

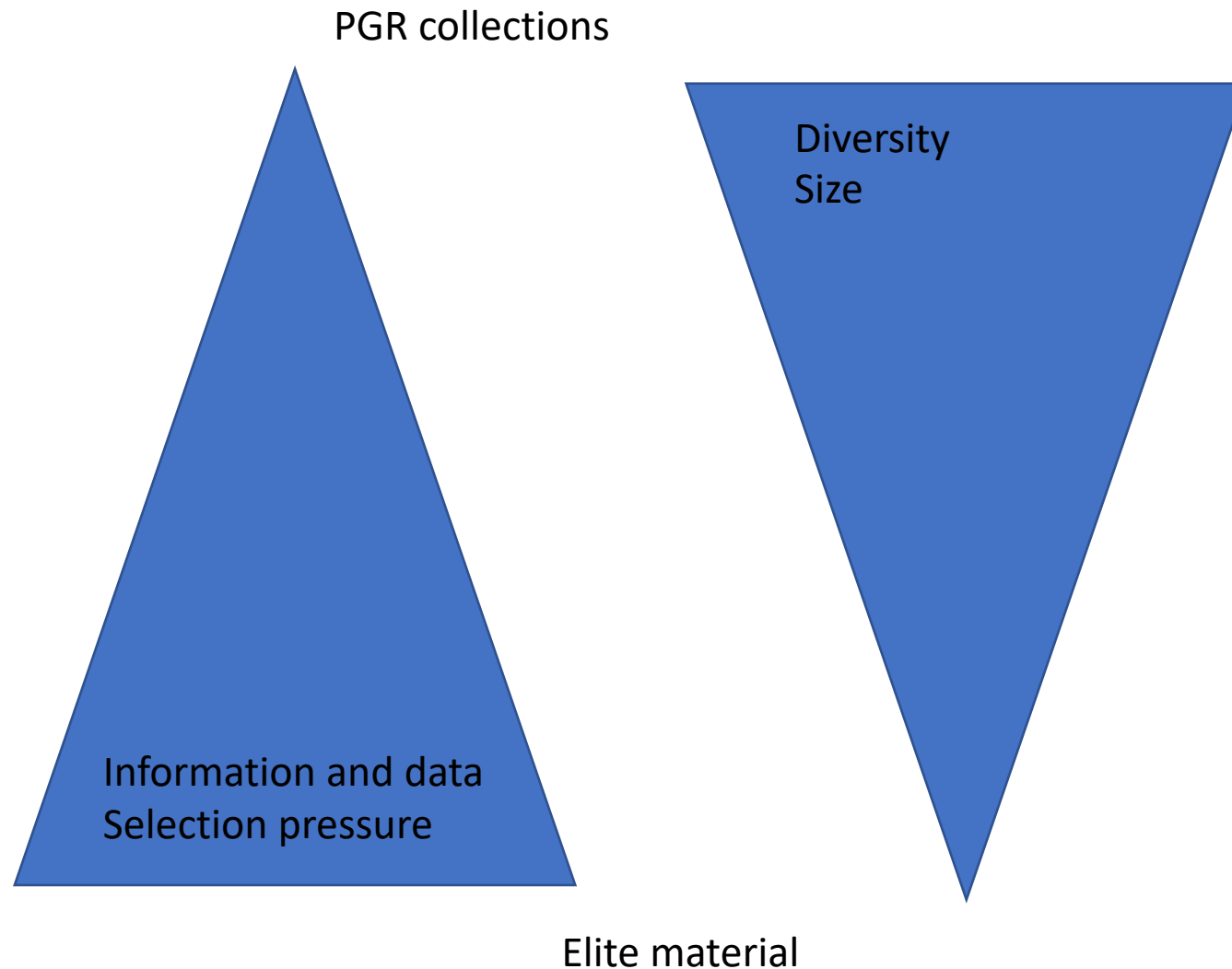
# FIGS: Combining phenotypic, environmental and genomic data to mine collections

