CHAPTER 8

Strategies for Finding and Using Highly Polymorphic DNA Microsatellite Loci for Studies of Genetic Relatedness and Pedigrees

JOAN E. STRASSMANN, CARLOS R. SOLÍS, JOHN M. PETERS,* AND DAVID C. QUELLER

Department of Ecology and Evolutionary Biology, Rice University, Houston, Texas 77005–1892

CONTENTS

*Present address: Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853.

Molecular Zoology: Advances, Strategies, and Protocols, Edited by Joan D. Ferraris and Stephen R. Palumbi.
ISBN 0-471-14461-4

SYNOPSIS

Detailed knowledge of the genealogical relationships between individuals is often essential to developing an understanding of the evolutionary significance of social interactions. Related individuals may cooperate in order to obtain indirect fitness benefits in situations in which unrelated individuals would not. Genetic information can also be critical for understanding the dynamics of social groups since it can indicate which individuals are reproducing, distinguish progeny of different parents, and reveal surreptitious reproduction. A codominant, neutral, single-locus class of genetic markers that are sufficiently polymorphic for relatedness estimation and parentage assignment has recently been identified and promises to revolutionize the field. These markers, known as DNA microsatellites, can be compared reliably across gels, which allows for comparisons among large numbers of individuals. They are short tandem repeats of simple motifs, such as AAT. The polymerase chain reaction can be used to amplify DNA regions that contain such repeats. These products are electrophoresed and then visualized, allowing precise determination of length polymorphisms. Polymorphic microsatellites have been used in gene mapping, population structure analyses, and most recently in assessment of relatedness and parentage. Microsatellites typically are found in rapidly evolving, noncoding DNA. Therefore flanking sequences used to prime the amplification of a microsatellite in one species are likely to work only in fairly closely related species. Although identifying microsatellites in a species of interest can be time-consuming, the techniques employed are well established and reliable. They involve constructing a partial genomic DNA library, identifying and sequencing clones containing microsatellites, designing polymerase chain reaction primers that flank the repeat region, amplifying the microsatellite in a population sample of individuals, and using those microsatellites that are polymorphic. Here we present straightforward protocols that should be applicable to all species.

INTRODUCTION

Evolutionary studies of social behavior benefit greatly from the availability of genetic markers capable of revealing parentage and other genetic relationships among individuals. Such markers are important because it is often neither feasible nor efficient to obtain relatedness information by observing individuals: females may mate with several males and copulations may be concealed. In social insects, the problem is particularly severe since even the egg layers cannot always be identified, and mating usually occurs outside the colony. In social species, there may also exist conflicts of interest among relatives, which can be understood with precise genetic data. Related individuals may cooperate in order to obtain indirect fitness benefits in situations in which unrelated individuals would not.

To date, a number of different techniques have been applied to the task of establishing genetic relatedness (Queller et al., 1993). Pedigrees can be based on inferring parentage from behavioral observations. This is the oldest technique. It is

very useful in some cases but may not detect surreptitious matings or identify mothers in groups where many females are reproducing. Genetic techniques have indicated that surreptitious matings are much more common than was previously assumed (e.g., Birkhead and Moeller 1992; Westneat et al., 1987). Allozymes have been used extensively in studies of genetic relatedness (e.g., Gadagkar, 1985; Queller et al., 1993; Strassmann et al., 1989). They have been very useful for establishing relatedness among classes of individuals across groups. However, low genetic variability and a paucity of loci make this technique unsuitable for parentage assignment or within-group relatedness determinations. DNA fingerprinting has proved to be a reliable technique for parentage exclusion within family groups, but it is not suitable for examining larger groups or across groups because of the high variability of minisatellite loci and the difficulties associated with making comparisons between gels. In addition, fingerprinting requires large amounts of high-quality DNA. Other techniques discussed elsewhere in this book, such as RAPDs and SSCP, have advantages and disadvantages for various applications. For parentage assignment and relatedness estimation, an ideal marker would be mendelian, codominant, selectively neutral, and highly polymorphic. Scores would be unambiguous enough to be evaluated against a standard across gels. DNA microsatellites meet these criteria (Litt and Luty, 1989; Smeets et al., 1989; Tautz, 1989; Weber and May, 1989).

In this chapter, we discuss how to find and use microsatellites. We describe in detail techniques that have been used successfully in our laboratory on wasps, ants, and fish (Hughes and Queller, 1993; Queller et al., 1993). Since many investigators will find microsatellites from the literature or from GenBank, we first discuss using microsatellites that have previously been identified, and then go on to techniques for locating microsatellites in a new organism.

USING MICROSATELLITES

Finding Microsatellites

The simplest means of identifying microsatellites is by using previously described sequences. Partial sequence information for a growing number of species is available through GenBank or in the literature (Stallings et al., 1991). This is particularly true for species related to those popular with geneticists, such as *Drosophila*, mice, and humans, as well as agriculturally important species such as cattle and chickens (Cornall et al., 1991; Beckmann and Weber, 1992; Edwards et al., 1992). For reasons to be explained later, a disadvantage to this approach is that most previously identified microsatellites are dinucleotide repeats.

