Computing Integrated Ratings from Heterogeneous Phenotypic Assessments: A Case Study of Lettuce Postharvest Quality and Downy Mildew Resistance

Ivan Simko,* Ryan J. Hayes, and Matthew Kramer

ABSTRACT

Comparing performance of a large number of accessions simultaneously is not always possible. Frequently, only subsets of all accessions are tested in separate trials with only some (or none) of the accessions overlapping between subsets. Using standard statistical approaches to combine data from such a sparsely populated accession × trial matrix is precluded if different rating scales are used to evaluate accessions in those trials. Here we compare two approaches that can compute an overall linear rating for the performance of all accessions across a set of trials, even for accessions that were never tested together and were rated on dissimilar scales. We use data from lettuce (*Lactuca sativa* L.) postharvest quality collected on 178 accessions in 18 trials and assessment of lettuce resistance to downy mildew (*Bremia lactucae* Regel) performed on 583 accessions in 53 trials. The projected values (PV) approach uses a combination of principal component analysis and resampling to merge trial results and calculates an overall rating from real values. In contrast, the rank-aggregation (RA) approach uses an extension of the Rasch model to combine rank-ordered data from individual trials. We found high correlation between ratings produced by the two approaches for the postharvest quality (r = 0.803) and the resistance to downy mildew (r = 0.748). Combining data from multiple experiments identified lettuce accessions with a high level of resistance to the disease and a slow rate of deterioration when processed for salad. The PV and RA approaches also allow combining data from different laboratories or databases.

I. Simko and R.J. Hayes, United States Department of Agriculture, Agricultural Research Service, U.S. Agricultural Research Station, 1636 E. Alisal St., Salinas, CA 93905; M. Kramer, United States Department of Agriculture, Agricultural Research Service, Biometrical Consulting Service, Henry A. Wallace Agricultural Research Center, 10300 Baltimore Ave., Beltsville, MD 20705. Received 17 Feb. 2012. *Corresponding author (ivan.simko@ars.usda.gov).

Abbreviations: AUDePS, area under the deterioration progress stairs; AUDPS, area under the disease progress stairs; IQR, interquartile range; MNV, mean of normalized values; PCA, principal component analysis; PV, projected values; RA, rank aggregation; RIL, recombinant inbred line.

The best way to compare accessions is to conduct a series of trials (experiments) that captures the range of environments of interest and to have every accession present in each trial. However, testing all accessions in all trials is usually not possible because of limiting factors, such as the availability of space (e.g., field, greenhouse, growth chamber, laboratory, etc.), time required for assessments, demands for equipment, labor, logistics, and economic requirements. Because of these limitations, separate experiments are typically performed on different subsets, with only a few (or no) overlapping accessions between subsets. This type of testing leads to a sparsely populated accession × experiment matrix (Simko and Piepho, 2011). If the objective is to compute mean scores for a set of accessions across a set of experiments and all evaluations were performed at matching scales, it is possible to adjust individual scores and combine them into overall adjusted means (Piepho, 2003). However, if the rating scales for different experiments differ, combining data from multiple experiments with only partially overlapping sets of accessions is

Published in Crop Sci. 52:2131–2142 (2012).
doi: 10.2135/cropsci2012.02.0111

more challenging. Here we show how such heterogeneous data can be combined using the methods of projected values (PV) (Ehlenfeldt et al., 2010) and rank aggregation (RA) (Simko and Linacre, 2010; Simko and Pechenick, 2010). These two approaches are applied to evaluations of postharvest quality and resistance to downy mildew conducted on lettuce accessions.

Salad-cut lettuce is a highly perishable product with a short shelf life. To extend shelf life of minimally processed lettuce, salad is typically kept at low temperature and in a modified atmosphere consisting of low O_2 and high CO_2 levels (Smyth et al., 1998; Kim et al., 2005). Nevertheless, significant phenotypic differences in shelf life (percent deteriorated tissue) were observed among lettuce cultivars stored under these modified atmospheric conditions (Hayes and Liu, 2008). These differences appear to be relatively independent of environment because similar rankings of cultivars were observed under different conditions. Given the importance of salad-cut products to the lettuce industry, cultivars, breeding lines, and populations need to be evaluated for the rate of deterioration after processing for salad. In the present work, we combine data from 18 trials that were used to evaluate 178 accessions for the rate of deterioration. Because it was not possible to test all accessions simultaneously, individual trials included only a subset of accessions that were evaluated on three rating scales.

Downy mildew disease, caused by the oomycete *Bremia lactucae* Regel, is a major threat to lettuce production worldwide. Control of the disease is usually based on a combination of fungicides and resistance genes (Crute, 1984, 1992). Two types of resistance genes against downy mildew are present in lettuce: qualitative resistance that is conferred by major genes that render the host incompatible with the pathogen and quantitative resistance that is usually conferred by multiple genes, each of minor effect (Grube and Ochoa, 2005; McHale et al., 2009). Unfortunately, qualitative resistance is not durable because *B. lactucae* develops new races that can defeat the resistance genes of lettuce (Crute, 1998). Quantitative resistance is usually more durable, but it is also more difficult to evaluate (Crute and Norwood, 1981; Grube and Ochoa, 2005). To detect material with good resistance in a particular growing area, a large number of trials need to be performed. Here we combine data from 53 trials that were used to evaluate 582 accessions for their resistance to downy mildew.

This study was motivated by a large database consisting of hundreds of accessions tested over several years either for their postharvest quality or resistance to downy mildew, but an overall rating of the accessions was precluded because the trials were performed on only partially overlapping subsets of accessions and evaluations were performed with different rating scales. The objectives of the present study were (i) to combine data from multiple trials into an overall rating of accessions, (ii) to compare results of the two approaches based on projected values and rank aggregation, and (iii) to identify accessions with both a slow rate of deterioration after processing for salad and a high resistance to downy mildew under field conditions.

MATERIALS AND METHODS

Plant Material

A set of accessions tested for postharvest quality (Supplemental Table S1) and/or resistance to downy mildew (Supplemental Table S2) included commercial cultivars, plant introductions from both Pullman, WA, and Salinas, CA, seed depositories, breeding lines and released germplasm from our breeding programs, and F_8 recombinant inbred lines (RILs) from the Salinas 88 × La Brillante mapping population (Hayes et al., 2011). Lettuce plants were seeded in two rows on 1 m wide beds and were thinned to obtain a spacing of about 30 cm between plants. All trials were grown in Salinas, CA, with a single exception of one experiment grown in Yuma, AZ, to assess shelf life. Standard commercial practices for the area were used for irrigation and fertilization; however, downy mildew infection was not controlled.

Postharvest Quality

Plants at harvest maturity were harvested, processed into salad, and packaged in 22.8 by 30.5-cm clear film bags as described in Hayes and Liu (2008). In total, 178 accessions from 18 trials were tested for postharvest quality. Bags of salad-cut lettuce were stored at 3.5°C and deterioration of material was evaluated by one of the three methods.

