Genetic purity and testing technologies for seed quality: a company perspective

J. S. C. Smith* and J. C. Register III

Pioneer Hi-Bred International, Inc., Research and Product Development, PO Box 1004, Johnston, IA 50131-1004, USA

Abstract

A high level of genetic purity in crop varieties must be achieved and maintained for agronomic performance as well as to encourage investment and innovation in plant breeding and to ensure that the improvements in productivity and quality imparted by breeders are delivered to the farmer and, ultimately, to the consumer. Traditionally, morphological comparisons have formed the basis for genetic purity evaluations. However, replicated field observations are time-consuming, expensive and unreliable. Morphology cannot provide information on the purity of specific genetic attributes that relate to grain quality or to pest or herbicide resistance bred into varieties. Biochemical assays, including isozymes, can distinguish varieties within several species. Isozymes have been routinely used in checking seed-lot purity in maize (*Zea mays* L.) for the past 20 years. Newer DNA-based technologies such as restriction fragment length polymorphisms and more recently developed methods that use the polymerase chain reaction can allow even more discriminative and faster identification of varieties. However, none of the DNA methods have replaced biochemical methods for seed purity assays, other than in a relatively select group of crops with very high seed value, due to their high datapoint cost. It will require further miniaturization, automation and enhanced capabilities to process numerous samples simultaneously before newly developed methods supplant biochemical methods for routine usage in purity testing. New varieties that have major genes for herbicide or insect resistance incorporated within them require purity assays during product development and following seed production of the commercial variety. Immunological or DNA sequence assays can be developed and automated systems are required to process hundreds of thousands of individuals. Ultra-high, micro-array technologies and single-molecule detection systems are now under development. These technologies offer the promise that adequate distinction and high sample throughput will be combined. New methods may eclipse the capabilities of biochemical methodologies, thereby potentially raising genetic purity standards and enabling farmers and consumers better to utilize and benefit from increasingly productive varieties that are bred from a more diverse base of genetic resources.

Keywords: AFLP, allele-specific hybridization, isozyme, maize, RAPD, RFLP, seed purity.

Introduction

Genetic purity within individual crop varieties is imperative for breeders so that: (i) intellectual property protection through Plant Variety Protection (PVP) or Utility Patents can be obtained and then subsequently maintained; (ii) varieties can be created with uniform appearance and agronomic performance that meet the demands of farmers, processors and consumers; (iii) varieties with stable genetic identities can be created so that agronomic performance can be as predictable as possible given unpredictable environmental fluctuations; and (iv) breeders can more completely and precisely characterize and measure genetic diversity so that genetic resources can be more thoroughly evaluated in terms of agronomic performance and thence more effectively utilized for the creation of improved varieties.

This paper provides a perspective on current and potential future approaches for evaluation of genetic purity. Although other species are noted, our focus is on the genetic purity of maize, and our perspective is that of a commercial seed producer.

*Correspondence
Fax: +1-515-270-4312
E-mail: smiths@phibred.com

Presented at the Symposium on Seed Biology and Technology, USDA National Seed Storage Laboratory, 1111 South Mason Street, Fort Collins, CO 80521-4500, USA, August 13–15, 1997

Abbreviations: AFLP = amplified fragment length polymorphism; ASH = allele-specific hybridization; CE = capillary electrophoresis; EDV = essentially derived variety; PCR = polymerase chain reaction; PIC = polymorphic index content; PVP = Plant Variety Protection; RAPD = random amplified polymorphic DNA; RFLP = restriction fragment length polymorphism; SSR = simple sequence repeat

Varietal distinction

Although varietal distinction is not the subject of this paper, it is related to varietal purity, so a brief mention of the history and status of technologies and approaches used for varietal distinction provides a useful context for discussion of varietal purity.

A new variety must be distinguishable from all previously known varieties of that species and be of uniform appearance and stable before it can be granted PVP or Utility Patent protection. Granting a Utility Patent also requires evidence of invention or unobviousness, utility and an enabling disclosure, usually in the form of a seed deposit that is available solely for the purposes of verifying the claims of the invention. A variety must be pure and stable for the same characteristics that are used for varietal distinction. Otherwise it would be simple to 'create' new varieties by purification without any effort in breeding to generate a more productive variety.

