OPINION

The evolution of molecular markers — just a matter of fashion?

Christian Schlötterer

In less than half a century, molecular markers have totally changed our view of nature, and in the process they have evolved themselves. However, all of the molecular methods developed over the years to detect variation do so in one of only three conceptually different classes of marker: protein variants (allozymes), DNA sequence polymorphism and DNA repeat variation. The latest techniques promise to provide cheap, high-throughput methods for genotyping existing markers, but might other traditional approaches offer better value for some applications?

Being able to distinguish between genotypes that are relevant to a trait of interest is a key goal in genetics. Often, this distinction is not based directly on the trait of interest, but on informative marker systems. A genetic marker provides information about allelic variation at a given locus. The first genetic map of *Drosophila melanogaster* was built by Sturtevant using phenotypic markers[1]. However, since these early attempts, the advent of molecular biology has allowed a repertoire of genetic markers to be developed. These molecular markers have been applied to many biological questions, ranging from gene mapping to population genetics, PHYLOGENETIC RECONSTRUCTION, paternity testing and forensic applications.

Over the years, advances in molecular biology have led to the introduction of many new types of molecular marker. The general view is that the rise and fall in popularity of these different marker types (FIG. 1) reflects the continuous improvement in the way in which we assay genetic variation; that is, the latest marker systems are the most informative ones. Nevertheless, in reviewing the history of molecular markers and their pros and cons, I argue that there are only a few conceptually different classes of marker and that recently developed high-throughput methods might not be unconditionally superior to more traditional approaches.

Allozymes

The first true molecular markers to be established were allozymes (a term that originates from a contraction of the phrase 'allelic variants of enzymes'). The principle of allozyme markers is that protein variants in enzymes can be distinguished by native gel electrophoresis according to differences in size and charge caused by amino-acid substitutions. To visualize the allozyme bands, the electrophoretic gels are treated with enzyme-specific stains that contain substrate for the enzyme, cofactors and an oxidized salt (for example, nitro-blue tetrazolium). Early studies that used simple starch gel electrophoresis in humans and *Drosophila* showed substantial polymorphism in natural populations of these species[2–5]. This surprisingly high amount of within-population polymorphism directly led to the neutral theory of molecular evolution, which states that most mutations are effectively neutral[6,7]. This important conceptual advance is probably the greatest legacy of this marker system. Since then, allozyme markers have been successfully used in a wide range of species. Owing to their cost effectiveness, many populations with large sample sizes are typically studied in allozyme surveys. Nevertheless, the number of informative marker loci is too small to use allozymes for mapping and ASSOCIATION STUDIES[8]. Furthermore, surveys of natural variation based on allozymes were often challenged by non-neutral evolution of some of the markers used (see, for example, REFS 9–11).

The arrival of DNA-based markers

One of the criticisms levelled at allozyme markers is that they are an indirect and insensitive method of detecting variation in DNA. A more direct molecular marker would survey DNA variation itself, rather than rely on variations in the electrophoretic mobility of proteins that the DNA encodes. Another important advantage that DNA-based markers have over allozymes is that they allow the number of mutations between different alleles to be quantified. Given these unambiguous advantages, the arrival of DNA manipulation techniques promoted a shift from enzyme-based to DNA-based markers.

"…the arrival of DNA manipulation techniques promoted a shift from enzyme-based to DNA-based markers."

RFLPs. The discovery and isolation of restriction endonucleases in the 1960s by Arber, Smith and Nathans set the grounds for a new class of genetic marker: RESTRICTION FRAGMENT LENGTH POLYMORPHISMS (RFLPs)[12]. These markers allowed DNA variation to be assayed because single base substitutions in the recognition sequence of a restriction enzyme changed the pattern of resulting restriction fragments. This technological improvement allowed, for the first time, the analysis of non-coding sequences or silent changes in a protein coding sequence, as well as the survey of changes

in coding sequences. The first DNA-based genetic maps and the first successful association study were based on RFLP markers[13]. Furthermore, RFLP analyses of mitochondrial DNA (mtDNA) and ribosomal DNA (rDNA) have been used for population genetics, biogeographical surveys and phylogenetics[14]. The greatest obstacle for an even wider range of applications of RFLP analysis was the requirement for a suitable hybridization probe to detect the polymorphism. So, despite an almost infinite number of available markers, technical reasons prevented their further exploitation.

Minisatellites. Similar to RFLP analysis, the first step of minisatellite analysis also involves digestion of genomic DNA with restriction enzymes. Nevertheless, they represent a conceptually different class of marker[15]. Minisatellites consist of tandem repeats that frequently show length polymorphism arising from unequal crossing over or GENE CONVERSION[16]. Hybridization of genomic DNA with a minisatellite core sequence produces a bar-code-like hybridization pattern. The extremely high polymorphism of minisatellites revolutionized the genetic identification of individuals — a technique that has been called DNA fingerprinting[17]. Despite their high polymorphism and the enormous success of their use in forensics and paternity testing, minisatellites were not widely used in population genetics and genome mapping. The complex banding pattern precluded the assignment of alleles to a given locus. Therefore, standard population genetic analyses could not be applied to minisatellite data. The use of minisatellites for mapping and association studies was limited by the non-random distribution of minisatellites on the genome. Later, single locus minisatellites were developed[18]. However, as this procedure is technically challenging and most markers require high-molecular DNA, many studies at present rely on other PCR-based markers.

