Case study: Genebank mining with FIGS, the Focused Identification of Germplasm Strategy

Dr Ken Street

International Center for Agricultural Research in the Dry Areas (ICARDA)

Abstract

As the world moves forward into an uncertain future our agro-ecosystems will come under increasing pressure, threatening our food security. In fact, climate change, dwindling water supplies, rising energy costs, the emergence of new pests and diseases, loss of arable land and population growth mean that our crop plants will need to yield more on less land, with fewer inputs under increasingly harsh conditions. For this reason, plant breeders will be forced to mine global plant genetic resources collections for variation that can be used to future-proof our crop plants. However, the genetic resources collections are large and we cannot afford to evaluate every accession in a collection as we hunt for desirable traits. The Focused Identification of Germplasm Strategy (FIGS) was developed to help unlock the variation in genebanks and make it more accessible to the plant breeding community. This paper explains how FIGS works and gives examples of how rare traits have been uncovered by using the technology.

In the course of my work at ICARDA I would have spent something like 30,000 km bumping around in old Russian 4-wheel-drives on collection missions. It was often rough going, and I frequently wondered if the journey would be worthwhile! That is what led me to the scientific aspect of my work: namely, to find better ways to use plant genetic resources from our genebanks.

The context for this case study is the serious situation facing our agro-ecosystems. We are going to have to grow more food on less land, with fewer inputs under harsher conditions as we move forward into the future. Therefore, without question, we are going to have to redefine the capabilities of our crop plants. That means we shall need a great deal of innovative plant breeding, very soon, and until the 'fabled' gene editing technology becomes available we are going to have to rely on the genetic diversity within the accessible gene pools.

Luckily, because there have been many plant collection missions in the past, there is already a great deal of useful diversity within the genebank system. The genebank shown in Figure 1 is the ICARDA genebank. It is very diverse, and its diversity is backed up in facilities like the one at Svalbard.

Challenges of finding the traits you need

It is excellent that the genebanks have captured so much diversity, but there is also a problem with that. Imagine you are a plant breeder and you are

---

This paper has been prepared from a transcript and the Powerpoint slides of the presentation.

Figure 1. The genebank at ICARDA (main) and the Svalbard vault (top left).
(ICARDA = International Center for Agricultural Research in the Dry Areas)

breeding for a rare trait, a rare adaptation, such as frost resistance during the reproductive phase. You have only enough resources to screen a couple of hundred accessions: not thousands; just a couple of hundred. However, the gene pool you are working with contains something like 315,000 accessions. How can you choose the accessions that will be useful?

We need rational ways of digging into these collections of genetic resources so we can pull out the useful variation. And that is where the Focused Identification of Germplasm Strategy (FIGS) becomes important. That is what FIGS does.

FIGS is user-driven. The user's request for a particular trait comes to the genebank, and we then attempt to put together a small set comprising useful germplasm. The selection process is based on some very well established principles – Darwinian Natural Selection. In a nutshell, to find a particular adaptive trait, FIGS searches collections from the environments where there is likely to have been selection pressure for that trait. It is very straightforward, but genebanks did not use this approach until quite recently.

Linking traits to selection pressures

There are lots of examples of eco-geographic variation for adaptive traits. One I have found really interesting is that apparently human nose shape is connected to the environments from which our ancestors came: the narrower the nose the colder the environment. These connections are thought to apply to plants as well, but do they apply in practice?

To answer that, I will now explain how FIGS works, in general terms without very much detail. Figure 2 shows the initial very simple filtering stage. The staff member looks at the environmental conditions at the collection sites from which the material in the genebank has come, and applies a filter to these parameters and importance.

