

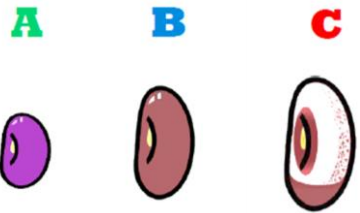


Digital functional phenomic descriptors featured from machine learning-driven image-based phenotyping improve the accuracy of classic descriptors: A case study on *Arachis* spp. and *Phaseolus* spp.

Conejo-Rodriguez F, Gonzalez-Guzman J , Ramirez-Gil J, Urban Milan and Wenzl P

# Characterization / evaluation in genebanks: Future scenarios

Genebank regenerate



Seed Increase

Distribution

Seeds viability

Seeds health



Seed regenerations



Descriptors selection

Traits optimization for identifying higher value germplasm [25]

Classification ML

Species  
Accessions  
Tolerance  
Nutrition

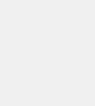
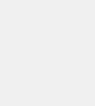
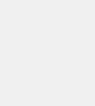
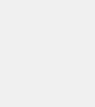
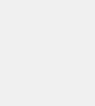
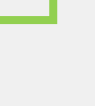
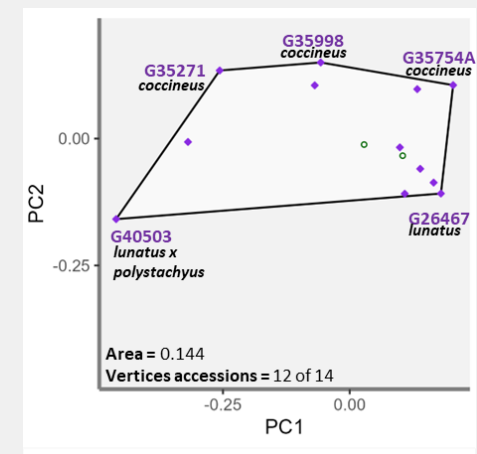
Phenotypic data  
(Morphology, physiology and agronomy)



HTP and spatial analysis

Sensors  
Image analysis  
Spectrometers

Genebank functional diversity



# OBJECTIVE

- Improve the characterization processes of bean and peanut accessions in the CIAT genebank through phenotypic descriptors based on digital images that contribute to accuracy in the discrimination and identification of redundancy among accessions using artificial intelligence models and propose a methodology that could be integrated into the genebanks workflow.





# Forage peanut and bean accessions

2



*A. paraguansensis*

11



*A. pintoii*

1



*Arachis spp*

1



*A. archeri*

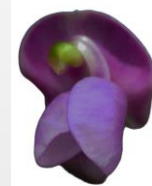
1



*A. repens*

16 accesiones  
*Maní*

1



*P. tuerckheimii*

1



*P. lunatus x P. polystachyus*

1



*P. dumosus x P. vulgaris*

4



*P. coccineus*

4



*P. dumosus*

3



*P. lunatus*

14 accesiones  
*Frijol*

Ten plant for  
accession

160 forage peaunut plant  
140 beans plant



# Classic descriptors



**Growth habit**  
Performance  
Height plant  
Leaf shape



**Growth habit**  
Leaf ratio  
Leaf length



**Days to flower**  
Flower color



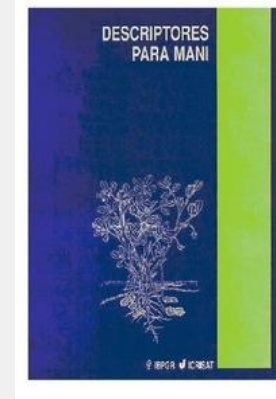
**Days to flower**  
Staminate flower color  
Wings flower colors



