Journal of Cereal Science 74 (2017) 19–27

Contents lists available at ScienceDirect

Journal of Cereal Science

journal homepage: www.elsevier.com/locate/jcs

Evaluation of wheat and emmer varieties for artisanal baking, pasta making, and sensory quality

Lisa Kissing Kucek [a], Elizabeth Dyck [b], June Russell [c], Liz Clark [d], Jeffrey Hamelman [e], Sharon Burns-Leader [f], Stefan Senders [g], Jenny Jones [h], David Benscher [a], Michael Davis [a], Greg Roth [i], Steve Zwinger [j], Mark E. Sorrells [a], J.C. Dawson [k, *]

[a] Cornell University, Plant Breeding and Genetics Section, School of Integrative Plant Science, 240 Emerson Hall, Ithaca, NY 14853, USA
[b] Organic Growers' Research and Information-Sharing Network, 1124 County Rd 38, Bainbridge, NY 13733, USA
[c] Greenmarket, GrowNYC, 100 Gold Street, Suite 3300, New York, NY 10038, USA
[d] Gimme! Coffee, 506 West State Street, Ithaca, NY 14850, USA
[e] King Arthur Flour, 135 US Route 5 South, Norwich, VT 05055, USA
[f] Bread Alone Bakery, 2121 Ulster Avenue, Lake Katrine, NY 12449, USA
[g] Wide Awake Bakery, 4361 Buck Hill Rd. S., Trumansburg, NY 14886, USA
[h] Gramercy Tavern, 42 E. 20th Street, New York, NY 10003, USA
[i] Pennsylvania State University, Department of Plant Science, College of Agricultural Sciences, 407 Agricultural Sciences and Industries Building, University Park, PA 16802, USA
[j] North Dakota State University, Carrington Research and Extension Center, 663 Hwy. 281 NE, PO Box 219, Carrington, ND 58421, USA
[k] University of Wisconsin-Madison, Department of Horticulture, 1575 Linden Drive, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 23 June 2016
Received in revised form
14 December 2016
Accepted 28 December 2016
Available online 2 January 2017

Keywords:
Wheat
Variety evaluation
Sensory analysis
Baking quality

ABSTRACT

Identifying varieties best suited to local food systems requires a comprehensive understanding of varietal performance from field to fork. After conducting four years of field trials to test which varieties of ancient, heritage, and modern wheat grow best on organically managed land, we screened a subset of varieties for bread, pastry, pasta, and cooked grain quality. The varieties evaluated were three lines of emmer (*T. turgidum* L. ssp. *dicoccum* Schrank ex Schübl) and eleven lines of common wheat (*Triticum aestivum* L.), including two modern soft wheat varieties, four soft heritage wheat varieties, four hard modern wheat varieties, and one hard heritage wheat variety. A diverse group of bakers, chefs, researchers, and consumers compared varieties for qualities of interest to regional markets. Participants assessed differences in sensory profiles, pasta making ability, and baking quality for sourdough, matzah crackers, yeast bread, and shortbread cookies. In addition to detecting significant differences among varieties for pasta, sourdough, and pastry quality, participants documented variation in texture and flavor for the evaluated products. By demonstrating which varieties perform best in the field, in the bakery, and on our taste buds, these results can support recommendations that strengthen the revival of local grain economies.

1. Introduction

Global consumers increasingly demand food that is organic (ERS, 2014) and local (Elbehri, 2007; Low et al., 2015), with fewer additives (Kaptan and Kayisoglu, 2015) and excellent sensory quality (Codron et al., 2005). Bread is a key component of changing consumer demand. Although wheat remains one of the most industrially consolidated food products (Hendrickson and Heffernan, 2007), local food systems are increasingly purchasing wheat produced on nearby farms and sprouting small-scale flour mills and bakeries that have been absent for decades (Halloran, 2015; Hills, 2012). This transition requires the identification of varieties that best support local grain economies.

Previous research has not identified the wheat varieties that are

* Corresponding author.
E-mail addresses: lkk26@cornell.edu (L. Kissing Kucek), edyck@ogrin.org (E. Dyck), jrussell@grownyc.org (J. Russell), liz.clark@gimmecoffee.com (L. Clark), Jeffrey.Hamelman@kingarthurflour.com (J. Hamelman), sharon.burnsleader@breadalone.com (S. Burns-Leader), stefsend@gmail.com (S. Senders), jjones@untitledatthewhitney.com (J. Jones), gwr@psu.edu (G. Roth), steve.zwinger@ndsu.edu (S. Zwinger), dawson@hort.wisc.edu (J.C. Dawson).

http://dx.doi.org/10.1016/j.jcs.2016.12.010
0733-5210/

best suited to the local grain markets of the United States. Local markets focus on organic production, low-extraction stone milling, artisanal sourdough baking, and consumer demand for unique taste. Consumers, bakers, and farmers involved in local and organic grain economies of the United States have also expressed interest in heritage and ancient wheat varieties (Packaged Facts, 2015), in part because some genotypes have demonstrated distinctive flavors and reduced impacts in individuals with wheat sensitivity (Kissing Kucek et al., 2015). The term heritage describes varieties of common wheat (*Triticum aestivum* L.) developed before the use of dwarfing genes in the 1950's, while modern wheat refers to varieties of common wheat developed after that time. Ancient wheat describes hulled relatives of wheat, such as emmer (*T. turgidum* L. ssp. *dicoccum* Schrank ex Schubl). The baking quality of heritage wheat varieties, however, are poorly documented. Moreover, few scientific studies have compared the sensory attributes of different varieties of heritage, ancient, and modern wheat.

Vindras-Fouillet et al. (2014) found significant differences in artisanal baking and sensory quality among eight farmer-selected wheat populations and one modern variety in France. Similarly, four varieties demonstrated different texture and appearance when baked into wholemeal bread in Germany (Ploeger et al., 2008). Starr et al. (2013) also documented significant differences in texture, appearance, aroma, and flavor of cooked grain from 20 wheat varieties grown in Northern Europe. None of the varieties assessed in these studies, however, are commonly grown in the United States. To inform local markets of the United States, this study compared varieties of organically grown heritage, modern, and ancient wheat for whole-grain technical parameters, artisanal bread baking, pasta making, pastry quality, and sensory attributes.

2. Materials and methods

2.1. Field methods

To identify varieties that may be best suited to organic production in the northeastern and northcentral United States, we evaluated 40 winter wheat, 24 spring wheat, and 16 spring emmer entries over four years (2012–2015) at three organically certified locations in Willsboro, NY, Freeville, NY, and Rock Springs, PA. Spring wheat and emmer entries were also tested on certified organic acreage in Carrington, ND. All entries were replicated three times and plot sizes varied from 3.78 to 8.91 square meters, depending on location. Agronomic results of these variety trials are published elsewhere (Sorrells, 2015).

2.2. Variety selection

A subset of varieties entered each of three quality evaluations: bread wheat varieties for sourdough baking and cooked grain, soft wheat varieties for matzah crackers [plural matzot], yeast bread, shortbread cookies, and cooked grain; and emmer varieties for pasta and cooked grain. Table 1 provides an overview of which varieties were included in each evaluation, and their technical parameters. During all baking, pasta making, and sensory evaluations, a randomly generated three-letter code masked the identity of each variety.

