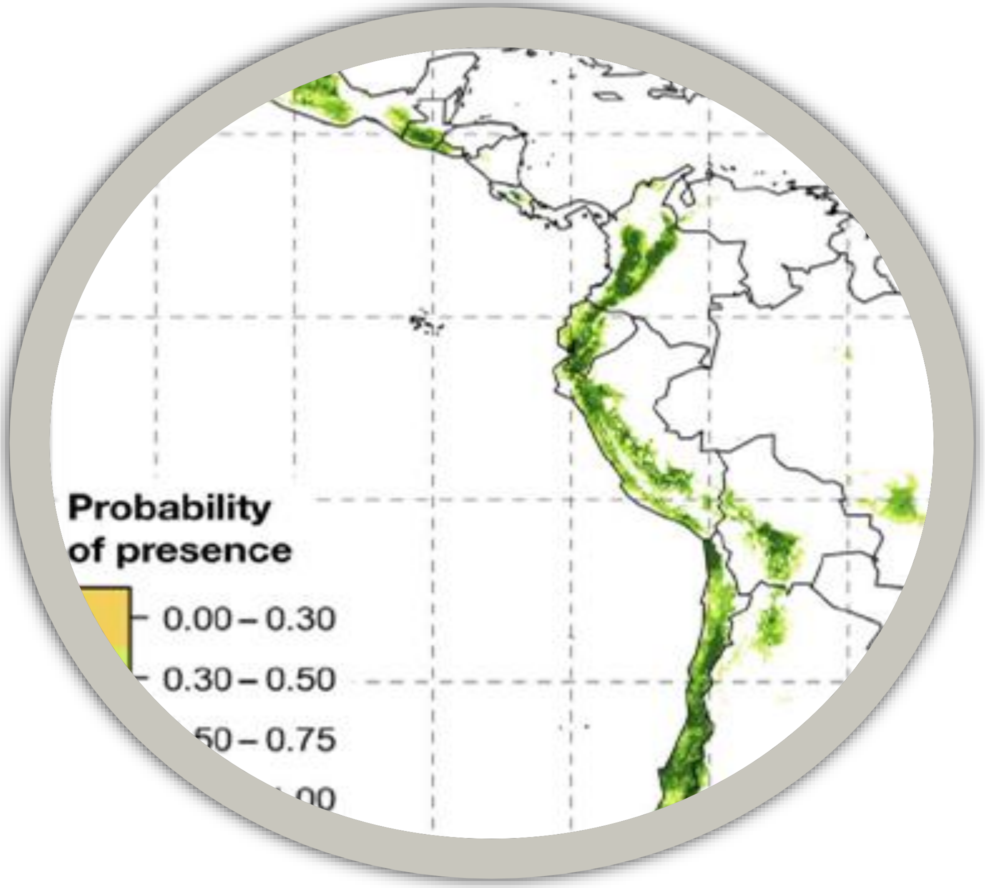
General introduction to gap analysis



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

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

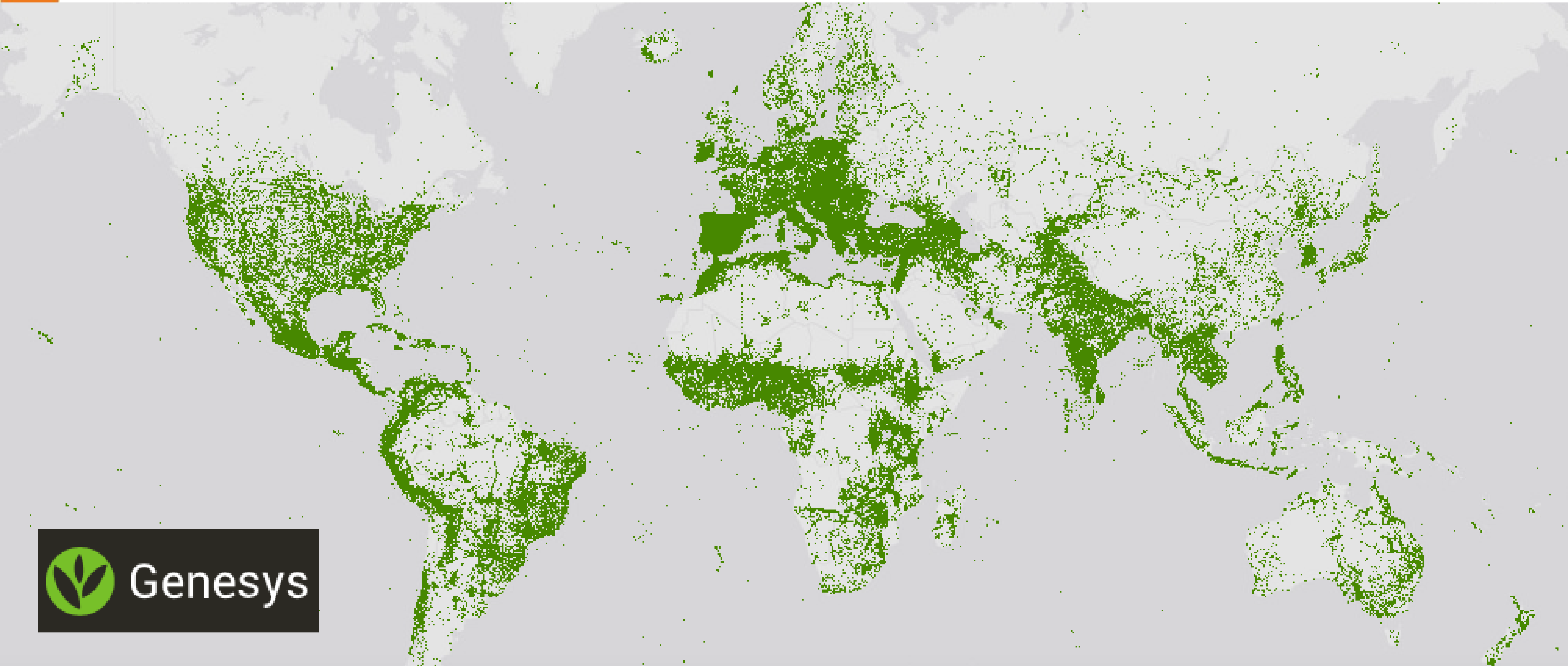
1,750 **genebanks**

maintain

7.2 **million** **accessions**

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

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

