

Genetic architecture of apple fruit quality traits following storage and implications for genetic improvement

Craig M. Hardner¹ · Kate Evans² · Chris Brien³ · Fred Bliss⁴ · Cameron Peace⁵

Received: 8 September 2015 / Revised: 31 January 2016 / Accepted: 19 February 2016 / Published online: 5 March 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Accurate prediction of genetic potential and response to selection in breeding requires knowledge of genetic parameters for important selection traits. Data from breeding trials can be used to obtain estimates of these parameters so that predictions are directly relevant to the improvement program. Here, a factor allocation diagram was developed to describe the sampling design used to assess the quality of fresh and post-storage (2 months) fruit from advanced selection trial in an apple breeding program from which models for analyses were developed. Genetic variation was the largest source of variation for the fruit size, red colour type, proportion of red skin colour and lenticels, and instrumentally assessed fruit diameter, mass, puncture force and titratable acidity. In contrast, residual variation was the largest for fruit shape, juiciness, sweetness, aromatic flavour, eating and overall quality, and instrumental crispness. Genetic effects for traits were

generally stable over fixed effects, except for a significant interaction with storage duration for firmness. Genetic correlations among traits were generally weak except between fruit mass (and diameter) and sensory size (0.98), titratable acidity and sensory acidity (0.97), puncture force and sensory firmness (0.96–0.90), crispness and juiciness (0.87), sweetness and aromatic flavour (0.84) and instrumental and sensory crispness (0.75). Predictions of the performance for seven commercial cultivars are presented. This study suggests that the Washington State apple production area can be treated as a single target environment and sufficient diversity exists to generate new elite cultivars. In addition, options for evaluating the efficiency of apple breeding are discussed.

Keywords Heritability · Repeatability · Breeding · Genetic correlation · Cultivar testing

Communicated by D. Chagné

✉ Craig M. Hardner
craig.hardner@uq.edu.au

Kate Evans
kate_evans@wsu.edu

Chris Brien
chris.brien@unisa.edu.au

Fred Bliss
fbliss@dcn.org

Cameron Peace
cpeace@wsu.edu

¹ Queensland Alliance for Agriculture and Food Innovation, University of Queensland, St Lucia, QLD 4072, Australia

² Washington State University (WSU-TFREC), 1100 N. Western Avenue, Wenatchee, WA 98801, USA

³ University of South Australia, GPO Box 2471, Adelaide, SA 5001, Australia

⁴ Department of Plant Sciences, University of California, One Shields Ave., Davis, CA 95616, USA

⁵ Department of Horticulture, Washington State University, Pullman, WA 99164, USA

Introduction

Knowledge about the genetic architecture of selection traits in a breeding program is required for accurate prediction of genetic effects (Henderson 1984) and selection response (Falconer and Mackay 1996). Parameters describing the quantitative genetic architecture of a trait include repeatability, which describes the relative influence of genetic factors on a trait, and genetic and environmental correlations, which describe how changes in one trait affect levels of another.

Accurate estimates of genetic parameters can be obtained from specialised field trials designed to minimise confounding of factors (e.g. Williams et al. 2002). However, these types of trials can be expensive for horticultural tree crops as the individual experimental unit is large and must be maintained for several seasons. In addition, germplasm used for estimation of genetic parameters might not reflect that of breeding populations to which they are to be applied. Multiple traits are often of interest and their assessment methods can be laborious or require specialised equipment. An alternative to specialised field trials is the use of data collected as an ongoing part of a breeding program so that genetic parameters are directly relevant to that program (Kumar et al. 2010; McKay et al. 2011). The aims of this study were to use data routinely collected from an applied breeding program to identify major sources of variation and correlations among fruit quality selection traits and examine the implications of the genetic architectures of these traits for the genetic improvement of apples.

Fruit quality, a combination of appearance and eating quality (texture, juiciness, sweetness, acidity and aroma), is a major driver of consumer choice in apples. Several studies (Dailliant-Spinnler et al. 1996; Hampson et al. 2000; Harker et al. 2003) have attempted to further define consumer preference. Texture appears to be the principal driver with consumers particularly focusing on crispness (Hampson et al. 2000). When fruit is above an acceptable texture threshold, other quality factors become more important determinants of consumer preference, with some studies (Dailliant-Spinnler et al. 1996) clearly showing consumer segmentation based on sweetness and acidity. No single set of quality attributes (e.g. firm, crisp, sweet and low acid) is liked by *all* consumers; the wide range of textures and sweetness/acidity ratios offered by the available range of apple cultivars and demanded by consumers is what helps distinguish apple from other fresh fruit products on the market (Bonany et al. 2014; Dailliant-Spinnler et al. 1996).

The Washington State University Apple Breeding Program (WABP) was used as the model for this study. This program was established in 1994 to develop new improved apple cultivars suited to the growing conditions of central Washington State (Evans 2013), a region that produces approximately 60 % of the US dessert apple crop. Ensuring consistent fruit quality following storage is an important consideration as the

crop is marketed throughout the year (Fellman et al. 2003). In the WABP, cultivar candidates pass through three phases of selection before any are identified for release as new cultivars (Evans 2013). In the first phase, each candidate is represented by a single tree; consequently, the amount of fruit, and therefore quality evaluation, is limited. ‘Advanced’ selections moving to phase 2 are vegetatively propagated and five trees of each are planted in a randomised trial at each of three locations representing different growing conditions in WA, with reference cultivars also planted to support performance comparisons. Replication of selections and higher individual tree production compared with the first phase means that in phase 2 it is possible to evaluate fruit from multiple harvest dates and after several durations of post-harvest storage. Phase 3 is a more extensive commercial evaluation of the most ‘elite’ selections, with approximately 50 to 75 trees planted at each of up to four locations.

The most accurate assessment of apple sensory traits is the human palate, particularly using trained sensory panellists in a well-controlled design (Forde et al. 2007). However, this approach is not feasible for large-scale testing required in applied breeding programs and commonly small teams of experts that have developed a deep understanding of fruit quality variability through many years of experience are employed (Brookfield et al. 2011; Hampson et al. 2000; Oraguzie et al. 2009). Alternatively, instrument-based evaluation methods might be less expensive and more repeatable than sensory approaches (Harker et al. 2002a, b) as instrument-based traits might be less affected by non-genetic sources of variation. Thus, selection efficiency could be improved if there is a strong genetic correlation between an instrument-based and sensory measure of a trait (e.g. Evans et al. 2010; Oraguzie et al. 2009), particularly for those traits that are difficult or expensive to assess sensorially or are prone to operator bias.

Methods

Field trial design

Data used were routinely collected as part of WABP phase 2 evaluation trials established between 2004 and 2011. Each year, advanced selections identified from WABP phase 1 seedling trials and commercial cultivars were planted as entries in trials at three orchard locations representing the range of the commercial production region of the Washington State apple industry (Table 1). Between 2004 and 2008, trials were established in the south of Washington State near Basin City, in the centre near Orondo (Columbia View) and in the north near Chelan (Fuller’s). From 2009, the southern trials were planted near Prosser and central trials near Wenatchee (Sunrise), with northern trials continuing near Chelan. Selections from phase 1 were not replicated among years,

Table 1 Congruence of genetic entries (selections and standard cultivars) among series (i.e., year of planting) by farm

Series	Farm	2004 B, C	F	2005 B, F	C	2006 B	C, F	2007 B, C, F	2008 B, F	2009 F	P, S	2010 F, P, S	2011 F	P, S
2004	B, C	16	15	2	2	2	2			2	3	3	3	3
	F		15	2	2	2	2			2	3	3	3	3
2005	B, F			13	12	1	1			3	3	3	3	3
	C				12	1	1			3	3	3	3	3
2006	B					9	8			1	2	2	2	2
	C, F						8			1	2	2	2	2
2007	B, C, F							2						
2008	B, F								2					
2009	F									27	27	4	4	4
	P, S										28	5	5	5
2010	F, P, S											20	5	5
2011	F												12	12
	P, S													13

B Basin City, *C* Columbia View, *F* Fullers, *P* Prosser, *S* Sunrise

although two to five commercial standards were usually planted along with the selections. A total of seven commercial cultivars were included in this study: ‘Braeburn’ (planted in 2004), ‘Coop15’ (planted in 2004), ‘Fuji’ (planted in 2004, 2006, 2009, 2010 and 2011), ‘Gala’ (planted in 2004, 2005, 2006, 2009, 2010 and 2011), ‘Golden Delicious’ (planted in 2005, 2009, 2010 and 2011), ‘Honeycrisp’ (planted in 2009, 2010 and 2011) and ‘Cripps Pink’ (planted in 2004, 2005, 2009, 2010 and 2011).

Assessment methods

Data were taken on fruit collected from field trials during 2005 through 2012. Of 16 ordinal sensory traits, fruit size (SIZE), fruit shape (SHAPE), proportion of red coloured flesh (PCOL), type of red colour flesh (TYPECOL), main ground colour of flesh (GRDCOL), extent of russetting (RUSS) and extent of lenticels (LENT) were appearance sensory traits (Table 2). Three were texture-related (firmness—FIRM, crispness—CRISP and juiciness—JUIC); three were taste-related (sweetness—SWEET, tartness—TART and aromatic flavour—AROM); and appearance quality (APPSUM), eating quality (EQ) and overall quality (OVERALL) were hedonic quality ratings.

The instrumentally assessed traits were fruit diameter (FRTD, cm), fruit mass (FRTM, g), total soluble solids concentration (TSS, °Bx, using a digital refractometer), titratable acidity (TA, mg/l, using an auto-titrator) and the texture traits of M1 (maximum pressure at a constant velocity and force required to penetrate the outer region of the apple flesh, kPa), M2 (maximum force required to penetrate the inner region between the outer flesh and the inner core, kPa) and CN

[a Fourier transformation of forces Evans et al. (2010)]. A Mohr® DigiTest instrument (Evans et al. 2010; Mohr and Mohr 2000) was used to assess fruit diameter (FRTD, mm), fruit mass (FRTM, g) and three measures of pressure resistance of flesh (M1, M2 and CN) in 2009 only on fruit from Basin City, Columbia View and Fullers and from 2010 on fruit from Basin City, Fullers, Prosser and Sunrise.

Statistical methods

The linear model for analysis was developed following the methods described in Brien and Bailey (2006) and Brien and Demetrio (2009) (Appendix). The design involves multi-stages of sampling, which can also be referred to as multi-phase design (e.g. Wilkinson et al. 2008). However, here we reserve the term ‘phase’ for the different steps in the breeding program. Initially, a factor allocation diagram was developed to describe the stages in the experimental design and all identifiable sources of variation from which experimental data were generated. Using the factor allocation diagram, a full linear model was developed to describe the sampling design and was reduced to an estimable model of convenience that recognises the sources of confounding.

