Published March 15, 2018

RESEARCH

DeltaGen: A Comprehensive Decision Support Tool for Plant Breeders

M. Z. Z. Jahufer* and Dongwen Luo

ABSTRACT

In this paper, we introduce a unique new plant breeding decision support software tool DeltaGen, implemented in R and its package Shiny. DeltaGen provides plant breeders with a single integrated solution for experimental design generation, data quality control, statistical and quantitative genetic analyses, breeding strategy evaluation, simulation, and cost analysis, pattern analysis, index selection, and underlying basic theory on quantitative genetics. Key analysis procedures in DeltaGen were demonstrated using three datasets generated from forage breeding trials in Australia, New Zealand, and the United States. Analyses of the perennial ryegrass seasonal growth data in Case Study 1 was based on residual maximum likelihood analysis and pattern analysis. A graphical summary of the performance of entries across locations was generated, and entries with specific and broad adaptation were identified. The quantitative genetic analysis and breeding method simulation procedures applied to the perennial ryegrass half-sib (HS) family data in Case Study 2 enabled estimation of quantitative genetic parameters, prediction of genetic gain, and calculation of costs per selection cycle. These results enabled comparison of three breeding methods, which also included genomic selection, and their simulation. Data from Case Study 3 were analyzed to investigate a multivariate approach to identify HS families of switchgrass with breeding values that would enable an increase in biomass dry matter yield (DMY) and cell wall ethanol (CWE) and a decrease in Klason lignin (KL). The Smith–Hazel index developed enabled identification of HS families with genetic worth for increasing DMY and CWE and reducing KL, in contrast with individual trait selection. Analysis of the datasets in all three case studies provides a snapshot of the key analyses available within DeltaGen. This software tool could also be used as a teaching resource in plant breeding courses. DeltaGen is available as freeware at http://agrubuntu.cloudapp.net/PlantBreedingTool/

M.Z.Z. Jahufer and D. Luo, AgResearch, Grasslands Research Centre, Private Bag 11008, Palmerston North, New Zealand. Received 28 July 2017. Accepted 26 Jan. 2018. *Corresponding author (zulfi.jahufer@agresearch.co.nz). Assigned to Associate Editor Jeffrey Endelman.

Abbreviations: A_pWF_{gsy-HS}, among and within half-sib family; BLUP, best linear unbiased predictor; CWE, cell wall ethanol; DMY, dry matter yield; FS, full-sib; ΔG, genetic gain; GEBV, genomic estimated breeding value; GS, genomic selection; HS, half-sib; HSPT, half-sib family with progeny testing; KL, Klason lignin; MANOVA, multitrait analysis of variance; REML, residual maximum likelihood; SH, Smith–Hazel; TAFE, Technical and Further Education.

Plant breeding is an integrative discipline that requires trial design, data analysis, and decisions to be optimized for efficiency and effectiveness. Despite the widespread use of computation in plant breeding, there are few convenient tools that bring together the necessary functions of experimental design, data analysis, and breeding method optimization in terms of genetic gain (ΔG) and resource factors in a way that is readily accessible to the breeder through a single interface.

Choice of an efficient breeding strategy in terms of achievable ΔG per cycle of selection is important. In designing the optimal structure of a breeding program with the resources available, there are important decisions to be made at the onset of cultivar development and revised during the breeding process. These decisions will depend on key information such as the genetic structure of breeding populations under selection (Moll and Stuber, 1974; Milligan et al., 1990). Estimates of genetic parameters, such as additive and nonadditive components of variance and narrow-sense heritability, enable prediction of expected ΔG using the range of conventional breeding methods available (Dudley and Moll, 1969; Falconer, 1989). The basic model used

Published in Crop Sci. 58:1118–1131 (2018).
doi: 10.2135/cropsci2017.07.0456

in the prediction of ΔG or response to selection is $R = ih^2\sigma_p$, where, i is the intensity of selection, h^2 is the appropriate estimate of heritability of the phenotypic values of individuals or families and σ_p is the phenotypic standard deviation among individuals or families (Falconer, 1989). Based on the type of selection unit—single plants, parental clones, or half-sib (HS) or full-sib (FS) among- and within-family progeny—several variations to this model are available (Hallauer and Miranda, 1988; Casler and Brummer, 2008).

Genotype × environment interactions are a fundamental problem when breeding for broad adaptation (Eisemann et al., 1990) and must be considered when planning breeding programs. Thus, the availability of estimates of the size and nature of genotype × environment interactions will influence the design of optimal multienvironment testing strategies (Nyquist, 1991; Cooper et al., 1993). Cooper et al. (1993) proposed a descriptive framework to study the influence of genotype × environment interactions on the response to selection. Three basic components associated with their approach were: (i) defining the population of target environments, (ii) analysis of the variation among genotypes, and (iii) prediction of the response to selection in the target population of environments.

In forage cultivar development programs, breeding material (e.g., HS or FS families and experimental varieties) is generally evaluated across multiple years, seasons, and locations. These trials generate large amounts of data that are often underused in terms of determining quantitative genetic information of value to the enhancement of selection and breeding strategy implementation. One of the major limitations to generating this information and its application in applied breeding programs is the lack of availability of user-friendly software to field breeders. Although there are robust multifaceted statistical software such as GenStat (Payne et al., 2009) and SAS (SAS Institute, 2011) available, these programs do not provide direct estimates of information vital to the enhancement of applied breeding strategies (e.g., rate of ΔG and associated cost per selection cycle, and summaries of genotype performance across target environments). There is often a lack of familiarity with the application of basic quantitative genetic theory to analyze field trial data, which complicates estimation of genetic parameters using analysis outputs from software programs such GenStat and SAS. Today, the integration of novel selection methods such as genomic selection (GS) (Meuwissen et al., 2001) into forage cultivar development programs further emphasizes the importance of quantitative genetic information. The application of this information to assess the relative efficiency of new selection techniques in comparison with existing breeding practices, in terms of rate of ΔG and associated costs per unit of gain, is crucial to planning breeding programs. The ability of breeders to make strategic decisions will be improved by the availability of decision support software that will provide a seamless progression from experimental design to analysis of field data (i) to generate best linear unbiased predictors (BLUPs) (White and Hodge, 1989; Galwey, 2006), (ii) to summarize large multilocation breeding trial data, (iii) to estimate additive genetic variation and associated interaction effects, and (iv) to compare the relative efficiencies among breeding strategies. Information significant to the design and implementation of more efficient cultivar development programs could be generated by software simulation of breeding strategies based on selection among and within genetic families, and different combinations of year, season, site, replicate, and sample numbers with associated costs per selection cycle.

One of the key objectives of any breeder is to increase the rate of ΔG. The development of plant breeding software capable of providing decision support (i) to improve the efficiency of experimental trial design and data analysis of phenotypic data from field trials, and (ii) to enable assessment of breeding strategies and implementing new selection techniques such as GS in terms of rate of ΔG per cycle of selection and associated cost, will offer value to breeders. This paper describes a new and unique decision support tool implemented in R (R Core Team, 2016), called DeltaGen. The objective of developing DeltaGen was to provide breeders with a user-friendly quantitative genetic platform that will enable a seamless progression through different stages and methods of field data analysis. This includes data summary and quality assessment, linear mixed model analysis, generation of BLUPs, best linear unbiased estimates, and variance-covariance components, estimation of quantitative genetic parameters, simulation of rate of ΔG, and calculation of associated cost per cycle of selection. DeltaGen also provides multivariate analysis methods that include pattern analysis of genotypic performance across environments, and index selection based on the Smith–Hazel (SH) model (Smith, 1936; Hazel, 1943). In making the tool of value to teachers and learners, an important feature of DeltaGen is that basic theoretical information at each stage of quantitative genetic analysis is also provided in the associated "Help" screen.

While most of the components in the tool we report on the development of in this paper are of value for plant breeding focused on any sexually reproducing species and trait or combination of traits, we will focus on examples in forage breeding and associated breeding strategies. Forage breeding is the domain of what breeders consider to be minor crops, and have often lagged behind in terms of software tools available to breeders, and realized levels of genetic improvement.

In this paper, key components of DeltaGen are demonstrated by analysis of three different sets of data generated

from forage breeding trials in Australia, New Zealand, and the United States. The data are analyzed using DeltaGen to show: (i) variance-covariance component analysis, (ii) prediction of ΔG using different breeding methods, including GS and its simulation, together with associated costs (NZ$) per cycle of selection, (iii) pattern analysis, (iv) multivariate analysis, and (v) SH index (Smith, 1936; Hazel, 1943) calculation and associated predicted ΔG. These analyses will be associated with three case studies: (i) an analysis of multilocation trials of perennial ryegrass (*Lolium perenne* L.) breeding lines for seasonal herbage growth across three key grazing environments in New Zealand, (ii) a quantitative genetic analysis of perennial ryegrass HS families, evaluated at one location in Australia to estimate potential ΔG in seasonal herbage growth using different breeding methods and their associated cost estimation, (iii) multivariate analysis of three traits—biomass dry matter yield (DMY), cell wall ethanol (CWE), and Klason lignin (KL)—in switchgrass (*Panicum virgatum* L.) based on HS family evaluation at two locations in the United States.

To validate the accuracy of key estimates such as variance components and their associated standard errors, the estimates from DeltaGen were compared with outputs from GenStat 7.1 (VSN International, 2003).

