CrossMark

Decision support for cost-efficient and logistically feasible marker-assisted seedling selection in fruit breeding

Daniel A. Edge-Garza · James J. Luby · Cameron Peace

Abstract Development of genetically improved new cultivars is expensive, especially for tree fruit which may take years for the organs of economic importance to appear for direct evaluation. A costly and time-consuming part of breeding programs is seedling selection, which involves identifying a few seedlings from among many thousands that have the genetic potential for desired performance levels. New DNA-based diagnostic tools exist that can assist with resource-use efficiency and prediction accuracy of seedling selection, the application of which is known as marker-assisted seedling selection (MASS). However, MASS by fruit breeders is not yet routine, partially because of the lack of decision support for identifying resource-efficient schemes. The MASS Efficiency Calculator version 1.0 described here can be used to model the costs and timing of a breeding program's traditional operations and combine them with the costs and logistics of DNA testing operations to identify cost-efficient MASS schemes. With this tool, users can explore various possible scenarios to identify conditions in which MASS is predicted to increase cost efficiency—without the need for complicated hand calculations or de novo spreadsheet creation. Users can determine whether enough time will be available for integrating DNA testing into traditional operations. Demonstration of the spreadsheet technology identified logistically feasible and cost-efficient schemes for several scenarios described in an earlier study reported in 2001 with a more realistic complex breeding design modeled here. In these scenarios, up to 43 % of operational costs were predicted to be saved over the first 6–8 years of breeding program operations. These results indicated that MASS can often b
e very cost efficient, enhancing resource-use efficiency under typical breeding situations with even a single DNA test. When using multiple DNA tests, maximal cost efficiency can be achieved by deploying DNA tests in sequence rather than together. Even when up to 60 % of DNA test results must be disregarded due to technical issues such as failed DNA extractions, substantial savings in resource use can be realized if a high proportion of seedlings are culled. For long-juvenility crops such as apple and grape, breeders have considerable flexibility in choosing when to implement MASS during seedling production while still achieving increased cost efficiency.

D. A. Edge-Garza · C. Peace (✉)
Department of Horticulture, Washington State University, Pullman, WA 99164, USA
e-mail: cpeace@wsu.edu

D. A. Edge-Garza
e-mail: daniel.edgegarza@email.wsu.edu

J. J. Luby
Department of Horticultural Science, University of Minnesota, St. Paul, MN 55108, USA

🗬 Springer

Introduction

In the development of superior new fruit cultivars by breeding, program operations include parent evaluation, parent selection, parent crossing, seedling raising, seedling evaluation, and seedling selection (Galletta 1975; Zhebentyayeva et al. 2012; Evans 2013; Ru et al. 2015). Ru et al. (2015) also describe advanced seedling selection and replicated trials following seedling selection. "Traditional seedling selection" starts from seed germination and continues to just before advanced selection (Peace and Norelli 2009; Badenes and Byrne 2012; Ru et al. 2015). An example of this definition for apple breeding would include the "seedling production" and "phase 1" stages of seedling trials described by Evans (2013). For such a tree fruit crop, the fruit targets for selection are not produced until seedlings are physiologically mature (Janick and Moore 1975; Luby and Shaw 2001; Dirlewanger et al. 2004; Ru et al. 2015). Raising seedlings through the juvenile phase, a period often extending 2–5 or more years for tree fruit, makes traditional seedling selection expensive even when dealing with only a few thousand genetically distinct plants. The long juvenile phase reduces operational efficiency because funds are unnecessarily wasted on raising genetically inferior seedlings before fruits are produced and evaluated (Luby and Shaw 2001; Edge-Garza et al. 2010; Karaagac et al. 2012; Ru et al. 2015). Therefore, maximizing cost efficiency of traditional seedling selection to identify potentially superior cultivars is an important factor on the minds of fruit breeders (Luby and Shaw 2001; Dreher et al. 2003; Morris et al. 2003).

New breeding tools exist that can increase efficiency of seedling evaluation and subsequent selection (Collard and Mackill 2008; Peace and Norelli 2009); one of these tools is marker-assisted seedling selection (MASS). As Luby and Shaw (2001) noted that breeders would not likely invest in a technology that provides no added value, a suggested added value through MASS is improved internal cost efficiency of a breeding program (Moreau et al. 2000; Dreher et al. 2003; Morris et al. 2003; Hospital 2009; Slater et al. 2013; Ru et al. 2015). However, few published analyses are available showing cost efficiency of MASS in fruit crops (Moreau et al. 2000; Luby and Shaw 2001; Collard and Mackill 2008; Bliss 2010; Edge-Garza et al. 2010), and the perceived costs and lack of available or reliable DNA tests and supporting technology are mentioned by some fruit breeders as reasons for not employing MASS in their programs (Byrne 2007; Collard and Mackill 2008; Gallardo et al. 2012; Yue et al. 2012).

Luby and Shaw (2001) concluded that MASS most likely will make economic sense in fruit breeding under the following conditions: (1) It is conducted for simply inherited traits; (2) the trait is expressed in the mature phase of crops with long juvenile periods; (3) the seedlings are screened early in the juvenile phase; (4) the cost of traditional screening is comparatively more expensive; (5) DNA tests for trait targets of interest are available with inexpensive marker technology and marker-locus(-trait) associations are highly robust; and (6) DNA tests are more accurate in trait-level prediction than the best of established traditional practices. Although satisfying one or more of these conditions is not necessary for MASS to be a viable component of new cultivar development, doing so for each should favor MASS adoption to provide internal economic efficiency in a breeding program.

So, almost 15 years later, why has MASS still not been widely adopted in fruit breeding? One reason, among others outlined by Ru et al. (2015), is that determining the internal economic benefits of MASS can be complicated. For example, each MASS case will be unique for each breeding program and for each possible alternative scenario within the same breeding program (Luby and Shaw 2001; Peace and Norelli 2009; Slater et al. 2013). Additionally, many variables are included in both the cost structure of MASS and traditional seedling selection (Moreau et al. 2000; Dreher et al. 2003). Luby and Shaw (2001) concluded that DNA tests will economically improve only few cases of cultivar development, basing conclusions on 1998 breeding operations costs of breeding programs for three crops (apple, grape, and strawberry) and break-even costs calculated for MASS implementation that included selection indexing for traits only partially explained by DNA tests. For simplicity and clarity, Luby and Shaw (2001) assumed that seedling numbers were not reduced due to a loss in seed viability (e.g., no germination) or early phenotypic selection (e.g., susceptibility to pervasive diseases or low seedling vigor), and used only one level of reliability for DNA tests (95 % per locus due to recombination or technical problems in determining marker genotypes). Those assumptions are easily

broken by the variability of scenarios within and among breeding programs and DNA tests. Also, cost analyses of MASS should include time considerations and other aspects of logistical feasibility (Dreher et al. 2003). These additional considerations can include the achievable throughput of and access to genotyping platforms, DNA extraction technology (e.g., throughput levels and requirements for sample storage), scalability and technical reliability of DNA tests, and available software for information management (Dreher et al. 2003; Xu and Crouch 2008; Peace and Norelli 2009; Edge-Garza et al. 2010).