It is often possible to use microsatellite loci that have been identified for another species. Microsatellite primer pairs frequently amplify sequences common to most closely related species in any one genus and have been found occasionally to work across families. Currently, the success with which specific microsatellite loci can be used on other species must be empirically determined on a case-by-case basis. One

study found that the relationship between repeat length and heterozygosity (Weber, 1990) present in the original species did not hold for new species (Moore et al., 1991). This suggests that all possible microsatellites should be explored in new species, not just those that were most polymorphic in the original species. Of 48 specific and polymorphic primers identified in cattle, 20 were variable in sheep, and none were variable in horses (though three amplified but were monomorphic) or humans (Moore et al., 1991). Arévalo et al. (1994) and Bhebhe et al. (1994) tried ten primers identified in goats (*Capra hircus*) on eleven species from eight other genera across the Bovidae. They found that four of the loci did not amplify product in any of the other species. Of the others, two were polymorphic in four new species, two were polymorphic in five species, one was polymorphic in six, and one was polymorphic in seven of the eleven species. Another locus that was monomorphic in goats proved to be polymorphic in cows (E. Arévalo, *personal communication*). One or more of the six dinucleotide microsatellites identified from the reed bunting (*Emberiza schoeniclus*) worked on species from other families, including Muscicapidae, Sturnidae, Paridae, Hirundinidae, Sylviidae, and Passeridae, although they were not evaluated for polymorphisms in the new species (Hanotte et al., 1994). One of two loci from barn swallows were polymorphic in both bank swallows and house martins (Ellegren, 1992). Three dinucleotide repeats isolated from brown trout (*Salmo trutta*) were also present and polymorphic (exhibiting new alleles) in the rainbow trout (*Oncorhynchus mykiss;* Estoup et al., 1993). Primers identified in humans are useful and polymorphic in both species of chimpanzees (Morin and Woodruff, 1992; Morin et al., 1994). We have found that the primers for 40 or so trinucleotide microsatellites that we have identified in three species of polistine wasps in two genera generally work well on species in the same subgenus. We have evaluated a sample of all the genera in the Vespidae and have found at least two of our 40 microsatellite primers produce a PCR product of the right size on all genera assayed (V. O. Ezenwa et al., *in preparation;* J. Strassmann et al., *unpublished data*). To what extent these will prove polymorphic in the new species, and therefore useful, is still unknown. Thus, once a large number of microsatellites are available for a given group, a much wider range of taxa may be accessible to study.

DNA Extraction

One of the primary advantages of PCR-based microsatellite analysis is that nanogram quantities of low-quality, partially degraded DNA can be used effectively, since fragments of only 100–300 base pairs are needed. By contrast, microgram quantities of high-quality DNA are necessary for traditional DNA fingerprinting. The minute quantities of DNA required for microsatellite analysis represent a significant advantage for researchers studying small organisms or aiming to carry out nondestructive genetic sampling of moderate-sized organisms. Microsatellite studies of genetic relatedness and parentage often require that many samples be analyzed. It is therefore important that the DNA extraction technique be easy and yield sufficient stable DNA to provide an adequate time window for the analysis. Con-

ventional DNA isolation techniques designed to give high molecular weight DNA thus may be unnecessarily expensive and time-consuming for this application.

Contaminants, including EDTA and possibly insect cuticle pigments, inhibit the PCR reaction, so techniques that completely remove or avoid the use of these compounds are best (Strassmann.1). Our preferred protocol (Strassmann.1) works quickly and easily for whole insects, tissues, and eggs. A second protocol used extensively in our laboratory may be more successful if removal of pigment contaminants and lipids proves to be problematic (Strassmann.2). Some questions require that DNA be amplified from ancient, tiny, or degraded samples, samples stored in ethanol, sperm or eggs, blood, feathers, hair, scales, and other unconventional samples. Such samples may require different extraction techniques. We extract DNA from tissues where we can obtain a few micrograms (Strassmann.1, Strassmann.2). When only a tiny amount of tissue is available, as is the case for hymenopteran sperm from female spermathecae, we simply lyse the cells to make the DNA available to the PCR primers and avoid precipitating the DNA (Strassmann.3).

Carry-over of DNA between samples is a potentially serious problem since PCR can geometrically amplify even single DNA molecules. We use a fresh grinding surface for each sample (Strassmann.1). For both large, hard wasps and softer larvae we use a bead beater (Biospec Products), which pulverizes the tissues in a grinding buffer with two sizes of glass beads, which are then discarded. For wasp eggs, we use individual, disposable pestles made by melting a pipetteman tip in an eppendorf tube (Strassmann.1). Of course, many tissues can be broken down chemically without grinding of any kind.

We store genomic DNA at 4°C unless we do not plan to use it for 2 months or more, in which case, it is stored at −70°C. We find that this storage strategy minimizes DNA breakage, which could be caused by repeatedly freezing and thawing DNA samples.

Conditions for PCR

We run all radioactively labeled (with ^{35}S) PCR reactions in 10-μl volumes with an oil overlay under fairly standard conditions (Strassmann.4). We have found 10 μl to be the optimal volume since 5-μl volumes result in a higher number of samples failing to amplify on the first run. Volumes greater than 10 μl increase cost (more *Taq* polymerase and ^{35}S label must be used) and are unnecessary, since only 4 μl of the reaction, including loading buffer, are typically run out on the gel in the next step.

Visualizing Microsatellite Products and Size Standards

Before microsatellite polymorphisms can be scored, they must be visualized. This is typically done by running the product out on a gel (Strassmann.5). Some laboratories run products out on native, nondenaturing, thin (0.6 mm), acrylamide gels and then stain them with ethidium bromide or use silver staining protocols. Another

means of visualizing length polymorphisms is through end-labeling one primer with ^{32}P. Although this method gives cleaner results for dinucleotides (because amplifying both strands increases stutter band numbers), it does, of course, use hazardous ^{32}P. If you plan to run thousands of samples, automated sequencing machines may be advantageous, though the use of fluorescent dyes makes this expensive. The most useful automated sequencing machines for microsatellite applications are those that read five different fluorescent labels per lane. These fluorescent labels are usually used to identify the four nucleotides and one size standard, but they can be used to label as many as nine different microsatellite products, if the microsatellites are of two nonoverlapping size classes. The microsatellite primers are end-labeled with the fluorescent labels (a different one for each same-sized microsatellite product to be run in that lane), which leaves one as a size standard. The resulting data can be imported directly into a program for analysis.

