This is a copy of the final published version of a paper published over 20 years ago in Trends in Ecology & Evolution.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/84082

Published paper

Methods that allow the detection and verification of genetic relationships among organisms studied in the field have long been sought by workers in evolutionary biology. Since its inception just four years ago, DNA fingerprinting has already begun to fulfill its promise as a widely applicable solution to this problem. Here, recent progress with fingerprinting is discussed in the context of some other DNA techniques. In particular, the use of molecular methods is revolutionizing the study of mating systems.

Ecological and evolutionary biologists often endeavours to determine genetic relationships and mating patterns among individual organisms studied in the field. For example, in a previous issue of TREE Birkhead described how males of bird species considered to be monogamous may frequently copulate with females other than their own mate (extrabond copulations). If such behaviour is adaptive, extrabond copulations should have a chance of leading to successful fertilization, and they could therefore affect substantially the reproductive success of males. Measures of a male's reproductive performance obtained by counting offspring in nests may be too low if a male's extrabond copulations are successful, or too high if he is a victim of this behaviour by other males.

In lekking species, where males provide no parental care, we would ideally like to be able to test sexual selection theory by comparing the fitness of different male phenotypes and behaviours. Until recently, the only component of fitness that it has been possible to compare is the number of observed matings achieved by each male, but this may be misleading as the number of successful fertilizations may be substantially different. In such cases we wish to be able to identify the true fathers of individual offspring.

The reproductive success of females as well as males might be positively or negatively affected by intraspecific brood parasitism, a behaviour recorded in many species (in which a female lays an egg in the nest of a conspecific). Brood parasitism may sometimes be detectable through the appearance of an unexpected additional or distinctive egg in a nest, but such observations may be confounded by parasitic females that remove an egg when laying their own.2,3

Heritability analyses—which are of increasing interest to behavioural ecologists4—and estimates of effective population size may also be confounded by the misallocation of offspring to parents. There has therefore been a longstanding need for methods that allow us to determine true biological relationships.

Genetic markers

There have been several attempts to use genetically polymorphic markers to verify true biological relationships and mating systems. Though some studies on wild populations have used morphological or chromosomal characteristics under simple genetic control (e.g. colour polymorphisms in Cepaea nemoralis snails5, platyfish Xiphophorus maculatus6, and lesser snow goose Chen c. caerulescens7; chromosomal inversions in Drosophila8), most have used biochemical polymorphisms detected by starch gel electrophoresis (e.g. Refs 9, 10 and 11).

Genetic methods have most often been applied to tests of paternity, since true mothers are often known or assumed. However, precise data have been difficult to obtain, even in those studies using electrophoresis in which many polymorphic genetic systems were available. If an incorrectly assigned male is to be detected, he has to have a genotype that is inconsistent with the female–offspring combination. In electrophoretic studies, the mean probability of detecting the misassignment of a male (the exclusion probability) was typically in the range 0.4–0.7, and so the corresponding probability of nondetection (the inclusion probability = 1 – exclusion probability) was still usually too large to allow true fathers to be detected with a high level of certainty. In just a few studies of wild populations, the number of possible fathers in each breeding group has on occasion been sufficiently small to allow paternity to be inferred by a process of elimination of all the non-fathers.2,3

There has therefore been a need for methods that produce a much larger exclusion probability. This probability is a function of the number of polymorphic loci examined and the number and frequencies of alleles detected at those loci. The use of protein polymorphisms has proved inadequate because the number of scorable polymorphic loci has usually been insufficient to compensate for the relatively low number of alleles, and low heterozygosity, at those loci. It is now apparent that the methods of DNA analysis that have been developed during recent years will often provide an alternative and successful solution to the problem of detecting more genetic variability.

