

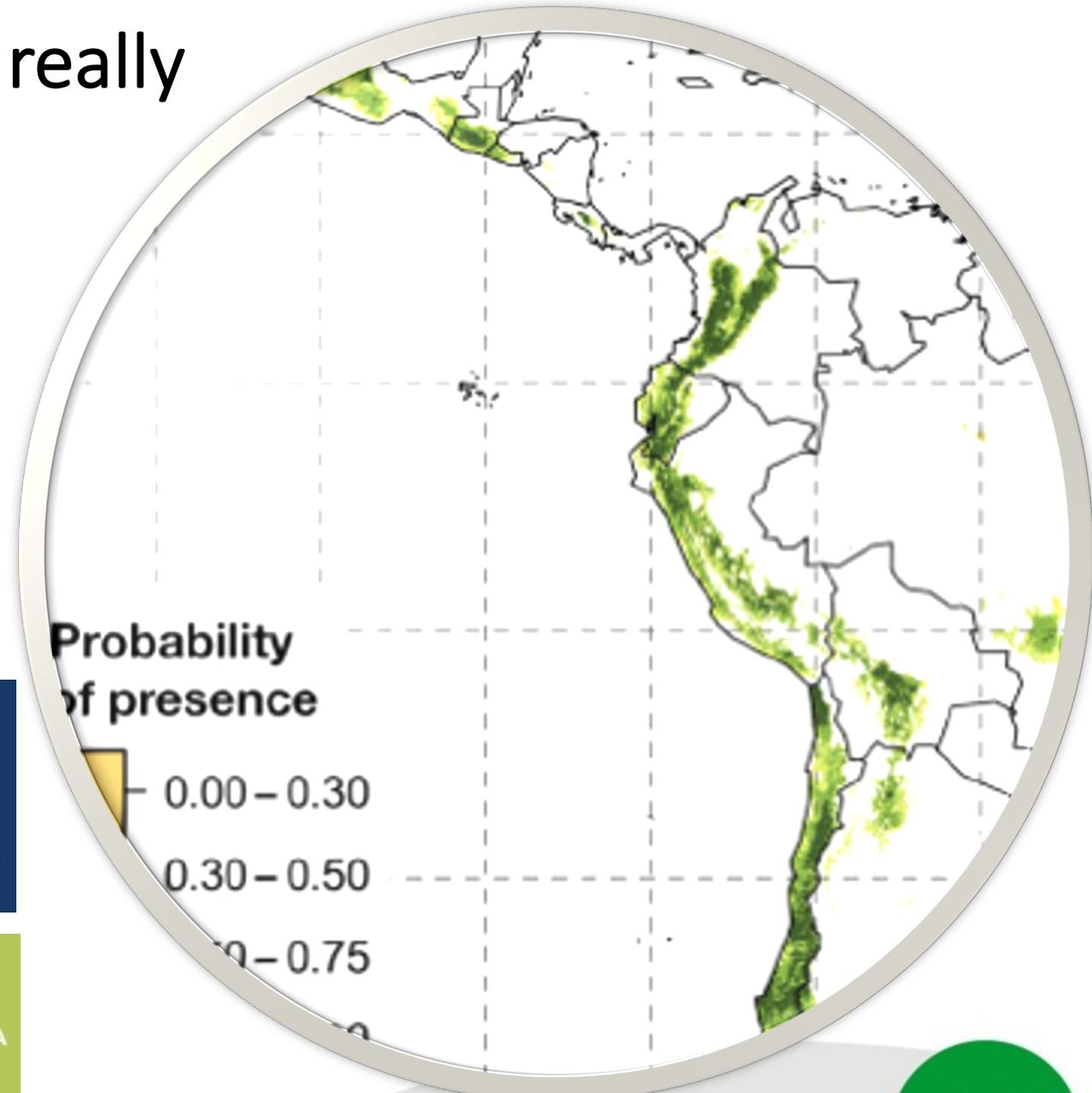
# How much genetic diversity is really still missing from genebanks?

A review of ~15 years of work on gap analysis

**Prof. Dr. Julian Ramirez-Villegas**

Principal Scientist, CGIAR

Professor, Wageningen University & Research



Building a sustainable future



**2020**

**maintain genetic diversity of seeds, cultivated plants, farmed and domesticated animals and their related wild species,**

**through**

**soundly managed and diversified seed and plant banks at national, regional and international levels,**

**and**

**and ensure access to and fair and equitable sharing of benefits arising from the utilization of genetic resources and associated traditional knowledge as internationally agreed**