**Days to harvest**  
100 seed weight  
Primary color seed  
Secondary color seed



**Forma de semilla**  
Brillo de semilla  
Peso de 100 semillas



Historic data genebank



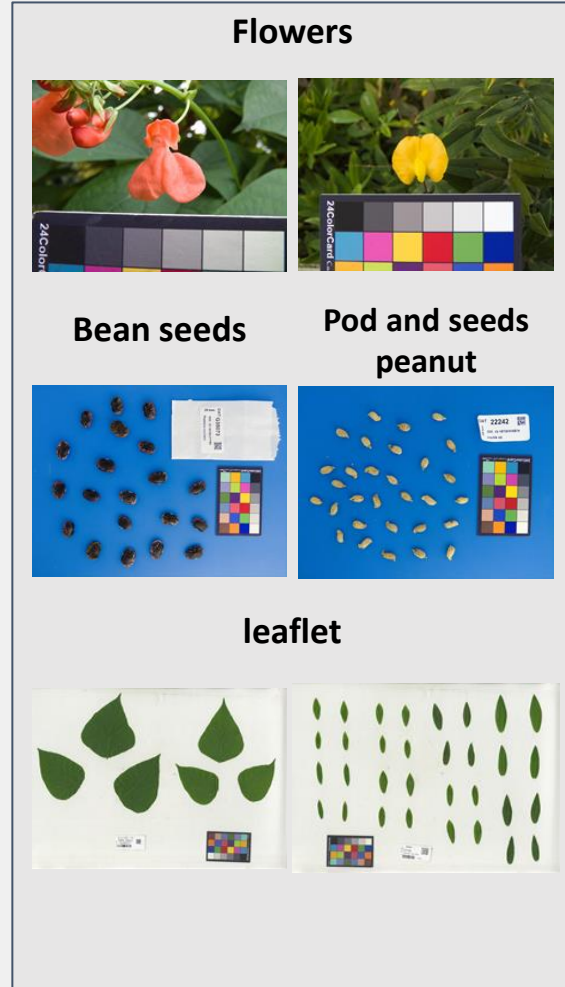
RHS  
Chart  
color

Phenomic characterization  
Images processing