Observation of many morphological traits in field growouts has been, and continues to be, the most widely used approach for describing varieties *de novo*, identifying varieties and monitoring purity. After all, morphologies of varieties will be omnipresent in agriculture, and their comparison encompasses features that will be seen by the farmer. The *Union pour l'Obtention des Protections des Vegetales* (UPOV, 1988) prescribes lists of descriptors that are mainly based upon a series of 30–50 morphological traits specific for each crop species. *De novo* descriptions of crop varieties for PVP are based upon characters and standards set out in these UPOV descriptor lists. Morphological descriptions of varieties that form part of the application for a Utility Patent also frequently are based in part upon these lists; unlike PVP certificates, though, Utility Patent applications require agronomic performance information.

Because the genetic basis of most morphological traits remains unknown, morphological characters provide, at best, an indirect means of assessing genetic purity. Morphological comparisons provide a poor or incorrect basis for measures of genetic identities and relationships among varieties. Morphological evaluation cannot be used to assess modifications of protein or starch quality in grain, and it cannot detect the presence of specific genes imparting resistance to pests, diseases or environmental stresses without hugely detailed and replicated field trials. Evaluating the purity of dominant genes that affect these traits is of increasing importance as varieties have these genetics 'stacked' into them. For example, if a parental line possessing a dominant herbicide-resistance gene shows 0.25% susceptible plants, it would be expected that progeny from a cross to a non-resistant parent will result in 5% susceptible plants ($\sqrt{0.0025}$). Obviously, technologies that provide improved purity assays for such genes are of significant value to the seed industry. Finally, morphological comparisons are prone to inaccuracy and are very time-consuming and expensive because many characters should optimally be evaluated in replications over multiple years and environments. Further details can be found in Smith (1992).

There now exists a variety of laboratory-based technologies (e.g. biochemical characterization using seed storage protein or isozyme polymorphisms and/or DNA polymorphisms) for distinguishing and identifying varieties *de novo*. Readers are referred to Wrigley (1995) for a general review of variety identification. Widely used isozyme methodologies are available in maize due to the pioneering work of C. W. Stuber and M. M. Goodman (see Stuber *et al.*, 1988). Because molecularly based methods often have greater discriminatory power and because they measure genetic composition more directly than does evaluation of morphological characteristics, it would seem straightforward to use these technologies exclusively for such work. There are, however, drawbacks that may not be obvious. For instance, purity analysis, which is a prerequisite for determination of uniformity and stability, requires profiling tens to hundreds of individuals per line (or sample); the costs typically associated with molecular methods would be prohibitive here. Another concern with molecular methods is that small differences between individuals within a variety can usually be found if one examines enough plants with a sufficient number of loci. Granting new varietal status on the basis of selecting within existing varieties for unique individuals, based on an expanded list of molecular characteristics, would undermine the intent of existing protection and could potentially have a host of undesirable consequences, including rendering PVP meaningless and reducing the genetic diversity within crop species. Nonetheless, use of these genetic methods is commonplace and widespread in the scientific community. Consequently, efforts to understand better the implications of their use in varietal distinction and purity have been undertaken by such groups as the Biochemical and Molecular Techniques Working Group of UPOV and the International Association of Plant Breeders (ASSINSEL).

These molecular and biochemical methods can measure genetic distances and are well suited for determination of genetic conformity between varieties. Consequently, biochemical and DNA characters will probably be used first for varietal identification (for varieties that have already been described *de novo* using morphological characters) and to help determine whether new varieties are essentially derived varieties (EDVs) in accordance with new PVP laws based upon the most recent UPOV

revision. The widespread use of molecular data for varietal distinction *de novo* will be postponed, or at least somewhat restricted, until more cost-effective methods have been developed and until national and international bodies have developed and implemented standards for their use. Despite this general voluntary moratorium, the Australian PVP Office has already accepted random amplified polymorphic DNA (RAPD) data as evidence of distinction, and the US PVP Office has used evidence from a single microsatellite band to grant varietal distinction for soybean (*Glycine max* (L.) Merr.).

Technologies used historically for varietal identification and purity testing

Despite the advantages of using the genetic approaches over morphological characters for purity testing, the practical utility of any technique will depend not only on the information content of the data produced, but also on how rapidly and cost-effectively those data can be generated. Reviews of biochemical methods for varietal identification are provided by Cooke (1984, 1988, 1995). The first use of genetic loci for assessment of varietal purity was through both observation of isozymes extracted from maize and profiles of seed storage proteins in wheat and maize in the 1980s (Smith and Weissinger, 1984; Smith and Wych, 1986). For maize, 24–30 samples, each comprising 100 individuals, can be surveyed using 15–20 isozyme loci in approximately 70 person-hours. It takes one week to progress from seed to data, and the data can be scored as alleles at mapped loci. Isozyme data can also be used to help select for uniformity during inbred line development resulting in an increase in inbred genetic purity and greater assurance of hybrid purity. Seed storage proteins have been used mainly to characterize varieties of potatoes (*Solanum tuberosum* L.) and major grain or forage crops. These protein assays allow for a less complete coverage of the genome than does the complement of isozyme loci, yet their discriminatory power affords a level of assured genetic purity equivalent to that provided by isozyme analysis. In this analysis, proteins are separated by electrophoresis in acid or alkaline gels or by isoelectric focusing in starch, agarose or acrylamide. The current state of the technology allows for a throughput of 25–50 individuals in 2–4 person-hours per electrophoresis unit.

