Use of a Quality Trait Index to Increase the Reliability of Phenotypic Evaluations in Broccoli

Zachary Stansell¹, Thomas Björkman¹, Sandra Branham², David Couillard², and Mark W. Farnham^{2,3}

U.S. Department of Agriculture, Agricultural Research Service, U.S. Vegetable Laboratory, Charleston, SC 29414; and the Department of Horticulture, Cornell University, Geneva, NY 14456

Additional Index words. *Brassica oleracea* var. *italica*, hybrids, selection index, relative-importance analysis, intraclass correlation coefficient

Abstract. Selection of superior broccoli hybrids involves multiple considerations, including optimization of head quality traits. Quality assessment of broccoli heads is often confounded by relatively subjective human preferences for optimal appearance of heads. To assist the selection process, we assessed five candidate head quality indices that make use of a set of individual and distinct ratings for traits such as head color, head smoothness, head size, head uniformity, and others. The head quality indices were tested for both a) the ability to reduce interobserver rating variability and b) the ability to emphasize specific attributes that display the greatest associations with overall horticultural quality of heads. Index development was based on datasets generated from quality evaluations by three independent raters of two replicated variety trials in Spring 2014. Relative-importance analysis was used to identify specific traits most associated with overall quality. Developed models were subsequently tested and compared using data collected by three raters evaluating two similar trials in Spring 2015. Head smoothness, head uniformity, head color, and holding ability were found to account for 78% of the model variation in overall head quality. Intraclass correlation coefficients (ICCs), which measure the degree of concordance among raters, were increased from 0.71 to 0.88 ($P < 0.05$) in one 2015 trial and from 0.67 to 0.80 ($P < 0.05$) in the second when comparing the simple overall quality assessment to the use of the index weighted by the most important individual head attributes. Thus, results showed that a quality index taking into account the relative importance of individual traits should enhance the identification of the best hybrids adapted to target conditions. This method can be used to improve concordance for subjective ratings in general.

Broccoli (*Brassica oleracea* L. var. *italica*) is the most economically important cole crop in the United States and currently has a farm-gate value of about \$900 million (USDA-NASS, 2015). Horticultural characteristics of mature broccoli heads have traditionally been difficult to quantify in breeding programs because of many potential quality defects (Farnham and Björkman, 2011a). Starting in 2010, a public-private collaboration [East Coast Broccoli Project; National Institute of Food and Agriculture (NIFA) Project No. 2010-51181-21062, Specialty Crops Research Initiative] has worked to

select and develop high-quality F₁ hybrid broccoli suitable for eastern environments and production, providing an alternative to current broccoli hybrids specifically developed for western production.

Breeding and selection for superior horticultural quality of broccoli heads under conditions like those occurring along the eastern seaboard that may include abiotic stress involves the manipulation of many genes conferring complex traits for which phenotypic differentiation can be difficult to assess (Farnham and Björkman, 2011a; Heather et al., 1992). The need to develop a consistent, stringent, and robust means of identifying suitable or unsuitable hybrids for the East became apparent during the course of conducting East Coast Broccoli trials at five regional test sites in South Carolina, North Carolina, Virginia, New York, and Maine. Although up to 10 attributes of heads produced by tested hybrids were evaluated in all regional trials, decisions to advance entries to subsequent trials have been based on assessments of mean overall quality scores. However, overall quality tends to be a subjective and difficult-to-define metric of horticultural acceptability that is hard to standardize when there are

many raters evaluating multiple trials. With a goal of overcoming this dilemma, we postulated that a quality trait index could be devised that would take into account a variety of individual quality attributes and provide a more robust measure of superior quality.

At the outset of conducting experiments described herein, we deemed it important to determine if evaluations by independently operating raters correlate with one another and if ratings will have high validity. Because horticultural quality assessments of broccoli heads are partially subjective, we also wanted to determine if a useful evaluation index based on greater human consensus could be constructed. In addition, it was recognized early that the elimination of redundant traits could be helpful in streamlining quality evaluations to save time and expense in conducting relatively large-scale and numerous quality trials.

The above considerations motivated us to devise an approach for analyzing evaluation trials by developing head quality phenotyping indices that could be compared using the ICC. The ICC, introduced by Fischer (Bartko, 1966), is a statistic used when measurements or observations are made on the same test subject by multiple raters, and it is a useful means of testing interobserver reliability and consistency (Fleiss and Cohen, 1973).