The experimental design is described as a four-stage design. The first stage (PLANTS) was composed of a series of entries (*G*) replicated by trees. The second stage (FRUIT SAMPLING) was composed of apples, collected at three harvests (*H*) over several seasons (*S*) from a single plant at a position within a plot (*P*) within a block at a farm (*F*). Each block was composed of a series of entries that were randomised to five non-contiguous plots of positions planted with a single tree. Timing of the harvest for the plant at a

Table 2 Details of fruit quality traits considered in this study including cultivar used as anchor and the perceived value of anchor

Trait	Class	Description	Scale		No. of values	Anchor	
			Min	Max		Cultivar	Value
SIZE	Appearance	Fruit size	1 (small)	5 (large)	11	Gala	3
SHAPE	Appearance	Fruit roundness	1 (flat)	5 (cylindrical)	10	Gala	3
PCOL	Appearance	Proportion red colour	1 (1 %)	5 (100 %)	10		
TCOL	Appearance	Red colour type	1 (blush)	3 (stripe)	6		
GCOL	Appearance	Main ground colour	1 (green)	3 (yellow)	6		
RUSS	Appearance	Extent of russetting	1 (severe)	5 (absent)	10	Gala	3
LENT	Appearance	Extent of lenticels	1 (large)	5 (absent)	10		
FIRM	Texture		1 (soft)	5 (firm)	11	Cripps Pink	4
CRISP	Texture		1 (chewy)	5 (crisp)	10	Honeycrisp	4
JUIC	Texture	Juiciness	1 (very dry)	5 (very juicy)	11	Honeycrisp	4
SWEET	Taste	Sweetness	1	5 (very sweet)	9	Gala	3
TART	Taste	Tartness	1	5 (very acid)	11	Gala	3
AROM	Taste	Aroma taste	1 (bland)	5 (very fruity)	12	Gala	3
APP	Summary	Appearance quality	1 (ugly)	5 (beautiful)	11	Gala	3
EQ	Summary	Eating quality	1 (yuck)	10 (outstanding)	15	Gala	6
OVERALL	Summary	Overall quality	1 (reject)	3 (advance)	4		

position was determined by weekly visual assessment of maturity, and all apples considered mature were collected. The third stage, *STORAGE*, was composed of fruit, within cases, within two storage rooms, within seasons. Within each season, apples were bulked across the trees within a plot for each block and two samples (one for sensory assessment and one for instrument-based assessment) of five fruit was allocated to a case. Storage treatment duration (*D*, fresh, i.e. no storage, or 2 months under regular air storage at approximately 2 °C) was applied to the storage rooms. The fourth stage (*EVALUATION*) describes the structure of assessment each season. For each annual assessment, each case of five fruit was assessed as a single unit with all cases from the same storage room (i.e. storage treatment) assessed at the same event.

The final estimable model for the analysis of individual trait observations for the unit of five fruits from each duration, harvest, season, entry, plot and farm was

$$y = F*(S + A)*H*D + G/(D*S*H) + F^P/(S*H) + F^P^D/(S + H) + e$$

where the symbol \wedge denotes the interaction between terms, / denotes nesting of terms (i.e. $A/B = A + A \wedge B$) and * implies a full expansion of terms (i.e., $A * B = A + B + A \wedge B$). The terms from the expansion of $F*(S + A)*H*D$ were treated as fixed and the rest as random. The factor age (*A*) was added as this effect could be separated from season for some terms as different series were planted at different years but were assessed in the same season. The error term, *e*, was the variation among

individual observations after the effect of the factors specified in the model had been removed. However, this model does not fully specify the suite of factors that can influence variation, as confounded terms were removed to achieve an estimable model (Appendix). The term cases within season-by-rooms, which describes the variation at the third stage, was completely confounded (and hence removed) with units within assessments-by-events in the fourth stage (retained). In addition, the effect of season in the third stage (removed) and the effect of assessment in the fourth stage (removed) were completely confounded with the effect of season in the second stage (retained). The term duration-by-harvest-by-season-by-plot-by-block-by-farm was removed as this was confounded with the error term. The term event-by-assessment was removed as it was confounded with season-by-duration. All terms involving series-by-farm and entry-within-series-by-farm were removed as they were confounded with plot-within-block-within-farm. Finally, the variation among entries might be inflated if there is variation between series, as series was removed to simplify the model and any variation among series is likely to be due to random sampling as different series were not selected for different targets.

The general mixed linear random model used to estimate variance components and test fixed effects for individual fruit quality traits under different storage durations was

$$y = Xb + Z_{it}u + r$$

where *y* was a vector of observations for the trait following 2 months storage, *b* was a vector of fixed effects for the trait to be estimated, *X* was an incidence matrix that mapped the

observations to the fixed effect model, \mathbf{u} was a vector of random effects, \mathbf{Z}_u was an incidence matrix that mapped the observations onto the random effects and \mathbf{r} was a vector of random residuals for each observation.

The variance of \mathbf{y} was given as

$$\text{var}(\mathbf{y}) = \mathbf{Z}_u \mathbf{U} \mathbf{Z}_u^T + \mathbf{R}$$

where \mathbf{y} and \mathbf{Z} are defined above, \mathbf{U} was the variance-covariance among random effects and \mathbf{R} was the variance-covariance among the different traits. The structure of \mathbf{U} was a block diagonal matrix, with each element being the variance-covariance matrix among the effects for each random term. As no relationship was accounted for among entries, all these individual term variance-covariance matrices were also diagonal. The \mathbf{R} matrix was also diagonal and assumed that the residual variance was homogenous across different structures.

Variance components for the random factors defined in the mixed model were estimated by restricted maximum likelihood (Patterson and Thompson 1971) with the software package ASReml (Gilmour et al. 2009). To examine the assumption of normality of residual required by the mixed linear model approach, the Shapiro-Wilk statistic (Shapiro and Wilk 1965) was calculated for the residuals.

Wald statistics were used to test fixed effects (Kenward and Roger 1997). The significance of random terms was tested using the likelihood ratio test (LRT) (Wilks 1938). However, for testing terms at the edge of the boundary space (e.g. if a variance component is 0), an adjusted LRT was used following Stram and Lee (1994). A value of $P=0.05$ was used for significance testing.

To compare the architecture of genetic and non-genetic variation of fruit quality traits, the ratio of the variance of individual significant random terms to the total phenotypic variance was calculated. The total phenotypic variance was calculated as the sum of the individual significant variance components. To evaluate the magnitude of the interaction among entry effects and storage duration, the average genetic correlation between storage durations was estimated as follows:

$$\hat{r}_{gD} = \frac{\hat{v}G}{(\hat{v}G + \hat{v}G.D)}$$

where $\hat{v}G$ is the estimated variance among the main effects of entry, and $\hat{v}G.D$ is the variance of the interaction between entry effects and storage duration. The average genetic correlation among season (i.e. \hat{r}_{gS}) was estimated in a similar manner. While the interaction between entry effects and farm was confounded with the plot-within-farm effect, the upper limit of the genetic correlation among farms was estimated as follows:

$$\hat{r}_{gF^*} = \frac{\hat{v}G}{(\hat{v}G + \hat{v}F.P)}$$

where $\hat{v}G.F$ was the estimate of the plot within farm variance.

To estimate genetic correlations among entry effects, the term D (storage duration) was removed and a term K was added to the model of convenience for single traits defined above:

$$y = K/[F^*(S + A)]*H + K^{\wedge}[G/(S*H)] + K^{\wedge}[F^{\wedge}P/(S + H)] + K^{\wedge}e$$

where the terms from the expansion $K/F^*(S + A)*H$ were treated as fixed and the remainder random. In this model, all random terms can be expressed in the form $K \cdot \text{subject}$ (where *subject* was all the other terms in the linear model for a single trait). A correlation-variance matrix was assumed for $K^{\wedge}G$ by extending the structure of the variance-covariance among random effects (\mathbf{U} , in the mixed models). The remaining random terms except the residual ($K^{\wedge}e$) were treated independent, i.e. uncorrelated, to reduce the number of parameters required for estimation and the instability of the REML estimation methods.

The structure of the variance-covariance matrix for $K^{\wedge}e$ was complex as the observational unit on which the sensory traits were assessed was different from that on which instrument-based traits were assessed. Hence, covariances among residuals for traits assessed on different units were constrained to 0, while covariances among those assessed on the same unit were estimated.

For multi-variate analyses that included more than three traits, the variance-covariance matrix of entry effects was approximated by a factor analytic (FA) parameterisation (Piepho 1997; Smith et al. 2001; Thompson et al. 2003) as it can be difficult to estimate parameters for an unstructured matrix that is the most general form of the variance-covariance among traits. The FA parameterisation reduces the number of parameters requiring estimation by modelling the main patterns in the variance-covariance matrix as loadings on hypothetical common factor/s (λ_q) and a specific variance not explained by the loadings on the common factor/s (ψ) for each trait. An FA structure was also used to model the residual variance-covariance matrix for analyses of traits that were assessed on the same unit. For the analyses of those assessed on different units, an unstructured variance-covariance matrix was used (with the appropriate covariances constrained to 0). The most parsimonious FA model required to model a variance-covariance matrix was determined by likelihood ratio testing as described above for non-boundary conditions. Degrees of freedom were estimated as the difference in the number of non-boundary parameters between the extended and reduced

model. For the multi-variate models of traits assessed on the same unit, the order of the factor analytic parameterisation of the variance-covariance model was alternatively increased for the entry and residual variance-covariance matrices.

Ordination and cluster analyses were used to investigate the patterns in the estimated genetic variance-covariance matrices. Loadings and scores estimated using factor analytic parameterisation of the variance-covariance matrices were rotated to the principal component representation (Smith et al. 2001) so that the loadings of the first factor account for the maximum amount of variation in the variance-covariance matrix. Prior to rotation, the FA loadings were divided by the *entry* variance to remove the effect of heterogeneity in scale among traits.

Hierarchical clustering was undertaken of the trait dissimilarity matrix estimated from the correlation matrix using Gowers transformation (Kempton 1984). Clusters were formed using Ward's minimum variance strategy (DeLacy et al. 1996). The leaf nodes of the cluster diagram were ordered using the Gruvaeus and Wainer algorithm in the R package seriation (Hahsler et al. 2008) so that each object in a cluster is adjacent to the most similar object in a neighbouring cluster.

Results

Individual trait architecture

The distribution of the residuals from individual trait analyses closely approximated a normal distribution. Shapiro-Wilk statistics were greater than 0.97 for all analyses.

The main effects of *Season*, *Harvest*, storage *Duration* and *Farm* were significant for most traits, while the effect of *Age* was less consistent (Table 3). Interactions between *Season* and storage *Duration*, and between *Season* and *Farm*, were also significant for most traits, while interactions between *Season* and *Harvest*, *Age* and *Farm* and storage *Duration* and *Farm* were significant for fewer traits. Interaction between *Age* and *Harvest* and three-way interactions were significant for only a few traits. There was no significant effect of the interaction between *Age* and storage *Duration*, *Farm*, *Age* and *Harvest*, and *Farm*, *Age*, *Harvest* and storage *Duration* for any trait. However, interpretation of the significance of fixed effects is difficult as fixed effects are confounded to some degree due to the unbalanced nature of the sampling design.

The main effect of *entry* was the largest source of phenotypic variation for SIZE, FRTD, FRTM, TCOL, PCOL, LENT, M1, M2 and TA, with residual variation being below 15 % of phenotypic variation for FRTD, FRTM, M1 and M2 (Table 4). In contrast, residual variation was the largest source of variation among observations for SHAPE, JUIC, CN, SWEET, AROM, EQ and OVERALL. The relative

magnitudes of genetic and residual variation were similar for GCOL, RUSS, FIRM, CRISP, TSS, TART and APPSUM. The proportion of phenotypic variance due to variation among *entries* for all instrumentally assessed traits was higher than that for the sensory assessment of the corresponding traits except for CN (i.e. compared to CRISP).

No significant interaction between *entry* and *Harvest*, *Plot* within *Farm* by *Harvest* and *Plot* within *Farm* by *Harvest* by storage *Duration* was detected for any trait (Table 4). In contrast, interactions between *Plots* within *Farms* and *Season*, and *Plots* within *Farms* by *Season* and *Harvest* accounted for more than 10 % of the phenotypic variation for PCOL, LENT, CRISP, JUIC, TART EQ, and OVERALL.

Although genetic correlation among *Seasons* was greater than 0.80 for all traits, *entry* effects for RUSS, SHAPE, APPSUM, FIRM and TSS were most reactive to seasonal differences (as indicated by the % variance explained by the *G.S* term and the genetic correlation among *Seasons*—*rgS*, Table 4). Proportion of phenotypic variation explained by *entry* effects was moderately correlated (0.59) with *rgS* across traits. The two-way interaction between *entry* and storage *Duration* was not significant for any appearance trait (Table 4); however, *entry* effects for texture traits (FIRM, M1, M2, CRISP, JUIC and CN) exhibited some sensitivity to storage (*G.D* and *rgD* in Table 4) as did TSS and EQ, with EQ, FIRM and M2 being the traits for which *entry* effects were the most reactive. There was also a significant three-way interaction between *entry*, storage *Duration* and *Season* for SWEET and AROM. Most traits exhibited significant *Farm*-by-*Plot* interaction. The genetic correlation among *Farms* (*rgF**, assuming all the variation in this term was due to interaction between *entries* and *Farms*) was lowest for SWEET (0.79), OVERALL (0.82), APPSUM (0.86), TSS (0.88) and FRTD (0.89).