1. Method of Hayes and Liu (2008). Evaluations were conducted when some bags of salad reached 100% deterioration. To evaluate the degree of deterioration, deteriorated pieces were collected and weighed. The percentage of deteriorated pieces out of all tissue was recorded as degree of deterioration on a scale of 0 to 100%.
2. Evaluations were conducted when a bag with the highest rate of deterioration reached approximately 50%. All bags were visually rated (without opening) on a scale of 1 to 5, in which 1 indicated no deterioration and 5 indicated 100% deterioration. Rating of bags was performed six times at weekly intervals and a mean score for each bag was calculated from these six evaluations.
3. Evaluations of bags were conducted at weekly intervals, starting the first week after processing and continuing until the last bag reached 100% deterioration. Rating of deterioration was visually performed without opening the bags on a rating scale ranging from 0 to 10 that corresponds to the estimated percentage of deteriorated tissue divided by 10. The area under the deterioration progress stairs (AUDePS) was calculated from individual evaluations to combine progress of deterioration into a single value (Simko and Piepho, 2012).

Resistance to the Downy Mildew Disease

Lettuce reaction to downy mildew was assessed on 582 accessions grown in 53 trials. Trials were not artificially inoculated with *B. lactucae*, as natural infection occurred in all experiments. Resistance to the disease was evaluated under field conditions by one of the three methods.

1. Method of Grube and Ochoa (2005). Evaluation of resistance was performed when the majority of plants in the field were at the harvest maturity. Disease symptoms were evaluated on 10 plants per plot as described by Grube and Ochoa (2005). The disease rating scale of 0 (no disease) to 5 (nearly 100% disease) combined both symptom severity and the number of symptomatic leaves per plant (Grube and Ochoa, 2005).
2. Single evaluation of disease at harvest maturity. Disease on plants was visually rated on a scale of 1 to 5, in which 1 indicated no disease and 5 indicated very heavy infection and plants with many dead leaves. An overall disease score was given to the whole plot (approximately 60 plants).
3. Continuous evaluations of disease progress. Evaluations started when the most susceptible accession was moderately infected and continued at weekly intervals until harvest maturity. Disease was rated on a scale of 0 to 5, in which 0 indicated no disease and 5 indicated nearly 100% infection. This approach was different than evaluation method A in that the overall disease score was visually estimated for the whole plot (approximately 15–30 plants) rather than counting the number of diseased leaves on selected plants. The area under the disease progress stairs (AUDPS) was calculated from individual evaluations to combine the disease progress data into a single value (Simko and Piepho, 2012).

Combining Data by the Projected Values Approach

Data from individual trials were combined into a single rating as described in Ehlenfeldt et al. (2010) with minor modifications. First, a core set of accessions and trials were selected. This core set consisted of accessions that were tested in at least two trials and their rating was neither the highest nor the lowest in all of their participating trials. If an accession was tested in just a few trials and was rated the best or worst in all of them, its score was unnaturally high or low if left in the core data set. Therefore these accessions were excluded from the core set and their rating was calculated later. Trials that were selected into the core set contained at least three accessions that were tested in at least two or more trials and were linked to other accessions through testing in some other trials. Gradual elimination of accessions and trials was performed until the core set was determined. The shelf life data core set contained 176 accessions and 17 trials. The core set of downy mildew data contained 575 accessions and 53 trials. After the core set was determined, the projected value approach (Ehlenfeldt et al., 2010) was applied to calculate estimates for all accessions in the core set. We outline the method below; the R code (R Development Core Team, 2008) is given in Supplemental File S3. This approach uses a combination of principal component analysis (PCA) and resampling. A PCA determines an axis that best goes through datasets of two randomly selected trials with at least a few accessions in common, thus allowing calculation of "projected" values (that is, projected to a common axis determined by the accessions in common in the two trials; accessions that are unique to one or the other of the two trials are projected onto this common axis using the parameters from the PCA so that the projected axis includes values for all accessions from the two trials). Random resamplings and calculations were then repeated 999 additional times. The whole process was then started again, but this time drawing random samples from the projected values; this was iterated until the scores stabilized. The number of iterations depended on the original data set; we found convergence typically occurred after a few hundred iterations (and sometimes far less). This entire procedure was repeated 30 times and the resulting scores used to calculate a mean and variance for each accession in the core set. Well-represented accessions that performed consistently in trials had small variances; rarer accessions, especially if they performed inconsistently in trials, had large variances. Finally, values for accessions that were originally eliminated from the core set were incorporated through linear regression into the common score axis. The main assumption underlying this method is that there is some underlying "true" axis for the trait of interest that the accessions lie on and that the "noise" (both inherent biological noise and that caused by any accession × trial interaction) can be averaged over. Note that the variance estimates produced represented the variability in final scores (which had several components, such as how often the accession appeared in trials) and did not represent the variability one saw in the field. The scores can be easily converted into rankings, with an accompanying measure of the uncertainty of the ranking.

The PV approach requires that correlation between trials is generally positive (although a few pairs of negatively correlated trials is acceptable, as might occur when one accession has a very unusual score in the trial, as long as these trials are positively correlated when paired with the other trials in the dataset). This allows PCA to merge the data from the trials, and as trials are added and resampling is performed, all accessions are forced into a linear relationship (Ehlenfeldt et al., 2010). The PV approach also requires that a pair of trials has at least three overlapping accessions to calculate a principal component axis (if not, the trials still get used, they just get merged with other trials until sufficient overlap of accessions exists to start merging the intermediate projected value scores). The resampling method used by PV allows estimating stability of rating and ranking results (Supplemental File S4) for each accession that was tested in at least two trials.

Combining Data by the Rank Aggregation Approach

The rank-aggregating approach uses a relative ranking of accessions in individual trials to calculate the overall performance rating (Simko and Linacre, 2010; Simko and Pechenick, 2010). Calculations were performed using the polytomous partial credit model for rank-ordered data (Linacre, 1992; Masters, 1982; Wright and Masters, 1982) that is an extension of the Rasch model (Rasch, 1993). The model assumes that there are as many score categories as accessions in each trial and that each score category is occupied by a single accession (or multiple accessions, if tied rankings are present) (Simko and Linacre, 2010; Simko and Pechenick, 2010). Overall performance of an accession was calculated from the accession's performance in multiple trials, the mean performance of other accessions included in those trials, and a category calibration:

$$\log (P_{kij}/P_{ki(j-1)}) = B_k - D_i - F_{ij},$$

in which P_{kij} is the probability that accession k tested in trial i is observed in category j, $P_{ki(j-1)}$ is the probability that accession k tested in trial i is observed in category $j - 1$, B_k is the performance of the accession k, D_i is the mean performance of the accessions included in trial i (the difficulty measure of trial i), and F_{ij} is the calibration of category j relative to category $j - 1$ in trial i. All analyses were performed using the Winsteps 3.65.0 computer program (Linacre, 2008). Winsteps calculates a central estimate of performance for each accession and calibrations for trials and categories based on performance of all accessions tested in these trials. The calculation is repeated several times and in each iteration, a new performance of an accession is calculated from the performance and calibration values obtained in the previous iteration. Calculations are repeated until the estimation procedure has reached an acceptable level of convergence, or the maximum number of iterations has been reached, or the estimates are not improving. Detailed explanations of the calculation and the parameters are described in the Winsteps manual. Only accessions that were tested in two or more trials were included in this data analysis.

