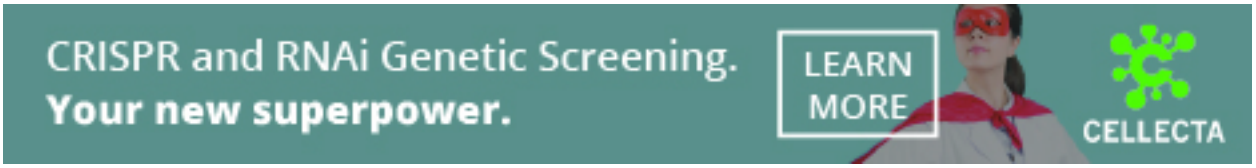
DNA fingerprinting by arbitrarily primed PCR.

M McClelland and J Welsh

Genome Res. 1994 4: S59-S65



DNA Fingerprinting by Arbitrarily Primed PCR

Michael McClelland and John Welsh

California Institute of Biological Research, La Jolla, California 92037

Fingerprinting of genomes with arbitrary primers[1,2] has proved a versatile method for detecting polymorphisms for genetic mapping, phylogenetics, and population biology. The method generates a fingerprint using arbitrarily selected primers under conditions where the primer will initiate synthesis on DNA even when the match with the template is imperfect. Some of these priming events occur in opposite strands. The most efficiently primed of these pairs of priming events compete with each other during amplification to produce a fingerprint of a few to >100 prominent PCR products. There have been more than 300 published papers that have used this strategy. In the last 4 years a multitude of variants and improvements have been developed. In this brief methods-oriented review, we will highlight some of these protocols and justify our current preferred protocol for DNA fingerprinting. Readers are also referred to some other recent reviews.[3–7] An accompanying article[8] will discuss RNA fingerprinting using arbitrarily primed PCR.[9,10]

INTRODUCTION TO FINGERPRINTING METHODS

In the literature much has been made of differences in primer length or in the gel system used for resolving PCR-based fingerprints. The fundamental principle that is being utilized does not change, but these differences do allow a number of subtle variations in the design of a fingerprinting experiment. When designing such an experiment, it is important to decide how many PCR products are wanted per lane and whether to use agarose, nondenaturing acrylamide, or denaturing acrylamide. One choice is to generate relatively few PCR products and to resolve these on an agarose gel. The more prominent of these products can then be scored quite reliably. Such fingerprints can be generated using 10-base primers and AmpliTaq in the manner first described by Williams et al.[2,5,11] Alternatively, a larger number of fragments can be generated and resolved on an acrylamide gel. There are a number of ways to achieve this:

1. Select, in a preliminary screen, 10-mers that yield a large number of fragments using AmpliTaq holoenzyme.
2. Use longer primers, such as 18-mers.[1]
3. Use primers that are biased toward more common sequences in the genome by performing a statistical analysis.
4. Very short primers used in very large amounts give extremely complex fingerprints.[12]
5. The protocol presented here uses arbitrary 10-mers in combination with the AmpliTaq Stoffel fragment (or KlenTaq). This enzyme generates almost twice as many PCR products as does AmpliTaq using the same primer and template DNA. Furthermore, this enzyme increases the number of arbitrary 10-mers that give productive fingerprints from ~75% to >90%.[13]

At least for primers of 10 bases or longer, the more complex the PCR fingerprint, the more reproducible it seems to be. However, too many products make the pattern difficult to interpret. Thus, reliable fingerprints can be obtained by screening arbitrary primers for moderately complex patterns and then using these on the population under study.

If a simple pattern is desired, then it is important to be aware of the "context effect". This situation arises for very simple patterns but not for complex patterns. The fingerprint is the result of a competition between many PCR products, and the fewer the winners the greater the effect of their presence and absence on the probability of other PCR products being amplified. For example, if two similar but nonidentical genomes give very simple fingerprints with the same primer and differ by one or more of these major products, then one must be concerned that the differences may affect the

4:S59–S65©1994 by Cold Spring Harbor Laboratory ISSN 1054-9805/94 $1.00

probability of amplification of other PCR products. One symptom of such a context effect in a mapping population would be nonparental products that appear as a result of the absence of a prominent polymorphism from one parent in some offspring. This phenomenon is a good reason to avoid using simple fingerprint patterns that contain <10 prominent PCR products.

