Mining for Useful Variation in the Watkins Wheat Collection

Luzie U. Wingen

luzie.wingen@bbsrc.ac.uk

Dept. Crop Genetics

John Innes Centre
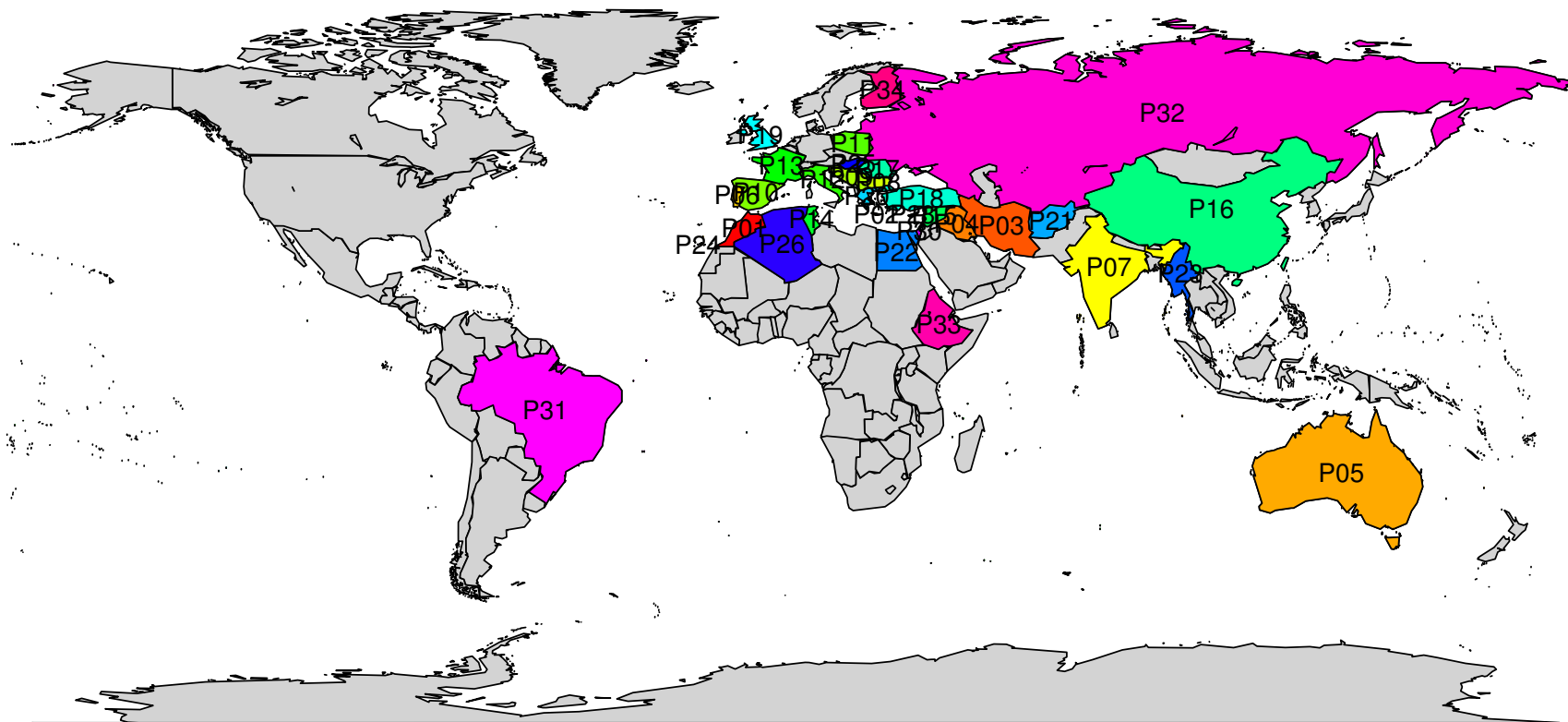
WGIN Stakeholder Meeting: 17th Nov. 2010

The Watkins Wheat Collection

- collection of wheat landraces (LR)
- by A. E. Watkins in the 1920s and 1930s
- from 32 countries
- originally over 7,400 accessions
- current living collection: just over 1,000 accessions
- phenotypically diverse

