Improvement of High-throughput Genotype Analysis After Implementation of a Dual-curve Sybr Green I-based Quantification and Normalization Procedure

Donald Livingstone III, Barbie Freeman, Cecile L. Tondo, and Kathleen A. Cariaga
USDA-ARS, Subtropical Horticulture Research Station, 13601 Old Cutler Road, Miami, FL 33158

Nora H. Oleas
Florida International University, Biology Department, Miami, FL 33199
Alan W. Meerow, Raymond J. Schnell, and David N. Kuhn
USDA-ARS, Subtropical Horticulture Research Station, 13601 Old Cutler Road, Miami, FL 33158

Additional index words. DNA quantification, Sybr Green I, microplate, automated, genotyping

Abstract. The ability to rapidly genotype a large number of individuals is the key to any successful marker-assisted plant breeding program. One of the primary bottlenecks in high-throughput screening is the preparation of DNA samples, particularly the quantification and normalization of samples for downstream processing. A rapid and simple Sybr Green I-based quantification procedure that can be performed in a 96-well format is outlined. In this procedure, a dual standard curve method is used to allow better resolution of dilute samples and to reduce fluorescence value variation between samplings. A method to quickly normalize samples, and the importance of normalization, is also explored. We demonstrate that successful fragment amplification of a Theobroma grandiflorum (Willd ex Spreng) Schum. population is increased from 70% to 98% when each DNA extract is quantified and normalized as opposed to quantifying only a subset and normalizing all the samples based on the average of that subset. Improved microsatellite amplification was also observed among individuals in the monocot genus Phaedranassa Herb. spp. Additionally, when our normalization method is applied to a Persea americana Mill. population, 97% of the samples normalized to 4 ng·μL⁻¹ amplify at least three of six microsatellite regions, whereas only 30% of the samples below 4 ng·μL⁻¹ (i.e., samples that could not be normalized) amplify at least three regions. We describe an undemanding method to quantify and normalize a large number of samples, which can be done manually or can be automated.

Modern advances in marker-assisted selection (MAS) have profoundly changed breeding methods not only in commodity crops, but also specialty and ornamental crops. Breeders are now able to obtain data for hundreds of molecular markers across thousands of individuals. This increase in throughput has created new obstacles that must be overcome to maintain a successful MAS breeding program. For example, the extraction, quantification, and normalization of high-quality genomic DNA from large numbers of individuals remains problematic and time-consuming (Lange et al., 1998). Traditional methods of quantifying DNA include measuring absorbance with a spectrophotometer or fluorometer (Labarca and Paigen, 1980; Le Pecq and Paoletti, 1966; Sambrook et al., 1989; Sela and Antignus, 1971). Previously in our laboratory, samples would be quantified using spectrophotometry. Because this process is time-consuming, only a small subset of samples would be quantified, their concentration averaged, a dilution factor calculated based on that calculated average, and all samples diluted globally. Unfortunately, spectrophotometric measurements are susceptible to contaminants such as polyphenolic compounds that are carried over from the DNA extraction process and can skew the DNA absorbance readings (Rengarajan et al., 2002; Singer et al., 1997). Although these compounds may not effect downstream applications, they do have a significant effect on concentration determination. Single-cell spectrophotometers provide consistent results from pure samples but are not amenable to high-throughput applications; additionally, concentrations of DNA extracts vary enough to hinder the calculation of an accurate estimated concentration based on a small sample subset. An alternative method to rapidly and accurately quantify dsDNA, that was not sensitive to contaminating ultraviolet-absorbing compounds, was desired. In recent years, cyanine dyes such as Sybr Green I (SG), which bind to double-stranded DNA, have successfully been used to quantify DNA samples in solution (Leggate et al., 2006; Rengarajan et al., 2002; Singer et al., 1997; Vitzthum et al., 1999). An SG-based method seemed to be capable of meeting the speed, ease, and accuracy desired (Leggate et al., 2006; Rengarajan et al., 2002; Vitzthum et al., 1999). A plate format assay was preferred over a single cuvette to allow higher throughput and to reduce variation between sample measurements. Previously published SG protocols used large volumes and measured samples individually (Rengarajan et al., 2002). To address high-throughput demands, these protocols were adapted to use small volumes and to measure multiple samples in a 96-well plate format (Leggate et al., 2006). In this report, we describe an SG-based, semi-automated, high-throughput method to quantify and normalize plant DNA extracts, and we also examine the effect of accurate normalization on downstream MAS applications.

