Data and Process-Driven Decision-Making To Increase Plant Breeding Efficiency

Michel Ragot, DELTAGee

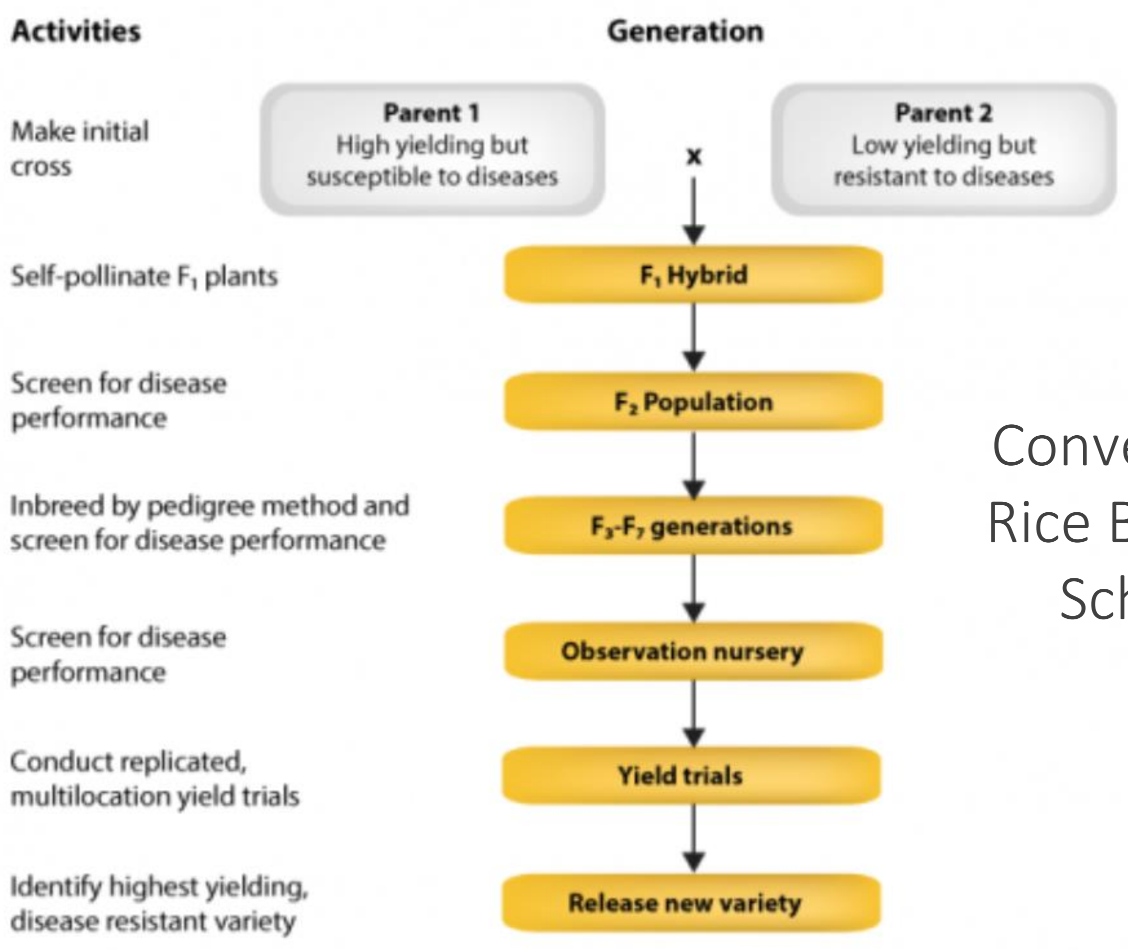
XIV International Rice Conference, Panama
June 10-14, 2024

BACKGROUND

Plant Breeding is a Process

Starting Point:
Germplasm

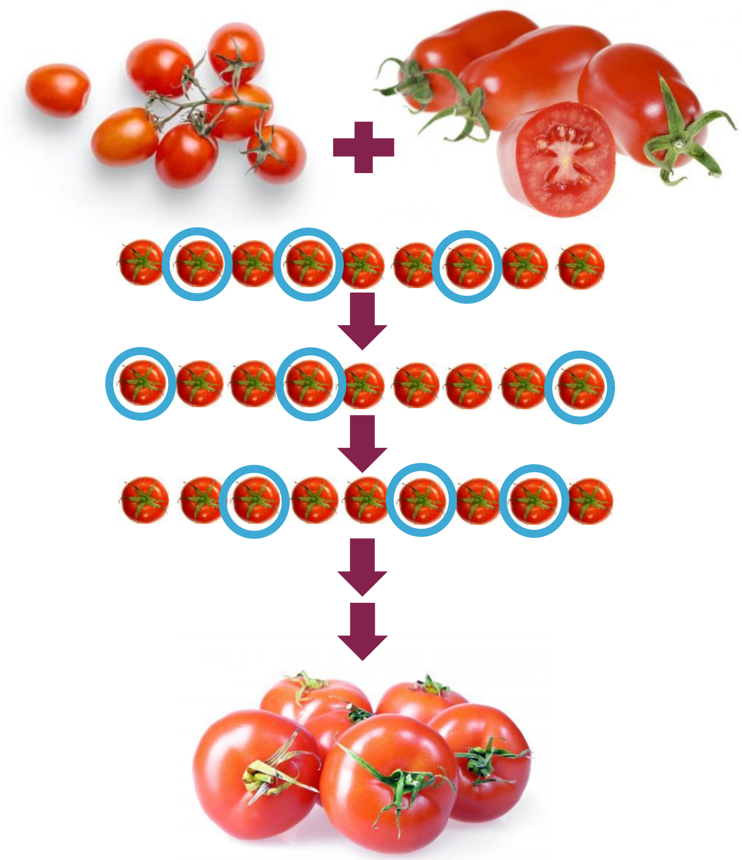
End Point:
Variety Release



Conventional
Rice Breeding
Scheme

A.Singh, A.Singh, A.A.Mahama; Rice Breeding (2023) *In:* W.Suza, K.Lamkey (Eds.); Crop Improvement; Iowa State University Digital Press. DOI: 10.31274/isudp.2023.138

How to Improve Plant Breeding Efficiency?



Improve:

- Selection intensity
- **Selection accuracy**
- Genetic variance
- Cycle time

$$\Delta G = \frac{ir\sigma_A}{y}$$

Decision-Making for Process Optimization

Risk of not getting there?
Time it will take?
Resources I will need?

Start

End

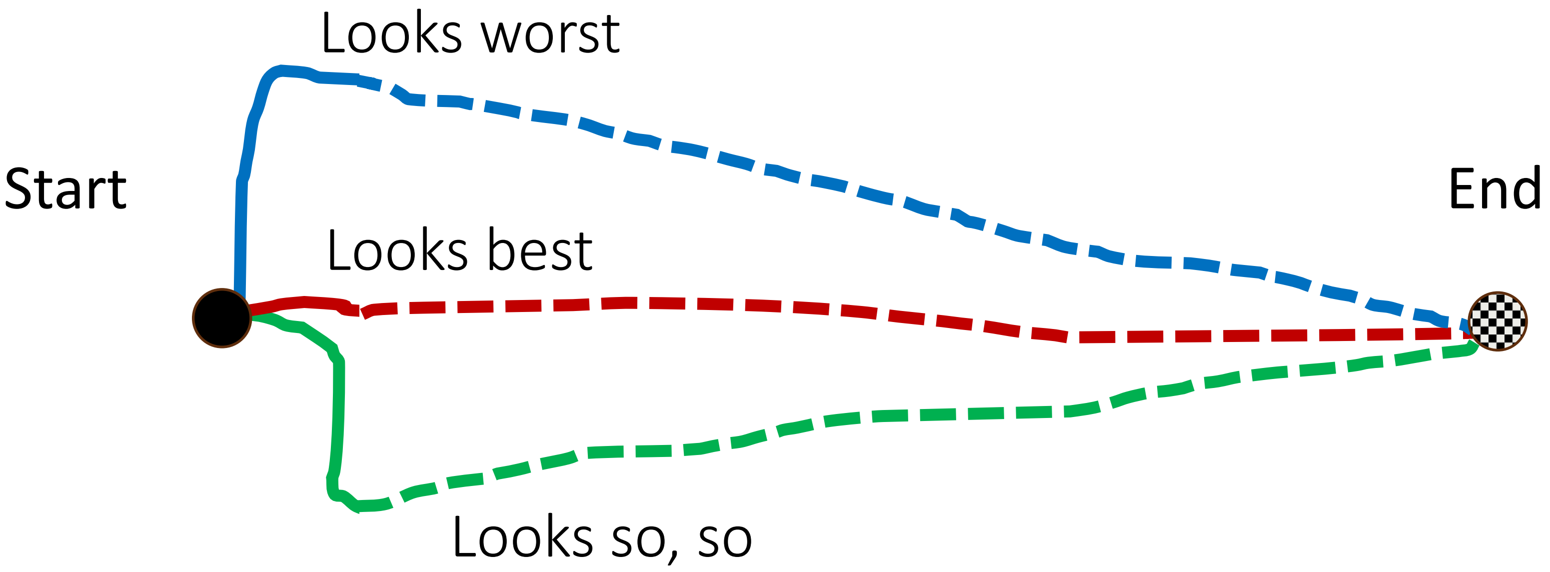
Decision-Making for Process Optimization