MATERIALS AND METHODS

Software

DeltaGen is a web-based application developed using the computer language R (R Core Team, 2016) and its package Shiny (Chang et al., 2017). Shiny provides the web framework for all the applications in DeltaGen. Using the R–Shiny combination has provided the opportunity for developing real-time analytical solutions and decision support tools for plant breeding in a single analytical package. Using this software platform enables advanced analytics, complex simulations, routine calculations, and interactive visualization to be conducted in DeltaGen.

The framework of DeltaGen is based on a step-by-step approach to data analysis (Fig. 1). This will allow users to follow an intuitive and logical process from basic analysis of field data to more in-depth quantitative genetic parameter estimation and simulation of breeding strategies. Once data are uploaded into DeltaGen, data quality checks (e.g., data distribution plots, pivot tables, and heat maps of field data, which provide graphical summaries of information) can be conducted and then progress to linear mixed model analysis, multitrait analysis of variance (MANOVA), and pattern analysis (a combination of cluster and principal component analyses), as required. As indicated in Fig. 1, both the linear mixed model and MANOVA analyses components are linked to the quantitative genetic simulation and the selection index components of DeltaGen. For analyses that require a graphical summary of genotype performance across environments, the genotype × environment two way BLUP data matrix (genotypes, entry numbers, and names shown in the first column followed by columns 2, 3, 4…..etc., consisting of environment-specific BLUP values for each entry) generated

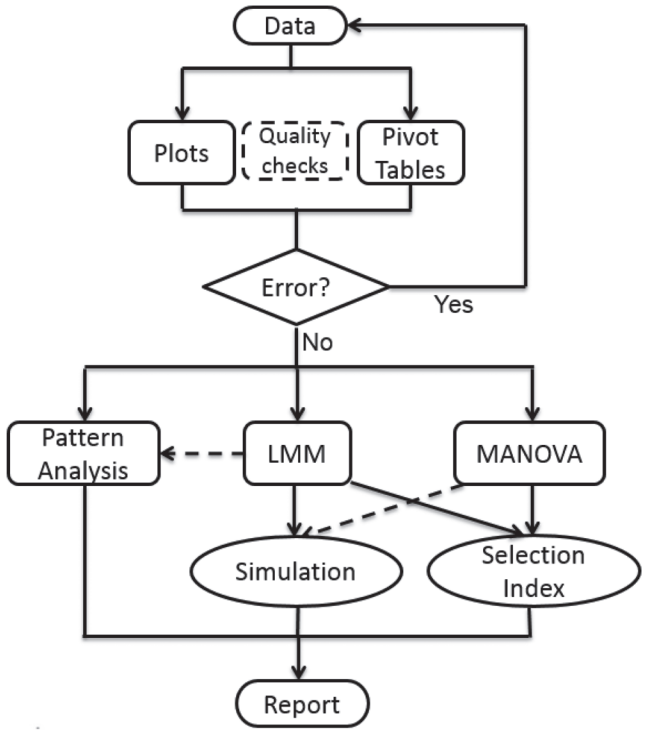

Fig. 1. Framework of the integrated data analysis components in DeltaGen. LMM, linear mixed models; MANOVA, multivariate analysis of variance.

from the linear mixed model analysis can be uploaded directly into the associated pattern analysis component. The link is indicated in Fig. 1. The links between different analysis components within DeltaGen provide a complete solution to carrying out basic breeding trial data analysis and further quantitative genetic and breeding method interrogation, as required.

DeltaGen has two "Help" options: the "Help" option on the main tool bar provides access to the "Quick start manual" (users guide), and the "Help" option in each of the data analysis windows provides information on the quantitative genetic models used and key associated references.

In this paper, the key data analytical components in DeltaGen are demonstrated using three field trial datasets, referred to as case studies.

Datasets

Case Study 1: Identification of Perennial Ryegrass Breeding Lines with Broad and Specific Adaptation

Data were generated from perennial ryegrass field trials planted to evaluate breeding lines at Palmerston North (40.20237° S, 175.28220° E), Ruakura (37.78° S, 175.32° E), and Kerikeri (35.22° S, 173.94° E), New Zealand, in late winter 2013. At each location, 107 entries (83 breeding lines and 24 commercial check cultivars) were planted in 1-m rows separated by 30-cm spacing, according to a randomized complete block design consisting of three replicates. The multilocation trial was conducted for a period of 3 yr under rotational sheep grazing, during which seasonal (spring, summer, autumn, and winter) growth was visually scored using a qualitative scale of 0 (no observed growth) to 9 (very high growth). The data analyzed are saved in the file "CaseStudy 1" as one of the example datasets provided in DeltaGen.

Case Study 2: Genetic Analysis of 90 Half-Sib Families Generated from a Breeding Population of Perennial Ryegrass

The field trial was performed at the experimental farm of the South West Technical and Further Education (TAFE) campus at Glenormiston, VIC (38.860° S, 142.5860° E, 114 m asl) in Australia. The trial was planted in spring 2007. The 90 HS families of perennial ryegrass containing AR1 endophyte (Easton and Fletcher, 2007) were hand sown into 1-m rows. The experimental layout was a row column design (John, 1987) with three replicates. The rows were 35 cm apart. The trial was conducted for a period of 3 yr. Seasonal growth during spring, summer, autumn, and winter was scored on a qualitative scale of 1 (poor) to 5 (high). After every growth score, the rows were uniformly defoliated using a lawn mower to a height of 4 cm above the soil surface to simulate grazing. The harvested foliage was removed from the trial area. The data analyzed are saved in the file "CaseStudy 2" as one of the example datasets provided in DeltaGen.

Case Study 3: Multivariate Analysis of Half-Sib Families of Switchgrass.

The data matrix used for multivariate analyses was generated from switchgrass HS family evaluation field trials conducted at two locations: Arlington (43.30° N, 89.35° W) and Marshfield (44.64° N, 90.13° W), WI, USA. A total of 147 HS families were sown in spring 2008 using a five-row drill planter that placed the seeds at a depth of 5 to 10 mm in the soil. The sward plot size was 0.9 × 1.1 m, with 0.9 m between rows and the adjoining columns. The data matrix used in the analysis presented is based on the traits biomass DMY (Mg ha^{-1}), CWE (mg g^{-1}), and KL (mg g^{-1}) measured in Years 1 and 2 of the trial. Detailed descriptions of the two locations and trait measurements are documented in Jahufer and Casler (2015), where a similar trial based on a different set of switchgrass HS families is reported. The data analyzed are saved in the file "CaseStudy 3" as one of the example datasets provided in DeltaGen.

Variance Component Analysis

Case Study 1

Variance component analyses of seasonal growth were conducted using the residual maximum likelihood (REML) (Patterson and Thompson, 1971, 1975; Harville, 1977) procedure in DeltaGen and GenStat 7.1. (VSN International, 2003). Linear mixed models were used for analysis of the data within individual locations and across all locations.

The linear mixed model used for analysis within individual locations across seasons and years was

$$Y_{ijkl} = M + g_i + s_j + (\mathrm{gs})_{ij} + \gamma_k + (\mathrm{gy})_{ik} + (\mathrm{gsy})_{ijk} + (\mathrm{sy})_{jk} + b_{jkl} + (\mathrm{gb})_{il} + \varepsilon_{ikjl} \quad [1]$$

where Y_{ijkl} is the value of an attribute measured from entry i in replicate l in season j of year k and $i = 1, ..., n_g$, $j = 1, ..., n_s$, $k = 1, ..., n_y$, and $l = 1, ..., n_b$, where g, s, γ, and b are entries, seasons, years, and replicates, respectively; M is the overall mean; g_i is the random effect of entry i, $N(0, \sigma^2_g)$; s_j is the fixed effect of season j; γ_k is the fixed effect of year k; b_{jkl} is the random effect of replicate l within season j, within year k, $N(0, \sigma^2_b)$; $(\mathrm{gs})_{ij}$ is the random effect of the interaction between entry i and season j, $N(0, \sigma^2_{gs})$; $(\mathrm{gy})_{ik}$ is the random effect of the interaction between entry line i and year k, $N(0, \sigma^2_{gy})$; $(\mathrm{gsy})_{ijk}$ is the random effect of the interaction between entry i, season j, and year k, $N(0, \sigma^2_{gsy})$; $(\mathrm{sy})_{jk}$ is the interaction between the fixed effects season j and year k; $(\mathrm{gb})_{il}$ is the effect of the interaction between entry line i and replicate l, $N(0, \sigma^2_{gb})$; and ε_{ijkl} is the residual effect for entry i in replicate l in season j, during year k, $N(0, \sigma^2_\varepsilon)$.

The linear mixed model used for analysis across locations, seasons, and years was

$$Y_{ijklm} = M + g_i + l_j + (\mathrm{gl})_{ij} + s_k + (\mathrm{gs})_{ik} + \gamma_l + (\mathrm{gy})_{il} + (\mathrm{sy})_{kl} + (\mathrm{gly})_{ijl} + (\mathrm{gsy})_{ikl} + b_{jklm} + (\mathrm{gb})_{im} + \varepsilon_{ijklm} \quad [2]$$

where Y_{ijklm} is the value of an attribute measured from entry i in replicate m at location j in season k of year l and $i = 1, ..., n_g$, $j = 1, ..., n_l$, $k = 1, ..., n_s$, $l = 1, ..., n_y$, and $m = 1, ..., n_b$, where g, s, γ, and b, are as described in Eq. [1] and l are locations. For a detailed definition of the model effects as in Eq. [2], please refer to Supplemental File 1.