2.2.1. Sourdough bread and cooked grain evaluation

For the sourdough baking and cooked grain evaluation, principal component analysis was used to select wheat varieties with a broad range of technical quality parameters (Supplementary Fig. 1). The seven selected varieties included heritage varieties ('Fulcaster' and 'Red Fife'), modern cultivars that were widely grown by organic farmers in the northeastern United States ('Warthog,' 'Fredrick,' and 'Glenn'), and other modern cultivars that had performed well in variety trials ('Appalachian White' and 'Tom'). A blend of 2012 (21%) and 2013 (79%) grain harvested at the Freeville, NY site was used for the sourdough evaluation.

2.2.2. Matzah cracker, yeast bread, shortbread cookie, and cooked soft wheat grain evaluation

To evaluate soft wheat varieties for matzah crackers, yeast bread, shortbread cookies, and cooked grain, five soft wheat varieties were selected: the heritage varieties 'Forward,' 'Pride of Genesee,' and 'Yorkwin' and two high-yielding modern varieties, 'Susquehanna' and 'Fredrick.' Grain for the soft wheat evaluation originated from the 2014 Freeville, NY harvest.

2.2.3. Pasta and cooked emmer grain evaluation

The pasta and cooked grain evaluation included the three emmer varieties 'Lucille,' 'North Dakota Common,' and 'Red Vernal,' all of which were high-yielding in field trials. Emmer grain was a blend of 45% grain from 2012 to 55% grain from 2014 Freeville, NY trial harvests.

2.3. Baking and pasta making evaluations

2.3.1. Sourdough bread evaluation

For the sourdough baking evaluation, grain was milled on an Osttiroler Getreidemuehlen tabletop stone mill (Rondella model), which has similar properties to stone mills commonly used by artisan millers in the United States. The unsifted flour rested at room temperature for 31 days before baking. A panel of eight artisan bakers from the northeastern and northcentral United States prepared and evaluated loaves of bread made from individual varietal flours. Baking methods followed a typical traditional sourdough recipe for the region (Supplementary Table 1). After developing a common ranking scale and vocabulary, bakers scored doughs individually throughout the baking process. Bakers varied levels of hydration, rest time, and mix time to allow each varietal flour to reach its full potential in bread making (Supplementary Table 1). Researchers measured circumference and weight of all baked loaves. For height, a subsample of five varietal loaves was cut in half and measured from the lowest to highest point. To calculate density, researchers determined loaf volume of three representative varietal loaves by displacement in flaxseed.

2.3.2. Matzah cracker, yeast bread, and shortbread cookie evaluation

For evaluations of soft wheat, grain was ground using the Osttiroler Getreidemuehlen tabletop stone mill (Rondella model) three days before the baking evaluation. Two regional millers sifted flour with a coarse mesh, obtaining 90–97% extraction rates. To test the yeast bread-baking quality of soft wheat varieties, a panel of nine bakers compared four soft wheat varieties ('Forward,' 'Pride of Genesee,' 'Yorkwin,' and 'Fredrick') to a hard spring wheat check with high baking quality, 'Red Fife.' Bakers used a yeast-based bread recipe typical for the region (Supplementary Table 1). The bakers changed mixing time, hydration, autolyse time, and number of folds as needed to optimize bread quality for each variety (Supplementary Table 1). To make matzah crackers, bakers followed the formula in Supplementary Table 2. Bakers also prepared shortbread cookies following the formula in Supplementary Table 3. After a consensus was developed on vocabulary, bakers individually scored doughs for all products.

2.3.3. Emmer pasta evaluation

For emmer pasta evaluations, grain was dehulled using a Codema lab-scale oat dehuller and ground four days before the

Table 1
Technical parameters of varieties included in the grain quality evaluations.

Evaluation[a]	Processing Evaluators	Sensory Evaluators	Habit/ Species	Variety type	Variety name	Class hardness and color	Yield[b] kg/ha	Test weight[b] kg/hl	Flour moisture %	Grain protein at 12% moisture	Falling number seconds	DON[d] ppm
Sourdough Bread and Cooked Grain	8 sourdough bakers	30 trained panelists	Winter wheat	Modern	Appalachian white	Hard white	3071	72.8	10.7	9.8	459.3	<0.5
				Modern	Fredrick	Soft white	3233	71.7	11.0	9.5	335.5	0.7
				Heritage	Fulcaster	Soft red	2766	72.9	10.6	10.5	393.5	<0.5
				Modern	Warthog	Hard red	3393	74.2	10.5	9.9	434.4	<0.5
			Spring wheat	Modern	Glenn	Hard red	2277	71.1	10.3	15	406.8	0.7
				Heritage	Red Fife	Hard red	1798	66.7	10.3	14.8	370	<0.5
				Modern	Tom	Hard red	2384	69.9	8.3	14.7	513.4	0.7
Matzah Cracker, Yeast Bread, Shortbread, and Cooked Grain	9 yeast-based bread bakers	11 trained panelists, 24 public preference tasters	Winter wheat	Heritage	Forward	Soft red	3040	72.5	9.1	13.0	403	0.8
				Modern	Fredrick	Soft white	3233	71.7	9.7	11.5	233	1.3
				Heritage	Pride of Genesee	Soft white	2801	72.8	9.2	13.3	311	0.7
				Modern	Susquehanna	Soft red	3307	69.3	9.6	11.1	301	<0.5
				Heritage	Yorkwin	Soft white	3078	71.6	8.9	12.8	308	1.0
Pasta and Cooked Grain	3 pasta makers	12 trained panelists, 26 public preference tasters	Spring emmer	Ancient	Lucille	Hard red	2494[c]	46.4[c]	12.1	14.2	545.4	0.7
				Ancient	North Dakota Common	Hard red	2499[c]	47.4[c]	11.9	14.6	492.8	0.6
				Ancient	Red Vernal	Hard red	2478[c]	46.8[c]	11.9	15.0	594.5	0.7

[a] Tested flour for the sourdough bread and cooked grain evaluation was a blend of 21% 2012 and 79% 2013 harvests from Freeville, NY; flour for the matzah cracker, yeast bread, shortbread cookie, and cooked soft wheat grain evaluations was harvested from Freeville, NY in 2014; flour for the pasta and cooked emmer grain evaluation was a blend of 45% 2012 and 55% 2014 harvests from Freeville, NY.
[b] Yield and test weight values are a mean of three sites over four years (2012–2015).
[c] Yield and test weight values for emmer are reported in the hull.
[d] Deoxynivalenol (DON) had a minimum detectable value of 0.5 ppm.

pasta making evaluation using a KoMo Fidibus 21 tabletop mill with a ceramic/corundum millstone that achieved a fineness of grind similar to commercially available emmer flour. The flour was not sifted. Three pasta makers evaluated varieties of emmer for pasta quality. Evaluators chose a 64% emmer-based pasta formula (Supplementary Table 4), which, in their experience, was the highest concentration of local emmer flour that could produce a functional dough. Pasta makers treated each varietal dough with additional quantities of Antico Molino Caputo 00 flour during rolling to create an ideal pasta feel (Supplementary Table 4). Since pasta makers evaluated pastas as a group, only one overall score was recorded per variety, and statistical analyses were not possible.

