

Towards more efficient mining of genetic variation in *ex situ* collections

Michael Mackay, Kenneth Street, Evgeny Zuev, Navreet Kaul Bhullar, Mustapha El Bouhssini, Jan Kanopka, Olga Mitrofanova



ITMI / COST Workshop 2009, Clermont-Ferrand



Backdrop

Utilizing PGR during past 100 years:

- Collection, observation, experimentation.
- PGR movements of 20th Century.
- Core collections – and derivatives.
- Mass screening – e.g. Ug99 in recent years.
- More recent genomic (and other ‘omic’) research.

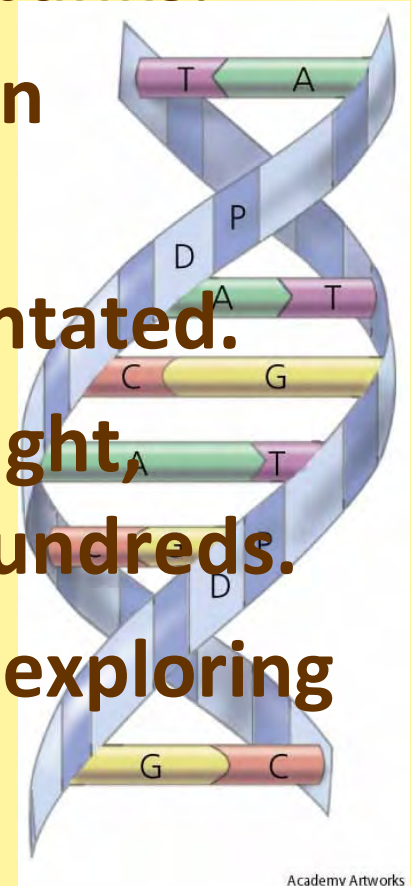
Cost / benefits?

What degree of success?



Upon reflection

- 6 million accessions in 1400 genebanks.
- Many successes, such as the Green Revolution, but at what cost?
- Core collections are not trait orientated.
- Depending on variation being sought, probably only need to evaluate hundreds.
- Perhaps there's a smarter way of exploring and exploiting PGR?



Towards efficiency

Use all available information to ‘focus’ in on the accessions most likely to possess the desired genetic variation.

- Traditional
- Derived info
- GIS/environment
- Intuition





Matching needs

Most PGR users want variation for specific traits:

- *“I am screening for drought adaptation– could you send me some landraces from Afghanistan and some other dry countries”...*
- *“I am looking for new salt tolerance genes can you send me some wild relatives from salty areas”...*

NOT

- *“I want a random sample of accessions to grow to see if I can find something useful”*

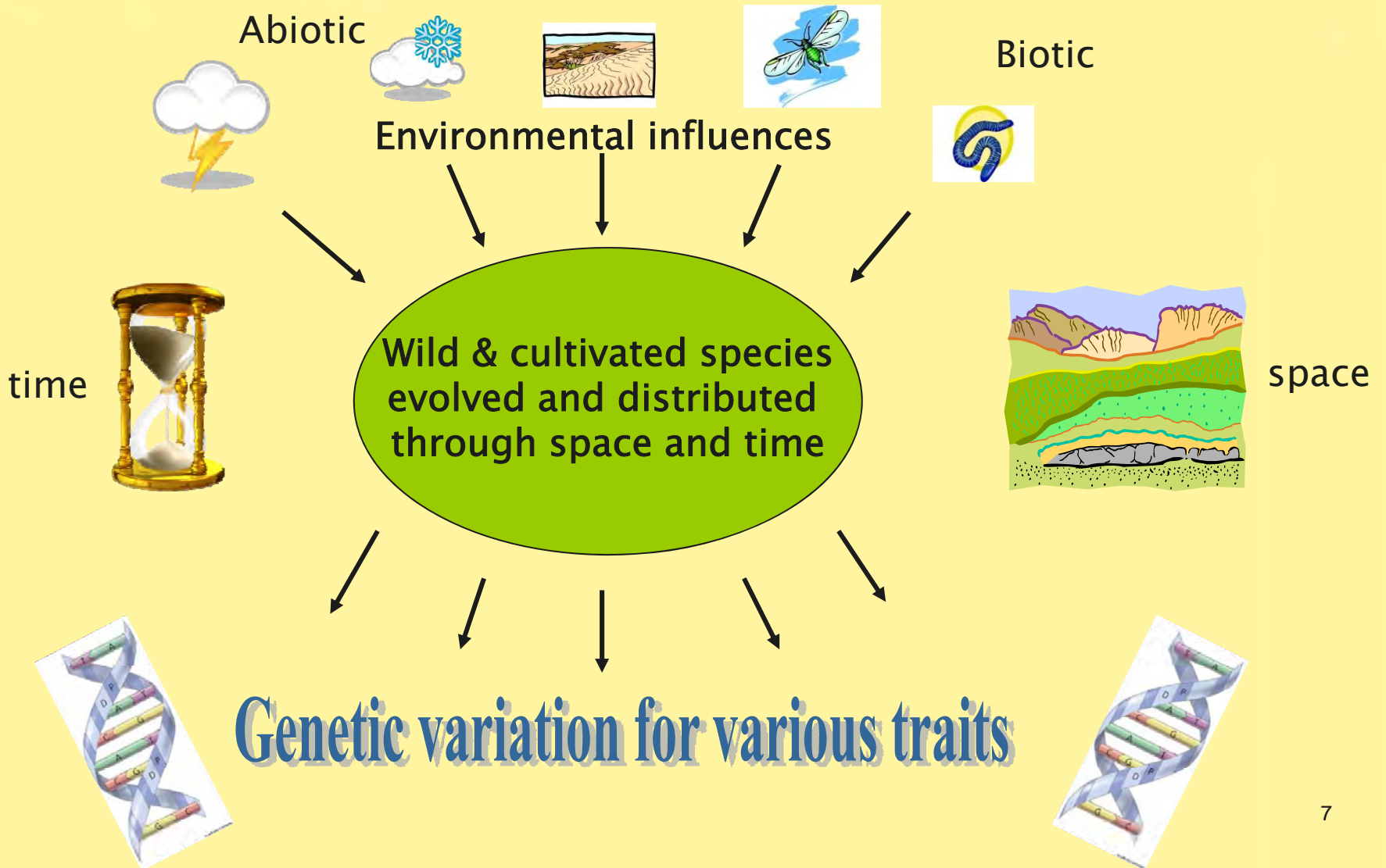


Stated purpose of barley germplasm requests to ICARDA over the last 20 years

(Source: K Street)

Purpose	Acc. sent
Disease screening	51656
Abiotic stress screening	18221
Diversity studies - molecular	4296
Agro-morphological screening	3442
Grain quality	1062
Crossing blocks	1024
Core collection backup and evaluation	285
Unspecified or not specific	43145
Total	122939