Routine use of new DNA-based technologies for high-throughput seedling selection in fruit breeding will be aided with refined assessments of cost efficiency and logistical feasibility for breeding scenarios that may be considered or encountered in a program. Because many variables are included in cost analyses (Luby and Shaw 2001), determining how cost-efficient MASS can be and when to apply the technology requires decision support (Groenendaal et al. 2004; Xu and Crouch 2008; Xu et al. 2012). Decision support has been developed for other complicated breeding tasks such as tracing desired alleles over multiple generations (Valente et al. 2013). The aim of this work is to develop decision support for MASS application in traditional seedling selection programs as described by Ru et al. (2015), by evaluating a single generation of alternative seedling selection scenarios under a wide range of fruit breeding situations with choices among the general features of DNA tests that a breeding program is likely to have access to. Here we describe and demonstrate MASS Efficiency Calculator version 1.0, a spreadsheet-based tool to model the cost efficiency and logistics of MASS, and reveal how MASS generally makes cost-efficient and logistical sense in many fruit breeding scenarios, improving resource-use efficiency.

Materials and Methods

Development of spreadsheet-based MASS decision support

Mathematical relationships among variables in the MASS cost equation of Luby and Shaw (2001) were modeled in a Microsoft Excel 2007 (Microsoft Corp.,

Redmond, Washington) spreadsheet and expanded to include additional variables that relaxed previous assumptions. All costs were in units of US dollars for illustration. Information about the breeding program stages for seedling testing was entered in section 1. Input variables included descriptions and corresponding costs of each stage of seedling selection (from generating seeds through raising and evaluating performance of seedlings to selecting those predicted to have superior genetic potential for target traits). Stages were defined and delimited by the beginning of a new expense for traditional operations (e.g., repotting seedlings to larger pots or transferring plants to a new location) and/or a reduction in plant number from routine and often unavoidable causes (such as lack of germination, culling due to poor vigor, and susceptibility to pervasive diseases). New input variables were duration of each stage and reduction proportion of seedlings in each stage. Section 2 ("Numbers of plants") included initial number of seeds entering the scenario under consideration and minimum number or proportion of seedlings that a breeder would be comfortable retaining after DNA testing regardless of actual DNA test predictions of genetic value. Descriptions of available DNA tests were entered in section 3. DNA testing input variables such as number of DNA tests, expected cull level for each test, and reliability of DNA testing in general (e.g., how closely linked the markers are to the trait locus) or its components (e.g., tissue sampling, DNA extraction, genotyping, and provision of results to the breeding program) were included from Luby and Shaw (2001). An additional input was number of markers included in each DNA test. In situations where the results of multiple markers were needed for an accurate cull decision (e.g., when two DNA markers flank a QTL or a multi-marker selection index is used for multi-locus traits), these markers were considered to constitute a single DNA test. Section 4 included costs of DNA testing, inputs of which were costs of equipment depreciation and costs and durations of components of DNA testing, including consumable and labor costs of obtaining each data point and incorporating any economies of scale as occurs when fixed costs are spread over an increased number of markers for certain marker platforms or service providers. Bypass inputs were added to override individual components of section 4 allowing users to enter fixed costs charged by external DNA testing service providers. Section 5 included inputs of

inflation rates, or any other pro-rated changes in operational costs of breeding. In section 6, durations were entered, including time required for DNA testing process according to seedling and marker numbers. These times were based on the historical average of a technician at WSU.

Outputs included total costs and costs per initial seed (enabling direct comparisons among strategies) for traditional seedling selection operations, for DNA testing alone, and for the most cost-efficient MASS scenario involving both DNA testing and later pheno-typing of retained seedlings. The spreadsheet was programmed to calculate and present: estimated cost savings in total and per initial seed for the optimal MASS scheme versus traditional selection; estimated cost–benefit ratio (MASS cost savings to DNA testing costs); estimated cost savings achievable by imple-menting DNA testing at various possible stages (where the most cost-efficient stage was designated as opti-mal); verification that sufficient time would be avail-able at a given stage to conduct the entire DNA testing process for the designated number of seeds/seedlings; and verification that a breeder-comfortable number of seedlings would remain after DNA testing. The spreadsheet was programmed to indicate whether the most cost-efficient DNA testing strategy would be sequential (i.e., culling seedlings after each test and subjecting only the retained seedlings to the next test) or simultaneous (i.e., running tests together by pooling assays or delaying culling according to each test until all test results are obtained). In situations of multiple DNA test use, a calculation was programmed to determine whether marginal savings provided by each DNA test were positive, with a red text warning given for DNA tests with negative marginal savings. The resulting decision support tool was called MASS Efficiency Calculator version 1.0.

Application of MASS decision support

Typical fruit breeding MASS cost efficiency out-comes were compared by modeling breeding opera-tion cost structures of the apple, grape, and strawberry breeding programs of Luby and Shaw (2001) in the MASS Efficiency Calculator version 1.0. Operational costs, inflation considerations, and timelines for each program were taken directly from Luby and Shaw (2001). Seedling reductions within certain stages unrelated to DNA testing (e.g., no germination, low

vigor, susceptibility to pervasive disease) were based on historical averages from the Washington Apple Breeding Program and kept constant across each breeding program for ease of comparison. For each program, seedling selection stages were defined using letters A through F (Fig. 1). For apple and grape, stage A was the time from the first availability of seed/seedling tissue (for use in DNA testing) to the first routine reduction in seedling number not asso-ciated with DNA testing. Stage A therefore began with obtaining seeds and ended with seedling germi-nation, which assumed the possibility of conducting DNA testing on seed tissue. Stage B began immedi-ately after stage A and ended at the next round of costs incurred through normal routine breeding oper-ations (e.g., transfer to nursery). Stage C began after the previous cost and extended to the next routine reduction in seedling number (culling due to low vigor) and assumed the possibility of conducting DNA testing on young leaf tissue. Stage D extended from stage C end with the next reduction in seedlings, which was culling due to early phenotypic selection (e.g., susceptibility to pervasive disease). Stage E followed, ending in field planting. Stage F began after field planting and ended when seedlings were removed at the end of seedling selection. The stages for strawberry were as follows: Stage A was the same as for apple and grape; stage B followed to the next round of costs for field establishment; stage C began with field establishment and lasted to the next round of seedling reductions; stage D followed these reductions through to culling due to early phenotypic selection; stage E followed early selection to the next cost for field evaluations; and stage F followed the evaluation costs through to seedling removal from the field.