Normalization of Ratings, Calculation of the Mean of Normalized Values, and Detection of Outliers

Both the projected value and the rank-aggregating methods produce ratings on latent scales. To compare results calculated by the two approaches ratings were normalized as

$$z = (X - \bar{X})/s,$$

in which X is the rating value to be normalized, $\bar{X}$ is the arithmetic mean of all rating values, and s is the standard deviation of all rating values. Because values estimated by PV and RA approaches were highly correlated, we averaged normalized ratings of PV and RA to calculate the mean of normalized values (MNV). The MNV was the measure we used to identify the best and the worst performing accessions.

Suspected outliers and outliers were identified through the interquartile range (IQR) approach (Tukey, 1977), in which IQR is the difference between the first and the third quartiles. Suspected outliers are values 1.5 × IQR either below the first quartile or above the third quartile. Values more than 3 × IQR below the first or above the third quartile were defined as outliers.

Mean Correlation Coefficient

To estimate a mean correlation coefficient among trials, individual correlations between pairs of trials were calculated and then transformed using Fisher's r-to-z transformation (Corey et al., 1998):

$$z_i = (1/2)\ln[(1 + r_i)/(1 - r_i)],$$

in which r_i is the correlation coefficient between a pair of trials. A weighted z' score was calculated from individual z_i scores and the respective sample size N_i:

$$z' = \Sigma(N_i z_i)/\Sigma N_i.$$

The average z' score was subsequently back transformed to obtain the mean r' value:

$$r' = [e^{2z'} - 1]/[e^{2z'} + 1].$$

Only pairs of trials that had at least 50 accessions in common were used in this analysis.

RESULTS

Rating of Lettuce Postharvest Quality

The rate of deterioration was evaluated at least twice on 178 accessions tested in 18 trials. However, because not every accession was tested in every trial, the accessions × trials matrix contained only 728 data points, which is approximately 22.7% of all data points in the complete matrix (178 × 18 = 3204). Normalized ratings for the projected value approach ranged from −2.43 (the best shelf life) to 1.81 (the worst shelf life) (Fig. 1). For the rank-aggregating approach, normalized ratings ranged from –2.26 to 2.57 (Fig. 1). Although ratings of accessions by the two approaches were not identical (Supplemental Table S1), the overall correlation between the PV and RA ratings was high ($r = 0.803$, $p < 0.0001$) (Fig. 2). The MNV of the two ratings was used to identify accessions with the best and the worst shelf life.

Approximately 70% of accessions had MNV in the range of–1 to 1; 13% (23 accessions) had MNV below –1 (the slowest rate of deterioration) whereas 17% (31 accessions) had MNV above 1 (the fastest rate of deterioration) (Fig. 1). A group of the best performing accessions (MNV < −1) tested in at least two trials included eight cultivars: Calmar, Autumn Gold, Tiber, Silverado, Glacier, Salinas 88, Darkland, and Salinas. With the exception of Darkland, which is a romaine lettuce, all other cultivars are iceberg type lettuces (Table 1). A morphologically more diverse group of 10 cultivars had the worst shelf life (MNV > 1) and consisted of four romaine lettuces (Bandit, Dark Green Romaine, Lee Tal, and Triple Threat), three Latin type lettuces (Pavane, Little Gem, and Barnwood Gem), two butterhead lettuces (Cobham Green and Tinto), and a single Batavia type lettuce (La Brillante). The overall best MNV of –2.21 was calculated for the iceberg breeding line RH09-1700 tested in two trials whereas the worst MNV of 1.89 was calculated for Latin-type cultivar Pavane tested in four trials (Fig. 3; Supplemental Table S1). No outliers or suspected outliers were identified in the PV, RA, and MNV data for shelf life (Fig. 1).

Rating Of Lettuce Resistance to Downy Mildew

Resistance to downy mildew was assessed at least twice on 582 accessions tested in 53 trials. The accessions × trial matrix contained only 1996 data points, which is approximately 6.5% of all data points in the complete matrix (582 × 53 = 30,846) (Fig. 4). Normalized ratings ranged from

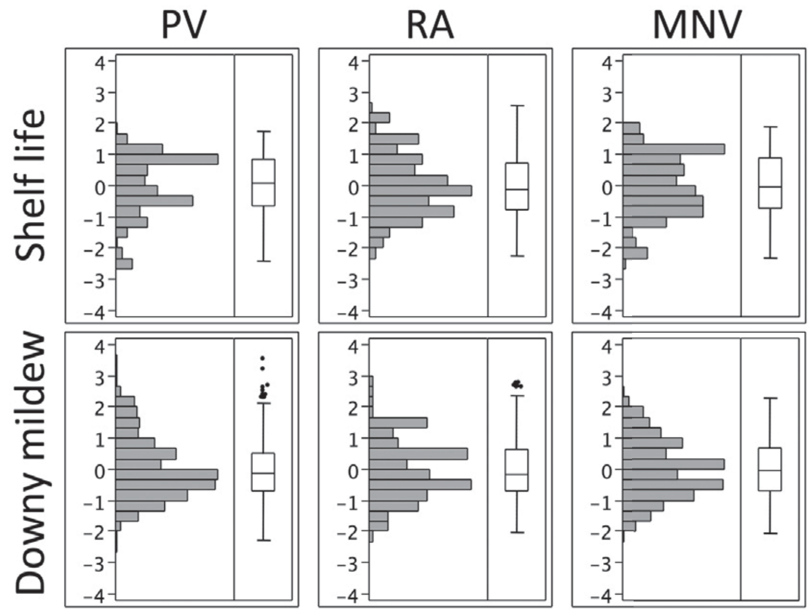

Figure 1. Distribution of the normalized values for shelf life (top row, 178 accessions) and downy mildew resistance (bottom row, 582 accessions) calculated with the projected values (PV) approach and the rank-aggregation (RA) approach. The last column shows distribution of the mean of normalized values (MNV) calculated from averaging the two approaches. The left part of each panel shows the frequency of normalized values and their distribution. The right part of each panel shows a boxplot diagram. The box identifies the first and the third quartile, the band in the middle of the box is the median. The ends of the whiskers represent the minimal and the maximal values. If suspected outliers were identified, these are plotted as small circles and the end of the respective whiskers represent the end of the 1.5 × interquartile range above the third quartile.

–2.34 (highest resistance) to 3.54 (highest susceptibility) for PV, from –2.13 to 2.79 for RA, and from –2.12 to 2.34 for MNV (Fig. 1). Thirteen accessions (2.2%) in the PV dataset and seven accessions (1.2%) in the RA dataset were identified as suspected outliers, all with a high susceptibility to downy mildew. Although suspected outliers identified in the PV and RA datasets were different, correlation between normalized ratings was high ($r = 0.748$, $p < 0.0001$) (Fig. 2). The MNV of the two ratings was used to identify accessions with the highest resistance and susceptibility to downy mildew.