For the purposes of fingerprinting, it is often best to use arbitrary primers in pairwise combinations.[14] A few primers can be used in a very large number of pairwise combinations. For example, 20 primers can be used in 380 different combinations. One potential disadvantage of using pairwise combinations of primers is that some of the products will have the same primer at both ends and these products may be shared between fingerprints using that primer. However, we have shown that the amount of overlap between fingerprints using a primer alone or in combination with a different primer is less than the 25% overlap that would be expected by applying simple statistics.[14] Most of the "winners" in the PCR have a different primer at each end. Among the possible explanations for this observation is the possibility that ssDNA products containing primer A and its complement, A′, can form a panhandle structure that is at a disadvantage for amplification when there is an alternative that cannot form panhandles.

One further advantage of using primers in pairwise combinations is the fact that the products can be directly sequenced using conventional PCR sequencing kits. However, these sequences are not always of the highest quality, perhaps because no part of a fingerprinting gel is entirely free of other products. Generally, it is best to clone the products first, as described.[14]

MOTIF SEQUENCES ENCODED IN PRIMERS

Completely arbitrary priming lies at one end of a spectrum of possible targeting strategies for fingerprinting. The other end of the spectrum uses primers derived from known perfect or near perfect dispersed repeats, for example, Alu–PCR,[15] tDNA–intergenic length polymorphisms[16] or REP–PCR.[17] In this spectrum lie a cornucopia of other repeats such as purine–pyrimidine motifs that have been successfully used to produce PCR fingerprints.[18] These microsatellite repeats are particularly useful because primers directed toward them reveal more polymorphisms than seen in the average arbitrarily primed fingerprint, making them particularly useful for detecting polymorphisms between closely related individuals.[18] Primer pairs directed toward tRNA genes are also useful because the tRNA gene clusters evolve more slowly than most of the rest of the genome, which is under less stringent selection pressure. The patterns produced by tDNA–ILP primers can be used to compare genomes at a higher taxonomic level than is possible with arbitrarily primed PCR.[16] Other primers biased to rarer conserved motifs, such as those that occur in some promoters or in gene families, have also been tried.[19] Finally, statistics can be applied to the known sequences in the DNA data base for a particular species or related species. This data can be used to develop primers that carry sequences that are rare or common in a genome. Such primers influence the number of PCR products in a fingerprint.

One might expect that primers directed against sequences such as purine–pyrimidine repeats or other dispersed repeats may generate significantly more reliable fingerprints than completely arbitrary primers because the interaction of the primer with the template will be better. In some cases, such as primers directed against tRNA motifs, we have confirmed a considerable bias by sequencing the products.[20,21] We also found that microsatellite repeat primers often amplify the expected microsatellite repeats. However, these microsatellite targets probably represent the minority of PCR products on the gel despite the fact that microsatellites are quite prevalent in most

genomes. The other products are arbitrarily primed at nonmotif locations.[18] Primers directed toward rarer motifs will have a correspondingly lower success rate.

The extent to which priming is biased to the expected motif depends on the number of good matches that occur in the genome. Arbitrary priming can occur at matches comprising as few as 5 matches out of 6 bases at the 3′ end of the primer. Thus, there is a huge number of potential PCR products competing with the putative targets of primers designed to match a particular dispersed motif. Among these targets, only the best primer–template matches are amplified, and these arbitrary matches can sometimes be quite good. For example, consider the frequency of a potential arbitrary PCR priming site; a 7- out of 10-base match with the primer that occurs twice in opposite orientations, spaced <1000 bases apart. Such pairs of potential priming sites occur ~200 times in a typical bacterial genome and almost half a million times in a higher eukaryotic genome. Among these potential arbitrarily primed PCR products, only the most efficient initial priming events and most efficient amplifying events will be visible. Thus, only in those few cases where the repeat is very abundant and almost perfect is a fingerprint generated by a motif primer likely to be derived primarily from the motif sites. Even in these cases there may be some products derived from among the best arbitrarily primed sites.

