



Data and its Management in Genebanks

*With emphasis on
characterization and
evaluation data*

RDA Workshop, 10 September 2009

A colorful illustration at the bottom of the slide depicting a woman's face on the left, surrounded by various agricultural products like corn, tomatoes, and grapes, and a bird on the right.

Improving lives through biodiversity research

Outline

- Scope is data linked to PGRFA accessions. Need to provide answers to the following questions:
 1. Why is PGRFA data important?
 2. How do we obtain it?
 3. How do we store and retrieve it?
 4. How do we use it?

What this is and is not about:

Is about:

- Practical situations in genebanks;
- Practical solutions;
- Need to interact with users;
- Need to rationally use PGRFA.

Is not about:

- Too much theory;
- Need to be at cutting edge;
- Working in isolation from PGRFA users;
- Watching accessions in genebanks.



Grounding exercise

I want to walk through a scenario that every genebank which provides a PGRFA service to users has to, or should, consider early in developing its 'mission statement' and 'vision'.

Because, without a clear focus on what you are trying to achieve, for whom, and by what general means, there is a real risk that genebanks can become more like museums than vibrant sources of new genetic variation.

The purpose for collecting, documenting and conserving is eventual utilization.

Of course knowledge is continually being gathered.

Original reasons for genebanks:

1. Collect & conserve PGRFA
2. Support for plant improvement.
3. Support for research.
4. Others????

No sense in PGR conservation unless there is a rational end use.

Scope

PGRFA - plant genetic resources for food & agriculture

Conservation: activities associated with storing and maintaining PGR.

Utilization: activities associated with identifying and incorporating novel alleles into new varieties for farmers.

Breeders: researchers/pre-breeders and plant breeders.

My scenario will mostly refer to bread wheat



PGRFA components



Marshall and Brown, 1981





Gene Discovery - reflection

How did we find new variation in the past?

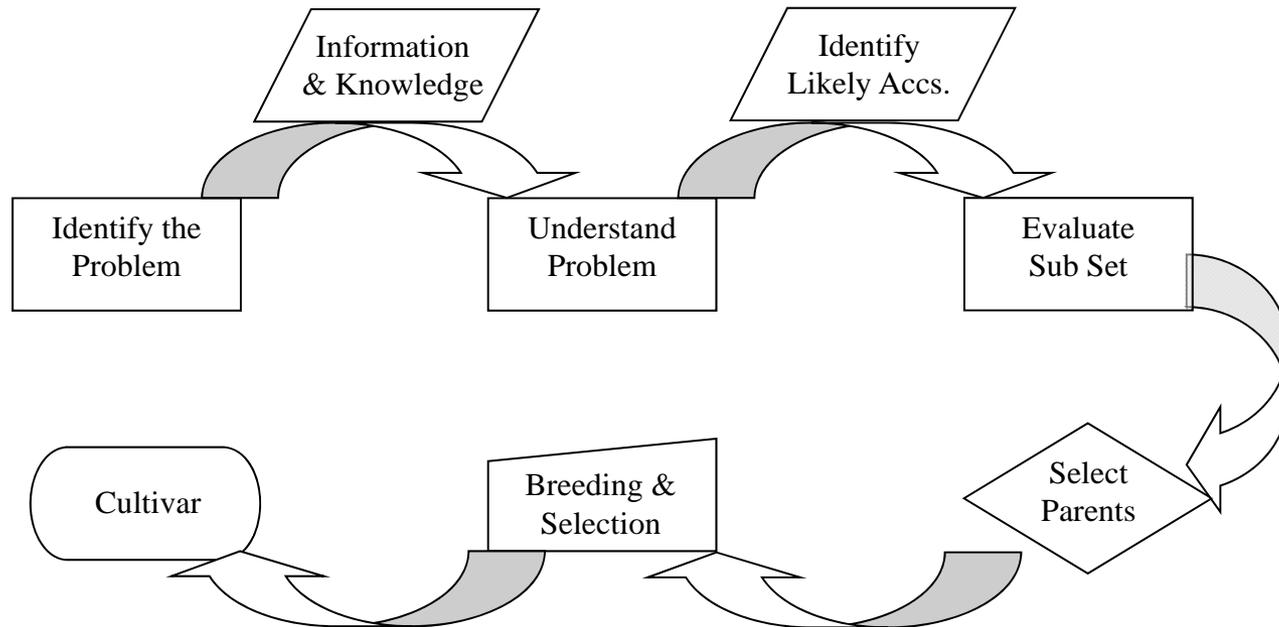
What methods have been used?

What is success rate?

Can we improve the efficiency with which we utilise
PGR?



