

Seed-X

Astonishing results in a variety of applications

Seed Breeding
& Research

Seed-X.com





Environmental Effect

Genetic identity vs. environmental effect. Different seed lots of same line under varying growing conditions, location and season, and affected seed phenotypes.

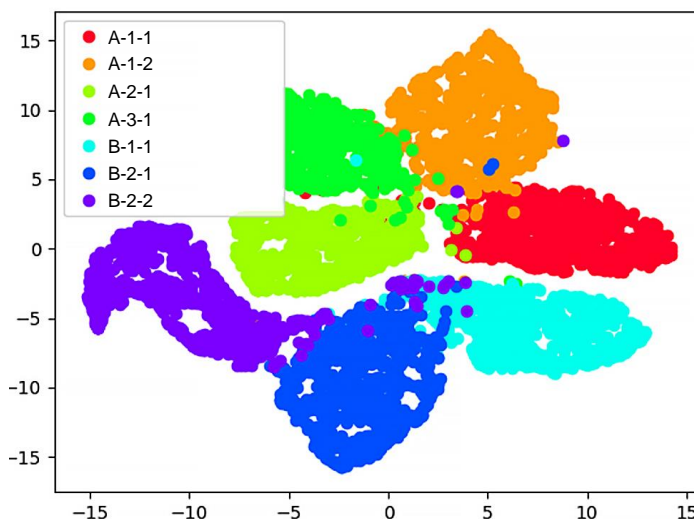
Target: Classification between 2 tomato seed types (F1) grown under different environments Genetic identity vs. environmental effect

Two tomato hybrids, A and B, were produced in three locations over four seasons. All seeds were treated according to common industry practices. Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology. For each line, images were split randomly into two groups: 90% training and 10% test. This process was repeated 10 times. Seed-X technology was used to predict seed for test set images.

For each seed image of these sets, Seed-X technology outputs the probability that the seed belongs to the trained hybrids. A similarity matrix was plotted to visualize the similarity between different varieties.

Two tomato hybrids A and B, were produced in three and four locations respectively. The above proximity distribution highlights:

1. A clear difference between line A and line B
2. Within each of the lines, the different seed lots share close phenotypes
3. A clear separation between the different lots (location/season).





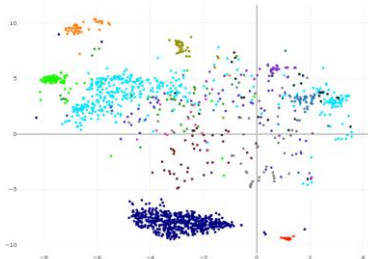
Proximity Distribution

Proximity distribution used scatter plots to visualize seed variance in a two-dimensional space.

Target: To identify connection between seed phenotype and the genetics of segregated populations or specific seeds within a segregated population.

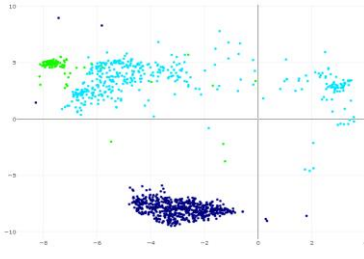
A tomato segregated family includes two parents, P1 and P2. Their F1, F2, eight F3 selections and two F4 selections were chosen. All seeds were treated according to common industry practices. Hundreds of seeds from each variety were analyzed using Seed-X unique GeNee™ technology. For each line, images were split randomly into two groups: 90% for training and 10% for testing. This process was repeated 10 times. Seed-X technology was used to predict origin of the test set seeds and to demonstrate relatedness to other family members.

Overall General distribution



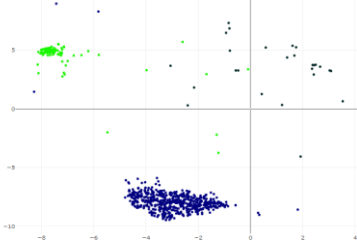
Overall distribution including the two parents, P1 and P2, their F1, F2, eight F3 selections, and two F4 selections

Highlight on P1, P2 and F2 Segregation



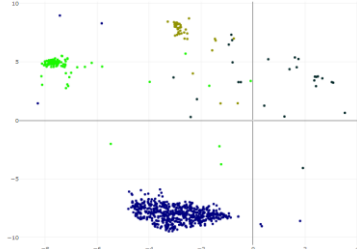
Parental lines represent two types of tomato plants: P1– small fruit size, good taste, P2- large fruit size, poor taste. F2 population segregates between the two parental lines

F3 Selection



F3 selected seeds distribute between the two parental lines, with a clear bias towards P1

F4 Selection



F4 seed distribution is significantly reduced, and the bias toward P1 is significant





Seed-X

Hybrid Classification

Confusion matrix of hybrid tomato seeds

Target: Classify tomato hybrids, based on seed phenotypes

Eight tomato hybrids, were grown in the same greenhouse, over the same season. All seeds were treated according to common industry practices. Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology. For each hybrid, images were split randomly into two groups: 90% training and 10% test. This process was repeated 10 times.