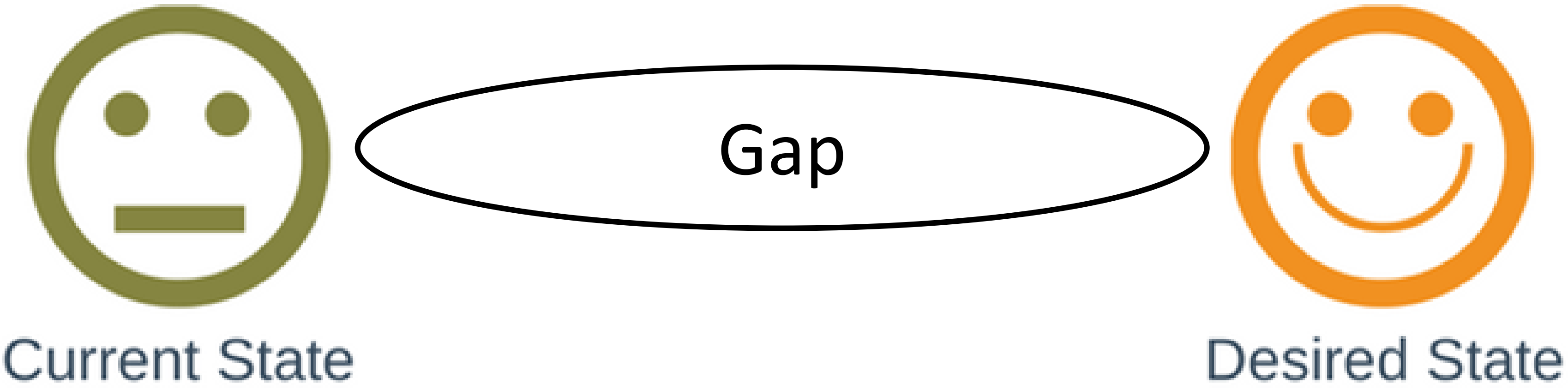


But are there gaps?

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?

simple

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)

complex

Data available / needed

Passport

Eval / characterization

Genetic / genomic

more available

less available

How do we go about quantifying this problem, then?