Zakaria Kehel  
z.kehel@cgiar.org

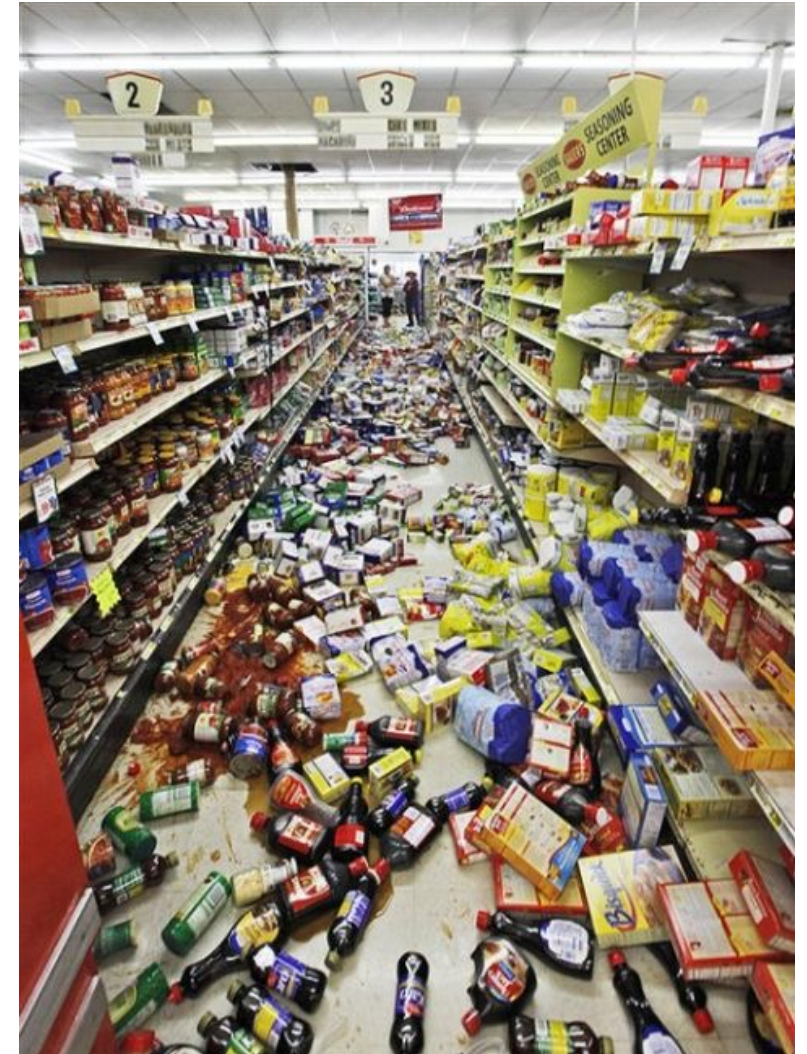


**Go Digital**

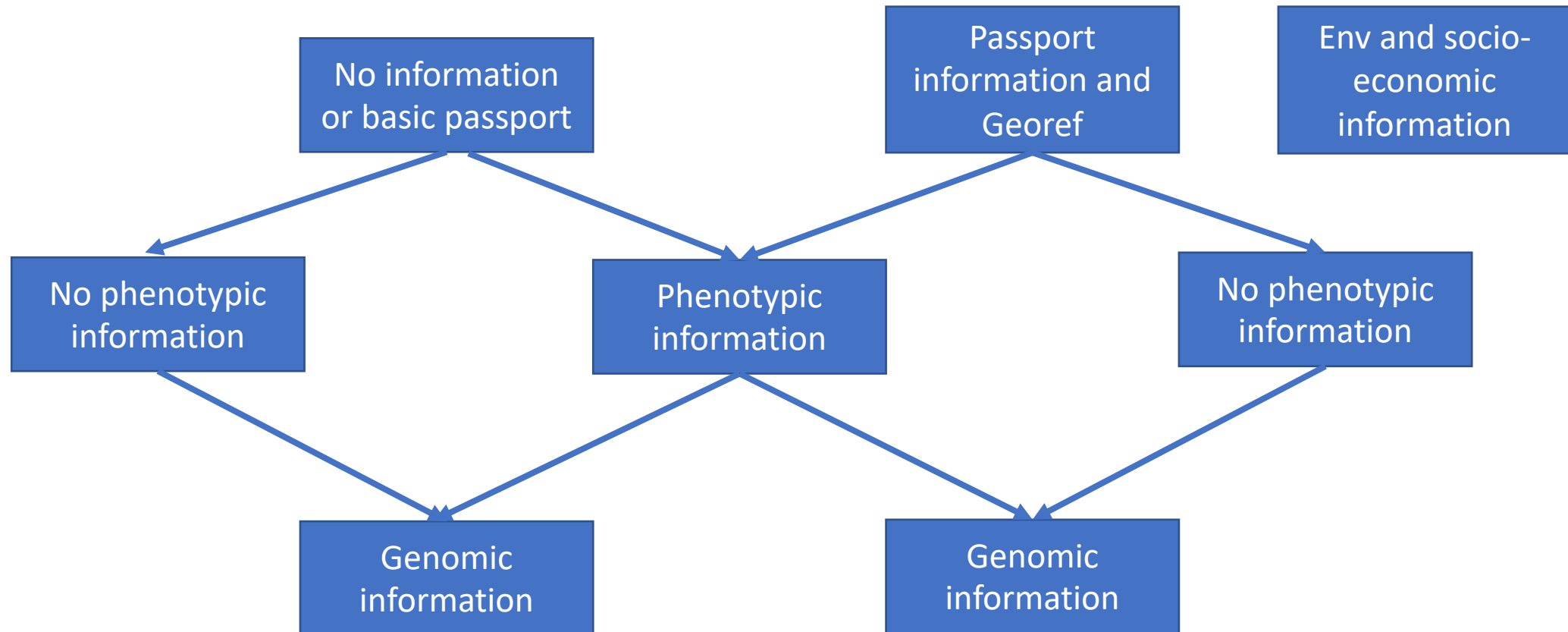
# Background



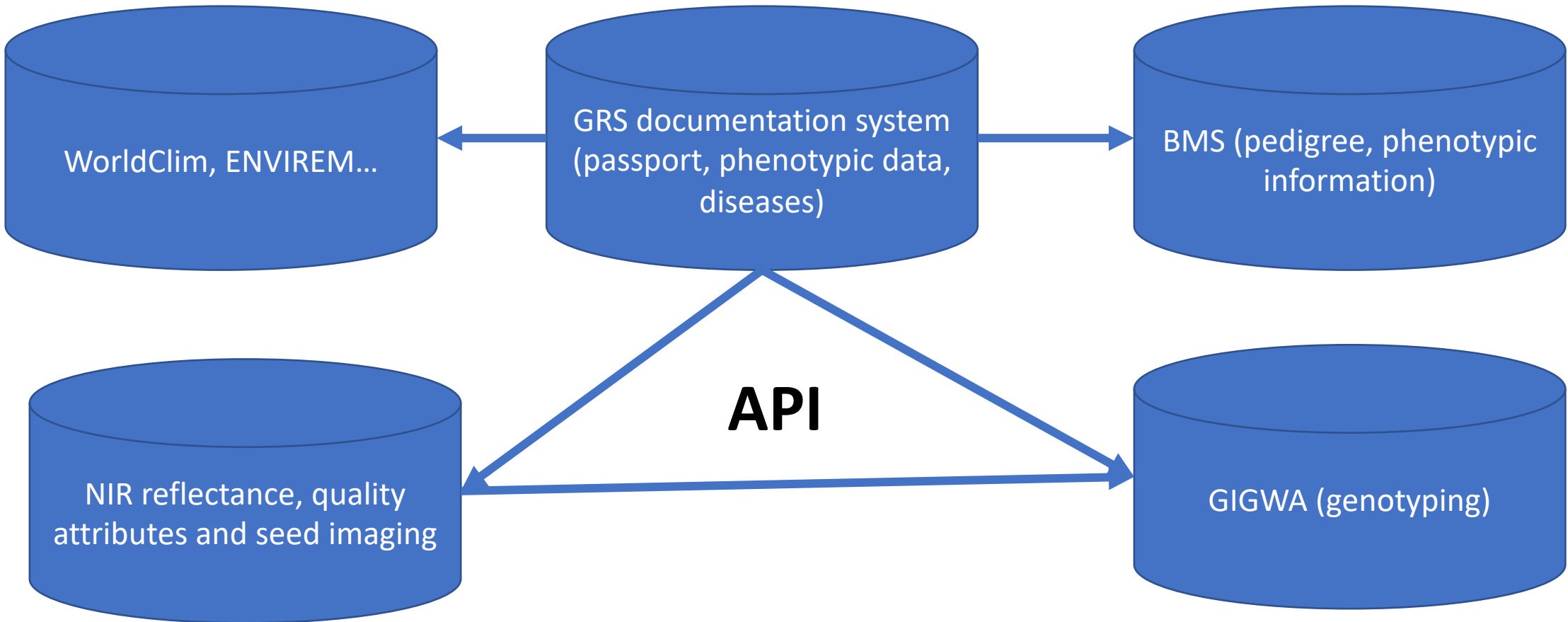
# Digital genetic resources



# What to do now?