Decision-making based on descriptive data: insights

Decision-Making for Process Optimization

Looks best

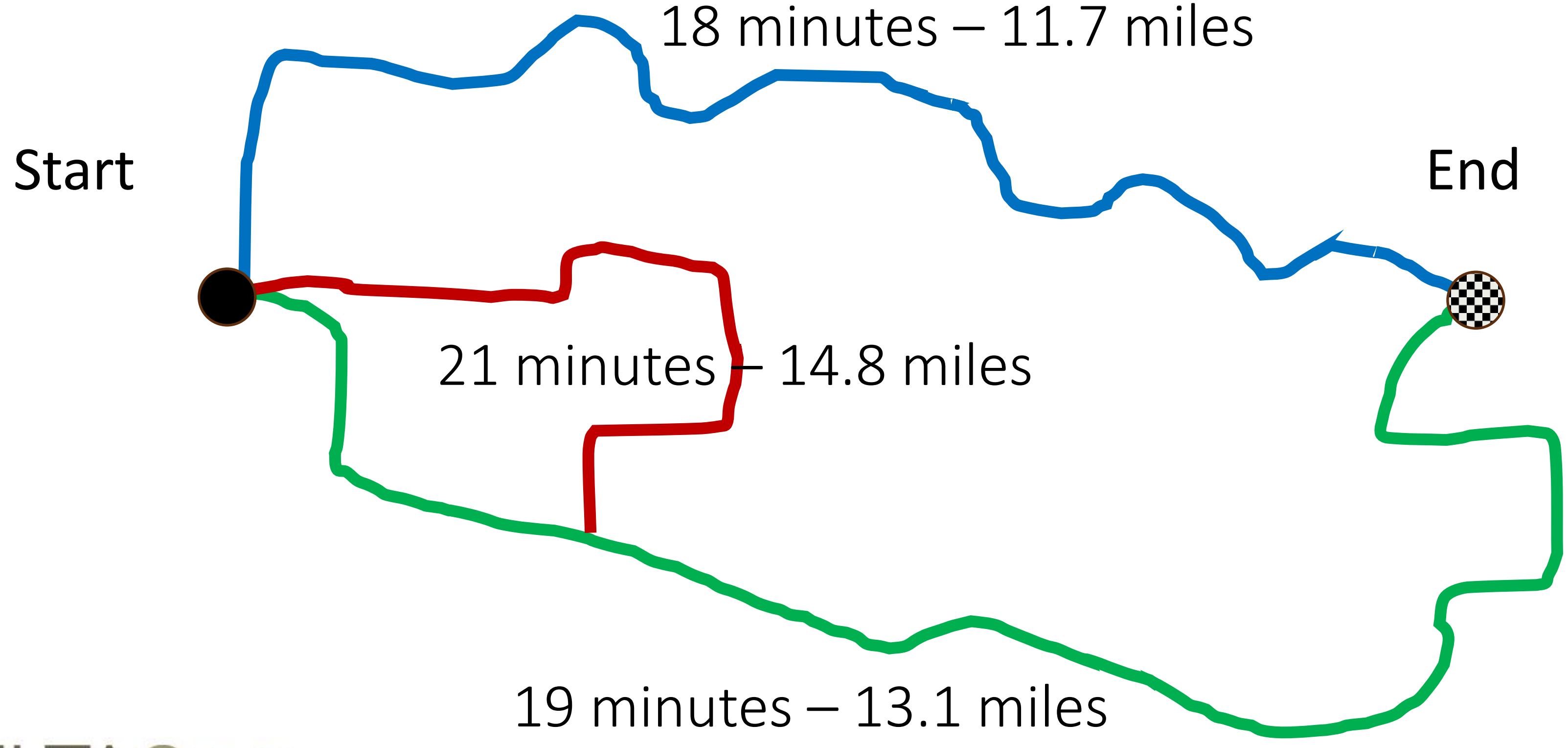
Decision-making based on descriptive data: insights



DELTAGee
Improving crops through better decisions

Decision-Making for Process Optimization

Forward-looking decision-making: best path



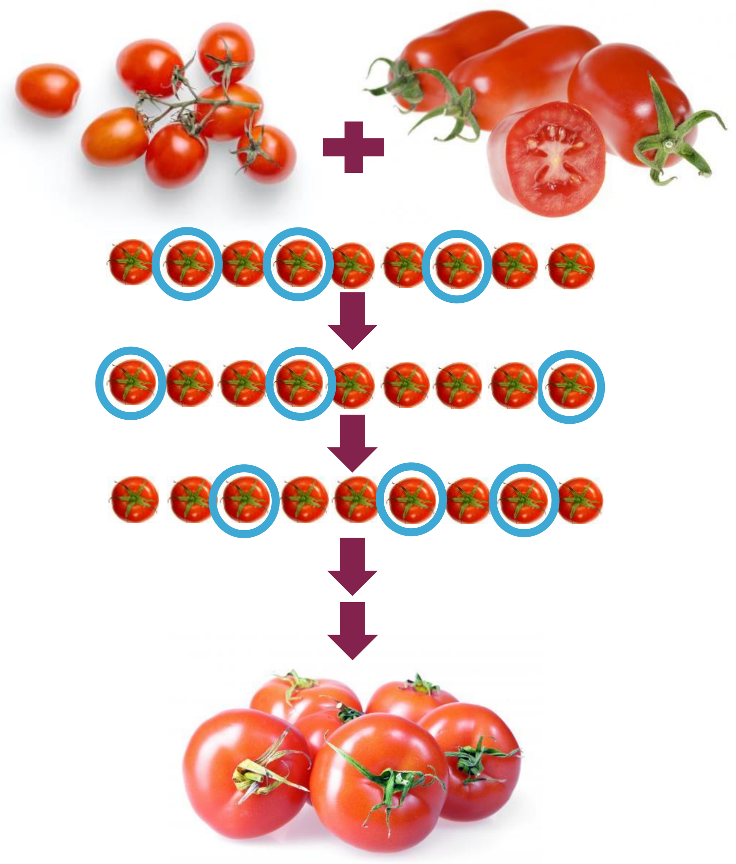
Decision-Making for Process Optimization

Forward-looking decision-making: shortest path

BACKGROUND

Plant Breeding is a Complex Process

With tens to hundreds of plants at each step (generations), there are **millions of potential "itineraries"** to go from starting point to end point

