Fig. Ib), *A. halleri* (Lanes 7–9, Fig. Ib), *Brassica oleracea* (Lanes 10,11, Fig. Ib) and, as an unrelated species, *Beta vulgaris* (Lane 12, Fig. Ib). Amplified products were sequenced, and SNPs were scored from the aligned sequences. Figure Ic gives the number of SNPs identified at synonymous or nonsynonymous sites (number of aligned base pairs in parenthesis): Values in the diagonal cells are number of SNPs in PHYE fragment among two genotypes within the species; values in the nondiagonal cells are number of SNPs differentiating the species (Fig. Ic). For instance, four synonymous and zero nonsynonymous sites were detected among two *A. thaliana* genotypes, and 16 synonymous and four nonsynonymous sites were detected between an *A. thaliana* and *A. lyrata* genotype.

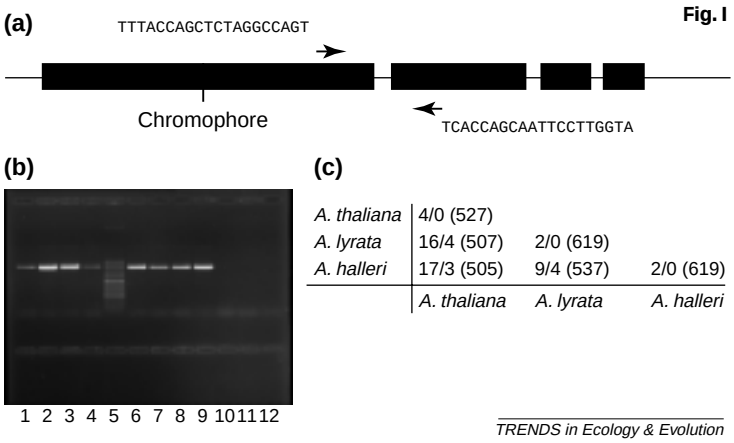

Fig. I

References

a Kuittinen, H, *et al.* (2002) Primers for 22 candidate genes for ecological adaptations in Brassicaceae. *Mol. Ecol. Notes* 2, 258–262

b Quail, P. (1997) An emerging molecular map of the phytochromes. *Plant Cell Environ.* 20, 657–665

SNPs are mendelian, co-dominant markers, that can be analyzed by any statistical method that is based on genotype frequencies. They can also be easily mapped in segregating populations [12].

The fragments sequenced can contain coding or noncoding DNA. The chances of finding SNPs are usually highest in noncoding and intergenic regions of the genome, because these are expected to be under less stringent selection than coding regions. Once polymorphisms in or around a target gene have been identified, efficient detection systems can be developed that enable screening of large numbers of genotypes. SNPs can be detected using specific primers in PCR reactions. In this case, the SNP is incorporated as the 3′ nucleotide in the primer. Several systems now offer high-throughput analysis of SNPs, based on newly developed technologies, including MALDI-TOF MASS SPECTROMETRY, use of new DNA-modifying enzymes, and specific forms of (mini-)sequencing. MULTIPLEX ANALYSIS for multiple variants in a single reaction is also possible, using DNA microarray technology (see *Nature Genetics* January 1999 supplement for an overview). Although expectations for high-throughput and multiplex analyses are high [13], large-scale use in a microarray format is still hampered by the lack of known SNPs in species other than model organisms, such as *Arabidopsis*.

Potential and limits of using functional markers

The success of a gene-targeting approach depends on four factors: (1) existing knowledge of the candidate genes for the traits of interest; (2) associations between the traits and markers; (3) transportability of the methods across species boundaries; and (4) development of statistical tools to handle the data.

Selection of candidates

Traits of interest include disease resistance, stress tolerance, growth rate, morphology and life-history traits, which are often quantitative traits that are likely to be affected by many genes. Target sequences for SSAP approaches can be selected from a gene family known to affect such traits. The first studies included functional domains of genes that are involved in disease resistance and plant development (NBS and MADS box domains, respectively, two families of transcription factors) [14]. The number of genes containing such domains can be high; for instance, 82 different MADS box-containing motifs were found in the *A. thaliana* genome [15], as well as 150 NBS-containing resistance gene analogues [11].

From quantitative trait loci (QTL) mapping it is clear, however, that our knowledge of candidate genes is incomplete. For instance, none of the four QTLs detected for obesity in mice mapped to the positions of the five candidate genes identified *a priori* [16]. By contrast, many of the QTLs for disease resistance in wheat mapped onto locations of some of the 50 known candidate loci for disease resistance [17]. More work is needed to identify the best targets for diversity assessment. This will involve mapping and association studies, but also the discovery of new candidate genes in model species by functional genomic approaches and mutagenesis screens.