# Data availability



# Focused Identification of Germplasm Strategy (FIGS)

## FIGS Approach

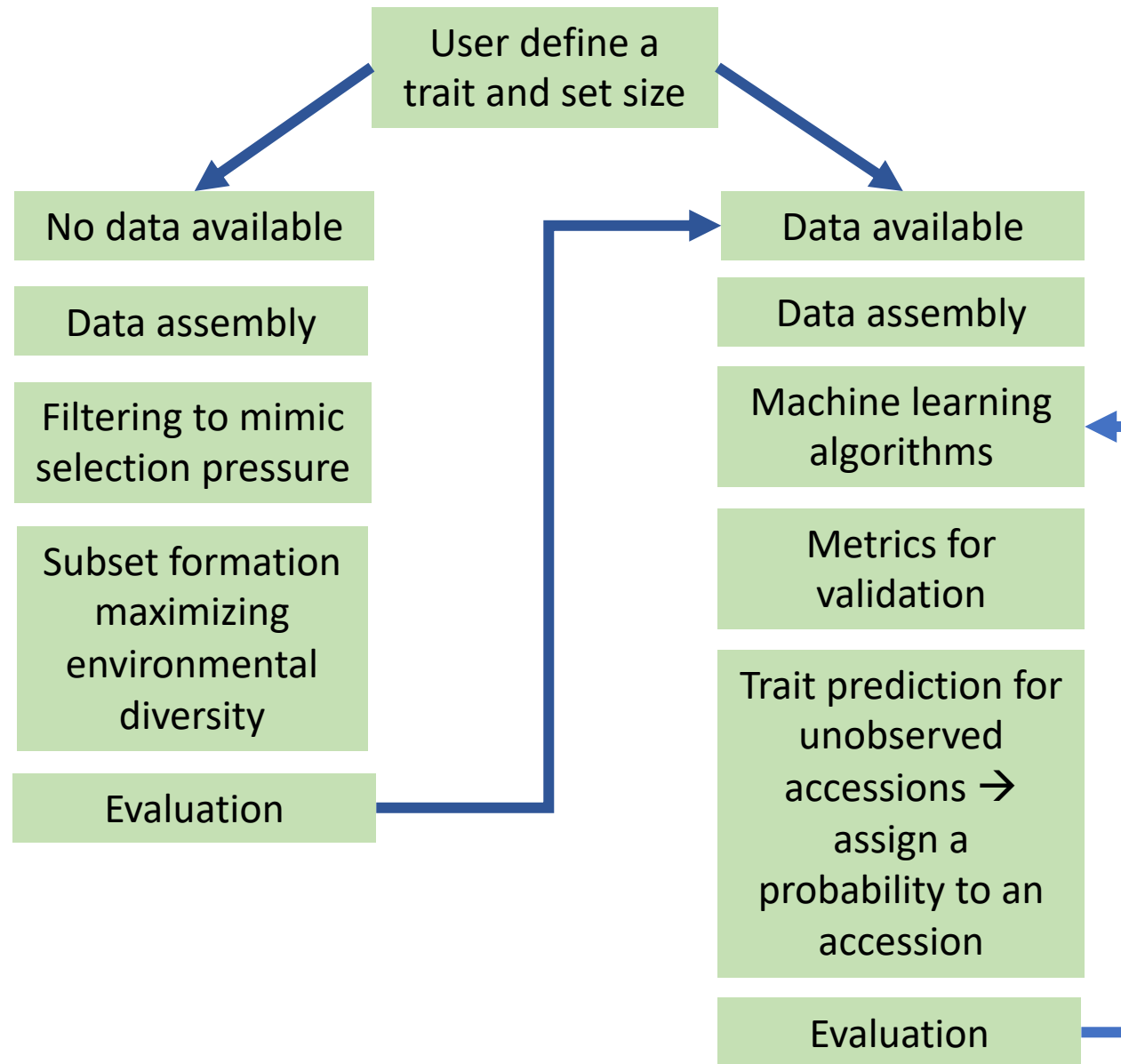


By **linking** traits, environments (and associated selection pressures) with genebank accessions (e.g. landraces and crop