An estimate of entry mean broad-sense heritability (h^2_b), (Falconer, 1989) across years, seasons, and locations was calculated by selecting the heritability option provided in DeltaGen. The broad-sense heritability (h^2_b) calculation was based on the equation

$$h^2_b = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_{gy}}{n_y} + \frac{\sigma^2_{gs}}{n_s} + \frac{\sigma^2_{gl}}{n_l} + \frac{\sigma^2_{gb}}{n_r} + \frac{\sigma^2_{gsy}}{n_s n_y} + \frac{\sigma^2_{gly}}{n_l n_y} + \frac{\sigma^2_\varepsilon}{n_y n_s n_l n_r}} \quad [3]$$

where n_y, n_s, n_l, and n_r are the number of years, seasons, locations, and replicates, respectively, and σ^2_g is the estimated genotypic variation among the 107 entries of perennial ryegrass.

Case Study 2

A linear mixed model analysis using the REML procedure was conducted to estimate additive genetic variance among the 90 HS families and associated interaction components. The linear mixed model used was

$$Y_{ijklmn} = M + f_i + \gamma_j + (\mathrm{fy})_{ij} + s_{jk} + (\mathrm{fs})_{ik} + b_{jkl} + r_{jklm} + c_{jkln} + \varepsilon_{ijklmn} \quad [4]$$

where Y_{ijklmn} is the value of an attribute measured from HS family i in row m and column n of replicate l nested in season k in year j and $i = 1, ..., n_f$, $j = 1, ..., n_y$, $k = 1, ..., n_s$, $l = 1, ..., n_b$, $m = 1, ..., n_r$, and $n = 1, ..., n_c$, where f, γ, s, b, r, and c are HS families, years, seasons, replicates, rows, and columns, respectively. For a detailed definition of the model effects as in Eq. [4], please refer to Supplementary File 1.

Analysis of variation among the 90 HS families (σ^2_f) provided an estimate of 1/4 additive genetic variation (σ^2_A) (Falconer, 1989). Selecting the heritability option provided in DeltaGen generated an estimate of narrow-sense

heritability (h^2_n) on a family mean basis across seasons and years (Nyquist, 1991):

$$h^2_n = \frac{\sigma^2_f}{\sigma^2_f + \frac{\sigma^2_{fs}}{n_s} + \frac{\sigma^2_{fy}}{n_y} + \frac{\sigma^2_\varepsilon}{n_s n_y n_r}}$$

where n_s, n_y, and n_r are the number of seasons, years and replicates, respectively.

Significance Testing of Genotypic and Family Variance Components

The REML analysis output in DeltaGen included estimates of ± standard error for all variance components. In addition, 95% confidence interval estimates for variance components are also provided. To further evaluate results generated from DeltaGen, the statistical significance of the variance components was further assessed using the likelihood ratio test (Holland et al., 2002; Galwey, 2006).

Fixed Effects Analysis

Please note that the *F* ratio values calculated for the fixed effects, *y*, *s*, *l*, and *y* × *s* interactions in Case Studies 1 and 2 were based on ratios, with the estimated error mean square as the denominator. This does not account for the hierarchical structure in the data and may result in reduced *p* values. If a complete fixed effects analysis, which also includes entries (*g*) and their interactions with *y*, *s*, *l*, and *r*, is required for data with complex hierarchical structures, as in Case Studies 1 and 2, a split-plot or split-split-plot type analysis is recommended (Steel and Torrie, 1981; Nyquist, 1991).

Predicting Genetic Gain and Simulation

The genetic analysis component within DeltaGen enables prediction of rate of Δ*G*, calculation of cost ($) per cycle of selection, and the opportunity to conduct associated simulation. The simulation option enables manipulation of the number of years, seasons, locations, replicates, and samples for GS accuracy. Selection intensity can also be varied in any simulation.

Once trial data are uploaded and analyzed using the univariate option, if the data have a HS or FS family structure, clicking on the "Simulation" option will open the "Breeding Strategies and Simulation" window. This provides access to ΔG prediction equations for a range of forage breeding strategies proposed by Casler and Brummer (2008): HS (half-sib family), HSPT (HS with progeny testing) and AWF-HS (among and within HS). In addition, the application of strategies based on correlation response to selection is also available: CR_Y-HS (correlated response to selection in primary trait *Y* while selecting for a secondary trait *X*) and CR_YWF_X–HS (correlated response to selection in a primary trait *Y* when within-family selection is based on a secondary trait *X*).

In DeltaGen the correlated response of a trait *Y* resulting from GS is estimated using the equation $\Delta G_Y = kch_X r_A \sigma_{AY}$, where *k* is the selection pressure, *c* is the parental control, h_X is the square root of heritability of trait *X* (in this case, GS), r_A is the GS accuracy, and σ_{AY} is the standard deviation of additive genetic variance for trait *Y* that is under selection. The GS accuracy is the Pearson's correlation coefficient between the true breeding values (BV) and their associated genomic estimated breeding values(GEBV) (Meuwissen et al., 2001; Heslot et al., 2012). This equation is a modification of the equation for Δ*G* resulting from correlated response to selection proposed by Falconer (1989), $CR_Y = kch_X h_Y r_A \sigma_{PY}$. By expanding h_Y, $CR_Y = kch_X r_A(\sigma_{AY}/\sigma_{PY})\sigma_{PY} = kch_X r_A \sigma_{AY}$. If we assume that $h_X = 1$ (Dekkers, 2007a, 2007b), $CR_Y = kcr_A \sigma_{AY}$ (Lorenz, 2013; Desta and Ortiz, 2014; Bassi et al., 2015).

In DeltaGen, this equation is also used to predict Δ*G* using a combination of among-family selection on phenotype and within-family selection using GS, based on a modification of the equation proposed by Casler and Brummer (2008) for among- and within-family selection, where within-family selection is conducted on a secondary trait *X*. Using the approach of a correlated response to GS, in the within-HS-family selection component, the modified Casler and Brummer (2008) equation used in DeltaGen is

$$A_pWF_{gsy-HS} = k_f c_f \frac{1/4\,\sigma^2_{AY}}{\sigma_{PF}} + k_w c_w h_X r_{A-XY} \frac{\sqrt{3}}{2}\sigma_{AY} \quad [6]$$

where A_pWF_{gsy-HS} is the predicted Δ*G* for trait *Y* using a combination of phenotypic among-HS-family selection and within-HS-family GS; σ^2_{AY} is the additive genetic variance for the trait *Y* under selection; σ_{AY} is standard deviation of additive genetic variance for trait *Y*, as the within-HS-family genetic variance is $3/4\sigma^2_A$, $\sqrt{3/4} = \sqrt{3}/2$; σ_{PF} is the among-family phenotypic standard deviation for trait *Y*, k_f and k_w and the among- and within-HS-family selection pressure, respectively; c_f and c_w are the among- and within-HS-family parental controls (for HS family selection, $c = 0.5$, and for FS family selection, $c = 1$; Casler and Brummer, 2008), respectively; h_X is the square root of heritability of trait *X* (GS) and is assumed to be 1; and r_{A-XY} is the GS accuracy. Please note that, in practice, the $h_X = 1$ assumption could result in overestimation of the correlated response in trait *Y*. There is an option in DeltaGen to change h_X if an appropriate estimate is available or for the purpose of simulation. In DeltaGen, all the breeding equations are also available for the analysis of FS family data.

The estimates of additive genetic variance and associated interactions from the genetic analysis of seasonal growth data in Case Study 2, based on HS families, were used to demonstrate the application of DeltaGen to compare predicted Δ*G* among the breeding strategies HS, HSPT, and A_pWF_{gsy-HS}. Simulations based on varying year and or replicate number in the HS and A_pWF_{gsy-HS} breeding strategies were also performed. For simulation using A_pWF_{gsy-HS}, an assumed genomic accuracy (r_{A-XY}) of 0.25 was used, based on the range $r_{A-XY} = 0.010$ to 0.315 for perennial ryegrass growth reported by Grinberg et al. (2016). The associated cost per cycle of selection for each breeding method was calculated. The inputs used for cost calculation are approximate estimates in New Zealand dollars. A guide to calculating the cost of GS (the cost of generating a single GEBV) is presented in the "Quick start manual" (user's guide) under "Help" in DeltaGen.

Case Study 3

The 147 switchgrass HS family × three-trait (DMY, CWE, and KL) data matrix was used to demonstrate multivariate analysis procedures in DeltaGen. The first step towards generation of

this two-way matrix was to conduct a variance component analysis for each trait separately, based on a completely random linear model, using the REML procedure. The linear model was

$$Y_{ijklmn} = M + f_i + y_j + (\mathrm{fy})_{ij} + l_{jk} + (\mathrm{fl})_{ik} + b_{jkl} + r_{jklm} + c_{jkln} + \varepsilon_{ijklmn} \quad [7]$$

where Y_{ijklmn} is the value of an attribute measured from HS family i in row m and column n of replicate l nested in location k in year j and $i = 1, ..., n_f$, $j = 1, ..., n_y$, $k = 1, ..., n_l$, $l = 1, ..., n_b$, $m = 1, ..., n_r$, and $n = 1, ..., n_c$, where f, y, l, b, r, and c are as described in Eq. [4]. For a detailed definition of the model effects as in Eq. [7], please refer to Supplemental File 1.