The question often arises whether molecular and biochemical methods for determining purity need to reflect genetic differences related to traditional morphological characters. There are really two issues which need to be addressed here.

First, for practical purposes, genetic approaches must optimally be able to identify seedlots that will express genetically-based morphological or chemical differences that would be of concern to the farmer and industrial customers, even if those differences have no agronomic significance. For example, in the case of maize, customers will not be satisfied with a crop that is genetically pure for a set of DNA or isozyme loci, but which expresses variability for plant height or kernel type that exceeds the normal bounds of experience or expectation. Because isozymes and genes affecting morphological traits are most usually coded by different and unlinked loci, a 'clean' isozyme profile will not necessarily correlate with morphological homogeneity. Despite this potential drawback, purity evaluations in maize using either isozymes or zein seed storage proteins have attained routine international usage over the past 20 years. Without a doubt, the combination of precision, speed and cost-effectiveness of these biochemical methods provides a practical way to ensure a high degree (albeit not absolute certainty) of genetic purity and thus a good guarantee of predictable and consistent varietal performance.

Second, a strict correlation between molecular and morphological differences will only be possible if there is tight linkage between the molecular marker loci and the loci that form the genetic basis for expression of the morphological traits, and if environmental factors do not significantly affect expression of the morphological traits. Varietal purity can usually be assured via molecular profiles to a level that is equivalent to morphological evaluation from field growouts provided the genome is sampled reasonably thoroughly by molecular markers that can adequately discriminate among elite adapted germplasm. One measure that is used to characterize the discriminatory ability of genetic methods is the polymorphic index content (PIC), which takes account not only of the number of alleles that are revealed among varieties for each locus, but also of the relative frequencies of these alleles. PIC values increase as the number of alleles increases and as the mean frequencies of these alleles become more similar. For maize, the PIC collectively provided by 20 routinely used isozyme loci or by zein profiles is sufficient to allow distinction between 75–85% of commonly used US inbred lines. Indeed, all commercial hybrids are made using unrelated inbred parents. They usually exhibit one to four polymorphic isozyme loci that distinguish between female selfs and true F_1 seed. Most instances of pollination by unrelated lines can also be detected with this level of PIC. Seedlots that are still segregating for one or more isozyme loci cause these loci to be of limited use at best for evaluating purity of hybrid seedlots made using these segregating parent stocks. Where instances of segregation for isozyme alleles in parental inbred lines are unknown, this lack of information will result

in incorrect interpretations for the cause of allelic variability in hybrid seedlots. However, with good breeding practices, residual segregation of isozymes can be kept fairly low. Isozymic data can also be used during the latter stages of inbreeding to help derive finished lines that do not segregate for these loci. The availability of several polymorphic loci for purity checking can also aid in the interpretation of profile data that deviate from expected F_1 profiles in the event that such residual segregation is encountered.

DNA-based technologies

Criteria for utility

Purity evaluations using biochemical methods based upon isozymes or seed storage proteins that are already in widespread and routine use provide benchmarks for PIC, throughput, reliability, labour requirements and costs that any replacement technology has to surpass before it can be considered a viable alternative. In our experience with maize, while it has not been difficult to find technologies providing a greater PIC than isozymes or seed storage proteins, the cost per datapoint of 5 US cents (costed at $50 per gel loaded with 100 individuals and assayed for 10 loci) for isozymes has, so far, proven impossible to beat. Whereas cost may not exert the same level of influence over other applications of genetic marker technologies, it has proven to be the primary barrier to adoption of new technologies for purity testing where hundreds of samples, each comprising seedlots of hundreds or even thousands of individuals, need to be assayed and the data recorded and evaluated in a matter of days or hours without significantly increasing the price of the seed. To date, even those DNA-based technologies with suitable throughput and robustness have fallen short in the area of cost-effectiveness. Whereas these assays may be eminently practical for providing data for protection of the intellectual property that is vested in a variety, they remain orders of magnitude outside the realms of cost and time that would be feasible for purity assays for most major field crops. For a recent review of molecular marker technologies, see Mohan *et al.* (1997).