Geographic Distribution

Afghanistan, Algeria, Australia, Brazil, Bulgaria, Burma, China, Cyprus, Egypt, Ethiopia, Finland, France, Greece, Hungary, India, Iran, Iraq, Italy, Morocco, Palestine, Poland, Portugal, Romania, Spain, Syria, Tunisia, Turkey, UK, USSR, Yugoslavia

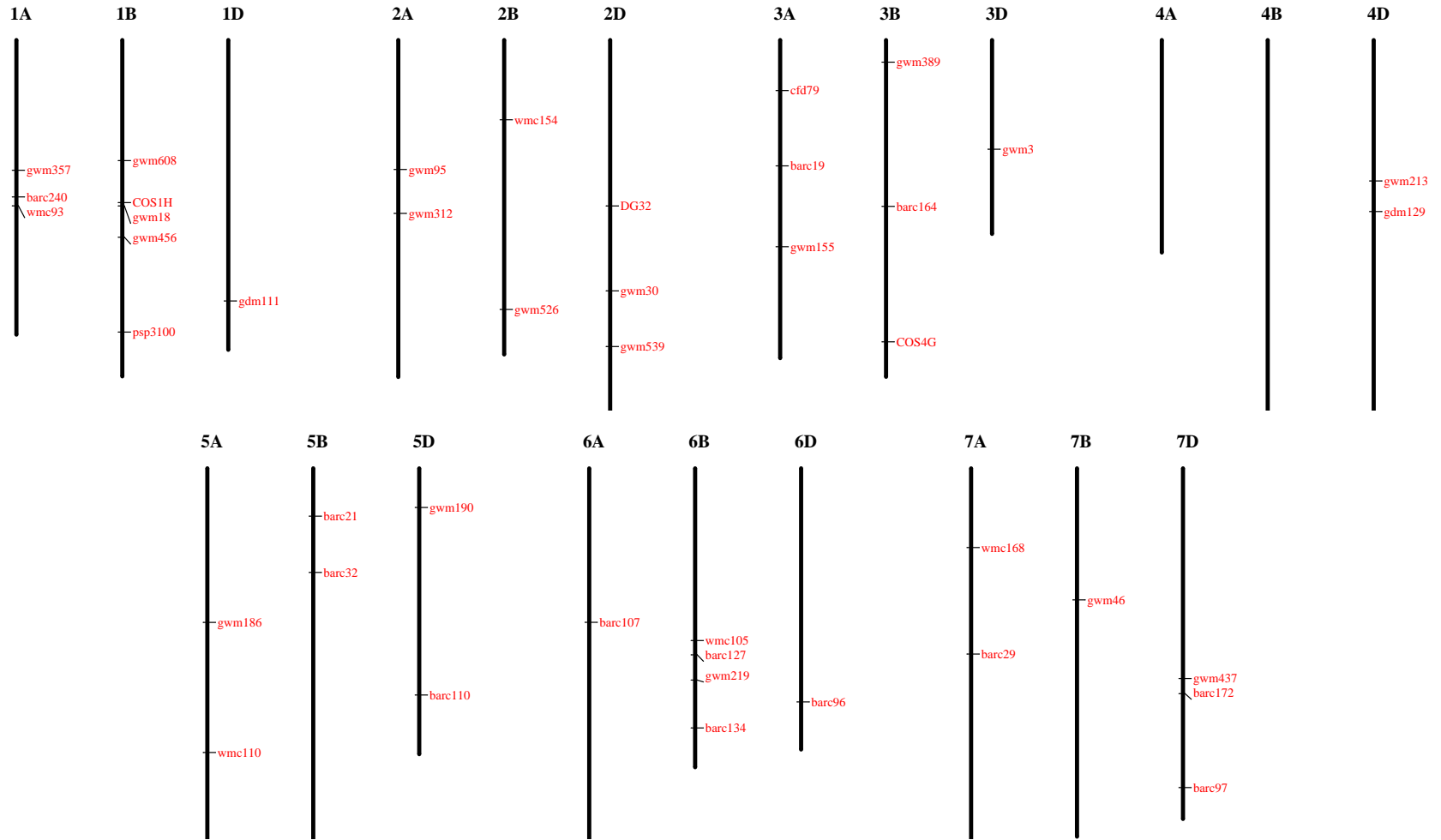# Phenotypes: Heading Date

Phenotypes: Height

Phenotypes: Awns, Waxiness

Phenotypes: Ear Colour

Genotyping

40 microsatellite markers (SSRs) and 3 gene-based markers (COS) on 21 wheat chromosomes.

