



# The ImageCLEF 2013 Plant Identification Task

Goëau H., Joly A., Bonnet P.,  
Boujema N., Barthelemy D. and Molino J.-f.

# The ImageCLEF 2013 Plant Identification Task

## ❖ Introduction

- ❖ Task resources & description
- ❖ Participants and results
  - SheetAsBackground
  - NaturalBackground
  - Details by views
- ❖ Conclusion

# Context & challenges

Biodiversity erosion & Global warming affects the **environment** as well as **agriculture** and **food security**

Accurate knowledge of **plants** (distribution and ecology) is essential for **sustainable agriculture** and **biodiversity conservation**

But accessing basic information about plants is still challenging



**Botanical data** is:

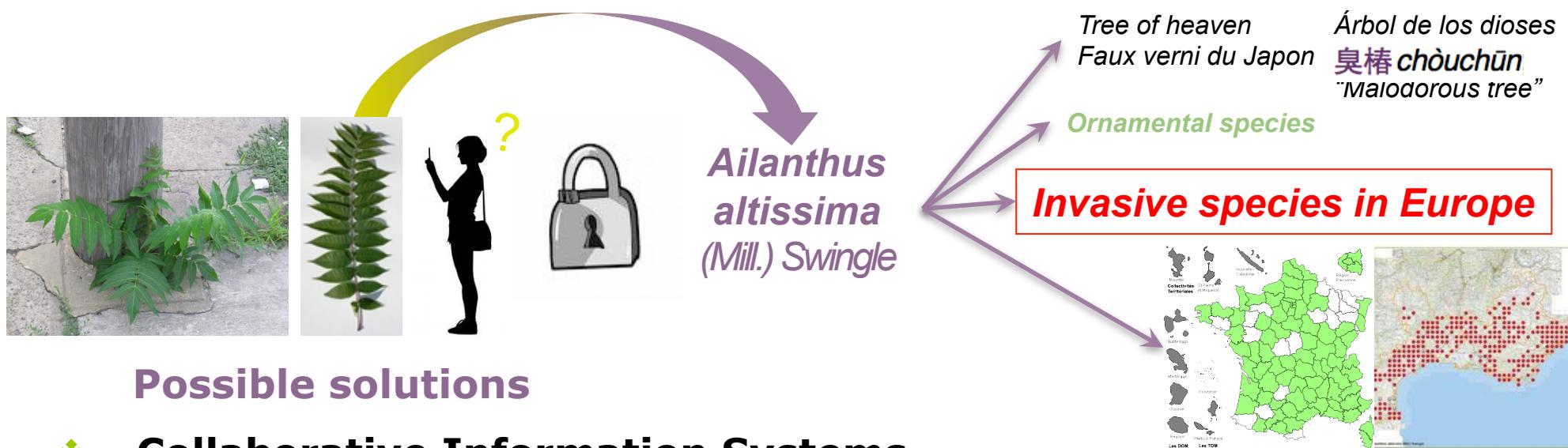
- ❖ *decentralized and heterogeneous*
- ❖ *complex (un-structured tags, empirical measurements,...)*
- ❖ *sparse and incomplete*
  - *huge & unknown number of species (300K ?)*
  - *"long tail distribution" (1 record per species !)*



# Towards bridging the taxonomic gap

As a consequence, **identifying plants is very difficult**

-> How to control plant's distribution and plant ecology ?



# The ImageCLEF 2013 Plant Identification Task

- ❖ Introduction

## ❖ **Task resources & description**

- ❖ Participants and results

- SheetAsBackground
  - NaturalBackground
  - Details by views

- ❖ Conclusion

# Task characteristics & Data creation

An **attractive** task as **simple** as possible but as **realistic** as possible

- ❖ *Simple* : 1 media type (image),  
2 mains categories *SheetAsBackground ; NaturalBackground*
- ❖ *Realistic* : collaborative data, numerous contributors

how to reduce bias between training data and real user's data ?



Let real users collect training data and botanists validate



Grow training data with an online Identification and Validation Tool

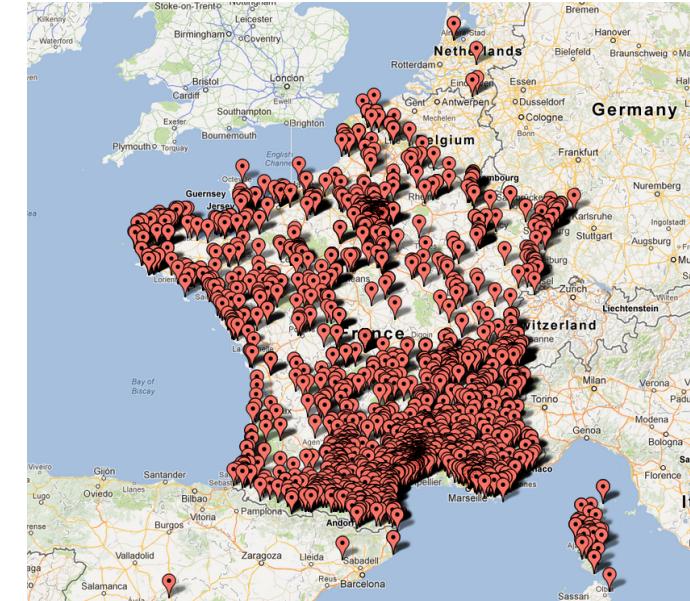
=**The PlantViews dataset**

A screenshot of the PlantViews dataset interface. It shows a field worker in the background and a digital interface in the foreground. The interface displays a list of leaf species with their names, families, and percentages: Acer campestre (20%), Acer negundo (15%), Acer pseudoplatanus (5%), Acer monspessulanum, and Acer opulus. A detailed view of a maple leaf is shown on the right. The interface includes a search bar, filters for rank, family, and genus, and legal notices at the bottom.

# Pl@ntViews dataset



	2011	2012	2013
Species NB	71	126	250
Images NB	5 400	11 500	26 077 (327 contributors)
View types	Leaves	Leaves	



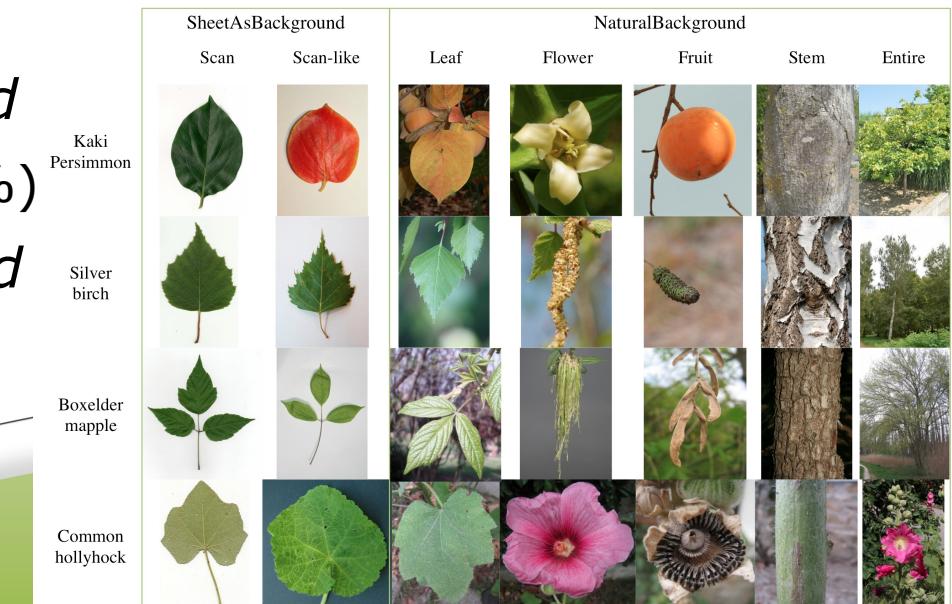
## ❖ 2 mains categories of images

### ❖ 11 031 img *sheetAsBackground*

- Leaf scan & scan-like (42%)

### ❖ 15 046 img *NaturalBackground*

- Leaf (16%)
- Flower (18%)
- Fruit (8%)
- Stem (8%)
- Entire plant (8%)



# Pl@ntViews dataset



## ❖ Metadata (XML)

- IndividualPlantID
- Data and time
- Image type “NaturalBackground” or not
- Content (Leaf, Flower, Fruit, etc.)
- Full taxon (*APGIII*)
- ClassID Species identifier
- Common name
- Author name
- Locality name
- GPS
- ...



