An Inquiry-Based Introduction to Molecular Biology

Foster Levy

An attraction of genetics instruction has traditionally been in its inquiry-based approach. More recently, the emphasis in genetics research and instruction has undergone a transition from examination of phenotypes to analyses of the macromolecules important for encoding (DNA) and decoding (RNA) information and regulation of information processing. With the incorporation of hands-on molecular biology into both secondary school and introductory level undergraduate curricula, the fascination with technology need not replace the inquiry-based approach long associated with genetics. Many students are currently introduced to molecular biology by running gels that allow them to "see" DNA. A demonstration of the technique of electrophoretic separation often takes precedence over the process of posing and answering questions using scientific reasoning. Kits are sold that include the hardware and reagents for running DNA gels and many include restriction endonucleases for fragmenting DNA. But, in the most common introductory molecular exercises, students "look" at uncut and cut DNA, and possibly measure fragment sizes, activities that are essentially observational, descriptive and intellectually passive for the students. The purpose of this report is to show that molecular biology, even in an introductory experience, can be presented in a more scientific framework. In fact, more can be done with less, in a more investigative setting!

To highlight investigative approaches, two themes will be emphasized: a deductive determination of the nature of the nucleic acids visualized in a gel, and a comparison of different genomes. Why should students be asked to "take it on faith" that what they view on a gel is DNA? With the use of simple DNA preparations to isolate total nucleic acids and inexpensive nucleases, students can examine and distinguish both DNA and RNA. A simple, inexpensive two-hour DNA mini-prep using higher plant material will provide high-quality DNA, preserve the various RNAs in usable form, and will not be destructive to the organism (Murray & Thompson 1980). The procedure requires no approval for animal experimentation and is safe as long as chemicals are handled with care. Similar mini-preps can be used to isolate nucleic acids from bacteria and plasmids (but these should not include steps that degrade RNA) (Sambrook et al. 1989).

Foster Levy is an Associate Professor in the Department of Biological Sciences, East Tennessee State University, Johnson City, TN 37614; e-mail: levyf@etsu.edu.

Identity of Nucleic Acids

Electrophoresing an untreated sample of a higher plant total nucleic acids preparation on a lane of a 0.8% agarose gel will show a large molecular weight band (> 20 kb), two or more small bands (approximately 1 to 2 kb), a very small (< 100 bp) blob, and a background smear that may span the gel lane (Figure 1). In an adjoining lane, students can visualize a sample from the same preparation after RNase

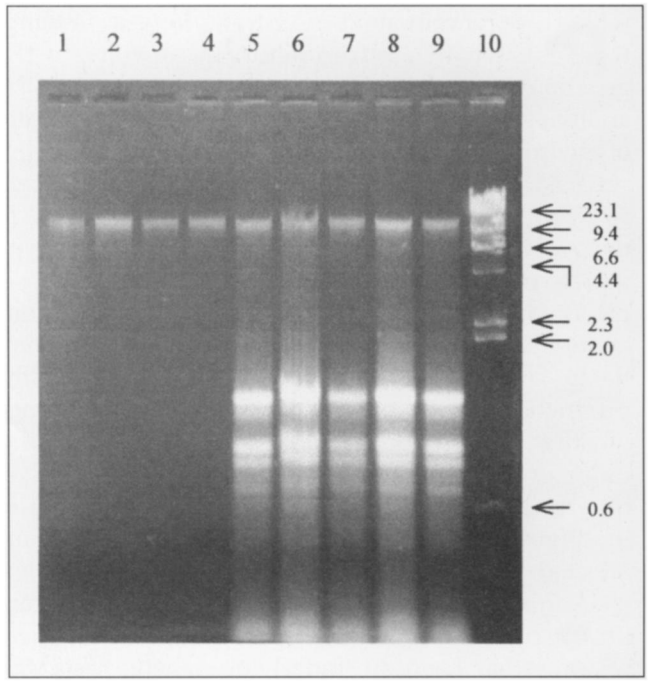

Figure 1. Total nucleic acid preparations from higher plants. Lanes 1-4, preps treated with RNase; lanes 6-9, untreated preps; lane 10, lambda *Hind* III digest with fragment sizes shown to the right in kilobase pairs. The DNA band is apparent in each lane of plant preps. Strong bands in the untreated preps (between the two lowermost lambda bands) represent ribosomal RNAs; the RNA blobs at the very bottom of untreated lanes correspond to tRNAs.

treatment. RNase is relatively inexpensive and fairly stable for long periods of time. Newly available forms of RNase can digest RNA in a matter of minutes so students can treat samples just prior to loading a gel. To ascertain sizes of DNA fragments on a gel, the location of bands of unknown size are compared to positions of bands with known fragment lengths. The gels depicted in Figures 1 and 2 use size standards consisting of fragments resulting from *Hin*d III digestion of the genome of bacteriophage λ.