Measures of Genetic Diversity

- allele richness

- average allele number

- number of rare alleles (< 2%)

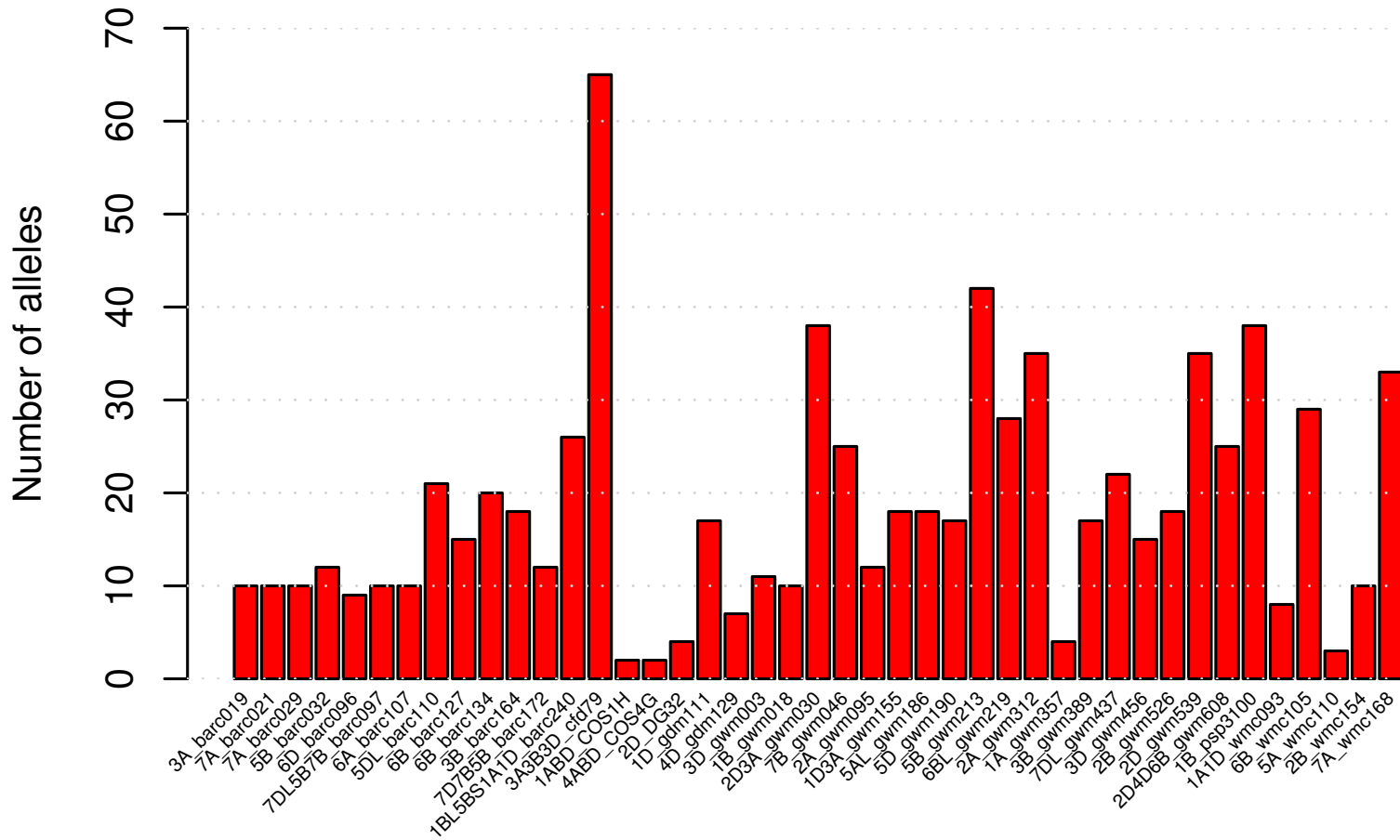- polymorphic information content PIC

$$PIC = 1 - \Sigma_i^n (p_i)^2$$
