



BioDT use case 4.1.2.1

crop wild relatives

Desalegn Chala – Natural History Museum, University of Oslo

2023 04 18
WP6 meeting



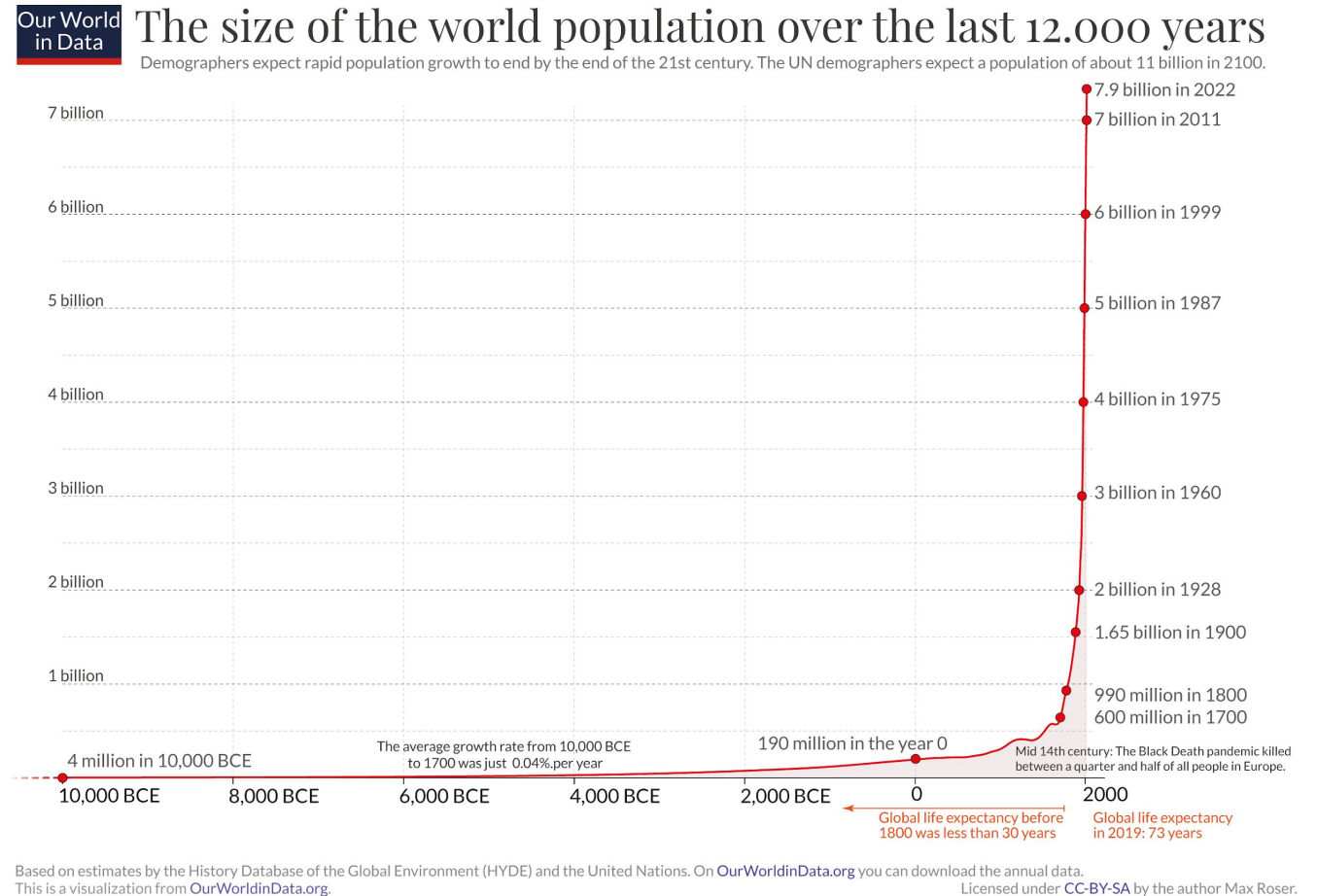
Funded by
the European Union

Outline

- Why crop wild relatives?
- Our plan – where are we heading?
- What have we done so far?
- Tasks waiting ahead

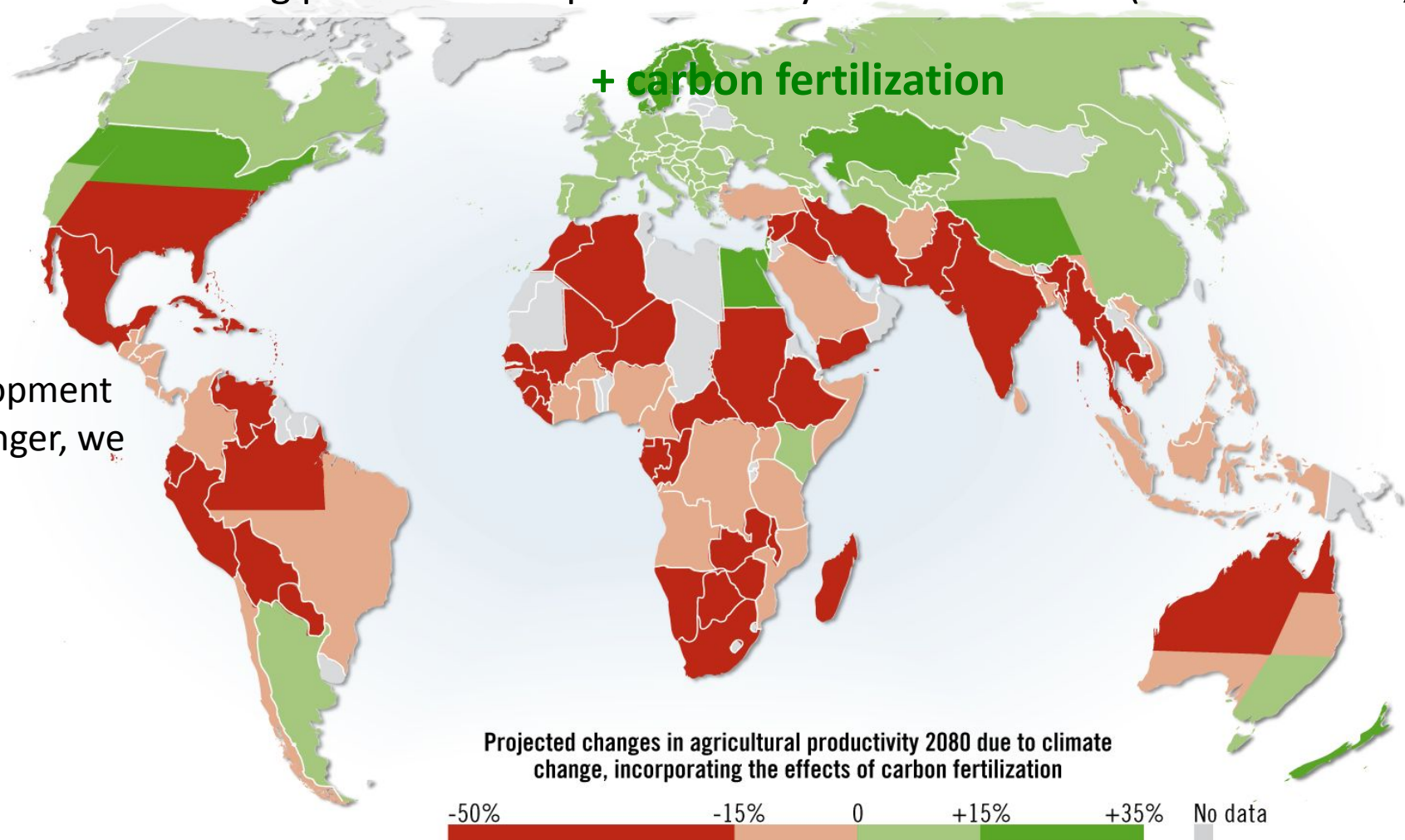


- The issue of food security is on top of the global agenda
- Food security is challenged by two main factors:
 - Population growth and
 - Climate change
- Human population has increased from one to seven billion just over the past 200 years
 - About 700% increment
- Expected to reach 11 billion by the end of this century



Environmental stresses are reducing potential food production by 2% each decade (IPCC AR5 WGII, 2014).

