Optimum Experiment Size for Screening Watermelon Cultigens for Fruit Resistance to Bacterial Fruit Blotch

James Daley* and Todd C. Wehner
Department of Horticultural Science, North Carolina State University, Raleigh NC 27695-7609
**jamd6133@outlook.com*

Introduction

Bacterial fruit blotch (BFB) is a serious seed-borne disease caused by *Acidovorax citrulli* (13) that leads to significant watermelon (*Citrullus lanatus*) production losses (6, 8, 12). Although there are no watermelon cultivars with resistance to BFB (1), numerous studies have identified resistance sources (1, 2, 3, 4, 5, 6, 7, 10, 14). In one such study, watermelon cultigens (plant introductions and cultivars) were screened for fruit resistance to BFB under field conditions (3). Unfortunately, the most resistant cultigens were plant introductions with undesirable horticultural traits (3).

Although fruit resistance screening may be simplified as we better understand the underlying mechanisms, introgression of resistance into elite cultivars and confirmation of resistance in other sources will likely require using the original screening method, which is labor intensive and significantly affected by environmental variation (3). However, in addition to method descriptions and observations, prior empirical variance component estimates provide an opportunity to optimize future resistance screening under similar conditions by adjusting the resource allocations to further mitigate extraneous variance (3).

In this article, we calculated simulated estimates of the variance of a cultigen mean ($\sigma^2_{Cul\ M\ sim}$) (15) and broad-sense heritability ($H^2_{B\ sim}$) using simulated allocation scenarios and experimentally derived variance component estimates to predict the efficiency of hypothetical fruit resistance screenings and provide a framework for optimizing future experiments.

Methods

In order to demonstrate optimization of future experiments, average variance component estimates (3) and scenario-dependent values for years, blocks per year (blocks), and replications per block (replications) were used to calculate $H^2_{B\ sim}$ and $\sigma^2_{Cul\ M\ sim}$ (Table 1) estimates over 10 resource allocation scenarios.

Results and Discussion

In the optimization scenarios, the greatest gains were achieved by increasing allocations in descending order: years > blocks > replications, and scenarios that maximized blocks over years had the fastest results over time (Table 2). The optimum scenario would minimize $\sigma^2_{Cul\ M\ sim}$, increasing the power to discriminate genotypes (15), and maximize $H^2_{B\ sim}$, improving genetic gain (utilizing narrow-sense heritability) (9), over the shortest time. Ultimately, the preferred metric depends on the research goals. Scenario 2 had the maximum number of years and the best $H^2_{B\ sim}$, 0.46, and $\sigma^2_{Cul\ M\ sim}$0.54, and scenario 4, which maximized replications, had the worst $H^2_{B\ sim}$, 0.25, and $\sigma^2_{Cul\ M\ sim}$1.39. Scenario 3 maximized blocks and had the most favorable $H^2_{B\ sim}$, 0.40, and $\sigma^2_{Cul\ M\ sim}$, 0.69, in the shortest time. The respective range of $H^2_{B\ Sim}$ and $\sigma^2_{Cul\ M\ sim}$ among the more balanced scenarios, 5-10, was 0.31 to 0.45 and 0.99 to 0.56. The $H^2_{B\ sim}$ and $\sigma^2_{Cul\ M\ sim}$ of the completely balanced scenario, 1, was 0.39 and 0.70, respectively. Simulated broad-sense heritability and $\sigma^2_{Cul\ M\ sim}$were strongly negatively correlated, Pearson's correlation coefficient, $r(8) = -.97$, $p < 0.001$ (11).

As variance components are further partitioned by adding more replications, blocks, and years, gains are diminishing, i.e., more expensive. In reality, adding years would likely be more expensive than adding blocks, which would be more expensive than adding replications. However, probably the heaviest toll would result from opportunity costs from long experiments that delay breeding decisions and research results (15). For example, the impractically long scenario 2 had the best metrics but would be the most time-expensive, incurring unclear costs beyond what it took to conduct the experiment. A more in-depth demonstration of cost analysis and allocation compromises to balance gains for multiple traits is given in a similar optimization by Swallow and Wehner (15). Customized to resources and experimental objectives, expanded optimization scenarios could be used to evaluate the effects of various allocations beyond the 10 scenarios provided here.