Input parameters for DNA testing used for all scenarios were based on typical service provider costs (assuming 100 % reliability of DNA extraction and genotyping is provided at these prices). Costs and timing for tissue sampling, data organization, and results interpretation were based on our experience. Although the spreadsheet allowed for inclusion of costs for culling inferior seedlings after receiving DNA test results, it was assumed that such costs were not different from those already included in the cost estimates provided by Luby and Shaw (2001) for traditional apple, grape, and strawberry breeding (2001).
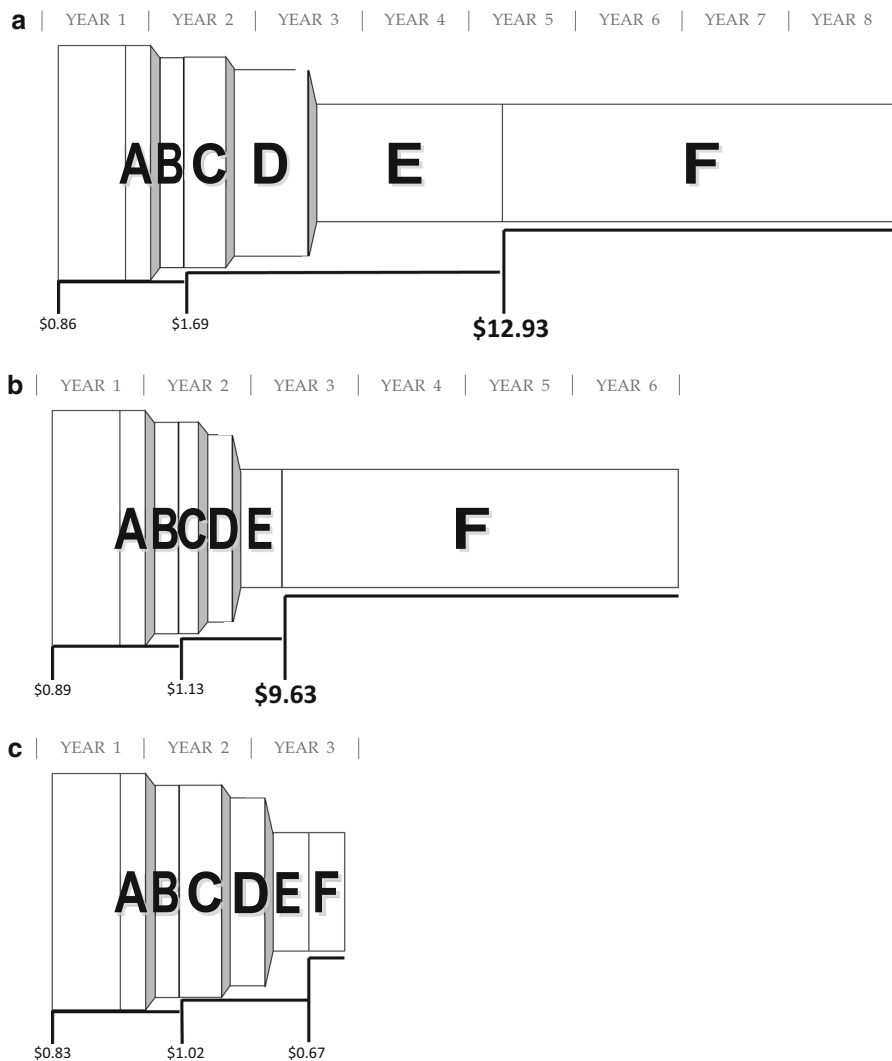
Fig. 1 Timeline and costs of traditional seedling selection for each of the apple (**a**), grape (**b**), and strawberry (**c**) breeding programs, with the possible DNA testing stages listed on each. For apple and grape: Stage A was between obtainment of seeds and seedling germination; stage B was between end of stage A and next round of costs incurred, stage C began after previous cost and ended after next routine reduction in seedling number, and stage D extended from stage C to culling due to early phenotypic selection. Stage E followed, ending in field planting. Stage F began after field planting and terminated at the end of seedling selection with the removal of the seedlings. For strawberry: Stage A was as for apple and grape. Stage B followed the next round of costs for field establishment. Stage C was from field establishment to the next round of seedling reductions. Stage D followed previous reductions to culling due to early phenotypic selection. Stage E followed early selection to the next cost in field evaluations, and stage F followed the evaluation costs through to seedling removal from the field. Costs for each breeding program from Luby and Shaw (2001)

Scenario 1: One DNA test, one marker, 50 % cull

Scenario 1 described a single DNA test with an expected cull proportion of 50 % (e.g., seedlings of an Aa × aa cross, where only aa seedlings are kept) and corresponded with "Case 2" of Luby and Shaw (2001). In this scenario, the test involved one marker. In this scenario and all others below, 10,000 seeds were hypothetically planted and the proportions of seedlings lost through routine breeding operations were 10 % each in stages A and C and 30 % in stage D (totaling 43 %).

Scenario 2: One DNA test, one marker, 75 % cull

Scenario 2 represented a single DNA test where 75 % cull was expected (such as the same test as scenario 1 used for a family derived from two heterozygous individuals for which the aa genotype is desired). This scenario corresponded with the first part of "Case 3" of Luby and Shaw (2001), again with one marker for the test.

Scenario 3: One DNA test, two markers, 75 % cull

Corresponding with the second part of "Case 3" from Luby and Shaw (2001), scenario 3 represented a DNA test requiring the results of two markers before a cull decision can be made for each seedling. Here, 75 % of seedlings were expected to be culled.

Scenario 4: Multiple DNA tests and markers

This scenario, recreated from "Case 4" of Luby and Shaw (2001), explored a complicated situation as would typically be encountered in real-world application, with the use of three DNA tests for three traits in total, each comprising a certain number of markers, on multiple families of various sizes. The first test, for the first trait, included one marker with an expected cull proportion of 50 %. The second test involved three markers required to predict performance of a second trait. One of these markers would be used to identify the presence of a desired AA or Aa genotype at one locus for a family of Aa × Aa comprising 68 % of the seedlings and for a second Aa × aa family comprising 32 % of the seedlings, and 33 % of all seedlings are culled. The other two markers would be flanking another locus containing a desired B allele, for which both families are Bb × Bb, and 25 % of the seedlings are culled. The combined three markers would therefore provide a total cull proportion of 50 % independent of the first test. The third test included four markers for a third trait, such as two pairs of markers flanking two trait-influencing loci, again with two families, the first and second families and genotypes in similar proportions as with the second test, and with a combined cull level of another 50 %. The total cull level of the three tests was therefore 87.5 %.

Special scenarios

Three special scenarios were explored for the Luby and Shaw (2001) apple program. In the first special

scenario, scenario 5, the cost efficiency of MASS was explored for a disease trait that would normally be evaluated by phenotype during stage D (nursery) of normal seedling selection operations. Therefore, only stages A through D were investigated for MASS opportunities. The DNA test included one marker and an expected cull proportion equal to that of phenotypic evaluation (30 %). In scenario 6, the cost efficiency of applying DNA markers to pyramid disease resistance alleles was investigated, considering the three DNA tests involved in scenario 4. Iterative costs were entered in stage D for a phenotyping protocol that would presumably give the same information as these DNA tests (i.e., ensure that the seedling inherited all disease resistance alleles) until a break-even cost was obtained. For scenario 7, several iterations of the sample size from scenarios 1 through 4 (7a through 7d, respectively) were entered to investigate how many more seeds could be sown by using saved costs from DNA testing for this particular purpose.