To meet the Sustainable Development Goal of bringing about zero hunger, we need to boost our food grain production by 70%

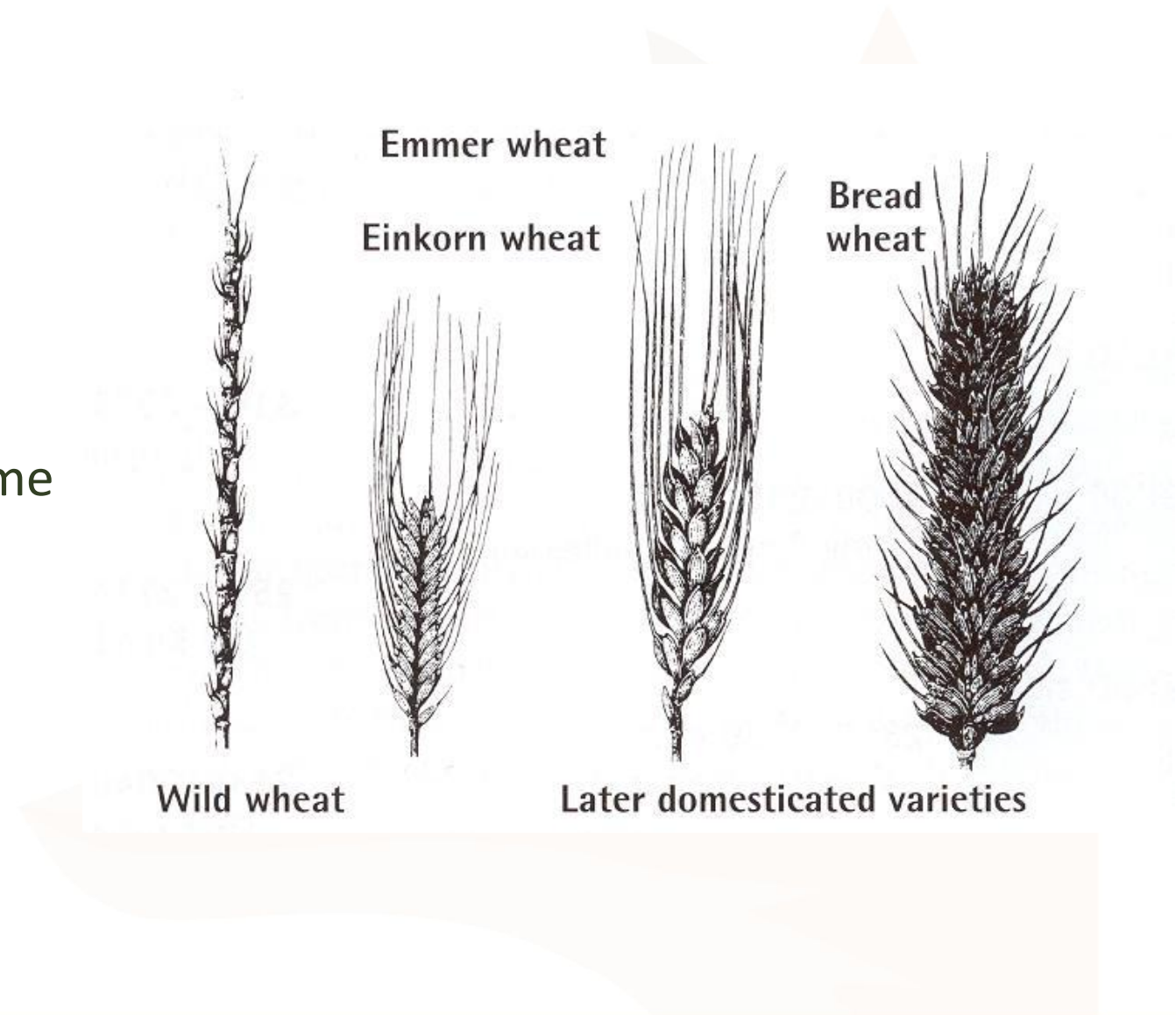


Map by Hugo Ahlenius, GRID-Arendal (2008).Source: Cline W. (2007, 2008). Global Warming and Agriculture.

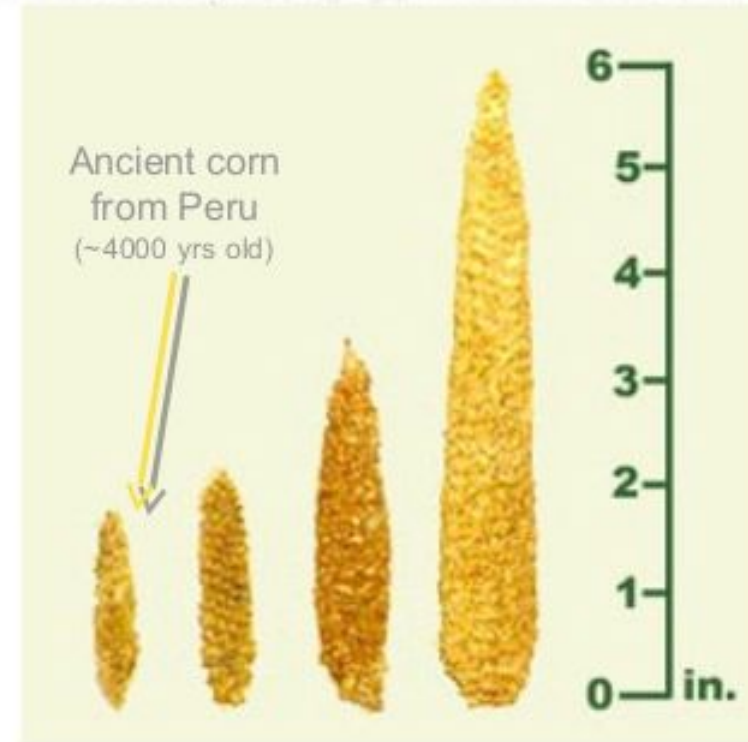


- **WE NEED CROPS:**
 - with higher yields
 - with higher nutritional values
 - adaptation to changing environments

- We were selecting the traits of our interest
 - Genetic homogenization
 - Loss of genetic diversity and some adaptive traits
 - To diseases and
 - Extreme conditions



- We were selecting the traits of our interest
 - Genetic homogenization
 - Loss of genetic diversity and some adaptive traits
 - To diseases and
 - Extreme conditions



- We were selecting the traits of our interest
 - Genetic homogenization
 - Loss of genetic diversity and some adaptive traits
 - For resistance to diseases and
 - Extreme conditions
- The domesticated gene pool (genetic diversity in crops and breeding lines) is limited by the “domestication bottleneck”