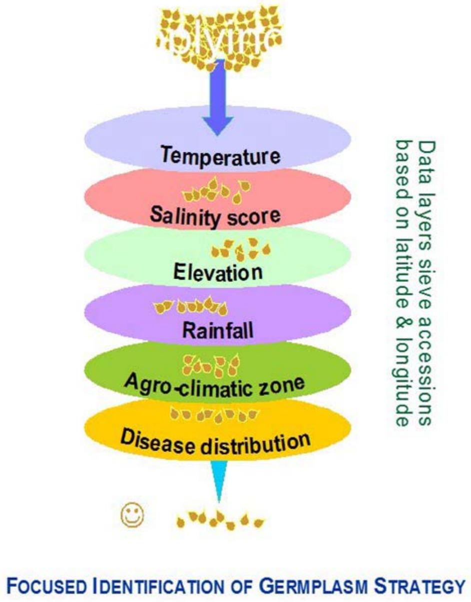


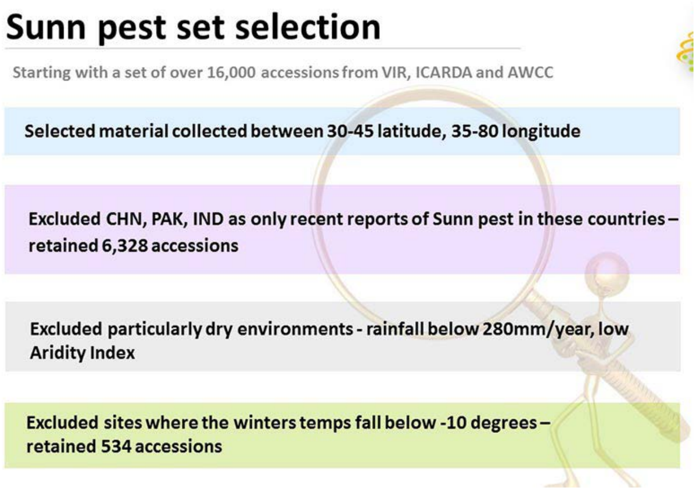
Figure 2. Diagram of the FIGS simple filtering method.

Figure 3. The Sunn Pest *Eurygaster integriceps* (10–12 mm long).

A good example, the Sunn Pest, shows how science does not have to be complicated to solve a problem; it can be quite simple. The Sunn Pest (Figure 3) devastates smallholders' plots right across the northern hemisphere. It is not a problem here in Australia ... yet. Dr Mustapha El-Bouhssini, of the ICARDA Entomology unit, had screened literally thousands of accessions of hexaploid wheat looking for resistance and found nothing. When he heard about FIGS he became really excited and asked us to put a set together.

The process we used to select a set is shown in Figure 4. First, we excluded any collection sites where there had been no record in the past of some pest being a problem; second, we excluded dry environments and very cold environments during winter time, the idea being that we wanted to select sites that would favour a high pest load.

When we presented Mustapha with the material we had assembled, he found ten sources of resistance in a small set size. It is a really important finding,



Sunn pest set selection

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

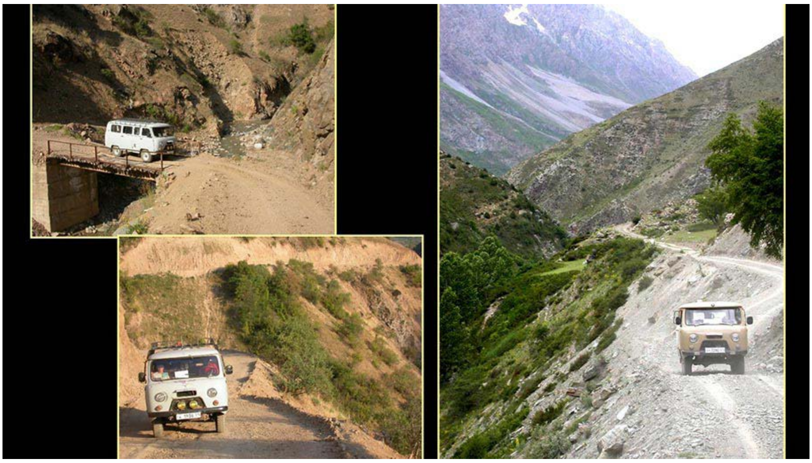
Selected material collected between 30-45 latitude, 35-80 longitude

Excluded CHN, PAK, IND as only recent reports of Sunn pest in these countries – retained 6,328 accessions