The Seed-X technology was used to predict the seed hybrid for the test set images. For each seed image of these sets, the Seed-X technology outputs the probability of the seed to belong to the trained hybrids.

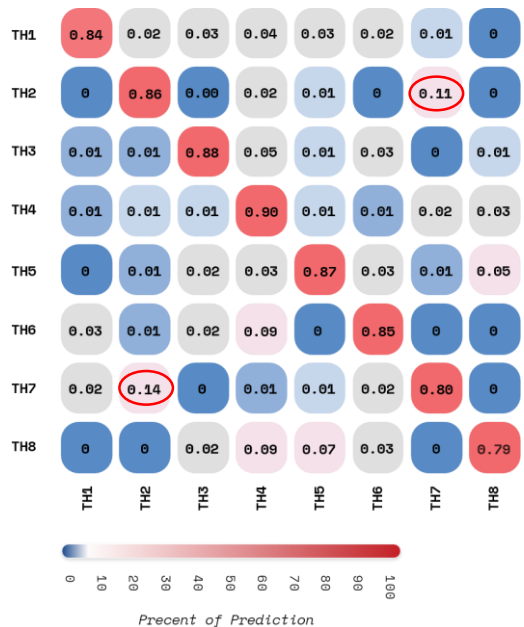
A confusion matrix was plotted to visualize the similarity between different varieties.

Confusion Matrix Results

In this Confusion Matrix, row labels indicate the real hybrids analyzed and column labels illustrate predicted hybrids. Numbers and colors assigned to each cell reflect the sum of seed images predicted as pertaining each hybrid.

For example, the first cell (first row, first column to the left) is marked with 126, which means that 84% (125/150) of "TH1" seeds of the "TH1" seeds (because they are in the "TH1" row) were indeed predicted as "TH1" (because they are in the "TH1" column). All values on the diagonal are correct predictions, where all other values are errors.

The confusions between TH2 and TH7 (marked by circles), prompted us to investigate the genetic connection between them. We found that these varieties share a common maternal parent.





Seed-X

Hybrid Classification

Confusion matrix of hybrid tomato seeds

Target: Classified tomato hybrids, based on seed phenotypes

Eight tomato hybrids, were grown in the same greenhouse, over the same season. All seeds were treated according to common industry practices. Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology. For each isolate, images were split randomly into two groups: 90% training and 10% test. This process was repeated 10 times.

The Seed-X technology was used to predict the seed isolate for the test set images. For each seed image of these sets, the Seed-X technology outputs the probability of the seed to belong to the trained hybrids.

A Confusion matrix was plotted to visualize the similarity between different varieties.

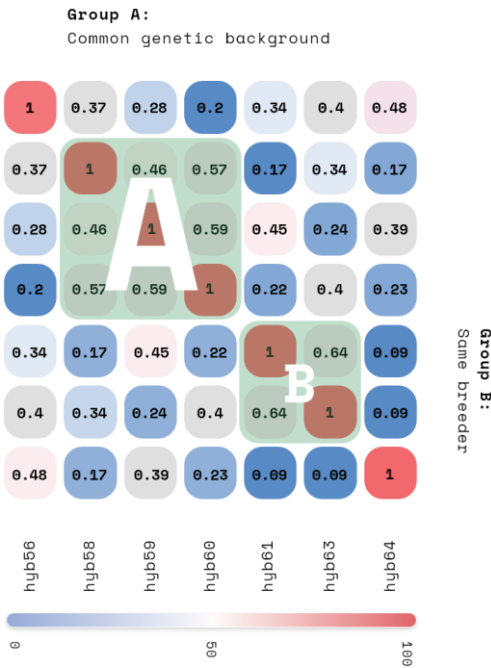
Confusion Matrix Results

This Confusion Matrix represents Seed-X GeNee™ envision distance analysis, comparing between tested hybrids.

The distance is represented in numbers, 0-1 scale, where 0 symbolizes no connection between the hybrids and 1 is exact match. The colors correlate to the number according to the color bar on the bottom. We highlight the groups of hybrids that show high Confusion.

Group A – Hyb58, 59 and 60; the maternal parent of these three hybrids originate from the same genetic background (M82).

Group B – we have no additional information regarding the genetic material, but these two lines come from the same breeder (might be considered to have the same genetic background).





Biotic Trait Detection

Isogenic Line Classification – Virus, Nematode and Fungus resistance

Target: Correlate seed phenotypes with biotic traits

Seeds of isogenic lines, presenting either resistance or susceptibility to different biotic stresses were grown under the same environmental conditions and in the same season.

All seeds were treated according to common industry practices. Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology.

For each variety, images were split randomly into two groups, 90% training and 10% test. This process was repeated 10 times. The Seed-X technology was used to predict seed isoline based on the test set images. The percentage of correct predictions for each line is presented at the table below.

