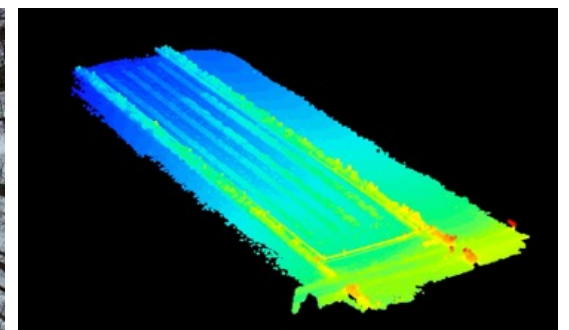
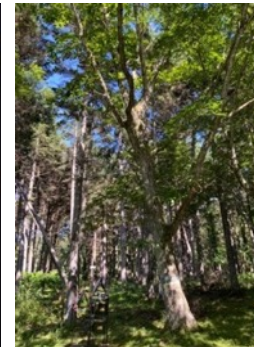




Application of Data Science to Accelerate Crop Improvement

Hiroyoshi Iwata

Grad Sch. Agricultural and Life Sciences, Lab. Biometry and Bioinformatics



Our laboratory

Graduate School of Agricultural
and Life Science, UTokyo

Department of Agricultural and
Environmental Biology

Lab. Biometry and
Bioinformatics

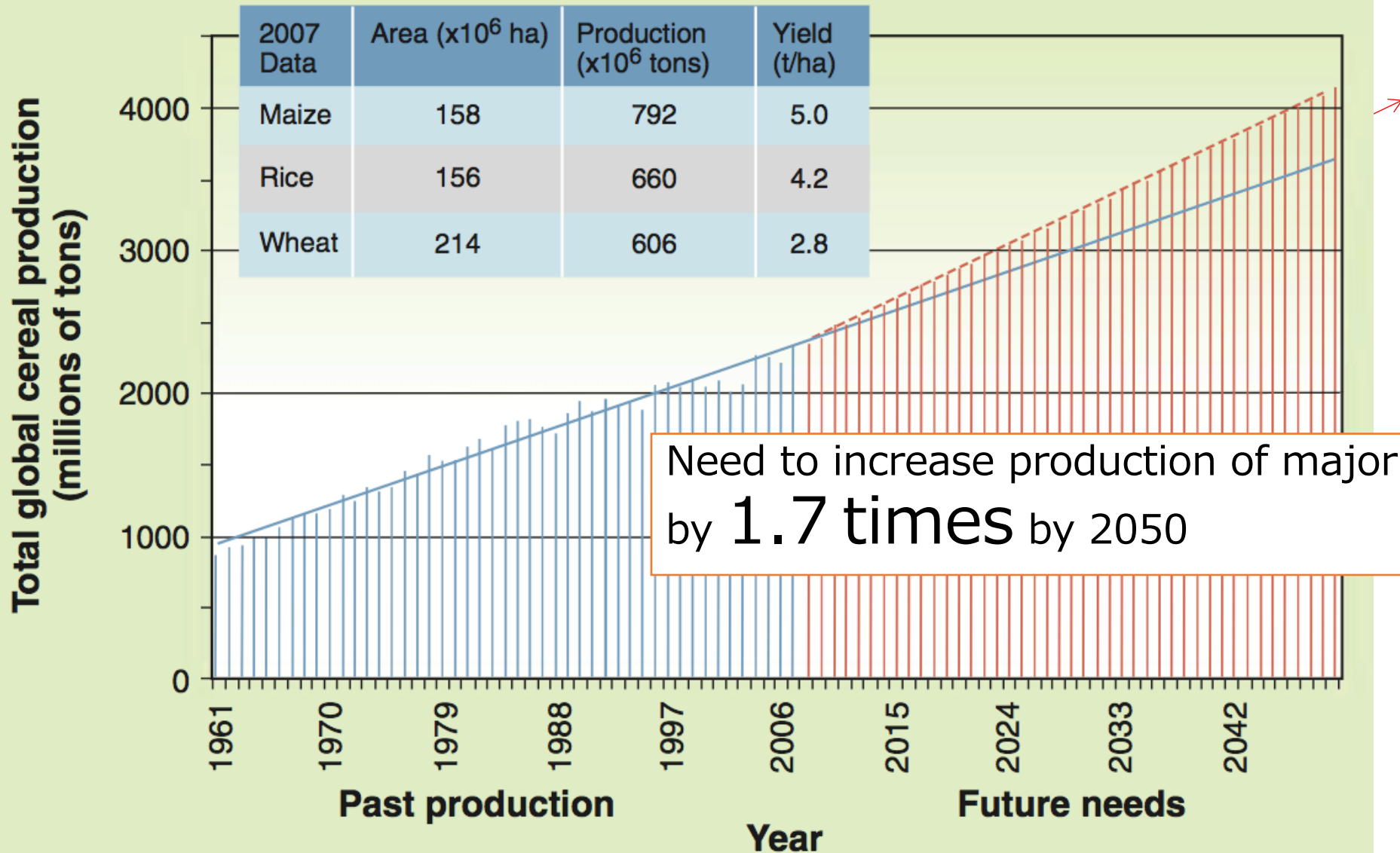
Special field:

Biometrics, Quantitative Genetics,
Statistical Genetics

Main objectives:

Improvement of efficiency in plan
genetic ability improvement (i.e.,
“plant breeding”)





Crop genetic improvement (breeding) must be accelerated!

Tester and Langridge (2010) Science 327: 818

Two big breakthroughs to accelerate breeding



High-throughput and low-cost DNA analysis



ARTICLE

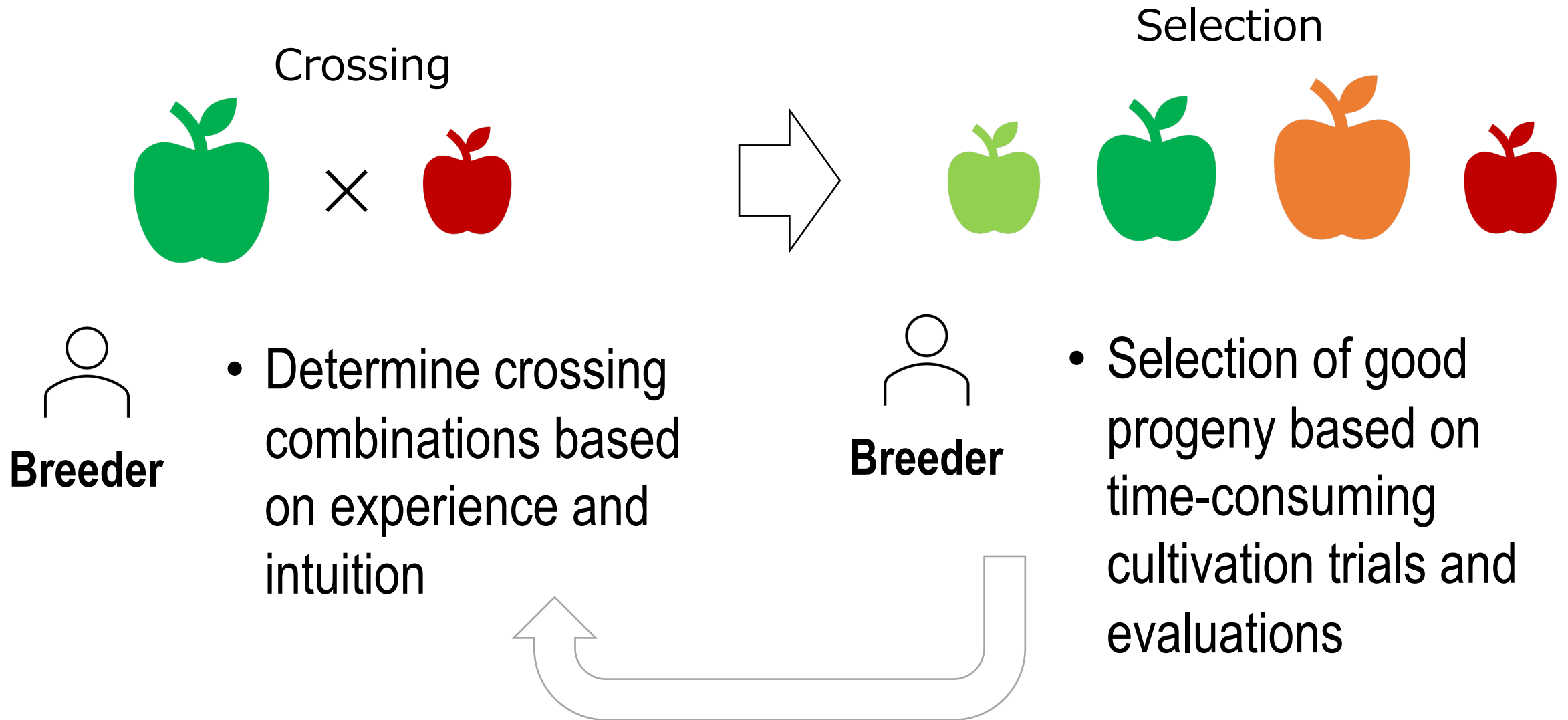
doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