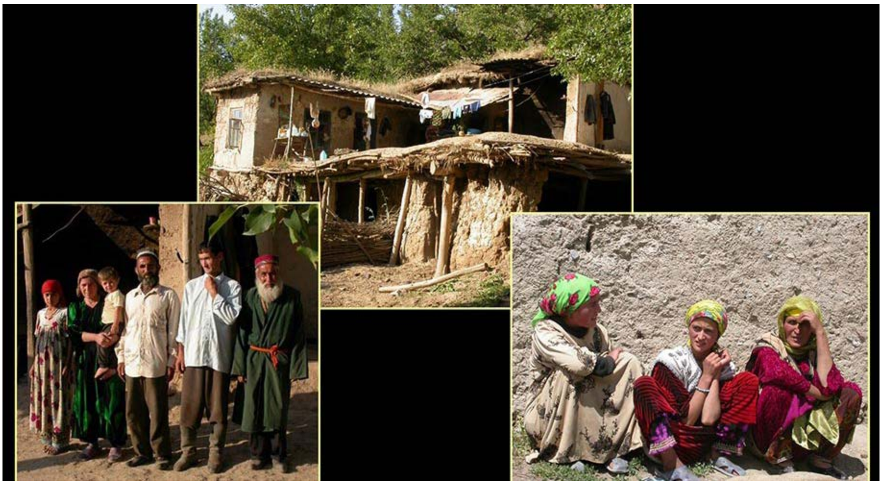
Excluded particularly dry environments - rainfall below 280mm/year, low Aridity Index

Excluded sites where the winters temps fall below -10 degrees – retained 534 accessions

Figure 4. Example to illustrate the FIGS selection process.



Above & below: Collection trips to Tajikistan (below) and Afghanistan finding 2 landrace accessions identified as resistant at juvenile stage and 8 landrace accessions (Afghanistan).



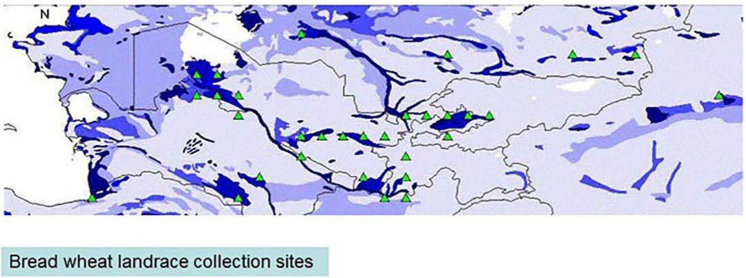


Figure 5. Map used in screening for salinity tolerance in bread-wheat landraces. The FIGS subset was chosen by mapping collection sites over salinity probabilities.

particularly for the people that ICARDA services, the smallholders. The trait is now in the breeding program and apparently varieties will be released shortly.

That example ties back to those collection projects in the old Russian 4WDs (photos previous page). Two of those sources of resistance were found in samples from a collection mission in Tajikistan in a little village that time had forgotten. These people were custodians of really important genetic variation – not just resistance to Sunn Pest but also resistance to the Syrian biotype of the Russian wheat aphid, a very virulent version, as well as to a suite of diseases.

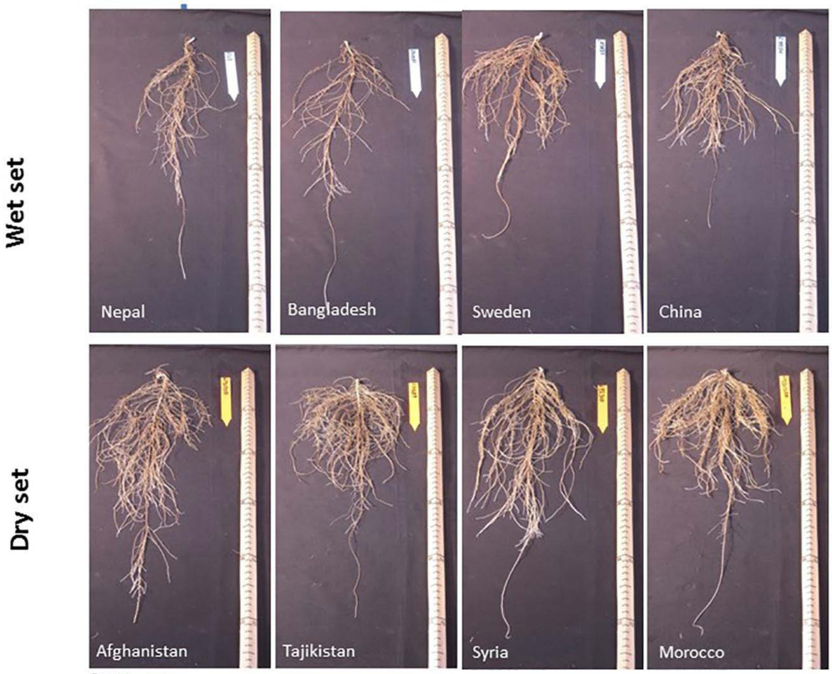
More examples: salinity tolerance and drought tolerance

The map above (Figure 5) shows the probability of encountering saline soils: the darker the colour the higher the probability. We mapped sites where wheat had been collected over this base map and on that basis we chose a subset of bread-wheat accessions to be tested for salinity tolerance by measuring plant uptake. When we measured the salt taken into the leaves of the subset after 10 days, compared to a core set of accessions, we found that 21% of the subset were resistant to salt uptake compared to only 3% of the core set.