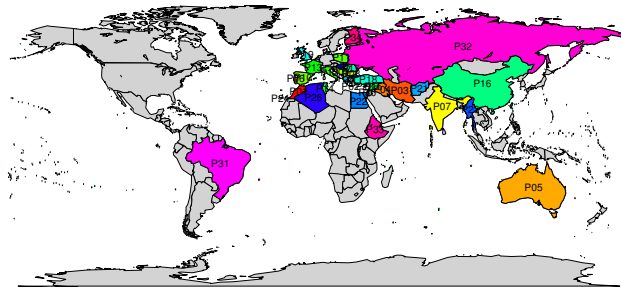
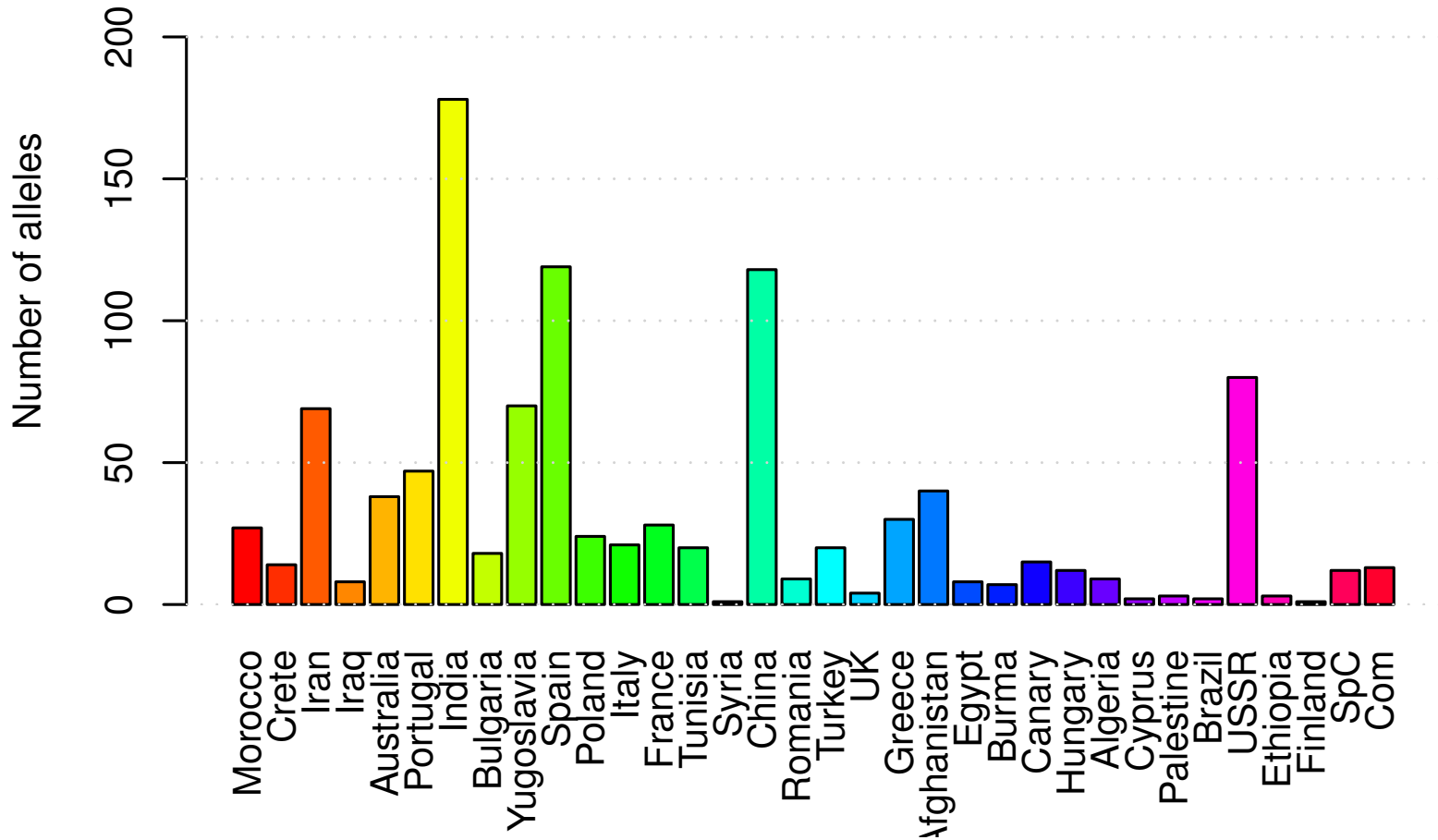- Shannon-Weaver diversity index I

$$I = -\Sigma_i^n (p_i ln p_i)$$

p_i is the frequency of the ith allele.