relatives) we can **focus** in on those accessions most likely to possess **trait specific genetic variation**.

Two methods resulting from the FIGS approach:

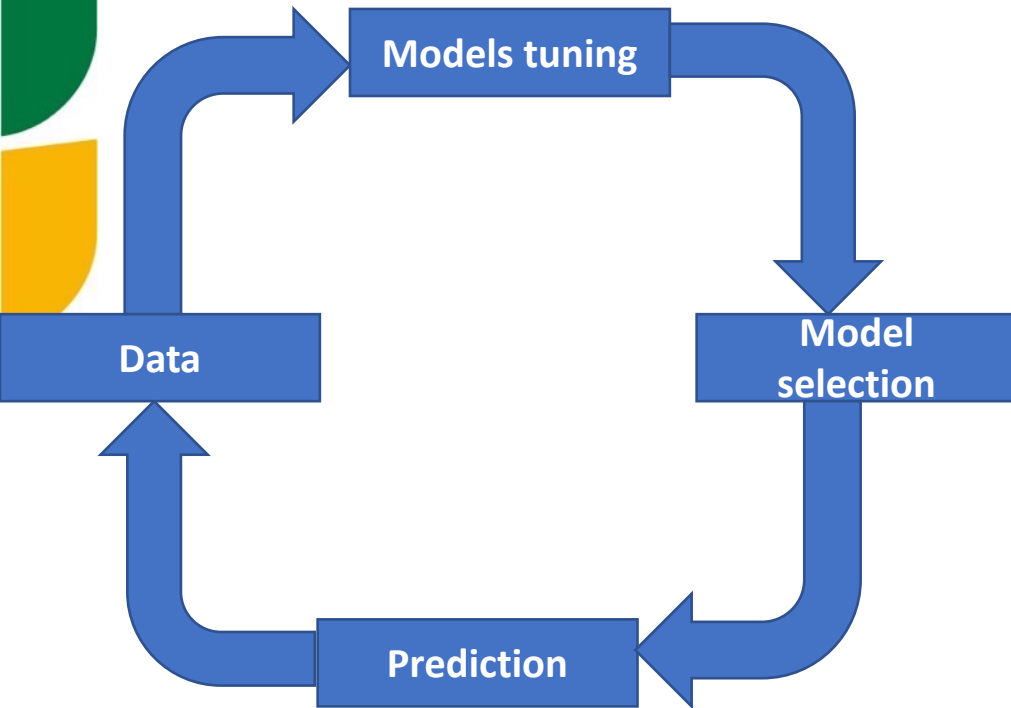
1. Filtering method
2. 'Predictive' modeling method

