

Test of an In Vivo Method to Detect Chloroplast Division in Crop Plants

Part III: Statistical Proofs of Observation and General Utility of the Method

Ping Zheng, Karim Ammar, Anne-Marie Michelle Girard, Steve Rodermel, David R. Thomas, Li Ning, James B. Callis, Gerry E. Edwards, and Larry Daley

This article is the third installment in a series about a novel spectrofluorometric method that allows for in vivo observation of the division of chloroplast populations in leaves of *Arabidopsis thaliana*.

This article provides a statistical approach allowing quantification of probability in two-dimensional in vivo fluorescence spectroscopy of biological samples.

When using spectroscopy for chemicals, the number of molecules, atoms, or other moieties — even when observed in complex matrices — is commonly very large, and their state rela-

tively homogeneous. During biological in vivo experiments, the moiety units explored are usually far larger in size, less concentrated, more variable, and found in far fewer numbers than common in chemicals or even biochemical work. Thus, in vivo fluorescent spectroscopy of the relatively small numbers, variability, and complexity of these biological moieties offers exceptional statistical challenges.

Chloroplast division is critical in plant development. A novel spectrofluorometric method that allows for in vivo observation of division of chloroplast populations in leaves of *Arabidopsis thaliana* mutants was described in part one of this series (1). A second paper followed that described preliminary confocal microscopic work and confirming investigations using amaryllis, wheat, sugarcane, narcissus, and other plants (2).

The technology of this series of articles holds promise in applied, as well as theoretical, areas of plant science. This article describes the novel statistical methods used for the initial discovery. We expect this statistical approach to have more general applications.

Chloroplasts, the green organelles found in plant cells, because of their photosynthetic functions are central to life on Earth and the leaves of higher plants are packed with them. However, the seeds from which these plants grow are small compared with the mature plant, and seeds commonly have few or no functional chloroplasts. Chloroplasts have a separate genome, which codes much of the chloroplast's function. Thus chloroplasts in plants must arise through repeated chloroplast division.

Chloroplast division occurs primarily in developing leaf mesophyll cells, where chloro-

Ping Zheng is a graduate student in the Department of Horticulture at Oregon State University (OSU, Corvallis, OR). She may be contacted by e-mail at zhengp@bcc.orst.edu.

Karim Ammar is a breeder in the Hybrid Wheat Program at International Maize Wheat Improvement Center, CIMMYT (Texcoco, Mexico); e-mail: k.ammar@cgiar.org.

Anne-Marie Michelle Girard is a faculty research assistant at the Center for Gene Research and Biotechnology at OSU; e-mail: girarda@bcc.orst.edu.

Steven Rodermel is a professor in the Botany Department at Iowa State University (Ames); e-mail: rodermel@iastate.edu.

David R. Thomas is professor emeritus in the Department of Statistics at OSU; e-mail: thomas@stat.orst.edu or thomasd541@msn.com.

Li Ning is a design engineer for Broadcom Corporation (Fremont, CA); (408)922-7523, e-mail: nli@broadcom.com.

James B. Callis is a professor in the Department of Chemistry, University of Washington (Seattle); (206) 543-1208, e-mail: callis@cpac.washington.edu.

Gerald E. Edwards is a professor in the Botany Department, Washington State University (Pullman); e-mail: edwardsg@wsunix.wsu.edu.

Larry Daley is a professor in the Department of Horticulture at OSU; e-mail: daley@bcc.orst.edu.

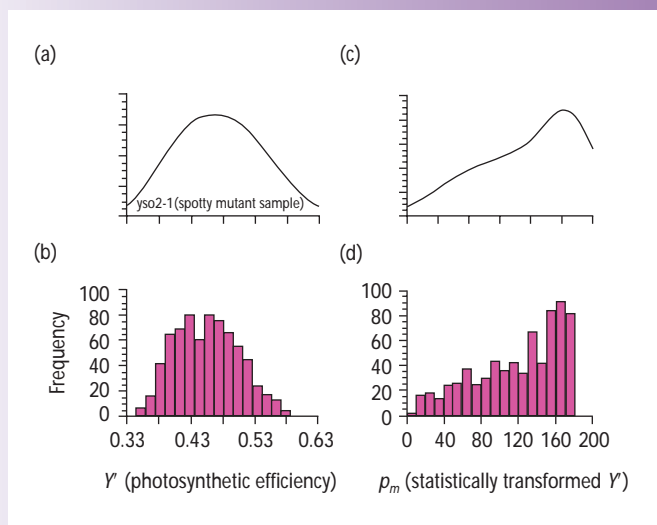


Figure 1. Ranked and statistically processed data showing distinct populations of chloroplast efficiencies in *Arabidopsis* spotty mutants. Left plots (a and b) show data before statistical transformation. Right plots (c and d) show statistically transformed data. Upper plots (a and c) show curve fitting to histograms in lower (b and d) plots.