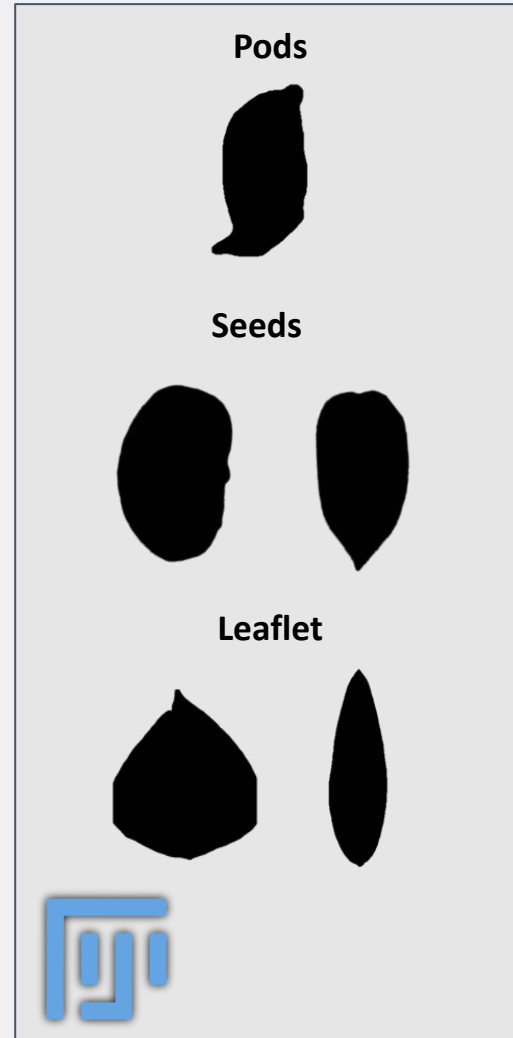
Image capture



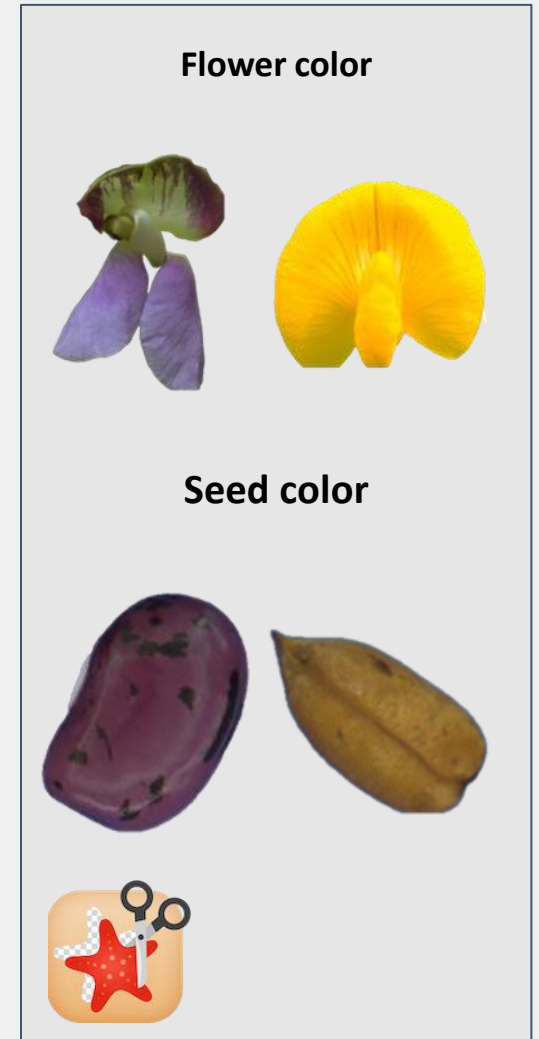
JPG image



Geometric morphometrics  
Outline analysis

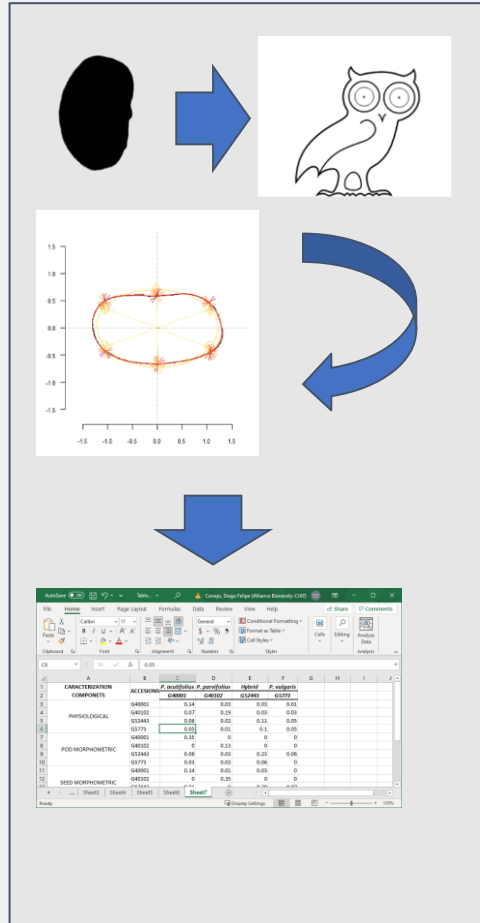


Color space RGB  
extracted

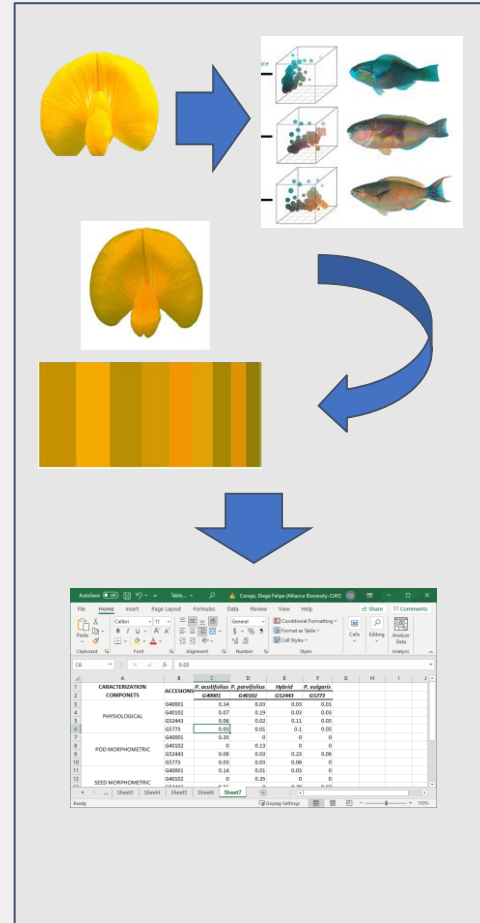