In another example, with Faba Bean, we used the FIGS method to put together two sets, looking for major gene traits associated with drought tolerance. As shown in Figure 6, the root system is obviously adapted to dry environments, but those tests used long-term monthly averages to define the climatic parameters.

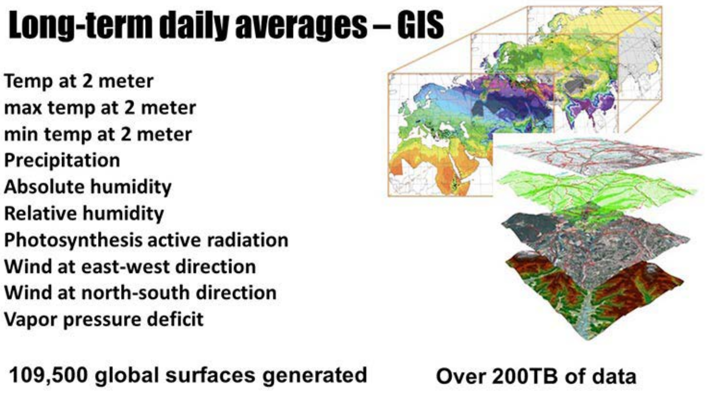
We wanted to explore the growing season a bit more deeply and refine the FIGS technology. So (using 'big data'), we constructed a series of GIS surfaces for a set of parameters (Figure 7) to construct long-term daily average surfaces. It involved generating over 100,000 surfaces, for the whole of the Northern Hemisphere, wherever our major crops developed – over 200 terabytes of data. From that we could construct, for every single collection site, or every single pixel on the map in Figure 8, a 'probable onset of growing period' for the crop in question. From that we could estimate the growth cycle of the crop, and then analyse the specific stages of crop development where the trait of origin is likely to be selected for. In this case, in the growing season, in the heading phase, we would apply our filter for frost tolerance, for example. That is what we are doing in a GRDC project at the moment.

As a result of using that slightly advanced FIGS methodology we have found useful sources of resistance for a whole suite of characters, which is very helpful to our growers of course.



© H Khazaei

Figure 6. FIGS selection for drought tolerance in Faba Bean.



Long-term daily averages – GIS

Temp at 2 meter
max temp at 2 meter
min temp at 2 meter
Precipitation
Absolute humidity
Relative humidity
Photosynthesis active radiation
Wind at east-west direction
Wind at north-south direction
Vapor pressure deficit

109,500 global surfaces generated

Over 200TB of data

Figure 7. Using 'big data' we generated long-term daily (rather than monthly) averages of climatic parameters

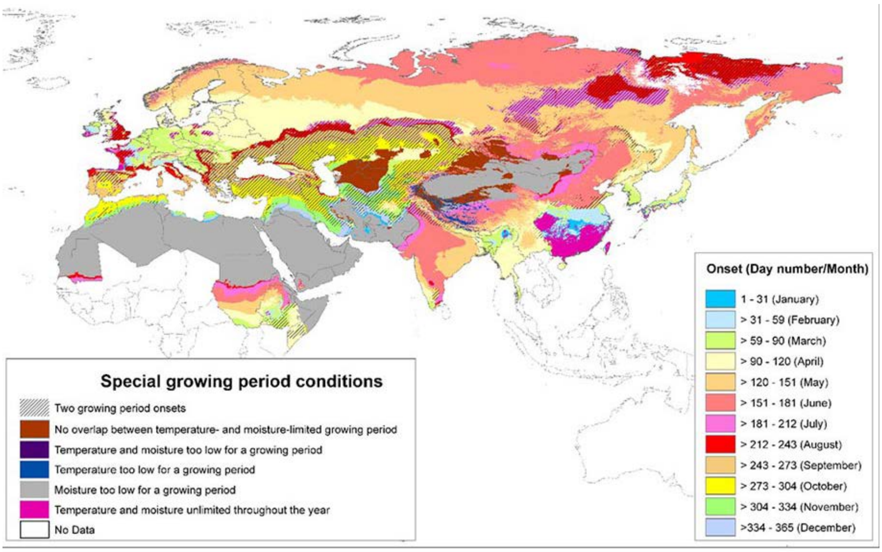


Figure 8. Map showing onset of growing periods limited by temperature and moisture.