In our laboratory, we run denaturing gels exactly like those used for sequencing and visualize the microsatellites by incorporating ^{35}S into the PCR reaction (Figs. 1 and 2; Strassmann.5). Our technique has a number of advantages. A sequencing reaction can be run as a size standard on the denaturing gel. ^{35}S is a very low-energy beta emitter and, as such, is much less hazardous than ^{32}P. Another advantage is that the label is incorporated during the PCR reaction, so no extra steps are required to label primers. There has been some concern about volatile decomposition products from ^{35}SdATP in PCR (Trentmann et al., 1995), which make it preferable to keep the thermocycler in a vented hood. In our own laboratory frequent wipe tests inside and outside the lid of the thermocycler have not indicated elevated levels of radiation, perhaps because our volumes are so small.

It is possible to run more than one locus in a lane, if the fragments are either of nonoverlapping size classes or are loaded at different times. In the former case, samples are simply mixed and then loaded together. In the latter case, sample sets are loaded about an hour apart depending on the size, and separate size markers are needed for each set. We do not run more than one locus in the actual PCR reaction because interactions occasionally occur between the primers in some samples.

We score alleles for microsatellite loci by assigning each allele a size based on the M13mp18 sequence run in each gel (Strassmann.5). Samples may exhibit a complex banding pattern in which an allele may appear as several bands, particularly if they are dinucleotide repeat motifs. Care must be taken to always select the same band for scoring.

For behavioral ecologists, trinucleotide microsatellites are typically superior to dinucleotides. While not necessarily as variable or abundant as dinucleotides, in most species (even those in the Hymenoptera, which are frequently deficient in allozyme variation), trinucleotides have more than ample variability for precise relatedness estimation and for parentage assignment. Once identified, trinucleotide microsatellites are much more easily and accurately scored than shorter repeat motifs. Trinucleotides can be internally labeled with ^{35}S, whereas dinucleotides must often be end-labeled with ^{32}P or other relatively high-energy isotopes to reduce shadow bands. This adds an extra enzymatic step to the DNA analysis, as well as presenting a greater safety hazard. However, for population geneticists interested

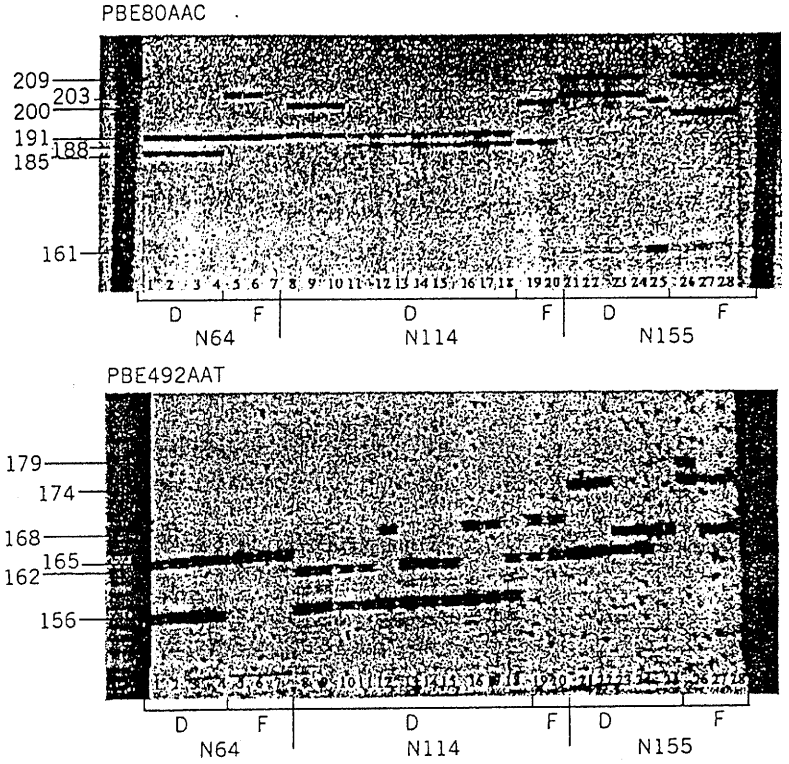

Figure 1. Trinucleotide microsatellite genotypes of *Polistes bellicosus* mothers and daughters on three colonies and two loci. Note the virtual absence of stutter bands from these trinucleotide repeats. The outside lanes are size markers from the C lane of an M13 sequence. For each colony, D indicates daughters and F indicates foundresses, or mothers. Lacking paternal genotypes, daughters cannot always be assigned unambiguously to one mother. However, it is clear from her score at PBE80AAC that daughter 25 is not the offspring of any of the foundresses on the colony. Daughter 25's score from PBE492AAT is consistent with this result, but this locus alone does not rule out any of the queens as mothers. Taken together, however, these data indicate that daughter 25 probably wandered over from another nest. (From Strassmann et al., 1995, by permission. http://www.rice.edu/wasps)

primarily in genome mapping, the relatively high frequency of dinucleotide repeats may be essential.