Alleles Per Locus
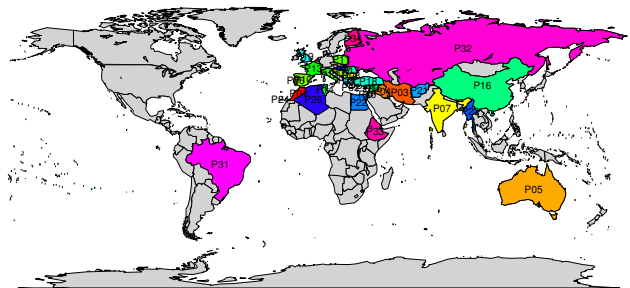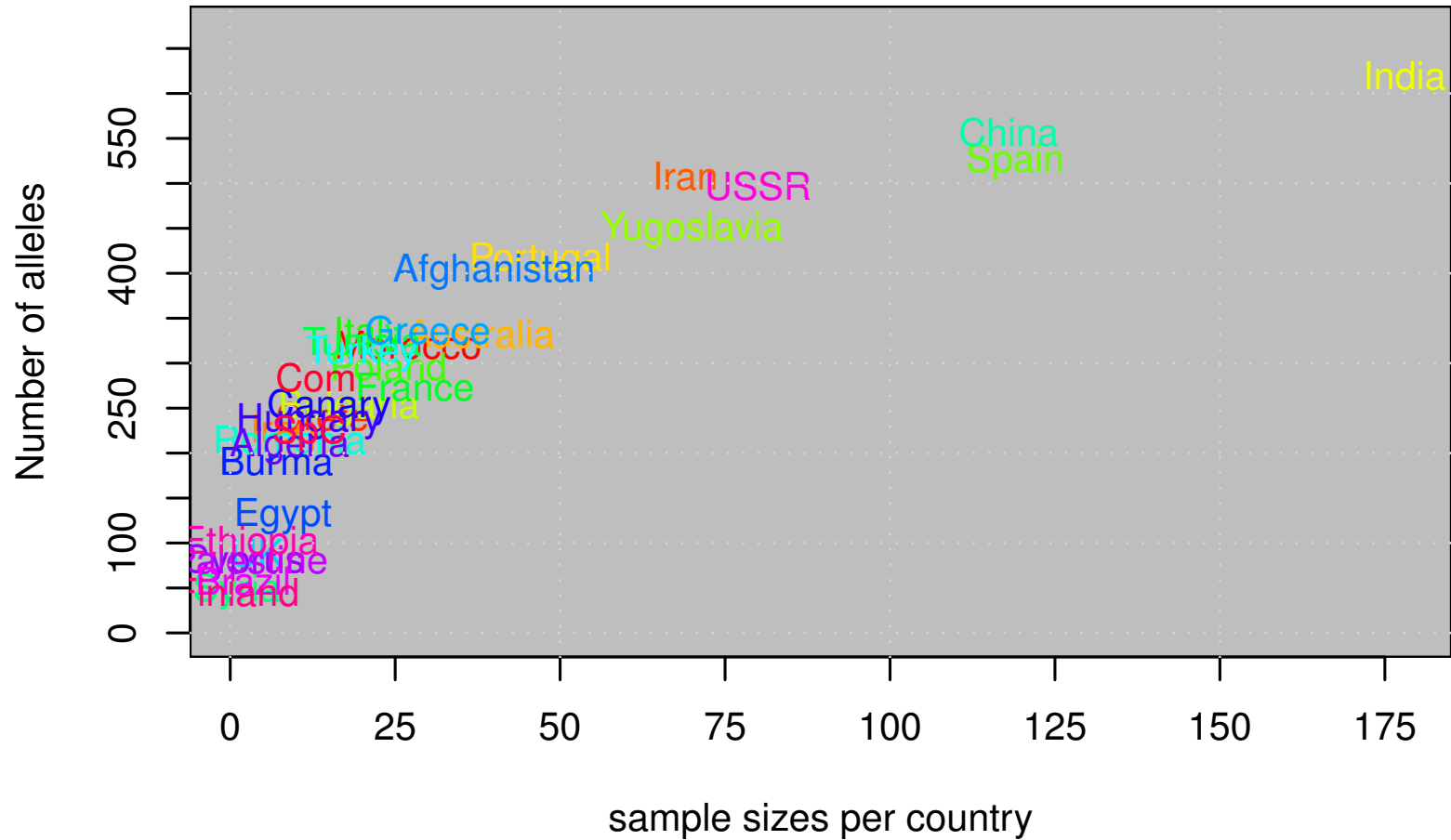
Alleles per Country