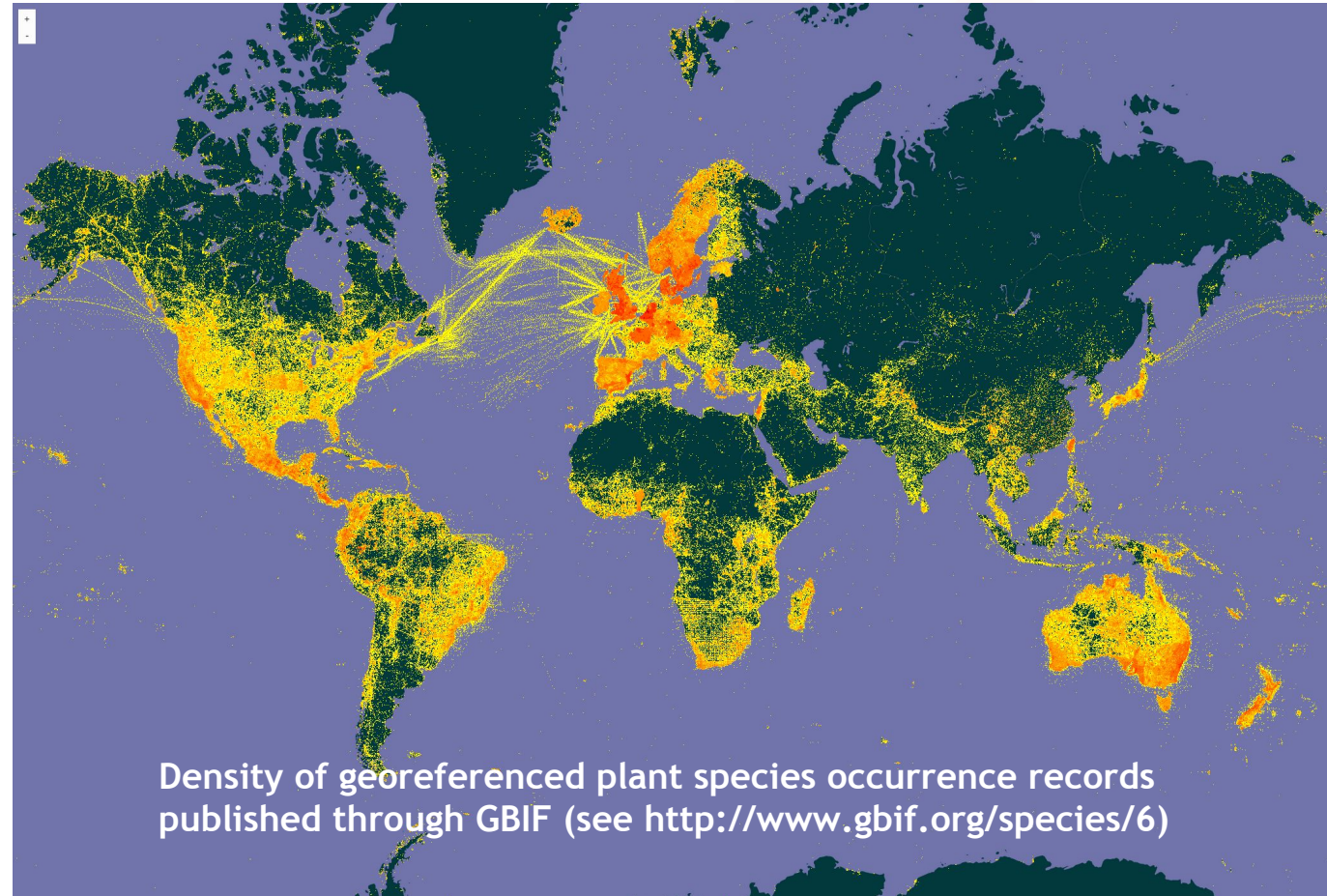


- Untapped genetic diversity can be found in:
 - Traditional cultivars
 - Landraces
 - **Crop Wild Relatives**

- Large genetic diversity (gene pool) is required to tackle current challenges
- Crop wild relatives provides an untapped resource for genetic diversity
- They have been through selective pressure as they have to go through several environmental conditions
- Some are adapted to:
 - Drought conditions
 - Water logged areas
 - Warmer temperature
 - Colder temperature
 - Saline soil
- Some are resistant to insect and fungal pests as well as other diseases
- They have important genetic resources that are needed to cope with the changing globe and to feed the ever growing population



- Crop wild relatives are wild plant species closely related to crops.
- By broad definition it refers to all taxa within the same genus as a crop (Maxted et al. 2006)
- They account for around 21% of the world flora (Maxted and Kell 2009)

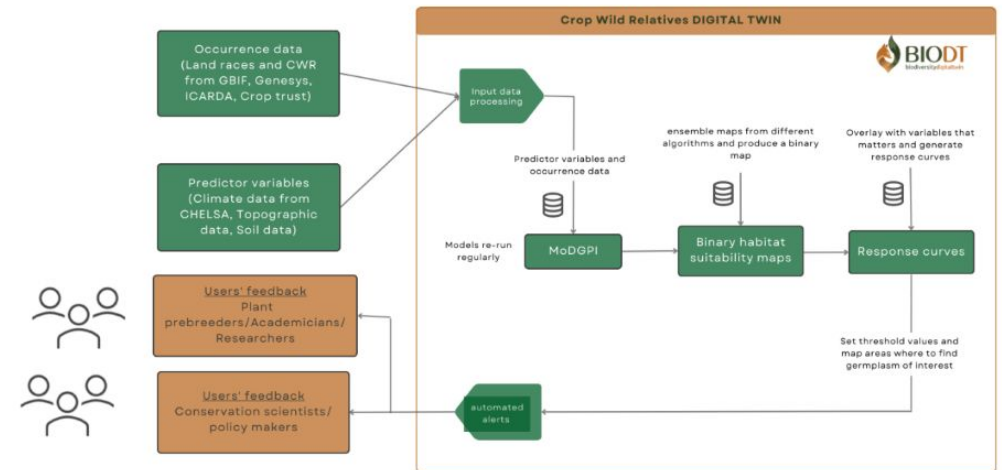


Objective

- Creating a modelling tool that facilitates the search for germplasm of interest from CWR and traditional cultivars gene pool to improve domesticated crops

MoDGP: modelling the distribution of germplasms of interest

- MoDGP uses different high performing species distribution modelling algorithms
 - GAM, RF, BRT and MaxEnt to produce habitat suitability maps of model targets.
- Different model settings and model replications
- Results from multiple model runs will be evaluated and maps from high performing ones will be selected.

