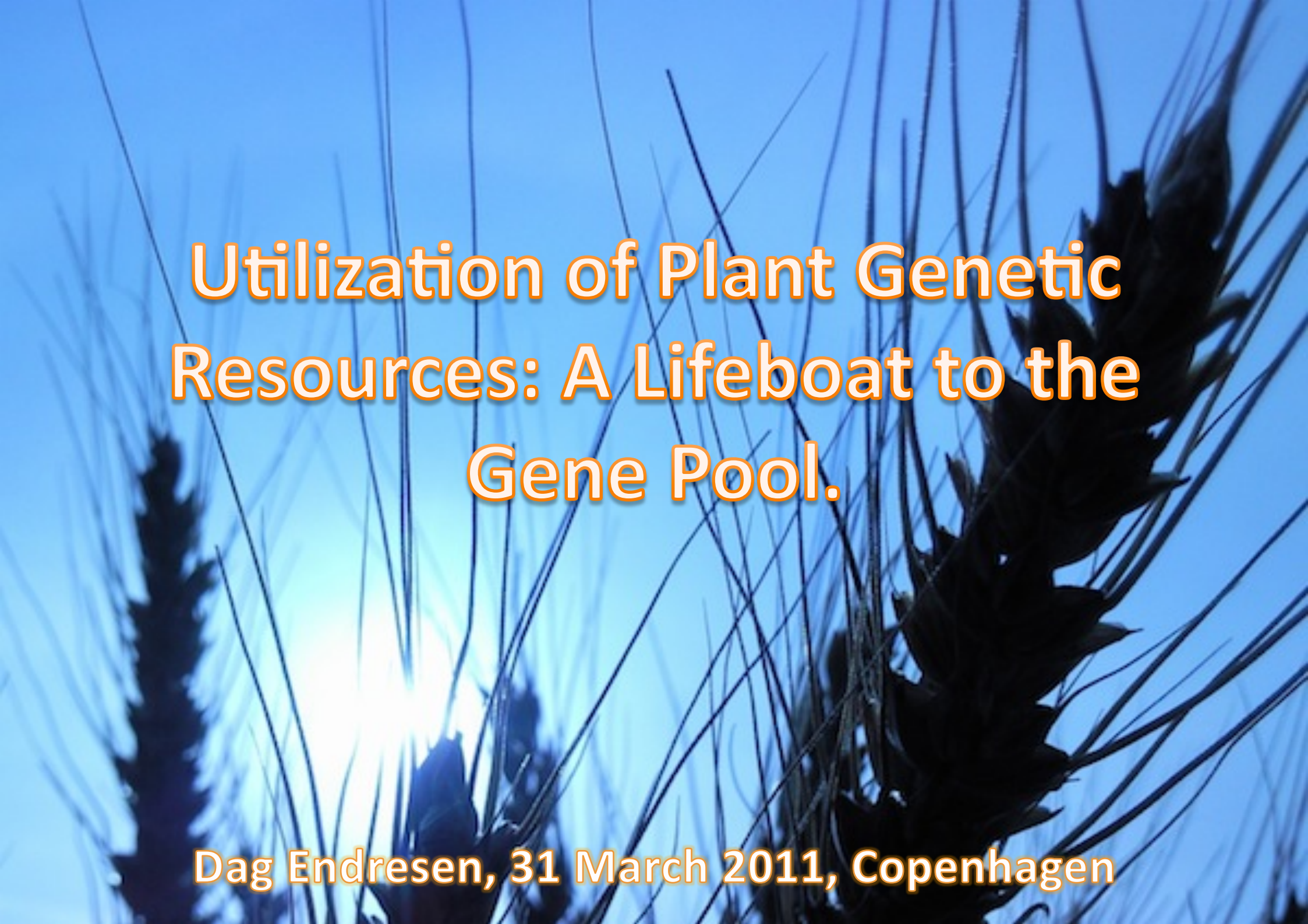


Utilization of Plant Genetic Resources: A Lifeboat to the Gene Pool.

Dag Endresen, 31 March 2011, Copenhagen

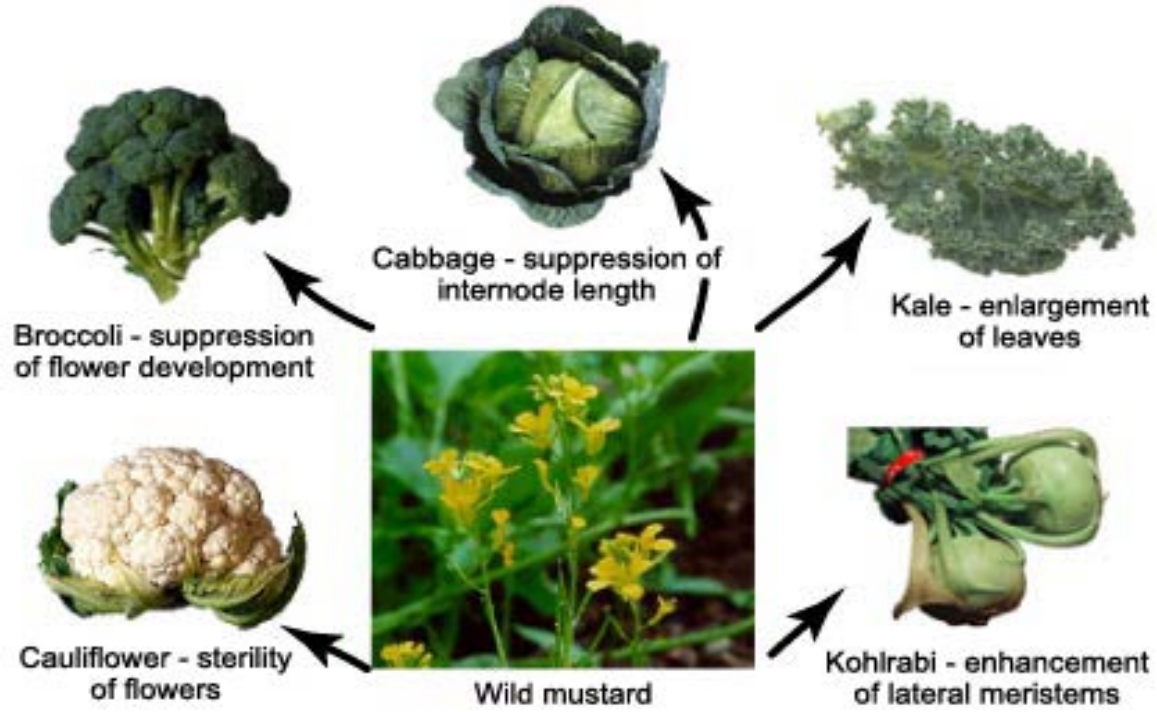
TOPICS:

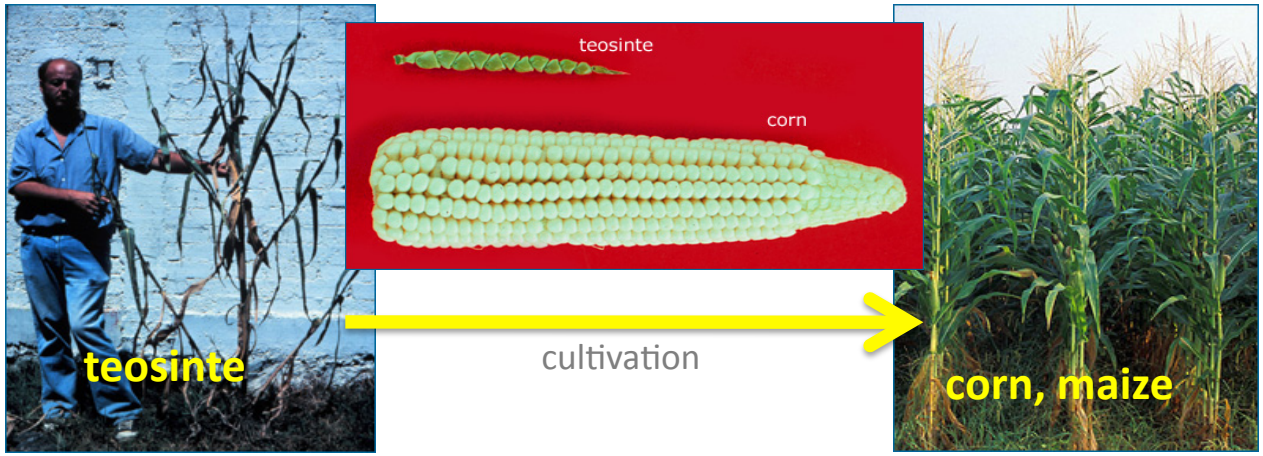
- Data mobilization and sharing
  - Darwin Core extension for genebanks
- Trait mining with FIGS
  - Predictive link between climate data and trait data
  - Case studies:
    - Morphological traits in Nordic barley
    - Biotic stress traits in wheat and barley
    - Blind prediction of stem rust, Ug99 in bread wheat landraces

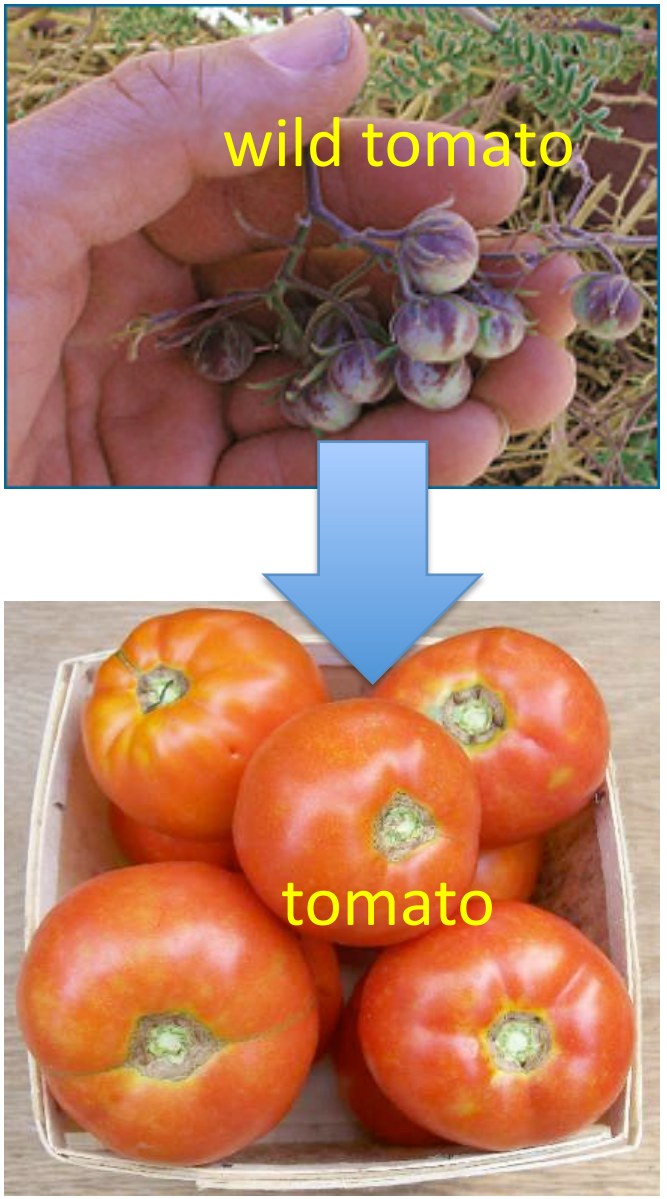
Wheat at Alnarp, June 2010

Domestication and cultivated plants:

Utilizing genetic potential from the wild







Ex situ Genebank collections for plant genetic resources

Seed drying room

Seed containers

Seed store

Household freezers

SESTO | Trait (C&E) | References | Utilities | User Manual | Links | Order material | [NordGen Home]

Genus | Taxons | Cultivars | Accessions | Dataset | [Search]

SESTO

2011

Taxon

Scientific Name	Authority	Family	English name	IT PGR Annex 1
Bromus inermis	Leyss.	Poaceae	Awnless Brome	FALSE

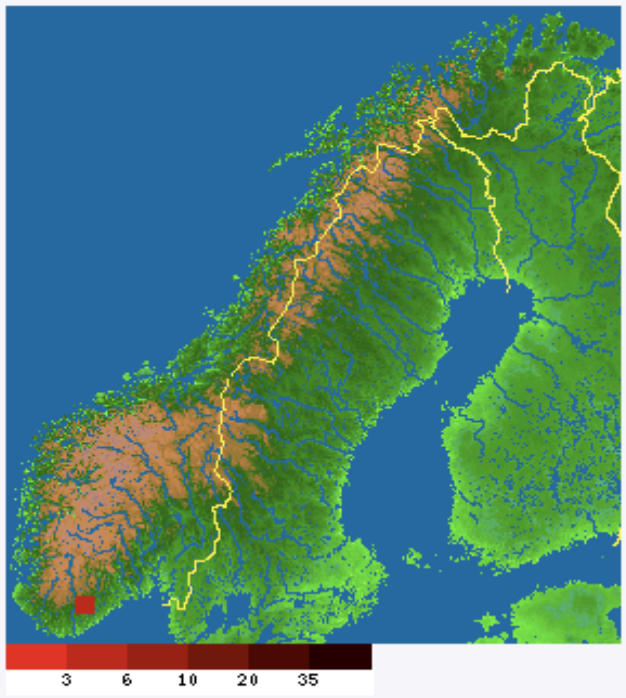
Accession

Accession name	Accession number	Mandate	GeneBank	Origin country	Cultivar type
15-8-65-2	NGB13366	ACC	NORDGEN	Norway	P