Data available / needed

Passport

Eval / characterization

Genetic / genomic

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

taxonomic accuracy and
precision aren’t ideal

not everything has coordinates

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

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

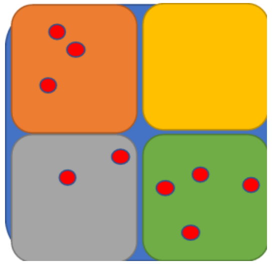
Is the **number of samples** in genebanks comparable to what is known about the taxon**?**

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

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

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

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



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

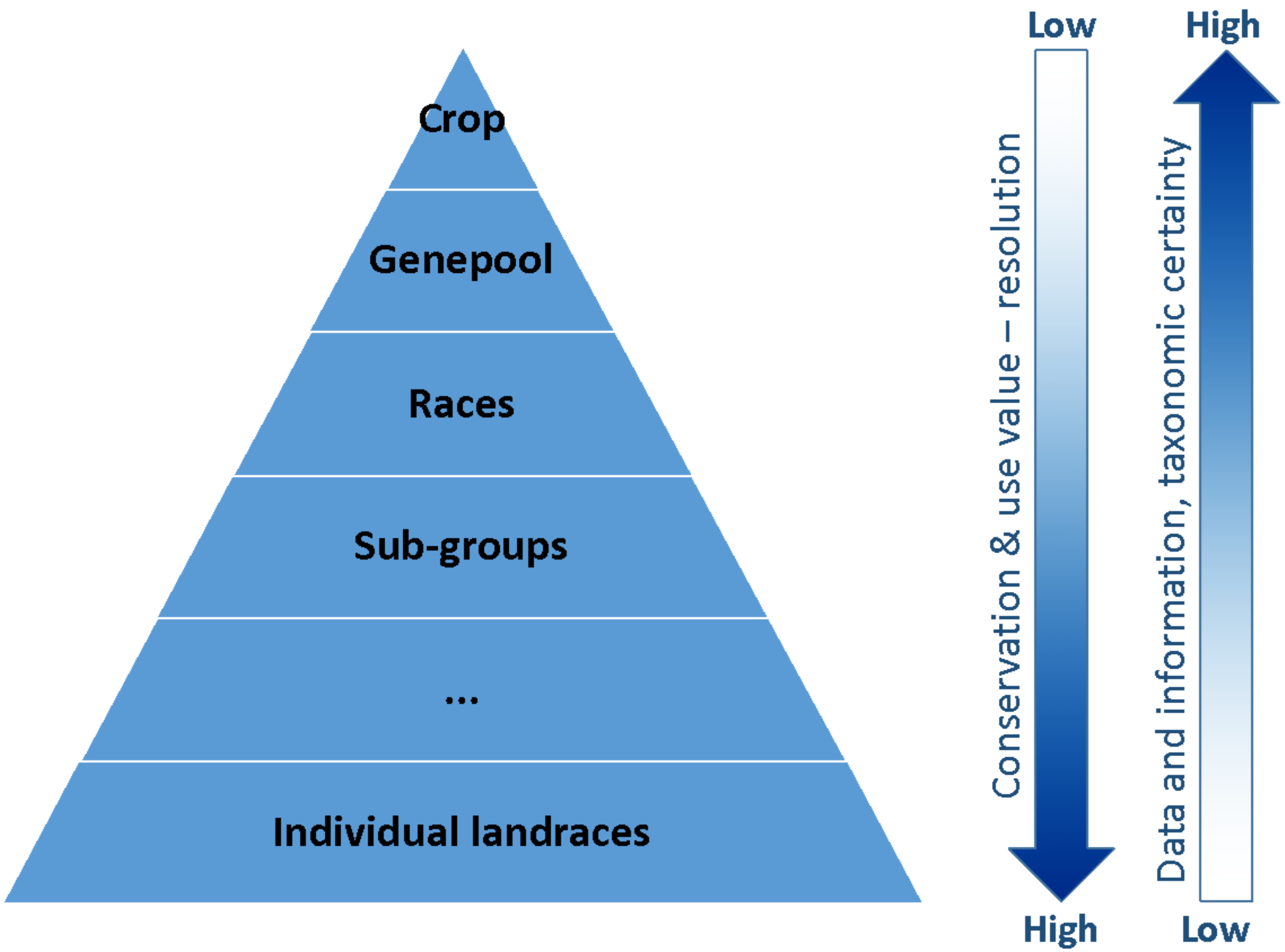
Ramirez-Villegas et al. (2010)