1066 landraces

11 *T. sphaerococcum* accs.

11 *T. compactum* accs.

Allele Number vs Sample Size

Genetic Diversity

	Gediflux (modern)	Watkins Collection
markers	42	43
accessions	511	1088
allele richness	443	793
average allele no	10.5	18.4
range	2-26	2-65
avg. rare allele no	5.5	11.1
range	0-16	1-52
PIC	0.649	0.731

Gediflux: genetic diversity of modern European winter wheat

Wheat Diversity Studies

material	acc. no.	marker no.	avg. allele no.	range	PIC	publication
IPK collection	998	24	18.1	4-46	0.77	Huang 2002
modern wheat	502	20	10.5	4-22	0.647	Röder 2002
Chinese LR	24	40	6.9	1-16	0.419	Wei 2005
Eur. 1840-2000	480	39	16.4	4-40	0.65	Roussel 2005
UK/USA/Aus 1845-2005	140	379	DArT	-	0.399	White 2007
INRA collection	3,942	38	23.9	7-45	0.74	Balfourier 2007
modern+traditional	1,057	178	DArT	-	0.44	Raman 2010
Watkins collection	1,088	43	18.4	2-65	0.73	

Population Structure

Genetic diversity is usually not evenly distributed.
How to detect the population structure?

- a measure of dissimilarity between two accessions e.g.

  - Euclidean distance $dE = \sqrt{\sum_i^m \sum_j^{n_i}(p_{ij} - q_{ij})^2}$

  - Cavalli-Sforza and Edwards distance
    $dCE = \sqrt{1/m \sum_i^m (1 - \sum_j^{n_i} \sqrt{p_{ij}q_{ij}})}$

  - Nei's genetic diversity index $dN = 1/m \sum_i^m (1 - \sum_j^{n_i} \sqrt{p_{ij}q_{ij}})$

- a clustering method to group accessions e.g.