A source of variation



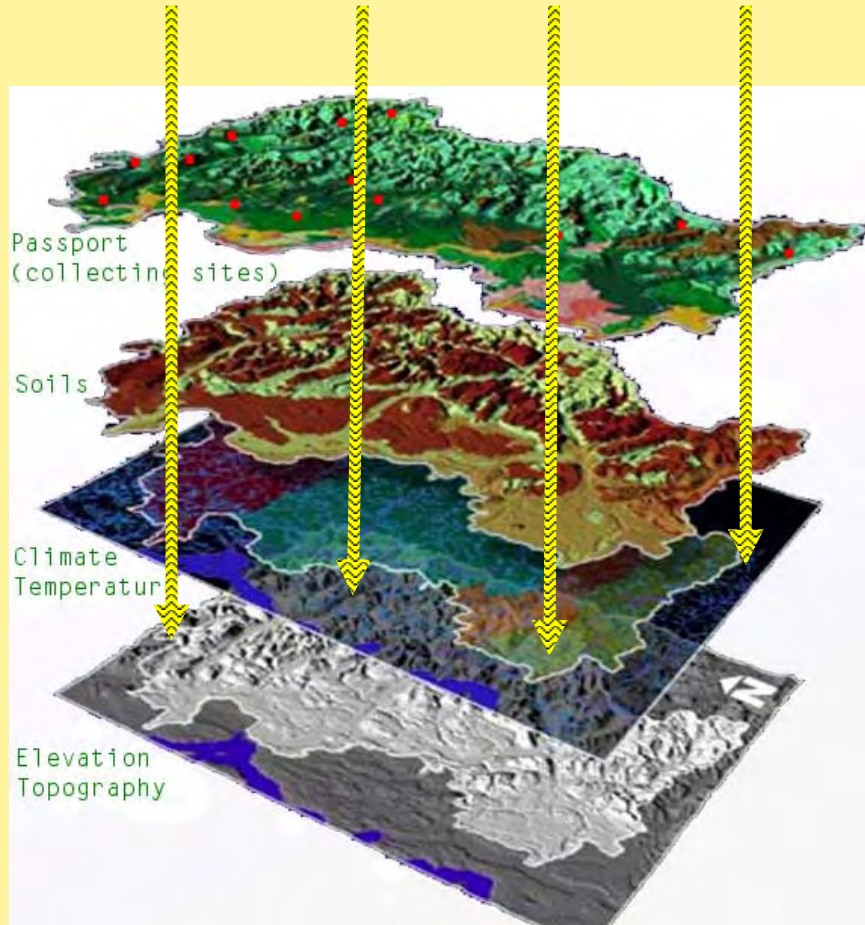


Mining natural variation

Use nature's variation – it's free!

- **There is a wealth of variation sitting in genebanks – largely untapped.**
- **By linking traits, environments (and associated selection pressures) with genebank accessions (e.g. landraces and crop relatives) we can 'focus' in on those accession most likely to possess specific genetic variation**

Linking environmental data to collection sites



choose accessions from environments that impose selection pressure for a given trait

Eg: for pests - select material from environments that favor a high pest load



The picture

**Genotypes x Environments x Time¹
= Genetic Variation**

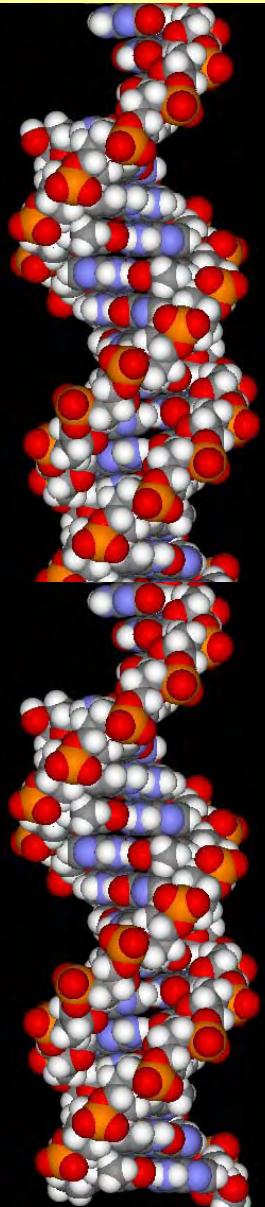
Can we use the same evolutionary principles in reverse to identify the environments that 'engender' trait specific genetic variation?

¹ plus some selection

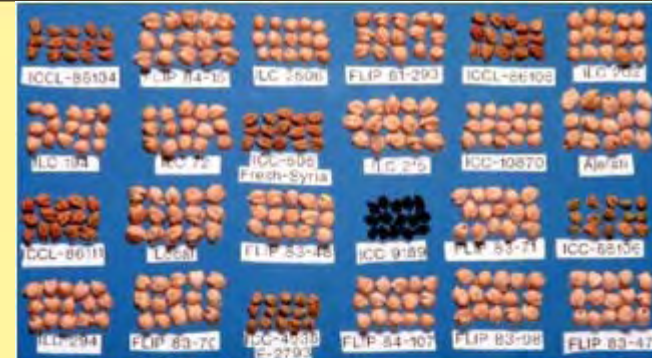
Examples of eco-geographic variation of traits linked to environmental influences

Species	Trait	Environment influence	Reference
<i>Trifolium repens</i>	Cyanogenesis	Low altitudes, high winter temp., low summer rain, spring cloudiness	Pederson, Fairbrother et al. 1996
Annual legumes	Seed dormancy, early flowering, high seed to pod ratio	Aridity	Ehrman and Cocks 1996
Bread wheat	Tolerance to Boron toxicity	Soil type	Mackay (1990)
Bread wheat	Russian Wheat Aphid (RWA) resistance	Altitude, winter temp, RWA distribution	Bohssini, et al accepted for publication 2008
<i>Triticum dicoccoides</i>	Drought resistance	Temperature, aridity	Peleg, Fahima et al. 2005
Durum wheat	Glume colour and beak length	Altitude	Bechere, Belay et al. 1996
<i>Triticum dicoccoides</i>	Heading date, culm length, biomass, grain yield and its Components	Climate, soil and water availability	Beharav and Nevo 2004
Durum wheat	Glutenin diversity	Precipitation, minimum January temperature, altitude.	Vanhintum and Elings 1991
Woody perennials	More efficient RUBISCO activity	temperature, aridity	Galmes et al, 2005
Barely	Hordatine accumulation (disease defence)	Water relations, temperature and edaphic factors	Batchu, Zimmermann et al. 2006

