Variety Evaluation for Sourdough Baking and Sensory Quality

Lisa Kissing Kucek¹, June Russell², Elizabeth Dyck³, Liz Clark⁴, David Benscher¹, Mark E. Sorrells¹, Julie Dawson⁵

¹Cornell University; ²Greenmarket, Grow NYC; ³Organic Growers Research and Information-Sharing Network; ⁴Gimme! Coffee; ⁵University of Wisconsin-Madison

We thank Stefan Senders of Wide Awake bakery for hosting and Jeffrey Hamelman of King Arthur Flour for facilitating the baking trial. We are grateful for the many bakers and tasters who dedicated time and effort in completing this evaluation.

Preliminary Data - 24 January 2015

This work is part of “Value-added Grains for Local and Regional Food Systems”, supported by the Organic Agriculture Research and Extension Initiative
Evaluation Process:

Wheat varieties were screened for use in local organic food systems.

Evaluate 146 varieties of wheat, spelt, emmer, and einkorn under organic management.

Analyze 37 varieties for protein, falling number, and vomitoxin.

Evaluate sensory profiles with 30 trained tasters.

Assess 7 varieties for sourdough baking quality with 8 bakers.
### Overview of Results

#### Table 1. Selected wheat variety performance for bread quality (green: higher scoring; red: lower scoring; *indicates statistical significance)

<table>
<thead>
<tr>
<th>Type</th>
<th>Variety</th>
<th>Market</th>
<th>Variety</th>
<th>Yield</th>
<th>Test Weight</th>
<th>Protein</th>
<th>Baking</th>
<th>Bread Height</th>
<th>Bread Taste</th>
<th>Crumb Texture</th>
<th>Surface Texture</th>
<th>Bread Ability to Dissolve</th>
<th>Bread Graininess</th>
<th>Bread Dryness</th>
<th>Whole Grain Taste</th>
<th>Whole Grain Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>Appalachian</td>
<td>Hard White</td>
<td>Modern</td>
<td>8 of 33</td>
<td>5 of 33</td>
<td>10.4</td>
<td>3.9*</td>
<td>6.5</td>
<td>5.2</td>
<td>6.7</td>
<td>5.6</td>
<td>20.3</td>
<td>5.1</td>
<td>4.5</td>
<td>3.3*</td>
<td>2.5*</td>
</tr>
<tr>
<td></td>
<td>Frederick</td>
<td>Soft White</td>
<td>Modern</td>
<td>11 of 33</td>
<td>22 of 33</td>
<td>9.7</td>
<td>5.5*</td>
<td>5.1*</td>
<td>5.5</td>
<td>7.9*</td>
<td>6.7*</td>
<td>20.7</td>
<td>5.6*</td>
<td>3.8*</td>
<td>4.7</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Fulcaster</td>
<td>Soft Red</td>
<td>Heritage</td>
<td>27 of 33</td>
<td>15 of 33</td>
<td>10.8</td>
<td>6.2</td>
<td>5.9</td>
<td>5.1</td>
<td>6.9</td>
<td>5.0</td>
<td>19.5</td>
<td>5.3</td>
<td>4.0</td>
<td>4.1</td>
<td>3.0*</td>
</tr>
<tr>
<td></td>
<td>Warthog</td>
<td>Hard Red</td>
<td>Modern</td>
<td>2 of 33</td>
<td>6 of 33</td>
<td>11.1</td>
<td>6.5</td>
<td>8.0*</td>
<td>4.8*</td>
<td>6.6</td>
<td>5.6</td>
<td>20.3</td>
<td>5.4</td>
<td>4.0</td>
<td>5.4*</td>
<td>2.7</td>
</tr>
<tr>
<td>Spring</td>
<td>Red Fife</td>
<td>Hard Red</td>
<td>Heritage</td>
<td>19 of 22</td>
<td>14 of 22</td>
<td>15.3</td>
<td>6.8</td>
<td>6.3</td>
<td>5.7*</td>
<td>6.9</td>
<td>4.8</td>
<td>21.9</td>
<td>4.7</td>
<td>4.8</td>
<td>4.0</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>Tom</td>
<td>Hard Red</td>
<td>Modern</td>
<td>1 of 22</td>
<td>2 of 22</td>
<td>16.7</td>
<td>7.6*</td>
<td>7.4*</td>
<td>5.4</td>
<td>6.5</td>
<td>3.9*</td>
<td>23.5</td>
<td>4.7</td>
<td>4.6</td>
<td>4.2</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>Glenn</td>
<td>Hard Red</td>
<td>Modern</td>
<td>8 of 22</td>
<td>1 of 22</td>
<td>16.0</td>
<td>7.7*</td>
<td>7.5*</td>
<td>5.3</td>
<td>5.4*</td>
<td>3.7*</td>
<td>27.8*</td>
<td>3.9*</td>
<td>5.6*</td>
<td>3.7</td>
<td>2.5*</td>
</tr>
</tbody>
</table>