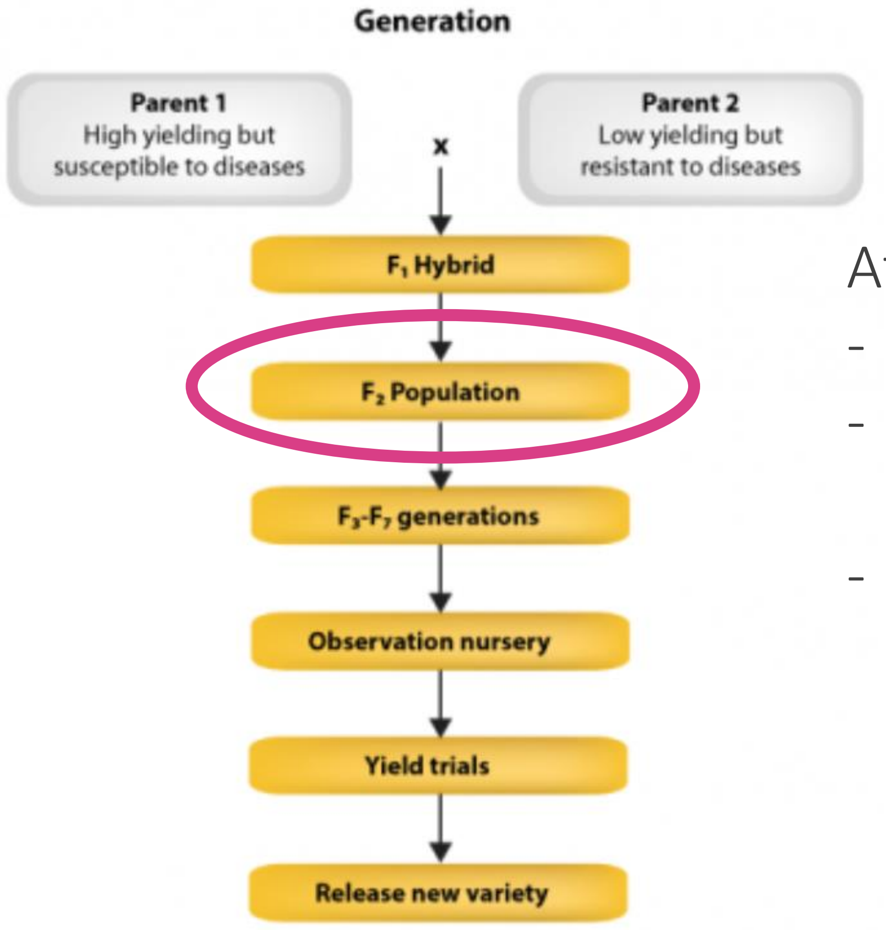


Starting Point:
Germplasm

End Point:
Variety Release

DELTAGee
Improving crops through **better decisions**

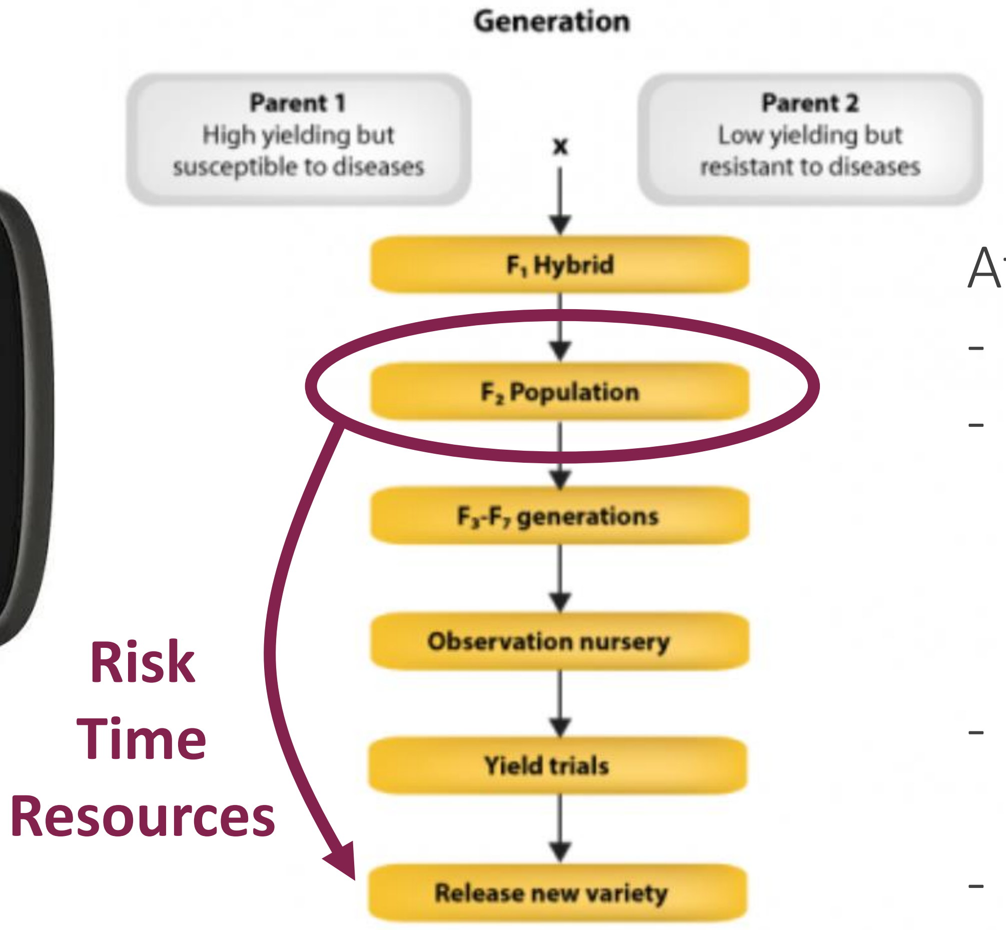
Current: Descriptive Data-Driven



At each step:

- Collect data
- **Select plants with best data**
- Insights

Future: Forward-Looking Data and Process-Driven



At each step:

- Collect data
- Calculate plants potential to deliver desired variety
- **Select plants with best potential**
- Shortest path

Future: Forward-Looking Data and Process-Driven

- Increases breeding efficiency very significantly:
  - Risk management
  - Timeline
  - Resources
- Is permitted by the availability of genotypic data on all selection candidates
- Can be implemented whenever a product objective can be defined by one or a set of genotypes (TI, MAS, GS)

Can We Improve Decision-Making in Plant Breeding?

1994

"Since the approximate map positions of the QTLs are known, it is possible to identify those individuals which, upon further crossing are most likely to produce the ideal genotype… This approach is a way to breed the most transgressive segregant, applying Mendelian genetics rather than phenotypic selection…" (Piet Stam) *

2003

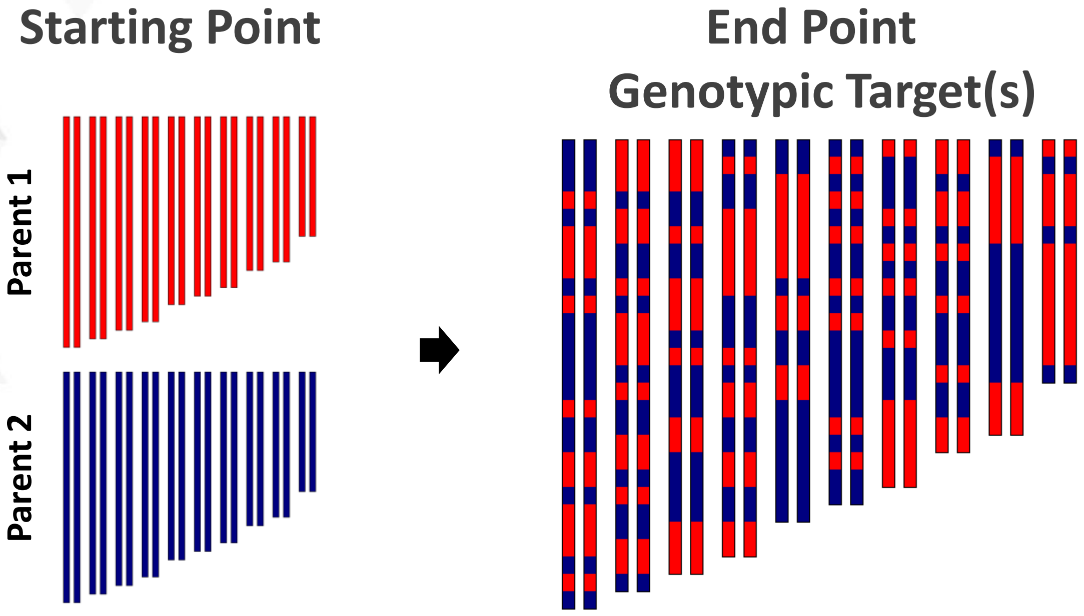
"Eventually, knowledge of the map positions of all loci of agronomic interest, the allelic variation at these loci, and their contribution to the phenotype should enable the breeder to design superior genotypes comprising a combination of favorable alleles at all loci… Software tools should enable us to determine the optimal route for generating those mosaic genotypes…" (Johan Peleman) **

2021

"Can machine learning be used to mimic the selection decisions made by breeders?... The purpose here would not be to replace breeders with a machine and an algorithm but rather to aid breeders in their work" (Rex Bernardo) ***

* P Stam (1994) Marker-assisted breeding. Eucarpia Biometrics
** JD Peleman and JR Van der Voort (2003) Breeding by design. Trends in plant science 8.7: 330-334.
*** R Bernardo (2021) Predictive breeding in maize during the last 90 years. Crop Science. 61: 2872– 2881. https://doi.org/10.1002/csc2.20529

Forward-Looking Decision-Making



Optimized decision making means:

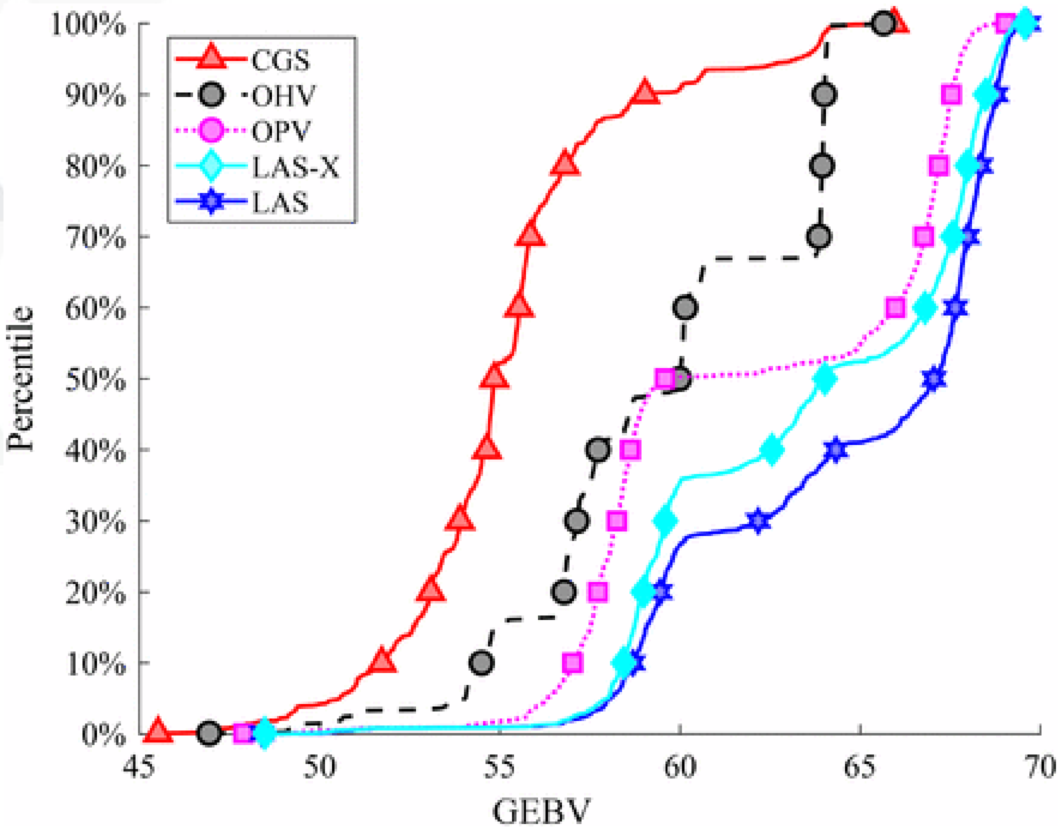
- Finding the “shortest” path from beginning to end
- Shortest in terms of time, resources, risk

