

Genetic diversity analysis of wild and landrace barleys using the BOPA1 Illumina genotyping platform

Andy Flavell



**Division of Plant Sciences
University of Dundee at SCRI**

CONTENTS

What happens when you select markers in elite barley germplasm and apply them in primitive and wild germplasm?

What is the relationship between genotype and geography for landrace and wild barley populations?

What is the relationship between wild and landrace barley populations derived from similar geographical locations?

GERMPLASM

CULTIVAR GERMPLASM

**137 elite European barley cultivars selected from the UK AGOUEB Project
- Luke Ramsay, Robbie Waugh *et al.***

LANDRACE GERMPLASM

**The ICARDA Syrian-Jordanian Landrace Collection (SJLC = 480 lines
derived from Jordan and Syria) and subsets of it - Stefania Grando**

WILD BARLEY GERMPLASM

**A subset of 131 individuals from the World Barley Diversity Collection
(WBDC) - Brian Steffenson (U. Minnesota)**

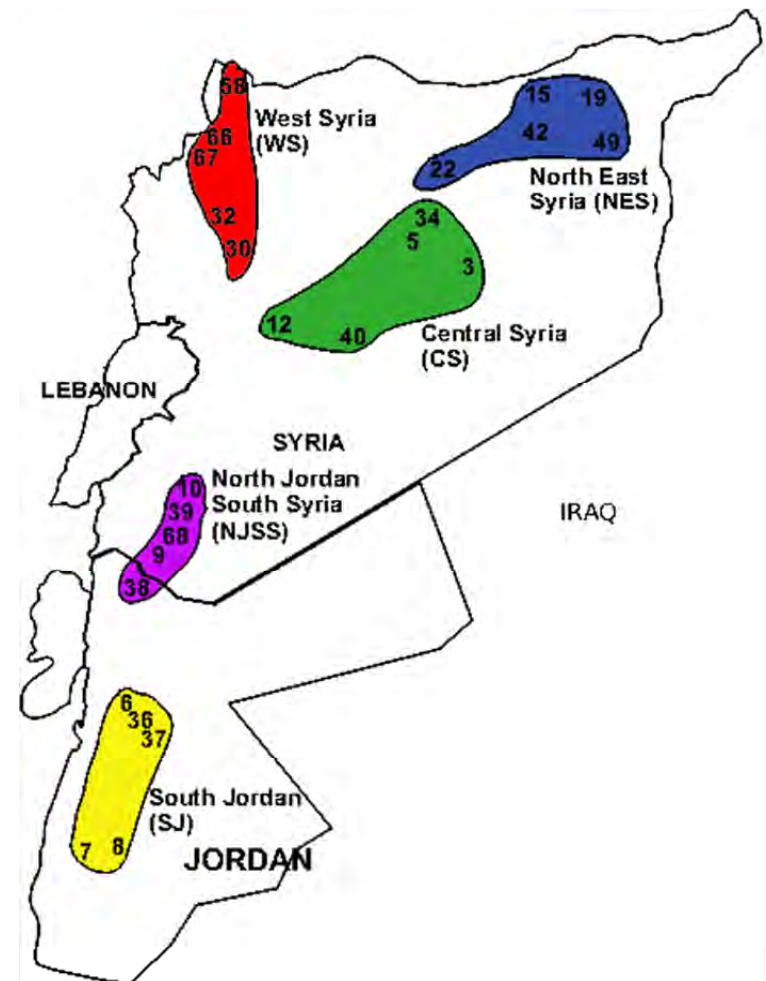
MARKERS

The Barley OPA1 Illumina marker set of 1536 SNPs and subsets of it

GERMPLASM

LANDRACE GERMPLASM

The ICARDA Syrian-Jordanian Landrace Collection (SJLC) 480 lines derived from Jordan and Syria



GERMPLASM

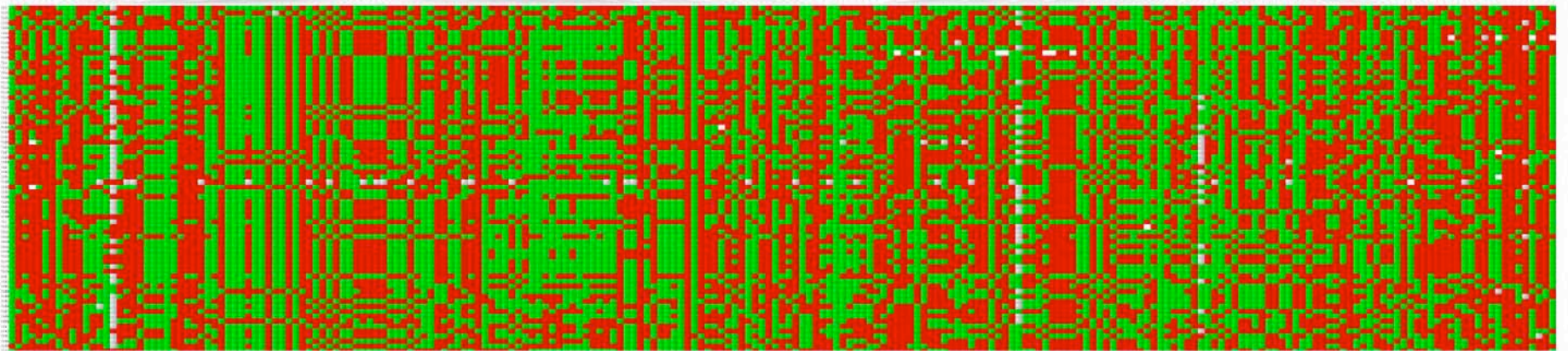
WILD BARLEY GERMPLASM

A subset of 131 individuals from the World Barley Diversity Collection (WBDC)

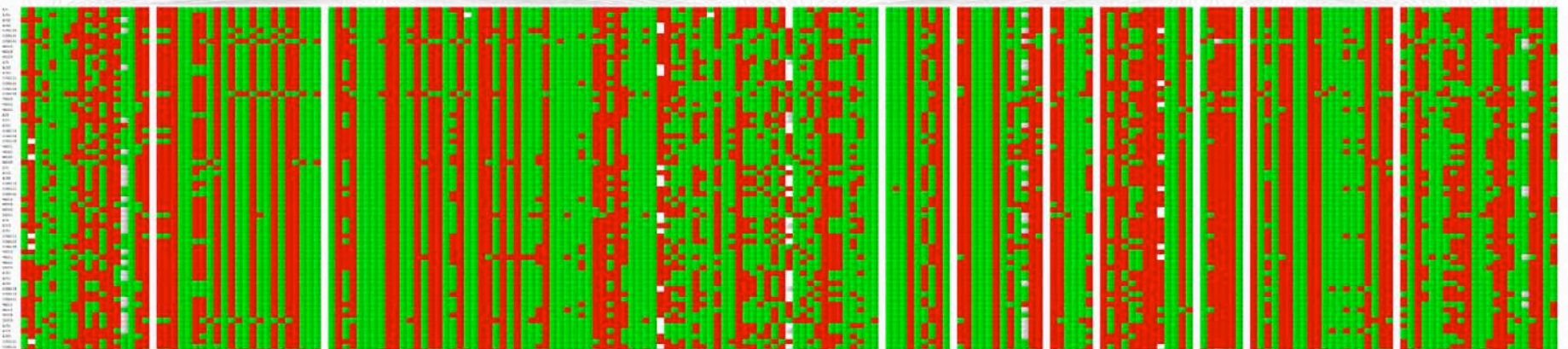


1. ASCERTAINMENT BIAS IN BARLEY ILLUMINA SNP MARKERS

UK Cultivars



Syrian-Jordanian Landraces (SJLC)