Problems & solutions

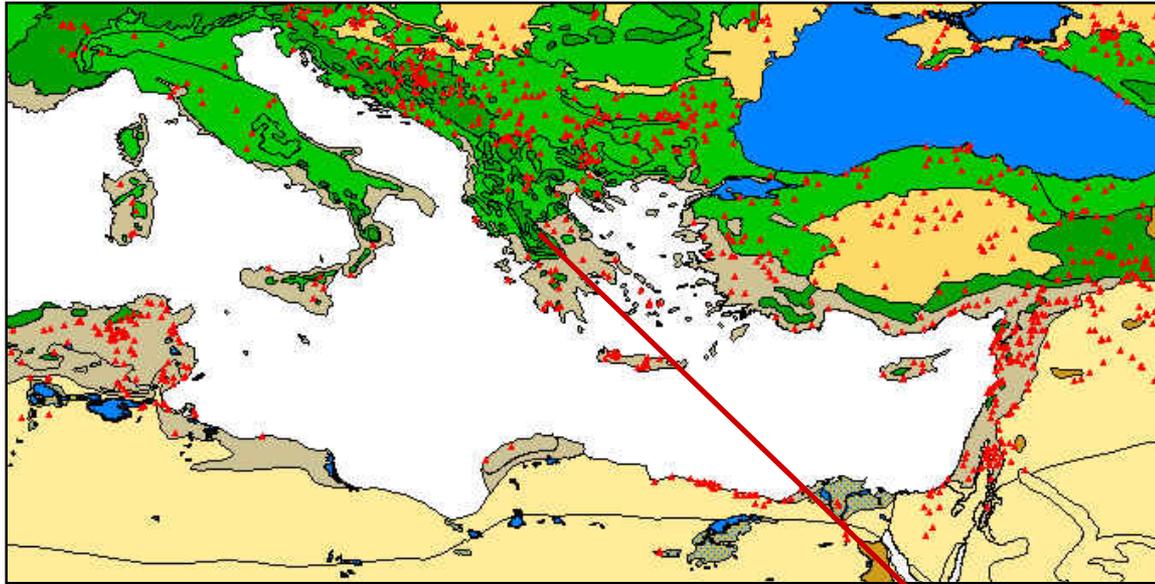


More recently

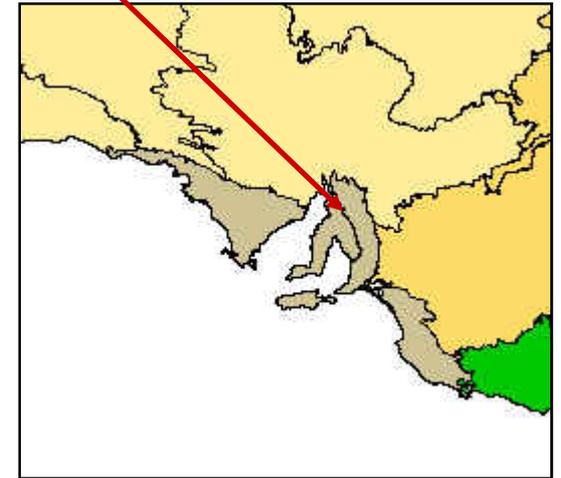
Late 20th Century examples:

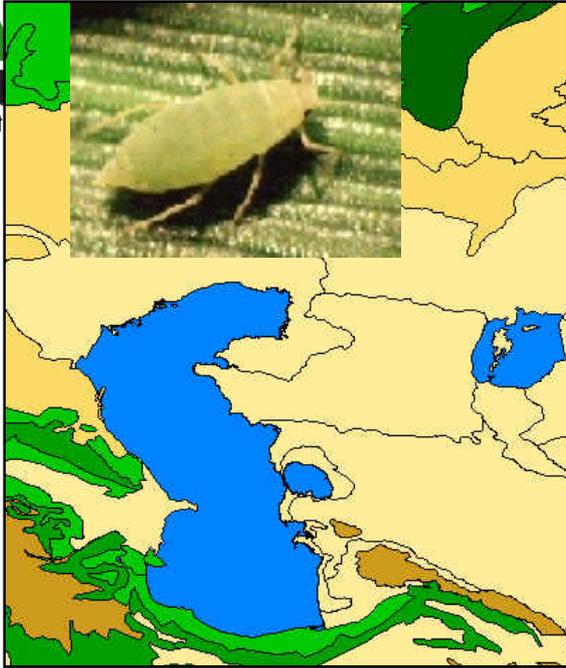
- CCN
- Boron tolerance
- RWA

A



Landraces from marine origin soils in Mediterranean region provided genetic variation for boron toxicity.