Requires:

- Whole process approach
- Ability to predict future
- Ability to take operational elements into consideration
- Ability to adjust in real time

FORWARD-LOOKING DECISION-MAKING

Line Breeding



Cumulated distribution functions of population maximum

CGS: Conventional Genomic Selection – Selection of individuals with the highest GEBV - DESCRIPTIVE

OHV: Optimal Haploid Value – Selection of individuals based on GEBV of best-possible DH progeny – PREDICTIVE 1 GENERATION

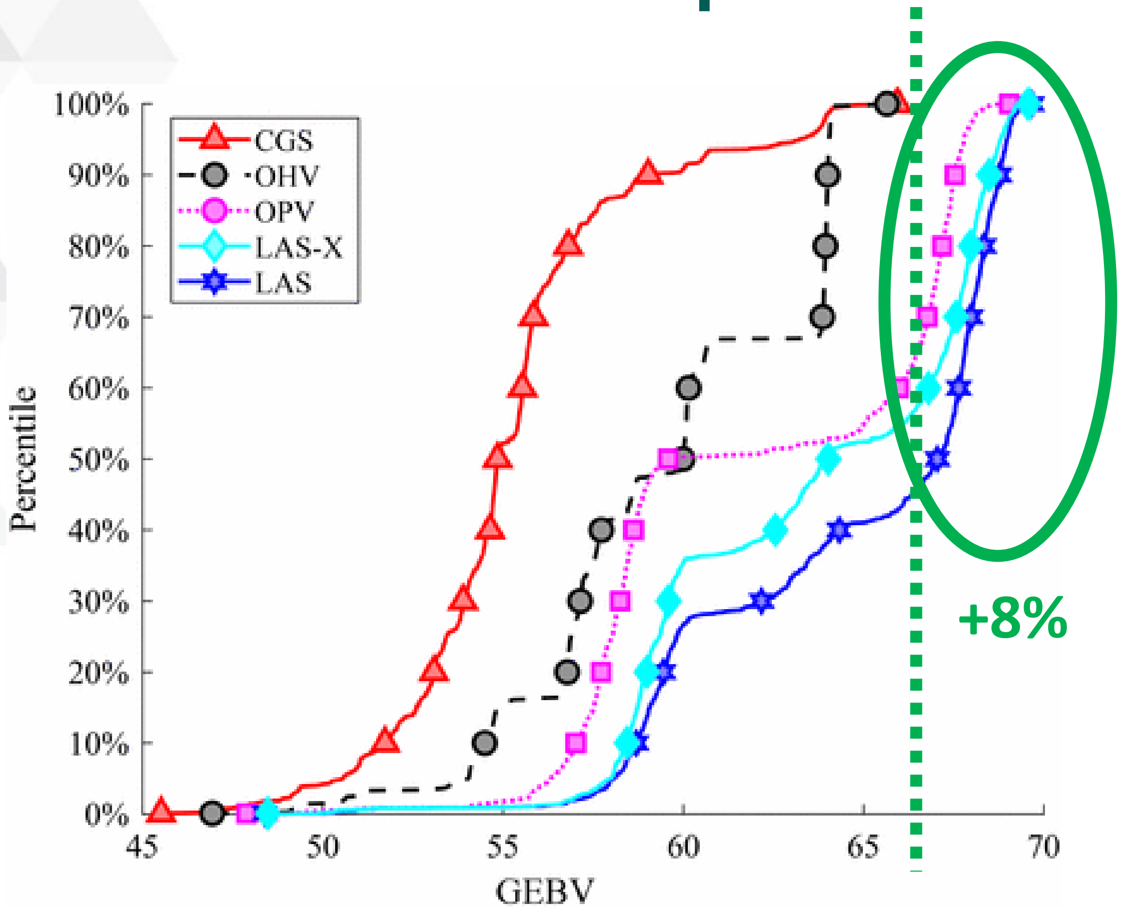
OPV: Optimal Population Value – Selection of set of individuals based on GEBV of best-possible progeny – PREDICTIVE INFINITE GENERATIONS

LAS: Look-Ahead Selection – Selection of set on individuals based on expected GEBV of best offspring with constraints – PREDICTIVE TERMINAL GENERATION

S Moeinizade et al. (2019) Optimizing selection and mating in genomic selection with a look-ahead approach: An operations research framework. G3 Genes|Genomes|Genetics, 9(7): 2123–2133. https://doi.org/10.1534/g3.118.200842

DELTAGee
Improving crops through better decisions

New Line Development



Cumulated distribution functions of population maximum

CGS: Conventional Genomic Selection – Selection of individuals with the highest GEBV - DESCRIPTIVE

OHV: Optimal Haploid Value – Selection of individuals based on GEBV of best-possible DH progeny – PREDICTIVE 1 GENERATION

OPV: Optimal Population Value – Selection of set of individuals based on GEBV of best-possible progeny – PREDICTIVE INFINITE GENERATIONS

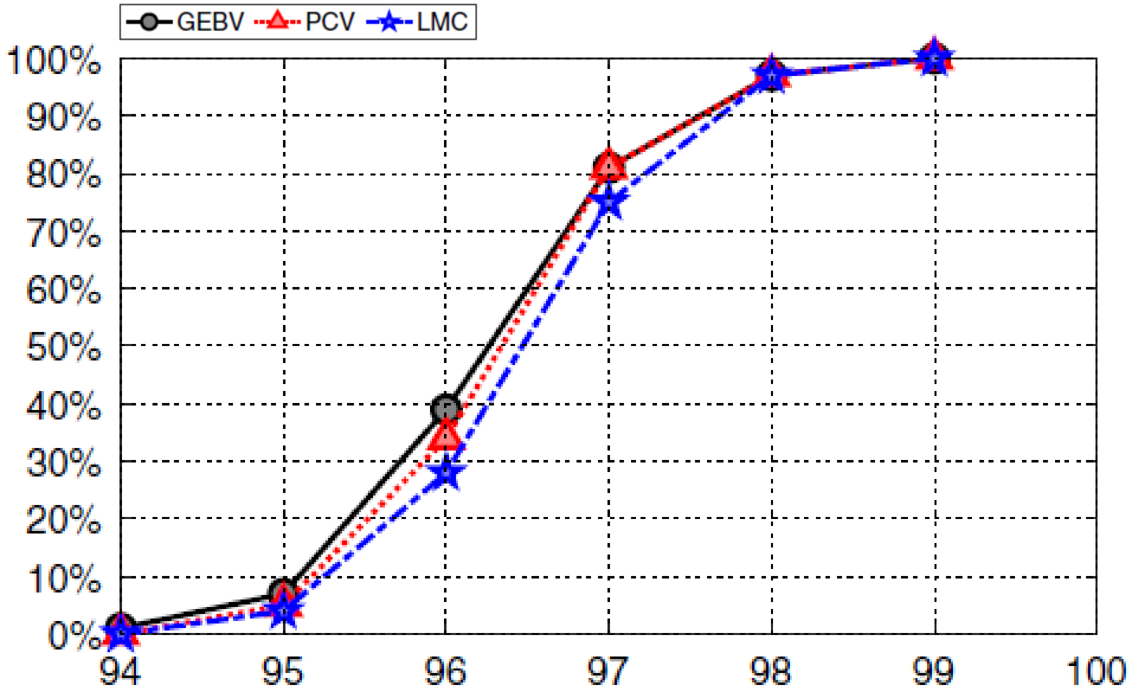
LAS: Look-Ahead Selection – Selection of set on individuals based on expected GEBV of best offspring with constraints – PREDICTIVE TERMINAL GENERATION

S Moeinizade et al. (2019) Optimizing selection and mating in genomic selection with a look-ahead approach: An operations research framework. G3 Genes|Genomes|Genetics, 9(7): 2123–2133. https://doi.org/10.1534/g3.118.200842

Forward-looking decision-making delivers higher performance levels than other approaches

DELTAGee
Improving crops through **better decisions**

Trait Introgression



Cumulated distribution functions of population maximum in BC3

GEBV: Background Selection – Selection of individuals with the highest recurrent parent genome recovery - DESCRIPTIVE

PCV: Predicted Cross Value – Selection of individuals based their probability of production of the "perfect" gamete – PREDICTIVE 1 GENERATION