Marc Moragues

Graphical genotypes for Chr. 5H

1. ASCERTAINMENT BIAS IN BARLEY ILLUMINA SNP MARKERS

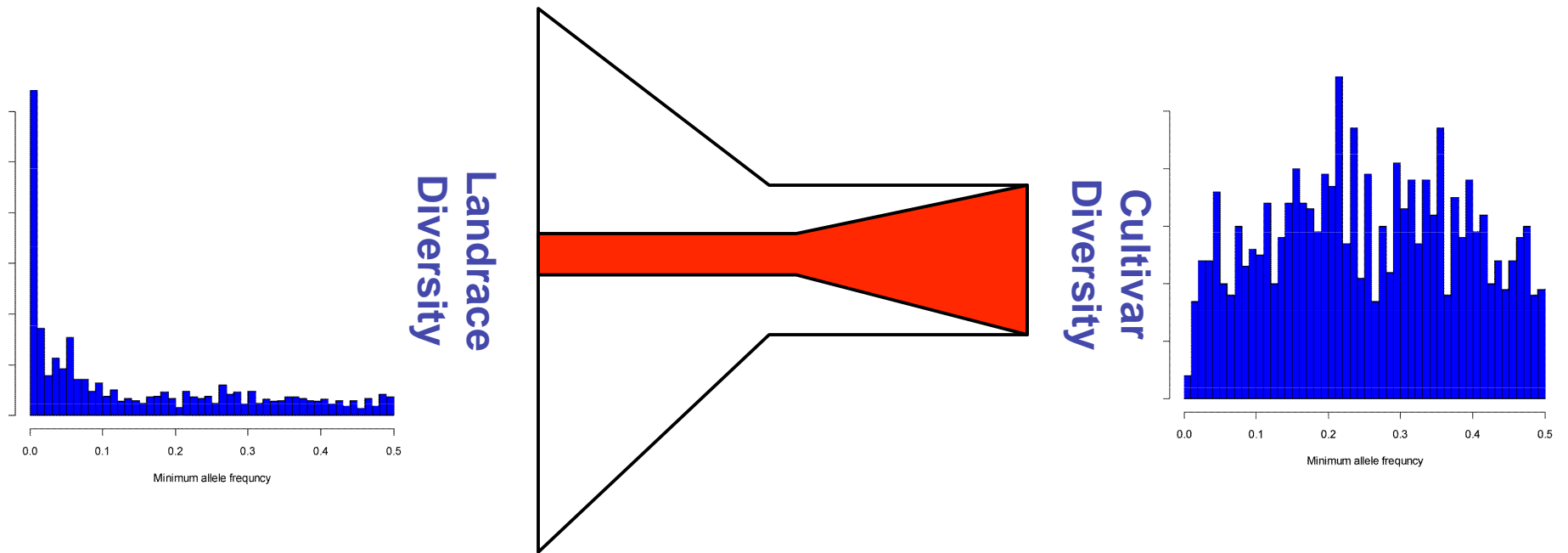
Hypothesis

The BOPA1 marker set reveals lower diversity in landraces because the markers were selected to be polymorphic in cultivar germplasm