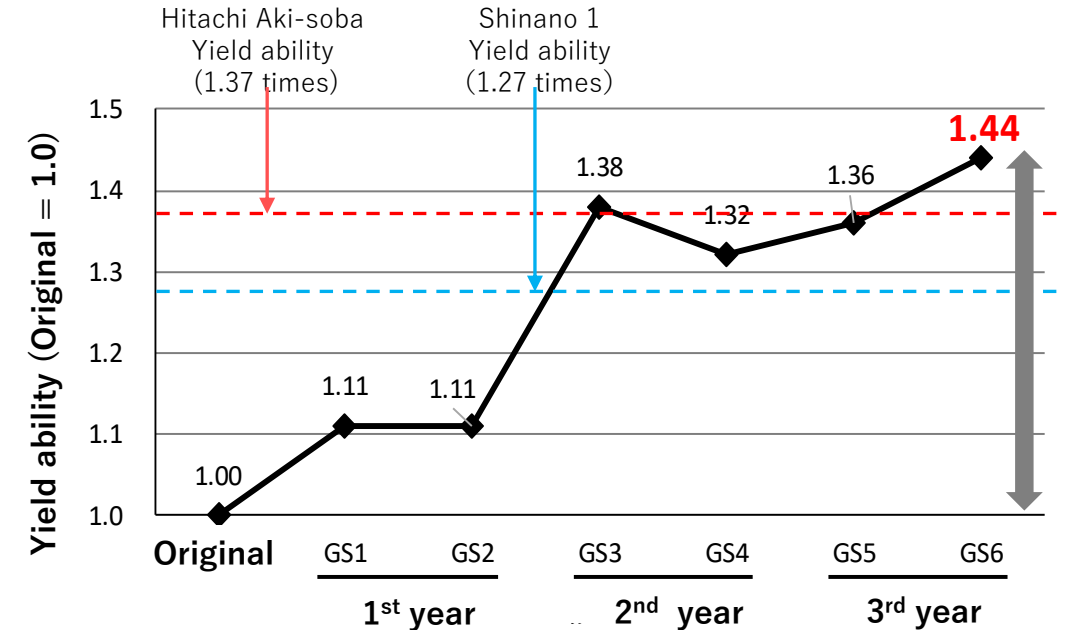
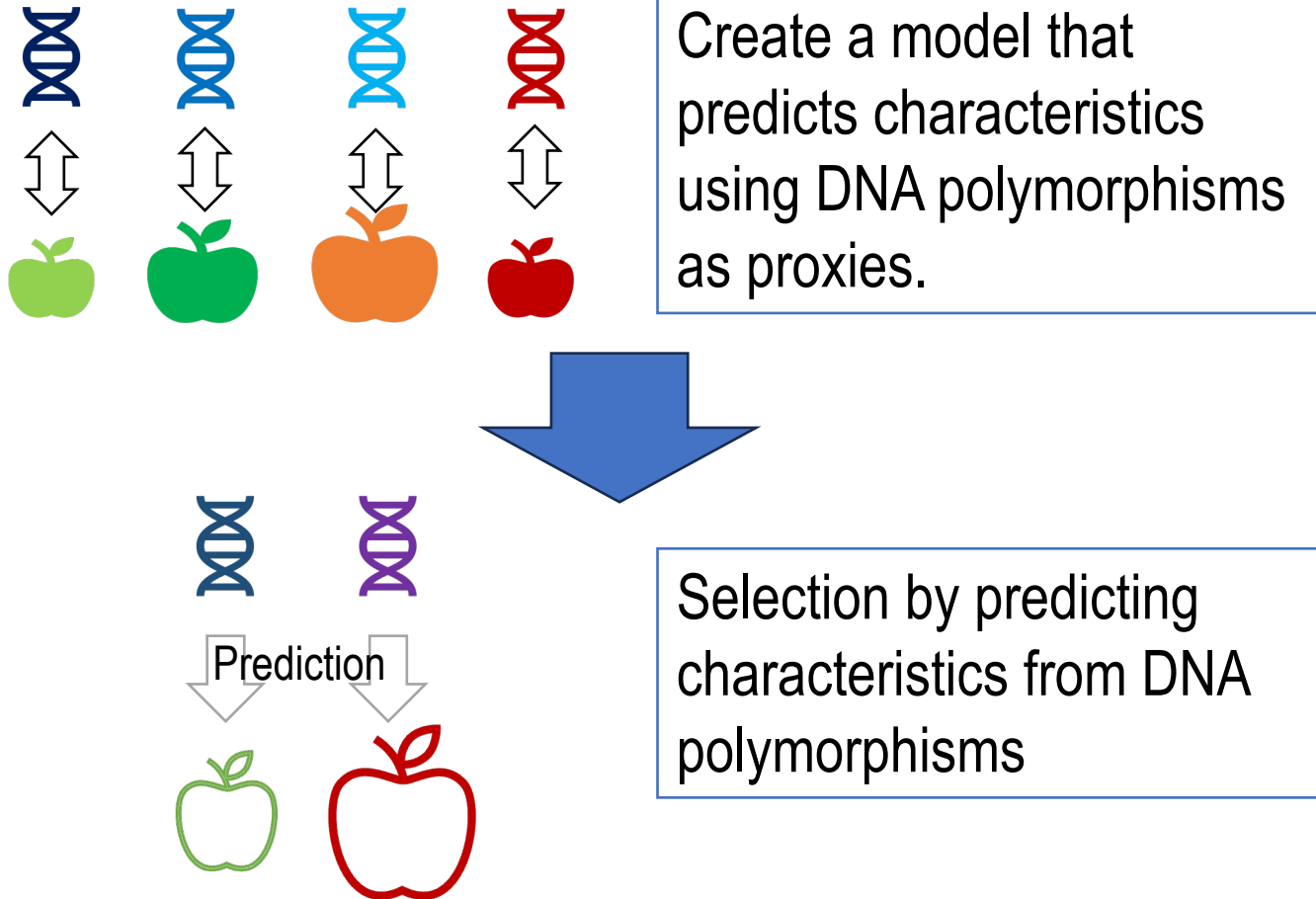
David Silver^{1*}, Aja Huang^{2*}, Chris J. Maddison³, Arthur Guez², Laurent Sifre¹, George van den Driessche¹, Julian Schrittwieser², Ioannis Antonoglou¹, Veda Panneershelvam¹, Marc Lanctot¹, Sander Dieleman¹, Dominik Grewe¹, John Nham², Nal Kalchbrenner¹, Ilya Sutskever², Timothy Lillicrap³, Madeleine Leach³, Koray Kavukcuoglu¹, Thore Graepel¹ & Demis Hassabis¹

