# Surfing the Genepool

## The Effective and Efficient Use of Plant Genetic Resources

Michael Mackay
*Faculty of Landscape Planning, Horticulture and Agricultural Science*
*Department of Plant Breeding and Biotechnology*
*Alnarp*

Doctoral Thesis
Swedish University of Agricultural Sciences
Alnarp 2011

Acta Universitatis agriculturae Sueciae
2011:90

Cover: Landrace and crop wild relative diversity, Azerbaijan, 2001. (Photo: Michael Mackay)

ISSN 1652-6880
ISBN 978-91-576-7634-4

Print: SLU Service/Repro, Alnarp 2011

# Surfing the genepool: Effective and efficient use of plant genetic resouces

## Abstract

Humankind has been left a wealth of plant genetic diversity; a gift of nature, evolution and domestication over many millennia. Until recent times we have used these plant genetic resources (PGR) in harmony with our environment. The 20th Century brought many changes to the world including new frontiers in science and agriculture. Plant breeding, as we know it today, began in 1900 with the rediscovery of Mendel's laws of inheritance published in 1866, following his experiments on plant hybridization.

Early plant breeders were quick to take advantage of landraces originating from different regions and hybridized them to aggregate more agronomically important traits into widely adapted modern cultivars. Nickolai Ivanovich Vavilov (1887-1943) was among the first scientists to recognize the distribution of plant characteristics was not random, but related to ecological and other environmental parameters. By the mid 20th Century PGR were being widely collected and conserved in *ex situ* genebanks.

The use of PGR in breeding was largely based on information sourced from other researchers and genebanks. Breeders specifically asked for evaluation data to effectively use genebank accessions.

The Focused Identification of Germplasm Strategy (FIGS) emerged as an approach to target accessions more likely to possess specific genetic variation sought by breeders. FIGS involves gathering available information and knowledge to facilitate the identification of candidate accessions. GIS, statistical and modeling techniques can then be used to select the candidates for evaluation based on understanding the trait by environment relationships.

This study concludes that rational approaches, such as FIGS, deliver more effective and efficient utilization of PGR by identifying 'best bet' subsets of accessions to address contemporary plant breeding challenges requiring novel genetic variation for adaptive traits.

*Keywords:* plant genetic resources, landraces, genebanks, adaptive traits, genetic variation, utilization, trait by environment relationship

*Author's address:* Michael Mackay, SLU, Department of Plant Breeding and Biotechnology, P.O. Box 101, SE-230 53 Alnarp, Sweden.
*E-mail:* mackaymcj@gmail.com

# Dedication

This work is dedicated to future generations, represented here by three generations of my own family. We need to think of these future generations as we plan our response to global food security. Working together is the only option to ensure the world is fed sustainably.

*Three rules of work: Out of clutter find simplicity; From discord find harmony; In the middle of difficulty lies opportunity.*

Albert Einstein

# Contents

# List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Mackay, M.C. (1990). Strategic planning for effective evaluation of plant germplasm. In Srivastava, J.P. & Damania, A.B. (Eds.) *Wheat genetic resources: meeting diverse needs*. John Wiley & Sons, Chichester, 21-25.

II Mackay, M.C. (1995). One core collection or many? In Hodgkin, T., Brown, A.H.D., Van Hintum, T.J.L., Morales, E.A.V. (Eds.) *Core Collections of Plant Genetic Resources*. John Wiley & Sons Ltd., Chichester, 199-210.

III Mackay, M.C., von Bothmer, R. & Skovmand, B. (2005). Conservation and Utilization of Plant Genetic Resources – Future Directions. *Czech Journal of Genetics and Plant Breeding Journal* 41(Special Issue), 335-344.

IV Bhullar, N., Street, K., Mackay, M., Yahiaoui, N. & Keller, B. (2009). Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. *Proceedings of the National Academy of Sciences,* 106**,** 9519-9524.

V Bhullar, N., Mackay, M. & Keller, B. (2010) Genetic Diversity of the Pm3 Powdery Mildew Resistance Alleles in Wheat Gene Bank Accessions as Assessed by Molecular Markers. *Diversity,* 2**,** 768-786.

VI El Bouhssini, M., Street, K., Amri, A., Mackay, M., Ogbonnaya, F.C., Omran, A., Abdalla, O., Baum, M., Dabbous, A. & Rihawi, F. (2010)

Sources of resistance in bread wheat to Russian wheat aphid (Diuraphis noxia) in Syria identified using the Focused Identification of Germplasm Strategy (FIGS). *Plant Breeding*, 130**,** 96-97.

VII Endresen, D.T.F., Street, K., Mackay, M., Bari, A. & De Pauw, E. (2011). Predictive Association between Biotic Stress Traits and Eco-Geographic Data for Wheat and Barley Landraces. *Crop Science*, 51**,** 2036-2055.

