

# Estimating Nucleotide Diversity From Random Amplified Polymorphic DNA and Amplified Fragment Length Polymorphism Data

Richard L. Borowsky

Department of Biology, New York University, Washington Square, New York, New York 10003

Received July 31, 2000

**A way to estimate the index of nucleotide diversity ( $\pi$ ) from band match frequencies in random amplified polymorphic DNA and amplified fragment length polymorphism data is described.  $\pi$  is shown to be a simple function of the proportion of mismatched bands between two individuals drawn at random from a population ( $\phi$ ) and the number of discriminating sites in the amplification system. The method is computationally and conceptually simple and avoids some of the assumptions inherent in other approaches: the relationship is independent of the base composition of the target DNA and avoids the bias inherent in estimations of allelic frequencies in dominant systems. Only two individuals from a population are needed to estimate  $\pi$ . This economy of material suggests utility of this approach in conservation genetics or other fields where obtaining large samples is impractical or undesirable.** © 2000 Academic Press

**Key Words:** nucleotide diversity; RAPD; AFLP.

## INTRODUCTION

Genomic variability is an attribute of all biological populations. In diploid organisms it is the basis for heterozygosity and individual physiological flexibility via heterosis, and in all organisms it is the basis for population genetic diversity and evolutionary potential. A fundamental and universally applicable measure of genomic variability is nucleotide diversity ( $\pi$ ). Nucleotide diversity is defined as "the average number of nucleotide differences per site between two randomly chosen DNA sequences" in a population (Nei and Li, 1979). In consequence, nucleotide diversity is of considerable interest to conservationists, population geneticists, and evolutionists and its efficient assessment in natural populations is a research priority.

Random amplified polymorphic DNA (RAPD) (Welsh and McClelland, 1990; Williams *et al.*, 1990) and amplified fragment length polymorphism (AFLP) data (Vos *et al.*, 1995) have great potential for efficient es-

timation of  $\pi$  because they sample large numbers of unlinked or loosely linked sites. Whereas they differ in technical detail, both methods produce individual specific profiles from template DNA by amplifying anonymous fragments from sites scattered throughout the genome. The individuality reflects nucleotide diversity in primer complementary sites (RAPD) or restriction sites and primer extensions sites (AFLP). Because both RAPD and AFLP scan loci throughout the genome and allow enormous amounts of information to be obtained with minimal effort, surveys with multiple primers can give robust population estimates of genetic diversity relatively unbiased by local genomic variation, compared to estimates that might come from sequence data.

Methods have been developed to estimate  $\pi$  from both types of data, although they have acknowledged drawbacks of analytical complexity, assumptions about the nucleotide composition of target DNA, and requirements for relatively large sample sizes. Clark and Lanigan (1993) provide a method for estimating nucleotide diversity from RAPD data, but its application to diploid species requires the estimation of allelic frequencies. Estimates of allelic frequencies for both RAPDs and AFLPs are biased because of dominance (Lynch and Milligan, 1994) and no single unbiased estimator has been found (Jorde *et al.*, 1999). The smaller the sample size, the greater the bias, and Clark and Lanigan (1993) state that their method of estimating  $\pi$  should be used when "many individuals from a single population are examined."

Innan *et al.* (1999) provide a method for estimating nucleotide diversity from AFLP data. Their method of estimation is based on assumptions about the GC content of the genome screened and involves a series of calculations including an equation too complex to solve by hand, although it can be implemented by computer. They estimate nucleotide diversity in wild yams, lentils, and soybeans from AFLP data using sample sizes ranging from 4 to 16 individuals.

In this paper, RAPD or AFLP band diversity is

shown to be a simple function of  $\pi$ . Although the function's value varies with the distribution of allelic frequencies among loci, it is remarkably stable over the extremes of symmetrical distributions.

Thus, band survey data can be converted easily to equivalent measures of nucleotide diversity. The method has the merits of computational and conceptual simplicity and a requirement for sample sizes as small as two individuals.

