CHAPTER 4

Characterization of Genetic Structure and Genealogies Using RAPD-PCR Markers: A Random Primer for the Novice and Nervous

RICHARD K. GROSBERG

Center for Population Biology, Section of Evolution and Ecology, University of California–Davis, Davis, California 95616

DONALD R. LEVITAN

Department of Biological Sciences, Florida State University, Tallahassee, Florida 32306-2403

BRENDA B. CAMERON

Center for Population Biology, University of California, Davis, California 95616

CONTENTS

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

SYNOPSIS

The genetic and genealogical structure of natural populations affords both a reflection of evolutionary history, as well as a set of constraints and possibilities for future evolutionary change. Evolutionary biologists, armed with a variety of recently developed, high-resolution genetic markers, now possess unprecedented capacity to portray structure and reconstruct genealogies. Unfortunately, fashion and novelty, rather than utility, often dictate the popularity of a given technique. The truth is that no single type of marker yet offers the capacity to reveal cheaply and powerfully genetic structure and genealogical relationships across a full spectrum of temporal and spatial scales. In this chapter, we evaluate the strengths (cheap, fast, easy, and polymorphic) and weaknesses (repeatability, comigration of bands, detectability, dominance, and nonmendelian inheritance) of randomly amplified polymorphic DNA (RAPD) markers for depicting population structure and genealogies. Overall, RAPD markers, because they are anonymous and expressed as dominant alleles potentially concealing substantial amounts of hidden variation, generally rate poorly

for their ability to portray breeding systems and phylogenetic relationships (especially above the level of populations). However, the available data suggest that in many taxa RAPD-PCR can easily generate hundreds of highly polymorphic, independent markers, the majority of which behave as mendelian alleles. Many of the markers may exhibit substantial variation in evolutionary rates. As such, RAPD markers offer evolutionary biologists still unmatched opportunities to evaluate intraspecific genetic structure and genealogical relationships in large samples that exhibit structure across diverse spatial scales.

INTRODUCTION

Evolutionary ecology rests on a foundation of understanding how variation in morphology, behavior, and resource allocation relates to fitness in different environments. The first step in building this foundation involves measuring phenotypic traits and analyzing the contributions that environmental, ontogenetic, and genetic processes make to the expression of these traits. The next, and still more challenging, step entails portraying the spatial and temporal scales over which environmental and genetic structure vary, so that character evolution can be set into a relevant selective context (Bell, 1992). The last step—one that has proved the most technically elusive—requires establishing how phenotypic variation correlates with reproductive success.

The measurement of reproductive success, achieved through either sexual reproduction or asexual propagation, fundamentally relies on the ability to assign offspring to parents. When females brood offspring or exhibit some form of parental care, establishing maternity can be relatively easy, even without genetic markers (but see Westneat et al., 1987). However, when females (or males) show no parental care, or communally care for young, direct assignment of parentage becomes substantially more difficult. In the extreme, when male and female gametes are released into the environment and motile propagules can disperse extensively (as in many sessile invertebrates and plants), assigning parentage and inferring population structure by direct observation are essentially impossible.

The development of allozyme markers in the 1960s opened the door to genetic characterization of population structure and reproductive success. Numerous individuals can be analyzed through allozyme electrophoresis at relatively low cost and with few technical impediments: that is one of the reasons why a number of us have stuck by what many consider an antiquated technique. However, 30 years of experience have shown that, in many taxa, the number of polymorphic loci and alleles per locus is often too low to characterize all but larger scale genetic patterns or to assign parentage with high confidence, especially when the parental pool is large or inbred. In addition, destructive sampling is usually obligatory, precluding analysis of reproductive success and genetic structure beyond a single generation.

Over the last decade, the development of new molecular tools that provide access to highly variable DNA markers, especially mini- and microsatellite VNTR (variable number of tandom repeat) loci (Burke et al., 1991) and loci of the major histocompatibility complex, has made high-resolution analysis of genealogies and

genetic structure possible, but still not easy or risk free (reviewed in Avise, 1994). Most of these methods involve complex and expensive protocols and demand expertise in the tools of molecular genetics; nearly all are time-consuming and expensive. Some of these techniques, such as single-locus and multilocus DNA fingerprinting, depend on the availability of large amounts of undegraded DNA (Weatherhead and Montgomery 1991) and may require destructive sampling. For population biologists, the recent development of microsatellite markers offers several distinct advantages over multilocus fingerprinting (Queller et al., 1993; also see Strassmann et al., Chapter 8, and Fleischer, Chapter 7, in this volume).

As long as cost is an object, time a constraint, or technical ability and motivation limiting, population ecologists must always face a trade-off between gathering high-resolution genetic information and sampling large numbers of individuals (Lessa and Applebaum, 1993). This trade-off is especially intrusive where associations among particular traits (or suites of traits), selective regimes, and fitness may be weak, but nonetheless highly significant, both in a statistical and evolutionary sense. The accurate portrayal of such associations necessarily involves large samples. Thus, although mini- and microsatellite loci, and nuclear introns, offer access to highly polymorphic genetic markers, in terms of both time and money, their expense highlights, rather than ameliorates, the trade-off between genetic resolution and large sample sizes.

The development five years ago of randomly amplified polymorphic DNA (RAPD) markers by Williams et al. (1990), along with arbitrarily primed (AP) markers by Welsh and McClelland (1990) and DAF (DNA amplification fingerprinting) by Caetono-Anollés et al. (1991), appeared to provide an escape from the choice between *Scylla* (high genetic resolution) and *Charybdis* (large sample size) for molecular ecologists (Hedrick, 1992): hundreds of apparently neutral polymorphic markers throughout the genome can be generated quickly (Bowditch et al., 1993); large numbers of individuals can be processed relatively inexpensively and without any prior DNA sequence information; radioisotopes are unnecessary; and—because the polymerase chain reaction is used to amplify the RAPD markers—little (but nevertheless undegraded) DNA is needed, and nondestructive sampling is feasible (Hadrys et al., 1992; Black, 1993). Moreover, different loci exhibit different degrees of polymorphism, hence RAPD markers should be useful for deciphering population structure across a range of spatial and genealogical levels of resolution.

A number of technical and analytical problems soon tempered the initial enthusiasm for RAPD markers (reviewed by Hadrys et al., 1992; Black, 1993), leading some workers to question their utility in population ecology and phylogeny reconstruction (e.g., Black, 1993). Some of the negative reactions toward RAPD-PCR are undoubtedly justified; some, however, are hyperbole. In this chapter, we critically assess the strengths and weaknesses of RAPD markers in the context of characterizing parentage and genetic structure in populations of sessile organisms. We avoid discussion of the use of RAPD markers for the reconstruction of higher level spatial and genealogical/phylogenetic relationships, in part because our own research concerns intraspecific patterns and processes, but also because we, like

others (Clark and Lanigan, 1993; Smith et al., 1994), doubt that these markers can generally yield useful information about transpecific relationships. Our approach is first to outline the basic protocols for generating, scoring, and interpreting RAPD markers. We then reexamine some of the problems encountered by us, and many others, in using RAPD markers and offer remedies to some of these problems. Finally, we evaluate the prospects of RAPD-PCR for characterizing population-level parameters in sessile and sedentary organisms with motile propagules by chronicling our own experiences with the technique.

WHAT ARE RAPDs?

Introduction

Randomly amplified polymorphic DNA markers consist of relatively short DNA fragments (about 200–2000 base pairs long), amplified via PCR by small (usually 10 bases in length) arbitrary (with a G + C content > 50%) primers. Typically, primers are used singly and must anneal to priming sites in opposite orientations in order for amplification to occur. Following convention, we term the pair of inverted priming sites, plus the intervening sequence of nucleotides, a RAPD locus, and the amplified product from a particular locus a RAPD marker. The resulting amplification product(s) can be size-separated electrophoretically on an agarose gel and visualized by ethidium bromide staining.

Depending on the number of inverted complementary priming sites in an individual's genome, and the lengths of the intervening DNA sequence, a given primer may amplify from 0 to 30 products. Hundreds of RAPD primers are commercially available (from Operon Technologies Inc., Alameda, CA), and when several primers are used independently, hundreds of polymorphic markers potentially can be identified. Nevertheless, different species (and even populations, e.g., Dawson et al., 1993; Huff et al., 1993) exhibit dramatically different degrees of polymorphism, both in terms of the proportion of RAPD loci that vary and the number of loci that amplify (Caetano-Anollés et al., 1991). Reasons for this variation in levels of polymorphism remain unclear.

Because RAPD-PCR primers are not designed to amplify a specific target sequence, the amplified loci are anonymous and presumably scattered throughout the genome (Williams et al., 1990, 1991; Tinker et al., 1993). RAPD loci carry the advantages that (1) there is no need for prior nucleotide sequence data for the taxa under study and (2) many of the loci may be acting as neutral markers. On the other hand, in the absence of information concerning the nature of the DNA being amplified by RAPD-PCR, there is no guarantee that any single primer will produce usable markers or that the variation is, in fact, neutral.

Most workers assume that RAPD loci amplify only if the priming sites perfectly match the oligonucleotide used (Hadrys et al., 1992), based on Williams et al. (1990), who showed that single nucleotide substitutions at a priming site can preclude amplification in tests with human, soybean, corn, and yeast DNA. However,

annealing temperatures in RAPD-PCR are typically quite low (35–45°C) and primers are relatively short (8–15 bp); consequently, primer–template complementarity could be less than perfect and still result in amplification of template DNA. Fortunately, it is possible to choose PCR conditions that minimize imperfect priming (see below).

Scoring RAPD Polymorphisms

The usual gel assay for genetic variation at RAPD loci is the presence or absence of a band of a specific molecular weight, amplified by a given primer, in an individual sample. Bands (i.e., putative loci) that appear in all individuals are considered monomorphic; those present in some individuals but not others are polymorphic in the sampled population. Bands of different size are usually considered to represent independent loci and are scored as independent traits. This cannot be the case in every instance. For example, both Martin et al. (1991) and Smith et al. (1994) showed through Southern blot hybridizations that different bands amplified by a single primer contain homologous sequences, suggesting that different bands are not necessarily independent traits. Such bands may represent homologous alleles at the same locus or gene duplications. In addition, there should be some concern that bands amplified by different primers may be homologous, once again due to duplications or the presence of multiple priming sites within putative loci.

In some cases judgment of the presence or absence of a band has a subjective component, as band intensity can quantitatively vary over a wide range (see "Problems with RAPD-PCR and Some Remedies: Variation in Band Intensity and Limits to Detection"). Although band intensity itself can be heritable (Hunt and Page, 1992; Heun and Helentjaris, 1993; Levitan and Grosberg, 1993), nonheritable factors, including PCR conditions, DNA concentrations, and associations with other loci (Hunt and Page, 1992; Heun and Helentjaris, 1993; Wilkerson et al., 1993; Smith et al., 1994), may cause considerable brightness variation (reviewed in Black, 1993).

As a matter of convenience, most workers assume that when a given primer amplifies bands of the same apparent molecular weight, the bands represent the same allelic state (the dominant allele); conversely, the absence of a band represents a single alternative allelic state (the null, or recessive, allele). Because RAPD bands are usually expressed as phenotypically dominant markers in diploids, it is generally impossible to distinguish between the genotypes of individuals homozygous and heterozygous for the dominant allele (see "Problems with RAPD-PCR and Some Remedies: Dominance"). Segregation analysis, however, can be performed in organisms such as hymenopterans, in which males are haploid (Fondrk et al., 1993; Shoemaker et al., 1994), in those plants with a macrogametophytic stage (Bucci and Menozzi, 1993), or in individual gametes.