VIII Bari, A., Street, K., Mackay, M., Endresen, D.T.F., De Pauw, E. & Amri, A. Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables (Submitted).

Papers I-VII are reproduced with the permission of the publishers.

The contribution of Michael Mackay to the papers included in this thesis was as follows:

I Solely responsible for the book chapter.

II Solely responsible for the book chapter.

III Researched and drafted manuscript and final paper collated with editorial input from co-authors.

IV Participated in the design and implementation of research, contributed to editing of the paper.

V Participated in the design and implementation of research, contributed to writing of the paper.

VI Contributed to research design.

VII Participated in design and contributed to writing of the paper.

VIII Participated in design and contributed to writing of the manuscript.

# Abbreviations

| | |
|---|---|
| AWCC | Australian Winter Cereals Collection, Tamworth, Australia |
| Bioversity | Bioversity International Rome, Italy. |
| C & E | Characterization and evaluation data |
| CGIAR | Consultative Group on International Agricultural Research |
| CIMMYT | International Maize and Wheat Improvement Centre, El Batán, Mexico |
| CWR | Crop wild relative |
| ECPGR | European Cooperative Programme for Plant Genetic Resources |
| EURISCO | European catalogue of *ex situ* genebank collections |
| FAO | Food and Agriculture Organization of the United Nations |
| FIGS | Focused identification of germplasm strategy |
| GCDT | Global Crop Diversity Trust |
| GENESYS | Accession-level information system for plant genetic resources |
| GIS | Geographic information system |
| GRDC | Grains Research and Development Corporation, Australia |
| GRIN | USDA-ARS NPGS Genetic Resources Information Network |
| GxE | Genotype by environment interaction |
| IARC | International agricultural research center |
| IBPGR | International Board for Plant Genetic Resources Rome, Italy. Renamed IPGRI in 1991 and Bioversity International in 2006. |
| ICARDA | International Centre for Agricultural Research in the Dry Areas, Aleppo, Syria |
| IPGRI | International Plant Genetic Resources Institute Rome, Italy (1991-2006). Bioversity International from 2006. |
| ITPGRFA | International treaty on plant genetic resources for food and agriculture |
| NPGS | USDA-ARS National Plant Germplasm System |
| PGR | Plant gene resources |
| SINGER | System-wide Information Network for Genetic Resources |
| TxE | Trait by environment relationship |
| USDA | United States Department of Agriculture |
| USDA-ARS | USDA Agricultural Research Service |
| VIR | N.I. Vavilov Research Institute of Plant Industry, St. Petersburg. |

# 1 Introduction

*We are now in a position where we must not only manage our crop plants, our domestic animals, our fisheries, our forests and range lands, but the whole globe is in our care, ready or not, competent or not. We are affecting the atmosphere, the oceans, the forests, rainforests, deserts, and even the climate. We are woefully unprepared for this awesome responsibility. This is an age of great knowledge and little wisdom, but we have no choice; we must blunder on. Who is in charge here? God helps us, we are!* – Source: The epilogue of Jack R Harlan's second edition of Crops & Man.

## 1.1 Mankind's early use of plants

Of the estimated period that mankind has lived on this earth some 90% of this time he existed as a hunter and gatherer. Since around 10,000 years ago, roughly 6% of the time man has existed on the planet, he began to domesticate plants and animals and thus live with agriculture (Lee and DeVore, 1968; as quoted in Harlan, 1992). It is only during the past hundred and fifty years, after naturalist Charles Darwin published his *"On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life"* in 1859 and Mendel's laws of inheritance were rediscovered in 1900 following his earlier reports of experiments with hybridizing peas, that mankind has been able to move into his present manner of existence with modern agriculture and industrialized societies.

## 1.2 Natural diversity in plants and their domestication

Given the billions of years that plants, animals and other organisms evolved under the prevailing conditions of natural selection, before mankind began to play a role in the selection process, the resulting amount of natural genetic variation is obviously vast. In the case of flowering plants, whose ancestors diverged from gymnosperms some 200 million years ago, only about 300 species, of the estimated more than 250,000 species of flowering plants, have been widely used in agriculture leaving an enormous quantity of unknown and untapped diversity yet to be identified and understood.

There are likely to be many more species used by man for other purposes, including food and medicine, but these will have been subject to very little, if any, selection by mankind. For example, many asexually propagated plants are used as a source of food, but these are virtually identical to their earlier ancestors and therefore unaffected by modern agricultural practices.

Even today, with all the scientific and technological advances made over the past century, we have probably only touched the ‘tip of the iceberg’ when it comes to fully understanding and exploiting the genetic diversity that exists right under our nose in the landraces and wild relatives of the handful of species we rely on so heavily as our staple food crops.

There are several phases of domestication (after Kingsbury, 2009):

### 1.2.1 ‘Protoculture’ phase

This phase is where mankind first began to manage nature by means of, for example, fire; the interdependence between mankind and nature first became apparent.