Donor AccNam	Donor AccNum	Donor country
BLADFAKS	15-8-65-2	Norway

Dataset: Ordinary collection

Accession details | Stored material | Collecting site | Evaluation | Transactions | Images



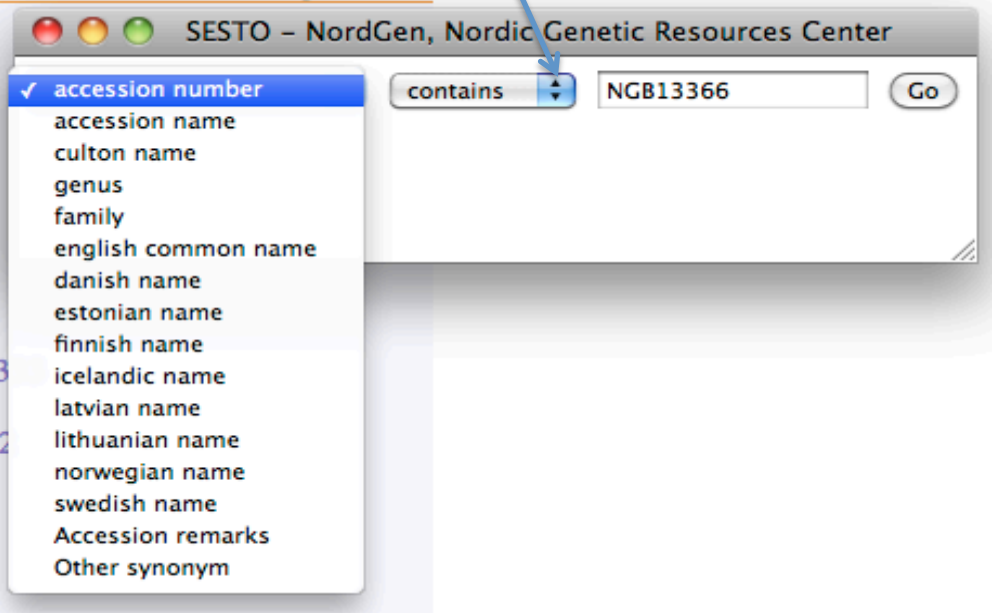
Change map size: [normal] [medium] [big] [large]
Change map region: [Sudan] [Scandinavia] [Europe] [Eurasia]
[Denmark] [Finland] [Iceland] [Norway] [Sweden]
[Estonia] [Latvia] [Lithuania] [Africa]

Map application © GPL 2.0 [NGB]

Collecting place for
accnum: NGB13366

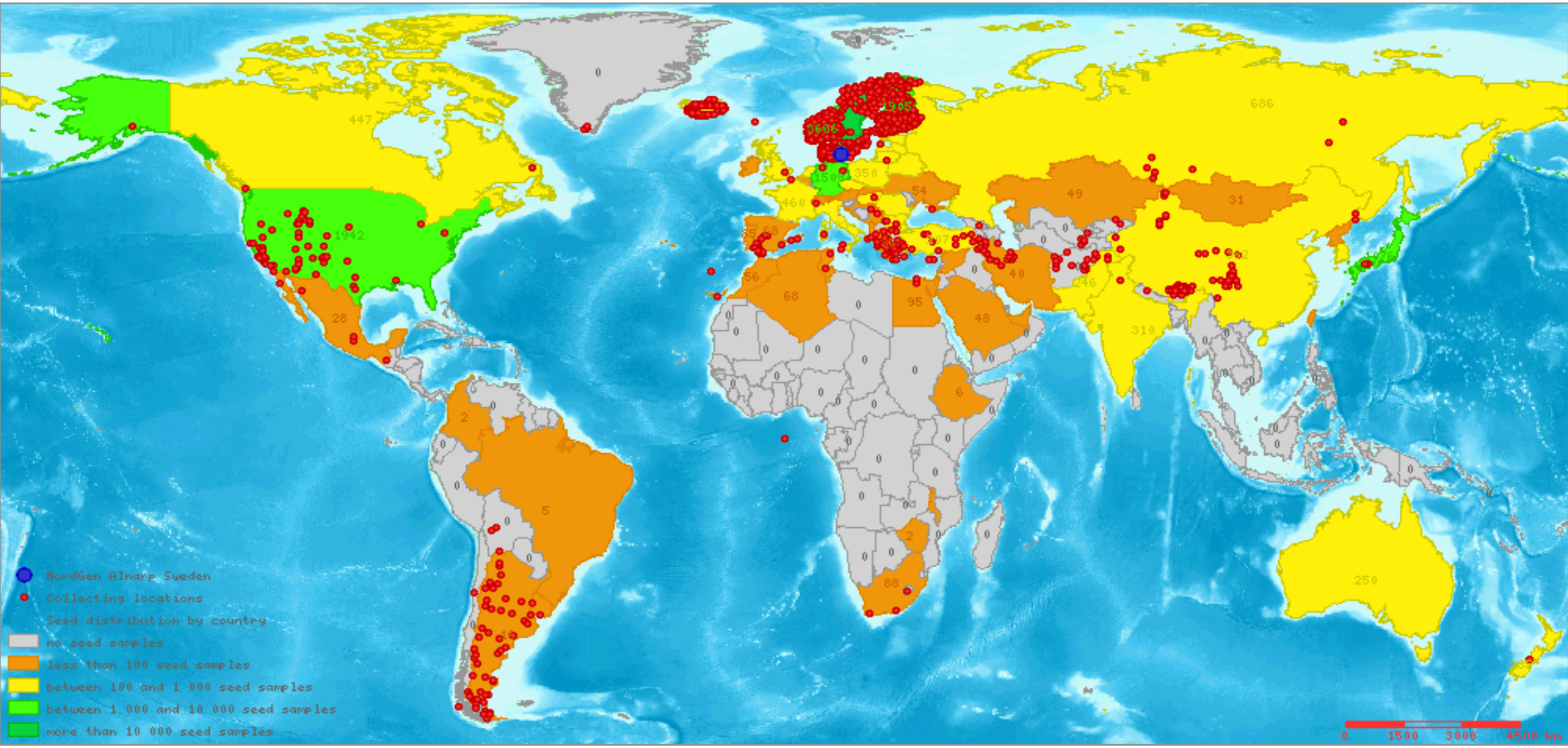
Place name: Birkenes
Collecting date: 19810901
Latitude: 58°33'--'' N
Longitude: 008°08'--'' E
Latitude decimal: 58.55
Longitude decimal: 8.133333333333
Elevation: 250 meter
Collector accession number: 15-8-65-2
:
Source population
:
Collecting Site Description
Commune: Birkenes
:
Collecting Site Habitat
Special habitat: ROA
Landform: SLO
Aspect direction: S
Slope percent: 3%

Total of 1 accessions [inside the map region]



http://sesto.nordgen.org

ORIGIN VERSUS USE (SEED REQUESTS)

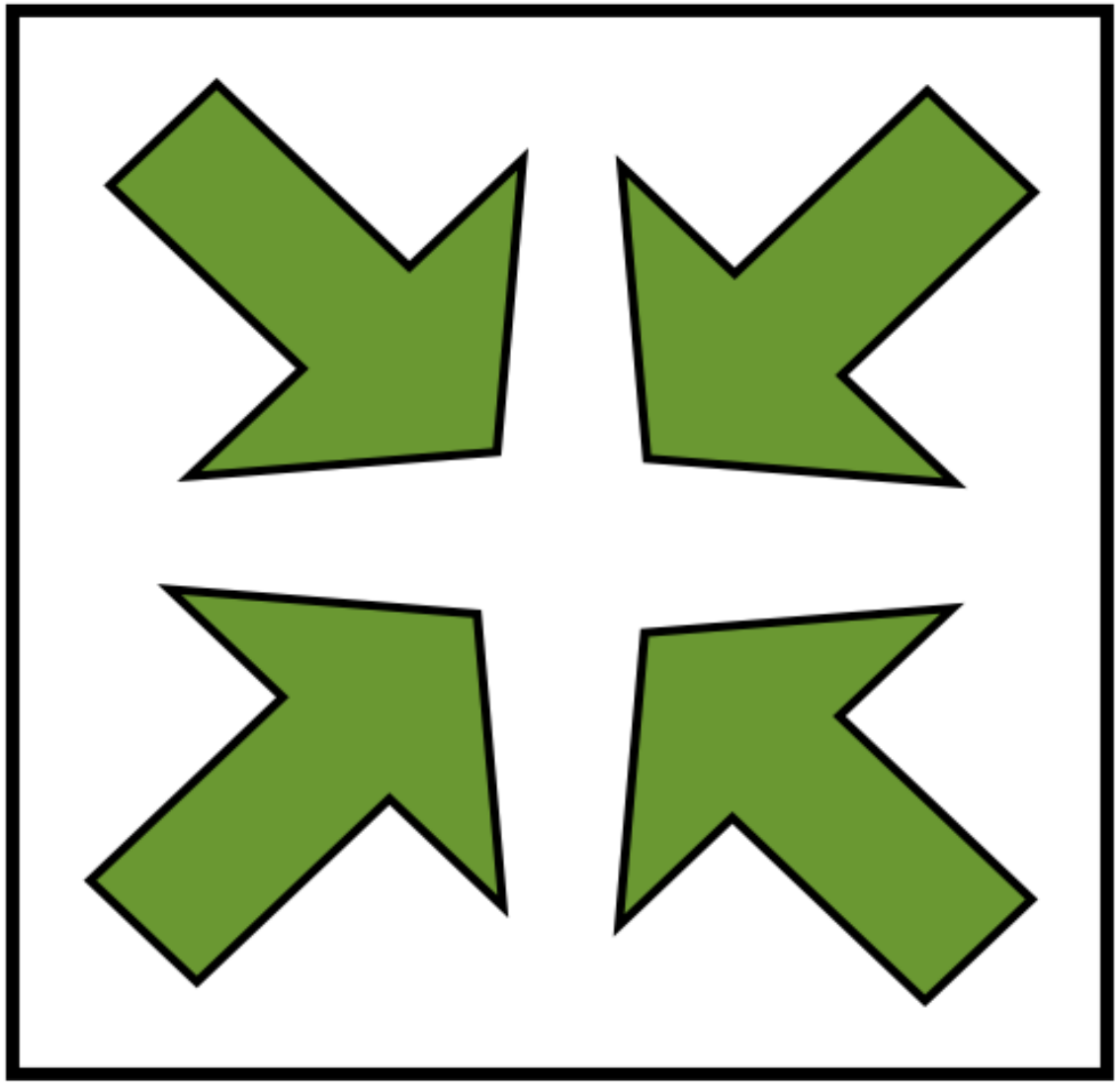


SESTO distribution and georeferenced accessions

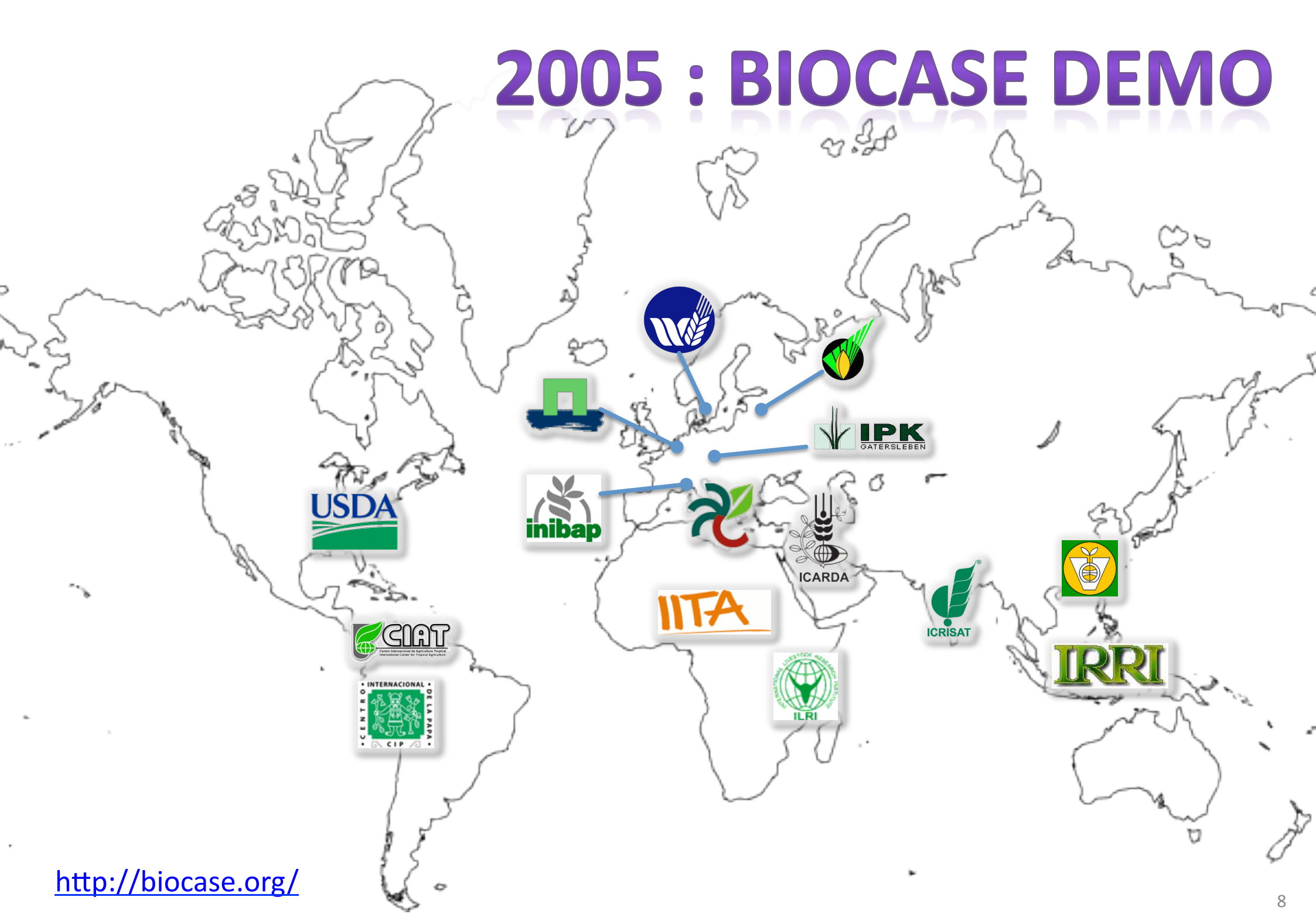
Red dots are the georeferenced collecting places
Countries are colored by accessions DISTRIBUTED