Because large numbers of populations could be typed with minimal effort, the method makes feasible large-scale species or even community comparisons. For example, this method was used to compare nucleotide diversities among 17 populations of cave and surface balitorid fishes and revealed a significant deficit of nucleotide diversity in the cave species (Borowsky and Vidthayanon, 2000).

### ESTIMATION METHOD

We define the Phenotypic Heterogeneity Index,  $\Phi$  or  $\phi$ , as the proportion of unmatched RAPD bands between two individuals randomly chosen from the population. This definition parallels that of  $\pi$ , the index of nucleotide diversity.

In the following, it is assumed that the distinctions between RAPD positive and RAPD negative alleles are due to single base pair substitutions in the priming sites. RAPD alleles do sometimes vary in band size, but this is relatively rare compared to presence/absence differences (Cognato *et al.*, 1995; Pillay and Kenny, 1995).

It is also assumed that nucleotide diversity is low enough so that there are only two allelic alternatives at each variable site. Traditionally, a RAPD or AFLP locus is recognized by the presence of an amplified fragment on a gel, indicating local complementarity to selecting sequences. Implicit in the stated assumption is an expansion of the definition of RAPD/AFLP loci to include regions that fall short of complementarity by a single base. This definition recognizes the existence of loci with band allele frequencies of zero, but limits them to relevant sequences that might have arisen from or given rise to band positive sequences through a single mutational step. In the following treatment, it is also assumed that genotypes at marker loci are in Hardy-Weinberg proportions (HWP). Whereas this is a reasonable assumption for stable populations, the method should be applied cautiously where there is reason to suspect that HWP might not obtain.

Considering only the *i*th RAPD site, where  $p_i$  is the frequency of the RAPD positive allele and  $q_i = (1 - p_i)$  is the frequency of the RAPD minus alternative, the site's contribution to the estimate of  $\pi$  is  $2q_i(1 - q_i)$ . Its contribution to  $\phi$  is two times the frequencies of the bandless and banded phenotypes,  $2q_i^2(1 - q_i^2)$ . Thus, for the *i*th locus,

$$\frac{\phi_i}{\pi_i} = q_i(1 + q_i) \quad (1)$$

and, averaged over all loci,

$$\phi = \frac{1}{n} \sum_{i=1}^n 2q_i^2(1 - q_i^2) \quad (2)$$

and

$$\pi = \frac{1}{n} \sum_{i=1}^n 2q_i(1 - q_i). \quad (3)$$

We require  $R = \phi/\pi$ . The value of  $R$ , however, will vary with the distribution of allelic frequencies among RAPD loci,  $\rho(q)$ , which, generally, is unknown. That is,

$$\pi = \int_1^u \rho(q) \cdot 2q(1 - q) dq \quad (4)$$

and

$$\phi = \int_1^u \rho(q) \cdot 2q^2(1 - q^2) dq \quad (5)$$

with lower and upper limits of integration  $1/2N$  and  $(2N - 1)/2N$ , where  $N$  is the breeding population size.

Using these relationships, we can calculate values of  $R$  over a range of reasonable  $\rho(q)$ : the simplest calculation is for the case in which all loci have an allelic frequency of  $q = 0.5$ . From Eq. [1],  $R = \frac{3}{4}$ . For the case in which the allelic distribution is uniform,  $\rho(q) = 1$ , we obtain  $\phi = 2/15$ ,  $\pi = \frac{1}{6}$ , and  $R = \frac{4}{5}$ , from evaluation of Eqs. [4] and [5]. For a U-shaped allelic distribution, the classic approximation (Wright, 1931, pp. 122-127),  $\rho(q) = a/q(1 - q)$ , is computationally simple and yields  $\phi = \frac{5}{6}a$ ,  $\pi = a$ , and  $R = \frac{5}{6}$ , when substituted in Eqs. [4] and [5]. For a symmetrical distribution in which half the loci have  $q \rightarrow 0$  and half have  $q \rightarrow 1$ , Eq. [1] gives the result  $R \rightarrow 1$ . Thus, over a wide range of symmetrical distributions,  $R$  is restricted to the range 0.75 to 1.0.