Linkage between marker and traits

The main difference between a random marker and a marker for a functional gene is the distance to the mutation causing the phenotypic effect in the trait of interest. Proximity is very important given the expected rate of decline of LINKAGE DISEQUILIBRIUM (D) per generation (t) owing to recombination (r) [i.e. $D_t = (1-r)^t D_o$ in a random-mating population [18]], so that in evolutionary time all associations except those with the

Box 3. Detecting markers that tag genes under selection

A marker tagging a gene under selection is expected to show a pattern of variation that is different from that of a neutral marker. One way to test this is by comparing the divergence among populations using Wright's F-statistics, separately for neutral markers and gene-targeted markers (Table I). F_{st}-r is an among-population divergence F_{st} estimate that is obtained from random markers. F_{st}-t is an F_{st} estimate for a marker tagging a gene with a putative ecological function. Merilä and Crnokrak [a] published a similar table, comparing F_{st} values for neutral genetic markers to Q_{st} values for quantitative traits.

Table I. Relationship between F_{st} values for gene-targeted and neutral markers

F_{st}-t > F_{st}-r:	Population divergence is higher for the gene-targeted marker: indicates divergent selection and local adaptation for the tagged gene
F_{st}-t = F_{st}-r:	Population divergence is similar for both categories: no indication that selection plays any role
F_{st}-t < F_{st}-r:	Population divergence is lower for the gene-targeted marker: the marker therefore tags a gene with a function that is required in all populations

Markers in the first class (F_{st}-t > F_{st}-r) are the most valuable markers for conservation purposes as well as for a genetic resource for breeding programmes. Markers in the second category (F_{st}-t = F_{st}-r) just follow the pattern of divergence because of migration and drift. Markers in the third category (F_{st}-t < F_{st}-r), although presumably under selection, require no special attention, because they will be maintained by any conservation strategy. They could, for instance, tag important house-keeping genes.

Reference

a Merilä, J. and Crnokrak, P. (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* 14, 892–903

nearest, physically linked markers will disappear rapidly. However, spurious associations might result from HITCHHIKING, FOUNDER EFFECTS, or population structure with restricted gene flow. Even for markers inside or very close to an ancient, mutated gene the linkage might have been broken by recombination (e.g. the association between markers in and around the gene encoding lipoprotein lipase, a putative factor contributing to coronary heart disease in humans [19]). The association between a marker and the phenotypic expression tagged by that marker will depend on the age of the mutation causing the effect, the frequency of recombination and presence of recombinational hotspots, the breeding and population structure, the extent of population expansion and the genealogical history of the genomic region. Nordborg *et al.* [20] showed that even in the predominant selfer *A. thaliana*, linkage disequilibrium decayed within 1 cM, or 250 kb, indicating that a dense map of random markers would be required to pick up an association with a focal trait.

Transportability

Sequence information and knowledge of candidate genes is currently available for only a few species. SSAP profiling and SNP development are both transferable from model species to related species, because they use conserved sequences for primer design. In the EU-MOTORS project (MOlecular TOols for Related Species) [21], markers for 22 different genes were developed, using primers designed from sequences in the *A. thaliana* data base. Markers were validated in five species, with 2–4 genotypes per species. Primer combinations were designed that worked well in the relatives of *A. thaliana* (*A. lyrata* and *A. halleri*), and sometimes in *Brassica oleracea*, with adjustments in PCR conditions. As expected, only a few primer pairs worked well in the unrelated species, *Beta vulgaris*. Amplification products were sequenced to confirm the homology of the genes. In *B. oleracea*, sequence results were sometimes difficult to interpret, probably because of gene duplication in this species. It should also be feasible to design universal primers that work for many dicots, as shown recently for a phytochrome B (PHYB) exon [22].

For the SSAP approach, the primer sequences used to tag NBS- and MADS-containing genes in tomato are also effectively transferable to the closely related potato and pepper, and also to more distant species, such as barley and lettuce (C.G. van der Linden *et al.*, unpublished).

Statistical methods

Markers in specific functional regions of the genome call for statistical analyses that consider the possibility that the regions might have experienced different selective regimes. For markers that are affected only by migration and drift, aggregated parameters for diversity, polymorphism, heterozygosity and population subdivision can be calculated combining the data for the different markers. New statistical approaches need to be developed to deal with multilocus profiles for large sets of genotypes using markers that are putative tags for functional variation, and that might not be neutral. Research can be directed to find associations between markers and expressed traits by experimental testing of a subset of all genotypes. The latter can be done in a segregating population derived from a cross (e.g. by QTL mapping), or by association studies of markers and trait values at the level of populations or genotypes. This will be a massive undertaking, but with high payoffs in our understanding of the link between genetic and phenotypic variation.