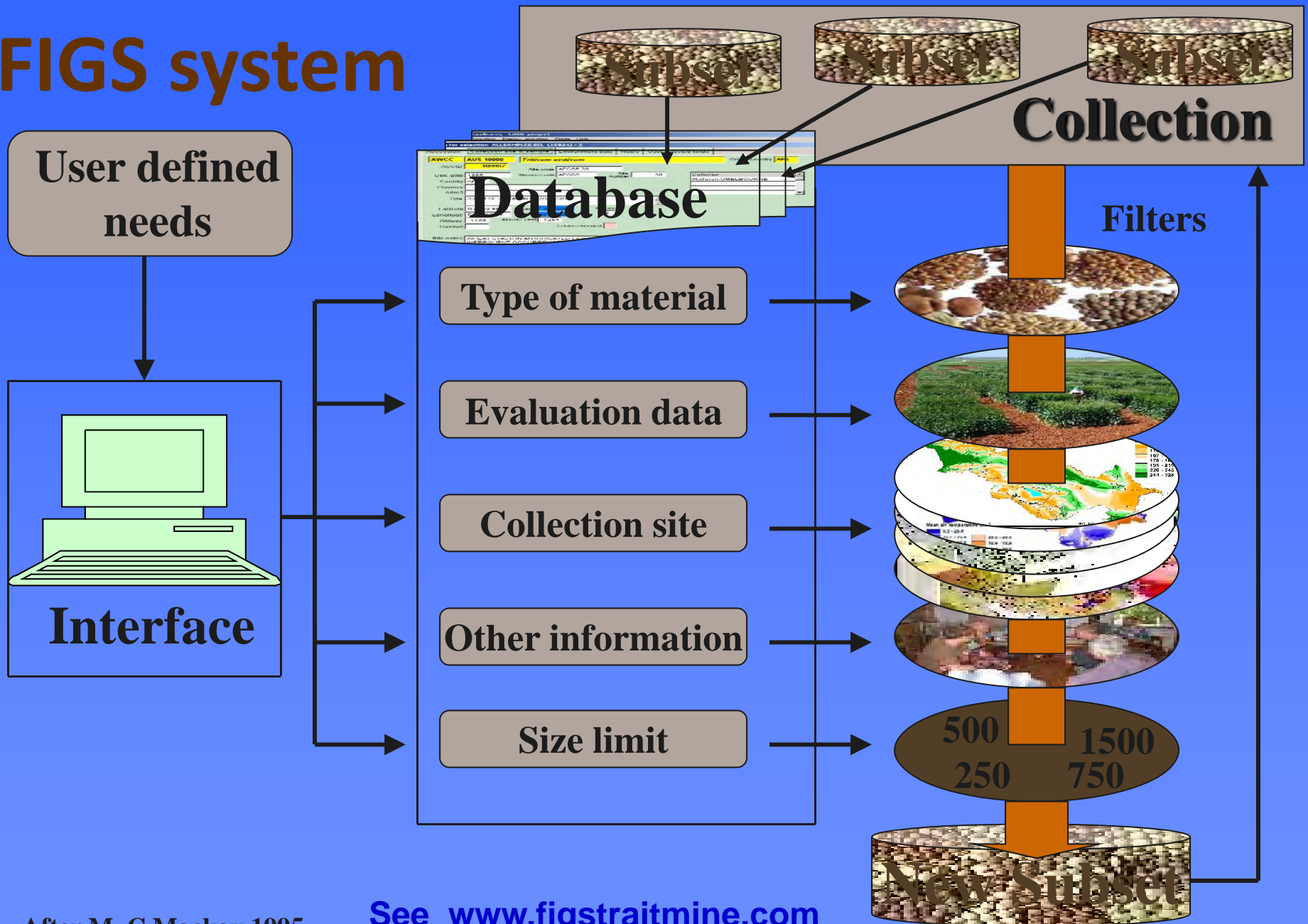
Focusing in on the 'best bet' accessions