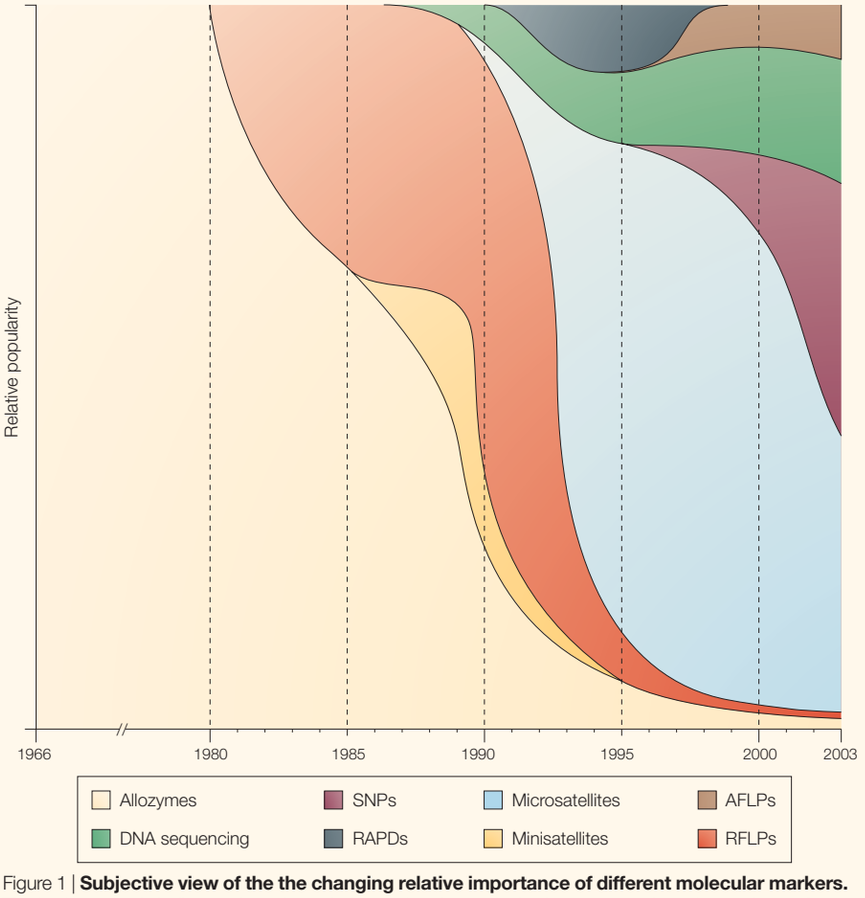

Figure 1 | **Subjective view of the the changing relative importance of different molecular markers.** The horizontal axis indicates time. At each time point, the vertical axis corresponds to the total use of molecular markers. If more than one molecular marker is used at a given time point, its relative importance is reflected by its proportion on the vertical axis. AFLP, amplified fragment length polymorphism; RAPD, randomly amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; SNP, single nucleotide polymorphism.

PCR-based markers

The main breakthrough of DNA-based molecular markers was driven by the invention of PCR[19]. For the first time, any genomic region could be amplified and analysed in many individuals without the requirement for cloning or isolating large amounts of ultrapure genomic DNA.

Microsatellites. The first widespread markers to take full advantage of PCR technology were microsatellites[20–22]. Similar to minisatellites, microsatellites are also tandemly repeated sequences, but their repeat motifs are shorter. With a typical repeat region that is smaller than 100 base pairs (bp), most microsatellite loci can be amplified by a standard PCR. Microsatellites are highly polymorphic, abundant and fairly evenly distributed throughout the EUCHROMATIC part of genomes. These properties have made microsatellites one of the most popular genetic markers for mapping, paternity testing and population genetics[23]. Microsatellites probably had the largest impact on constructing genetic maps in a wide range of species. In human genetics, the development of microsatellite technology was a key step towards the subsequent positional cloning of numerous monogenic human disease genes (see, for example, REFS 24–26). The high sensitivity of PCR also made microsatellites the method of choice for forensics, non-invasive sampling studies and the analysis of small social insects[27].

The high mutation rate of microsatellites allowed a more detailed analysis of their mutation pattern. Microsatellites gain and lose repeat units by DNA-replication slippage, a mutation mechanism that is specific to tandemly repeated sequences[28,29]. However, it quickly became clear that microsatellites have a complex mutation pattern, which creates difficulties for populations-genetic analyses. Furthermore, technical problems, such as PCR artefacts (STUTTER BANDS), complicate the automated scoring of microsatellite alleles. Finally, despite a high number of microsatellite loci in the genomes of most eukaryotes, the density of informative microsatellite loci could be too low for some mapping applications[30,31].