In neutral systems, equilibrium allelic distributions are determined by the effective population size ( $N_e$ ) and mutation rate ( $\mu$ ). When  $4N_e\mu$  is smaller than unity, the two forces will result in classic Wrightian U-shaped distributions, with most loci fixed for one allele or another (Wright, 1931). Such distributions are likely for RAPDs, where the targets for mutation are on the order of  $10^1$  bases, mutation rates per base are on the order of  $10^{-9}$  (Li and Graur, 1991), and  $N_e \ll 10^7$ .

Thus, for most organisms, a reasonable value of  $R$  will be close to  $\frac{5}{6}$ .

It is necessary to correct for a bias in estimation of  $\phi$  before it can be used to estimate  $\pi$ . The relationship between  $\phi$  and  $\pi$  derived above is theoretical and assumes a complete knowledge of the RAPD loci amplified in the system. In practice, however, RAPD loci are only detectable when at least one individual sampled exhibits a band. Thus, in small samples a proportion of RAPD loci go undetected. This inflates the estimate of  $\phi$  and necessitates a correction of its empirically determined value ( $\phi_e$ ) based on the proportion of loci undetected.

The proportion of loci undetected ( $\eta$ ) is estimated as

$$\eta = \frac{\int_1^u \rho(q) \cdot q^{2k} \cdot dq}{\int_1^u \rho(q) \cdot dq}, \quad (6)$$

where  $k$  is the number of individuals sampled. The corrected value of  $\phi = \phi_e (1 - \eta)$ .

The value of  $\eta$  is only slightly dependent upon population size when  $k$  is small and nearly independent of it when  $k$  is large. For cases in which  $\rho(q) = a/q(1 - q)$  and only two individuals are compared ( $k = 2$ ),  $\eta = 0.445$  in very large populations and 0.38 when the population size is  $10^3$ . Where  $\rho(q) = 1$ ,  $\eta = 0.2$  over a broad range of population sizes.

Thus, assuming a U-shaped allelic distribution, a sample size of 2, and  $\eta = 0.38$ ,

$$\pi = \frac{6}{5} \cdot \frac{\phi_e \cdot (1 - \eta)}{m} \approx \frac{3}{4} \cdot \frac{\phi_e}{m}, \quad (7)$$

where  $m$  is the number of bases per band screened in the process. For RAPDs amplified with single primers,  $m$  is two times the number of bases in the primer. For RAPDs amplified with primer pairs,  $m$  is the sum of the bases in both primers. For AFLP data,  $m$  is the total number of bases in both restriction sites plus the bases in the selective extensions.

$\phi_e$  is a binomial variable and its variance can be estimated as  $\phi_e \cdot (1 - \phi_e)/n$ , where  $n$  is the total number of bands scored. Alternatively, if several independent estimates of  $\phi_e$  are available for a population, these can be used to estimate the parameter's variance. The second procedure is more conservative and was used in the example below, because it does not assume that each band represents a separate locus. Standard errors of estimate for  $\pi$  are obtained from those of  $\phi_e$ .

### ALLELIC DISTRIBUTIONS AT RAPD LOCI IN NATURAL POPULATIONS

Two technical difficulties hinder the empirical determination of allelic distributions at RAPD and AFLP

loci in natural populations. (1) Because banding alleles are dominant, allelic frequency estimates from diploids are biased (Jorde *et al.*, 1999). This bias can be eliminated by typing haploid stages or decreased in diploid data by statistical correction and increasing sample size. (2) The second problem is detection of loci with rare banding alleles. Because a RAPD/AFLP locus is recognized by the presence of a band, loci with low band allele frequencies may go undetected. Loci, as defined in the section on estimation method, with band allele frequencies of 0.0 will be undetectable in surveys of single populations. Thus, even if allelic distributions are U-shaped, surveys of single populations will reveal J-shapes because of a bias against detection of loci with low band allele frequencies. A partial solution to this problem is to survey multiple populations that have diverged to the extent that loci undetectable in some can be seen in others where band allele frequencies are higher. Despite these difficulties, preliminary data exist that support the expectation that RAPD/AFLP loci have U-shaped allelic distributions.