plast division is essential to mesophyll development. The first article in this series lists the environmental factors that influence chloroplast division as cell size, nuclear ploidy, and the environment of the leaf, especially light quality (1).

Chloroplast accumulation in developing leaves enormously affects photosynthetic production, yet control of the chloroplast division process itself is one of the least understood areas of chloroplast biology (3). This situation calls for improved technology, such as that which we have contributed (1, 2).

When we were trying to distinguish between the observations, we made comparing *Arabidopsis* mutants and *wild type* (natural or unmutated) plants, we found that in both there were multiple populations of chloroplast efficiencies. Puzzled, we were forced to develop novel statistical tools to distinguish between really divided and spurious divided populations.

Experimental

Fluorometric analysis. Fluorometric analysis followed the procedure described by Ning et al. (4). Our laboratory developed an imaging fluorometer/spectrophotometer (4–10) that can quantify the fluorescence characteristics of leaves in two-dimensional space. This instrument detects changes related to the photosynthetic function and pathology of plants. Our imaging fluorometer uses a charge-coupled device (CCD) as the detector element, acquiring spectra for 31,680 positions per sample (12-bit, Lynxx, PC CCD Imaging System, SpectraSource Instruments, Westlake Village, CA).

Plant material. The material was grown in the greenhouses adjacent to the laboratory, sampled with permission from some of the material grown by other researchers at Oregon State University (Corvallis), or shipped by express mail from other researchers across the country.

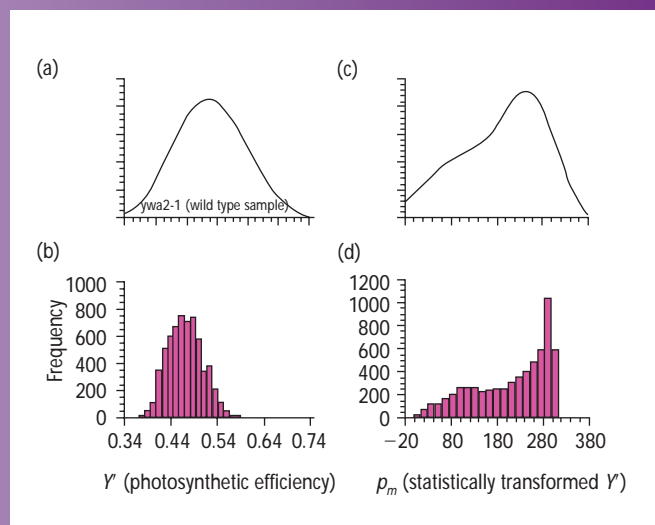


Figure 2. Ranked and statistically processed data showing distinct populations of chloroplast efficiencies in wild-type (nonmutant) *Arabidopsis*. Left plots (a and b) show data before statistical transformation. Right plots (c and d) show statistically transformed data. Upper plots (a and c) show curve fitting to histograms in lower (b and d) plots.

The *Arabidopsis thaliana* plants used for these experiments were prepared and grown (11) by Steve Rodermel at Iowa State University of Science and Technology (Ames). These materials consisted of grown *Arabidopsis* plants. The *A. thaliana* germplasm used was: V28, Spotty, and wild type. Plants were shipped, carefully protected, and sent by express mail service from Iowa to the Oregon laboratory site. After the plants arrived at the Corvallis, Oregon, site, they were carefully unpacked and allowed to grow overnight to permit recovery of photosynthetic function from shipping stresses.

Leaves were sampled and then analyzed by the imaging fluorometer (1, 2). We noted the growth details with care; these *Arabidopsis thaliana* leaves were taken at 0, 4, and 8 days after arrival from Iowa. Mutants and their growing conditions have been described previously (1).

When we grew our own plants we used OSU soil mix, Osmocote (Sierra Chemical, Milpitas, CA) slow-release fertilizer in amounts and release times appropriate for each species, and the greenhouse crew watered each plant twice a day and controlled any pathogens and pests that could cause problems.