RAPDs, ISSRs, IRAPs and AFLPs. Another class of PCR-based markers relies on the use of PCR primers, which can bind to multiple sites in the genome. This can be achieved by using either short PCR primers (RANDOMLY AMPLIFIED POLYMORPHIC DNAS, RAPDs)[32], PCR primers that are complementary to repetitive elements such as microsatellites (INTER-SIMPLE-SEQUENCE-REPEATS, ISSRs)[33] or retrotransposons

(INTER-RETROTRANSPOSON AMPLIFIED POLYMORPHISMS, IRAPs)[34]. Alternatively, restriction fragments could be amplified by adding linkers and subsequent selective amplification (AMPLIFIED FRAGMENT LENGTH POLYMORPHISMS, AFLPs)[35,36]. In all cases, PCR amplification yields multiple bands that show a presence/absence variation among individuals. The principal advantage of RAPDs, ISSRs and AFLPs is that they do not require *a priori* knowledge about primer sequences in the target species. In particular, for species with large and/or poorly investigated genomes, AFLP analysis was successfully used to construct genetic maps (see, for example, REFS 37–41). Nevertheless, these markers can be unreliable. This is particularly true of RAPDs, which have been shown to be notoriously difficult to reproduce[42]. Consequently, some journals (see, for example, *Molecular Ecology* in the online links box) discourage diversity studies based on RAPDs. Whether or not the same will apply to AFLPs remains to be seen.

DNA-sequence polymorphism

The markers that have been discussed up to this point have used various strategies to survey aspects of variation in the genome. However, I argue that all marker systems discussed here assay variation in one of three types of genetic marker: protein variants (allozymes), DNA sequence polymorphism (RFLPs, AFLPs) and DNA repeat variation (minisatellites, microsatellites). So, if we look at marker systems in this way, the use of sequence polymorphisms as markers has been common ever since the first RFLP studies were done, and since then, many methods have been used to detect DNA sequence variation (BOX 1). So, the high-throughput methods to detect DNA sequence polymorphisms that have recently been developed are not a conceptual advance: they are merely a larger-scale way of doing things. The recent popularity of single nucleotide polymorphisms (SNPs) has led to an unfortunate heterogeneity in the use of the term SNPs. Rather than following the general trend to call all sequence variations SNPs, I differentiate between SNPs (genotyped using high-throughput techniques) and sequence variants assayed through DNA sequencing.

SNPs. The main advantage of SNPs is their high potential for an automated high-throughput analysis at moderate cost[43]. A classic SNP marker focuses on a single nucleotide position in the genome. Within a population, however, most nucleotide positions are invariant. So, *a priori* information about the presence of allelic variation at a given genomic position is required. Various strategies have been pursued to identify SNPs. The simplest approach is a screen of expressed sequence tags for polymorphic sites (see, for example, REF. 44). The most comprehensive way to identify SNPs throughout the genome is the generation of whole shotgun genome sequences[45], using a pool of individuals as donors for the genomic DNA to be sequenced. A more efficient strategy is reduced representation shotgun sequencing[46], in which only a reduced fraction of the genome (for example, separated by size fractionation after a restriction digest) can be sequenced for multiple individuals. All SNP isolation strategies result in a notable bias for various parameters, such as F_{ST}, allele frequency distribution and LINKAGE DISEQUILIBRIUM (LD)[47–49]. However, recent studies have shown that this ASCERTAINMENT BIAS can be accounted for in a range of conditions[47–49].

In addition to fine mapping, SNPs hold great promise for LD mapping of complex genetic traits. The most ambitious and highly-debated project is the SNP-based construction of a haplotype map of the human genome (HapMap). In addition to LD mapping studies, SNPs are also useful for the inference of past demographic events, such as population expansion or ADMIXTURE[50]. Apart from the ascertainment problem of SNPs, this genetic marker system also suffers from other shortcomings. First, as they are usually biallelic, the information content of a single SNP is limited, particularly if one of the two alleles occurs at a low frequency. Second, the development of a set of SNP markers is time- and cost-intensive. Third, SNPs might be located at hypermutable sites[51], which violates the assumption that they are biallelic — often made when analysing SNPs for population-genetics purposes[48]. In fact, often more than two alleles are observed for a single SNP (for example, >7% were observed for *Drosophila*[52]). However, it should be noted that several methods for estimating population parameters, such as migration and recombination, that take multiple hits into account[53–55], could also be applied to SNPs.

DNA sequencing. Sequence determination for a given genomic region in multiple individuals (often referred to as 'resequencing' in human population studies) provides the most fine-grained genetic information. Despite not being a marker in the narrow sense, DNA-sequence analysis should also be included in

Box 1 | Methods for the detection of DNA sequence variation

Many methods have been developed to assay variation in DNA sequences. DNA sequencing and high-throughput SNP genotyping are the most widely used of these methods but there are many others. Most of these methods rely on differences in the DNA sequences being reflected in differences in the chemical properties of the molecule. The most common way in which a change in these properties is detected is through a difference in electrophoretic mobility. Some of these other methods used to detect sequence variation are described below.