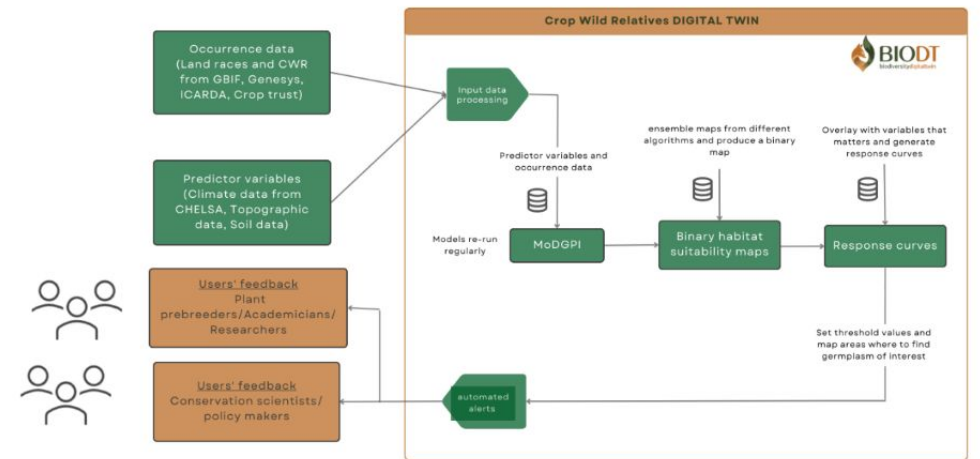


MoDGP (Modelling the distributions of germplasms of interest)

- It uses multiple algorithms of species distribution models to produce habitat suitability of landraces and crop wild relatives
- Maps from high performing algorithms will be ensemble and classified into suitable and unsuitable classes
- Values of environmental variables that predict germplasm of interest will be extracted at each of the grid cells (pixels) that present suitable classes and pixel counts will be plotted as a response curve and threshold of values will be decided
- user interface enables users to make modifications to the threshold values and make decisions on which germplasm to test or to collect

MoDGP: modelling the distribution of germplasms of interest

- Combining the maps through an ensemble approach and producing a binary map with suitable/unsuitable classes
- The suitable class of the binary maps will be overlaid with the environmental variable that predict a germplasm of interest.
- Each crop wild relatives will be ranked based on their range of tolerances to these environmental factors.
- For model targets with high tolerance to these factors, geographic areas where plants presenting the desired genotypes are growing will be mapped and provided.

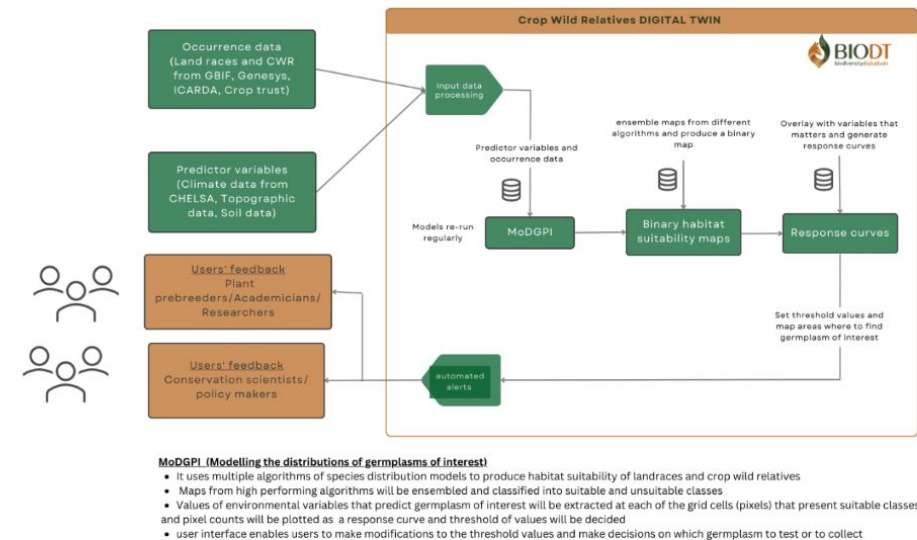


MoDGPI (Modelling the distributions of germplasms of interest)

- It uses multiple algorithms of species distribution models to produce habitat suitability of landraces and crop wild relatives
- Maps from high performing algorithms will be ensemble and classified into suitable and unsuitable classes
- Values of environmental variables that predict germplasm of interest will be extracted at each of the grid cells (pixels) that present suitable classes and pixel counts will be plotted as a response curve and threshold of values will be decided
- user interface enables users to make modifications to the threshold values and make decisions on which germplasm to test or to collect

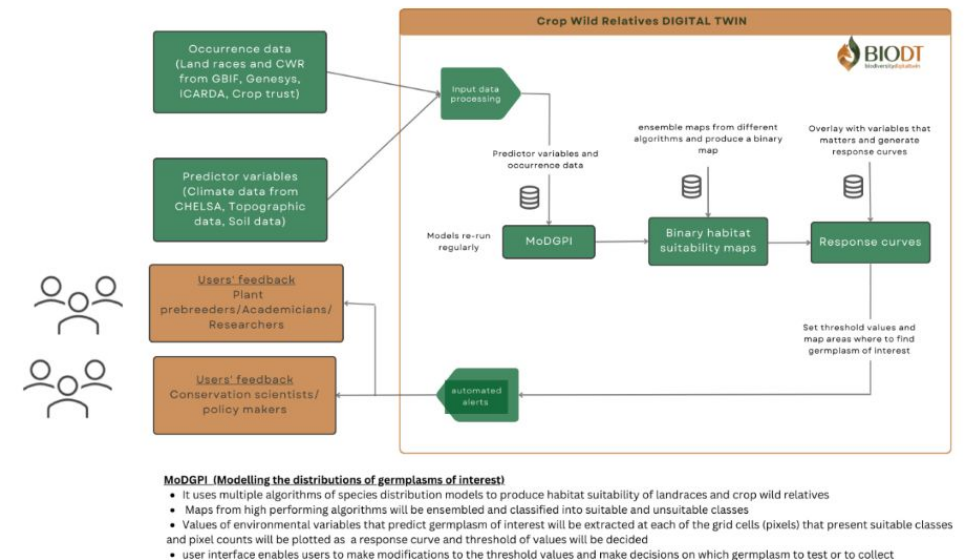
MoDGP: modelling the distribution of germplasms of interest

- Input:
- I) Occurrence data from GBIF and potentially also from other sources such as ICARDA, Genesys and crop trusts.
- II) Environmental variables as such as climate, soil and topographic data are used as response variables in raster format.

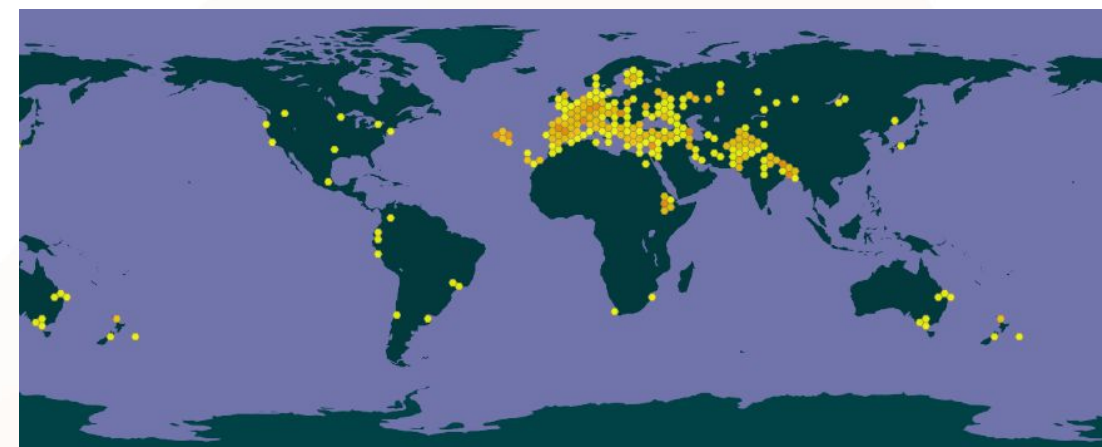


MoDGP: modelling the distribution of germplasms of interest

- Output:
- three outputs:
 - I) habitat suitability maps of each crop wild relatives and landraces;
 - II) tolerances to environmental factors that are of importance to the traits of interest in the form of response curves, with cutoff thresholds;
 - III) areas that potentially present genotypes of interest, in the form of maps.
 - IV) access accessions from gene banks and ranks them (for landraces)