Error Checking

Eliminating all sources of error is essential when using microsatellites for parentage assignment. Thus all aspects of sample preparation, gel loading, and scoring must

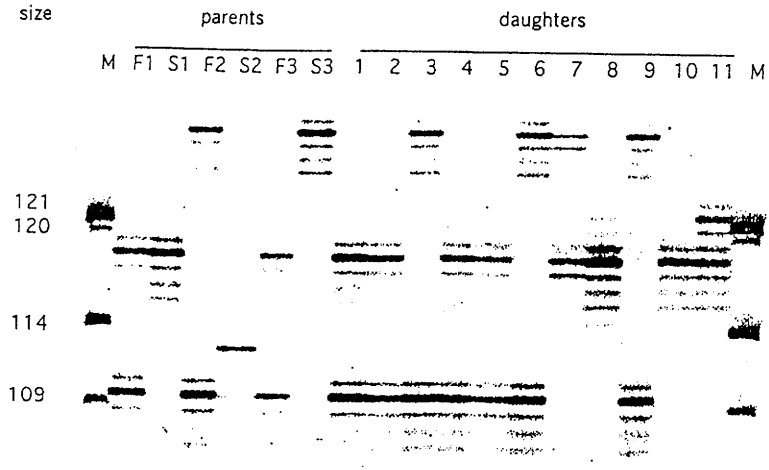

Figure 2. Microsatellite genotypes of *Polistes annularis* mothers, sperm from their spermathecae and daughters. The two outside lanes (M) are size markers (the C lane from an M13 sequence) labeled in nucleotides. The interior lanes contain amplified genotypes for locus PAN111bAAT from one colony. Three parent sets are shown, both foundress (F) and her stored sperm (S). Each sperm lane shows a single allele (each allele shows one strong band plus several weak stutter bands characteristic of microsatellites). The remaining 11 lanes are daughters. Lanes 1, 2, 4, 5, 8, and 10 show daughters of the first foundress and lanes 3, 6, 7, and 9 show daughters of the third foundress; note that these are distinguishable only because of the sperm. Lane 11 matches none of the collected foundresses and is the daughter of an uncollected, presumably dead, foundress. (From Peters et al., 1995, by permission.)

be done in ways that minimize error. Incorrect genotyping of an individual could result either in false exclusion of the true parent or in false acceptance of a nonparent. While both types of error are equally serious, to illustrate the importance of minimizing errors, we will focus here on errors leading to the false exclusion of the true parent. The analysis that follows could apply to errors in tube labeling, in adding the wrong DNA to the tube, in loading, in scoring, or in entering data.

Suppose we score n loci and the error rate per locus per individual is e. Suppose further that a fraction k of these errors leads to false exclusion of the true parent (some errors will not result in false exclusion; for example, true parent is ab, offspring is bb, but we score offspring as aa). When we score n loci, there are $2n$ chances to make an error for every parent–offspring pair. The probability of false exclusion, F, for a single pair is one minus the probability that no false exclusion error will happen $(1 - ke)$ in $2n$ tries or

$$F = 1 - (1 - ke)^{2n}$$

The goal is clearly to make e as low as possible, but how low does it need to be? Suppose $k = 0.5$ and suppose we are scoring $n = 8$ loci (a fairly typical number). If our error rate, e, is 0.05, then 33% of the true parents will be excluded, which is clearly unacceptable. If we lower the error rate to 0.01, we exclude 8% of the true parents, and this is still too high. Perhaps we might consider a 1% rate acceptable (although even this is probably too high for reasons given in the next paragraph). Solving for e, we get 0.0012. In other words, we need to strive for an error rate of less than 1 in 1000.

The impact of errors on the accuracy of results is even more serious in the case of errors affecting many individuals. For example, if one skips a sample in loading a gel, all the samples loaded thereafter will be assigned to the wrong individual. A similar cascade effect can result from an error in entering data. Also, any error in a parent's genotype may affect all of its offspring, so a single error could produce incorrect results for an entire family.

Of course, F would be reduced by having lower k or n in equation 1. Obviously, however, this is not an option since high values of k and n are the very things that we need for parentage studies to succeed. If we are to succeed in our goal of correctly excluding individuals as parents, we must therefore screen many variable loci, each of which provides us with a good chance of being able to discriminate among individuals.

A number of procedures will reduce or eliminate error. The best check would be to do everything twice independently and then compare the results. But it is impractical to do all the molecular work twice. Tubes must be labeled carefully and ordered in ways to minimize mistakes. Checks can be made by running subsamples twice. We reduce mixups between people by supplying everyone with her/his own working box of primers and other reagents.

PCR reaction setup is a stage at which errors are both possible and serious. If a tube is skipped, the outcome may be errors in all subsequent tubes. The setup that we find minimizes errors is one that puts the genomic DNA and the PCR tubes in the same column in the same rack, so genomic DNA is only added to PCR tubes directly in front of it.

Gel loading is another point where mixups are possible and serious. Samples should be processed and loaded onto gels in numerical or alphabetical sequence, to prevent errors when recording loading patterns and make "skipped" samples immediately evident. Use a rectangular heat block to denature DNA (not round floaters). Each tube should be moved to a separate rack as loaded, to prevent double loading. Indicate skipped lanes, and locations of markers, by leaving gaps in the separate rack and mark them in pen on the front glass of the sequencing rig. After loading is complete, tube order should be rechecked.

Once an autoradiograph is produced, double scoring is inexpensive and feasible. In our laboratory, at least two people independently score each autorad and enter the data into computer files. The data sets are then compared in an Excel spreadsheet, using a function that detects nonmatching cell entries and enters a 0 or a 1 in a new matrix, depending on whether the cell contents match. Discrepancies are checked,

and rerun if necessary. For unbiased scoring, all ink marks from previous scoring are removed from autorads with alcohol and each individual independently labels the ladder, sample IDs, and allele sizes. Since each locus has its own particular morphology, we provide scorers with a band summary sheet for each locus, drawn up after discussion of band characteristics at that locus (which band is to be scored, location of faint shadow bands, etc.).