# FIGS pathways



# Predictive characterization

## *Durum wheat Grain filling period example*

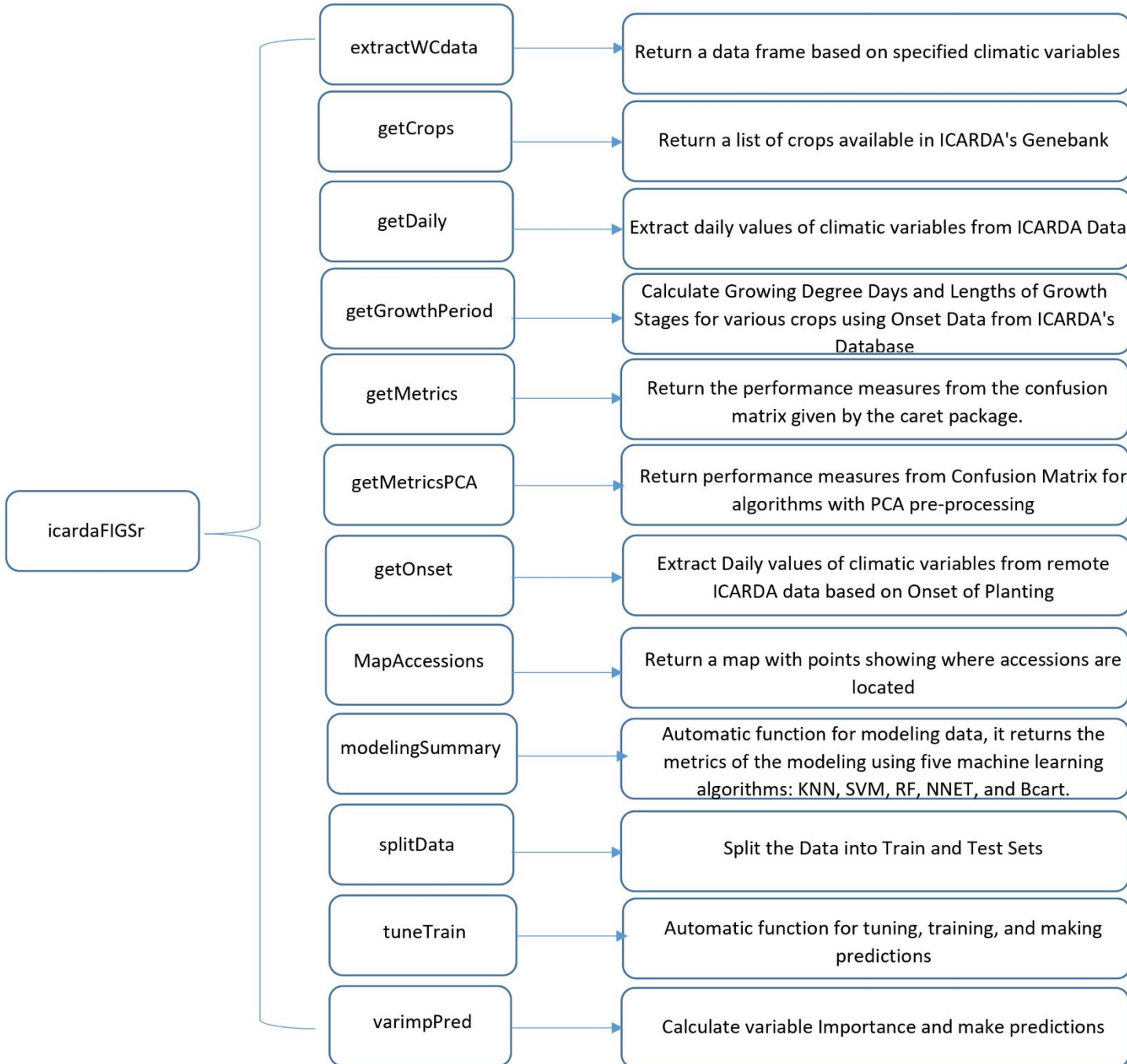


**FIGS modeling pathway**

Performance Measures	k-Nearest Neighbours	Random Forest	Support Vector Machine
Accuracy	0.834	0.838	0.817
95% CI	(0.799, 0.865)	(0.804, 0.868)	(0.781, 0.849)
No Information Rate	0.762	0.762	0.762
P-Value [Acc > NIR]	3.58E-05	1.37E-05	0.001423371
Kappa	0.563	0.557	0.467
Sensitivity	0.722	0.675	0.54
Specificity	0.869	0.889	0.903
True Positive	91	85	68
True Negative	351	359	365
False Positive	53	45	39
False Negative	35	41	58