DNA methods

Four different classes of variable DNA sequence have been used in the study of mating behaviour: random restriction fragment length polymorphisms (RFLPs), minisatellite DNAs (detected as DNA fingerprints), sex-linked sequences and mitochondrial DNA. The random RFLP and DNA fingerprinting methods are the most generally applicable, and both have been used first in field studies of birds. This partly reflects the frequent choice of birds for population biological research, but birds are also especially convenient subjects because avian erythrocytes are nucleated (unlike those of mammals) and a very small drop of blood therefore provides an adequate quantity of DNA.8,9,10

The random RFLP method, applied to the detection of non-kin in the lesser snow goose7,8,11, involves the specific detection of

DNA Fingerprinting and Other Methods for the Study of Mating Success

Terry Burke
The procedures for DNA fingerprinting and RFLP analysis share many features. An individual’s DNA is isolated (A) from a tissue sample (e.g. mammalian leucocytes or avian erythrocytes) and cleaved with a restriction endonuclease that cuts the molecules at each position containing a specific short sequence of nucleotides (B). The resulting fragments are separated according to size by gel electrophoresis (C). Once separated, the double-stranded DNA fragments are denatured into their component single strands and permanently transferred to a nylon or nitrocellulose filter membrane by Southern blotting (D). Specific regions of the DNA are then detected and characterized by using a previously cloned DNA sequence as a probe (E1 and E2). The probe DNA is radioactively labelled (F), denatured into single strands, and then bound (hybridized) in solution to those restriction fragments on the membrane that contain complementary sequences (G). These restriction fragments are finally revealed as bands on an autoradiograph (H), by exposing it to the hybridized filter (H).

**RESTRICTION FRAGMENT LENGTH POLYMORPHISMS**

If the sequence containing a restriction endonuclease recognition sequence differs among individuals, then when this restriction site (J) is absent one large fragment will be detected instead of two smaller ones. Several such sites may vary within the region detected by the probe. This variability is described as a restriction fragment length polymorphism (RFLP). In addition to the variability at restriction sites, RFLPs may also arise due to the variability in length between sites that can be produced by the insertion or deletion of DNA sequences. In a study of lesser snow geese by Quinn and colleagues, the probe DNAs were obtained by cloning random pieces of those DNA sequences present in the genome in only a single copy. The choice of restriction enzymes used in the search for RFLPs affects the number and relative variability of RFLPs that are detected. The enzymes used in this study included two (TaqI and MspI) that contain the relatively mutable CpG dinucleotide in their recognition sequences, and this helped these workers to find several particularly variable—and therefore useful—RFLPs.

**REFERENCES**

RFLP analysis and DNA fingerprinting reveal differences in the nature of the selected RFLPs, containing minisatellite DNA variable in length due to variation in the dinucleotide repeat units; there are therefore more alleles at such a locus than occur in the loss and gain of restriction sites. Units at many minisatellite loci include polycore sequences (E2), leading to bands that show mendelian inheritance. Analysis, however, these bands cannot be detected with polycore sequences (E2), leading to bands that show mendelian inheritance. However, these bands cannot be detected.

In the interpretation of the fingerprints (right), diagnostic bands specific to each adult, and the parent of origin for bands in offspring, are indicated with coloured lines:

- Solid black lines indicate bands of apparently similar mobility in any two of the adults, and in the offspring they indicate bands that could be from either parent.
- Grey lines indicate bands where scoring in at least one adult is confounded by an intense band nearby, and in offspring they indicate that at least one parent has the band.
- Diagnostic bands are used for initial assignment of chicks to males, then all other bands are checked for their presence in the assigned adults. The mean probability of a male other than the father containing all the paternal bands is very small (<10^-9).

Fingerprints reproduced with permission from Nature.
polymorphisms at one locus at a time. The alternative approach, DNA fingerprinting, simultaneously detects polymorphisms at multiple loci, and has been applied in particular to the analysis of paternity in the house sparrow (Passer domesticus)\(^\text{15,20}\), dunnock (Prunella modularis)\(^\text{21}\) and long-finned pilot whale (Globicephala balaena)\(^\text{22}\). Though the two methods detect distinct classes of DNA sequence, they use closely related technology. Both methods are described in the accompanying centre-page diagram.

The locus-specific RFLP and multilocus DNA fingerprinting approaches each have their own advantages and disadvantages. The main advantage of fingerprinting is that the available probes can be applied to many diverse species, whereas at least 80% of random RFLP probes (in birds \(^\text{17}\) probably derive from non-coding unique sequences which are unlikely to be conserved among species. Isolating suitable probes therefore requires a large initial investment. Locus-specific analysis is technically slightly easier than fingerprinting, but many probes have to be used in combination to obtain a useful exclusion probability, and this prolongs the analysis. However, where there are difficulties in obtaining large families for evaluating fingerprinting systems (see below), the locus-specific approach offers the advantage that allele frequencies and disequilibria can be measured by pooling data from different individuals and families. The RFLP method has been described in detail elsewhere (Ref. 17, and see centre-page diagram).