### 1.2.2 ‘Agrotechnical’ phase

During 11,000 BCE [1]– 8,300 BCE there was an ‘agrotechnical’ phase evidenced through the discovery of sickle blades, mortars, pestles, storage pits and some permanent settlements in the Fertile Crescent - the earliest region of domestication. In a seed based system this suggests the early harvesting of plants for seeds and their subsequent processing for food and storage for later use or replanting. In the case of non-seed based systems, so called ‘vegiculture’, an example is the Australian Aboriginal who collected many yams, but always left some tubers to ensure a food supply for the next season.

---

[1]. BCE means “Before the Common Era” and equates to BC.

### 1.2.3 ‘Domestication revolution’ phase

A ‘domestication revolution’ phase (8,300 BCE – 5,500 BCE) where cultivation of wild forms of grasses and their gradual domestication is more established. Examples include emmer wheat (*Triticum dicoccon*) and barley (*Hordeum vulgare*) in southern parts of the Fertile Crescent whilst einkorn wheat (*Triticum monococcum* var. *boeticum*), barley, legumes and flax (*Linum usitatissimum*) were examples in the northern parts. Morrell and Clegg (2007) surmise a second barley domestication event 1,500 to 3,000 km east of the Fertile Crescent based on haplotype frequency differences at a number of loci between the geographic regions.

The domestication act of ‘vegiculture’ was the transplanting of selected tubers or, in the case of *Musa*, suckers.

## 1.3 Beginnings of plant breeding

The actual act of collecting some seeds and storing them for planting in the next growing season is an unconscious act of plant breeding because it involves sexual reproduction, genetic recombination and subsequent segregation. At some stage the criterion by which a particular head of grain is chosen in preference to another, for seeding the following year, is a human decision that guides the future genetic composition and thus traits, of the cultivated species.

In the case of seed based domestication processes the ‘selection decisions’ are often reported as being based on preference for traits such as non-shattering, more determinate growth, more seed production, higher seedling vigor, more rapid and even germination, and loss of seed dispersal mechanisms. Crops that have been cultivated over longer timeframes are usually the most different to their ancestral forms. There are, however, always exceptions and Kingsbury (Ibid.) suggests that asexually propagated species such as the pomegranate (*Punica granatum*), fig (*Ficus carica*), olive (*Olea europaea*) and grape (*Vitis vinifera*) have been cultivated for many millennia but do not reflect the same extent of genetic recombination as their sexually propagated counterparts.

## 1.4 Modern plant breeding

From the start of the 20th century, utilization of plant genetic resources (PGR) became a more deliberate process whereby early breeders began to collect and hybridize various cultivated plant species to further enhance productivity.

Landraces, or local varieties, underpinned many of the early cultivars because they had been subject to both natural and passive human assisted selection over long time periods; many were still selections from heterogeneous local varieties. By the 1920s, N. I. Vavilov had recognized and illustrated the concept of ‘centres of diversity’ that established the association between diversity and eco-geographic distribution (Vavilov, 1926). Vavilov and his colleagues subsequently collected a large range of useful landraces and species from diverse environments all over the world that fed into the early Soviet breeding programs (Vavilov, 1992). In his ecographical surveys of cereals and other crops, Vavilov later demonstrated linkages between adaptive traits and environmental conditions as well as the need to start with the right material in the hybridization and selection process (Vavilov, 1957).

Breeders and geneticists largely exploited diversity in their working collections. Modern day breeders’ working collections contain a wide range of material from locally adapted cultivars (usually originating from landraces with additional desirable alleles progressively introgressed), various sources of new variation addressing the current breeding objectives and the resultant breeding lines subsequently developed through hybridization and selection. The sources of material in these collections include local cultivars, accessions from genebanks, material shared by other breeding programs or otherwise obtained during, for example, scientific meetings or their own collecting missions. In addition to landraces, crop wild relatives (CWR) have provided sources of, for example, stem rust resistance in bread wheat where genes from *Thinopyrum elongatum*, *Triticum timopheevi* and *T. dicoccum* were successfully introgressed (Knott, 1961; Macindoe and Walkden Brown 1968). However, as Frankel and Bennett (1970) pointed out, these early collections were managed in a casual manner that did not harness the full potential of the available diversity. By the 1960s it was recognized that landraces and CWR, along with their diverse genetic variation, were being replaced by new cultivars as more recently demonstrated through genealogical analysis (Martynov and Dobrotvorskaya, 2006).