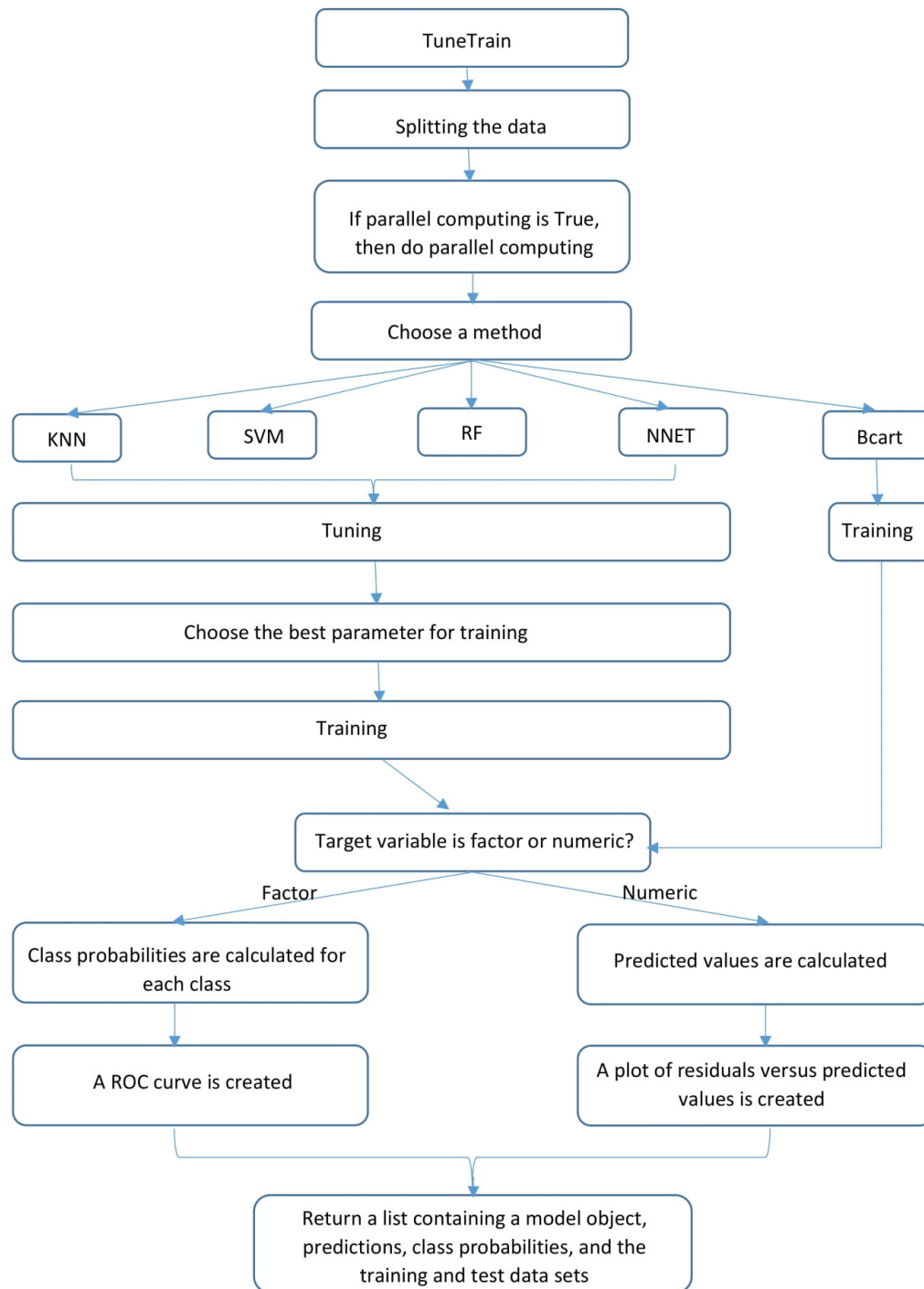
High accuracy showing that there is a strong relationship between GFP and longterm climatic conditions





# Tools: R package

# Tools: R package



- **TuneTrain** is an automatic function for tuning and training data, it returns a list containing a model object, predictions, class probabilities, and the training and test data sets
- **TuneTrain**(data, y, p = 0.7, method = method, parallelComputing = F, length = 10, control = "repeatedcv", number = 10, repeats = 10, process = c('center', 'scale'), summary= multiClassSummary,positive, ...)