Selection indices have been frequently tested in a variety of agronomic crops, including soybean (*Glycine max* L. Merr.; Bouchez and Goffinet, 1990; Byth et al., 1969), oats (*Avena sativa* L.; Eagles and Frey, 1977), peanuts (*Arachis hypogaea* L.; Chandra et al., 2003), and wheat (*Triticum aestivum* L.; Sharma and Duveiller, 2006). These studies were primarily focused on selection of low genotype-by-environment traits, such as yield (Byth et al., 1969; Chandra et al., 2003) or on multiple yield-related trait selection (Sharma and Duveiller, 2006). We are unaware of the use of a selection index in horticultural crops to improve evaluation of a relatively subjective quality trait.

With the above considerations in mind, the specific objectives of this work were to 1) compare trait scores made by different individual raters on the same plots and trials, 2) determine which individual traits were most associated with overall quality, 3) define unique linear trait indices (e.g., linear combinations of trait values) that might be used as alternative measures for identifying the best hybrids, and 4) determine if the indices increase rater consistency for ranking hybrids. Our ultimate goal was to develop a reliable and optimal linear combination of rating scores that can effectively identify the best adapted broccoli hybrids for East Coast production.

Materials and Methods

Plant materials and growth conditions. Broccoli hybrids included in East Coast Broccoli field trials were all experimental

Received for publication 12 June 2017. Accepted for publication 19 Sept. 2017.

This study was funded by the United States Department of Agriculture, Agricultural Research Service Project No. 6080-21000-018-00D and the National Institute of Food and Agriculture, Project No. 2016-51181-25402.

¹Graduate Research Assistant and Associate Professor at Cornell.

²Research Geneticist, Biological Science Technician, and Research Leader, respectively with the USDA-ARS.

³Corresponding author. E-mail: Mark.Farnham@ars.usda.gov.

Table 1. Descriptions of rating scores for traits collected during east coast broccoli trials (ECBT). All plots were rated on a 1 (worst) to 5 (best) scale.

Plot uniformity:	This is an indication of the homogeneity of appearance and relative maturity of all plants in a plot. 5 = all plants highly uniform, 4 = plants uniform, with few minor differences, 3 = moderate variability, 2 = poor uniformity, 1 = plants more dissimilar than alike.
Head extension:	This is an indication of the broccoli head apex position at market maturity relative to the topmost foliage of the plant. 5 = head apex is very high (e.g., four to six inches) above foliage, 4 = head apex above (e.g., two to three inches) foliage, 3 = head apex at same position of top foliage, 2 = slightly below (e.g., two to three inches) top foliage, 1 = head apex buried deep in foliage.
Holding ability:	This provides an indication/rating of the degree to which heads retain their mature shape when the size surpasses a diameter of 15 cm. This rating is made $\approx 5-7$ d after other data are collected at maturity. 5 = keeps shape as head increases in size up to a week after maturity, 4 = mostly keeps shape as heads increase in size, 3 = moderate holding, 2 = poor holding, 1 = no holding once it reaches maturity.
Dome:	This is a rating of the relative degree of convex shape. 5 = high dome, 4 = medium dome, 3 = slight dome, 2 = flat head, 1 = concave head. See illustrated cross sections on the ECBT rating chart.
Bead size:	Provides rating of the mean size of unopened buds at commercial maturity. 5 = fine beads <1.1 mm diameter, 4 = small beads (1.1–1.4), 3 = medium beads (1.4–1.7), 2 = large beads (1.7–2.0 mm), 1 = extralarge beads (>2.0 mm). Also, see illustration of small beardedness on ECBT rating chart.
Bead uniformity:	Indicates the degree of bead uniformity at commercial maturity of the head. 5 = beads highly uniform across head surface; 4 = slight nonuniformity of beads across surface; 3 = acceptable, but marginal uniformity of beads; 2 = significant variability of bead size and appearance across head surface, with significant cat-eye of beads; 1 = highly variable bead size and appearance. Also, see illustration of uniform bead on ECBT rating chart.
Head firmness:	Rating of the relative firmness of a head; heads that offer little to no physical resistance when touched receive a lower firmness score. 5 = very firm or hard when fingers are pressed on top of the head with no give on pressing; 4 = generally firm with only slight resistance when pressed; 3 = less firm and head surface is displaced more when pressed; 2 = unacceptable firmness, with head surface loose and easily displaced when touched; 1 = very loose head with open space visible through surface of head.
Head color:	This is a rating of the head color which can vary significantly from yellow to green to blue green depending on the environment. 5 = deep blue green, 4 = green to blue green, 3 = light green, 2 = lime, 1 = yellow. See colors illustrated on ECBT rating card.
Head uniformity and smoothness:	Indicates degree of overall regularity of curd/bead surface. Smoother, more uniform heads receive higher scores. 5 = very smooth, even surface across the head surface; 4 = mostly uniform surface with only slight irregularity; 3 = acceptable uniformity across surface with more moderate variability across the surface; 2 = significant lack of smoothness across head surface with a more lumpy appearance; 1 = distorted head surface.
Overall quality:	Perceived overall quality of broccoli hybrid. Ratings below three are not considered marketable. 5 = excellent quality with all individual ratings at the 4 to 5 level; 4 = very good quality individual ratings at the 3 to 4 level; acceptable quality (e.g., marketable) broccoli with individual ratings of mostly 3 and observable flaws; 2 = poor quality broccoli with individual ratings below 3; 1 = not generally recognizable as broccoli with most attributes distorted.