Correlation of entry effects among traits following storage

Sensory traits

The most parsimonious multi-variate model of variation among the 16 sensory traits following 2-month storage was a five-factor FA structure for the *entry* trait variance-covariance model and a six-factor FA structure for the residual trait variance-covariance structure. Rotated loadings of the factors accounted for 67 % of the variation in *entry* effects and 39 % of the residual variance. Loadings on factors explained all of the variation in RUSS, SWEET, APPSUM, EQ and OVERALL; 99 % of variation in CRISP; 90 % of the variation in JUIC; 86 % of the variation in AROM; 76 % of the variation in GCOL; 66 % of the variation in TART; 59 % of the variation in FIRM and 43 % of variation in SIZE, but less than 25 % of the variation in SHAPE, TCOL, PCOL and LENT.

Table 3 Significance (***) indicates $p < 0.001$, ** indicates $0.001 < p < 0.01$, blank cell indicates $p > 0.05$ of fixed effects (*S* season, *A* age, *H* harvest, *D* storage duration, *F* farm) and interactions from the analysis of individual fruit quality traits (*App* appearance, *Text* texture, *Sum* summary) assessed by sensory (*Sens*) or instrumentally assessed (*Instr*) methods as part of the WABP phase 2 trials. Full details of the traits given in the text and Table 2. Interactions A.D and F.A.H and F.A.H.D not included for brevity (see text for results)

Trait	Class	Type	S	A	H	D	F	F.S	F.A	F.H	F.D	S.H	A.H	S.D	H.D	S.H.D	F.S.H	F.A.H	F.S.D	F.A.D	F.H.D	F.S.H.D	
SIZE	App	Sens	***		**	***	***	***	**					***									
FRTD	App	Instr	0.02	0.02	***	***	***	***	0.02		***									**	**	**	**
FRTM	App	Instr	***		***	**	***	***			***		**	***			0.02	0.01	0.01		0.02	0.03	0.03
SHAPE	App	Sens	**		***	***	**	0.01			0.02			***				0.03					
GCOL	App	Sens	***		***	***	***	**	**					***									
TCOL	App	Sens	***		***	***	0.02	**	**					***									
PCOL	App	Sens	***	0.03	***	0.02	***	***		**		***	**	***							0.04		0.04
LENT	App	Sens	***		***	***	***	***	**					***									
RUSS	App	Sens	***	**	**	***	***	0.02	**					***									
FIRM	Text	Sens	***	0.02	***	***	***	***	***			0.02		***			**						
CRISP	Text	Sens	***		***	***	***	***	0.03	0.03	0.04	**		**	0.05								
JUJC	Text	Sens	***		***	***	***	***			0.03	**		0.01	**				**				
M1	Text	Instr	***		***	***	0.03	***			***	***		0.01	**		0.04						
M2	Text	Instr	***		***	***	***	***	**	**	***	**		***	***						0.01		
CN	Text	Instr	***	**	***	***	***	***			***	***		***	***								
SWEET	Taste	Sens	0.02		***	0.04	***	***	0.02			***		**	***				**	**	0.05		
AROM	Taste	Sens	***		***	**	***	***	0.02	**		0.02		***	***				0.01	**			
TSS	Taste	Instr	***	**	***	***	***	***	***		***	0.03		***	***				0.04	0.03			
TART	Taste	Sens	***	**	***	***	***	***	0.02					***	***								
TA	Taste	Instr	***	***	***	***	***	***	**					0.02	0.04								
APPSUM	Sum	Sens	***	0.03	***	***	***	***	0.03					**	0.03				0.01	**			
EQ	Sum	Sens	***		***	***	***	***		0.01	0.04	***		***	***		0.02					0.01	0.02
OVERALL	Sum	Sens	***		**	***	***	***				**		**	***								

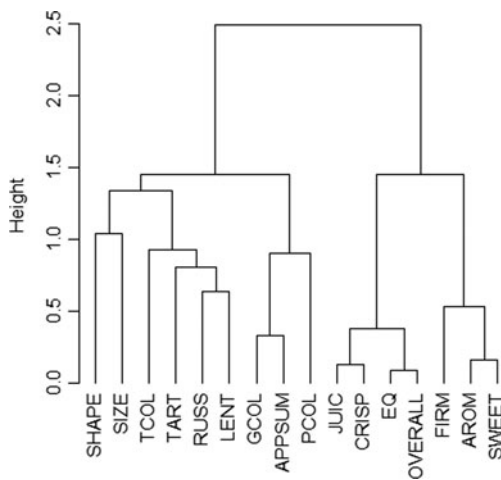


Fig. 1 Dendrogram from a hierarchical cluster analysis of the entry variance-covariance matrix of 16 sensory traits using Ward’s minimum variance strategy for cluster agglomeration. Order of leaf nodes determined so that traits most similar are on the edge of neighbouring clusters

Cluster analysis of the *entry* sensory trait dissimilarity matrix (Fig. 1) and biplot of the first three rotated dimensions (Fig. 2) reflected the patterns in the correlation matrix (Table 5). The first rotated factor discriminated traits based on their correlation with a hypothetical ‘average’ trait. Loadings of the second factor contrasted *entry* effects for GCOL, TART, LENT, APPSUM, RUSS and AROM against those for JUIC, CRISP and SIZE. Loadings of the third factor contrasted *entry* effects for SWEET, AROM and FIRM against those for APPSUM, RUSS and SIZE. The near-zero correlation among *entry* effects for any appearance trait (i.e. SIZE, SHAPE, GCOL, TCOL, PCOL, LENT and RUSS) was evident in the dendrogram as distant branching among these traits (Fig. 1).

Entry effects for sensory texture traits CRISP and JUIC were strongly correlated (Table 6) and clustered tightly in the dendrogram (Fig. 1) with an acute angle between vectors

for these traits in the biplot (Fig. 2). However, there was little genetic correlation between FIRM and CRISP, or FIRM and JUIC, and almost all pairwise genetic correlations among an appearance trait and a texture trait were close to zero, except for a weak negative correlation between SIZE and FIRM. This result appears in the dendrogram as distant branching between the clusters containing these traits and perpendicular vectors in the biplot.

Entry effects for SWEET and AROM were strongly correlated (Table 5) and, similar to the pattern for CRISP and JUIC, were tightly clustered in the dendrogram (Fig. 1) and in the biplot (Fig. 2). *Entry* effects for TART were not correlated with other taste traits. There was a weak positive association between SWEET and FIRM and consequently between AROM and FIRM, shown in a clustering of FIRM with these taste traits and a similar trajectory of vectors in the biplot. Genetic correlations were weakly negative between SIZE and SWEET and between SIZE and AROM.

There was a strong correlation between *entry* effects for EQ and OVERALL (which were also tightly clustered in the dendrogram), a weak correlation among APPSUM and OVERALL and almost no correlation between *entry* effects for APPSUM and EQ (Table 5). *Entry* effects for APPSUM were moderately correlated with GCOL, and weakly with RUSS, but uncorrelated with other traits. *Entry* effects for EQ were strongly correlated with those for CRISP, moderately with those for JUIC texture and AROM taste and weakly with FIRM texture, SWEET taste and GCOL appearance. There was a moderate correlation among *entry* effects for the OVERALL sensory summary trait and those for texture CRISP and JUIC and weakly with AROM taste, GCOL appearance and FIRM texture.

Instrumentally assessed traits

FRTD and FRTM were highly correlated (Table 6). *Entry* effects for FRTD were moderately negatively correlated with

Fig. 2 Biplot of first 3 rotated loadings for 19 sensory fruit quality traits (*aZ* SIZE, *aH* SHAPE, *aG* GCOL, *aP* PCOL, *aT* TCOL, *aL* LENT, *aR* RUSS, *tF* FIRM, *tC* CRISP, *tJ* JUIC, *fS* SWEET, *fA* AROM, *fT* TART, *AQ* APPSUM, *EQ* EQ, *OQ* OVERALL) and scores for 78 selections (+) and 7 commercial cultivars (*br* ‘Braeburn’, *cp* ‘Cripps Pink’, *c15* ‘Coop15’, *fu* ‘Fuji’, *ga* ‘Gala’, *gd* ‘Golden Delicious’, *hc* ‘Honeycrisp’)

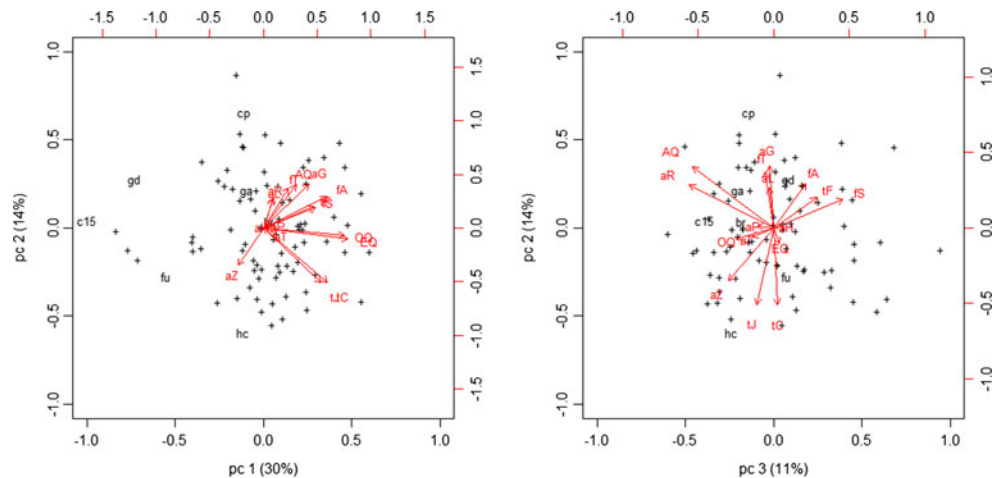


Table 5 Genetic correlation matrix for 16 sensory apple fruit quality traits assessed following short-term storage

	SHAPE	GCOL	TCOL	PCOL	LENT	RUSS	FIRM	CRISP	JUIC	SWEET	AROM	TART	APPSUM	EQ	OVERALL
SIZE	-0.04	-0.31	0.02	0.03	-0.20	0.02	-0.44	0.04	0.13	-0.48	-0.48	-0.31	-0.03	-0.26	-0.10
SHAPE		-0.03	0.00	-0.03	0.02	-0.02	0.03	-0.02	-0.01	0.06	0.03	0.00	-0.08	0.00	-0.05
GCOL			-0.06	0.22	0.08	0.03	0.35	0.06	-0.01	0.45	0.53	0.25	0.66	0.41	0.52
TCOL				-0.03	0.08	0.25	0.05	0.16	0.21	-0.02	0.08	0.09	0.05	0.18	0.18
PCOL					-0.08	-0.06	0.00	0.11	0.06	-0.02	0.01	0.02	0.26	0.11	0.21
LENT						0.36	0.17	-0.16	-0.09	0.16	0.26	0.23	0.11	0.08	0.03
RUSS							-0.05	-0.23	0.05	-0.11	0.18	0.24	0.51	0.06	0.21
FIRM								0.32	0.18	0.53	0.58	0.39	0.04	0.59	0.40
CRISP									0.87	0.22	0.26	0.01	-0.07	0.82	0.75
JUIC										0.24	0.28	-0.13	-0.01	0.74	0.71
SWEET											0.84	-0.03	0.00	0.56	0.35
AROM												0.22	0.26	0.68	0.55
TART													0.27	0.28	0.24
APPSUM														0.24	0.57
EQ															0.91

those for M1, weakly correlated with those for CN and weakly negatively correlated with those for TSS. A similar correlation pattern existed for FRTM. *Entry* effects for M1 and M2 were highly correlated, and there was a weak genetic correlation between M1 and TSS (and FRTM and FRTD as mentioned above). There was only a very weak correlation between *entry* effects for CN and M1 and for CN and M2, and *entry* effects for TA were not correlated with entry effects for any other trait.