In those cases where the motif is poorly conserved or where the motif is rare, it might seem logical to attempt to increase the stringency of the motif–PCR. However, such attempts, for example, hot start, have generally failed because the fingerprints become unreliable at higher stringency. Increasing the stringency for a poorly matched or rare motif still does not allow motif sites to dominate versus the best arbitrary events. Eventually, as the stringency is increased, the whole fingerprint becomes unreliable. Nevertheless, there is nothing to be lost by using a primer with a particular sequence. At worst, the fingerprint will be arbitrary.

WHAT GEL TO USE?

The reasons some researchers prefer agarose gels and simpler fingerprints are that they believe agarose gels are easier to set up and because agarose gels can be stained with ethidium, thereby avoiding radioactivity. When beginning a project on fingerprinting, it may be wise to run products on an agarose gel to ensure that the amplification has been robust. It is disappointing to run an acrylamide gel and later find that an error was made in the experiment. However, after the technology is mastered, the use of a protocol that generates a relatively complex fingerprinting pattern and the extra work of running a denaturing acrylamide gel are rewarded by much more and better resolved data than can be obtained using agarose. When such data are scored, there is a greater chance that PCR products of a particular size are homologs because there is almost single-base resolution of fragments. In addition, the use of denaturing gels eliminates the problem of uneven amplification of the two strands that, on a native gel, yield a double-stranded product and a fainter variable single-stranded product for the strand that is in excess.

We have also separated denatured DNA fingerprints on a native acrylamide gel. This strategy[22] combines single-stranded conformation polymorphisms (SSCP)[23] with PCR-based fingerprinting. An example is shown in Figure 1. SSCP gels allow increased confidence in scoring the usual presence/absence polymorphisms that result from arbitrarily primed PCR because each strand for these polymorphisms occurs at different parts of the gel, allowing two opportunities for scoring. Furthermore, SSCPs that could not be scored by other gel systems, that is, those that carry internal sequence polymorphisms,