ICARDA FIGS Data Extraction Climatic Data Analysis Trait Analysis

Passport Data World Climatic Data

Table Map Plot

Extract Accessions

Get Accessions by Crop Name  
 Get Accessions by IG  
 Upload External Data

Select a crop  
Crop

Select country(ies)  
Countries

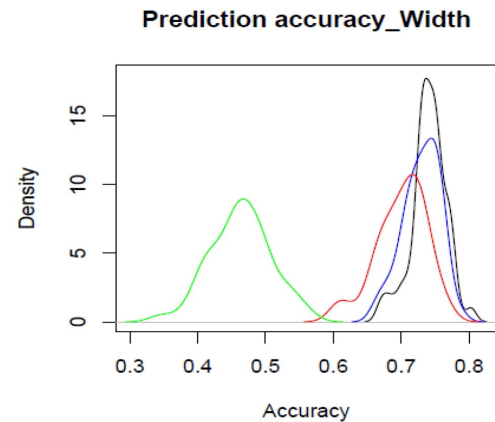
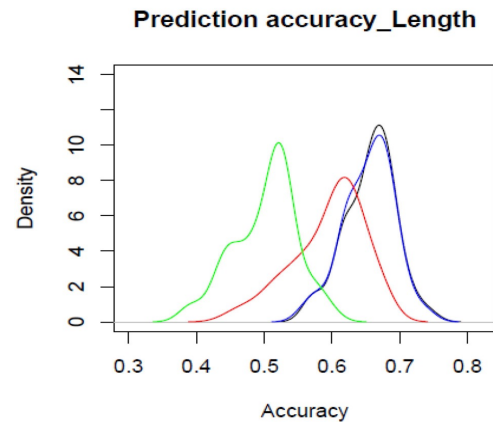
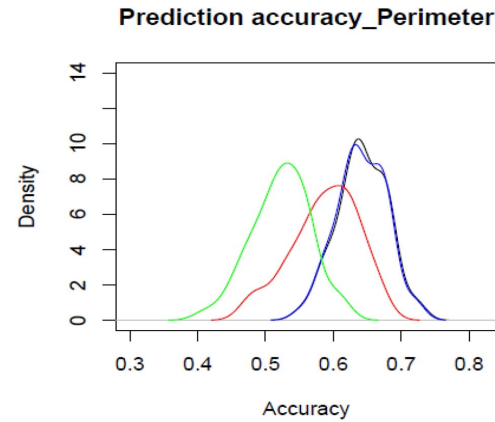
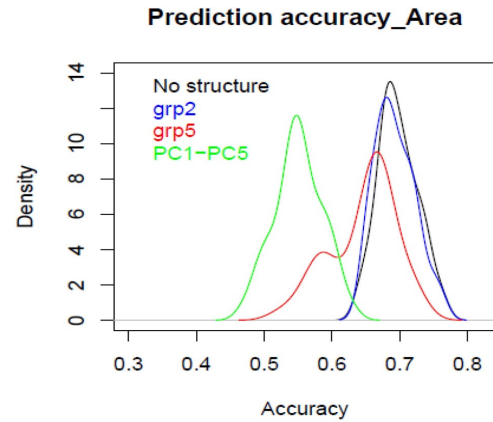
coordinates  
 DOI  
 Availability  
 Other IDs

Get Accessions

Download Accessions (.csv)