The estimated variance components for the traits were used to calculate h^2_n on a family mean basis across locations and years (Nyquist, 1991). Predicted ΔG was also calculated based on the HS breeding strategy.

Pattern Analysis

DeltaGen provides an option for graphically summarizing large entry × trait and entry× location two-way data matrices using a combination of cluster analysis and principal component analysis—a method of analysis termed "pattern analysis" (Cooper and Hammer, 1996). Cluster analysis of the entry × multiple trait or location matrix, produced from the analysis of variance, is used to generate entry groups. This is followed by principal component analysis (ordination) of the same entry× multiple trait or location matrix to generate a biplot. Each of the entry groups identified from clustering are assigned a different color and superimposed on the biplot. The result is a graphical summary of information within the entry × multiple trait or location matrix.

In DeltaGen, pattern analysis can be conducted within the univariate models option to summarize trials across multiple years, seasons, and location, and also directly under the "Pattern Analysis" option provided on the main control bar. The "Pattern Analysis" option is for the analysis of two-way entry × multiple trait data matrices.

In DeltaGen, cluster analysis is performed using a hierarchical agglomerative classification procedure with squared Euclidean distance as a measure of dissimilarity (Burr, 1968, 1970; Wishart, 1969), and the Hartigan clustering algorithm (Hartigan, 1975) is used as the grouping strategy. Principal component analysis is conducted according to Jolliffe (2002).

Before conducting cluster analysis, the data are standardized to remove scaling effects (Cooper and DeLacy, 1994) using the "Standardization" option in DeltaGen.

In Case Study 1, pattern analysis was conducted on the entry × location BLUP matrix to summarize performance of the 107 perennial ryegrass entries (83 breeding lines and 24 commercial check cultivars) across the locations Palmerston North, Ruakura, and Kerikeri.

Multivariate Analysis

In Case Study 3, a MANOVA was conducted using a linear model similar to Eq. [1], but completely random, which included HS families, years, locations, replicates, HS × year and HS × location interactions. The MANOVA output consisted of matrices of sums of cross products and sums of mean cross products for the different factors in the linear model. Variance-covariance and genetic correlation matrices for the three traits DMY, CWE, and KL were also generated (results are presented in Supplemental File 2).

Selection Index

DeltaGen computes a selection index based on the SH model. A SH index was computed using the switchgrass multitrait data from 147 HS families in Case Study 3. The objective was to identify superior HS families on a multitrait scale, showing genetic worth (I) associated with the potential to simultaneously increase DMY and CWE (strong positive correlation with ethanol) and decrease KL, targeting fermentation platforms (Jahufer and Casler, 2015).

The SH index is used to identify genetic families on their individual genetic worth (I) or breeding value, based on a set of chosen traits. The SH index equation is

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{A}\mathbf{w} \quad [8]$$

where $\mathbf{P}$ and $\mathbf{A}$ are phenotypic and additive genetic variance-covariance matrices, respectively, and $\mathbf{b}$ and $\mathbf{w}$ are vectors of index coefficients and economic weightings. Definition of an economic value for a plant trait and converting it to a realistic weighting can be complicated. In DeltaGen, a subjective approach (Baker, 1974) based on identifying optimum sets of weightings (w) (Christophe and Birot, 1983; Dean et al., 1986) is used. In DeltaGen, different sets of w can be entered manually, and multiple iterations using the SH index equation can be performed. The optimum set of w will be associated with index coefficients (b) that will generate HS family indices (I) resulting in the desired predicted ΔG at a specific selection pressure, estimated according to Van Vleck et al. (1987) as

$$\Delta G_Y = k_f c r_{IY} \sigma_{PY} \quad [9]$$

where ΔG_Y is the predicted ΔG for individual trait Y in the index; k_f is the among-HS-family selection pressure; c is the parental control; r_{IY} is the correlation between the calculated SH indices I for the individual HS families and their BLUP values for trait Y; and σ_{PY} is the among-HS-family phenotypic standard deviation of the BLUP values for trait Y.

With regard to the 147 HS families of switchgrass, expected ΔG for each trait was estimated at 20% selection pressure ($k = 1.4$).

RESULTS

Case Study 1

Comparison of the variance components estimated using DeltaGen and GenStat indicate essentially identical results (Table 1). There was significant ($P < 0.05$) genotypic variation among the 107 entries for mean seasonal herbage growth across years within each location, and also across all locations, seasons, and years. The importance of conducting multilocation trials across multiple years for perennial ryegrass is highlighted by the significant ($P < 0.05$) genotype × location (σ^2_{gl}) and genotype × location × year (σ^2_{gly})

Table 1. Case Study 1: genotypic (σ^2_g), genotype × year interaction (σ^2_{gy}), genotype × season interaction (σ^2_{gs}), genotype × location interaction (σ^2_{gl}), genotype × season × year (σ^2_{gsy}) interaction, genotype × location × year (σ^2_{gly}) interaction, genotype × replicate interaction (σ^2_{gb}), replicates within seasons and years ($\sigma^2_{y/s/b}$), replicates within locations, seasons, and years ($\sigma^2_{y/s/l/b}$), and pooled error (σ^2_ε) variance components, their associated standard errors (±SE), and line mean broad-sense heritability (h^2_b), estimated from seasonal herbage growth (scored on a 0–9 scale) of the 107 perennial ryegrass entries evaluated at Palmerston North (PN), Ruakura (RU), and Kerikeri (KE) across seasons and years. The level of significant differences for each of the fixed effects terms of year, season, location, and their interaction are also presented.

Source of variation	PN across years and seasons		RU across years and seasons		KE across years and seasons		Across years, seasons, and locations	
	DeltaGen	GenStat	DeltaGen	GenStat	DeltaGen	GenStat	DeltaGen	GenStat
σ^2_g	0.464 ± 0.078	0.464 ± 0.078	3.068 ± 0.481	3.068 ± 0.481	0.559 ± 0.131	0.559 ± 0.131	0.932 ± 0.172	0.932 ± 0.172
σ^2_{gy}	0.077 ± 0.017	0.077 ± 0.017	0.420 ± 0.066	0.420 ± 0.066	0.408 ± 0.063	0.408 ± 0.063	0.005 ± 0.023	0.005 ± 0.023
σ^2_{gs}	0.094 ± 0.018	0.094 ± 0.018	0.007 ± 0.039	0.007 ± 0.039	0.000	0.000	0.077 ± 0.012	0.077 ± 0.012
σ^2_{gl}	–	–	–	–	–	–	0.532 ± 0.068	0.532 ± 0.068
σ^2_{gsy}	0.073 ± 0.018	0.074 ± 0.018	0.349 ± 0.045	0.349 ± 0.045	0.266 ± 0.034	0.266 ± 0.034	0.043 ± 0.010	0.043 ± 0.010
σ^2_{gly}	–	–	–	–	–	–	0.407 ± 0.036	0.407 ± 0.036
σ^2_{gb}	0.111 ± 0.014	0.111 ± 0.014	1.034 ± 0.088	1.034 ± 0.088	0.667 ± 0.063	0.667 ± 0.063	0.246 ± 0.023	0.246 ± 0.023
$\sigma^2_{y/s/b}$	0.184 ± 0.043	0.184 ± 0.043	0.012 ± 0.005	0.012 ± 0.005	0.359 ± 0.085	0.359 ± 0.085	–	–
$\sigma^2_{y/s/l/b}$	–	–	–	–	–	–	0.291 ± 0.041	0.291 ± 0.041
σ^2_ε	0.870 ± 0.017	0.870 ± 0.017	1.403 ± 0.023	1.403 ± 0.023	2.169 ± 0.040	2.169 ± 0.040	1.910 ± 0.019	1.910 ± 0.019
h^2_b	0.821 ± 0.025	NA†	0.870 ± 0.030	NA	0.602 ± 0.032	NA	0.743 ± 0.038	NA
Fixed terms	***F* values**							
Year	21.6***	21.6	172.9***	172.8	87***	26.7	123***	123
Season	5.1*	5.1	56.9***	56.8	15***	3.7	NS‡	NS
Location	–	–	–	–	–	–	15***	15
Season × year	7.5*	7.5	NS	NS	30***	30	10***	10

*, *** Significant at the 0.05 and 0.001 probability levels, respectively.

† NA not applicable.

‡ NS, not significant.

interaction variance components estimated for seasonal herbage growth. These estimates indicated that the relative performance of the 107 entries changed across the three locations and 3 yr. There was also significant ($P < 0.05$) interaction of the entries with seasons and years (Table 1). The significant ($P < 0.05$) genotypic variation for herbage growth among the 107 entries across locations, seasons, and years indicated the potential for selection of lines with broad adaptation across the target population of environments represented by Palmerston North, Ruakura, and Kerikeri, which are key perennial ryegrass evaluation sites in New Zealand. The high line mean broad-sense heritability (h^2_b) estimates from DeltaGen provide a rough estimate of the potential genetic variation available for selecting lines for specific or broad adaptation for seasonal herbage growth across the three locations.