To make any profiling technology suitable from a practical perspective for use in genetic purity evaluation, several key factors must be present:

(1) The degree of discrimination among varieties must be at least 75% and preferably higher.

(2) Unlinked loci (preferably loci from different chromosome arms) should be assayed.

(3) Using maize (with 10 chromosomes) as the model, 10 unlinked polymorphic loci are sufficient for purity testing of hybrids, but 20 unlinked loci would be an improvement. For more precise estimates of genetic distances, at least 60–100 loci (covering all chromosome arms) must be assayed. Such estimates will be needed to help determine the identities of EDVs for protection under PVP and to help ascertain whether varieties fall under the scope of protection for varieties for which Utility Patents have been issued.

(4) The data must be reproducible and easily obtained, scored and databased.

(5) Sample throughput and cost have to be at least 2- to 5-fold better than current approaches to warrant a change to new technology. Since our current throughput in a single laboratory using isozymes is 25 samples each of 100 individuals, then desirable capacity would be 50–250 samples per day at a cost of $5–$25 per sample (or 0.5–2.5 cents per datapoint).

(6) Methods must be standardized across laboratories on an international basis if they are to become internationally recognized procedures. For example, the International Seed Testing Association has published guidelines and standards for conducting electrophoretic assays of wheat varieties (ISTA, 1992). The publication of these and other guidelines facilitates the routine achievement and demonstration of high-quality standards as exemplified, for example, by ISO9000 accreditation. These systems of accreditation allow the simultaneous assurance of high quality standards with efficient provision of seed to customers by avoiding complex and often ineffective regulatory controls.

(7) Finally, although it is not necessary, it would be advantageous to utilize a technology that is both sufficiently cost-effective for use in purity testing and capable of providing more detailed genetic data for fingerprinting.

Restriction fragment length polymorphisms (RFLPs)

Detection of RFLPs was the first DNA-based method that revealed numerous polymorphisms that were inherited in a simple Mendelian fashion and which could be genetically mapped. Maize provides an excellent example of the high discriminatory power that can be obtained using RFLP technology (Weber and Helentjaris, 1989). Data from 60–80 loci that collectively survey each chromosome arm allow all but the most closely related inbred lines to be distinguished, thus essentially enabling 'fingerprinting' of inbred and hybrid lines. For many species (e.g. soybean), RFLPs provided a higher level of discriminatory power than previously available using biochemical methods (Smith and Smith, 1992; Smith, 1995).

Despite these advantages, RFLP analysis is slow and requires relatively large amounts of plant material, intensive labour support and much laboratory space, making it very expensive. For

example, during 3 months, two full-time staff could profile 200–300 individual samples (two or three seedlots of 100 individuals each) using 20 RFLP markers at a cost of $8000 to $12 000. Using 20 isozyme loci, however, two full-time staff could instead profile 25 samples of 100 individuals during each of 4 days per week and so profile 1200 seedlots during this same 3 months. Furthermore, RFLPs do not reveal much variation in some crops, such as tomato (*Lycopersicon esculentum*) and wheat (*Triticum aestivum*) (Smith and Smith, 1992; Smith, 1995), and so at present RFLPs cannot be used to evaluate purity in these crops. For these reasons, widespread use of RFLPs for purity analysis has not yet been adopted.

Polymerase chain reaction (PCR)-based methods

PCR-based methods offer new opportunities for genetic purity analysis since small amounts of (typically less pure) DNA are required, and profiles can be obtained more quickly than with RFLPs (in days rather than weeks). PCR-based methods such as RAPDs (Williams *et al.*, 1990), amplified fragment length polymorphisms (AFLPs; Vos *et al.*, 1995) or microsatellites (also called simple sequence repeats (SSRs) (Powell *et al.*, 1995; Brown *et al.*, 1996) are more cost-effective than RFLPs (at 20 cents per datapoint versus $1–2 or more per datapoint for RFLPs) and faster (days for SSRs, weeks for RFLPs) but are subject to the relatively high expense of thermostable polymerases and the time, personnel and space needed to run and score numerous gels. Start-up costs, especially if they include both the identification of SSRs and their multiplexing into sets of markers, are considerable (1–2 years of full-time effort for two or more people). Although the start-up costs are high, potential savings through use of automation and reduction in reaction volumes exist for PCR-based methods not available for RFLPs.