Rapid technical advancement in data sciences

Conventional Plant Breeding



Plant Breeding with genomic selection (GS)

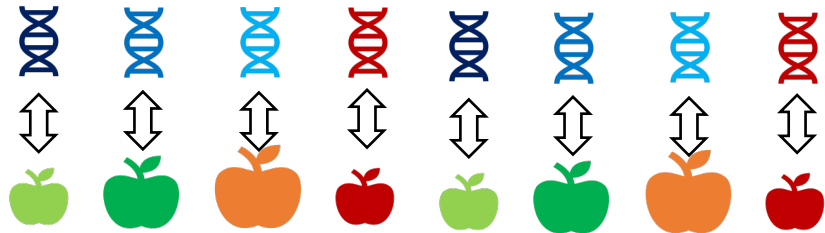


Genomic selection experiment in buckwheat (蕎麦)

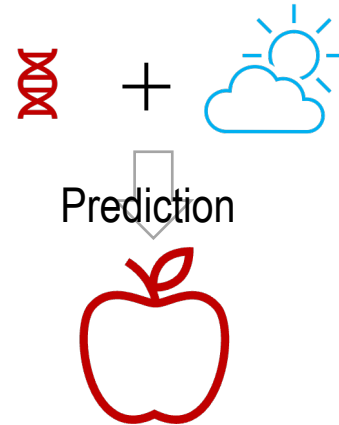
Yabe et al. (2018) Front. Plant Sci. 9: 276

- GS eliminates time-consuming cultivation tests and evaluations
- GS can speed up the genetic improvement

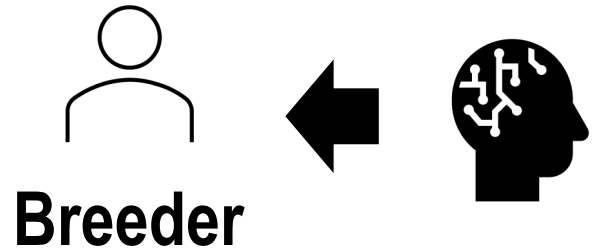
Challenges in increasing the efficiency of genomic selection (GS)



Collect a large number of data that can be used for model building (automatic and semi-automatic measurements)



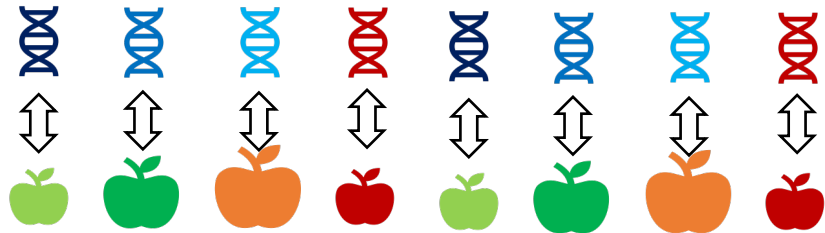
Make the model capable of proxying not only genomic but also environmental data



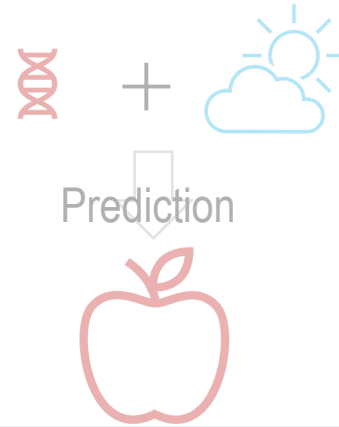
Making breeders' decisions data-driven, e.g., through data science optimization methods

We are currently working primarily on these topics.

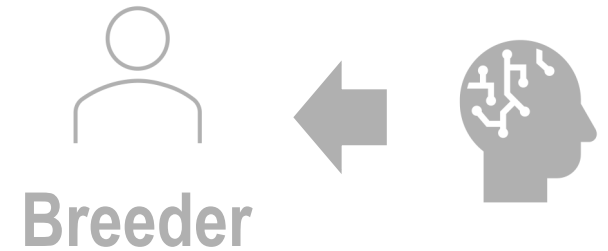
Challenges in increasing the efficiency of genomic selection (GS)



Collect a large number of data that can be used for model building (automatic and semi-automatic measurements)

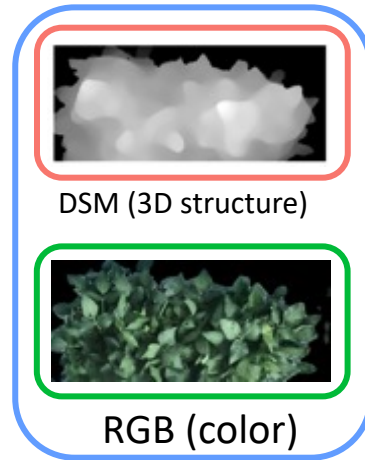


Make the model capable of proxying not only genomic but also environmental data

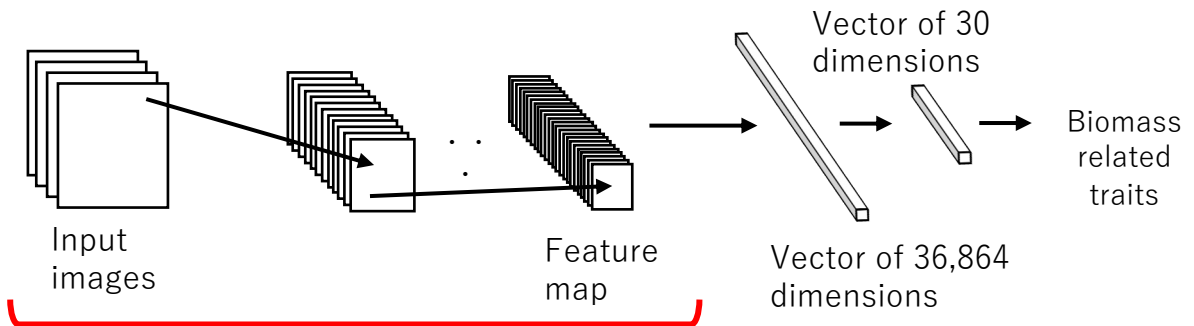


Making breeders' decisions data-driven, e.g., through data science optimization methods