Denaturing gradient gel electrophoresis
This method is based on the principle that when a double-stranded PCR fragment migrates through a gradient of denaturing solvents, the conformation of the fragment changes to a partially denatured form[81,82]. This conformation change notably reduces the mobility of the fragment, resulting in a sequence-specific position in the gel. PCR fragments that differ in sequence have specific denaturation conditions that allow the discrimination between sequence variants.

Temperature gradient gel electrophoresis
This method is based on the same principle as denaturing gradient gel electrophoresis, but the denaturing gradient is achieved through a temperature gradient[83].

Single-strand conformation polymorphism
This technique relies on variation in electrophoretic mobility of secondary structures formed by single-stranded DNA fragments[84,85].

Heteroduplex analysis
This method assays genetic-variation-based differences in electrophoretic mobility between heteroduplex molecules (DNA molecules that are formed from strands from two different alleles) and homoduplex molecules[86].

PCR-RFLP, cleaved amplified polymorphic sequences
This technique is based on the principle that PCR fragments are digested with a restriction enzyme that is sized by gel electrophoresis. As for standard restriction fragment length polymorphisms (RFLPs), base-pair substitutions in the restriction site lead to changes in the patterns of restriction fragments.

this discussion: it has a long-standing history in population genetics and has been frequently used to infer selection and demographic patterns. The first sequencing studies pre-date the PCR era and were based on the *Adh* region that was cloned from 11 *D. melanogaster* individuals[56]. Since the arrival of PCR, the number of sequence surveys has continuously increased. Although this method was initially time-consuming and expensive, recent advances in sequencing technology permit sequence analysis of many DNA fragments for many individuals[51–53]. In comparison to SNPs, DNA sequencing offers the advantage of no ascertainment bias. Furthermore, DNA sequencing provides complete information from the analysed region. The trade-off is that invariant sites are also needlessly sequenced. For species with low polymorphism, sequencing a large fraction of invariant sites might seem uneconomic, but given appropriate analysis tools, the pattern of invariant sites could be informative[49]. One further advantage of DNA sequences is that the analytical framework for demographic analyses and neutrality tests is already well-developed[57].

Marker choice — facts or fashion?

Early studies using molecular markers were heavily influenced by the availability of methods and laboratory equipment. Today, however, the situation has changed markedly. For a large range of genetic model organisms, the full repertoire of genetic markers is available. For non-model organisms, however, markers still need to be developed. Both marker development and analysis can be easily out-sourced if the required equipment is not available in a particular laboratory, so the decision of which marker to use should primarily be made on the basis of objective grounds. The strengths and weaknesses of different markers are summarized in TABLE 1 and are discussed below.

Inference of demographic processes. Studies of natural populations have recently moved away from the purely descriptive analysis of amounts of variability and genetic differentiation to more sophisticated analyses that are used to attempt to infer the demographic past (see, for example, REFS 58–64). As these methods rely on the allele distribution, they need to account for new mutations and possible ascertainment biases. The high mutation rate of microsatellites could be used to infer past demographic events[58–61], but uncertainty about mutation rates and pattern greatly complicate the analysis[28,29]. SNP-based approaches heavily depend on an appropriate correction for the ascertainment bias[47–49]. Unfortunately, direct comparisons of both markers using the available repertoire of analytical approaches are still scarce, so, at present, it is not clear which marker is most useful for the inference of demographic processes. Nevertheless, the safest approach for estimates of demographic history is DNA sequencing of multiple genomic regions.

Paternity testing and forensics. Both forensic applications and paternity testing require informative and reliable markers. SNPs and microsatellites are two excellent choices. Although SNPs might be preferable for routine applications, owing to their higher automation potential, the use of microsatellites is advised for non-model-organism studies, as their isolation is easier and a smaller number of loci are required, owing to the higher information content of a single marker[65].

Association studies. Mapping studies require a large number of markers distributed over the genome. For experimental crosses, AFLPs, microsatellites and SNPs are good markers, and the choice of marker will largely depend on the scale of the study and availability of markers. LD mapping is a highly promising mapping strategy that searches in a population sample for an association between a disease-causing locus and a marker. Interestingly, in most instances in which LD has been used successfully, strong linkage and LD were detected using microsatellites. This is consistent with the decision of several researchers to switch from SNP-based LD mapping back to microsatellite-based mapping[66]. Also, theoretical studies indicate that microsatellites are much more powerful than SNPs, even if more SNPs are analysed[67].