Genebank material primarily originating from the Nordic region – seed requests primarily from the same region

DATA SHARING



2005 : BIOCASE DEMO



http://biocase.org/

2010 : IPT INSTALLATIONS FOR EURISCO

- EURISCO
- NordGen (Nordic countries)
- IPK Gatersleben (Germany)
- BLE (Germany)
- WUR CGN (The Netherlands)
- CRI (Czech Republic)
- VIR (Russian Federation)
- Baltic (Estonia, Latvia, Lithuania)
- Bioversity-Montpellier (France)

GBIF Publishing Toolkit
http://ipt.gbif.org

2009: THE DARWIN CORE EXTENSION FOR GENEBANKS

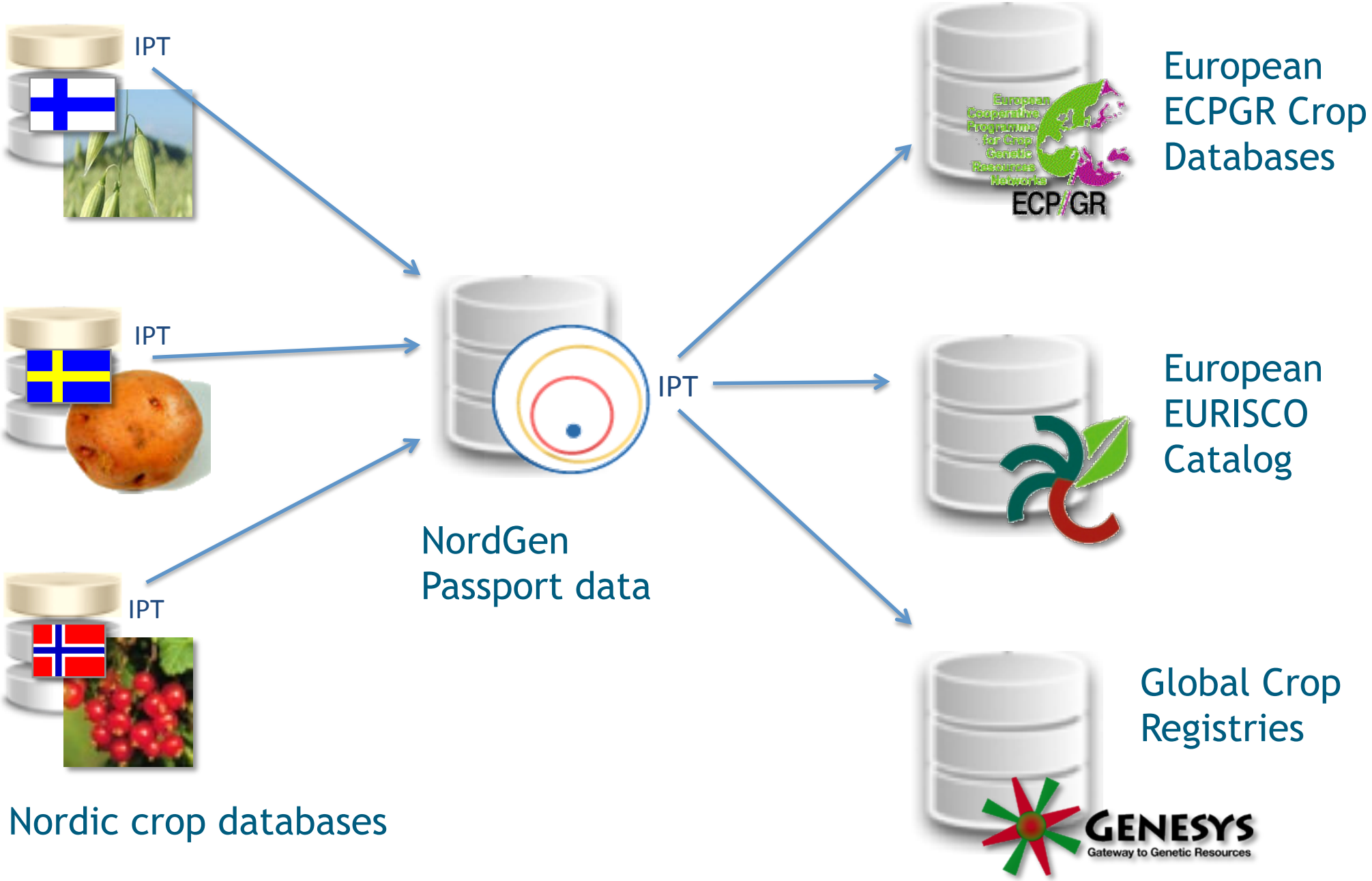
- "MCPD in Darwin Core"
- Includes the new terms for crop trait experiments developed as part of the European EPGRIS3 project
- Includes a few additional terms for new international crop treaty regulations

http://code.google.com/p/darwincore-germplasm

http://rs.nordgen.org/dwc

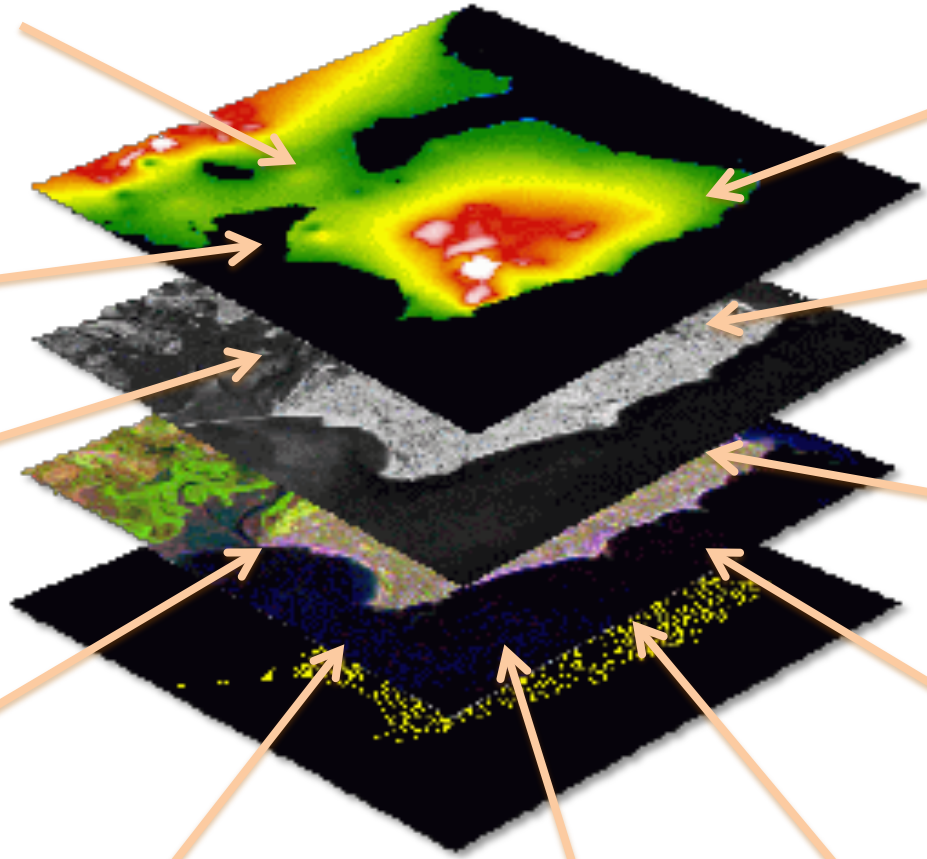
Endresen, D.T.F. and H. Knüpffer (2011). The Darwin Core extension for genebanks opens up new opportunities for sharing genebank datasets. *Submitted to Biodiversity Informatics.*

POSSIBLE UPGRADED GENEBANK NETWORK MODEL IN EUROPE



(NB! Proposal, not currently implemented using the GBIF IPT)

MOVING TOWARDS… GLOBAL INTEGRATION OF INFORMATION



Genebank datasets

OSGeo
Spatial data

ICARDA

IUCN
Threatened species

Bioversity International
Crop standards

GBIF
Global Biodiversity Information Facility

CBD

FAO
FIAT PANIS

CMS
Migratory species

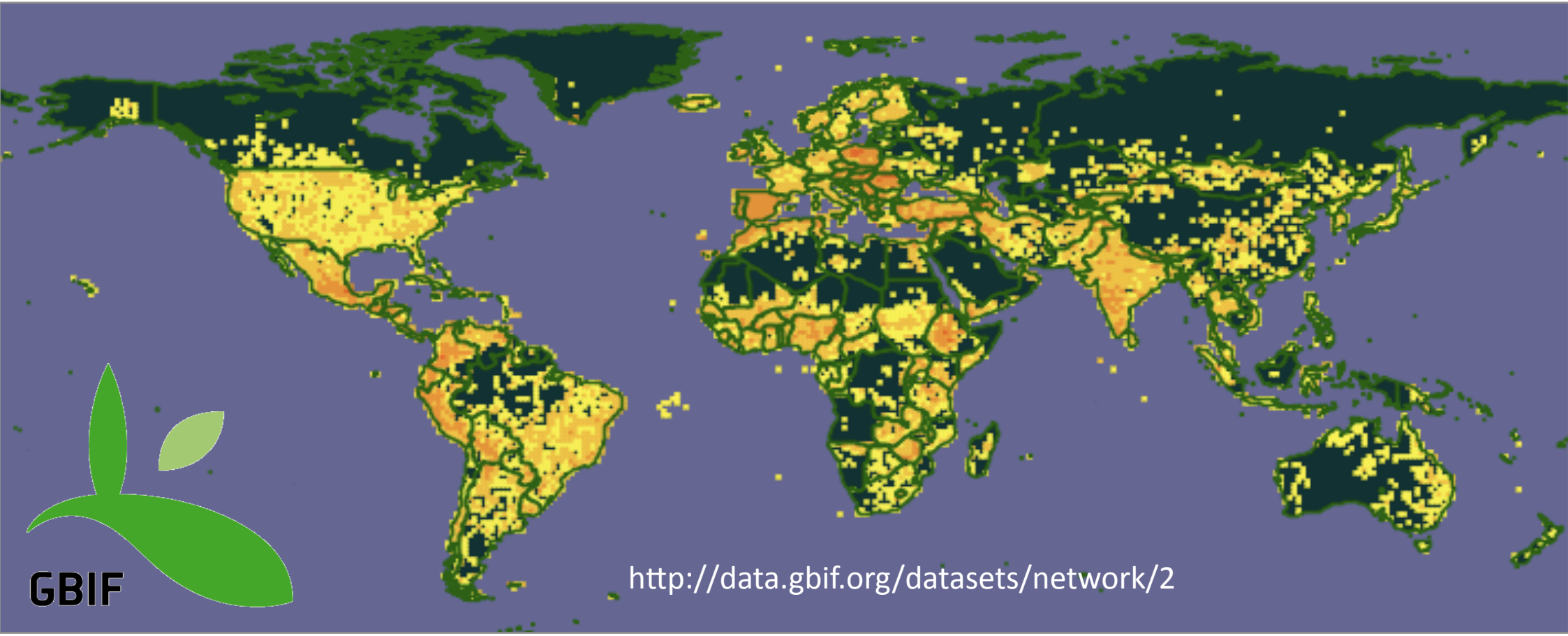
Legislation and regulations etc.