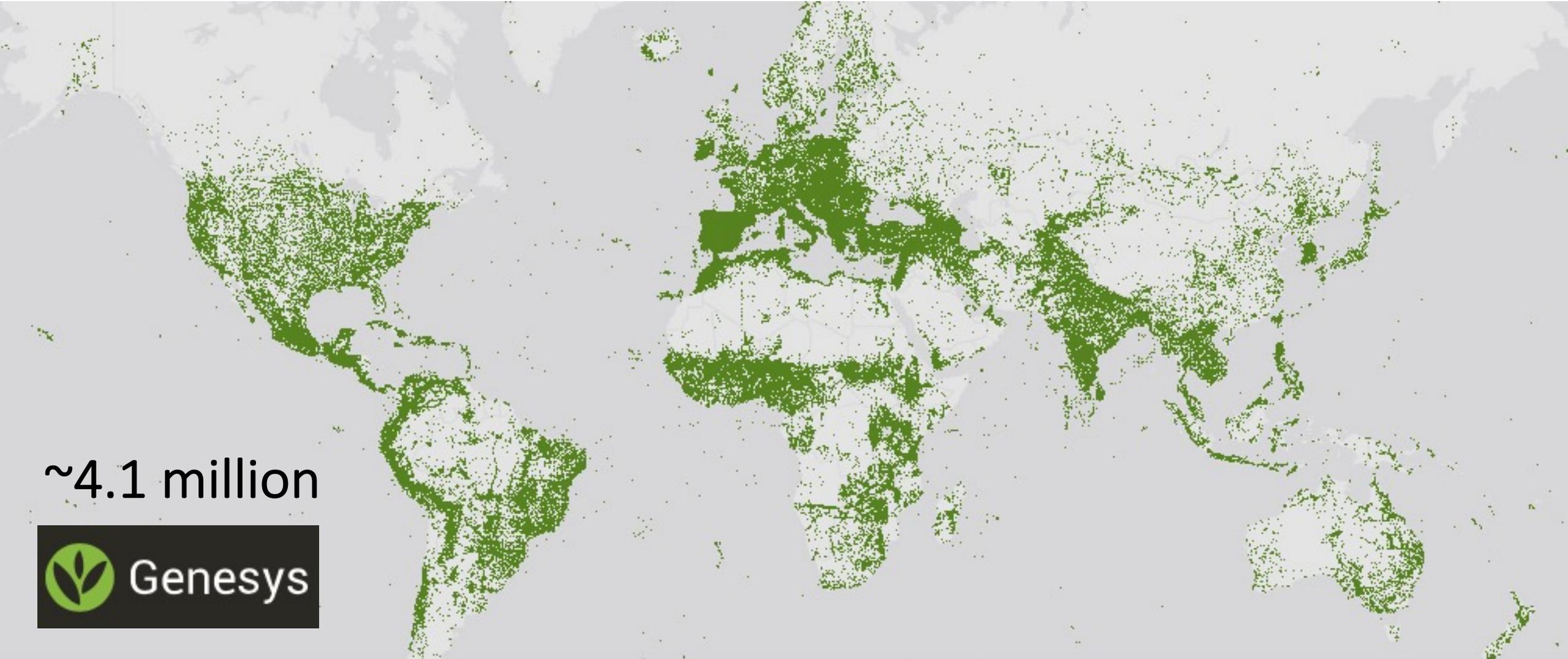
**1,750** genebanks

maintain

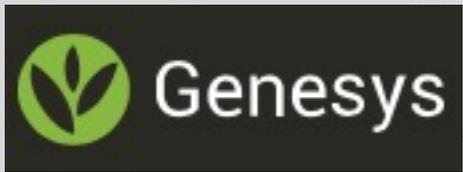
**7.2** million accessions

**2.8** million are (thought to be) unique

How the 'visible' part of this global system looks like



~4.1 million



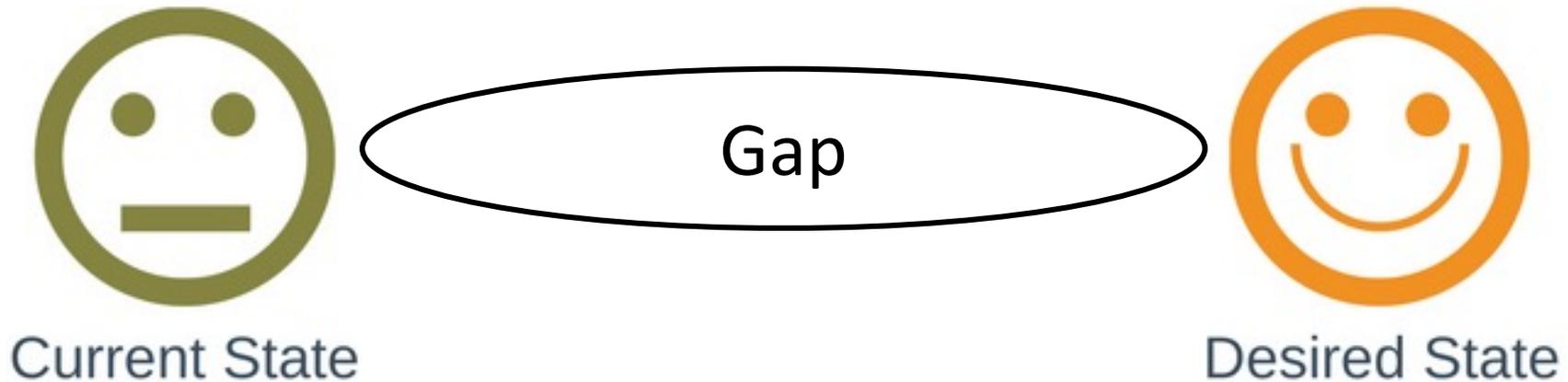
But are there gaps?

and if so, how big, and where?

# What is a gap?

Something that is missing

Typically this "something" takes us from a (current) undesired state to a (future) desired state



# If there is anything missing, then, what could this be?

Taxa/taxon

Population(s) or environmental range of given taxa

Specific sample(s) / accession(s)

Specific individual(s) in sample or population

Specific trait(s) or trait combination(s)

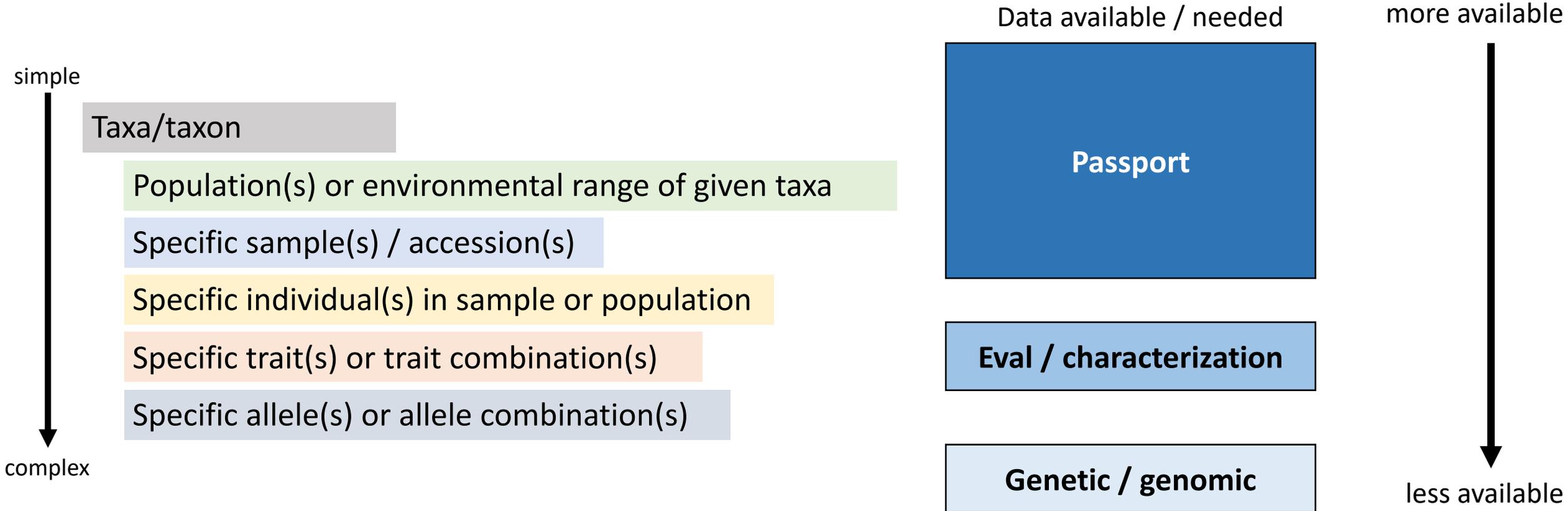
Specific allele(s) or allele combination(s)

simple



complex

# How do we go about quantifying this problem, then?



# How do we go about quantifying this problem, then?

Data available / needed



Taxonomic gaps

Population(s), “environmental”

Specific sample(s) / accession(s)

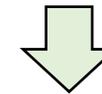
Trait gaps

Allelic / genetic gaps

## Geography

Collecting (our main way to fill any gaps) happens in specific locations,

so it all needs to go back to identifying those locations where gaps exist



Everything in the natural world has a detectable geographic pattern

**but...**

passport data isn't complete

not everything has coordinates

coordinates can be inaccurate, lack precision,  
or just plain wrong

taxonomic accuracy and  
precision aren't ideal

locality information isn't very complete either

there are duplicates

**Many more “buts” after...**

Let's take a look at the problem,



**Wild**



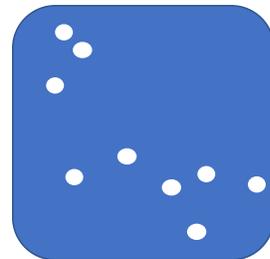
**Domesticated**

# First, we want to know which taxa are well sufficiently conserved

Is the **number of samples** in genebanks comparable to what is known about the taxon?

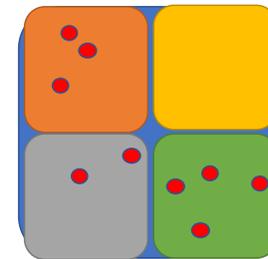
$$\text{SRS} = \frac{\# \text{ samples in genebanks}}{\# \text{ total taxon observations}}$$

Is the **geographic distribution** of these samples representative?



$$\text{GRS} = \frac{\text{white area}}{\text{blue} + \text{white area}}$$

Is the **environmental distribution** of these samples representative?