Fig. 1. Scoring chart for east coast broccoli trials (ECBT). Illustrated from top to bottom are: bead size (BS), head shape (HS), bead uniformity (BU), and head color (HC). Plots were rated on a 1 (worst) to 5 (best) scale. Traits not included on this rubric are head extension (HE), plot uniformity (PU), head firmness (HF), overall quality (OQ), and holding ability (HA). A color version of this figure is shown in the online edition of *HortScience*.

submissions or standard checks. In 2014, 39 hybrids were evaluated in each of the two tests, including two experimental hybrids from Seminis Vegetable Seeds (Woodland, CA), 10 experimental hybrids from Syngenta (Gilroy, CA), three from Bejo Zaden (Oceano, CA), nine from the USDA-ARS program at Charleston, SC, seven from the Cornell program at Geneva, NY, and eight control hybrids [Marathon, Green Magic and Green Gold developed by Sakata Seed (Morgan Hill, CA), Ironman and Castle Dome developed by Seminis, Bay Meadows, Amadeus, and Everest developed by Syngenta, and Belstar developed by Bejo Zaden]. A total of 48 tested hybrids in 2015 included different hybrids from the same cooperating

programs as those represented in 2014. This set included 20 experimental hybrids from Seminis, five experimental hybrids from Syngenta, nine from the USDA-ARS, three from Cornell, and four from Oregon State University. All of the same control hybrids except 'Amadeus' were used in 2015. All entries included in both years were numbered by a group at Cornell so that the individuals evaluating the trials described in this study had no knowledge of the actual identity or origin of any of the entries being rated.

In the spring of 2014, seeds were planted into 200 cell Speedling trays (Speedling Incorporated, Sun City, FL) in Fafard 3B (Conrad Fafard Inc., Agwam, MA) growth

media on 10 Feb. for the first 2014 trial and 25 Feb. for the second. Seedlings were grown in a greenhouse under natural light at about 24/18 °C day/night until transplanting. Seedlings were transplanted into a field of Yoncos loamy sand (fine loamy mixed, thermic Albaqualf) at the U.S. Vegetable Laboratory (Charleston, SC) on 14 Mar. for the first trial and on 31 Mar. for the second. Both trials were planted in the same field on raised beds formed on bare ground with a single row planted at the center of each bed. Between-row spacing was 91 cm and within-row plant spacing was 15 cm. The experimental design for each trial was a randomized complete block design with three replications. Individual plots were made up of 20 plants of each of the 39 hybrid entries. Two applications of calcium nitrate (337 kg·ha⁻¹) were applied with a mechanical side dresser, and weekly applications of liquid 20N–20P–20K (15.0 kg·ha⁻¹) were applied through drip irrigation during the course of each experiment. Any additional cultural practices were as previously described by Farnham and Björkman (2011b).

In 2015, two similar trials were established in a different field at the Charleston site with the same soil type as used in 2014. Experimental design, field layout, and plot maintenance were all conducted the same as described for 2014. Seed of the 48 entries tested in 2015 were planted in the greenhouse on 5 Feb. and 19 Feb. Transplanting dates for the two 2015 trials were 12 Mar. and 30 Mar.