Approximately 70% of accessions had MNV in the range from –1 to 1; 14% (83 accessions) had MNV below –1 (the highest resistance) whereas 16% (91 accessions) had MNV above 1 (the highest susceptibility) (Fig. 1). Twenty-four cultivars were in the group of accessions with highest resistance (MNV < –1). This group contained lettuces from five horticultural types: Batavia (eight), leaf (six), butterhead (six), stem (three), and iceberg (one). From examined cultivars, the highest level of field resistance to downy mildew was observed in Holborn Standard (Batavia) (–2.11) tested in 14 trials, Iceberg (Batavia) (–1.99) tested in 37 trials, and Lolla Rossa (leaf-type) (–1.81) tested in eight trials. The overall lowest score of –2.12 was observed for the leaf-type accession 04G642 that did not show any disease symptoms in the two trials it was tested in. The group of accessions with the highest susceptibility to downy mildew (MNV > 1) included 22 cultivars from six types of lettuces: iceberg (nine), romaine (four), Latin (three), butterhead (three), stem (two), and Batavia (one). The highest susceptibility was recorded for cultivars Pavane (Latin) (2.24) tested in seven trials, Celtuce (stem) (2.22) tested in three trials, Vista Verde (iceberg) (2.12) tested in four trials, Da Ye Wo Sun (stem) (1.90) tested in five trials, and Sturgis (romaine) (1.86) tested in three trials. The overall highest score indicating the highest level of susceptibility to downy mildew was calculated for the romaine-type PI 491224 (2.34) tested in four trials.

We observed large differences among horticultural types of lettuces in their reaction to *B. lactucae* infection. If cultivars with an MNV < –1 are labeled “resistant” and with an MNV > 1 “susceptible,” lettuce types with the highest ratio of resistant to susceptible cultivars were leaf (six vs. zero), Batavia (eight vs. one), butterhead (six vs. three), and stem (three vs. two) lettuces. The lowest ratios were observed for iceberg (one vs. nine), Latin (zero vs. three), and romaine (zero vs. four) types (Table 1).

Accessions Tested for Both Traits

One hundred and forty-eight accessions were tested for both traits in at least two trials. The correlation was not significant between the MNV for shelf life and downy mildew ($r = -0.150$, $p = 0.0698$) (Fig. 3). Some accessions showed a desirable combination of both traits (good shelf life and high resistance to downy mildew) whereas others performed poorly when tested for these traits. The group of accessions with low MNV (favorable combination) for both traits included Batavia cultivars, Holborn Standard, Iceberg, and Batavia Reine de Glaces, stem-type cultivar

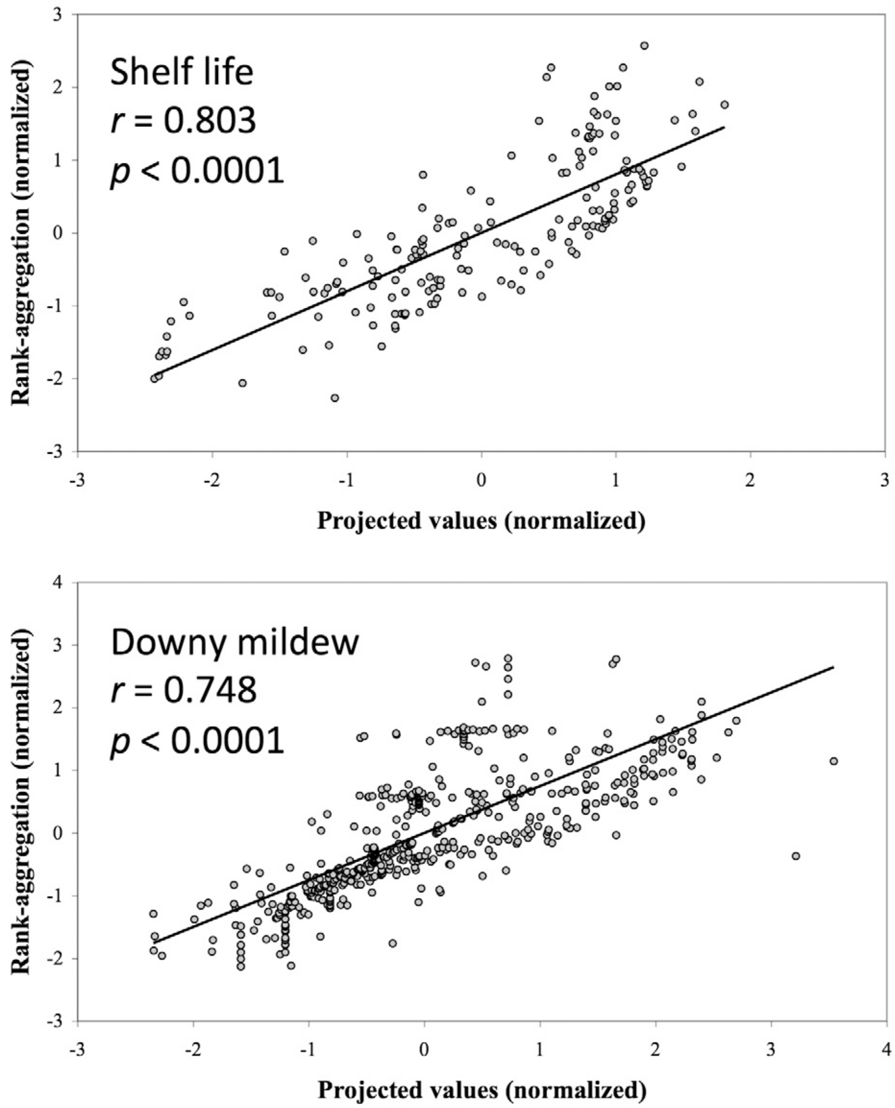

Figure 2. Correlation between the normalized values calculated by the projected values approach and the rank-aggregation approach. The top panel shows values for 178 accessions tested for shelf life and the bottom panel shows values for the 582 accessions tested for resistance to downy mildew.

Table 1. Distribution of the postharvest quality and the resistance to downy mildew in lettuce cultivars from different horticultural types.

Type	Good shelf life MNV† < –1	Poor shelf life MNV > 1	High resistance MNV < –1	High susceptibility MNV > 1
Batavia	0‡	1	8	1
Butterhead	0	2	6	3
Iceberg	7	0	1	9
Latin	0	3	0	3
Leaf	0	0	6	0
Romaine	1	4	0	4
Stem	0	0	3	2
Total	8	10	24	22

†MNV, mean of normalized values.

‡Values correspond to the number of cultivars from each horticultural type that belongs to the respective category. The number of cultivars that belong to the intermediate category (MNV from –1 to 1) is not shown.