Summary of results

Crop	Trait Group	Sub Trait	Precision*	Sensitivity**
Tomato	Virus	TSWV resistance	1	0.95
Tomato	Virus resistance	TYLCV resistance	1	0.96
Tomato	Virus	TMV resistance	0.96	0.94
Tomato	Nematode	Gene 1	0.97	0.92
Tomato	Nematode resistance	Gene 2	0.96	0.9
Melon	Fungus resistance	Powdery mildew resistance	0.97	0.89

* Precision: Rate of seeds-of-interest of all seeds predicted as such (including false positives)

** Sensitivity: Rate of seeds-of-interest predicted as such of all seeds-of-interest (including false negatives)





Consumer trait Detection

Isogenic Line Classification - Fruit Color, Shape, Size, Structure

Target: Correlate seed phenotypes and consumer traits

Seeds of isogenic lines, presenting different consumer traits were grown under the same environmental conditions and in the same season. All seeds were treated according to common industry practices. Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology.

For each isoline, images were split randomly into two groups: 90% training and 10% test. This process was repeated 10 times. The Seed-X technology was used to predict seed isoline based on the test set images. The percentage of correct predictions for each isoline is presented here.

Summary of results

Crop	Trait Group	Sub Trait	Precision*	Sensitivity**
Tomato	Fruit color	Purple/red	0.99	0.83
Tomato	Fruit color	Orange/red	0.95	0.69
Tomato	structure	Round/oval	0.93	0.9
Tomato	structure	Compact plant	0.82	0.71
Melon	Fruit color	Pale	0.96	0.96
Melon	Flesh color	Flesh	0.99	0.99
Melon	Structure	Shape	0.9	0.81
Melon	Taste	PH	0.91	0.81
Pepper	Color	Red/yellow	0.99	0.96
Pepper	Fruit size	Small/large	0.99	1
Pepper	Structure	Determinate	0.92	0.89

* Precision: Rate of seeds-of-interest of all seeds predicted as such (including false positives)

** Sensitivity: Rate of seeds-of-interest predicted as such of all seeds-of-interest (including false negatives)

Seed-X

Astonishing results in a variety of applications

Seed Production

Seed-X.com





Seed-X

Hybrid Classification

Confusion matrix of hybrid tomato seeds

Target: Classify tomato hybrids, based on seed phenotypes

Eight tomato hybrids, were grown in the same greenhouse, over the same season. All seeds were treated according to common industry practices. Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology. For each hybrid, images were split randomly into two groups: 90% training and 10% test. This process was repeated 10 times.

The Seed-X technology was used to predict the seed hybrid for the test set images. For each seed image of these sets, the Seed-X technology outputs the probability of the seed to belong to the trained hybrids.

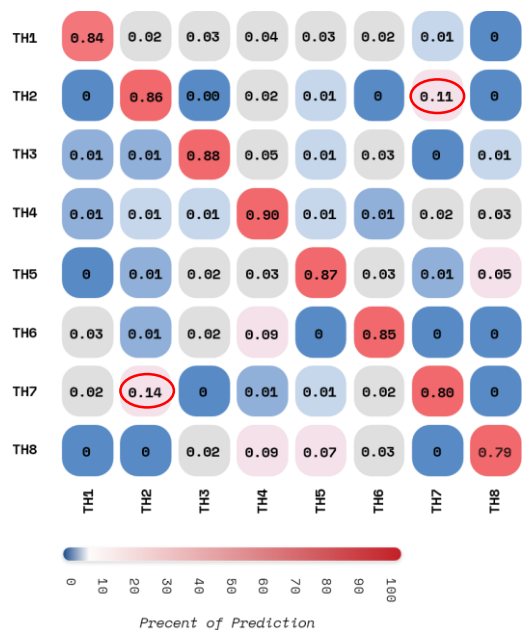
A confusion matrix was plotted to visualize the similarity between different varieties.

Confusion Matrix Results

In this Confusion Matrix, row labels indicate the real hybrids analyzed and column labels illustrate predicted hybrids. Numbers and colors assigned to each cell reflect the sum of seed images predicted as pertaining each hybrid.

For example, the first cell (first row, first column to the left) is marked with 126, which means that 84% (125/150) of "TH1" seeds of the "TH1" seeds (because they are in the "TH1" row) were indeed predicted as "TH1" (because they are in the "TH1" column). All values on the diagonal are correct predictions, where all other values are errors.

The confusions between TH2 and TH7 (marked by circles), prompted us to investigate the genetic connection between them. We found that these varieties share a common maternal parent.





Seed-X

Hybrid Classification

Confusion matrix of hybrid tomato seeds

Target: Classified tomato hybrids, based on seed phenotypes

Eight tomato hybrids, were grown in the same greenhouse, over the same season.

All seeds were treated according to common industry practices.

Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology.

For each isolate, images were split randomly into two groups: 90% training and 10% test.

This process was repeated 10 times.

The Seed-X technology was used to predict the seed isolate for the test set images.