- ❖ Focused
- ❖ Identification of
- ❖ Germplasm
- ❖ Strategy



FIGS system

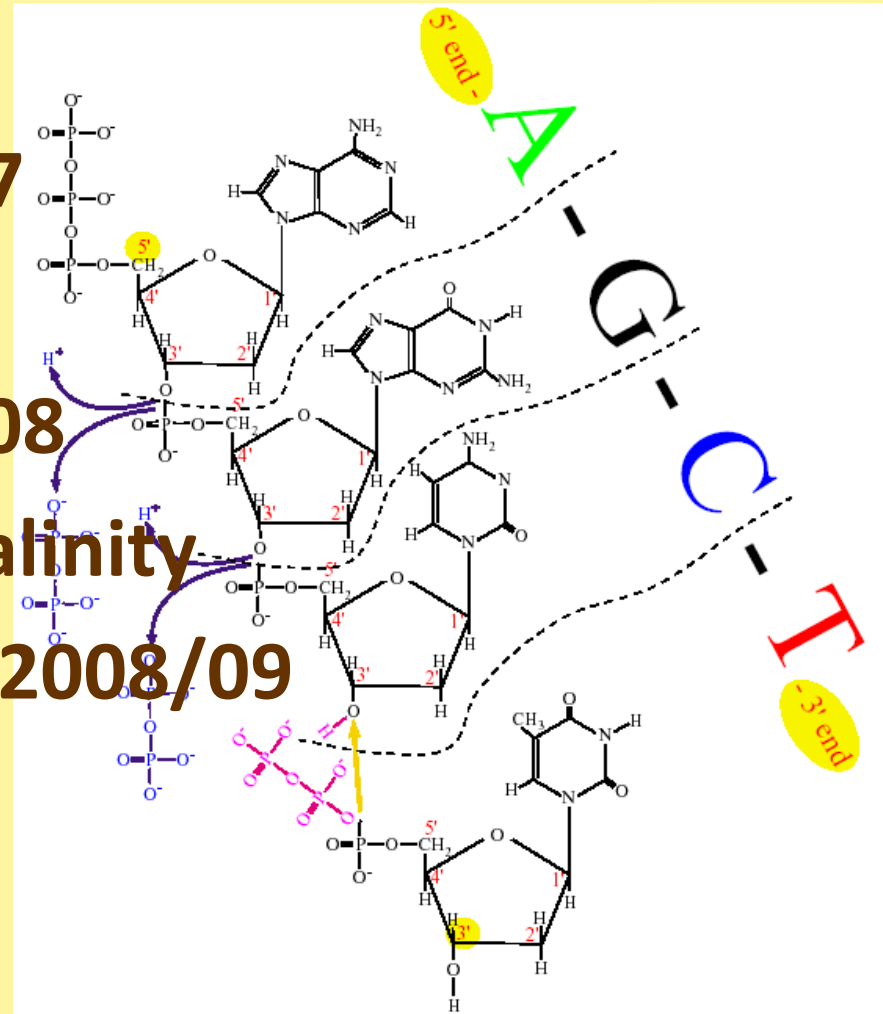


After M. C Mackay 1995

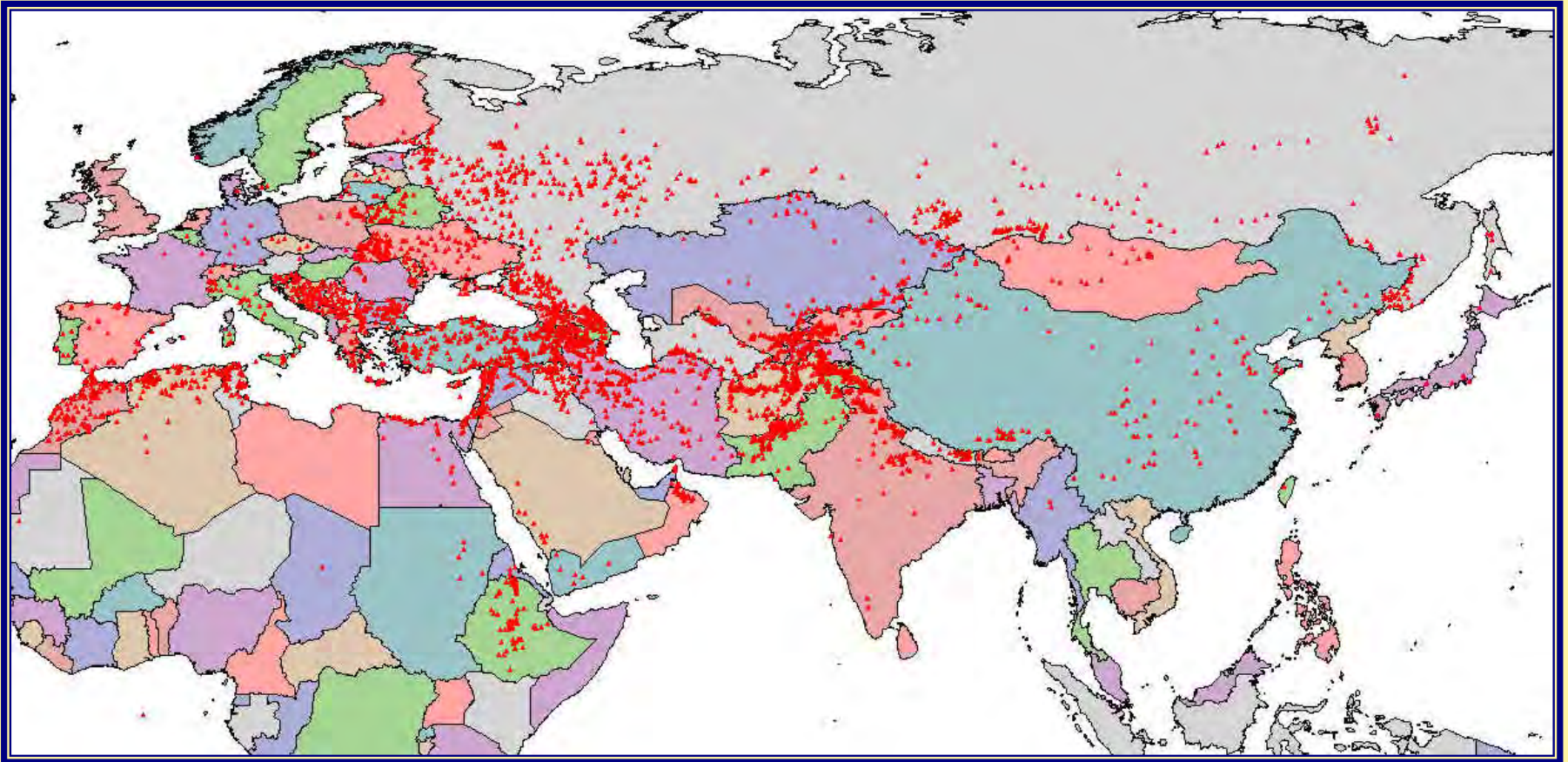
See www.figstraitmine.com

Some examples

- Boron toxicity in 1980s
- RWA in 1990s and 2007
- Sunn pest ca. 2007
- Powdery mildew in 2008
- Promising results for salinity
- Ug99 desktop study in 2008/09

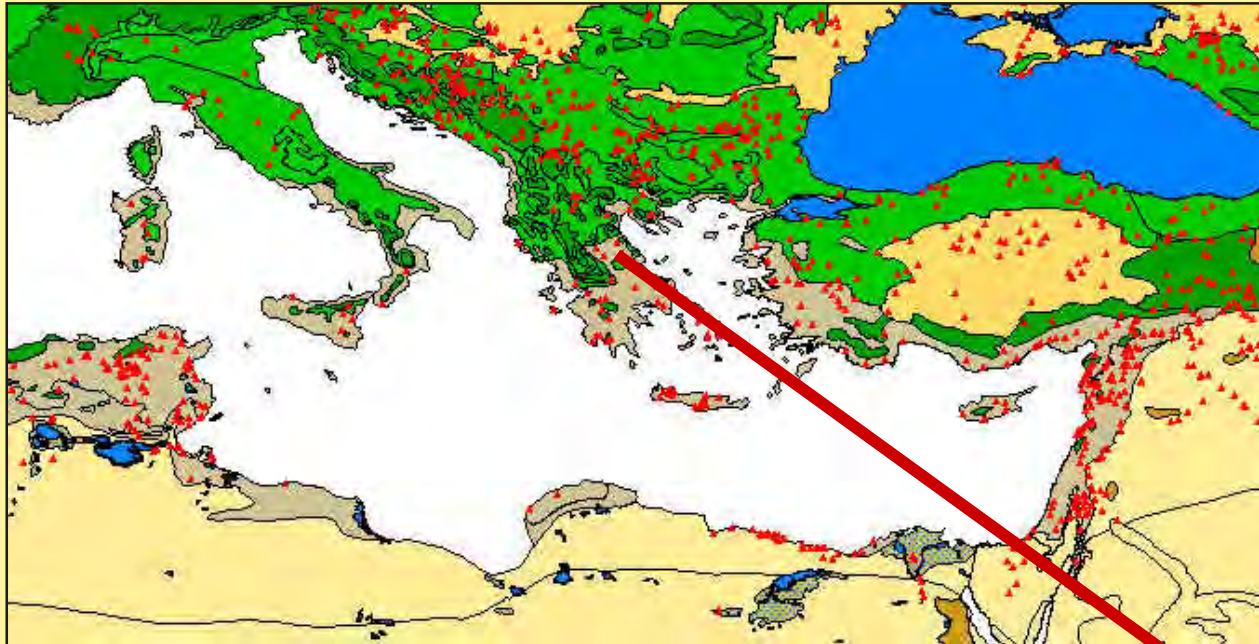


Distribution of bread wheat landrace collection sites – VIR, ICARDA, AWCC



17,000 accessions and 6,300 sites

The Boron Story



Wheat landraces from marine origin soils in Mediterranean region provided all the genetic variation needed to produce boron tolerant varieties





- ICARDA's entomology unit has screened 1000s of hexaploid wheat accessions for resistance to the Syrian RWA bio-type without success.
- The Genetic Resource unit put together a best-bet FIGS set of 500 accessions for screening.

12 accessions showed high to moderate resistance



This is the first time we have found RWA resistance for the Syrian biotype in hexaploid wheat at ICARDA. Marker analysis indicate new genes.

Sunn Pest



No sources of juvenile plant resistance previously found in hexaploid wheat

15% of ICARDA collection screened without result

A FIGS set of 534 (3.9% of collection) accessions was developed for screening

Sunn pest set selection



Starting with a set of over 16,000 accessions from VIR, ICARDA and AWCC

- Selected material collected between 30-45 latitude, 35-80 longitude.
- Exclude CHN, PAK, IND as only recent reports of Sunn pest in these countries – retained 6328.
- Randomly selected only one accession per site – retained 2830 accessions.
- Excluded particularly dry environments - rainfall below 280mm/year.
- Excluded sites where the winters temps fall below -10 degrees – 1502 accessions remaining.