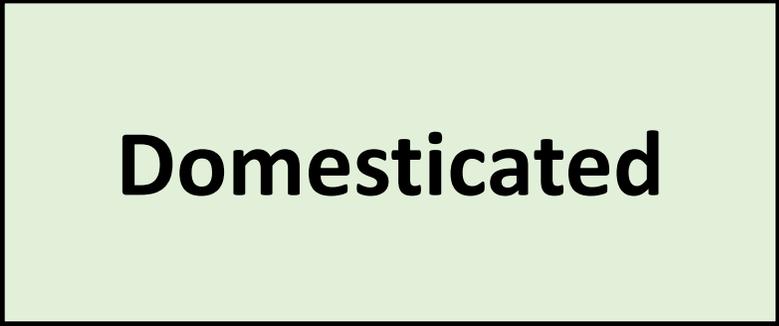
$$\text{ERS} = \frac{\# \text{ squares w/ red dot}}{\text{total} \# \text{ squares}}$$

Second, for those which aren't, we would like to know where the gaps are

Let's take a look at the problem, then

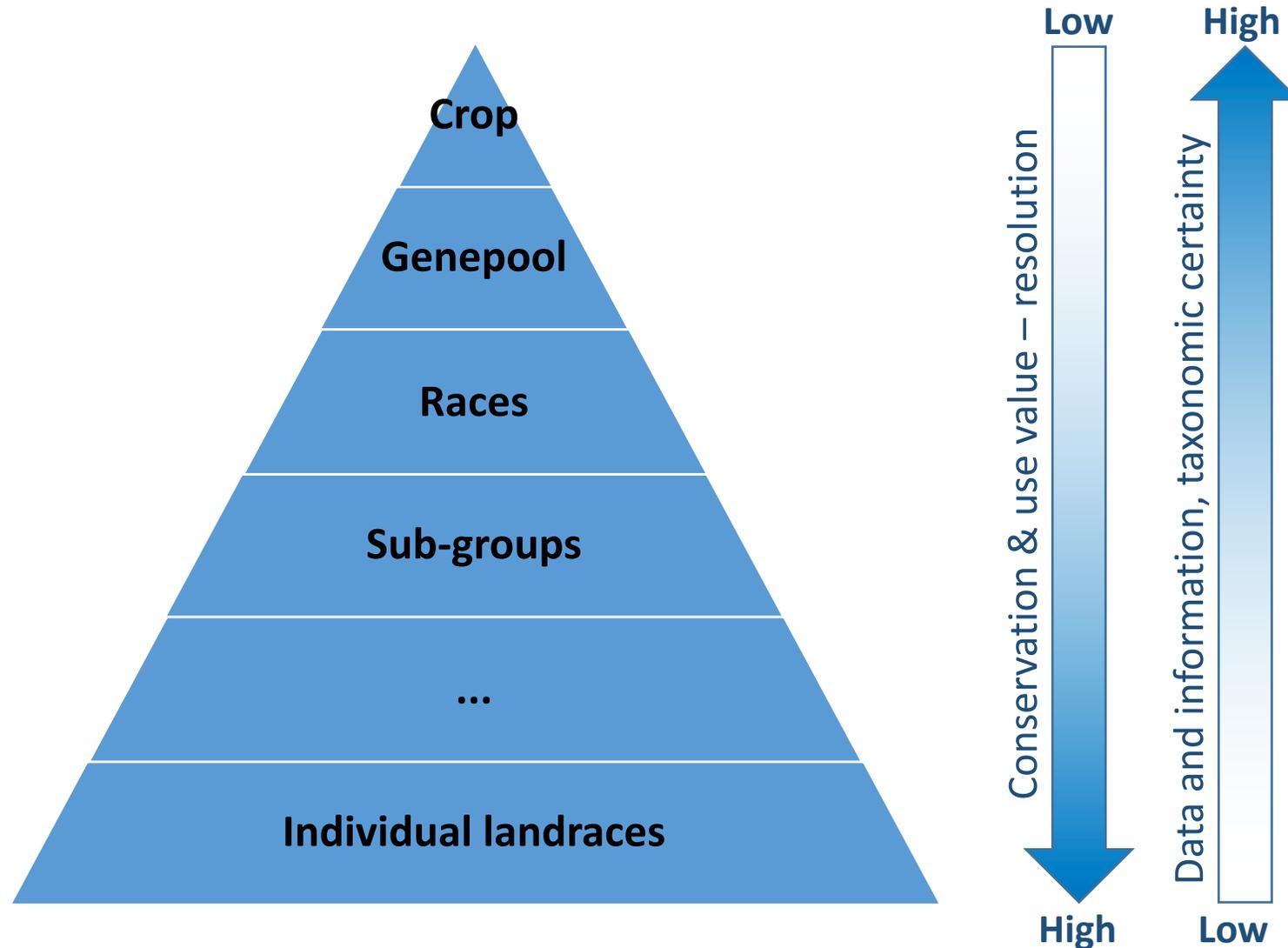


Wild



**Domesticated**

First, we want to structure the landrace accessions using what is known about their diversity

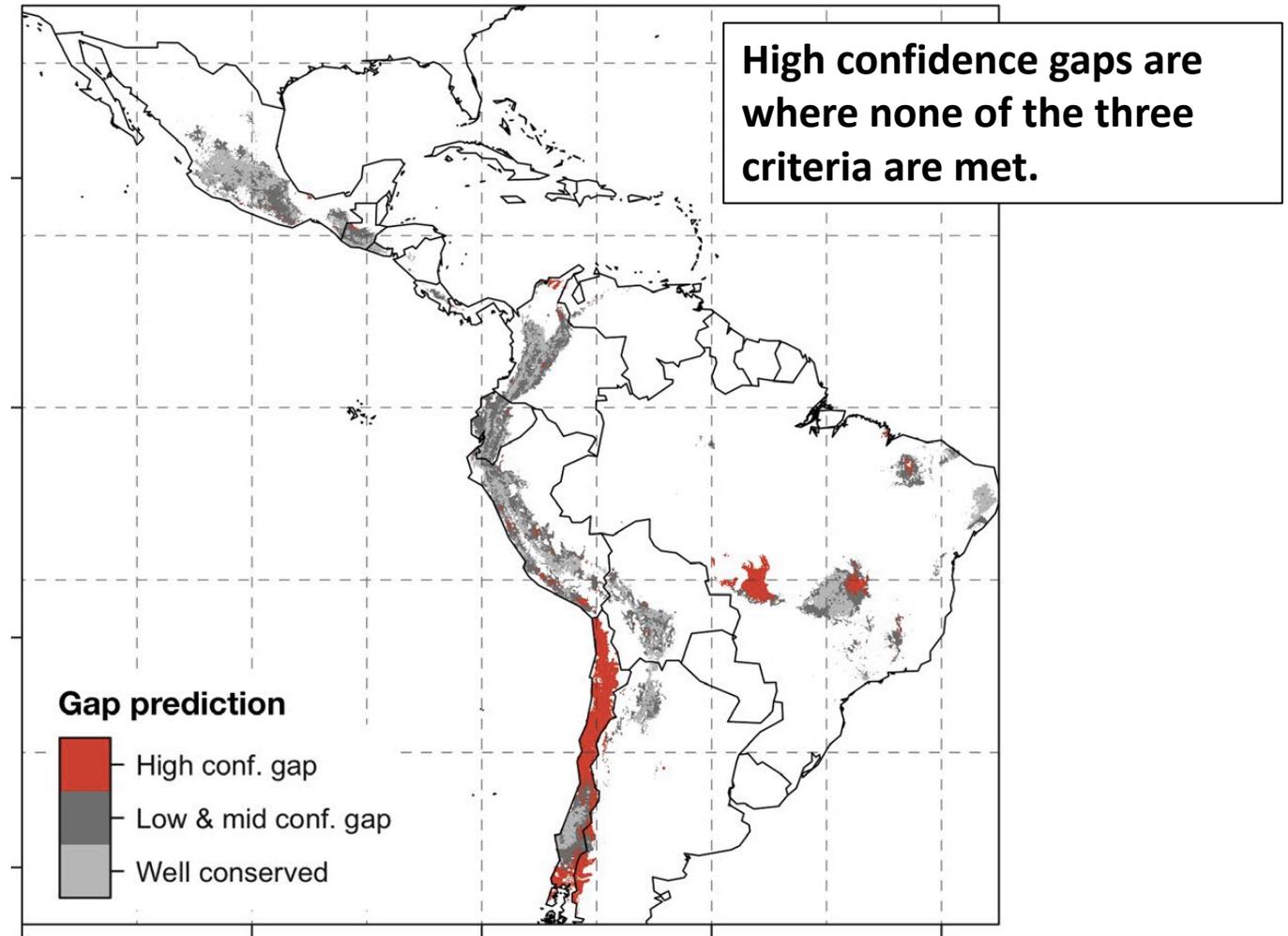


Second, we want to understand which locations in their geographic distribution are a gap

Is the **site easily accessible** from any other existing collection site?

Is the **site well within the network** of existing accessions?

Is the **site environmentally similar** to any other accession in a given climate type?