## 1.5 Recognition of value of PGR and growing significance of *ex situ* genebanks – 1960s-1980s

The 1967 Technical Conference on the Exploitation, Utilization and Conservation of Plant Genetic Resources conducted by the Food and Agriculture Organization of the United Nations (FAO) (Bennett, 1968) called for more rigor in PGR conservation and utilization. The resulting activities over the subsequent 25 years saw collecting missions mounted, genebanks established in many countries or regions and studies into conservation, distribution, species relationships, gene content and genetic diversity undertaken. By 2010 there were an estimated 7.4 million accessions held in some 1,750 genebanks around the globe (FAO, 2010). However, as the financial commitments to PGR collections mounted, questions were raised as to whether these investments were realizing significant returns.

Earlier, Vavilov (1957) recognized the need to start with the right material for the breeding process. Frankel (1977) proposed that effective utilization depends on characterization, evaluation and documentation of all accessions in a collection. This approach was later questioned (Marshall and Brown, 1981) and altered in favor of breeders evaluating material in their target environments (Frankel, 1984). In implementing this approach, decisions about how many and which accessions from a collection should be screened for a given breeding objective have to be made. To put this into perspective, screening all the available wheat accessions today for a particular trait would necessitate dealing with well over 856,000 accessions worldwide (FAO, 2010); even if one was to eliminate 40% as possible duplicates, there is still far too much material to evaluate. For example, if it cost USD$100 to evaluate each accession for a particular trait, the cost of screening 400,000 amounts to USD$40 million. Obviously this is beyond the scope of any individual breeding programme and could only be addressed by a coordinated international effort. This hypothetical example clearly demonstrates the challenge faced – how can one identify those accessions held in genebanks that are more likely to contain novel genetic variation for the specific adaptive trait the breeders needs to address?

## 1.6 Value of information about PGR recognized

The availability of information about characteristics of PGR stored in *ex situ* genebanks has been accepted as a pre-requisite for their utilization (Frankel, 1977; Khoury *et al.*, 2010). But what is the nature of this information, where does it come from and how is it used? There are a number of categories of data associated with the documentation of PGR in genebanks including passport, characterization, evaluation, environmental and genetic/molecular data. These categories, together with data standards, will be discussed in the context of their significance for utilization.

### 1.6.1 Passport data

This is the information used to identify, for example, a genebank accession. It describes when and where the original sample was collected, its unique identifiers (such as variety name, breeders' designation, and collector's identifier), its ancestry or pedigree, if it is a modern variety, and other associated information. The FAO/IPGRI Multi-crop passport descriptors (Alercia *et al.*, 2001) provide a comprehensive list of the passport data standards which the majority of genebanks attempt to use in documenting their germplasm.

Passport data is usually accumulated by plant collectors, genebanks and breeders. Attention to detail and long term documentation is more critical to the genebank community than to the breeding community because plant varieties and associated breeding objectives change more quickly than do the goals of the genebank. It is worth noting that there is a tendency amongst researchers (including breeders) to abbreviate germplasm sample names which can later result in their real identity becoming questionable. An example of this is a sample provided by the Australian Winter Cereals Collection (AWCC) to a barley breeder in the 1980s labeled with several unique identifiers including a quarantine code '2TR86'; several years later the breeder enquired as to the origin of 'TR2' as he was registering a selection made from it for commercial release. Accurate information is very important because the linkages observed between ecogeographic regions and plant characteristics, passport data about the origin of landraces, for instance, can provide an insight into the possibility of selecting germplasm, based on its geographic origin, to find further expression of particular traits (Vavilov, 1957).

### 1.6.2 Characterization data

This includes those descriptors that allow rapid morphological distinction between genotypes. They are often highly heritable, easily visible and expressed the same in all environments. Examples include seed characteristics such as color, presence or absence of pubescence on leaves and petal color.

This type of information was routinely collected by botanists and some schools of plant breeders to further categorize genotypes into sub-groups including 'botanical variety'. In the modern era genebanks are probably the most active accumulators of morphological characterization data. Collecting characterization data when multiplying or regenerating genebank accessions is a useful quality control procedure; if the data changes between one regeneration cycle and the next, genebank staff are immediately alerted to a potential mistake in labeling, cross contamination or some other aspect of their sample handling activities.

The value of characterization data is, in itself, not huge. However, plant breeders often will want to limit the scope of germplasm they obtain from genebanks based on characterization data. One example could be that a wheat breeder in a certain region, where only hard white grained kernels are produced, might prefer not to include accessions with soft or brown kernels when broadening the genetic base of the breeding program.

### 1.6.3 Evaluation data

Descriptors for evaluation are often traits that are expressed differently in different environments. They often relate to characteristics like productivity, reaction to biotic and abiotic stresses, agronomic attributes and time to maturity. It is this category of data that has been considered essential for utilization, indeed without such data new genetic variation is even more difficult to discover. Evaluation data and utilization have a ‘chicken and egg’ relationship (which came first, the chicken or the egg?) in that utilization cannot occur without evaluation, but evaluation isn’t usually undertaken without the need to identify novel genetic variation for utilization.