More complicated FIGS

As well as using the simple filtering method in FIGS, we are also able to evaluate collection site data with multiple trait states, and we are developing statistical non-linear models to capture a relationship between the trait state and the environment of origin, using machine-learning. The outcome is a model that can predict, or discriminate between, the different trait states. Given genebank data and the collection site data, it will generate a map of the likelihood of finding the traits we want (Figure 9 for example). We have published this in various journals.

In FIGS in the future we hope to incorporate molecular data into the process. We have very large quantities of molecular data. Also, we want to improve the

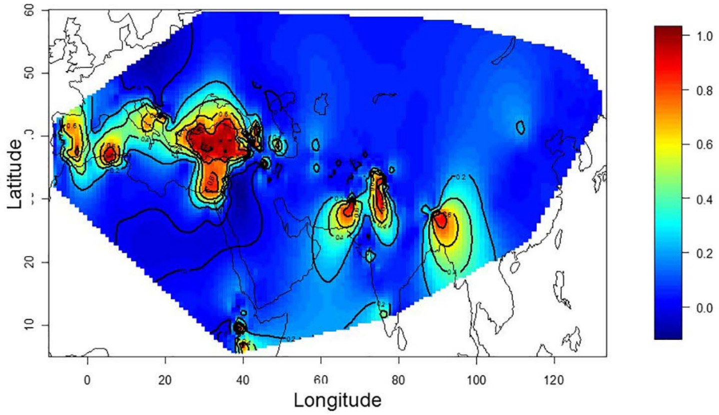


Figure 9. A typical output from FIGS predictivie modelling:
areas where resistance is likely to occur are shown in dark red.

Figure 10. Acknowledgements

accuracy of the onset data, and, most important, we want to develop a software application so anyone can do this process.

Acknowledgements

GRDC has backed its vision and given wonderful support over the years. ACIAR supported the collection missions. Michael Mackay (centre of Figure 10) came up with the concept of FIGS in the first place. This work has been done at ICARDA, and I also want to acknowledge the scientists currently working on FIGS.

Further reading

Bari A., Street K., Mackay M., Endresen D.T.F., De Pauw E. & Amri A. (2012). Focused Identification of Germplasm Strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. *Genetic Resources and Crop Evolution* **59**:1465–1481.

Bari A., Amri A., Street K., Mackay M., De Pauw E., Sanders R., Nazari K., Humeid B., Konopka J. & Alo F. (2014). Predicting resistance to stripe (yellow) rust in wheat genetic resources using focused identification of germplasm strategy. *Journal of Agricultural Science.* doi:10.1017/S0021859613000543.

Endresen D.T.F., Street K., Mackay M., Bari A. & De Pauw E. (2011). Predictive association between biotic stress traits and ecogeographic data for wheat and barley landraces. *Crop Science* **51**:2036–2055.

Endresen D.T.F., Street K., Mackay M., Bari A., De Pauw E., Nazari K. & Yahyaoui A. (2012). Sources of resistance to stem rust (Ug99) in bread wheat and durum wheat identified using Focused Identification of Germplasm Strategy (FIGS). *Crop Science* [Online first].

Khazaei H., Street K., Bari A., Mackay M. & Stoddard F.L. (2013). The FIGS (Focused Identification of Germplasm Strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0063107

Dr Ken Street is a genetic resources scientist who has spent most of his scientific career involved in the collection and conservation of cereal and food legume agro-biodiversity. Based at the ICARDA genebank, situated in Aleppo Syria, Dr Street was responsible for undertaking numerous plant collection missions throughout Central and West Asia and the trans-Caucasus. In addition to his plant collecting missions, Dr Street was also involved in research aimed at improving the efficiency with which we mine genetic resource collections for useful traits. The outcome of this collaborative work was the Focused Identification of Germplasm Strategy (FIGS) that has been successfully deployed to discover rare crop traits that will have a positive impact in farmers' fields.