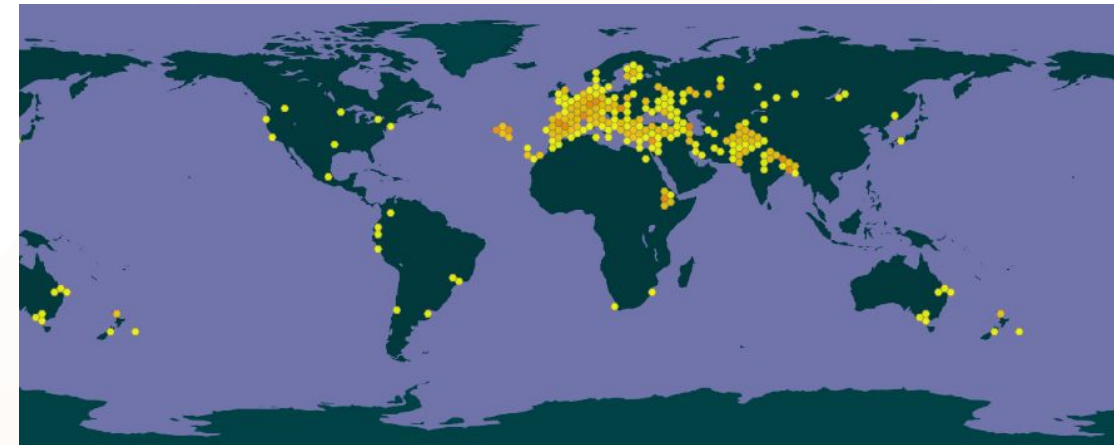
- Belongs to the family Fabaceae
- It can withstand extreme environments from drought to flooding
- Cropped after the main cropping season
- If crops fail, the same farm lands can be covered by it
 - This makes it a climate smart species
- Thus grasspea is often the only alternative to starvation when other crops fail



GBIF

Global Biodiversity
Information Facility

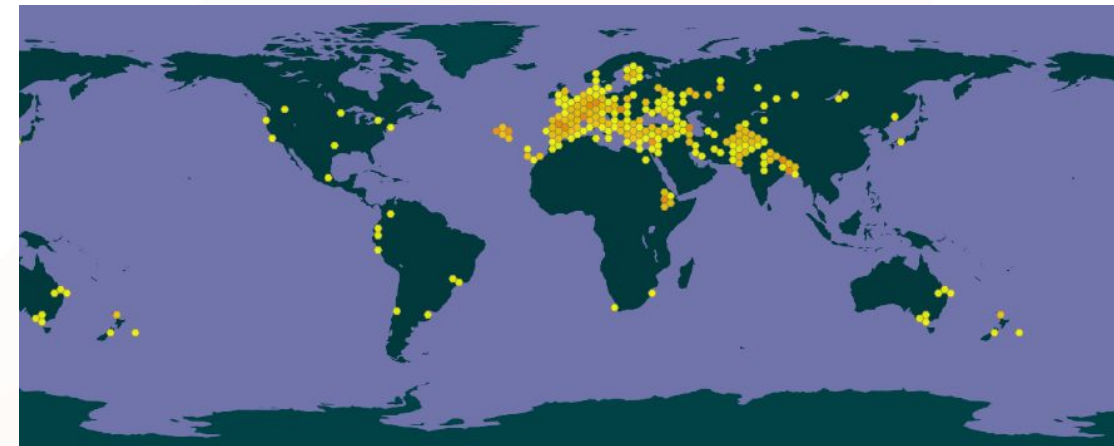
- It fixes nitrogen and serve as a natural nitrogen fertilizer
 - Can grow on degraded land
- It is very rich in protein
- It also serves as a fodder
- Considered as a super crop to beat protein malnutrition in the future



GBIF

Global Biodiversity
Information Facility

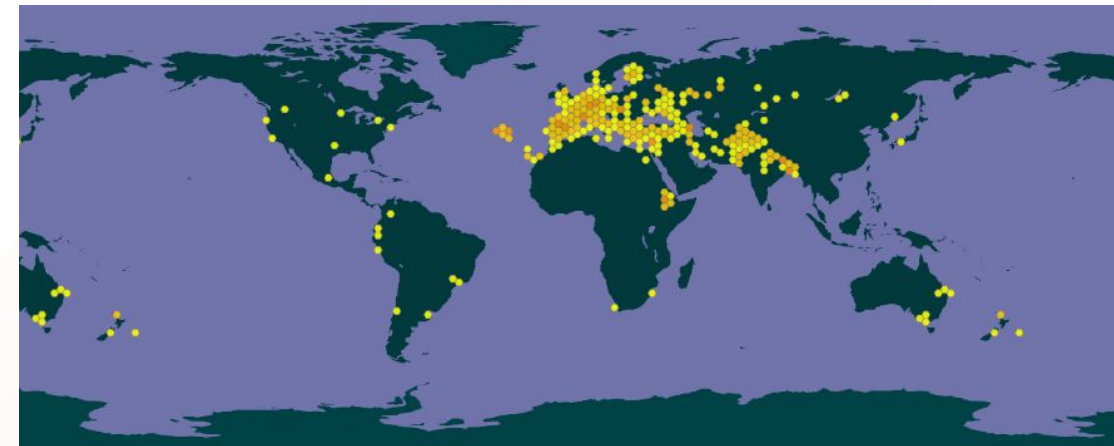
- presents a fascinating paradox – it is not only a life saver but also a destroyer as well
- It has a neurotoxin chemical
- when eaten as a large part of the diet over a long period, it can cause
 - Permanently paralyze of the lower limbs in adults - lathyrism
 - Brain damage in children
- This is often the case during famine periods