Sensitivity analyses

Several parameters of MASS Efficiency Calculator version 1.0 were varied to explore their influence on cost efficiency at each stage. Constant parameters were those corresponding to the costs and timeline of apple breeding from Luby and Shaw (2001). Estimated savings for each stage of DNA testing were recorded for each input variable level within breeding-appropriate parameters. The first sensitivity analysis explored the effects of increasing the proportion of seedlings lost due to routine breeding operations at stages A, C, and D when the cull level was either 50 % (as in scenario 1) or 75 % (as in scenario 2). Routine losses were varied by 10 % increments between 0 and 100 %, one stage at a time, while fixing reduction proportions in the other two stages at their values described in scenario 1. The second sensitivity analysis investigated the influence of DNA test reliability when seedling reductions at stage D were 30 and 40 % and the cull level of DNA tests were 50 and 75 %, assuming that the base cost did not include obtaining new samples to replace those yielding no results or to make up for false negative genotypes. DNA test reliability was varied between 0 and 100 % in 10 % increments. The third sensitivity analysis examined the relationship between MASS cost efficiency and fixed or variable costs of DNA testing per seedling. DNA testing variable costs (genotyping through information

delivery) associated with use of multiple tests (each with a 50 % cull level) or fixed costs (tissue sampling and DNA extraction) were increased in $1 increments until all previously identified cost-efficient MASS stage became cost inefficient. Seedling reductions from routine operations were set as described in scenario 1. For the fourth sensitivity analysis, the average change per year in traditional breeding program costs (inflation) was varied between 0 and 5 % in 1 % increments, to assess the impact on cost savings for the breeding program.

Results

Development of spreadsheet-based MASS decision support

The MASS Efficiency Calculator version 1.0 enabled comparisons among a wide range of possible seedling selection scenarios by allowing simultaneous consideration of changes in many variables influencing cost efficiency. The Calculator can be downloaded at its permanent location of www.rosaceae.org/breeders_toolbox/MASScalc; scenario 1 is presented upon opening. General instructions are also included in a separate worksheet to enable users to explore the scenarios of this study and other scenarios as desired.

Application of MASS decision support

In the typical fruit breeding scenarios investigated, the MASS Efficiency Calculator version 1.0 identified several cost-efficient windows for MASS for the apple and grape cases, but the hypothetical strawberry breeding program did not achieve positive cost savings from MASS due to its relatively inexpensive traditional seedling selection costs. For all scenarios, stage B (one of the stages before breeder-controllable losses in seedling numbers) and stage E (after breeder-controlled losses) were identified as the most cost efficient to implement DNA tests for MASS. Stage F was always cost inefficient and was therefore excluded from further consideration.

Scenario 1: One DNA test, one marker, 50 % cull

The Calculator indicated that stage E was the most cost-efficient DNA testing window. At this stage, the

hypothetical apple breeding program would save 24 % ($25,200; Table 1) of traditional operating costs within the first 8 years by incorporating DNA testing into seedling selection operations. Stage E was also identified as the most cost-efficient window for grape DNA testing, with a projected savings of 16 % ($12,400; Table 1) over the first 6 years of traditional operations. Stage B was the second most cost-efficient stage to apply DNA tests for both apple and grape (Table 1). Stages A, C, and D also achieved positive savings for both breeding programs, ranging from $14,400 to $19,100 for apple and $1600 to $4300 for grape (results not shown). Given the low traditional operation costs for the 3-year selection cycle of strawberry breeding in this scenario, employing MASS would not improve cost efficiency; rather it would cost at least an additional 74 % ($16,100; Table 1) and so no cost-efficient DNA testing stage was identified for this breeding program.

Scenario 2: One DNA test, one marker, 75 % cull

Increasing the cull level of the one-marker DNA test to 75 % from scenario 1's 50 % increased the savings for both the apple (42 % of traditional costs, $43,500) and grape (34 %, $26,000) breeding programs at stage E (Table 1). However, for apple, an additional $1100 (1 %, for a total of 43 %) would be saved if DNA testing occurred in stage B (Table 1). The greater cull proportion of this scenario still did not render MASS cost efficient for the strawberry breeding program (Table 1). Enough time was indicated by the spreadsheet for implementing MASS in any of the predicted cost-efficient testing stages of scenarios one and two.

Scenario 3: One DNA test, two markers, 75 % cull

The addition of a second marker to a DNA test but the same overall cull level (scenario 3 compared to scenario 2) reduced the cost efficiency of MASS (36 %, $37,200 saved for apple; 25 %, $19,700 for grape; Table 1). Also, with a more expensive test than for scenario 2, the tool indicated later use of DNA testing (stage E for both apple and grape). However, greater cost efficiency compared to traditional seedling selection was achieved regardless of the DNA testing stage used (stage B in Table 1; stages A, C, and D not shown). The Calculator indicated enough time was available to implement such a test into all possible

Table 1 Cost-efficient MASS scenarios for three fruit breeding programs as identified by the MASS Efficiency Calculator version 1.0

Characteristics of DNA tests					Savings								
					Apple			Grape			Strawberry		
Scenario	No. of tests	No. of markers	Proportion culled (%)	Cost of DNA testing	Stage B	Stage E	Proportion (%)	Stage B	Stage E	Proportion (%)	Stage B	Stage E	Proportion (%)
1	1	1	50	$3.22	$22,500	**$25,200**	24	$ 7,000	**$12,400**	16	−$22,000	−$16,100	−74
2	1	1	75	$3.23	**$44,600**	**$43,500**	43	$23,200	**$26,000**	34	−$18,800	−$15,200	−70
3	1	2	75	$4.33	$34,500	**$37,200**	36	$13,000	**$19,700**	25	−$28,900	−$21,500	−99
4	3	8	87.5	$6.20	$30,300	**$35,800**	34	$ 5,900	**$16,000**	21	−$42,500	−$31,600	−145

MASS savings provided to the apple, grape, and strawberry breeding programs described by Luby and Shaw (2001) were calculated assuming a 43 % reduction in the number of seedlings not associated with DNA testing (as described for scenario 1), before which the most cost-efficient stage was identified as stage B, and after which was stage E. Savings for the most cost-efficient stage for each scenario is shown in bold except for strawberry which had negative savings for all scenarios. For scenario 4, the first, second, and third DNA tests had 1, 3, and 4 markers, respectively, with each test sequentially culling 50 % of the seedlings for a combined cull level of 87.5 %. For apple and grape: stage A was between obtainment of seeds and seedling germination; stage B was between end of stage A and when the next round of costs were incurred; stage C began after the previous cost and ended after next routine reduction in seedling number; stage D extended from the end of stage C to culling due to early phenotypic selection; stage E followed, ending in field planting; and stage F began after field planting and terminated at the end of seedling selection with the removal of the seedlings. For strawberry: stage A was the same as for apple and grape; stage B followed to the next round of costs for field establishment; stage C was from field establishment to the next round of seedling reductions; stage D followed previous reductions to culling due to early phenotypic selection; stage E followed early selection to the next cost in field evaluations; and stage F followed the evaluation costs through to seedling removal from the field

cost-efficient DNA testing stages for apple and grape, despite the increased time required for DNA testing through the addition of the second marker.