Data processing. Program for data analysis (curve deconvolution). One of the most computation-intensive problems is calculating maximum-likelihood estimates for a mixture of normal distributions. This process can be thought of as parametric estimations of likelihood functions. Maximizing these likelihood functions is complicated by singularities and numerous spurious maximizers. Here we used the Expectation Maximization (EM) algorithm, which is a technique for finding maximizers of likelihood functions (12). The program running the EM was written using the MATLAB (The Math Works, Natick, MA) numerical software package.

This iterative EM algorithm is extremely reliable and usually finds the most appropriate maximizer from most rea-

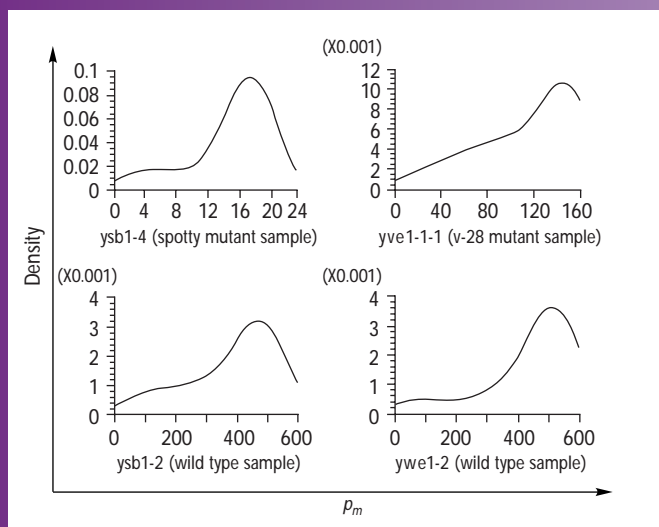


Figure 3. Statistically processed curves showing dual populations of chloroplast efficiencies within mutant and wild type leaves.

sonable initial guesses. However, it is very slow in cases where there is considerable overlap between component normal distributions.

We set up a model where the resultant distribution is considered as a mixture of normal distributions. To do this we wrote a novel program in MATLAB using the established formulas and mathematical conventions. In these formulas we consider the unknown distribution as a mixture of a number of subdistributions from 1 . . . m , where m is the number of normal components in the mixture.

A t - variable normal density $p(x|\mu, \Sigma)$ with t - variable mean vector μ and $t \times t$ symmetric, positive definite covariance matrix Σ , is defined for t - variable real x by

$$p(x | \mu, \Sigma) =$$

$$(\exp(-(x - \mu)^T \Sigma^{-1} (x - \mu) / 2)) / ((2\pi)^{t/2} |\Sigma|^{1/2})$$

Here the mixture has a t - variable normal distribution $p_m(x|\Theta)$, and is defined by the following equation:

$$p_m(x | \Theta) = \alpha_1 p(x | \mu_1, \Sigma_1) + \dots + \alpha_m p(x | \mu_m, \Sigma_m) \quad (1)$$

In the equation form $\alpha_1 p(x | \mu_1, \Sigma_1)$, the vertical line means "given." The notation $\{\alpha_1 p(x | \mu_1, \Sigma_1)\}$ means: given the values for μ_1 and Σ_1 , the equation defines the probability that the x occurs. Here x is observed random samples; Θ collectively refers to all the individual component mean and covariance parameters, along with the mixing proportions $\alpha_1, \alpha_2, \dots, \alpha_m$, which must sum unity (that is 1) and have values between 0 and 1.

The theory of maximum-likelihood states that good estimates of the unknown parameter Θ_0 can be obtained from the log-likelihood function $L(\Theta)$. The function $L(\Theta)$ is defined for a given t - variable sample $\{x_1, \dots, x_n\}$ and density of the form in (1), by