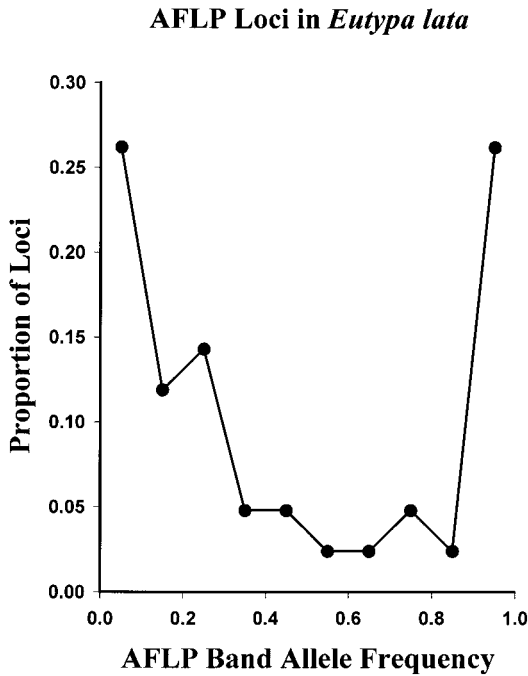
Peros *et al.* (1997) scored AFLP variation in 55 isolates of the fungus *Eutypa lata*. Because haploid material was typed and the sample size was large, allelic frequencies were determined without bias resulting from dominance and rare banding classes were detected. Band homology was tested by Southern hybridization and some cases of allelic size variability were reported, but this was considered a minor source of variation in the data set. Allelic frequencies of polymorphic loci were taken from Table 2 of Peros *et al.* (1997) and the unreported numbers of invariant loci were estimated from their Fig. 2. The allelic distribution of AFLP loci in *E. lata* is plotted in Fig. 1. It is clearly U-shaped.

Allelic distributions of RAPD loci were assessed by population surveys of species from two families of freshwater fishes. Population allelic frequencies were calculated from phenotypic frequencies using a correction for estimation bias in dominant systems due to Haldane and Li (Li, 1961). This correction is almost identical in magnitude to that proposed by Jorde *et al.* (1999).

Figure 2 show the allelic distribution of RAPD loci in a single population of the Mexican tetra, *Astyanax mexicanus* (Characidae). In this case, bands in the "zero class" were identified by their presence in neighboring populations. The U-shaped nature of the distribution is apparent.

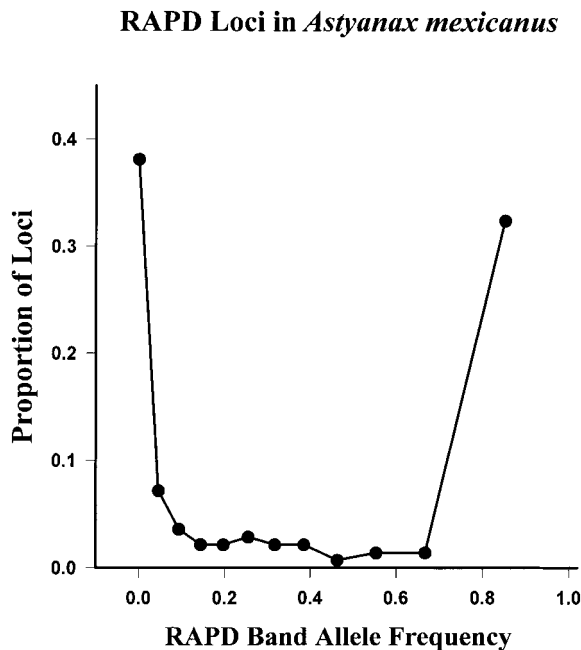
Figure 3 shows the distribution of loci in five populations of the poeciliid fish *Xiphophorus variatus*. The distribution is truncated on the left side because of the low number of populations biased against detection of zero and rare band loci. Nevertheless, the J-shaped distribution is almost U-shaped and distinctly elevated at both extremes.