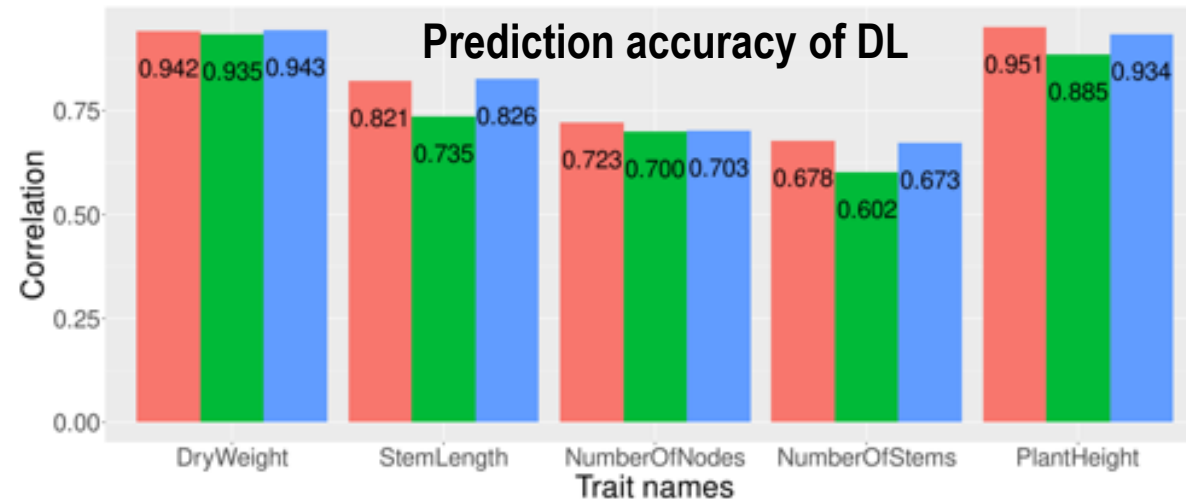
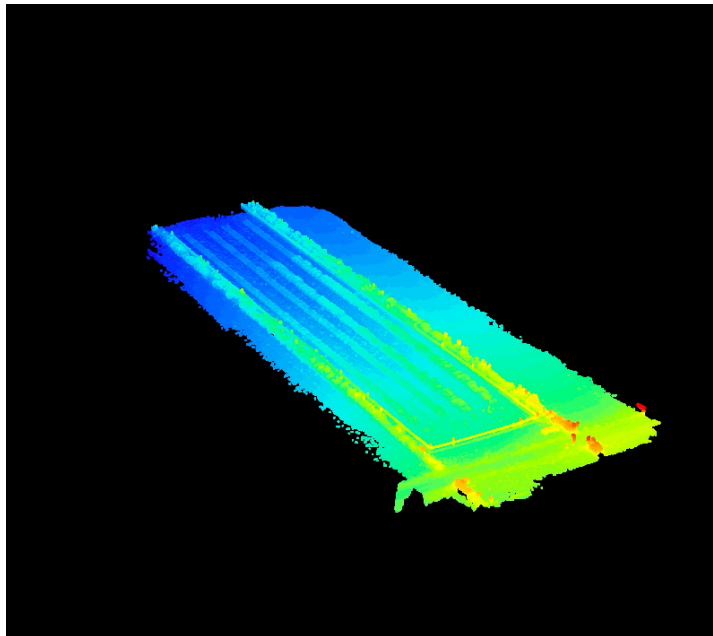
High-throughput data collection in the field



Prediction of biomass-related traits using deep learning



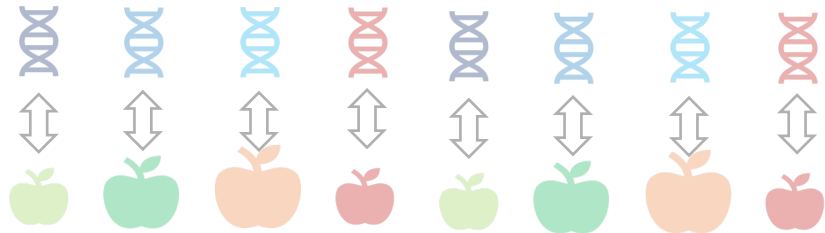
特徴量抽出



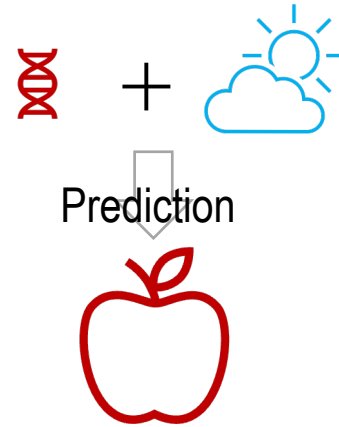
Less accurate than manual measurement, but superior in that it is non-destructive and allows measurement of large numbers of samples over time

Okada et al. (in prep)

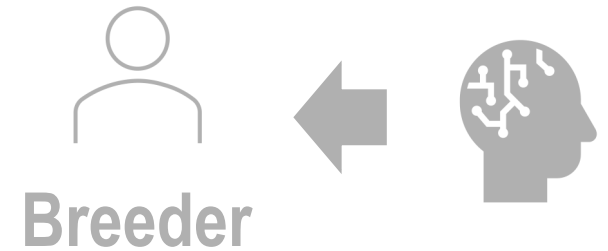
Challenges in increasing the efficiency of genomic selection (GS)



Collect a large number of data that can be used for model building (automatic and semi-automatic measurements)



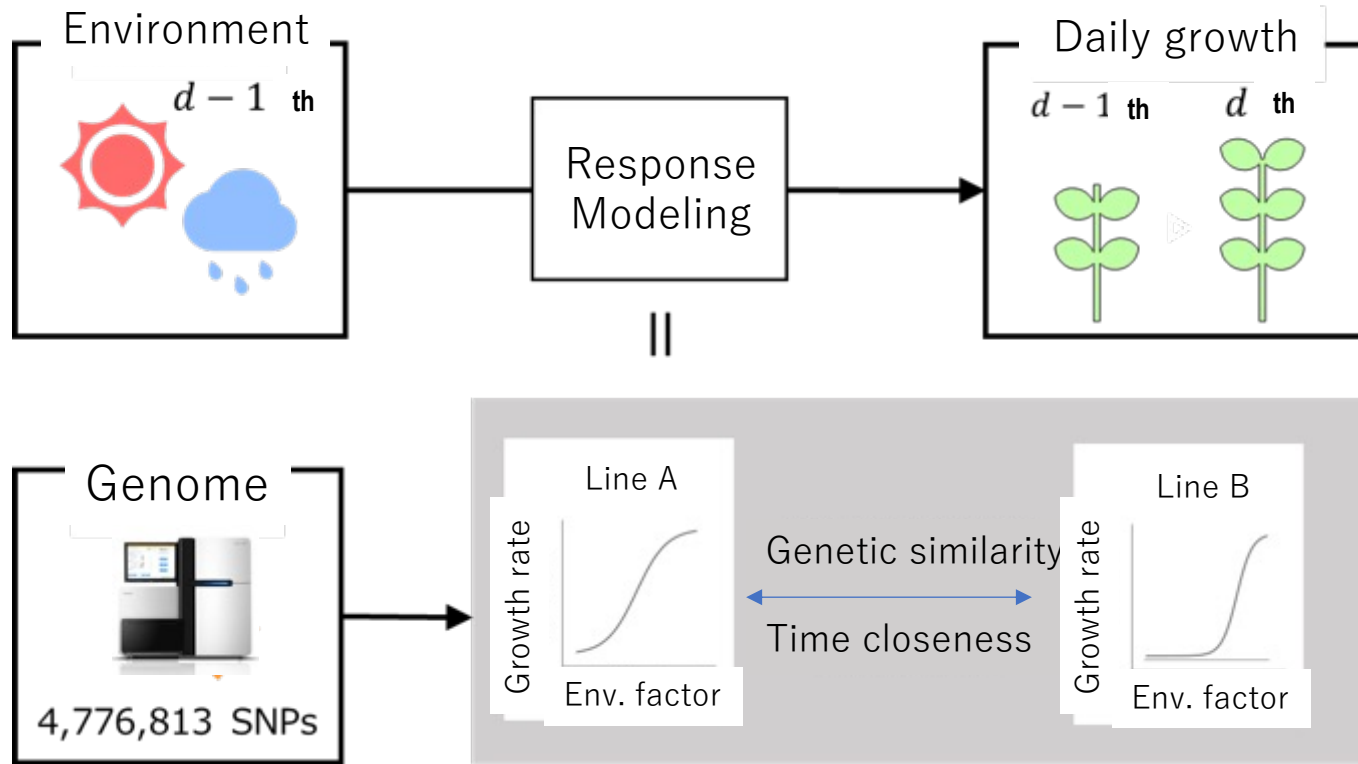
Make the model capable of proxying not only genomic but also environmental data



