

# DATA *Nugget*

## A plant breeder's quest to improve perennial grain

Featured scientist: Hannah Stoll (she/her) from the University of Minnesota

### Research Background:

Kernza<sup>®</sup> is a new grain crop that is similar to wheat. It can be ground into flour and used in bread, cookies, crackers and more! Unlike wheat, the rest of the plant can be eaten by livestock such as cattle. Another difference is that Kernza<sup>®</sup> is a **perennial**, meaning it grows in the ground for multiple years, whereas **annual** wheat only grows for one year. However, the challenge is that annual wheat makes more grain and is easier to harvest and sell. This means farmers currently prefer growing annual wheat over Kernza<sup>®</sup>.

One way to address this mismatch between annual and perennial crops is through **selective breeding**. This is when humans select individual plants with traits that are desirable for a specific reason. This group of individuals are strategically bred together. The breeder's goal is to shift the traits over generations. Scientists have only been working on breeding Kernza<sup>®</sup> for the past few decades; in comparison, humans started selecting annual wheat traits over 10,000 years ago! That is a lot of time to get the traits we are looking for.



Hannah collecting data on Kernza<sup>®</sup>.

Kernza<sup>®</sup> breeders are working on improving the same traits that have already been improved in annual wheat, including larger seed size. Kernza<sup>®</sup> scientists follow two main steps to breed plants 1) they select the best individuals from the population and 2) they intercross those individuals to create the next generation, or **breeding cycle**. With each breeding cycle, plant breeders see a slight improvement in the traits they selected.

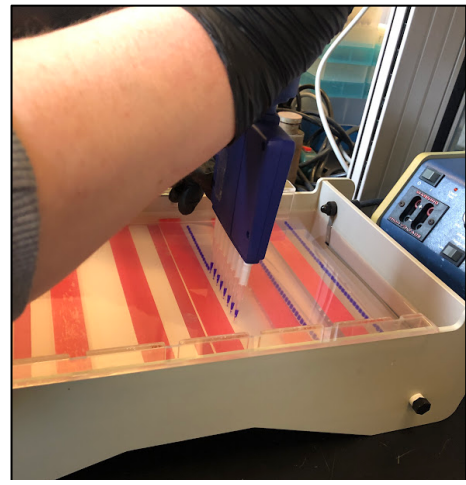
Breeders can select plants based on phenotypes, genotypes, or both. Historically, plant breeders have selected based on desired **phenotypes**, or visible traits, only. Modern plant breeding can take advantage of the fact that we can now look at **genotypes**, or the genetic makeup, of individual plants quickly and at low costs. Scientists can use this information to make quicker breeding

improvements, so we don't have to wait another 10,000 years for high-yielding Kernza®!

Hannah is a scientist currently working on Kernza®. Hannah's passion for plant breeding was ignited during her high school years. She discovered the captivating world of genetics in her AP Biology class. It was then that she first realized the potential for breeding crop plants to make them more productive and viable for human consumption.

Hannah decided to join other scientists who work on Kernza® at the University of Minnesota. Here, scientists have completed four breeding cycles and are about to start the fifth. Hannah wanted to see whether different genetic makeups (genotypes) lead to differences in seed size (phenotypes). Her goal was to look at each plants' phenotype and genotype for seed size.

To genotype a plant, scientists collect a small piece of leaf tissue, extract the DNA, and send the DNA to a lab for **sequencing**. This process tells scientists the genetic makeup that ultimately leads to the traits that we see. Specifically, sequencing data identifies **nucleotides**, or genetic building blocks of each plant's DNA. Plants have thousands of genes, which are made up of the DNA nucleotides A, T, C, and G. Sequencing data can be recorded in several ways. One common way is as **SNP data**, or **Single Nucleotide Polymorphism** data. You can think of SNP data as the recipe for proteins. In a SNP dataset, each SNP represents a *difference* in a **nucleotide**. Similar to using a different ingredient in a recipe, different nucleotides can result in a different phenotype.



DNA samples being tested for quality before genetic sequencing.

By looking at SNP data, plant breeders can identify differences in genotypes that lead to certain phenotypes. Hannah started by evaluating 1,000 Kernza® plants from the first four breeding cycles. Data on phenotypes had already been recorded for these plants. Hannah then collected SNP data to determine their genotypes as well. She was looking for a pattern between genotypes and phenotypes. If she sees that different genotypes have different phenotypes, scientists can then rely on genotypes to select individuals to breed in future breeding cycles.

***Scientific Question:*** What genotype should a plant breeder select if they are trying to increase seed size in the next generation?

***What is the hypothesis?*** Find the hypothesis in the Research Background and underline it. A hypothesis is a proposed explanation for an observation, which can then be tested with experimentation or other types of studies.