534 accessions screened for Sunn Pest at ICARDA



25 10:02AM



Success



**8 landrace accessions from
Afghanistan and**

**2 from Tajikistan identified as
resistant at juvenile stage**

**Now developing mapping
populations**

Verified under controlled
conditions



FIGS powdery mildew set



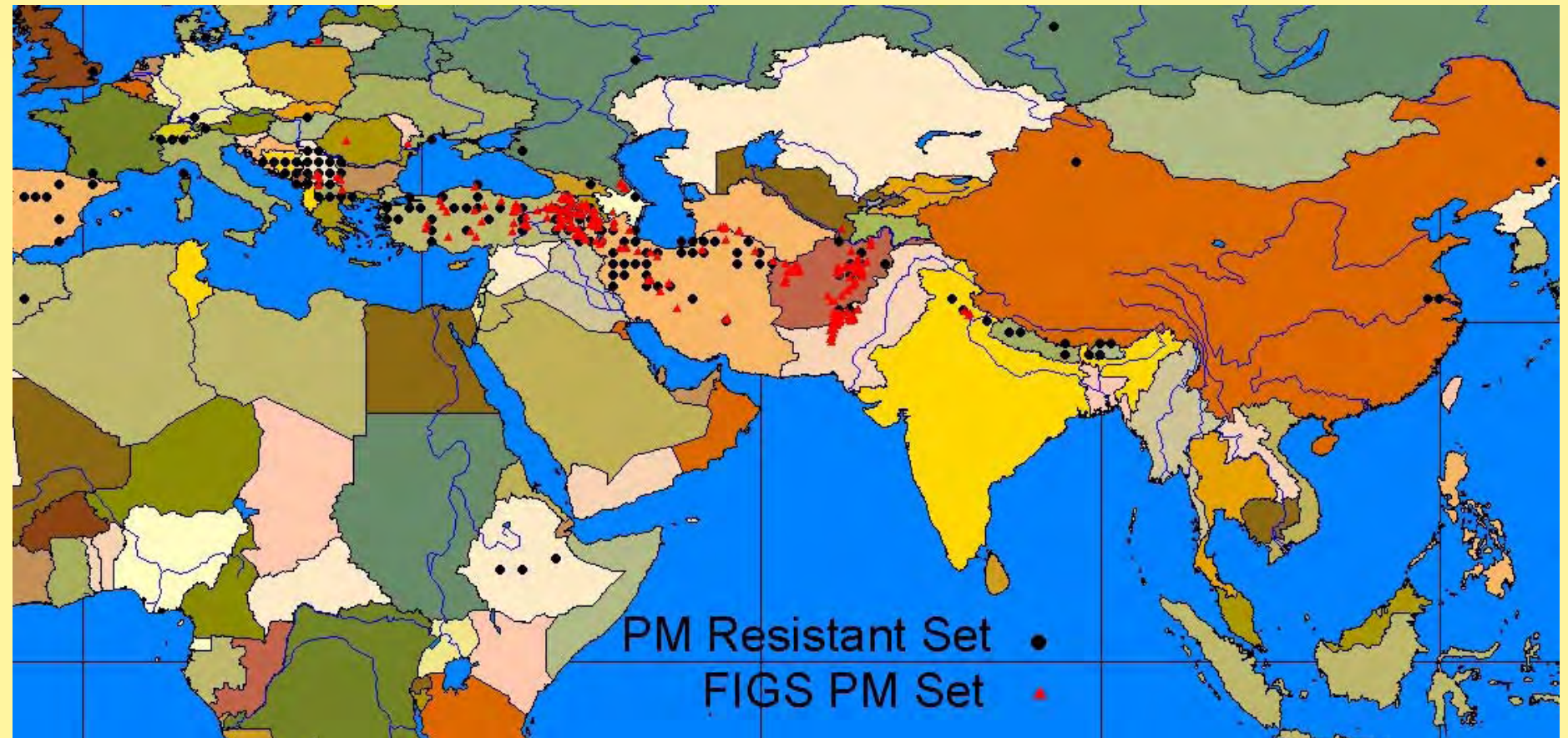
- **Approached by the University of Zurich to put together a powdery mildew set FIGS set**

Method

1. **Obtained collection site data for a set of 400 PM resistant accessions from the USDA (Harold Bockelman).**
2. **Statistically compared the collections sites of the resistant set and the Yemen test set**
3. **Chose accessions from sites with a high degree of similarity to sites where resistance has been found before.**



FIGS powdery mildew set



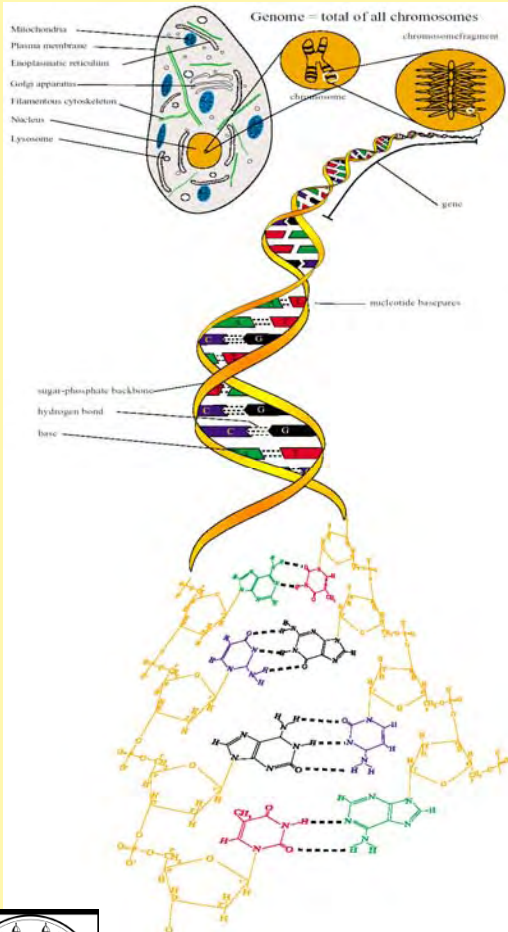
FIGS PM set



- **Results of screening**
 - Starting with a total pool of 16000 accessions collected from 6159 sites, the FIGS process chose 1320 accessions collected from 420 sites.
 - Accessions infected with 4 powdery mildew isolates which were avirulent or virulent to the known *Pm3* alleles.
 - Of these 420 sites, 40% yielded accessions that were resistant to the isolates used – 211 accessions

Example of Impact: Allele mining

- 16,000 bread wheat landraces
- 1,300 chosen using FIGS¹ method
(Passport + evaluation + GIS + environment data)
- Phenotyping showed 211 accessions either R or IR
- For *Pm3*:²
 - 100 years classic genetics = 7 alleles
 - FIGS + MoBo + 2 years = 7 new alleles
 - At least two have new race specificity



¹ FIGS = Focused Identification of Germplasm Strategy (various, incl. Mackay *et al*, manuscript in preparation)

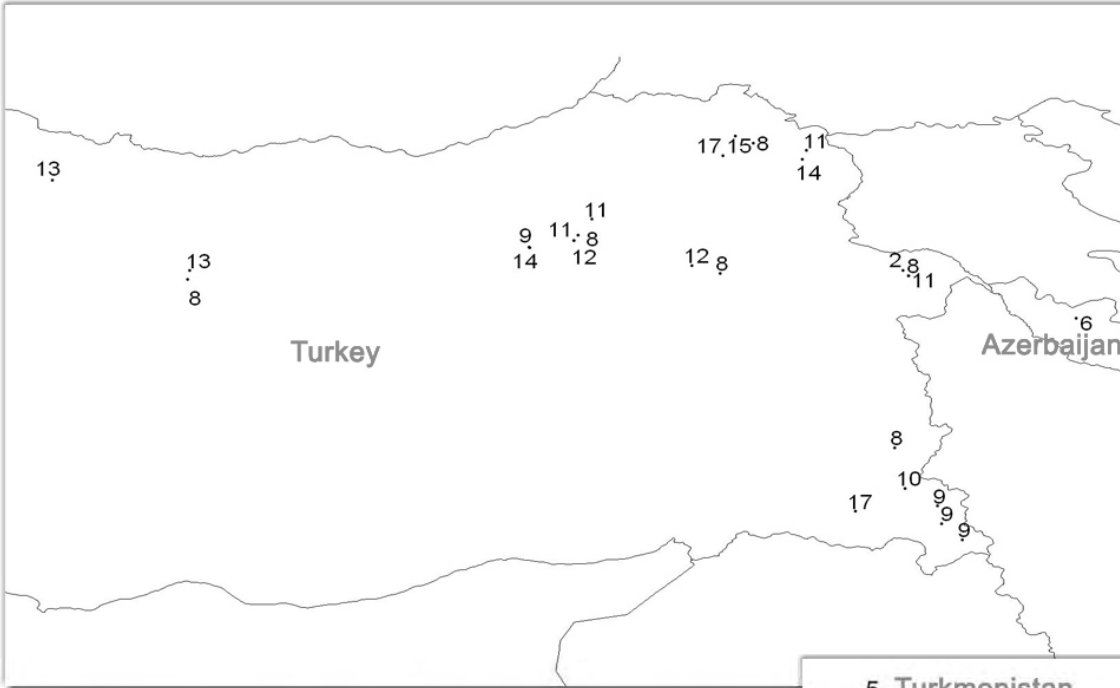
² Kaur K; Street K; Mackay M; Yahiaoui N; Keller B (2008). Allele mining and sequence diversity at the wheat powdery mildew resistance locus *Pm3*. 11th IWGS, 24-29 Aug., Brisbane)



