

# Molecular Cytometry for Characterization of Plant Biodiversity

David W. Galbraith, University of Arizona, Department of Plant Sciences/BIO5 Institute, Tucson AZ 85721

## Abstract

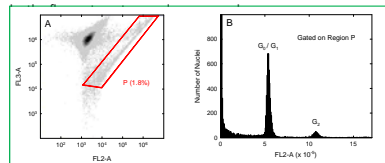
2010 has been declared the UN International Year of Biodiversity. It is clear that the survival of the human race depends entirely on biological diversity and our understanding of how to manipulate this diversity. Preparing a complete inventory of existing eukaryotic species takes on particular urgency, since we have taxonomic names for only 14% of an estimated 12 million species, since we know little about those identified, and since, yearly, thousands of species become extinct. Molecular cytometry has a key role to play in providing an inventory of existing and undiscovered species, and in providing a detailed molecular understanding of them. Two properties of organisms can be accessed using cytometric techniques: flow cytometry to determine genome sizes, and NextGeneration and Gen3 DNA sequencing to define the genome information content. Further development of these platforms envisages unprecedented accumulation of information at very reasonable costs. We are exploring application of these cytometric techniques to provide an inventory of the estimated 500,000 species of flowering plants. Flow cytometry provides a rapid means to characterize the genome sizes of plants over the entire range of C-values for which information exists. The methods are compatible with plants of disparate form, phenotype, growth habit, and storage location. NextGeneration short-read sequencing provides a cost-effective wealth of information that can be directly employed for taxonomic and phylogenetic purposes, and will identify valuable variation for agricultural improvement. Reductions in platform costs and increases in information output should allow us to provide a complete inventory of the flowering plants within a three year period. This will require unprecedented international cooperation, as well as local taxonomic expertise. Given the urgent need to achieve this inventory, allocation of resources toward this goal should be a top priority for funding agencies.

## Flow Cytometry and C-value Measurement

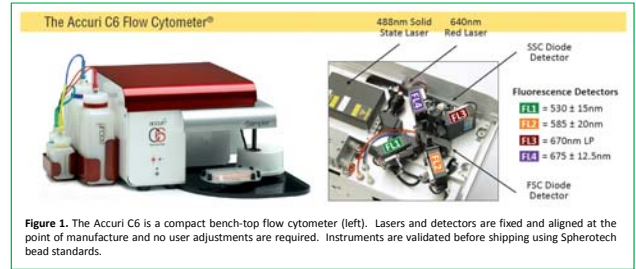
Flow cytometry involves the analysis of individual cells or organelles as they pass through the focus of an intense light source, typically a laser. Their fluorescence and light scatter properties are then quantified and the data stored for display and analysis. Staining with fluorochromes of defined specificity allows measurement of specific cellular properties. One example is the use of DNA-specific fluorochromes (DAPI, Hoechst, propidium iodide/RNase) for the measurement of nuclear DNA contents, otherwise termed the C-value. Somatic cells from diploid organisms have a 2C nuclear DNA content, and the range in the angiosperms is impressively large.

Flow cytometric C-value measurements in higher plants are performed using homogenates prepared by chopping (Galbraith *et al.*, Science 220:1049-1051 (1983)). After staining with DAPI or PI/RNase, the homogenate is filtered and analyzed with laser illumination at 350/405nm (DAPI) or (530/585nm (PI)). Recent developments in flow cytometric hardware have resulted in desktop instruments such as the Accuri C6 (Figure 1), which are available at a very reasonable price. The Accuri takes additional advantage of a 24-bit Analog-to-Digital Converter, which divides the fluorescent signal histograms into  $1.9 \times 10^7$  bins. This provides the machine with an exceptional dynamic range.

Using the Accuri C6, flow cytometric analysis of diploid tissues, such as leaves of *Pisum sativum*, identifies populations of nuclei having well-defined G0/G1, and G2 peaks (Figure 2). Analysis of homogenates is complicated by the presence of a large amount of subcellular debris derived from organelles. This debris scatters light and, since in most cases it is also autofluorescent, is therefore readily detected by the flow cytometer as a large amorphous region. The PI-stained nuclei, in contrast, occupy a discrete region (red box) representing the correlated fluorescence of PI distributed between the two fluorescence detectors which detect roughly the lower and upper halves of the PI emission spectrum. Gating on this region (a small minority of the total particles detected) provides a uniparametric histogram (Figure 2B) from which the nuclear C-value can be calculated. This is done using standard plants of known 2C DNA contents (Table 1), which can be co-chopped with the plants of unknown value or run consecutively. Sample processing is rapid, particularly if an autosampler is employed, and as many as twelve species per hour can be routinely analyzed.