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

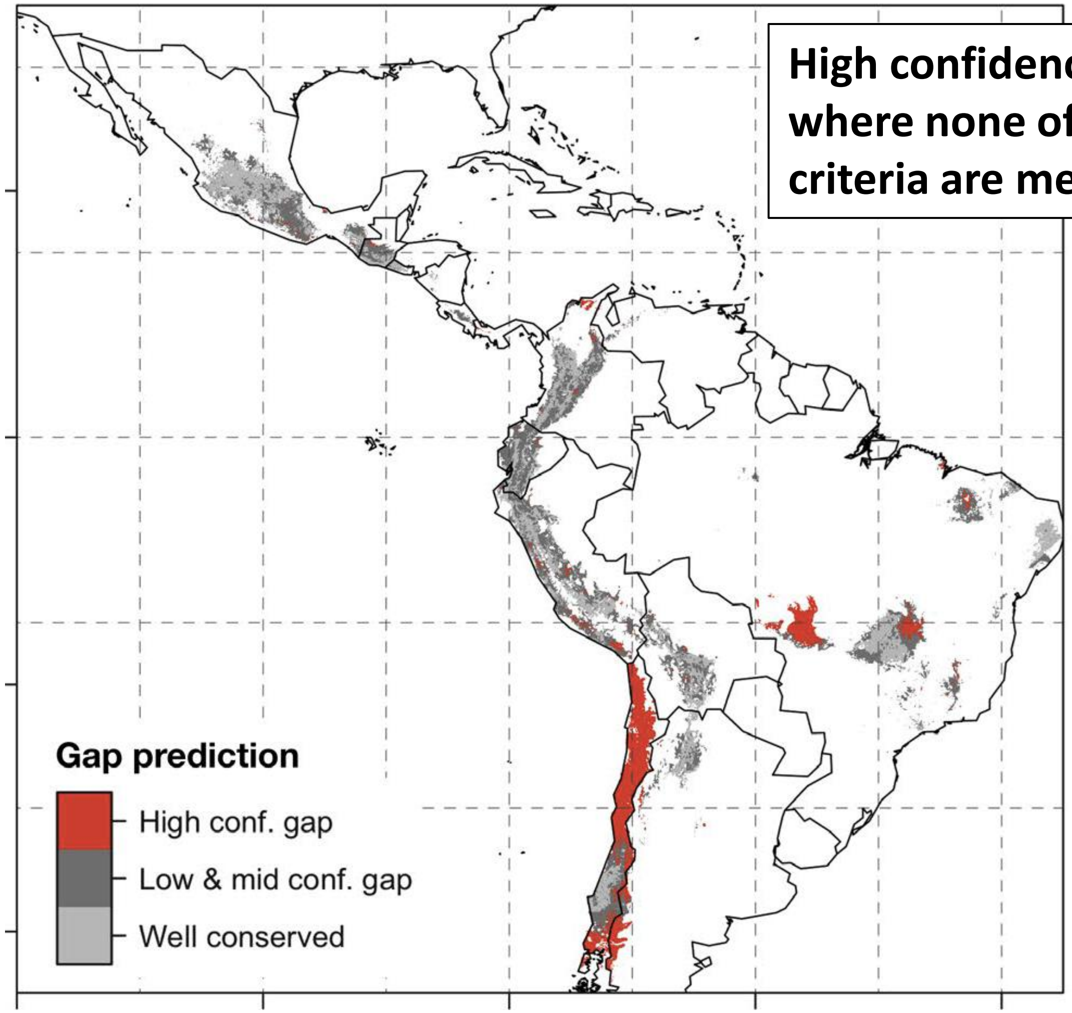


Second, we want to understand which locations in their geographic distribution are a gap

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?



High confidence gaps are where none of the three criteria are met.

Ramirez-Villegas et al. (2020)