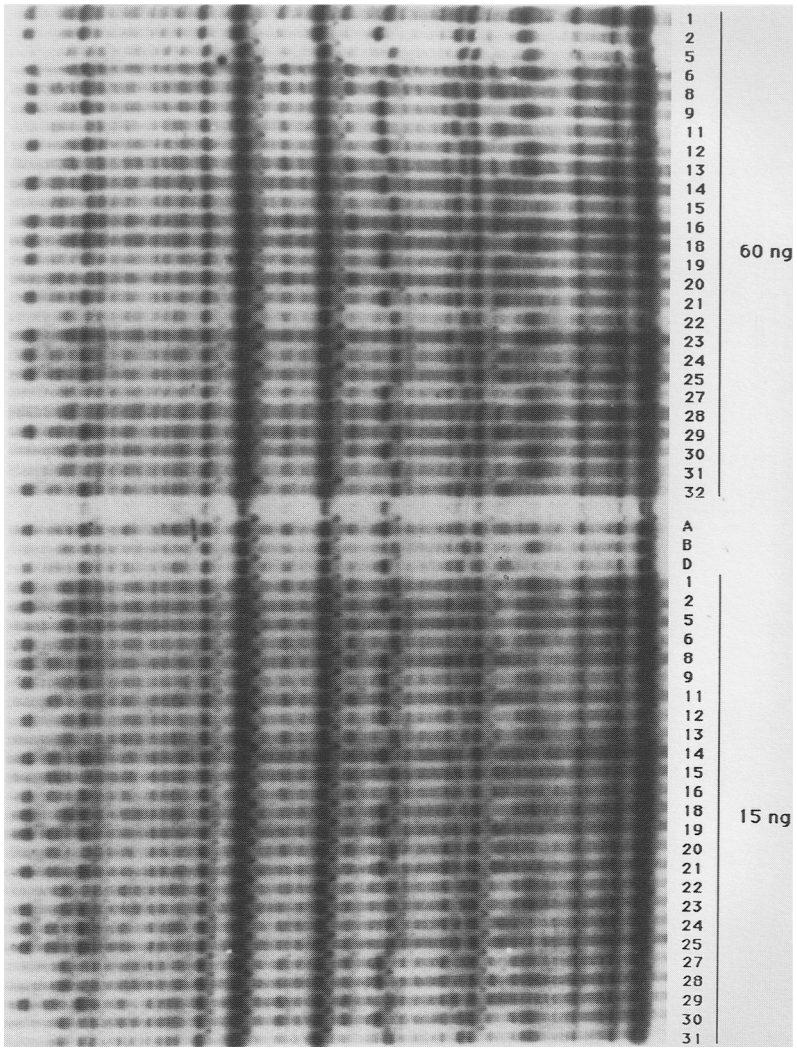


FIGURE 1 AP–PCR–SSCP for genetic mapping. Fifteen nanograms and 60 ng of genomic DNA from mouse strains in the C57BL/6J (B) vs. DBA (D) recombinant inbred mapping population were fingerprinted using the pair of primers S2 (5′-CCTCTGACTG) and S1 (5′-GAGGTCCACA) (Operon Tech. Inc., Alameda, CA) and the conditions described in the protocol. The fingerprints were resolved on 5% MDE + 5% glycerol, dried, then autoradiographed. Hydrolink-MDE was purchased from Baker, Phillipsburg, NJ. The numbers above each lane indicate the recombinant inbred lines. "A" indicates another mouse strain, A/J. Only the top half of the gel is shown. Fragments visualized range from ~400 bases to 800 bases. Numerous presence/absence and length polymorphisms are visible.

were also detected. It is worthwhile to use SSCP gels to increase throughput and reliability of scoring when mapping by PCR fingerprinting. However, because each strand occurs in a different part of the gel, SSCP is probably not appropriate for population biology and phylogenetics.

REPRODUCIBILITY

The issue of reproducibility is of much concern because some failures in PCR-based fingerprinting have been published and may discourage others from using the method. One concern is that the patterns may vary from day to day or from lab to lab (e.g., Ref. 24). The other concern is unreliability within the same experiment on the same day.

The problem of intraexperiment variability has been overstated. First, al-

most all these problems are the result of inadequately prepared DNA. The easiest way to find out if the DNA is of sufficient quality is to perform twofold serial dilutions of the DNA over a wide range from ~200 ng to 200 pg. If the DNA does not produce reliable fingerprints over a number of twofold dilutions with a number of different primers, then the DNA quality is suspect. Second, primers that give moderately complex patterns should be used, as described earlier. High-quality fingerprints have been obtained from genomes in every kingdom and over a wide spectrum of G+C contents. Thus, a failure of this kind must be attributed to inadequate DNA or reagents. We have reached the somewhat surprising conclusion that the reliability of the fingerprints seems to derive primarily from its complexity rather than the quality of the match of the primer with the template.

One cannot know if the difference between two genomic fingerprints is real if the experiment is not controlled for DNA quality and quantity. Thus, every experiment must include fingerprinting for at least two concentrations of genomic DNA for each individual. Any differences between individuals that do not occur at both genomic DNA concentrations are rejected. Too many concentration-dependent differences should lead to concern about the DNA or reagents.

Day to day variation and interlab variation is a more genuine concern. Such variation occurs because all PCR-based fingerprinting varieties are sensitive to the buffer conditions, enzyme quality, and the primer preparation. Although this is easy to control in a particular experiment, it is harder to control between experiments. The ratio of intensities among products within a single fingerprint lane may vary from day to day. However, as long as the fingerprint pattern is complex, the variation does not extend to variability in the presence or absence of bands from day to day and the ratio of intensities between lanes does not change. These differences between experiments and between experimenters are generally rather subtle so they can be accommodated by the simple expedient of fingerprinting DNA from reference strains on each gel.