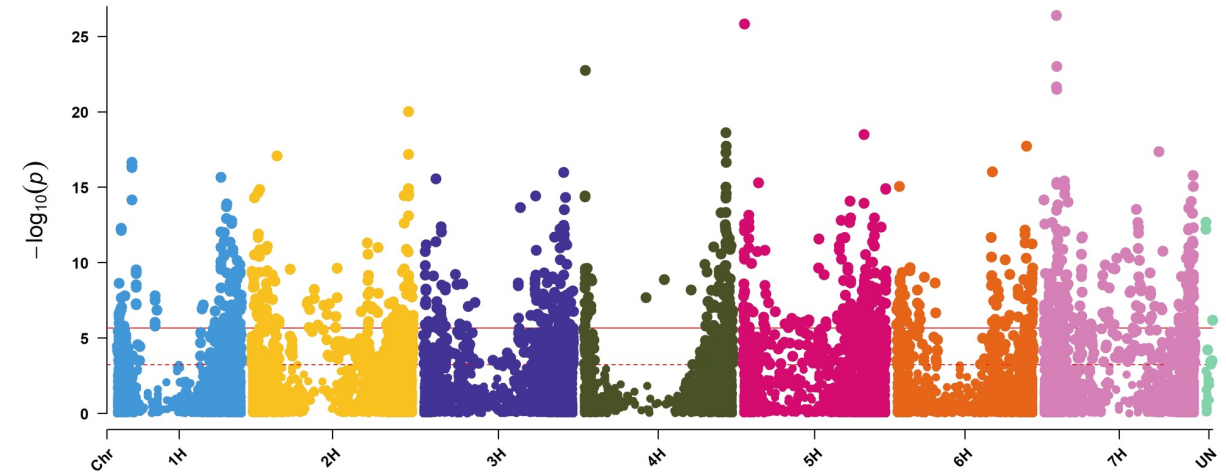
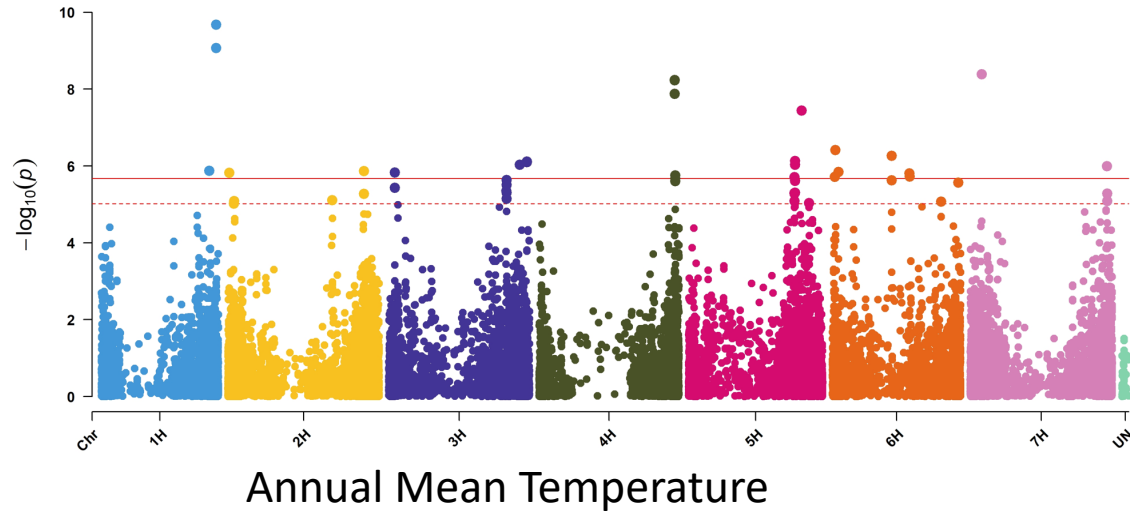
Download

# Adding value Genetic resources: FIGS augmented with Genomic selection for wheat quality example



		Area	Perimeter	Length	Width	ColCha1	ColCha2	ColCha3
All	Mean	0.70	0.64	0.66	0.74	0.69	0.71	0.66
	Maximum	0.76	0.73	0.74	0.80	0.78	0.80	0.73
	Minimum	0.64	0.55	0.57	0.67	0.60	0.62	0.59
All-Env	Mean	0.75	0.70	0.70	0.78	0.75	0.76	0.72
	Maximum	0.78	0.75	0.77	0.81	0.79	0.80	0.75
	Minimum	0.66	0.59	0.59	0.69	0.61	0.63	0.60
Env	Mean	0.06	0.04	0.03	0.10	0.03	0.05	0.04
	Maximum	0.20	0.19	0.22	0.29	0.20	0.25	0.24
	Minimum	-0.12	-0.19	-0.16	-0.07	-0.16	-0.13	-0.13
grp1	Mean	0.73	0.76	0.79	0.65	0.58	0.63	0.58
	Maximum	0.82	0.86	0.88	0.75	0.78	0.80	0.71
	Minimum	0.64	0.67	0.69	0.49	0.42	0.47	0.40
grp2	Mean	0.67	0.54	0.51	0.73	0.73	0.74	0.69
	Maximum	0.74	0.67	0.65	0.80	0.79	0.79	0.75
	Minimum	0.58	0.46	0.40	0.65	0.63	0.64	0.61

- GS shows efficiency in predicting seed morphology in Genebank wheat collection
- Combining FIGS and GS showed better predictability
- Increase knowledge on quality attributes which was bottleneck in the past



- eWAS revealed several regions for adaptation to Temp, Rain and soil
- Understand adaptation of CRW to different environments
- Increase knowledge to identify sources of adaptation to abiotic stress from PGR

- Organizing, curating and connecting different datasets
- Take advantage of data mining and machine learning technologies
- Increasing predictability with what is available





**Thank you!**  
**Questions/Comments?**