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

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

Ramirez-Villegas et al. (2022) in Nature Plants

CWR priorities for conservation

71.1% (765): High priority for collecting

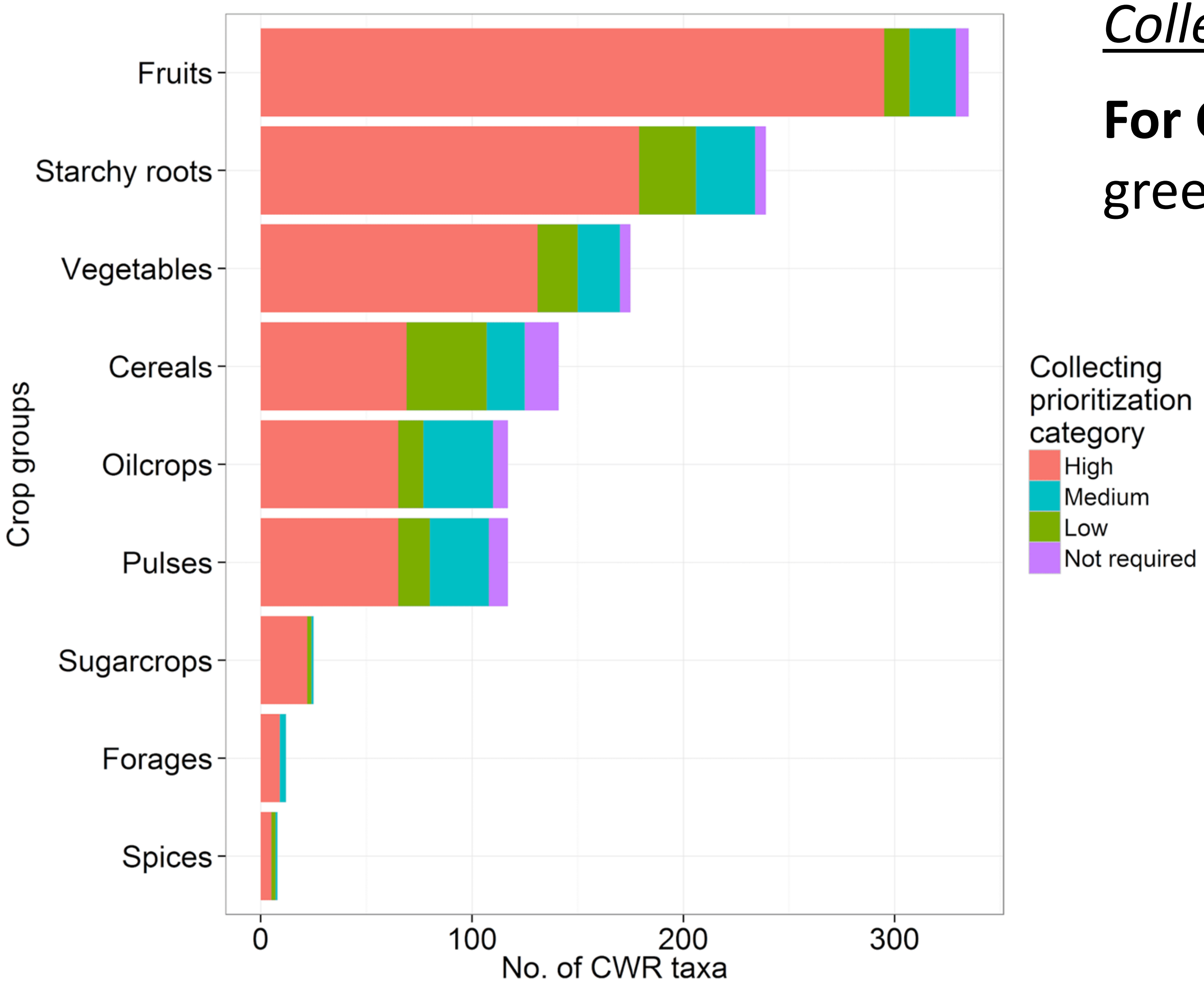
13.8% (148): Medium priority for collecting

11.0% (118): Low priority for collecting

4.2% (45) Sufficiently represented in genebanks

1,076
CWR taxa analyzed

Castaneda-Alvarez et al. (2016)