For each seed image of these sets, the Seed-X technology outputs the probability of the seed to belong to the trained hybrids.

A Confusion matrix was plotted to visualize the similarity between different varieties.

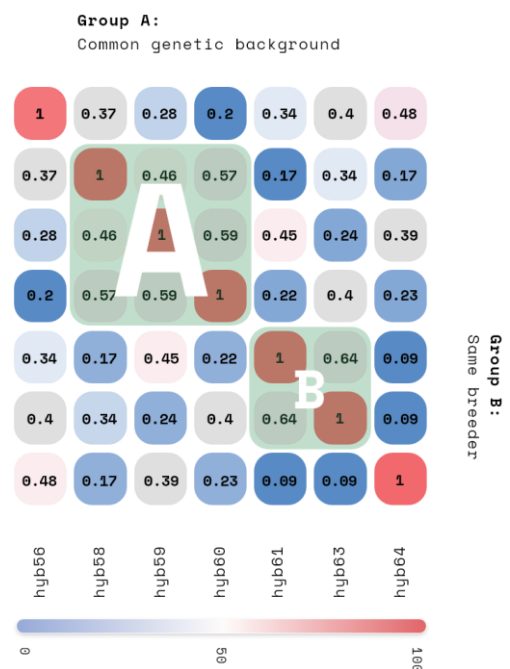
Confusion Matrix Results

This Confusion Matrix represents Seed-X GeNee™ envision distance analysis, comparing between tested hybrids.

The distance is represented in numbers, 0-1 scale, where 0 symbolizes no connection between the hybrids and 1 is exact match. The colors correlate to the number according to the color bar on the bottom. We highlight the groups of hybrids that show high Confusion.

Group A – Hyb58, 59 and 60; the maternal parent of these three hybrids originate from the same genetic background (M82).

Group B – we have no additional information regarding the genetic material, but these two lines come from the same breeder (might be considered to have the same genetic background).





Seed-X

Environmental Effect

Genetic identity vs. environmental effect. Different seed lots of same line under varying growing conditions, location and season, and affected seed phenotypes.

Target: Classification between 2 tomato seed types (F1) grown under different environments

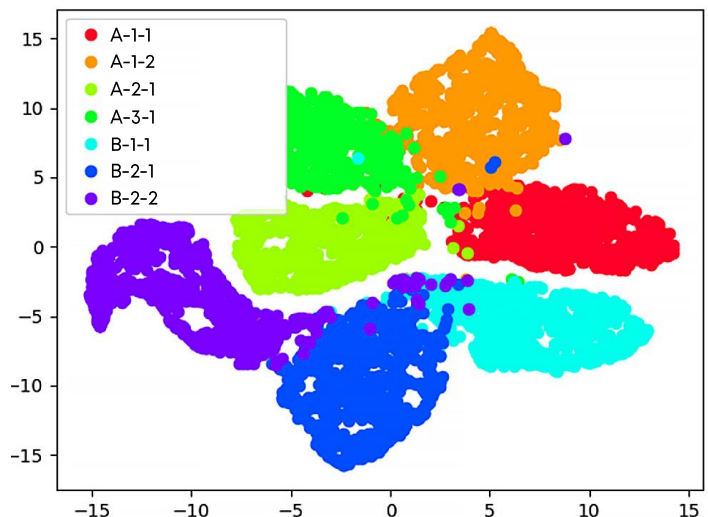
Genetic identity vs. environmental effect

Two tomato hybrids, A and B, were produced in three locations over four seasons. All seeds were treated according to common industry practices. Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology. For each line, images were split randomly into two groups: 90% training and 10% test. This process was repeated 10 times. Seed-X technology was used to predict seed for test set images.

For each seed image of these sets, Seed-X technology outputs the probability that the seed belongs to the trained hybrids. A similarity matrix was plotted to visualize the similarity between different varieties.

Two tomato hybrids A and B, were produced in three and four locations respectively. The above proximity distribution highlights:

1. A clear difference between line A and line B
2. Within each of the lines, the different seed lots share close phenotypes
3. A clear separation between the different lots (location/season).





Seed-X

Impurity Detection

Hybrids & Open Pollinated (OP) Tomato, Pepper, Melon, Wheat, Corn, Soybean, Lettuce, Cucumber

Target: Detect the level of off-type seeds contamination during seed production.

Demonstrate the potential to identify off-type seeds within a batch of a specific (known) variety

For each crop (Hybrid or OP), different varieties were grown in the same field (same conditions & season) and were compared to each other. All seeds were treated according to common industry practices. Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology.

For each variety, images were split randomly into two groups; 90% training and 10% test. This process was repeated 10 times.

The Seed-X GeNee™ technology was used to predict seed variety (of all trained varieties) for test set images. Prediction precision and sensitivity measures are given below.