Scientific Data:

**Complete Table 1a below, using the sequencing data for each plant to record its genotype.**

*Note: The table contains a subset, or smaller part, of data from Hannah's research. Hannah found that two nucleotides in the genotype were important for predicting the phenotype of Kernza® seed size. Differences in SNP 01 and SNP 04 significantly impacted the phenotype values. Data for these SNPs are included below.*

Table 1a. Phenotype and sequencing data for seed size from breeding cycles 1-4.

Plant ID	Breeding Cycle	Seed area (mm)	SNP 01	SNP 04	Genotype
1_1	1	5.4	A	G	AG
2_1	3	6.4	A	C	AC
3_1	2	7.5	T	A	TA
4_1	3	7.7	T	A	
5_1	2	6.9	A	C	
6_1	4	7.1	T	G	
7_1	1	6.6	A	C	
8_1	4	7.9	T	C	
9_1	3	6.8	A	C	
10_1	2	5.6	A	G	
11_1	1	6.1	A	A	
12_1	4	6.7	A	C	
13_1	1	6.7	T	G	
14_1	3	7.2	T	A	
15_1	4	7.6	T	C	
16_1	3	7.0	T	G	
17_1	1	6.5	A	C	
18_1	4	7.4	T	A	
19_1	2	7.3	T	A	
20_1	3	7.9	T	C	
21_1	3	7.6	T	A	
22_1	2	7.8	T	C	
23_1	1	7.0	T	G	
24_1	4	8.0	T	C	
25_1	2	6.3	A	A	
26_1	4	7.7	T	A	
27_1	1	5.7	A	G	
...	...	...	...	...	
1000_1	2	6.9	T	G	

Now, complete Table 1b with the average seed area for each genotype.

Table 1b. Summary of seed area phenotype based on genotype for breeding cycles 1-4.

Genotype	Average Seed Area (mm)	Sample Size	Standard Deviation	Standard Error
AA	6.20	2	0.14	0.10
AC			0.19	0.08
AG			0.15	0.09
TA			0.20	0.07
TC			0.15	0.07
TG			0.15	0.07

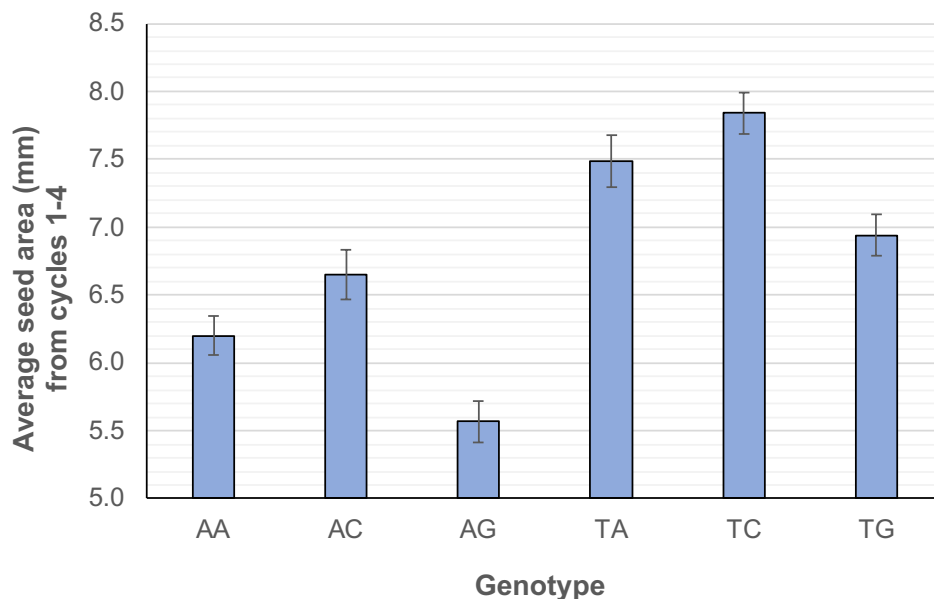
Next, use the data in Tables 1a and 1b to answer the scientific question.

What data will you graph to answer the question?

Independent variable: \_\_\_\_\_

Dependent variable: \_\_\_\_\_

Below is a graph of the data: Identify any changes, trends, or differences you see in your graph. Draw arrows pointing out what you see and write one sentence describing what you see next to each arrow.



Name \_\_\_\_\_

*Interpret the data:*

Make a claim that answers the scientific question, what genotype should a plant breeder select if they are trying to increase seed size in the next generation?

What evidence was used to write your claim? Reference specific parts of the tables or graph.

Explain your reasoning and why the evidence supports your claim. Connect the data back to what you learned about how DNA sequencing for genotypes can inform selective breeding.

Name \_\_\_\_\_

Did the data support Hannah's hypothesis? Use evidence to explain why or why not. If you feel the data are inconclusive, explain why.

*Your next steps as a scientist:* Science is an ongoing process. What new question(s) should be investigated to build on Hannah's research? How do your questions build on the research that has already been done?