When evaluating idealized scenarios, researchers need to consider the severity of missing data on their outcomes. The simulation calculations for 3 years, 3 experiments, and 1 replication (not shown) that matched the resource allocation for the actual screening experiments (3), projected 0.47 $H^2_{B\,sim}$ and 0.52 $\sigma^2_{Cul\ M\ sim}$. By using 2 years, 4 blocks, and 1 replication or 1 year, 8 blocks, and 1 replication similar metrics may have been achieved and time and resources saved. However, these simulated results are better than the realized average of 0.343 $H^2_{B\,sim}$ and 0.868 $\sigma^2_{Cul\ M\ sim}$ for the actual dataset that included the screening experiments and additional testing (3). This discrepancy was because of unexpected missing data that led to average harmonic means of 2.6 years, 4.8 blocks, and 5.2 replications (3), whereas, with no missing data, each cultigen would have been replicated at least 9 times over 9 blocks and 3 years just for the screening experiments. Considering the equations for $H^2_{B\,sim}$ and $\sigma^2_{Cul\ M\ sim}$, missing data adversely affects the metrics by decreasing the denominators for the cultigen interaction variances. In order to mitigate projected attrition and achieve their objectives while remaining within budget for time and resources, researchers must either conduct their experiments more efficiently or over-allocate resources, and increase costs, to compensate for missing data.

Post hoc optimization scenarios can be used to guide future experiments. Of course, by using prior estimates to predict future experimental outcomes, as presented here, we are making the assumptions that our conditions are typical, future variance component estimates will be the same, and the variance components can be infinitely partitioned. Indeed, the value of using prior data to predict outcomes will only be known following actual testing and post hoc analysis. Naturally, this additional data can then be used to further refine future experiments. While inherently flawed, using empirical variance component estimates to shape future experimental design outcomes is preferable to designing experiments based merely on resources and observations.

We simulated alternative resource allocation scenarios and calculated $H^2_{B\,sim}$ and $\sigma^2_{Cul\ M\ sim}$ in order to identify optimized testing conditions for screening for resistance to BFB in watermelon fruit and to illustrate a simple exercise to optimize useful metrics based on prior data. Scenarios 1, 3, 7, and 9 had favorable $H^2_{B\,sim}$ and $\sigma^2_{Cul\ M\ sim}$ that could be attained by running screening experiments for 2 years or less. These scenarios provide a guide for researchers and breeders to design more efficient experiments and trials based on available resources and variance component estimates.

Literature Cited

1. Branham, S. E., A. Levi, M.L. Katawczik, and W.P. Wechter. 2019. QTL mapping of resistance to bacterial fruit blotch in *Citrullus amarus*. *Theor Appl Genet*, 132, 1463-1471. https://doi.org/10.1007/s00122-019-03292-6.
2. Carvalho, F. C., L.A. Santos, R.C. Dias, R.L. Mariano, and B.S. Elineide. 2013. Selection of watermelon genotypes for resistance to bacterial fruit blotch. *Euphytica*, 190, 169-180. https://doi.org/10.1007/s10681-012-0766-1.
3. Daley, J. and T.C. Wehner. (In press). Screening for Resistance to Bacterial Fruit Blotch in Watermelon Fruit. *Crop Science.*
4. Goth, R.W. and R.E. Webb. 1981. Resistance of commercial watermelon (*Citrullus lanatus*) to *Pseudomonas pseudoalcaligenes* subsp. *citrulli*. *Plant Disease*, 65, 671-672.
5. Hopkins, D.L. and A. Levi. 2008. Progress in the Development of Crimson Sweet-type Watermelon Breeding Lines with Resistance to *Acidovorax avenae* Subsp. c*itrulli*. In M. Pitrat (Ed.), *Proceedings of the IXth Eucarpia Meeting on Genetics and Breeding of Cucurbitaceae*, 157-162. Avignon, France: INRA.
6. Hopkins, D.L. and C.M. Thompson. 2002. Evaluation of *Citrullus* sp. Germ Plasm for Resistance to *Acidovorax avenae* subsp. *citrulli*. *Plant Disease*, 86, 61-64. https://doi.org/10.1094/PDIS.2002.86.1.61.
7. Hopkins, D.L., C.M. Thompson, and G.W. Elmstrom. 1993. Resistance of watermelon seedlings and fruit to the fruit blotch bacterium. *HortScience*, 28, 122-123. https://doi.org/10.21273/HORTSCI.28.2.122.
8. Latin, R.X. and D.L. Hopkins, D. L. 1995. Bacterial fruit blotch of watermelon - the hypothetical exam question becomes reality. *Plant Disease*, 79, 761-765.
9. Lush, J. L. 1945. *Animal breeding plans*. Iowa State University Press, Ames.
10. Ma, S. and T.C. Wehner. 2015. Flowering stage resistance to bacterial fruit blotch in the watermelon germplasm collection. *Crop Science*, 55, 727-736. https://doi.org/10.2135/cropsci2014.01.0071.
11. R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
12. Schaad, N.W., E. Postnikova, and P. Randhawa. 2003. Emergence of *Acidovorax avenae* subsp. *citrulli* as a crop threatening disease of watermelon and melon. In: N.S. Iacobellis, et al. (Eds.), *Pseudomonas syringae* and related pathogens, 573-581. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-0133-4_63.
13. Schaad, N. W., E. Postnikova, A. Sechler, L.E. Claflin, A.K. Vidaver, J.B. Jones, I. Agarkova, A. Ignatov, E. Dickstein, and B.A. Ramundo. 2008. Reclassification of subspecies of *Acidovorax avenae* as *A. avenae* (Manns 1905) emend., *A. cattleyae* (Pavarino, 1911) comb. nov., *A. citrulli* Schaad et