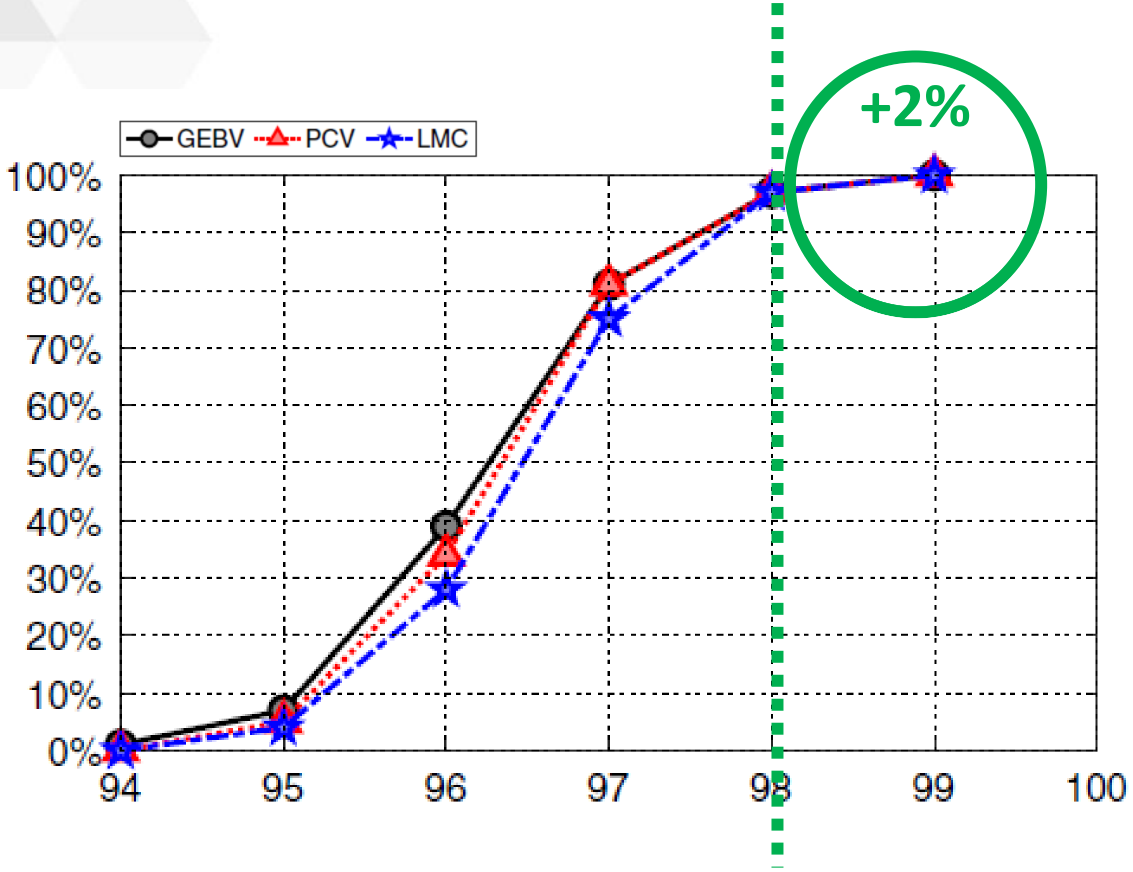
LMC: Look-Ahead Monte Carlo – Selection of individuals based on their predicted genetic distribution of BC progeny after multiple generations with constraints – PREDICTIVE MULTIPLE GENERATIONS

S Moeinizade et al. (2021) A look-ahead Monte Carlo simulation method for improving parental selection in trait introgression. Sci Rep 11, 3918.

DELTAGee
Improving crops through **better decisions**

FORWARD-LOOKING DECISION-MAKING

Trait Introgression



Cumulated distribution functions of population maximum in BC3

GEBV: Background Selection – Selection of individuals with the highest recurrent parent genome recovery - DESCRIPTIVE

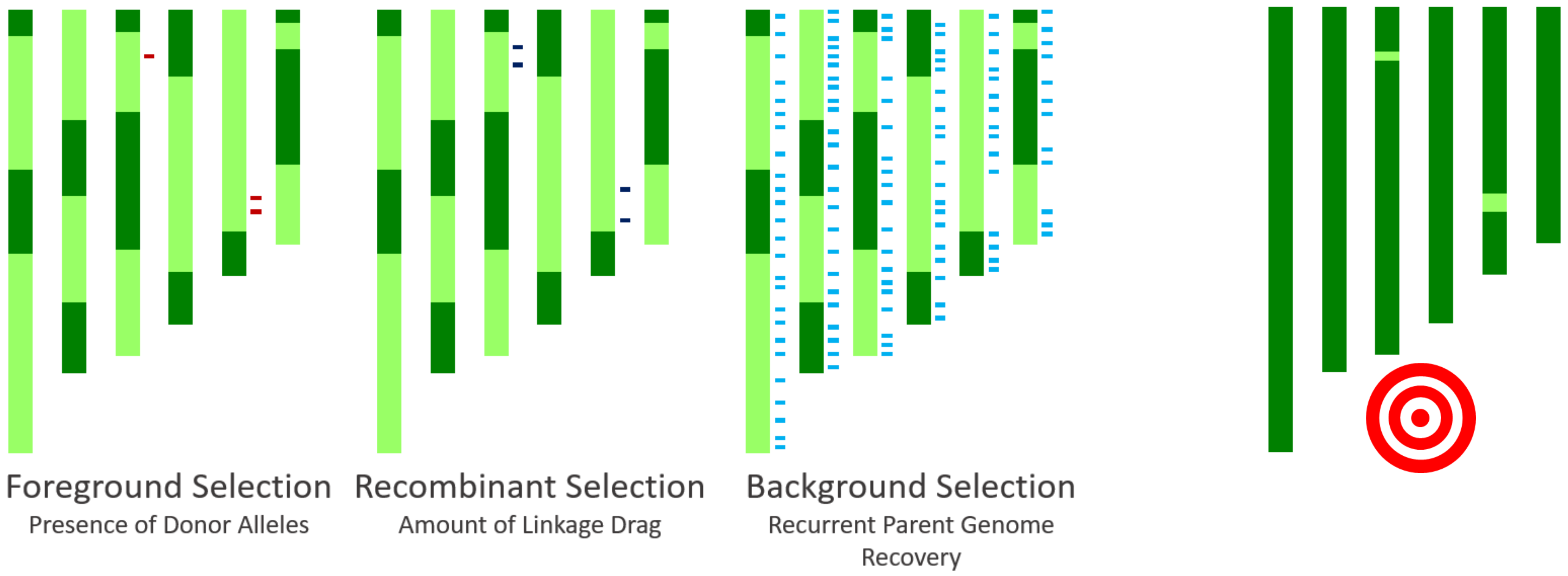
PCV: Predicted Cross Value – Selection of individuals based their probability of production of the "perfect" gamete – PREDICTIVE 1 GENERATION

LMC: Look-Ahead Monte Carlo – Selection of individuals based on their predicted genetic distribution of BC progeny after multiple generations with constraints – PREDICTIVE MULTIPLE GENERATIONS

S Moeinizade et al. (2021) A look-ahead Monte Carlo simulation method for improving parental selection in trait introgression. Sci Rep 11, 3918.