Summary of results

Crop	Type	# of tested lines	*Precision		**Sensitivity	
			Average	Range	Average	Range
Tomato	Hybrid	7	0.97	0.93-0.99	0.87	0.76-0.96
Melon	Hybrid	3	0.99	0.98-0.99	0.99	0.98-0.99
Wheat	OP	6	0.97	0.95-1	0.91	0.88-0.94
Pepper	Hybrid	10	0.98	0.92-1	0.96	0.82-1
Corn	Hybrid	7	0.99	0.92-1	0.92	0.81-1
Soybean	OP	4	0.98	0.94-1	0.94	0.92-0.95
Lettuce	OP	2	0.99	0.98-0.990	0.99	0.98-0.99
Cucumber	Hybrid	3	0.99	0.98-1	0.95	0.93-0.96

* Precision: Rate of seeds-of-interest of all seeds predicted as such (including false positives)

** Sensitivity: Rate of seeds-of-interest predicted as such of all seeds-of-interest (including false negatives)



Seed-X

Hybrids vs. Self-Identification

Tomato, Pepper, Corn

Target: Detect the level of self-pollination during hybrid seed production

Samples of hybrid seeds (from a field of hybrid production), and their female parental lines (self-pollinated) were taken from a greenhouse/field of hybrid production.

All seeds were treated according to common industry practices.

Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology.

10% of seed images of each variety were split and served as test sets.

Seed-X Genee™ technology predicted seed variety (of all trained varieties) for test set images.

Prediction precision and sensitivity measures are given below.

Summary of results

Crop	# of tested pairs	Precision*		Sensitivity**	
Tomato	3	0.99	0.98-1	0.90	0.87-0.95
Pepper	3	0.99	0.98-1	0.89	0.8-1
Corn	2	0.92	0.91-0.94	0.85	0.80-0.9

* Precision: Rate of seeds-of-interest of all seeds predicted as such (including false positives)

** Sensitivity: Rate of seeds-of-interest predicted as such of all seeds-of-interest (including false negatives)





Seed-X

Identify Ploidy level

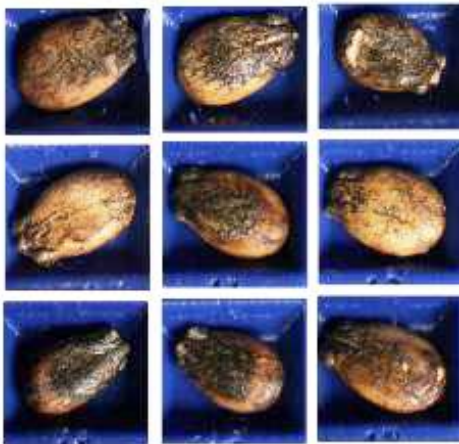
Tetraploid vs. Triploid in Watermelon Seeds

Target: Differentiate hybrid triploid watermelon seeds from its self seeds (tetraploids)

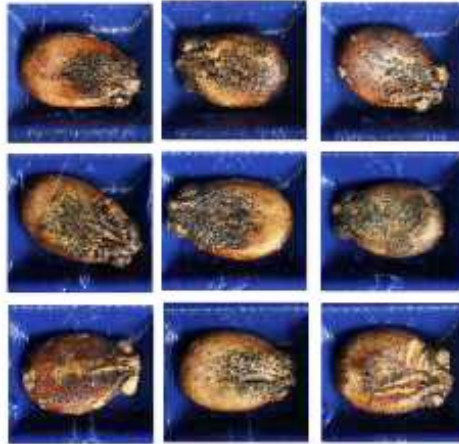
Seedless watermelon hybrids (triploids) are currently produced using manual pollination. Sorting tetraploid seeds from triploid will facilitate bee pollination for production of these hybrids. Samples of triploid hybrid seeds and their maternal tetraploid (self) lines were collected from a field of hybrid production. All seeds were treated according to common industry practices.

Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology. For each variety, images were split randomly into two groups: 90% training and 10% test. This process was repeated 10 times. Seed-X GeNee™ technology platform was used to distinguish the self seeds from the hybrids for each self-hybrid pair. Correct prediction rate was >98%.

Triploid Seeds (Watermelon)



Tetraploid Seeds (Watermelon)



Summary of results

Date quantity	Tetra Vs. Triploid
Data quantity	X
Random choice probability	50%
Current results	>98%



Seed-X

GMO Identification

Distinguishing between GMO and Non-GMO Corn Kernels

Target: Identification of GMO kernels

Samples of GMO corn kernels and their non-transgenic lines were taken from a field of hybrid production. All seeds were treated according to common industry practices.

Thousands of kernels from each variety were analyzed using Seed-X unique GeNee™ technology. 10% of kernel images of each variety were split and served as test sets.

Seed-X Genee™ technology predicted seed variety for test set images. Prediction precision and sensitivity measures are given below.