# Two global multi-crop studies (+ 20 other papers)

Castaneda-Alvarez et al. (2016) in Nature Plants

Analyzed 1,076 wild taxa related to 81 domesticated crops

Ramirez-Villegas et al. (under revision) in Nature Plants

Analyzed 72 “landrace groups” of 25 domesticated crops

# CWR priorities for conservation

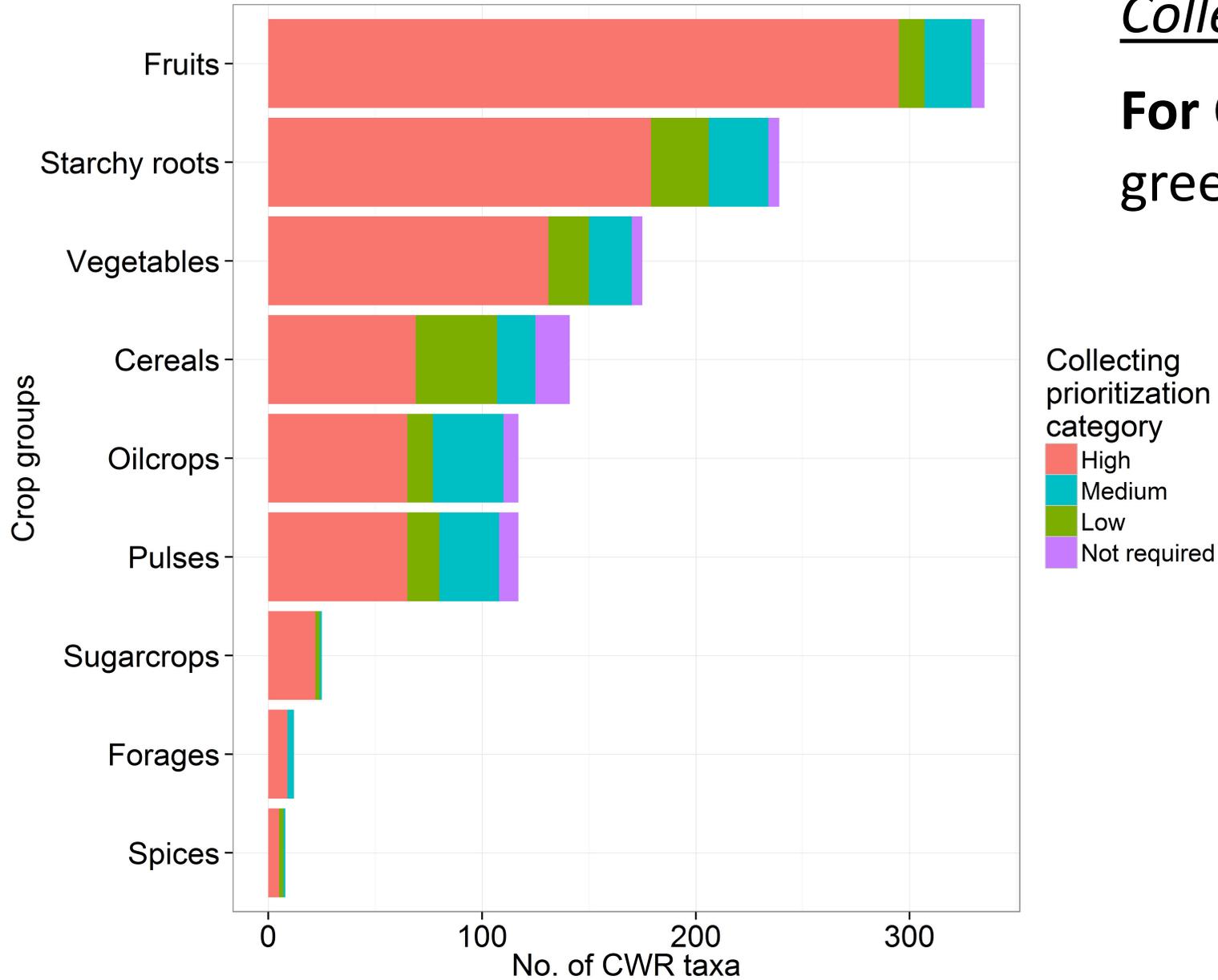
**71.1% (765):** High priority for collecting

**13.8% (148):** Medium priority for collecting

**11.0% (118):** Low priority for collecting

**4.2% (45)** Sufficiently represented in genebanks





Collecting –have we done enough?

**For CWR;** lots of red, not so much green and purple

So, no, not really. We haven't done enough, across the board.

## Collecting – have we done enough?

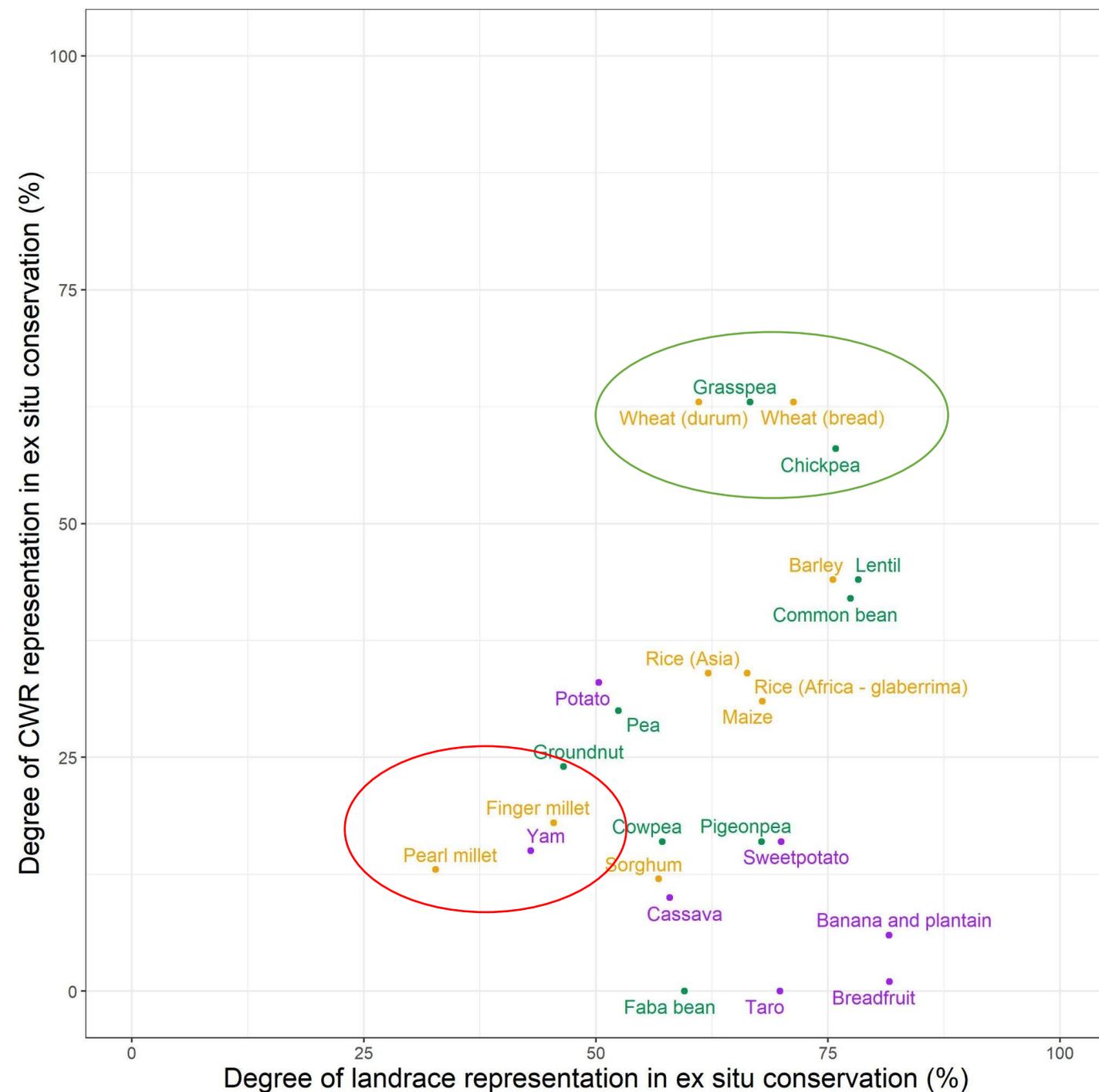
**For landraces; well, it depends.**

Grasspea, wheat, chickpea,  
relatively well conserved for both

But, millets, yam, groundnut are  
clearly poorly preserved ex-situ for  
both CWR and landraces