Forward-looking decision-making results in cleaner/faster line conversions

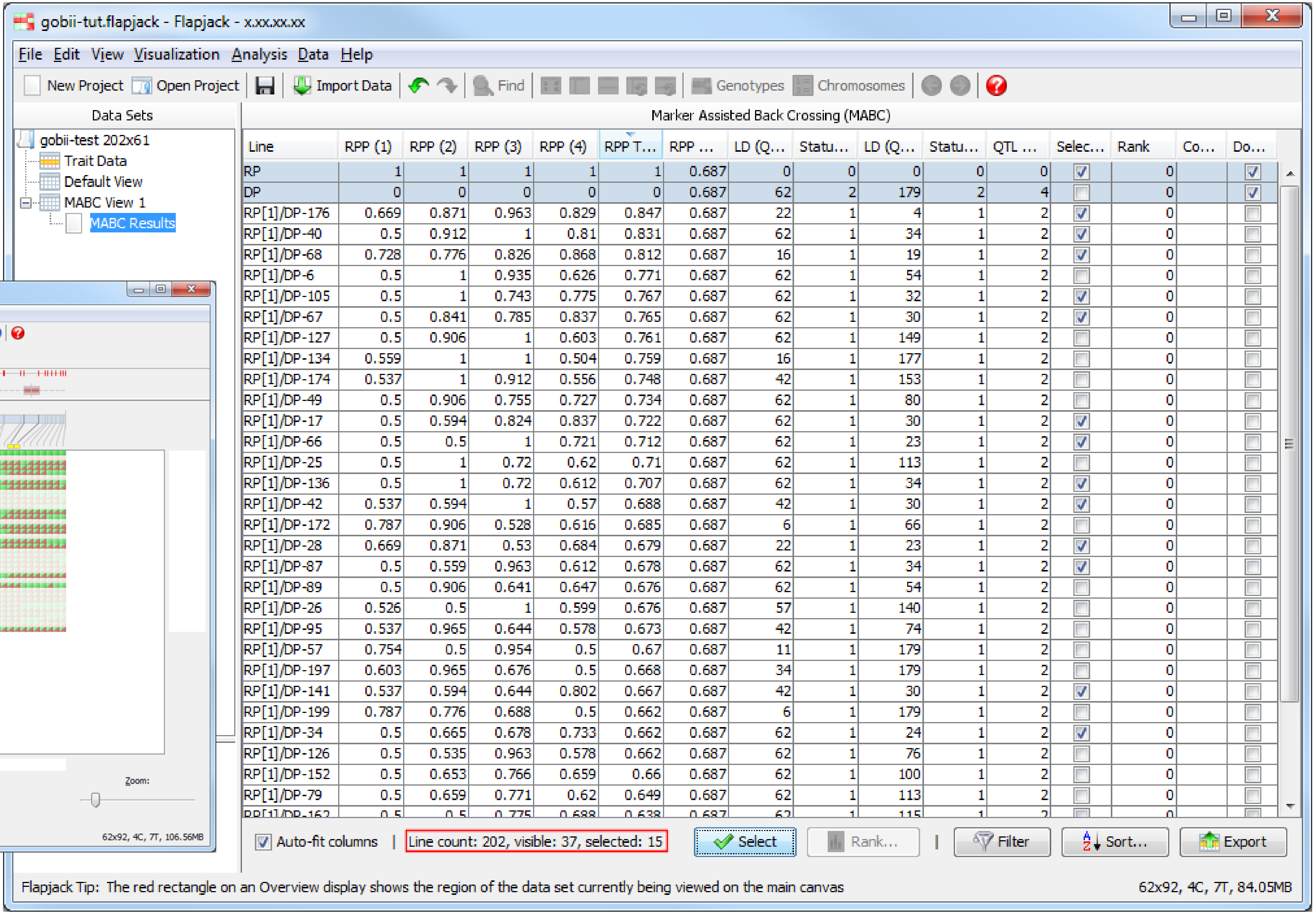
General Trait Introgression

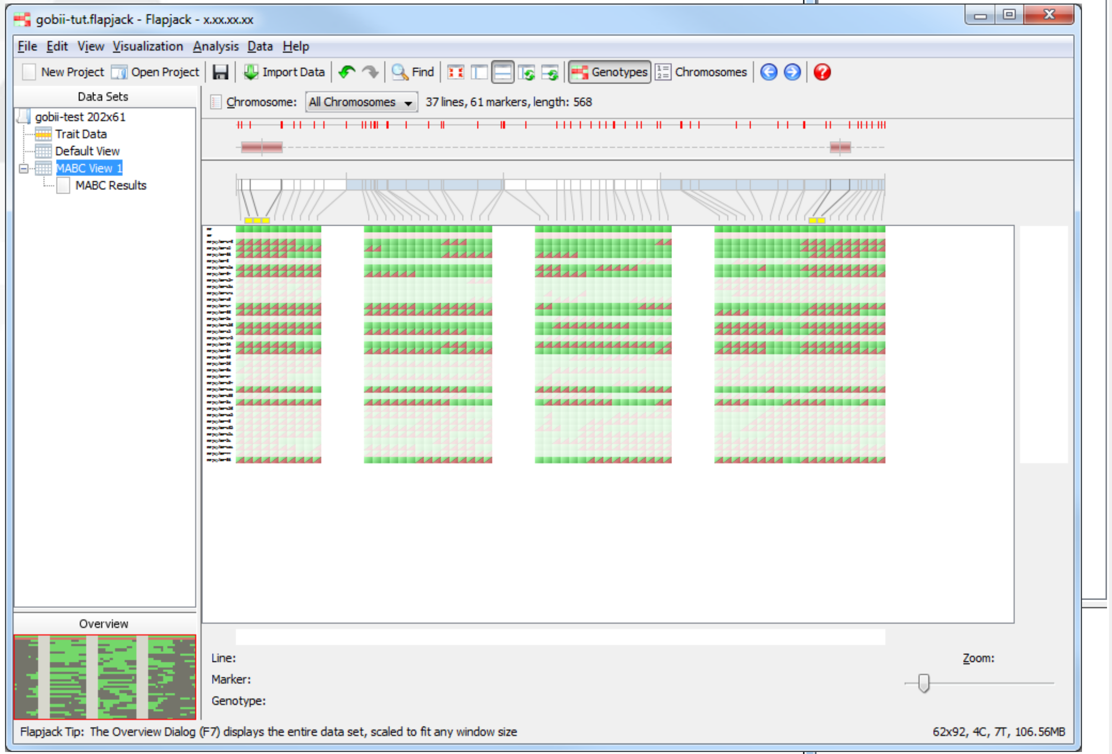


- **Difficult and time-consuming compromise** between recombination and background
- Selection of plants based on their own, **“per se” attributes**
- **No insights**/visibility into how to proceed with selected plants

TRAIT INTROGRESSION

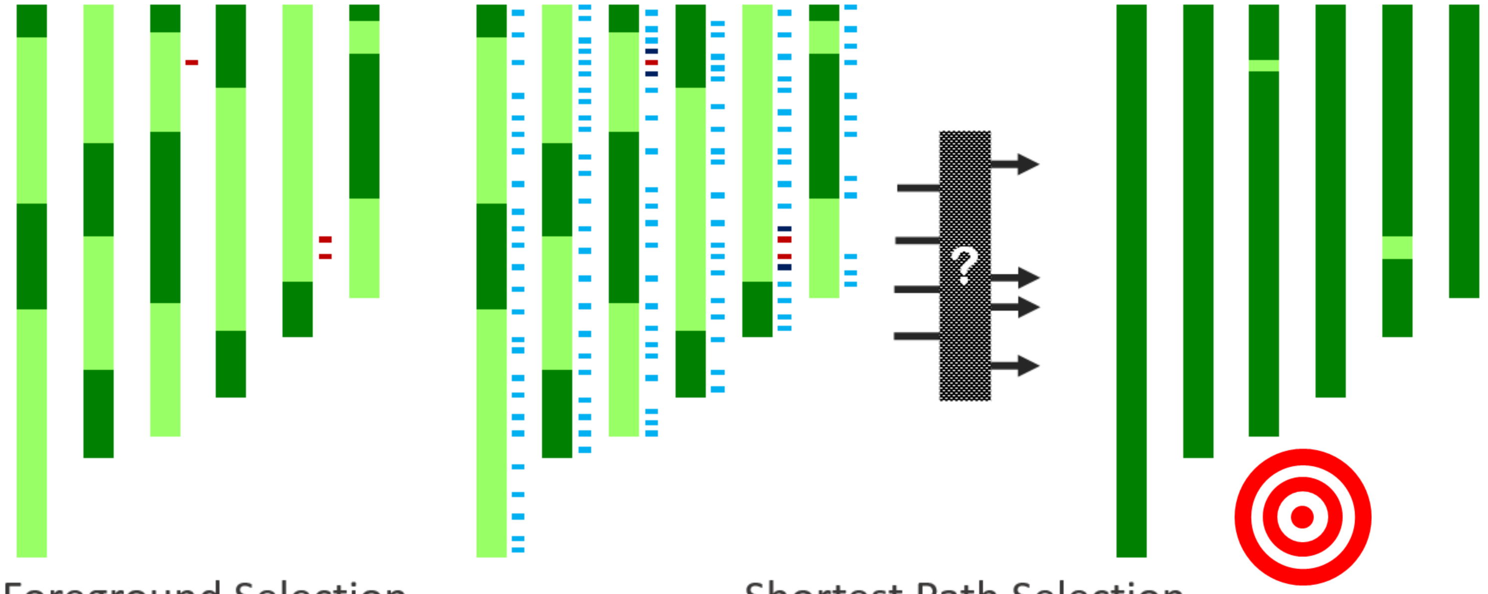
General Trait Introgression Software





I Milne et al. (2010) Flapjack – graphical genotype visualization. Bioinformatics 26(24), 3133-3134.