Hitchhiking mapping. In the wake of the recent genome projects, it has become feasible to perform genome-wide screens for chromosomal regions that bear the signature of selection[68]. Most of these hitchhiking mapping studies compared microsatellite variability across the genome to identify putative targets of selection (see, for example, REFS 69–71). Two of these studies were able to verify the non-neutral evolution of the identified regions by DNA-sequence analysis[72,73]. Similarly, candidate regions were also identified by a recent SNP genome scan[74]. The analysis of

Table 1 | **Comparison of different molecular markers**

Marker	Advantages	Disadvantages
SNPs	• Low mutation rate • High abundance • Easy to type • New analytical approaches are being developed at present • Cross-study comparisons are easy; data repositories already exist	• Substantial rate heterogeneity among sites • Expensive to isolate • Ascertainment bias • Low information content of a single SNP
Microsatellites	• Highly informative (large number of alleles, high heterozygosity) • Low ascertainment bias • Easy to isolate	• High mutation rate • Complex mutation behaviour • Not abundant enough • Difficult to automate • Cross-study comparisons require special preparation
Allozymes	• Cheap • Universal protocols	• Requirement for fresh or frozen material • Some loci show protein instability • Limited number of available markers • Potentially direct target of selection
RAPDs and derivatives	• Cheap • Produces a large number of bands, which can then be further characterized individually (for example, converted into single locus markers)	• Low reproducibility • Mainly dominant • Difficult to analyse • Difficult to automate • Cross-study comparisons are difficult
DNA sequencing	• Highest level of resolution possible • Not biased • Cross-study comparisons are easy; data repositories already exist	• Still significantly more expensive than the other techniques

RAPD, randomly amplified polymorphc DNA; SNP, single nucleotide polymorphism.

"To evaluate the usefulness of SNPs in non-model organisms, it is necessary to consider the investment in marker development, as well as the SNP-typing costs."

all available SNPs for two loci known to be selected showed that even selection events that did not result in the fixation of the selected allele could be detected when a large number of loci and individuals are used[75].

Non-model organisms. Despite a steadily increasing number of completed genomes (even including non-classic model organisms, such as the honey bee and dog), it is clear that for a large number of species, the closest relative with a sequenced genome will always be too distantly related to take advantage of it for marker development. So, the decision of which marker to use will be heavily biased by the costs of marker development. In comparison to microsatellites, two- to four-times as many SNPs are required to obtain comparable results for mapping and identification of individuals[76,77]. Consequently, the accuracy of a microsatellite-based genetic map is higher compared with a map based on the same number of SNP markers[78]. To evaluate the usefulness of SNPs in non-model organisms, it is therefore necessary to consider the investment in marker development, as well as the SNP-typing costs.

Microsatellites are obtained from a genomic library by screening for the presence of a microsatellite motif in the clones. Once isolated, a microsatellite above a certain repeat number can be assumed to be polymorphic[79]. Isolation of SNPs, however, requires the identification of polymorphic sites in a set of sequences. To do this, homologous regions must be sequenced from multiple chromosomes: not a straightforward task if the starting point is a clone library. The amount of sequencing required depends on the SNP isolation approach. Either the homologous sequence must be sequenced from an independent clone (the probability of doing so for two independent libraries is small and depends on the success of enrichment for similar sequences). Alternatively, an anonymous DNA region could be isolated and used to construct PCR primers. These would be used to amplify the homologous region, which is subsequently sequenced for the identification of SNPs. In species with low amounts of variability (such as humans), many PCR products need to be sequenced before a SNP is detected.

One essential question, therefore, is whether the cost-effectiveness of high-throughput SNP assays can compensate for the higher costs of SNP-marker isolation. Many studies in non-model organisms, particularly in conservation-genetic studies, involve only small to moderate sample sizes, which render microsatellites a more cost-effective marker.

The future of molecular markers

All genetic markers depend crucially on the underlying mutation processes that generate variation. These processes are complex and still poorly understood — irrespective of which marker is studied. The trade-off is either to ignore the complexity of the mutation process by using an approximate mutation model (for example, microsatellites) or to neglect rate heterogeneity and recurrent mutations (for example, SNPs). Of course, in both cases it is possible to generate more complex and realistic models from the data. For SNPs, this would mean that parameters for each SNP would need to be estimated, as their mutation rate could differ by orders of magnitude. I suggest that for most applications, it will be better to allow some model violations rather than to attempt to estimate complex models from the data.

For most biological questions that are addressed, the strategy that is used will not matter too much — particularly if a larger number of loci is used. It is more important to avoid systematic bias in the data. DNA sequencing is, therefore, the preferred choice for most marker applications that attempt to characterize existing variation in populations. In particular, when advances in sequencing technology further reduce the costs and increase the throughput, DNA-sequence analysis will gradually replace traditional markers in population surveys. Technical improvements that provide information about the phase of mutations (that is, the assignment of alleles to the two homologous chromosomes), are highly desirable[80].

Christian Schlötterer is at the Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien, Josef Baumann Gasse 1, 1210 Wien, Austria.
e-mail: Christian.schloetterer@vu-wien.ac.at

doi:10.1038/nrg1249

Glossary

AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP). This marker system uses DNA fragments that are ligated to complementary adaptor oligonucleotides and subsequent rounds of PCR amplification using primers that are complementary to the adaptor sequences. The multiple rounds of amplification reduce the complexity of the PCR product population so that the amplified fragments can be easily resolved by gel electrophoresis.