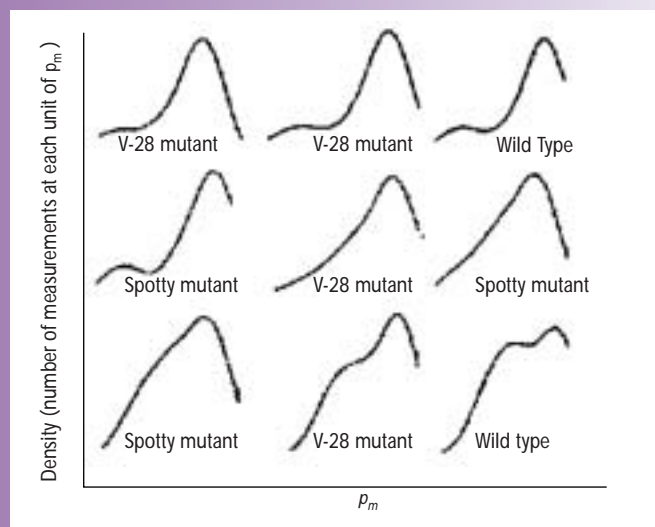


Figure 4. A larger set of data also showing statistical analysis from second variegated mutant (V-28) (adapted from reference 3).

$$L(\Theta) = \log(p_m(x_1 | \Theta)) + \log(p_m(x_2 | \Theta)) + \dots + \log(p_m(x_n | \Theta))$$

The chosen maximizer is denoted here by Θ^* . The approach considered here results from applying different optimization techniques to find Θ^* by maximizing $L(\Theta)$.

Optimization algorithms are iterative in nature. These algorithms will, in theory, generate an infinite sequence $\{\Theta_{(r)}\}$ of approximations to Θ^* . In this sequence each successive $\Theta_{(r)}$ tends to more closely approximate Θ^* .

The procedure starts with supplying an initial, and usually rough, estimate $\Theta_{(1)}$ of Θ^* . The main check used in these tests compares the new iterate $\Theta_{(r+1)}$ with the old iterate $\Theta_{(r)}$ at each step. When there is very little difference between successive iterates the process is near completion. This is because this circumstance usually means that the accuracy of $\Theta_{(r+1)}$ as an approximation to Θ^* will not be much improved by further iteration.

For the above reason the implemented iteration is terminated as soon as two successive iterates are very similar. For $i = 1, \dots, m$, where m is the number of normal components in the mixture, the following calculations are done for $\alpha_{i(r+1)}$, $\mu_{i(r+1)}$, and $\Sigma_{i(r+1)}$:

$$p(x_k | \mu_{i(r)}, \Sigma_{i(r)})$$

Where r is the iteration, $i = 1, \dots, m$, $k = 1, \dots, n$

$$p_m(x_k | \Theta_{(r)}) = \sum \alpha_{i(r)} p(x_k | \mu_{i(r)}, \Sigma_{i(r)})$$

where $i = 1$

$$W_{ik(r)} = \alpha_{i(r)} p(x_k | \mu_{i(r)}, \Sigma_{i(r)}) / p_m(x_k | \Theta_{(r)})$$

$$A_{i(r)} = \left(\sum_{k=1}^n W_{ik(r)} \right)$$

$$i_{(r+i)} = A_{i(r)} / n$$

$$\mu_{i(r+1)} = \left(\sum_{k=1}^n x_k W_{ik(r)} \right) / A_{i(r)}$$

$$\sum i_{(r+1)} = \left(\sum_{k=1}^n (x_k - \mu_{(r+1)})^T W_{ik(r)} \right) / A_{i(r)}$$

The property of EM is such that $L(\Theta_{(r+1)}) \geq L(\Theta_{(r)})$ for each iteration, so that progress towards finding a maximizer is always being made.

Results and Discussion

For this particular series of experiments seeking differences in photosynthetic efficiency between the mutant and the normal parts of these *Arabidopsis*, we reprocessed the photosynthetic efficiency data (1, 2). We ranked the data, with respect to efficiency on the x -axis with the y -axis as the number of pixels at that efficiency range.

This generated normal curves of the efficiency of the chloroplasts within each image. Our original intent was, in this way, to measure the proportion of mutated tissue and the amount of normal tissue.

In these mutants we often, but not always, saw two populations with different photosynthetic efficiencies. That made us very happy because it seemed to explain the differences between dark green and much lighter white or yellow sectors of the mutation plant's leaves (for example, unprocessed spotty mutant data presented in Figure 1). Preliminary examination of data made us realize that these divided efficiency populations of chloroplast suggested that this did not exist in wild type plants (unprocessed data in Figure 2).

To obtain statistical proof of this difference, we tested the data statistically by a de novo rigorous statistical analysis for subpopulations (see Experimental). This statistical approach surprisingly showed us that these subpopulations existed in both wild type and mutant plants (processed data in Figures 3 and 4). This is attributed to the energy costs of chloroplast division (1, 3, 13).