Results the baking and sensory evaluation were based on only one method and baking session.
Sourdough Baking Trial Results
8 bakers evaluated 7 varieties in replicate

- **Glenn and Tom**: top categories for baking quality, height, and weight
- **Warthog**: intermediate for baking and weight, top category for height
- **Red Fife and Fulcaster**: intermediate in most categories
- **Appalachian White**: second lowest for baking, poor weight
- **Frederick**: lowest for baking, height, and weight

**Type III ANOVA with Satterwaite approximation**

\[ H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7; \alpha \leq 0.05 \]

\[ \gamma_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \]

- \( \gamma_{ijk} \): response for variety i and baker j
- \( \mu \): overall mean response
- \( \alpha_i \): fixed effect of variety i
- \( \beta_j \): random effect of baker j
- \( \varepsilon_{ij} \): experimental error associated with response i,j

To validate model assumptions, errors and random effects were checked for normal distribution, homogeneous variance, and independence.

There were significant differences in scores among varieties at \( p < 0.0001 \). \( n = 1567 \)
Scores for all major categories (mixing, floor time, make-up, proof, proofing condition, proofing tolerance, loaf, loaf-cuts, crumb, crumb-texture, and crumb-alveolage (p<0.0001)
**Figure to the right:**
There were significant differences in height among varieties at \( p < 0.0001 \). \( n = 35 \)

**Figure below:**
There were significant differences in weight and circumference among varieties at \( p < 0.0001 \). \( n = 81 \)

**Error bars are 95% CI. Letters are Tukey’s HSD.**

Not shown:
Volume (\( p = 0.109 \)) and density (\( p = 0.33 \)) of loaves were not significantly different among varieties. \( n = 21 \)
### Taste Intensity

(1 = no flavor, 10 = intense) n=380

<table>
<thead>
<tr>
<th>Variety</th>
<th>Taste Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>App. White</td>
<td>ab</td>
</tr>
<tr>
<td>Frederick</td>
<td>ab</td>
</tr>
<tr>
<td>Fulcaster</td>
<td>ab</td>
</tr>
<tr>
<td>Glenn</td>
<td>a</td>
</tr>
<tr>
<td>Red Fife</td>
<td>ab</td>
</tr>
<tr>
<td>Tom</td>
<td>ab</td>
</tr>
<tr>
<td>Warthog</td>
<td>b</td>
</tr>
</tbody>
</table>

Error bars are 95% CI. Letters are Tukey’s HSD 95% CI.

There were significant differences in taste intensity among varieties at p=0.021. Subject accounted for 17.68% of variation.

---

**Type III ANOVA with Satterwaite approximation**

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk} \]

\( Y_{ijk} \): response for variety \( i \), rep \( j \), order \( k \), and subject \( l \)

\( \mu \): overall mean response

\( \alpha_i \): fixed effect of variety \( i \)

\( \beta_j \): fixed effect of rep \( j \)

\( \gamma_k \): random effect of subject \( k \)

\( \epsilon_{ijk} \): experimental error associated with response \( i,j,k \)

To validate model assumptions, errors and random effects were checked for normal distribution, homogeneous variance, and independence.

---

**Sourdough Sensory Evaluation**

30 tasters evaluated 7 varieties over 2 replicates

- **Glenn**: smoothest surface texture category and most delicate crumb texture, longest time to dissolve, lowest graininess, highest moisture
- **Red Fife**: taste intensity higher than Warthog, earthier flavors
- **Tom**: smoothest surface texture category, largest air bubble size
- **Fulcaster**: second highest air bubble size
- **Appalachian White**: smallest air bubble size
- **Warthog**: taste intensity lower than Red Fife
- **Frederick**: roughest surface texture, most hearty crumb texture, highest graininess, driest bread

There were significant differences in taste intensity among varieties at \( p=0.021 \). Subject accounted for 17.68% of variation.
There were significant differences among varieties at $p<0.0001$. Subject accounted for 11% of variation.

Although there were no significant differences between varieties for any aromatics categories, replicate number influenced aromatics. Replicate 2 produced significantly higher values for whole sample aromatics ($p=0.0134$), crust aromatics ($p=0.0242$), and crumb aromatics ($p=0.0341$). Subject accounted for 22.18%, 39.66%, and 34.01% of variation, respectively.
Sourdough Sensory Evaluation

Texture: ability to dissolve
(seconds to dissolve in mouth)

There were significant differences in time to dissolve among varieties at $p<0.0001$. Subject accounted for 56.47% of variation.

Air Bubble Size
((centimeters))

There were significant differences in reported average air bubble size among varieties at $p<0.0001$. Subject accounted for 16.53% of variation.
Sourdough Sensory Evaluation

**Graininess:** amount of small particles (1=no graininess, 10 = overwhelming graininess)  
$n=397$

There were significant differences in graininess among varieties at $p<0.0001$. Subject accounted for 42.46% of variation.

**Dryness:** saliva taken from tongue (1 = very dry, 10 = moist)  
$n=414$

There were significant differences in dryness among varieties at $p<0.0001$. Subject accounted for 32.81% of variation.