Evaluation data is collected by numerous groups including genebanks, breeders and basic researchers. If an evaluation program is initiated and identifies useful genetic variation the germplasm containing the variation is used as a parent, thereby initiating another cycle of the breeding process.

Unfortunately evaluation data is rarely returned to the genebank from which accessions were originally obtained unless it was part of a collaborative arrangement with researchers or breeders. Most online information systems do not include much evaluation data and a real challenge of the future will be to capture such data and make it accessible for use in meeting subsequent plant improvement challenges.

### 1.6.4 Environmental data

It has only been in recent times that environmental (climatic and edaphic) data could be linked to accessions because of the development of geographic information systems (GIS). Prior to GIS, any environmental data had to be manually collected at the time of sampling and meticulously recorded and stored. GIS can not only provide estimates of environmental parameters at multiple points in time, but also allows the rapid harvesting of estimates for many such parameters to associate them with any geo-reference linked to a germplasm sample.

### 1.6.5 Genetic and molecular data

This category of data has become one of the most promising to facilitate the utilization of PGR. Tanksley and McCouch (1997) saw two strategies for evaluating and using germplasm. The ‘old paradigm’ of looking for the phenotype where one would “screen entries from a gene bank for a clearly defined character and the ‘new paradigm’ of looking for genes using “genetic linkage maps based on molecular markers” and the use of QTLs (quantitative trait loci).

In the molecular age of plant breeding huge quantities of information are being generated as the tools for probing the genome continue to evolve. It offers increasingly more affordable options for fingerprinting accessions and rapid turnaround with techniques such as marker assisted selection (MAS). This category of data is generated in molecular biology laboratories and stored in databases designed to facilitate rapid access as well as providing functionality for maintenance and updating of information through innovative interfaces. DNA sequencing is now commonplace and the need for programs to query sequence libraries resulted in the development of fast algorithms such as BLAST® and FASTA.

### 1.6.6 Data standards

Data must adhere to standards if it is to be universally interpreted and used. Consider language as an analogy; if one was to bring one thousand people together to discuss their different cultures and each of these people spoke a different language, imagine how little knowledge would actually be shared.

The same is true for information about PGR. The International Board for Plant Genetic Resources (IBPGR) and its successive institutions (International Plant Genetic Resources Institute and Bioversity International) have produced well over 100 publications of crop descriptors, or standards (http://www.bioversityinternational.org, verified 05/08/2011). Other groups, such as Biodiversity Information Standards provide further standards for describing biological data (http://www.tdwg.org/standards/, verified 05/08/2011). Even in 2011 there are many challenges faced in sharing data that can be attributed to a lack of standardization. One example is found in the *ex situ* plant collections maintained in Europe that publish their passport information in EURISCO (http://eurisco.ecpgr.org/home_page/home.php) that lists the following data in the 'Genus' passport field for the man-made allotetraploid triticale (*X triticosecale*): '?Triticosecale', 'Trirticosecale', 'Triticale', 'Triticisecale', 'Ttiticosecale', 'X Triticosecale', 'xTriticale' and 'Xtriticosecale'. While this is an exaggerated example, the point is that data is far more useful when standards, such as the FAO/IPGRI Multi-Crop Passport Descriptors (Alercia *et al.*, 2001), or MCPD, are used when sharing information about genebank accessions.

# 2 Scientific expectations of PGR use

> *All my life I have tried to pluck a thistle and plant a flower wherever the flower would grow in thought and mind* - Abraham Lincoln.

## 2.1 Plant improvement users.

Agriculture around the globe is interdependent and collections of diversity are fundamental to future productivity (Qualset and Shands, 2005). However, despite their obvious importance, genebank activities such as collection, conservation, documentation and characterization continue to struggle for support. Those who use PGR should be significant players in changing this situation. Breeders, as defined in section 2.2, are the primary users of PGR. The advent of plant breeding as a science over a century ago was based on genetic variation and the selection of agronomic traits to improve the productivity of the world’s crops and food supply system.

There are also more academic and educational pursuits involving the use of PGR. These might include simply trying to understand the nature and distribution of genetic variation, gene flow, genetic expression and interspecific hybridization, along with other scientific endeavors.

## 2.2 Expectations for breeding purposes

The more traditional plant breeding programmes of the 20th century involved scientists from a range of disciplines – genetics, physiology, pathology, chemistry and so on. Typically, the breeding program had a target region for which it attempted to develop higher yielding cultivars with broad agronomic adaptability and a suite of other characteristics required by farmers to maximize the return on their investment. It was often the allied discipline scientists who foraged through the more diverse germplasm to identify new genetic variation for inclusion in a breeding programme. They not only identified useful new variation, but also began introgressing it into the more adapted material within the breeding programme before handing it over to the plant breeder. These were what we today refer to as pre-breeders. To avoid confusion plant breeders, pre-breeders, and other scientists involved in plant improvement research, will hereafter be referred to simply as ‘breeders’ unless there is a need to distinguish between them.