Sources of Polymorphism and Amplification Artifacts

Not all variation at RAPD loci should be detectable simply in terms of the presence or absence of a given band, and not all mutations will necessarily lead to a shift

from presence to absence (or *vice versa*). Polymorphism, detectable or not, can arise in numerous ways. For example, an amplified band can be turned into a null in one of at least three ways: (1) either one or both priming sites can be lost due to mutations that reduce or eliminate priming; (2) an insertion may occur between priming sites that is so long that the fragment between sites is no longer amplifiable by PCR; or (3) sequences may be rearranged so that the priming sites are no longer complementary. A null can be transformed into an amplifiable marker when a mutation occurs at one or both inverted priming sites that makes those sites recognizable by a complementary primer, a deletion occurs between sites that shortens the intervening fragment so that it can be amplified by PCR, or by sequence rearrangements. Finally, an insertion or deletion between inverted priming sites of an amplifiable locus may alter the length of the amplified region between the priming sites, producing a length polymorphism of codominant alleles, rather than a presence/absence polymorphism.

Classes of mutations causing a gain or loss of a RAPD band appear straightforward to detect, but the mutational events themselves will be difficult to distinguish, at least without sequencing a locus or Southern blotting using amplified product as a genomic probe (Hadrys et al., 1992; see Smith et al., 1994 for an exemplary study). Mutations that produce length polymorphisms are far more difficult to detect without formal genetic analysis, largely because most RAPD primers amplify multiple loci, most of which are assumed to produce markers of different molecular weights (but see Hunt and Page, 1992, and "Problems with RAPD-PCR and Some Remedies: Comigration of Bands"). In practice, as with multilocus DNA fingerprinting, it is often impossible to distinguish bands of different molecular weights representing alleles at the same locus from bands representing alleles at nonhomologous loci (Lynch, 1988). For this reason, most workers simply ignore length polymorphisms ascribable to a single locus and score presence *versus* absence. Nevertheless, thanks to the haplo-diploid genetics of Hymenoptera, Hunt and Page (1992), as well as Shoemaker et al. (1994), identified a number of codominant alleles from single loci (also see Williams et al., 1990; Martin et al., 1991; Roehrdanz et al., 1993).

There are other potential problems in scoring RAPD variation solely in terms of presence or absence of a band. Southern blot analysis (using probes derived from amplified RAPD bands) of amplification products demonstrates that although some comigrating bands in different conspecifics are, indeed, homologous (Hadrys et al., 1992; Smith et al., 1994), some clearly are not (Smith et al., 1994). Overall, few studies have undertaken such time-consuming steps; hence, for the time being, we simply lack the data to judge whether this form of cryptic variation is widespread within and among populations.

In principle, therefore, conventional presence/absence scoring of RAPD markers likely underestimates levels of genetic variation (Apostol et al., 1993; Rossetto et al., 1995) and may provide, at best, a hazy reflection of underlying genotypic frequencies. For these reasons alone, our view is that RAPD markers should be scored as biallelic systems only when appropriate sequencing, blotting, or inheritance studies support such an assumption. Whether each RAPD marker should be scored as an independent character remains an open question.

There is an additional nongenetic concern when using RAPD markers to analyze

population genetic parameters and phylogenies: presence/absence variation may be due to amplification artifacts (Ellsworth et al., 1993; "Problems with RAPD-PCR and Some Remedies: Repeatability of Banding Patterns Within and Among Individuals") or the insensitivity of detection methods ("Problems with RAPD-PCR and Some Remedies: Variation in Band Intensity and Limits to Detection"). These amplification and detection artifacts may be difficult to distinguish, in part because they may often have similar causes (including imperfect primer–template annealing, interlocus interactions, and effects of copy number). The existence of such variation, if pervasive, can confound parent–offspring relationships and lead to overestimation of levels of allelic diversity and genetic distances. However, unless particular populations or families express nongenetic variation differentially, this "noise" should be of similar magnitude in different populations or families.

Population Genetic, Genealogical, and Phylogenetic Implications of Dominance and Cryptic Variation: A Preliminary Overview

If a population is in Hardy–Weinberg equilibrium, and allelic variation at a locus comprises two, and only two, allelic states, then the inability to distinguish dominant homozygotes from heterozygotes does not pose a serious barrier to the analysis of parentage and population structure. This is because the frequency of the recessive allele, particularly if it is common (>0.1) and the sample size is large, can be reckoned from the frequency of null/null homozygotes (Lynch and Milligan, 1994). Allelic and genotypic frequencies can then be analyzed by conventional software to calculate F-statistics and their analogues [e.g., BIOSYS-1 (Swofford and Selander, 1989), Genetic Data Analysis (Lewis and Zaykin, 1996), or AMOVA (Excoffier et al., 1992)] or relatedness (e.g., Reeve et al., 1992). If, however, there is more than one null allele, then it will be risky to calculate allelic and genotypic frequencies. To the extent that cryptic variation represents a general phenomenon, the implications for RAPD-based population genetic studies are clear: indices of genetic structure, gene flow, and mating patterns based on RAPD markers scored as biallelic character states in Hardy–Weinberg equilibrium should be interpreted cautiously.

The potential for cryptic variation in the null allelic class, and comigration of nonhomologous alleles, should also temper enthusiasm for using RAPD markers for transpecific comparisons of population genetic parameters, and even for intraspecific comparisons of variation exhibited by RAPD markers to other nuclear or mitochondrial markers (Liu and Furnier, 1993; Haig et al., 1994; but see Peakall et al., 1995). Applications of cladistic phylogenetic reconstruction (e.g., parsimony) that treat shared bands (or nulls) as synapomorphies (e.g., Welsh et al., 1992; Landry et al., 1993) are invalid if there are more than two allelic states (Smith et al., 1994). Even a purely phenetic approach to reconstructing genealogies, based on overall similarity in band-sharing patterns, can be confounded by the problem of cryptic variation, depending on the similarity metric and whether there is bias in the number of cryptic alleles in the presence *versus* absence (e.g., Rossetto et al., 1995). However, on balance, there are more ways to lose a RAPD band than to gain one, and shared presence alleles are more likely to be homologous than shared

nulls. Thus, as we discuss in "Problems with RAPD-PCR and Some Remedies: Nonmendelian Variation," similarity indices based on sharing of expressed bands alone (e.g., Nei and Li, 1979) should be more robust than those that weigh shared presence alleles equally with shared nulls (e.g., Black, 1993; Rossetto et al., 1995), especially when used to build phenetic trees.

PROBLEMS WITH RAPD-PCR AND SOME REMEDIES

Repeatability of Banding Patterns Within and Among Individuals

Background. A high-resolution genetic marker is only useful when each can be reliably and repeatably amplified from a given individual, and each element is heritable (see "Problems with RAPD-PCR and Some Remedies: Nonmendelian Variation"). As with any high-resolution, PCR-based technique (including direct sequencing of PCR products), different amplification conditions can yield different amplification products (Palumbi et al., 1991). Susceptibility to such amplification artifacts is potentially a very serious problem for RAPD-PCR, because the amplified loci are generally anonymous, the primer–template annealing conditions are comparatively permissive, undegraded DNA must be used, and a wide range of amplification parameters can affect banding patterns (reviewed in Schweder et al., 1995).

Not surprisingly, many RAPD-based studies report some difficulties in obtaining repeatable results in terms of band number, molecular weight, and brightness (reviewed in Ellsworth et al., 1993; Bielawski et al., 1995; Schweder et al., 1995). Most problems seem to involve (1) different banding patterns arising from replicate samples of the same individual or (2) bands appearing and disappearing haphazardly in a series of supposedly identical amplifications.

Schweder et al. (1995) recently garnered a fearsome list of potential causes of unreliable and inconsistent amplifications. These confounding variables act in part by influencing the specificity and efficiency of primer–template interactions during the initial phases of PCR amplification. The most frequently cited causes include (1) the identity or model of thermocycler (Klein-Lankhorst et al., 1991; Williams et al., 1991; Fani et al., 1993; Penner et al., 1993; but see Haig et al., 1994); (2) annealing temperatures (Welsh and McClelland, 1990; Welsh et al., 1991; Ellsworth et al., 1993; Levi et al., 1993); (3) denaturation, priming, and extension times and ramping profiles (Yu and Pauls, 1992; Levi et al., 1993; systematically examined by Schweder et al., 1995); (4) primer and template combinations *and* concentrations (Welsh and McClelland, 1990; Hadrys et al., 1992; Ellsworth et al., 1993; Levi et al., 1993; Micheli et al., 1994; Muralidharan and Wakeland, 1993; Williams et al., 1993; Smith et al., 1994); (5) dNTP concentration (Levi et al., 1993; Williams et al., 1993); (6) Mg^{2+} concentration (Ellsworth et al., 1993; Levi et al., 1993; Williams et al., 1993); and (7) *Taq* polymerase concentration (Levi et al., 1993; Williams et al., 1993) and source (Aldrich and Cullis, 1993; Fani et al., 1993; Levi et al., 1993; Williams et al., 1993). Additionally, some primers give notoriously

inconsistent results—perhaps because of nonspecific annealing to the template DNA (e.g., Bielawski et al., 1995)—whereas others are extremely reliable. Finally, the quality of a DNA sample may change over time because of contamination or degradation (Black et al., 1992), a problem we have faced with unextracted tissue stored at −80°C, and with apparently clean DNA stored at −20°C.

Inconsistency at the second level can arise because each extraction starts from a different piece of tissue and then follows an independent extraction (and perhaps amplification) trajectory. Variation in (1) the presence (or concentration) of pathogens, symbionts, or other contaminants (e.g., Micheli et al., 1994) in different tissues; (2) somatic mutations; or (3) extraction procedures and reagents can all produce initially different DNA samples at different concentrations. Subsequent differences in sample processing and storage (e.g., number of freeze–thaw cycles) can also introduce variation. For the reasons noted in the previous paragraph, such variation can affect banding patterns.

Some Remedies. We initially encountered many problems with consistency of amplifications at all the levels identified above. The first step toward success with RAPD-PCR is ensuring that the template DNA is consistently undegraded and uncontaminated by endonucleases, polysaccharides, and other garbage that would interfere with amplification by PCR. Our initial analyses of our extracted DNA on mini-agarose gels showed that it varied in quality, at times being quite degraded. Sometimes it could be fully digested by restriction endonucleases; other times it resisted digestion. We finally settled on the extraction procedures presented in Grosberg.1, using 2% hexadecyltrimethyl ammonium bromide (CTAB). CTAB is a buffer/detergent that seems to work extremely well in combination with standard proteinase-K plus phenol:chloroform extraction protocols, especially with mucusy organisms like many invertebrates and plants (Rogers and Bendich, 1988).

The next step is to optimize the relative concentrations and components of the PCR reaction mixture and the amplification conditions for your thermocycler. An essential step is to quantify the amount of DNA in a sample (preferably using a fluorometer, rather than a spectrophotometer), so that primer–template ratios can be optimized and consistently maintained. In setting up our protocols, we manipulated virtually all the variables listed in "Problems with RAPD-PCR and Some Remedies: Repeatability of Banding Patterns Within and Among Individuals"; however, we always used the same thermocycler and source of purified water. We ultimately found a reaction formula and amplification profile that allowed a wide range of DNA concentrations (three orders of magnitude) to be used without altering banding patterns or intensities for all the primers we used to generate data sets.