Field evaluations. As plants approached maturity, each experiment was examined three times per week, and individual plots were selected for evaluation. Plots were deemed ready to rate when about 50% to 60% of heads reached commercial maturity as crown-cut broccoli (e.g., with head

Table 2. Comparisons of different rater pairs for overall quality scores of heads in planting 1, planting 2, and also pooled for both plantings in 2014. Spearman correlation coefficients are indicated for each pairwise (AB, AC, or BC) rater comparison.

Comparison	Planting 1	Planting 2	Pooled
	<i>r</i>		
A/B	0.61	0.74	0.71
A/C	0.71	0.66	0.66
B/C	0.52	0.57	0.52

Fig. 2. Normalized relative-importance coefficients as predictors of overall head quality rating computed using all trait rating scores compiled by individual raters **A**, **B**, and **C** from 2014 trials. Each trait, including head uniformity and smoothness (US), bead uniformity (BU), head color (HC), head shape (HS), holding ability (HA), plot uniformity (PU), head firmness (HF), bead size (BS), head extension (HE), explained as a portion of R^2 for each rater. Vertical bars represent 95% confidence intervals.

diameter between 10.2 and 12.7 cm). Average days from transplanting to plot maturity ranged from 54 to 78 d in year 1 and 49

to 84 d in year 2. Once identified, the plots were evaluated independently by each of the three trained raters (designated A, B

and C), all of whom were members of the Charleston broccoli breeding project and experienced broccoli workers. In addition, all trial raters were trained in advance during project meetings and in the field by evaluating test materials in trials conducted before those described herein. Assessment of head quality of all hybrids tested in Charleston followed standardized procedures used by all raters in all East Coast Broccoli Project trials. Specifically, when plants in a plot were deemed ready (e.g., with enough heads at the crown cut stage), ten traits, including plot uniformity, head shape, head smoothness or uniformity, head color, head firmness, head extension, bead size, bead uniformity, overall quality, and holding ability were evaluated on a plot basis. All traits were scored using an ordinal scale from 1 to 5 (1 = worst, 5 = best). In general, a score of below three for a given trait was considered not suitable for market.

Plot uniformity was rated by examining all plants in the plot at maturity and assessing relative similarity of individuals to one another. Holding ability was assessed 5–7 d after the primary rating date indicating the relative ability of a given entry to retain good quality characteristics as the head continued to increase in diameter past the designated crown-cut size. The remaining traits were assessed as an average score of all heads at the mature size in the plot on the date the plot was rated. Criteria used when assigning ratings (Table 1) were provided to each rater in addition to a scoring help sheet that they used while rating plots (Fig. 1). All raters had previous experience scoring broccoli heads in field trials.

Index development. Data from the 2014 trials were used to design five quality trait indices (linear combinations of individual trait values). Index 1 was the overall quality score given to a plot by each rater, i.e., the value often used for hybrid selection in the East Coast Broccoli Project. Index 2 was calculated as if all evaluated traits have equal value. Index 3 was subjective, using weights based on expert opinion about stakeholders' requirements. Index 4 was based on explaining maximum variance, weighting the traits that contributed most to the model R^2 for predicting overall quality ratings higher, and those contributing little, much less (Johnson and Lebreton, 2004). Index 5 was a relative-importance model using the best subset of predictor variables associated with overall quality to minimize extraneous data collection.

Specific coefficients of relative importance for selection indices 4 and 5 were generated using the relaimpo package (Groemping and Matthias, 2013) in R (R Development Core Team, 2015) in conjunction with the "lmg" setting which partitions variance and sums to total R^2 by averaging over orders. Bootstrapping was used to estimate 95% confidence intervals of each predictor's relative importance by bootstrapping under random

model assumptions using 1000 iterations with the function boot. Variable selection for index 5 (relative-importance subset) was accomplished with the leaps (Lumley, 2017) function included in regression subset package leaps. Leaps provide an exhaustive search for the best predictor variables using a branch-and-bound algorithm and was applied to the pooled 2014 data using an adjusted R^2 method.