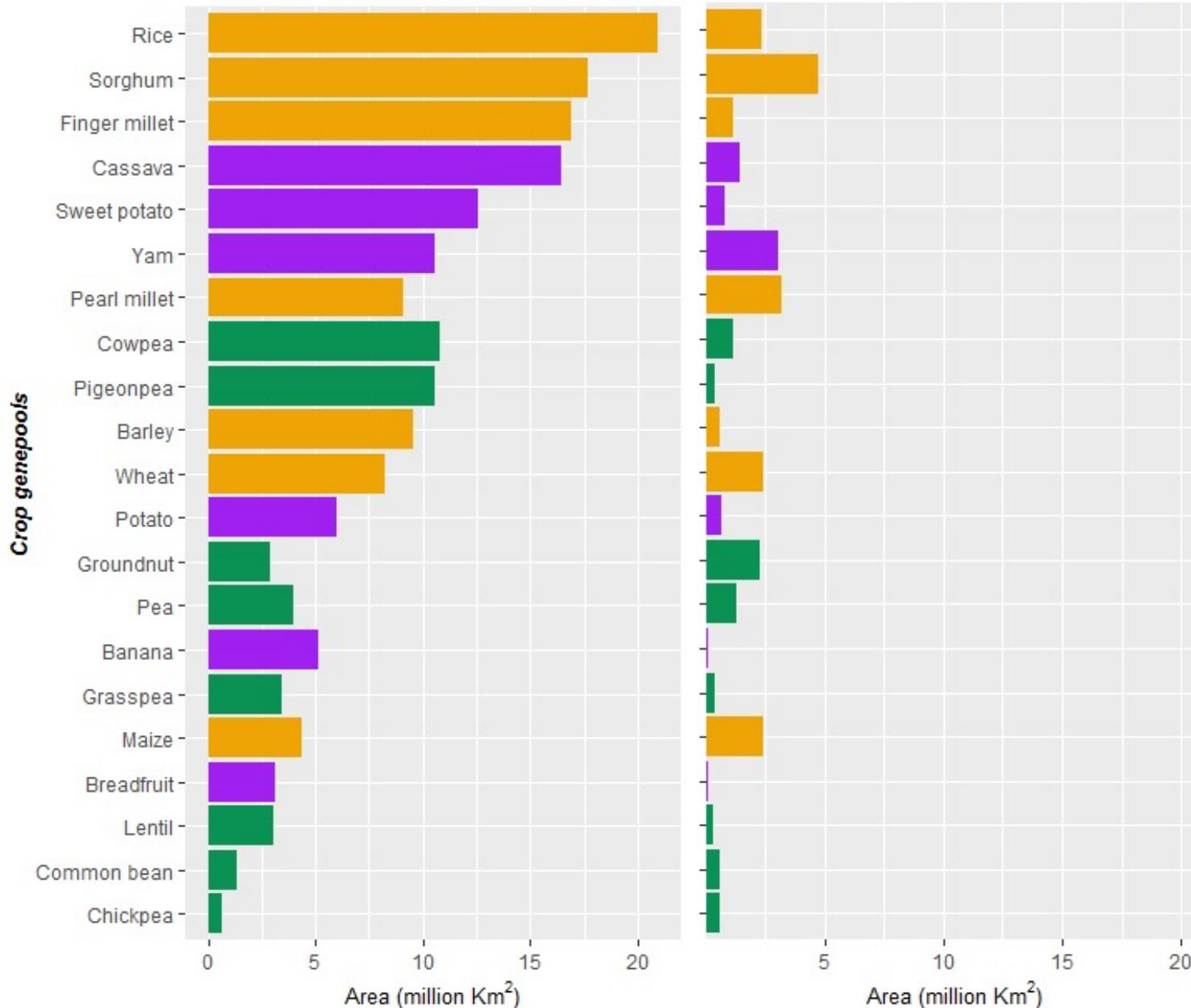
But on average genebanks conserve the  
equivalent to **63% of the landrace  
geographic distributions**

At the crop level, landrace conservation  
levels are not particularly correlated with  
crop wild relative (CWR) conservation