Summary of results

Line	GMO		Non-GMO	
	Precision*	Sensitivity**	Precision	Sensitivity
A	0.92	0.8	0.87	0.89
B	0.97	.87	0.97	0.91
C	0.97	0.88	0.97	0.97
E	0.98	0.92	0.99	0.96
F	0.96	0.98	0.98	0.96
G	0.98	1.00	1.00	0.98
Average	0.96	0.91	0.96	0.93

* Precision: Rate of seeds-of-interest of all seeds predicted as such (including false positives)

** Sensitivity: Rate of seeds-of-interest predicted as such of all seeds-of-interest (including false negatives)



Biotic Trait Detection

Isogenic Line Classification – Virus, Nematode and Fungus resistance

Target: Correlate seed phenotypes with biotic traits

Seeds of isogenic lines, presenting either resistance or susceptibility to different biotic stresses were grown under the same environmental conditions and in the same season.

All seeds were treated according to common industry practices. Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology.

For each variety, images were split randomly into two groups, 90% training and 10% test. This process was repeated 10 times. The Seed-X technology was used to predict seed isoline based on the test set images. The percentage of correct predictions for each line is presented at the table below.

Summary of results

Crop	Trait Group	Sub Trait	Precision*	Sensitivity**
Tomato	Virus	TSWV resistance	1	0.95
Tomato	Virus resistance	TYLCV resistance	1	0.96
Tomato	Virus	TMV resistance	0.96	0.94
Tomato	Nematode	Gene 1	0.97	0.92
Tomato	Nematode resistance	Gene 2	0.96	0.9
Melon	Fungus resistance	Powdery mildew resistance	0.97	0.89

* Precision: Rate of seeds-of-interest of all seeds predicted as such (including false positives)

** Sensitivity: Rate of seeds-of-interest predicted as such of all seeds-of-interest (including false negatives)





Seed-X

Consumer trait Detection

Isogenic Line Classification - Fruit Color, Shape, Size, Structure

Target: Correlate seed phenotypes and consumer traits

Seeds of isogenic lines, presenting different consumer traits were grown under the same environmental conditions and in the same season. All seeds were treated according to common industry practices. Thousands of seeds from each variety were analyzed using Seed-X unique GeNee™ technology.

For each isoline, images were split randomly into two groups: 90% training and 10% test. This process was repeated 10 times. The Seed-X technology was used to predict seed isoline based on the test set images. The percentage of correct predictions for each isoline is presented here.

Summary of results

Crop	Trait Group	Sub Trait	Precision*	Sensitivity**
Tomato	Fruit color	Purple/red	0.99	0.83
Tomato	Fruit color	Orange/red	0.95	0.69
Tomato	structure	Round/oval	0.93	0.9
Tomato	structure	Compact plant	0.82	0.71
Melon	Fruit color	Pale	0.96	0.96
Melon	Flesh color	Flesh	0.99	0.99
Melon	Structure	Shape	0.9	0.81
Melon	Taste	PH	0.91	0.81
Pepper	Color	Red/yellow	0.99	0.96
Pepper	Fruit size	Small/large	0.99	1
Pepper	Structure	Determinate	0.92	0.89

* Precision: Rate of seeds-of-interest of all seeds predicted as such (including false positives)

** Sensitivity: Rate of seeds-of-interest predicted as such of all seeds-of-interest (including false negatives)



Seed-X

Germination Prediction

Germination Enrichment in Tomato

Target: Detect non-germinate seeds in low germination seed lot

Seed-X received an unsellable tomato seed batch with a germination rate of 58%.

GeNee™ technology was trained to distinguish between germinating and non-germinating seeds by analyzing seed images captured by GeNee™ Detect. Once trained, GeNee™ technology was eligible to predict germination. Germination prediction scores were assigned per seed and conjointly with a tunable Decision Threshold used to generate two models balancing precision (percentage of germinated seeds in the final batch) and sensitivity (percentage of seeds that were marked).

Phenotypes			
Prediction	Germinating	Non-Germinating	Sum
Germinating	79	7	86
Non-Germinating	2	66	68
Sum	81	73	154