Although these three examples do not prove that all

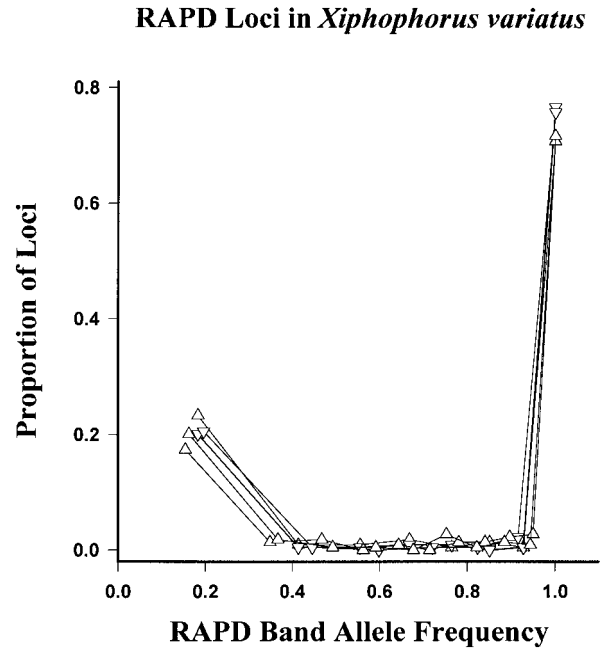


**FIG. 1.** Allelic distributions of AFLP loci in a population of the fungus *Eutypa lata*; data from Peros *et al.*, (1997).

RAPD/AFLP distributions are U-shaped, they give empirical support to the theoretical expectation. Clearly, more research is needed to establish the population genetics of RAPD/AFLP loci.



**FIG. 2.** Allelic distributions of RAPD loci in the Caballo Moro population of the Mexican Tetra, *Astyanax mexicanus* (Characidae). Data from a screen of 11 individuals using two primers and visualized on silver-stained PAGE gels (data from Espinasa and Borowsky, 2000).



**FIG. 3.** Allelic distributions of 169 RAPD loci in the poeciliid fish *Xiphophorus variatus*. Unpublished data from a survey of five small populations (40 individuals total) screened with six primers. RAPD fragments were visualized on agarose gels with ethidium bromide staining.

## EXAMPLES

Innan *et al.* (1999) estimate  $\pi$  in the wild yam, *Dioscorea tokoro*, using their method on AFLP data, as  $5.5 \cdot 10^{-3}$ . Data that they cite on sequences of coding and intron regions from nuclear genes give an average estimate of  $\pi = 4.0 \cdot 10^{-3}$ . Reanalyzing their AFLP data with Eq. [7] above gives an estimate of  $\pi = 2.3 \cdot 10^{-3}$ . Thus, all three methods are in close agreement.

Estimates of  $\pi$  based on RAPDs for seven different populations of five different species of balitorid hill stream loaches from Thailand are presented in Table 1. Using the methods outlined above,  $\pi$  is estimated to vary among these species from about 1 to  $10 \cdot 10^{-3}$ . Although only a single primer pair was used to obtain the data, and relatively few RAPD bands were scored (98–145), the estimates are relatively consistent among replicates of a population and among populations of a species. Also, the estimates of  $\pi$  are close to others made with RAPD/AFLP in eukaryotes (average =  $13.5 \pm 3.2 \cdot 10^{-3}$ ; data from Harada *et al.*, 1994; Martinez-Torres *et al.*, 1997; Silveira *et al.*, 1998; Innan *et al.*, 1999; Verovnik *et al.*, 1999).

## DISCUSSION

The method outlined above for estimating  $\pi$  from RAPD and AFLP data is straightforward and avoids some of the complications inherent in other methods

TABLE 1

**Indices of RAPD Band Sharing ( $\phi_e$ ) and Estimates of Nucleotide Diversity ( $\pi$ ) in Freshwater Fishes from Thailand (Balitoridae)**