Random amplified polymorphic DNA and amplified fragment length polymorphisms (RAPDs and AFLPs)

Neither RAPD nor AFLP requires extensive or costly marker development as no up-front knowledge of target DNA sequences is necessary, so these are the least expensive PCR-based methods to initiate in a genetic purity testing program. However, these characteristics, which make these two approaches attractive to implement from a technology perspective, have been reported to provide irreproducible and sometimes unexpected results (Riedy *et al.*, 1992 ; Heun and Helentjaris, 1993).

Of these methods, RAPD is the least accepted because of the relatively low degree of complementarity between primer and target DNA sequence, making the test difficult to standardize because of differing results from separate profilings even in the same laboratory. The relatively low primer annealing stringency used in RAPD analysis apparently also results in lack of amplification of some parental bands in the F_1 hybrid (Heun and Helentjaris, 1993). Overall, lack of reproducible results greatly compromises the accuracy and the practicality of using RAPDs for purity analysis. The ability to obtain robust data from RAPDs can be increased provided great care is taken to monitor DNA cycling, to standardize chemistry and DNA concentration. Satisfactory repeatability can usually be attained for samples amplified within at least the same laboratory provided immense care is taken and check samples are included to evaluate thoroughly variations in amplification that do not have a genetic basis. However, usually it will be necessary to include standard genotypes of assured high purity along with each set of individuals of the variety being assayed to facilitate correct data interpretation. While RAPD data are easy to generate, hence the large number of publications using this methodology, challenges with the quality and interpretation of such data mean that this method should be used with great care and only when no other method is practically feasible. Some of these problems can be alleviated by cloning and sequencing RAPD products, redesigning longer PCR primers, and then converting these random markers to sequence-characterized amplified regions, or SCAR markers (Paran and Michelmore, 1993). However, the effort and expense required to do this then offsets the major advantage of RAPD technology.

AFLP technology appears to circumvent most of these difficulties by using high stringency PCR primer annealing conditions to known DNA sequences that are ligated onto restriction fragments. AFLPs, however, continue to be a problem in that inter-laboratory variability remains an issue for purity testing as does the fact that, across a range of germplasm, identical reactions can (reproducibly) generate genetically different markers (John Arbuckle, personal communication). Finally, neither RAPDs nor AFLPs produce data that can be unambiguously and readily scored as co-dominant alleles at mapped loci. This limitation poses a practical problem for genetic purity assays since outcrosses can then remain undetected.

Simple sequence repeats (SSRs) or microsatellites

In contrast to RAPDs and AFLPs, SSR technology is initially expensive to implement, and the method must be independently initiated for most individual species. DNA sequences from the target species must be obtained and screened for (usually) di-, tri-, or tetra-nucleotide repeat motifs occurring in tandem arrays. The regions flanking each tandem array are

then sequenced, and primers designed for amplification of the intervening repeat region. Regardless of whether the starting point is enriched DNA repeat libraries or sequence databases, primer design and marker screening for polymorphisms among varieties of the species of interest are required. This is because (i) some SSRs will be more polymorphic than others across the varieties of interest; (ii) map information must be obtained and subsequently considered; and (iii) assessment of the robustness of amplification reactions must be carried out empirically. The results of such efforts usually are a set of highly informative, multiallelic and co-dominant markers. Because PCR product size is used to define alleles in this technology, size separation of PCR products is required. Depending on the sizes of the PCR products analysed, separation using agarose, MetaPhor (FMC BioProducts) or polyacrylamide can be used. If a single marker is analysed in each lane, the cost and throughput for SSR analysis are dependent on the number of thermocyclers, electrophoresis units and people devoted to the effort. While the throughput can be greater than that for RFLPs, it still remains several orders of magnitude below that of isozymes. Such costs may be acceptable for very high value seed-crops such as tomato where the need for genetic purity is at a premium, and use of SSRs in this fashion has already become routine by some tomato breeders and seed producers.

Throughput and cost-effectiveness can be improved by using PCR primers tagged with fluorescent dyes and by investing in the hardware and software that is necessary to allow for the capture and analysis of profile data (Oetting *et al.*, 1995; Mitchell *et al.*, 1997). Companies such as Applied Biosystems Inc./Perkin Elmer and Li-Cor have adapted their high-throughput DNA-sequencing hardware and analysis software for this purpose and have commercial systems. Because multiple dyes are available with this technology, additional up-front effort into designing primers permitting amplification and resolution of multiple markers in single PCR reactions (multiplexing) and gel lanes is rewarded with a greater throughput (Oetting *et al.*, 1995; Mitchell *et al.*, 1997). We (Sullivan, H.R., Cook, D.A. and Register, J. C., III, unpublished data) and other laboratories (Mitchell *et al.*, 1997) have had repeated success multiplexing at least 10 markers into single robust PCRs that can be separated simultaneously in a single electrophoresis lane. But the up-front costs to achieve multiplexing alone (with a set of useful SSRs already provided) are considerable. Starting with a candidate set of 100 maize SSRs, creating a multiplexed set of 50 SSRs can take approximately 6 months of full-time work.

