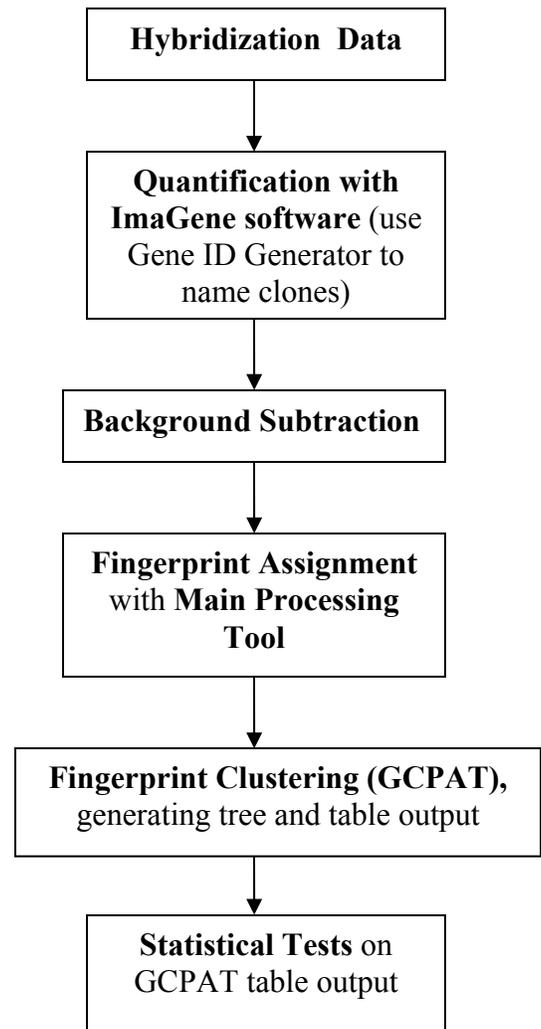


Overview of OFRG Data Analysis

- 1. Hybridization Data:** This will consist of image files (.TIF files) generated by the program that runs your phosphoimager/ autoradiograph scanner. There should be one image file per probe. Since the robotic gridding runs (each of which generates twelve membranes) take a long time, we usually take one membrane from each and hybridize these all to the reference probe. Therefore there may be multiple reference probe images.
- 2. Quantification with ImaGene software:** ImaGene is a program designed to quantify the intensity of arrayed spots in images. And is available from BioDiscovery, Inc. (100 North Sepulveda Blvd., Suite 1230, El Segundo, CA 90245 www.biodiscovery.com). For our 9,600 clone arrays we have found it necessary to have a Gene ID file to use with ImaGene for creating image templates. We generate Gene ID files using the **Gene ID Generator**, an Excel macro.
- 3. Background Subtraction:** For this step to proceed you will need your quantified data for each probe in a tabulated form identical to the output generated by ImaGene v. 5.6, with one data file per folder. The name of each data file and the name of its folder must be identical. When using multiple reference probe images (one image per set of 12 membranes), we organize the data into subsets so that all the probes on membranes from one run plus the reference probe for that run are background subtracted and assigned fingerprints as a separate unit. The data from each subset is later combined for fingerprint clustering. **Background Subtraction** is an Excel macro.
- 4. Fingerprint Assignment:** The “Flag” column in the **Background Subtraction** data sheets indicates which data points are to be analyzed and which will be ignored due to missing data (ie, no binding with reference probe). When more than one reference probe image is generated, the Flag column for those images will not always be identical. To generate a Flag column representative of all the reference probe images combined, the **Flag Column Tool**, an Excel macro, is used. This representative Flag column should be used when analyzing each subset of **Background Subtraction** data.



The **Main Processing Tool** is used to assign fingerprints. This program (an Excel macro) will: (1) generate averaged (where there is more than one hybridization per experimental probe) ratio values of hybridized probes to the reference probe, (2) separate control clone image data into clones which theoretically should bind and those which should not for each probe, and generate statistical parameters for each group for each probe, then (3) assign fingerprint values of 0, 1 or N for each clone and probe value based on the statistical parameters generated in step 2.

To generate theoretical fingerprints for the control clones, the **Theoretical Fingerprint Generator** (an Excel macro) should be used. You will need two FASTA data files saved in a text format: one containing the complete clone insert gene sequences and one containing the probe set to be used. The order of probes in the FASTA file must match the order of the probes in the data entered into the **Main Processing Tool**.

When analyzing subsets of data with this program, the data can be combined in a separate Excel file containing clone names and assigned fingerprint values for all probes. At this point you can sort the probes into a desired order. Be sure to save this file in a tab-delimited text file format.

5. **Fingerprint Clustering:** This is accomplished with the Java-based GCPAT (Greedy Clique Partitioning pAckage Tool) program. There are two forms of output: a fingerprint UPGMA tree, where clones with similar fingerprints will cluster together, and a taxonomic table, where detailed data describing the clones belonging to each fingerprint group is generated. To print trees out, it is useful to have Adobe Acrobat (Adobe Systems, Inc.) installed on your machine.
6. **Statistical Tests:** We use SAS v.8.02 (SAS Institute, Inc. Cary NC 27513, <http://www.sas.com/>) programs for these analyses unless otherwise noted. Analyses include:
 - (1) One way ANOVAs on the total number of clones found per treatment, or on fingerprint richness, diversity, or evenness in each treatment.
 - (2) Stepwise discriminant analysis, to identify fingerprint clusters in which the distribution of clones per treatment varies (you need to have
 - (3) Stepwise regression or correlation analyses, to identify fingerprint clusters in which the distribution of clones per treatment varies with some measurement that varies between treatments (for example, the ability of soils in different treatments to suppress a plant parasitic nematode).