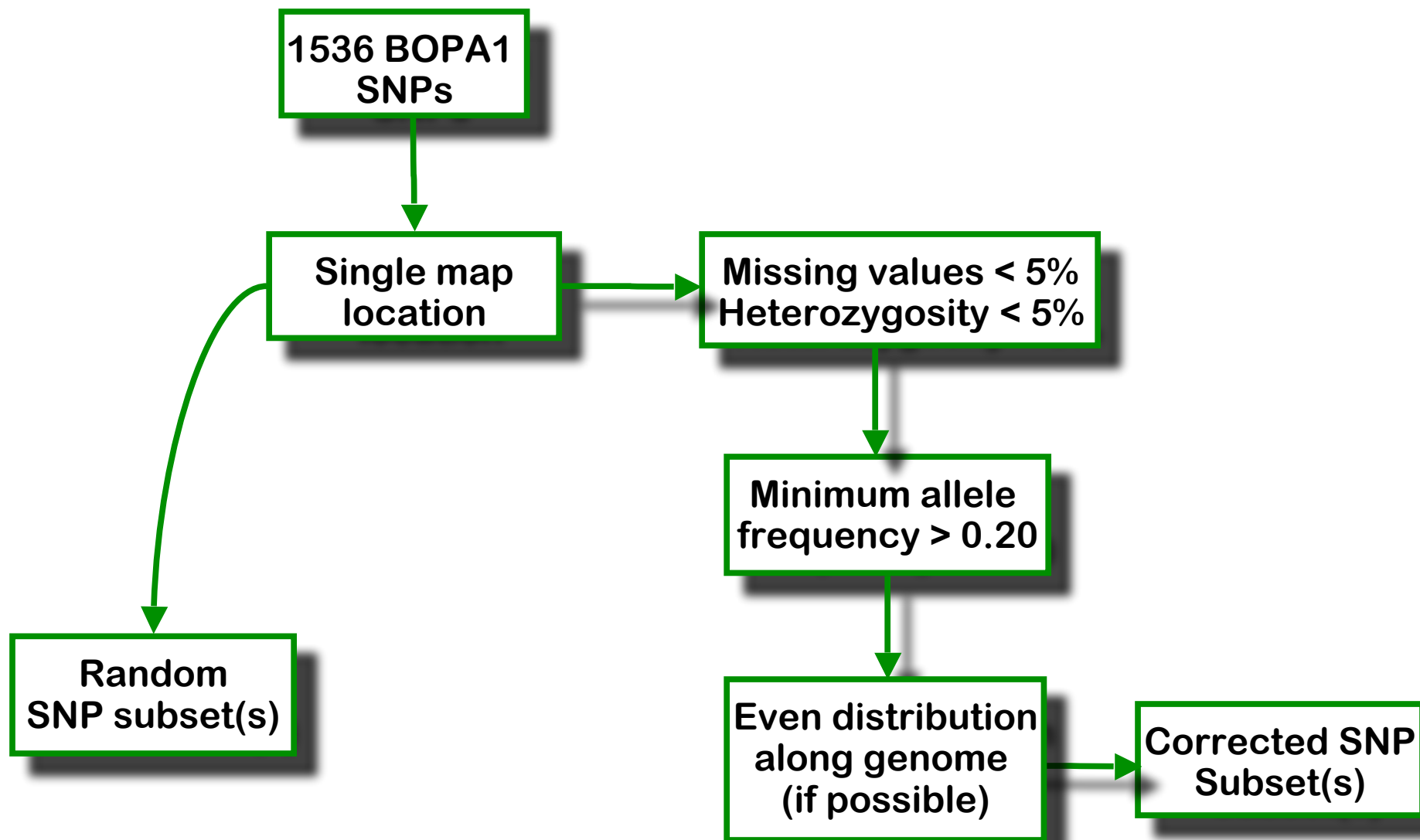
Evidence

Microsatellite analysis and gene sequencing of the SJLC landraces & cultivars reveals higher diversity in the landraces (Joanne Russell, Katie Caldwell)

1. ASCERTAINMENT BIAS IN BARLEY ILLUMINA SNP MARKERS



DESIGNING SNP SUBSETS TO CORRECT FOR ASCERTAINMENT BIAS



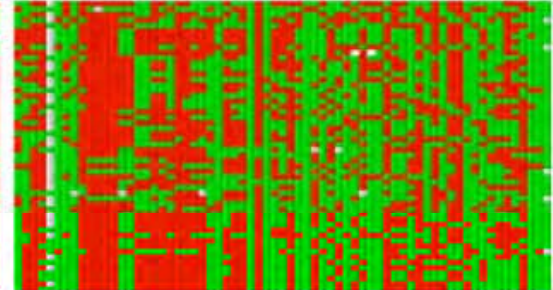
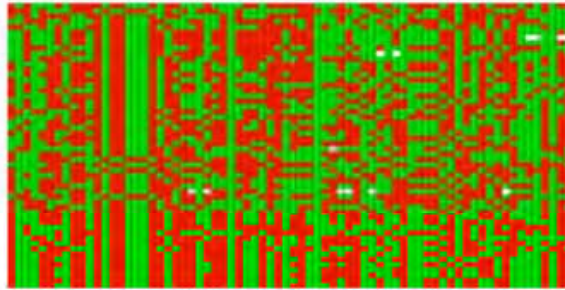
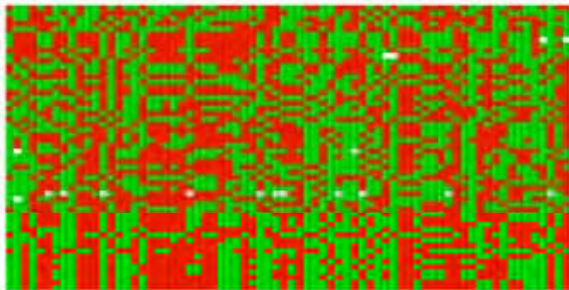
Graphical Genotypes for 384 SNP subsets optimised for cultivars or landraces or unoptimised

Cultivar-optimised SNPs

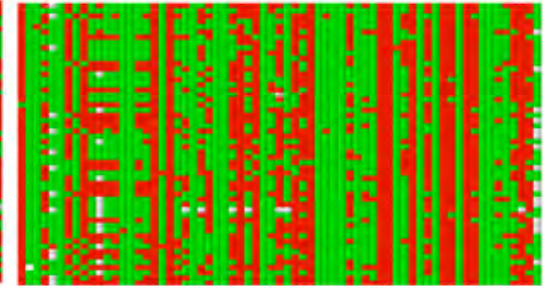
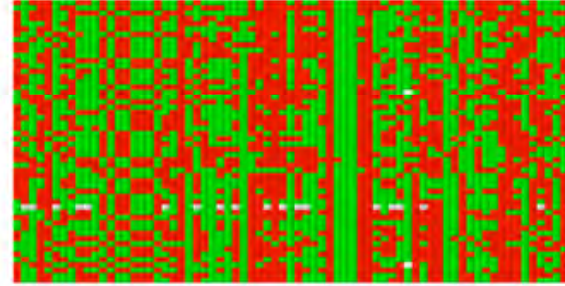
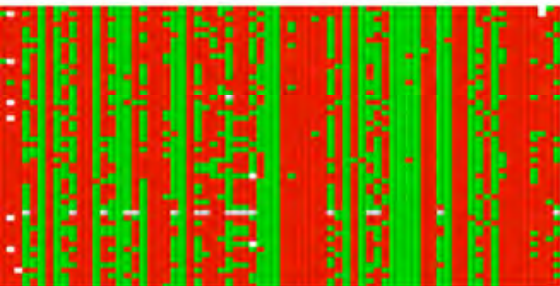
Landrace-optimised SNPs