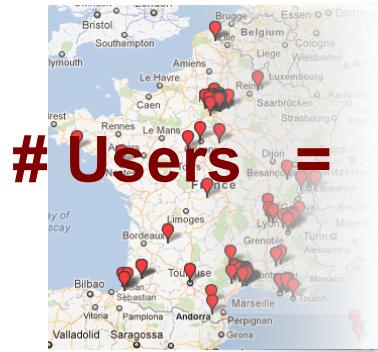
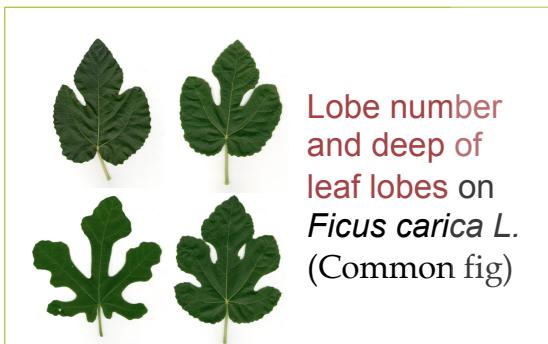
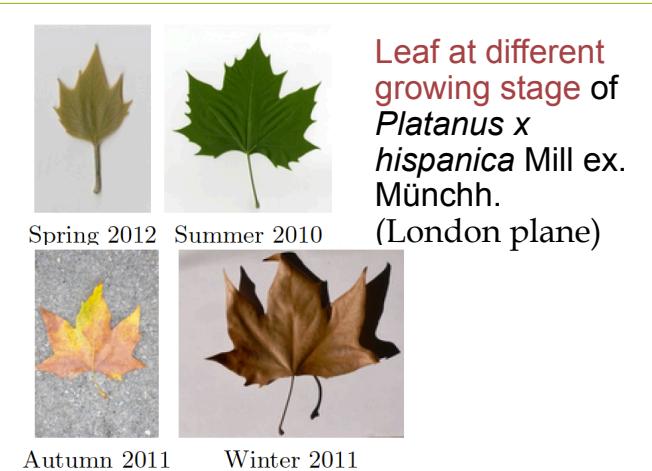
```
<?xml version="1.0" encoding="UTF-8"?>
<Image>
  <FileName>5438.jpg</FileName>
  <Date>03/11/10</Date>
  <Type>pseudoscan</Type>
  <Author>Pierre Bonnet</Author>
  <Organization>Inra</Organization>
  <Content>Leaf</Content>
  <IndividualPlantId>215</IndividualPlantId>
  <Taxon>
    <Regnum>Plantae</Regnum>
    <Class>Equisetopsida C. Agardh</Class>
    <Subclass>Ginkgoideae Engl.</Subclass>
    <Superorder>Ginkgoideae Engl.</Superorder>
    <Order>Ginkgoales Gorozh.</Order>
    <Family>Ginkgoaceae Engl.</Family>
    <Genus>Ginkgo L.</Genus>
    <Species>Ginkgo biloba L.</Species>
  </Taxon>
  <ClassId>Ginkgo biloba</ClassId>
  <VernacularNames>Ginkgo</VernacularNames>
  <Locality>France, Montpellier</Locality>
  <GPSLocality>
    <Longitude>3.8794472217559814</Longitude>
    <Latitude>43.60817337036133</Latitude>
  </GPSLocality>
</Image>
```

## ❖ A unique dataset !

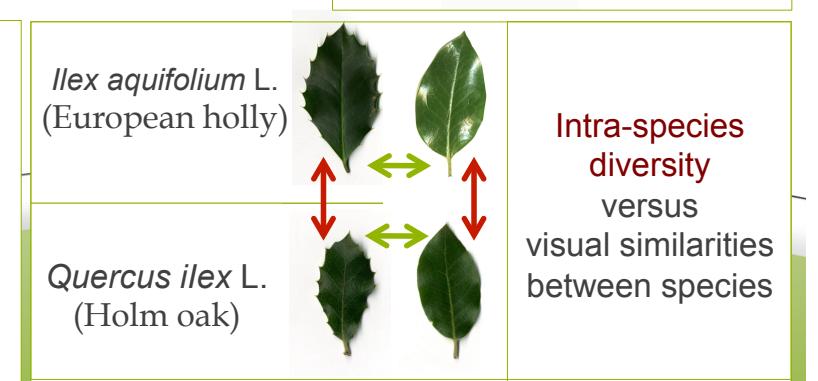
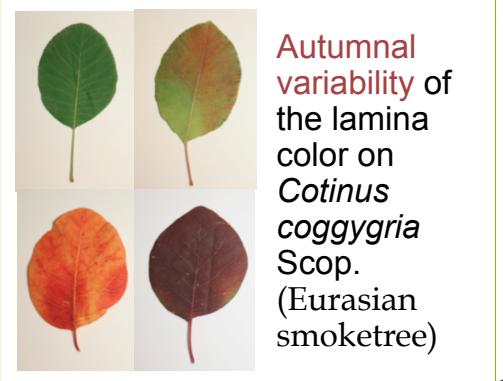
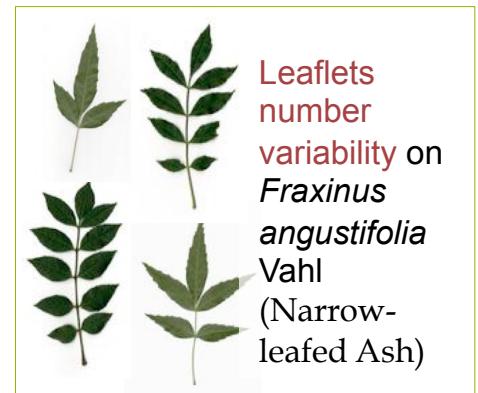
- Pictures of views of species, coming from **different individual plants**
- **Hundreds of contributors**, with numerous devices
- Pictures at **different periods** of the year (over 3 years)
- **Taxonomic validation** by a network of botanists

# PI@ntViews dataset

## Leaf diversity



# localities  
# seasons  
# environments  
# climate  
# ecosystems  
# devices



# PI@ntViews dataset

## *Flower diversity*



### COLOR



Brown



White



Green



Rose



Blue



Yellow

### Symmetry



Radial



Bilateral

### Structure



4



5



6



>>6

Number of petals

### Orientation



Face



Profil

### Size



small



middle



big

# PI@ntViews dataset

*Fruit, Stem and Entire plant*



## Fruit types



Achene



Berry



Capsule



Cone



Drupe



Folicle



Legume



Samara



Silique

With an important diversity of colors, shape, texture, orientation, etc.



## Stem types



very young



young



adult



old

The visual diversity of the bark of the *Robinia pseudoacacia*.

## Entire plant



From large trees to small herbs

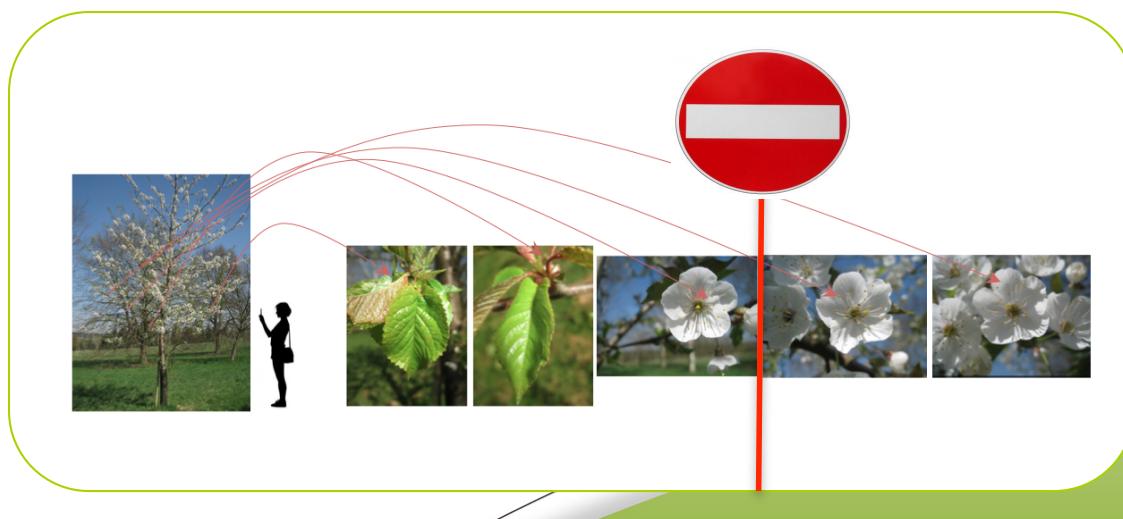
# Task description

## ❖ Retrieval task

- For each test picture, a list of ranked species
- Separate scores for the 2 mains categories
- Free training strategy

		Images	Plants	Authors	Species
SheetAsBackground	Train	9781	732	36	126
	Test	1250	150	14	70
NaturalBackground	Train	11204	2553	176	244
	Test	3842	2454	229	238
Entire	Train	1455	955	104	234
	Test	694	567	107	177
Flower	Train	3521	1328	127	233
	Test	1233	970	142	203
Fruit	Train	1387	512	64	156
	Test	520	302	77	103
Leaf	Train	13285	1046	73	210
	Test	2040	420	68	143
Stem	Train	1337	629	38	131
	Test	605	408	35	77
All	Train	20985	11204	176	250
	Test	5092	3842	229	241

## ❖ Plant-based splits:



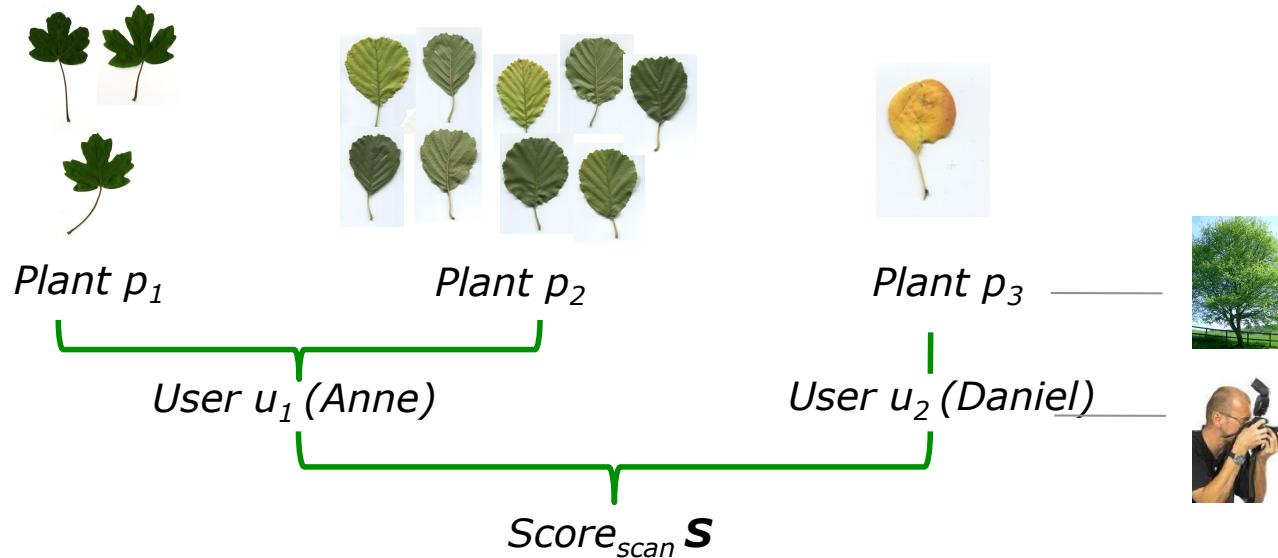
A plant centred approach  
& a multiple organ queries