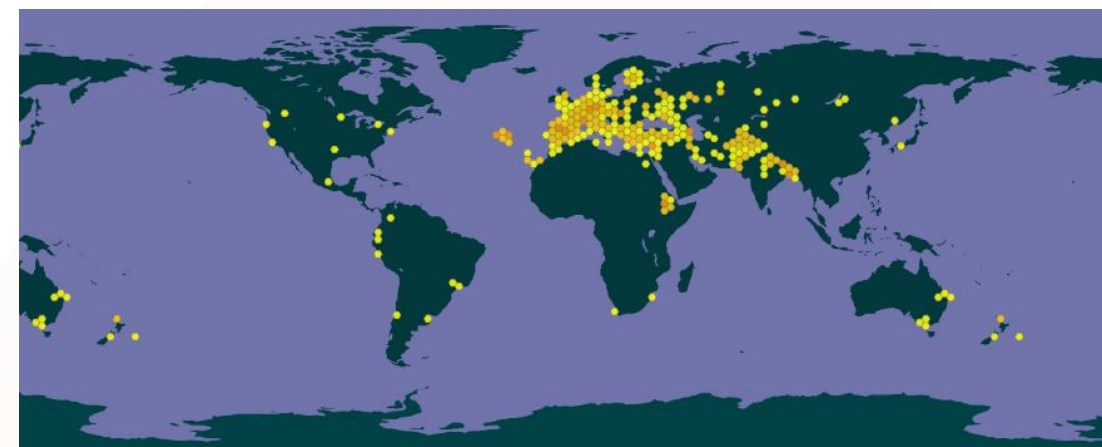
GBIF

Global Biodiversity
Information Facility

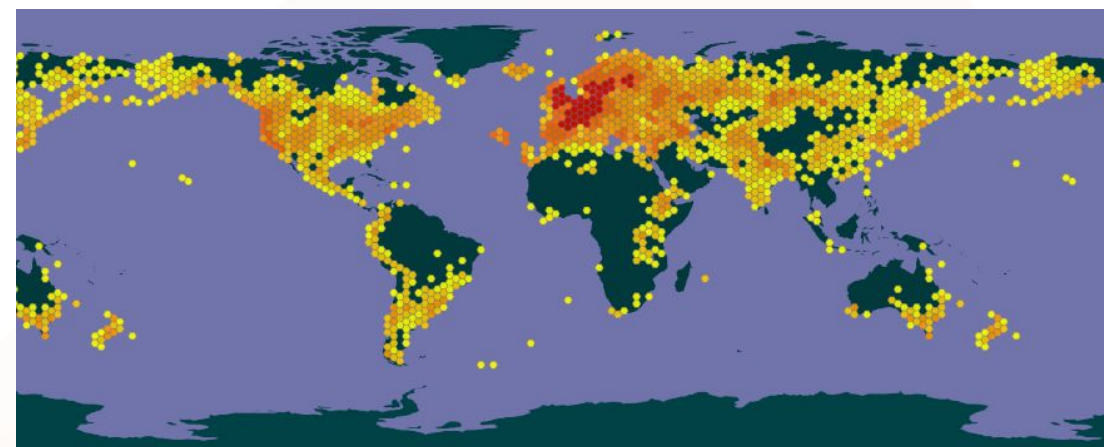
- The toxicity of grasspea is affected by different factors such as
 - water stress,
 - soil zinc content,
 - salinity
- Toxicity increases with water and zinc deficiency and with salinity.
- This means the same genotype may present different levels of toxicity under different environmental conditions, complicating the matter.

**GBIF**Global Biodiversity
Information Facility

- We hypothesize that
 - 1) grasspea landraces and wild relatives growing in dry areas and in zinc deficient and acidic soils are most likely efficient in water, zinc and sodium uptake; and
 - 2) through improving these efficacies of grasspea, it is possible to minimize production of the neurotoxin.
- Thus, we are aiming to model the habitat suitability of grasspea wild relatives and identify genotypes with adaptation to these environmental factors

**GBIF**Global Biodiversity
Information Facility

- Quite widely distributed
 - Diverse land races and farmers' varieties
- More than 180 herbaceous species belonging to the same genus
 - Diverse wild relatives
 - Easy to experiment with them
- Huge genetic resources
 - Certain varieties from western Asia have a low level of neurotoxin



Model targets

- 86 species out of about 180 species have reasonably good data
- with occurrence data ranging from 42 to ~20,000
- pseudo-absence/absence: 1e5 randomly selected points where the model target is absent but other species are present

```
Abs_1s <- lapply(unique(uniLoc2$species), FUN = function(x){
  NotSpecies_df <- uniLoc2[uniLoc2$species != x, ]
  SampledAbs <- sample(1:nrow(NotSpecies_df), size = 1e5, replace = FALSE)
  Report_df <- NotSpecies_df[SampledAbs,c("decimalLongitude", "decimalLatitude")]
})
```



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
42	134	386	6853	1980	222059