CGIAR
Global crop collections

European Cooperative Programme for Crop Genetic Resources Networks
ECP/GR
Crop collections in Europe

GENESYS
Gateway to Genetic Resources
Global crop system

POTENTIAL OF THE GBIF TECHNOLOGY



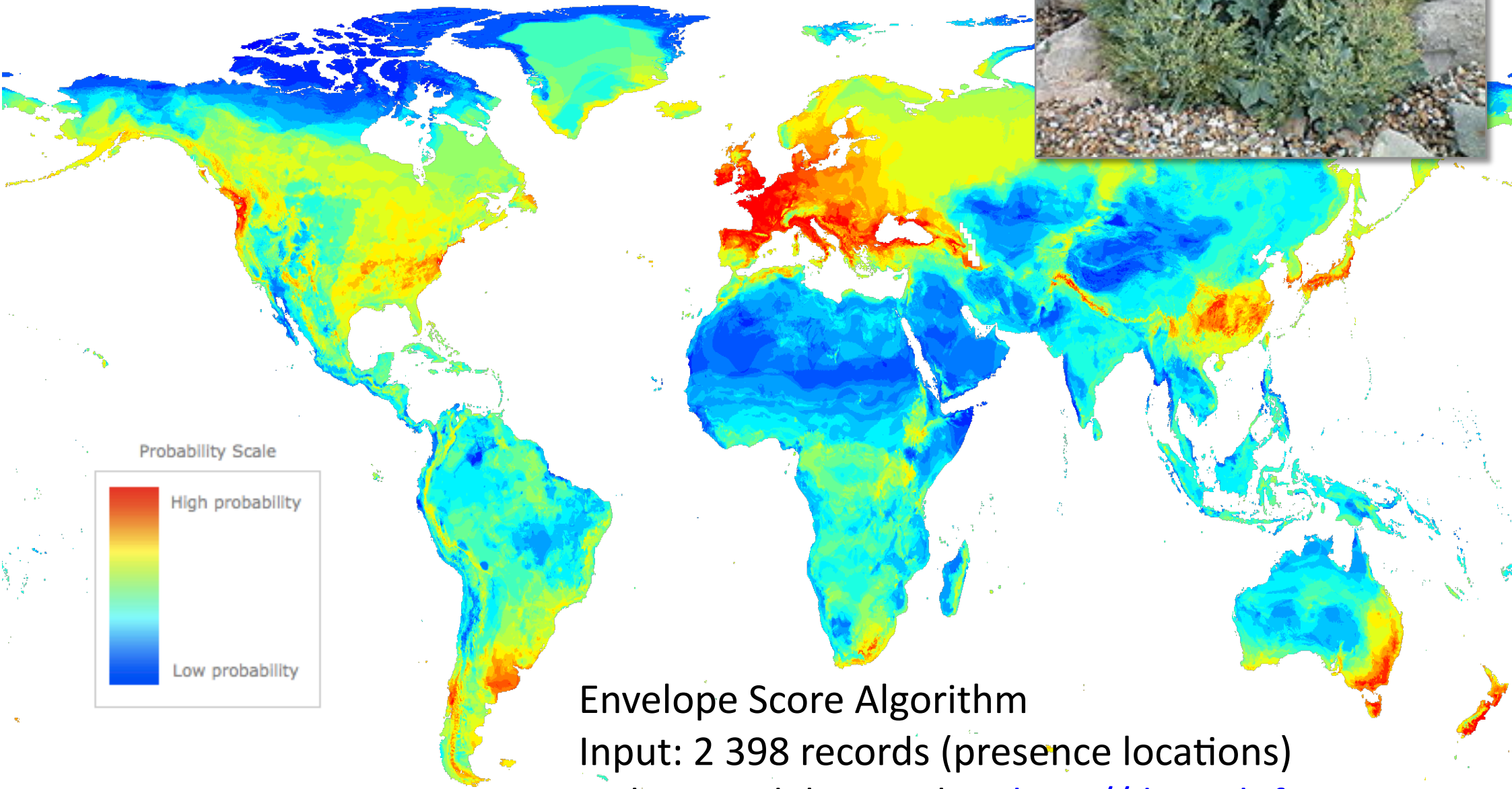
Using GBIF/TDWG technology (and contributing to its development), the PGR community can more easily establish specific PGR networks without duplicating GBIF's work.

The compatibility of data standards between PGR and biodiversity collections made it possible to integrate the worldwide germplasm collections into the biodiversity community (TDWG, GBIF).

GAP ANALYSIS

- Identify gaps in the gene bank collections
- Maximize the conserved genetic diversity

SEA KALE (*CRAMBE MARITIMA* L.)
NORDGEN STUDY: JUNE 2010



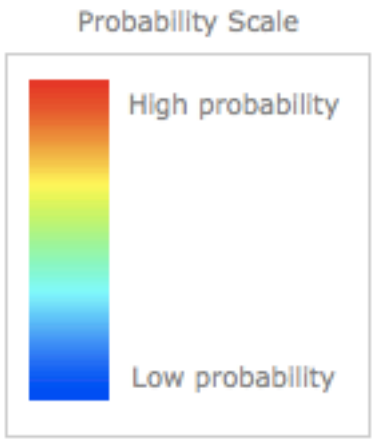
Envelope Score Algorithm
Input: 2 398 records (presence locations)
Online modeling tool at: http://data.gbif.org

WORMWOOD (*ARTEMISIA ABSINTHIUM* L.)
NORDGEN STUDY: JUNE 2010


Wormwood (*Artemisia absinthium* L.)

Species distribution model
(7 364 records)

Using the Maxent desktop software.



GAP ANALYSIS TO COMPLEMENT GENEBANK COLLECTIONS

Objectives of Gap analysis:

- Advice the planning of new collecting/gathering expeditions
- Identification of relevant areas were the crop species is predicted to be present
- Focus on areas least well represented in the genebank collection (maximize diversity)

See for example http://gisweb.ciat.cgiar.org/GapAnalysis/ for more information.

TRAIT MINING

- Eco-geographic data analysis
- Focused Identification of Germplasm Strategy (FIGS)
- Identify useful target traits for crop improvement

A NEEDLE IN A HAY STACK

- Scientists and plant breeders want a few hundred germplasm accessions to evaluate for a particular trait.

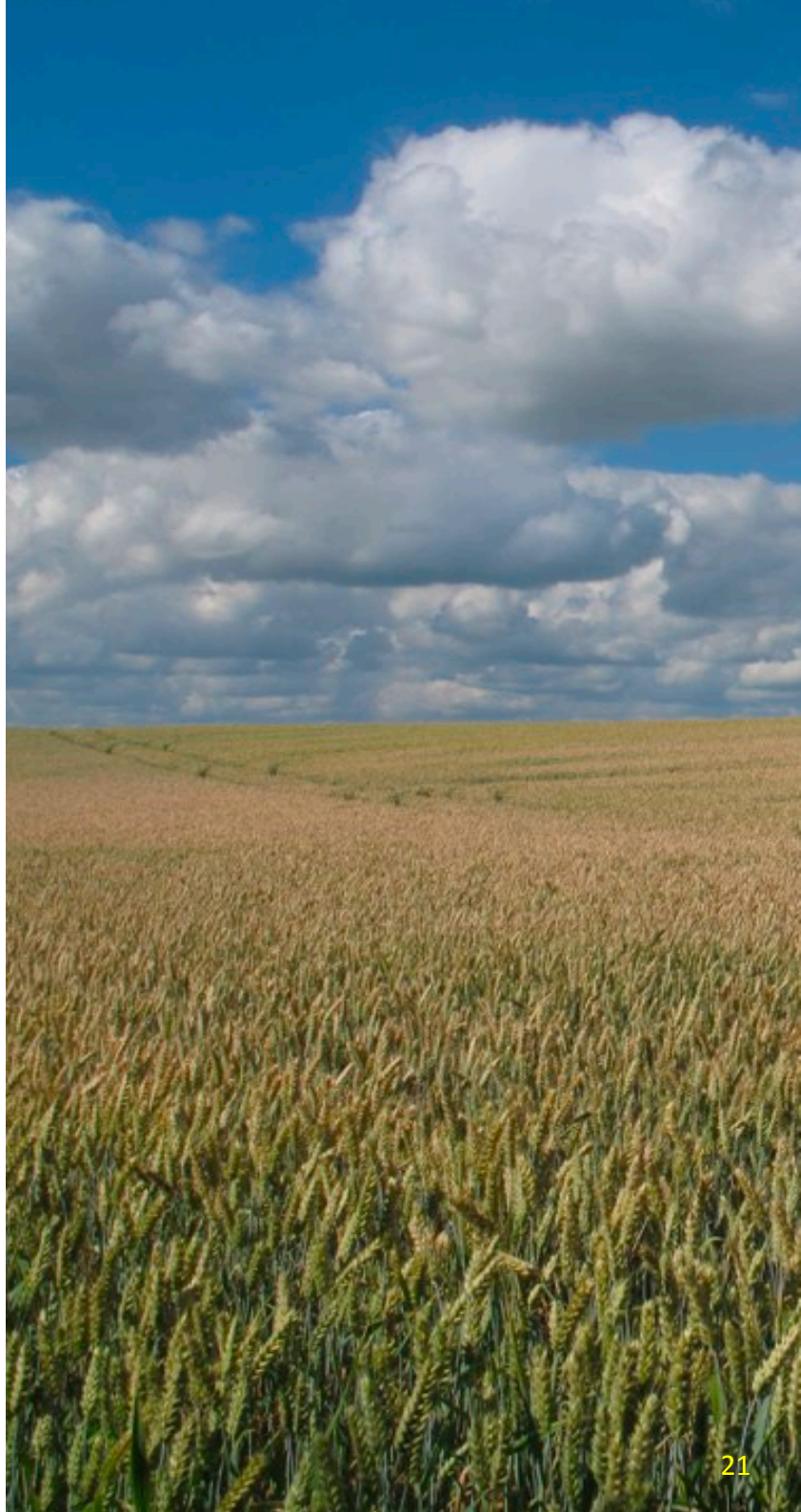
- How does the scientist select a small subset likely to have the useful trait?

Slide topic by Ken Street, ICARDA
FIGS team

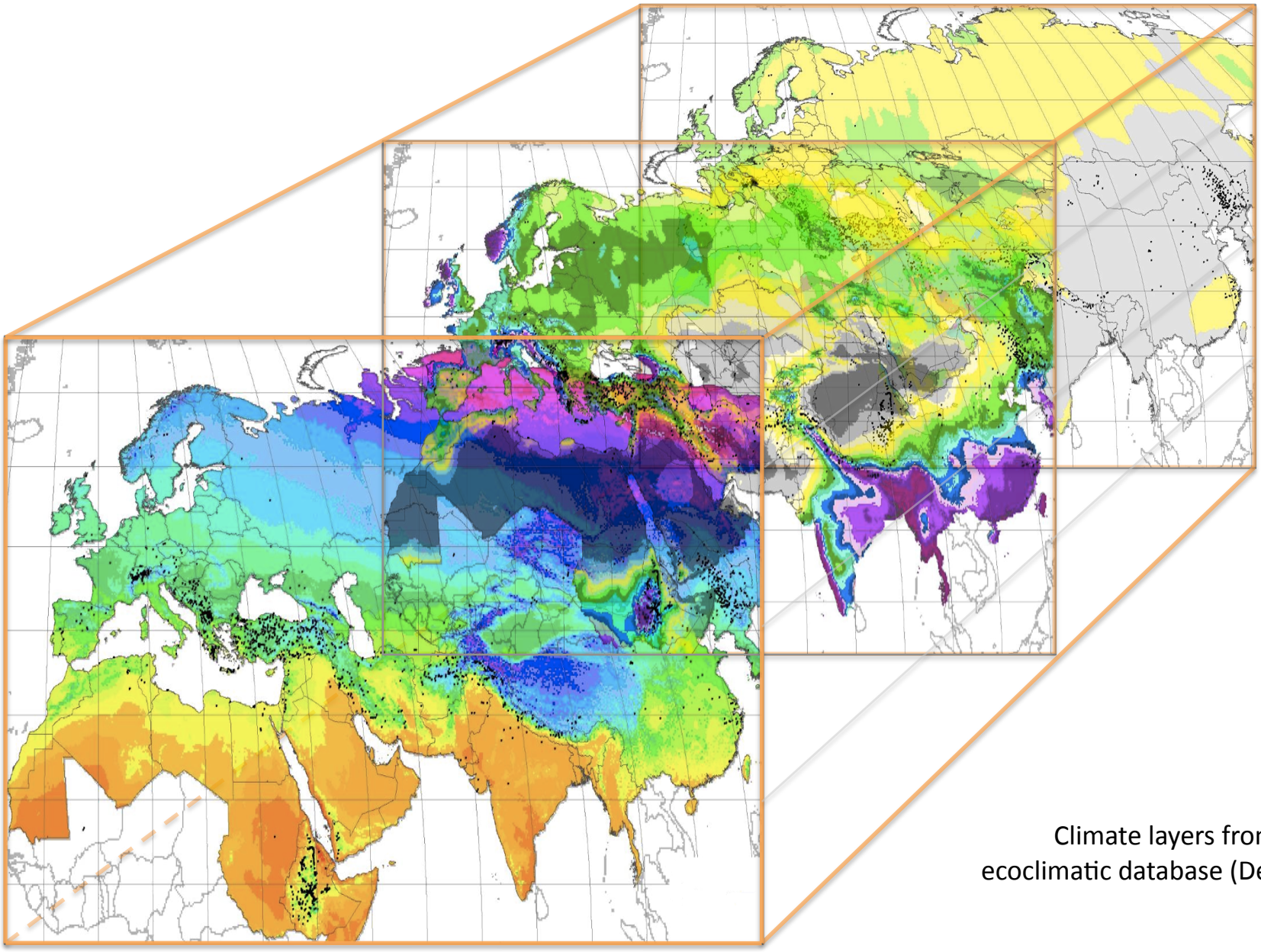

Challenges for utilization of plant genetic resources
* Large gene bank collections
* Limited screening capacity

OBJECTIVES OF FIGS

- Using climate data for prediction of crop traits BEFORE the field trials.
- Identification of landraces with a higher probability of holding an interesting trait property.