Balady Banha, and the recombinant inbred line RH08-0112 originating from the Salinas 88 × La Brillante mapping population (Fig. 3). In the group that included accessions with both poor shelf life and high susceptibility to downy mildew were three Latin-type cultivars (Pavane, Little Gem, and Barnwood Gem), two romaine accessions

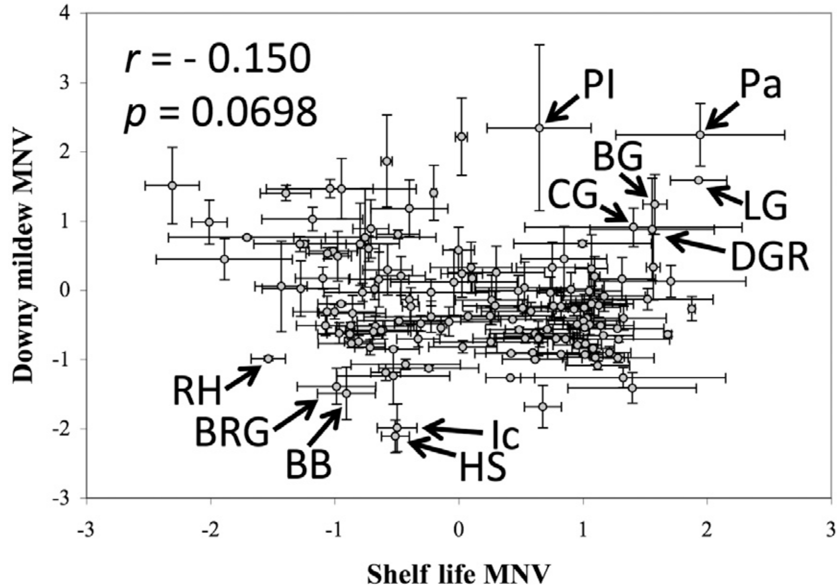

Figure 3. Scatter plot of the mean of normalized values (MNV) for 148 accessions that were tested for both shelf life and resistance to downy mildew. Circles show MNV and whiskers indicate the normalized values calculated by the projected value approach and the rank-aggregating approach. For example, the normalized values for shelf life of accessions PI 491224 are projected values (PV) = 0.22, rank aggregation (RA) = 1.06, and MNV = 0.64. Therefore the circle on *x*-axis is at 0.64 and the whiskers extent from 0.22 to 1.06. Similarly, the values for the disease resistance are PV = 3.54, RA = 1.15, and MNV = 2.34; therefore, the circle on *y*-axis is at 2.34 and the whiskers indicate interval from 1.15 to 3.54. Accessions having desirable values for both traits (good shelf life and high resistance) are close to the lower left corner (–3, –3 value). Accessions having undesirable values for both traits (poor shelf life and high susceptibility) are close to the upper right corner (3, 4 value). Abbreviations for accessions are RH, RH08-0112 (recombinant inbred line); BRG, Batavia Reine de Glaces (Batavia type); BB, Balady Banha (stem type); HS, Holborn Standard (Batavia type); Ic, Iceberg (Batavia type); PI, PI 491224 (romaine type); Pa, Pavane (Latin type); BG, Barnwood Gem (Latin type); LG, Little Gem (Latin type); CG, Cobham Green (butterhead type); and DGR, Dark Green Romaine (romaine type).

(PI 491224 and Dark Green Romaine), and a single butterhead cultivar (Cobham Green) (Fig. 3; Table 1).

DISCUSSION

Rating of Lettuce Postharvest Quality

The mean correlation among postharvest quality trials was $r' = 0.705$ ($p < 0.001$), indicating a limited accession × trial interaction. Combining phenotypic observations from 18 trials allowed identification of accessions with either slow or fast rate of deterioration. Most of the best performing cultivars (seven out of eight) were iceberg type lettuces (Autumn Gold, Calmar, Glacier, Salinas, Salinas 88, Silverado, and Tiber) with firm, compact heads and overall high phenotypic (Simko et al., 2011) and genetic (Simko, 2009) similarity. There were three Latin cultivars and two butterhead cultivars in the group of accessions with the fastest deterioration (MNV > 1) and none in the group with the slowest deterioration (MNV < –1). Combined results from 18 trials indicated that both of these types had generally poor shelf life when processed for salad (Supplemental Table S1).

Based on F_8 RILs from the Salinas 88 × La Brillante mapping population tested in multiple trials, we approximated the number of days to 100% deterioration. Although differences among trials existed, accessions with the fastest deterioration (MNV > 1) reached 100% on average in about 25 to 30 d whereas accessions with the best shelf life (MNV < –1) reached 100% deterioration in approximately 65 to 90 d. Our results confirmed very large phenotypic variability in the postharvest quality of lettuce (Hayes and Liu, 2008) that can be used to develop material with improved shelf life (Simko et al., 2010).

One of the shelf life evaluation methods (method C) was used to observe deterioration process until profound cell lysis was evident (Fig. 5). Such highly deteriorated salad cannot be used for human consumption; however, evaluating the dynamics of deterioration allows the study of mechanisms that are not evident from assessments at a single time point. For example, two different genes may autonomously initiate the beginning of deterioration at a similar time but at a dissimilar progress rate. Thus, detailed analysis of deterioration progress allows detection of genes involved in postharvest quality (R.J. Hayes and I. Simko, unpublished data, 2012) that can be targeted for marker-assisted selection. Overall, the AUDePS of deterioration significantly correlated ($r = 0.874$, $p < 0.0001$) with the number of days that salad was suitable for retail (rating of 0 on a scale of 0–10). Moreover, the AUDePS score significantly correlated ($r = 0.841$, $p = 0.0045$) in a blind experiment with evaluation of postharvest quality performed by Fresh Express (Salinas,

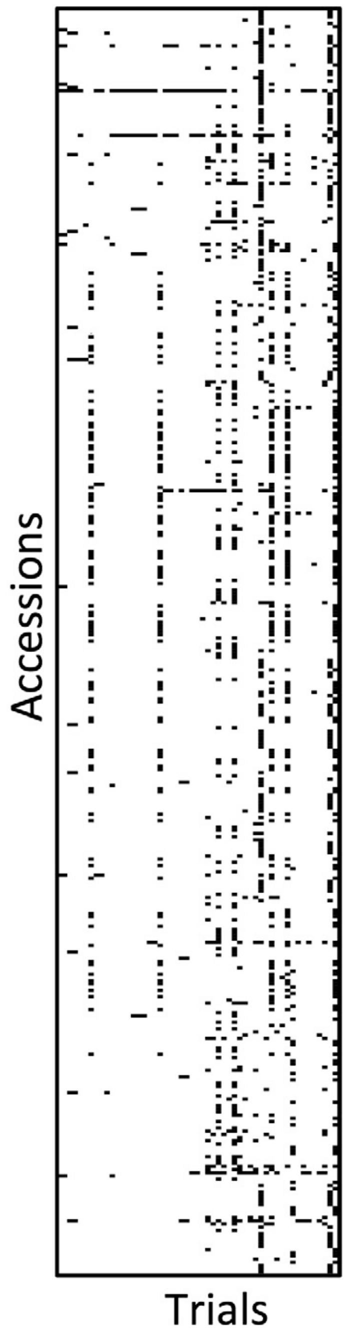

Figure 4. Distribution of 1996 data points that were used to calculate overall ratings of lettuce resistance to downy mildew. Columns represent 53 trials, rows show information for 582 accessions, and black squares indicate distribution of data points. If every accession were tested in every trial, the complete matrix would contain 582 × 53 = 30,846 data points.