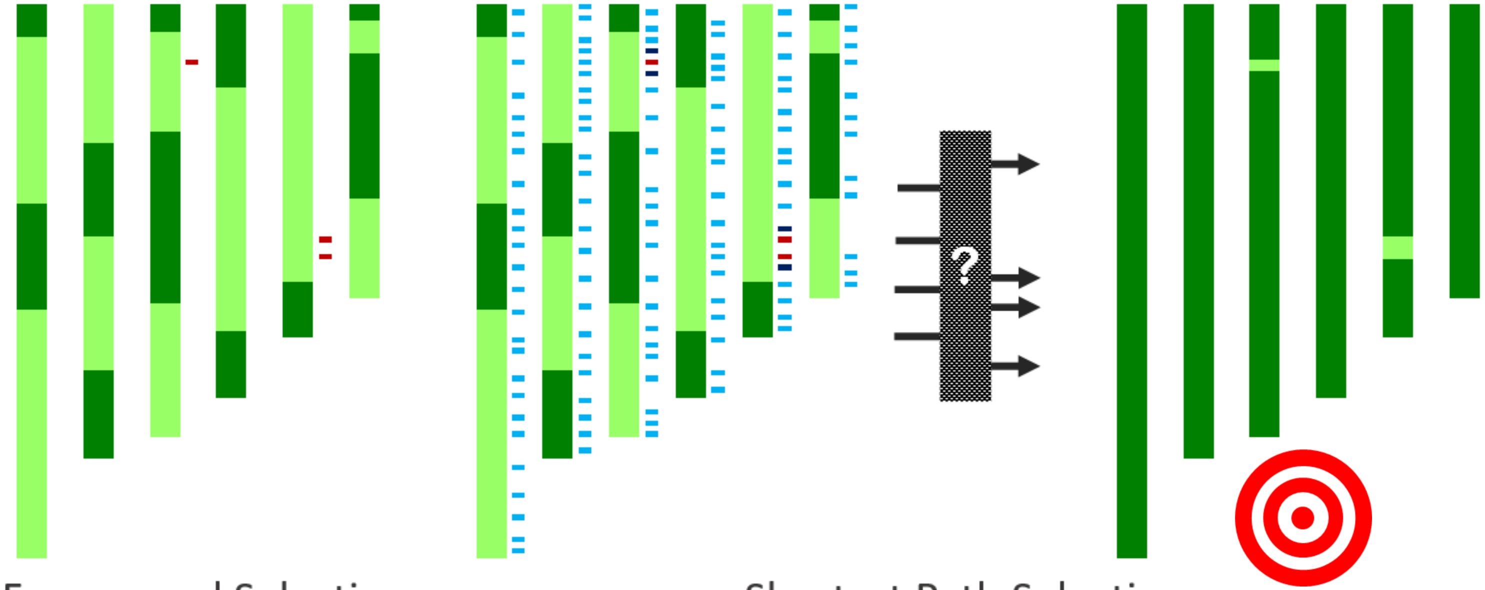
Optimized Trait Introgression



Foreground Selection
Presence of Donor Alleles

Shortest Path Selection
Highest Probability to Deliver Target Genotype

Optimized Trait Introgression

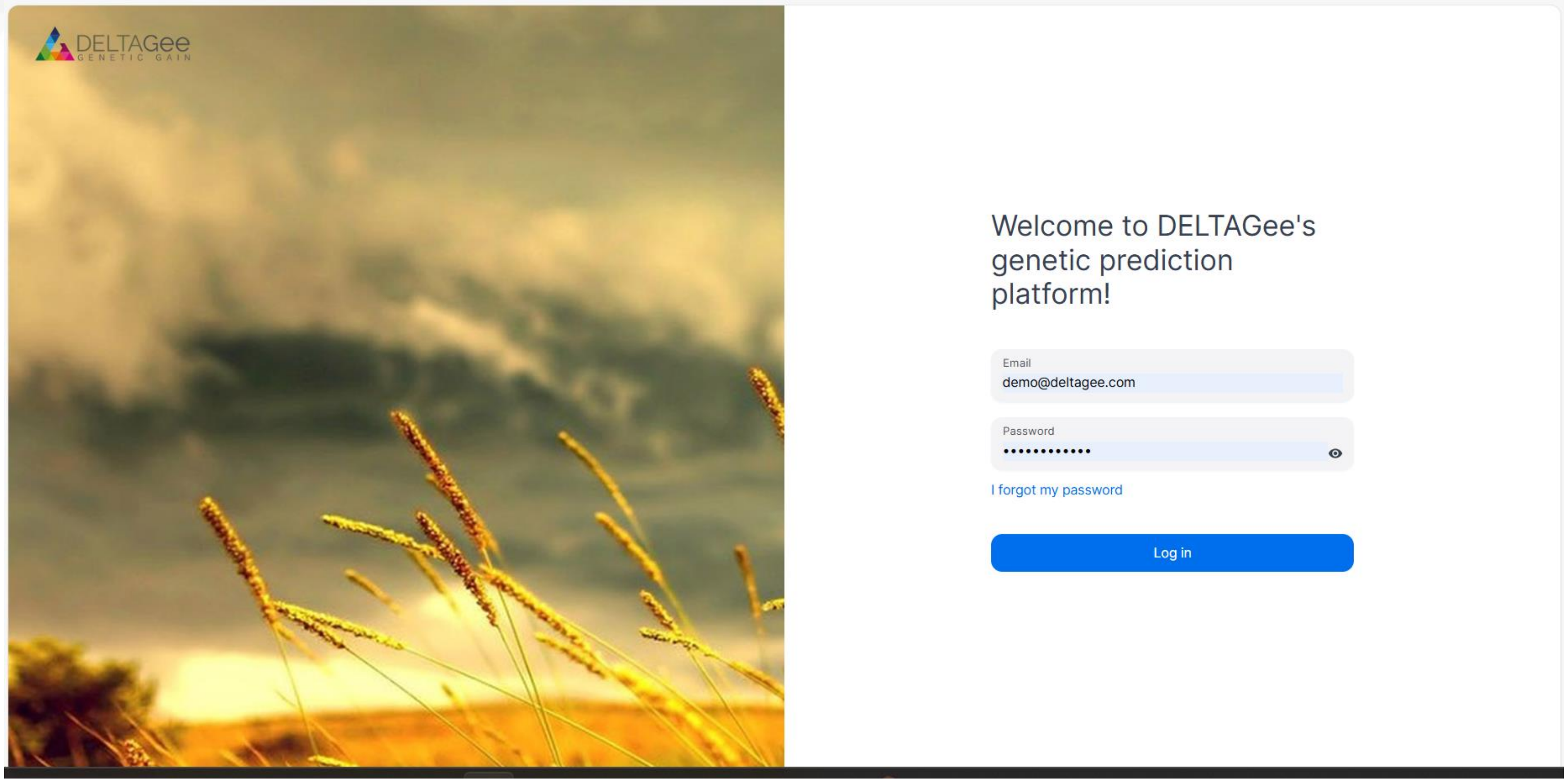


Foreground Selection
Presence of Donor Alleles

Shortest Path Selection
Highest Probability to Deliver Target Genotype

- **No compromise** between recombination and background
- Selection of plants based on their **future potential to deliver** the target genotype
- Ability to re/define upcoming steps (population sizes) to **secure success**

Optimized Trait Introgression - On-Demand Services



Optimized Trait Introgression - On-Demand Services

PLANT ID	RISK AT BC1S1 Faster alternative	RISK AT BC2S1 Faster alternative	RISK AT BC3S1 User defined protocol	RPGR	Total LD (%)	Total LD (cM)	Nb recomb (/ max nb rec)	Nb missing DP	Nb het mkr	Nb het segments	Nb homoz REC mkr	Nb homoz DON mkr	
ProjXt3_Plt856	100.00%	98.08%	0.21%	82.42%	6.43%	75.42	2	0	37	16	80	0	See more
ProjXt3_Plt55	100.00%	99.26%	0.42%	78.31%	1.72%	20.12	2	0	55	16	62	0	See more
ProjXt3_Plt222	100.00%	99.27%	1.36%	82.30%	4.05%	47.445	1	0	33	13	84	0	See more
ProjXt3_Plt119	100.00%	99.95%	1.93%	76.88%	3.59%	42.07	2	0	50	21	67	0	See more
ProjXt3_Plt486	100.00%	99.31%	2.13%	81.75%	4.05%	47.445	1	0	43	15	74	0	See more
ProjXt3_Plt425	100.00%	99.35%	2.17%	81.10%	4.05%	47.445	1	0	36	14	81	0	See more
ProjXt3_Plt367	100.00%	99.87%	2.17%	77.37%	3.30%	38.64	2	0	55	21	62	0	See more
ProjXt3_Plt192	100.00%	99.81%	2.94%	82.16%	7.72%	90.47	1	0	43	13	74	0	See more
ProjXt3_Plt227	100.00%	99.64%	3.15%	80.34%	6.89%	80.795	1	0	38	16	79	0	See more
ProjXt3_Plt476	100.00%	99.80%	3.36%	81.19%	10.21%	119.675	1	0	46	15	71	0	See more
ProjXt3_Plt900	100.00%	99.79%	3.39%	80.14%	4.05%	47.445	1	0	40	14	77	0	See more
ProjXt3_Plt97	100.00%	99.97%	5.39%	78.74%	6.89%	80.795	1	0	47	15	70	0	See more
ProjXt3_Plt379	100.00%	99.98%	5.87%	75.22%	1.98%	23.205	2	0	55	22	62	0	See more

‹ 1 2 3 4 5 ··· 57 ›

Optimized Trait Introgression - On-Demand Services

Pre-selected scenarios

User defined protocol

Population sizes: 1125 in BC2, 800 in BC3, 180 in BC3S1

Risk at BC3S1: 0.80%

Plants to select	BC2 progeny size to develop
ProjXt3_Plt856	300
ProjXt3_Plt55	300
ProjXt3_Plt222	300
ProjXt3_Plt119	225