Pattern analysis of the 107-entry × three-location BLUP matrix of seasonal herbage growth, generated from REML analysis in DeltaGen, was conducted using the "Pattern Analysis–Cluster" and "Pattern Analysis–PCA" options within the univariate analysis menu. These analyses generated dendrograms of location and line grouping (results presented in Supplemental File 3) and the biplot presented in Fig. 2. This biplot provided a graphical summary of seasonal growth performance of the 107 entries across the three evaluation sites Palmerston North, Ruakura, and Kerikeri. The correlation structure among the three sites, based on performance of the 107 entries, was indicated by the angles between the directional vectors. The association between Palmerston North, Ruakura, and Kerikeri was positive (angles between the directional vectors are at <90°). However, Palmerston North and Ruakura had a stronger positive correlation (Fig. 2). Of the three entry groups generated from cluster analysis, Group 1, with 58 members, consisted of entries with above-average seasonal herbage growth, especially those within the 95% probability error ellipse (Mandel, 2013). Breeding lines 53, 54, 81, and 99 in Group 1 showed above-average performance for mean seasonal growth across all three sites (Fig. 2). These lines showed higher performance than any of the commercial check cultivars. Lines 57 and 58 showed good potential specific adaptation to Palmerston North and Ruakura, and lines 90 and 92 showed good potential specific adaptation to Kerikeri. However, another option would be to combine (polycross) elite genotypes from all the eight breeding lines and develop a population with improved mean seasonal herbage growth across all three locations.

Case Study 2

The mixed model REML analysis of seasonal herbage growth data collected across 3 yr at Glenormiston

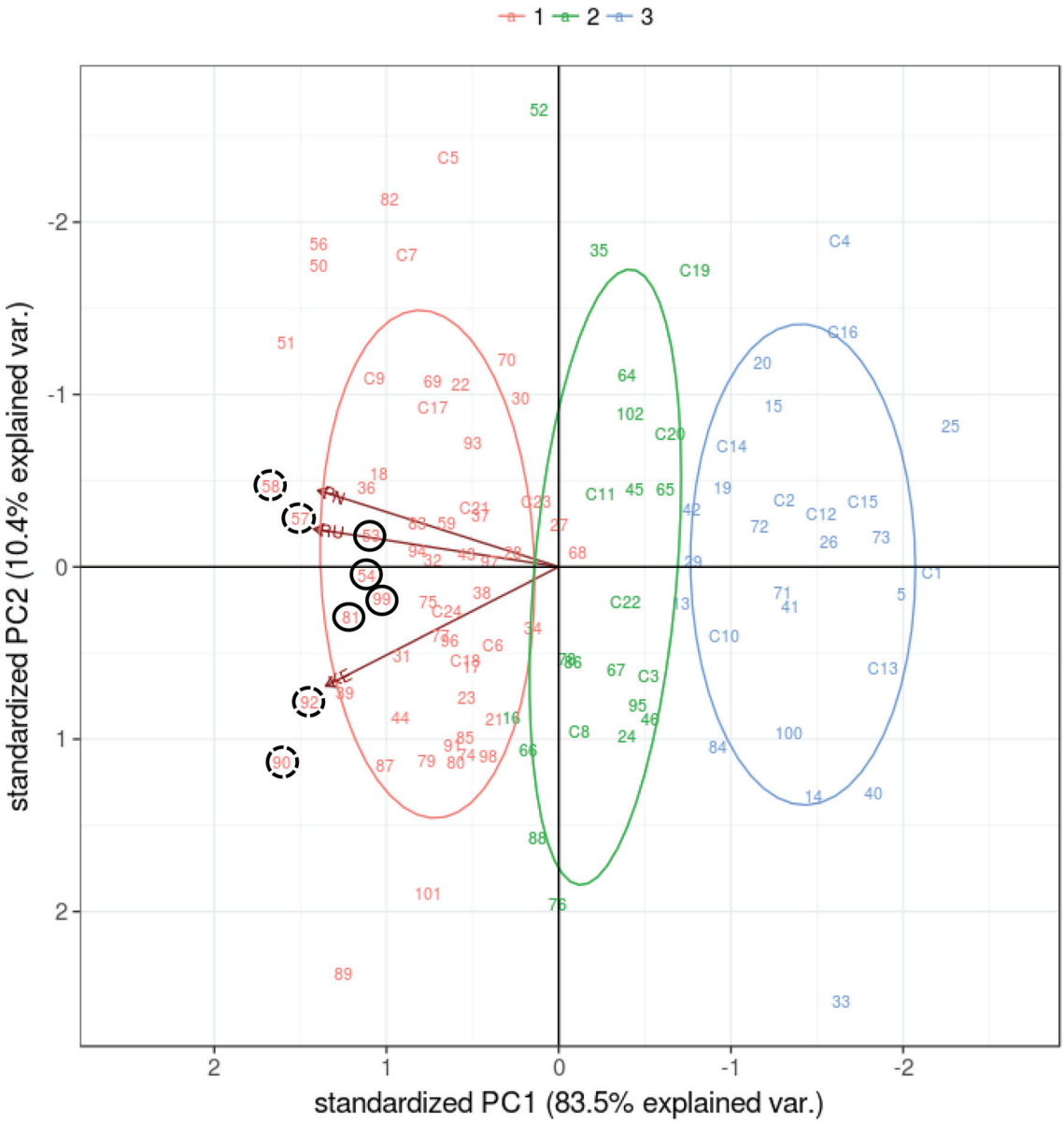

Fig. 2. Case Study 1: biplot generated from pattern analysis of the standardized 107 perennial ryegrass entries (83 breeding lines and 24 commercial check cultivars) × location best linear unbiased predictor (BLUP) matrix for seasonal herbage growth scores, evaluated under grazing for 3 yr at Kerikeri, Palmerston North, and Ruakura. Principal components 1 and 2 (PC1 and PC2) account for 94% of the total variation. The three groups from cluster analysis are indicated by different colors. The breeding lines are indicated by numbers, and the commercial checks from C1 to C24. The error ellipses for each of the three groups indicate a 95% confidence level. Of the 24 commercial check cultivars, 15 are in Groups 2 and 3. Dotted circles indicate specific adaptation; solid circles indicate broad adaptation.

indicated significant ($P < 0.05$) additive genetic variation (σ^2_f) among the 90 HS families of perennial ryegrass (Table 2). There was also significant ($P < 0.05$) HS family × season (σ^2_{fs}) and HS family × year (σ^2_{fy}) interaction. Although the magnitude of additive variation indicated the potential for selection among the HS families for the genetic improvement of seasonal herbage growth, the significant interactions (σ^2_{fs} and σ^2_{fy}) suggest a change in the relative performance of the 90 families across both seasons and years. The estimated narrow-sense heritability on a HS-family-mean basis was intermediate. It should be noted that all the genetic estimates were based on data collected at a single location, not permitting the estimation of possible family × location interaction if the trial was performed across multiple locations. All the variance component estimates from DeltaGen were the same as those derived using GenStat (Table 2).

The estimates of significant additive genetic, family × season, and family × year interaction variance components from the perennial ryegrass HS family analysis were applied in the simulation component of DeltaGen to predict ΔG using equations for three breeding methods (Table 3). The numbers of seasons and years were not changed. The costs for both the field trial and GS are

Table 2. Case Study 2: among-half-sib (HS)-family (σ^2_f), family × season (σ^2_{fs}) interaction, family × year (σ^2_{fy}) interaction, replicates within seasons within years ($\sigma^2_{y/s/b}$), rows within replicates within seasons within years ($\sigma^2_{y/s/b/r}$), columns within replicates within seasons within years ($\sigma^2_{y/s/b/c}$), and experimental error (σ^2_ε) variance components, associated standard errors (±SE), and narrow-sense heritability (h^2_n), estimated from data of the 90 half-sib families of perennial ryegrass evaluated at Glenormiston for seasonal herbage growth (scored on a 1–5 scale) across 3 yr. The level of significant ($P < 0.05$) differences for each of the fixed effects terms of year and season within year are also presented.

Sources of variation	DeltaGen	GenStat
σ^2_f	0.064 ± 0.020	0.064 ± 0.019
σ^2_{fs}	0.022 ± 0.009	0.022 ± 0.008
σ^2_{fy}	0.063 ± 0.014	0.063 ± 0.013
$\sigma^2_{y/s/b}$	0.013 ± 0.030	0.013 ± 0.030
$\sigma^2_{y/s/b/r}$	0.084 ± 0.013	0.084 ± 0.013
$\sigma^2_{y/s/b/c}$	0.111 ± 0.016	0.111 ± 0.017
σ^2_ε	0.763 ± 0.018	0.763 ± 0.019
h^2_n	0.573 ± 0.087	NA†
Fixed term	***F* value**	
Year	7.33*	7.33
Season within year	3.86*	3.87

* Significant at the 0.05 probability level.

† NA, not applicable.

shown at the bottom of Table 3. Applying a selection pressure of 20% resulted in predicted ΔG of 5.21 and 10.42% for the HS and HSPT breeding methods, respectively. This is not surprising, as in the HSPT method, we go back to saved clones of the original parents to combine individuals with good combining ability, and the parental control is 1, in comparison with 0.5 for the HS breeding method. However, progeny from the cross resulting from HSPT is one generation behind the progeny from the HS breeding method. It will require another polycross, which also makes it more expensive (Table 3). The additional polycross may be conducted using glasshouses or off-season nurseries (Casler and Brummer, 2008). The breeding method A_PWF_{gs-HS} using the combination of among-HS-family phenotypic selection and within-family GS, although more expensive than the HS method by $92,500, would provide an extra 4.39% gain in seasonal herbage growth, based on the assumed accuracy ($r_{A\text{-}HS}$) of 0.25. Genomic selection enabled use of the within-family 3/4 additive genetic variation. In addition a higher within-family selection pressure of 5% was applied, as each of the selected 18 HS families were represented by a random sample of 100 seedlings (Table 3).