Can't be split  
all images in training or test dataset

# Task description

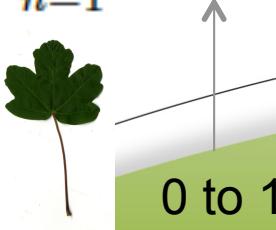
## Score

- ❖ **Unbalanced** real-world data



- ❖ **Normalized** Average Score

$$S = \frac{1}{U} \sum_{u=1}^U \frac{1}{P_u} \sum_{p=1}^{P_u} \frac{1}{N_{u,p}} \sum_{n=1}^{N_{u,p}} s_{u,p,n}$$



0 to 1

$$S_{u,p,n} = 1 / \text{RankOfCorrectSpecies}$$

*To not give too much importance to individual plants with numerous images*

*To not give too much importance to users who contribute a lot*

*« As a new user of the plant identification system, what is the score I can expect? »*

$$\neq \sum_{\text{all test pictures}} S_{u,p,n}$$

# The ImageCLEF 2013 Plant Identification Task

- ❖ Introduction
- ❖ Task resources & description

## ❖ **Participants and results**

- SheetAsBackground
- NaturalBackground
- Details by views

- ❖ Conclusion

# Participation and Methods

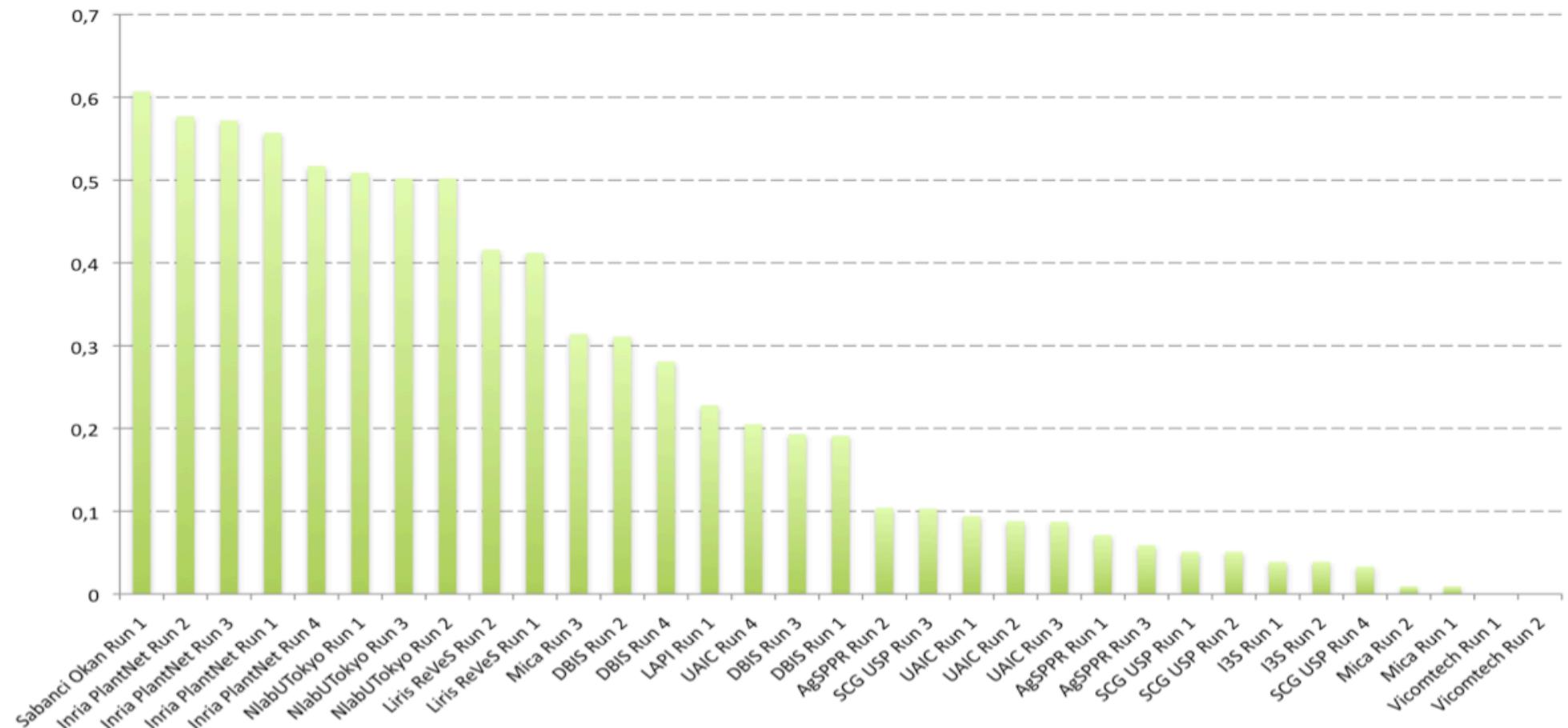
2011 : 8 teams / 20 methods  
 2012 : 10 teams / 30 methods  
 2013 : 12 teams / 33 methods

Groups	Runs	Features	Prediction based on
 AGSPPR	3	SPACT, SIFT, global shape	SVM
 BTU DBIS	4	Global color histograms EOH, Tamura, CEDD, FCTH.	SVM
 I3S	2	SIFT + BOW	SVM
 INRIA	4	Shape: Triangular, Directional Fragment, Context Corners, LBP/SURF/Fourier/EOH/HSV/wRGB, date	Matching + nn SVM
 LAPI	1	Curve partitioning	Linear Discriminate Analysis
 LIRIS ReVes	2	Lab/Gabor/SURF Hu/Zernike/centered moments, geoloc.	Naïve distance based
 MICA	3	GIST/SURF + BOW	SVM
 Nlab	3	SIFT variation + FisherVector	Linear Logistic Regression
 SABANCI-OKAN	1	Numerous shape and texture desc. wHSV, date.	SVM
 SCG USP	3	LBP, Fractal, Gabor, geometrical	Linear Discriminate Analysis
 UIAC	4	Joint composite descriptor, geoloc.	Nn rules, naïve Bayesian
 VicomTech	2	Trace transform, Shape relationship	SVM