FOCUSED IDENTIFICATION OF GERMPLASM STRATEGY



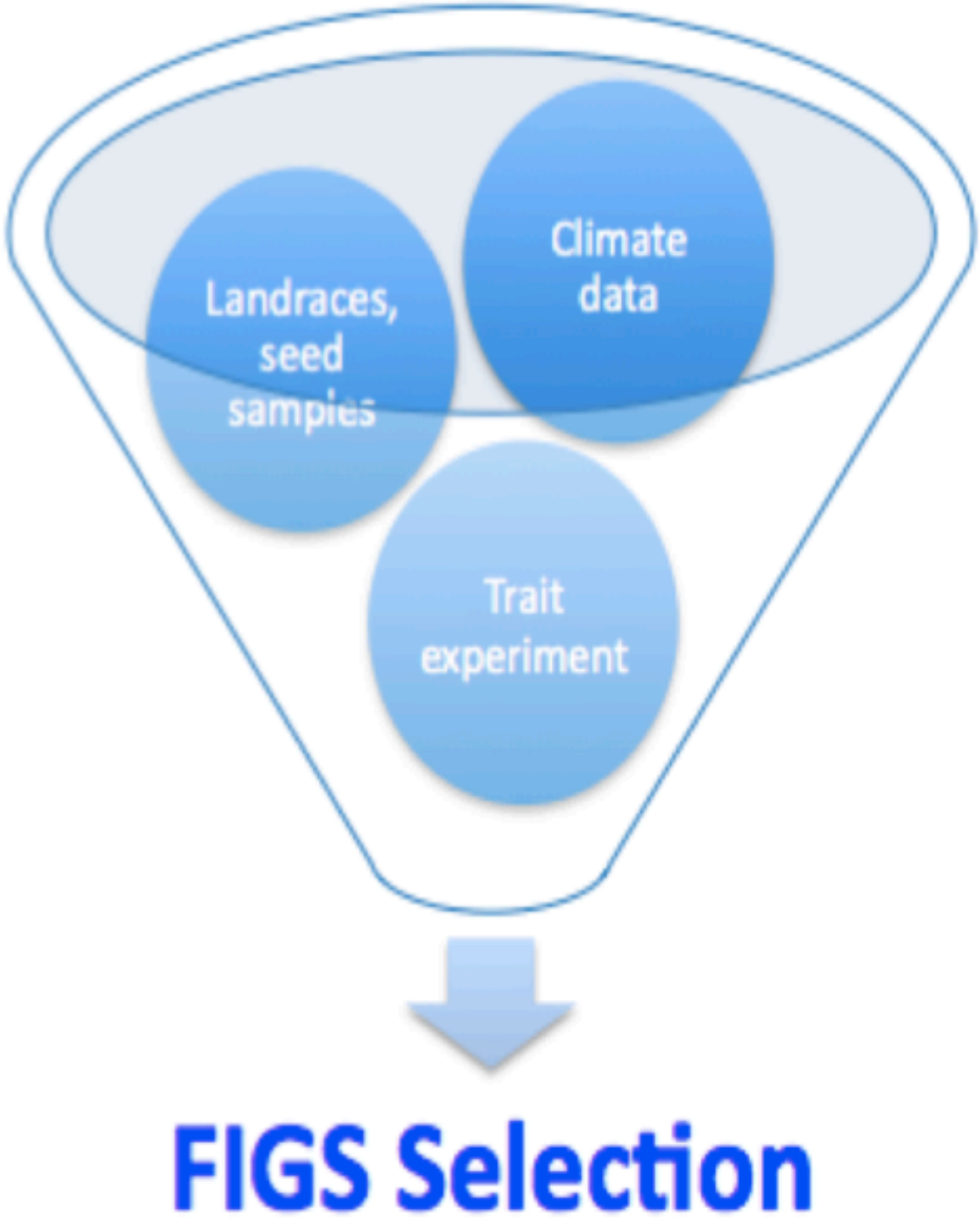
Climate layers from the ICARDA ecoclimatic database (De Pauw, 2003)

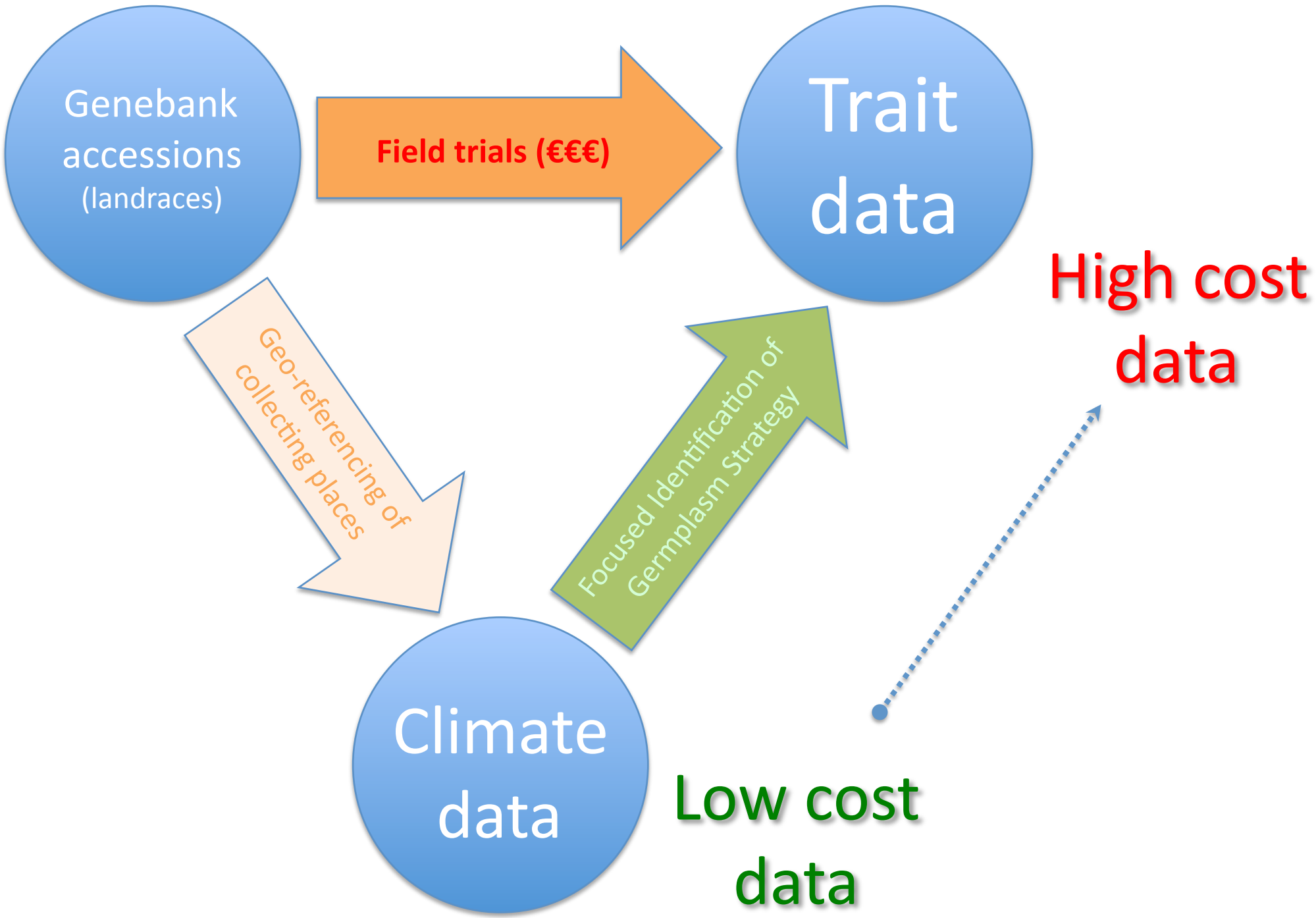
ASSUMPTION:

The climate at the original source location, where the crop landrace was developed during long-term traditional cultivation, is correlated to the trait property.

AIM:

To build a computer model explaining the crop trait score from the climate data.




Genebank accessions (landraces)
Field trials (€€€)
Trait data
Geo-referencing of collecting places
Focused Identification of Germplasm Strategy
Climate data
High cost data
Low cost data

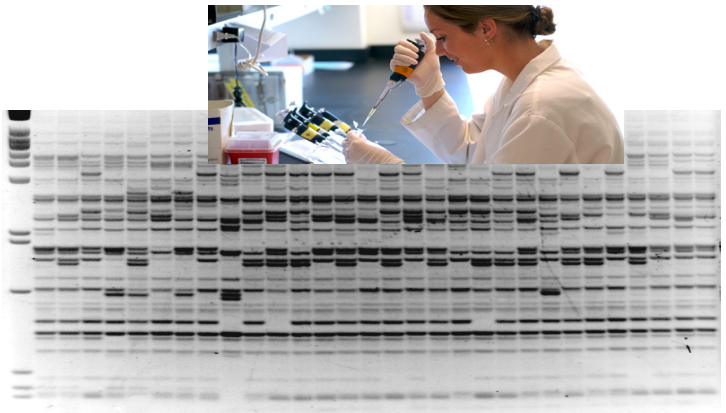
CLIMATE EFFECT DURING THE CULTIVATION PROCESS

Wild relatives are shaped by the environment

Primitive cultivated crops are shaped by local climate and humans

Traditional cultivated crops (landraces) are shaped by climate and humans

Modern cultivated crops are mostly shaped by humans (plant breeders)



Perhaps future crops are shaped in the molecular laboratory…?

PREDICTIVE LINK BETWEEN ECO-GEOGRAPHY AND TRAITS

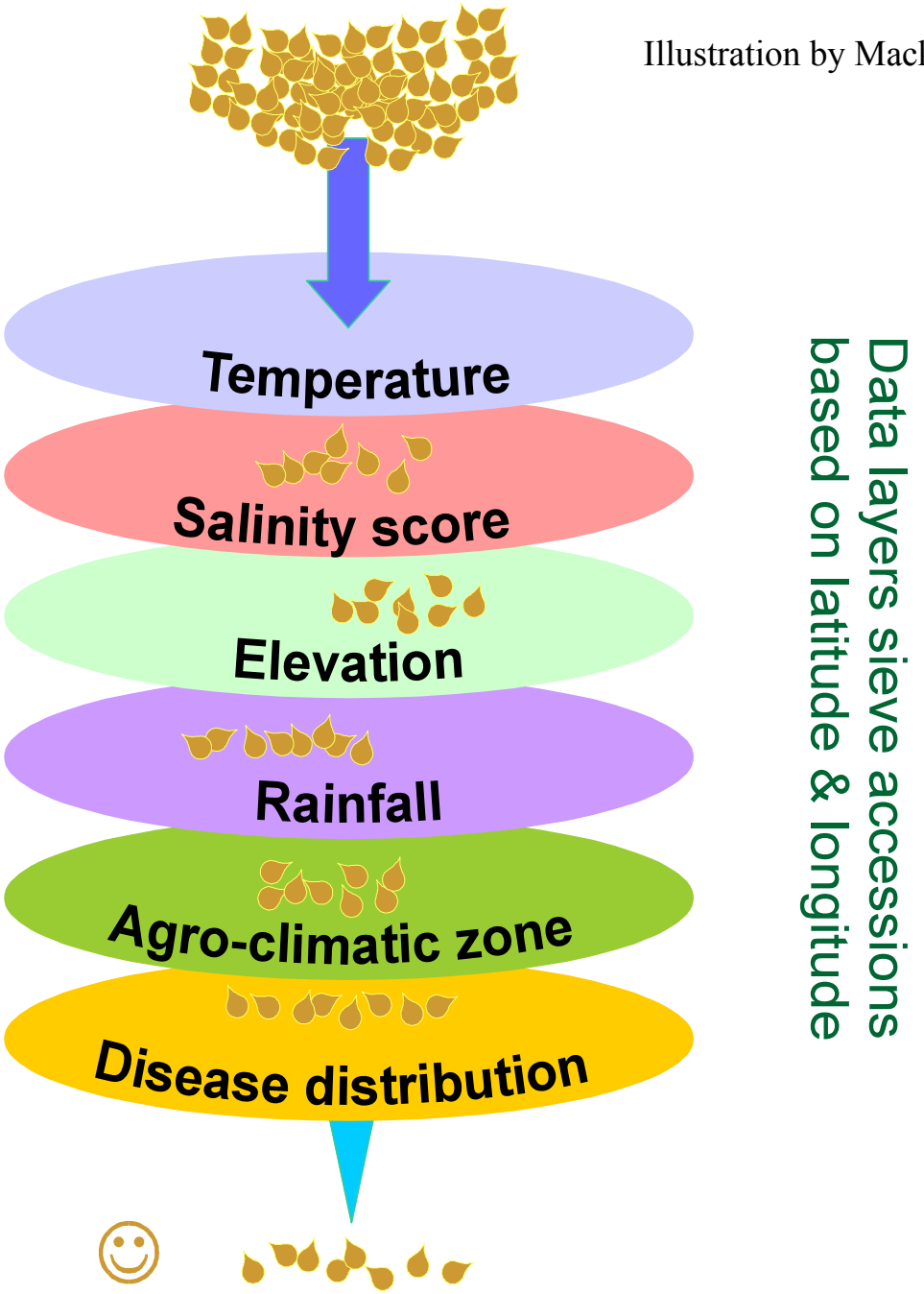
It is possible that the human mediated selection of landraces will contribute to the link between ecogeography and traits.

During traditional cultivation the farmer will select for and introduce germplasm for improved suitability of the landrace to the local conditions.