Correlation among sensory and instrumentally assessed fruit quality traits

There was a near perfect genetic correlation ($rg=0.99$) between SIZE and FRTD and between SIZE and FRTM. The genetic correlation between FIRM and M1, FIRM and M2 ($rg > 0.9$, Table 7) and TA and TART ($rg=0.97$, Table 7) were also very strong. The genetic correlation between CN and CRISP was moderate ($rg=0.75$), while the genetic correlation between TSS and sensory SWEET was weaker ($rg=0.57$).

Table 6 Estimated correlation matrix among entry effects for seven instrumentally assessed fruit quality traits in phase 2 trials of the Washington Apple Breeding Program

	FRTM	M1	M2	CN	TSS	TA
FRTD	0.98	-0.62	-0.35	0.44	-0.42	-0.15
FRTM		-0.54	-0.27	0.50	-0.40	-0.13
M1			0.82	0.18	0.47	0.26
M2				0.39	0.37	0.23
CN					-0.23	-0.01
TSS						0.17

Performance of commercial cultivars and selections

The biplot of rotated loadings and scores from the multivariate analysis of 19 sensory fruit quality traits suggests that commercial cultivars ‘Cripps Pink’ and ‘Honeycrisp’ contrasted for traits that defined the loadings of second rotated factor as they were on opposite sides of the origin almost parallel with the axis of the second rotated factor. ‘Cripps Pink’ was located along the line of the positive vectors for GCOL, TART and LENT and diagonally opposite CRISP and JUIC, with ‘Honeycrisp’ opposite in the expression of these traits. There was a tendency for fruit from ‘Cripps Pink’ to be smaller and express more intense aromatic taste, with lower levels of LENT and RUSS. ‘Gala’ and ‘Golden Delicious’ showed attributes similar to ‘Cripps Pink’, but with lower intensity, and ‘Fuji’ was more similar to ‘Honeycrisp’. Fruit quality of ‘Braeburn’ was intermediate for most traits. The selections from the WABP covered the full range of possible fruit quality expression. Several selections had higher fruit quality than commercial cultivars examined, particularly

Table 7 Estimated genetic correlations among (a) instrumentally assessed traits M1, M2 and CN and sensory traits FIRM, CRISP and JUIC; and (b) instrumentally assessed traits TSS and TA and sensory traits SWEET, AROM and TART

	FIRM	CRISP	JUIC
(a)			
M1	0.96	0.17	0.01
M2	0.90	0.34	0.20
CN	0.41	0.75	0.71
(b)	AROM	SWEET	TART
TSS	0.56	0.57	0.25
TA	0.05	-0.29	0.97

for GCOL, SWEET, AROM, APPSUM, EQ and OVERALL. In addition, there were some selections for which CRISP were equal to or higher than 'Honeycrisp'.

Discussion

Use of advanced selection trials for estimation of genetic parameters

This study demonstrates the value of data collected as part of an operational breeding program for investigating the genetic architecture of important selection traits in the breeding program. Phenotypic data for fruit quality in tree fruit are expensive to generate and information on the genetic architecture must be relevant for the breeding germplasm and target environment. Because data used in this study were collected directly from breeding germplasm in phase 2 trials, estimates are directly relevant for selection and other investigations of this breeding phase. Similarly, because selections were assessed over multiple locations, including grower sites, across the Washington State apple industry growing region, results presented here are directly relevant for this target environment.

A factor allocation diagram (Brien and Demetrio 2009; Brien et al. 2011) was utilised to develop an appropriate model of convenience, as the sampling design that generated the data was complex. This approach supported clear identification of sources of variation and potential confounding factors so that significant sources of variation could be detected, hypotheses generated for the cause of these sources of variation and alternative designs developed to remove confounding and bias, and to improve accuracy, if required. While the sampling design was unbalanced, unbiased estimates of genetic parameters were obtained because of the mixed model approach adopted (Henderson 1975).

The entry-by-farm effects were confounded with plot-within-farm effects in the sampling design used in this study because only a single experimental unit for each entry was sampled at each farm. Entries were randomised to non-contiguous multi-tree plots within a farm; however, as apples from the multiple trees per plot were combined into a single unit from which fruits are sub-sampled, the experimental unit for entry was not replicated at a single site. Nevertheless, the main effect of entries across farms could be estimated as the samples for each entry were collected from multiple farms. It is important to realise that the multiple sampling of a single experimental unit within a farm is not the replication needed to estimate the effect of an entry as the samples are not independent (Hurlbert 1984). Repeated sampling of the same experimental unit only improves the precision of the estimate of the experimental unit but does not provide any independent replication within a farm from which to separate the entry effect from other within farm non-treatment effects. This 'pseudo-

replication' (Hurlbert 1984) is rife in horticultural literature, where repeated measures are incorrectly used as the error to test differences among entries and does not permit inferences to be made about the effect of treatments.

The design of the field trials used here was not optimum for separating seasonal and age effects, particularly in the interaction with entry. An alternative design that would allow this separation would be replication of entries across series. While age is a repeatable factor, season might be interpreted as a random factor unless environmental covariates can be used to predict the interaction of entries with season.

Many of the sensory traits were assessed using an ordinal rating scale. Such responses might introduce difficulties in interpretation as the increments between categories might not be equal and means of these responses do not have the same meaning as those of continuous responses (Munzel and Bandelow 1998). However, the approximate normal distribution of the residuals from our analysis of these types of traits suggested that they behave as if they were normal, particularly when there are a large number of samples and the central limit theorem applies (Land and Shepard 1987).

While these results are relevant for selections in phase 2 trials, how closely they approximate the genetic architecture of unselected phase 1 seedlings is not clear. If there has been intense directional selection in the phase 1 trials, our estimates of heritability might underestimate that of segregating populations due to the Bulmer effect (Bulmer 1971). Similarly, our estimates of genetic correlations might be biased estimates of those for unselected populations. Nevertheless, our results indicate that genetic variation in phase 2 trials for fruit quality traits is large compared to other sources of variation. In addition, if selection in phase 1 targets multiple directions and the selection trajectory of individual traits differs among these directions, genetic variation might not be as strongly affected compared to selection for a single target.

Comparison of genetic parameter estimates with published literature

This study presents a comprehensive examination of repeatability and interaction of total (i.e. additive and non-additive) genetic effects across locations, seasons, harvests and storage durations for fruit quality traits in apple breeding germplasm. Only a reduced set of factors have been examined in previous studies of genetic architecture of fruit quality in this crop (Abbott et al. 1984; Alspach and Oraguzie 2002; Bavay et al. 2013; Bonany et al. 2013, 2014; Brookfield et al. 2011; Corollaro et al. 2013, 2014a; Costa et al. 2012; Daillant-Spinnler et al. 1996; Dever et al. 1995; Durel et al. 1998; Echeverria et al. 2008; Hampson et al. 2000, 2007, 2009; Hampson and McKenzie 2006; Harker et al. 2006; Karlsen et al. 1999; King et al. 2000; Kouassi et al. 2009; Kuhn and Thybo 2001; Kumar et al. 2010, 2011, 2013; McKay

et al. 2011; Miller et al. 2004, 2005a, 2007; Nybom et al. 2013; Oraguzie et al. 2009; Oraguzie et al. 2001; Seppa et al. 2013). By combining data across these factors, general trends are observable and inflation of sources of variation due to confounding is reduced.

Many of the previous studies use incorrect designs or analyses to estimate and test genetic entry effects on fruit quality traits. Some confound genetic effect with other potential sources of variation. Often there is no independent replication of genetic treatment, and interactions with repeated factors (such as season or samples drawn from a population pooled across other factors) are used as the error term. These design flaws lead to invalid statistical tests and inflated estimates of heritability or repeatability (Hurlbert 1984). Nevertheless, estimates of entry effects from samples pooled across other factors would be accurate if sampling is not biased with respect to these factors, but differences between entries cannot be tested. Correlations among entry effects estimated from pooled samples might also represent unbiased estimates of genetic correlations. In contrast, means presented in Miller (2005b, 2007) were based on multiple replicates of trees of each cultivar at and across locations; however, these estimates might be biased as location is confounded with assessor (see discussion below).

While this study reports genetic parameters for clonally replicated individuals, other studies (Alspach and Oraguzie 2002; Durel et al. 1998; Kouassi et al. 2009; Kumar et al. 2010; Oraguzie et al. 2001) report unbiased estimates of additive genetic parameters from segregating populations. Genetic parameters published in King et al. (2000) were based on clonal replication of full-sibs from a single family across locations and therefore included additive and non-additive effects but do not include genetic variance among families.

Repeatability of genetic effects

Our estimates of a moderate to high repeatability for many apple fruit quality traits concur with the published results of Corollaro et al. (2014a). There was little difference among estimates of narrow-sense heritability reported by Durel et al. (1998) for sensory size (0.33), russetting (0.36), firmness (0.33), juiciness (0.34) and attractiveness (similar to our APPSUM, 0.37) assessed at a single location. Similar to our results that repeatability of appearance traits was generally higher than for texture and taste, estimates of narrow-sense heritability in Kouassi et al. (2009) and Kumar et al. (2010) were relatively higher for proportion of over-colour and fruit size compared to juiciness, sweetness, aromatic flavour and global taste (similar to our EQ).

The higher repeatability of tartness compared to firmness, juiciness, aromatic flavour and sweetness reported here is also generally consistent with narrow-sense heritability estimates in Kouassi et al. (2009), although Kumar et al. (2010) reported

a relatively low heritability (0.19) for tartness. Also, while our results suggest that repeatability of crispness is moderate and comparable to fruit size and tartness, Kouassi et al. (2009) reported that heritability of crispness was considerably lower than for fruit size and tartness. The generally higher repeatability observed here for the instrumentally assessed traits compared to sensory traits agrees with narrow-sense heritability estimates reported in Kouassi et al. (2009).

Our estimates of low repeatability of sweetness, juiciness, aromatic flavour and, to some degree, firmness were mainly a consequence of the high level of residual variation for these traits. Residual variation was due to inconsistency among individual observational units that were not related to differences attributable to season, harvest, or storage effects or specific combinations of these factors. A large residual variation estimate might be due to inconsistency among assessors (Alspach and Oraguzie 2002; Brookfield et al. 2011; Murray et al. 2001; Oraguzie et al. 2009) and/or the inherent variability among fruit for these traits (Brookfield et al. 2011; Dever et al. 1995; Visser and Verhaegh 1978). In other studies and similar to our findings, consistency among sensory panellists was highest for tartness, with sweetness the most variable trait, and juiciness, firmness and crispness intermediate (Corollaro et al. 2014a; Echeverria et al. 2008; Oraguzie et al. 2009). Perception of sensory sweetness might be inhibited or masked by other texture or taste traits (Baldwin et al. 2008; Echeverria et al. 2004, 2008; Harker et al. 2002b, 2006; Visser and Verhaegh 1978), or there might be differences among assessors in their ability to discriminate among highly sweet fruit (Oraguzie et al. 2009). However, fruit variability rather than inconsistency among assessors has been identified as the greater source of variation for sweetness (Bavay et al. 2013; Echeverria et al. 2008) and to some extent acidity (Bavay et al. 2013).