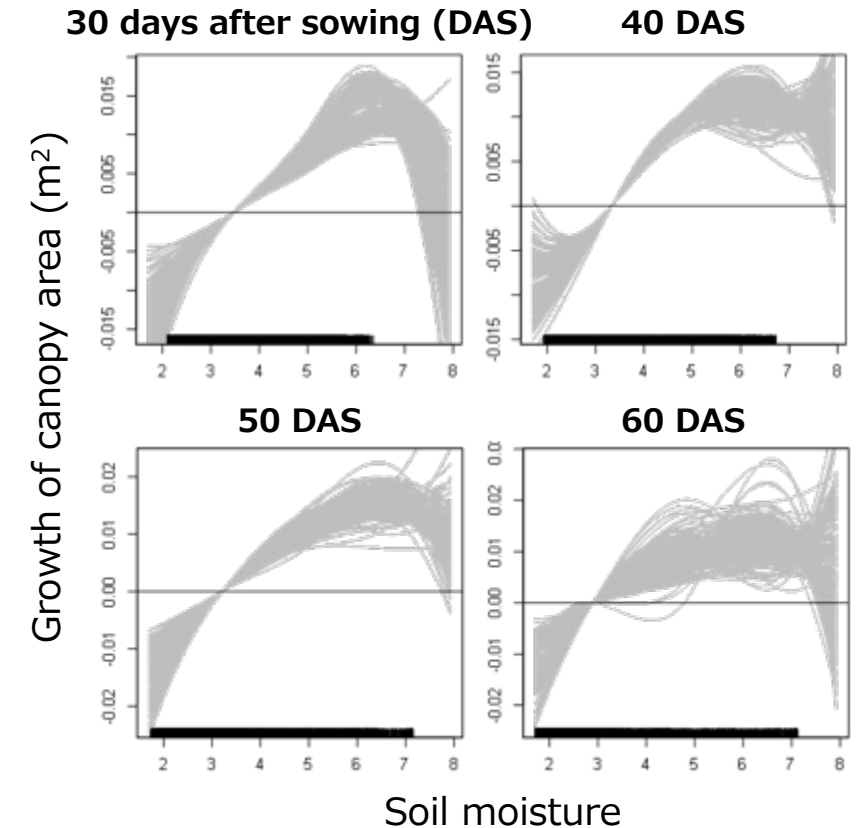
Making breeders' decisions data-driven, e.g., through data science optimization methods

Genomic and environmental modeling - Predicting daily growth

Model the environmental response of the growth with additive splines (assumed to be smooth with respect to genome and time of observation)



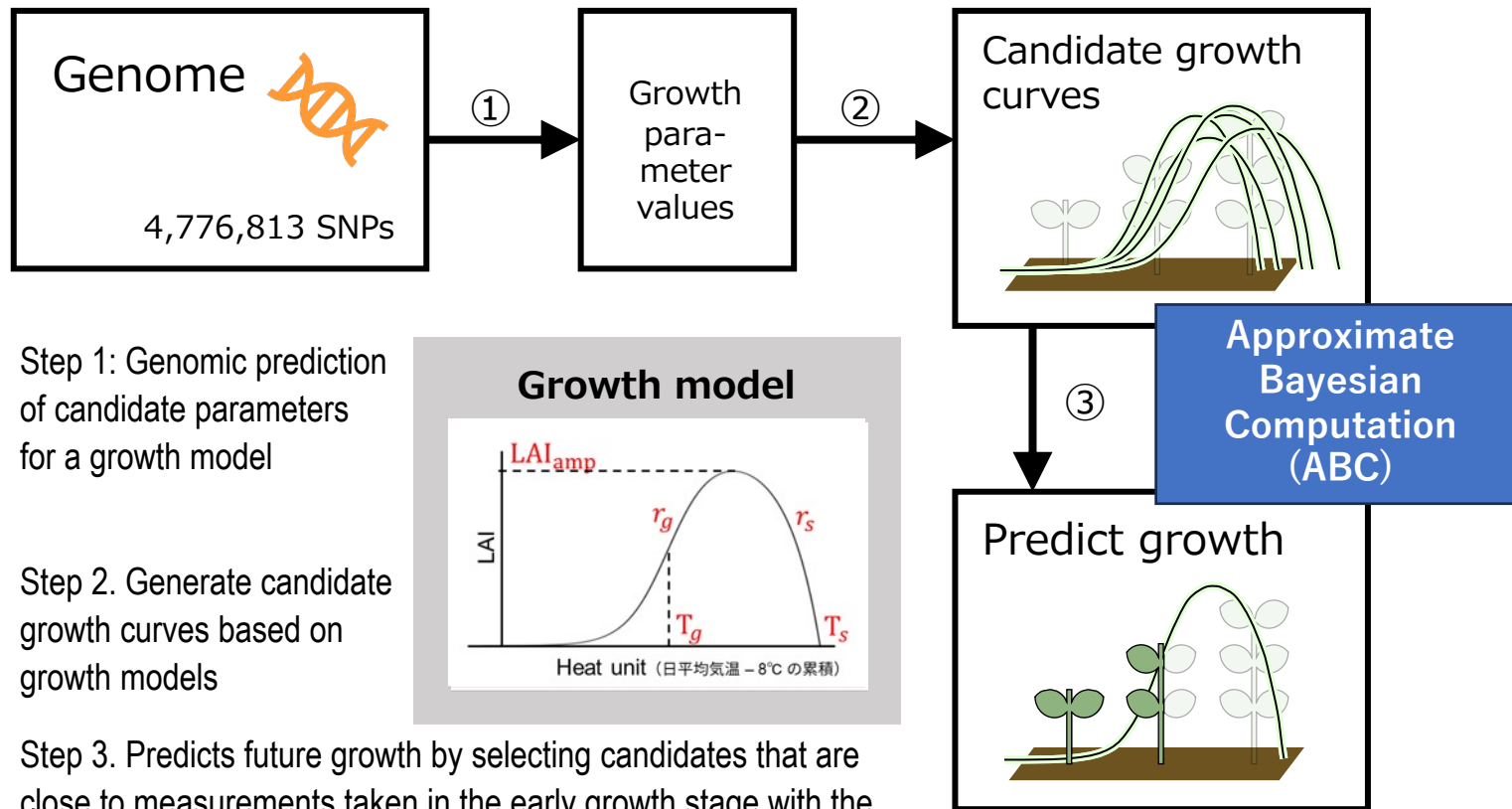
Response curves of all varieties to an environmental factor (soil moisture)



Toda et al. (Theor Appl Genet, under review)

Genomic and Environmental Modeling: Predicting future growth

Integrate growth models, genomic prediction models, and Bayesian methods for data assimilation