After RNase treatment, all but the largest band disappear (Figure 1). A simple comparison of untreated and RNase-treated lanes should lead to a determination of the bands corresponding to RNA and an inference that the RNase has specificity to certain nucleic acids. The use of a stain such as ethidium bromide minimizes the likelihood that molecules other than nucleic acids are visualized on the gel. However, those with more resources may include a protease treatment to experimentally eliminate proteins as a source of the gel bands. With knowledge of the bands that correspond to RNA, knowledge-based reasoning is used to infer the types of RNA visualized in each region of the gel. The strong bands between 1 to 2 kb indicate the presence of (1) many copies of each RNA and (2) different sized RNAs—these bands most likely represent ribosomal RNAs (Figure 1). The diffuse blob between 75 to 150 bp is comprised of very small RNAs that suggests a heterogeneous mix consisting of the different transfer RNAs (Figure 1). Which of the RNAs visualized in the untreated lane is unaccounted for? The smear. Ask students what size RNA is expected for the remaining type of RNA, messenger RNA. This discussion should include reasoning that many different messages are produced in a cell and these messages are not expected to be of uniform size, hence the smear is an indicator of the heterogeneity of message sizes. The observation that the smear disappears after RNase treatment eliminates the possibility that the smear represents degraded DNA.

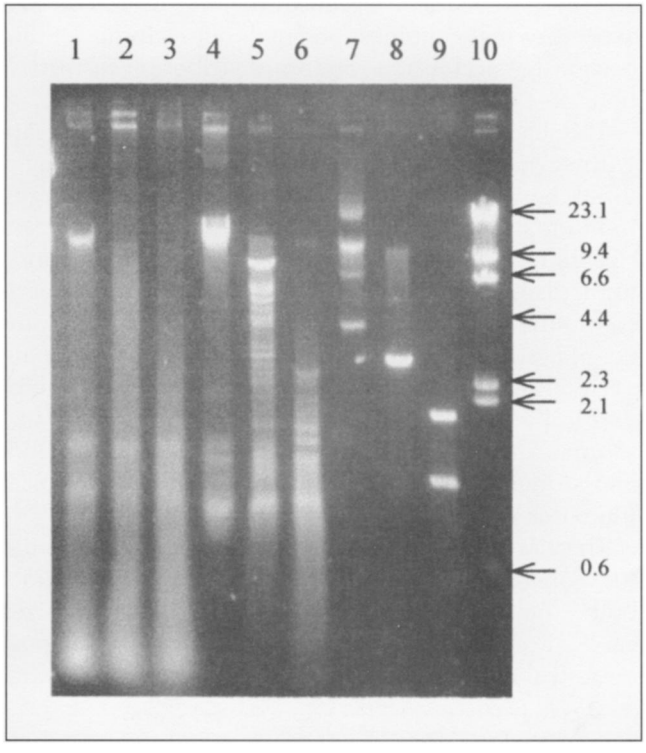

Figure 2. Three genomes, each represented by an uncut sample followed by six- and four-cutter digestions. Lanes 1-3, plant total nucleic acids; lanes 4-6, chloroplast DNA; lanes 7-9, plasmid DNA; lane 10, lambda *Hin*d III digest with fragment sizes shown to the right in kilobase pairs. The first lane of each genome sample is untreated, the second and third lanes are *Hin*d III (six-cutter) and Rsa I (four-cutter) digests respectively. Multiple bands in the uncut plasmid sample (lane 7) are due to differential migration of relaxed, coiled and supercoiled circular DNAs. Restriction enzyme digestions did not include RNase treatments.

The comparison of untreated and RNase-treated samples provides positive evidence for identifying RNA and leaves the large molecular weight band intact. However, negative evidence is not sufficient to conclude that what remains after RNase treatment must be DNA. A third sample from the same DNA prep can be treated with DNase. It is important to use an RNase-free DNase to preserve the RNA. In a lane containing a sample treated with DNase, students will see the loss of the large molecular weight band and persistence of the RNA bands identified in lane 2. At this point, there is positive evidence for identification of all components of the original prep. From these simple treatments, students will learn how to systematically refine hypotheses and the "look-see" nature of DNA gels can be converted into an investigatory procedure.