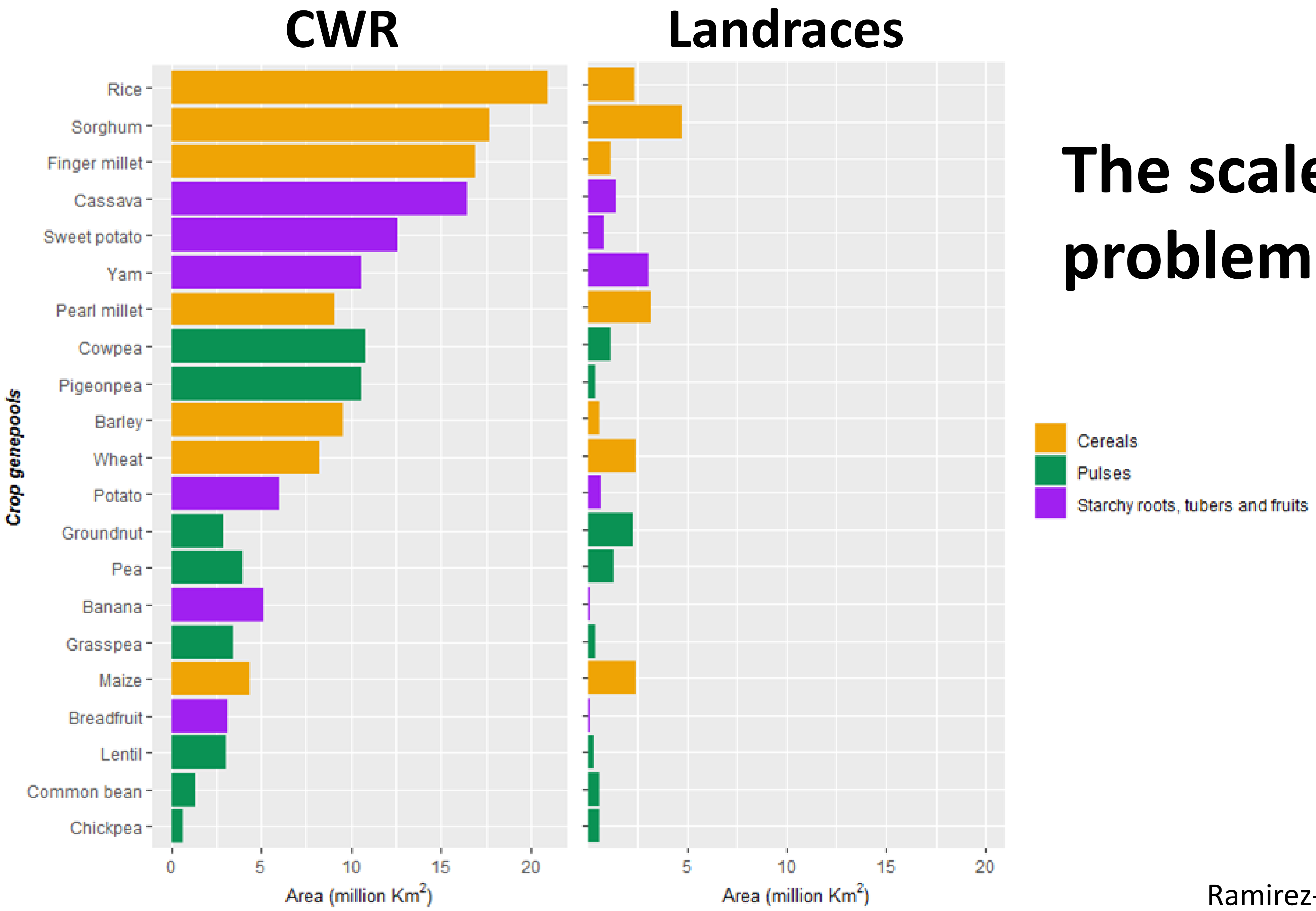


Collecting –have we done enough?

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

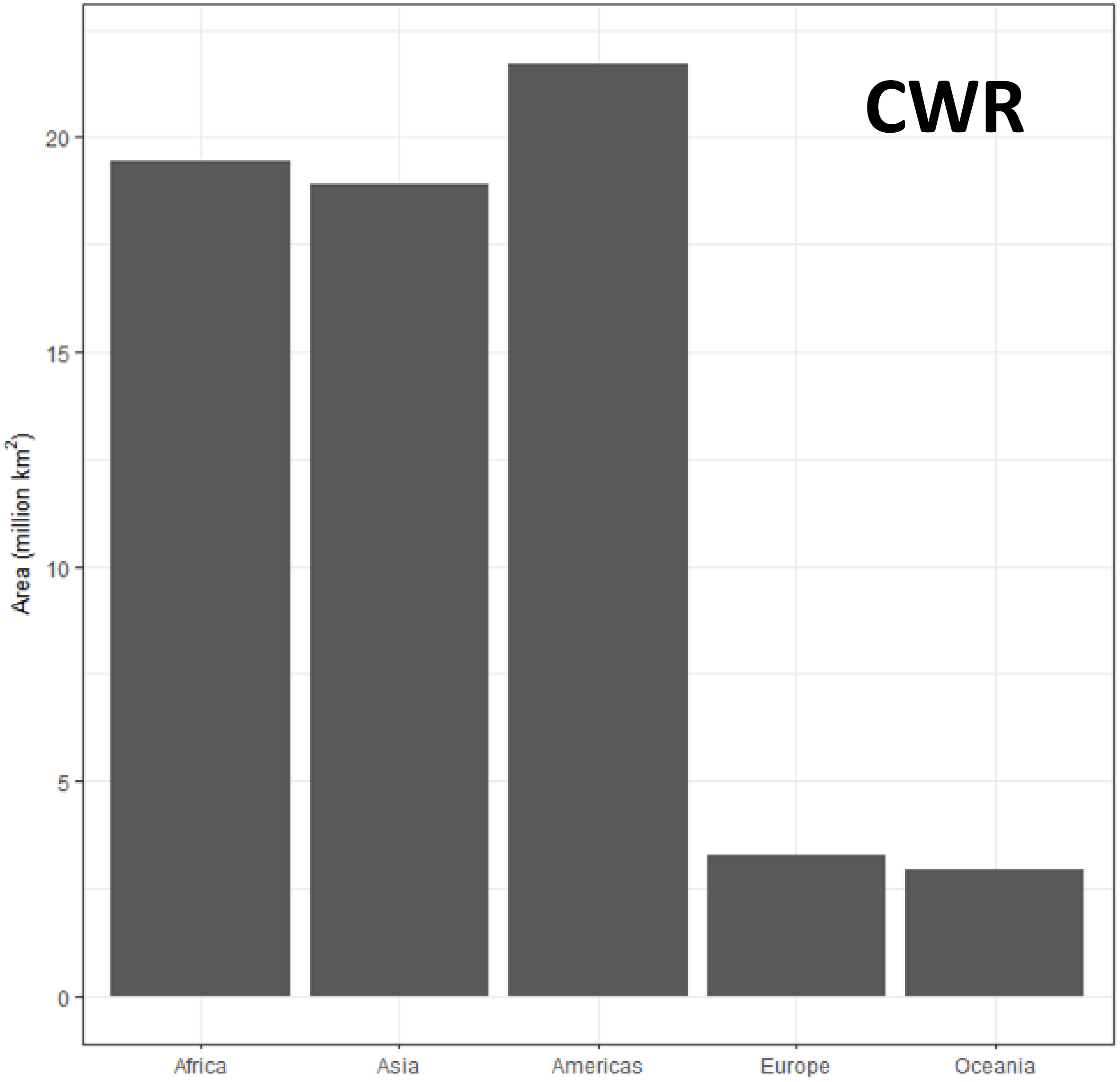
So, no, not really. We haven't done enough, across the board.