As previously stated, the sampling design of this experiment did not permit formal separation of the interaction between entry and location ($G \times E$) from variation among plots-within-farm. Nevertheless, the relatively small magnitude of the plot-within-farm variance suggests that the ranking of entries was stable across farms, although the relatively large interaction of plot-within-farm with seasons might indicate that interaction between entries and farms occurs under specific seasonal conditions. More definitive results would be produced if analyses were based on data with multiple independent samples of entries within a farm.

Similar to our findings, no, or only slightly significant, $G \times E$ for russetting, crispness, juiciness and titratable acidity for 23 cultivars planted across 12 diverse North American locations was reported in Miller et al. (2004, 2005a, b). Low $G \times E$ for russetting was also reported in Alspach and Oraguzie (2002), although significant $G \times E$ for crispness, juiciness and acidity was reported by others (Alspach and Oraguzie 2002; Miller et al. 2007). So too, Miller et al.

(2004, 2005) reported significant $G \times E$ for fruit diameter and weight, general appearance and overall acceptance (in contrast with our results), although Alspach and Oraguzie (2002) found negligible $G \times E$ in fruit weight (similar to our findings).

While we did not find evidence for $G \times E$ for over-colour, firmness and tartness, significant $G \times E$ was reported for these traits in Miller et al. (2004, 2005b). We detected a large plot-within-farm effect for sweetness and total soluble solids, possibly indicating the presence of $G \times E$; however, no significant $G \times E$ for these traits was reported in Miller et al. (2004, 2005b), although significant $G \times E$ was reported in a separate study (Miller et al. 2007). Hence, it is difficult to determine if $G \times E$ is apparent for these traits or if their within-farm variability was greater compared to other traits.

Differences in the magnitude of environmental variation and diversity of germplasm might explain some of the inconsistency between our results and those cited. Five locations within the Washington State apple industry production region were studied, compared to a broader range of environments by Miller et al. (2004, 2005b, 2007). Alternatively, as different assessors were used at each location in Miller et al. (2004, 2005b, 2007), significant $G \times E$ reported in those studies might be due to different perceptions or preferences of the sensory trait among assessors (Murray et al. 2001). In our study, the same assessors were generally used for assessment of all sample units across locations and seasons. $G \times E$ detected by Alspach and Oraguzie (2002) might be a consequence of more diverse germplasm that includes interspecific hybrids and unimproved seedlings.

Despite collating data from eight locations, Kouassi et al. (2009) assumed common additive genetic and residual variances across locations and hence did not investigate $G \times E$. Thus, if $G \times E$ was present, the residual would be inflated and heritability underestimated. While King et al. (2000) report $G \times E$ for some fruit quality traits in apple, they used an incorrect residual structure.

Total soluble solids, russetting, hardness and fruit shape were traits for which mean entry effects appeared to be most sensitive to seasonal variation, with entry effects across farms, years, harvests and storage durations for sweetness, aromatic flavour and tartness being the most stable. However, significant third-order interactions suggest that entry effects for these taste traits might be sensitive to specific season-by-storage duration effects and possibly with farm-by-seasons effects. Care is required when trying to identify the cause of these interactions, firstly, because age is confounded with season in the interaction with entry. In addition, the effect of 'season' in our analysis included all factors that might vary among years, including possible farm management or post-harvest conditions.

Alspach and Oraguzie (2002) reported lower stability among seasons for fruit quality traits than what we found. However, family effects for fruit weight and proportion of

over-colour were more highly correlated across sampling years than family effects for russetting and firmness. In contrast, however, correlations of family effects across seasons for juiciness, sweetness, tartness and crispness were low (0.3). These observations suggest that assessor variation among years might be the cause of the lower genetic correlations among years for taste and texture traits. McKay et al. (2011) reported no significant interaction between season and family effects for sensory crispness, juiciness and firmness, although this factor was significant for instrumentally assessed firmness. Correlations between 15 cultivar means assessed over 2 years for sugar content and malic acid content Brown and Harvey (1971) suggested that cultivar means for acid content were more stable ($r=0.85$) than sugar content (0.61).

Despite a significant general effect of storage on most fruit quality traits, the relative performance of entries was only sensitive to storage duration for texture traits, similar to the findings of others (Kouassi et al. 2009; Kuhn and Thybo 2001; Seppa et al. 2013). Costa et al. (2012) defined three groups of cultivars based on response of texture following storage: (i) declining, (ii) stable and (iii) slightly improving. The significant three-way interaction of entry-by-season-by-storage for sweetness and aromatic flavour indicates that sensitivity of entries to storage duration was inconsistent among seasons for these traits.

Storage life of apples varies among cultivars/selections and is dependent on many factors, e.g. accurate determination of optimum harvest date and rate of respiration (Jackson 2003). As climacteric fruit, ripening of apple is accelerated by the presence of ethylene; cultivars and selections with low ethylene production rates generally have more potential for longer storage duration (Sunako et al. 1999). Rate of fruit softening has been linked to the presence of specific alleles of the *1-aminocyclopropane-1-carboxylic acid synthase* gene in the ethylene biosynthesis pathway (Costa et al. 2005; Oraguzie et al. 2004; Zhu and Barritt 2008) and the *polygalacturonase* gene, a cell-wall degrading enzyme (Longhi et al. 2013).

It could be expected that traits that exhibit high genotype-by-environment interaction might also exhibit lower repeatability of mean genetic effects (e.g. Gitonga et al. 2014), simply due to numerical constraints of partitioning phenotypic variation. However, this hypothesis is not supported by this study, which found a negative correlation between the proportion of phenotypic variation explained by mean genetic effects and the magnitude of the genetic correlation between storage conditions and the limited extent of potential genetic-based interaction among locations.

Genetic correlations among fruit quality traits

The low genetic correlations among the sensory fruit quality traits examined here indicate that most traits under consideration are influenced by different non-pleiotropic loci that

either are not linked and have not been selected intensely in the same direction. Inconsistencies between estimates reported here and elsewhere might be due to differences in the germplasm studied, particularly as cultivars are selected to improve multiple traits.

Published values for genetic correlations among fruit appearance traits in apple (Alspach and Oraguzie 2002; Durel et al. 1998; Kouassi et al. 2009; Oraguzie et al. 2001) generally reflect our estimates, except for a moderate negative genetic correlation of -0.56 between proportion for red over-colour and russetting (Oraguzie et al. 2001), which we and Alspach and Oraguzie (2002) report as close to zero. The weak negative genetic correlation between fruit size and fruit firmness is well known and most likely due to the more densely packed cells present in small apples (Pitts et al. 1997).

Estimates of the genetic correlations among sensory taste and texture traits are inconsistent among studies. The high genetic correlation between sensory crispness and juiciness reported here is similar to other studies (Brookfield et al. 2011; Kouassi et al. 2009; Kuhn and Thybo 2001; Kumar et al. 2010), except for Alspach and Oraguzie (2002) who found a weak correlation (0.31). A direct link between the release of juice as the cell wall fractures, typical in a crisp apple, has been described (Echeverria et al. 2008; Harker et al. 2006).

The low genetic correlation we found between firmness and crispness, while supported by Alspach and Oraguzie (2002) and Miller et al. (2005b), contrasts with the results from others (Kouassi et al. 2009; Kuhn and Thybo 2001; Kumar et al. 2010) who report strong (>0.9) genetic correlation. Miller et al. (2005b) suggest that a low genetic correlation indicates that a fruit with a firm texture is not necessarily crisp. Differences might be due to the characteristics of the germplasm. Much of the material in this study was derived from 'Honeycrisp', which is known not only for its particular crisp texture but also its relatively low firmness (Bedford 2001; McKay et al. 2011).

Our estimate of genetic correlation close to zero between sensory tartness and sweetness agrees with Alspach and Oraguzie (2002) and that of Brown and Harvey (1971) for malic acid and percentage sugar. However, others (Kouassi et al. 2009; Kuhn and Thybo 2001; Kumar et al. 2010) have reported a moderate to large negative genetic correlation between these traits (Kouassi et al. 2009; Kuhn and Thybo 2001; Kumar et al. 2010). A low genetic correlation between tartness and sweetness is supported by recent molecular genetics studies. Genetic control of acidity in apple has been ascribed to the *Ma* locus (Maliepaard et al. 1998) on linkage group (LG) 16 with a second locus on LG 8 (Liebhard et al. 2003; Kenis et al. 2008). Several loci influencing sugar content in apple have recently been identified (Guan et al. 2015), but only one was found on LG 16 and was small and not repeatable over years.

The presence of little genetic correlation between juiciness and sweetness previously reported (Alspach and Oraguzie 2002; Echeverria et al. 2008; Kouassi et al. 2009) agrees with our results; however, others (Kumar et al. 2010) reported a high correlation. Kumar et al. (2010) also reported a moderate positive genetic correlation between sweetness and firmness (in agreement with our results), while other studies (Alspach and Oraguzie 2002; Kouassi et al. 2009; Kuhn and Thybo 2001) identified no correlation. Echeverria et al. (2008) suggested that low correlation of sweetness with other traits might be due in part to variability in the articulation of sweetness among assessors. Our finding of a strong genetic correlation between sweetness and aromatic flavour is supported by Kuhn and Thybo (2001). Studies in tomato (Baldwin et al. 2008) have shown that the presence of added sugars in tomato puree enhances the perception of other flavour attributes. This phenomenon might explain our results that indicated a strong genetic correlation between sensory sweetness and aromatic flavour if elevated levels of sweetness in apple enhance the perception of aromatic flavour. Alternatively, the high genetic correlation might be a consequence of pleiotropic effects of genes influencing these two traits, possibly due to large effect of the genes, parallel selection, linkage or a consistent bias with respect to maturity across harvest when sampling fruit from multiple cultivars.

Genetic correlations between component sensory traits and quality summary ratings of fruit quality might indicate the importance to selection of individual traits. Traits that are weakly correlated with summary ratings might either not be important in selection decisions or only be important for some targets if multiple targets are being pursued. The moderate genetic correlation between appearance quality and ground colour of the fruit and russetting, and virtually no genetic correlation with any other appearance trait, suggests that overall appearance quality is determined predominantly by variation in these two traits. In contrast, Kouassi et al. (2009) reported a low genetic correlation between attractiveness and ground colour and a moderate correlation between fruit size and attractiveness, which was also reported in Durel et al. (1998). Assessors in those other studies evidently considered other visual cues than ground colour to be important contributors to attractive appearance or else did not have similar genetic variation for ground colour to the WABP's phase 2 germplasm.

Moderate positive genetic correlations of sensory crispness, juiciness, aromatic flavours, firmness and sweetness with hedonic eating quality suggest that these traits are the major determinants of selection for eating quality in this program. Similarly, Kouassi et al. (2009) reported moderate estimates (0.61–0.71) of additive genetic correlation between sensory crispness and global taste preference and between juiciness and global taste preference after 2 months of storage; however, that study also reported virtually no

genetic correlation between global taste and sweetness, in contrast with our results. Our results are in agreement with Hampson et al. (2000) who suggested that firmness is redundant in explaining consumer preference as it does not explain any more variation than crispness and that sweetness, aromatics and sourness contributed equally to variation in flavour preference. Sweetness, firmness, juiciness and acidity are major traits influencing consumer preference (Bonany et al. 2014; Dailliant-Spinnler et al. 1996), but there are two major preference segments among consumers: one preferring sweeter, crisper apples and the other preferring juicy, acidic apples. The lower importance of acidity in our results might be due in part to the complexity of the interaction between sweetness and acidity influencing flavour preferences in apple (Nybom 1959).

The strong genetic correlation between overall quality and eating quality, and the lower correlation with appearance summary, suggests that eating quality is the major driver of directional selection at the phase of the breeding program studied here. Clearly, correlations of sensory traits with hedonic quality ratings depend on preferences of the assessors and targets of the selection program.