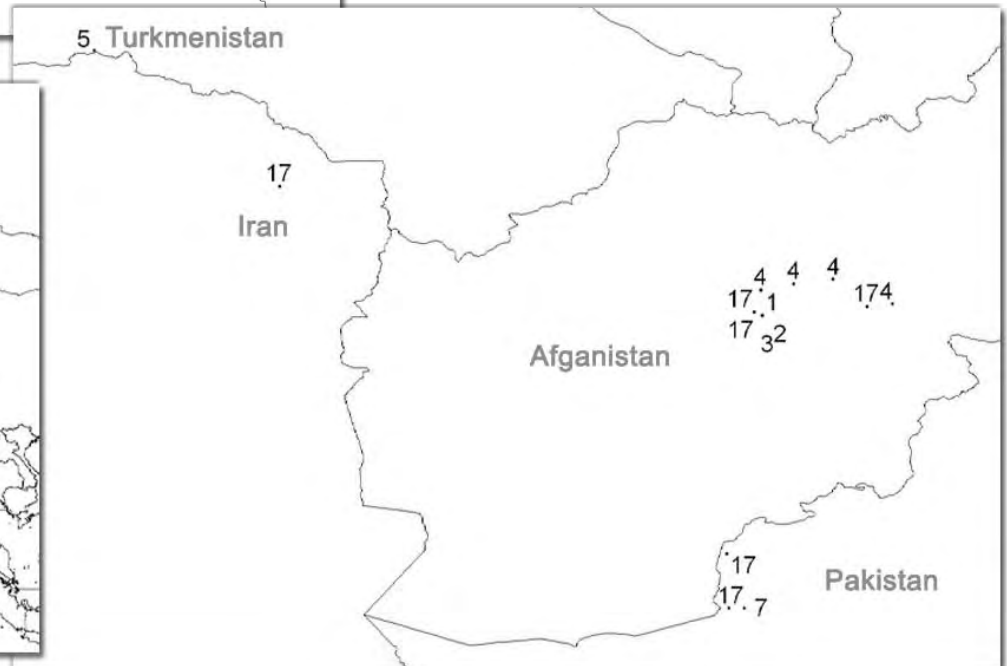
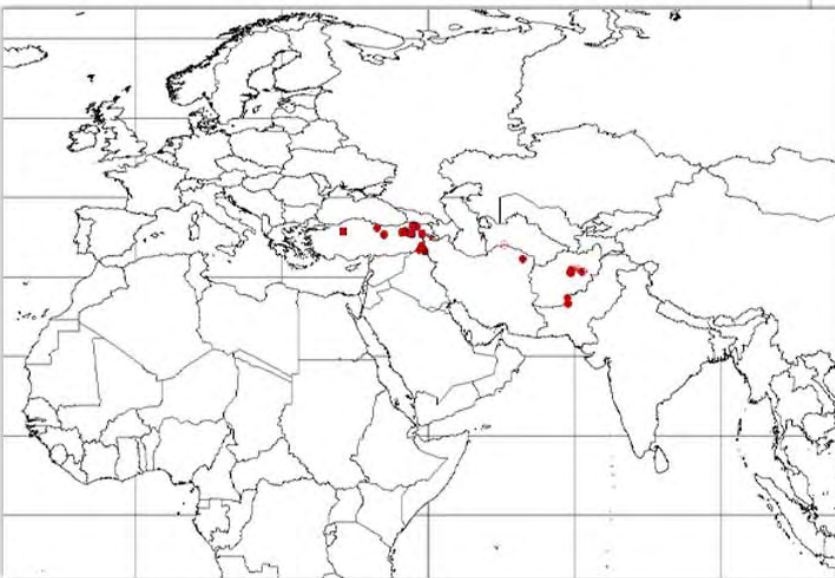
Distribution of new *Pm3* alleles

Pm3 alleles

- | | | | |
|---|-----------|----|-----------|
| 1 | Pm3 10963 | 10 | Pm3 42281 |
| 2 | Pm3 13636 | 11 | Pm3 42416 |
| 3 | Pm3 14442 | 12 | Pm3 42469 |
| 4 | Pm3 14475 | 13 | Pm3 42525 |
| 5 | Pm3 23728 | 14 | Pm3 42868 |
| 6 | Pm3 31594 | 15 | Pm3 42920 |
| 7 | Pm3 41606 | 16 | Pm3 9939 |
| 8 | Pm3 42255 | 17 | Pm3CS |
| 9 | Pm3 42277 | | |



5 Turkmenistan





Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the *Pm3* resistance locus

Navreet K. Bhullar^a, Kenneth Street^b, Michael Mackay^c, Nabila Yahiaoui^{a,1}, and Beat Keller^{a,2}

^aInstitute of Plant Biology, University of Zurich, 8008 Zurich, Switzerland; ^bInternational Center for Agricultural Research in the Dry Areas, Aleppo, Syria; and ^cBioversity International, 00057 Maccarese, Rome, Italy

Communicated by Jeffrey L. Bennetzen, University of Georgia, Athens, Georgia, April 16, 2009 (received for review December 27, 2008)