Unoptimised (random) SNPs

Cultivars



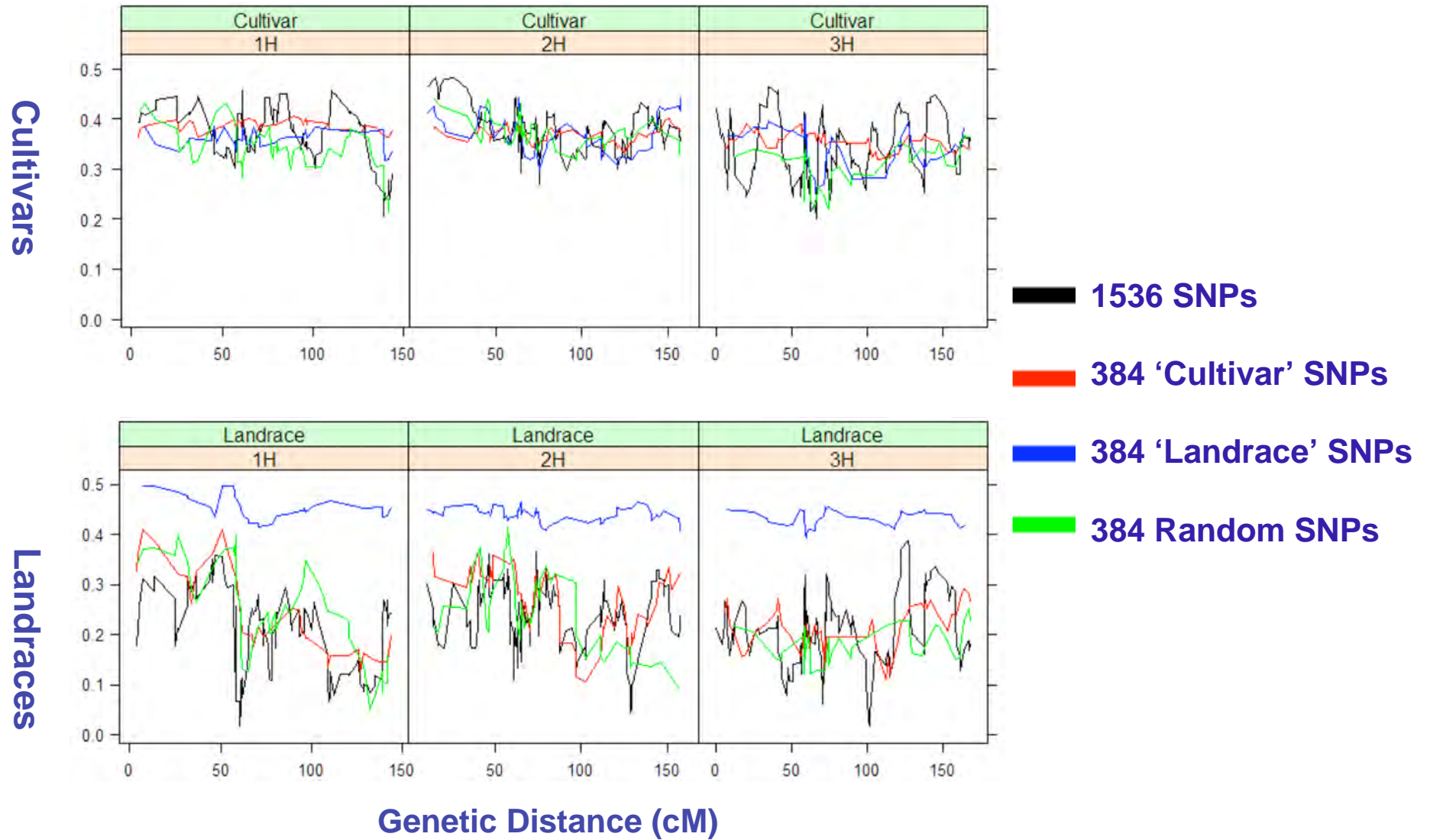
Landraces



Diversity estimated with different SNP subsets

	1536 SNPs	96 'Landrace' SNPs	384 'Landrace' SNPs	96 'Cultivar' SNPs	384 'Cultivar' SNPs	96 random SNPs	384 random SNPs
Assayed in Landraces	0.21	0.50	0.44	0.23	0.23	0.21	0.21
Assayed in Cultivars	0.36	0.37	0.35	0.37	0.37	0.35	0.35

DISTRIBUTION OF DIVERSITY ALONG CHROMOSOMES



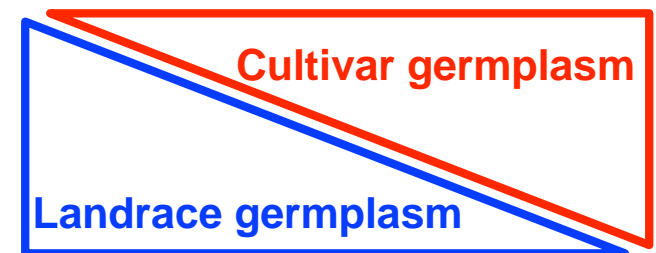
ROBUSTNESS OF POPULATION STRUCTURE ESTIMATES

Mantel test between similarity matrices

	1536 SNPs	'Landrace' 96 SNPs	'Landrace' 384 SNPs	'Cultivar' 96 SNPs	'Cultivar' 384 SNPs	Random 96 SNPs	Random 384 SNPs
1536 SNPs	-	0.88	0.97	0.83	0.91	0.91	0.98
'Landrace' 96 SNPs	0.64	-	0.9	0.76	0.82	0.81	0.87
'Landrace' 384 SNPs	0.82	0.9	-	0.81	0.87	0.88	0.95
'Cultivar' 96 SNPs	0.96	0.67	0.81	-	0.94	0.79	0.83
'Cultivar' 384 SNPs	0.89	0.73	0.86	0.89	-	0.86	0.91
Random 96 SNPs	0.92	0.59	0.73	0.8	0.91	-	0.91
Random 384 SNPs	0.98	0.65	0.84	0.88	0.94	0.89	-

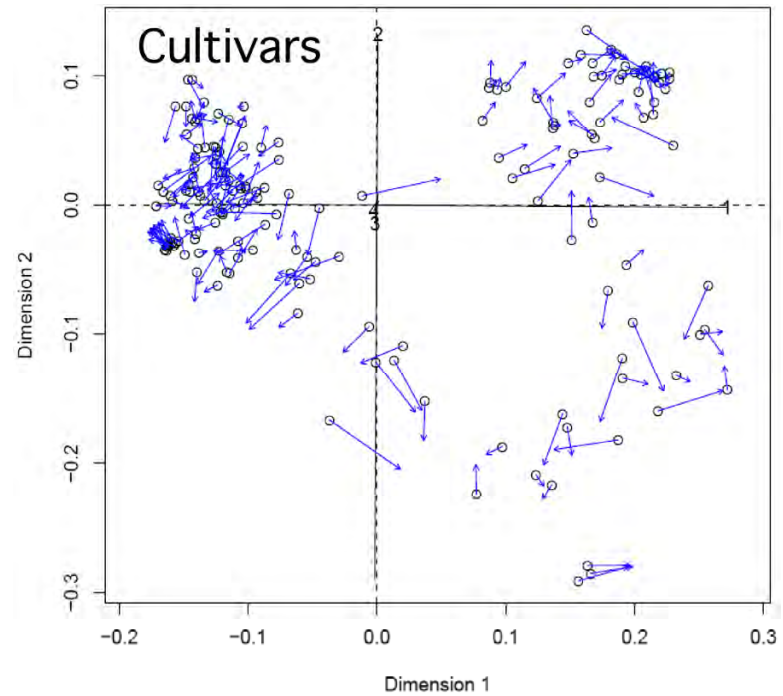
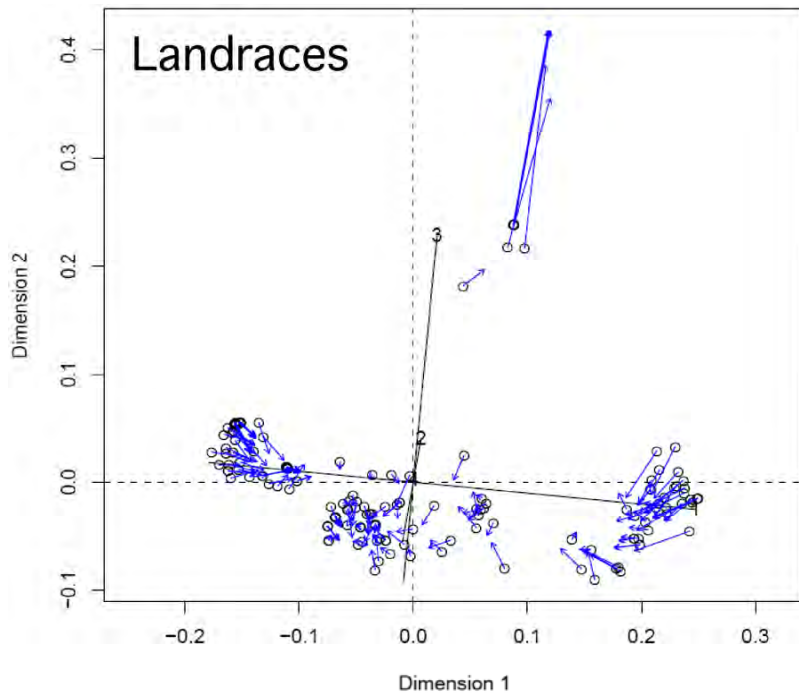
Conclusions:

1. Using 96 SNPs badly erodes the quality of deduced population structure in landrace germplasm
2. 384 SNPs produce more reproducible structures than 96 SNPs and should be OK in many applications
3. For cultivars all 384 SNP subsets work OK and all the random 96 subsets aren't bad either



ROBUSTNESS OF POPULATION STRUCTURE ESTIMATES

Procrustes rotation analysis



L-384 -> 1536

Data resolution coefficient statistics (DR) for all SNP sets

SNP set	Cultivars	Landraces
1536 SNPs	0.974	0.973
384 'Landrace' SNPs	0.912	0.954
96 'Landrace' SNPs	0.694	0.856
384 'Cultivar' SNPs	0.929	0.906
96 'Cultivar' SNPs	0.790	0.665
384 Random SNPs	0.911	0.902
96 Random SNPs	0.738	0.718

1. Using 96 SNPs yields poor quality data resolution in all germplasm
2. 384 SNPs produce DR statistics approaching 1536 SNPs
- 384 is a good number of SNPs for routine diversity applications
3. 1536 SNPs yield modest improvement in diversity applications and we suggest that the extra cost may often not be justified.

Genetic Diversity data

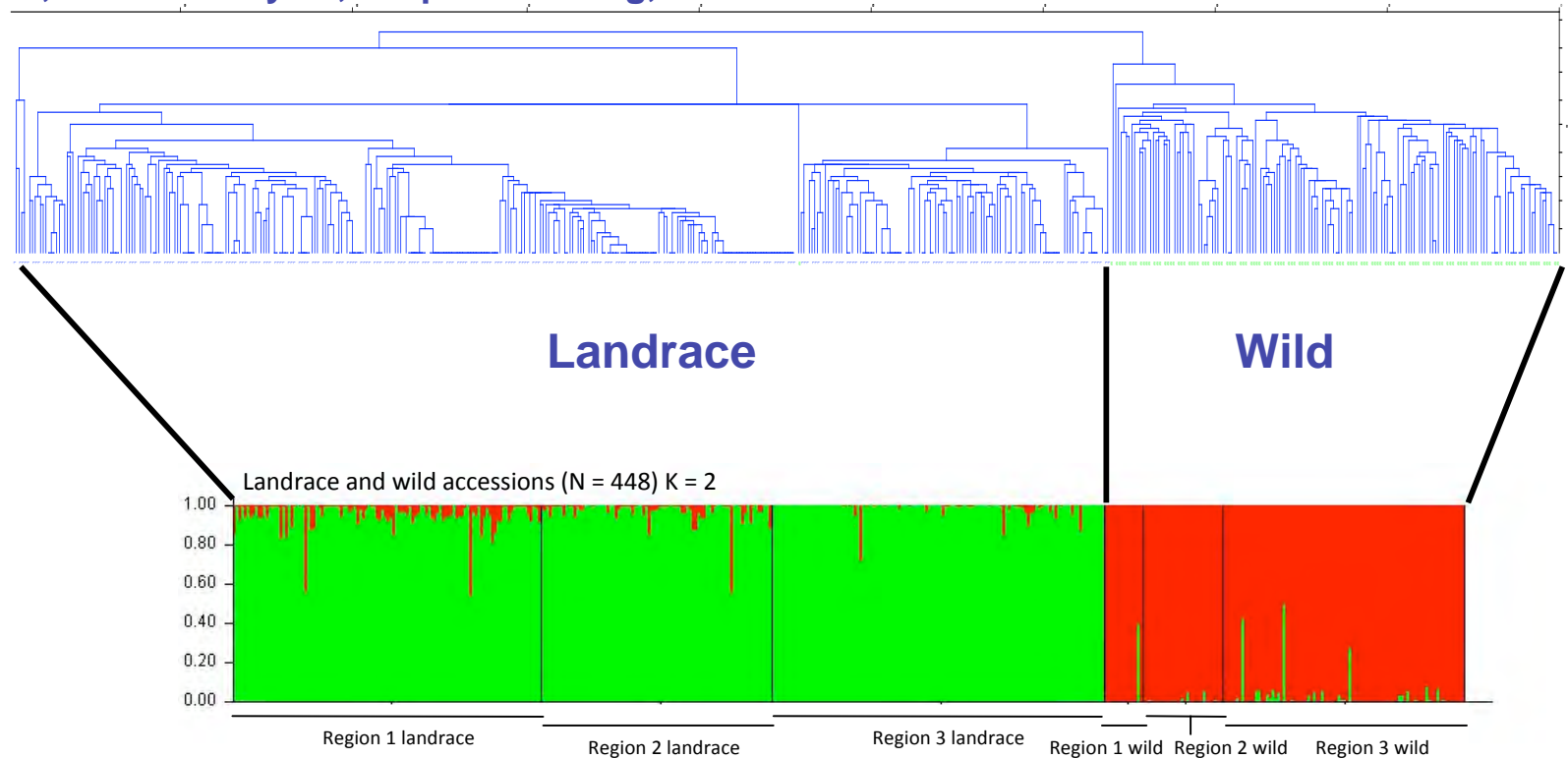
	N	Polymorphic loci (from 1135 scored)	'H' (Nei unbiased diversity)
Total accessions	448	1074	
Landrace accessions	317	914	
Region 1	112	812	0.173
Region 2	84	787	0.158
Region 3	121	701	0.120
Mean values across Regions	105.6	767	0.150
Wild accessions	131	1006	
Region 1	14	540	0.136
Region 2	29	769	0.192
Region 3	88	958	0.218
Mean values across Regions	43.7	756	0.182