FIGS:

Origin of FIGS:
Michael Mackay (1986, 1990, 1995)

Illustration by Mackay (1995)



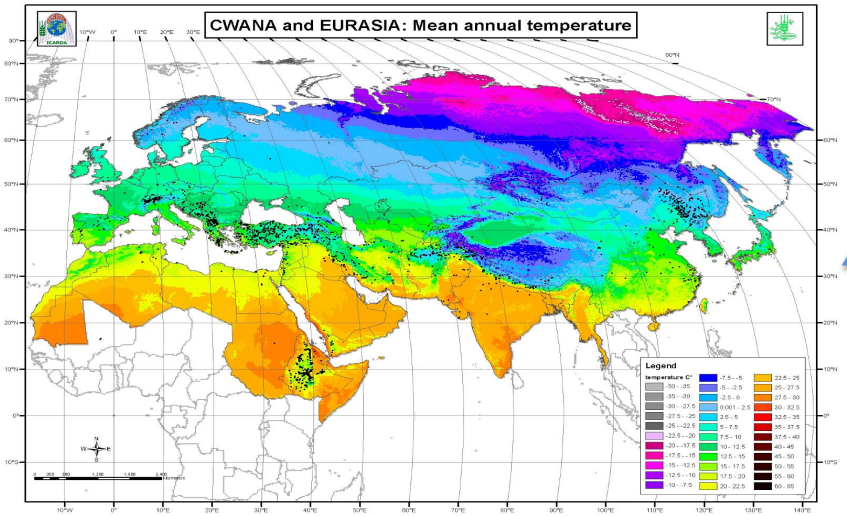
FOCUSED IDENTIFICATION OF GERMPLASM STRATEGY

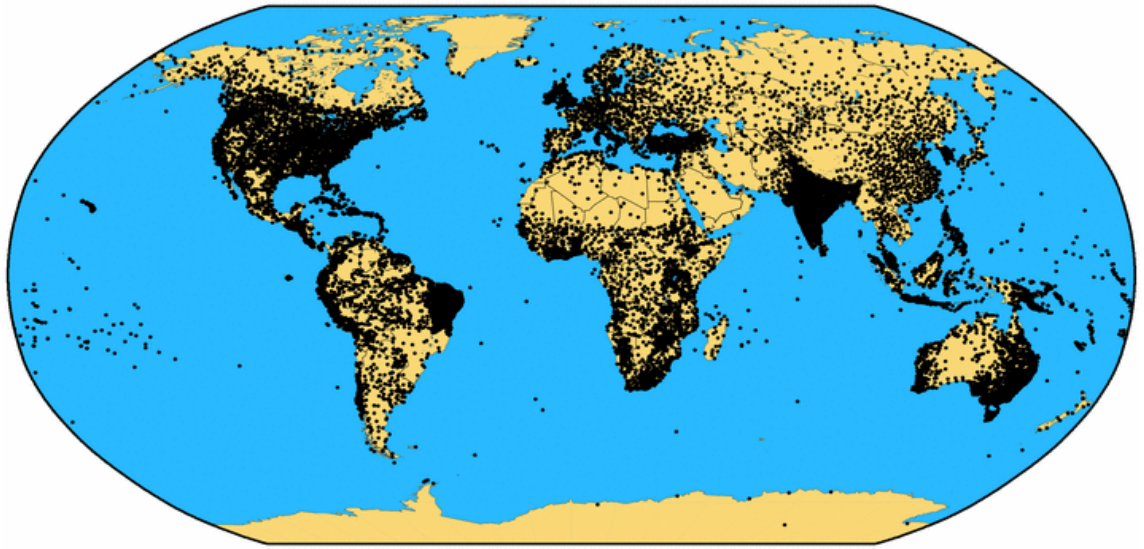
CLIMATE DATA – WORLDCLIM

The climate data can be extracted from the WorldClim dataset.
http://www.worldclim.org/
(Hijmans et al., 2005)

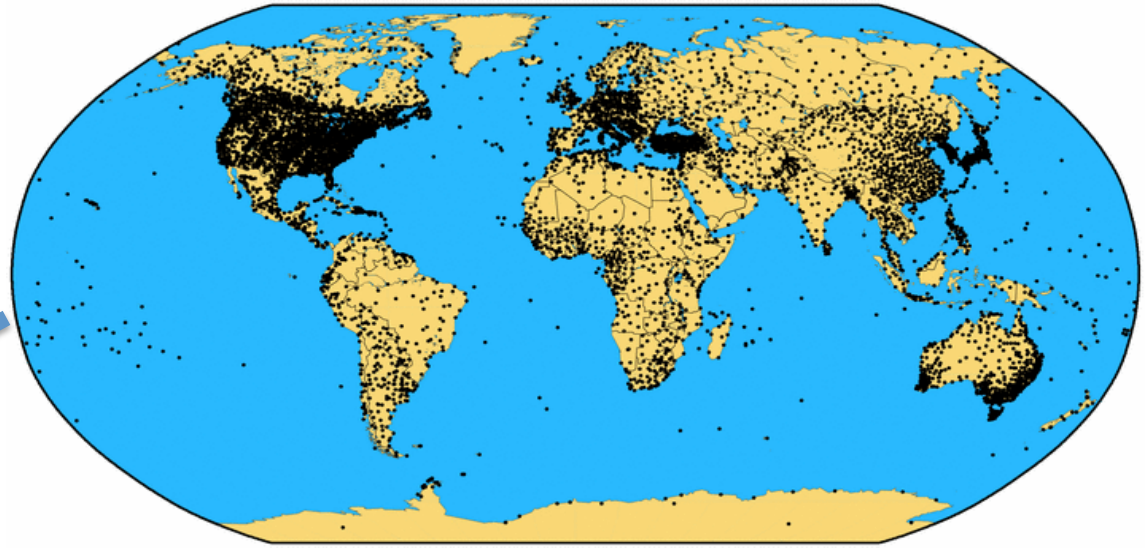
Data from weather stations worldwide are combined to a continuous surface layer.

Climate data for each landrace is extracted from this surface layer.





Precipitation: 20 590 stations



Temperature: 7 280 stations

CLIMATE DATA

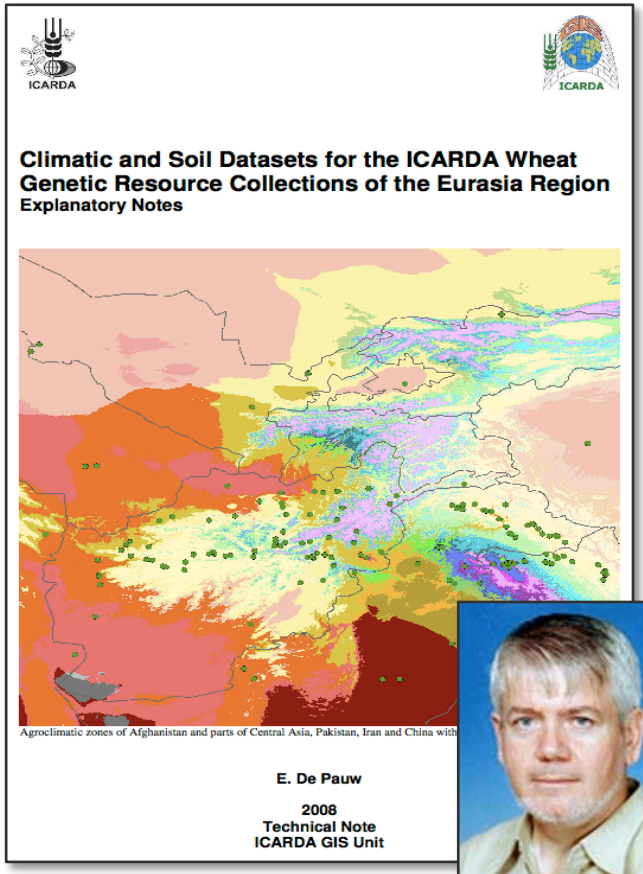
Layers used in this study:

- Precipitation (rainfall)
- Maximum temperatures
- Minimum temperatures

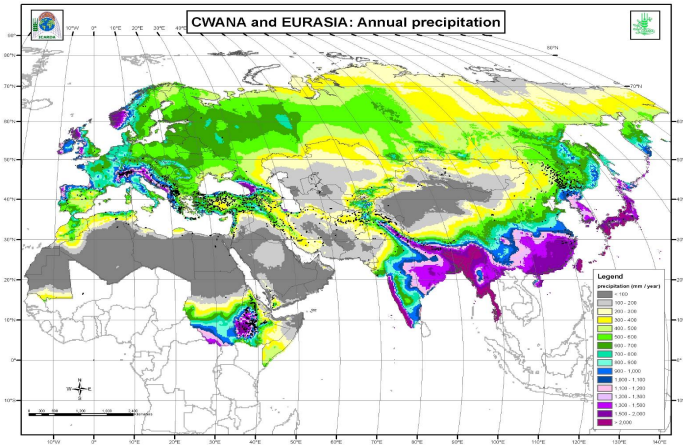
Some of the other layers available:

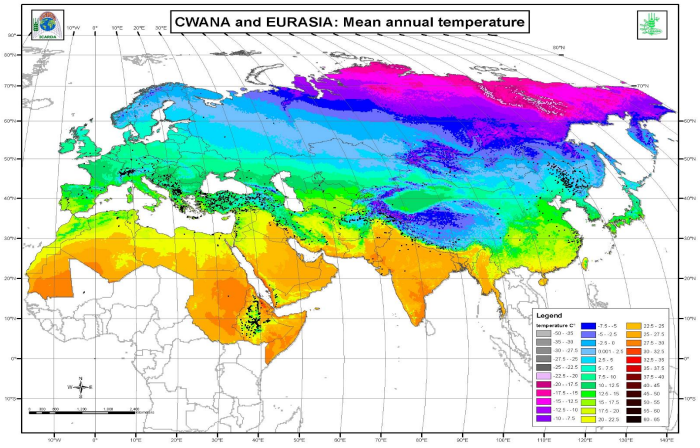
- Potential evapotranspiration (water-loss)
- Agro-climatic Zone (UNESCO classification)
- Soil classification (FAO Soil map)
- Aridity (dryness)

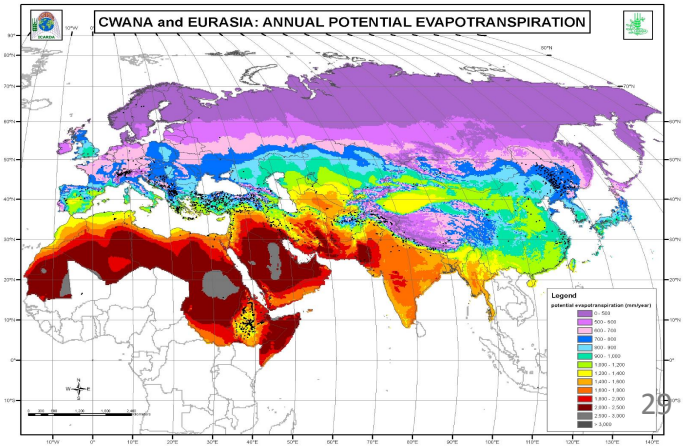
(mean values for month and year)



Eddy De Pauw
(ICARDA, 2008)







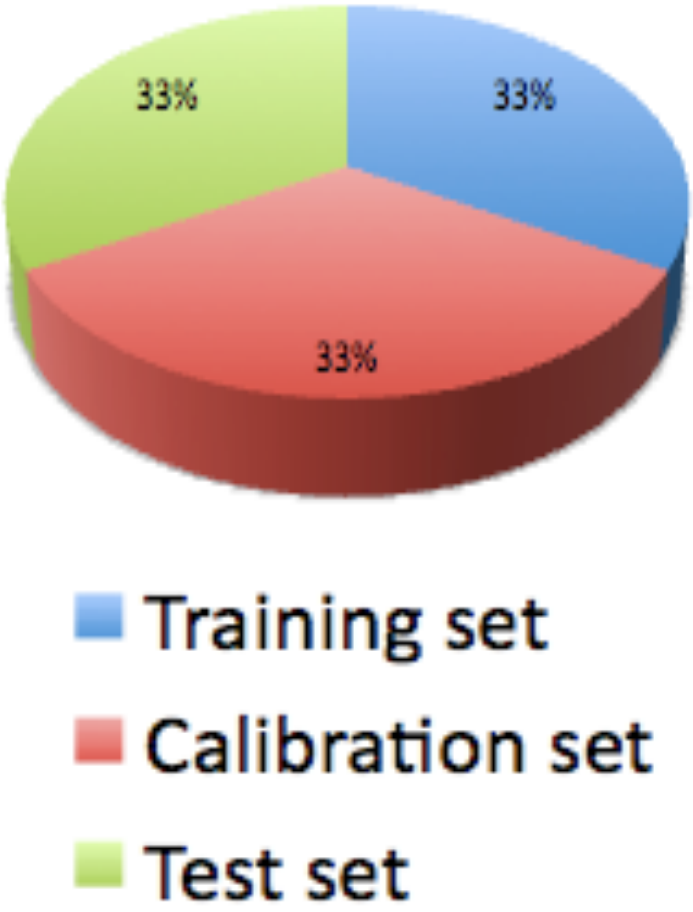
LIMITATIONS OF FIGS

- Landraces and wild relatives
  - The link between climate data and the trait data is required for trait mining with FIGS. Modern cultivars are not expected to show this predictive link (complex pedigree).

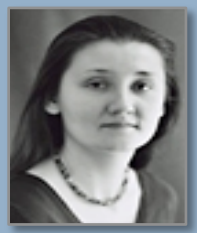
- Georeferenced accessions
  - Trait mining with FIGS is based on multivariate models using climate data from the source location of the germplasm. To extract climate data the accessions need to be accurately georeferenced.