ADMIXTURE
Gene flow between differentiated populations.

ASCERTAINMENT BIAS
Systematic bias introduced by the criteria used to select individuals and/or loci in which genetic variation is assayed; a pronounced problem for single nucleotide polymorphism (SNP) and restriction fragment length polymorphism (RFLP) analyses.

ASSOCIATION STUDY
A study that aims to identify the joint occurrence of two genetically encoded characteristics in a population. Often, an association between a genetic marker and a phenotype (disease) is assessed.

EUCHROMATIN
Part of an interphase chromosome that stains diffusely; less condensed than the heterochromatin.

F_{ST}
Wright's among-population fixation index. A measure of the extent of population subdivision.

GENE CONVERSION
Non-reciprocal exchange of genetic material among chromosomes.

INTER-RETROTRANSPOSON AMPLIFIED POLYMORPHISM (IRAP). DNA fragments found between adjacent, oppositely oriented retrotransposons, amplified through PCR, separated by gel electrophoresis and scored for the presence or absence of fragments.

INTER-SIMPLE-SEQUENCE-REPEAT (ISSR). DNA fragments found between adjacent, oppositely oriented microsatellites, amplified through PCR, separated by gel electrophoresis and scored for the presence or absence of fragments.

LINKAGE DISEQUILIBRIUM
Haplotype frequencies in a population that differ from expectations based on a random combination of alleles at each locus.

PHYLOGENETIC RECONSTRUCTION
Attempt to reconstruct the ancestral relationship among species or populations.

RANDOMLY AMPLIFIED POLYMORPHIC DNA (RAPD). A marker system that relies on the use of short PCR primers.

RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP). A fragment length variant that is generated through the presence/absence of a restriction enzyme recognition site. Restriction sites could be gained/lost by base substitutions, insertions or deletions.

STUTTER BANDS
Artifacts that occur by DNA-replication slippage during the PCR amplification of microsatellites. Most stutter bands are shorter than the actual microsatellite allele.