A PROTOCOL FOR USING 10-MERS IN PAIRWISE COMBINATIONS

DNAs should be of similar quality and of at least 10-kb average length. Surprisingly, it is mainly consistency in DNA quality that is most important, so for example, one can use crude lysates of bacterial DNA if all the strains are treated in the same way.[1,3]

Equipment and Reagents

Thermal Cycler (Perkin-Elmer 9600 model)

PCR tubes (Microamp, Perkin-Elmer)

2× AP–PCR Reaction Mixture (20 mM Tris at pH 8.3, 20 mM KCl, 10 mM $MgCl_2$, 0.2 mM each dNTP, 0.1 U/μl of *Taq* polymerase Stoffel fragment (Perkin-Elmer, Branchburg, NJ), 0.1 μCi/μl [α-^{32}P]dCTP and 0.8 μM primer.

Two arbitrary 10-mer primers (Genosys, Woodland, TX)

Method

All procedures can be set up on ice or at room temperature, as convenient. Reactions are prepared at two DNA concentrations. It is wise to initially titrate the DNA over two orders of magnitude to find the concentrations that give the most robust fingerprints. For mammalian DNA, the best fingerprints are usually obtained in the range of 5–50 ng per 20 μl of reaction volume. The optimal concentration of the primer must be determined empirically. For most 10-mers, the optimum is around 0.4 μM. Note, however, that we have

not been able to fingerprint bacterial genomes using pairs of primers except at high template concentrations (>1 μg). Perhaps, for simple genomes the primer dimers compete effectively. We recommend that only one primer be used at 0.4 μM for small genomes.

1. DNA is prepared at 2× final concentration and 10 μl distributed to tubes.
2. 10 μl of 2×AP–PCR Reaction Mixture is added. The volume proportions of DNA to reaction mixture can, of course, be changed, as can the total final reaction volume as long as the final concentrations of reaction components are kept the same. Many laboratories run 10-μl reactions.
3. The thermocycling profile for 10-mers is: 94°C for 1 min, 35°C for 1 min, 72°C for 2 min for 40 cycles.
4. The products are diluted 1:4 in 80% formamide containing 10 mM EDTA and tracking dye, heated to 65°C for 15 min and 2 μl electrophoresed through a denaturing sequencing-type polyacrylamide gel or an MDE gel (Baker, Phillipsburg, NJ). Sequencing gels are electrophoresed in 1.0× TBE at 50 W for 3 hr at 65°C. MDE gels are electrophoresed in 0.5× TBE at 7 W for 18 hr at room temperature. The gel may be dried and exposed to X-ray film without hindering the ability to clone fragments from the gel. Alternatively, the radioactive label can be omitted entirely and native agarose gels followed by ethidium bromide staining can be used. Note also that *Taq* polymerase Stoffel fragment seems to give much better fingerprints, with more bands and greater primer quality/concentration independence. Also, when primers are used in pairwise combinations to generate largely unique patterns, the cost of generating many fingerprints is greatly reduced.
5. If desired, bands can be excised from the gel with a razor blade, eluted, reamplified, and cloned (see accompanying article on RNA fingerprinting).

ACKNOWLEDGMENTS

We thank Rhonda Honeycutt for helpful discussions and comments on the manuscript. This work was supported in part by National Institutes of Health grants AI34829, NS33377, and HG00456 to M.M. and AI32644 to J.W.