Another route is to filter out random patterns, so that only the selectively interesting patterns remain. Filtering could make use of different kinds of data. First, one could use CANONICAL CORRESPONDENCE ANALYSIS to relate genetic marker data to parameters associated with sampling site. Such parameters could include potential selective factors, but also geographical or climatic data. It might prove difficult to distinguish between geography and climate, because geographical location will often be associated with differences in environmental conditions, leading to spurious correlations. Another approach would be to find markers for which the pattern of variation is very different from the pattern of variation observed or expected for neutral DNA. This could be an indication of selection at the tagged DNA region. One could compare multilocus profiles of genotypes based on targeted genes with profiles based on random fragments (e.g. AFLPs), and test whether the divergence among populations is

Box 4. Comparison of different ways to assess genetic diversity within species

Genetic diversity within species can be assessed using variation in neutral markers (NM), gene-targeted markers (GT), gene expression (expressed sequence tags, EST) or in quantitative measurements in experiments under controlled conditions (quantitative traits, QT). Each method has its advantages and disadvantages (Table I).

Table I. Comparison of characteristics of different methods for assessing genetic variation within species

Advantages/Disadvantages	NM[a]	GT	EST	QT
Required investment in marker development	Low	Low/medium	High	n/a
High-throughput screening	Yes	Yes	No	No
Genome coverage	Medium	Low/medium	High	High
Statistics well developed	Yes	No	No	Yes
Information on history, drift	Yes	Limited	No	No
Information on selection, functional variation	Limited	Yes	Yes	Yes
Experiments must be under controlled conditions	No	No	Yes	Yes
Feasible for most organisms	Yes	Yes	No	No

[a]Abbreviations: EST, expressed sequence tags; GT, gene-targeted markers; NM, neutral markers; QT, quantitative traits.

The initial investment in marker development is generally higher for the gene-targeted approach [at least for single nucleotide polymorphisms (SNPs)] than for neutral markers, although some neutral markers, such as microsatellites, also need to be developed specifically for the species studied. EST profiling requires a cDNA library on a microarray against which the fragments from the different genotypes can be hybridized. These microarrays are commercially available for some species, but need to be assembled *de novo* for most. Once developed, neutral as well as functional markers can often be converted to a (semi-) high-throughput system. However, as yet, gene expression microarrays cannot be used in a high-throughput setting. Although efficient screening of many cDNA fragments for a limited number of genotypes and/or accessions is possible, diversity studies often require testing large samples of individuals under many different conditions, which is not possible with the technology that is currently available

Genome coverage indicates whether a large part of the relevant variation is covered. For random markers, this is debatable. With respect to GTs, SNP markers rely on the availability of sequence data of candidate genes. As long as our knowledge of candidate genes is still limited, genome coverage will be rather low. For the targeted approach using conserved domains genome coverage is better, but it is still limited by the choice of gene family/conserved region and the participation of other genes/gene families in specifying the trait of interest. For ESTs, the coverage can be very high but it is dependent on the specific conditions under which the organisms are grown. The same holds for QTs: genotypes that differ in phenotype under the tested conditions will be picked up, regardless of how many genes affect the phenotype or where the genes are located. However, it could be that different genotypes have the same phenotype but with different combinations of genes.

Statistics are well developed for neutral markers and quantitative traits, and are under development for gene-targeted markers and EST profiles.

Neutral markers give the best information on demographic processes such as the evolutionary history, population structure, and drift, especially if one can combine data from nuclear markers and maternally inherited cytoplasmic markers. Yet gene-targeted markers may be more indicative for variation at loci that determine the ecological and functional distribution of the species.

ESTs and QTs require extensive experimentation to acquire the expression profiles (ESTs) and quantitative measurements on the performance of species (QT) under ecologically relevant conditions. Clearly, this is restricted to species that are amenable for experimentation and have a relatively short life span. Moreover, deciding what the relevant conditions are might not be easy.

different from random markers, for instance using Wright's F-statistics (Box 3). Most of these options are largely unexplored.

Alternatives to gene targeting

The use of variation in and around functional genes has to be weighed against other available methods, each with their own potential and limitations (Box 4). Moreover, technological developments might further expand the range of possibilities. Large-scale sequencing of genes in many individuals will become faster and cheaper, enabling detailed statistical analyses of the genes studied [23]. The selection of genes to sequence is as important here as it is for SSAP or SNP analyses. Also, more efficient methods to screen gene expression are becoming available [24], and if one is interested in functional variation it seems logical to concentrate at the RNA, or protein, level [25]. However, there are still many problems (Box 4), such as obtaining expression data under conditions that are relevant and representative for the species in its native habitat.