Scenario 4: Multiple DNA tests and markers

The cost of DNA testing per sample increased from $3.22 in scenario 1 to more than $10.00 in scenario 4 due to the extra markers per DNA test and extra tests used. However, assuming that the tests were run in sequential order, the average cost due to reduction in seedlings occurring with each test was $6.22 (Table 1). MASS Efficiency Calculator version 1.0 indicated that only for apple at stage E could savings still be achieved by running the DNA tests simultaneously (data not shown). However, at all stages of the apple program, greater savings were always associated with sequential rather than simultaneous use of multiple DNA tests. Calculator outcomes indicated that the most cost-efficient MASS scheme was to begin with the test providing the largest savings on its own, followed by the next best test used only on those seedlings retained after the first test, and so on. With this strategy, the apple program would maximize cost savings in stage E at $35,800 (Table 1). Sequential DNA testing in the grape program would save $16,000 by testing at stage E, rather than cost $11,500 for simultaneous testing. No stage of grape breeding was identified where simultaneous DNA test use would be cost efficient, and only stages B, E (Table 1), and D (data not shown) were cost efficient for sequential DNA testing. At best, the strawberry program would spend (not save) an additional $31,600 with sequential testing at stage E (Table 1). Although the Calculator indicated that enough time would be available at the most cost-efficient stages to screen the seedlings with the DNA tests, it also indicated that not enough time would be available to screen all seedlings at stage B (if stage E was not possible) for the grape breeding program, because the duration of this window was too short at only 8 weeks.

Special scenarios

In scenario 5, no cost-efficient stages were identified. Instead, it would cost an extra $24,018 (99 %) or $29,038 (120 %) if seedlings were genotyped in stage B or stage D, respectively. In scenario 6, MASS would be cost inefficient unless the costs for phenotyping for pyramided disease-resistant seedlings were higher than $4.69 per seedling. For scenario 7a, the breeder could plant 3170 (32 %) more seeds. Seed plantings could be increased by 7490 (75 %), 5570 (56 %), and 6250 (63 %) for scenarios 7b–7d, respectively.

Sensitivity analyses

The optimal stage to apply DNA tests was sensitive to differences in the proportion of seedlings lost for reasons associated with routine seedling reductions at stage D (e.g., losses due to pervasive disease) (Fig. 2). Similar results were obtained from varying the proportions of seedlings lost in stages A and C; applying MASS at these stages decreased the cost efficiency at a faster rate than at those stages where MASS was applied after the losses. The Calculator consistently identified later stages with maximized cost efficiency for applying DNA tests, if the proportion of seedlings lost was more than 30 %. The Calculator identified earlier cost-efficient stages for applying DNA tests when higher test cull levels were considered.

In single DNA test scenarios, variation in DNA test reliability did not change the most cost-efficient stage (Fig. 3). The gain in savings at any particular stage with an increase in technical reliability was not linear, and for a higher culling level of DNA tests (i.e., 75 %) and lower routine seedling reductions in traditional breeding (i.e., 30 %), the Calculator still identified cost-efficient MASS schemes for apple breeding (Fig. 3) even if DNA testing reliability was low (i.e., 60 %).

DNA testing costs minimally affected the stage at which to apply DNA testing for cost-efficient MASS for maximum savings. Although other stages rapidly became cost inefficient as fixed costs elevated, stage E remained cost efficient even after these costs elevated above $5 per sample (Fig. 4a). If fixed costs of DNA testing remained constant (i.e., $2.10), cost efficiency became a function of the variable DNA testing costs (Fig. 4b). Stage E also remained cost efficient after variable costs elevated above $2.50 (Fig. 4b). In both cases, the Calculator indicated that using DNA tests in sequence would always be more cost efficient than using them together (Fig. 4). The Calculator consistently identified the same optimal cost-efficient stage to apply DNA tests (stage E), regardless of the inflation value. Increases in inflation proportionally increased savings over the long term (Fig. 5).
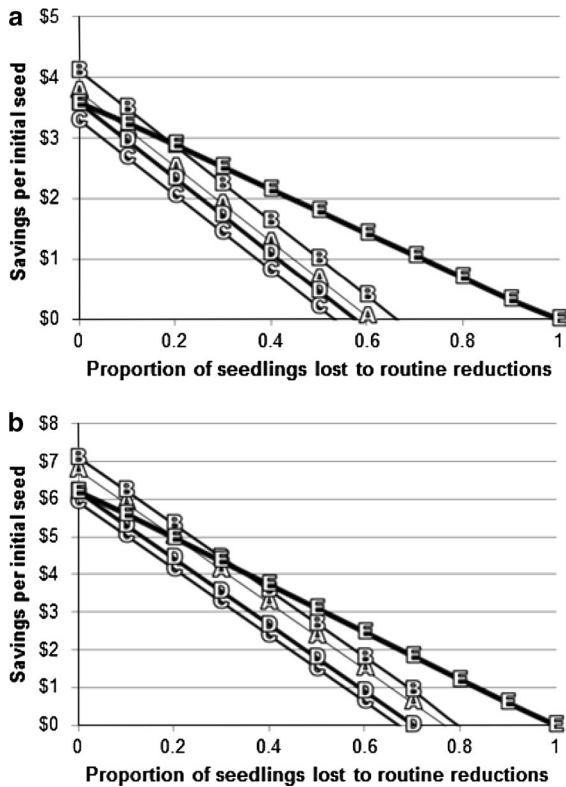
Fig. 2 Sensitivity to routine seedling reductions at stage D of the apple breeding program for identified optimal windows of cost-efficient MASS schemes for **a** scenario 1 (one DNA test, one marker, 50 % expected cull value) and **b** scenario 2 (one DNA test, one marker, 75 % expected cull value) as routine seedling reductions increase. Stages are indicated by *letters* on each line. Stage A was between obtainment of seeds and seedling germination; stage B was between end of stage A and when the next round of costs were incurred (i.e., transfer to nursery); stage C began after the previous cost and ended after next routine reduction in seedling number; stage D extended from the end of stage C to culling due to early phenotypic selection; stage E followed, ending in field planting
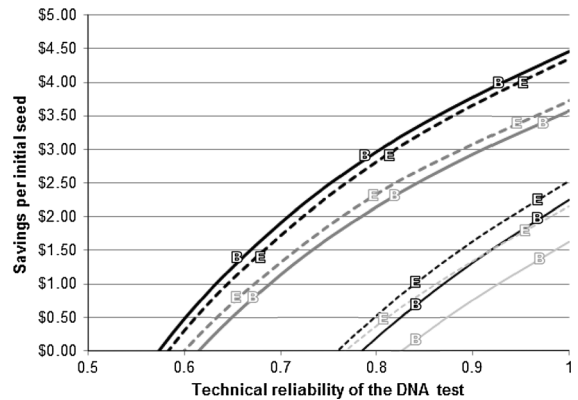