2.4. Sensory evaluations

Trained panels conducted descriptive analysis of sourdough bread, matzot, pasta, and cooked grains of the test varieties. To screen out nontasters, i.e., those who lack taste receptor(s) for one or more basic tastes, all prospective panel members took a blind taste test of sour, sweet, and salty solutions, each at low, medium, and high concentrations. For the soft wheat and emmer evaluations, prospective panelists were also tested on bitter solutions. To qualify as a panelist, each taster needed to accurately identify all taste groups and correctly label at least 78% of concentrations.

2.4.1. Sourdough bread and cooked grain evaluation

The panel for the sourdough bread and cooked grain tasting consisted of six professional bakers who participated in the baking evaluation and 24 consumers in the Ithaca, NY area who regularly purchase local sourdough bread. Training on flavor attributes (Supplementary Table 5) and on visual and texture characteristics (Supplementary Table 6) was held for 6 h over two days. For the evaluation, bread made from each variety was cut into 7.62 cm diameter slices that included both crust and crumb. Slices were kept under cellophane until consumed. Whole grains of each variety were cooked using a 2:1 ratio of water to grain until *al dente*, drained, and refrigerated until served in 30 mL portions. Panelists tasted two replicates of the bread samples and one replicate of the cooked grains. Using a randomized complete block design, each panelist received one sample at a time. The tasting of both bread and cooked grain samples was completed in four and a half hours.

2.4.2. Matzah cracker and cooked soft wheat grain evaluation

The matzah cracker and cooked grain sensory panel consisted of seven students and two faculty members of the Culinary Institute of America and two research team members. Training in distinguishing ten flavors (Supplementary Table 7) and visual and mouthfeel characteristics (Supplementary Table 8) was conducted in 9 h over three days. For the matzah evaluation, each panelist was simultaneously given four, 11 cm diameter matzot. Cooked grain was prepared as stated in Section 2.4.1. Each panelist was simultaneously presented with four 30 mL containers filled with cooked grain. Two replications of each evaluation were completed, with panelists alternating between evaluating matzot and cooked grain. No time limit was given for the evaluations, but all panelists completed the evaluations within 3 h.

A preference tasting of cooked grain samples of the four soft wheat varieties was also held during an event on local grains that was open to the public. No training was held. Instead, 24 participants were each simultaneously presented with four cooked grain samples and were given written instructions that asked them to rank the samples according to preference and then answer questions on flavor attributes and their willingness to purchase.

2.4.3. Pasta and cooked emmer grain evaluation

Five instructors at The Natural Gourmet Institute, two food journalists, and five members of the research team completed a descriptive sensory analysis of varietal pasta and cooked emmer grain. Training in distinguishing ten flavors (Supplementary Table 7) and visual and mouthfeel characteristics (Supplementary Table 9) was conducted in 6 h over one day. For the pasta evaluation, each panelist was simultaneously given three 30 mL cups filled with varietal pastas. Cooked emmer grain was prepared and evaluated as stated in Sections 2.4.1 and 2.4.2, respectively. Two

replications of each evaluation were completed within 3 h.

A preference tasting of cooked emmer grain samples of four emmer varieties was also held during a by-invitation only event. Twenty-six participants evaluated samples in a manner similar to the public preference tasting described in Section 2.4.2. The emmer tasted in this evaluation was grown in a different environment (Rock Springs, PA 2014) than the grains tasted by the trained sensory panel.

2.5. Statistical analysis

Statistical analyses were completed in R [version 3.2.2] (R Core Team, 2015), package 'lme4' [version1.1-10] (Bates et al., 2015). The model below, similar to that used by Vindras-Fouillet et al. (2014), incorporated the effects of variety, panelist, order and their subsequent interactions. A reduced model was used if there was not a second replicate (*e.g.*, sensory evaluation of cooked grain) or an order term (*e.g.*, baking trials). For continuous responses, an analysis of variance (ANOVA) allowed the detection of differences among varieties, using a significance threshold of $P < 0.05$. A Satterthwaite approximation facilitated the analysis of unequal variances (Satterthwaite, 1946). As a consequence of unbalanced data, package 'lmerTest' [version 2.0–32] (Kuznetsova et al., 2016) calculated either Type III ANOVA when interactions were significant, or Type II ANOVA to increase power when interactions were not significant. Tukey's honestly significant difference (HSD) made pairwise comparisons of varieties through the package 'multcomp' [version1.4-2] (Hothorn et al., 2008). To validate model assumptions, errors and random effects were checked for normal distribution, homogeneous variance, and independence.

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \alpha_i{:}\beta_j + \alpha_i{:}\gamma_k + \alpha_i{:}\delta_l + \beta_j{:}\gamma_k + \varepsilon_{ijkl}$$

Y_{ijkl}: response for variety i, panelist j, replicate k, and order l
μ: overall mean response
α_i: fixed effect of variety i
β_j: random effect of panelist j
γ_k: fixed effect of replicate k
δ_l: fixed effect of order l
$\alpha_i{:}\beta_j$: random interaction of variety i by panelist j
$\alpha_i{:}\gamma_k$: fixed interaction of variety i by replicate k
$\alpha_i{:}\delta_l$: fixed interaction of variety i by order l
$\beta_j{:}\gamma_k$: random interaction of panelist j by replicate k
ε_{ijkl}: experimental error associated with response i,j,k,l

For binomial responses, logistic regression models evaluated whether variety was significantly associated with the log odds of preference or flavor presence in a sample. A likelihood ratio test compared models with and without variety used in the model below. Varieties were determined to be significantly different using least-square means at a significance level of $P < 0.05$. To validate model assumptions, the number of observations multiplied by the sample probability mean for each response needed to be greater than five. Results were graphed using the R package 'plotrix' [version 3.6–1] (Lemon, 2006).

$$Y_{ijkl} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{j2} + \beta_3 x_{k3} + \beta_4 x_{l4}$$

Y_{ijkl}: log odds of success (*e.g.* a flavor used to describe a sample, or selection of preference)
B_0: intercept log odds of reference sample and replicate
β_1: partial slope associated with variety
x_{i1}: fixed variable of variety i
β_2: partial slope associated with panelist
x_{j2}: random variable of panelist j
β_3: partial slope associated with replicate
x_{k3}: fixed variable of replicate k
β_4: partial slope associated with order
x_{l4}: fixed variable of order l

3. Results

3.1. Baking and pasta evaluations

3.1.1. Sourdough baking evaluation

There were significant differences among scores for varietal performance throughout the sourdough baking process, including mixing, floor time, make-up, proofing, loaf, and crumb quality (Fig. 1). The three spring wheat varieties ('Glenn,' 'Tom,' and 'Red Fife') received the highest overall baking performance scores. Although the winter wheat varieties had lower protein content than the spring varieties (Table1), the overall baking score for 'Warthog' was not significantly different than for 'Red Fife' ($P = 0.1730$). Bakers thought that all varieties made satisfactory loaves, except for the soft wheat 'Fredrick,' which was difficult to manage. Bakers recognized early in the process that 'Fredrick' was a soft wheat and even considered using a loaf pan for baking since the preshape did not look viable. There were also significant differences in loaf measurements among varieties (Table 2). In terms of loaf height, 'Glenn,' 'Tom,' and 'Warthog' made the highest loaves, and 'Fredrick' the lowest ($P < 0.0001$). The circumferences of 'Glenn' and 'Appalachian White' loaves were smaller than 'Fulcaster,' 'Tom,' and 'Red Fife' ($P = 0.0088$). The loaves made of 'Glenn,' 'Red Fife,' 'Tom,' and 'Warthog' were heavier than those of 'Appalachian White' and 'Fredrick' ($P < 0.001$). Loaf volume and density did not differ significantly among varieties ($P = 0.1085$ and 0.3367, respectively).