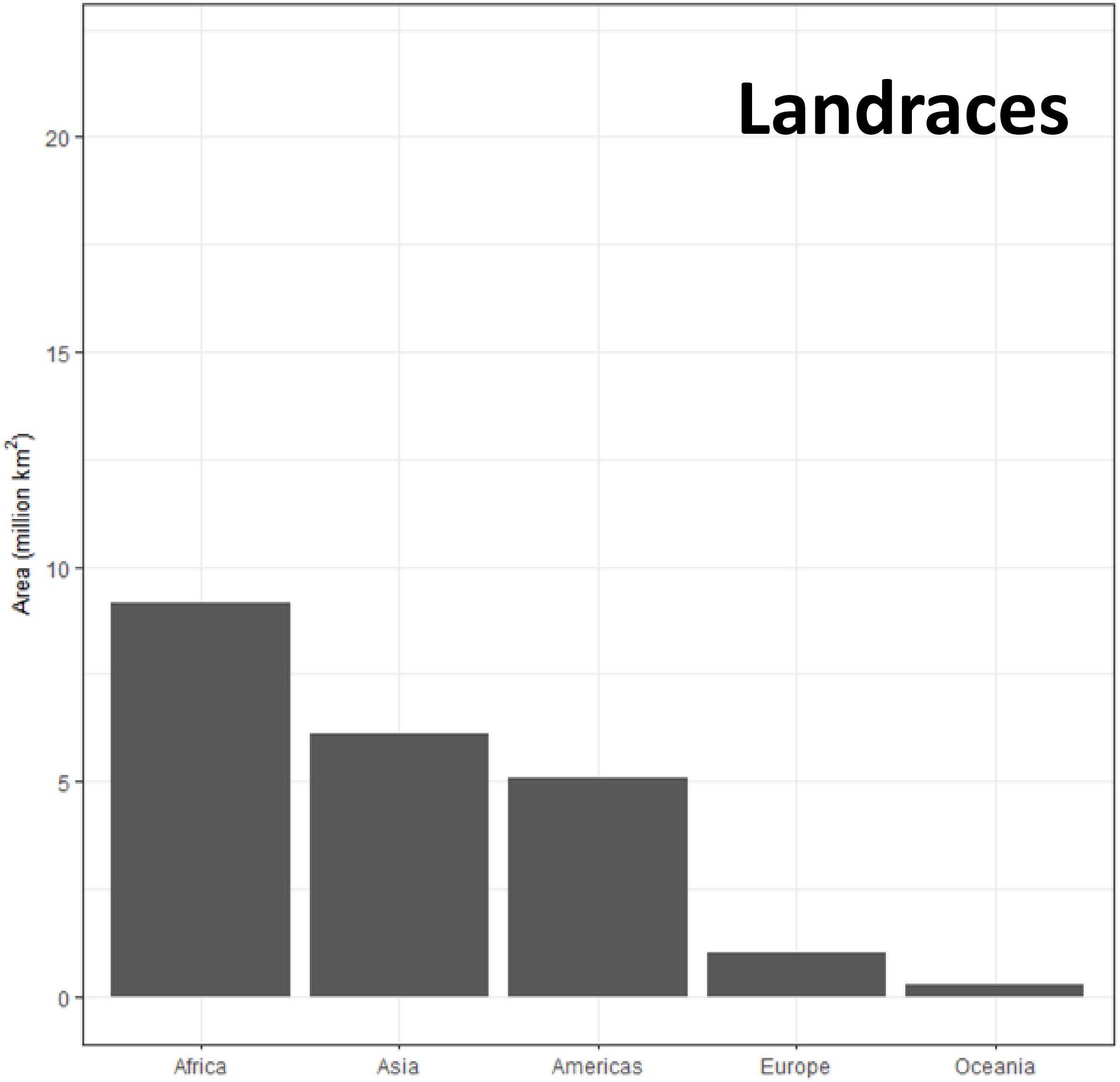
Castaneda-Alvarez et al. (2016)