  - Principal Component Analysis (PCA)
  - Bayesian Model-based Clustering (STRUCTURE)
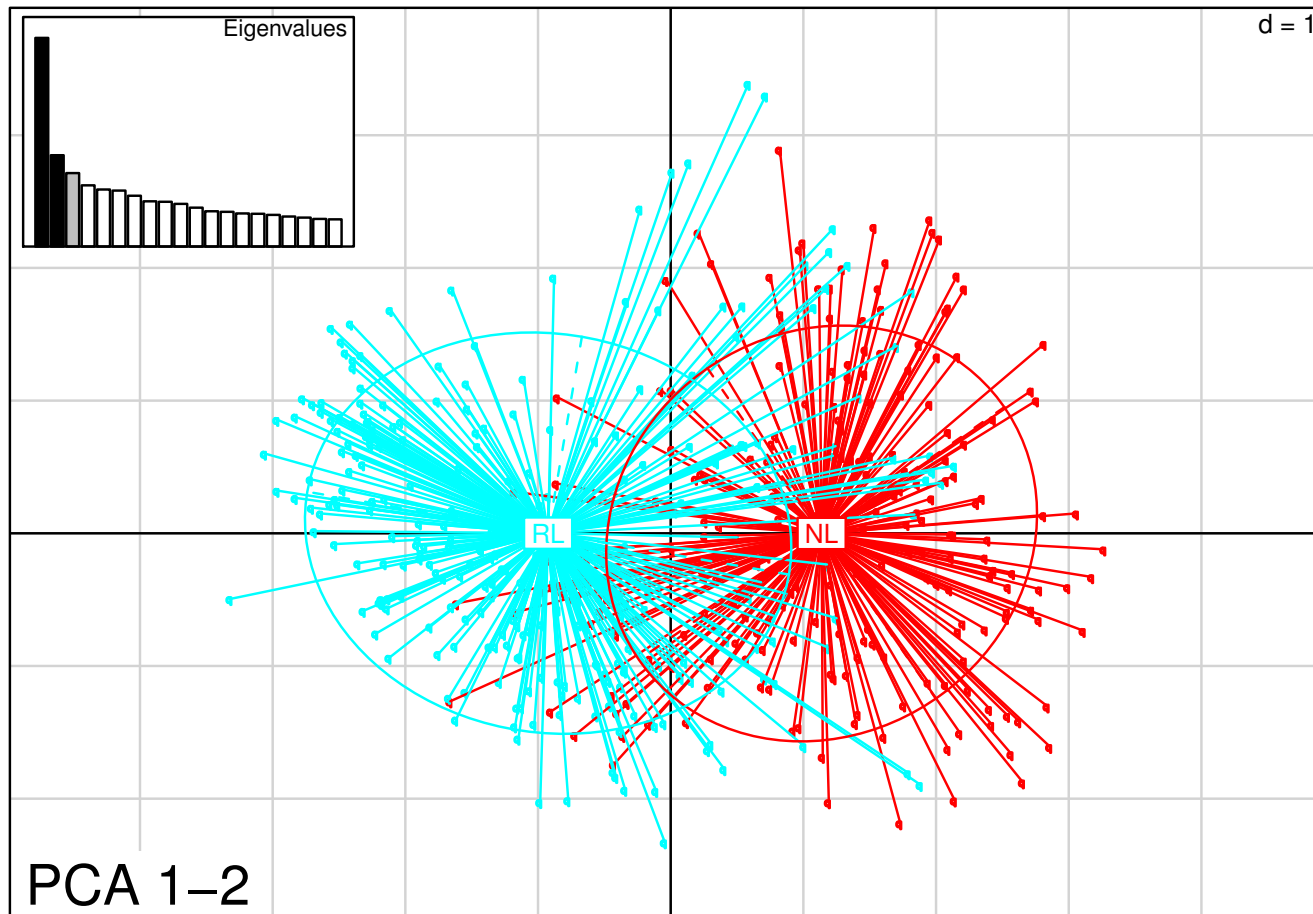 - Hierarchical Clustering