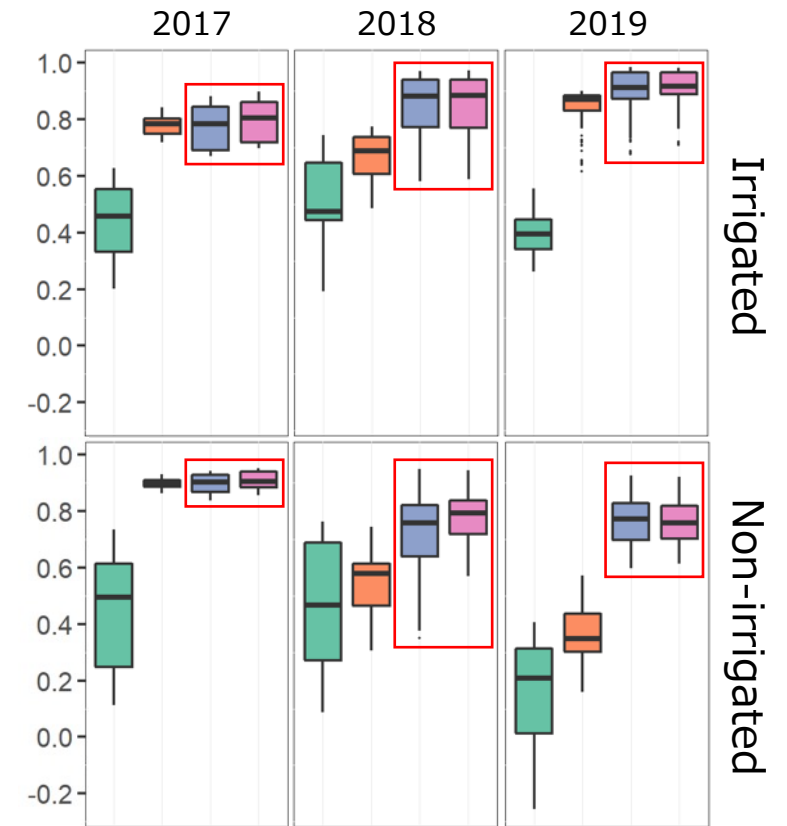


Step 1: Genomic prediction of candidate parameters for a growth model

Step 2. Generate candidate growth curves based on growth models

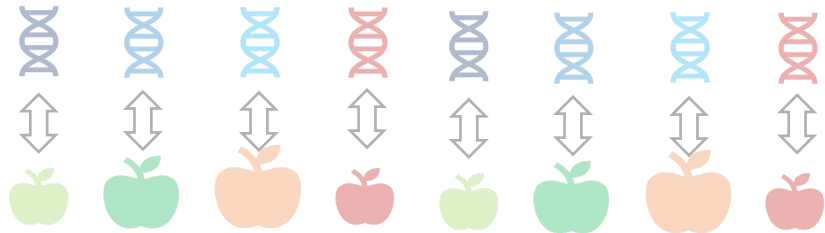
Step 3. Predicts future growth by selecting candidates that are close to measurements taken in the early growth stage with the ABC algorithm

Improved accuracy in predicting future growth

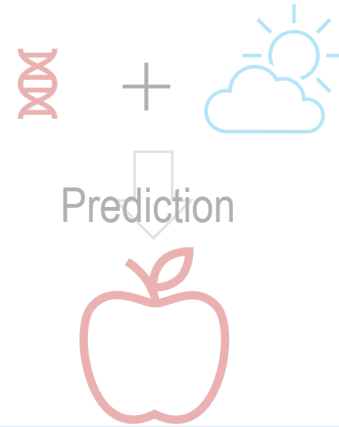


Toda et al. (2022) Frontiers in Plant Sci 13: 828864

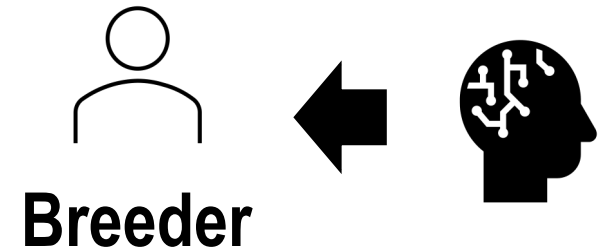
Challenges in increasing the efficiency of genomic selection (GS)



Collect a large number of data that can be used for model building (automatic and semi-automatic measurements)



Make the model capable of proxying not only genomic but also environmental data



Making breeders' decisions data-driven, e.g., through data science optimization methods

GS breeding Game

Plant Breed Game

PlantBreedGame

Information
Apimeta simulans, a species with a bright future!
 Recently discovered on the borders of the upper valley of the Aghromonpe, *Apimeta simulans* belongs to the *Statisticaceae* genus. It produces flowers which contain

Simulation game developed in France for genomic selection breeding

sequencing of 20 individuals: a high-density chip with 10000 SNP markers and a low-density one with 3333 SNP markers. KASPar genotyping can also be developed for single SNPs.

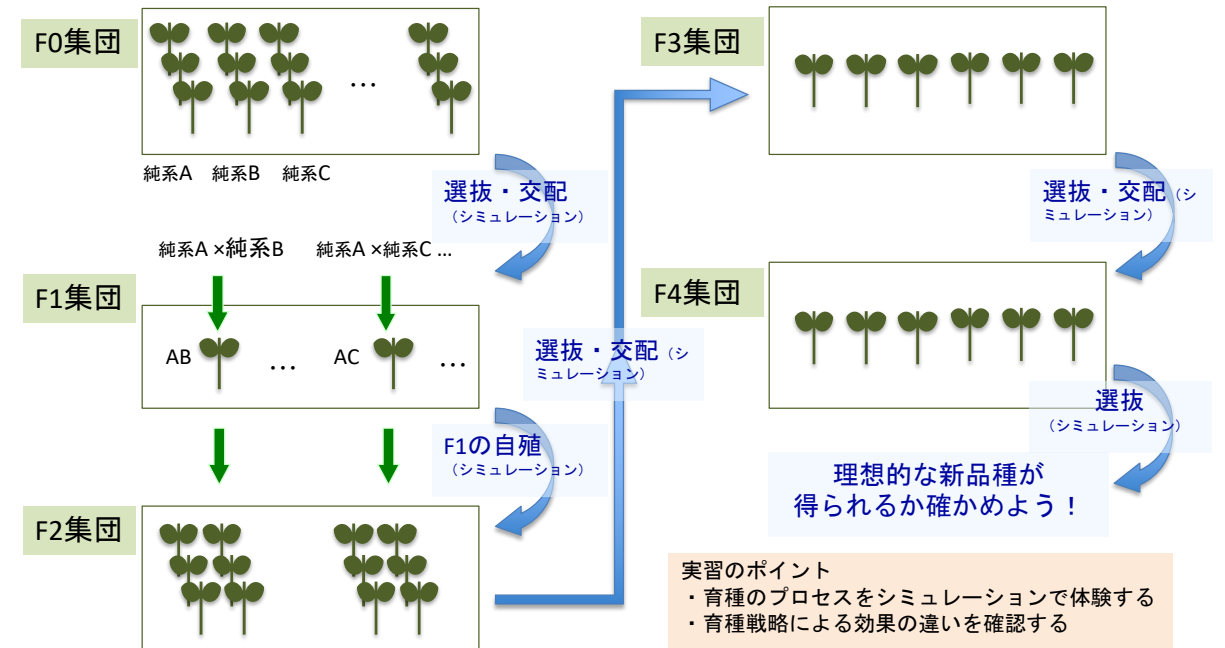
The species was domesticated recently. Despite the dangers in the uninhabited Aghromonpe valley, several sampling campaigns were conducted. As a result, numerous accessions were gathered into a genetic resources collection, from which 825 lines were derived.

Available data
 These lines were planted and phenotyped on the only experimental site consisting of 300 plots. Starting in 2005, each year for 10 years, 150 lines were planted, in 2 plots each. In addition, most lines were planted two successive years. Each year, the trial hence includes 75 lines already tested in the previous year, and 75 new

The goal of the breeding is the improvement of of yield of "sepmetin" in *Apimeta simulans* (*Statisticaceae*)

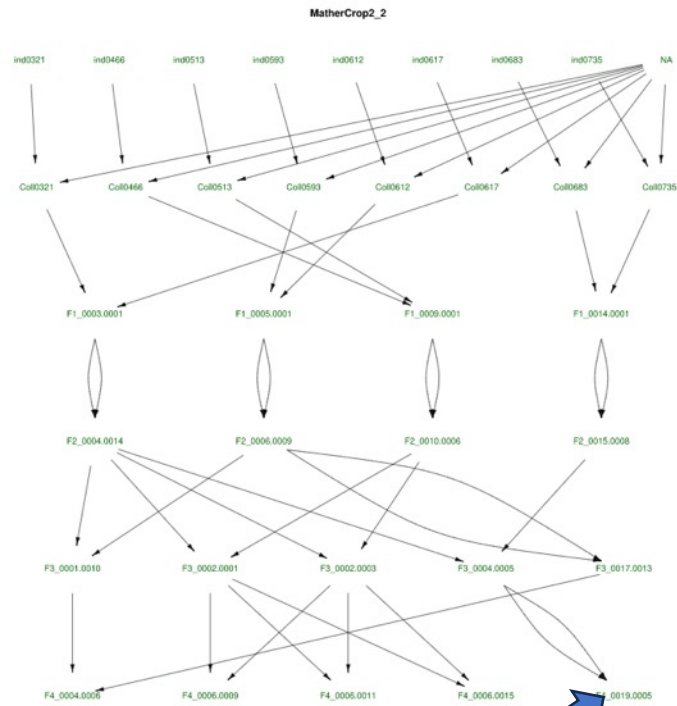
In our laboratory (Lab. Biometry and Bioinformatics), every year we have a student experiment program using the game.