CA), a major salad-processing company (R.J. Hayes and I. Simko, unpublished data, 2009).

Rating of Lettuce Resistance to Downy Mildew

Resistance of lettuce to downy mildew is conferred either by major genes (qualitative resistance), or minor genes (quantitative resistance), or a combination of the two (Grube and Ochoa, 2005; McHale et al., 2009). Minor genes usually provide less resistance but the level of resistance conferred is similar against most of the races of *B. lactucae* (Crute and Norwood, 1981). Major genes, on the other hand, usually provide complete resistance against specific races of downy mildew but are rendered completely ineffective by virulent races of the pathogen (Crute, 1998; Grube and Ochoa, 2005). We did not observe any pathogen race × accession interactions in which an accession would be disease free in one trial but highly susceptible in another trial. The mean correlation between trials was $r' = 0.342$ ($p < 0.001$). All isolates of downy mildew collected from our field trials were fully avirulent to *Dm17* and *Dm38* resistance genes (data not shown).

The most resistant accession from those that were tested in at least two trials was 04G642 (MNV = –2.12) that is used as a differential to identify avirulence genes in *B. lactucae* isolates. Resistance in this line is based on *Dm17*, a major dominant gene that still provides resistance against most of the downy mildew races detected in the Salinas-growing region (Michelmore et al., 2011). If races that defeat the *Dm17* gene were to appear in our trials, resistance in 04G642 would likely be substantially or completely reduced. On the other hand, the high level of resistance observed in the cultivars Holborn Standard, Iceberg, Lolla Rossa, Grand Rapids, and some others appears to be conferred mostly by a race-nonspecific combination of genes with minor effects. In almost all trials, small amounts of infection were observed on these cultivars, but infection was usually limited to relatively few, small lesions with a minimal sporulation of the pathogen. Our observations are in line with previous studies that did not find any evidence of race-specific resistance in cultivars Grand Rapids and Iceberg (Norwood et al., 1983; Grube and Ochoa, 2005).

Comparison of different horticultural types of lettuces revealed that their resistance to *B. lactucae* infection substantially varies. Some lettuce types, such as leaf and Batavia, comprise several cultivars with high resistance while no cultivar from iceberg, Latin, and romaine type was included in this group (Table 1). The single exception was the iceberg cultivar Ice Cube with a relatively high resistance to downy mildew (Supplemental Table S2). However, this cultivar is a "mini" lettuce that was not developed for a large-scale commercial production. Because genetic and phenotypic variability within romaine and especially iceberg types is relatively small (Simko, 2009, Simko et al., 2009), our analysis indicated that novel genes for quantitative resistance against downy mildew need to be introgressed into these genepools from other horticultural types or lettuce species. Therefore we are developing iceberg and romaine breeding lines with resistance genes originating from several Batavia, stem, and leaf-type accessions.

Figure 5. Example of postharvest deterioration that is accompanied by a profound cell lysis. Material was processed for salad and stored at 3.5°C for 4 wk. Breeding line SM09A (left) (Simko et al., 2010) shows slow deterioration while PI 491224 that is a parent of SM09A shows a very high rate of deterioration with most of the cells already disrupted.

Comparison of the Results from the Projected Values and Rank-Aggregation Methods

When testing all accessions simultaneously is not practicable or possible, combining phenotypic data from multiple trials into the overall rating is an effective approach. We used two relatively new approaches: PV (Ehlenfeldt et al., 2010) and RA (Simko and Linacre, 2010; Simko and Pechenick, 2010) as well as averaging across them to identify accessions with the slowest rate of deterioration after processing for salad and the highest resistance to downy mildew.

Rank-aggregation based methods have previously been used for combining data from biological experiments recorded on different rating scales (Simko and Linacre, 2010; Simko and Pechenick, 2010; Lin, 2011). This approach requires a certain level of difference in rankings in different trials to construct linear measures from ranked data. Perfectly ordered observations do not provide enough information to calculate distances between accessions (Linacre, 1992; Simko and Linacre, 2010). The advantage of RA-based methods is that they are invariant to transformation and normalization (as long as relative ordering is preserved), independent of rating scales, distribution free, robust to outliers, and require only two accessions per trial as minimum (DeConde et al., 2006; Conlon et al., 2007; Simko and Pechenick, 2010; Lin, 2011). However, when ranked data are used to calculate an overall rating, some information is inevitably lost compared with effect size-based methods (Choi et al., 2003; Lin, 2011), such as PV.

Although the PV and RA methods use different approaches to calculate overall rating, we found a high correlation between the final estimates. The correlation between the PV and RA ratings calculated for shelf life data on 178 accessions was r = 0.803 ($p < 0.0001$) whereas the correlation for downy mildew data on 582 accessions was r = 0.748 ($p < 0.0001$). Similar results were reported for approaches used in web metasearch, where rank-based methods performed comparably to score-based methods (Renda and Straccia, 2003) even though the underlying assumptions differed. The largest differences in ratings estimated by the PV and RA methods were observed when an accession performed very well or very poorly in some trial(s). In such cases, ratings based on PV were usually more distal on a normalized scale whereas those based on RA were closer to the center of a normalized scale. For example, the largest difference between the PV and RA normalized values for resistance to downy mildew was observed for the cultivar White Paris (PV 3.21, RA –0.36) and for the accession PI 491224 (PV 3.54, RA 1.15). White Paris, tested twice, performed very poorly in one of the trials. Therefore, it was not included in the core set of accessions and its PV value was calculated from linear regression on the common score axis. Plant introduction 491224 was tested four times and it also performed very poorly in one of the trials. The difference between the PV and RA ratings reflected the fact that RA did not take into consideration actual values, only relative rankings. However, the opposite situation, when PV values were closer to the center of a normalized scale than RA values, can also happen. For example, the largest difference between the PV and RA normalized values for shelf life was observed for the cultivars Cobham Green (PV 0.519, RA 2.268) and Tinto (PV 0.486, RA 2.136). Both of these cultivars were tested twice and consistently ranked among the worst performing accessions in those trials. However, their shelf life values were not substantially different from other poorly performing accessions; therefore, the absolute values of RA for both cultivars were higher than their respective PV values.

In the present work, only accessions tested in at least two trials were used for analyses. However, 155 more accessions were tested only once for shelf life, and 1266 more accessions were tested only once for resistance to downy mildew (data not shown). Although observations from only a single experiment do not provide enough information, these data were used to identify accessions with potentially high resistance to downy mildew and good shelf life. These results need to be confirmed in additional experiments. When estimates based on a single experiment were added to datasets, the correlation between the PV and RA ratings calculated for shelf life data on 333 accessions was $r = 0.677$ ($p < 0.0001$). The correlation for downy mildew data on 1848 accessions increased to $r = 0.790$ ($p < 0.0001$), indicating a good match for estimates calculated by the PV and RA approaches.