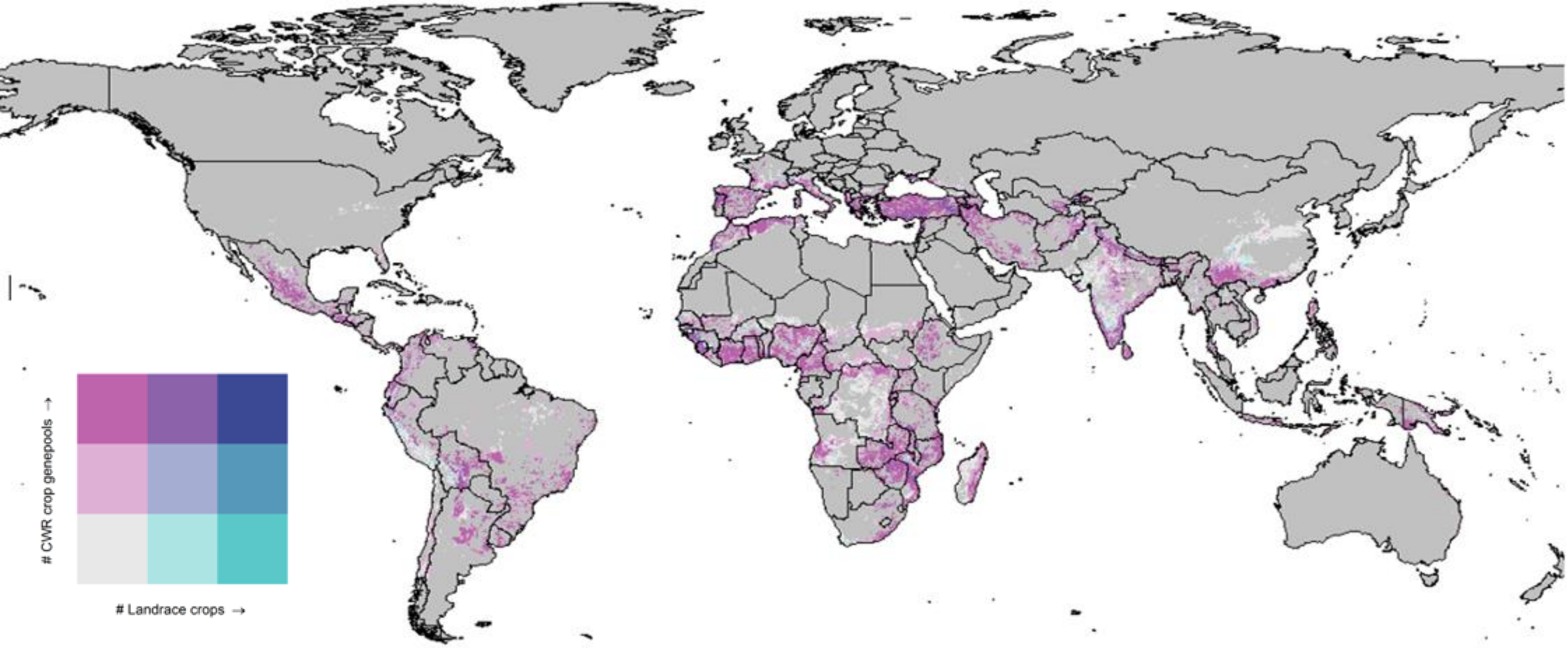


The scale of the problem

Ramirez-Villegas et al. (in prep)


CWR
Area (million km2)
0
5
10
15
20
Africa
Asia
Americas
Europe
Oceania


Landraces
Area (million km2)
0
5
10
15
20
Africa
Asia
Americas
Europe
Oceania



Ramirez-Villegas et al. (in prep)


1984
Mansoura
El-Mahalla
El-Kubra
Damietta Branch (Nile)
5 km
2021
Mansoura
El-Mahalla
El-Kubra
Damietta Branch (Nile)
It can be late
5 km