al., 1978) comb. nov., and proposal of *A. oryzae* sp nov. *Systematic and Applied Microbiology*, 31, 434-446. https://doi.org/10.1016/j.syapm.2008.09.003.

14. Sowell, G. and N.W. Schaad. 1979. *Pseudomonas pseudoalcaligenes* subsp. *citrulli* on watermelon: seed transmission and resistance of plant introductions. *Plant Disease Reporter*, 63, 437-441.
15. Swallow, W.H. and T.C. Wehner. 1989. Optimum allocation of plots to years, seasons, locations, and replications, and its application to once-over-harvest cucumber trials. *Euphytica*, *43*, 59-68. https://doi.org/10.1007/BF00037897.

Table 1. Equations for simulated broad-sense heritability ($H^2_{B\ Sim}$) and variance of a cultigen mean ($\sigma^2_{Cul\ M\ sim}$).

Equation 1: $$H^2_{B\ sim} = \frac{\sigma^2_{Cultigen}}{\sigma^2_{Cultigen} + \frac{\sigma^2_{Cultigen \times Year}}{\text{Years}} + \frac{\sigma^2_{Block(Year) \times Cultigen}}{\text{Years X Blocks}} + \frac{\sigma^2_{Error}}{\text{Years} \times \text{Blocks} \times \text{Replications}}}$$

Equation 2: $$\sigma^2_{Cul\ M\ sim} = \frac{\sigma^2_{Cultigen \times Year}}{\text{Years}} + \frac{\sigma^2_{Block(Year) \times Cultigen}}{\text{Years X Blocks}} + \frac{\sigma^2_{Error}}{\text{Years} \times \text{Blocks} \times \text{Replications}}$$

Note. Average variance component estimates (3):
$\sigma^2_{Cultigen} = 0.454$; $\sigma^2_{Cultigen \times Year} = 0.172$; $\sigma^2_{Block(Year) \times cultigen} = 0.804$; $\sigma^2_{Error} = 3.320$. Blocks refer to blocks per year; replications refer to replications per block.

Table 2. Simulated broad-sense heritability ($H^2_{B\ Sim}$) and variance of a cultigen mean ($\sigma^2_{Cul\ M\ sim}$) optimization scenarios using different allocations of eight plots over years, blocks per year (blocks), and replications per block (replications).

	Allocations			Estimates	
Scenario	Years	Blocks	Replications	$H^2_{B\ Sim}$[z]	$\sigma^2_{Cul\ M\ sim}$[y]
1	2	2	2	0.39	0.70
2	8	1	1	0.46	0.54
3	1	8	1	0.40	0.69
4	1	1	8	0.25	1.39
5	4	2	1	0.45	0.56
6	4	1	2	0.41	0.66
7	2	4	1	0.43	0.60
8	2	1	4	0.33	0.90
9	1	4	2	0.37	0.79
10	1	2	4	0.31	0.99
$H^2_{B\ Sim}$ and $\sigma^2_{Cul\ M\ sim}$ Pearson's correlation coefficient: -.97 ($p < .001$)					

[z]Table 1, equation 1.
[y]Table 1, equation 2.