DATA FOR THE TRAIT MINING MODEL

- **Training set**
  - For the initial calibration or training step.
- **Calibration set**
  - Further calibration, tuning step
  - Often cross-validation on the training set is used to reduce the consumption of raw data.
- **Test set**
  - For the model validation or goodness of fit testing.
  - External data, not used in the model calibration.



MORPHOLOGICAL TRAITS IN NORDIC BARLEY LANDRACES



Field observations by Agnese Kolodinska Brantestam (2002-2003)

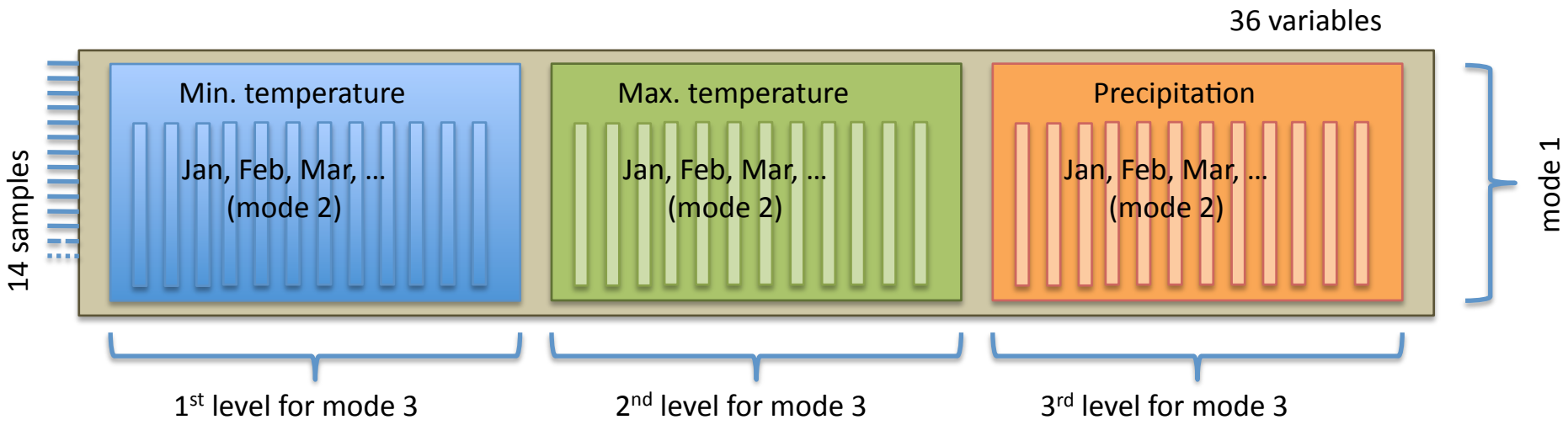
Multi-way N-PLS data analysis, Dag Endresen (2009-2010)

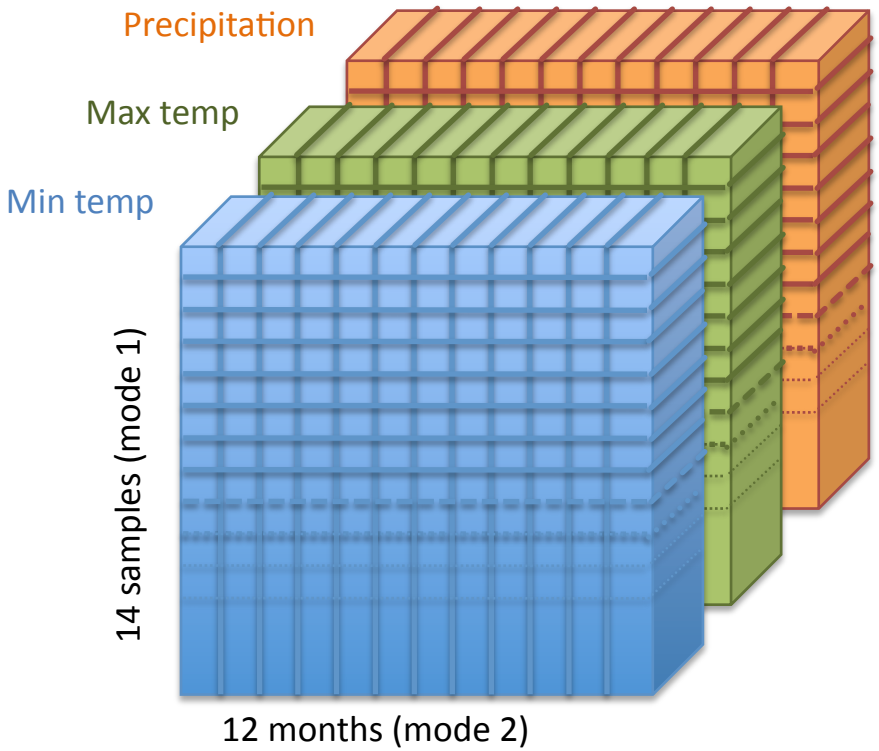

Priekuli (LVA)

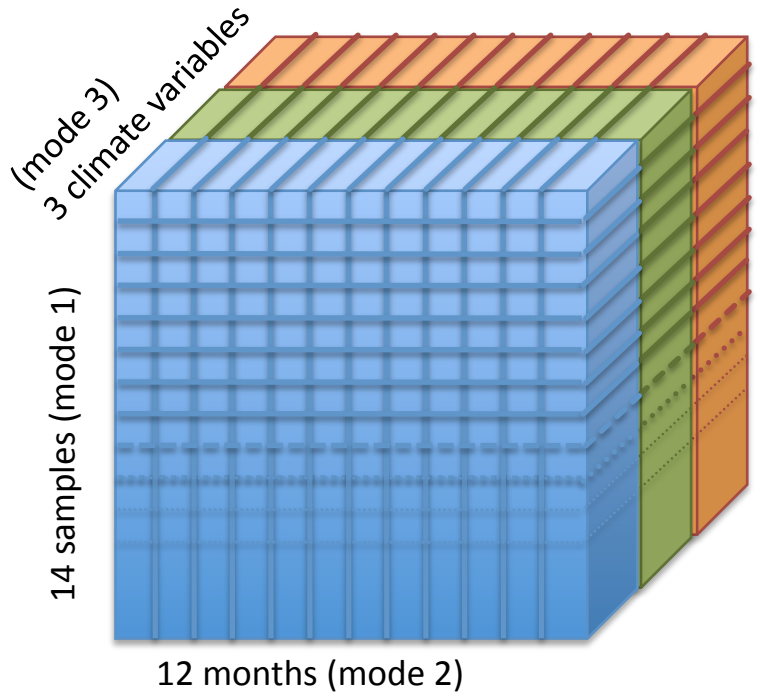

Bjørke (NOR)


Landskrona (SWE)

MULTI-WAY DATA STRUCTURE (N-PLS)







MULTI-WAY N-PLS RESULTS NORDIC BARLEY LANDRACES

Experiment Site	Year	Heading days	Ripening days	Length of plant	Harvest index	Volumetric weight	Thousand grain weight
LVA	2002[1]	n.s.	n.s.	n.s.	n.s.	***	n.s.
LVA	2003	***	n.s.	**	**	***	n.s.
NOR	2002	-	*	**	***	**	n.s.
NOR	2003	**	***	***	*	*	n.s.
SWE	2002	**	***	n.s.	**	*	n.s.
SWE	2003[2]	n.s.	**	n.s.	n.s.	**	n.s.

*** Significant at the 0.001 level (p-value)

** Significant at the 0.01 level

* Significant at the 0.05 level

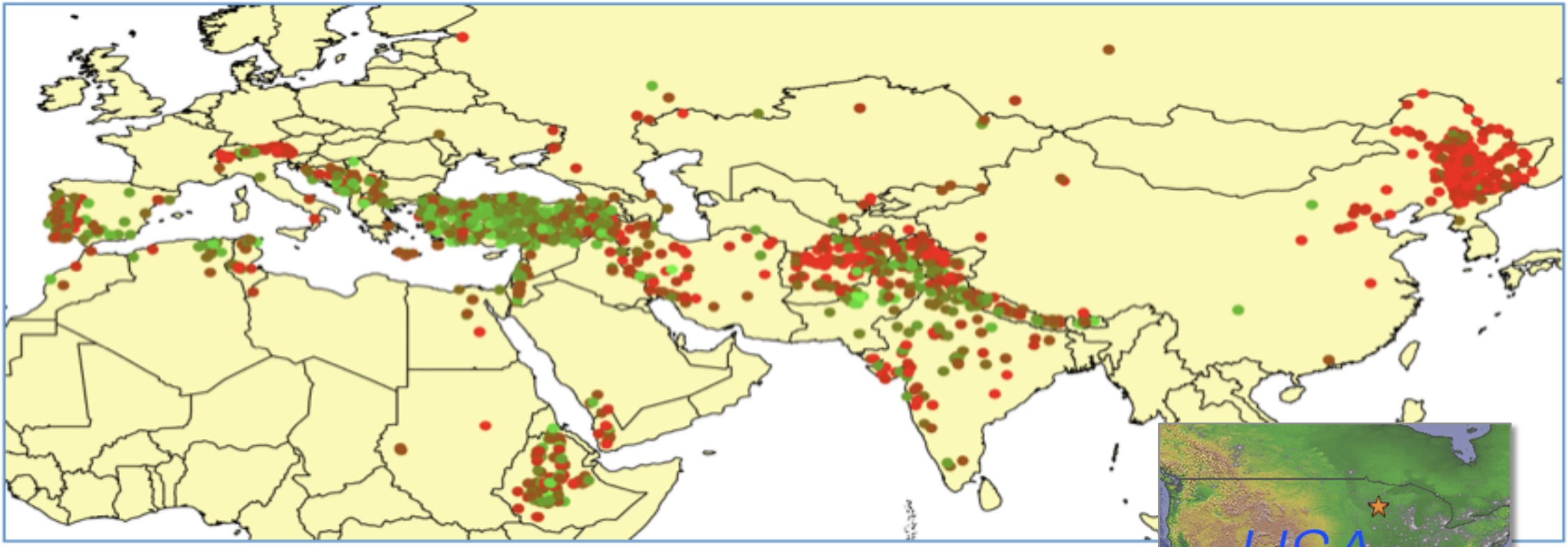
n.s. Not significant (at the above levels)

[1] LVA 2002 Germination on spikes (very wet June)

[2] SWE 2003 Incomplete grain filling (very dry June)

Endresen, D.T.F. (2010). Predictive association between trait data and ecogeographic data for Nordic barley landraces. Crop Science 50: 2418-2430. DOI: 10.2135/cropsci2010.03.0174