Genome Comparisons

The purpose of this exercise is to compare the relative sizes of various genomes and to examine the expected and observed effects of two different restriction endonuclease treatments on each genome. This exercise is novel in that restriction endonucleases provide the means to accomplish a goal as opposed to viewing restriction fragments as an end in itself. Relatively large quantities of DNA representing at least four different genomes (nuclear, bacterial, viral, plasmid) can be readily prepared with minor commitment of time using basic equipment or they can be purchased inexpensively. Using fresh leaf tissue from most plants, total nucleic acid preps (Murray & Thompson 1980) can be conducted entirely within microfuge tubes and these preps yield approximately 2 to 5 μg of DNA (approximately 100 ng is sufficient for a gel lane so each prep can supply at least 20 gel lanes). Moreover, 10 to 20 preps can be completed in a half day, giving sufficient DNA for several years

for most classroom uses. Similar yields are possible using standard bacterial (*E. coli*) and plasmid mini-prep procedures (Sambrook et al. 1989). DNA from a viral genome, that of bacteriophage lambda, can be purchased ($50 will supply enough DNA for many years). Moreover, lambda DNA serves a dual purpose: first, as an example of a viral genome, a feature which is often overlooked, and second, the common usage as a size marker on gels. I have been fortunate in having access to chloroplast and mitochondrial DNA in addition to these other four genomes. Researchers in your vicinity may be willing to share small quantities of organellar DNAs. Each prep can be treated with RNase because RNA has already been examined in the previous exercise.

Each genome will be examined uncut and digested with each of two kinds of restriction enzymes; one whose recognition sequence is six base pairs in length ("six-cutter") and one with a four base pair recognition sequence ("four cutter") (Figure 2). There is tremendous variation in sizes among these genomes (approximately 3 kb to many million base pairs), but it is the comparisons that are particularly instructive. The plant genome is so large that cutting with a restriction enzyme followed by electrophoresis and ethidium staining produces a DNA smear (Figure 2). By including RNase in restriction digests, the resultant smear must be DNA, unlike the mRNA smear observed in untreated samples. Students should be told the approximate size of the genome (10^7 to 10^8 bp) and asked to calculate the expected number of fragments for each enzyme. This is done by *assuming* each nucleotide base occurs with equal frequency in the genome (1 of 4) and that the four nucleotide bases are randomly distributed throughout the genome. Deviations from these assumptions do not alter the results significantly but they do stimulate important discussion (see below). Under the equality and even distribution assumptions, the expected incidence of a recognition site for a six-cutter enzyme is $(1/4)^6$ or approximately once per 4000 bp. A genome of 10^7 bp cut every 4×10^3 bp yields an expectation of 2500 cut sites. Ask students to draw, in pencil, just 100 bands on a gel "lane" on a piece of paper—they should quickly realize this many bands result in a smear. Similarly, digesting the large genome with the four-cutter produces an even higher expected number of fragments and a resultant smear. However, the expected distance between four-cutter recognition sites is less, $(1/4)^4$ or one site per 256 bp.

The design of the exercise, with an *a priori* expectation based on theory followed by observation, converts the laboratory from strictly observational to a hypothesis testing exercise. Deviations from the expectations must then be used to question and modify the assumptions. Students can estimate the mean size of fragments from each restriction digest by visually locating the midpoint of the smear, or measuring the length of the smear with a ruler and defining a midpoint. Comparison of the midpoint to a size standard, such as lambda DNA digested with *Hin*d III, provides an estimate of the average size of a restriction fragment after digestion with the respective enzymes. They can compare the means (the six-cutter mean should be larger than that of the four-cutter mean) to each other and to theoretical expectations. Two typical results, deviations from expectations as well as variation around the mean sizes, are topics worthy of discussion because they highlight two key aspects of real genomes: potential biases in nucleotide base usage and a non-random distribution of restriction sites.

The bacterial genome should also yield smears with each enzyme digest and the mean sizes should differ between enzymes. But even with a mini-gel, individual bands may be visible in the 1 to 7 kb size range. The ability to observe individual bands immediately suggests a smaller genome in comparison to the nuclear plant genome. However, the number of bands is too large to expect an accurate estimate of total genome size. Moreover, the resolution of "typical" 0.8% gels does not permit accurate size estimates for DNA fragments larger than 15 to 20 kb and students must be made aware of these technical limitations.