RNA expression can also be used as a starting point to develop DNA markers; for example, finding SNPs by sequencing expressed sequence tags from different genotypes to obtain gene-targeted markers would be an alternative to the approaches explained above. Given the variation in expression levels and difficulties arising from RNA processing and the presence of introns, this could be as difficult as developing DNA markers directly.

Future developments

Assessment of genetic diversity using markers for specific genomic targets is technologically feasible, although our knowledge of potential candidates is still very incomplete. SSAP profiling can be readily applied to any species sharing the conserved motif. For large-scale profiling of gene bank accessions, we expect this to become an attractive strategy. The current rate of growth in sequence information will facilitate SNP detection for more homologous genes, although SNPs are not shared among species and need to be developed for each species.

Whether markers for functional variation are useful will depend primarily on the goals of the diversity study. For the reconstruction of historical processes, neutral variation is suitable. For ecological genetic studies, or for the assessment of variation in wild relatives of crop plants or livestock, it could be valuable to have information about variation in specific genes, such as those affecting resistance, or growth and reproduction under stressful condition. For the protection of endangered species, however, it seems risky to replace random markers that cover a large part of the genome with markers targeted in a small subset of genes. Furthermore, if the threat comes from genome-wide accumulation of deleterious genes and/or inbreeding depression, rather than a loss of particular (ecotypic) variation, assessment of diversity by random markers could even be preferable.

Moritz [26] defined MANAGEMENT UNITS based on EVOLUTIONARILY SIGNIFICANT UNITS (ESUs) for the conservation of genetic variation within species, the latter aiming to encompass the historic, evolutionary diversity of a taxon rather than the current distribution of populations and alleles. Likewise,

Glossary

Adapter: short double-stranded DNA fragment of known sequence that is attached to the ends of unknown DNA fragments to facilitate their amplification and/or cloning.
Canonical correspondence analysis: multivariate ordination technique that investigates variation in a set of quantitative data by linking it to a second set of external, explanatory variables. For instance, genetic (marker) data could be linked to environmental conditions at the site of origin to detect which genes could be under selection.
Conserved motif: DNA sequence that is (nearly) identical between similar genes found in different species or between related genes within one species. Conserved motifs often represent a specific conserved function, such as the MADS box family of transcription factors.
Evolutionarily significant unit: a collection of individuals of a given species with a unique evolutionary history that makes them worthwhile to protect. The uniqueness is often inferred from phylogenies based on molecular genetic data.
Founder effects: characteristics of a population that can be traced back to the origin of the genotypes that founded it.
Functionally significant unit: a population or group of populations of a given species that harbour adaptations that are not found elsewhere and allow them to function in their habitat.
Hitchhiking: a gene linked to a gene under selection might show a similar pattern of variation, because of a shared history and/or close linkage.
Inosine: a (rare) purine base that does not form hydrogen bonds with other DNA bases. Inosines act as neutral bases in annealing of DNA strands, preventing mismatches between sequences that are not 100 per cent identical.
Linkage disequilibrium (LD): also called gametic disequilibrium. Difference between the observed and expected frequencies of combinations of alleles at different loci in gametes. This is somewhat confusing, because genes need not be linked to be in gametic disequilibrium.
MALDI-TOF mass spectrometry: Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry is a technique that allows the identification (size determination) of (large) biomolecules.
Management unit: area containing species or communities that are protected in order to preserve the biodiversity it contains.
Microsatellite: DNA tract consisting of short, tandemly repeated DNA sequences (i.e. GAGAGAGAGA or TCATCATCATCATCA. The number of repeats is often variable between individuals of a species.
Multiplex analysis: analysis of multiple markers in a single reaction or assay.
Retroelement: a transposable element that transposes in the genomic DNA by its ability to synthesize DNA from an RNA template using reverse transcription.
SAMPL: Selective Amplification of Microsatellite Polymorphic Loci is an AFLP based molecular marker technique in which the AFLP reaction is anchored towards microsatellite DNA.
SSAP: Sequence-Specific Amplified Polymorphism. A PCR-based molecular marker technique in which the PCR products are anchored to the Long Terminal Repeats of retro-elements.

we could define FUNCTIONALLY SIGNIFICANT UNITS (FSUs) based on differences in allelic frequencies for genes with important ecological functions. In fact, this would be more in line with the original definition of ESUs [27], as units containing significant adaptive variation [1]. Management practices in conservation could be primarily directed at maintaining different FSUs rather than the genetic diversity at large. Indeed, Hedrick and Parker assessed major histocompatibility complex diversity in natural populations of endangered *Poeciliopsis* guppies [28]. Populations were very divergent, and the population that had experienced most bottlenecks was monomorphic. They suggest using this information on this gene in management strategies to protect the species, thus giving weight to the presence of ecologically relevant genetic variation.