Using maize, preliminary results suggest that 10 SSR loci could collectively provide a PIC approximately equivalent to that provided by 20 isozyme loci currently used for genetic purity assessments. Using currently available technology (with the capacity to run two gels at a time), anywhere from 60 to 150+ seedlots could be profiled with the same 2–3 person-months of effort necessary to screen 1200 seedlots using isozymes or two or three seedlots with RFLPs. This wide estimate of sample throughput is due to the differing capacities of currently available SSR separation technologies. Because this field is evolving rapidly, these estimates will probably be superseded in a few months' time. Although the high-end estimates for SSR throughput bring the technology to within an order of magnitude of isozymes, the cost per datapoint still remains an additional factor favouring isozymes. Taken together, this level of cost-effectiveness and throughput still restricts the use of SSRs for genetic purity assays to high-value seed-crops such as tomato. It is not yet feasible to use this methodology to assay genetic purity in crops such as maize or sorghum.

One emerging technology that could have a significant impact on both the throughput and cost of applying SSRs is capillary electrophoresis (CE). CE has already proven useful in DNA sequencing applications (Lu *et al.*, 1994) and, analogous to what has occurred with the gel-based technologies, adaptation to SSRs should be possible. The primary drawback with CE has been that single capillary units do not allow for sufficient sample throughput. Multicapillary units (up to 96 capillaries) are likely to be commercialized in the near future. When such hardware (and the appropriate analytical software) becomes available, its compatibility with automated sample loading offers the potential for dramatic throughput increases over current technologies and may widen the scope for practical use of SSRs for purity testing.

Future technologies

Requirements

As should be clear from this discussion, given a satisfactory PIC content coupled with reliable data acquisition and interpretation, the suitability of any technology for genetic purity applications becomes dependent on cost-effectiveness. Two components contribute to cost-effectiveness: the monetary cost of applying the technology (reagents, consumables, etc.) and sample throughput. As noted above for SSRs, there are three primary ways to reduce these costs: (i) miniaturization, which reduces use of consumables and reagents; (ii) automation of all aspects of the process from sample extraction through data acquisition, databasing, interpreting and reporting;

and (iii) parallel processing (i.e. simultaneous processing of many samples). Although the minimum cost-effectiveness threshold will vary across crop species (according to the value of the seed), cost has the single greatest impact on the suitability of a technology for genetic purity applications in general. The ability to achieve cost reduction only comes after very large investments into multiplexing PCRs, electrophoretic or hybridization assays and the purchase of expensive equipment ($100 000 to $200 000). Thus, prospects for new automated procedures are only available for organizations with large budgets.

Whereas the principles underlying miniaturization, automation and parallel processing can be applied to varying degrees for nearly any technology (as evidenced in the discussion above on SSRs), other technologies are evolving using these principles as their starting point. Much of this development is coming from the Human Genome Project. However, the long-term applications are even more widespread.

Array technologies

Arrays with 96 positions (8 x 12) have become the standard for industrial sample-processing for technologies as diverse as PCR and immunoassays. For genetic markers, allele-specific hybridization, in which binding of a sequence-specific probe (often an oligonucleotide) to a DNA target (often following PCR amplification) depends on the presence of a specific allele (Wallace *et al.*, 1979; Conner *et al.*, 1983), can be developed around such a format. Allele-specific hybridization employing dot blots has been demonstrated (Koenraadt and Jones, 1992; Erlich *et al.*, 1995), as has a dot blot/hybridization-based variation of RAPD technology (Penner *et al.*, 1996). Commercially available hardware facilitates creation of dot-blot arrays with up to 384 dots in the same physical space as a standard 8 x 12 array. If one moves away from membranes to other supports, such as glass slides or silicon chips, array densities of >10 000 dots cm^{-2} have been demonstrated (Schena *et al.*, 1995; Chee *et al.*, 1996). Although such array technologies are still in their infancy for plant genetic studies, their potential for cost-effectiveness will surely drive evaluation of their suitability for use in a wide variety of areas, including genetic marker, and thus genetic purity applications.