Fig. 3 Influence of the general reliability of DNA testing (e.g., how closely linked the markers are to the locus trait) or its components (e.g., tissue sampling, DNA extraction, genotyping, and provision of results to the breeding program) on the maximum cost efficiency. Stages shown are B (*solid lines*) and E (*dashed lines*). DNA test cull levels were 50 % (*thinner lines*) or 75 % (*thicker lines*) each with an accompanying routine seedling reduction at stage D of 30 % (*black lines*) or 40 % (*gray lines*). Stage B was between seed germination and when the next round of costs were incurred (i.e., transfer to nursery); stage E followed early phenotypic selection and ended at field planting

Discussion

This study indicated that MASS can make internal economic and logistical sense in many typical fruit breeding scenarios. The general feature of cost-efficient scenarios is that the cost of conducting DNA testing operations is lower than the cull proportion from this DNA testing multiplied by the cost of maintaining and evaluating inferior seedlings from the time that DNA testing identifies their genetic inferiority through to when traditional phenotyping would lead to the same decision. Where DNA tests are

available, MASS can be implemented for single or multiple traits at several breeding operational stages to improve cost efficiency in a breeding program. The MASS Efficiency Calculator version 1.0 provides the user with decision support for identifying when and how to integrate DNA testing into traditional breeding operations to maximize the cost efficiency of seedling selection. The timing of stages in which to implement cost-efficient MASS is primarily dependent upon the logistics of each breeding program's traditional operational structure, associated costs, and proportions of routine seedling losses due to non-DNA testing causes.

MASS can be cost efficient for fruit breeding. As observed in cost efficiency calculations of the four scenarios investigated (Table 1), the hypothetical apple and grape breeding programs satisfied the general conditions necessary for cost-efficient MASS as explained by Luby and Shaw (2001). Conditions providing maximal cost efficiency were found to vary by situation, as Luby and Shaw (2001) originally predicted. However, DNA testing costs alone minimally affected at what stage applying DNA tests would maximize cost efficiency (Fig. 4). Rather, the cull level of the DNA tests and routine seedling
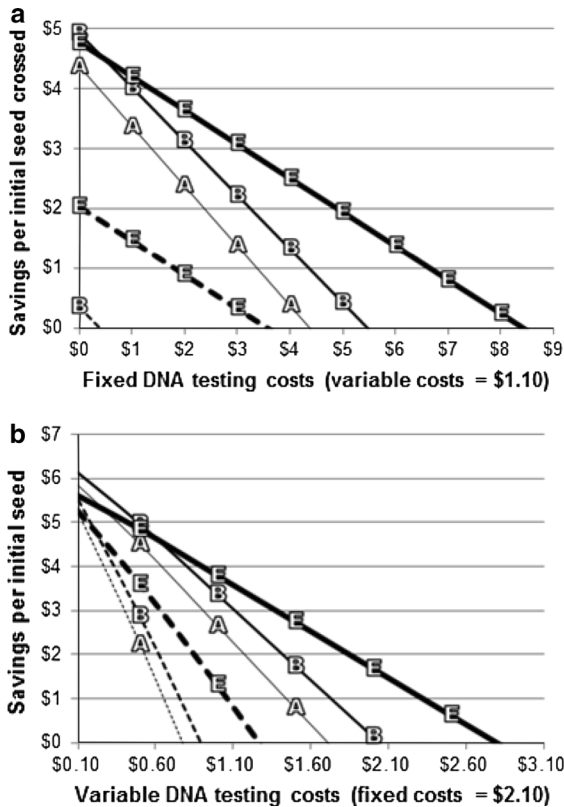
Fig. 4 a Identified maximum cost-efficient MASS opportunity in the apple breeding program for scenario 4 as fixed DNA testing costs (sampling and DNA extraction costs per sample) increased in expense. Variable costs (genotyping, scoring, analysis, and delivery costs per data point) for this scenario remained at $1.10 per data point. b The MASS window with maximum cost efficiency in the apple breeding program for scenario 4 as a function of increased variable DNA testing costs. DNA testing costs were fixed at $2.10 per sample. Scenario 4 = three DNA tests, eight markers, 87.5 % cull. Costs were modeled for DNA tests that were run both together (dashed lines) and sequentially (solid lines). Stages are indicated by letters on each line. Stage A was between obtainment of seeds and seedling germination; stage B was between end of stage A and when the next round of costs were incurred (i.e., transfer to nursery); stage C began after the previous cost and ended after next routine reduction in seedling number; stage D extended from the end of stage C to culling due to early phenotypic selection; stage E followed, ending in field planting; and stage F began after field planting and terminated at the end of seedling selection with the removal of the seedlings. For strawberry: Stage A was the same as for apple and grape; stage B followed to the next round of costs for field establishment; stage C was from field establishment to the next round of seedling reductions; stage D followed previous reductions to culling due to early phenotypic selection; and stage E followed early selection to the next cost in field evaluations
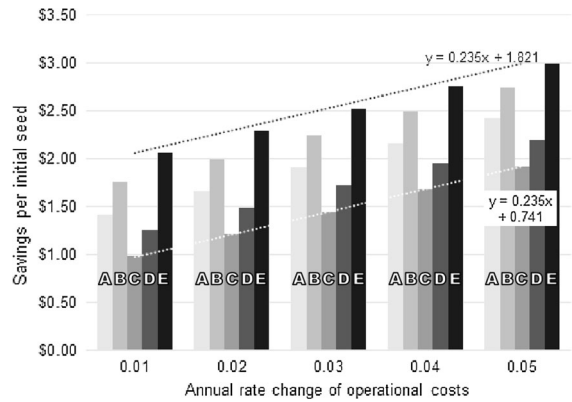

Fig. 5 Savings projected for each MASS stage of scenario 1 as a function of changes in traditional seedling selection costs over time (e.g., inflation and operational procedure changes). Stages are indicated on the bars. Stage A was between obtainment of seeds and seedling germination; stage B was between end of stage A and when the next round of costs were incurred (i.e., transfer to nursery); stage C began after the previous cost and ended after next routine reduction in seedling number; stage D extended from the end of stage C to culling due to early phenotypic selection; stage E followed, ending in field planting; and stage F began after field planting and terminated at the end of seedling selection with the removal of the seedlings. For strawberry: stage A was the same as for apple and grape; stage B followed to the next round of costs for field establishment; stage C was from field establishment to the next round of seedling reductions; stage D followed previous reductions to culling due to early phenotypic selection; stage E followed early selection to the next cost in field evaluations

reductions chiefly governed at what stage to conduct DNA testing for maximum cost efficiency (Table 1, Fig. 2). The strawberry program had such inexpensive traditional seedling selection operations that MASS would provide no better internal cost efficiency (Table 1). When considering traits that can be phenotyped early, such as the disease screen in scenario 5, internal cost efficiency provided by MASS was not improved over traditional seedling selection in apple breeding either. This outcome was because the costs of apple traditional seedling selection for early-phenotyped traits were similar (only $0.03 more) to the traditional seedling selection for strawberry. Also, the apple and grape breeding programs both might have shown similar cost inefficiency over traditional seedling selection if the DNA technology used was relatively higher in cost. Example technologies with such high per-sample DNA testing costs include whole

genome selection technology and SNP array technologies that simultaneously assay hundreds to thousands of data points. However, in real-world situations, the costs of DNA testing and phenotype-based selection can change over time (Luby and Shaw 2001). As fixed or variable DNA testing costs become cheaper, or the costs of other breeding program operations become higher, additional opportunities for MASS arise, allowing for increased flexibility for MASS application (Luby and Shaw 2001; Dreher et al. 2003; Ru et al. 2015). This study indicated that breeding programs with characteristics similar to those for the hypothetical apple and grape programs may find substantial savings through MASS, because their long juvenile periods means that raising inferior seedlings is costly for traditional breeding operations. Further, in similar breeding scenarios, the use of additional DNA tests would improve resource efficiency even more over traditional selection by culling additional seedlings early in breeding operations.