- Number of occurrence – 1159 unique points

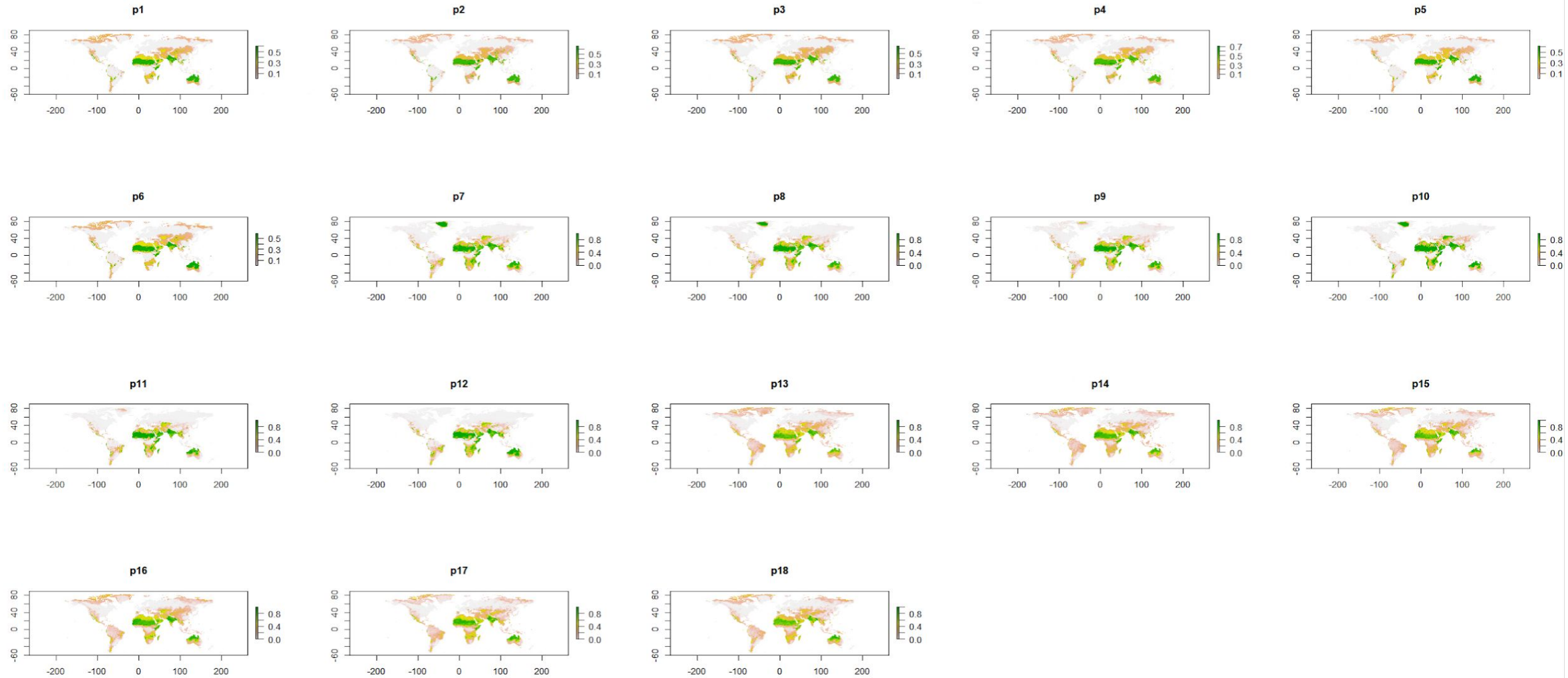
```
m = sdm(presence~., d, methods = c ("maxent","gbm","GAM","RF"),
  replications = c("sub", "boot"), test.p = 25, n = 3,
  parallelSetting = list(ncore = 4, method = "parallel"))
```

```
class                                : sdmModels
=====
number of species                    : 1
number of modelling methods          : 4
names of modelling methods           : maxent, brt, gam, rf
replicate.methods (data partitioning) : subsampling,bootstrap
number of replicates (each method)   : 3
total number of replicates per model : 6 (per species)
test percentage (in subsampling)      : 25
-----
model run success percentage (per species) :
-----
method      presence
-----
maxent      :      100 %
brt         :      100 %
gam         :      100 %
rf          :      100 %

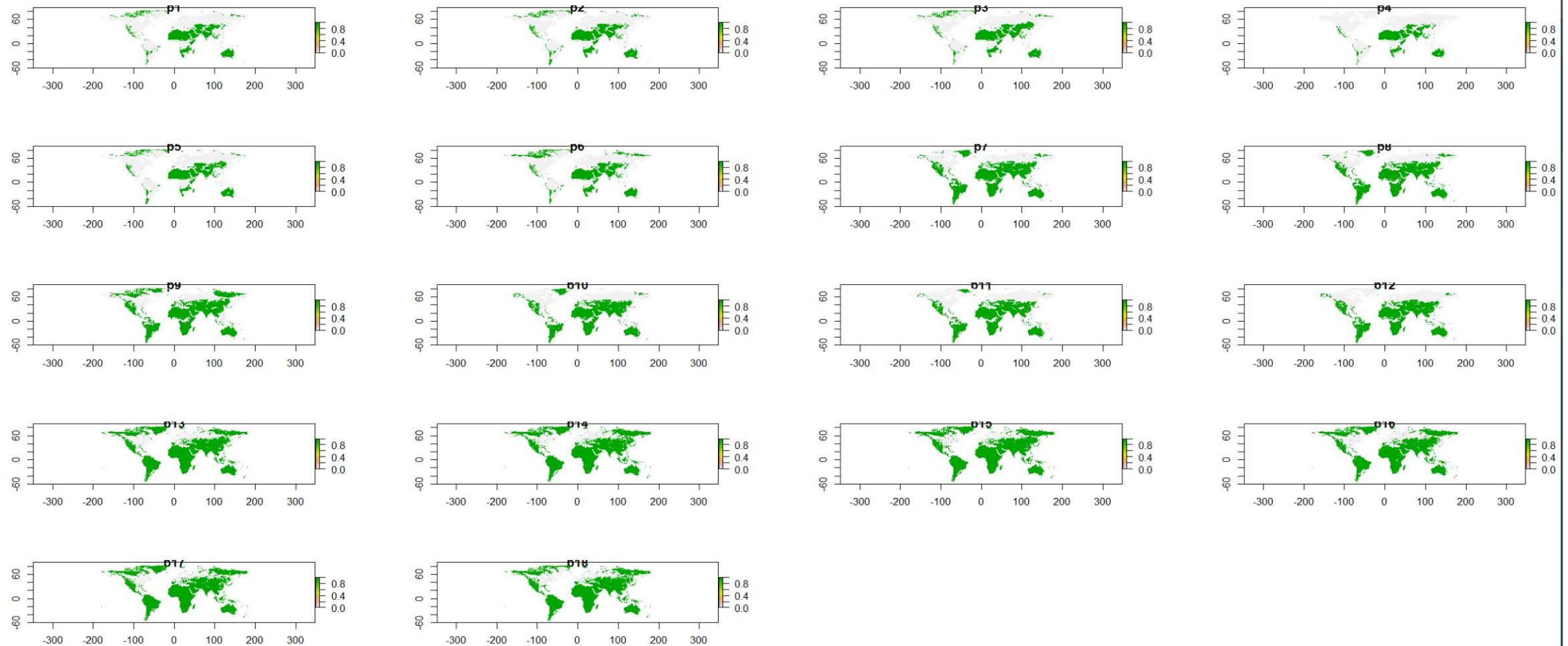
#####
model Mean performance (per species), using test dataset (generated using partitioning):
-----

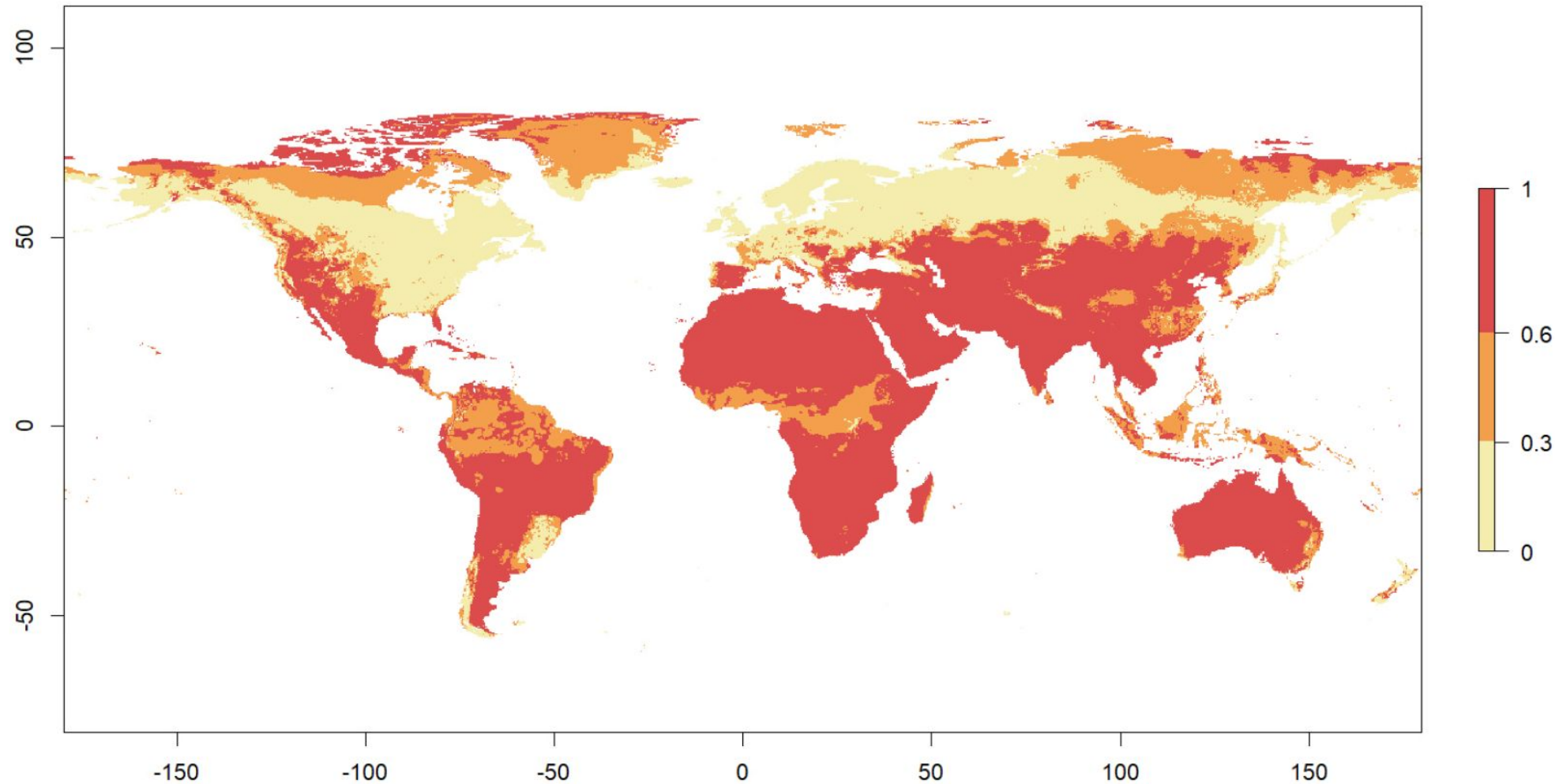
## species   :  presence
=====

methods      :      AUC      |      COR      |      TSS      |      Deviance
-----
maxent       :      0.92      |      0.36      |      0.71      |      0.29
brt          :      0.9       |      0.48      |      0.68      |      0.09
gam          :      0.92      |      0.47      |      0.71      |      0.08
rf           :      0.96      |      0.7       |      0.81      |      0.06
```