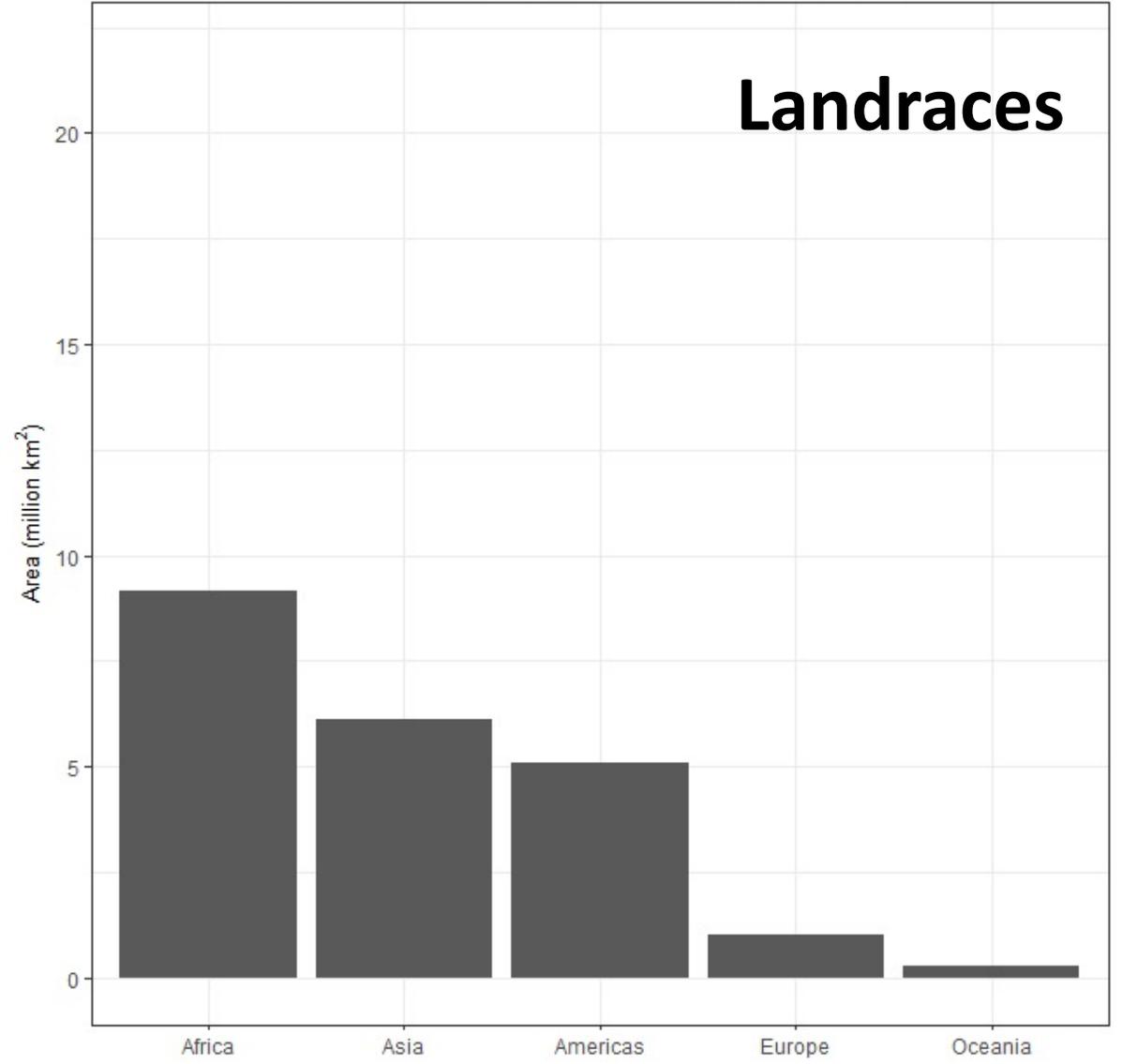
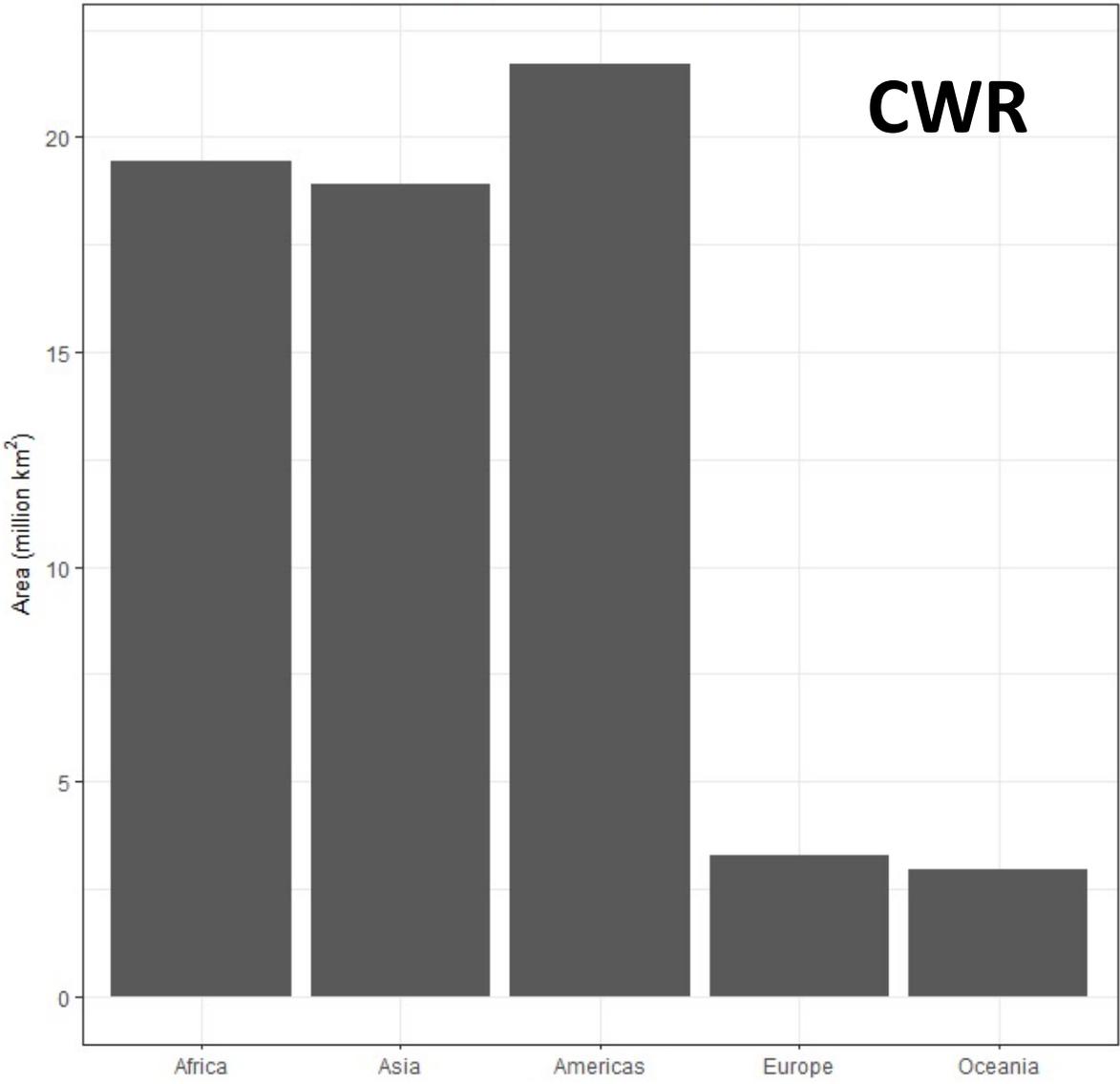
# CWR