As a check on the success of the above measures to eliminate error, we choose a random sample of individuals and primers for repeat runs. We also rerun cases where one locus gives results that are not consistent with the other loci for that individual. For example, if scoring of ten loci consistently assigns parentage to one individual, while scoring of an eleventh locus excludes that individual as a parent, we rerun that locus.

Analyzing Microsatellite Data

Microsatellites can provide a great deal of information on genetic relatedness and parentage assignment. Queller and Goodnight (1989) discuss estimation of Grafen's relatedness statistic (Grafen, 1985). This technique has been implemented in a user-friendly Macintosh program, Relatedness 4.2b (Goodnight and Queller 1994; currently obtainable, with its instruction manual, from our website, http:/www.rice.edu/wasps). This program allows the calculation of genetic relatedness in groups, between specific individuals, and jackknifes for error estimation. It also allows for population structure and inbreeding estimates.

If there are few candidate parents and progeny, genotypes can be lined up and compared by hand to assign the progeny to the appropriate parents. It is also desirable to estimate the power of the analysis by determining the frequency of a match due to chance alone. For more complex situations, maximum likelihood techniques that assign individuals to the most likely parents may be appropriate. Goodnight and Queller (*in preparation*) are working on such programs.

FINDING YOUR OWN MICROSATELLITES

Though somewhat tedious and time-consuming, there are no difficult steps involved in identifying microsatellites in a new species (Queller et al., 1993). Here, we follow the general approach of Hughes and Queller (1993) with some modifications. Though the protocols are very standard to molecular biologists, we nevertheless provide many of them because they are unlikely to be so familiar to behavioral ecologists. For the behavioral ecologist, inexperienced in molecular biology, the steps necessary for generating new microsatellites will involve learning quite a number of unfamiliar techniques. It is unlikely that all the steps will proceed smoothly at the first attempt, so it is important to not simply follow the protocols but to understand exactly what is going on at each step and know what is most likely to fail. The biochemical background need not be acquired before embarking on the project, but can be assimilated as you go by reading reference texts and by comparing

protocols and browsing frequently through the newsgroup bionet.molbio.methds-reagnts. As you do this, you will see that there is a strong historical component to many protocols and newer ones may omit many steps or ingredients. Some of these shortcuts may be worth the time they save, others not. You should also watch the advertisements in the front of publications like *Nature* for new proprietary products that may greatly facilitate cloning and probing, though I would wait for assurance that they work from newsgroup members.

The general strategy for identifying microsatellites involves plating out on nylon membranes a partial genomic library of DNA fragments 300–600 base pairs long, inserted into plasmids that have been introduced into *Escherichia coli* cells. Replica membranes are made and probed with oligonucleotides such as $(AAT)_{10}$. Positive clones are picked up and grown in culture. Then the plasmid DNA is digested with a restriction enzyme to cut out the insert. This digest is run on an agarose gel. A Southern transfer of that gel is probed with the oligonucleotides to verify the positives and to estimate insert size. Positive clones are sequenced. Those with at least eight uninterrupted, identical, trinucleotide repeats are frequently variable in Hymenoptera (Hughes and Queller 1993; J. Strassmann et al., *unpublished data*). For these, PCR primer pairs are designed that amplify the repeat region. PCR conditions are optimized on nonradioactive samples run out on agarose gels and stained with ethidium bromide. A sample of 30 or 40 unrelated individuals is screened using radiolabeled PCR products run out on denaturing polyacrylamide gels to determine the degree of heterozygosity for the locus. The best loci can have 20 or more alleles, with the most common one occurring at a frequency of less than 0.20.

Much of the critical work is done by enzymes that act in very specific ways only under specific conditions, so any steps involving enzymes should be carried out precisely. If a procedure fails completely, bad enzyme or poor conditions for the enzyme are likely suspects. Although it is a rare occurrence, inactive or underactive enzymes have been supplied by all the companies with which we have dealt. The enzyme has usually been replaced at no cost. It is therefore often worth the extra effort to do controls that ascertain enzyme activity and prudent to remain with a supplier with whom you have had good experience.

Another general practice that can facilitate molecular work is to keep track of how much DNA you have by consistently running lambda standards of several known amounts next to your experimental DNA on agarose gels. We do this almost to the exclusion of other means of DNA quantification because it is easy, cheap, and also gives size information.

Extra steps in the overall process that provide additional information are often worth the trouble. For example, we do not begin sequencing immediately after picking up positives from a library. First, we cut out the inserts from the plasmid, run them on a gel, photograph it, and then probe a Southern transfer of that gel. This procedure verifies the presence of a microsatellite (if the Southern from that clone also yields a positive), gives an indication of the size of the insert, and reveals whether or not there were multiple inserts in one clone (they will be cut apart by the restriction enzyme and so show up as multiple bands in the lane). This information allows us to prioritize the clone candidates for sequencing.

Since many of the early steps in microsatellite development can be carried out on a large number of library membranes almost as quickly as on a few, we favor making and simultaneously probing very large libraries. The positives can then be picked up and frozen for eventual sequencing. Excess ligation reactions can also be frozen away to save time in making a larger library, if this proves necessary later. The libraries that generated most of our microsatellites came from two species of wasps probed simultaneously on a total of 30 large (132 mm) membranes averaging 1500 (range 1189–1772) colonies per membrane from one library (12 membranes total), and 656 (range 415–876) colonies per membrane from the other library (18 membranes total) (J. Strassmann et al., *unpublished data*). Probing these nearly 30,000 colonies required only an hour or two per day more than it would have taken to probe just one or two membranes. If too few colonies have been probed to generate enough useful microsatellites, the entire procedure must be repeated. This is time-consuming and frustrating. It is better to have more positives than you are likely to need frozen away for use. This is essential because many positives will have repeats that are stuck to the vector (so a flanking primer cannot be designed), have too many errors, or are too short.