Materials and Methods

Preparation of DNA samples. Solutions of known concentrations of salmon sperm DNA (Stratagene, La Jolla, CA) were created by weighing out dried DNA and suspending in TE (10 mM Tris-HCl pH 7.5, 1 mM EDTA) buffer. Plant DNA extracts were prepared from mature leaves harvested from 1843 Persea americana Mill., four Theobroma cacao L., and 55 Theobroma grandiflorum (Wild ex Spreng) Schum. trees and 96 Phaedranassa Herb. spp. individuals using FastDNA SPIN Kit (M P Biomedical, Solon, OH). These DNA extracts represent the Lauraceae (P. americana), Malvacaeae (T. cacao and T. grandiflorum), and Amaryllidaeae (Phaedranassa) families.

Quantification assay. Our SG assay was initially based on published procedures (Leggate et al., 2006; Rengarajan et al., 2002). Previously, SG assays consisted of 100 μL of 100x Sybr Green I (Lonza, Rockland, ME) combined with 100 μL of DNA sample in TE resulting in a 200 μL reaction of 50x SG. The reaction was then incubated at room temperature in the dark for 5 min before fluorescence was measured in a plate reader. We altered the final reaction volumes and SG concentrations to determine the effect of volume and concentration on fluorescence detection. First, final reaction volumes were changed to 200 μL, 100 μL, 50 μL, and 30 μL. Each reaction consisted of a 1-μL DNA sample, half volume 100x SG (50x SG
final volume was water. Next, SG concentrations were varied by combining the 1-μL DNA sample, 24 μL water, and 25 μL of 100×, 80×, 60×, 40×, or 20× SG (50×, 40×, 30×, 20×, or 10× SG final concentration). The different Sybr Green I solutions were made by diluting a 10,000× SG stock in TE. For all optimization experiments, DNA standards from 0 to 48 ng·μL⁻¹ were used as samples, measured in triplicate, and the average with so was plotted (Fig. 1). Our optimized assay has a final reaction volume of 50 μL and an SG concentration of 30×. The optimized assay contains 24 μL of water, 1 μL of the DNA sample, and 25 μL of 60× SG (diluted in TE buffer). When fluorescence from DNA samples is off scale, samples are diluted 1:20 in water and 1 μL of the dilution is used in the quantification protocol. All samples are prepared in a 96-well plate with the last three wells reserved for standards. A 0 ng·μL⁻¹ standard (blank) is read for the purpose of subtracting out background fluorescence. Additionally, each plate is read twice. The first uses a 48 ng·μL⁻¹ salmon sperm DNA standard to set the autogain levels, and for the second read, the gain is adjusted to a 12 ng·μL⁻¹ salmon sperm DNA standard.

Creation of standard curves. Samples of salmon sperm DNA ranging from 0 ng·μL⁻¹ to 48 ng·μL⁻¹ were read in triplicate on a Bio-tek FLx 800 microplate fluorescence reader (Bio-tek Instruments, Winooski, VT) with 485 nm/528 nm excitation/emission wavelength filters. The gain sensitivity on the plate reader was set to autoadjust to the 48 ng·μL⁻¹ sample; this is the broad range curve. A second standard curve was run with the same samples with the autoadjust set to the 12 ng·μL⁻¹ sample. This narrow range curve allows better separation of the fluorescence values of samples with low concentration. Average fluorescence values were plotted versus standard concentration and a curve fit was to the points. A curve, with an $R^2$ value of 0.9980, was fit to the values for the 48 ng·μL⁻¹ curve. A line was used as the fit for the 12 ng·μL⁻¹ standard curve, because it provided an $R^2$ value of 0.9896 and was more consistent over the range of concentrations of our samples. Plots, curve fits, and $R^2$ values were all performed with Excel (Microsoft, Redmond, WA). Concentration was plotted on the Y-axis to allow for easier determination of unknown concentrations from the fluorescence values obtained.