The lack of a strong genetic correlation among instrumentally assessed firmness, TSS and TA agrees with published results (Corollaro et al. 2014a; Kouassi et al. 2009). The high genetic correlations between sensory firmness and puncture force and between TA and sensory acidity, and lower genetic correlation between TSS and sensory sweetness, are not unexpected (Evans et al. 2010; Harker et al. 2002a, b; Kouassi et al. 2009; Miller et al. 2005b; Oraguzie et al. 2009). Low correlation between TSS and sensory sweetness might be due to lower repeatability of sensory sweetness (as discussed above), or sensory sweetness might be a more complex character than can be described by TSS. In contrast to our strong genetic correlation between measure of puncture force and firmness, several papers (Chauvin et al. 2010; Harker et al. 2002a) have demonstrated that humans might be more sensitive to variation in firmness than can be detected by penetrometers when fruits are soft. In this study, few fruits with a very soft texture were observed.

While other studies (Evans et al. 2010) have demonstrated a high phenotypic correlation between the DigiTest parameter CN and sensory crispness, results here indicate a weaker genetic correlation, which might be due to the lower repeatability of CN or difference in germplasm considered. Others (Brookfield et al. 2011) reported a stronger correlation (0.7) between puncture force and crispness than here (0.17, 0.34).

Performance of commercial cultivars

This study provides unbiased estimates of contrasting performance among several commercial cultivars using extensive sampling across locations, seasons and harvests which has

not been undertaken previously. On average and after confounding effects were removed, ‘Cripps Pink’ was predicted to produce small attractive fruit with yellow ground colour and low incidence of russetting and lenticels, low crispness and juiciness and tart flavour in the Washington State production environment. Fruit produced by ‘Honeycrisp’ were predicted to be less attractive and larger, tending to have greener ground colour with higher incidence of russetting and lenticels, crisper and juicier, but with a slightly less firm texture and low tartness. Fruit from ‘Fuji’ were predicted to be similar to ‘Honeycrisp’ but slightly smaller, with less crispness and juiciness. ‘Gala’ and ‘Golden Delicious’ were predicted to be more similar to ‘Cripps Pink’ than fruit of ‘Braeburn’, ‘Coop15’, ‘Fuji’ and ‘Honeycrisp’. In addition, fruit from ‘Golden Delicious’ were predicted to be firmer and more aromatic in taste than the other cultivars examined, with ‘Gala’ less so.

The relative performance of ‘Cripps Pink’ agrees with that described in Miller et al. (2005b) and Miller et al. (2007) who reported that it tends to produce smaller and more acidic fruit with low russetting relative to other cultivars. Aprea et al. (2012) defined five groups of cultivars based on flavour attributes. In their analysis, ‘Braeburn’ was at the centre of the variation, similar to the pattern of our results, but ‘Cripps Pink’, ‘Fuji’ and ‘Golden Delicious’ were more tightly clustered than we observed, perhaps because a greater range of traits included in this study.

Implications for genetic improvement

The general agreement of our results with those from other studies suggests that the genetic architecture of traits described here are applicable to other programs. On the other hand, correlations between sensory traits and hedonic ratings might be specific to individual programs.

The lack of any substantial genotype-by-storage duration interaction for most fruit quality traits except firmness suggests that selection could be made at either storage duration (i.e. at harvest or after 2 months). The Washington State apple industry relies on long-term (>2 months) storage to maintain out-of-season fruit supply (Fellman et al. 2003). Kouassi et al. (2009) reported that genetic correlations between fruit quality traits at 2 and 4 months were higher than between fresh and 2 months. This, along with our results, suggests that assessment of fruit at 2 months, rather than both at harvest and after 2 months, is an efficient method for identifying elite germplasm for the Washington apple industry, although consideration of the effect of storage on other traits not examined here might be required.

The absence of evidence of large $G \times E$ for most traits suggests that the apple production area of Washington State can be treated as a single target environment for new cultivar selection, assuming that locations used in this study are

representative. However, significant main effects of location, season and harvest indicate that fruit will be variable across this region. It is difficult to predict how new cultivars selected using data from trials examined here will perform in production environments beyond Washington State. While significant $G \times E$ was reported in some studies (Miller et al. 2005b, 2007), there is uncertainty if this interaction has a biological cause (see above) or significance (Alspach and Oraguzie 2002; Kumar et al. 2010). Given that apple is a perennial crop, quantitative information and understanding of patterns in $G \times E$ across current production zones will facilitate more efficient selection and deployment of apple cultivars to these target environments.

The high diversity of fruit quality traits among the selections from phase 1, and the superior performance of some of these selections relative to commercial cultivars, indicates that potential exists for new elite cultivars to be delivered through selection in the WABP. High repeatability of the fruit quality traits and the lack of strong antagonistic genetic correlations among them suggest that attributes can be efficiently combined to generate new cultivars producing highly desirable fruit.

The high residual variation observed in this study suggests that prediction accuracy might be improved through sampling designs that reduce this source of variation. If among-fruit variability is the primary source of this variation, increasing the number of fruit per sample might improve consistency among observational units. Nevertheless, our study employed five fruit per unit and others have suggested that four (McKay et al. 2011) or two (Alspach and Oraguzie 2002) fruit are sufficient. On the other hand, the influence of residual variation on accuracy of entry effects in this design might not be large as observations were averaged across seasons, locations and harvests.

Our study also supports the use of a small team of expert assessors to evaluate sensory fruit quality, particularly as each unit was generally evaluated by the same multiple assessors across locations, seasons and harvests to reduce among-assessor variation, and the variability among units was averaged out by using multiple units to estimate entry means. Brookfield et al. (2011) concluded that an expert team of three assessors was more discriminating than instrument-based measures of fruit quality. Randomisation is required to remove bias (Piepho et al. 2013), and sources of variation among assessors that are confounded with particular factors (e.g. season) can be removed by including these as an effect in the analysis. Pre-evaluation training of assessors (Hampson et al. 2000; Oraguzie et al. 2009) was employed in this study to maximise consistency among assessors. In addition, assessment of traits on a neutral scale reduces confounding of the sensory perception of assessors with their preferences that would tend to inflate inconsistency (Corollaro et al. 2014b; Meilgaard et al. 1999) and reduce accuracy.

The design of assessment methods is a compromise between cost and accuracy of identifying elite individuals (Brennan et al. 1998). The benefit of reduced costs from reducing sampling intensity should be balanced against increased opportunity cost of failing to identify truly elite individuals. The lack of evidence for a strong interaction between entry and location, seasons and harvests for any fruit quality trait suggests that less intensive assessment can be used to predict entry performance with little reduction in accuracy. For example, Kumar et al. (2010) suggest that breeding at two locations and testing advanced selections at multiple locations are appropriate given the low $G \times E$ observed for most apple fruit quality traits. However, the significant main effects of location and season indicate that it is important to control for these factors rather than use sampling designs (e.g. Bonany et al. 2014) where genetic effects are confounded with any of these factors. The mixed model approach (Henderson 1975) used in this study is the method of choice to account for unbalance of entries across seasons (and ages). While assessment of fruit quality from multiple locations by a team of assessors might be expensive, unbalanced or split plot designs can be used to improve efficiency (Forde et al. 2007).

Our results of very high (>0.95) correlation of fruit diameter (and mass) with sensory fruit size, M1 with firmness, and TA with tartness, suggest that the instrument-based assessment methods are measuring the corresponding sensory trait and might be efficient for evaluating genetic potential of entries for fruit size and sensory firmness and tartness. Others (Chauvin et al. 2010; Harker et al. 2002a) suggested that puncture force might not accurately evaluate sensory firmness of soft fruit. In contrast, our results suggest that entry effects for sensory sweetness are not well predicted by TA or TSS, in agreement with Hampson et al. (2000). Thus, direct sensory evaluation of sweetness is required for accurate selection of this trait in phase 2 of the WABP and cannot be reliably replaced with an instrumental measure.

Estimates of genetic architecture as presented here can be used to evaluate efficiency of genetic improvement programs. A detailed analysis of breeding efficiency is beyond the scope of this study but would ideally include consideration of trait assessment costs and consider alternatives that fit within constraints of the operational program.

Conclusions

This study demonstrated the value of formal analysis of data collected on fruit quality from advanced selection trials in an apple breeding program. While the sampling

design was relatively complex and some fixed effects were confounded, replication of experimental units across locations supported unbiased estimation of genetic and non-genetic effects, their interactions and genetic correlations among traits. Genetic parameters reported here generally agree with other published literature. In general, genetic effects for fruit quality traits appeared quite stable over the locations, seasons, harvests and storage durations examined. Genetic architecture of fruit quality trait parameters suggested that considerable genetic gain to generate new cultivars that produce highly desirable fruit is possible through directional selection. The genetic parameter estimates can be used to evaluate alternative trial designs. While variation among units was large, its effect on prediction accuracy of entry effects is likely to be negligible due to multiple sampling across locations, season and harvests. Further research is required to compare costs of assessment with potential genetic gains to optimise investment in genetic improvement.

Acknowledgments The Washington Tree Fruit Research Commission funded in part the Washington State University Apple Breeding Program and the analysis of the data collected as part of the program. We also appreciate the assistance of trial collaborators and technical staff that contributed to the collection of the extensive data used in this study.

Data archiving statement The phenotypic data has been deposited in the Genomic Database for Rosaceae (GDR: <https://www.rosaceae.org>) with the accession reference tfGDR1024.

Appendix

Experimental design

A factor allocation diagram (following Brien and Bailey 2006, Brien and Demetrio 2009 and Brien et al. 2011) was developed to describe the experimental design from which the data was generated and aid in the development of the linear mixed model for the analysis of individual and multiple fruit quality traits. In these diagrams, both the treatment factors and experimental units of the experimental design were displayed as belonging to objects. Panels were used to group factors associated with objects into the tiers and display the nesting and crossing structures of factors within a tier. Tiers are defined (Brien 1983) as a set of factors for which all levels of these factors are observable. Panels were connected by lines to display the allocation of the factors in a tier to experimental units in a second tier. Following Brien et al. (2011), solid lines indicate random allocation and dashed lines were used to represent systematic allocation. In the following description, the names of stages are capitalised and italicised, the names of the factors are italicised, the letter in the factor name used as the abbreviation in the development of the linear model is capitalised and the first mention of a factor is bold.

The experiment design is a four-stage design that requires five tiers for description (Fig. 3). The first tier (*PLANTS*) was composed of *bgt* trees described by *t* *Trees* of *g* *entries* (*G*) in *b* *sEries*.

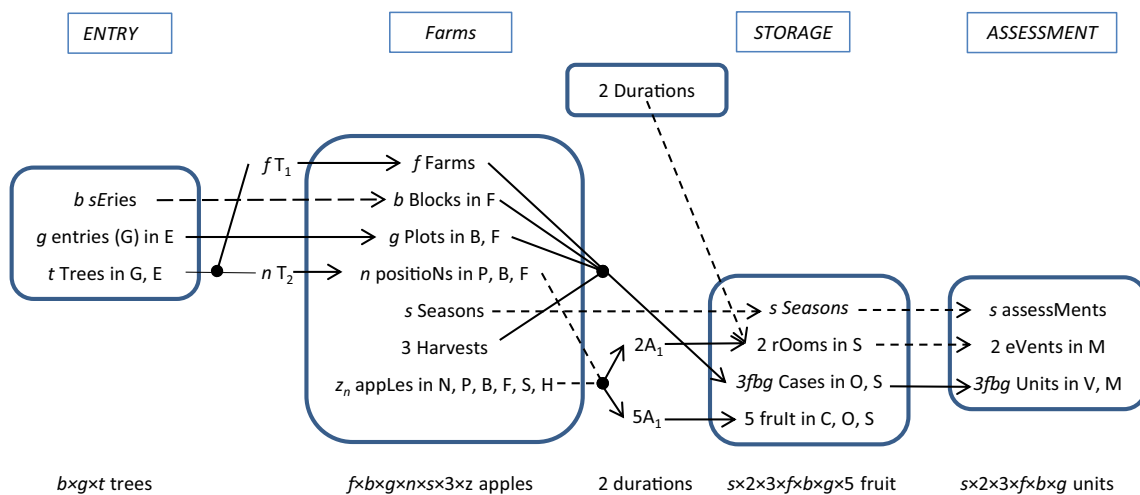


Fig. 3 Factor allocation diagram for the evaluation of apple fruit quality traits assessed from the phase 2 trials in the Washington State University Apple Breeding Program

The second tier (*FIELD SAMPLE*) involved the allocation of the factors describing trees to those describing apples. It contained *fbgns3z* apples described by *n* *positionNs*

in *g* *Plots* in *b* *Blocks* in *f* *Farms*. The *sEries* in tier 1 were systematically allocated to the *b* *Blocks* within each of the *f* *Farms* in tier 2. The *g* *entries* within *sEries* in tier 1 were

randomly allocated to the g *Plots* within *Blocks* in tier 2. Thus, there was no replication of *entries* across *Plots* within a *Block*. In addition, there was little replication of *entries* across *sEries* as there was practically no replication of *entries* across *Blocks* within a *Farm*. To fully describe the randomisation of *Trees* in tier 1 to *positioNs* in tier 2, two pseudo-factors (Yates 1936; Brien and Bailey 2006) were created. Pseudo-factors define random grouping of subsets of objects to support randomisation to other objects. The pseudo-factor T_1 indexed the random allocation of *Trees* within *entry* to *Farms*, and the pseudo-factor T_2 indexed the random allocation of *Trees* within *entry* to *positioNs* within *Plots*.