# Results - Key to succeed

## SheetAsBackground

Sabanci-Okan, Turkey  
Inria PlantNet, France  
NlabUTokyo, Japan

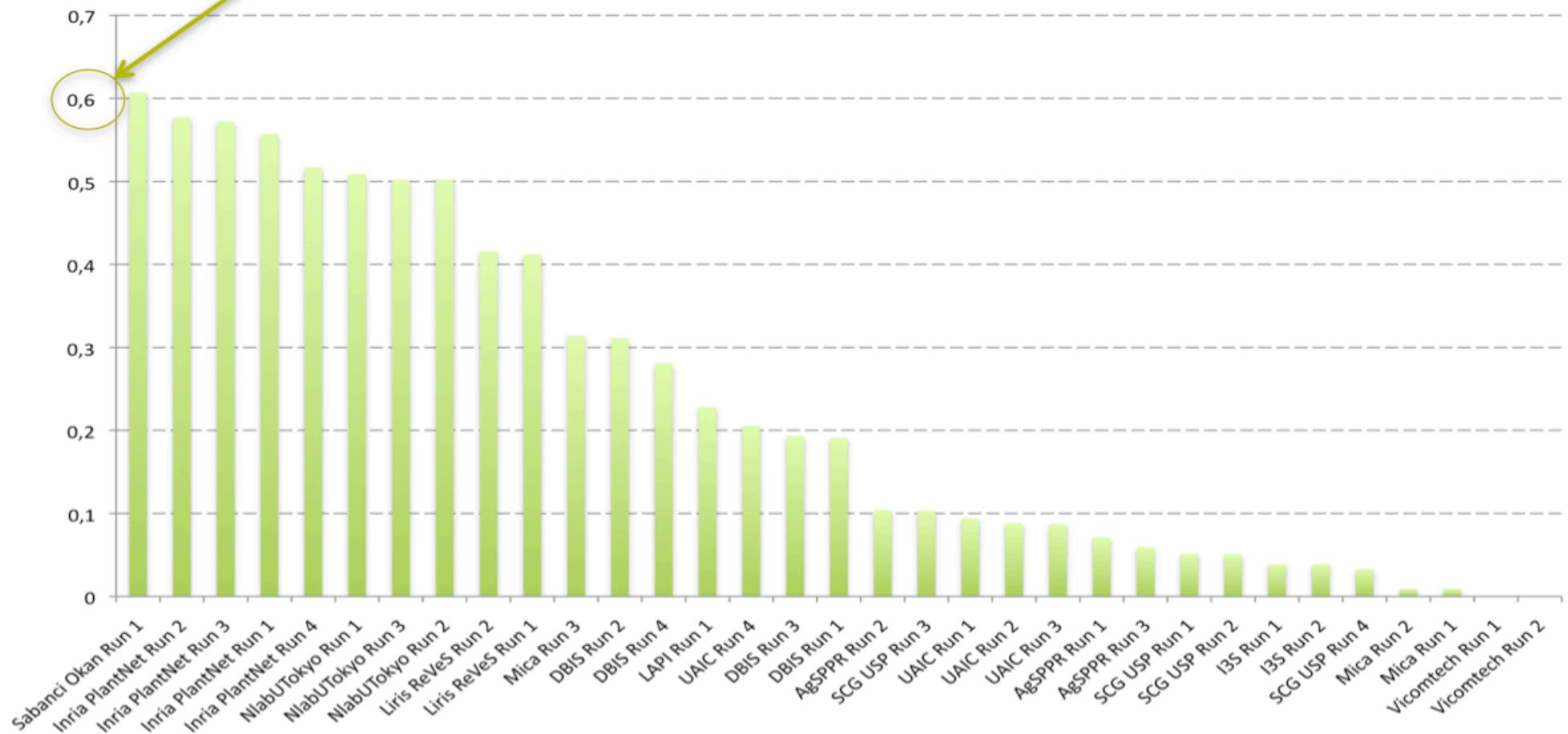


# Results - Key to succeed

## SheetAsBackground

*Improvement comparing  
to 2012 task in spite of*  

- more species,
- more complex pictures



**Sabanci-Ökan, Turkey**  
**Inria PlantNet, France**  
**NlabUTokyo, Japan**

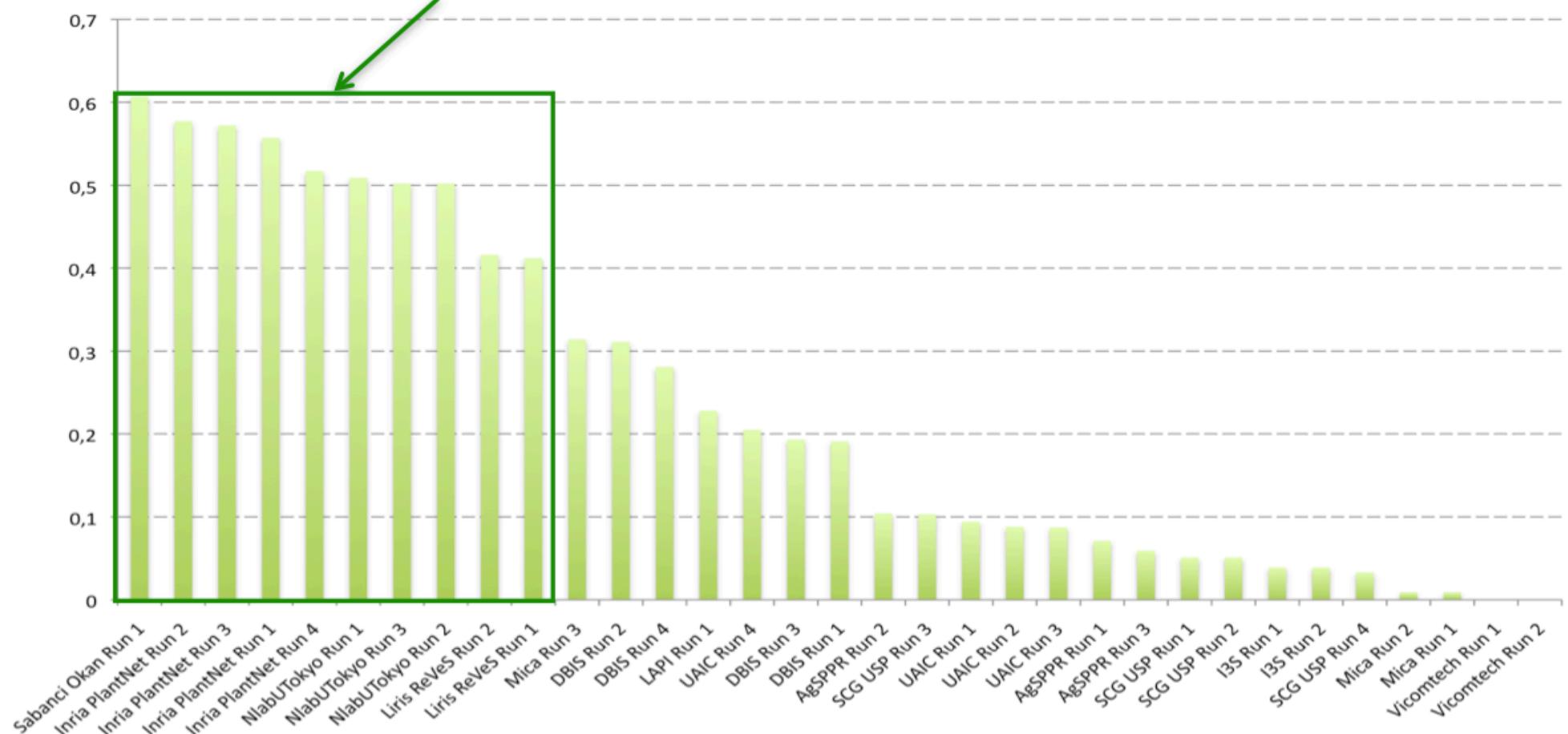
# Results - Key to succeed

## SheetAsBackground



*Individual-plants not split during preliminary evaluation:*  
- better choice of features  
- avoid over-fitting problems

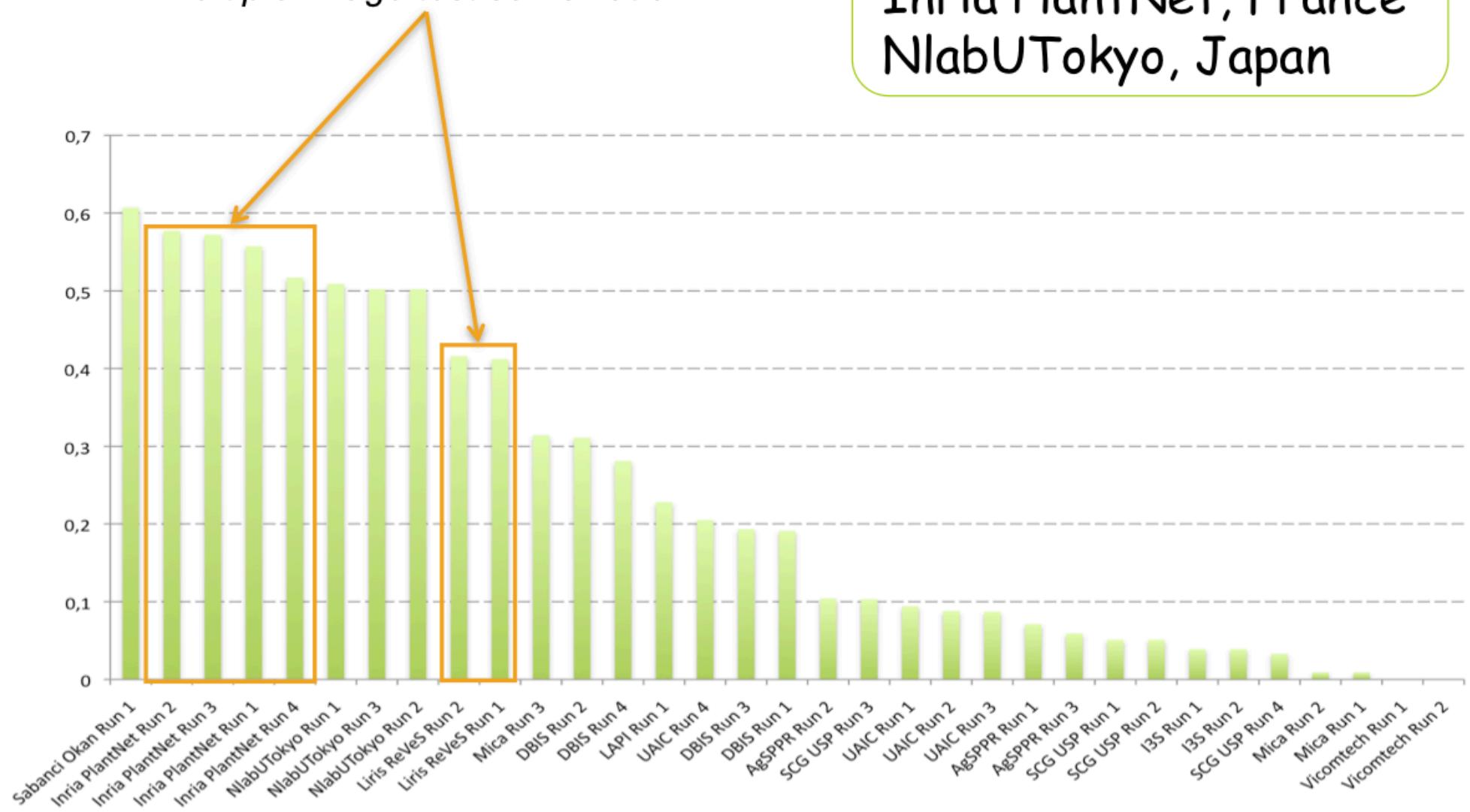
Sabanci-Okan, Turkey  
Inria PlantNet, France  
NlabUTokyo, Japan