To the extent that components of breeding program operations and DNA testing can be accurately quantified, MASS Efficiency Calculator version 1.0 can be used to define the range of conditions under which a breeding program could expect MASS to provide internal cost efficiency. The Calculator, as demonstrated through four general scenarios and sensitivity analyses of several important parameters, can explicitly define such conditions in terms of (1) specific stages of seedling selection, (2) characteristics and costs of DNA tests, (3) time considerations, (4) together versus sequential DNA testing, and (5) reliability of DNA tests. The MASS Efficiency Calculator version 1.0, therefore, improves upon the break-even costs provided by Luby and Shaw (2001) for the MASS schemes investigated by allowing users to dynamically model varying costs associated with MASS, both in traditional breeding operations and DNA testing operations. Additional capabilities of the Calculator are to (1) readily examine cost structures for single and multiple DNA tests, (2) identify alternative MASS strategies if not enough time is available, and (3) modify input parameters (e.g., DNA testing costs, sample throughput, and technological reliability) to evaluate impacts on cost efficiency or available time. Users can readily compare existing opportunities and identify new cost-efficient strategies that effectively integrate DNA testing into traditional breeding operations. This study reinforced the utility

of decision support tools, as previously suggested by Dreher et al. (2003) and demonstrated by Valente et al. (2013). Developing tools such as the MASS Efficiency Calculator version 1.0 eliminates laborious hand calculations or the need for de novo spreadsheet creation for individual breeding scenarios.

Conducting multiple DNA tests sequentially rather than together is a potentially complicated operation, the value of which was quantified in this study. While sequential use was observed to always be more cost efficient than applying them together (Fig. 4), sample handling for sequential DNA testing may involve additional costs not modeled in the Calculator. Sample consolidation may also depend on the logistics of using a particular genotyping technology and carry concerns about data tracking errors (Dreher et al. 2003; Xu and Crouch 2008; Ru et al. 2015). In situations where sample handling costs are not negligible, especially if additional quality control measures are introduced to reduce data error concerns, costs could be modeled and incorporated into the Calculator.

Available time to perform DNA testing plays a major role in the cost efficiency of MASS. In each of the scenarios presented, enough time was available for the stages that were most cost efficient. Some stages, however, did not allow for sufficient time for testing to be completed by one technician (e.g., stage B of scenario 4 in grape). The availability of enough time to conduct DNA testing, and its effect on cost efficiency, is directly related to the throughput levels achieved by personnel and the DNA extraction and genotyping platforms used (Dreher et al. 2003; Xu and Crouch 2008). Regarding personnel, some of the flagged time constraints in DNA testing from the Calculator could be met by staggering the workload among multiple technicians. Another strategy for MASS application under time constraints is to spread the DNA testing across multiple cost-efficient stages, testing as many seedlings as time allows in the most cost-efficient stage, followed by the next most cost-efficient stage. Both strategies can be modeled using MASS Efficiency Calculator version 1.0. Regarding DNA testing platforms, low-throughput single-tube DNA extraction methods and/or those requiring tissue freeze-drying and relatively low-cost genotyping platforms such as agarose and polyacrylamide gel electrophoresis are suitable for small sample numbers but are not logistically feasible for tens of thousands of seedlings with a short several-week window for DNA testing

(Dreher et al. 2003; Collard and Mackill 2008; Xu and Crouch 2008). In such situations of many samples and little time, which is typical of fruit breeding programs where MASS is to have real impact, breeders require access to high-throughput platforms such as the silica bead method for DNA extraction (Edge-Garza et al. 2014) and cost-efficient genotyping platforms such as ABI DNA Analyzers (Applied Biosystems/Life Technologies, Grand Island, NY). Therefore, DNA tests will have practical application in fruit breeding only if they are developed for time-efficient platforms.

Results from each of the scenarios presented in this study indicated that if expensive seedling selection operation costs exist in traditional breeding, then less-expensive DNA testing could be conducted at any time in the breeding process up to when seedling-handling operation costs are incurred. Therefore, MASS does not require that DNA tests be applied as early as possible in seedling selection for crops with long juvenile periods, as previously suggested by Luby and Shaw (2001). This finding further suggests the need for decision support, such as that provided by MASS Efficiency Calculator version 1.0, to identify optimal MASS schemes.

Improving components of reliability for any DNA test available for a breeding program will help maximize cost efficiency. Increases in reliability of DNA tests increased the proportion of seedlings identified with genetic superiority, and thus, the savings projected (Fig. 3). Luby and Shaw (2001) and Dreher et al. (2003) also showed increased potential savings through increased marker and technical reliability. However, if a DNA test is projected to identify a high proportion (e.g., 75 %) of seedlings to cull, its use will likely improve cost efficiency even at low reliability (e.g., 60 %), as indicated from the sensitivity analyses (Fig. 3) and demonstrated in application in the Washington apple breeding program (Edge-Garza et al. 2010). Conversely, sensitivity analyses (Fig. 3) indicated that increasing the reliability of a DNA test from, for example, 90–95 % would be unlikely to significantly increase savings over the long term. The Calculator allows ready determination of whether the use of a DNA test, given its reliability, is a cost-efficient selection tool. The added value of improved reliability can also be readily calculated. Users can thereby maximize cost efficiency with available DNA tests and platforms while, if calculated to be worthwhile, simultaneously direct

efforts to increase marker reliability (e.g., by reducing the linkage distance between markers and trait-including loci) and/or improve technical reliability of the DNA testing platforms.

MASS Efficiency Calculator version 1.0 can also provide indirect support for decisions. For example, while the Calculator identified a switch to a later stage to use DNA tests if routine seedling losses were to approach from 30 to 40 % (Fig. 2b), in reality there may be less interest in having DNA tests applied later in the breeding cycle and more focus on better crossing decisions to keep the routine reductions in seedlings below 40 %. The extra 1 % that could be gained by switching to stage B in scenario 2 of the apple example (Table 1) could be sacrificed in favor of standardizing the application of DNA tests in one stage. Ru et al. (2015) described that such trade-offs to standardize sample collection and application of DNA tests were explored for MASS application in the Washington apple breeding program. The Calculator can also be used indirectly to estimate the number of additional seeds that could be sown by using the estimated savings from DNA testing for that purpose. For example, as observed in scenario 7, MASS could assist the apple program to accommodate 32–75 % more seeds.