Others have not been so fortunate. To them, we first recommend taking whatever time is necessary to develop collection, extraction, and DNA quantification protocols that are simple and reliable. Once this step is overcome, it is worth consulting Williams et al. (1993), Micheli et al. (1994), and Bielawski et al. (1995), for suggestions on optimizing consistency in the amplification step of RAPD-PCR. Although we have not tried it ourselves, among the more interesting recent suggestions is the use of a single-stranded binding protein [gene protein (Gp) 32] in the PCR

cocktail to enhance the specificity of primer annealing, and thereby increase the intensity and repeatability of RAPD amplification products (Bielawski et al., 1995).

The bottom line concerning repeatability *within* individual genotypes is simple: at the outset, develop procedures ensuring that replicate tissue extractions provide identical banding patterns for a particular primer–template combination. If the same set of protocols gives reliable amplifications across a set of primer–template combinations, then do not change any extraction and amplification protocols, including ramping times and PCR machine (unless you are interested in adding to the list of potential woes, or are a molecular biologist who actually seeks to understand the causes of the problems). Use positive (i.e., samples with known banding patterns) and negative (i.e., blanks) controls throughout a study. In other words, don't ever gloat, don't cut corners, and don't underestimate the fickleness of PCR. When consistency problems arise, be prepared to start over, first checking your reagents, your DNA, your thermocycler, and your horoscope.

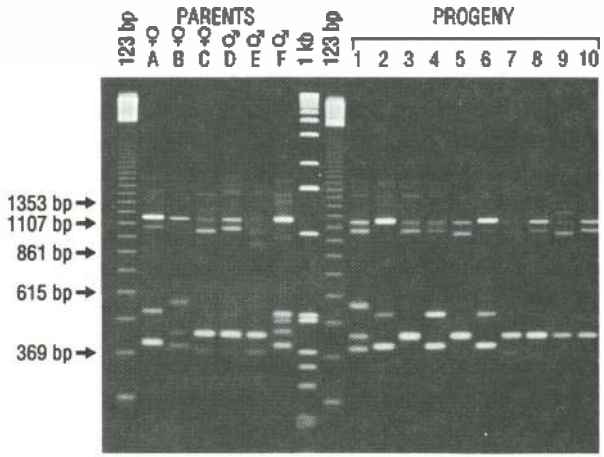
To our knowledge, there is only one way to *ensure* that bands are both repeatable and reliable mendelian markers *among* genotypes: perform crosses to ensure that each and every band is transmitted in a manner consistent with mendelian inheritance. Clearly, such a task would be impossible for large numbers of bands. In "Problems with RAPD-PCR and Some Remedies: Nonmendelian Variation," we offer a statistical approach to this problem, although it will not satisfy those requiring a guarantee of absolute heritability.

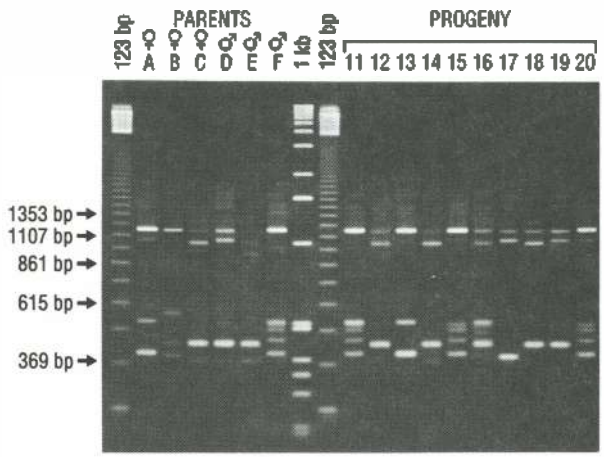
Comigration of Bands

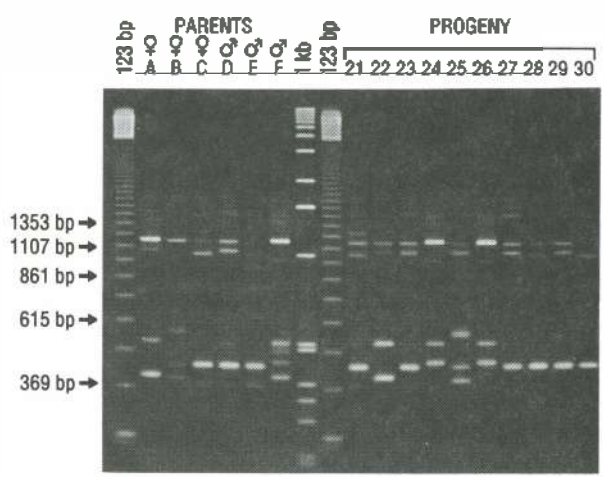
Background. RAPD bands can be numerous and tightly spaced on an agarose gel (Fig. 1) and can be very difficult to distinguish. In principle, the amplification products of distinct RAPD loci, or even different alleles at the same locus, could be of indistinguishable molecular weight (see "What are RAPDs: Sources of Polymorphism and Amplification Artifacts"). The probability of this problem increases as the number and density of bands and the number of sampled individuals increase.

Comigration of homologous (but different) alleles, or nonhomologous alleles, *in the same individual* (i.e., the same lane on a gel) will produce a composite band that comprises two, or more, amplification products. Such composites will yield an underestimation of the number of alleles per locus, heterozygosity, and the frequency of polymorphic loci and will diminish the power to discriminate among individual genotypes and populations. If extensive, comigration of both homologous and nonhomologous products will also confound cladistic analyses (Smith et al., 1994).

Some Remedies. If the goal is simply to increase the resolving power of RAPD-PCR, then there are at least four courses of action. The first three involve attacking the problem at the level of the gel and minimizing scoring errors: agarose gel electrophoresis, in combination with ethidium bromide staining, is notoriously insensitive to revealing small differences (5–10 base pairs, at best) in molecular weight of amplified DNA, especially when bands are slightly blurry. An expensive, time-consuming, and toxic remedy is to use acrylamide (because of its higher


123 bp
PARENTS
♀ A
♀ B
♀ C
♂ D
♂ E
♂ F
1 kb
123 bp
PROGENY
1 2 3 4 5 6 7 8 9 10
1353 bp
1107 bp
861 bp
615 bp
369 bp


123 bp
PARENTS
♀ A
♀ B
♀ C
♂ D
♂ E
♂ F
1 kb
123 bp
PROGENY
11 12 13 14 15 16 17 18 19 20
1353 bp
1107 bp
861 bp
615 bp
369 bp


123 bp
PARENTS
♀ A
♀ B
♀ C
♂ D
♂ E
♂ F
1 kb
123 bp
PROGENY
21 22 23 24 25 26 27 28 29 30
1353 bp
1107 bp
861 bp
615 bp
369 bp

resolving power) and silver-staining or radioactive labeling to visualize PCR products. Alternatively, and much less expensively, one can simply run longer gels (Smith et al., 1994). We (Levitan and Grosberg, 1993), like Hunt and Page (1992), also found that the use of relatively low concentration agarose gels (0.6%), coupled with the addition of a cross-linker (1.0% Synergel; Diversified Biotech, Newton Center, MA), greatly reduces blurriness, allowing us to resolve bands that differ by as little as 10 base pairs. The fourth course of attack is to increase the number of primers used to screen a sample, and thereby reveal more polymorphic loci. This brute force strategy has the advantage of simplicity; it should also increase statistical power for the estimation of relatedness, as it is more likely (although not guaranteed) to reveal polymorphism at independent loci (Queller and Goodnight, 1989; Reeve et al., 1992).

If, on the other hand, the goal is to characterize accurately levels of polymorphism, breeding systems, or phylogenetic relationships, then it is essential to establish whether comigrating bands in different individuals are homologous. There are at least three approaches to addressing this question. Smith et al. (1994) suggest digesting comigrating RAPD amplification products with one (or more) restriction endonucleases, then re-running the digested products on an agarose gel. Digestion of homologous bands should produce the same number and sizes of products. Two additional approaches, both considerably more powerful, expensive, and time-consuming, involve sequencing amplified bands, or use of amplified products as probes in blots of genomic DNA (Hadrys et al., 1992; Smith et al., 1994). These approaches should be most successful when comigration occurs across samples. However, as discussed above, if comigration occurs within lanes too, then neither probing nor sequencing will necessarily clarify just how much cryptic variation there is. To date, Southern hybridizations suggest that at least some comigrating bands represent amplification products from nonhomologous loci (Smith et al., 1994; but see Martin et al., 1991; Hadrys et al., 1992).

Variation in Band Intensity and Limits to Detection

Background. RAPD bands are usually scored as discrete presence/absence characters; yet, (1) some bands consistently amplify more strongly than others and (2) a given band in different individuals may quantitatively vary in intensity from barely visible to unmistakably brilliant. Furthermore, bright bands tend to amplify more

Figure 1. RAPD-PCR products amplified from *Hydractinia symbiolongicarpus* by Operon Technologies primer F-07 on a 0.6% agarose (plus 1.0% Synergel), ethidium bromide-stained gel. The number, intensity, resolution, and spacing of amplification products depend on the primer–template combination (compare lanes labeled Progeny 1 to 10). The amplification products are from the three matings described in the text (♀A × ♂F; ♀B × ♂D; ♀C × ♂E) and their offspring (♀A × ♂F: 2, 4, 6, 11, 13, 15, 20, 22, 24, 26; ♀B × ♂D: 1, 3, 8, 10, 17, 19, 23, 25, 27, 29; ♀C × ♂E: 5, 7, 9, 12, 14, 16, 18, 21, 28, 30). Comparison of banding patterns in parents and offspring shows that all the scoreable bands (see text) on this gel behave as mendelian markers. DNA ladders (123 bp and 1 kb) and base pair sizes provide standardized reference points.

reliably than faint ones (*personal observations;* also see Black, 1993). Because it may be difficult to detect a weakly amplified product on an ethidium bromide-stained agarose gel, a locus could be scored as a null/null homozygote when, in fact, a weakly amplified allele is present. For example, in five of 47 strains of *Xanthomonas,* Southern analysis revealed an amplification product undetectable on ethidium bromide-stained agarose gels (Smith et al., 1994).

Some variation in band intensity appears to be both repeatable and heritable (Hunt and Page, 1992; Heun and Helentjaris, 1993; Levitan and Grosberg, 1993). Heritable variation in band intensity may be due to variation in (1) the number of copies of loci amplifiable by a particular primer–template combination, (2) copy number of tandemly repeated amplifiable sequences within a locus, or (3) additive interactions among loci (Hunt and Page, 1992; Smith et al., 1994). However, due to priming site competition and other nonadditive interactions among loci (Smith et al., 1994), as well as the effects of amplification conditions on primer–template affinity, not all variation in band intensity is likely to be repeatable and heritable.

Finally, even when the intensity of a specific RAPD band is both repeatable and heritable, the question remains of how to compare two individuals, one of which expresses a band strongly, the other of which expresses the same band weakly. One problem with scoring such intensity differences is that intensity itself is a notoriously difficult trait to quantify, especially on ethidium bromide-stained gels. Second, in no case is it known how differences in intensity relate to differences in underlying genotype.