3.1.2. Matzah cracker, yeast bread, and shortbread cookie soft wheat evaluation

In the evaluation of soft wheat varieties for yeast bread baking, the hard wheat check included in the evaluation, 'Red Fife,' received a significantly higher overall baking performance score than 'Pride of Genesee' ($P = 0.0396$) (Fig. 2). 'Fredrick,' which scored lowest in overall baking performance in the sourdough trial, did not score significantly lower than 'Red Fife' when made into yeast-based bread ($P = 0.9968$). Both varieties tore less during proofing than the soft heritage wheat varieties ($P = 0.0077$).

In the production of matzah crackers, 'Forward' was rated as better than 'Pride of Genesee' ($P = 0.0024$) (Supplementary Fig. 2). 'Pride of Genesee' had insufficient extensibility compared with all other varieties ($P = 0.0201$). 'Yorkwin' and 'Susquehanna' needed more hydration than the other two varieties, which could reduce production costs by requiring less flour in the final product.

As a shortbread, 'Yorkwin' received a higher ranking than the other varieties for overall shortbread baking quality ($P = 0.0060$) (Supplementary Fig. 3). 'Pride of Genesee' tended to have excessive stickiness during mixing, when compared to the top-rated variety, 'Yorkwin' ($P < 0.001$). On the other hand, 'Pride of Genesee' could potentially lower production costs by absorbing less butter. Bakers tended to prefer the flavor of 'Forward' more than 'Susquehanna,' although the difference did not meet the threshold of significance ($P = 0.0615$).

3.1.3. Pasta making evaluation

The pasta makers rated 'Lucille' and 'Red Vernal' as better than

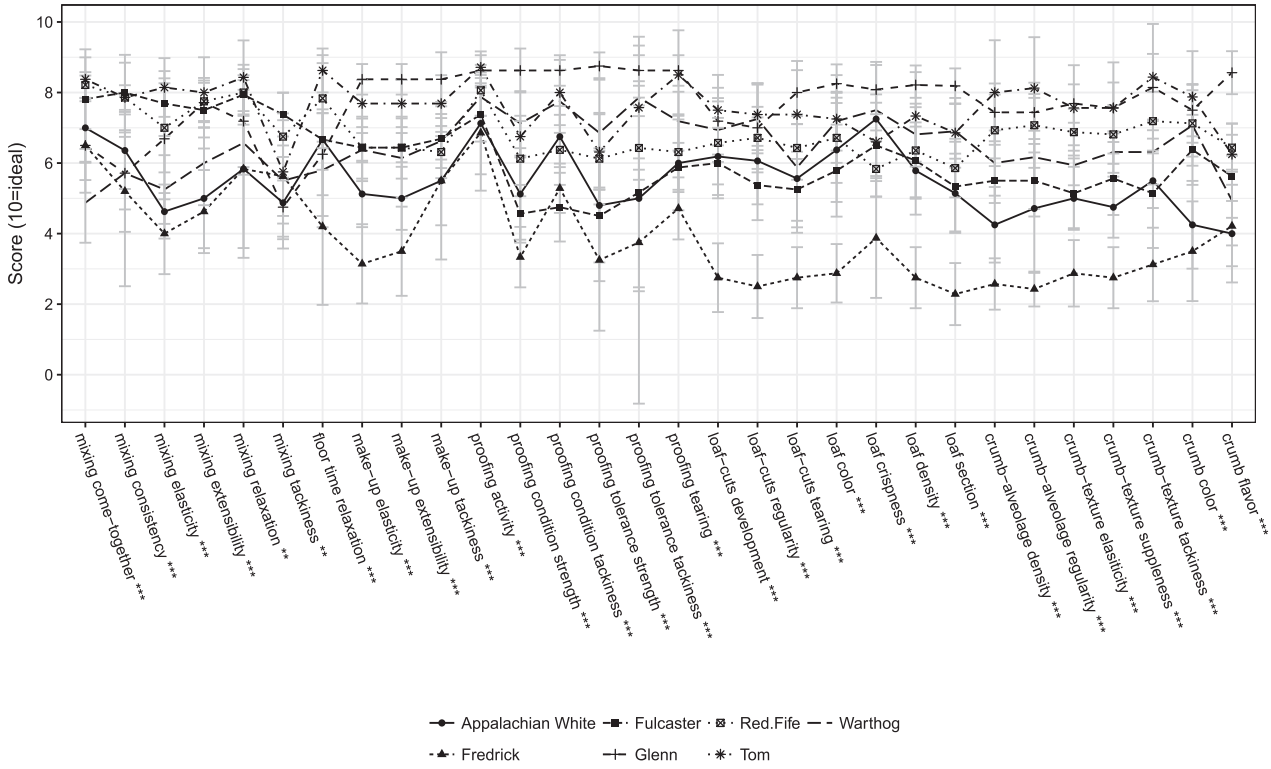

Fig. 1. Varietal performance as rated by bakers in the sourdough bread baking evaluation (** indicates a significant difference at $P < 0.01$ and *** indicates a significant difference at $P < 0.001$). Error bars show 95% confidence intervals.

'North Dakota Common' for pasta making. 'Lucille' and 'Red Vernal' received overall scores of seven out of ten, while 'North Dakota Common' scored four out of ten. 'Lucille' had the best technical performance, as it was strong and easy to roll out and cut with the machine. 'Red Vernal' produced the best texture and had the most preferred flavor by the pasta chefs. 'North Dakota Common' produced a very tacky dough, which demanded additional flour and more time in the pasta roller to obtain the right texture.

3.2. Sensory evaluations

3.2.1. Sourdough bread and cooked grain evaluation

There were significant differences among varietal sourdough for surface texture, texture of crumb, size of air bubbles, graininess, dryness, and ability to dissolve (Table 2). Although panelists assigned 'Red Fife' the highest and 'Warthog' the lowest flavor intensity, the difference was only significant if order of tasting was removed from the model ($P = 0.0278$). 'Fulcaster' had lower odds of being described with bitter flavors, particularly when compared to 'Glenn' ($P < 0.0001$) (Fig. 3b). Variety also influenced the odds of nutty flavors in a sample ($P = 0.0498$). Rather than variety, replicate impacted the aroma and sour flavor of samples. Overall aromatics of bread samples ($P = 0.0085$), wheat aroma of crumb ($P = 0.0410$), and odds of sour flavor ($P = 0.0218$) were higher in the second replicate than the first.

When tasted as cooked grain, the trained panelists recorded differences among varieties for flavor intensity (Table 2). 'Warthog' had higher flavor intensity than 'Appalachian White' ($P < 0.001$), 'Glenn' ($P < 0.001$), 'Red Fife' ($P = 0.0040$), and 'Fulcaster' ($P = 0.0271$). When describing cooked grain samples, variety was significantly associated with the likelihood of dairy flavors ($P = 0.0291$), with 'Fredrick' having the highest odds being described with dairy flavors (Fig. 3a). While there were no significant differences in cooked grain dryness among varieties, panelists rated the first sample they tasted as moister ($P = 0.04338$).