During the past half century plant breeding has progressively become more commercialized in most developed countries, such as the USA (Qualset and Shands, 2005), many European nations and Australia. This has also taken place at a lower rate in some developing countries that have heavily invested in plant breeding and allied sciences. Thailand is one such country where significant investment in rice and maize from multinational companies has occurred recently (FAO, 2011a). Another example is Uruguay where many local companies have been acquired by multinational seed, chemical, or biotechnology companies (FAO, 2011b). For the most part the largest investments have been in the ‘big’ crops (including maize, rice, soybean, canola and wheat) and other remaining, so called, minor crops have struggled to compete under the prevailing market forces. One consequence of this shift is that the genetic base of these big crops has not broadened to any significant extent unless there has been a clear and pressing need to introgress new alleles from non-adapted germplasm to protect, for example, the crop from a newly emerged virulent disease. Furthermore, the general diversity of these crops across large areas has been significantly reduced thus predisposing them to future calamities such as the Irish potato famine of the 18th century and, about 40 years ago, the epidemic of Southern corn leaf blight in North America that resulted in huge economic losses to farmers and associated industries (Mann, 1997).

Today’s breeders only tend to approach genebanks for new variation in reaction to new challenges facing their cultivars in the target production region. A recent example is the reaction to the emergence of the Ug99 form of stem rust in wheat, first identified in Uganda in 1999. Several approaches to identifying accessions with possible resistance to Ug99 will be discussed in Chapter 6.

## 2.3 General research expectations.

A research interest in PGR will always be present because scientists continually seek to identify previously undescribed natural phenomena involving genetic variation in plants. Take, for example, the recent description of a rat using plant derived poison as a defense against predators (Kingdon *et al.*, 2011). The African crested rat acquires the poison from the bark of the *Acokanthera*, or 'poison-arrow' tree, and stores it in specialized, wick-like hairs that are exposed to predators when attacked; the rat itself seems immune to the poison. There will certainly be new compounds and adaptive genetic variation found in landraces and wild relatives in the future, and PGR will be the primary source of these future discoveries. Research is also a basis for new knowledge that can be useful for breeding purposes.

Exotic plant germplasm always attracts interest from scientists of many disciplines. It is recognized that there is an enormous wealth of untapped genetic variation held in *ex situ* collections (Qualset and Shands, 2005) to interest, for example, plant physiologists, pathologists and molecular biologists.

Scientists variously try to understand and describe the distribution of genetic diversity, identify locations rich in diversity but currently unrepresented in genebanks for future collection, and describe the relationship between diversity and the ecosystems in which it resides. Existing information in genebanks enables such studies to be undertaken and future research and collecting missions to be planned.

## 2.4 Strategies for sampling germplasm

The need to collect, conserve, document and use PGR is universally accepted with limited ongoing research, especially to support the first three of these activities. Research into practical sampling procedures for use of PGR, however, seems to have been largely overlooked.

Vavilov (1957, 1992) recognized the relationship between numerous traits and the eco-geographic origins of genetic resources; subsequent Soviet plant breeding exploited this knowledge to some extent. However, other than systematically evaluating whole collections or chancing upon accessions with desirable alleles, the only other approach to more effectively utilize PGR in the 20th Century was the core collection concept, which was developed for the different purpose of concentrating all genetic variation in the whole collection into a smaller (5-10% of the whole) subset.

Little attention has been given to how novel genetic variation might be found more effectively and efficiently in germplasm not agronomically adapted to a breeder’s target region. Most plant breeding texts do not offer much direction to the selection of parental material other than recognizing that PGR are an important source of new variation and these can be obtained from genebanks as well as other sources. Brown and Caligari (2008) is a recent exception where three general types of parents and their source are identified:

1. Unadapted material which possesses characters not available in more cultivated material which can come from plant introductions and genebanks;
2. Adapted material, such as current or obsolete cultivars, that might be obtained from other breeding programmes; and
3. Breeding lines selected from within the breeding programme.

They go on to describe parental selection as a cyclic process within the breeding programme and identify phenotypic and genotypic evaluation as two methods of deciding which parents to use in a breeding scheme.

Earlier, Tanksley and McCouch (1997) discussed the need for strategies to sample exotic germplasm. Recognising such large numbers of accessions in genebanks and the limited resources available, they realized the need for an effective and efficient way to choose which accessions to sample to increase the chance of finding novel and useful variation. Their solution was to use the ‘new paradigm’ of sampling for genetic composition rather than selection based on the phenotype. They state that “*accessions with DNA profiles most distinct from that of modern germplasm are likely to contain the greatest number of novel alleles*”. Such accessions are likely to be the CWR that have been shaped and evolved by nature over millions of years and landraces, which have benefited from the assistance of mankind during their domestication and subsequent selection. Of the known biological status of accessions estimated to be held in *ex situ* collections, about 60% are classified as landraces or wild/weedy species (FAO, 2010). In the case of wheat, if about 60% (or 500,000) of the estimated 856,168 *ex situ* genebank accessions are the potential source of novel genetic variation, and one is seeking a very rare allele, will the time and resources be available to evaluate such large numbers?