A number (z_n) of *appLes* were collected from each of the *Trees* at the n *positioNs* within a *Plot* for 3 *Harvests* (early, middle, late) over s *Seasons* (Fig. 3). Timing of first *Harvest* of all *positioNs* within a *Plot* in each *Season* was determined by weekly visual assessment of maturity, and all apples considered mature at the time of each *Harvest* were collected. Hence, *Harvest* has the same meaning across *Seasons*.

The third tier (*STORAGE*) involved the allocation of the factors describing apples to those describing fruit. It is based on *s6fbg5* fruit described by four factors, *fruIt* within *Cases* within storage *Room* within *Seasons*. All *appLes* collected in each s *Seasons* in tier 2 were systematically allocated to each of the s *Seasons* in the third tier. Within each *Season*, *appLes* were bulked across the n *Trees* at the n *positioNs* within each *Plot* by *Block* by *Farm* in tier 2 and randomly allocated to the *3fbg Cases* in the third tier. To describe the allocation of *appLes* in tier 2 to the 2 storage *Rooms* and 5 *fruIt* within each *Case* in the third tier, two pseudo-factors were created; A_1 that indexes the random allocation of *appLes* to the *Rooms*, and A_2 that indexes the random allocation of *appLes* to the 5 *fruIt* within *Cases*. At each *Season* in the third tier, the treatment storage *Duration* from the *DURATIONS* (fourth) tier was applied to the storage *Rooms*.

The fifth tier (*EVALUATION*) describes the structure of assessment of the *Cases* from the third tier to the units in the fifth tier. All *Cases* within each *Season* in the third tier were systematically allocated to the s *assessMents* in tier 5 and all the *3fbg Cases* within each storage *Room* in a *Season* in tier 3 were systematically allocated to an *eVent* within an *assessMent* in tier 5. Within each *eVent*, the *Cases* were randomly allocated to the *3fbg Unit* for evaluation. Thus, individual apples were not evaluated.

Linear model for individual trait analysis

The modified version of the method outlined by Brien and Bailey (2006) and Brien and Demetrio (2009) for the

development of a mixed linear model form of a factor allocation diagram was followed for the analysis of this multi-tiered experiment. The initial activity was to include all identified sources of variation in the model, then reduce the model to an estimable model of convenience by removing confounded terms.

1. Determine intra-tier model composed of intra-tier random (IR) and intra-tier fixed models (IF).

The factor allocation diagram presented and described above identifies the objects (treatments and observational unit) and tiers of the experiment (see 1a and 1b in Brien and Demetrio 2009). The observational unit is the collective 5 fruit within each of the $f*s*e*y*3*2$ *Units*. The intra-tier formulas were formed by collecting all terms in the tiers to which no factor was randomised in the intra-tier fixed model (IF) and all terms in the other tiers in the intra-tier random model (IR):

$$IR : [(F/B/P/N)*S_2*H]/L + S_3/O/C/I + M/V/U$$

$$IF : E/G/T + D$$

where the letters are the abbreviations defined in the description of the experimental design and Fig. 1.

2. Convert the intra-tier models to homogenous (HR) random and fixed (F) models.

Preliminary HR and F models were defined by setting the IR model to the HR model and the IF model to the F model. These were expanded by adding inter-tier interactions (2a in Brien and Demetrio 2009) E/G with F , S_2 and H , and E/G with D , and $[(F/B/P/N)*S_2*H]$ with D .

$$HR' : [(F/B/P/N)*S_2*H]/L + S_3/O/C/I + M/V/U + [(F/B/P/N)*S_2*H]^D$$

$$F' : E/G/T + (E/G//T) \wedge [(F*S_2*H)] + (E/G/T) \wedge D$$

Next, the HR and F models were augmented with the additional factor *Age* that was not taken into account in the initial factor allocation diagram (2b in Brien and Demetrio 2009). For a single series, *Age* is completely confounded with *Season*; however, as different *sEries* were planted at different years, the effect of *Age* and *Season* can be separated for some terms. Only *Age* and the interactions with F , D and H were added as S_2^A and S_3^A were confounded with E , and $F^A B^A$ is confounded with $F^A B$:

$$HR'' : [(F/B/P/N)*S_2*H]/L + S_3/O/C/I + M/V/U + [(F/B/P/N)*S_2*H]^D$$

$$F'' : E/G/T + D + (E/G/T)^{[(F*S_2*H)]} + (E/G/T)^D + A/(F*H*D)$$

Finally, the factors in the model were then designated as fixed or random and the terms swapped between the HR and F models so that only fixed terms appear in the F model and random terms appeared in the HR model (2c and 2d, in

Brien and Demetrio 2009). The factors F, H, S_2, A, S_3, M and D were taken to be fixed and the remaining terms random. After interchanging the appropriate terms, the full model becomes:

$$HR : \left[\left(\frac{F \wedge B/P/N}{S_2 * H} \right) / L + S_3 \wedge O/C/I + M \wedge V/U + E/G/T + (E/G/T)^{[(F * S_2 * H)]} + (E/G/T)^D \right. \\ \left. + \left[(F \wedge B/P/N)^{(S_2 * H)} \right]^D \right]$$

$$F : D * F * S_2 * H + S_3 + M + A / (F * H * D)$$

3. Construct random and fixed models of convenience (CR and CF) by removing terms that were confounded (3c in Brien and Demetrio 2009).

The HR and F model presented above are over-parameterised for the available data. Here we focus on achieving an estimable model. Firstly, preliminary CR and CF models were set to the HR and F model derived above. The models were then fully expanded. The terms $E \wedge G \wedge T, F \wedge B \wedge P \wedge N, [(F \wedge B \wedge P \wedge N)^{(S_2 \wedge H)]/L$ and $S_3 \wedge O \wedge C \wedge I$ and interactions with these terms were removed as there is no data to estimate these.

The term $S_3 \wedge O/C$ was removed as it is confounded with $M \wedge V/U$. The terms S_3 and M were removed from the fixed model as they are confounded with S_2 . The term $F \wedge B \wedge P \wedge S_2 \wedge H \wedge D$ was removed as it is confounded with $M \wedge V \wedge U$. The term $M \wedge V$ was removed as it is confounded with $Y_2 \wedge D$. All terms involving $E \wedge F$ and $E \wedge G \wedge F$ were removed as they are confounded with $F \wedge B \wedge P$. Finally, to simplify the model, E and $F \wedge B$ were removed from the model recognising that if differences exist among these terms, removing E will inflate the variance among entries (G) and removing $F \wedge B$ will inflate the variance among $F \wedge B \wedge P$. The final models of convenience were thus:

$$CR : G/(D*S*H) + F \wedge P/(S*H) + F \wedge P \wedge D/(S + H) + M \wedge V \wedge U$$

$$CF : F*(S + A)*H*D$$

References

Abbott JA, Watada AE, Massie DR (1984) Sensory and instrument measurement of apple texture. *J Am Soc Hortic Sci* 109:221–228

Alspach PA, Oraguzie N (2002) Estimation of genetic parameters of apple (*Malus domestica*) fruit quality from open-pollinated families. *N Z J Crop Hortic Sci* 30:219–228

Aprea E, Corollaro ML, Betta E, Endrizzi I, Dematte ML, Biasioli F, Gasperi F (2012) Sensory and instrumental profiling of 18 apple cultivars to investigate the relation between perceived quality and odour and flavour. *Food Res Int* 49:677–686

Baldwin EA, Goodner K, Plotto A (2008) Interaction of volatiles, sugars, and acids on perception of tomato aroma and flavor descriptors. *J Food Sci* 73:S294–S307

Bavay C, Symoneaux R, Maitre I, Kuznetsoya A, Brockhoff PB, Mehinagic E (2013) Importance of fruit variability in the assessment of apple quality by sensory evaluation. *Postharvest Biol Technol* 77:67–74

Bedford D (2001) Honeycrisp. *Compact Fruit Tree* 34:98–99

Bonany J et al (2013) Consumer eating quality acceptance of new apple varieties in different European countries. *Food Qual Prefer* 30:250–259

Bonany J et al (2014) Preference mapping of apple varieties in Europe. *Food Qual Prefer* 32:317–329

Brennan JP, Kingwell RS, Thomson FM, Cullis BR (1998) Optimal strategies for regional cultivar testing. *Aust J Agr Resour Econ* 42:209–226