56%
Initial Germination Rate

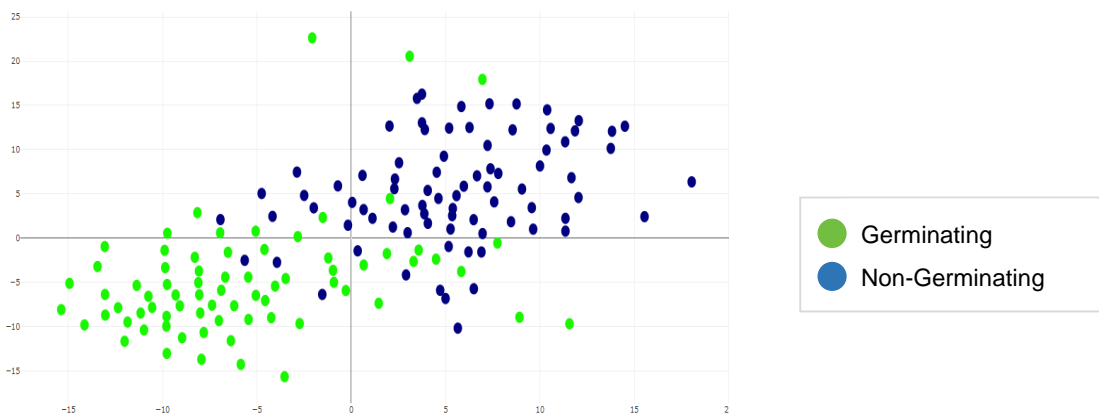
$$86/154=0.56$$

92%
Germination After
Seed-X Refinement

$$79/86=0.92$$

98%
Sensitivity*

$$79/81=0.98$$



* Sensitivity: Rate of seeds-of-interest predicted as such of all seeds-of-interest (including false negatives)



Seed-X

Germination Prediction

Germination Enrichment in Pepper

Target: Detect non-germinate seeds in low germination seed lot

Seed-X received an unsellable pepper seed batch with a germination rate of 82%.

GeNee™ technology was trained to distinguish between germinating and non-germinating seeds by analyzing seed images captured by GeNee™ Detect. Once trained, GeNee™ technology was eligible to predict germination. Germination prediction scores were assigned per seed and conjointly with a tunable Decision Threshold used to generate two models balancing precision (percentage of germinated seeds in the final batch) and sensitivity (percentage of seeds that were marked).

Phenotypes			
Prediction	Germinating	Non-Germinating	Sum
Germinating	141	8	149
Non-Germinating	44	28	72
Sum	185	36	221

84%
Initial Germination
Rate

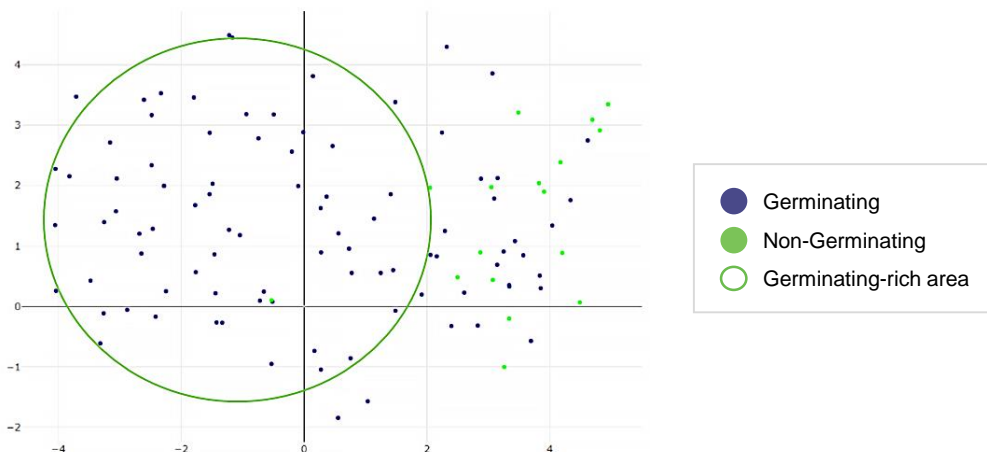
$$185/221=0.84$$

95%
After Seed-X
refinement

$$141/149=0.95$$

76%
Sensitivity*

$$141/185=0.76$$



* Sensitivity: Rate of seeds-of-interest predicted as such of all seeds-of-interest (including false negatives)



SeedX

Germination Prediction

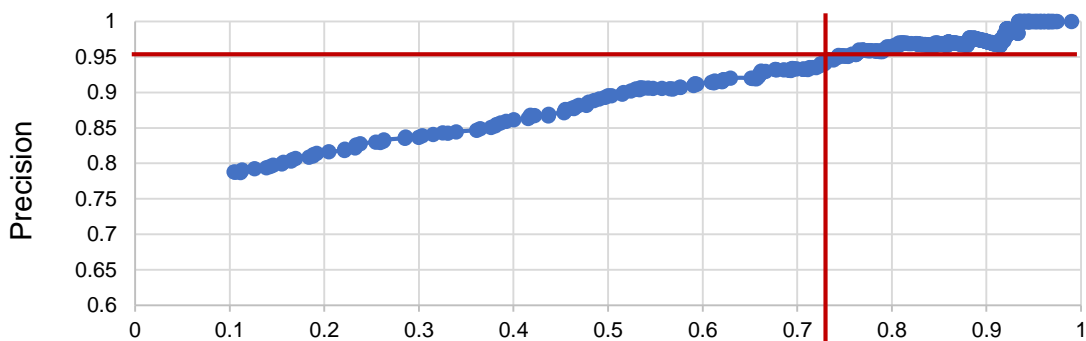
Precision vs. Sensitivity

Example: Increasing 62% germination to 95% is coupled with 100% to 85% reduction in sensitivity, translated to 44% waste