The structure may reveal historic trends. It helps selecting a
core collection, a smaller set with maximum diversity.
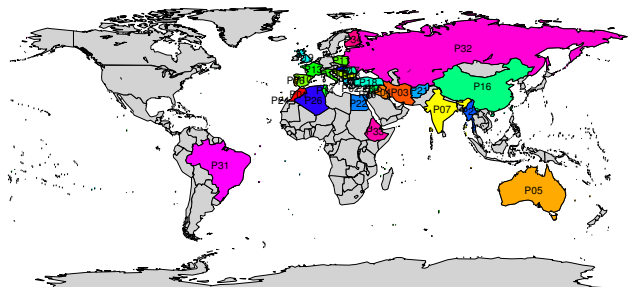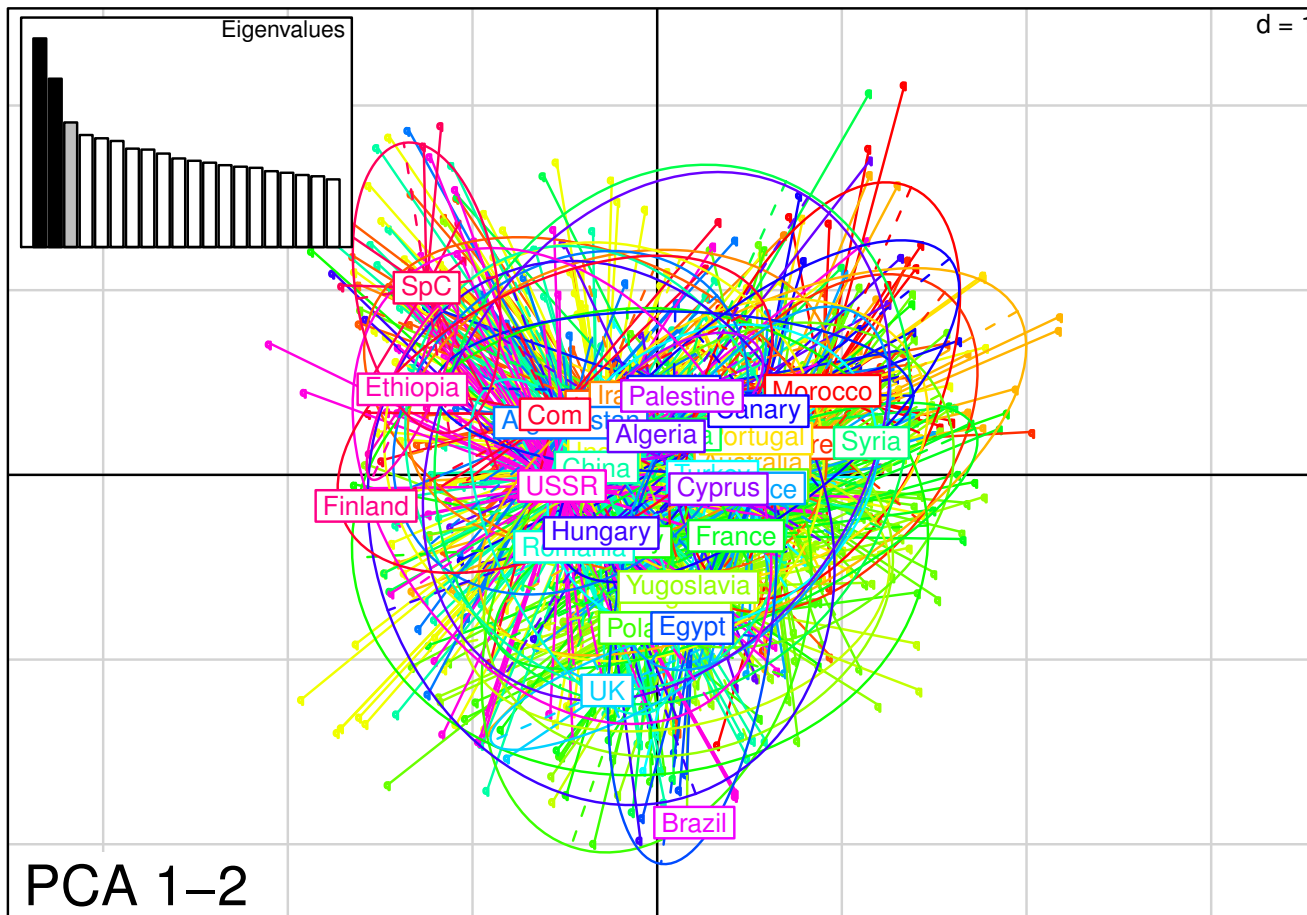
Gediflux - PCA

Gediflux: genetic diversity of modern European winter wheat:
NL: UK National List (late 1990s)
RL: EU-recommended List (1945-2000)

Watkins PCA 1-2

No complete separation of groups.
No clear geographical association.

Further Plans

- further characterisation of the population structure

- comparisons to other resources (e.g. Gediflux, Organic Wheat Project)

- define a core collection
  a resource which can be distributed to other labs for their efficient evaluation

- biparental populations for gene identification
  SSD in Paragon (PxW34, PxW141, PxW209, PxW292, PxW352, PxW468, PxW481, PxW729)

- physiological screening within BBSRC Pre-Breeding Programme esp. for NUE

- association analysis for gene identification

- publish result on WGIN website http://www.wgin.org.uk/

Stocks for future projects

Acknowledgement

Richard Goram, Lorelei Bilham, John Snape„ Michelle Leverington-Waite, Sue Freeman, Simon Griffiths, Simon Orford„
Lesley Fish, Leodie Alibert