Some Remedies. One could classify bands according to their brightness, analyze the inheritance of brightness classes of a particular band, and score different brightness categories as different alleles at the same locus. This option is worth pursuing in haplo-diploid taxa (e.g., Hunt and Page, 1992; Shoemaker et al., 1994), because the association between genotype and phenotype can be discerned in the haploid gender, or perhaps when relatively few loci are being analyzed. However, if it is desirable to score numerous loci, then this approach will be impractical, at best.

There are at least two other realistic options. The first makes use of every polymorphic band: provided it is detectable, a band is scored as present, regardless of its intensity and absent if it is undetectable. This procedure does not discriminate between genetic and nongenetic causes of variation in banding intensity and assumes that there are no population-specific biases in the expression of brightness variation, including the pattern of false negatives (i.e., undetectable, but nevertheless amplified, products).

The alternative is to score only those bands that are bright when present, with little or no quantitative variation from this state to absence (i.e., undetectability). This scoring procedure appears to lack some of the subjective bias of classifying all bands, regardless of intensity, as present or absent. Unfortunately, as the number of sampled individuals increases, so too does the likelihood that brightness polymorphism for a band will be exhibited in the sample. In our own studies, we were in several instances quite confident that we were scoring a clear presence/absence polymorphism for a given band, only to discover that as we expanded our sample to

include individuals from other populations, the band exhibited quantitative variation in brightness. One then must decide whether to move on to other loci, or to develop an unbiased scoring protocol. This approach may severely underestimate overall levels of genetic variation, because it discards bands (loci?) exhibiting quantitative variation for band intensity. However, unless there is population-specific variation in the number of loci exhibiting such variation, comparisons of genetic similarity may not be severely biased.

Dominance

Background. RAPD alleles are expressed as dominant markers, complicating the inference of underlying genotypes. If either allele at a locus carries an amplifiable fragment, then the phenotype will be a visible band, but—assuming the locus is biallelic—the underlying genotype could be either +/+ or +/−. If the absence of a band represents a single null allele, individuals that lack a particular band should be homozygous (i.e., −/−) for the null allele at that putative locus. The frequency of that allele in a sample (in Hardy–Weinberg equilibrium) can then be calculated from the square root of the proportion of individuals that lack a particular band. It should therefore be possible to estimate allelic and genotypic frequencies reasonably accurately at a RAPD locus, provided that (1) the sample is large (thereby minimizing error in the estimate of the true allelic frequencies); (2) the frequency of the null category exceeds 0.1; (3) the population is in Hardy–Weinberg equilibrium; and (4) the locus in question is biallelic (Lynch and Milligan, 1994). [If the sampled population is in Hardy–Weinberg equilibrium, it is possible to calculate the frequency (and variance) of the expressed marker allele (if it represents a single allele), even if the null category conceals multiple alleles.] If all of these conditions are met, then these frequencies can be used to calculate F-statistics (and their corresponding errors; see Lynch and Milligan, 1994) and to characterize likelihoods of parentage. If they are not, then it will be virtually impossible to estimate allelic and genotypic frequencies.

Satisfaction of the first two conditions is simply a matter of scoring a large sample, and using only those loci with reasonably high frequencies of nulls. Although most studies assume Hardy–Weinberg equilibrium, they fail to demonstrate it. In our view, but without independent knowledge of whether sampled populations and loci conform to Hardy–Weinberg equilibrium, it seems risky to assume that they are.

The available theoretical (see "What are RAPDs: Sources of Polymorphism and Amplification Artifacts") and empirical evidence suggests that the final condition may not be generally realized. Southern blots have shown that a presence band can conceal more than a single allelic variant (Smith et al., 1994). Our own data (Levitan and Grosberg, 1993) suggest that the same is true for the null allelic category at numerous loci in the hydrozoan *Hydractinia symbiolongicarpus.*

We reached this conclusion through simple-minded probabilistic reasoning, rather than sequencing null alleles. Assume that RAPD loci are truly neutral, and that each locus is represented by two alleles, one expressed, the other null. At some loci,

one of the alleles would be more common than the other, whereas the reverse would be true at other loci. However, for a large sample of loci, the distribution of allelic frequencies across loci should approximate a binomial, with a mean of 0.5, because there is no reason to believe that one allele should be systematically more common than the other (i.e., the ratio of presence to absence alleles should be approximately 50:50). In *H. symbiolongicarpus,* we estimated the mean frequency of the null allele, averaged across 133 polymorphic loci, to be 0.809 (lower 95% CI = 0.785; upper 95% CI = 0.831) in the parental generation (n = 6 individuals), and 0.818 (lower 95% CI = 0.789; upper 95% CI = 0.845) in their 30 offspring. This implies that null alleles actually outnumber presence alleles by several times, and that the absence phenotype comprises a pool of alleles that all share the characteristic of not being amplifiable by RAPD-PCR. If this is generally true, and presence phenotypes also encompass numerous alleles, we see no obvious way to estimate RAPD allelic frequencies from presence/absence data.

Some Remedies. There appear to be three alternatives, all of which entail substantial compromises. The "see-no-evil" approach is to hope that there really are only two alleles per locus and that the population sample is in Hardy–Weinberg equilibrium. A slightly better "hear-no-evil" variant would be to use the binomial test we discussed above, with the caveat that the approach only gives a general picture for many loci, not for a specific locus. For the connoisseur of molecular genetics and evolution, the "speak-no-evil" choice would be to use Southern blotting (using probes derived from amplified markers) to reveal all cryptic variation in both scorable classes at all loci, then use this information rather than presence/absence data to estimate allelic and genotypic frequencies. This converts RAPD loci to anonymous single-copy loci (*sensu* Karl and Avise, 1993), thereby losing the major advantages of RAPDs.

The third option is to accept that cryptic variation lies concealed in both the presence and absence categories and acknowledge that RAPD markers are generally ill-suited to the calculation of allelic and genotypic frequencies. Despite this, it is still possible to estimate frequencies of different bands and use those frequencies in either an AMOVA (Excoffier et al., 1992) or a G-test to analyze genetic structure. In this approach, different band frequencies indicate different gene frequencies in different populations. Similarly, although high-resolution analysis of parentage requires considerably more RAPD markers than would be the case for codominant markers (Milligan and McMurry, 1993), this is unlikely to be a limiting feature of RAPD-PCR, except perhaps in highly inbred populations.

Nonmendelian Variation

Background. We refer to repeatable variation in RAPD-PCR products that does not correspond to underlying genetic variation as nonmendelian variation. We exclude from this designation those sources of banding variation that are not repeatable (see "Problems with RAPD-PCR and Some Remedies: Repeatability of Banding Patterns"). Nonmendelian inheritance of RAPD markers, whatever its source, po-

tentially confounds analyses of parentage based on inclusion or exclusion techniques and, at the population level, leads to overestimates of the amount of allelic variation and heterozygosity at RAPD loci. Because RAPD loci are anonymous, and RAPD banding patterns reflect the inheritance of multiple loci and only dominant alleles, nonmendelian inheritance of RAPD markers can be very difficult to detect. Fortunately, most analyses of the transmission genetics of RAPD markers suggest that the majority of bands behave as mendelian alleles (Williams et al., 1990; Carlson et al., 1991; Martin et al., 1991; Welsh et al., 1991; Hunt and Page, 1992; Bucci and Menozzi, 1993; Fondrk et al., 1993; Levitan and Grosberg, 1993). Several studies, however, report the amplification of bands from offspring DNA that did not appear in either putative parent (Riedy et al., 1992; Hunt and Page, 1992; Fondrk et al., 1993).

There are at least three major sources of nonparental bands in offspring, which are often confounded (e.g., Riedy et al., 1992). The first and most obvious (but rarely admitted) is that the putative parents are not the true genetic parents, an option best tested through remating. The second (and, in our opinion, the most common) source comprises laboratory or field artifacts, such as contamination, degradation of template DNA, and inconsistent technique, that promote the spurious appearance (or disappearance) of bands, at least some of which may be repeatably and reliably amplified. Finally, because RAPD-PCR uses fairly permissive annealing conditions, it can produce heteroduplexes (Hadrys et al., 1992; Hunt and Page, 1992), which form when homologous alleles also contain a nonhomologous region. This yields a hybrid fragment that differs in gel migration rate from the actual length of either allele at a locus. RAPD bands may also disappear due to varying levels of competition among primer binding sites, such that efficient amplification depends on the identity and complexity of the remainder of the genome (Hunt and Page, 1992). Indeed, Smith et al. (1994) identified several cases in the phytobacterium genus *Xanthomonas* in which the presence of one fragment appeared to preclude efficient amplification of another fragment.

Some Remedies. The only definitive remedy is to conduct segregation analyses on all bands, separate those markers that behave as mendelian alleles from those that do not, and eliminate all nonsegregating bands from subsequent analyses. In theory, if all scored bands behave as mendelian markers, then it should be possible to analyze parentage using any combination of exclusion, inclusion, or likelihood algorithms (Meagher, 1986; Thompson, 1986; Meagher and Thompson, 1987; Devlin et al., 1988). Similarly, all such markers would provide useful information for the calculation of higher-order genealogical relationships (reviewed in Reeve et al., 1992). Nevertheless, there are several significant drawbacks to this approach. First, it is not clear that establishing the fidelity of markers with a subset of individuals guarantees that a nonhomologous comigrating band will never appear. Second, the number of polymorphic markers that can be assayed, the frequencies of those markers, and the sampling variances of the frequencies fundamentally constrain the power to reconstruct genealogies (Westneat et al., 1987; Lynch, 1988; Burke, 1989; Reeve et al., 1992). Even if it were possible to conduct the necessary matings and

identify the subset of markers that behave as mendelian alleles, it would be especially burdensome to establish such behavior for large numbers of dominant markers in diploid individuals.

The less palatable (but more practical) option is to accept that a comprehensive segregation analysis is either impossible, or not worth the trouble, and devise analytical techniques that reduce (or ignore) errors caused by nonmendelian bands. The virtue of this option is that it makes use of essentially all bands, the high-resolving power of numerous markers being one of the strengths of RAPD-PCR. However, nonmendelian bands can wreak havoc with parentage analyses based on inclusion (the search for parent-specific diagnostic bands in progeny) or exclusion (the search for marker bands that can eliminate one or more individuals from a pool of parents).

A straightforward alternative for characterizing genealogical relationships and genetic structure in a sampled population is first to use all data to build a similarity (or dissimilarity) matrix for all individuals based on the degree of band-sharing, and then to analyze the similarity matrix with a cluster analysis or other tree-building algorithm. This yields a hierarchical depiction of patterns of similarity across all loci, which should mirror genealogical relationships, if band-sharing indices reflect underlying genetic distances. Cluster analyses have the advantage of pooling information that alone would be insufficient to establish genealogical relationships. For example, a marker shared by all members of a sibship, but not private to those individuals in that sibship alone, cannot alone definitively establish membership in that group; however, it can add to the weight of data supporting a particular set of genetic relationships (Apostol et al., 1993). Likewise, if a few nonparental bands appear in offspring, clustering and other tree-building (e.g., neighbor-joining) algorithms will weigh those bands against the high percentage of bands inherited as mendelian markers.

On the other hand, the clustering approach lacks the ability to discriminate classes of relatives of similar relatedness (e.g., full-sibs *versus* parent–offspring), and (especially with overlapping generations) it may be very difficult to discern discrete classes of relatedness. It is possible to plot indices of band-sharing against relatedness inferred from known pedigrees, to calibrate the relationship between band-sharing and relatedness (Piper and Rabenold, 1992; Apostol et al., 1993).