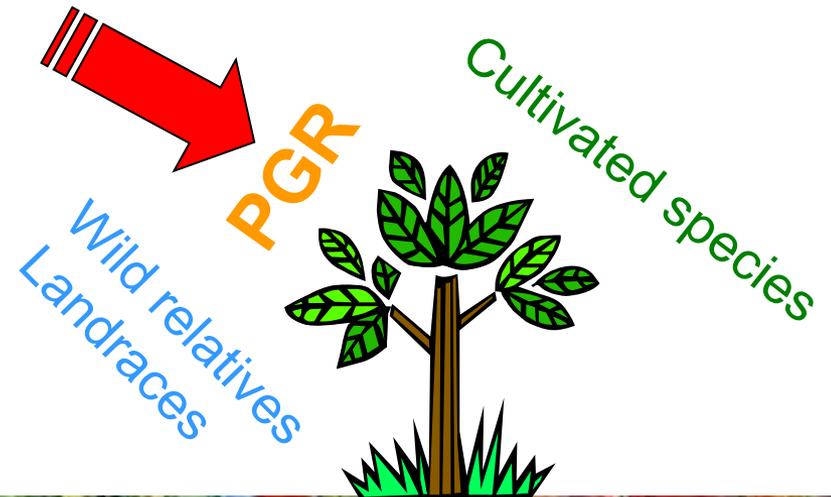
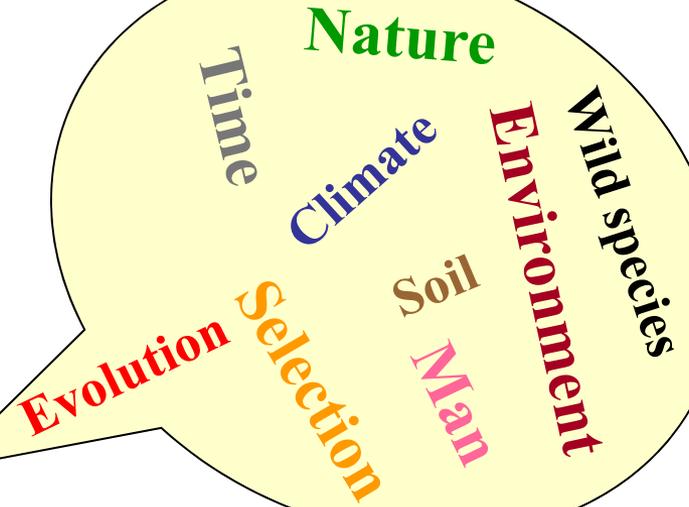
Russian wheat aphid is endemic in parts of southern Russia.

In early 1990s AWCC chose 42 landrace accessions from this region for testing in France.

One hexaploid *T. vavilovii* accession was found to have resistance, which has been backcrossed into Australian agronomic backgrounds.

Perspective

Ingredients



Accident and / or chance

Nature & man work together to provide variation:

First appeared around 6,500 years ago

Spread from Caucasus region

- migration
- trade
- war

Landraces “evolved” over millennia

Distributed across Eurasia

A

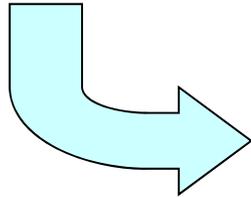
Attempts at design

Sir Otto Frankel proposed the *Core Collection* concept in 1984.

- Largely in response to criticism of minimal utilisation of large *ex situ* collections
- Intended to represent the general variation within a collection in a 5-10% subset

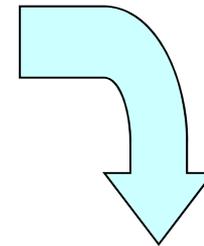
Utilization & core collections

Main Collection
25,000 accs



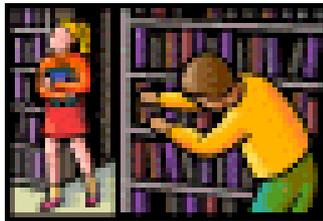
Method

$$t = \frac{\bar{w} - \omega}{\sigma_w / \sqrt{N}} = \frac{\bar{w} - \omega}{\sqrt{\frac{\sum_{i=1}^N (w_i - \bar{w})^2}{N(N-1)}}$$

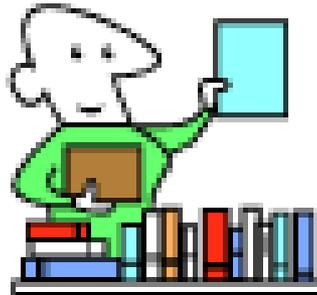


Core
Collection
5% - 10%

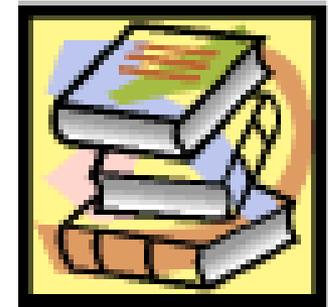
Library core collection analogy



Sampling



Evaluating



Selecting

**Emphasis on using core collection
for utilization - not other purposes**

**If no other information is available
then a core collection is fine to use**

A question

How effective or efficient would it be to use this method to find a book on genebank management in a general library?