In this study, use of the MASS Efficiency Calculator, like the analyses of Luby and Shaw (2001), was limited to considering only the internal cost efficiency of using DNA tests for seedling selection in a breeding program. The Calculator did not directly attempt to estimate or internalize potential external costs or benefits of employing MASS. External costs not considered include development of genomics technologies and DNA tests, trialing of DNA extraction and genotyping platforms, and confirmation of trait predictive ability of DNA tests for a breeding program's germplasm. Some of these costs might be substantially defrayed by public sector investment (Dreher et al. 2003), while others might be borne largely by the breeding program. In any case, such prior research costs are independent of cost efficiency achievable with available DNA tests for which a breeding program has confidence in the extent of information provided on genetic potential of seedlings. External benefits not included were potential value to producers, distributors, and consumers of cultivars that otherwise might not have been produced or produced more slowly without the use of DNA tests. Nevertheless, the Calculator can be used to indirectly estimate some external cost/benefit situations, as

demonstrated in scenario 6 for disease allele pyramiding that is often described as a valuable use of DNA information (e.g., Bus et al. Bus et al. 2009; Ru et al. 2015). In that scenario, MASS would be cost inefficient if an equivalent and cheap phenotyping protocol existed. However, such protocols do not exist in situations where more than one resistance allele provides the same phenotype (Evans and James 2003; Collard et al. 2005; Tester and Langridge 2010). Where phenotyping protocols for the resistance allele of each locus are available, the typical laborious progeny testing required (Tartarini et al. 1999; Bliss 2010) would likely cost more than the break-even point for phenotyping found in this scenario, emphasizing the value of DNA testing.

Many seedling selection scenarios among fruit breeding programs are similar to those of the hypothetical apple and grape breeding programs in this study. The MASS Efficiency Calculator version 1.0 can provide decision support to effectively compare cost efficiency among alternative seedling selection scenarios and therefore enhance breeding resource-use efficiency. Future improvements in MASS decision support will focus on the addition of new Calculator functionalities. Meanwhile, the current version can be used to explore cost-efficient scenarios for seedling selection in any breeding program with multiple possible stages of seedling culling and costs.

References

Badenes ML, Byrne DH (eds) (2012) Fruit breeding. Springer, New York

Bliss FA (2010) Marker-assisted breeding in horticultural crops. Acta Hortic 859:339–350

Bus VGM, Esmenjaud D, Buck E, Laurens F (2009) Application of genetic markers in rosaceous crops. In: Folta KM, Gardiner SE (eds) Genetics and genomics of Rosaceae. Springer, New York, pp 563–599

Byrne DH (2007) Molecular marker use in perennial plant breeding. Acta Hortic 751:163–167

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B 363:557–572

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196

Dirlewanger E, Graziano E, Joobeur T, Garriga-Calderé F, Cosson P, Howad W, Arús P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci 101:9891–9896

Dreher K, Khairallah M, Ribaut J-M, Morris M (2003) Money matters (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. Mol Breed 11:221–234

Edge-Garza D, Zhu Y, Peace C (2010) Enabling marker-assisted seedling selection in the Washington Apple Breeding Program. Acta Hortic 859:369–373

Edge-Garza DA, Rowland TV Jr, Haendiges S, Peace C (2014) A high-throughput and cost-efficient DNA extraction protocol for the tree fruit crops of apple, sweet cherry, and peach relying on silica beads during tissue sampling. Mol Breed 34:2225–2228

Evans K (2013) Apple breeding in the Pacific Northwest. Acta Hortic 976:75–78

Evans KM, James CM (2003) Identification of SCAR markers linked to Pl-w mildew resistance in apple. Theor Appl Genet 106:1178–1183

Gallardo RK, Nguyen D, McCracken V, Yue C, Luby J, McFerson JR (2012) An investigation of trait prioritization in rosaceous fruit breeding programs. HortScience 47: 771–776

Galletta GJ (1975) Blueberries and cranberries. In: Janick J, Moore JN (eds) Advances in fruit breeding. Purdue University Press, West Lafayette, pp 154–196

Groenendaal H, Galligan DT, Mulder HA (2004) An economic spreadsheet model to determine optimal breeding and replacement decisions for dairy cattle. J Dairy Sci 87: 2146–2157

Hospital F (2009) Challenges for affective marker-assisted selection in plants. Genetica 136:303–310

Janick J, Moore JN (eds) (1975) Advances in fruit breeding. Purdue University Press, West Lafayette

Karaagac E, Vargas AM, de Andrés MT, Carreño I, Ibáñez J, Carreño J, Martínez-Zapater JM, Cabezas JA (2012) Marker assisted selection for seedlessness in table grape breeding. Tree Genet Genomes 8:1003–1015

Luby JJ, Shaw DV (2001) Does marker-assisted selection make dollars and sense in a fruit breeding program? HortScience 36:872–879

Moreau L, Lemarie S, Charcosset A, Gallais A (2000) Economic efficiency of one cycle of marker-assisted selection. Crop Sci 40:329–337

Morris M, Dreher K, Ribaut J-M, Khairallah M (2003) Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. Mol Breed 11:235–247

Peace CP, Norelli J (2009) Genomics approaches to crop improvement in the Rosaceae. In: Folta K, Gardiner S (eds)

Genetics and genomics of Rosaceae. Springer, New York, pp 19–53

Ru S, Main D, Evans K, Peace C (2015) Current applications, challenges, and perspectives of marker-assisted seedling selection in Rosaceae tree fruit breeding. Tree Genet Genomes 11:8. doi:10.1007/s11295-015-0834-5

Slater AT, Cogan NOI, Forester JW (2013) Cost analysis of the application of marker-assisted selection in potato breeding. Mol Breed 32:299–310

Tartarini S, Gianfranceschi L, Sansavini S, Gessler C (1999) Development of reliable PCR markers for the selection of the *Vf* gene conferring scab resistance in apple. Plant Breed 118:183–186

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327:818–822

Valente F, Gauthier F, Bardol N, Blanc G, Joets J, Charcosset A, Moreau L (2013) OptiMAS: a decision support tool for marker-assisted assembly of diverse alleles. J Hered 104:586–590

Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Sci 48: 391–407

Xu Y, Yanli L, Chuanxiao X (2012) Whole-genome strategies for marker-assisted plant breeding. Mol Breed 29:833–854

Yue C, Gallardo RK, McCracken V, Luby J, McFerson J, Liu L, Iezzoni A (2012) Technical and socio-economic challenges to setting and implementing priorities in North American rosaceous fruit breeding programs. HortScience 47:1320–1327

Zhebentyayeva T, Ledbetter C, Burgos L, Lácer G (2012) Apricot. In: Badenes ML, Byrne DH (eds) Fruit breeding. Springer, New York, pp 415–458