**Figure 2.** Flow cytometric analysis of homogenates prepared from *Pisum sativum* (pea) seedling tissue. (A) Biparametric contour plot of FL2-A (585/40nm) versus FL3-A (>670nm) fluorescence emission. (B) Uniparametric histogram of F2-A fluorescence, gated on region P1.



**Figure 1.** The Accuri C6 is a compact bench-top flow cytometer (left). Lasers and detectors are fixed and aligned at the point of manufacture and no user adjustments are required. Instruments are validated before shipping using Spherotech bead standards.

**Table 1.** Names of plant species, source of seeds, 2C DNA contents, and attribution.

Species	Common name	2C DNA content (pg)	DNA content attribution	Source of plants and hyperlinks
<i>Arabidopsis thaliana</i> ecotype Columbia	Thale cress	0.32	Kew C-value Database (Bennett and Leitch, 2004)	Lehle Seeds*
<i>Raphanus sativus</i> cv. Saxa	Radish	1.11	Olomouc website (Dolezel <i>et al.</i> , 2007)	J. Dolezel <sup>†</sup>
<i>Medicago sativa</i> L. cv. Cimarron	Alfalfa	3.50	Kew C-value Database (Bennett and Leitch, 2004)	Lehle Seeds*
<i>Pisum sativum</i> L. cv. Citrad	Garden pea	9.09	Olomouc website (Dolezel <i>et al.</i> , 2007)	J. Dolezel <sup>†</sup>
<i>Secale cereale</i> L. cv. Dankovske	Rye	16.19	Olomouc website (Dolezel <i>et al.</i> , 2007)	J. Dolezel <sup>†</sup>
<i>Triticum aestivum</i> L. line 812	Wheat	34.65	Kew C-value Database (Bennett and Leitch, 2004)	Lehle Seeds*
<i>Alstroemeria aurea</i> Grah.	Lily of the Incas	80.90	Kew C-value Database (Bennett and Leitch, 2004)	Inter-American Products*

\* <http://www.arabidopsis.com/main/cat/seeds/teaching/Crop/ipsa.html>  
<sup>†</sup> <http://mccz.szb.czu.cz/research/protocols.php?protocol=41n4>  
<http://www.interamericanproducts.com>

## Nuclear DNA Content Measurement across all Angiosperms

### Dynamic Range:

Approximately 1-2% of angiosperm species have entries in the RBG Kew C-value database (<http://data.kew.org/cvalues/>). These values range from 0.20 pg (*Fragaria viridis* Duch.) to 254.80 pg (*Fritillaria assyriaca* Baker). The dynamic range of the ADC in the Accuri C6 is greater than this biological range of values, implying the machine should be able to accommodate measurements from any and all angiosperm species. This prediction was tested (Figure 3) using species ranging in 2C DNA content from 0.32pg (*Arabidopsis thaliana*) to 80.9 pg (*Alstroemeria aurea*). Well-resolved 2C DNA peaks were identified for all species, and the predicted DNA content scaled linearly with PI fluorescence over the entire range.

### Throughput and Cost Considerations:

Taken together, setting a goal of determining the C-values for the remaining 98% of the angiosperms appears reasonable, at least in terms of the Accuri C6 measurement platform. Running the homogenization and flow analysis pipeline at twelve samples per hour, for ~280,000 species would require about 24,000 hours. Amortized over 20 instruments (a capital investment of ca. \$700,000), this reduces to 1,200 hours per instrument, which is clearly feasible.

Caveats would include problems of sample collection and identification, of the uncovering of new species and genera, and of samples recalcitrant to the chopping method. The small platform size of the Accuri C6, coupled to its minimal electrical and liquid input requirements, would allow this instrument to be placed in laboratories close to the sample sources, thereby avoiding the impact of laws regulating biodiversity export.