Index comparison. The five indices were compared using ICC which is defined as:

$$ICC = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\epsilon}^2}$$

For this computation, σ_{α}^2 is the variance of the test subject and σ_{ϵ}^2 the variance among observers. ICCs are used to describe the degree to which members of groups resemble each other with values ranging from 0 to 1

(Donner and Koval, 1980) and higher values demonstrating a stronger correlation between rater scores. Because multiple raters make observations on the same subject, we deemed the ICC as an appropriate tool to compare the proposed evaluation indices and ultimately confirm the reliability and concordance of the different indices used.

ICCs were calculated for the five indices using the irr package (Gamer et al., 2012) in R. ICC calculation assumptions implemented a two-way model (e.g., every genotype is rated by every rater in each experiment). ICC “type” parameter was “concordance” because the ability to reliably differentiate between high from low quality characteristic hybrids was preferred (raters will agree that high performers will always be high, low performers will always be low) contrary to anchoring to an absolute rating scale (i.e., all

raters will agree on an absolute score for a particular hybrid). The unit parameter was selected as “single” because data were not averaged across replication. A one-sided test ($H_1: r > r_0$) was performed to compare the differences in index performance. Ninety-five percent confidence intervals were also computed for index ICC.

Results and Discussion

Interobserver reproducibility for individual traits. Three different raters scored the overall quality of heads differently for many hybrid plots in 2014 based on results of three different pairwise comparisons (Table 2). All three possible rater comparisons (A/B, A/C, and B/C) were positively correlated whether examined by individual plantings or when pooled, but many individual scores varied between the raters with coefficients ranging from 0.52 to 0.74. Correlations of scores between A and B and between B and C were always better than between A and C. Even though all raters were trained to conduct the scoring in a similar fashion at the outset, the raters gave different scores operating independently. These observations illustrate the imperfect and subjective nature of the rating designed to assess overall quality of broccoli heads.

Overall, quality ratings of heads have been used extensively by the East Coast Broccoli Project as the primary criterion for choosing the best hybrids for advanced trials. Based on many personal communications, the authors know a similar type rating is also commonly used by public- and private-sector broccoli breeders. Concerns about rater variability in scoring such a subjective trait were validated by the significant rater variability found in the 2014 trials. Correlation coefficients among rater scores for overall quality were always positive and greater than 0.50; however, even when this trait was scored by well-trained raters, as here, correlation coefficients never exceeded 0.75. These observations give credence to the idea that an index making use of other trait scores, several of which are less subjective to rate, might provide a means to develop a better selection tool.

Trait correlation with overall quality. Relative-importance analysis of the contribution of individual traits to overall quality was performed for the 2014 trials for each individual rater (Fig. 2) and also when pooled across all three raters (Fig. 3).

Fig. 3. Normalized relative-importance coefficients as predictors of overall head quality rating computed using all trait rating scores compiled by three individual raters and pooled for 2014 trials. Each trait, including head uniformity and smoothness (US), bead uniformity (BU), head color (HC), head shape (HS), holding ability (HA), plot uniformity (PU), head firmness (HF), bead size (BS), head extension (HE), explained as portion of R^2 . Vertical bars represent 95% confidence intervals for the relative coefficient. Coefficients were used in constructing indices relative importance (RI) and relative importance subset (RI*).

Table 3. Weights used to compute five different selection indices (linear combinations) of broccoli head traits including head extension (HE), plot uniformity (PU), head color (HC), head shape (HS), head uniformity or smoothness (US), head firmness (HF), bead size (BS), bead uniformity (BU), overall quality (OQ), and holding ability (HA).

Selection index	Trait									
	OQ	HE	PU	HC	HS	US	HF	BS	BU	HA
1) OQ	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2) Average	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
3) Weighted	0.16	0.04	0.04	0.16	0.08	0.16	0.08	0.08	0.16	0.04
4) Relative importance	0.00	0.01	0.05	0.14	0.12	0.32	0.01	0.00	0.23	0.12
5) Relative importance subset	0.00	0.00	0.00	0.19	0.00	0.41	0.00	0.00	0.27	0.13

Table 4. Results of leaps analysis showing adjusted R^2 of the subsets of variables, including head extension (HE), plot uniformity (PU), head color (HC), head shape (HS), head uniformity or smoothness (US), head firmness (HF), bead size (BS), bead uniformity (BU), overall quality (OQ), and holding ability (HA), that increasingly predict OQ.