After estimation of ΔG for the three breeding methods and their associated costs per selection cycle, a set of simulations was conducted using the HS and A_PWF_{gs-HS} strategies. The objective was to examine the effect of reducing the number of years and replicates on the efficiency of the two breeding methods in terms of predicted ΔG and cost ($). Simulation based on altering the number of replicates and years was performed for both HS and A_PWF_{gs-HS} breeding methods. Two simulations were conducted for each breeding method. For both breeding methods, reducing the number of years and replicates by one had a similar reduction in ΔG: 0.69% for HS and 0.70% for A_PWF_{gs-HS} (Table 3). For both breeding methods, by reducing the number of years and replicates, the cost per selection cycle was reduced by $28,400 each.

Case Study 3

The analysis of variance using both DeltaGen and GenStat programs generated the same estimates of variance

Table 3. Case Study 2: predicted rates of genetic gain (ΔG) per selection cycle and associated costs for mean seasonal herbage growth based on data from the evaluation of 90 half-sib families of perennial ryegrass evaluated across 3 yr at Glenormiston. Selection intensity = 1.40 (20%) and 2.06 (5%). Parental control 0.5 for half-sib families and 1 as both parents are selected.

Breeding method†	No. of replicates	No. of years	Selection pressure		Assumed selection accuracy based on HS families	ΔG per selection cycle	Cost per selection cycle
			Among-family	Within-family			
			——— % ———			%	NZ$
HS	3	3	20	–	–	5.21	91,120‡
HSPT	3	3	20	–	–	10.42	101,120‡
A_PWF_{gs-HS}	3	3	20	5	0.25	9.60	183,620‡§
			Simulation				
HS	3	2	20	–	–	4.78	64,080
HS	2	2	20	–	–	4.52	62,720
A_PWF_{gs-HS}	3	2	20	5	0.25	9.17	156,580
A_PWF_{gs-HS}	2	2	20	5	0.25	8.90	155,220

† HS, half-sib family selection; HSPT, HS with progeny testing; A_PWF_{gs-HS}, among-HS-family phenotypic selection and within-family genomic selection.

‡ Field trial costs: cost of scoring one row per growth assessment = $0.50 replicate^{-1} = $500 yr^{-1} location^{-1} = $25,000 yr^{-1}, other (seedling establishment, polycrossing, etc.) = $10,000.

§ Costs associated with genomic selection based on a cost of $50 for generating each genomic estimated breeding values ; total number of seedlings for genotyping by sequencing was 18 (20% of 90 HS) by 100 (number of seedlings sampled per selected HS); $2500 for other associated costs. The cost for one cycle of A_PWF_{gs-HS} was the sum of the field trial costs to select the top 20% of HS families, plus the genomic selection costs for selecting the top 5% of individuals within each selected family.

Table 4. Case Study 3: among-half-sib (HS)-family (σ^2_f), among years (σ^2_y), family × location (σ^2_{fl}) interaction, family × year interaction, (σ^2_{fy}), locations within years ($\sigma^2_{y/l}$), replicates (b) within locations within years ($\sigma^2_{y/l/b}$), rows within replicates within locations within years ($\sigma^2_{y/l/b/r}$), columns within replicates within locations within years ($\sigma^2_{y/l/b/c}$), and experimental error (σ^2_ε) variance components, their associated standard errors (±SE), and narrow-sense heritability (h^2_n), estimated from data of the 147 half-sib families of switchgrass evaluated at Arlington and Marshfield for the three traits biomass dry matter yield (DMY, Mg ha^{-1}), cell wall ethanol (CWE, mg g^{-1}) and Klason lignin (KL, mg g^{-1}), measured in 2011 and 2012. Predicted genetic gains (ΔG) in absolute values and percentages are also given. Selection intensity = 1.76 (10%) and parental control = 0.5. Means of the 147 HS families for the traits DMY, CWE, and KL, were 10.06 Mg ha^{-1}, 56.80 mg g^{-1}, and 98.77 mg g^{-1}, respectively.

	DMY		CWE		KL	
Sources of variation	DeltaGen	GenStat	DeltaGen	GenStat	DeltaGen	GenStat
σ^2_f	0.207 ± 0.098	0.207 ± 0.099	0.428 ± 0.185	0.428 ± 0.186	5.744 ± 2.450	5.744 ± 2.451
σ^2_y	0	0	0	0	0	0
σ^2_{fl}	0.171 ± 0.116	0.171 ± 0.117	0	0	0	0
σ^2_{fy}	0	0	0	0	0.344 ± 2.819	0.344 ± 2.820
$\sigma^2_{y/l}$	6.942 ± 4.367	6.942 ± 4.368	17.871 ± 12.047	17.871 ± 12.048	97.477 ± 65.625	97.477 ± 65.625
$\sigma^2_{y/l/b}$	0.335 ± 0.237	0.335 ± 0.238	3.522 ± 2.200	3.522 ± 2.201	15.530 ± 11.92	15.530 ± 11.92
$\sigma^2_{y/l/b/r}$	0.045 ± 0.058	0.045 ± 0.059	4.848 ± 0.678	4.848 ± 0.679	52.992 ± 7.345	52.992 ± 7.346
$\sigma^2_{y/l/b/c}$	1.147 ± 0.194	1.147 ± 0.195	4.890 ± 0.746	4.890 ± 0.747	40.451 ± 6.417	40.451 ± 6.417
σ^2_ε	4.363 ± 0.177	4.363 ± 0.178	11.420 ± 0.443	11.420 ± 0.444	117.530 ± 4.781	117.530 ± 4.782
h^2_n	0.316 ± 0.120	NA†	0.310 ± 0.095	NA	0.366 ± 0.118	NA
ΔG (%)	0.225 (2.24)	NA	0.321 (0.56)	NA	–1.275 (–1.29)	NA

† NA, not applicable.

components for all sources of variation for the traits DMY, CWE, and KL (Table 4). There was significant ($P < 0.05$) additive genetic variation among the 147 HS families for all three traits. There was no significant ($P > 0.05$) variation for family × year or family × location interaction for any of the traits, indicating that the relative performance of 147 HS families was stable across the 2 yr, and also across the two locations, Arlington and Marshfield (Table 4). The narrow-sense heritability (h^2_n) estimates for the traits were below intermediate levels. However, the estimates indicated the potential for the genetic improvement of all three traits. The breeding strategy simulation option in DeltaGen was used to calculate ΔG (%) for each trait using the HS family selection breeding strategy equation. Selection of the top 10% HS families as parents would result in predicted ΔG (%) of 2.25, 0.56, and −1.29 for the traits DMY, CWE, and KL, respectively (Table 4). The negative is associated with selection to decrease KL, selection of the bottom 10% families.

Estimates of phenotypic correlation between DMY and CWE, DMY and KL, and CWE and KL were 0.25, −0.19, and −0.26, respectively (for all estimates, $P < 0.05$). Both DeltaGen and GenStat generated similar values (results not presented). Genetic correlation for the traits were estimated using the MANOVA option in DeltaGen, which provided a complete analysis that included results of sums of cross products, mean cross products, variance-covariance, and finally the HS family genetic correlation matrix; DMY and CWE, DMY and KL, and CWE and KL were 0.11, −0.32, and −0.15, respectively.

Smith–Hazel Index

As part of the computation of index coefficients in **b** using the SH index, DeltaGen provides detailed output of the associated components—$\mathbf{P}^{-1}$ (inverse of phenotypic variance-covariance) and **A** (additive genetic variance-covariance) matrices—for the SH equation (Table 5). Changing the **w** values will change the index coefficients in **b**. The final **w** values after multiple iterations using the SH index, to identify a set associated with the desired ΔG (%), were 2, 2, and −1 (Table 5).

The comparison of the selection differentials (S) resulting from single- and multi-trait selection is presented in Table 6. As would be expected, the single-trait (individual) selection approach resulted with the highest S for all three traits: DMY, CWE, and KL. However, for each individual trait selection, the selection gains for the other associated traits in the top 10% of the selected HS families were much lower in comparison with their individual S values. In comparison, the S values for DMY, CWE, and KL in the top 10% HS families selected using the SH index were relatively higher. This indicated that the individual HS families within the SH selected families had higher breeding values across all three traits. Estimates of predicted ΔG (Table 6) show that if selection is focused on improving only a specific trait such as DMY, individual trait selection would be the preferred method. However, using a multitrait approach, such as the SH index, will result in concurrent genetic improvement of all traits, increasing DMY and CWE and, as required, reducing KL expression.