## NextGen DNA Sequencing

### Role in Plant Biodiversity Measurement:

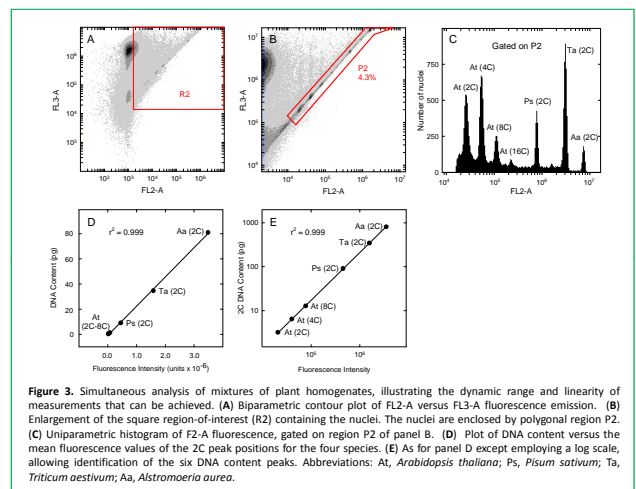
Discovering and cataloging the C-values for all angiosperms would provide a valuable dataset for identifying environmental, physiological, and ecological factors that affect this fundamental cellular parameter. It would also provide valuable information concerning one ultimate goal of Plant Biodiversity measurement, the full sequencing of the nuclear genomes of all angiosperms.

### Sequence Output, Sample Throughput, and Cost:

Two companies (Life Technologies and Solexa/Illumina) have developed NextGen sequencing technologies that produce large numbers ( $4 \times 10^9$ ) of short reads (75-100 base; i.e. 400 Gb) from a single run. The cost per lane is ~\$1,000; the cost of library preparation is ~\$300 per sample, and is particularly labor intensive. A single instrument run takes around one week, and can currently produce, per week, the equivalent of two human genomes at 30X coverage. Applied to plants, the preferential selection of species having small C-values, coupled to the use of multiplexing, are important considerations from the viewpoint of sample throughput and cost. NextGen sequence multiplexing, which has already been shown to be feasible to the level of several hundred samples per lane, inevitably reduces the genomic coverage. The extent to which reduced coverage provides a useful readout of sequence diversity will therefore need careful analysis. Genomes of large C-value species are likely to contain similar numbers of genes as compared to genomes of small C-value species. Since the C-value differences are likely to represent amplification of repetitive sequence families, redundant sequencing of these families may be something to avoid.

### Gen3 Sequencing:

Given the historically exponential rate of growth of sequencing capacity, and the emergence of Gen3 sequencing platforms (including Helicos, Pac Bio, Ion Torrent, Complete Genomics, etc.), we can readily anticipate 1000 human genomes will be sequenced by the end of 2010. "In 2011, the worldwide capacity for whole human genome sequencing will easily reach 50,000, data based on orders that have already been placed" (Richard Resnick, <http://blog.genomequest.com/2010/07/implications-of-exponential-growth-of-global-whole-genome-sequencing-capacity/>). By 2013, Resnick predicts a world-wide capacity will exist to sequence 1 million human genomes. This implies that the concept of fully sequencing all angiosperm species is not speculative, and also fits within the timeframe described above for completing C-value measurements.



**Figure 3.** Simultaneous analysis of mixtures of plant homogenates, illustrating the dynamic range and linearity of measurements that can be achieved. (A) Biparametric contour plot of FL2-A versus FL3-A fluorescence emission. (B) Enlargement of the square region-of-interest (R2) containing the nuclei. The nuclei are enclosed by polygonal region P2. (C) Uniparametric histogram of F2-A fluorescence, gated on region P2 of panel B. (D) Plot of DNA content versus the mean fluorescence values of the 2C peak positions for the four species. (E) Plot of DNA content versus the mean fluorescence values of the six DNA content peaks. Abbreviations: At, *Arabidopsis thaliana*; Ps, *Pisum sativum*; Ta, *Triticum aestivum*; As, *Alstroemeria aurea*.

## Conclusions

Technology development has conventionally been viewed as secondary to the goal of addressing questions of biological significance. The recent acceleration in technology development in molecular cytometry, and most notably in DNA-based sequencing instrumentation, has rendered this viewpoint antiquated, if not completely obsolete. Given the challenge of identifying and cataloging global biodiversity, collaborative mechanisms should be established and funding resources defined, that can fully leverage new technologies of molecular cytometry to address this challenge.

**Acknowledgments:** I thank Georgina Lambert for valuable technical assistance, and Accuri Cytometers for loan of the C6 flow cytometer.