In our laboratory, approximately six months of full-time effort was required to generate a minimum of 10 highly polymorphic trinucleotide microsatellite loci for an entirely new species in which microsatellites are not particularly rare. Though this represents a substantial time investment, the questions of parentage and kinship that can be addressed with microsatellites make it worthwhile.

Isolating High Molecular Weight DNA for Library Construction

When constructing a library, it is necessary to isolate high molecular weight DNA to ensure that the fragments for insertion have appropriate overhanging ends. There are many possible ways of obtaining high molecular weight DNA. One of the most important steps is the initial breakdown of tissues: overly vigorous grinding shears DNA. Grinding in liquid nitrogen is a simple and efficient means of isolating moderate amounts of high-quality genomic DNA from a variety of tissues (Strassmann.1, Strassmann.6).

Restriction Digests and Size Selection

The size of DNA to be inserted into the plasmid should not be longer than can be sequenced using forward and reverse primers that complement sites in the vector. However, if inserted DNA is too short, the chance is increased that repeats will fall so close to the edge of the inserted DNA that flanking PCR primers cannot be designed. Genomic DNA inserts between 300 and 600 base pairs are a good compromise. Although it may be difficult to obtain a readable sequence of 600 bases without the use of automatic sequencing, the use of forward and reverse primers will easily reveal any microsatellites in an insert of this size. If necessary, an internal primer can be designed to accurately obtain the rest of the sequence on the opposite side of the microsatellite. This primer can then be used as one of the two PCR primers, if a compatible primer can be designed on the other side of the repeat.

To cut up the genomic DNA for insertion into plasmids, we typically use SAU3A, a restriction enzyme with a four base pair recognition sequence that yields a large number of small fragments in the size range of interest. This is a very good choice of enzyme since it produces fragments with ends compatible with the sticky ends left by BamHI, a six cutter commonly chosen to cut open the plasmid. After digesting the high molecular weight DNA to completion, select the appropriate size fragments by running the DNA out on an agarose gel (Strassmann.7). If microsatellites are likely to be rare in your species, you may enrich the genomic DNA for microsatellites using techniques discussed by Fleischer in this volume. Bear in mind, however, that the extra steps entailed rely on the success of several enzymes (Kandpal et al., 1994).

Plasmid Selection, Cutting, and Dephosphorylation

The plasmid is processed at the same time that the genomic DNA is digested (Strassmann.7). We have used both pUC19 and pBluescript SK+ plasmids with success. It is important to identify three restriction sites within the polylinker region of the chosen plasmid. The restriction sites on either side of the insertion site are required when the insert is cut out in order to verify the presence of a microsatellite and determine the insert's length. Since these latter two restriction enzymes must work simultaneously, it is critical that they function well under the same buffer and temperature conditions. In pUC19 we used HindIII and EcoRI. In SURE cells we used HindIII and SACI. Details of restriction enzyme conditions and compatibility can be found in suppliers' catalogues. The middle site must create sticky ends in the plasmid polylinker complementary to the genomic DNA overhangs so that the genomic DNA and the plasmid are capable of base pairing during the ligation reaction. We use BamHI to cut the plasmid and produce sticky ends matching those of the genomic DNA cut with SAU3A. The cut end of the plasmid must then be dephosphorylated so that when ligase is added in the next step, the plasmid will not reanneal before receiving the insert. Since dephosphorylation sometimes fails, we use two different phosphatases in succession, and ensure that they are fresh. The success of the dephosphorylation can be checked by adding ligase in the absence of any insert before proceeding. Alternatively, this check can be carried out as a control to the simultaneous checking of insert to vector ratios.

After cutting and dephosphorylation, the plasmids are run through an agarose gel and the linear plasmid band is cut out of the gel and purified. If the digestion of the plasmid was complete, further purification by running the plasmid through a gel and selecting the linear band is probably unnecessary.

Ligation of Insert into the Plasmid

This is a critical step that depends on the success of the restriction enzyme digestion, on the dephosphorylation of the plasmid, and on the quantification of vector and insert DNA (Strassmann.8). Since the plasmid is about 2900 bases long and the insert averages 400 bases, a 1:1 sticky end ratio (i.e., one insert for each plasmid) would require a DNA concentration with approximately seven times as much vector

as insert. However, it is not always possible to quantify vector and insert DNA very accurately, so we usually try a number of different ligation ratios and have generally had the most success with two to four times as many inserts as plasmids (Strassmann.8). Another check on the ligase may be done by setting up a tube of plasmid that has been cut but not dephosphorylated. The success of the ligation cannot easily be determined until transformation takes place.

Transformation and Plating Out of Cells

Competent cells can either be purchased or made by the researcher (Hanahan, 1983; Sambrook et al., 1989:1.76–1.81) and should be frozen in 100- or 200-μl aliquots at −70°C in 1.5-ml microfuge tubes. We usually make our own competent cells, but it is probably not worth the effort for someone who does not do this routinely. Catching the growing cells at the optimum point for competency can be a little tricky. We recommend a healthy, simple strain such as DH5alpha (from GibcoBRL) or XL1Blue (from Stratagene Inc., La Jolla, CA). We have also used SURE (from Stratagene) cells that the manufacturers claim reduce recombination. However, these cells were harder to grow, making many subsequent steps difficult. In any case, we have not found recombination to be a problem for microsatellites. Transformation itself is quite simple (Strassmann.9).

To avoid duplication, transformed cells should be grown for no more than 45 minutes and should then be plated out. Since this is a test of the initial ligation reaction, plating can be done on small (100 mm) agar plates to which antibiotic has been added. The antibiotic used must, of course, be one to which the chosen plasmid is resistant. We usually plate out two or three replicates of each ligation reaction, using 50 μl of transformed cells (Strassmann.10). The number of colonies on each plate should then be counted. There should be very few on the no-insert control plate, and many on one or more of the test ligations.