1. Sturtevant, A. H. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* **14**, 43–59 (1913).
2. Hubby, J. L. & Lewontin, R. C. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* **54**, 203–215 (1966).
3. Harris, H. Enzyme polymorphism in man. *Proc. R. Soc. Lond. B* **164**, 298–310 (1966).
4. Lewontin, R. C. & Hubby, J. L. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595–609 (1966).
5. Johnson F., Kanapi C., Richardson R., Wheeler M. & Stone W. An analysis of polymorphisms among isozyme loci in dark and light *Drosophila ananassae* strains from American and Western Samoa. *Proc. Natl Acad. Sci. USA* **56**,119–125 (1966).
6. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
7. King, J. L. & Jukes, T. H. Non-Darwinian evolution. *Science* **164**, 788–798 (1969).
8. Hillis, D. M., Moritz, C. & Mable, B. K. *Molecular Systematics* (Sinauer Associates, Sunderland, 1996).
9. Berry, A. & Kreitman, M. Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogster* on the east coast of North America. *Genetics* **134**, 869–893 (1993).
10. Hudson, R. R., Bailey, K., Skarecky, D., Kwiatowski, J. & Ayala, F. J. Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. *Genetics* **136**, 1329–1340 (1994).
11. Krieger, M. J. & Ross, K. G. Identification of a major gene regulating complex social behavior. *Science* **295**, 328–332 (2002).
12. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331 (1980).
13. Kerem, B. *et al.* Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**, 1073–1080 (1989).
14. Avise, J. C. *Molecular Markers, Natural History and Evolution* (Chapman and Hall, New York, 1994).
15. Jeffreys, A. J., Wilson, V. & Thein, S. L. Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**, 67–73 (1985).
16. Jeffreys, A. J. *et al.* Human minisatellites, repeat DNA instability and meiotic recombination. *Electrophoresis* **20**, 1665–1675 (1999).
17. Gill, P., Jeffreys, A. J. & Werrett, D. J. Forensic application of DNA 'fingerprinting'. *Nature* **318**, 577–579 (1985).
18. Armour, J. A. L., Povey, S., Jeremiah, S. & Jeffreys, A. J. Systematic cloning of human minisatellites from ordered array charomid libaries. *Genomics* **8**, 501–512 (1990).
19. Saiki, R. K. *et al.* Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354 (1985).
20. Litt, M. & Luty, J. A. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**, 397–401 (1989).
21. Tautz, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* **17**, 6463–6471 (1989).
22. Weber, J. L. & May, P. E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–396 (1989).
23. Goldstein, D. & Schlötterer, C. *Microsatellites: Evolution and Applications* (Oxford Univ. Press, Oxford, 1999).
24. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
25. de Gortari, M. J. *et al.* A second-generation linkage map of the sheep genome. *Mamm. Genome* **9**, 204–209 (1998).
26. Sakamoto, T. *et al.* A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* **155**, 1331–1345 (2000).
27. Queller, D. C., Strassmann, J. E. & Hughes, C. R. Microsatellites and kinship. *Trends Ecol. Evol.* **8**, 285–288 (1993).
28. Schlötterer, C. Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365–371 (2000).
29. Ellegren, H. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* **16**, 551–558 (2000).
30. Tóth, G., Gáspári, Z. & Jurka, J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**, 967–981 (2000).
31. Dieringer, D. & Schlötterer, C. Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Res.* **13**, 2242–2251 (2003).
32. Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. & Tingey, S. V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**, 6531–6535 (1990).
33. Zietkiewicz, E., Rafalski, A. & Labuda, D. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20**, 176–183 (1994).
34. Kalendar, R., Grob, T., Regina, M. T., Suoniemi, A. & Schulman, A. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.* **98**, 704–711 (1999).
35. Zabeau, M. & Vos, P. Selective restriction fragment amplification: a general method for DNA fingerprinting. *European Patent* 0 534858 A1 (1993).
36. Vos, P. *et al.* AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414 (1995).
37. Parsons, Y. M. & Shaw, K. L. Mapping unexplored genomes: a genetic linkage map of the Hawaiian cricket *Laupala*. *Genetics* **162**, 1275–1282 (2002).
38. Castiglioni, P. *et al.* An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. *Genetics* **149**, 2039–2056 (1998).
39. Cervera, M. T. *et al.* Dense genetic linkage maps of three Populus species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics* **158**, 787–809 (2001).
40. Menz, M. A. *et al.* A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP, RFLP and SSR markers. *Plant Mol. Biol.* **48**, 483–499 (2002).
41. Remington, D. L., Whetten, R. W., Liu, B. H. & O'Malley, D. M. Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. *Theor. Appl. Genet.* **98**, 1279–1292 (1999).
42. Schierwater, B. & Ender, A. Different thermostable DNA polymerases may amplify different RAPD products. *Nucleic Acids Res.* **21**, 4647–4648 (1993).
43. Chen, X. & Sullivan, P. F. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J.* **3**, 77–96 (2003).
44. Picoult-Newberg, L. *et al.* Mining SNPs from EST databases. *Genome Res.* **9**, 167–174 (1999).
45. Weber, J. L. & Myers, E. W. Human whole-genome shotgun sequencing. *Genome Res.* **7**, 401–409 (1997).
46. Altshuler, D. *et al.* An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407**, 513–516 (2000).
47. Wakeley, J., Nielsen, R., Liu-Cordero, S. N. & Ardlie, K. The discovery of single-nucleotide polymorphisms — and inferences about human demographic history. *Am. J. Hum. Genet.* **69**, 1332–1347 (2001).
48. Nielsen, R. & Signorovitch, J. Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theor. Popul. Biol.* **63**, 245–255 (2003).
49. Kuhner, M. K., Beerli, P., Yamato, J. & Felsenstein, J. Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics* **156**, 439–447 (2000).
50. Brumfield, R. T., Beerli, P., Nickerson, D. A. & Edwards, S. V. The utility of single nucleotide polymorphism in inferences of population history. *Trends Ecol. Evol.* **18**, 249–256 (2003).
51. Meyer, S., Weiss, G. & von Haeseler, A. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* **152**, 1103–1110 (1999).
52. Schlötterer, C. & Harr, B. Single nucleotide polymorphisms derived from ancestral populations show no evidence for biased diversity estimates in *Drosophila melanogaster*. *Mol. Ecol.* **11**, 947–950 (2002).
53. McVean, G., Awadalla, P. & Fearnhead, P. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–1241 (2002).
54. Kuhner, M. K., Yamato, J. & Felsenstein, J. Maximum likelihood estimation of recombination rates from population data. *Genetics* **156**, 1393–1401 (2000).
55. Beerli, P. & Felsenstein, J. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proc. Natl Acad. Sci. USA* **98**, 4563–4568 (2001).
56. Kreitman, M. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**, 412–417 (1983).
57. Kreitman, M. Methods to detect selection in populations with applications to the human. *Annu. Rev. Genom. Hum. Genet.* **1**, 539–559 (2000).
58. Chikhi, L., Bruford, M. W. & Beaumont, M. A. Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. *Genetics* **158**, 1347–1362 (2001).
59. Reich, D. E., Feldman, M. W. & Goldstein, D. B. Statistical properties of two tests that use multilocus data sets to detect population expansions. *Mol. Biol. Evol.* **16**, 453–466 (1999).
60. Wilson, I. J. & Balding, D. J. Genealogical inference from microsatellite data. *Genetics* **150**, 499–510 (1998).
61. Beaumont, M. A. Detecting population expansion and decline using microsatellites. *Genetics* **153**, 2013–2029 (1999).
62. Kuhner, M. K., Yamato, J. & Felsenstein, J. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* **140**, 1421–1430 (1995).
63. Harpending, H. C. *et al.* Genetic traces of ancient demography. *Proc. Natl Acad. Sci. USA* **95**, 1961–1967 (1998).
64. Wall, J. D., Andolfatto, P. & Przeworski, M. Testing models of selection and demography in *Drosophila simulans*. *Genetics* **162**, 203–216 (2002).
65. Glaubitz, J. C., Rhodes, O. E. & Dewoody, J. A. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Mol. Ecol.* **12**, 1039–1047 (2003).
66. Terwilliger, J. D., Haghighi, F., Hiekkalinna, T. S. & Goring, H. H. A bias-ed assessment of the use of SNPs in human complex traits. *Curr. Opin. Genet. Dev.* **12**, 726–734 (2002).
67. Ohashi, J. & Tokunaga, K. Power of genome-wide linkage disequilibrium testing by using microsatellite markers. *J. Hum. Genet.* **48**, 487–491 (2003).
68. Schlötterer, C. Hitchhiking mapping — functional genomics from the population genetics perspective. *Trends Genet.* **19**, 32–38 (2003).
69. Kayser, M., Brauer, S. & Stoneking, M. A genome scan to detect candidate regions influenced by local natural selection in human populations. *Mol. Biol. Evol.* **20**, 893–900 (2003).
70. Kohn, M. H., Pelz, H. J. & Wayne, R. K. Natural selection mapping of the warfarin-resistance gene. *Proc. Natl Acad. Sci. USA* **97**, 7911–7915 (2000).
71. Payseur, B. A., Cutter, A. D. & Nachman, M. W. Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. *Mol. Biol. Evol.* **19**, 1143–1153 (2002).
72. Harr, B., Kauer, M. & Schlötterer, C. Hitchhiking mapping — a population based fine mapping strategy for adaptive mutations in *D. melanogaster*. *Proc. Natl Acad. Sci. USA* **99**, 12949–12954 (2002).
73. Vigouroux, Y. *et al.* Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc. Natl Acad. Sci. USA* **99**, 9650–9655 (2002).
74. Akey, J. M., Zhang, G., Zhang, K., Jin, L. & Shriver, M. D. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12**, 1805–1814 (2002).
75. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
76. Gill, P. An assessment of the utility of single nucleotide polymorphism (SNPs) for forensic purposes. *Int. J. Legal Med.* **114**, 204–210 (2001).
77. Kruglyak, L. The use of a genetic map of biallelic markers in linkage studies. *Nature Genet.* **17**, 21–24 (1997).
78. Leal, S. M. Genetic maps of microsatellite and single-nucleotide polymorphism markers: are the distances accurate? *Genet. Epidemiol.* **24**, 243–252 (2003).
79. Schlötterer, C. in *Molecular Genetic Analysis of Populations: A Practical Approach* 2nd edn (ed. Hoelzel, A. R.) 237–261 (Oxford Univ. Press, Oxford, 1998).
80. Stephens, M. & Donnelly, P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**, 1162–1170 (2003).