Although clustering based on overall similarity should be relatively insensitive to the occasional appearance of nonmendelian bands, particularly with numerous (say, >100) polymorphic markers, detection of lower-order groupings may be most sensitive to nonmendelian "noise." For this reason, and to appease purists, it is always imperative to minimize contamination, degradation, and amplification artifacts. It is also worth carefully inspecting the raw data matrix to determine whether the presence (or absence) of one band, or a few bands, defines a particular grouping. This can be done by observing how PAUP (Swofford, 1991), or some other cladistic method, assigns character state transitions on the tree generated by distance methods. However, most cladistic tree-building algorithms become glacially mired if the data set includes numerous (say, >20–25) individuals or higher level taxa.

The choice of a similarity metric is also not trivial; in particular, several com-

monly used indices make very different assumptions about weighting of presence *versus* absence information. For example, Nei and Li's (1979) similarity index considers only band presence, not absence; whereas many other commonly used indices incorporate both presence and absence data (e.g., Apostol et al., 1993; Black, 1993; Rossetto et al., 1995). The former index ignores band absence when estimating similarity; we prefer it because we suspect that on average shared bands are more likely than nulls to represent homologous alleles. In the case of rare or endangered organisms, others (e.g., Rossetto et al., 1995) oddly advocate the use of indices that include both the presence and absence of bands, arguing that the risk of overestimating genetic similarity acts as a "safety margin" in genetic conservation strategies. We suggest trying a variety of indices, many of which are conveniently available in the RAPDistance software package (Armstrong et al., 1994), and comparing distance matrices using programs such as DIPLOMO (DIstance PLot MOnitor) analysis (Weiller and Gibbs, 1995).

Finally, different phenetic tree-building algorithms, because they use different agglomeration procedures and make different assumptions about rooting and evolutionary rates, may yield different tree topologies from identical similarity matrices (Felsenstein, 1993). We recommend using a variety of phenetic tree-building methods (e.g., UPGMA, Ward's minimum variance, and neighbor-joining), analyzing the statistical robustness of the resulting trees with some sort of randomization test (e.g., bootstrapping), and comparing the topologies of the trees generated by different algorithms. If the trees differ beyond the details, you should carefully consider why.

TWO CASE STUDIES WITH COLONIAL HYDROZOANS

Background

Against this problem-laden background, critics, sceptics, and population biologists may wonder why anyone would step into the breach of RAPD-PCR. One trivial answer is that all genetic markers have their problems, but their advocates are generally not terribly forthright. For us, the answer is fairly simple: RAPD-PCR gives ready access to literally hundreds of characters that appear to evolve at rather different rates and that can be assayed in hundreds or thousands of individuals, all at relatively low cost in time and materials.

In this section, we act as advocates rather than critics of RAPD-PCR and show how RAPD markers can be used to characterize parent–offspring and higher-order genetic relationships, as well as population structure. We focus on our own studies of two congeneric species of colonial marine hydrozoans, in which the temporal association between adults and their gametes, or adults and offspring, is so brief that genetic methods must be used to deduce parentage and to analyze population structure.

The first study centers on *Hydractinia symbiolongicarpus,* which lives in protected nearshore waters along the New England coast of the United States. We considered three questions in this species: Can we use RAPD markers to (1) associ-

ate parents with their offspring; (2) identify sibships correctly; and (3) characterize the genetic composition of chimerical colonies? The second study concerns two questions that involve portraying genetic structure in natural populations of *H. milleri*, an inhabitant of the Pacific Northwest (from northern California, northward at least to Vancouver Island): (1) at what level of resolution can we identify genetic structure, and (2) to what extent is our ability to discern structure limited by variation in the marker?

In addition to living on opposite coasts and distinct habitats, these two species of gonochoric hydroids have life histories that differ in several key aspects. *Hydractinia symbiolongicarpus* colonizes gastropod shells occupied by the hermit crab *Pagurus longicarpus* (Buss and Yund, 1989). Males release their sperm into the water, whereas females release their sinking eggs onto the benthos (Bunting, 1894; Ballard, 1942). Fertilized eggs rapidly develop into demersal larvae (Müller, 1969; Weis and Buss, 1987; Yund et al., 1987; Shenk and Buss, 1991). These larvae settle onto passing hermit crab shells, metamorphose, and develop into colonies consisting of hundreds to thousands of polyps, linked to one another by a common gastrovascular system (reviewed in Yund et al., 1987). Two or more larvae frequently colonize a single shell, and when they do, there is the potential for intense intraspecific competition for space (Yund et al., 1987). As the colonies come into contact, one of two outcomes generally ensues. Either the colonies fuse via their gastrovascular systems, forming a physiologically and morphologically unified genetic chimera; or one or both colonies produce highly modified extensions of the gastrovascular system, termed hyperplastic stolons (Ivker, 1972, and references therein). Large batteries of nematocysts cover these specialized stolons, which can inflict fatal damage to an opponent (Buss et al., 1984). The hyperplastic response can involve a substantial investment of somatic tissue, decreasing somatic growth and reproductive output by up to 50% (R.K. Grosberg and L.W. Buss, *unpublished observations*). Although *H. symbiolongicarpus* colonies themselves are sessile, the hermit crabs that bear them appear to mix extensively; thus populations ought to exhibit little genetic structure. In fact, allozyme markers revealed a high degree of polymorphism, but no evidence of genetic structure or nonrandom mating, among *H. symbiolongicarpus* colonies from different hermit crabs at the same geographic locality.

Colonies of *Hydractinia milleri*, in contrast to *H. symbiolongicarpus*, inhabit large boulders on wave exposed shores, low in the intertidal. In terms of expected genetic structure, their reproductive biology differs from *H. symbiolongicarpus* in one significant respect: although males release sperm into the water, females brood their embryos (*personal observations*). The brooded larvae eventually take up residence among the polyps of their mother, before crawling away to found a new colony. Colonies are typically 5 (about 10 polyps) to 100 mm across (several thousand polyps) but occasionally reach 200 mm or more. Colonies generally form aggregations consisting of from 10 to 100 colonies, with discrete aggregations separated by anywhere from 50 cm to hundreds of meters of apparently habitable, but uncolonized, substrate.

Unlike *H. symbiolongicarpus*, natural interactions between conspecific *H. mil-*

leri do not apparently elicit the production of hyperplastic stolons, or any other obvious sign of aggression. Colonies seem to coexist with stable boundaries for several years. One explanation for this lack of aggression is that neighboring colonies are so closely related that they lack sufficient genetic disparity to evoke such behavior. Indeed, interactions between colonies taken from different localities do elicit the production of nonfeeding, hypertrophied polyps (tentaculozooids) along colony margins. Taken together, the reproductive biology, dispersion, and lack of aggression exhibited by *H. milleri* colonies suggest that aggregations consist of a small number of founding individuals and their offspring, which, if persistent, may become genetically differentiated and inbred with respect to nearby aggregations. If both sperm and larval dispersals are primarily restricted to the bounds of an aggregation, then an aggregation provides a conveniently small subpopulation for investigating parent–offspring relations and correlates of reproductive success.

For both species, we needed a genetic marker that would distinguish closely related individuals, yet allow us to assay thousands of individuals. Allozymes revealed a rich store of genetic variation in *Hydractinia symbiolongicarpus,* allowing us to distinguish all sampled individuals based on their multilocus genotypes. Similar attempts using allozyme markers to distinguish individuals of *H. milleri* failed. Moreover, we became interested in analyzing the parentage of larvae, and in nondestructively sampling natural populations through time. This required that we employ not only a highly polymorphic marker, but one that could be amplified by PCR. At about this time, the first papers describing RAPD-PCR and its variations began to appear. We jumped on the bandwagon. In the next two sections we describe how we developed RAPD markers in these two species, and we summarize some preliminary results. Levitan and Grosberg (1993) fully report the data summarized here for *Hydractinia symbiolongicarpus*.

Study I: Parentage in *Hydractinia symbiolongicarpus*

DNA Extraction. At the time we began these studies, there were no protocols for extracting high molecular weight DNA from hydroids. Conventional proteinase-K/phenol:chloroform extraction techniques with various detergents (e.g., SDS and Triton-X-100) were undependable, and the DNA was often severely degraded. Finally, we followed botanical advice (Milligan, 1992) and added CTAB (hexadecyltrimethyl ammonium bromide) to the extraction buffer. Electrophoresis on minigels, as well as fluorometric analysis, revealed consistent yields of high-quality DNA.

Optimization of PCR-Amplification Conditions. The first step is to develop a PCR cocktail that gives repeatable amplification of mendelian bands using the minimum amount of template DNA (often in short supply) and DNA polymerase (the most expensive reagent, by far). This involves careful adjustment of the amount of template DNA in each extraction, followed by optimization of the concentrations of template DNA, RAPD primer, dNTPs, DNA polymerase, and salts ($MgCl_2$ and KCl).

The second step is the amplification process itself. A typical PCR reaction

involves about 30–50 amplification cycles, each consisting of three phases: denaturation of template DNA, primer annealing, and polymerization/extension. RAPD-PCR differs from other types of amplification procedures in that only a single primer is generally used to amplify template DNA, and primer annealing takes place at a relatively low temperature (35–50°C). The duration and temperature of each phase can and should be adjusted so as to minimize the probability of nonspecific priming (the highest feasible annealing temperature) and maximize repeatability. If possible, the effects of transition profiles (i.e., "ramping times") between phases should also be explored (see references in "Problems with RAPD-PCR and Some Remedies: Repeatability of Banding Patterns Within and Among Individuals"). Finally, we found it worthwhile in terms of repeatability to start the reaction with a single cycle of prolonged denaturation.

Electrophoresis and Detection of Amplification Products. Of the various electrophoretic media currently available, agarose gels, at concentrations between 0.5% and 2.0%, are by far the most popular for RAPD-PCR. An alternative medium is polyacrylamide, preferred by some for RAPD-PCR because of its apparently superior ability to resolve differences in molecular weight. However, acrylamide is more expensive, cumbersome, and more toxic (in its nonpolymerized form) than agarose.

In our experience, agarose concentration can strongly influence the migration patterns and resolution of RAPD bands. We were often disappointed by the thickness and blurriness of bands we obtained across a range of agarose concentrations, making it difficult to resolve differences among bands migrating at similar rates. We then tried adding a cross-linker (1% Synergel) to our gels and decreased the agarose concentration to 0.6%. This gave us narrow, sharp bands. On the other hand, gel thickness and size had little effect on our ability to detect RAPD variation.

Screening for RAPD Polymorphism. In our study of *H. symbiolongicarpus* we screened 56 primers on three male and three female individuals, which we then mated in three pairs to generate offspring for an inheritance study. We identified 13 primers that exhibited polymorphism, produced sharp banding patterns, and amplified between five (enough to provide genetic information) and 15 bands (the greatest number that we were comfortable scoring). Of the other primers we screened, 19 failed to reveal polymorphism; 14 amplified less than 5 polymorphic bands; two produced more than 15 polymorphic bands; and eight amplified only blurred bands. All told, we used 23% of the screened primers for further analyses, a greater success rate than for *Hydractinia milleri*.