Next, plate out the transformations that yielded the most colonies on large (132 mm) nylon membranes (Hybond, Amersham) for the library (Strassmann.11). This should be done as soon as possible after selecting the successful ligation reactions and determining the quantity needed to result in about 1000 evenly distributed colonies per membrane (Strassmann.10). Plating out on large nylon membranes rather than directly on agar facilitates producing clear replicate membranes for probing, as well as making it easier to identify and pick up specific colonies.

After growing the initial plates for 10–14 hours, the colonies should appear as small, discrete raised bumps. At this point, make replicate filters for probing (Strassmann.12). It is important to make holes through the replicate membrane and the original so that the two can be lined up later. After the copy is made, allow colonies on both membranes to grow another 3 or 4 hours until they again appear as raised bumps. At this time, a reserve copy of the library can be made and stored at −70°C using the technique of Dreyer et al. (1991) (Strassmann.12).

Probing, Selection, and Sequencing of Positives

Once the library is made and a replicate is denatured, we probe it with a trinucleotide repeat containing 10 repeats that has been labeled with ^{32}P using the

enzyme terminal deoxynucleotidyl transferase (TdT) (Strassmann.13; Rosenberg et al., 1990). Membranes can be probed with more than one repeat at a time, if their melting temperatures are similar.

At the point of picking up positives from plates, the researcher has usually accumulated a number of autoradiographs probed with different repeat oligonucleotides (Strassmann.14). If it is not possible to pick up all positives on the same day, it is best to pick all positives from a subset of filters. We pick up positives by touching a toothpick to the clone and dropping it into a culture tube containing about 3 ml of LB/ampicillin medium. Keep track of what possible repeats are on each autorad so that you probe the Southern with the appropriate oligonucleotides.

Unfortunately, we have not found any correlation between intensity of the positive signal and length of the microsatellite repeat. A very dark positive on both library membrane and Southern is no more likely to yield a long repeat than is a less dark positive. Variability in signal intensity probably results from factors other than repeat length influencing binding of the probe to the clone.

Sequence inserts of the best positives either by hand, using the protocols provided with Sequenase, or with an automated sequencer. Much time in this step can be saved by contracting the sequencing out to a reliable university sequencing facility.

Designing PCR Primers, Optimizing, Checking for Polymorphisms

Once you have identified repeat regions that have eight or more uninterrupted, identical trinucleotide repeats, or somewhat more not quite perfect repeats, it is time to design primers that amplify across the repeat. Though it can be done by hand, this is probably best done by entering the entire sequence into a program like Oligo or MacVector and using the program to help design the primers. The primers should have very similar lengths and melting temperatures and should be noncomplementary so they do not anneal to one another or form hairpin loops. They should produce products of 100–200 base pairs in length. Every effort should be made to find primer pairs that produce products in the shorter end of this range, since they will run faster on a gel and will be easier to score, though having nonoverlapping sizes is advantageous for multiplexing on the same gel. Longer products can be run on a 4% acrylamide gel, though it is somewhat more difficult to handle than the standard 6% gel.

The synthesized primers are then used in a polymerase chain reaction at the predicted annealing temperature and run out on agarose gels to test for the presence of correctly sized product. If the run fails to yield product at this temperature, the PCR should be repeated at higher and lower temperatures in 2.5 degree steps until a product of the correct size is obtained. In attempts to obtain product, it is also sometimes helpful to vary magnesium chloride concentrations. To assess the variability (i.e., the number of alleles) at the locus in question, the primers should then be tested on a sample of 30 unrelated individuals in a radiolabeled PCR reaction that is run out on a denaturing gel. We have found that over 80% of the loci chosen following the above guidelines have four or more alleles and are useful for studies of kinship.

Development of a good set of polymorphic trinucleotide microsatellite loci for a species provides a powerful tool for addressing many exciting questions involving kinship and parentage. The answers to such questions are then easily obtained, since the processes involved in genotyping individuals using microsatellite primers are simple enough that new students can often produce usable data within a week of entering the laboratory (Strassmann.1 to .5).

ACKNOWLEDGMENTS

We thank Madhusudan Choudhary and Colin Hughes for helping us work out microsatellite methods. We thank Elisabeth Arévalo for many insightful discussions of specific protocol steps and Jeanne Zeh for greatly improving the clarity of the manuscript. We also thank David Zeh and Jeremy Field for comments on the manuscript. We thank the W. M. Keck Center for Computational Biology for the use of shared equipment. Our own work on microsatellites has been supported by the National Science Foundation (NSF IBN-9210051, NSF BSR-9021514).

REFERENCES

Arévalo E, Holder D, Derr J, Bhebhe E, Linn R, Ruvuna F, Davis S, Taylor J (1994): Caprine microsatellite dinucleotide repeat polymorphisms at the SR-CRSP-1, SR-CRSP-2, SR-CRSP-3, SR-CRSP-4, and SR-CRSP-5 loci. *Anim Genet* 25:645.

Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J, Struhl K (1990): *Current protocols in molecular biology.* New York: Wiley.

Beckmann J, Weber J (1992): Survey of human and rat microsatellites. *Genomics* 12:627–631.

Bender WP, Spierer D, Hogness S (1983): Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in *Drosophila melanogaster. J Mol Biol* 168:17–33.

Bhebhe E, Kogi J, Holder D, Arévalo E, Derr J, Linn R, Ruvuna F, Davis S, Taylor J (1994): Caprine microsatellite dinucleotide repeat polymorphisms at the SR-CRSP-6, SR-CRSP-7, SR-CRSP-8, SR-CRSP-9, and SR-CRSP-10 loci. *Anim Genet* 25:646.