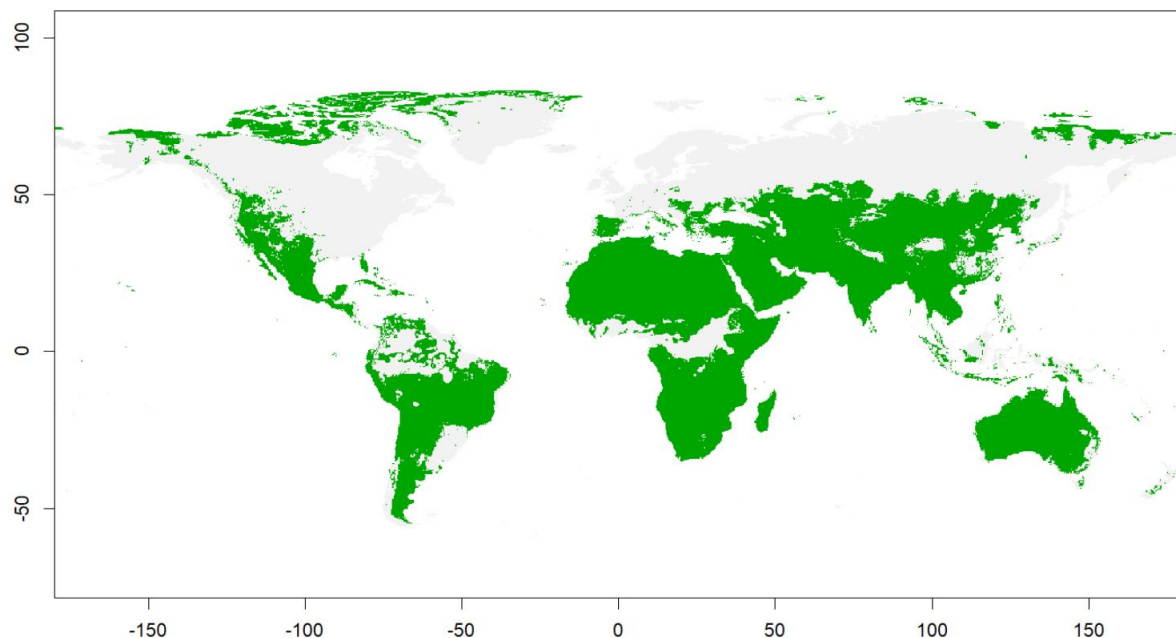



Binary suitable/unsuitable maps (*Lathyrus sativus*)

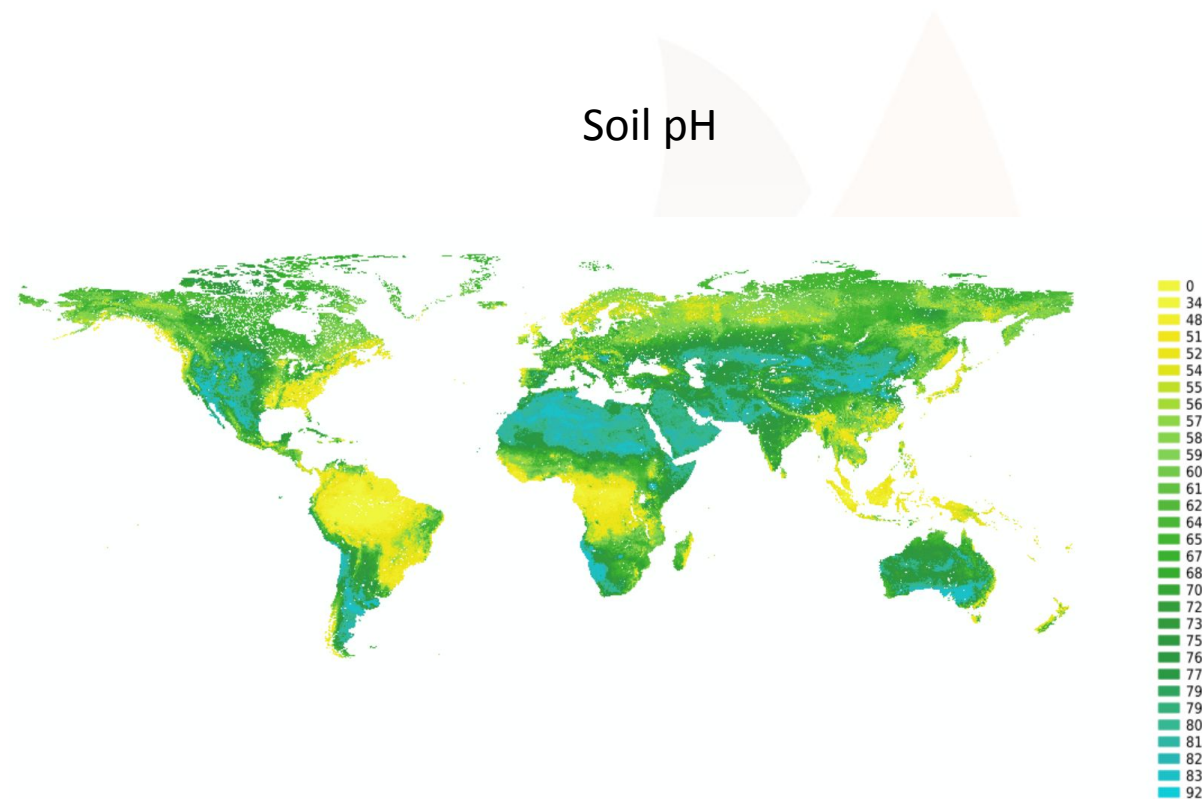




Habitat suitability



Soil pH



Habitat suitability for all the 86 species
 Extracting soil zinc content, soil moisture,
 pH values for suitable pixels
 Comparing ranges of tolerances
 Proposing candidate accessions to be tried (for landraces)
 Proposing candidate populations to be tried (for CWR)

We are planning to provide graphical user interphase to enable users to make their alternative decisions

We are aiming to make it transferable across traits and crops

Additional tasks

- CAPFITOGEN - Ecological Land Characterization (ELC) approach
 - Ecotype mapping to assist conservation
- Predictive characterization
- FIGS - Focused Identification of Germplasm Strategy
 - Very useful when evaluation data is available.
- MoDGP

- We can use the Digital Twin service for:
 - Automation of data flow (data fusion from distributed data sources)
 - Dynamic model updating (e.g. feeding in updated environmental data, followed by new model iteration)
 - Automated model uncertainty analysis (comparisons with real-life data)
 - Automated alert for new genetic diversity with predicted interesting target genetic properties (desired alleles)



Questions and comments

- It can make a nice study by itself
- Where is the digital twin part of it?
 - What is a digital twin? A digital version of an object/a process per microsecond? per minute, per day, per seven days, per month, per six months, per year? per year? Or what?
 - The engineering and factory definition ...we should have a definition from biodiversity point of view.
 - careful differentiation between industry & science DTs
- Projecting to future climate ...
 - We should focus on the central objective –the contemporarily available resource to improve cropsclimate change implication may be helpful but it comes with a lot of uncertainties...
 - how do DTs feedback already at the present to the real-world? DT should give feedback to real world.
- Do you exclude cities and some land covers where crops are not growing
 - The wild relatives can thrive everywhere.
 - idea: think of different layers of feedback (e.g. direct vs. indirect; information, alerts, ...)



BIODT
biodiversitydigitaltwin

 @BiodiversityDT

 BioDT



Funded by
the European Union