3.2.2. Matzah cracker and cooked soft wheat grain evaluation

The trained panel found differences in woody ($P = 0.0297$) flavor intensity among varietal matzot, with 'Susquehanna' receiving the woodiest flavor (Fig. 3d). 'Susquehanna' also had lower odds of earthy flavors than 'Yorkwin' ($P = 0.0123$) and 'Pride of Genesee' ($P = 0.0233$). Replicate, rather than variety, influenced the fresh flavor intensity ($P = 0.0320$) and odds of bitter flavor ($P = 0.0013$), with higher values in the first replicate. There were no significant differences among varieties in visual texture, shape, roughness, graininess, firmness, and cohesion (Table 2). Order significantly influenced texture responses, with samples tasted first receiving heavier texture ratings ($P = 0.0137$).

The trained panel found significant differences in dryness and texture of cooked grain from soft wheat varieties (Table 2). 'Susquehanna' was moister and less chewy than 'Pride of Genesee' and 'Yorkwin' ($P = 0.0410$ and 0.0374, respectively). There were no significant differences in flavor characterization of cooked grain among the varieties (Fig. 3c). Order was associated with fresh flavor ($P = 0.0184$), with highest odds when tasted first, and lowest when tasted last. Replicate influenced the odds that a sample would be described as warming sweet ($P = 0.0386$) and fresh ($P = 0.0010$), with higher likelihood when tasted during the first replicate.

Participants in the public cooked grain tasting concurred with the findings from the trained panel, selecting 'Susquehanna' as the moistest variety ($P < 0.0001$), and 'Pride of Genesee' and 'Yorkwin' as the chewiest varieties ($P = 0.0030$). Among varieties ranked for personal preference by tasters, 'Pride of Genesee' was the most preferred, while 'Yorkwin' was the least preferred variety ($P = 0.0015$). Moreover, tasters indicated that they would be more

Table 2
Results of processing and sensory evaluations. Black and grey cells indicate values that tend to be more and less preferred, respectively. Letters show Tukey's significant difference at $P < 0.05$. NE indicates not evaluated, *** indicates a significant difference at $P < 0.0001$, ** indicates a significant difference at $P < 0.001$, * indicates a significant difference at $P < 0.05$, and + indicates a difference at $P < 0.10$.

Evaluation	Variety Details		Processing and Sensory Results											
Sourdough Bread and Cooked Grain	Growth Habit	Variety name	Sourdough Baking Score 10=ideal	Bread Height cm	Bread Weight grams	Bread Taste Intensity 10=intense	Average Air Bubble Size cm	Surface Texture 10=rough	Crumb Texture 10=hearty	Bread Graininess 10=grainy	Bread Dryness 10=moist	Bread Cohesion seconds to dissolve	Cooked grain Flavor 10=intense	Cooked grain Dryness 10=moist
			***	***	***	P=0.0798+	***	***	***	***	***	***	***	P=0.9595
	Winter Wheat	Appalachian White	5.5d	6.5b	599de	5.2a	2.0ab	5.6a	6.7b	5.1ab	4.5ab	20.3b	3.3c	4.8a
		Fredrick	3.9e	5.1c	589e	5.5a	2.1b	6.7a	7.9a	5.6a	3.8b	20.7b	4.7ab	4.9a
		Fulcaster	6.2c	5.9b	612cd	5.1a	2.5ab	5.0ab	6.9b	5.3a	4.0ab	19.5b	4.1bc	4.8a
		Warthog	6.5bc	7.5a	615c	4.8a	2.4ab	5.6ab	6.6b	5.4ab	4.0ab	20.3b	5.4a	4.9a
	Spring Wheat	Glenn	7.7a	8.0a	652a	5.3a	2.3ab	3.7b	5.4c	3.9b	5.6a	27.8a	3.7c	4.6a
		Red Fife	6.8b	6.3b	622bc	5.7a	2.0ab	4.8ab	6.9ab	4.7ab	4.8ab	21.9ab	4bc	4.8a
		Tom	7.6a	7.4a	631b	5.4a	3.0a	3.9ab	6.5b	4.7ab	4.6ab	23.5ab	4.2abc	4.7a
Matzah Cracker, Yeast Bread, Shortbread, and Cooked Grain	Growth Habit	Variety name	Matzah Making Score 10=ideal	Short-bread Score 10=ideal	Yeast Bread Baking Score 10=ideal	Matzah Visual Texture 1=smooth	Matzah Visual Shape 10=jagged	Matzah Roughness 10=rough	Matzah Graininess 10=grainy	Matzah Firmness 10=hard	Matzah Cohesion seconds to dissolve	Cooked grain Preferred 1=best	Cooked grain Texture 10=chewy	Cooked grain Dryness 10=moist
			P=0.0166*	***	P=0.0327*	P=0.1131	P=0.4487	P=0.5123	P=0.0834.	P=0.1924	P=0.5456	***	***	***
	Winter Wheat	Forward	7.1a	6.4b	7.2ab	5.6a	5.0a	5.0a	5.6a	5.6a	16.0a	2.4ab	5.0ab	4.2b
		Fredrick	NE	NE	7.7ab	NE	NE	NE	NE	NE	NE	NE	NE	NE
		Pride of Genesee	5.2b	6.9b	6.0b	4.6a	4.7a	4.7a	5.9a	5.4a	15.2a	2.0b	6.5a	3.9b
		Susquehanna	6.4ab	5.6b	NE	4.7a	4.7a	4.7a	5.1a	5.4a	15.5a	2.4b	4.8b	5.2a
		Yorkwin	6.4ab	8.9a	7.2ab	5.2a	5.2a	4.2a	5.6a	6.2a	16.3a	3.0a	6.5a	3.6b
	Spring	Red Fife	NE	NE	7.9a	NE	NE	NE	NE	NE	NE	NE	NE	NE
Pasta and Cooked Grain	Growth Habit	Variety name	Pasta Making Score 10=ideal	Pasta Preferred probability	Pasta Sheen 10=shiny	Pasta Surface Stickiness 10=sticky	Pasta Texture Starchiness 1=starchy	Pasta Roughness 10=rough	Pasta Graininess 10=grainy	Pasta Firmness 10=chewy	Pasta Cohesion seconds to dissolve	Cooked grain Preferred probability	Cooked grain Texture 10=chewy	Cooked grain Dryness 10=moist
			NA	P=0.0888+	P=0.1284	P=0.7818	P=0.4005	P=0.0157*	***	***	P=0.0003**	P=0.0915+	***	P=0.4076
	Spring Emmer	Lucille	7	0.42a	5.2a	4.5a	5.1a	4.6a	3.9b	4.5b	11.1ab	0.19a	5.4b	4.7a
		North Dakota Common	6	0.19a	5.9a	4.6a	5.6a	3.5b	3.6b	3.6c	10.2b	0.42a	6.3a	4.6a
		Red Vernal	7	0.27a	4.8a	4.2a	4.9a	5.0a	5.7a	6.2a	13.5a	0.15a	6.2ab	4.3a

likely to purchase 'Pride of Genesee' than 'Yorkwin' ($P = 0.0146$). There were also significant differences in the selection of the most flavorful variety ($P = 0.0015$), with 'Yorkwin' having the lowest odds. The order in which cooked grain samples were tasted impacted preference. Samples tasted first were most preferred, while those tasted last were least preferred ($P = 0.0118$).

