A rapid method for determining clone frequency in complex populations using PCR and the Poisson distribution

Danny C. Alexander
Calgene, Inc., 1920 Fifth St., Davis, CA 95616, USA

Submitted November 5, 1990

There are instances in which it is desirable to know the frequency of a particular cloned nucleic acid species in a complex population. For instance, before undertaking isolation of a clone from a library, it is helpful to know how many clones should be screened; also, the frequency of a clone in a cDNA population may be taken as a first approximation of mRNA frequency (keeping in mind possible cloning bias). In our case we are interested in determining the frequency of specific clones before and after cDNA enrichment experiments. The typical way to score clone frequency is by filter hybridization using nucleic acid probes. This is laborious and time consuming, and is often complicated by artifactual spots on the filters.

Bloem and Yu (1) described a way to use the polymerase chain reaction (PCR) to decrease the complexity of a cloned population prior to screening by filter hybridization. After eluting phage heads from a filter replica of plated phage, PCR was performed to determine if the clone of interest was in the population on that plate; further subdividing a positive plate and repeating the analysis narrowed the search for the clone by several orders of magnitude. This suggested to us a simple way to score clone frequency using PCR. By diluting an E. coli library to a complexity such that the presence of a clone of interest falls within the range of the Poisson distribution (2), a simple plus/minus test can be applied to replicate growths subjected to PCR with appropriate primers. The null term of the Poisson distribution and the library titer are used to calculate the frequency of the clone of interest. The method will be described for libraries in E. coli cells, but there should be several ways to apply it to phage libraries, most simply by filter elutions as described by Bloem and Yu, or by phage growths in liquid culture.

For plasmid or cosmid libraries in E. coli we make 10-fold serial dilutions of the bacterial suspension in Luria broth containing the appropriate antibiotic. Using a 12 place pipetor the cells of each dilution are aliquoted (200 µl each) to the 1 ml wells of a deep microtitre plate (Beckman). Either 12 or 24 wells are inoculated for each dilution (of course the accuracy of the result is increased with larger numbers of replicates). A flat cover is attached (not air-tight) and the plate is shaken overnight in a tilted position in a 37°C shaker. The range of the dilutions are generally such that the most concentrated wells are inoculated with approximately 10⁶ cfu/well, and the most dilute wells with approximately 10³ cfu/well. Aliquots are simultaneously plated on LB plates with antibiotic in order to determine the exact titer.

PCR is performed (94°C 1 min, 50°C 2 min, 72°C 3 min, 25 cycles) in 50 µl assays using as template 5 µl of preboiled culture, then products are analysed by agarose gel electrophoresis. It is most efficient to first perform PCR with a single aliquot of each dilution; the most dilute inoculum which scores positive, and the least dilute inoculum which scores negative in this single well test are chosen for further analysis. All the replicate aliquots of these two dilutions are then used for PCR and the number of nulls scored. The Poisson distribution predicts nulls as

$$P_0 = e^{-x}$$

where Po is the fraction of nulls and x represents the frequency of the clone in question. This rearranges to

$$x = \ln (1/P_0)$$

Dividing x by the inoculation titer for that dilution gives the frequency in the original population. For instance, if there are four nulls out of twelve wells tested, and the wells were inoculated at 10⁶ cfu/well, then the frequency of the clone is 1.1 divided by 10⁶, or 1.1 x 10⁻⁶; this is a frequency of about one in 9,000. To test this approach a tobacco leaf cDNA library was spiked with a rabbit β-globin cDNA clone into which a lac operator sequence had been inserted at the internal EcoRI site. This clone is easily detectable as a blue colony when harbored in a lac+ host (e.g. MM294) in the presence of X-gal. Determination of the 'blue-globin' clone frequency by direct plating gave an estimate of 0.40% (26/6,500), compared to an estimated frequency of 0.39% by PCR quantification (8 nulls/12 wells at an inoculum of 1.05 x 10² cfu/well).

Reconstruction experiments have shown that without optimization most primer sets can detect the cloned DNA in 20 or fewer E. coli cells bearing pUC-type plasmid vectors. Overnight growths in LB usually reach titers of approximately 10⁹/ml; this is 5 x 10⁶ cells in 5 µl, so clones at a frequency of one in 250,000 are easily detectable (assuming no growth bias). At this sensitivity a unique clone in a library of 5 x 10⁶ independent clones should be detected once in a sample of 24 replicate wells, each inoculated with 2.5 x 10⁴ cells. This is much more sensitive than filter hybridization, which at this level becomes quite impractical because of the high noise level (false positive spots).

As stated above, the detection limits described are using standard conditions with no optimization of primer sequence or reaction conditions; optimized PCR amplifications have been described in which single molecules are detected (3). Thus it may be possible under optimal conditions to detect the DNA from a single E. coli cell, or even a single phage or naked DNA...
molecule. This would obviate the need for any type of intermediate amplification (in this case achieved by overnight growth).

Primer choice should be considered in the light of the type of information desired. For instance, the detected frequency of a particular species in a cDNA library will vary depending on the position of the 5'-most primer; a primer set with both primers near the 3' end will score most clones of the species, whereas a set including a primer at the 5' end would score only full-length clones. Primers might also be designed to score only particular members of a multi-gene family. It should also be possible to test multiple genes in the same PCR reactions, providing the products are easily distinguished by gel electrophoresis.

Application of this method is greatly facilitated by the use of 96 well microtiter plates. All steps from the growths through the PCR and gel electrophoresis can be performed in the microtiter format using multi-channel pipettes. Such improvements in efficiency allow the frequency determination of many genes in a short time without the need for radioactive probes or laborious filter hybridizations.

ACKNOWLEDGEMENT

Thanks to Richard Michlemore for reviewing the manuscript.

REFERENCES