# Results - Key to succeed

## *SheetAsBackground*

*Multiple-image test combination*

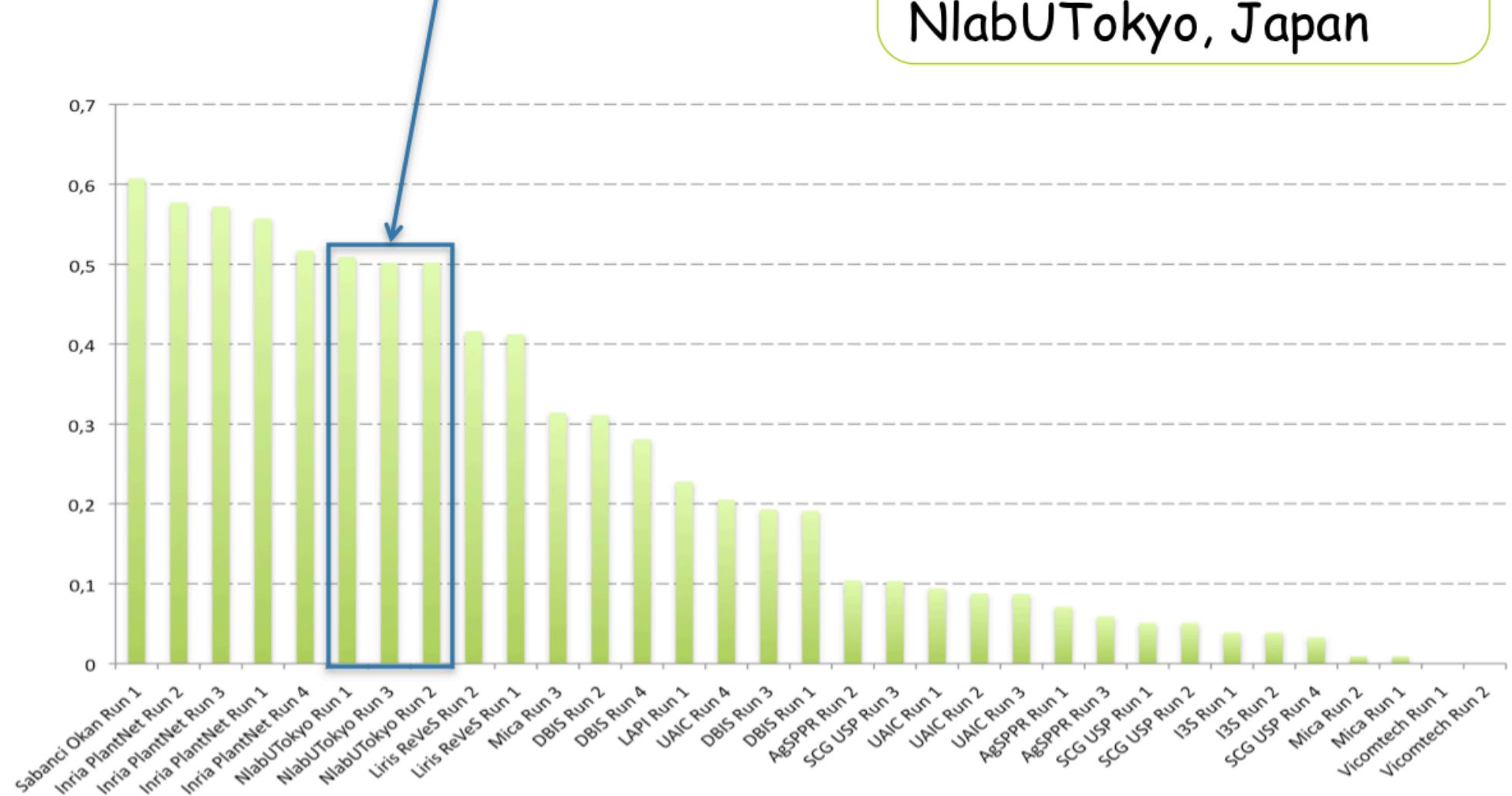


**Sabanci-Ökan, Turkey**  
Inria PlantNet, France  
NlabUTokyo, Japan

# Results - Key to succeed

## *SheetAsBackground*

Non-shape-based approach

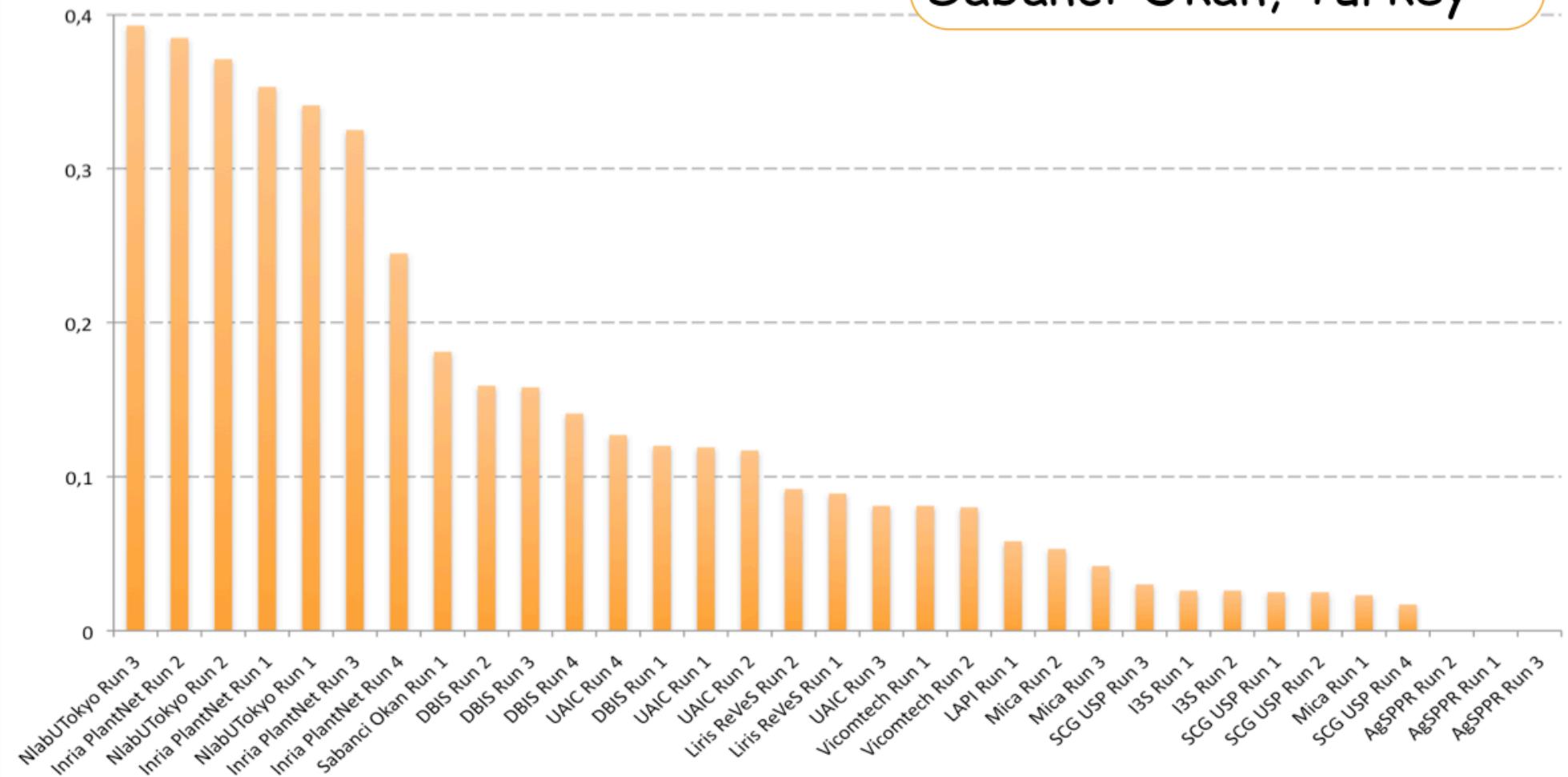


**Sabanci-Ökan, Turkey**  
Inria PlantNet, France  
NlabUTokyo, Japan

# Results - Key to succeed

## NaturalBackground

NlabUTokyo, Japan  
Inria PlantNet, France  
Sabanci-Okan, Turkey

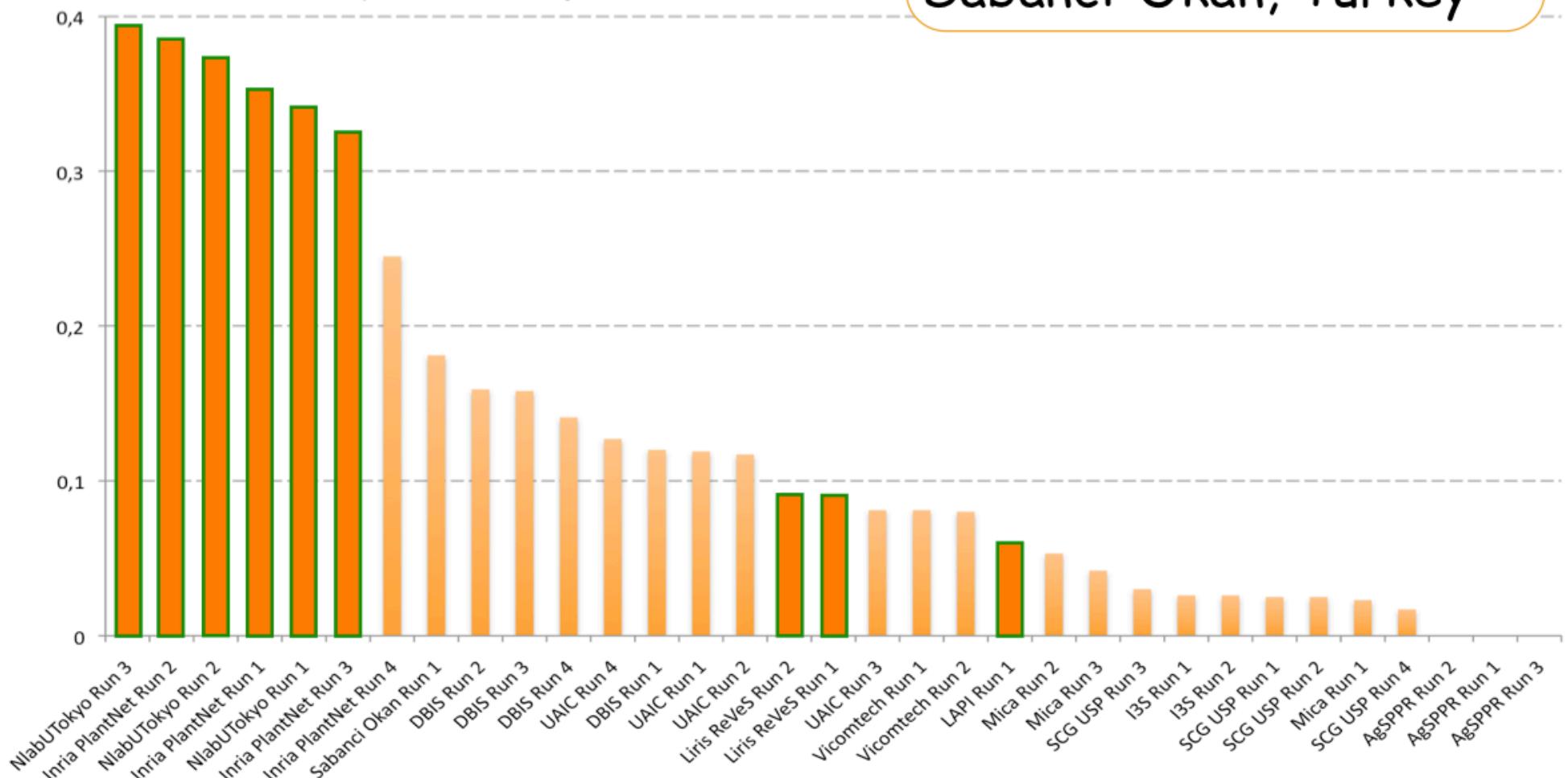