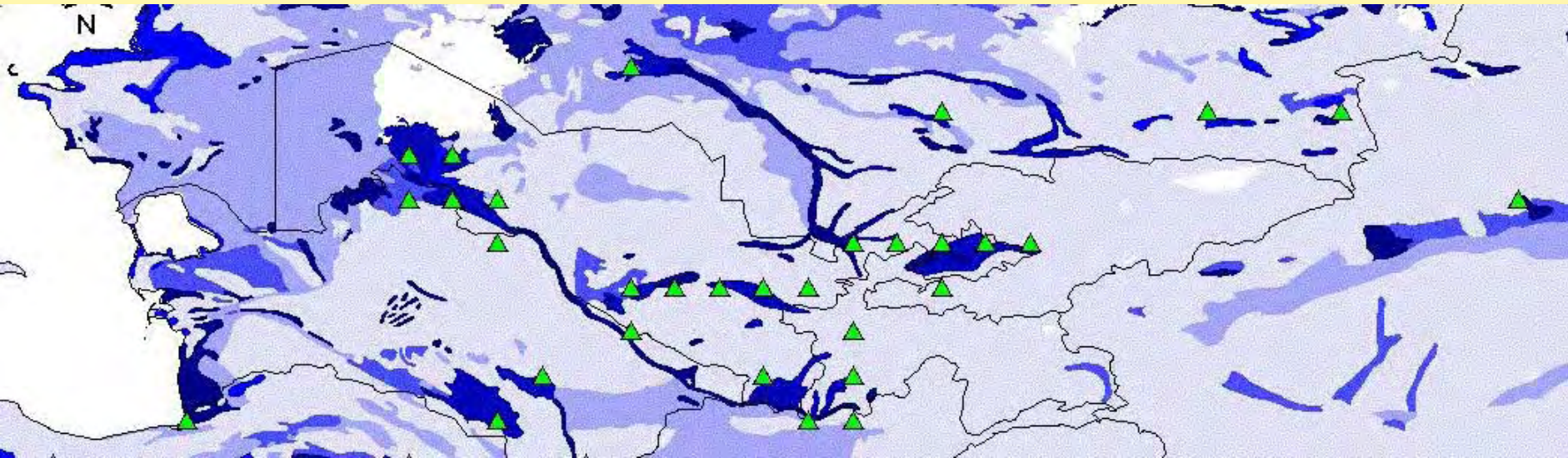
The continuous improvement of crop plants is essential for agriculture in the coming decades and relies on the use of genetic variability through breeding. However, domestication and modern breeding have reduced diversity in the crop germplasm. Global gene banks conserve diversity, but these resources remain underexplored owing to a lack of efficient strategies to isolate important alleles. Here we describe a large-scale allele-mining project at the molecular level. We first selected a set of 1,320 bread wheat landraces from a database of 16,089 accessions, using the focused identification of germplasm strategy. On the basis of a hierarchical selection procedure on this set, we then isolated 7 resistance alleles of the powdery mildew resistance gene *Pm3*, doubling the known functional allelic diversity at this locus. This targeted approach for molecular utilization of gene bank accessions reveals landraces as a rich resource of new functional alleles. This strategy can be implemented for other studies on the

Genetic variation is caused by allelic diversity at the genetic loci contributing to a particular trait. Allele mining is a relatively underexplored method to identify new alleles at a known locus. However, it is being used in important plant species, such as maize and barley (ref. 10; N. Stein, et al., personal communication). Because the first wheat disease-resistance genes have been cloned (11–16), the sequence information of these genes should allow the analysis of genetic diversity at these loci and the identification of new alleles through allele mining. *Pm3*, existing in 7 functionally distinct alleles (*Pm3a* to *Pm3g*), is the only wheat powdery mildew resistance gene cloned to date (13, 14, 16). In addition to the alleles from the bread wheat gene pool, a new functional allele recently has been described in a wild tetraploid wheat accession (17). All *Pm3* alleles encode coiled-coil (CC), nucleotide binding site (NBS), and leucine-rich repeat (LRR) proteins. The high sequence conservation

Screening for salinity tolerance in bread wheat

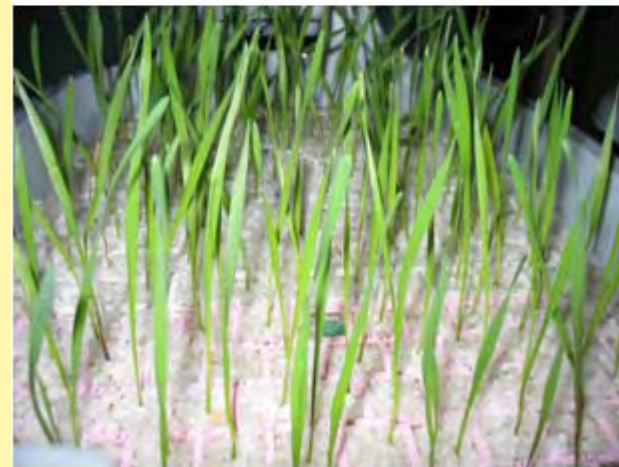


FIGS subset chosen by mapping collection sites over salinity probabilities



A core set was constructed, using published methods, to maximize both genetic variation (measured with agro-morphological data) and eco-typic origin (using 51 agro-climatic parameters)

Hydroponic system for salt-tolerance evaluation (VIR)



**Method after R.Munns et al, 2003.
Na and K content measured in 3rd
leaf after salinity treatment
imposed for 10 days**



57



13



Preliminary results

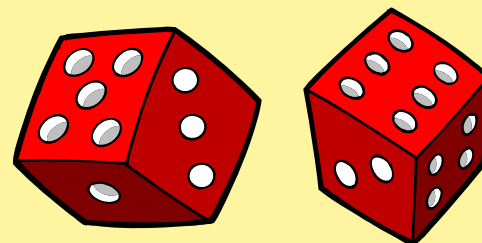
Set	Leaf Na content range in accessions mg/g dry weight (min – max)	Average leaf Na content mg/g dry weight	Number of accessions with Na content below 3 mg/g dry weight		Total accessions evaluated
			accessions	%	
salt	1.16 – 22.46	5.42	67	21	320
core	1.57 – 51.26	10.63	10	3	319