# 3 Sampling accessions for utilization

> *A practical botanist will distinguish at the first glance the plant of the different quarters of the globe and yet will be at a loss to tell by what marks he detects them.* Carolus Linnaeus (1707-1778).

## 3.1 Two challenges identified

With the estimated 7.4 million genebank accessions held *ex situ*, how does the researcher find the particular accessions that are most likely to possess the genetic variation required to address the challenges faced by farmers in their field? This is the primary question that this thesis seeks to explore. Information about the characteristics of genebank accessions is a limiting factor in their use; an important underlying question is how can such information be gathered, aggregated and made available to facilitate its use in utilizing PGR?

Table 1. Estimate of total number of accessions held for the seven largest crops and the three largest genebanks for each crop (Source: FAO, 2010). The code in round brackets is the FAO institute code[1] for each genebank listed.

| **Genus (crop)** | **Total accessions** | **1st Genebank** | **2nd Genebank** | **3rd Genebank** |
|---|---|---|---|---|
| *Triticum* (wheat) | 856 168 | CIMMYT (MEX002) | NSGC (USA029) | ICGR-CAAS (CHN001) |
| *Oryza* (rice) | 773 948 | IRRI (PHL001) | NBPGR (IND001) | CNRRI (CHN121) |
| *Hordeum* (barley) | 466 531 | PGRC (CAN004) | NSGC (USA029) | CENARGEN (BRA003) |
| *Zea* (maize) | 327 932 | CIMMYT (MEX002) | BPGV-DRAEDM (PRT001) | NC7 (USA020) |
| *Phaseolus* (bean) | 261 963 | CIAT (COL003) | W6 (USA016) | CNPAF (BRA008) |
| *Sorghum* (sorghum) | 235 688 | ICRISAT (IND002) | S9 (USA016) | ICGR-CAAS (CHN001) |
| *Glycine* (soybean) | 229 944 | ICGR-CAAS (CHN001) | SOY (USA033) | RDAGB-GRD (KOR011) |

[1] Refer to FAO WIEWS site http://apps3.fao.org/wiews/wiews.jsp?i_l=EN. (Verified 03/10/2011).

An understanding of how breeders describe and search for the genetic variation they require is necessary before trying to answer this question. Peeters and Williams (1984) found that breeders would like to obtain information, in order of priority, about the following traits from genebanks:

1) pest/disease,
2) adaptation/stress,
3) traits to improve productivity/yield,
4) grain/seed quality,
5) maturity.

These are mostly evaluation traits that are environmentally sensitive and many genebanks are not in a position to store such data so it can be easily accessed and used because genebanks differ in mandates. Some have a broader agenda and undertake research activities, evaluation and/or diversity studies whilst others have a narrower mandate and provide a service function to the germplasm users. Genebanks with a mandate for research and evaluation are likely to have the necessary facilities to manage such data, but many others are less likely to have such capacity.

Mackay (1986) reported the outcome of a study to gain an understanding of how wheat breeders used the AWCC during the early 1980s. The AWCC, for example, could have been described as a service genebank because its mandate stopped at characterization data, with the exception of plant height and maturity. Any other evaluation data held was collected in collaboration with genebank users. This data included traits such as reactions to pests and diseases.

In many cases the type of information breeders wanted is not actually collected by genebanks, but by the research community using the genebank, including pre-breeders and breeders. Much of this evaluation information never finds its way back to the genebank and remains unavailable for use in querying accessions; this has been a challenge up until the present time. One possible reason for this is that once an evaluation experiment is completed, the results are often published without reference to the actual accession identifiers of the germplasm sourced from a genebank. There will always be exceptions, perhaps one being the collections held by the N I Vavilov Research Institute of Plant Industry's in St Petersburg, Russia (VIR); in these collections there was an enormous amount of evaluation work undertaken over a considerable period of time in the 20th Century. In this case the information was usually collated into hard copy 'catalogs', before the digital age, which were then widely distributed. The digitization of this information has now commenced (Sergey Alexanian, personal communication) but is still not available to facilitate the selection of accessions for inclusion in breeding programs.