- Brien CJ (1983) Analysis of variance tables based on experimental structure. *Biometrics* 39:53–59
- Brien CJ, Bailey RA (2006) Multiple randomizations. *J R Stat Soc Ser B-Stat Methodol* 68:571–599
- Brien CJ, Demetrio CGB (2009) Formulating mixed models for experiments, including longitudinal experiments. *J Agric Biol Environ Stat* 14:253–280
- Brien CJ, Harch BD, Correll RL, Bailey RA (2011) Multiphase experiments with at least one later laboratory phase I. Orthogonal designs. *J Agric Biol Environ Stat* 16:422–450
- Brookfield PL, Nicoll S, Gunson FA, Harker FR, Wohlers M (2011) Sensory evaluation by small postharvest teams and the relationship with instrumental measurements of apple texture. *Postharvest Biol Technol* 59:179–186
- Brown AG, Harvey DM (1971) Nature and inheritance of sweetness and acidity in cultivated apple. *Euphytica* 20:68
- Bulmer MG (1971) Effect of selection on genetic variability. *Am Nat* 105:201–211
- Chauvin MA, Ross CF, Pitts M, Kupferman E, Swanson B (2010) Relationship between instrumental and sensory determination of apple and pear texture. *J Food Qual* 33:181–198
- Corollaro ML et al (2013) Sensory profiling of apple: methodological aspects, cultivar characterisation and postharvest changes. *Postharvest Biol Technol* 77:111–120
- Corollaro ML et al (2014a) A combined sensory-instrumental tool for apple quality evaluation. *Postharvest Biol Technol* 96:135–144
- Corollaro ML, Gasperi F, Grappadelli LC (2014b) An overview of sensory quality of apple fruit. *Journal of the American Pomological Society* 68:141–157
- Costa F et al (2005) Role of the genes Md-ACO1 and Md-ACS1 in ethylene production and shelf life of apple (*Malus domestica* Borkh.). *Euphytica* 141:181–190
- Costa F et al (2012) Texture dynamics during postharvest cold storage ripening in apple (*Malus x domestica* Borkh.). *Postharvest Biol Technol* 69:54–63
- Dailliant-Spinnler B, MacFie HJH, Beyts PK, Hedderley D (1996) Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the southern hemisphere. *Food Qual Prefer* 7:113–126
- DeLacy IH, Basford KE, Cooper M, Bull JK, McLaren CG (1996) Analysis of multi-environment trials: an historical perspective. In: Cooper M, Hammer GL (eds) *Plant adaptation and crop improvement*. CAB International, Wallingford, pp 39–124
- Dever MC, Cliff MA, Hall JW (1995) Analysis of variation and multivariate relationships among analytical and sensory characteristics in whole apple evaluation. *J Sci Food Agric* 69:329–338
- Durel CE, Laurens F, Fouillet A, Lespinasse Y (1998) Utilization of pedigree information to estimate genetic parameters from large unbalanced data sets in apple. *Theor Appl Genet* 96:1077–1085
- Echeverria G, Lara I, Fuentes T, Lopez ML, Graell J, Puy J (2004) Assessment of relationships between sensory and instrumental quality of controlled-atmosphere-stored ‘Fuji’ apples by multivariate analysis. *J Food Sci* 69:S368–S375
- Echeverria G, Graell J, Lara I, Lopez ML, Puy J (2008) Panel consonance in the sensory evaluation of apple attributes: influence of mealiness on sweetness perception. *J Sens Stud* 23:656–670
- Evans K (2013) Apple breeding in the Pacific Northwest. *Eucarpia Fruit Breeding and Genetics*. *Acta Horticulturae* 976:75–78
- Evans K, Brucher L, Konishi B, Barritt B (2010) Correlation of sensory analysis with physical textural data from a computerized penetrometer in the Washington State University Apple Breeding Program. *HortTechnology* 20:1026–1029
- Falconer DS, Mackay TFC (1996) *An introduction to quantitative genetics*, 4th edn. Longman, Essex
- Fellman JK, Rudell DR, Mattinson DS, Mattheis JP (2003) Relationship of harvest maturity to flavor regeneration after CA storage of ‘Delicious’ apples. *Postharvest Biol Technol* 27:39–51
- Forde CG, Williams ER, O’Riordan PJ (2007) Use of multi-stratum design and analysis in a sensory trial. *Food Qual Prefer* 18:614–618
- Gilmour AR, Gogel B, Cullis BR, Thompson R (2009) *ASReml user guide release 3.0*. VSN International Ltd, Hemel Hempstead
- Gitonga VW, Koning-Boucoiran CFS, Verlinden K, Dolstra O, Visser RGF, Maliepaard C, Krens FA (2014) Genetic variation, heritability and genotype by environment interaction of morphological traits in a tetraploid rose population. *BMC Genet* 15:32
- Guan Y, Peace C, Rudell D, Verma S, Evans K (2015) QTLs detected for individual sugars and soluble solids content in apple. *Mol Breed* 35:1–13
- Hahsler M, Hornik K, Buchta C (2008) Getting things in order: an introduction to the R package seriation. *J Stat Softw* 25:1–34
- Hampson CR, McKenzie DL (2006) Sensory characteristics of 21 new apple cultivars after short-term cold air storage. *Journal of the American Pomological Society* 60:9–19
- Hampson CR, Quamme HA, Hall JW, MacDonald RA, King MC, Cliff MA (2000) Sensory evaluation as a selection tool in apple breeding. *Euphytica* 111:79–90
- Hampson C et al (2007) Performance of apple cultivars in the 1999 NE-183 regional project planting III. Fruit sensory characteristics. *Journal of the American Pomological Society* 61:115–126
- Hampson CR, McNew R, Cline J, Embree C, Zandstra J, Wilson K (2009) Regional differences in performance of Canadian-bred apple cultivars and implications for breeding. *Can J Plant Sci* 89:81–91
- Harker FR, Maindonald J, Murray SH, Gunson FA, Hallett IC, Walker SB (2002a) Sensory interpretation of instrumental measurements 1: texture of apple fruit. *Postharvest Biol Technol* 24:225–239
- Harker FR, Marsh KB, Young H, Murray SH, Gunson FA, Walker SB (2002b) Sensory interpretation of instrumental measurements 2: sweet and acid taste of apple fruit. *Postharvest Biol Technol* 24:241–250
- Harker FR, Gunson FA, Jaeger SR (2003) The case for fruit quality: an interpretive review of consumer attitudes, and preferences for apples. *Postharvest Biol Technol* 28:333–347
- Harker FR, Amos RL, Echeverria G, Gunson FA (2006) Influence of texture on taste: Insights gained during studies of hardness, juiciness, and sweetness of apple fruit. *J Food Sci* 71:S77–S82
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447
- Henderson CR (1984) *Applications of linear models in animal breeding*. University of Guelph, Guelph
- Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. *Ecol Monogr* 54:187–211
- Jackson JE (2003) Eating quality and its retention. In: Jackson JE (ed) *The Biology of apples and pears*. Cambridge University Press, Cambridge, pp 341–383
- Karlsen AM, Aaby K, Sivertsen H, Baardseth P, Ellekjaer MR (1999) Instrumental and sensory analysis of fresh Norwegian and imported apples. *Food Qual Prefer* 10:305–314
- Kempton RA (1984) The use of biplots in interpreting variety by environment interactions. *J Agric Sci* 103:123–135
- Kenis K, Keulemans J, Davey MW (2008) Identification and stability of QTLs for fruit quality traits in apple. *Tree Genet Genomes* 4:647–661
- Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53:983–997
- King GJ et al (2000) Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill.). *Theor Appl Genet* 100:1074–1084
- Kouassi AB et al (2009) Estimation of genetic parameters and prediction of breeding values for apple fruit-quality traits using pedigreed plant material in Europe. *Tree Genet Genomes* 5:659–672

- Kuhn BF, Thybo AK (2001) Sensory quality of scab-resistant apple cultivars. *Postharvest Biol Technol* 23:41–50
- Kumar S, Volz RK, Alspach PA, Bus VGM (2010) Development of a recurrent apple breeding programme in New Zealand: a synthesis of results, and a proposed revised breeding strategy. *Euphytica* 173: 207–222
- Kumar S, Volz R, Weskett R (2011) Genetic architecture of fruit quality traits in *Malus x domestica* (Borkh.) compared between own-rooted seedlings and vegetative propagules on ‘M. 9’ rootstock. *Tree Genet Genomes* 7:1079–1088
- Kumar S, Garrick DJ, Bink M, Whitworth C, Chagne D, Volz RK (2013) Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC Genomics* 14
- Land DG, Shepard R (1987) Scaling and ranking methods. In: Piggot J (ed) *Sensory analysis of foods*. Elsevier, New York, pp 155–185
- Liebhart R, Koller B, Gianfranceschi L, Gessler C (2003) Creating a saturated reference map for the apple (*Malus x domestica* Borkh.) genome. *Theor Appl Genet* 106:1497–1508
- Longhi S, Cappellin L, Guerra W, Costa F (2013) Validation of a functional molecular marker suitable for marker-assisted breeding for fruit texture in apple (*Malus x domestica* Borkh.). *Mol Breed* 32: 841–852
- Maliapaard C, Alston FH, van Arkel G, Brown LM, Chevreaux E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG, den Nijs APM, Periam N, Rikkerink E, Roche P, Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Vrieling-van Ginkel M, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theor Appl Genet* 97:60–73
- McKay SJ, Bradeen JM, Luby JJ (2011) Prediction of genotypic values for apple fruit texture traits in a breeding population derived from ‘Honeycrisp’. *J Am Soc Horticult Sci* 136:408–414
- Meilgaard M, Civille GV, Carr BT (1999) *Sensory evaluation techniques*, 3rd edn. CRC, Boca Raton
- Miller S et al (2004) Performance of apple cultivars in the 1995 NE-183 Regional Project planting: II. Fruit quality characteristics. *Journal of the American Pomological Society* 58:65–77
- Miller S et al (2005a) Performance of apple cultivars in the 1995 NE-183 Regional Project planting: III. Fruit sensory characteristics. *Journal of the American Pomological Society* 59:28–43
- Miller SS et al (2005b) Effect of cultivar and site on fruit quality as demonstrated by the NE-183 Regional Project on apple cultivars. *HortTechnology* 15:886–895
- Miller S et al (2007) Performance of apple cultivars in the 1999 NE-183 Regional Project planting II. Fruit quality characteristics. *Journal of the American Pomological Society* 61:97–114
- Mohr CL, Mohr BC (2000) Apple maturation state and condition determination method in plantation, involves determining position of plunger moved through data points in each zones of fruit, to compute data related to viscoelastic properties of fruit. US6643599-B1
- Munzel U, Bandelow B (1998) The use of parametric vs. nonparametric tests in the statistical evaluation of rating scales. *Pharmacopsychiatry* 31:222–224
- Murray JM, Delahunty CM, Baxter IA (2001) Descriptive sensory analysis: past, present and future. *Food Res Int* 34:461–471
- Nyborg N (1959) On the inheritance of acidity in cultivated apples. *Hereditas* 45:332–350
- Nyborg H, Afzadi MA, Sehic J, Hertog M (2013) DNA marker-assisted evaluation of fruit firmness at harvest and post-harvest fruit softening in a diverse apple germplasm. *Tree Genet Genomes* 9:279–290
- Oraguzie NC, Hofstee ME, Brewer LR, Howard C (2001) Estimation of genetic parameters in a recurrent selection program in apple. *Euphytica* 118:29–37
- Oraguzie NC, Iwanami H, Soejima J, Harada T, Hall A (2004) Inheritance of the Md-ACS1 gene and its relationship to fruit softening in apple (*Malus x domestica* Borkh.). *Theor Appl Genet* 108:1526–1533
- Oraguzie N, Alspach P, Volz R, Whitworth C, Ranatunga C, Weskett R, Harker R (2009) Postharvest assessment of fruit quality parameters in apple using both instruments and an expert panel. *Postharvest Biol Technol* 52:279–287
- Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* 58:545–554
- Piepho HP (1997) Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics* 53:761–766
- Piepho HP, Mohring J, Williams ER (2013) Why randomize agricultural experiments? *J Agron Crop Sci* 199:374–383
- Pitts M, Drake S, Cavalieri R (1997) Apple size and length of storage affects firmness. <http://postharvest.tfrec.wsu.edu/pages/PC97D>. Accessed 07 Aug 2014
- Seppä L, Peltoniemi A, Tahvonen R, Tuorila H (2013) Flavour and texture changes in apple cultivars during storage. *Lwt-Food Science and Technology* 54:500–512
- Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). *Biometrika* 52:591
- Smith A, Cullis B, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57:1138–1147
- Stram DO, Lee JW (1994) Variance-components testing in the longitudinal mixed effects model. *Biometrics* 50:1171–1177
- Sunako T, Sakuraba W, Senda M, Akada S, Ishikawa R, Niizeiki M, Harada T (1999) An allele of the ripening-specific 1-aminocyclopropane-1-carboxylic acid synthase gene (ACS1) in apple fruit with a long storage life. *Plant Phys* 119:1297–1304
- Thompson R, Cullis B, Smith A, Gilmour A (2003) A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. *Aust N Z J Stat* 45:445–459
- Visser T, Verhaegh JJ (1978) Inheritance and selection of some fruit characteristics of apple. 1. Inheritance of low and high acidity. *Euphytica* 27:753–760
- Wilkinson RI et al (2008) Effects of 1-methylcyclopropane on firmness and flesh browning in Pink Lady (TM) apples. *Journal of Horticultural Science & Biotechnology* 83:165–170
- Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann Math Stat* 9:60–62
- Williams ER, Matheson AC, Harwood CE (2002) *Experimental design and analysis for tree improvement*. CSIRO Publishing, Melbourne
- Yates F (1936) A new method of arranging variety trials involving a large number of varieties. *J Agric Sci* 26:424–455
- Zhu YM, Barritt BH (2008) Md-ACS1 and Md-ACO1 genotyping of apple (*Malus x domestica* Borkh.) breeding parents and suitability for marker-assisted selection. *Tree Genet Genomes* 4:555–562