1135 BOPA1 loci scored (includes monomorphic markers, excludes markers with low quality scores)

Region 3 has lowest diversity for landrace material and highest diversity for wild barley (from much more accessions)

Wild and landrace barleys are very clearly resolved into separate populations by BOPA1 marker analysis

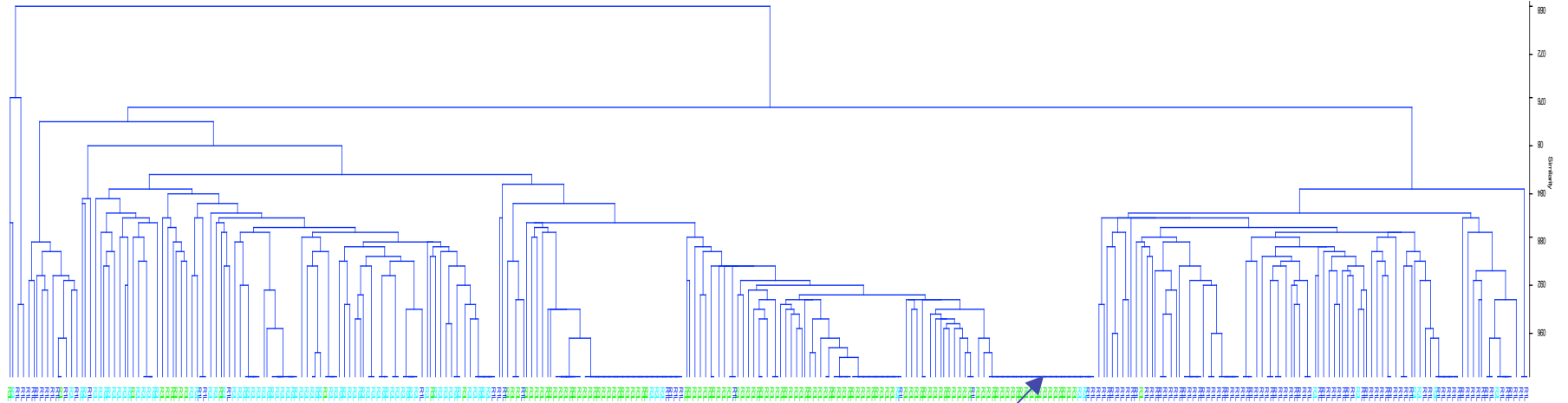
PAST, cluster analysis, simple matching, UPGMA