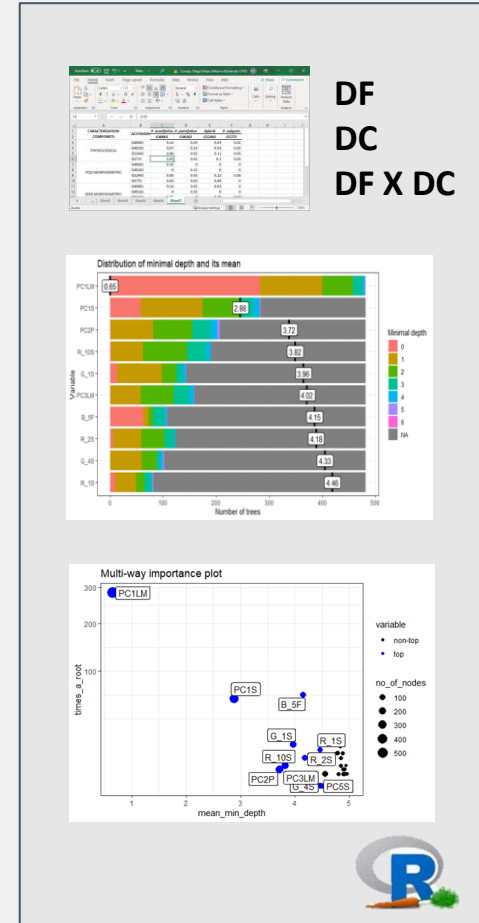
### Outlines análisis Geometric morphometrics



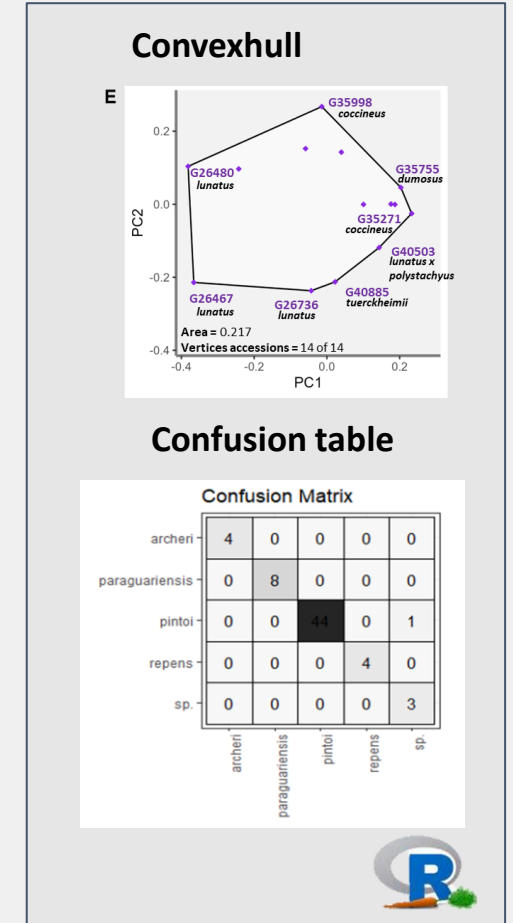
### Colorimetrics cluster - k medias



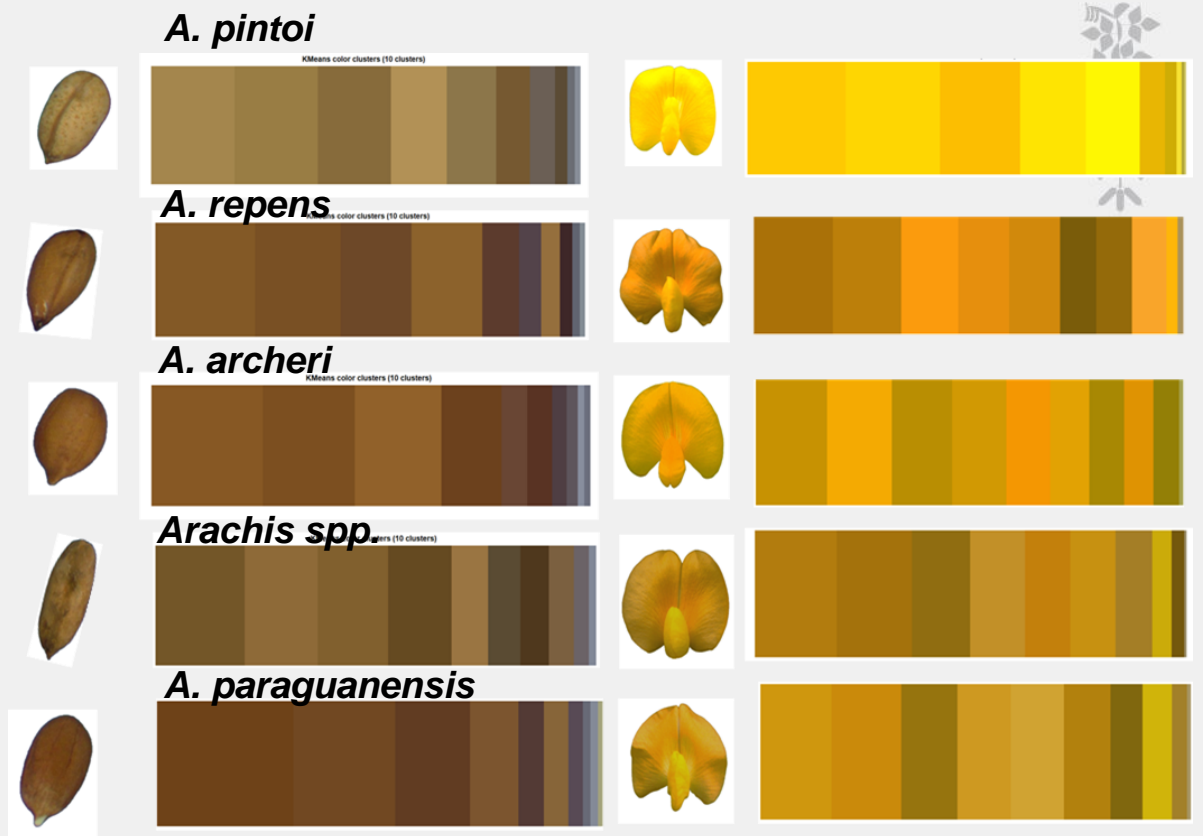
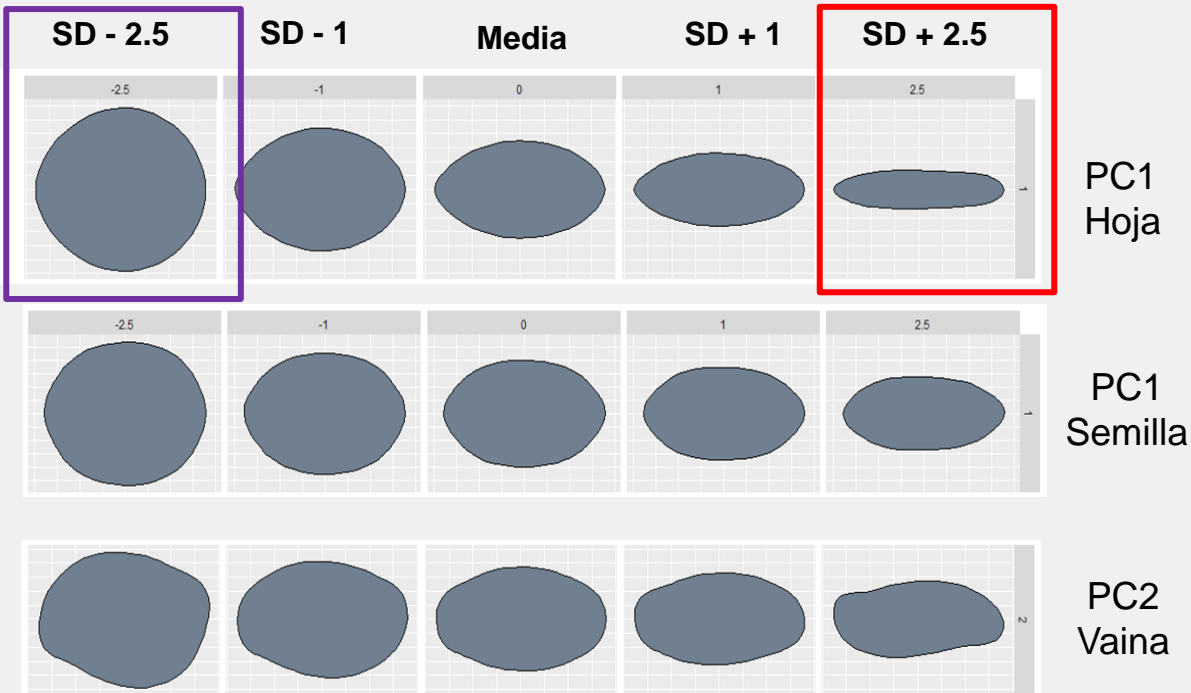
### Descriptors selection: Random forest