Birkhead T, Møller AP (1992): *Sperm competition in birds.* London: Academic Press.

Bruford MW, Hanotte O, Brookfield JF, Burke T (1992): Single-locus and multilocus DNA fingerprinting. In: Hoelzel, A.R. (ed). *Molecular genetic analysis of populations.* Oxford: Oxford University Press, pp 225–269.

Cornall R, Aitman T, Hearne C, Todd J (1991): The generation of a library of PCR-analyzed microsatellite variants for genetic mapping of the mouse genome. *Genomics* 10:874–881.

Dreyer K, Opalka B, Schulte-Holthausen H (1991): An improved method for long-term storage of colony filters. *Trends Genet* 7:178.

Edwards A, Hammond H, Jin L, Caskey C, Chakraborty R (1992): Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241–253.

Ellegren H (1992): Polymerase-chain-reaction (PCR) analysis of microsatellites—a new approach to studies of genetic relationships in birds. *Auk* 109:886–895.

Estoup A, Presa P, Kreig F, Vaiman D, Guyomard R (1993): $(CT)_n$ and $(GT)_n$ microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). *Heredity* 71:488–496.

Gadagkar R (1985): Evolution of insect sociality—a review of some attempts to test modern theories. *Proc Indian Acad Sci* 94:309–324.

Goodnight KF, Queller DC (1994): *Relatedness* 4.2b. Pascal Program for Macintosh computers. Available at http://www.rice.edu/wasps.

Grafen A (1985): A geometric view of relatedness. *Oxford Surveys of Evolutionary Biology* 2:28–89.

Gustincich S, Manfioletti G, Del Sal G, Schneider C, Carninci P (1991): A fast method for high-quality genomic DNA extraction from whole human blood. *Biotechniques* 11:298–301.

Hanahan D (1983): Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166:557–580.

Hannotte O, Bruford MW, Burke T (1992): Multilocus DNA fingerprints in gallinaceous birds: general approach and problems. *Heredity* 68:481–494.

Hanotte O, Zanon C, Pugh A, Greig C, Dixon A, Burk T (1994): Isolation and characterization of microsatellite loci in a passerine bird: the reed bunting *Emberiza schoeniclus*. *Mol Ecol* 3:529–530.

Hughes C, Queller DC (1993): Detection of highly polymorphic microsatellite loci in a species with little allozyme polymorphism. *Mol Ecol* 2:131–137.

Hughes CR, Queller DC, Strassmann JE, Davis SK (1993): Relatedness and altruism in *Polistes* wasps. *Behav Ecol* 4:128–137.

Kandpal R, Kandpal G, Weissman S (1994): Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region specific markers. *Proc Natl Acad Sci USA* 91:88–92.

Litt M, Luty J (1989): 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.

Moore S, Sargeant L, King T, Mattick J, Georges M, Hetzel D (1991): The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics* 10:654–660.

Morin P, Wallis J, Moore J, Woodruff D (1994): Paternity exclusion in a community of wild chimpanzees using hypervariable simple sequence repeats. *Mol Ecol* 3:469–478.

Morin P, Woodruff D (1992): Paternity exclusion using multiple hypervariable microsatellite loci amplified from nuclear DNA of hair cells. In: Martin R, Dixson A, Wickings E (eds). *Paternity in primates: genetic tests and theories*. Basel: Karger, pp 63–81.

Paetkau D, Strobeck C (1994): Microsatellite analysis of genetic variation in black bear populations. *Mol Ecol* 3:489–495.

Peters JM, Queller DC, Strassmann JE, Solís CR (1995): Maternity assignment and queen replacement in a social wasp. *Proc R Soc Lond B* 260:7–12.

Queller DC, Goodnight KF (1989): Estimating relatedness using genetic markers. *Evolution* 43:258–275.

Queller DC, Strassmann JE, Hughes CR (1988): Genetic relatedness in colonies of tropical wasps with multiple queens. *Science* 242:1155–1157.

Queller DC, Strassmann JE, Hughes CR (1993): Microsatellites and kinship. *Trends Ecol Evol* 8:285–288.

Rosenberg HF, Ackerman SJ, Tenen DG (1990): An alternative method for labeling oligonucleotide probes for screening cDNA libraries. *Biotechniques* 8:384.

Sambrook J, Fritsch EF, Maniatis T (1989): *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Smeets HB, Ropers H, Wieringa B (1989): Use of variable simple sequence motifs as genetic markers: application to the study of myotomic dystrophy. *Hum Genet* 83:245–251.

Stallings R, Ford A, Nelson D, Torney D, Hildebrand C, Moyzis R (1991): Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics* 10:807–815.

Strassmann JE, Hughes CR, Queller DC, Turillazzi S, Cervo R, Davis SK, Goodnight KF (1989): Genetic relatedness in primitively eusocial wasps. *Nature* 342:268–270.

Tautz D (1989): Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17:6463–6471.

Thomas SM, Moreno RF, Tilzer LL (1989): DNA extraction with organic solvents in gel barrier tubes. *Nucleic Acids Res* 13:5411.

Trentmann SM, Knaap E van der, Kende H (1995): Alternatives to ^{35}S as a label for the differential display of eukaryotic messenger RNA. *Science* 267:1186–1187.

Weber J (1990): Informativeness of human $(dC\text{-}dA)_n$. $(dG\text{-}dT)_n$ polymorphisms. *Genomics* 7:524–530.

Weber J, May P (1989): Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44:388–396.

Westneat DF (1987): Extra-pair fertilizations in a predominantly monogamous bird: genetic evidence. *Anim Behav* 35:877–886.