Scoring Protocol. We first tried using a state-of-the art optical imaging program to analyze banding patterns from the gel photographs. For several reasons, it quickly became apparent that using these programs would be more time-consuming than hand scoring the gels. The primary problem is that RAPD bands are often tightly clustered. To avoid having the program pool narrowly separated bands into the same category, the range of sizes used to classify any single band must be very small.

However, as we reduced this "binning" range, the small and inevitable variation in band migration introduced by otherwise trivial amounts of gel warping caused the program to score identical bands as different. Even with three or four DNA ladders per gel as standards, slight gel warping caused the misidentification of many bands; eventually, we gave up on this kind of automation. Instead, we created separate enlarged templates depicting the positions (i.e., molecular weights) in a "synthetic" lane of all bands amplified by each primer in the whole sample. We could then quickly score each gel by recording the subset of bands found in each lane. Because some bands consistently amplified strongly in some individuals or subpopulations, but weakly in others, we initially scored all bands that repeatably amplified (see next section), regardless of intensity.

Analysis of Repeatability. We excised duplicate tissue samples from each of the six parental colonies and 30 offspring used in the inheritance study (described in the next section) and extracted the DNA independently in each sample. We amplified these samples using the full set of primers and compared banding patterns between duplicates and from the same sample run on different days. Once we had optimized extraction and amplification conditions, we found no variation for a given primer–template between independent extractions or days (Levitan and Grosberg, 1993; Fig. 1). It is also worth noting that in our initial screening of the six parental genotypes using 56 primers, DNA concentration in the amplification cocktail ranged from 1 to 30 ng/μl in duplicate extractions (largely because we started with different amounts of tissue). Over this range of variation in DNA concentration, we found no discernible effect on banding pattern in the duplicated extractions.

Inheritance of Markers. To explore whether RAPD markers behave as mendelian alleles in *Hydractinia*, we mated three randomly paired male and three female colonies collected haphazardly from a large population at Barnstable Harbor, MA. We then analyzed ten randomly selected offspring from each mating, for a total of 30 offspring (Fig. 1). We coded and scrambled the 30 F_1 DNA samples so that we could not discern their relationships to each other or to their parents through the time of band scoring and analysis. We screened the six adults and 30 offspring with 13 primers, generating a total of 156 markers of which 133 were polymorphic. Only four nonparental markers appeared in the offspring.

Rather than attempting to determine whether every polymorphic marker behaved as a mendelian allele, we once again adopted a statistical approach to characterize the overall transmission properties of the majority of bands. The analysis is based on binomial expectations for the transmission of a dominant allele to offspring in a sibship. From a phenotypic perspective, matings can involve parents in which both express a band, or only one expresses a particular band. From a genetic perspective, a parent expressing a particular band can either be $+/+$ or $+/-$ at that locus, whereas a parent lacking a band will be $-/-$. So when both parents in a mating express a marker, the mating could be one of three types: (1) $+/+ \times +/-$; (2) $+/+ \times +/+$; or (3) $+/- \times +/-$. Offspring from the first two types of mating should all inherit the band, whereas on average 75% of the offspring from the second type of

mating should inherit the band. When one parent expresses a band, but the other does not, the mating could be one of two types: (1) $+/+ \times -/-$ or (2) $+/- \times -/-$. In the first case, all offspring would inherit the band; in the second, on average 50% of the offspring should inherit the band.

When accumulated across many loci, the distribution of numbers of offspring in our sibships of ten that inherited a band when both parents expressed the band should have two modes: one at seven or eight, reflecting $+/- \times +/-$ matings, and one at ten, reflecting $+/+ \times +/-$ and $+/+ \times +/+$ matings. Similarly, for matings in which only one parent expressed the band, there should be modes at five and ten offspring, the former reflecting $+/- \times -/-$ matings, and the latter reflecting $+/+ \times -/-$ matings. These are the distributional patterns we found empirically (Levitan and Grosberg, 1993). This outcome by no means guarantees that each and every band will be transmitted as a mendelian allele; but it does support the assumption that the majority of bands must be transmitted as mendelian alleles.

Genealogical Analysis of Banding Patterns. We analyzed the banding pattern data in parents and offspring with three methods: exclusion, inclusion, and phenetic clustering. Exclusion involves searching for markers in offspring that are absent from some, but not all, individuals in the population of potential parents. This protocol is most effective when one parent (usually the mother) is known. For example, if an offspring carries a particular marker, and the mother lacks that marker, then all males who also lack the marker can be excluded from the pool of potential fathers. However, if nonparental bands occur in the offspring, then true genetic parents may be incorrectly excluded from consideration. In the best of circumstances, only a single potential father will carry the marker and remain in the pool. More commonly, several prospective fathers will remain (Lewis and Snow, 1992): likelihood methods can then be used to assign relative probabilities of paternity to these individuals (Meagher, 1986; Devlin et al., 1988; Roeder et al., 1989).

Inclusion requires identification of a marker (or better, markers) unique to a single parent, and then assigning offspring carrying those markers to that parent. Unlike exclusion, parentage assignment by inclusion can at least partially overcome the appearance of nonparental bands in offspring, as long as several markers can be used to associate an offspring with a particular parent. However, inclusion suffers from the danger that an unsampled potential parent outside the hypothetical parental pool has a set of markers identical to a sampled prospective parent within the pool. Thus parentage assignment by inclusion will be most powerful when there is a high degree of polymorphism and when the pool of potential parents can be circumscribed.

Because it is relatively easy to survey potential parents and offspring for hundreds of polymorphic markers using RAPD-PCR, phenetic tree-building methods (given the assumptions discussed in "Problems with RAPD-PCR and Some Remedies: Nonmendelian Inheritance") offer an alternative approach to the analysis of genetic relationships. Tree-building methods group individuals in a hierarchical framework based on overall similarity of banding patterns. Even if there are some nonparental bands in offspring, the "signal" of well-behaved mendelian bands

should overwhelm the "noise" of nonmendelian bands, provided enough mendelian characters (i.e., markers) are considered. We constructed a presence/absence matrix for all 156 bands in the six parents and 30 offspring, and then calculated Nei and Li's (1979) similarity index for all pairwise combinations. We analyzed this similarity matrix using a variety of tree-building algorithms, all of which yielded concordant topologies.

In comparing inclusion and exclusion approaches to parentage analysis in this simple situation, we found that inclusion was slightly more powerful than exclusion, but both accurately assigned parents to offspring using the 13 polymorphic primers (Levitan and Grosberg, 1993). However, with both methods the presence of nonparental markers made some assignments ambiguous. The clustering approach was far and away the most powerful technique (also see Apostol et al., 1993). It correctly and significantly associated all offspring with their true parents and grouped together each sibship without error or ambiguity. We plan to explore the power of clustering further, as we expand the number of prospective parents and offspring that we include in the analysis.

Analysis of Chimeric Colonies. Fusion between *Hydractinia symbiolongicarpus* colonies may be a reasonably common event (Yund et al., 1987), especially when sibling larvae cosettle on the same shell. Fusion potentially confers substantial fitness benefits, as well as costs, which together influence the evolution of specificity in self/nonself, or allorecognition, systems (reviewed in Grosberg, 1988). For example, fusion produces a larger chimeric individual, which may enjoy increased resistance to predation and enhanced competitive ability (Buss, 1990). On the other hand, because most clonal invertebrates do not sequester their germ lines, fusion opens several avenues to intergenotypic competition between cell lines inhabiting the same soma (Buss, 1990). Thus, in order minimally to assay the fitness costs and benefits of intergenotypic competition, it is necessary to characterize the representation of both genotypes in somatic and reproductive tissues.

Unfortunately, when two colonies fuse and form a genetically chimeric individual, after a short period of time it is often impossible visually to distinguish the two genotypes. In the case of *Hydractinia*, because the extensive polymorphism in the self/nonself recognition system almost necessarily limits fusion to closely related individuals (our own data), the marker must be highly polymorphic, so that tissues of close relatives can be genetically distinguished. We therefore began to explore the utility of RAPD-PCR markers for characterizing the genetic composition of somatically fused colonies of *Hydractinia symbiolongicarpus*.

In these preliminary studies, we first clonally replicated sibling colonies and paired them to determine which sibling combinations fused. At the same time, we began an extensive RAPD-PCR screening of these siblings to identify markers that would unambiguously diagnose each genotype in a fusible pair. We then established chimeric colonies by allowing compatible siblings to fuse. At roughly monthly intervals postfusion, we clipped one polyp (or a few polyps) from beneath 20–40 random positions on a grid placed over the colony. We then extracted DNA in the manner described in Grosberg.1 for small tissue samples and amplified the DNA

using a primer that would allow us to discern to which component genotype the sampled polyp(s) belonged. Figure 2 shows a diagram of one such chimera some 540 days postfusion, with the corresponding gel. This figure clearly illustrates that RAPD markers can be used to identify the genotypic identity of components of a chimera, and that intergenotypic chimeras can remain fairly stable over sustained periods.

We also are beginning to use RAPD markers to portray the gametic composition of chimeras. This is, however, a more complex undertaking than characterizing somatic composition, both because RAPD markers are dominant (thus only half the gametes from a heterozygous parent will receive the marker) and gametes usually carry only a single copy of the parental DNA. Although several protocols exist for extracting and amplifying very small amounts of DNA (e.g., Landry et al., 1993; Grosberg.1), at this point, we are not convinced we can reliably amplify the DNA from a single gamete. We therefore fertilize ova of unknown genotypic identity with sperm from a male carrying a diagnostic marker and assay the resulting (multicellular) larvae. Conversely, we use sperm of unknown genotypic identity to fertilize ova with a diagnostic marker and assay the larvae for the presence of one or the other diagnostic bands of the chimera. This procedure is most efficient when both components of the chimera are homozygous for the presence of their diagnostic marker band; but even when they are not, 50% of the progeny should carry the marker, and the analysis should be unbiased.

Study II: Population Structure in *Hydractinia milleri*

Screening for RAPD Polymorphism. Using the same procedures described above for *Hydractinia symbiolongicarpus,* we extracted DNA from five *H. milleri* colonies collected from a single aggregation from Tatoosh Island, Washington. We screened these samples with 101 primers. Of these primers, nine failed to amplify any loci; 35 gave bands that were too smeared to score; 23 amplified clear but invariant bands; and 34 gave clear polymorphic banding patterns. We chose the 14 most promising primers (using the criteria in "Parentage in *Hydractinia symbiolongicarpus:* Screening for RAPD Polymorphism") and screened those on 57 individuals from an additional site at Waadah Island, Washington (approximately 10 km from Tatoosh Island). We repeated this screening three times and chose the best eight primers that produced repeatable, polymorphic bands. Thus, of the original 101 primers that we screened, slightly more than 8% proved useful. We then extended the geographic scope of our sampling to include two sites in northern California and ultimately screened over 200 colonies, carrying 278 polymorphic bands.