### Classification groups and accessions



# Phenomics forage peanut characterization



*A. pinto*  
CIAT 20826



*A. paraganensis*  
CIAT 22636

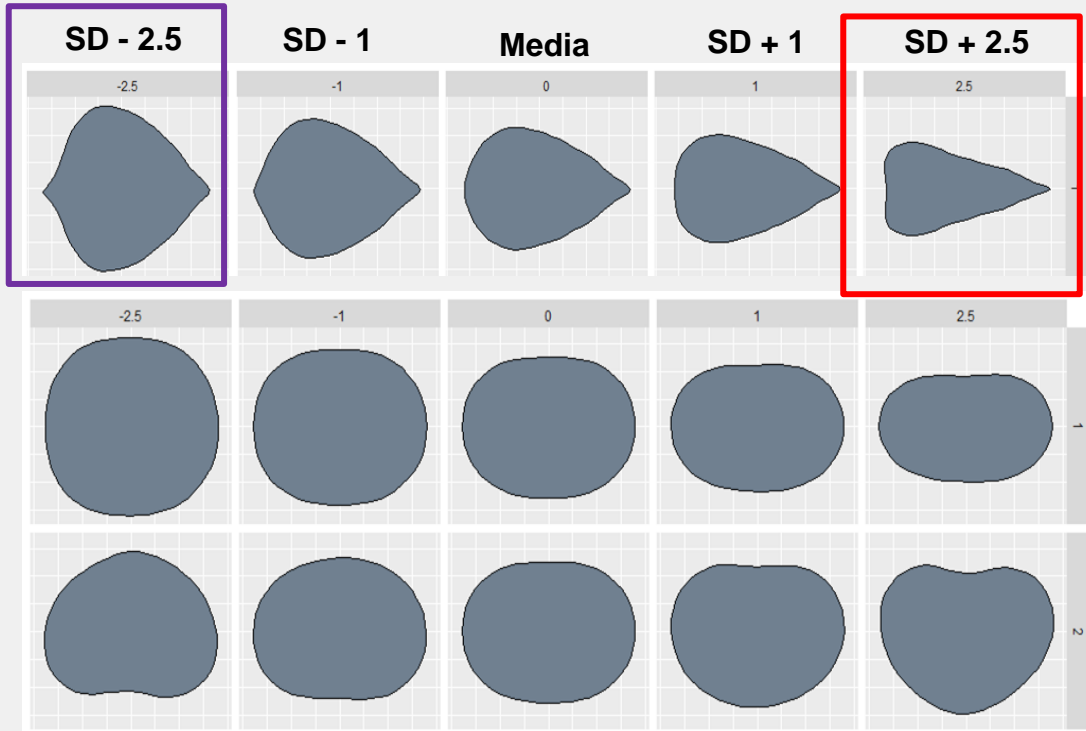


Outlines analysis

Color cluster accession



# Phenomics bean characterization



*P. dumosus*  
CIAT G35586



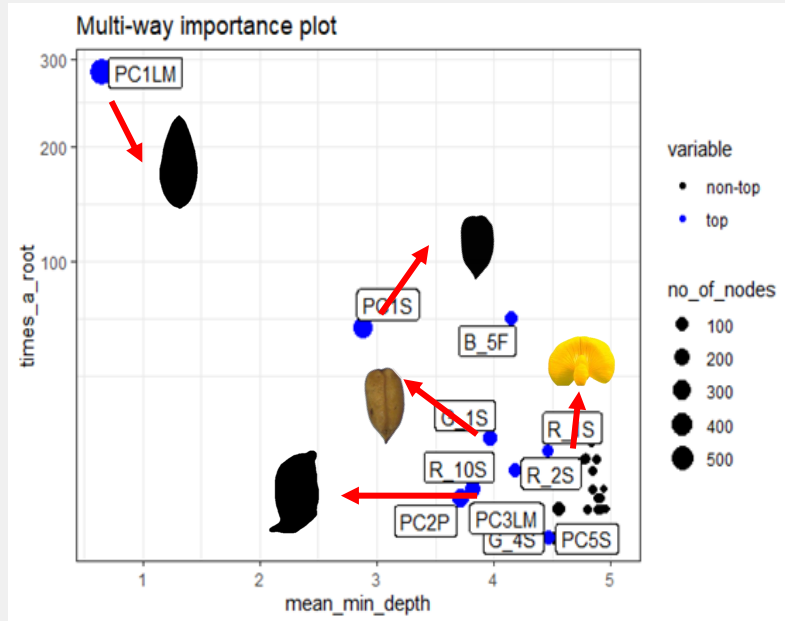
*P. lunatus*  
CIAT G26736



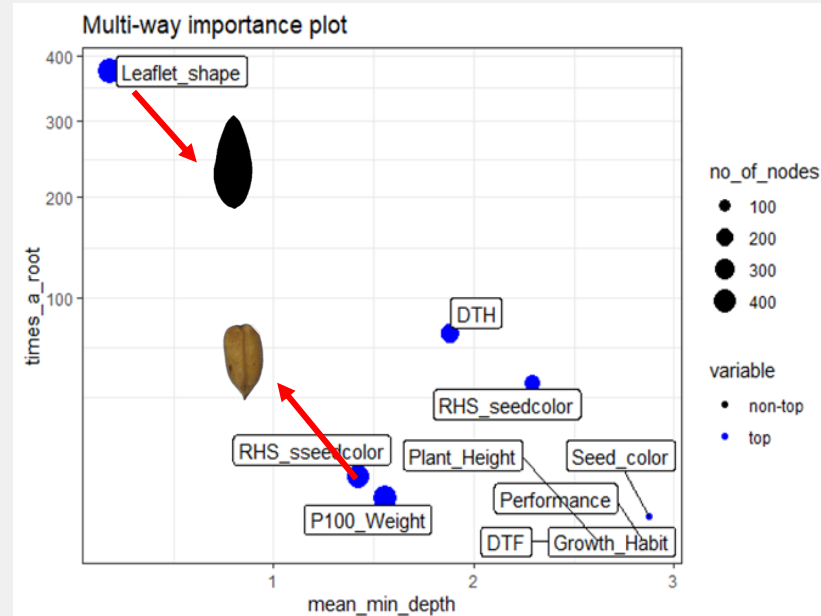
# Descriptors selection in peanut using machine learning



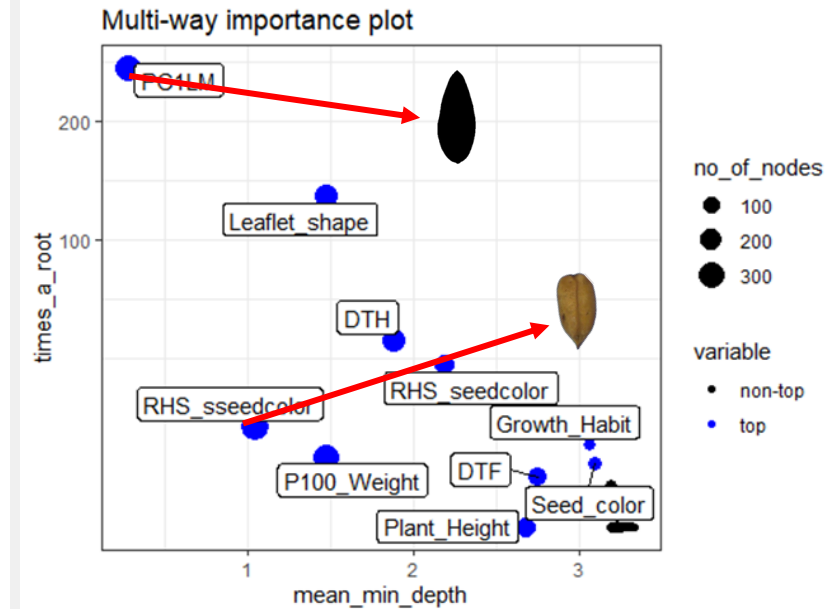
## Phenomic descriptors



## Classic descriptors



## Phenomic + Classic



Leaf and seed morphometrics is important discriminant descriptor in accession classification