REFERENCES

1. Welsh, J. and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18:** 7213–7218.
2. Williams, J.G., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18:** 6531–6535.
3. Welsh, J. and M. McClelland. 1993. The characterization of pathogenic microorganisms by genomic fingerprinting using arbitrarily primed polymerase chain reaction (AP-PCR). In *Diagnostic molecular microbiology*. (ed. D.H. Persing, T.F. Smith, F.C. Tenover, and T.J. White), pp. 595–602. ASM Press, Washington, DC.
4. Bowditch, B.M., D.G. Albright, J.G. Williams, and M.J. Braun. 1993. Use of randomly amplified polymorphic DNA markers in comparative genome studies. *Methods Enzymol.* **224:** 294–309.
5. Williams, J.G., M.K. Hanafey, J.A. Rafalski, and S.V. Tingey. 1993. Genetic analysis using random amplified polymorphic DNA markers. *Methods Enzymol.* **218:** 704–740.
6. Clark, A.G. and C.M.S. Lanigan. 1993. Prospects for estimating nucleotide divergence with RAPDs. *Mol. Biol. Evol.* **10:** 1096–1111.
7. Yu, K.F., A. Van Denze, and K.P. Pauls. 1993. Random amplified polymorphic DNA (RAPD) analysis. In *Methods in plant molecular biology and technology* (ed. B.R. Glick and J.E. Thompson). CRC Press, Inc., Boca Rotan, Fl, USA; London, U.K.
8. McClelland, M. and J. Welsh. 1994. RNA fingerprinting by arbitrarily primed PCR. *PCR Methods Applic.* (this issue)
9. Welsh, J., K. Chada, S.S. Dalal, D. Ralph, R. Cheng, and M. McClelland. 1992. RNA fingerprinting by arbitrarily primed PCR. *Nucleic Acids Res.* **20:** 4965–4970.

10. Liang, P. and A. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257:** 967–971.
11. Williams, J.G., R.S. Reiter, R.M. Young, and P.A. Scolnik. 1993. Genetic mapping of mutations using phenotypic pools and mapped RAPD markers. *Nucleic Acids Res.* **21:** 2697–2702.
12. Caetano-Anolles, G., B.J. Bassam, and P.M. Gresshoff. 1992. Primer-template interactions during DNA amplification fingerprinting with single arbitrary oligonucleotides. *Mol. & Gen. Genet.* **235:** 157–165.
13. Sobral, B.W.S and R.J. Honeycutt. 1993. High output genetic mapping of polyploids using PCR-generated markers. *Theor. Appl. Genet.* **86:** 105–112.
14. Welsh J. and M. McClelland. 1991. Genomic fingerprinting with AP-PCR using pairwise combinations of primers: Application to genetic mapping of the mouse. *Nucleic Acids Res.* **19:** 5275–5279.
15. Nelson, D.L., S.A. Ledbetter, L. Corbo, M.F. Victoria, R. Ramirez-Solis, T.D. Webster, D.H. Ledbetter, and C.T. Caskey. 1989. Alu polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci.*. **86:** 6686–6690.
16. Welsh, J. and M. McClelland. 1991. Genomic fingerprints produced by PCR with consensus tRNA gene primers. *Nucleic Acids Res.* **19:** 861–866.
17. Versalovic, J., K. Thearith, and J.R. Lupski. 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic. Acids. Res.* **19:** 6823–6831.
18. Welsh, J., C. Petersen, and M. McClelland. 1991. Polymorphisms generated by arbitrarily primed PCR in the mouse: Application to strain identification and genetic mapping. *Nucleic Acids Res.* **19:** 303–306.
19. Birkenmeier, E.H., U. Schneider, and S.J. Thurston. 1992. Fingerprinting genomes by use of PCR primers that encode protein motifs or contain sequences that regulate gene expression. *Mamm. Genome* **3:** 537–545.
20. McClelland, M., C. Peterson, and J. Welsh. 1992. Length polymorphisms in tRNA intergenic spacers detected using the polymerase chain reaction can distinguish streptococcal strains and species. *J. Clin. Microbiol.* **30:** 1499–1504.
21. Welsh, J. and M. McClelland. 1992. PCR-amplified length polymorphisms in tRNA intergenic spacers for categorizing staphylococci. *Mol. Microbiol.* **6:** 1673–1680.
22. McClelland, M., H. Arensdorf, R. Cheng, and J. Welsh. 1994. Arbitrarily primed PCR fingerprints resolved on SSCP gels. *Nucleic Acids Res.* **22:** 1770–1771.
23. Hayashi, K. 1991. PCR-SSCP: A simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods Applic.* **1:** 34–38.
24. Meunier, J.R. and P.A.D. Grimont. 1993. Factors affecting reproducibility of amplified polymorphic DNA fingerprint. *Res. Microbiol.* **144:** 373–379.