The take home message

Message:

Recognise problem, gather knowledge & PGR based on data, then evaluate etc.

Knowledge comes from ?

Data types:

- Passport: identification of the accession
- Characterization: gross morphological traits not usually environmentally sensitive.
- Evaluation: Environmentally sensitive traits.
- Environmental: Climatic & edaphic.
- Genetic: Genotypic data, sequences, etc.
- Derived – from above.
- Expert opinion: Knowledge of experts without ‘hard supporting data’.

- Data adds value – usually the more data the greater the potential value of an accession.
- Without data the accessions in genebanks are virtually worthless.
- Useful data can come from many sources, as indicated in the previous slide.



First: Characterization

- What is characterization data?
 - Often gross morphological traits.
 - Give examples
- Can you give me an example of characterization data you are familiar with?
 - How do you describe it?
 - How do you measure it?
 - What do you use it for?
 - Can other people understand it?



Next: Evaluation

- How does it differ from characterization?
 - Environmentally sensitive.
 - Use categories in GRIN – morphology, phenology, pest & diseases (biotic), abiotic, etc.
- But similar in how to manage?

Next: Evaluation

What form does C & E data usually come in?

- Raw vs Summary data.
 - How often will raw data be necessary for utilization? Publication? Other reason?
 - Value in terms of utilization?
- Associated statistics.
 - Standard deviation
 - Mean & range for experiment
- Necessity for meta-data.
 - Methods – used to measure or collect data. E.g. for drought tolerance; how many methods can you think of?
 - Experiment – location (geo-reference), soil, scientist, institute, rainfall, others such as duration, planting date, fertilizer, irrigations etc.)

Next: Evaluation

Intellectual property: Something we all need to become aware of and consider in our management of data about PGRFA.

- Copyright. Deals with economic (e.g. publishing) and moral (e.g. attribution to author) rights. *Rights over look and feel.*
- Patent. More difficult to understand. Deals with situations where a person or company has exclusive rights to use a certain discovery or system. This can be sub-licensed.

I am no expert – but I intend to become very knowledgeable.

There will be more guidance on this area later in the course

Descriptors for wild and cultivated

Rice

(*Oryza* spp.)



IRRI



CHARACTERIZATION

7. Plant descriptors

Colour: To keep colour description simple, both Bioversity and IRRI colour codes for colour descriptors have been listed. The IRRI colour coding system uses one code for one colour regardless of the descriptor. To ensure consistent recording of colour states, the use of standard colour charts is recommended; use the IRRI colour code in conjunction with Annex II to look up the corresponding colour in the Methuen or RHS standard charts.

Source: Rice Descriptors

Quantitative characters: For all quantitative descriptors, it is recommended to use actual measurements. Coded scores for *Oryza sativa* are provided as an alternative where resources are insufficient to take actual measurements.

Taxonomic applicability: Descriptors listed here are applicable for wild and cultivated species, unless otherwise specified.

A table summarizing descriptor cross-references to other documentation systems (e.g. UPOV) and their recording stage can be found in Annex III.

Source: Rice Descriptors

GRIN-Global – Data Standards

Characterization and Evaluation data...

The collection of crop descriptors used by the USDA National Plant Germplasm System (NPGS) is created through a collaboration between a Crop Germplasm Committee (CGC) and the curator responsible for maintaining that crop within the NPGS.

Inputs:

- IPGRI/Bioversity Descriptors are used as the starting point for crop specific descriptors
- Needs of the researchers
- Historical data already collected on the crop is considered in order to minimize the possibility of losing data



GRIN-Global – Data Standards

Impact on GRIN Schema...

Constraints:

- GRIN Descriptors are crop specific (and GRIN handles all crops)
- GRIN Descriptors are evolving over time to meet changes in research

Result:

The GRIN system uses a row-wise storage methodology (database normalized) for characterization and evaluation data

- Very flexible (system can handle any descriptors)
- More work to load data (handled through programming)
- More work to retrieve data (handled through programming)

So, what's all this mean?

- A lot of work on genebank management activities has already been done;
- Is it all relevant to all the objectives of a genebank?
- Good management can only facilitate effective utilization.
- So, now we have some characterization data, what do we do with it????

Using the data

Schemas for managing data:

1. GRIN
2. Global Portal

www.legjournal.com

The ultimate objective:

Utilization

Focused Identification of Germplasm Strategy