# Results - Key to succeed

## NaturalBackground



*Individual-plants not split  
during preliminary evaluation*



**NlabUTokyo, Japan**

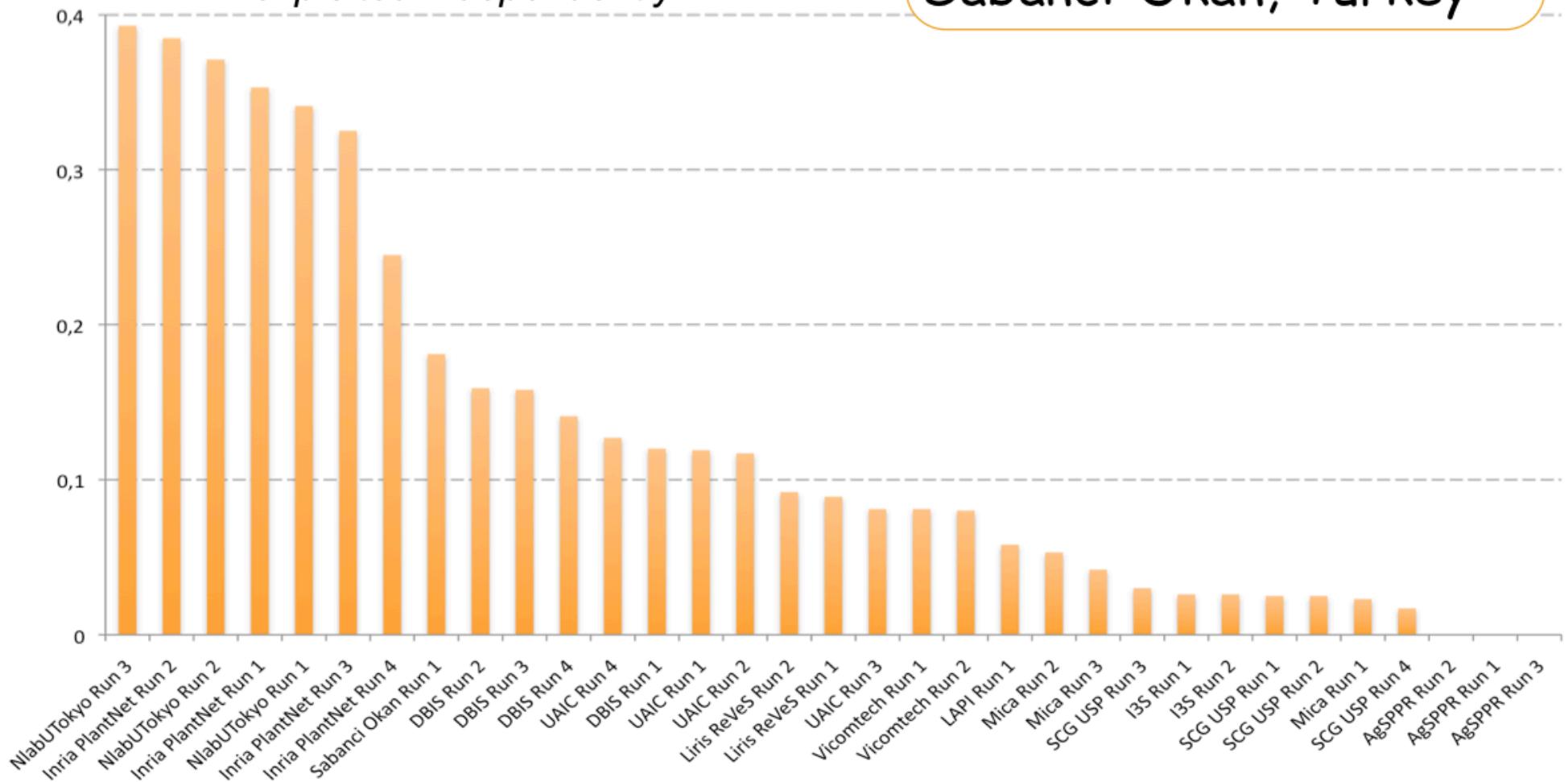
**Inria PlantNet, France**  
**Sabanci-Okan, Turkey**

# Results - Key to succeed

## NaturalBackground



*Subcategories generally exploited independently...*



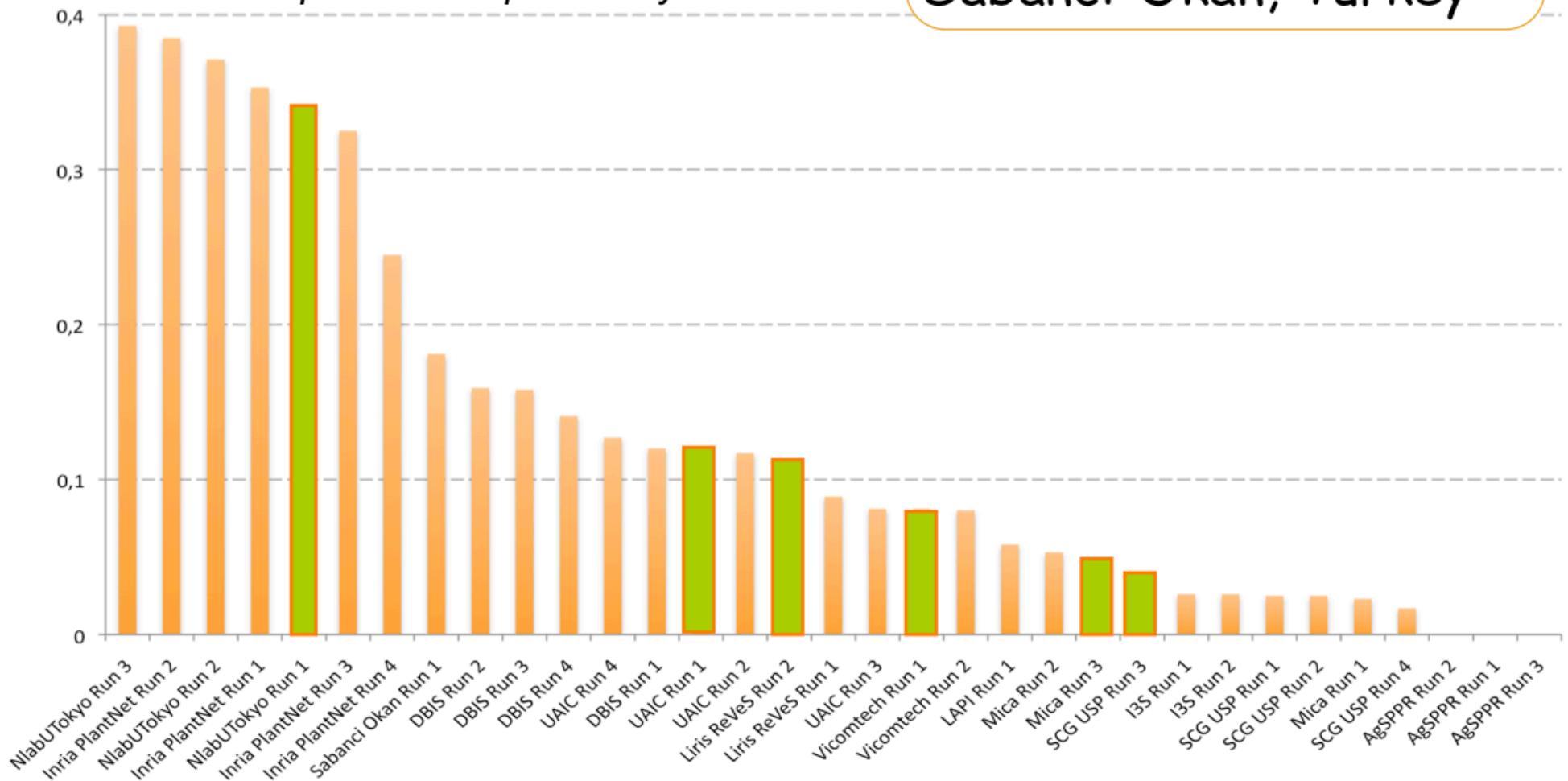
**NlabUTokyo, Japan**

**Inria PlantNet, France**  
**Sabanci-Okan, Turkey**

# Results - Key to succeed

## NaturalBackground

.... Subcategories not  
exploited independently.