# Descriptors selection in beans using machine learning

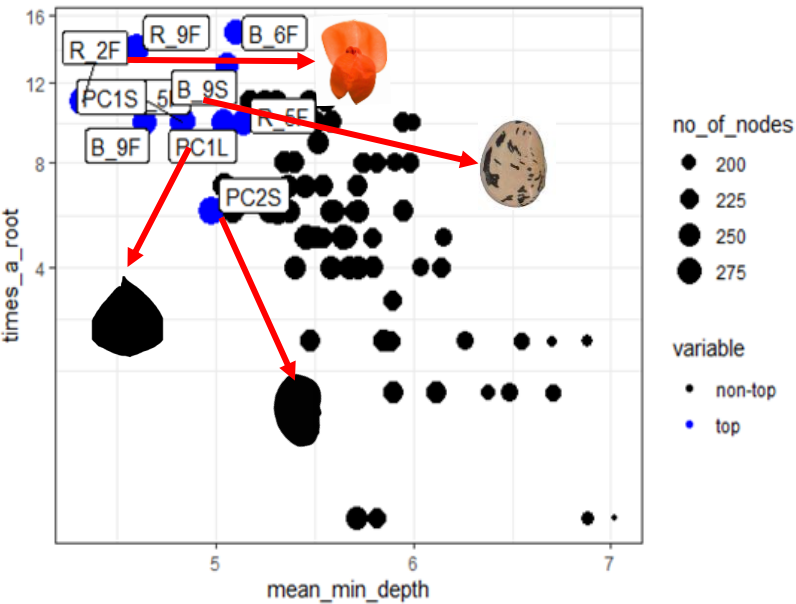


## Phenomic descriptors

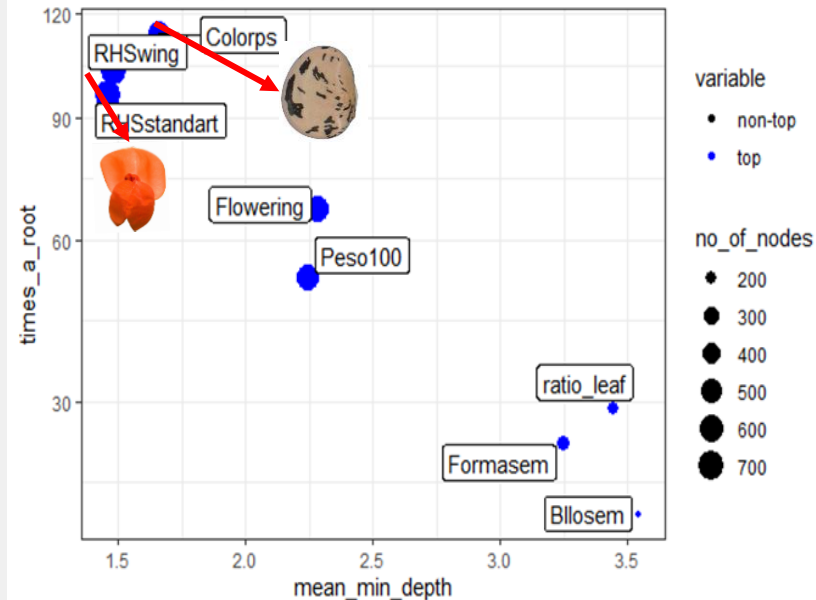
## Classic descriptors

## Phenomics + Classic

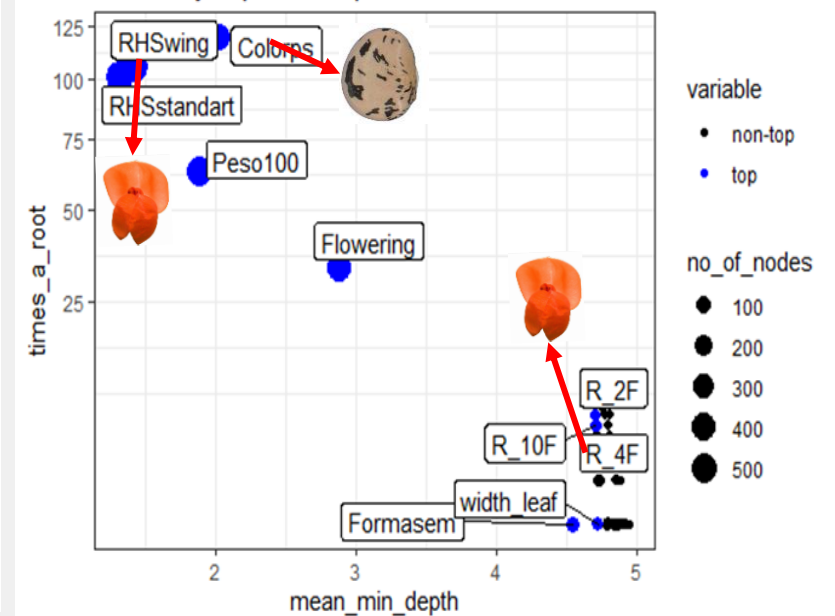
Multi-way importance plot



Multi-way importance plot



Multi-way importance plot



The descriptors in beans that show the greatest discrimination are flower and seed color. The classic flower color descriptor from the RHS color chart is most important.