Limitations of the Projected Values and Rank-Aggregation Methods

Although both the PV and RA methods can calculate ratings and rankings from sparse accession × trait matrices, there are certain limitations that need to be considered before combining heterogeneous phenotypic data across multiple experiments. These are

1. All accessions that are to be placed into a common linear scale need to be compared, either directly or indirectly, through intermediary accessions.
2. Both approaches assume transitivity of data; if accession A outperforms accession B, and accession B outperforms accessions C, it is assumed that accession A also outperforms accession C.
3. Results of evaluations based on different rating scales have to be providing similar information. Data that were used to combine both shelf life and downy mildew evaluations into the respective overall ratings originate from different ratings scales. However, results of evaluations based on different rating scales (for both traits) were strongly correlated when comparative testing was performed in selected trials (I. Simko, unpublished data, 2010). Before combining data from different rating scales, it is critical to ensure that all rating scales evaluate the same trait. If individual rating scales provide dissimilar information, combining data from these different scales should not be performed.
4. Negative correlation between a pair of trials. Combining a pair of trials with many overlapping accessions that are negatively correlated is a conceptual problem. Essentially, that means that accessions that perform well in one trial may perform relatively poorly in another trial and vice versa. If such difference in relative rankings exists, calculating an overall ranking from negatively correlated trials may have limited meaning. However, if combining trials with negative correlation is deemed necessary, RA approach can be used.
5. Accession × trial interactions can significantly affect the relative ranking of accessions. If the interaction is strong, it is not possible to estimate values for missing observations reliably. For example, assume that resistance to downy mildew was tested on a set of accessions, each with a different single resistance gene. If different trials were infected with races of *B. lactucae* having different avirulence genes, a significant accession × trial interaction would occur. If the interaction effect is large, it will mask the additive effect of accession and make apparent resistance a function of which trial the accession appeared in (i.e., the same accession was highly rated in one trial and poorly rated in another). This situation did not happen often in our trials, which had significant and positive mean correlations (shelf life $r' = 0.705$, $p < 0.001$, and downy mildew $r' = 0.342$, $p < 0.001$). Therefore, observations from individual trials could be combined into the overall rating. However, it is imperative to proceed with caution when combining data from trials where a strong accession × trial interaction might exist, such as disease-resistance trials (example of a strong interaction that precludes combining of data is shown in Simko and Linacre, 2010).
6. An effect of consistently the best or worst performing accessions. When an accession performed consistently the best (or worst) in all trials it was tested in, then calculation of the overall rating for this accession would not be possible by the RA approach. To avoid this kind of problem, a small value is added to each ranking before calculations (Simko and Linacre, 2010; Simko and Pechenick, 2010). This modification has a relatively small effect on the overall rating scores, and as the number of comparisons among accessions increases the effect becomes negligible. The PV approach avoids instability caused by consistently the best (or the worst) performing accessions by eliminating these values from the core set. Once estimates are calculated from the core set, the eliminated accessions are added back to the dataset and their overall rating is calculated.

CONCLUSIONS

In recent years, there has been considerable interest in adopting methods for combining data from different biological trials or experiments (Lin, 2011). If the objective is to compute an overall linear rating that estimates performance of all accessions across a set of trials, such ratings can be constructed even for accessions that were never tested together and their evaluations were performed on different scales. Combining data from multiple experiments into a single rating provides more reliable results than individual studies (Lin, 2011). For example, marker–trait associations tend to be more significant when calculated from combined data than from each of the individual trials (Simko and Pechenick, 2010). Both PV and RA statistical approaches can also be used to combine data from different laboratories (accession × laboratory matrix) or databases (accession × database matrix), thus improving the major bottleneck in obtaining phenotypic data for large-scale studies (e.g., association mapping).

We combined evaluations of lettuce postharvest quality and field resistance to downy mildew from multiple trials using two approaches, PV and RA. There was high correlation between the two approaches for both postharvest quality and resistance to downy mildew. Combining data from multiple experiments allowed us to detect accessions with high disease resistance and a slow rate of deterioration. Accessions with desirable traits identified in this study are being incorporated into our breeding programs. In addition, our analysis indicates that there are substantial differences among horticultural types of lettuce for both traits. For example, iceberg cultivars tend to have a slow rate of deterioration and high susceptibility to downy mildew; Latin-type cultivars are generally highly susceptible to downy mildew and deteriorate quickly after processing for salad whereas several Batavia cultivars show good resistance to downy mildew. The shortage of well-performing cultivars from some horticultural types indicates that, to achieve improvement in shelf life and/or resistance to downy mildew, novel genes (or alleles) need to be introgressed into these types, for example, from other well-performing types or wild lettuce species.

Supplemental Information Available

Supplemental material is available at http://www.crops.org/publications/cs.

Supplemental Table S1. Rating 178 accessions for the rate of deterioration after processing for salad. Columns show the number of trials each accessions was tested, normalized ratings calculated by the projected values (PV) and rank-aggregation (RA) methods, and the mean of the two normalized values (MNV). Lower ratings indicate a slower rate of deterioration.

Supplemental Table S2. Rating 582 accessions for the resistance to downy mildew. Columns show the number of trials each accessions was tested, normalized ratings calculated by the projected values (PV) and rank-aggregation (RA) methods, and the mean of the two normalized values (MNV). Lower ratings indicate a higher resistance.

Supplemental File S3. The R-code for calculating combined ratings by the projected values approach. More information about the code can be obtained directly from Matthew Kramer (matt.kramer@ars.usda.gov).

Supplemental File S4. Example of resampling results calculated for cultivars Hilde, Holborn Standard, and Ice Cube by the projected values approach. Panels on the left show projected rankings while those on the right show projected ratings for resistance to downy mildew before normalization. Cultivar Hilde was tested in six trials, Holborn Standard in 14 trials, and Ice Cube in five trials. Lower ranking and rating indicate a higher resistance to the disease. Bars that are close to each other (e.g., for Holborn Standard rankings or Ice Cube ratings) indicate small variability of resampled estimates.

Acknowledgments

The authors thank Amy Atallah and Jose Orozco for assistance with evaluations of the trials. The California Leafy Greens Research Program and the Arizona Department of Agriculture, Agricultural Consultation and Training, have funded a portion of this project. The views and findings presented are the Grantee's and do not necessarily represent those of the State or the Arizona Department of Agriculture. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

References

Choi, J.K., U. Yu, S. Kim, and O.J. Yoo. 2003. Combining multiple microarray studies and modeling interstudy variation. Bioinformatics 19:i84–i90. doi:10.1093/bioinformatics/btg1010

Conlon, E.M., J.J. Song, and A. Liu. 2007. Bayesian meta-analysis models for microarray data: A comparative study. BMC Bioinformatics 8:80.