Error bars are 95% CI  
Letters are Tukey’s HSD  
95% CI
Tom (p=0.024), Red Fife and Warthog (p=0.073) lowered the odds for nutty flavors.

Warthog lowered the odds for yeasty flavors (p=0.060).

Fulcaster lowered the odds for bitter flavors (p=0.042).

Red fife increased the odds for earthy flavors (p=0.035).

Wald χ² test binomial distribution
H₀: β_i=0; α≤0.10
Y_ijk = β_0 + β_1x_i1 + β_2x_i2 + β_3x_i3

Y_i: log odds of a flavor used for sample
β_0: intercept log odds App.White rep 1
β_1: partial slope associated with variety
x_i1: fixed variable of variety i
β_2: partial slope associated with rep
x_i2: fixed variable of rep i
β_3: partial slope associated with taster
x_i3: random variable of taster i

To validate model assumptions, n*π>5 and n*(1-π)>5. n: number of observations; π: sample probability mean. See final slide for more details.
Cooked Whole Grain Sensory Evaluation

30 tasters evaluated 7 varieties over one replicate

- **Warthog**: most intense flavor, sweeter and less grainy/seedy flavors
- **Red Fife and Fulcaster**: nuttier flavors
- **Frederick**: yellow color and more dairy flavors
- **Glenn**: less nutty and less sweet flavors
- **Tom**: intermediate in all categories
- **Appalachian White**: least intense flavor

Type III ANOVA with Satterwaite approximation

\[ H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7; \alpha \leq 0.05 \]

\[ y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk} \]

- \( y_{ijk} \): response for sample i, order k, subject l
- \( \mu \): overall mean response
- \( \alpha_i \): fixed effect of variety i
- \( \beta_j \): fixed effect of order j*
- \( \gamma_k \): random effect of subject k
- \( \varepsilon_{ijk} \): experimental error associated with response i,j,k

To validate model assumptions, errors and random effects were checked for normal distribution, homogeneous variance, and independence.

*Only included in the dryness model. See last slide for details.

There were significant differences in flavor intensity of varieties at \( p < 0.0001 \). Subject accounted for 41.63% of variation.
Whole grain dryness was not significantly different by variety (p=0.946). The first sample tasted (order=1) was reported to have significantly higher moisture (p=0.0434).

There were significant differences in grain size of varieties at p<0.0001. Subject accounted for 60.02% of variation.
Warthog lowered odds for grainy/seedy flavors (p=0.091). Glenn lowered (p=0.035), while Red Fife (p=0.036) and Fulcaster (p=0.074) increased odds for nutty flavors. Glenn significantly lower odds for sweet flavors (p=0.067). Frederick increased odds for dairy flavors (p=0.002).

Wald χ² test binomial distribution

H₀: β₁=0; α≤0.10

Yᵢjₓk = β₀ + β₁x₁ + β₂x₂ + β₃x₃

Yᵢ: log odds of a flavor used for sample
B₀: intercept log odds
β₁: partial slope associated with variety
x₁: fixed variable of variety i
β₂: partial slope associated with order
x₂: fixed variable of order i
β₃: partial slope associated with taster
x₃: random variable of taster I

To validate model assumptions, n*π>5 and n*(1-π)>5. n: number of observations; π: sample probability mean.

*Indicates that a variety significantly influenced the probability of a particular flavor being used to describe a whole grain.
Analyses completed in R and JMP. Order, an ordinal variable from 1 to 7, was a candidate to be included as a covariate in all models. However, order was not linearly related to the responses evaluated in the models, and consequently, violated the assumptions of an ANCOVA model. Despite randomization, some samples are overrepresented in certain orders (e.g. Red Fife in order 2). When the model was run for samples that were balanced, there was not a significant effect for order. Order was only included as a fixed effect in the analysis of whole grain dryness, to interpret deviations found between the first sample tasted and all other orders.

**Baking evaluation R code:**
```r
model=lmer(Y~Variety+(1|baker))
summary(bakemodel)
anova(bakemodel,Type=3)
```

**Bread sensory R code:**
```r
model=lmer(Y~Variety+Rep+(1|Subject))
summary(model)
anova(model, Type=3)
```

**Bread flavor descriptors R code:**
```r
model=glmer(Y~Variety+Rep+ (1|Subject),
control=glmerControl(optimizer="bobyqa",optCtrl=list(maxfun=100000)),family="binomial",data=sens)
summary(model)
anova(model)
```

**Whole Grain taste and size R code:**
```r
Ymodel=lmer(Y~Variety+(1 |Subject))
```

**Whole Grain dryness R code:**
```r
model=lmer(Y~Variety+Order+(1 |Subject))
```

**Whole Grain flavor descriptors R code:**
```r
model=glmer(Y~Variety+(1 |Subject), control=glmerControl(optimizer="bobyqa",optCtrl=list(maxfun=100000)),family="binomial",data=sens)
summary(model)
anova(model)
```