3.2.3. Pasta and cooked emmer grain evaluation

There were significant differences among emmer varieties for pasta roughness, graininess, firmness, and cohesion (Table 2). Shininess, surface stickiness, and starchiness of texture were not significantly different among varieties. Although there was no significant difference in preference for variety, 'Lucille' tended to have higher odds of being preferred than the other two varieties ($P = 0.0894$). 'Red Vernal' was described as having earthier flavor ($P = 0.0101$), and less fresh flavor ($P = 0.0434$) than the other two varieties (Fig. 3e). When panelists chose the least and most prominent flavors to describe each variety [data not shown], 'Red Vernal' had higher odds of nutty being described as the most prominent flavor ($P = 0.0034$), and herbaceous as the least prominent flavor ($P = 0.0242$).

When comparing emmer varieties tasted as cooked grain, the panel rated 'Lucille' as more delicate and less chewy than 'North Dakota Common' (Table 2). There were no significant differences among varieties for cooked grain flavor intensity ($P = 0.4406$) or dryness ($P = 0.4076$). 'Lucille' was also most likely to have nutty described as the most prominent flavor ($P = 0.0197$) [data not shown]. 'North Dakota Common' was more likely to be preferred than 'Lucille,' although the difference was not significant ($P = 0.0915$).

In the untrained public tasting of cooked grain, 'Lucille' was more likely to be rated with the highest flavor intensity than the other varieties ($P = 0.0004$). Although tasters were twice as likely to seek out 'Lucille' for purchase than 'North Dakota Common,' the difference was not significant ($P = 0.2150$).

4. Discussion

4.1. Baking evaluations

Varieties differed in baking quality for sourdough bread, yeast bread, matzah crackers, and shortbread cookies. However, the ranking of a variety differed among products. 'Forward' was the top scoring variety for making matzah crackers, yet it fell in the lowest ranked category for shortbread cookies. Although 'Fredrick' performed poorly in sourdough baking, it was not the lowest performer for yeast breads. Consequently, artisan bakers will not find one variety that performs best for all types of baked goods.

Many heritage wheat varieties that are classified as soft may be semi-hard. Bakers in the northeastern United States have wondered whether soft heritage varieties could, therefore, be appropriate for bread baking. Among the four soft heritage wheat varieties

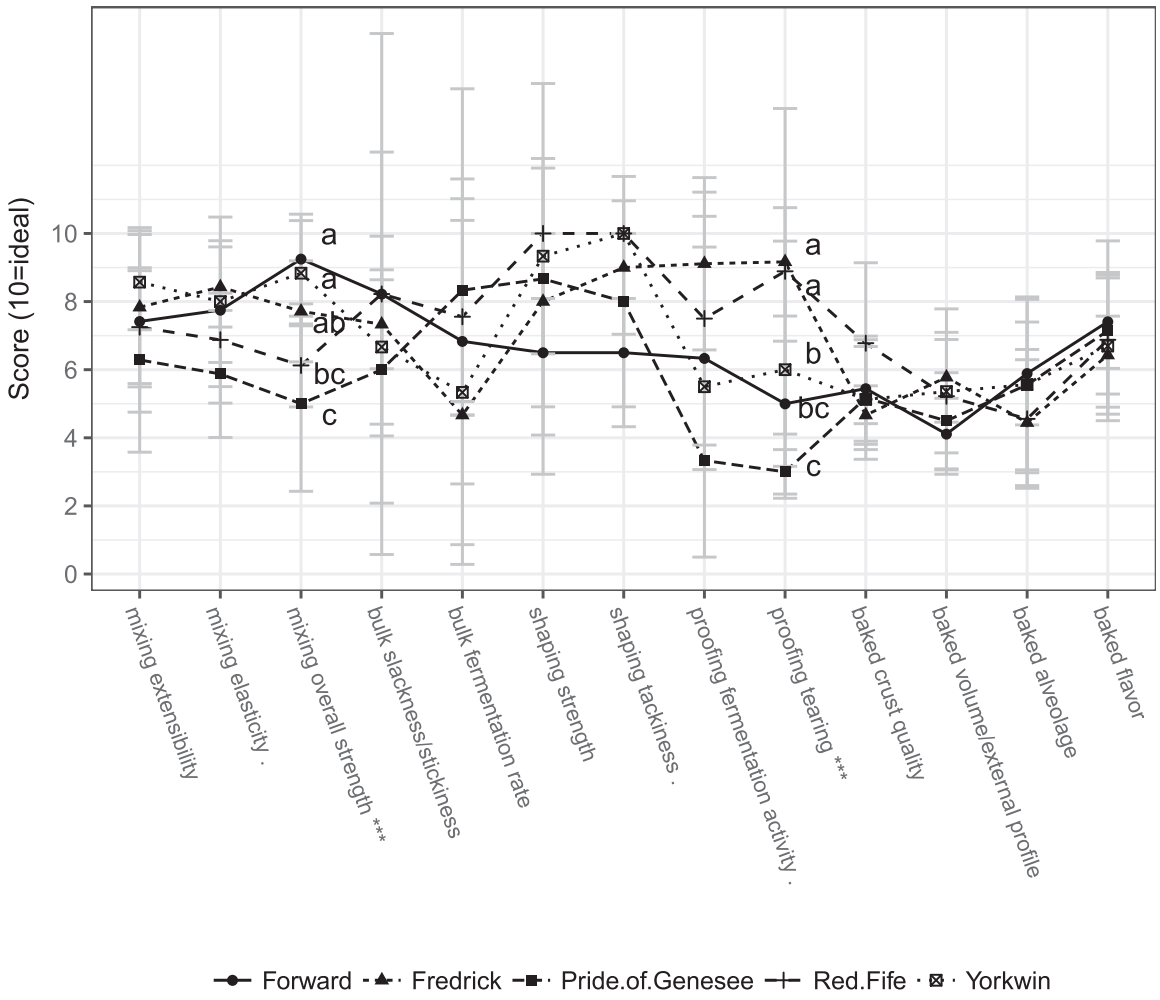

Fig. 2. Varietal performance as rated by bakers in the yeast bread baking evaluation (. indicates a difference at $P < 0.10$ and *** indicates a significant difference at $P < 0.001$). Error bars show 95% confidence intervals.

included in our evaluations, many did contain relatively high concentrations of protein (Table 1). However, the soft heritage wheat varieties included in this study represented a low to moderate spectrum of baking quality. In the sourdough trial, the soft heritage wheat, 'Fulcaster,' received intermediate scores for baking. It ranked better for overall baking, bread height, and weight than the soft modern wheat, 'Fredrick,' yet had lower bread height and wider circumference than the hard modern winter wheat, 'Warthog.' When baked into yeast bread, three soft heritage varieties ('Forward,' 'Pride of Genesee,' and 'Yorkwin') received intermediate or low scores. Two of these soft heritage varieties did not significantly differ in overall baking scores from the high-quality baking check. However, all three varieties excessively tore when compared to both the high quality ('Red Fife') and low quality ('Fredrick') baking checks.