Population Sampling Design. Because *H. milleri* females brood demersal, crawl-away larvae, it is likely that most sibling larvae will settle close to their mother. Consequently, populations of *H. milleri* ought to be more genetically structured than, say, *H. symbiolongicarpus* (as well as free-spawning invertebrates with feeding larvae, or plants with wind-dispersed propagules). But it is also possible that waves carry some larvae a considerable distance, where they successfully

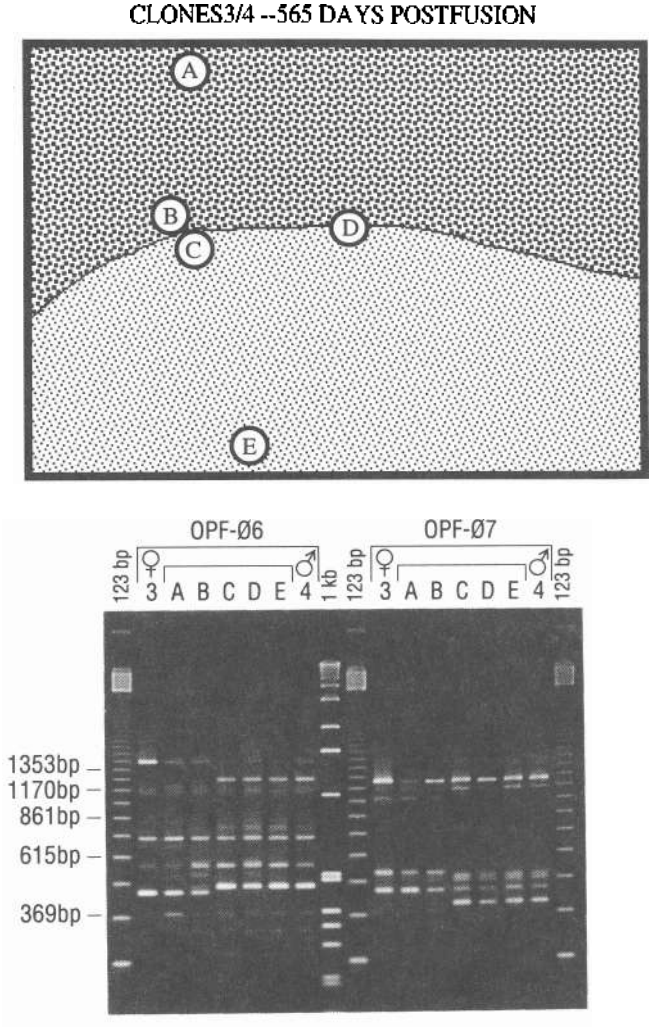


Figure 2. *Top panel:* Diagrammatic representation of a chimeric colony of *Hydractinia symbiolongicarpus* 565 days postfusion. The different shading patterns denote the apparent extent of the two strains (♀3 in the upper section and ♂4 in the lower section) that form the chimera, inferred from the phenotypic differences in the structure of their corresponding gastrovascular networks. At each of the lettered positions (A–E), we sampled two polyps and analyzed them for the presence of strain-specific, diagnostic RAPD markers. *Bottom panel:* Amplification products from two RAPD primers (OPF-06 and OPF-07) applied to DNA extracted from positions A–E. Lanes denoted ♀3 and ♂4 show amplification products from pure cultures of the constituent strains. The amplification products show that samples A and B have profiles characteristic of ♀3, whereas samples C–E have profiles like that of ♂4.

metamorphose. Thus the spatial scale over which *H. milleri* populations ought to exhibit genetic structure is difficult to predict.

We used a stratified sampling design in which we mapped the positions of over 200 colonies in 17 aggregations at three localities, two about 10 km apart in northern California (Doran Rocks: 73 colonies in seven aggregations; Coleman Beach: 73 colonies in six aggregations), and one about 1000 km northward near the mouth of the Straight of San Juan de Fuca in Washington State (Waadah Island: 54 colonies in four aggregations). For analysis of genetic structure, we carefully removed 3–5 polyps from each colony; we also removed brooded embryos from females for analysis of individual reproductive success. At the Waadah Island site, we mapped the colonies over a 3-year period, monitoring patterns of recruitment, mortality, fission, and fusion. This allowed us to compare the genetically based picture with an independent set of demographic observations, and thereby assess the power of the RAPD markers to discriminate clonemates from close relatives.

Data Analysis and Population Structure. We assayed each colony with the battery of eight primers. We then computed Nei and Li's (1979) similarity index for all pairs of colonies and analyzed the resulting similarity matrix using a variety of tree-building algorithms: all gave virtually, but not perfectly, identical tree topologies.

At the broadest sampling scale, the cluster analysis accurately assigns each individual colony to its correct site. As expected, individuals from the Doran Rocks and Coleman Beach sites are far more similar to each other than either is to samples from the Waadah Island site. At the level of aggregations within sites, the cluster analysis correctly grouped >90% of colonies from the same aggregation together. In most cases, "misassigned" colonies grouped with members of the nearest aggregation. However, about 2–3% of individuals grouped by themselves, suggesting that they immigrated from outside the area we sampled. Taken together, these data suggest that most dispersal occurs on scales of millimeters to centimeters (within aggregations). Some dispersal occurs on the order of tens of meters (among aggregations within a site), and perhaps farther, but not so often as to overwhelm the persistence of substantial genetic structure at very fine spatial scales.

There is an additional noteworthy pattern in the relationships among colonies in the Waadah Island aggregations: we could not distinguish two pairs of individuals. Either the 278 RAPD markers could not resolve these genotypes, perhaps because they represent pairs of close relatives, or the two pairs represent clonemates derived by asexual fragmentation. We addressed this question by comparing the genetically based inference of clonality to the time-series maps of the colonies in the Waadah Island aggregations. The chronologies of the two pairs of colonies show that both pairs fragmented 3 years before we sampled the aggregations for genetic analysis. Thus the array of markers could distinguish clonemates from close kin and further showed that cloning, although rare in *H. milleri,* does occur.

Overall, this analysis of genetic structure shows that genetic similarity declines with distance separating colonies, even at very fine spatial scales. The picture reassuringly reflects the reproductive ecology of *H. milleri,* with its limited potential for gene flow.

CONCLUSIONS

Our principal aim in this chapter was to assess some of the positive and negative attributes of RAPD markers in terms of their power to characterize genetic relationships and structure. In our view, RAPD markers, because they are anonymous and expressed as dominant alleles that may conceal substantial amounts of cryptic genetic variation, are generally ill-suited to the characterization of breeding systems, the calculation of population genetic parameters such as F-statistics, or the inference of phylogenetic relationships above the species level. Granted, with a lot of work, it could be possible to render a small number of markers useful for these purposes; but it seems far more profitable not to fight the shortcomings of RAPD-PCR.

On the other hand, RAPD-PCR can easily generate hundreds of potentially independent markers, the majority of which appear to behave as neutral mendelian alleles. In many cases, RAPD loci harbor sufficient polymorphism to reveal genetic structure at very fine spatial and temporal scales, but not so much as to obscure structure at broader geographic (and presumably temporal) scales. With some caution in both the generation and interpretation of these markers, they hold as yet unmatched promise for characterizing hierarchical genetic structure and close genealogical relationships in large samples, especially when population structure spans a range of spatial scales. For those willing to wait with bated breath, there can be little doubt that high-resolution markers—newer, more fashionable, and almost certainly better than RAPDs—will soon be developed and publicized as the final solution to the daunting problems of characterizing genetic structure and reproductive success. Until that time actually arrives, we urge the novice, nervous, or antagonistic population ecologist to look critically and fairly at all potentially useful genetic markers, including RAPDs, to assess their value for the problem at hand, and to employ the technique that works.

ACKNOWLEDGMENTS

NSF grants to RK Grosberg and DR Levitan allowed us to develop RAPD-PCR markers and supported our research on the population biology of colonial marine invertebrates. Funds from the Agricultural Experiment Station of the University of California also supported this research. J Ferraris and S Palumbi goaded us into writing this chapter; the reader can judge if they were right. We thank several reviewers, who prefer to remain anonymous, for their useful comments. The American Society of Zoologists and the Alfred P Sloan Foundation funded the symposium that was the original venue for this work.

REFERENCES

Aldrich J, Cullis CA (1993): RAPD analysis in flax: optimization of yield and reproducibility using Klen*Taq* 1 DNA polymerase, Chelex 100 and gel purification of genomic DNA. *Plant Mol Biol Rep* 11:128–141.

Apostol BL, Black WC IV, Miller BR, Reiter P, Beaty BJ (1993): Estimation of the number

of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. *Theor Appl Genet* 86:991–1000.

Armstrong JS, Gibbs AJ, Weiller G (1994): The RAPDistance package. Obtainable via Anonymous ftp or WWW. ftp://life.anu.edu.au/pub/RAPDistance; http://life.anu.edu.au/molecular/software/rapd.html.

Avise JC (1994): *Molecular markers, natural history, and evolution*. New York: Chapman and Hall.

Ballard WW (1942): The mechanism of synchronous spawning in *Hydractinia* and *Pennaria*. *Biol Bull* 82:329–339.

Bell G (1992): Five properties of environments. In: Grant PR, Horn HS (eds). *Molds, molecules, and Metazoa*. Princeton, NJ: Princeton University Press, pp 33–56.

Bielawski JP, Noack K, Pumo DE (1995): Reproducible amplification of RAPD markers from vertebrate DNA. *Bio techniques* 18:856–860.

Black WC (1993): PCR with arbitrary primers: approach with care. *Insect Mol Biol* 2:1–6.

Black WC, DuTeau NM, Puterka GJ, Nechols JR, Pettorini JM (1992): Use of the random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) to detect DNA polymorphisms in aphids (Homoptera: Aphididae). *Bull Entomol Res* 82:151–159.

Bowditch BM, Albright DG, Williams JGK, Braun MJ (1993): Use of randomly amplified polymorphic DNA markers in comparative genome studies. In: Zimmer EA, White TJ, Cann RL, Wilson AC (eds). *Methods in enzymology, Volume 224. Molecular evolution: producing the biochemical data*. San Diego, CA: Academic Press, pp 294–309.

Bucci G, Menozzi P (1993): Segregation analysis of random amplified polymorphic DNA (RAPD) markers in *Picea abies*. *Mol Ecol* 2:227–232.

Bunting M (1894): The origin of sex-cells in *Hydractinia echinata* and *Podocoryne*. *J Morphol* 9:203–236.

Burke T (1989): DNA fingerprinting and other methods for the study of mating success. *Trends Ecol Evol* 4:139–144.

Burke T, Hanotte O, Bruford MW, Cairns E (1991): Multilocus and single locus multisatellite analysis in populations biological studies. In: Burke T, Dolf G, Jeffreys AJ, Wolff R (eds): *DNA fingerprinting: approaches and applications*. Basel: Birkhäuser, pp 154–168.

Buss LW (1990): Competition within and between encrusting invertebrates. *Trends Ecol Evol* 5:352–356.

Buss LW, Yund PO (1989): A sibling species group of *Hydractinia* in the northeastern United States. *J Mar Biol Assoc UK* 69:857–875.

Buss LW, McFadden CS, Keene DR (1984): Biology of hydractiniid hydroids. 2. Histocompatibility effector system mediated by nematocyst discharge. *Biol Bull* 167:139–158.

Caetano-Anollés G, Bassam BJ, Gresshoff PM (1991): DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Biotechnology* 9:553–557.

Carlson JE, Tulsieram LK, Glaubitz JC, Luk VWK, Kauffeldt C, Rutledge R (1991): Segregation of random amplified DNA markers in F_1 progeny of conifers. *Theor Appl Genet* 83:194–200.

Clark AG, Lanigan CMS (1993): Prospects for estimating nucleotide divergence with RAPDs. *Mol Biol Evol* 10:1096–1111.