DISCUSSION

Analysis of data from the three case studies has demonstrated key analysis procedures that can be successfully performed using DeltaGen. Comparison of the results of REML analysis of data from the three case studies using

Table 5. Case Study 3 components: inverse of phenotypic variance-covariance matrix (P^{-1}), additive genetic variance-covariance matrix (A), weighting coefficients (w) used in the Smith–Hazel index model to calculate the index coefficients (b) for the traits dry matter yield (1.247), cell wall ethanol (–0.006), and Klason lignin (–0.266).

b		P^{-1}			A			w
1.247	=	1.402	−0.017	−0.017	0.207	0.031	−0.345	2
−0.006		−0.017	0.596	0.098	0.031	0.421	−0.223	2
−0.266		−0.017	0.098	0.054	−0.345	−0.223	5.629	−1

DeltaGen and GenStat indicated essentially identical outputs from both software programs.

Using DeltaGen to analyze data from Case Study 1 enabled the combination of REML analysis and pattern analysis to be conducted and generate a graphical summary of the performance of perennial ryegrass entries across the locations Palmerston North, Ruakura, and Kerikeri. The biplot showing three entry clusters and the association among the locations (directional vectors) enabled identification of entries with specific and broad adaptation. A clear picture of the relative performance of the 83 breeding lines and 24 check cultivars was presented in the biplot. This demonstrated the value of DeltaGen when applied to an analysis of forage breeding line evaluation trials conducted across years, seasons, and locations.

Analysis of HS family data of perennial ryegrass from Case Study 2 tested the key quantitative genetic analysis procedures in DeltaGen. Although the basic REML analysis generated estimates of additive, genetic, and associated interaction components, automatic transfer of these estimates into the breeding simulation component of DeltaGen provided the option of using this information in any of the breeding equations available in the program. Based on the estimates of narrow-sense heritability (h^2_n) and associated variance components, predicted ΔG and associated costs ($) per selection cycle were calculated and compared among three breeding strategies: HS, HSPT, and $A_P WF_{gs-HS}$. Simulation of HS and $A_P WF_{gs-HS}$ was also performed by varying year and replicate numbers. The ability to evaluate GS-based breeding strategies against conventional methods in DeltaGen, in terms of predicted ΔG and associated costs, using actual field trial generated data from training populations and associated estimates of GS accuracy, will be valuable to plant breeders.

The data from Case Study 3 were analyzed to investigate a multivariate approach to identify HS families of switchgrass with breeding values that would enable increasing the traits DMY and CWE and decreasing KL. The SH index constructed enabled identification of HS families that would result in simultaneous genetic improvement in all three traits, in contrast with individual selection. In a similar study, Jahufer and Casler (2015) used the SH index to identify HS families of switchgrass with a combination of high DMY and CWE and decreased KL. The SH index in DeltaGen will be a useful selection tool for perennial ryegrass breeders using the forage value index (Chapman et. al., 2017) based on yield, persistence, and quality traits to develop cultivars optimized for the New Zealand dairy industry.

Another significant component in DeltaGen, not described in this paper, is experimental trial design. DeltaGen generates a range of field trial designs: completely randomized, randomized complete block, factorial, and row-column (repeated check plots can also be included). Details of this component are included in the program help menu "Quick start manual."

Analysis of the datasets in the three case studies provided a snapshot of the key analyses available within DeltaGen. The aim of developing DeltaGen is to provide plant breeders with a single integrated solution for experimental design generation, data quality control, statistical and quantitative genetic analyses, breeding strategy evaluation, index selection, and underlying basic information on quantitative genetics. It should be noted that although default options are provided for fixed and random effects in the linear mixed model analysis option, there are no constraints to entering alternative models, by using the "Add" option, based on the factors within the uploaded

Table 6. Case Study 3: selection differentials based on the difference between the mean of the 10% selected half-sib families and the mean of the total 147 families for each trait: biomass dry matter yield (DMY, Mg ha^{-1}), cell wall ethanol (CWE, mg g^{-1}) and Klason lignin (KL, mg g^{-1}). The bolded values indicate selection differential values for the primary traits. Selection differential values for individual traits based on Smith–Hazel (SH) index selection are also presented. The predicted genetic gain (ΔG) is on individual trait and multitrait selection (SH index).

Selection of the top 10% HS families based on individual traits and the SH index	Selection differential		
	DMY	CWE	KL
	%		
Selection on DMY only	**4.33**	−0.06	0.31
Selection on CWC only	−0.13	**0.96**	−1.17
Selection on KL only	0.27	0.54	**−2.16**
Selection on SH index	2.84	0.51	−1.69
	Predicted ΔG		
	%		
On individual trait selection	2.24	0.57	−1.29
Response of individual traits based on SH index	1.42	0.22	−0.91

data. This augments the flexibility of mixed linear model analysis in DeltaGen. Another valuable component in DeltaGen is the option to evaluate the relative efficiencies of different breeding strategies on rate of estimated ΔG and associated costs on an annual and per-selection-cycle basis.

The availability of decision support software to field breeders as tactical tools will help enhance the efficiency of cultivar development programs, especially with the integration of marker-assisted selection technology such as GS. A number of software tools for quantitative genetics and plant breeding research have been developed (Podlich and Cooper, 1998; Wang and Pfeifferz, 2007; Iwata and Jannink, 2011; Mi et al., 2014; Lin et al., 2016; Faux et al., 2016; Yabe et al., 2017). Sun et al. (2011) discussed the importance of computer simulation to provide decision support to plant breeding programs and reviewed a range of software applications. Direct application of these software to applied field breeding programs as tactical tools would be challenging, as they are mostly simulation based and often require specialist knowledge to operate. Software platforms such as QU-GENE" (Podlich and Cooper, 1998) and Selectiongain (Mi et al., 2014) will be valuable as strategic breeding tools. However, their application and effectiveness in breeding programs for perennial cross-pollinating species is yet to be determined. Programming of the breeding strategy application modules within QU-GENE to simulate breeding methods for perennial cross-pollinating species (Casler and Brummer, 2008) will make this software an effective strategic tool to enhance the efficiency of forage cultivar development programs. Integration of the quantitative genetic estimates generated from software such as DeltaGen into QU-GENE will enhance the precision and applicability of prediction outputs to field breeding programs.

In addition to using DeltaGen as a decision support tool, the software also has applicability as a teaching resource in plant breeding courses, as the "Help" windows provide associated theory and important references.

DeltaGen is available as freeware at the URL http://agrubuntu.cloudapp.net/PlantBreedingTool/. The program will continue to be updated and improved. All additions and updates will be uploaded to this link. We recommend that the link is used through Google Chrome.

Conflict of Interest

The authors declare that there is no conflict of interest.

Supplemental Material Available

Supplemental material for this article is available online.

Acknowledgments

This work was funded by Pastoral Genomics (PG+), a joint venture cofunded by DairyNZ, Beef+Lamb New Zealand, Dairy Australia, AgResearch, New Zealand Agriseeds, Grasslands Innovation, DEEResearch, and the Ministry of Business, Innovation and Employment, New Zealand. We wish to thank our AgResearch colleagues Mr. Brent Barrett and Dr. Tony Conner for their continued encouragement and support to develop DeltaGen; Dr. Mingshu Cao for his contribution to the programming of MANOVA; Dr. Andrew Griffiths, Dr. Marty Faville, and Dr. Siva Ganesh for their advice on GS; Dr. Ken Dodds and Dr. John Koolaard for their advice on statistics; Dr. Jingli Lu for help with software server establishment; and Dr. Jeanne Jacobs for support with stakeholder collaboration. Thanks to Dr. Neil Coombes, New South Wales Department of Primary Industry, Australia, and Dr. Mario D'Autuono, Department of Agriculture and Food, Western Australia, for their advice on Digger. Thanks to Dr. Luis Apiolaza, New Zealand School of Forestry, University of Canterbury, for his advice on index selection. Thanks to Dr. Alan Stewart, PGG Wrightson Seeds, New Zealand, and Dr. Rex Clements, Glenormiston campus, South West Institute of TAFE, Victoria, Australia, for providing the datasets for Case Studies 1 and 2, respectively. Thanks to Dr. Michael Casler, USDA-ARS, Madison, WI, for advice on quantitative genetics and providing the switchgrass data used in Case Study 3. We thank the PG+ International Science Advisory Committee for their continuous support. We also wish to thank Dr. Jeffrey Endelman, Associate Editor, Crop Science, and the two reviewers for their valuable comments and suggestions that helped improve the quality of this paper.

References

Baker, R.J. 1974. Selection indexes without economic weights for animal breeding. Can. J. Anim. Sci. 54:1–8. doi:10.4141/cjas74-001

Bassi, F.M., A.R. Bentley, G. Charmet, R. Ortiz, and J. Crossa. 2015. Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). Plant Sci. 242:23–26. doi:10.1016/j.plantsci.2015.08.021

Burr, E.J. 1968. Cluster sorting with mixed character types. I. Standardization of character values. Aust. Comput. J. 1:97–99.

Burr, E.J. 1970. Cluster sorting with mixed character types. II. Fusion strategies. Aust. Comput. J. 2:98–103.

Casler, M.D., and E.C. Brummer. 2008. Theoretical expected genetic gains for among-and-within-family selection methods in perennial forage crops. Crop Sci. 48:890–902. doi:10.2135/cropsci2007.09.0499

Chang, W., J. Cheng, J.J. Allaire, Y. Xie, and J. McPherson. 2017. Shiny: Web application framework for R. R package version 1.0.0. R Found. Stat. Comput., Vienna. https://CRAN.R-project.org/package=shiny (accessed 12 Feb. 2018).