Species	$n_k$	$\phi_k$	$\phi_e \pm SE$	$\pi \pm SE$
<i>Balitora burmanica</i>	113	13.3		
	98	8.2		
	116	5.2	$8.9 \pm 2.4$	$3.1 \pm 0.9$
<i>Homaloptera smithi</i>	127	26.8		
	124	30.3		
	123	22.8	$26.6 \pm 2.2$	$9.4 \pm 1.0$
<i>Schistura poculi</i> (pop. 1)	122	11.5		
	124	21.0	$16.3 \pm 4.7$	$5.8 \pm 1.7$
<i>Schistura poculi</i> (pop. 2)	116	12.7		
	137	21.2		
	125	20.0	$18.0 \pm 2.7$	$6.4 \pm 0.9$
<i>S. maepaiensis</i>	113	4.4		
	132	6.1	$5.3 \pm 0.8$	$1.9 \pm 0.3$
<i>S. robertsi</i> (pop. 1)	145	4.1		
	128	2.3	$3.3 \pm 0.9$	$1.2 \pm 0.4$
<i>S. robertsi</i> (pop. 2)	105	7.6		
	105	7.6	$7.6 \pm 0.0$	$2.7 \pm 0.0$

Note.  $n_k$  and  $\phi_k$  are the number of bands scored in the  $k$ th independent pair of individuals in a population and the value of  $\phi_e$  for the pair. " $\phi_e$ " is the population average of  $\phi_k$ , weighted by  $n_k$ ;  $\phi_k$ ,  $\phi_e$ , and their standard errors times  $10^2$ ;  $\pi$  and its standard error times  $10^3$ .  $^{33}\text{P}$ -labeled fragments were resolved on 4.5% sequencing gels and visualized by autoradiography.

already suggested. In particular, it avoids the necessity to correct for bias in estimations of allelic frequencies and makes no assumptions about the nature of the mutation process or about the base composition of the genome screened.

Because RAPD and AFLP are both able to screen large numbers of unlinked or loosely linked loci, the method provides a genome-wide estimate of  $\pi$ , rather than the locus-specific or region-specific estimates generally obtained from sequence or restriction data. In addition, estimates of variation obtained from large numbers of loci may be very accurate, even with small sample sizes (Nei, 1978). Surveys of balitorid hill stream loaches using 11 primers and over 1000 loci per population give estimates of  $\pi$  that are consistent and stable using as few as two individuals (Borowsky and Vidthayanon, 2000).

No method is better than the assumptions that it incorporates and those underlying this analysis should be examined. One important assumption is that genotypes at marker loci are in Hardy-Weinberg proportions. HWP is a reasonable expectation for RAPD and AFLP loci because, in such large numbers scattered throughout the genome, the vast majority of the variants are effectively neutral. Also, RAPD/AFLP variation generally causes no external phenotype that can be the basis for assortative mating. Nevertheless, small and locally structured populations may not exhibit HWP and this method should be applied with

caution in such cases or in any case where sampling may not be random with respect to genetic variation.

The assumption of a U-shaped distribution of allelic frequencies is also incorporated into Eq. [7], but preliminary data support this theoretical prediction. Also, the exact shape of the distribution is of little consequence to the ratio of  $\phi$  to  $\pi$ , at least for symmetrical distributions.

Finally, the method assumes that most variation at RAPD/AFLP loci is bi-allelic (+/-) with one band allele and one bandless allele. Apparent examples of size variation at such loci are documented in the literature (Cognato *et al.*, 1995; Peros *et al.*, 1997; Pillay and Kenny, 1995) and familiar to all who work with these techniques. Nevertheless, size variation is relatively rare compared to +/- variation, at least for RAPDs.

Furthermore, apparent examples of size variation are typically documented by DNA hybridization. Hybridization can prove sequence similarity, but not homology. It can never prove allelic identity. A definitive way to test the hypothesis of allelic identity of band size variants is genetic mapping. Here, RAPD and AFLP have proved so useful as mapping markers precisely because most of the variation is +/- . For example, in our mapping work with *Astyanax mexicanus* (unpublished), only 2 of 136 mapped RAPD loci showed perfect linkage (in repulsion). This is a low enough proportion of band size variation ( $1/135 = 0.008$ ) to ignore in calculations of  $\pi$ .