Developments in molecular genetics and bioinformatics will provide us with better tools to identify candidate genes and boost the use of the approaches that we have discussed here. Genetic profiling will enable breeders to select accessions quickly for improvement of existing varieties. Ecologists might be able to identify the genes that determine a species' distribution, and the selective pressures during the evolutionary history of plants and animals. Finally, it might help managers in conservation biology to identify those FSUs (or ESUs) that contain genetic variation that is worthwhile protecting.

Acknowledgements
We thank Peter van Dijk, Ineke Booy, Tracey Parrish, Outi Savolainen, Montserrat Aguadé, René Smulders and the reviewers for comments and suggestions. This work was partially supported by EU Grants BIO4–98–0299 and BIO4–98–0332.

References

1 Ryder, O.A. (1986) Species conservation and systematics: the dilemma of subspecies. *Trends Ecol. Evol.* 1, 9–10
2 Luikart, G. and England, P.R. (1999) Statistical analysis of microsatellite DNA data. *Trends Ecol. Evol.* 14, 253–256
3 Sunnock, P. (2000) Efficient genetic markers for population biology. *Trends Ecol. Evol.* 12, 199–203
4 Reed, D.H. and Frankham, R. (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* 55, 1095–1103
5 Merilä, J. and Crnokrak, P. (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* 14, 892–903
6 FAO (1998) *The State of the World's Plant Genetic Resources for Food and Agriculture*, FAO
7 Karp, A. *et al.* (1998) *Molecular Tools for Screening Biodiversity*, Kluwer Academic Press
8 Waugh, R. *et al.* (1997) Genetic distribution of BARE-1 retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol. Gen. Genet.* 253, 687–694
9 Vos, P. *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414
10 Witsenboer, H. *et al.* (1997) Identification, genetic localization, and allelic diversity of selectively amplified microsatellite polymorphic loci in lettuce and wild relatives (*Lactuca* spp.). *Genome* 40, 923–936
11 The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
12 Cho, R.J. *et al.* (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.* 23, 203–207
13 Lemieux, B. *et al.* (1998) Overview of DNA chip technology. *Mol. Breed.* 4, 277–289
14 Hayes, A.J. and Saghai Maroof, M.A. (2000) Targeted resistance gene mapping in soybean using modified AFLPs. *Theor. Appl. Genet.* 100, 1279–1283
15 Riechmann, J.L. *et al.* (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105–2110
16 Keightley, P.D. *et al.* (1998) Test of candidate gene quantitative trait locus association applied to fatness in mice. *Heredity* 81, 630–637
17 Faris, J.D. *et al.* (1999) Candidate gene analysis of quantitative disease resistance in wheat. *Theor. Appl. Genet.* 98, 219–225
18 Falconer, D.S. and MacKay, T.F.C. (1996) *Introduction to Quantitative Genetics* (4th edn), Addison-Wesley
19 Templeton, A.R. *et al.* (2000) Recombinational and mutational hotspots within the human lipoprotein lipase gene. *Am. J. Hum. Genet.* 66, 69–83
20 Nordborg, M. *et al.* (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 30, 190–193
21 KuittInen, H. *et al.* (2002) Primers for 22 candidate genes for ecological adaptations in Brassicaceae. *Mol. Ecol. Notes* 2, 258–262
22 Simmons, M.P. *et al.* (2001) Phylogeny of the Celastraceae inferred from phytochrome B gene sequence and morphology. *Am. J. Bot.* 88, 313–325
23 Rozas, J. and Rozas, R. (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15, 174–175
24 Marshall, A. and Hodgson, J. (1998) DNA chips: An array of possibilities. *Nat. Biotechnol.* 16, 27–31
25 Kozian, D.H. and Kirschbaum, B.J. (1999) Comparative gene-expression analysis. *Trends Biotechnol.* 17, 73–78
26 Moritz, C. (1994) Defining evolutionarily significant units for conservation. *Trends Ecol. Evol.* 9, 373–375
27 Crandall, K.A. *et al.* (2000) Considering evolutionary processes in conservation biology. *Trends Ecol. Evol.* 15, 290–295
28 Hedrick, P.W. and Parker, K.M. (1998) MHC variation in the endangered Gila topminnow. *Evolution* 52, 194–199