DNA fingerprinting

A 'minisatellite' (also known as a hypervariable region, HVR, or a variable number tandem repeat locus, VNTR) is a DNA sequence (usually less than 20 000 base pairs) comprising multiple copies of a short sequence ('tandem repeat unit') of typically less than 65 base pairs. A minisatellite's organization is therefore similar to that of 'satellite' DNA, which also consists of multiple repeat sequences but on a much more extensive scale. Satellite DNAs are non-coding and so, presumably, are most minisatellites, though two are known to form parts of coding sequences\(^\text{23,24}\).

The key advance that led to the development of DNA fingerprinting came when, in the course of an analysis of the human myoglobin gene, Alec Jeffreys and colleagues discovered a new family of minisatellite sequences that had in common a 'core' sequence of about 12 nucleotides\(^\text{25-27}\). They showed that those minisatellites that consisted of multiple repeats of this core sequence could be used as probes to detect simultaneously the hypervariable minisatellites at many separate loci (see centre-page diagram). The theoretical probability of the same set of DNA fragments being detected in two humans is so small that every human except identical twins is expected to have a unique pattern, and the pattern of bands obtained on an autoradiograph is therefore described by analogy as a DNA 'fingerprint'\(^\text{25-27}\). Two slightly different poly-core probes, 33.6 and 33.15, were found which each could be used to obtain distinct fingerprints\(^\text{26,27}\).

Individual minisatellite loci are considered 'hypervariable' because they include the most polymorphic sequences ever detected. The hypervariability is the result of a high mutation rate for the loss or gain of repeat units. At the most variable locus\(^\text{28}\) a sample of 79 humans was found to have at least 77 different alleles, with an estimated heterozygosity of 97% and mutation rate of 0.003 per gamete\(^\text{29}\). We know little about the evolution of these minisatellites, though similarity between the core sequence and the chi cross-over initiator site in E. coli suggests that the core may play a role in a recombination process that leads to frequent length mutations\(^\text{29}\).

Other probes have since been discovered that can also detect highly variable complex band patterns\(^\text{30-32}\), and at least one of these has been shown to detect individual-specific fingerprints\(^\text{30}\). In order to prove that fingerprints are individual-specific it is necessary to show that a large proportion of the detected minisatellites are in linkage equilibrium. In practical terms, this means showing that they are inherited independently and that they therefore belong to unlinked loci. Proving a familial relationship also relies upon the assumption of independence among bands (though the exclusion of parentage may still be possible).

The simplest way to demonstrate independence is to analyse the segregation of bands in a large family (typically two parents with ten or more offspring). This analysis has so far been carried out in only a small number of animal species (human\(^\text{33}^\), dog\(^\text{34}\), cat\(^\text{34}\), mouse\(^\text{35}\), house sparrow\(^\text{15,36}\) and dunnock\(^\text{21}\)) and, in all cases except one\(^\text{30}\), only with the probes discovered by Jeffreys. In all of these species except the mouse the bands were found to be predominantly independent. The probability of the false inclusion of non-relatives as fathers by using two core probes is typically\(^\text{15,21,33,34}\) in the range 10^-2 to 10^-8; a single core probe is often adequate and close relatives can usually also be excluded\(^\text{37}\).

In the mouse, although there is a high degree of variability among inbred strains, large numbers of minisatellite fragments (\(<10^2\)) were found to be co-inherited and, therefore, closely linked\(^\text{35}\). This result emphasizes the importance of carrying out an evaluation of each restriction enzyme/probe combination prior to its application to studies of relatedness. This evaluation should first include the comparison of different enzyme/probe combinations to find those that produce the most potentially informative combinations of resolvable bands and the degree of band sharing among random individuals\(^\text{38}\).

Applications

There is evidence that minisatellite probes detect variable complex band patterns in a wide range of vertebrate species\(^\text{15-26}\), and even in plants\(^\text{39,40}\), but it remains to be seen whether at least one system can always be found that is informative for testing relationships. In view of the finding in the one plant species investigated in detail - rice - that minisatellites detected by human-derived probes are dispersed in the genome\(^\text{39}\), the signs that such probes will have very wide applicability are promising. Whether the fingerprinting system will provide sufficient statistical power will, however, depend very much on the precise questions...
and the parameters of the system (numbers of bands, degrees of band sharing and independence). In particular, we do not yet know whether relationships will be easily resolved in situations where dispersal is low and populations are relatively inbred.