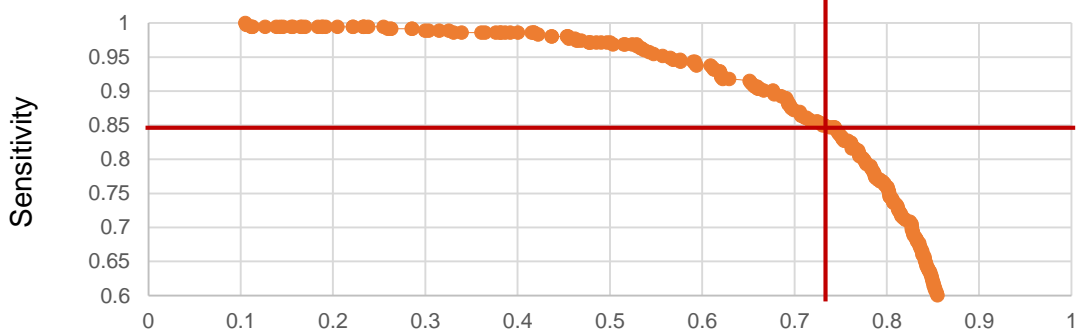
Calculation

- In 62% germination, of every 100 seeds only 62 germinate. Sensitivity is 100% and waste is 0%.
- Setting 95% precision results with 85% sensitivity (see plot below)
 - 85% Sensitivity – of 62 germinating seeds, 53 are recovered ($0.85 \times 62 = 53$)
 - 95% precision – of every 100 seeds, 56 seeds ($53 / 0.95 = 56$) are collected, resulting in a waste of 44%

Precision as a function of Decision Threshold



Sensitivity as a Function of Decision Threshold



Decision Threshold

* Precision: Rate of seeds-of-interest of all seeds predicted as such (including false positives)

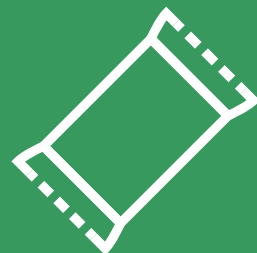
** Sensitivity: Rate of seeds-of-interest predicted as such of all seeds-of-interest (including false negatives)

Seed-X

Astonishing results in a variety of applications

Food Production

Seed-X.com





SeedX

Determine quality traits in bread wheat grains

Concentration and composition of grain protein and microelements

Target: Use Seed-X non-destructive screening to determine quality traits in bread wheat grains

EcoBreed and Seed-X collaborate to improve grain quality using with the goal of improving wheat grain quality using Seed-X unique GeNee™ technology and EcoBreed proprietary germplasm. Grain protein and microelements concentration and composition is an important quality measure which defines nutritional value.

The concentration and composition of grain protein and microelements constitute an important quality measure of nutritional value.

A set of unique landraces, comprising a wide range of nutritional quality measures (protein and microelements content), was selected: 26 bread wheat lines, 21 unique landraces and 5 commercial varieties. This genetic collection was grown under the same conditions by EcoBreed and characterized for different quality traits, such as protein content and different microelements, including Mg (Magnesium), Zn (Zinc) and K (Potassium).

Grains of all lines were captured and used to train Seed-X seed-level AI-driven phenotyping system to identify unique signatures in each line. These signatures were then used to classify the lines across 10 clusters (see table and figure below). These clusters were found to credibly correlate with lines' quality measures, i.e., lines within each cluster are ascribed with nutritional values that are more similar to those held by external lines. .

Summary of results [Next Page](#)



Determine quality traits in bread wheat grains

Concentration and composition of grain protein and microelements

Summary of results

Cluster 1 includes 4 of the commercial varieties present in this study. It can be seen that it presents low quality - low protein units 11.9, and low Mg 1026 ug per seed. On the other hand, this cluster has large grain size that correlates with high yield. **Cluster 3** includes 5 landraces. It presents medium quality - protein 14.2, and Mg 1370 mic g and medium grain size. **Cluster 8 and 10** - each includes 2 landraces, both clusters present small grain size with a high level of Mg, over 1500 mic g, cluster 8 present high level of protein while cluster 10 holds moderate to high protein levels. .

Cluster Number	# of lines in the cluster	Ug per Seed	Protein (%DW) Units	Seed area mm
1	4	1026	11.9	19.3
2	2	1139	13.0	18.8
3	5	1370	14.2	15.2
4	1	1344	14.3	15.9
5	2	1338	16.8	15.4
6	1	1412	21.2	18.6
7	7	1460	15.1	14.0
8	3	1539	16.4	14.3
9	2	1480	16.9	20.3
10	2	1536	15.5	14.4

Table 1: Summary of the different clusters created using the seed images average quality parameters

Dendrogram

Cluster 1:
Low Protein, Low Mg.,
Large Grain

Cluster 3:
Protein Medium,
Mg. Medium, Grain
Medium

Cluster 8:
Protein High, Mg. High,
Grain Small

Cluster 10:
Protein Medium, Mg. High,
Grain Small