81. Fischer, S. G., & Lerman, L. S. DNA fragments differing by single base-pair substitutions are seperated in denaturing gradient gels: correspondence with melting theory. *Proc. Natl Acad. Sci. USA* **80**,1579–1583 (1983).
82. Myers, R. M., Lumelsky, N., Lerman, L. S. & Maniatis, T. Detection of single base substitutions in total genomic DNA. *Nature* **313**, 495–497 (1985).
83. Rosenbaum, V. & Riesner, D. Temperature gradient gel electrophoresis: thermodynamic analysis of nucleic acids and proteins in purified formand in cellular extracts. *Biophys. Chem.* **26**, 235–246 (1987).
84. Hayashi, K. PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods Appl.* **1**, 34–38 (1991).
85. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. Detection of polymorhisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl Acad. Sci. USA* **86**, 2766–2770 (1989).
86. White, M. B., Carvalho, M., Derse, D., O'Brien, S. J. & Dean, M. Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* **12**, 301–306 (1992).

Acknowledgements
I am very thankful to R. Achmann, B. Harr, M.-T. Hauser, R. Nielsen, T. Leeb, C. Vogl and two anonymous reviewers for helpful discussion/comments on earlier versions of the manuscript. The laboratory of C.S. is supported by grants from the Fonds zur Förderung der Wissenschaftlichen Forschung and the European Union.

Competing interests statement
The author declares that he has no competing financial interests.

Online links

DATABASES
The following terms in this article are linked online to:
FlyBase: http://flybase.bio.indiana.edu
Adh

FURTHER INFORMATION
HapMap: http://www.hapmap.org
Molecular Ecology:
http://www.blackwellpublishing.com/journals/mec/submiss.htm
Access to this interactive links box is free online.

ONLINE

Christian Schlötterer received his doctorate from the Universität München and pursued postdoctoral training at the University of Chicago. He joined the Veterinärmedizinische Universität Wien in 1995, where he is now an associate professor at the Institut für Tierzucht und Genetik. He has a long-standing interest in the development of molecular markers and their applications to a range of biological questions. His current research interests encompass genome evolution and population genetics.

URLs

FlyBase

Adh
http://flybase.bio.indiana.edu/.bin/fbidq.html?FBgn0000055

Further information

Molecular Ecology:
http://www.blackwellpublishing.com/journals/mec/submiss.htm

HapMap:
http://www.hapmap.org