Number of predictors	Adjusted R^2	Trait								
		HE	PU	HC	HS	US	HF	BS	BU	HA
1	0.62					*				
2	0.71			*		*				
3	0.75			*		*			*	
4	0.78			*		*			*	*
5	0.80			*	*	*			*	*
6	0.80		*	*	*	*			*	*
7	0.80		*	*	*	*			*	*
8	0.81	*	*	*	*	*	*		*	*

Presence of "*" in a given row indicates the inclusion of the trait (column) in a given predictive model.

Table 5. Intraclass correlation coefficients (measuring rater consistency) for two different 2015 plantings using a two-way model of five different selection indices computed with up to ten broccoli head trait ratings.

Selection index	Planting 1	Planting 2—intraclass correlation coefficient ^z
Overall quality	0.67 a	0.71 a
Trait average	0.70 ab	0.80 b
Weighted average	0.74 bc	0.83 bc
Relative importance	0.80 d	0.88 d
Relative importance subset	0.78 cd	0.86 cd

^zCoefficients followed by the same letter are not significantly different ($P < 0.05$) from one another. Three raters made 144 plot observations for a total of ten heading traits. The 95% CI for the relative-importance index in plantings 1 and 2 were 0.75–0.85 and 0.84–0.91, respectively.

Relative-importance coefficients from the pooled data were used to develop the relative-importance selection index 4 and 5 (Table 3). Head uniformity and smoothness explained the greatest proportion of the variation in overall quality for all three raters (32% pooled), followed by bead uniformity (22% pooled). Head color, holding ability, and head shape were consistently important to overall quality but their rank order changed dependent on the specific rater. The relative importance of plot uniformity was inconsistent and varied from 2% to 7%. The remaining traits were not found to explain significant variation in overall quality in this study. Indeed, less than 3% of variation in the overall quality scores was explained by these other head attributes when examined on an individual or pooled basis.

Leaps analysis was applied to the pooled 2014 data to search for the traits most predictive of overall quality to develop a relative-importance model reduced to an optimized subset of the measured traits (e.g., index 5). Leaps results (Table 4) indicated that 78% of variance in overall quality was explained by four predictor variables; head uniformity and smoothness (41%), bead uniformity (27%), head color (19%), and holding ability (13%), which corresponds closely to the variables that explained the most variation in overall quality in the relative-importance analysis of the pooled data. The leaps results provided in row 5 of Table 3 are the specific coefficients to use in the relative-importance subset selection index (Table 3).

Effectiveness of the index selection approach. Rater consistency of the five different selection indices was compared by intraclass correlation analysis using scores from each of the two 2015 plantings. The four

computed indices were all more effective than overall quality scoring in reducing rater variability. However, the best was the relative-importance index, which had an ICC = 0.80 for the first trial in 2015 and ICC = 0.88 for the second (Table 5). Both of these values were significantly higher ($P < 0.05$) than the ICCs for overall quality scoring, the mean index, or the weighted index. The relative-importance subset index was slightly less effective than the relative-importance index with ICC = 0.78 for the first trial and ICC = 0.86 for the second, but the two indices were not significantly different from one another. Use of the overall quality score or the selection index using an average value across all traits resulted in a significantly lower rater consistency than with those based on relative-importance models.

Broccoli breeders typically deem smoothness of the head surface and uniformity of bead size as critical factors determining head quality because these two traits are often adversely affected by high temperature spikes and other stress conditions (Björkman and Pearson, 1998; Farnham and Björkman, 2011b). The effectiveness of the two relative-importance indices confirms the important role that head smoothness and bead uniformity play in determining head quality. Moreover, the utility of the relative-importance subset is especially important to note as it expands opportunities for improvement of broccoli quality through automated phenotyping. Phenotyping is the current limiting factor in both very large scale trial systems and genomic selection programs. In both cases, fewer observations of more genotypes might actually prove effective to accomplishing objectives.

Automated high-throughput phenotyping platforms have the potential to increase the accuracy, precision, and throughput of trait measurements, whereas decreasing costs through reduced labor requirements (Cobb et al., 2013). The subjective nature of the “overall quality” trait prevents the use of automated phenotyping systems. However, the more discrete traits of the relative importance subset (i.e., head smoothness, bead uniformity, head color, and holding ability) capture the most variation in overall quality while being better suited to the possibility of automated phenotyping. Likewise, in cases where genomic selection is conducted and phenotyping throughput is often a limiting factor, a streamlined approach provided by a subset of traits could be advantageous.