Corey, D.M., W.P. Dunlap, and M.J. Burke. 1998. Averaging correlations: Expected values and bias in combined Pearson *r*s and Fisher's *z* transformations. J. Gen. Psychol. 125:245–261. doi:10.1080/00221309809595548

Crute, I.R. 1984. The integrated use of genetic and chemical methods for control of lettuce downy mildew (*Bremia lactucae* Regel). Crop Prot. 3:223–241. doi:10.1016/0261-2194(84)90057-7

Crute, I.R. 1992. The role of resistance breeding in the integrated control of downy mildew (*Bremia lactucae*) in protected lettuce. Euphytica 63:95–102. doi:10.1007/BF00023915

Crute, I.R. 1998. The elucidation and exploitation of gene-for-gene recognition. Plant Pathol. 47:107–113. doi:10.1046/j.1365-3059.1998.00232.x

Crute, I.R., and J.M. Norwood. 1981. The identification and characteristics of field resistance to lettuce downy mildew *(Bremia lactucae* Regel). Euphytica 30:707–717. doi:10.1007/BF00038799

DeConde, R.P., S. Hawley, S. Falcon, N. Clegg, B. Knudsen, and R. Etzioni. 2006. Combining results of microarray experiments: A rank aggregation approach. Stat. Appl. Genet. Mol. Biol. 5:i-23.

Ehlenfeldt, M.K., J.J. Polashock, A.W. Stretch, and M. Kramer. 2010. Ranking cultivated blueberry for mummy berry blight and fruit infection incidence using resampling and principal components analysis. HortScience 45:1205–1210.

Grube, R.C., and O.E. Ochoa. 2005. Comparative genetic analysis of field resistance to downy mildew in the lettuce cultivars 'Grand Rapids' and 'Iceberg'. Euphytica 142:205–215. doi:10.1007/s10681-005-1683-3

Hayes, R.J., and Y.B. Liu. 2008. Genetic variation for shelf-life of salad-cut lettuce in modified-atmosphere environments. J. Am. Soc. Hortic. Sci. 133:228–233.

Hayes, R.J., L.K. McHale, G.E. Vallad, M.J. Truco, R.W. Michelmore, S.J. Klosterman, K. Maruthachalam, and K.V. Subbarao. 2011. The inheritance of resistance to Verticillium wilt caused by race 1 isolates of *Verticillium dahliae* in the lettuce cultivar La Brillante. Theor. Appl. Genet. 123:509–517. doi:10.1007/s00122-011-1603-y

Kim, J.G., Y. Luo, Y. Tao, R.A. Saftner, and K.C. Gross. 2005. Effect of initial oxygen concentration and film oxygen transmission rate on the quality of fresh-cut romaine lettuce. J. Sci. Food Agric. 85:1622–1630. doi:10.1002/jsfa.2158

Lin, S. 2011. Rank aggregation methods. WIREs. Comput. Stat. 2:555–570.

Linacre, J.M. 1992. Objective measurement of rank-ordered objects. In: M. Wilson, editor, Objective measurement: Theory into practice. Ablex, Norwood, NJ. p. 195–209.

Linacre, J.M. 2008. Winsteps computer programs for Rasch-model, version 3.65.0. Winsteps, Beaverton, OR.

Masters, G.N. 1982. A Rasch model for partial credit scoring. Psychometrika 47:149–174. doi:10.1007/BF02296272

McHale, L.K., M.J. Truco, A. Kozik, T. Wroblewski, O.E. Ochoa, K.A. Lahre, S.J. Knapp, and R.W. Michelmore. 2009. The genomic architecture of disease resistance in lettuce. Theor. Appl. Genet. 118:565–580. doi:10.1007/s00122-008-0921-1

Michelmore, R.W., M.J. Truco, and O.E. Ochoa. 2011. Breeding crisphead and leafy lettuce. California leafy green research program. April 1, 2010 to March 31, 2011. California leafy green research program, Salinas, CA. http://calgreens.org/control/uploads/Michelmore_Breeding_Crisphead_and_Leafy_Lettuce.pdf (accessed 26 March 2012).

Norwood, J.M., I.R. Crute, A.G. Johnson, and P.L. Gordon. 1983. A demonstration of the inheritance of field resistance to lettuce downy mildew (*Bremia lactucae* Regel.) in progeny derived from cv. Grand Rapids. Euphytica 32:161–170. doi:10.1007/BF00036876

Piepho, H.P. 2003. Model-based mean adjustment in quantitative germplasm evaluation data. Genet. Resour. Crop Evol. 50:281–290. doi:10.1023/A:1023503900759

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org (accessed 14 Nov. 2011).

Rasch, G. 1993. Probabilistic models for some intelligence and attainment tests. MESA Press, Chicago, IL.

Renda, M.E., and U. Straccia. 2003. Web metasearch: Rank vs. score based rank aggregation methods. In: Proceedings of the 2003 Association for Computing Machinery (ACM) Symposium on Applied Computing, Melbourne, FL. 9–12 Mar. 2003. ACM, New York, NY. p. 841–846.

Simko, I. 2009. Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.). J. Hered. 100:256–262. doi:10.1093/jhered/esn072

Simko, I., R.J. Hayes, K.V. Subbarao, and R.G. Sideman. 2010. SM09A and SM09B: Romaine lettuce breeding lines resistant to dieback and with improved shelf life. HortScience 45:670–672.

Simko, I., R.J. Hayes, M.J. Truco, and R.W. Michelmore. 2011. Mapping a dominant negative mutation for triforine sensitivity in lettuce and its use as a selectable marker for detecting hybrids. Euphytica 182:157–166. doi:10.1007/s10681-011-0407-0

Simko, I., and J.M. Linacre. 2010. Combining partially ranked data in plant breeding and biology: II. Analysis with Rasch model. Commun. Biometry Crop Sci. 5:56–65.

Simko, I., and D.A. Pechenick. 2010. Combining partially ranked data in plant breeding and biology: I. Rank aggregating methods. Commun. Biometry Crop Sci. 5:41–55.

Simko, I., D.A. Pechenick, L.K. McHale, M.J. Truco, O.E. Ochoa, R.W. Michelmore, and B.E. Scheffler. 2009. Association mapping and marker-assisted selection of the lettuce dieback resistance gene *Tvr1*. BMC Plant Biol. 9:135. doi:10.1186/1471-2229-9-135

Simko, I., and H.P. Piepho. 2011. Combining phenotypic data from ordinal rating scales in multiple plant experiments. Trends Plant Sci. 16:235–237. doi:10.1016/j.tplants.2011.02.001

Simko, I., and H.P. Piepho. 2012. The area under the disease progress stairs (AUDPS): Calculation, advantage, and application. Phytopathology 102:381–389. doi:10.1094/PHYTO-07-11-0216

Smyth, A.B., J. Song, and A.C. Cameron. 1998. Modified atmosphere packaged cut iceberg lettuce: Effect of temperature and O_2 partial pressure on respiration and quality. J. Agric. Food Chem. 46:4556–4562. doi:10.1021/jf980208s

Tukey, J.W. 1977. Exploratory data analysis. Addison-Wesley, Reading, MA.

Wright, B.D., and G.N. Masters. 1982. Rating scale analysis. MESA Press, Chicago, IL.