Structure analysis, K=2

The wild population is less structured than the landraces - longer branches

Clustering within landraces



Region 2

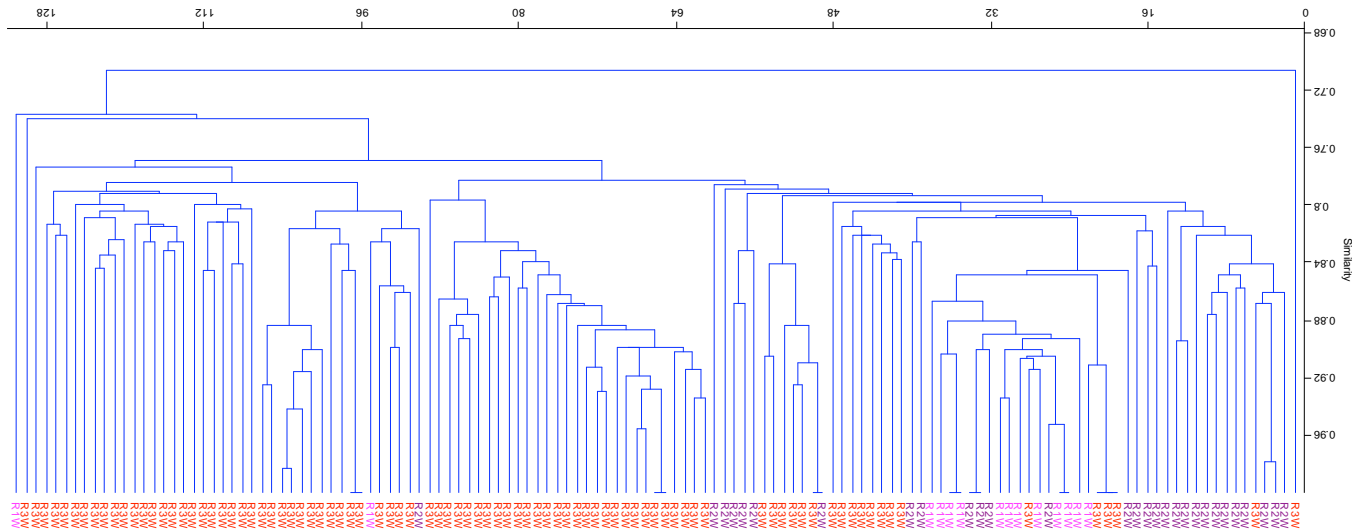
Region 3

Region 1

Strong region-specific clustering

**Identical accessions are evident
(only 63% of the landrace population is unique)**

Clustering within wild barleys

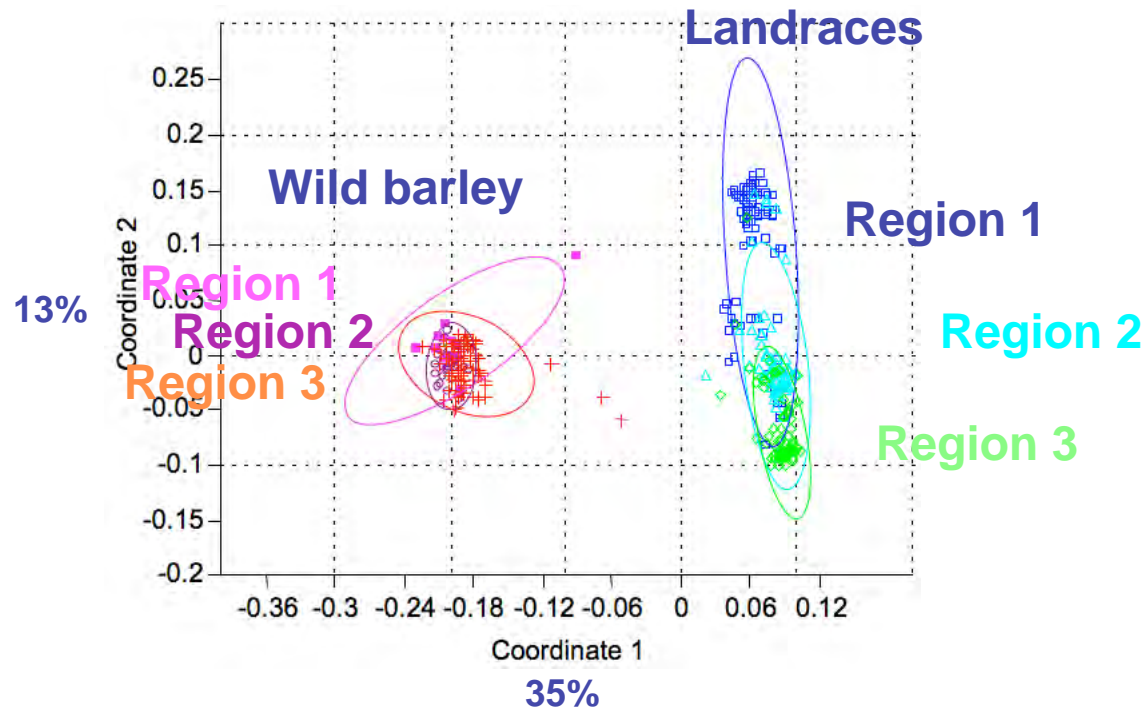


More modest clustering with respect to geographical region

The wild population is less structured than the landraces

94% of all accessions are unique

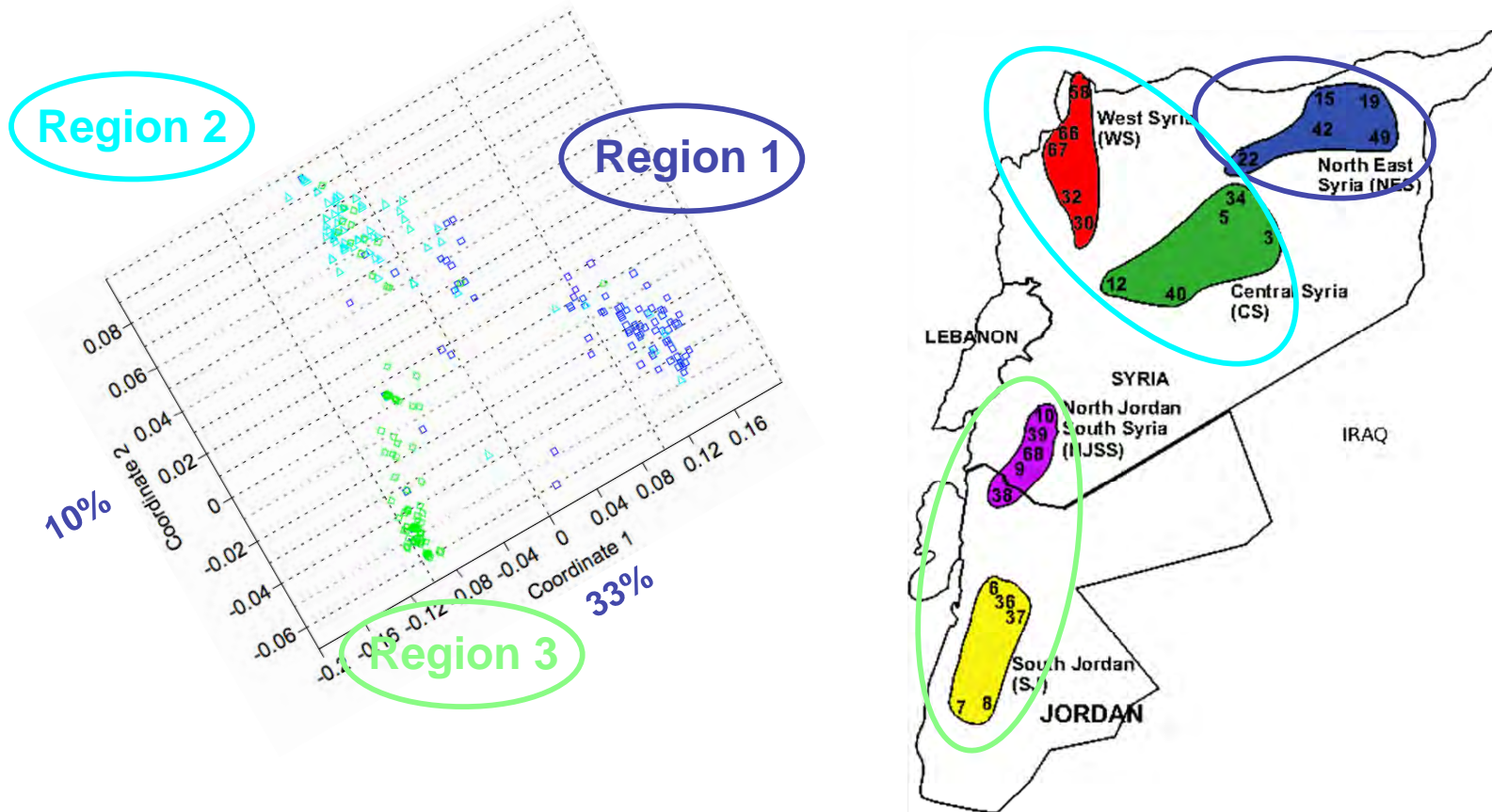
Relationship between wild and landrace barley populations derived from similar geographical locations - principal coordinate analysis



Very little overlap is evident between the genetic diversity of wild and landrace germplasm from the western Fertile Crescent -a few samples show evidence of introgression