Breeding Scheme in a Student Experiment

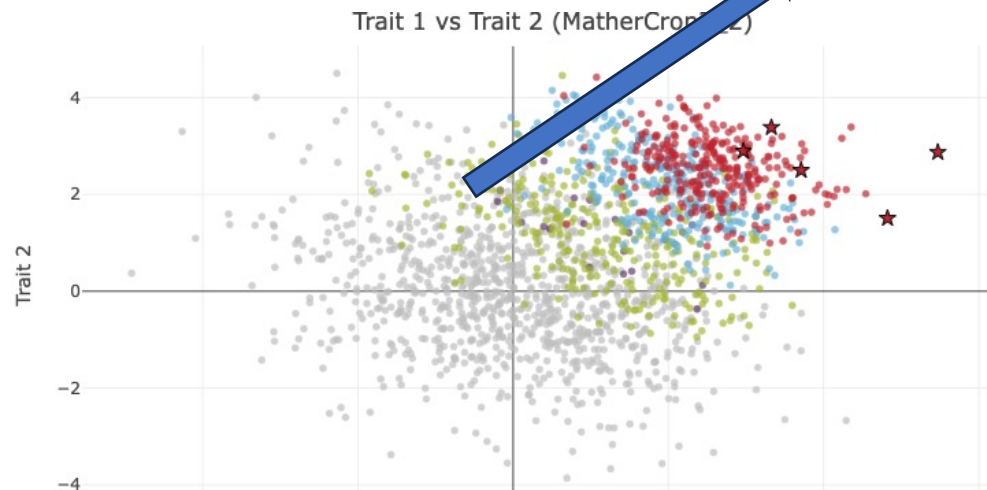


Four generations of selection and crossbreeding for simultaneous improvement of "yield" and "sepmetin" content

Group #3
Mather
Crop

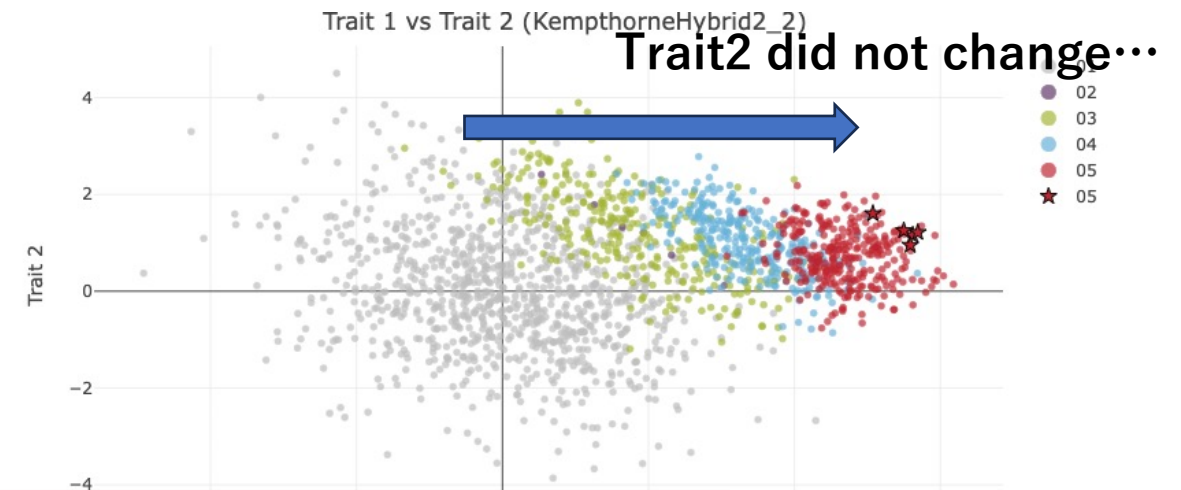
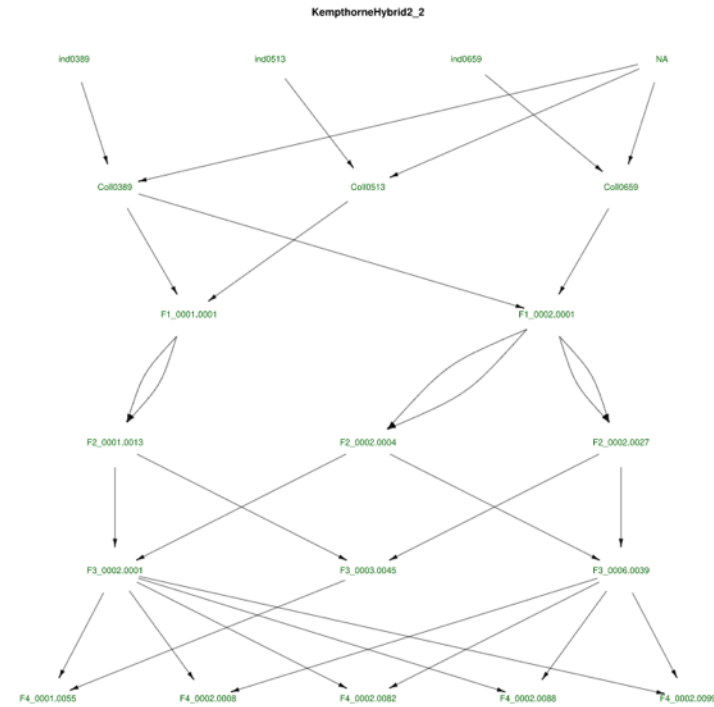


Good!



Results of the game

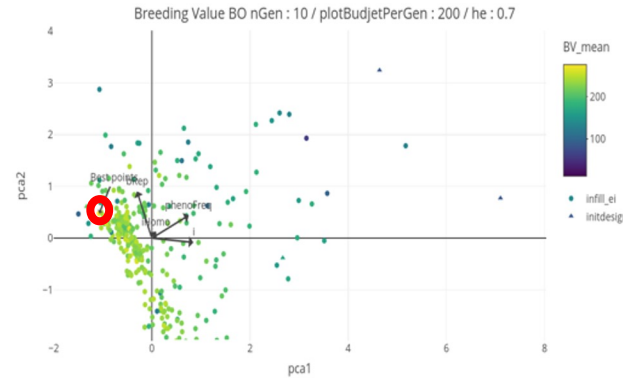
Group #2
Kempthorn
Hybrid



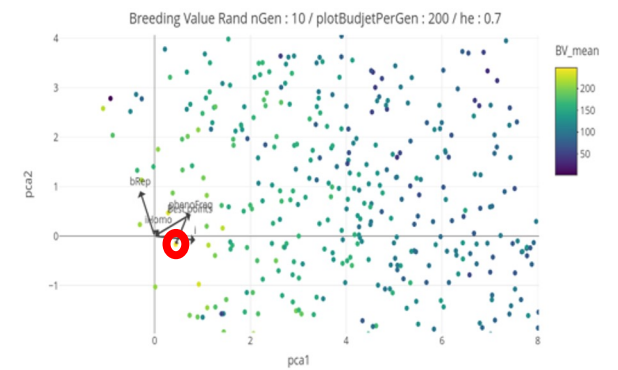
The same predictive model was used in the game.
But the results are very different!