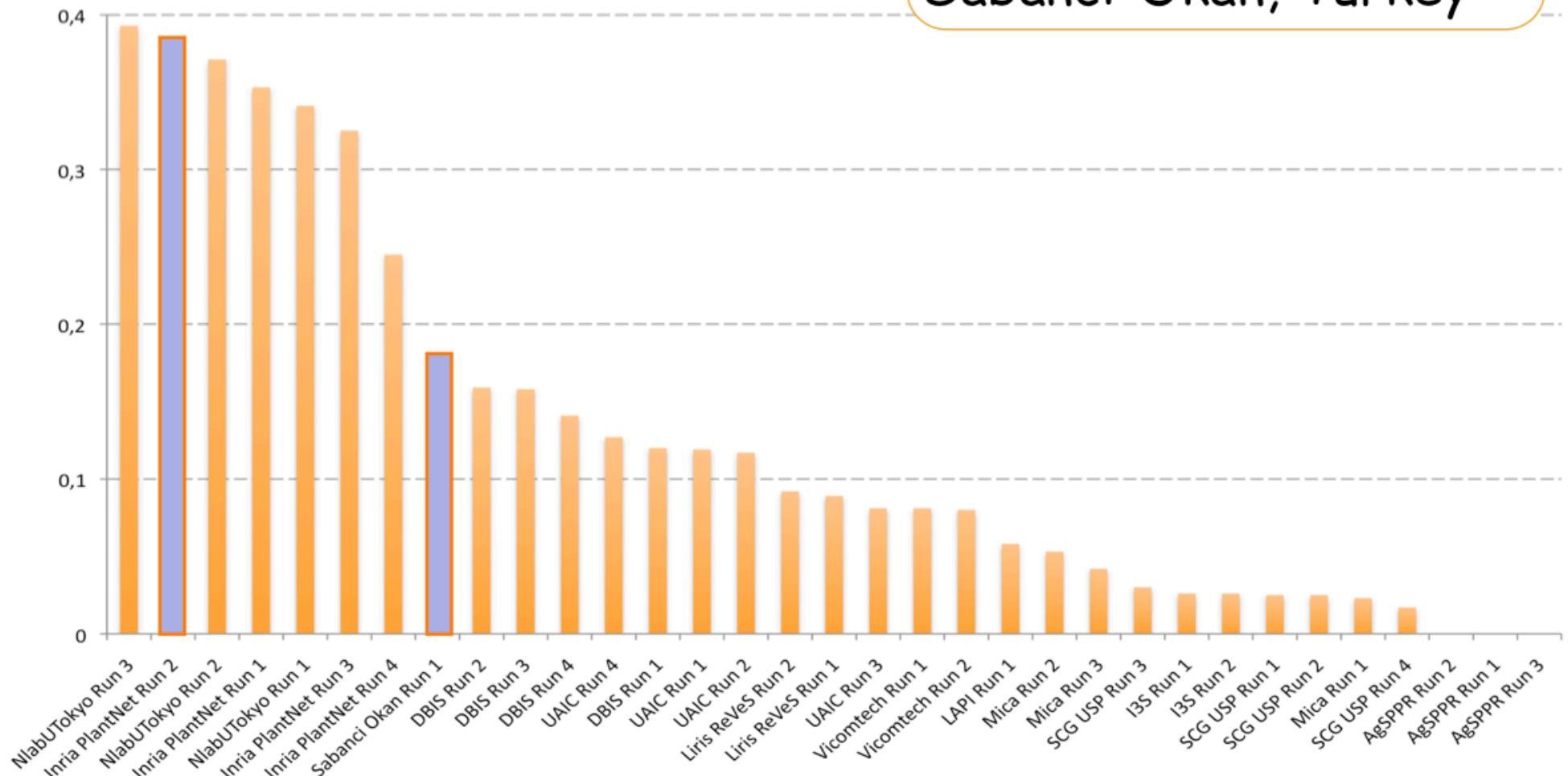


NlabUTokyo, Japan  
Inria PlantNet, France  
Sabanci-Okan, Turkey

# Results - Key to succeed

## NaturalBackground

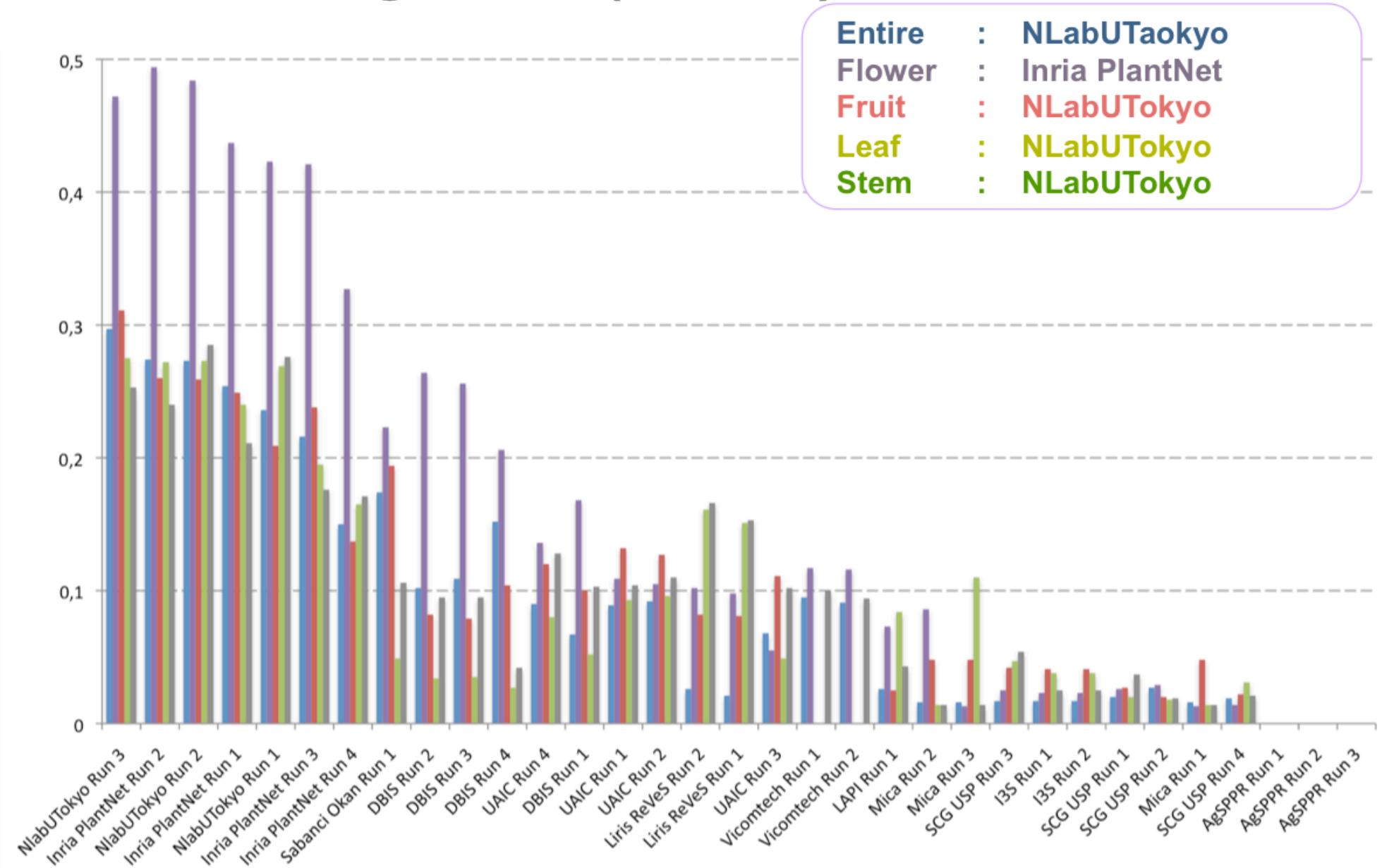
Use of date



NlabUTokyo, Japan  
Inria PlantNet, France  
Sabanci-Okan, Turkey

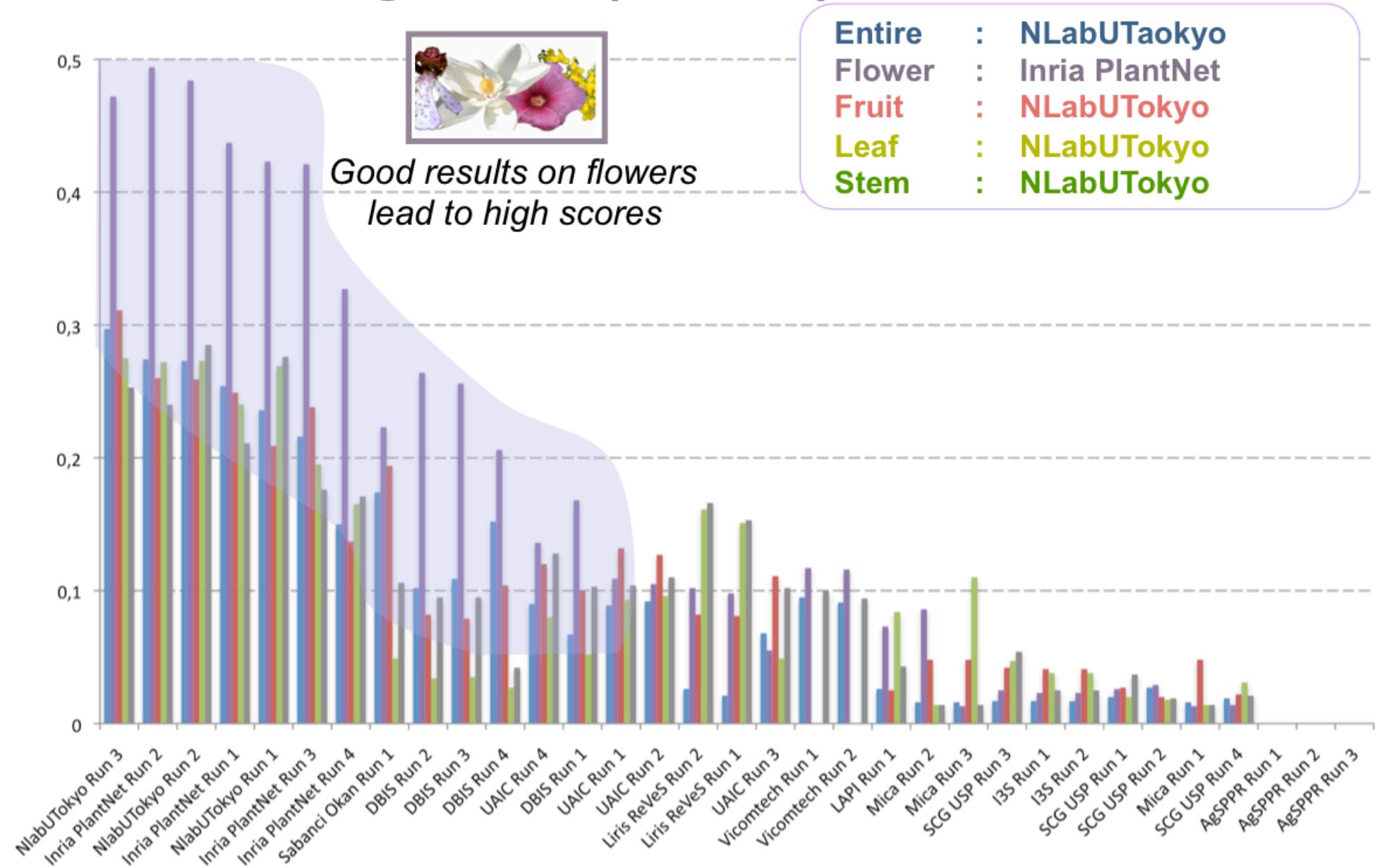
# Results – Complementary remarks

## NaturalBackground (details)



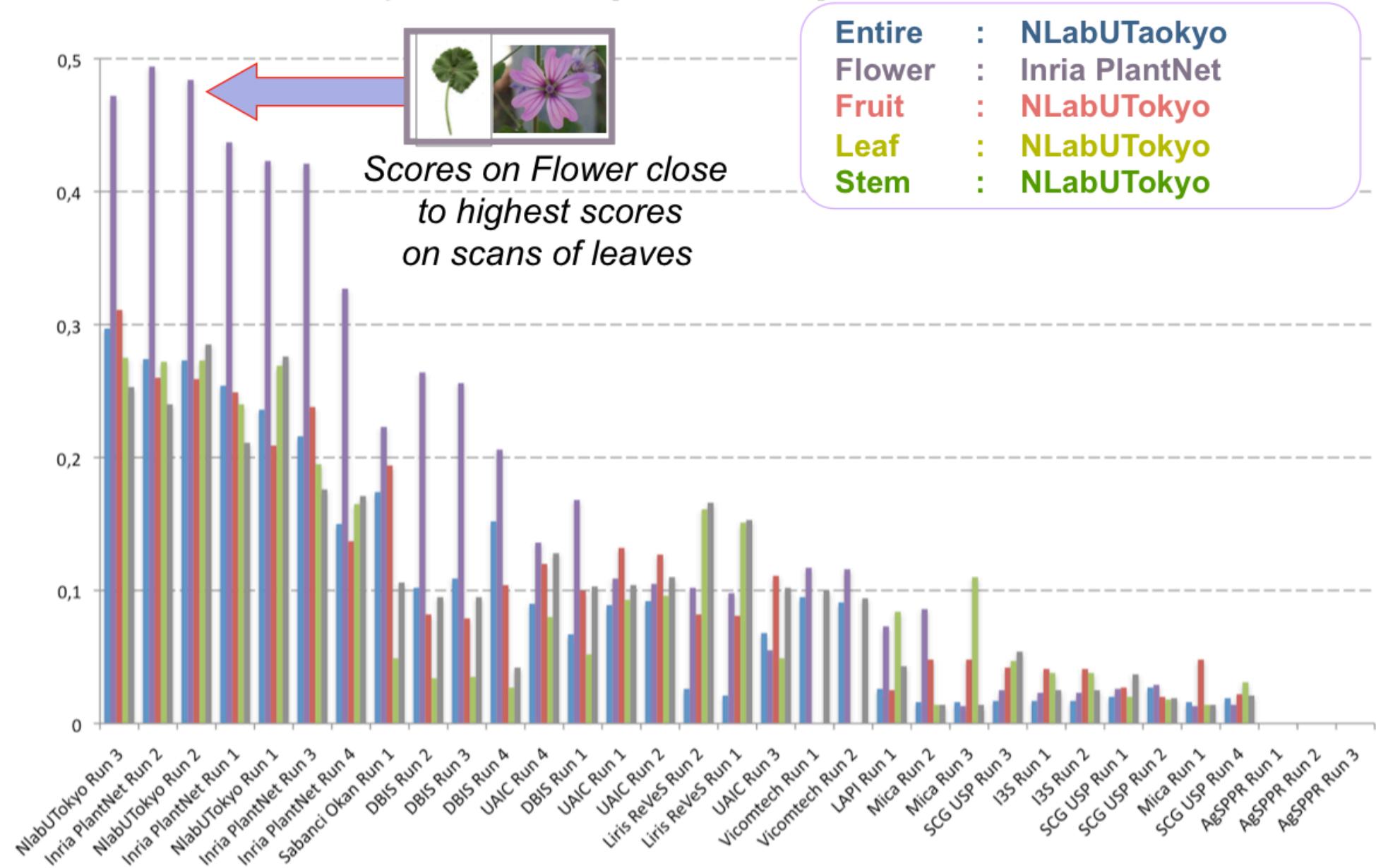
# Results – Complementary remarks

## NaturalBackground (details)



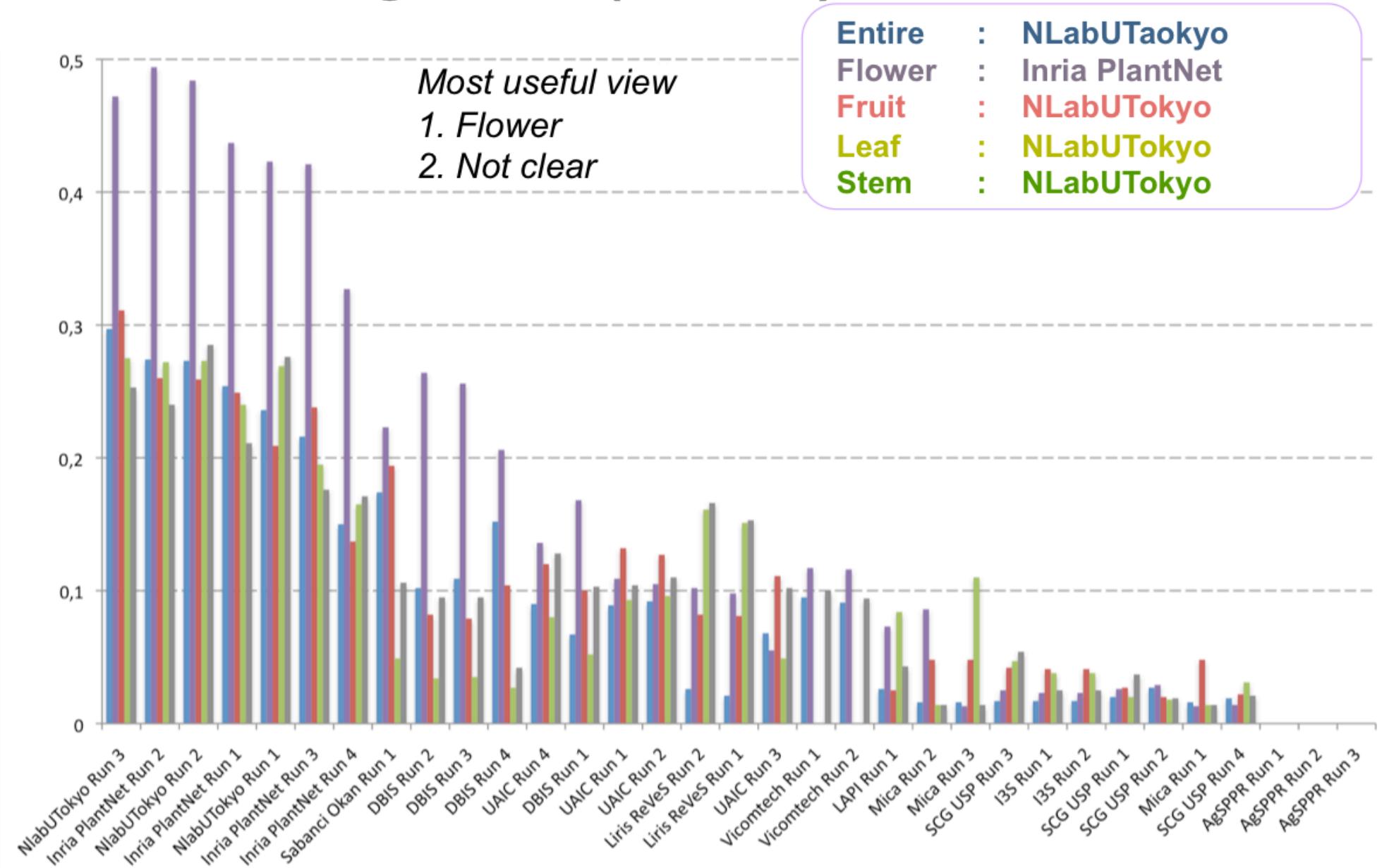
# Results – Complementary remarks

## NaturalBackground (details)



# Results – Complementary remarks

## NaturalBackground (details)



# The ImageCLEF 2013 Plant Identification Task

- ❖ Introduction
- ❖ Task resources & description
- ❖ Participants and results
  - SheetAsBackground
  - NaturalBackground
  - Details by views

## ❖ Conclusion

# Conclusion

- ❖ **Good participation** → Increasing interest of the community
- ❖ Still **good results** on *Leaf-SheetAsBackground*  
Performances with a realistic number of species (3000 to 5000 sp.) ?
- ❖ Real **difference** between  
*Leaf-SheetAsBackground* and *Leaf-NaturalBackground*
- ❖ Challenging unconstrained pictures  
Very **promising results** on *NaturalBackground* (mainly for flowers)
- ❖ No method above the others for the 2 categories
- ❖ Training strategy is essential
- ❖ Metadata:
  - + Impact of dates
  - Not clear GPS impact
  - Still unused data (taxonomic context, common names, Exif, ...)

# Perspectives and issues

- ❖ Task evolution through a new Lab, close to ImageCLEF, dedicated to environmental data management
- ❖ Growing task, with
  - More data, depending on collaborative contributions
  - More species (up to 500 or 1 000 sp.)
  - More content types (branch images)
  - More metadata (data quality evaluation, ...)
  - More multi-modal information  
(multi-lingual species names, species distribution, species description, ...)



A scenic view of a botanical garden, likely Kirstenbosch National Botanical Garden in Cape Town, South Africa. The foreground is filled with various succulent plants, including large aloe vera-like species and smaller, colorful flowers. In the middle ground, there are more diverse plantings and a few people walking through the gardens. A large, rugged mountain rises in the background under a clear blue sky.

Thank you!!