STEM RUST IN WHEAT LANDRACES

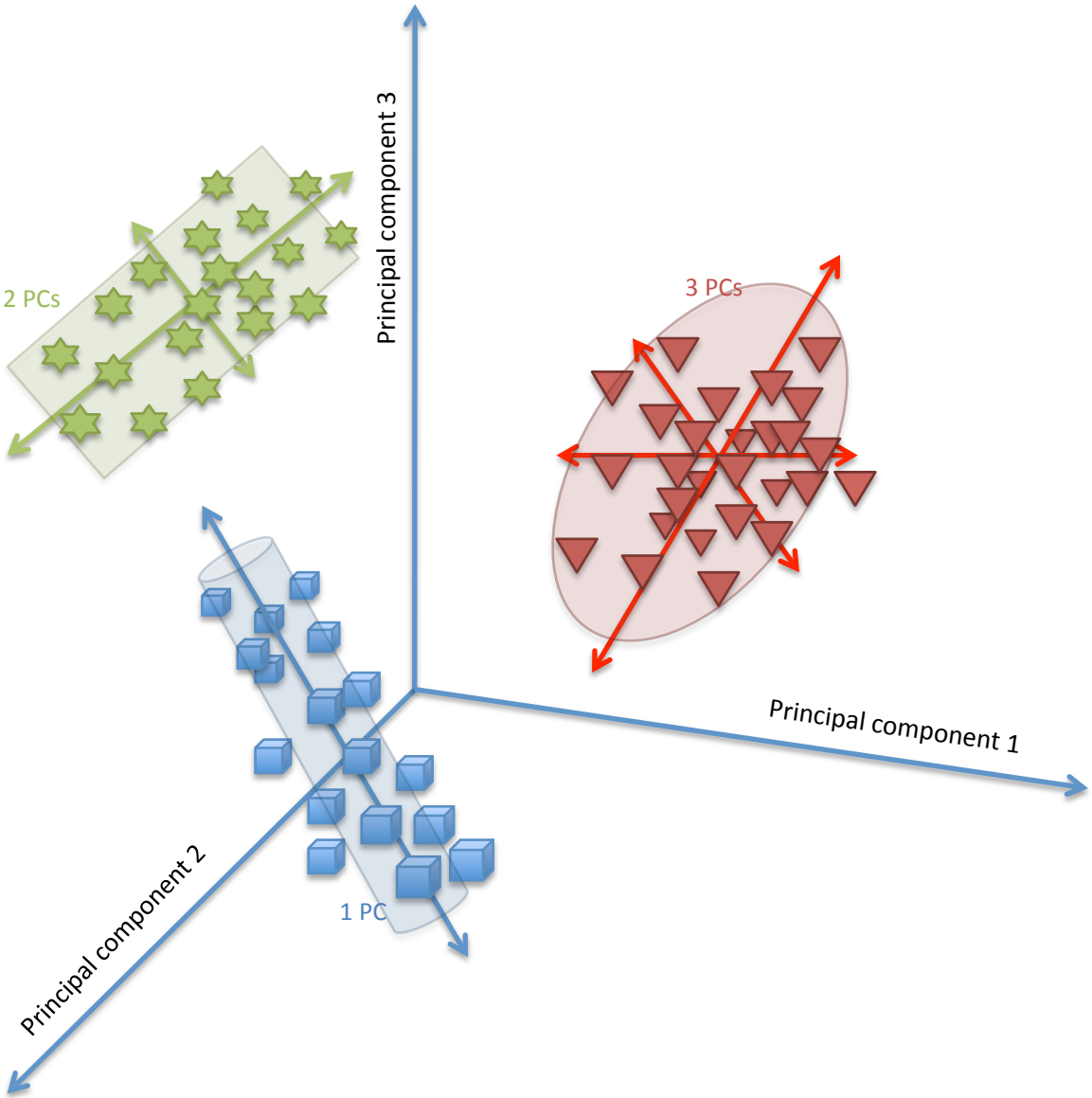


Field experiments made in Minnesota by Don McVey

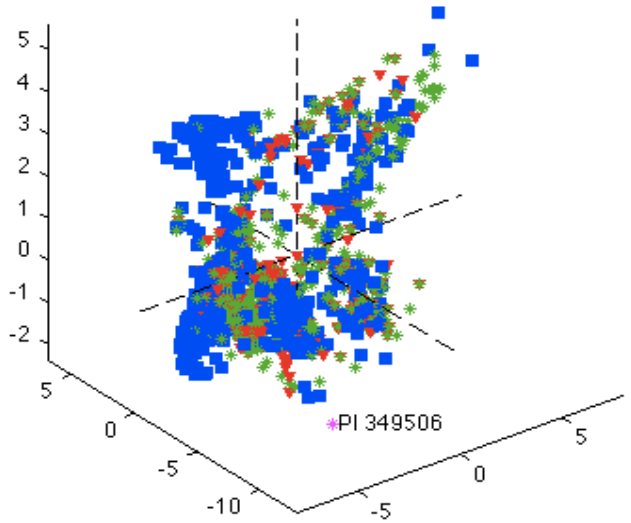
Green dots indicate collecting sites for resistant wheat landraces and red dots collecting sites for susceptible landraces.

USDA GRIN, trait data online:
http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?65049

SIMCA ANALYSIS (PCA MODEL FOR EACH CLASS)



Example from the stem rust set:



Resistant samples

Intermediate

Susceptible

Illustration modified from Wise et al., 2006:201 (PLS Toolbox software manual)

CLASSIFICATION PERFORMANCE

- Positive predictive value (PPV)
  - PPV = True positives / (True positives + False positives)
  - Classification performance for the identification of resistant samples (positives)

- Positive diagnostic likelihood ratio (LR+)
  - LR+ = sensitivity / (1 – specificity)
  - Less sensitive to prevalence than PPV

MULTIVARIATE SIMCA RESULTS STEM RUST IN WHEAT

Dataset (unit)	PPV	LR+	Estimated gain
Stem rust (accession)	**0.54** (0.50-0.59)	**3.07** (2.66-3.54)	**1.95** (1.79-2.09)
Random	**0.29** (0.26-0.33)	**1.04** (0.90-1.20)	**1.03** (0.91-1.16)
(28 % resistant samples)			
Stem rust (site)	**0.50** (0.40-0.60)	**4.00** (2.85-5.66)	**2.51** (2.02-2.98)
Random	**0.19** (0.13-0.26)	**0.94** (0.63-1.39)	**0.95** (0.66-1.33)
(20 % resistant samples)			

PPV = Positive Predictive Value; LR+ = Positive Diagnostic Likelihood Ratio

Endresen, D.T.F., K. Street, M. Mackay, A. Bari, E. De Pauw (submitted). Predictive association between biotic stress traits and ecogeographic data for wheat and barley landraces. *Crop Science, conditionally accepted 6 Feb 2011, revision 1 submitted.*

MULTIVARIATE ANALYSIS STEM RUST IN WHEAT

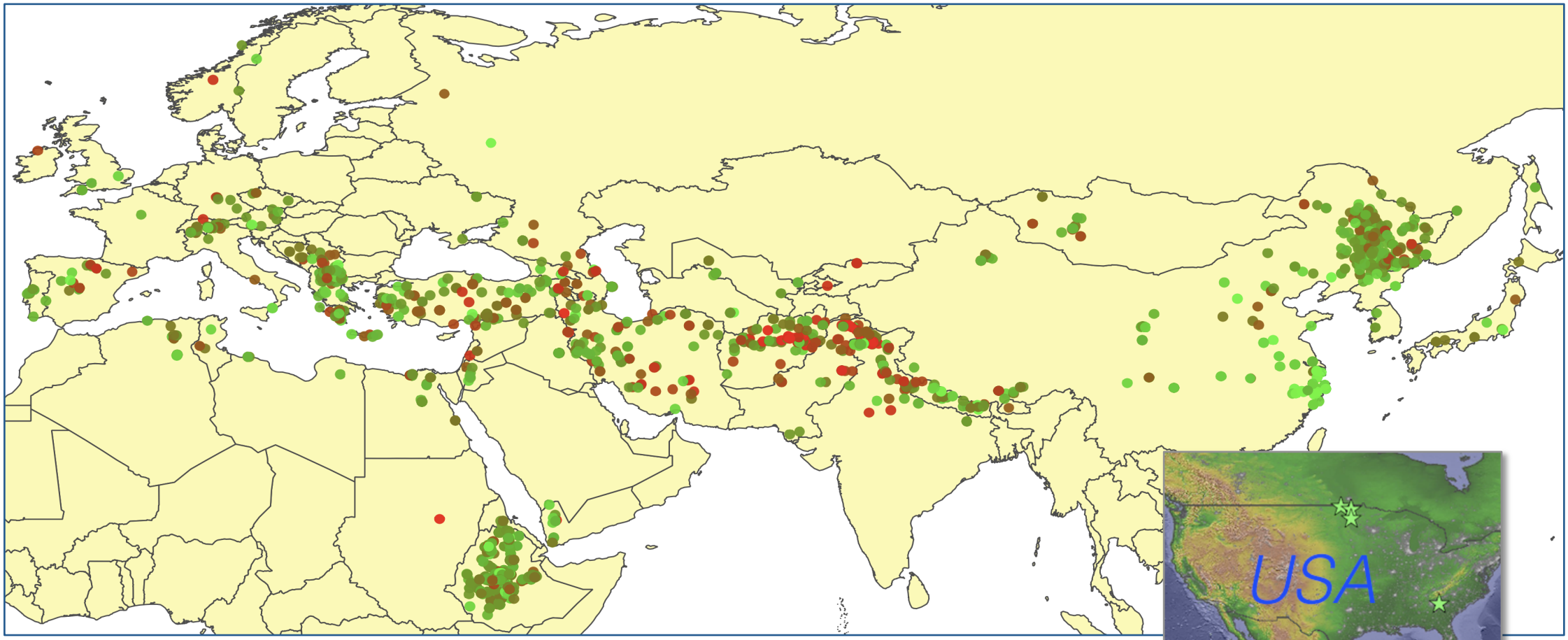
Classifier method	AUC	Cohen's Kappa
Principal Component Regression (PCR)	**0.69** (0.68-0.70)	**0.40** (0.37-0.42)
Partial Least Squares (PLS)	**0.69** (0.68-0.70)	**0.41** (0.39-0.43)
Random Forest (RF)	**0.70** (0.69-0.71)	**0.42** (0.40-0.44)
Support Vector Machines (SVM)	**0.71** (0.70-0.72)	**0.44** (0.42-0.45)
Artificial Neural Networks (ANN)	**0.71** (0.70-0.72)	**0.44** (0.42-0.46)

AUC = Area Under the ROC Curve (ROC, Receiver Operating Curve)

Bari, A., K. Street, , M. Mackay, D.T.F. Endresen, E. De Pauw, and A. Amri (submitted). Focused Identification of Germplasm Strategy (FIGS) detects wheat stem rust resistance linked to environment variables. *Recently submitted to GRACE, March 2010*.

Abdallah Bari (ICARDA)

NET BLOTCH IN BARLEY LANDRACES



Field experiments made in Minnesota, North Dakota and Georgia in the USA

Green dots indicate collecting sites for resistant wheat landraces and red dots collecting sites for susceptible landraces.

USDA GRIN, trait data online:
http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?1041

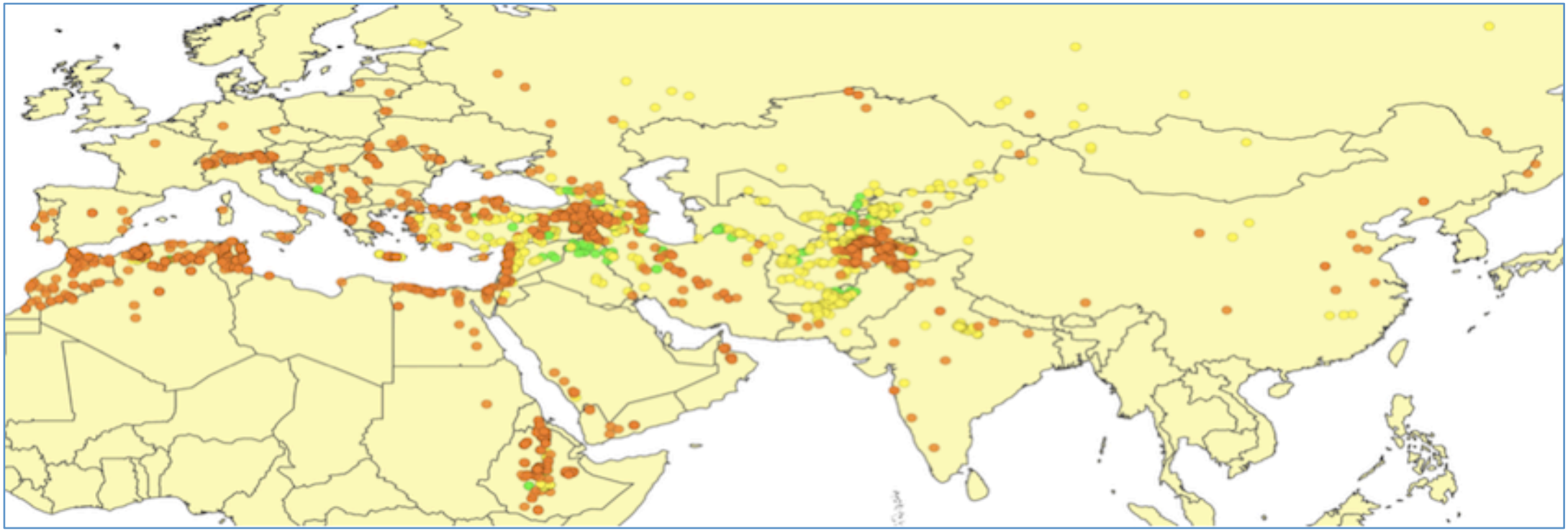
MULTIVARIATE SIMCA RESULTS NET BLOTCH IN BARLEY

Dataset (unit)	PPV	LR+	Estimated gain
Net blotch (accession)	**0.54** (0.48-0.60)	**1.75** (1.42-2.17)	**1.35** (1.19-1.50)
Random	**0.40** (0.35-0.45)	**0.99** (0.84-1.17)	**0.99** (0.87-1.12)
(40 % resistant samples)			

PPV = Positive Predictive Value; LR+ = Positive Diagnostic Likelihood Ratio

Endresen, D.T.F., K. Street, M. Mackay, A. Bari, E. De Pauw (submitted). Predictive association between biotic stress traits and ecogeographic data for wheat and barley landraces. *Crop Science, conditionally accepted 6 Feb 2011, revision 1 submitted.*

MULTIVARIATE SIMCA RESULTS STEM RUST (UG99) IN WHEAT



Ug99 set with 4563 wheat landraces screened for Ug99 in Yemen 2007, **10.2 % resistant** accessions. The true trait scores for 20% of the accessions (825 samples) were revealed. We used trait mining with SIMCA to select 500 accessions more likely to be resistant from 3728 accession with true scores hidden (to the person making the analysis). The FIGS set was observed to hold **25.8 % resistant** samples and thus **2.5 times higher** than expected by chance.

Endresen, D.T.F., K. Street, M. Mackay, A. Bari, E. De Pauw (*draft manuscript*). Sources of resistance in wheat to stem rust (Ug99) identified using Focused Identification of Germplasm Strategy (FIGS).

FACULTY OF LIFE SCIENCES
UNIVERSITY OF COPENHAGEN

PhD thesis

Dag Terje Filip Endresen

Utilization of Plant Genetic Resources
A Lifeboat to the Gene Pool

Academic advisor: Dvora-Laiô Wulfsohn and Brian Grout

Submitted: 9 February 2011

A LIFEBOAT TO THE GENE POOL

- PDF available from:
  - http://db.tt/lZMpwgJ
- Available from Libris (Sweden)
  - ISBN: 978-91-628-8268-6

ISBN 978-91-628-8268-6
9 789162 882686 >

Thanks for listening!

PhD dissertation
31 March 2011

Department of Ecology and Agriculture,
Faculty of Life Sciences,
Copenhagen University

Dag Terje Filip Endresen
dag.endresen@nordgen.org