Bayesian optimization for breeding schemes

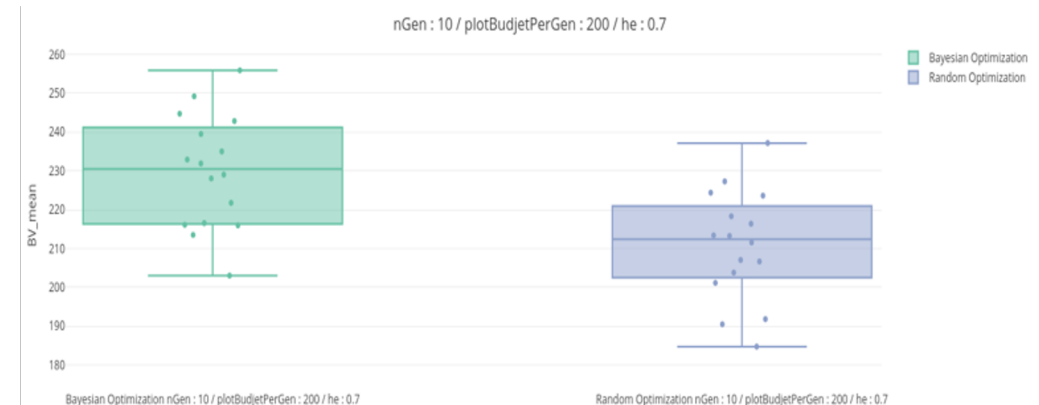
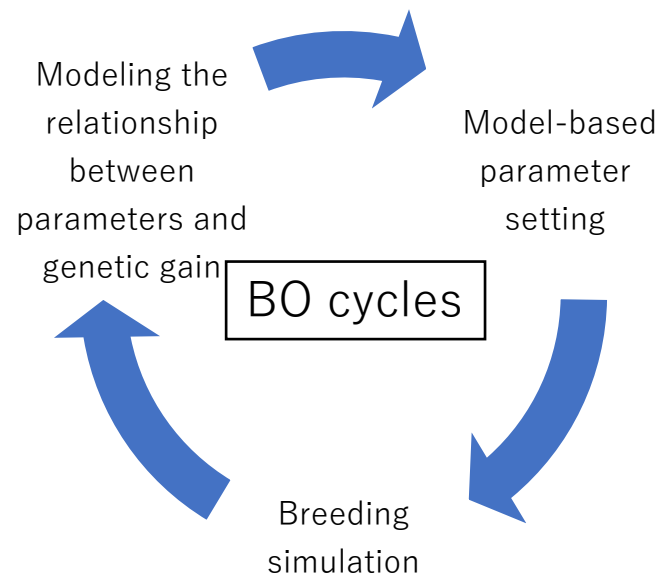
- Parameters to be optimized
 - Selection Strength**
 - budgetary allocation**
 - Frequency of phenotypic measurements**
- Restrictions on Breeding
 - total budget
 - Number of selected generations
 - Initial population size
 - genetic effect



The search for the best use of optimization algorithm



The search without using optimization algorithm



Use of optimization algorithms for greater effectiveness

Startups providing breeders with powerful tools to ensure stable food production



Our startup from UTokyo

Cooperative startups

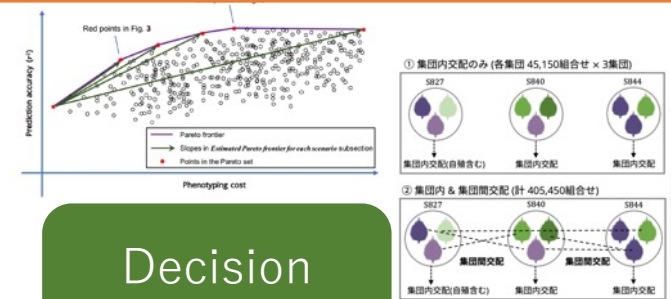


Implement data collection functions



Data Collection

Enhanced decision support capabilities



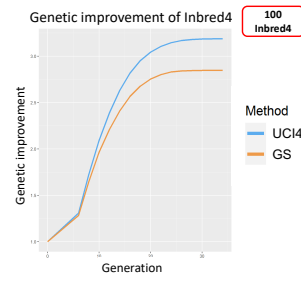
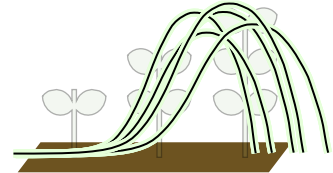
Decision Support

Data storage and linkage



Enhanced data storage and linkage functions

Data Analysis



Enhanced data analysis capabilities

Workshop on Plant biology

using remote sensing, modelling, genetics and genomics
to apply our knowledge for the better future.

Wed. 18th October, 09:00-15:00 (CEST)
UZH Irchel Campus, Y10-G-03/04 or online*

*Registration is required to obtain the Zoom link for online participation

<https://forms.gle/x2FQ7Mrms9rYs1wJA>



Funding support



Co-funded by the
UZH Global Strategy and
Partnerships Funding Scheme

Organizing committee

Reiko Akiyama
Chengchoon Ang
Kazuki Fukushima
Masaomi Hatakeyama
Lukas Rohrer
Aki Morishima
Yasuhiro Sato
Taro Takahashi
Kenji Yip Tong
Misako Yamazaki

Prof. Kentaro Shimizu

Morning Session I
Chair person: PD Andreas Hund

JST	CEST	Speaker	Affiliation	Title of the talk
16:00	09:00	Prof. Kentaro Shimizu	UZH	Opening remarks
16:05	09:05	Prof. Hiroyoshi Iwata	U o Tokyo	Modeling genome-phenome relationships to accelerate plant breeding
16:45	09:45	Yasuhiro Sato	UZH	Genome-wide neighbor effects predict genotype pairs that reduce herbivory in mixed planting
17:00	10:00	Afef Marzougui	ETH	Genotype-by-environment interaction studies using multi-scale phenotyping
17:20	10:20	Shunsuke Yoshioka	KU	A high-throughput phenotyping by UAVs of a wheat NAM population and exploration of effective traits for predicting yield components

Morning Session II
Chair person: Prof. Hiroyoshi Iwata

JST	CEST	Speaker	Affiliation	Title of the talk
18:00	11:00	Prof. Merry Schuman	UZH	Leveraging remote sensing to assess plant genetic variation
18:25	11:25	Haruka Sano	UoT	UAV measurements of growth and phenology in larch and their genetic dissection
18:40	11:40	Haozhou Wang	UoT	Virtual Broccoli Farmland Implementation by Drone-based Phenotyping and Cross-scale Data Fusion
18:55	11:55	Reiko Akiyama	UZH	Tool for monitoring plants in natura
19:10	12:10	Katharina Jung	UZH	Exploring an Asian NAM population for rust resistance and drought tolerance
19:25	12:25	Kengo Sakurai	UoT	Random regression for modeling soybean plant response to irrigation changes using time-series multispectral data

Afternoon Session
Chair person: Prof. Kentaro Shimizu

JST	CEST	Speaker	Affiliation	Title of the talk
20:45	13:45	Dr. Andreas Hund	ETH	Development of a global wheat full semantic segmentation dataset
21:05	14:05	Matthias Heuberger	UZH	How retrotransposons shape the evolution of wheat centromeres
21:20	14:20	Naoto Benjamin Hamaya	UZH	Unravelling the genetic architecture of male reproductive traits in Asian bread wheat
21:35	14:35	Aline Galatea Heger	UZH	molecular characterization of a non-conventional powdery mildew resistance protein in wheat: Pm4
21:50	14:50	Prof. Kentaro Shimizu	UZH	Closing remarks

UZH: University of Zurich
UoT: University of Tokyo
ETH: ETH Zürich