# Accessions classification in peanut and beans



**A** Confusion Matrix

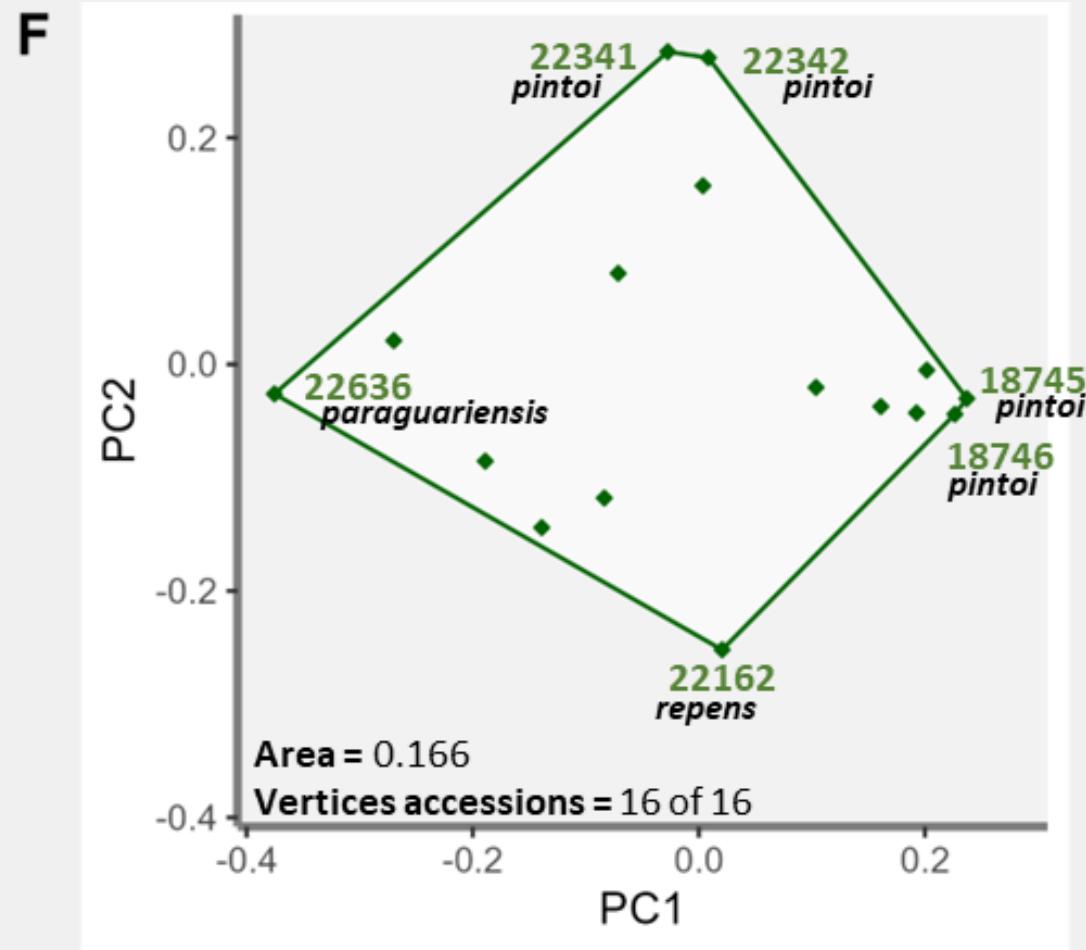
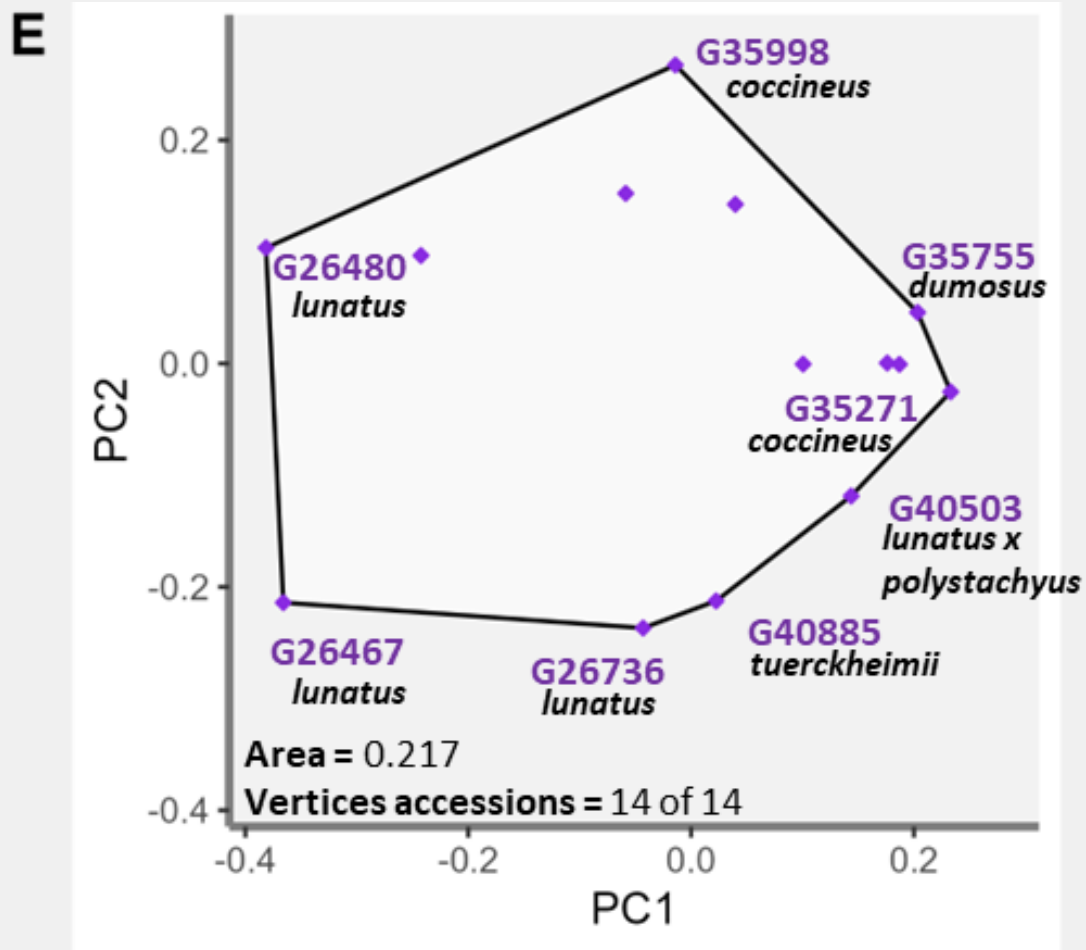
G24764B <i>xp. damasus</i> <i>x p. vulgaris</i>	2	0	0	0	2	0	1	0	0	0	0	0	1	0
G26467 <i>lanatus</i>	0	2	0	0	0	0	0	0	0	0	0	0	0	0
G26480 <i>lanatus</i>	0	2	4	0	0	0	0	0	0	0	0	0	0	0
G26736 <i>lanatus</i>	0	0	0	4	0	0	0	0	0	0	0	0	0	0
G35178 <i>damasus</i>	0	0	0	0	1	0	0	0	0	0	0	0	0	0
G35271 <i>coccineus</i>	0	0	0	0	0	4	0	0	0	0	0	0	0	0
G35580 <i>coccineus</i>	0	0	0	0	1	0	1	0	0	0	0	1	0	0
G35586 <i>damasus</i>	0	0	0	0	0	0	1	2	0	0	0	0	0	2
G35754A <i>coccineus</i>	0	0	0	0	0	0	0	1	4	0	0	2	0	0
G35755 <i>damasus</i>	0	0	0	0	0	0	1	0	0	3	0	0	0	0
G35998 <i>coccineus</i>	0	0	0	0	0	0	0	0	0	0	4	0	0	0
G36211 <i>coccineus</i>	0	0	0	0	0	0	0	0	0	1	0	1	0	0
G40503 <i>xp. lanatus x p.</i> <i>polysachyus</i>	0	0	0	0	0	0	0	0	0	0	0	0	2	1
G40885 <i>tuerckheimii</i>	2	0	0	0	0	0	0	1	0	0	0	0	1	1
	G24764B <i>xp. damasus</i> <i>x p. vulgaris</i>	G26467 <i>lanatus</i>	G26480 <i>lanatus</i>	G26736 <i>lanatus</i>	G35178 <i>damasus</i>	G35271 <i>coccineus</i>	G35580 <i>coccineus</i>	G35586 <i>damasus</i>	G35754A <i>coccineus</i>	G35755 <i>damasus</i>	G35998 <i>coccineus</i>	G36211 <i>coccineus</i>	G40503 <i>xp. lanatus x p.</i> <i>polysachyus</i>	G40885 <i>tuerckheimii</i>