The viral and plasmid genomes provide opportunities for more quantitative analyses because after digestion, both generate a finite number of bands whose sizes can be accurately estimated (Figure 2). Students must be told the lambda genome is so well characterized

that the size of each restriction fragment is known and this information provides a size standard from which to estimate fragments of unknown size from other genomes. Students can use lambda DNA digested with *Hin*d III to estimate total viral genome size by summing the lengths of the individual fragments. Students should expect the cumulative size of fragments would sum to the genome size. However, the band representing uncut lambda DNA (50,000 bp) is outside the range of resolution for a 0.8% gel (a limitation of the technique). Therefore, the sum of the fragment sizes provides the best estimate of genome size. If time permits, students can be taught how to use lambda sizes to generate a size/distance plot that can be used to estimate sizes of the plasmid bands (Winchester & Wejksnora 1996). It is important to remember that most plasmids which are used as cloning vectors have only one recognition site for many enzymes and that if more bands are desired, appropriate enzymes must be chosen. Refer to biotechnology catalogs for detailed plasmid information.

There is no reason why an introduction to molecular biology should be presented without the scientific framework inherent in less technical aspects of biology. When preceded by a unit on classical genetics, as usually occurs in either introductory biology or genetics courses, students will have had exposure to the use of the χ^2 statistic to test observed segregation ratios against expectations. Both viral and plasmid DNAs yield finite numbers of restriction fragments that permit precise statistical tests of the goodness of fit of the observed number of fragments to the expectation. For example, the lambda genome at approximately 50,000 bp has an expectation of 12.5 fragments after six-cutter digestion (50,000 / 4,000 = 12.5), but after *Hin*d III digestion, only 8 fragments are observed (in fact, the 0.1 kb fragment is rarely seen on 0.8% gels and students must be told about the limits of technical resolution). The resultant χ^2 of 2.53 (df = 1) is not significant, which suggests the assumption of one recognition sequence per 4000 bp is valid.

Logistics

In a lab that meets once per week, enzymatic treatments of DNAs (RNase, DNase, restriction endonucleases) can be set up the first week followed by a week in which gels are run. Five lanes are required to analyze each genome (untreated, RNase treated, DNase treated, six-cutter digested, four-cutter digested). Therefore two genomes can be accommodated on a typical mini-gel with 10 lanes (Table 1). Because the lambda genome acts as a size standard and produces an easily analyzed pattern, each gel (group of students) should include the set of treatments for the lambda genome and a set of treatments for any one of the other genomes. There should be ample time for discussion which can begin while gels are running. Groups analyzing different genomes will need to share results with each other and all should be included in a discussion of the reasoning leading to interpretations. Moreover, technical limitations that prevent exact size determinations for the larger genomes (nuclear, bacterial) should be discussed. These limitations include an inability to count individual fragments because (1) too many bands are crowded together on the gel (2) non-identical fragments of similar size migrate to similar positions and (3) size estimates on 0.8% gels are inaccurate above 15 kb and below 0.2 kb. Consequently, large genomes cannot be accurately sized using the methods employed but relative sizes can be inferred from examination of the "smears" of bands.

Table 1. Sample lane assignments on a gel for analysis of two (bacterial, phage λ) genomes.

Lane	*Genome*	*Treatment*
1	bacterial	untreated
2	bacterial	RNase
3	bacterial	DNase
4	bacterial	six-cutter + RNase
5	bacterial	four-cutter + RNase
6	phage λ	untreated
7	phage λ	RNase
8	phage λ	DNase
9	phage λ	six-cutter + RNase
10	phage λ	four-cutter + RNase

In classrooms already conducting introductory DNA labs, the exercises outlined will not significantly increase the time or expense. However, the relatively easy manipulations of treating with RNase and DNase facilitate the use of reasoning skills in molecular genetics. In contrast, the genome size comparison involves qualitative comparative thinking, while testing observed sizes against expectations in restriction digests incorporates quantitative hypothesis testing.

Acknowledgment

Development of these exercises was facilitated by a National Science Foundation-ILI grant.

References

Murray, M.G. & Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research, 8,* 4321–4325.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning, A Laboratory Manual,* 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Winchester, A.M. & Wejksnora, P.J. (1996). *Laboratory Manual of Genetics,* 4th ed. Dubuque, IA: Wm. C. Brown Publishers.