# Landraces

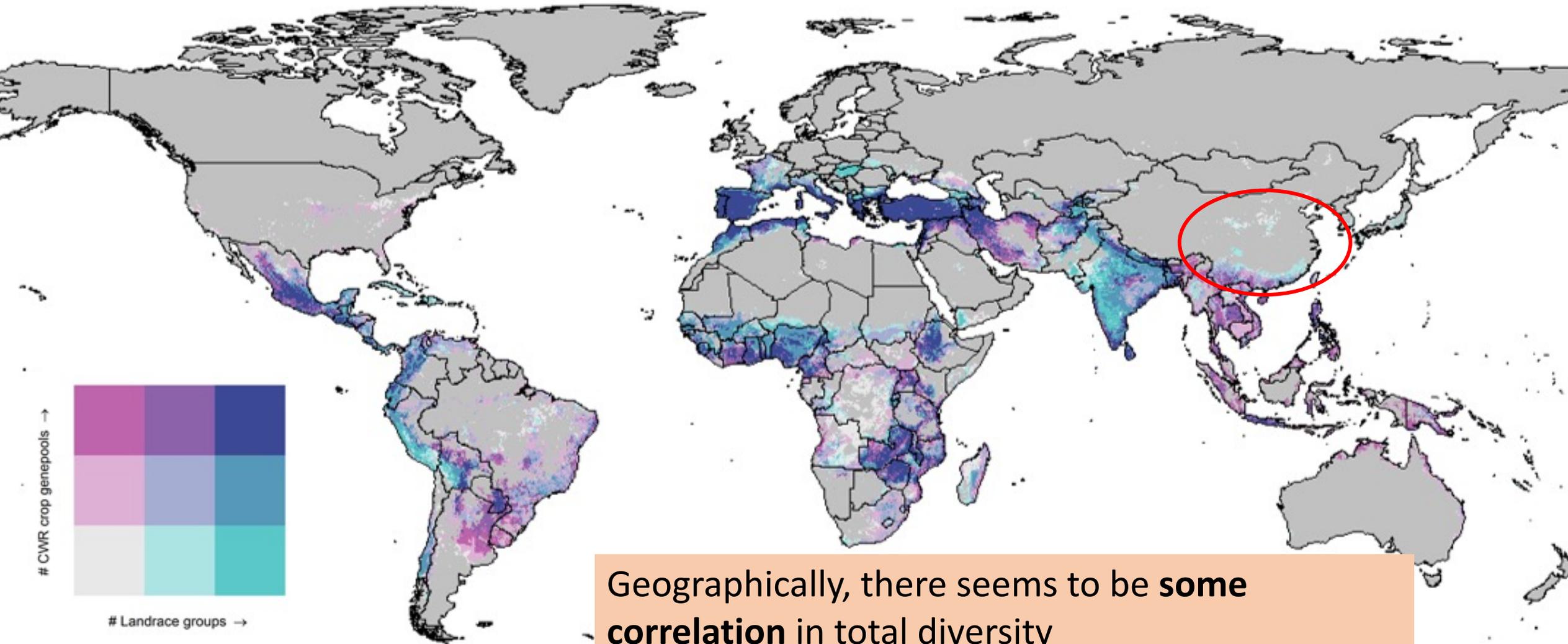


# The scale of the problem

# Where are then the gaps?

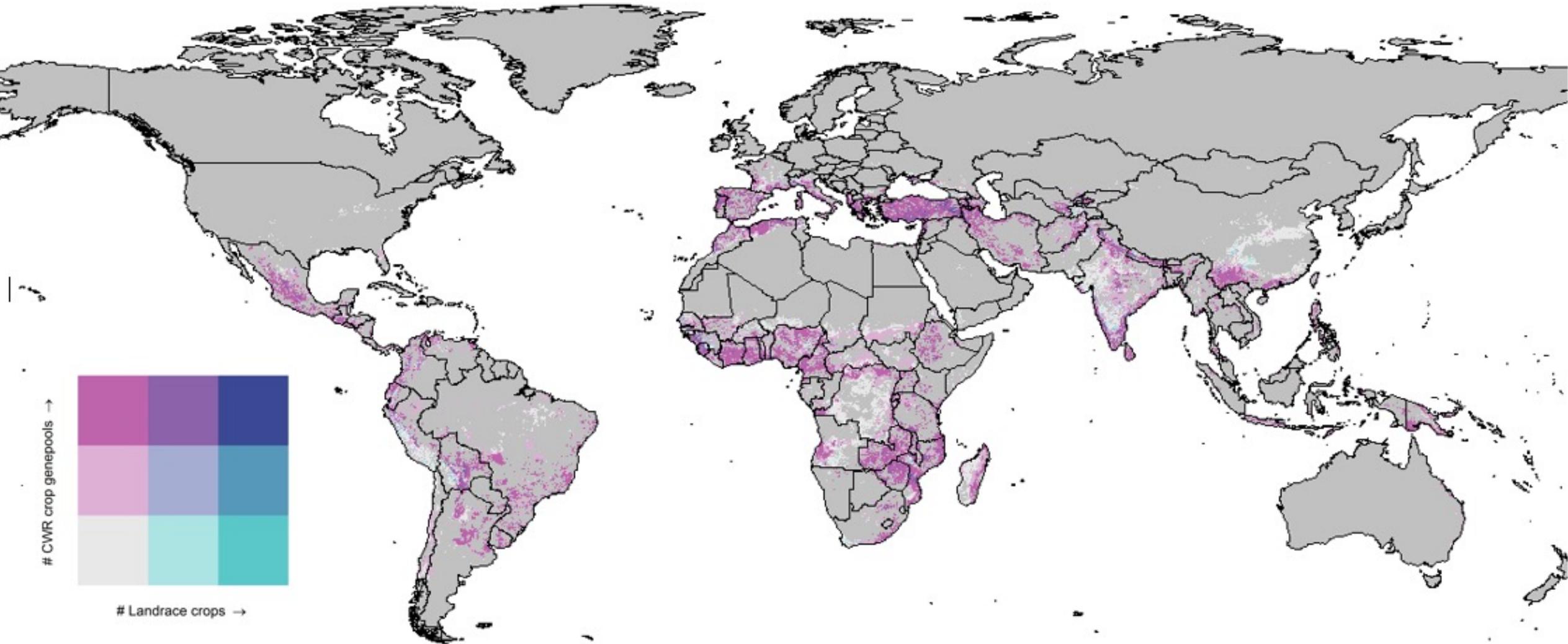


# Where are then those gaps? –first, total diversity



Geographically, there seems to be **some correlation** in total diversity

# Where are then those gaps? –predicted gaps



# With these results, we can now ask many other practical / research questions

What are the environmental patterns that are most often missing?

What diversity patterns emerge and at what scales? Are these consistent with domestication / diversity centers?

Which diversity/domestication centers have most gaps?

What about gaps in specific crop gene pools? Can we see gaps for “Mexican highland” maize?

What countries are most priority based on ‘uncollected’ area?

# We haven't done enough, but should we be collecting more?

## The conservation perspective

Diversity is at risk, it is unpreserved, and we need to preserve it.

Thus, we need to collect, urgently.

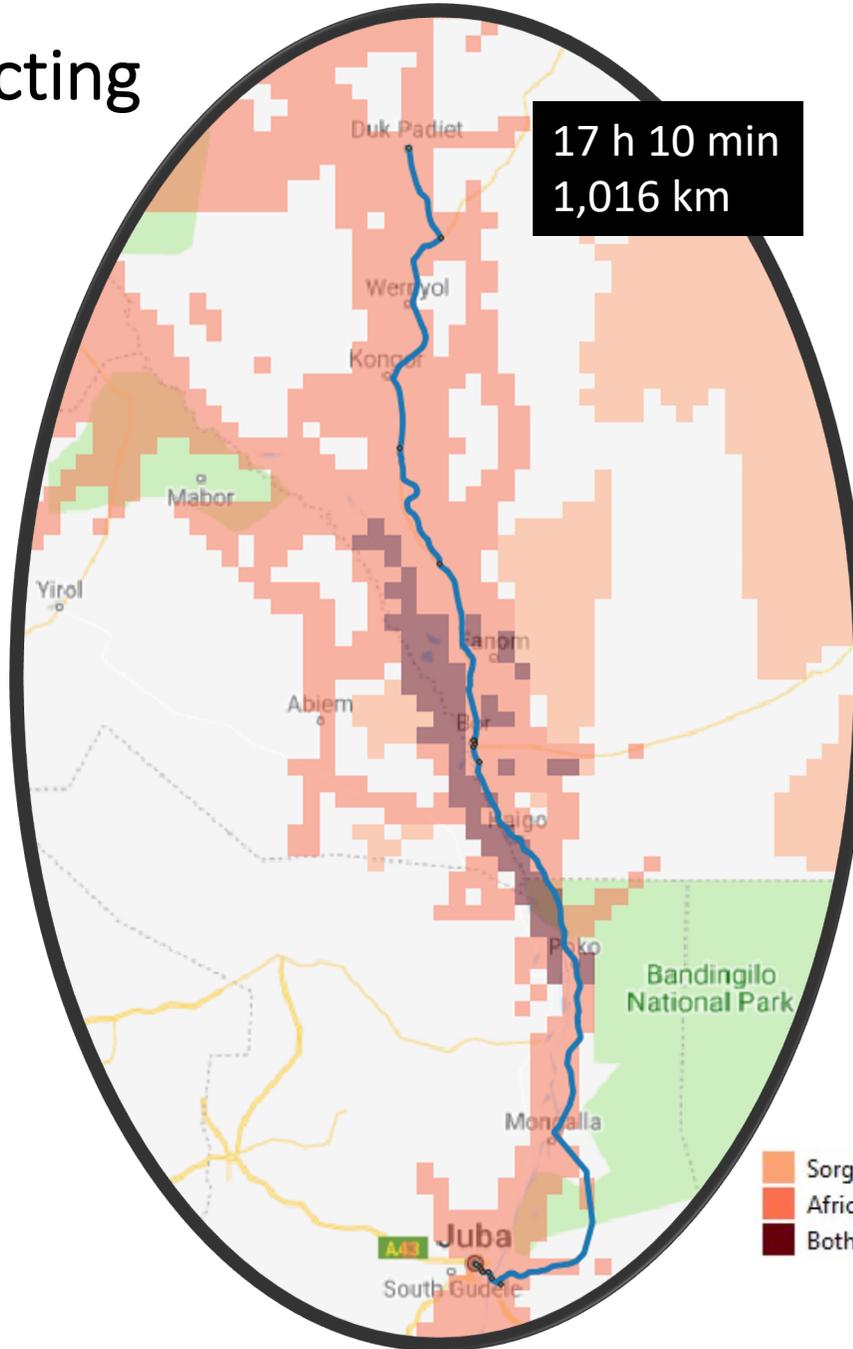
## The (measured) use perspective

Large parts of the collections are still unused by researchers and general public.