4.2. Sensory evaluations

Our results show that wheat and emmer varieties can differ in sensory characteristics, especially in terms of texture and mouthfeel attributes. Flavor differences among varieties were also detected, but tended to be subtler. However, sensory characteristics and preference for a variety often changed depending on what product was being evaluated. In all three sensory evaluations, the variety with the highest preference or taste intensity as a cooked grain received the lowest rating as a processed product. 'Warthog' was rated with the most intense flavor as a cooked grain, but received the lowest rating for flavor intensity when tasted as a sourdough bread. Similarly, the trained emmer taste panel gave 'North Dakota Common' the highest preference as a cooked grain and the lowest preference as a pasta. The least preferred cooked soft wheat grain was 'Yorkwin,' although this variety was most preferred when tasted as a varietal matzah cracker [data not shown]. A significant interaction between variety and product statistically demonstrates this point. For an emmer variety, the likelihood for preference, bran, nutty, fresh and earthy flavors depended on whether the variety was tasted as a cooked grain or pasta ($P < 0.05$). Similar to Section 4.1, selecting the best variety depends on what product will be made from that variety.

Preference and overall flavor were correlated. There was a significant and positive correlation ($P < 0.0001$, $r = 0.557$) between odds of the variety being most flavorful and preference rating for cooked grain of soft wheat. There was also a significant and positive correlation ($P = 0.01473$, $r = 0.2341$) between the odds of the most intense and most enjoyable flavor. While preference is influenced by sensory factors beyond flavor, such as texture (Heiniö et al., 2016), the association between flavor and preference was also found in tomato (Baldwin et al., 1998) and carrot (Simon et al., 1980).

The order in which samples were tasted did influence many

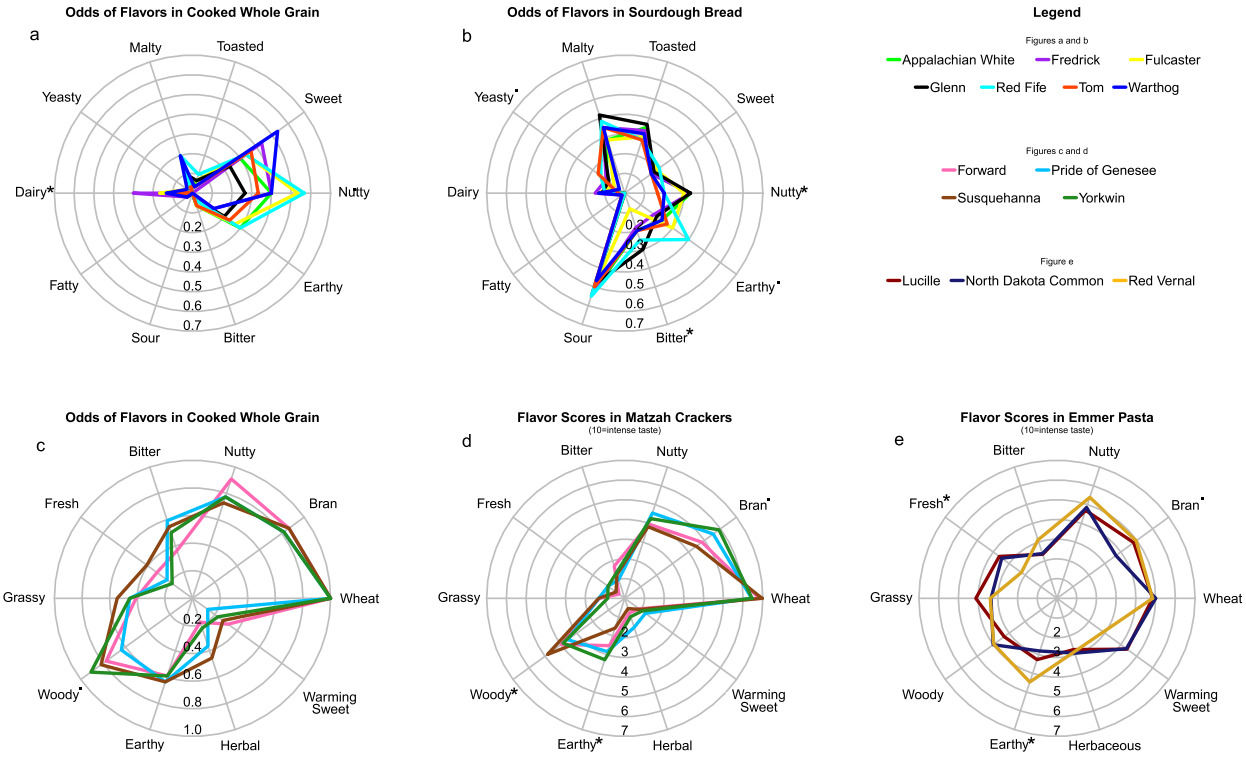

Fig. 3. Varietal effect on the likelihood that certain flavors would be used to describe (a) cooked wheat grain (b) sourdough breads, and (c) cooked soft wheat grain; and the intensity of certain flavors in (d) matzah crackers and (e) emmer pasta (. indicates an association at $P < 0.10$ and * indicates a significant association at $P < 0.05$).

sensory responses, particularly the assessment of preference. The sample that was tasted first tended to be evaluated differently than other samples for preference, fresh flavor, and some texture components. This finding concurs with the documented "first sample effect" in sensory science (Stone et al., 2012), and emphasizes the importance of an experimental design that balances the placement of varieties in the first and last orders.

4.3. Inference from results

The complexity and diversity of wheat processing complicate the evaluation of genotypes for baking and sensory quality. Interpretation of our results may be limited, since all material was derived from one site (Freeville, NY). Moreover, the flour extraction rates (85%–100%) and baking methods used in this study will not always match the practices of regional millers and bakers. It becomes expensive and time consuming to add additional treatments to baking and sensory evaluations, such as including varieties grown under multiple field conditions; using flour with varying extraction rates; and changing fermentation cultures, time, and temperature. Although the presented experimental design did not allow the assessment of genotype by environment, genotype by milling technique, and genotype by baking method interactions, inference from previous studies can illuminate the potential impact of these interactions on our results. Little is known about genotype by environment interactions on sensory characteristics in wheat, but results from other species indicate that there may be an effect. Significant genotype by environment interactions were detected for sweetness, bitterness, and roasted flavors in peanut (*Arachis hypogaea* L.) (Pattee et al., 1997) and protein, sucrose, citric acid, and malic acid in common bean (*Phaseolus vulgaris* L.) (Florez et al., 2009). Previous studies also showed that genotype by baking technique influenced quality. In Katina et al. (2006), longer sourdough fermentation enhanced roasted and pungent acid flavors, while use of *S. cerevisiae* reduced roasted flavors by metabolizing the amino acids associated with those flavors. The authors also demonstrated that longer fermentation increased loaf volume. Genotype by milling technique, however, may exert the largest influence on quality parameters. In Kihlberg et al. (2004), milling technique influenced bread quality more than the environment where the wheat was grown. Katina et al. (2006) documented more bitterness and aftertaste in bread made from higher bran flour. In their study, ash content influenced sourdough bread flavor more than temperature, length of fermentation, and type of sourdough culture.

To assess genotype by environment and genotype by milling technique interactions in the present study, the results can be compared to another variety evaluation completed by Mallory et al. (2014, 2015). The study tested similar varieties and sourdough baking techniques, but used different growing environments (Alburgh, VT 2010–2012) and milling techniques (85–95% extraction) than the present paper (Freeville, NY 2012–2014 and 100% extraction, respectively). In both evaluations, bakers felt that all varieties made satisfactory loaves of bread, apart from 'Fredrick.' The top-rated spring and winter wheat varieties in both evaluations included 'Glenn' and 'Warthog.' However, in Mallory et al. (2014), 'Tom' displayed excessive dough extensibility and low volume, while the bakers in the current study gave 'Tom' the best score for dough extensibility and second highest score for loaf volume. In another contrast, 'Appalachian White' received the second highest baking score in Mallory et al. (2015), while bakers in the present study rated it second lowest. The differences in variety rankings between the evaluations confirm that genotype by environment and/or genotype by milling interactions influence sourdough baking quality for organically grown wheat.