AFLP bands can also be lost by mutations that create restriction sites internal to the original fragment under consideration. Because AFLP is often performed with restriction enzymes that recognize four-base sequences, this is a potentially significant source of variation. In essence, this mechanism for the creation of "bandless" alleles increases the effective number of bases screened by the process. The method of estimating  $\pi$  proposed here should be applied cautiously to AFLP data until the magnitude of this potential source of variation has been assessed. This caveat does not apply to RAPD data.

In summary, RAPD and AFLP techniques are able to extract considerable information about genome-wide nucleotide variability. Analyzed judiciously, such data can facilitate the rapid and minimally obtrusive assessment of genetic variability and evolutionary potential in endangered species and other taxa where sample sizes are a limiting factor.

## ACKNOWLEDGMENTS

This work was supported by grants from the National Science Foundation (INT9605200), the H. R. Axelrod Foundation, and The Research Challenge Fund (NYU). Technical work by Juliya Starobinets is gratefully acknowledged. Rob Desalle and Dan Tranchina provided helpful comments on earlier versions of this paper.

## REFERENCES

- Borowsky, R., and Vidhayanon, C. (2000). Nucleotide diversity in populations of balitorid cave fishes from Thailand.
- Clark, A. G., and Lanigan, C. M. S. (1993). Prospects for estimating nucleotide divergence with RAPDs. *Mol. Biol. Evol.* **10**: 1096–1111.
- Cognato, A. I., Rogers, S. O., and Teale, S. T. (1995). Species diagnosis and phylogeny of the *Ips grandicollis* group (Coleoptera: Scolytidae) using random amplified polymorphic DNA. *Ann. Entomol. Soc. Am.* **88**: 397–405.
- Espinasa, L., and Borowsky, R. (2000). Eyed cave fish in a karst window. *J. Cave Karst Studies*, in press.
- Harada, K., Kinoshita, A., Shukor, N. A. A., Tachida, H., and Yamazaki, T. (1994). Genetic variation estimated in three *Shorea* species by RAPD analysis. *Jpn. J. Genet.* **69**: 713–718.
- Innan, H., Terauchi, R., Kahl, G., and Tajima, F. (1999). A method for estimating nucleotide diversity from AFLP data. *Genetics* **151**: 1157–1164.
- Jorde, P. E., Palm, S., and Ryman, N. (1999). Estimating genetic drift and effective population size from temporal shifts in dominant gene marker frequencies. *Mol. Ecol.* **8**: 1171–1178.
- Li, C. C. (1961). "Human Genetics: Principles and Methods," McGraw-Hill, New York.
- Li, W.-H., and Graur, D. (1991). "Fundamentals of Molecular Evolution," Sinauer, Sunderland, MA.
- Lynch, M., and Milligan, G. G. (1994). Analysis of population genetic structure with RAPD markers. *Mol. Ecol.* **3**: 91–99.
- Martinez-Torres, D., Carrio, R., Latorre, A., Simon, J. C., Hermoso, A., and Moya, A. (1997). Assessing the nucleotide diversity of three aphid species by RAPD. *J. Evol. Biol.* **10**: 459–477.
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.
- Nei, M., and Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**: 5269–5273.
- Pillay, M., and Kenny, S. T. (1995). Anomalies in direct pair-wise comparisons of RAPD fragments for genetic analysis. *Biotechniques* **19**: 694–698.
- Silveira, E. B., Al-Janabi, S. M., Magalhaes, B. P., Carvalho, L. J., and Tigan, M. S. (1998). Polymorphism of the grasshopper *Schistocerca pallens* (Thunberg) (Orthoptera: Acrididae) and its natural pathogen *Metarhizium flavoviride* (Gams and Rozsypal) (Hyphomycetes), revealed by RAPD analysis. *Anais Soc. Entomol. Brasil* **27**: 91–99.
- Verovnik, R., Trontelj, P., and Sket, B. (1999). Genetic differentiation and species status within the snail leech *Glossiphonia complanata* aggregate (Hirudinea: Glossiphoniidae) revealed by RAPD analysis. *Archiv. Hydrobiol.* **144**: 327–338.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**: 4407–4414.
- Welsh, J., and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18**: 7213–7218.
- Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A., and Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**: 6531–6535.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.