Chapman, D.F., J.R. Bryant, M.E. Olayemi, G.R. Edwards, B.S. Thorrold, W.H. McMillan, et al. 2017. An economically-based evaluation index for perennial and short-term ryegrasses in New Zealand dairy farm systems. Grass Forage Sci. 72:1–21. doi:10.1111/gfs.12213

Christophe, C., and Y. Birot. 1983. Genetic structures and expected genetic gains from multitrait selection in wild populations of Douglas fir and Sitka spruce. II. Practical application of index selection on several populations. Silvae Genet. 32:173–181.

Cooper, M., and I.H. DeLacy. 1994. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment trials. Theor. Appl. Genet. 88:561–572. doi:10.1007/BF01240919

Cooper, M., I.H. DeLacy, and R.L. Eisemann, 1993. Recent advances in the study of genotype × environment interactions and their application to plant breeding. In: B.C. Imrie and J.B. Hacker, editors, Proceedings of the 10th Australian Plant Breeding Conference, Gold Coast, QLD. Vol. 1. Univ. of Queensland, Brisbane. p. 116–131.

Cooper, M., and G.L. Hammer, editors. 1996. Plant adaptation and crop improvement. CABI, Wallingford, UK.

Dean, C.A., P.P. Cotterill, and R.L. Eisemann. 1986. Genetic parameters and gains expected from selection in *Pinus caribaea* var. *hondurensis* in northern Queensland, Australia. Silvae Genet. 35:229–236.

Dekkers, J.C.M. 2007a. Marker-assisted selection for commercial crossbred performance. J. Anim. Sci. 85:2104–2114. doi:10.2527/jas.2006-683

Dekkers, J.C.M. 2007b. Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124:331–341. doi:10.1111/j.1439-0388.2007.00701.x

Desta, Z.A., and R. Ortiz. 2014. Genomic selection: Genome-wide prediction in plant improvement. Trends Plant Sci. 19:592–601. doi:10.1016/j.tplants.2014.05.006

Dudley, J.W., and R.H. Moll. 1969. Interpretation and use of estimates of heritability and genetic variances in plant breeding. Crop Sci. 9:257–262. doi:10.2135/cropsci1969.0011183X000900030001x

Easton, H.S., and L.R. Fletcher. 2007. The importance of endophyte in agricultural systems-changing plant and animal productivity. In: A.J. Popay and E.R. Thom, editors, Proceedings of the 6th International Symposium of Fungal Endophytes of Grasses. Grassland Res. Practice Ser. 13. N. Z. Grassland Assoc., Christchurch, New Zealand. p. 11–18.

Eisemann, R.L., M. Cooper, and D.R. Woodruff. 1990. Beyond the analytical methodology better interpretation and exploitation of genotype-by-environment interaction in plant breeding? In: M.S. Kang, editor, Genotype-by-environment interaction and plant breeding. Louisiana State Univ., Baton Rouge. p. 108–117.

Falconer, D.S. 1989. Introduction to quantitative genetics. Longman Sci. Tech., New York.

Faux, A.M., G. Gorjanc, R.C. Gaynor, M. Battagin, S.M. Edwards, D.L. Wilson, et al. 2016. AlphaSim: Software for breeding program simulation. Plant Genome 9. doi:10.3835/plantgenome2016.02.0013

Galwey, N.W. 2006. Introduction to mixed modelling: Beyond regression and analysis of variance. John Wiley & Sons, West Sussex, UK. doi:10.1002/9780470035986

Grinberg, N.F., A. Lovatt, M. Hegarty, A. Lovatt, K.P. Skøt, R. Kelly, et al. 2016. Implementation of genomic prediction in *Lolium perenne* (L.) breeding populations. Front. Plant Sci. 7:133. doi:10.3389/fpls.2016.00133

Hallauer, A.R., and J.B. Miranda. 1988. Quantitative genetics in maize breeding. Iowa State Univ. Press, Ames.

Hartigan, J.A. 1975. Clustering algorithms. John Wiley & Sons, New York.

Harville, D.H. 1977. Maximum likelihood approaches to variance component estimation and related problems. J. Am. Stat. Assoc. 72:320–340. doi:10.2307/2286796

Hazel, L.N. 1943. The genetic basis for constructing selection indexes. Genetics 28:476–490.

Heslot, N., H.P. Yang, M.E. Sorrells, and J.-L. Jannink. 2012. Genomic selection in plant breeding: A comparison of models. Crop Sci. 52:146–160. doi:10.2135/cropsci2011.06.0297

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martinez. 2002. Estimating and interpreting heritability for plant breeding: An update. Plant Breed. Rev. 22:9–112. doi:10.1002/9780470650202.ch2

Iwata, H., and J.-L. Jannink. 2011. Accuracy of genomic selection prediction in barley breeding programs: A simulation study based on the real single nucleotide polymorphism data of barley breeding lines. Crop Sci. 51:1915–1927. doi:10.2135/cropsci2010.12.0732

Jahufer, M.Z.Z., and M.D. Casler. 2015. Application of the Smith–Hazel selection index for improving biomass yield and quality of switchgrass. Crop Sci. 55:1212–1222. doi:10.2135/cropsci2014.08.0575

John, J.A. 1987. Row: Column designs In: Cyclic designs: Monographs on statistics and applied probability. Springer, Boston. doi:10.1007/978-1-4899-3326-3_5

Jolliffe I.T. 2002. Principal component analysis. Springer series in statistics. 2nd ed. Springer, New York. doi:10.1007/b98835

Lin, Z., N.O.I. Cogan, L.W. Pembleton, G.C. Spangenberg, J.W. Forster, B.J. Hayes, and H.D. Daetwyler. 2016. Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. The Plant Genome 9. doi:10.3835/plantgenome2015.06.0046

Lorenz, A.J. 2013. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. G3: Genes, Genomes, Genet. 3:481–491. doi:10.1534/g3.112.004911

Mandel, M. 2013. Simulation based confidence intervals for functions with complicated derivatives. Am. Stat. 67:76–81. doi:10.1080/00031305.2013.783880

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Mi, X., H.F. Utz, F. Technow, and A.E. Melchinger. 2014. Optimizing resource allocation for multistage selection in plant breeding with R package *Selectiongain*. Crop Sci. 54:1413–1418. doi:10.2135/cropsci2013.10.0699

Milligan, S.B., K.A. Gravois, K.P. Bischoff, and F.A. Martin. 1990. Crop effects on broad-sense heritabilities and genetic variances of sugarcane yield components. Crop Sci. 30:344–349. doi:10.2135/cropsci1990.0011183X003000020020x

Moll, R.H., and C.W. Stuber. 1974. Quantitative genetics-empirical results relevant to plant breeding. Adv. Agron. 26:277–313. doi:10.1016/S0065-2113(08)60874-3

Nyquist, W.E. 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10:235–322. doi:10.1080/07352689109382313

Patterson, H.D., and R. Thompson. 1971. Recovery of interblock information when block sizes are unequal. Biometrika 58:545–554. doi:10.1093/biomet/58.3.545

Patterson, H.D., and R. Thompson. 1975. Maximum likelihood estimation of components of variance. In: L.C.A. Corsten and T. Postelnicu, editors, Proceedings of the 8th International Biometrical Conference, Bucharest, Romania. Acad. Repub. Soc. Rom., Bucharest. p. 197–207.

Payne, R.W., S.A. Harding, D.A. Murray, D.M. Soutar, D.B. Baird, A.I. Glaser, et al. 2009. The guide to GenStat release 12, Part 2: Statistics. VSN Int., Hemel Hempstead, UK.

Podlich, D.W, and M. Cooper. 1998. QU-GENE: A simulation platform for quantitative analysis of genetic models. Bioinformatics 14:632–653.

R Core Team. 2016. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna, Austria.
SAS Institute. 2011. Base SAS 9.3 procedures guide. SAS Inst., Cary, NC.
Smith, F.H. 1936. A discriminate function for plant selection. Ann. Eugen. 7:240–250. doi:10.1111/j.1469-1809.1936.tb02143.x
Steel, G.D.R., and J. Torrie. 1981. Principles and procedures of statistics A biometrical approach. McGraw-Hill, New York.
Sun, X., T. Peng, and R.H. Mumm. 2011. The role and basics of computer simulation in support of critical decisions in plant breeding. Mol. Breed. 28:421–436. doi:10.1007/s11032-011-9630-6
Van Vleck, L.D., E.J. Pollak, and E.A.B. Oltenacu. 1987. Genetics for the animal sciences. W.H. Freeman and Co., New York.
VSN International. 2003. GenStat for Windows: Release 7.1. VSN Int., Oxford, UK.
Wang, J., and W.H. Pfeifferz. 2007 Simulation modeling in plant breeding: Principles and applications. Agric. Sci. China 6:908–921. doi:10.1016/S1671-2927(07)60129-1
White, T.L., and G.R. Hodge. 1989. Predicting breeding values with applications in forest tree improvement. Forestry Sci. Ser. 33. Kluwer Academic, Boston, MA.
Wishart, D. 1969. Algorithm for hierarchical classifications. Biometrics 25:165–170. doi:10.2307/2528688
Yabe, S., H. Iwata, and J.-L Jannink. 2017. A simple package to script and simulate schemes: The breeding scheme language. Crop Sci. 57:1347–1354. doi:10.2135/cropsci2016.06.0538