4.4. Recommendations for high-throughput evaluations

Descriptive sensory analysis is costly and time-intensive. Moreover, the number of tested varieties is limited, since panelists can only handle a small number of samples before reaching sensory fatigue (Stone et al., 2012). Plant breeding, which handles large numbers of genotypes, requires more high-throughput sensory analysis methods. Our results suggest that unreplicated designs could improve throughput. For 58 continuous responses included in the sensory analysis, only four had a significant interaction between variety and replicate (wheat aroma of crust, matzot nutty flavor intensity, pasta grassy flavor intensity, and pasta fresh flavor intensity). These results indicate that unreplicated or partially replicated designs may generate accurate data for most sensory descriptors, thereby allowing the evaluation of more varieties at lower cost.

Acknowledgements

We thank Wide Awake Bakery, King Arthur Flour, Gramercy Tavern, The Natural Gourmet Institute, Bread Alone Bakery, and the Culinary Institute of America for hosting the baking and sensory evaluations. We thank the many bakers and tasters who dedicated time and effort to participate in this research. We also thank Lynn Johnson for her guidance on statistical analysis. Support was provided by USDA Organic Research and Extension grant # 2011-51300-30697, USDA Sustainable Agriculture Research and Education grant #LNE12-318, Hatch Project 149-430, 149-449, and by USDA-NIFA-AFRI grants, award numbers 2009-65300-05661 and 2011-68002-30029.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jcs.2016.12.010.

References

Baldwin, E.A., Scott, J.W., Einstein, M.A., Malundo, T.M., Carr, B.T., Shewfelt, R.L., Tandon, K.S., 1998. Relationship between sensory and instrumental analysis for tomato flavor. J. Am. Soc. Hortic. Sci. 123, 906–915.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48.

Codron, J., Grunert, K., Giraud-Heraud, E., Soler, L., Regmi, A., 2005. Retail sector responses to changing consumer preferences: the European experience. In: Regmi, A., Gehlhar, M. (Eds.), New Directions in Global Food Markets. Economic Research Service, Washington, pp. 32–46.

Elbehri, A., 2007. The Changing Face of the U.S. Grain System: Differentiation and Identify Preservation Trends. Economic Research Service. Economic Research Report Number 35.

ERS, 2014. Organic Market Overview. http://www.ers.usda.gov/topics/natural-resources-environment/organic-agriculture/organic-market-overview.aspx. Accessed 30 March 2016.

Packaged Facts, 2015. Food Formulation Trends: Ancient Grains and Sprouted Ingredients. Rockville.

Florez, A., Pujolà, M., Valero, J., Centelles, E., Almirall, A., Casañas, F., 2009. Genetic and environmental effects on chemical composition related to sensory traits in common beans (Phaseolus vulgaris L.). Food Chem. 113, 950–956.

Halloran, A., 2015. The New Bread Basket. Chelsea Green Publishing, White River Junction.

Heiniö, R.L., Noort, M.W.J., Katina, K., Alam, S.A., Sozer, N., de Kock, H.L., Hersleth, M., Poutanen, K., 2016. Sensory characteristics of wholegrain and bran-rich cereal foods - a review. Trends Food Sci. Technol. 47, 25–38.

Hendrickson, M., Heffernan, W., 2007. Concentration of Agricultural Markets. Department of Rural Sociology, Columbia.

Hills, K.M., 2012. Improving Baking Quality and Local Markets for Wheat Grown in Western Washington. Washington State University, Pullman.

Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. Biometrical J. 50, 346–363.

Kaptan, B., Kayisoglu, S., 2015. Consumers attitudes towards food additives. Am. J. Food Sci. Nutr. Res. 2, 21–25.

Katina, K., Heiniö, R.L., Autio, K., Poutanen, K., 2006. Optimization of sourdough process for improved sensory profile and texture of wheat bread. LWT - Food Sci. Technol. 39, 1189–1202.

Kihlberg, I., Johansson, L., Kohler, A., Risvik, E., 2004. Sensory qualities of whole wheat pan bread - influence of farming system, milling and baking technique. J. Cereal Sci. 39, 67–84.

Kissing Kucek, L., Veenstra, L.D., Amnuaycheewa, P., Sorrells, M.E., 2015. A grounded guide to gluten: how modern genotypes and processing impact wheat sensitivity. Compr. Rev. Food Sci. Food Saf. 14, 1–18.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2016. Package 'lmerTest.' R Package Version 2.0-32.

Lemon, J., 2006. Plotrix: a package in the red light district of R. R-News 6, 8–12.

Low, S.A., Adalja, A., Beaulieu, E., Key, N., Martinez, S., Melton, A., Perez, A., Ralston, K., Stewart, H., Suttles, S., Vogel, S., Jablonski, B.B.R., 2015. Trends in U.S. Local and Regional Food Systems: Report to Congress. Economic Research Service.

Mallory, E., Darby, H., Molloy, T., Cummings, E., 2014. Maine and Vermont Organic Spring Wheat Variety Trial Results 2010-2013. University of Vermont Extension and University of Maine Cooperative Extension, Orono.

Mallory, E., Darby, H., Molloy, T., Cummings, E., 2015. Maine and Vermont Organic Winter Wheat Variety Trial Results 2010-2013. University of Vermont Extension and University of Maine Cooperative Extension, Orono.

Pattee, H.E., Isleib, T.G., Giesbrechp, F.G., 1997. Genotype-by-environment interaction in sweet and bitter sensory of peanut. Peanut Sci. 24, 117–123.

Ploeger, A., Roger, M., Lueck, L., Leifert, C., 2008. Sensory Evaluation of Processed Wheat from a Defined Field-trial. In: 16th IFOAM Organic World Congress. Modena, Italy, pp. 1–4.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Satterthwaite, F.E., 1946. An approximate distribution of estimates of variance components. Biom. Bull. 2, 110–114.

Simon, P.W., Peterson, C.E., Lindsay, R.C., 1980. Genetic and environmental influences on carrot flavor. J. Am. Soc. Hortic. Sci. 105, 416–420.

Sorrells, M.E., 2015. Small Grains Cultivar Testing: Organic Trials. https://plbrgen.cals.cornell.edu/research-extension/small-grains/cultivar-testing. Accessed 30 March 2016.

Starr, G., Bredie, W.L.P., Hansen, Å.S., 2013. Sensory profiles of cooked grains from wheat species and varieties. J. Cereal Sci. 57, 295–303.

Stone, H., Beibaum, R., Thomas, H.A., 2012. Test strategy and the design of experiments. Sensory Evaluation Practices, fourth ed. Elsevier Inc., pp. 117–165

Vindras-Fouillet, C., Ranke, O., Anglade, J., Taupier-letage, B., Véronique, C., Goldringer, I., 2014. Sensory analyses and nutritional qualities of hand-made breads with organic grown wheat bread populations. Food Nutr. Sci. 5, 1860–1874.