Calculations. A spreadsheet was created to allow quick and accurate calculation of concentration based on measured relative fluorescence unit (RFU) values. This spreadsheet was designed to work with data exported from the KC junior software package (Bio-tek Instruments) that controls the Bio-tek FLx 800 microplate reader. This spreadsheet is available on the Internet (http://www.shrstropicalcrops.org/static/dna_quant_template_bt.xlsx) and can be modified to function with RFU values produced with other instruments. Raw fluorescence values are exported from the plate reader into this spreadsheet. The user then enters the names of each sample and any dilution factors used. The spreadsheet then determines the appropriate fluorescence values (gain autoadjusted to 48 ng·μL⁻¹ or 12 ng·μL⁻¹) and corresponding standard curve to use for each sample individually. If, when measured with the autogain set to 48 ng·μL⁻¹, the unknown sample’s fluorescence value is greater than the value of the 12 ng·μL⁻¹ standard, then the broad standard curve is used. If this value is below the 12 ng·μL⁻¹ standard, then the narrow standard curve and autogain values are used. The fluorescence value of the blank, as read at the appropriate autogain level, is subtracted from each sample, and the equation defining the appropriate standard curve is used to convert raw fluorescence values into concentration. Calculated concentrations are then used to create a working solution spreadsheet, which calculates the amount of sample and water that needs to

![Fig. 1. Effect of varying total reaction volume and Sybr Green I (SG) concentration. Quantification assays using standards of known concentrations were run in triplicate, varying total reaction volume (A) or SG final concentration (B). The gain setting was autoadjusted to the largest standard. Average fluorescence value (RFU) is plotted against concentration, and so for each average is represented as error bars for differing volumes (A) and SG concentrations (B).](image-url)
be combined to normalize sample concentrations to a user-defined value. This working solution sheet can be used with a liquid handling system to automate the normalization of samples or can serve as a guide for manual dilution.

Genetic marker analysis. Genetic marker analysis was performed using primers and polymerase chain reaction (PCR) conditions as previously described for *Phaedrana-ssa* (Oleas et al., 2005) and *P. americana* (Borrone et al., 2007). Genetic marker analysis for *T. grandiflorum* and *T. cacao* was performed using primers and conditions as described for *T. cacao* (Kuhn et al., 2008). Amplification success was defined as the percentage of samples providing scorable genetic marker analysis data.

To determine the amount of DNA template necessary for PCR reactions to achieve optimal results on the ABI Prism 3730 DNA Analyzer (Applied Biosystems, Foster City, CA), a concentration gradient was examined using different concentrations of DNA from four *P. americana* and four *T. cacao* individuals. DNA was quantified with SG and diluted for DNA template concentrations ranging from 0.004 to 4 ng·µL⁻¹. PCR reactions were performed using fluorescently labeled primers SHRSPa023 (Borrone et al., 2007) for *P. americana* and EST5371 (Kuhn et al., 2008) for *T. cacao*. PCR conditions for *P. americana* were followed as described by Borrone et al. (2007), whereas those for *T. cacao* were followed as described by Kuhn et al. (2008). PCR products were resolved by electrophoresis on an ABI Prism 3730 DNA analyzer. After it was determined that each allele was within the correct size range, peak height in RFU was recorded from the DNA analyzer. After examining these results, our final optimized quantification assay consists of combining 1 µL of DNA sample, 24 µL water, and 25 µL of 60× SG (30× final concentration). These assay conditions were used in the creation of a standard curve with salmon sperm DNA. Standards ranging in concentration from 0 to 48 ng·µL⁻¹ were quantified in quadruplicate, an average was calculated, the points plotted (Fig. 2A), and a curve fit. To account for any variation in fluorescence values between reads, the plate reader was set to autoadjust the gain to the highest concentration on the curve (48 ng·µL⁻¹). This sets the fluorescence value of that sample near the maximum allowed by the plate reader. By including the 48 ng·µL⁻¹ salmon sperm DNA standard on every plate read and autoadjusting the gain to that standard, any variation in the plate reader’s measurements will be eliminated by always calibrating the 48 ng·µL⁻¹ standard to the same maximum fluorescence value. Including an autoadjusting standard on every plate read allows the standard curve to be reused by correcting for any read-to-read variation.