It is worth noting that, unlike the technologies described above, most assays developed around array technologies will be dependent upon dominant alleles, or +/−. While information from such an assay is not as informative as information from co-dominant systems on a per locus basis, the ability to assay several thousand loci may compensate for this difference.

Single-molecule detection

Another approach being developed, in a collaboration between scientists at Pioneer Hi-Bred International, Inc. and Los Alamos National Laboratory, avoids the cost and reliability problems associated with PCR amplification and instead relies on detection and analysis of single molecules of DNA (Castro and Williams, manuscript submitted). In this approach, sequence-specific probes labelled with fluorescent dyes are mixed with restricted genomic DNA. This mixture is size-separated using CE, and individual probe–target complexes detected using a laser-based ultrasensitive fluorescence system. Results equivalent to those obtained with RFLPs could be produced using this technology but with far greater throughput and far less cost. Although efforts to implement this technology are even less advanced than those described above for array technologies, this approach does hold promise as a co-dominant multiallelic alternative to array-based methodologies for genetic marker studies.

Varietal purity of single gene or multiple gene 'stacked' conversions including transgenic varieties

Varieties have already been commercialized and harvested that have traits for herbicide or insect resistance that have been introduced by molecular marker-assisted backcrossing and/or by transformation. Presence of these genes needs to be assayed both during the product development process and in the final stage of purity assay before sale to the farmer. Varieties that contain these traits demand a significant price premium, and expectations from the farmer that the product will be of the highest quality are raised to new levels. Less than 0.1% contamination for a herbicide-resistance gene is very desirable to achieve. Assays are of three types: (i) a herbicide spray, or insect, fungal or toxin application to seedlings; (ii) a laboratory assay based on the expressed gene via an ELISA test; or (iii) a laboratory assay based on PCR DNA amplification. Laboratory assays can be developed in approximately 2–3 person-months since the necessary sequence data are usually available from which to develop nucleic acid hybridization or ELISA assays. A large plant-breeding organization will typically be conducting several hundred thousand assays each year to support its product development and seed production activities. Laboratory automation is an absolute necessity to monitor seed purity with the current production of gene-enhanced varieties. Many more varieties with more stacked traits will soon be available thereby increasing the requirement for purity assays still further.

Conclusions

The seemingly old-fashioned, well tried and tested biochemical methods have remained the premier methods for genetic purity testing in the face of the introduction of newer molecular methods, including those using PCR amplification of DNA. Biochemical methods have endured because newer technologies have not matched the cost-effectiveness of isozymes or seed storage proteins. For some crops, however, biochemical methods have not proven sufficiently discriminative, and so the newer molecular methods have gained increasing acceptance. Nevertheless, even in crops for which genetic purity can be reliably and cost-effectively assayed using biochemical methods, a method(s) of generic utility with greater discriminatory ability would markedly help to assure even higher standards of seed quality to the customer. Therefore, efforts to evaluate technologies with greater PIC than the traditional biochemical methods continue, with the focus being on ways to make these methods as cost-effective as possible. Of the current technologies, SSRs are the only possible alternative to the biochemical methods, but isozymes remain at least an order of magnitude more cost-effective. Use of automated CE for size separation may narrow that gap within the foreseeable future. Looking to a 5–10-year time-frame, array-based technologies and novel approaches such as single-molecule detection hold even greater promise of surpassing the standards that have been set by isozyme and seed storage protein analysis which have endured the last two decades. Varieties with herbicide or insect resistance have dramatically increased the need to assay hundreds of thousands of individuals quickly and reliably. Assays can usually be developed within 2–3 months that are then dependent upon automated laboratory procedures to help ensure the provision of varieties with a high level of purity for these traits.

References

Brown, S.M., Szewc-McFadden, A.K. and Kresovich, S. (1996) Development and application of simple sequence repeat (SSR) loci for plant genome analysis. pp 147–159 *in* P. P. Jauhar (Ed.) *Methods of genome analysis in plants.* Boca Raton, Florida, CRC Press.

Chee, M., Yang, R., Hubbell, F., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S. and Fodor, S.P.A. (1996) Accessing genetic information with high-density arrays. *Science* **274**, 610–614.

Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R. L. and Wallace, R.B. (1983) Detection of sickle cell β^s-globin allele by hybridization with synthetic oligonucleotides. *Proceedings of the National Academy of Sciences of the USA* **80**, 278–282.

Cooke, R.J. (1984). The characterization and identification of crop cultivars by electrophoresis. *Electrophoresis* **5**, 59–72.

Cooke, R.J. (1988) Electrophoresis in plant testing and breeding. *Advances in Electrophoresis* **2**, 171–261.