Apart from individual identification for forensic analysis, the identification of identical twins, monitoring bone marrow transplants and studies of tumours, DNA fingerprinting has so far mainly been applied to the analysis of paternity even, on occasion, in the absence of a sample from the father (partial patern al fingerprints can be inferred from mother-offspring comparisons, and the paternity of siblings compared). Paternity analysis is simplest when the assumption can be made that the mother is correctly identified, and this can first be tested. Mutations are common at hypervariable minisatellite loci and it is therefore necessary to allow for this in the analysis.

One of the first population studies to use DNA fingerprinting was concerned with the relationship of mating and parental care behaviours to paternity in the dunnock. This species has a remarkably flexible mating system. Polyandry (where a female shares a territory with two males) is the most commonly observed system. The two males have a dominance relationship (alpha male is dominant to beta male), and either or both may help to feed the female's brood. DNA fingerprinting showed that a male was much more likely to feed the brood if he had sired some of the nestlings (Table 1), though he showed no preference for, and presumably could not identify, his own offspring. A beta male was also more likely to feed a brood if he had had some exclusive access to the female during the pre-laying period, and his proportion of exclusive access was related significantly to the number of bands shared by the female. Paternity analysis has become valuable to studies of genetic interactions between species, the data obtained concerning the relative mating propensities of the sexes in hybridizing populations may well improve our understanding of some of the problems of sexual selection.

Table 1. Association between paternity (as ascertained by DNA fingerprinting), feeding of nestlings and exclusive access to the female during the mating period for polyandrous beta male dunnocks.

<table>
<thead>
<tr>
<th>Beta male's paternity</th>
<th>Male fed young</th>
<th>Male did not feed young</th>
<th>Male had some exclusive access to female</th>
<th>Male had no exclusive access to female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broods where he had paternity</td>
<td>14</td>
<td>3</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Broods where he had no paternity</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

* P = 0.001

** Fisher's exact probability test.

** * Where both alpha and beta males had had some exclusive access, the mean paternity split was 55% alpha: 45% beta. Data from Ref. 21.

Sex-linked and mitochondrial DNA

The two other kinds of DNA analysis that have provided data concerning mating behaviour have rather more specialized applications. One of these has involved the use of the extensive variability present in the ribosomal RNA gene family located on the mtDNA, and the similarity between mtDNA and the similarly clonally inherited Y-chromosome markers is proving valuable to studies of genetic interactions between species. The data obtained concerning the relative mating propensities of the sexes in hybridizing populations may well improve our understanding of some of the problems of sexual selection.

Practical aspects

The application of DNA technology would be perceived as expensive by most ecologists, but the scientific value added to a project by good data on genetic relationships will often more than justify this expense. The major direct cost is labour, but consumables are also expensive; a trained technician would probably produce up to 1000 DNA fingerprints per year, using about £5000 ($9000) worth of consumables. Many samples may also need to be run more than once, for example during evaluation of the system and if a single probe/enzyme combination should prove inadequate. The need for access to expertise and an expensive set of equipment makes collaboration with a molecular biological
The future are unlikely to be closely associated with coding sequences and are opening the way for the application of many long-term studies it may pay dividends in the future if blood retrospective analyses. The impetus of the new genetic methods, which are already being applied in many laboratories worldwide, will shortly lead to many new and valuable insights into animal behaviour. Perhaps one of the most pleasing trends to come out of this activity will be the bridging of the growing gulf between those who focus on whole organisms and those with more reductionist interests.

Acknowledgements
Our work on DNA fingerprinting is supported by the NERC and SERC.

References
19 Quinn, T.W., Davies, J.C., Cooke, F. and White, D.N. (1987) in press

To obtain extra copies of the colour centre-page diagram accompanying this article, write to: TREE Centre-page Diagrams, TREE Subscriptions, Elsevier Science Publishers Ltd, Crown House, Linton Road, Barking, Essex IG11 8JU, UK.

Quote reference number E3. All orders should be prepaid.

Price per six copies (minimum order quantity): £8.00 (plus 15% VAT in the UK) or US$15.00. There is a 15% discount on orders of 100 and over. Prices valid until December 31st 1989.