Thus we embarked on a wider examination of mutant and wild type spectra using the same statistical deconvolution (Figure 4). With this data we began to clearly see two populations of chloroplasts with different efficiencies not only in other mutants but also in wild type *Arabidopsis* plants (Figure 4). This binary distribution did not occur in all samples and thus clearly was not some general property of statistical approach (for example, see Figure 4). All of this was surprising, quite disturbing, very puzzling, and not especially tidy, for it blew our initial hypothesis clear out of the water.

Thus we came to the surprising conclusion that this phenomenon, of which we were not sure of at the time but now know was related to chloroplast division, occurred in both mutant and wild type plants.

This showed that we had unwittingly developed a method to statistically show chloroplast division in plants (1, 3).

Conclusions

Statistically we demonstrated that distinct populations of chloroplast efficiency are real phenomena, occurring both in wild type and in variegated mutant *Arabidopsis*. That these populations resulted from the measurement of chloroplast division also became clear. Thus, a statistical method that we thought only applied to a particular use has far more general applications in deconvoluting biological data.

Once this method was applied and the phenomenon was found to be real, it was possible to consider that the large number of pixels sampled was sufficient to fulfill statistical requirements without further statistical analysis.

General use of this method allows statistical verification of bimodal and polymodal distribution phenomena from data obtained by CCD cameras and other devices.

Acknowledgment

The authors thank Don Powell and his family, owners of Garland Nursery (Corvallis, OR), for their kind donation of numerous *Amaryllis* bulbs.

References

1. P. Zheng, C. Wetzel, K. Ammar, A.-M.M. Girard, S. Rodermel, D.R. Thomas, L. Ning, J.B. Callis, G.E. Edwards, and L. Daley, *Spectroscopy* 17(4), 16–25 (2002).
2. E.J. Robertson, N.R. Baker, and R.M. Leech, *Plant Cell Environ.* 16, 809–818 (1993).
3. P. Zheng, C. Wetzel, K. Ammar, A.-M.M. Girard, S. Rodermel, D.R. Thomas, L. Ning, J.B. Callis, G.E. Edwards, and L. Daley, *Spectroscopy* 17(9), 14–18 (2002).
4. L. Ning, G.E. Edwards, G.A. Strobel, L.S. Daley, and J.B. Callis, *Appl. Spectrosc.* 49(10), 1381–1389 (1995).
5. L. Ning, G.E. Edwards, and L. Daley, Oregon State University Patent, U.S. Patent No. 5,854,063, for Method and Apparatus for Spectrophotometric Observation of Plants; issued Dec. 29, 1998.
6. L. Ning, G.E. Edwards, and L. Daley, Oregon State University Patent, U.S. Patent No. 5,981,958, for Method and Apparatus for Detecting Pathological and Physiological Change in Plants; issued Nov. 9, 1999.
7. L. Ning, R. Ozanich, L.S. Daley, and J.B. Callis, *Spectroscopy* 9(7), 41–48 (1994).
8. L. Ning, W.J. Bowyer, A.M. Chozinski, T. Buban, A. Azarenko, G.E. Edwards, L.S. Daley, G.A. Strobel, and J.B. Callis, *Spectroscopy* 11(9), 44–49 (1996).
9. L. Ning, W.J. Bowyer, L.S. Daley, E.H. Piepmeier, G.A. Strobel, and J.B. Callis, *Spectroscopy* 11(3), 34–44 (1996).
10. L. Ning, W.J. Bowyer, L.S. Daley, E.H. Piepmeier, G.A. Strobel, and J.B. Callis, *Spectroscopy* 11(4), 68–74 (1996).
11. *Arabidopsis: The Complete Guide*, D. Flanders and C. Dean, eds., Norwich, U.K., Cambridge Laboratory Agriculture and Food Research Plant Molecular Biology *Arabidopsis* Programme (1991).
12. J.W. Davenport, M.A. Pierce, and R.J. Hathaway, A Numerical Comparison of EM and Quasi-Newton Type Algorithms for Computing MLE's for a Mixture of Normal Distributions, in *PCmpScSt20 (Computer Science and Statistics Proceedings of the 20th Symposium on the Interface)*, E.J. Wegman, Ed., Alexandria, Virginia, American Statistical Association (1988).
13. F. Doran and R.A. Cattolico, *Plant Physiol.* 115, 773–781 (1997). ■