Cooke, R.J. (1995) Gel electrophoresis for the identification of plant varieties. *Journal of Chromatography* **698**, 281–299.

Erlich, H.A., Horn, G., Saiki, R.K. and Mullis, K.B. (1995) Process for detecting specific nucleotide variations and genetic polymorphisms present in nucleic acids. US Patent No. 5,468,613.

Heun, M. and Helentjaris, T. (1993) Inheritance of RAPDs in F_1 hybrids of corn. *Theoretical and Applied Genetics* **85**, 961–968.

ISTA (1992) *ISTA handbook of variety testing – Electrophoresis handbook: variety identification* (R. Cooke, Ed.). Zurich, International Seed Testing Association.

Koenraadt, H. and Jones, A.L. (1992) The use of allele-specific oligonucleotide probes to characterize resistance to benomyl in field isolates of *Venturia inaequalis*. *Phytopathology* **82**, 1354–1358.

Lu, H., Arriaga, E., Chen, D.Y. and Dovichi, N.J. (1994) High-speed and high-accuracy DNA sequencing by capillary electrophoresis in a simple, low cost instrument. Two-color peak-height encoded sequencing at 40°C. *Journal of Chromatography Analysis* **680**, 497–501.

Mitchell, S.E., Kresovich, S., Jester, C.A., Hernandez, C.J. and Szewc-McFadden, A.K. (1997) Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. *Crop Science* **37**, 617–624.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T.G., Yano, M., Bhatia, C.R. and Sasaki, T. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding* **3**, 87–103.

Oetting, W.S., Lee, H.K., Flanders, D.J., Wiesner, G.L., Sellers, T.A. and King R.A. (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics* **30**, 450–458.

Paran, I. and Michelmore, R.W. (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics* **85**, 985–993.

Penner, G.A., Lee, S.J., Bezte, L.J. and Ugali, E. (1996) Rapid RAPD screening of plant DNA using dot blot hybridization. *Molecular Breeding* **2**, 7–10.

Powell, W., Machray, G.C. and Provan, J. (1995) Polymorphisms revealed by simple sequence repeats. *Trends in Plant Science* **1**, 215–222.

Riedy, M.F., Hamilton, W.J. and Aquadro, C.F. (1992) Excess of non-parental bands in offspring from known primate pedigrees assayed using RAPD PCR. *Nucleic Acids Research* **20**, 918.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.

Smith, J.S.C. (1992) Plant breeders' rights in the USA; changing approaches and appropriate technologies in support of germplasm enhancement. *Plant Varieties and Seeds* **5**, 183–199.

Smith, J.S.C. (1995) Identification of cultivated varieties by nucleotide analysis. pp 131–150 *in* C. W. Wrigley (Ed.) *Identification of food-grain varieties.* St Paul, Minnesota, American Association of Cereal Chemists.

Smith, J.S.C. and Smith, O.S. (1992) Fingerprinting crop varieties. *Advances in Agronomy* **47**, 85–140.

Smith, J.S.C. and Weissinger, H.H. (1984) Rapid monitoring of purity in seed lots of hybrid maize: modifications of current technologies. *Maize Genetics Cooperation Newsletter* **58**, 103–105.

Smith, J.S.C and Wych, R.D. (1986) The identification of female selfs in hybrid maize: a comparison using electrophoresis and morphology. *Seed Science and Technology* **14**, 1–8.

Stuber, C.S., Wendel, J.F., Goodman, M.M. and Smith, J.S.C. (1988) *Techniques and scoring procedures for starch gel electrophoresis of enzymes from maize (Zea mays L.).* Technical Bulletin 286. Raleigh, North Carolina, North Carolina Agricultural Research Service, North Carolina State University.

UPOV (1988) *International convention for the protection of new varieties of plants.* Geneva, International Union for the Protection of New Varieties of Plants.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407–4414.

Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1979) Hybridization of synthetic oligodeoxyribonucleotides to FC174 DNA: the effect of single base pair mismatch. *Nucleic Acids Research* **6**, 3543–3557.

Weber, D. and Helentjaris, T. (1989) Mapping RFLP loci in maize using B–A translocations. *Genetics* **121**, 583–590.

Williams, J.G.K., Kubelik, A.R., Livak, K. J., Rafalski, J.A. and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**, 6531–6535.

Wrigley, C.W. (1995) *Identification of food-grain varieties.* St Paul, Minnesota, American Association of Cereal Chemists, Inc.

Received 5 November 1997, accepted
5 February 1998