![Fig. 2. Standard curves created from the average of three separate readings. Known concentrations of salmon sperm DNA were measured in triplicate, and the average was plotted against the fluorescence value measured. Trend lines were fit and the resulting equation used to calculate concentration. The gain of the plate reader was set to autoadjust to the 48 ng·µL⁻¹ standard for the broad curve (A). The gain for the narrow curve was set by autoadjusting to the 12 ng·µL⁻¹ standard (B). Two curves are used to allow for better resolution of values.](image-url)
However, a side effect of adjusting the gain in this manner is that the fluorescence values of samples at the lower end of the standard curve differ by very little, thus making it difficult to distinguish 4 ng mL$^{-1}$ from 6 ng mL$^{-1}$, for example. To address this issue, a second standard curve was created in the same manner with the gain autoadjusted to the 12 ng mL$^{-1}$ salmon sperm DNA standard as the highest value (Fig. 2B). This increased the resolution of the fluorescence values at the lower end of the standard curve, allowing better differentiation of more dilute samples. Fluorescence data must be obtained for both gain settings, and therefore every sample is measured twice, autoadjusting for the appropriate standard, both of which must be included with the samples along with a blank (water) sample. Both data sets are exported to the Excel spreadsheet, which identifies the most appropriate values to use for each unknown sample measured as described in the “Materials and Methods.” The Excel spreadsheet then subtracts the zero reading, converts any “negative” concentrations to zero, and reports the calculated concentration after adjusting for any user-defined dilution of the DNA sample read. Next, the spreadsheet calculates the amount of DNA sample and water that need to be combined to normalize all the samples to the same volume and concentration. This normalization sheet can serve as a guide for manual dilution of DNA. Additionally, we have used a PerkinElmer MultiPROBE II PLUS HT EX (Downers Grove, IL) liquid handling system to prepare the quantification assay and to make working solutions by directly using the data generated in the normalization worksheet.

Using the quantification method described, we estimated DNA concentrations from leaf extractions of 1839 *P. americana* trees and 55 *T. grandiflorum* trees. Surprisingly, the concentrations of the individual DNA extractions ranged from 0 ng mL$^{-1}$ to 900 ng mL$^{-1}$ (average of 112 ng mL$^{-1}$) and standard deviation of 125 ng mL$^{-1}$) from identical amounts of tissue processed by the same experienced technician using the same kit. Genotype may play a role, because some individuals from the *P. americana* F1 population repeatedly gave low concentrations of DNA, even after three to four extraction attempts. The large variation in sample concentrations dramatically affects individual samples after a global dilution factor is applied based on the average of a subset of samples, template dilutions of 1:20 were prepared and used for PCR reactions, which resulted in 70% successful amplification. DNA was extracted a second time from the same 55 trees, quantified, and normalized to 4 ng mL$^{-1}$ for PCR reactions. This normalized DNA resulted in 98% successful amplification. Furthermore, in a data set comprised of 96 individuals from *Phaedranaassa* species, microsatellite amplification success with three primers was 66.7% when samples were diluted 1:10 immediately after extraction. Alternatively, amplification success increased to 97.6% when the same *Phaedranaassa* samples were first normalized to 4 ng mL$^{-1}$ after quantification with the SG method. These data show the ability to successfully perform genotyping reactions is significantly greater when the individual samples are first quantified and normalized.

The increase in genotyping success after normalization is further demonstrated by a large-scale genotyping project performed in our laboratory. This project involved genotyping 1839 avocado (*P. americana*) individuals using six microsatellite markers. After extraction, the DNA was subjected to SG quantification and normalized to 4 ng mL$^{-1}$. Unfortunately, some samples, despite numerous extractions, were not concentrated enough to normalize to 4 ng mL$^{-1}$. Thus, two sample subsets were created: one normalized to 4 ng mL$^{-1}$ and a second that could not be normalized. Samples in the second subset were then be attempted on those 600 samples with the markers that provided no data. This process would add 7 to 10 d to the genotype analysis time. Conversely, the time required to quantify and normalize is no more than 1 d. Rerun to produce usable data, which is expensive and significantly reduces throughput. To determine the optimal concentration of DNA to be used in microsatellite analysis, samples from *T. cacao* and *P. americana* were first quantified using SG, and then a concentration gradient was established for each. Microsatellite analysis was performed using DNA across these gradients and peak height was plotted as a function of concentration (Fig. 3). Samples normalized to 4 ng mL$^{-1}$ produced consistently scorable peak height signals. RFU values were consistently lower for avocado than for cacao, which is likely the result of differing specific activities of the primers tested.

The effect of normalization to 4 ng mL$^{-1}$ was examined by comparing two attempts at genetic marker analysis on *T. grandiflorum*. DNA was extracted from 55 trees and, based on the average of a subset of samples, template dilutions of 1:20 were prepared and used for PCR reactions, which resulted in 70% successful amplification. DNA was extracted a second time from the same 55 trees, quantified, and normalized to 4 ng mL$^{-1}$ for PCR reactions. This normalized DNA resulted in 98% successful amplification. Furthermore, in a data set comprised of 96 individuals from *Phaedranaassa* species, microsatellite amplification success with three primers was 66.7% when samples were diluted 1:10 immediately after extraction. Alternatively, amplification success increased to 97.6% when the same *Phaedranaassa* samples were first normalized to 4 ng mL$^{-1}$ after quantification with the SG method. These data show the ability to successfully perform genotyping reactions is significantly greater when the individual samples are first quantified and normalized.