Conclusions

This article describes an effort to develop a more effective and general selection tool for a relatively difficult-to-score trait of a horticultural crop, and we recommend it as a logical application for a crop such as broccoli, one of many horticultural crops for which quality is as paramount as the actual yield of the crop (Janick, 2005). We used overall quality broccoli ratings to inform selection index design in this study, but they were not included as a weighted component of the relative-importance evaluation indices. Using multiple, succinct, and clearly defined traits, such as head smoothness, bead uniformity, and head color, to model horticultural quality and compute an index, we showed that the elements of human subjectivity can be reduced in assessing quality. Moreover, rater-drift that could arise between field seasons might also be reduced by using quantitative indices of well-defined traits. Indeed, a rater may be susceptible to the “contrast effect”; specifically, rating an intermediate quality broccoli much higher when the rest of the experimental units are of a much lower quality, or vice versa, in scenarios where overall quality rating is assessed. By modeling horticultural quality more objectively, environment-by-environment comparisons could prove more reliable. Most importantly, our results indicate that the effects of rater variability and overly-subjective rating can be minimized when employing a relative-importance index to identify superior hybrids.

Literature Cited

- Bartko, J.J. 1966. The intraclass correlation coefficient as a measure of reliability. *Psychol. Rep.* 19:3–11.
- Björkman, T. and K.J. Pearson. 1998. High temperature arrest of inflorescence development in broccoli (*Brassica oleracea* var. *italica* L.). *J. Expt. Bot.* 49:101–106.
- Bouchez, A. and B. Goffinet. 1990. Evaluation of selection index: Application to the choice of an indirect multitrait selection index for soybean breeding. *Theor. Appl. Genet.* 79:261–267.
- Byth, D.E., B.E. Caldwell, and C.R. Weber. 1969. Specific and non-specific index selection in soybeans, *glycine max* l. (Merrill). *Crop Sci.* 9:702–705.
- Chandra, S., S.N. Nigam, A.W. Cruickshank, A. Bandyopadhyaya, and S. Hardcrishna. 2003. Selection index for identifying high-yielding groundnut genotypes in irrigated and rainfed environments. *Ann. Appl. Biol.* 143:303–310.
- Cobb, J.N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 2013. Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* 126: 867–887.
- Donner, A. and J. Koval. 1980. The estimation of intraclass correlation in the analysis of family data. *Biometrics* 36(1):19–25.
- Eagles, H.A. and K.J. Frey. 1977. Repeatability of the stability-variance parameter in oats. *Crop Sci.* 17:253–256.
- Farnham, M.W. and T. Björkman. 2011a. Breeding vegetables adapted to high temperatures: A case study with broccoli. *HortScience* 46:1093–1097.
- Farnham, M.W. and T. Björkman. 2011b. Evaluation of experimental broccoli hybrids developed for summer production in the eastern United States. *HortScience* 46:858–863.
- Fleiss, J.L. and J. Cohen. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ. Psychol. Meas.* 33:613–619.
- Gamer, M., J. Lemon, and I.F.P. Singh. 2012. irr: Various coefficients of interrater reliability and agreement. R package version 0.84.
- Groemping, U. and L. Matthias. 2013. Relaimpo: Relative importance for linear regression in R: The package relaimpo. *J. Stat. Softw.* 17(1): 1–27.
- Heather, D.W., J.B. Sieczka, M.H. Dickson, and D.W. Wolfe. 1992. Heat tolerance and holding ability in broccoli. *J. Amer. Soc. Hort. Sci.* 117:887–892.
- Janick, J. 2005. Horticultural plant breeding: Past accomplishments, future directions. *Acta Hort.* 694:61–65.
- Johnson, J.W. and J.M. LeBreton. 2004. History and use of relative importance indices in organizational research. *Organizational Research Methods* 70:238–257.
- Lunley, T. 2017. Leaps: Regression subset selection. R package version 3.0.
- R Development Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Sharma, R.C. and E. Duveiller. 2006. Farmer participatory evaluation confirms higher grain yields in spring wheat using a selection index for spot blotch resistance, maturity and kernel weight. *Euphytica* 150:307–317.
- U.S. Department of Agriculture, National Agricultural Statistics Service. 2015. Vegetables 2015 Summary.