Table and screening by Irina Koserava – VIR, St Petersburg

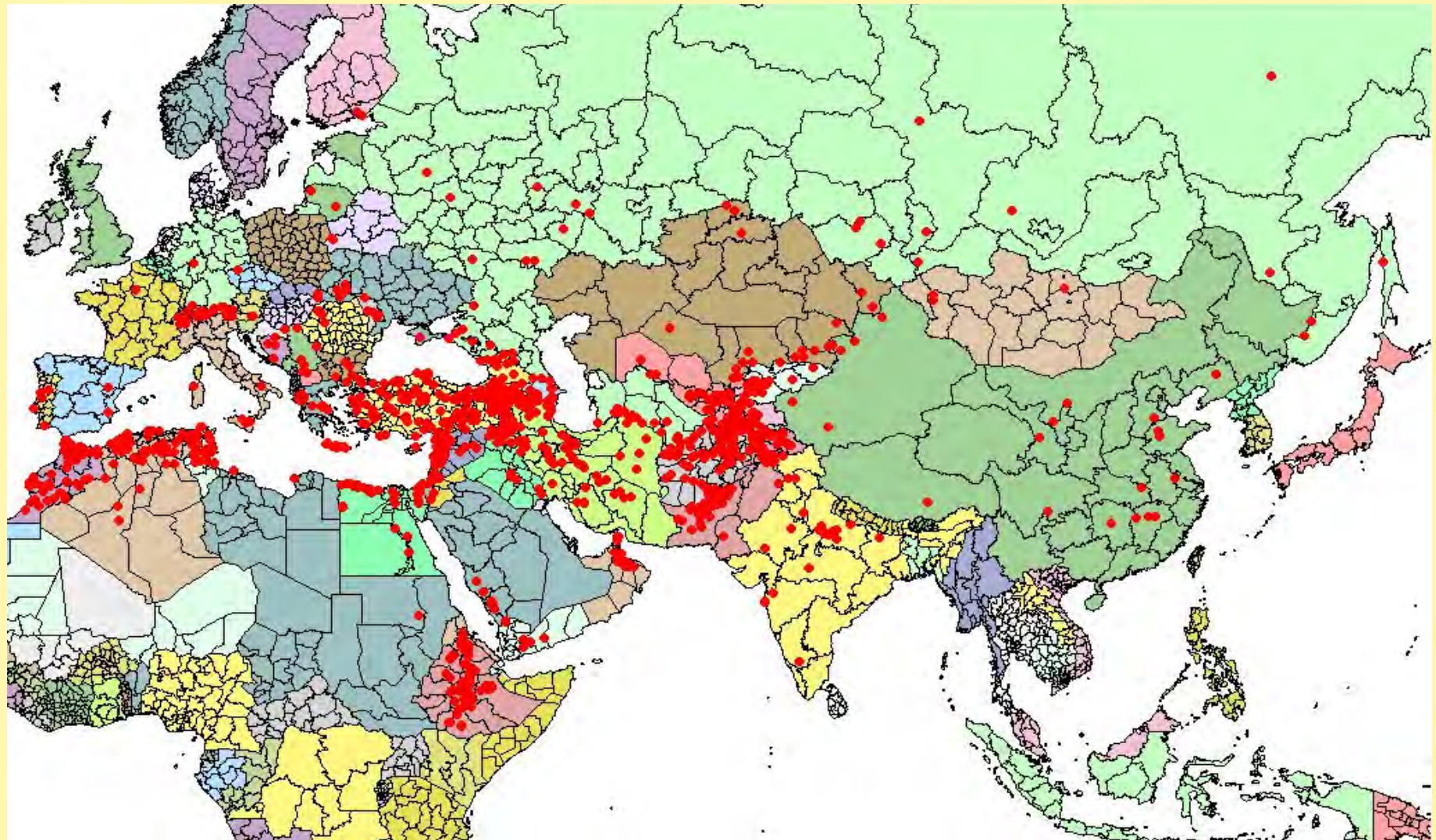
Proof of concept study



- In 2008, obtained data for 5400 bread wheat landraces screened in Yemen with a UG 99 isolate.
- The set was selected to give a wide eco-geographical spread.
- Strategy – screen as many as possible to increase the odds.



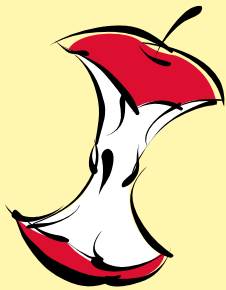
Distribution of Yemen test-set



Proof of concept study



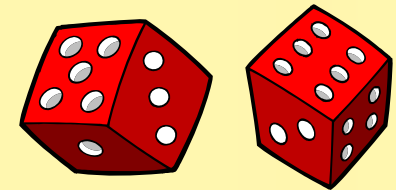
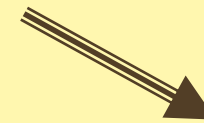
accessions screened in
Yemen
5400



Core set
540



FIGS sets
440- 540



Random set
540

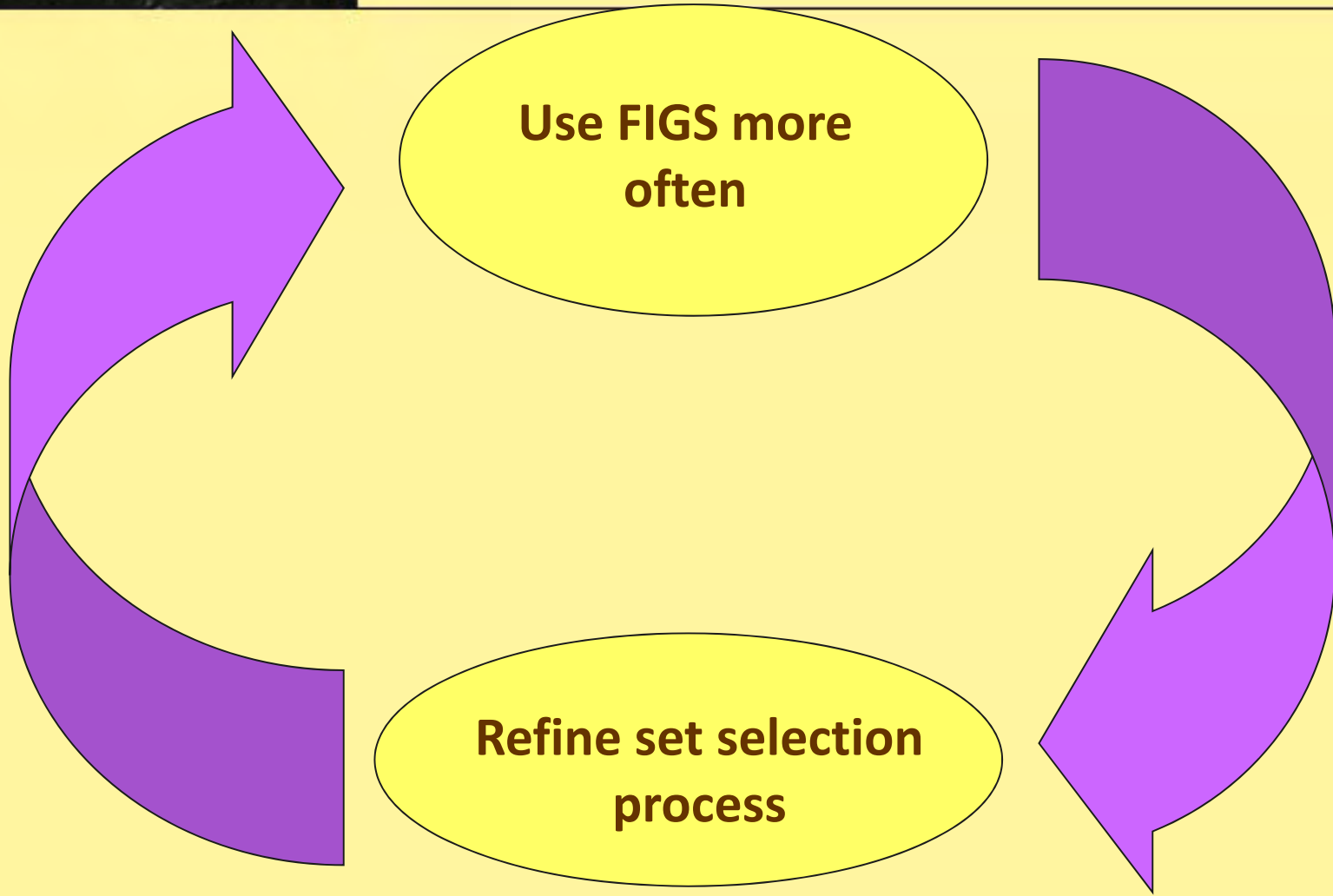
Note: sets constructed without looking at screening scores

Results



<u>Method</u>	<u>% resistant accessions within set</u>
Core	9.7%
Random	10.7%
FIGS-set 1 (PCA on total set)	13.8%
FIGS-set 2 (Stratified clustering)	17.8%

FIGS is heuristic



Acknowledgements



People and institutions

ICARDA:

Eddy DePauw, Mustapha Bohsini



USDA-ARS:

Harold Bockelman



VIR:

Irina Kosareva, Pjotr Strelchenko



University of Zürich:

Beat Keller



Funding

GRDC



AWCC / NSW DPI:

Greg Grimes





Thank you