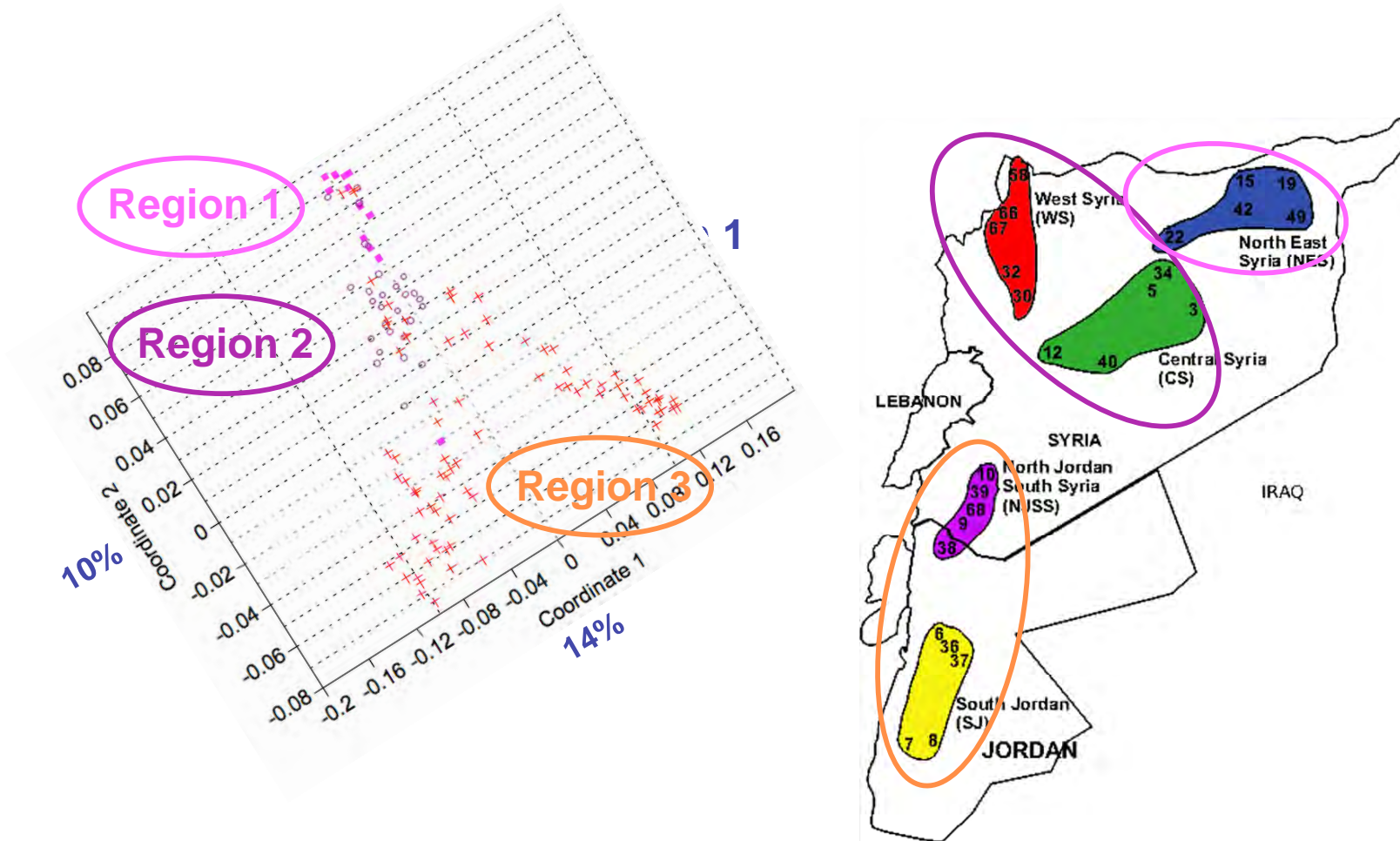
The BOPA1 SNP set reveals more diversity in the SJLC landraces than the WBDC subset wild barleys - maybe another case of bias?

Relationships within landrace barley populations and relationship to geography



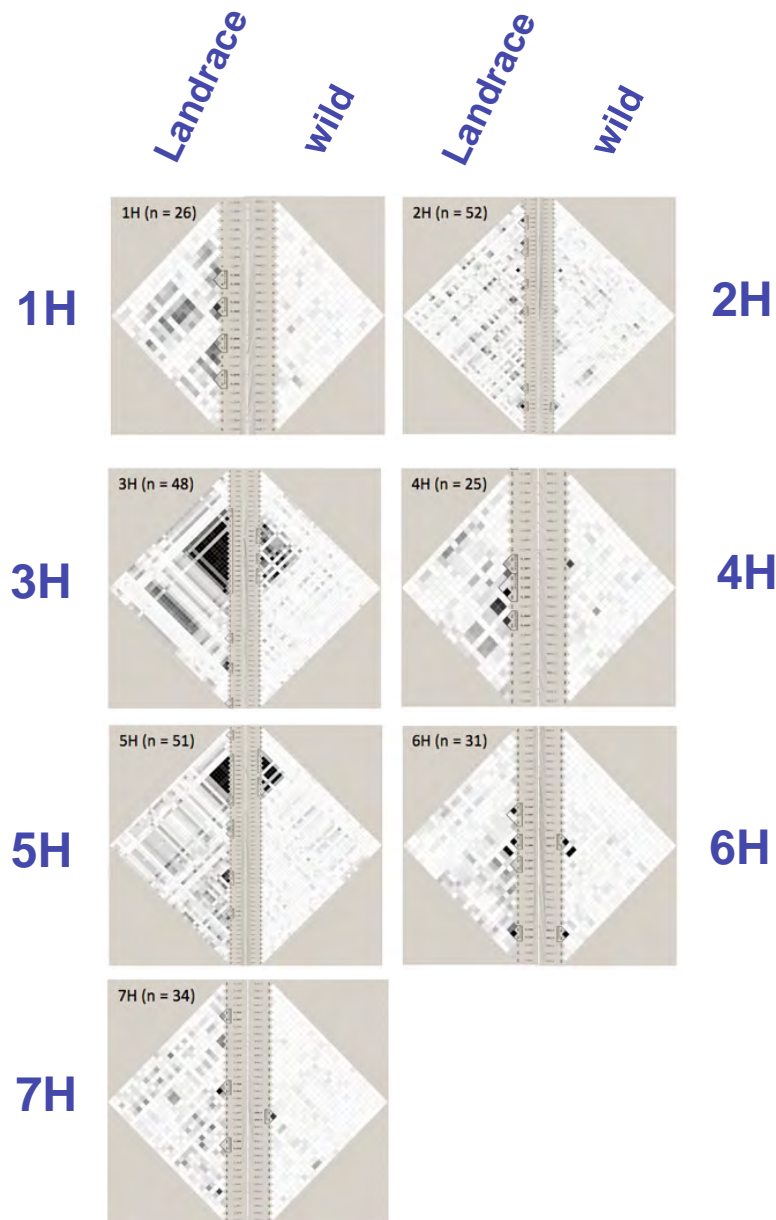
Excellent correlation between first two axes (=43% of the variation) and origin location

Relationships within wild barley populations and relationship to geography



Weaker correlation between first two axes (=24% of the variation) and origin location

Linkage disequilibrium by chromosome



LD decays more rapidly in wild barley than landraces

Strong chromosome effect, with particular regions (centromeric?) showing strong LD conservation

CREDITS

University of Dundee

Marc Moragues

Xin Xu

SCRI

Joanne Russell

Jordi Comadran

Luke Ramsay

Iain Milne

David Marshall

Robbie Waugh

ICARDA

Stefania Grando

U. Minnesota

Joy Roy

Brian Steffenson

The World Agroforestry Centre

Ian Dawson

The EXBARDIV Consortium

Alessandro Tondelli

Luigi Cattivelli

Christina Ingvarlsen

Soeren Rasmussen

Ruslan Kalendar

Alan Schulman

Florian Schnaithmann

Klaus Pillen

Rajiv Sharma

Benjamin Kilian

CREDITS



ERA-PG



Generation Challenge Programme



The UK AGOUeB Project