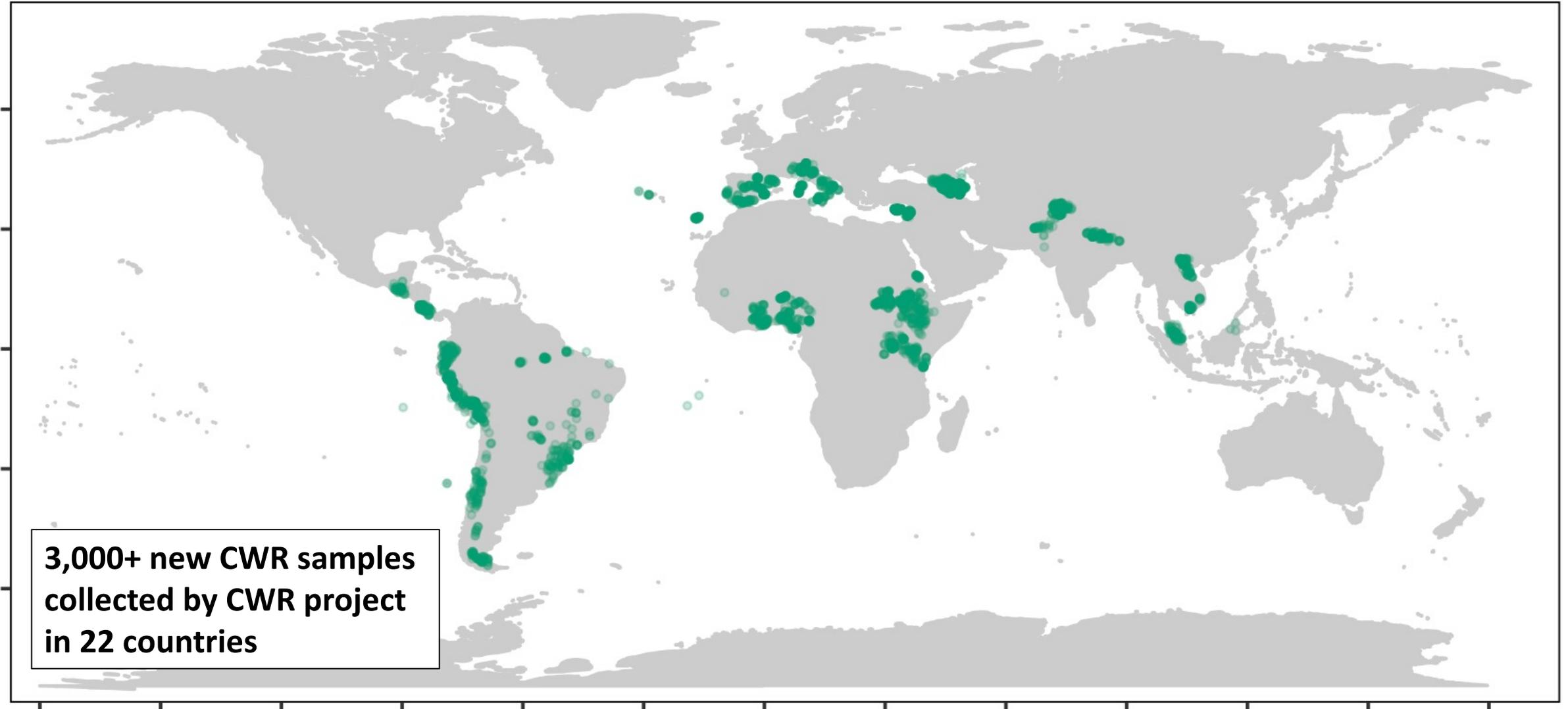
Before investing in more collecting we may need to first understand if we have what we need (for e.g., breeding climate-adapted crops).

But... does collecting really work, though? Does it fill gaps?

# Gap analysis can support collecting

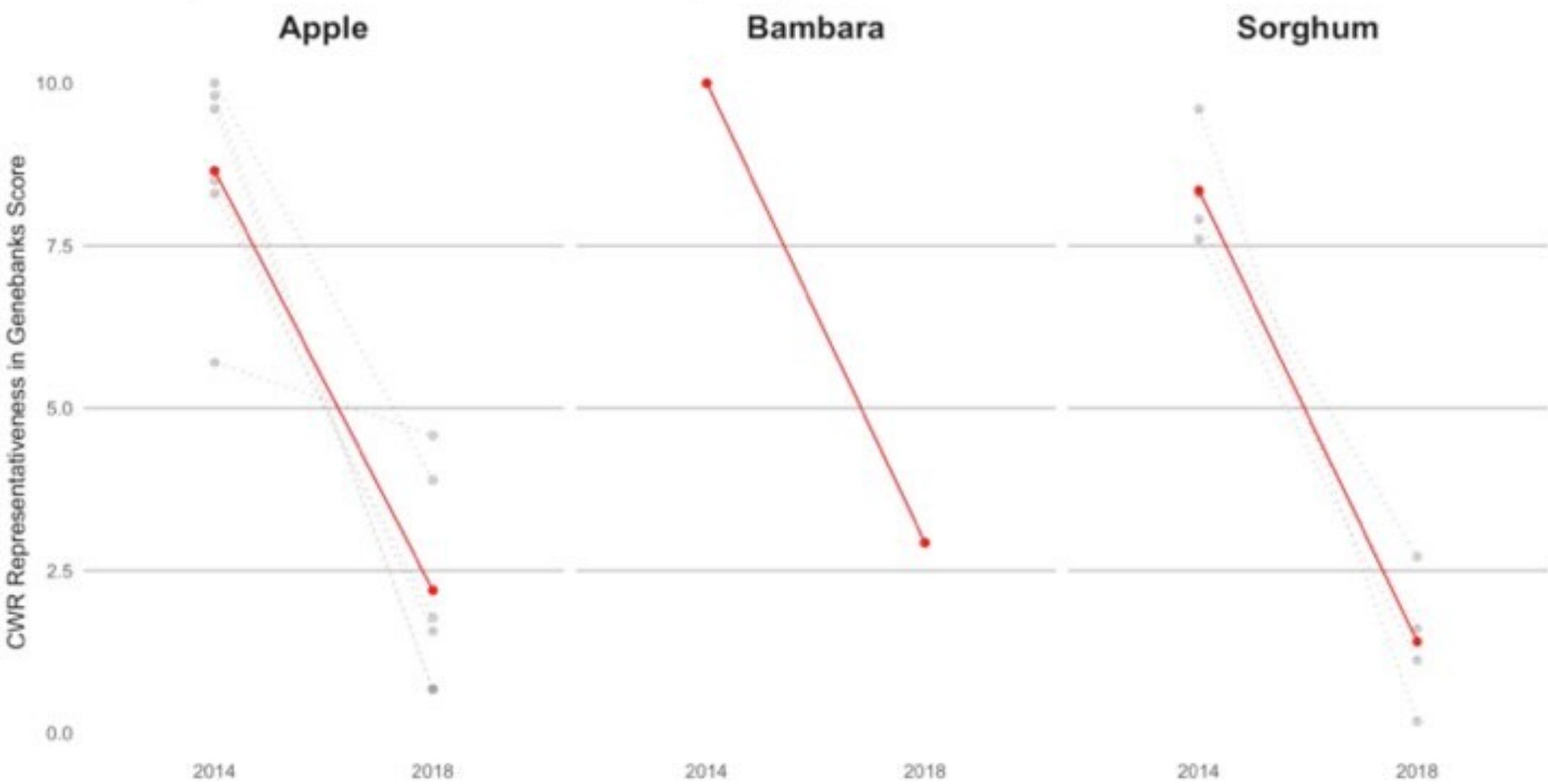


# If well targeted, collecting does fill gaps



**3,000+ new CWR samples  
collected by CWR project  
in 22 countries**

# If well targeted, collecting does fill gaps



# Conclusions

- After ~15 years of hard work, we have methods to predict conservation gaps in wild and domesticated plant genetic resources.
- Despite “salvage” efforts that collected wild and domesticated plant genetic diversity (mainly) in the 70s and 80s, gaps exist.
  - **Roughly two thirds** of CWR taxa are in need of conservation. We also know very little about them, and they may be endangered.
  - **Landraces are much better represented**, with 63% of their geographic area currently represented in genebanks
- Collecting, if done smartly, can help filling those gaps.  
**But** the scale of the collecting needed appears to be very large.

# Conclusions

- Lastly, our systems / knowledge ought to improve
  - **Data updates and data quality** need improvement. This would greatly improve the accuracy of any estimates of “where to collect”
  - **Data management during and after collecting** can lead to us not knowing whether gaps are filled or not, and therefore creates inefficiency risks
  - **New methods** that mine genomic data could help us get past ecogeography and any related assumptions about the diversity that we have and that we don't, as well as its potential use
  - We need ways of **monitoring risks of genetic erosion** especially in “gap areas”, so that this further helps prioritizing collecting

1984



2021