Dawson IK, Chalmers KJ, Waugh R, Powell W (1993): Detection and analysis of genetic variation in *Hordeum spontaneum* populations from Israel using RAPD markers. *Mol Ecol* 2:151–160.

Devlin B, Roeder K, Ellstrand NC (1988): Fractional paternity assignment: theoretical development and comparison to other methods. *Theor Appl Genet* 76:369–380.

Ellsworth DL, Rittenhouse KD, Honeycutt RL (1993): Artifactual variation in randomly amplified polymorphic DNA banding patterns. *Biotechniques* 14:214–216.

Excoffier L, Smouse PE, Quattro JM (1992): Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.

Fani R, Damiani G, Di Serio C, Gallori E, Grifoni A, Bazzicalup M (1993): Use of random amplified polymorphic DNA (RAPD) for generating DNA probes for microorganisms. *Mol Ecol* 2:243–250.

Felsenstein J (1993): *PHYLIP (Phylogeny Inference Package). Version 3.5C*. Seattle: Department of Genetics, University of Washington.

Fondrk MK, Page RE, Hunt GJ (1993): Paternity analysis of worker honey bees using random amplified DNA (RAPD). *Naturwissenschaften* 80:226–231.

Grosberg RK (1988): The evolution of allorecognition specificity in clonal invertebrates. *Q Rev Biol* 63:377–412.

Hadrys H, Balick M, Schierwater B (1992): Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol Ecol* 1:55–63.

Haig SM, Rhymer JM, Heckel DG (1994): Population differentiation in randomly amplified polymorphic DNA of red-cockaded woodpeckers *Picoides borealis*. *Mol Ecol* 3:581–595.

Hedrick P (1992): Shooting the RAPD's. *Nature* 355:679–680.

Heun M, Helentjaris T (1993): Inheritance of RAPDs in F_1 hybrids of corn. *Theor Appl Genet* 85:961–968.

Huff DR, Peakall R, Smouse PE (1993): RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloe dactyloides* (Nutt.) Engelm.]. *Theor Appl Genet* 86:927–934.

Hunt GJ, Page RE (1992): Patterns of inheritance with RAPD molecular markers reveal novel types of polymorphism in the honey bee. *Theor Appl Genet* 85:15–20.

Ivker FB (1972): A hierarchy of histo-incompatibility in *Hydractinia echinata*. *Biol Bull* 143:162–174.

Karl SA, Avise JC (1993): PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. *Mol Biol Evol* 10:342–361.

Klein-Lankhorst RM, Vermunt A, Weide R, Liharska T, Zabel P (1991): Isolation of molecular markers from tomato (*L. esculentum*) using random amplified polymorphic DNA (RAPD). *Theor Appl Genet* 83:108–114.

Landry BS, Dextraze L, Boivin G (1993): Random amplified polymorphic DNA markers for DNA fingerprinting and genetic variability assessment of minute parasitic wasp species (Hymenoptera: Mymaridae and Trichigrammatidae) used in biological control programs of phytophagous insects. *Genome* 36:580–587.

Lewis PO, Snow AA (1992): Deterministic paternity exclusion using RAPD markers. *Mol Ecol* 1:155–160.

Lewis PO, Zaykin D (1996): *Genetic data analysis* (computer software). Sunderland, MA: Sinauer.

Lessa EP, Applebaum G (1993): Screening techniques for detecting allelic variation in DNA sequences. *Mol Ecol* 2:119–129.

Levi A, Rowland LJ, Hartung JS (1993): Production of reliable randomly amplified polymorphic DNA (RAPD) markers from DNA of woody plants. *HortScience* 28:1188–1190.

Levitan DR, Grosberg RK (1993): The analysis of paternity and maternity in the marine hydrozoan *Hydractinia symbiolongicarpus* using randomly amplified polymorphic DNA (RAPD) markers. *Mol Ecol* 2:315–326.

Liu Z, Furnier GR (1993): Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. *Theor Appl Genet* 87:97–105.

Lynch M (1988): Estimation of relatedness by DNA fingerprinting. *Mol Biol Evol* 5:584–599.

Lynch M (1990): The similarity index and DNA fingerprinting. *Mol Biol Evol* 7:478–484.

Lynch M, Milligan B (1994): Analysis of population genetic structure with RAPD markers. *Mol Ecol* 3:91–100.

Martin GB, Williams JGK, Tanksley SD (1991): Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. *Proc Nat Acad Sci USA* 88:2336–2340.

Meagher TR (1986): Analysis of paternity within a natural population of *Chamaelirium luteum*. 1. Identification of most-likely male parents. *Am Nat* 128:199–215.

Meagher TR, Thompson EA (1987): Analysis of parentage for naturally established seedlings of *Chamaelirium luteum* (Liliaceae). *Ecology* 68:803–812.

Micheli MR, Bova R, Pascale E, D'Ambrosio E (1994): Reproducible DNA fingerprinting with the random amplified polymorphic DNA (RAPD) method. *Nucleic Acids Res* 22:1921–1922.

Milligan BG (1992): Plant DNA isolation. In: Hoelzel AR (ed). *Molecular genetic analysis of populations. A practical approach*. New York: IRL Press, pp 59–88.

Milligan BG, McMurry CK (1993): Dominant vs. codominant genetic markers in the estimation of male mating success. *Mol Ecol* 2:195–284.

Müller W (1969): Auslosung der Metamorphose durch Bakkterien bei den Larven von *Hydractina echinata*. *Zool Jahrb Abt Anat Ontog Tiere* 86:84–95.

Muralidharan K, Wakeland EK (1993): Concentration of primer and template qualitatively affects products in random amplified polymorphic DNA PCR. *Bio techniques* 14:362–363.

Nei M, Li WH (1979): Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Nat Acad Sci USA* 74:5267–5273.

Palumbi SR, Martin A, Romano S, McMillan WO, Stice L, Grabowski G (1991): *The simple fool's guide to PCR (Version 2.0)*. Honolulu: Department of Zoology and Kewalo Marine Laboratory, University of Hawaii.

Peakall R, Smouse PE, Huff DR (1995): Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloë dactyloides*. *Mol Ecol* 4:135–147.

Penner GA, Bush A, Wise R, Kim W, Domier L, Kasha K, Laroche A, Scoles G, Molnar SJ, Fedak G (1993): Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods Applications* 2:341–345.

Piper WH, Rabenold PP (1992): Use of fragment-sharing estimates from DNA fingerprinting to determine relatedness. *Mol Ecol* 1:69–78.

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

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

Reeve HK, Westneat DF, Queller DC (1992): Estimating average within-group relatedness from DNA fingerprints. *Mol Ecol* 1:223–232.

Riedy MF, Hamilton WJ, Aquadro CF (1992): Excess non-parental bands in offspring from known pedigrees assayed using RAPD PCR. *Nucleic Acids Res* 20:918.

Roeder K, Devlin B, Lindsay BG (1989): Application of maximum likelihood methods to population genetic data for estimation of individual fertilities. *Biometrics* 45:363–379.

Roehrdanz RL, Reed DK, Burton RL (1993): Use of polymerase chain reaction and arbitrary primers to distinguish laboratory-raised colonies of parasitic Hymenoptera. *Biol Control* 3:199–206.

Rogers SO, Bendich AJ (1988): Extraction of DNA from plant tissues. *Plant Mol Biol Manual* A6:1–10.

Rossetto M, Weaver PK, Dixon KW (1995): Use of RAPD analysis in devising conservation strategies for the rare and endangered *Grevillea scapigera* (Proteaceae). *Mol Ecol* 4:321–329.

Schweder ME, Shatters RG Jr, West SH, Smith RL (1995): Effect of transition interval between melting and annealing temperatures on RAPD analyses. *Biotechniques* 19:38–42.

Shoemaker DD, Ross KG, Arnold ML (1994): Development of RAPD markers in two introduced fire ants, *Solenopsis invicta* and *S. richteri*, and their application to the study of a hybrid zone. *Mol Ecol* 3:531–539.

Shenk MA, Buss LW (1991): Ontogenetic changes in fusibility in the colonial hydroid *Hydractinia symbiolongicarpus*. *J Exp Zool* 257:80–86.

Smith JL, Scott-Craig JS, Leadbetter JR, Bush GL, Roberts DL, Fulbright DW (1994): Characterization of random amplified polymorphic DNA (RAPD) products from *Xanthomonas campestris* and some comments on the use of RAPD products in phylogenetic analysis. *Mol Phylogenet Evol* 3:135–145.

Swofford DL (1991): *Phylogenetic Analysis Using Parsimony (PAUP). Version 3.0s.* Champaign, IL: Illinois Natural History Survey.

Swofford DL, Selander RB (1989): *BIOSYS-1. Release 1.7.* Champaign, IL: Illinois Natural History Survey.

Thompson EA (1986): *Pedigree analysis in human genetics*. Baltimore: Johns Hopkins University Press.

Tinker NA, Fortin MG, Mather DE (1993): Random amplified polymorphic DNA and pedigree analysis in spring barley. *Theor Appl Genet* 85:976–984.

Weatherhead PJ, Montgomery RD (1991): Good news and bad news about DNA fingerprinting. *Trends Ecol Evol* 6:173–174.

Weiller G and AJ Gibbs (1995): DIPLOMO: The tool for a new type of evolutionary analysis. *CABIOS* 11: in press.

Weis VM, Buss LW (1987): Biology of hydractiniid hydroids. 5. Ultrastructure of metamorphosis in *Hydractinia echinata*. *Postilla* 200:1–24.

Welsh J, McClelland M (1990): Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* 18:7213–7218.

Welsh J, Honeycutt RJ, McClelland M, Sobral BWS (1991): Parentage determination in maize hybrids using arbitrarily primed polymerase chain reaction (AP-PCR). *Theor Appl Genet* 82:473–476.

Welsh J, Pretzman C, Postic D, Saint Girons I, Baranton G, McClelland M (1992): Genomic fingerprinting by arbitrarily primed polymerase chain reaction resolves *Borrelia burgdorferi* into three distinct phyletic groups. *Int J Syst Bacteriol* 42:370–377.

Westneat DF, Frederick PC, Wiley RH (1987): The use of genetic markers to estimate the frequency of successful alternative reproductive tactics. *Behav Ecol Sociobiol* 21:35–45.

Wilkerson RC, Parsons TJ, Albright DG, Klein TA, Braun MJ (1993): Random amplified polymorphic DNA markers readily distinguish cryptic mosquito species (Diptera: Culicidae: Anopholes). *Insect Mol Biol* 1:205–211.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990): DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18:6531–6535.

Williams JGK, Kubelik AR, Rafalski JA, Tingey SV (1991): Genetic data analysis with RAPD markers. In: Bennett JW, Lasure LL (eds). *More gene manipulations in fungi*. San Diego, CA: Academic Press, pp 433–439.

Williams JGK, Hanafey MK, Rafalski JA, Tingey SV (1993): Genetic analysis using random amplified polymorphic DNA markers. *Methods Enzymol* 218:704–740.

Yu K, Pauls KP (1992): Optimization of the PCR program for RAPD analysis. *Nucleic Acids Res* 20:2606.

Yund PO, Cunningham CW, Buss LW (1987): Recruitment and post-recruitment interactions in a colonial hydroid. *Ecology* 68:971–982.

Weiller GF, Gibbs A (1993): *DIPLOMO: Distance Plot Monitor version 1*. Computer Program distributed by the Australian National University. Available via WWW: http://life.anu.edu.au/molecular/software/rapd.html