Alternative protocol

Population sizes: 600 in BC2, 460 in BC3, 180 in BC3S1

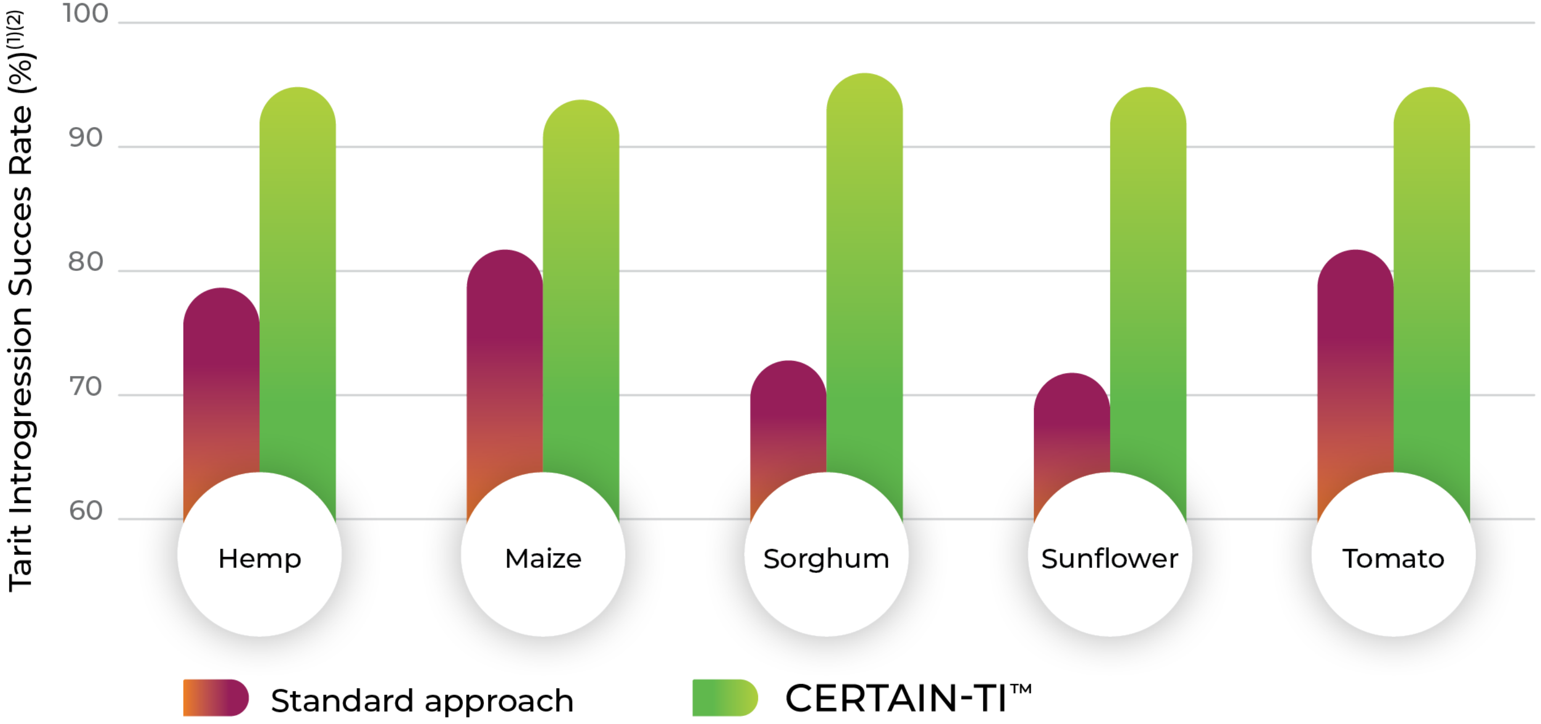
Risk at BC3S1: 4.91%

Plants to select	BC2 progeny size to develop
ProjXt3_Plt856	300
ProjXt3_Plt55	300

Revenue Creation through Operational Savings

Decision-Making is Key in Trait/Product Launch

Revenue Creation through Product Launch and Market Capture

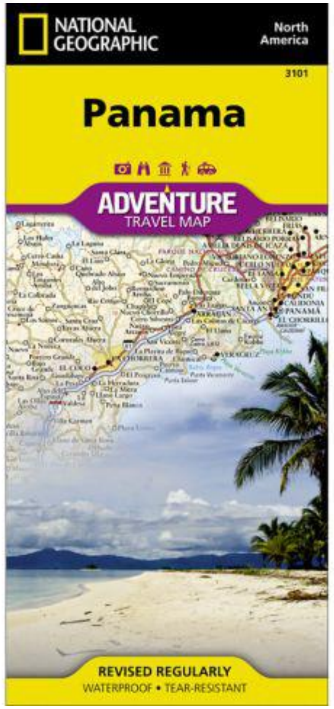


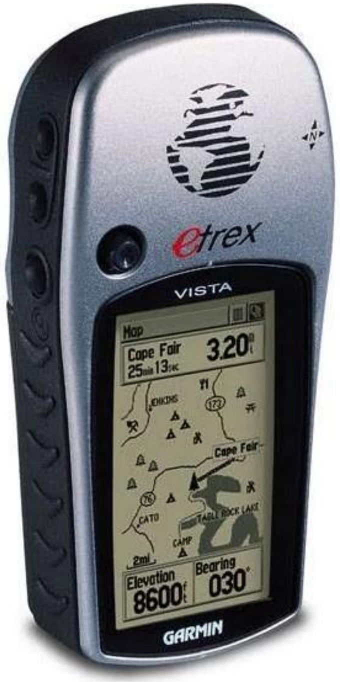
(1) Success rate is defined as the percentage of trait introgression projects that met both quality and time requirements

(2) Side-by-side comparisons with identical protocols within crop

DELTAGee
Improving crops through **better decisions**

"It's like having a Breeding Navigator"





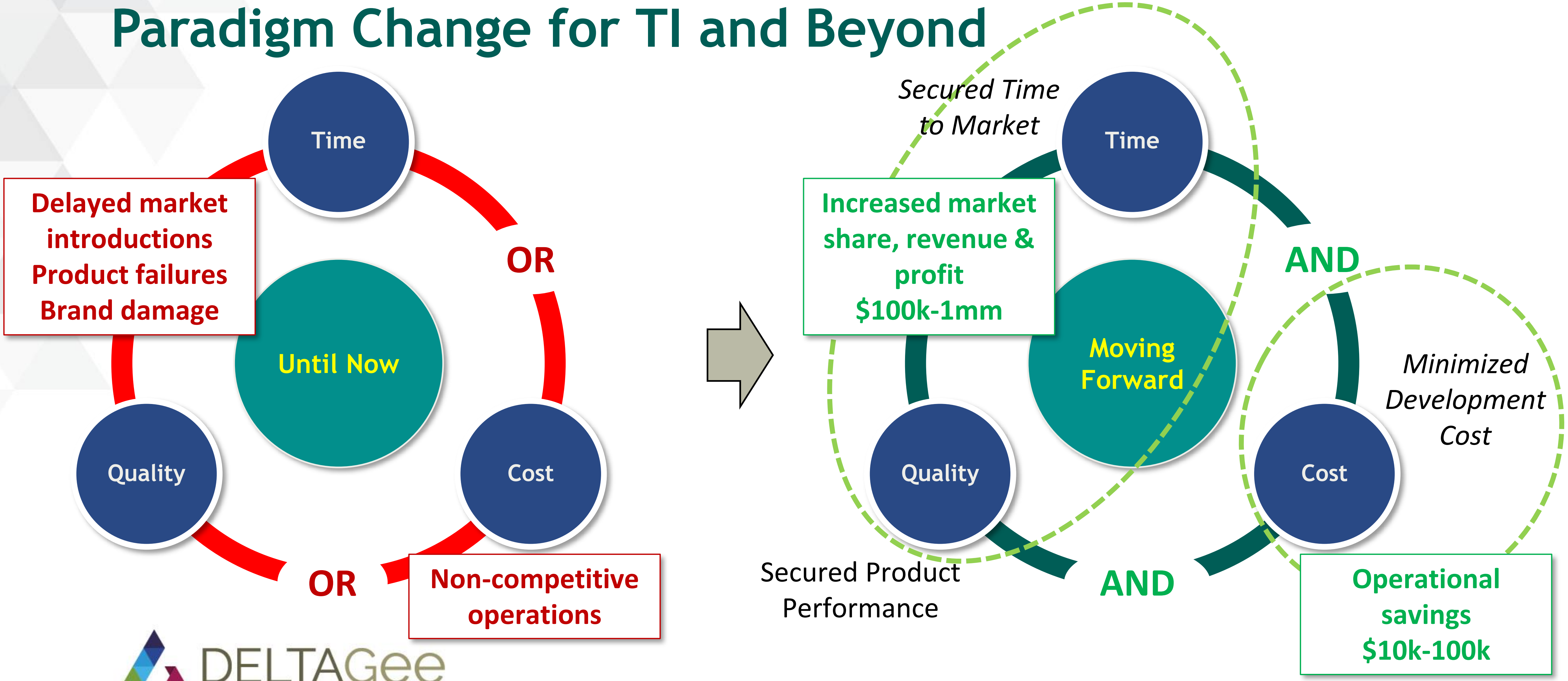
Data — — — **Data and Process**

From where we are... with high accuracy... to how to get where we want to go

Non-Mathematical/Computing Challenges to Forward-Looking/Data and Process-Driven Decision-Making

- Clear description of product objective
  - Product profile and breeding objectives
- Relevant marker-trait models
  - Diagnostic markers for simple traits (Trait Introgression)
  - Predictive models for complex traits (Genomic Selection)
- Access to genome-wide genotyping
  - Cost of useful data points
  - Data generation turn-around time

Paradigm Change for TI and Beyond



TAKE-HOME POINTS

Get the Most out of Your Data: Better Decisions

- In order to deliver what our societies need, breeding has to be product/objective-driven – Let's really make it that way; hoping is not enough
- There is more information in your data than you think – extract it all
- Making better decisions out of the same data is a very cheap way to improve breeding efficiency – the cost is in generating the data
- Who would want to get on a road trip without a navigation system?

Thank you…

Gracias…

DELTAGee
Improving crops through **better decisions**

deltagee.com