**B** Confusion Matrix

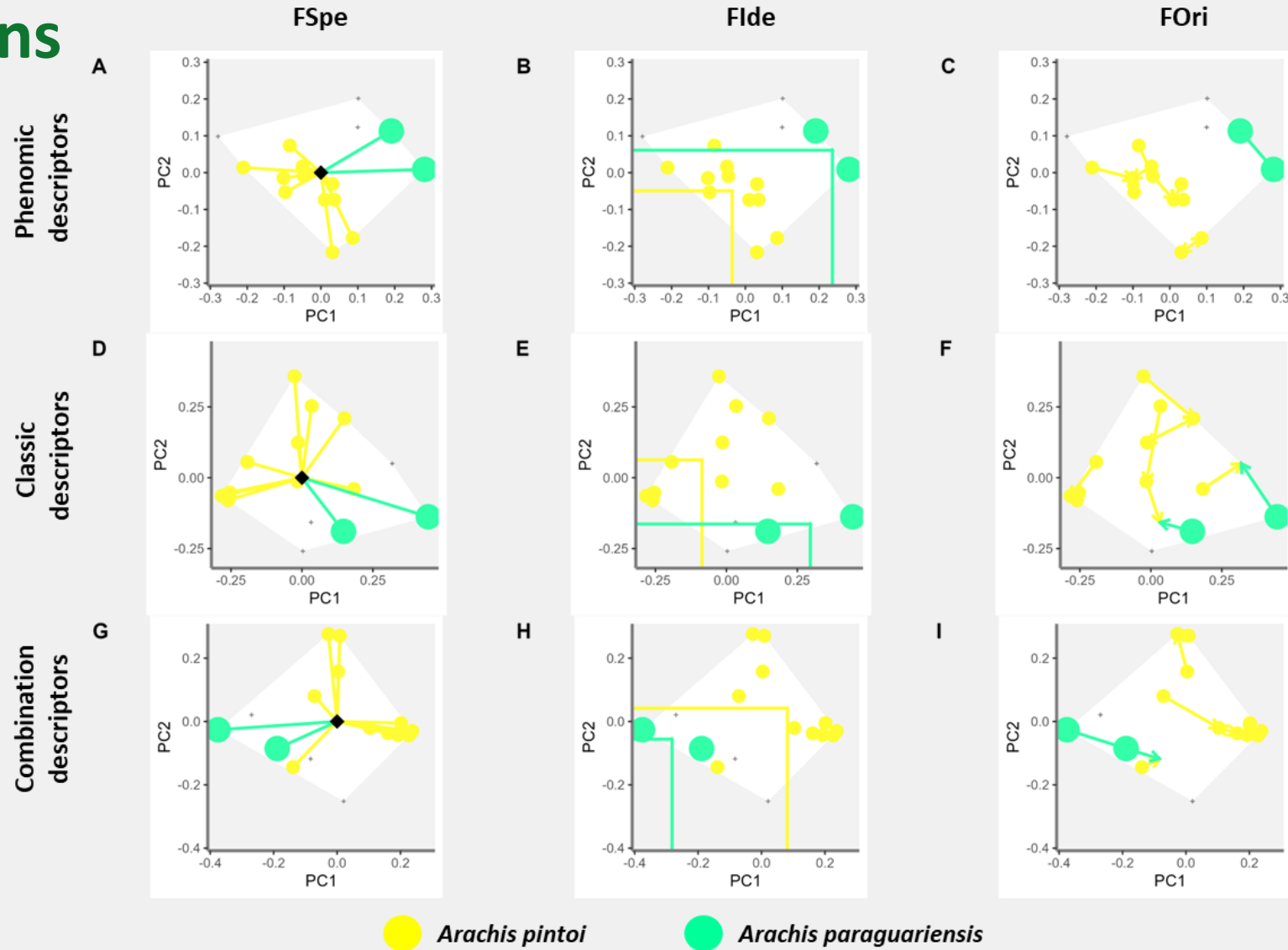
18745 <i>pinto</i>	3	0	1	0	1	0	0	0	0	0	0	0	0	0	0	
18746 <i>pinto</i>	1	4	0	1	1	0	0	0	1	0	0	0	0	0	0	
18748 <i>pinto</i>	0	0	3	0	0	0	0	0	0	0	0	0	1	0	0	
20826 <i>pinto</i>	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
22151 <i>pinto</i>	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	
22152 <i>pinto</i>	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	
22159 <i>pinto</i>	0	0	0	2	0	0	2	0	1	0	0	0	0	0	0	
22162 <i>repps</i>	0	0	0	0	0	0	1	4	0	0	0	0	0	0	0	
22176 <i>pinto</i>	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	
22226 <i>archer</i>	0	0	0	0	0	0	0	0	0	4	0	0	0	0	1	
22231 <i>pinto</i>	0	0	0	0	0	0	0	1	0	0	2	0	0	1	0	
22242 <i>sp</i>	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	
22341 <i>pinto</i>	0	0	0	0	1	2	0	0	0	0	0	0	1	1	0	
22342 <i>pinto</i>	0	0	0	0	0	0	0	0	0	0	2	0	1	1	0	
22633 <i>paraguensis</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	
22636 <i>paraguensis</i>	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	
	18745 <i>pinto</i>	18746 <i>pinto</i>	18748 <i>pinto</i>	20826 <i>pinto</i>	22151 <i>pinto</i>	22152 <i>pinto</i>	22159 <i>pinto</i>	22162 <i>repps</i>	22176 <i>pinto</i>	22226 <i>archer</i>	22231 <i>pinto</i>	22242 <i>sp</i>	22341 <i>pinto</i>	22342 <i>pinto</i>	22633 <i>paraguensis</i>	22636 <i>paraguensis</i>



# Functional diversity increases when the types of descriptors are combined



# Functional diversity indices drive selection of phenotypically key accessions



# Conclusions

The use of combined phenomic and conventional descriptors increases the accuracy in the classification of Phaseolus and Arachis species under our working conditions.

ML tools allow optimizing the rapid and efficient characterization of genebanks by selecting phenomic and conventional descriptors according to their importance.

The use of phenomic descriptors is likely to decrease the time and subjectivity of traits that are used in conventional descriptors while strengthening control points in genetic quality at the bank workflow.

The addition of images of flower, fruit and other phenomic descriptors taken in the characterization during the regeneration process helps to clarify taxonomic doubts and to make functional diversity analysis.



# Perspectives

In order to standardize image capture and processing, it is necessary to obtain computer infrastructure (storage and processing), photographic infrastructure (professional cameras and lighting systems) and remote sensors (multispectral and hyperspectral).

Currently, the genebank has personnel trained in the capture of images and characterization data; therefore, it is important to develop integrated processes for germplasm regeneration, conservation and health unit operations based on high-performance characterization strategies and to automate them as much as possible.

The capture and storage of images and spectral signatures can be part of digital characterization libraries that are associated with the genetic information of the digital bank, which in turn can be made public to users and have the validity to be used in association studies such as GWAS.





# Thanks