Table 1. Percent of avocado individuals that successfully amplified at a given number of microsatellite loci after normalization.

| Sample concn | Number of individuals | Number of loci with successful amplification |
|--------------|-----------------------|---------------------------------------------|
|              | 0 | 1 | 2 | 3 | 4 | 5 | 6   |
| Less than 4 ng mL$^{-1}$ | 63 | 23.81 | 30.16 | 15.87 | 7.94 | 14.29 | 3.17 | 4.76 |
| 4 ng mL$^{-1}$ | 1,776 | 0.34 | 1.18 | 1.07 | 2.36 | 6.14 | 20.66 | 68.24 |
| All samples | 1,839 | 1.14 | 2.18 | 1.58 | 2.56 | 6.42 | 20.07 | 66.07 |
In our laboratory, reagent cost for microsatellite analysis is $0.26 per reaction, whereas the reagent cost for the SG assay is only $0.07 per sample. Thus, quantifying all 1839 samples would cost $130, whereas the cost to rerun 600 samples with six markers would be $936. These costs do not include the additional labor costs associated with repeated microsatellite analysis. Therefore, not only does quantification and normalization improve amplification success, but the time and cost associated with quantification is substantially lower than having to attempt reanalyzing individuals that gave poor results as a result of low template concentrations.

The advantages of quantifying and normalizing all samples in an investigation cannot be underemphasized. The SG method described allows for quick, simple, and accurate quantification and normalization of a large number of samples in a 96-well format, and the method is easily automated through the use of a liquid handling system. Implementation of this normalization procedure has vastly improved the results of downstream applications, including microsatellite amplification and high-quality sequencing data across a variety of plants. It is our belief that any high-throughput plant genotyping facility can greatly benefit by incorporating this quantification and normalization step into the daily workflow.

**Literature Cited**

Borrone, J.W., R.J. Schnell, H.A. Violi, and R. Ploetz. 2007. Seventy microsatellite markers from *Persea americana* Miller (avocado) expressed sequence tags. Mol. Ecol. Notes 7:439–444.

Kuhn, D.N., J.C. Motamayor, A.W. Meerow, J.W. Borrone, and R.J. Schnell. 2008. SSCP markers provide a useful alternative to microsatellites in genotyping and estimating genetic diversity in populations and germplasm collections of plant specialty crops. Electrophoresis 29:1–14.

Labarca, C. and K. Paigen. 1980. A simple, rapid, and sensitive DNA assay procedure. Anal. Biochem. 102:344–352.

Lange, D.A., S. Penuela, R.L. Denny, J. Mudge, V.C. Concibido, J.H. Orf, and N.D. Young. 1998. A plant DNA isolation protocol suitable for polymerase chain reaction based marker-assisted breeding. Crop Sci. 38:217–220.

Le Pecq, J.-B. and C. Paolletti. 1966. A new fluorometric method for RNA and DNA determination. Anal. Biochem. 17:100–107.

Leggate, J., R. Allain, L. Isaac, and B. Blais. 2006. Microplate fluorescence assay for the quantification of double stranded DNA using SYBR Green I dye. Biotechnol. Lett. 28:1587–1594.

Oleas, N.H., A.W. Meerow, and J. Francisco-Ortega. 2005. Isolation and characterization of eight microsatellite loci from *Phaedranassa tunguragousa* (Amaryllidaceae). Mol. Ecol. Notes 5:791–793.

Rengarajan, K., S.M. Cristol, M. Mehta, and J.M. Nickerson. 2002. Quantifying DNA concentrations using fluorometry: A comparison of fluorophores. Mol. Vis. 8:416–421.

Sambrook, J., E.F. Fritsch, and T. Maniatis. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, New York, NY.

Sela, I. and Y. Antignus. 1971. Spectrophotometric determination of nucleic acids and nucleoproteins at the far-UV region. Anal. Biochem. 43:217–226.

Singer, V.L., L.I. Jones, S.T. Yue, and R.P. Haugland. 1997. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. Anal. Biochem. 249:228–238.

Vitzthum, F., G. Geiger, H. Bisswanger, H. Brunner, and J. Bernhagen. 1999. A quantitative fluorescence-based microplate assay for the determination of double-stranded DNA using SYBR Green I and a standard ultraviolet transilluminator gel imaging system. Anal. Biochem. 276:59–64.