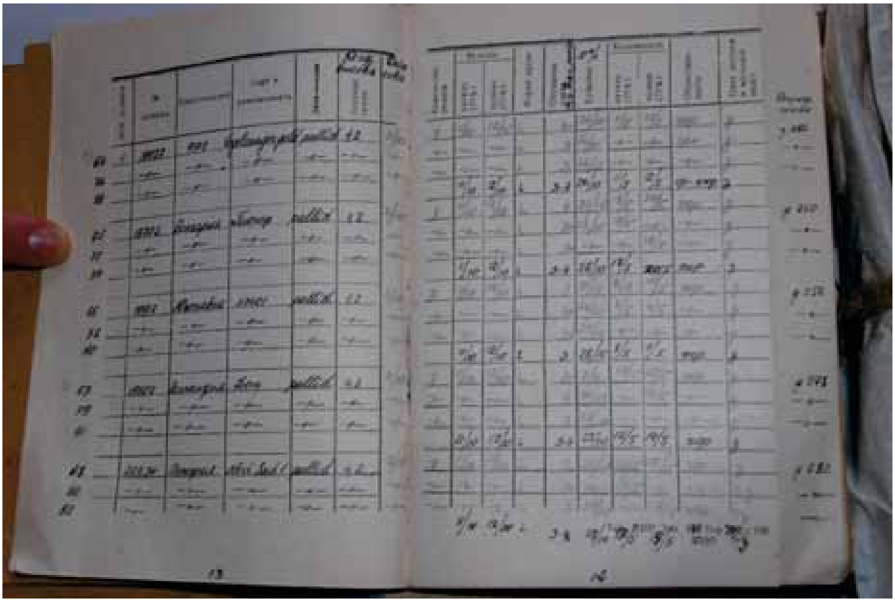

Figure 1. VIR has many volumes of data that can be digitized and used for identifying accessions of potential value to breeding programmes.

There are two challenges identified here:

1. How could one identify the genebank accessions that are most likely to contain the genetic variation being sought by breeders (Chapters 3 and 4)?
2. How could the required information be gathered and linked back to the accessions held in genebanks (Chapter 5)?

The main approaches to 'minimize the noise', or converge genetic variation in a smaller subset, are to either:

1. eliminate those accessions possessing agronomically unattractive traits, or
2. identify accessions most likely to possess novel variation for the specific adaptive traits sought by the breeder.

## 3.2 The core collection concept

The core collection was proposed as a way to work with fewer accessions that would represent, "*with a minimum of repetitiveness, the genetic diversity of a crop species and its relatives*" (Frankel, 1984). The concept arose largely in response to the questions being asked regarding difficulties in actually utilizing large collections (Frankel pers. comm., 1986). There are numerous examples of methodologies to develop core collections (see Hodgkin *et al*., 1995) which in practice tend towards limiting the size of the core collection to around 10% (Brown, 1989a, b) or, based on a different model, 20~30% (Yonezawa, 1995) of the original collection size. However, in terms of utilization *per se*, the contribution of core collections has not been clearly demonstrated. The literature suggests much of the focus has been on methods (or sampling strategies) to establish core collections (Holbrook *et al*., 1993; Ortiz *et al*. 1998; Hu *et al*., 2000; Malosetti and Abadie, 2001) and analyzing the diversity within core collections (Casler, 1995; Tohme, Gonzalez *et al*., 1996; Bartish *et al*., 2000; Fu *et al*., 2005).

The literature also suggests a number of core collection variants to enhance the efficiency of capturing diversity and promoting utilization. These include specialized core collections (Brown and Spillane 1999), mini-core subsets/collections (for example, Upadhyaya and Ortiz, 2001), nested core collections (McKhann *et al*., 2004) and composite core collections (Furman 2006). While numerous articles mention utilization, many do so in terms of opportunities to improve utilization (Diwan *et al*., 1994; Bisht *et al*., 1998) rather than demonstrate methods of identifying new sources of genetic variation via core collections - or modified versions thereof.

## 3.3 A case study to determine how genebank accessions can be used more effectively

For utilization of wheat genetic resources, Mackay (1986) describes various aspects of using the AWCC. These include the need for users to have an appreciation of the descriptors used to document genebank accessions, which were categorized into two main groupings – 'curator' and 'other' descriptors.

The former were used primarily for identification and description purposes whilst the latter are more useful for choosing genebank accessions for evaluation and included origin, biological status (refer to the FAO/IPGRI Multi-Crop Passport Descriptors, Alercia *et al.*, 2001), growth habit, kernel colour, time to flowering and plant height.

In the AWCC study Mackay (1986) found some descriptors were 'time-limited in value'. These might include reaction to fungal diseases where the race of the pathogen can determine if a particular accession is rated as resistant, susceptible or somewhere in between.

The concept of the 'predictive' value of descriptors was introduced. A combination of geographic origin, growth habit, grain colour, ear emergence and plant height are an example of descriptors that could be used to infer the type of environment in which an accession might have evolved or undergone passive and/or active selection.

At the time the AWCC was considering two methods for enhancing utilization:

- the 'predictive' approach to establish a small 'core' of accessions, and
- a co-ancestry sieve to eliminate closely related genotypes from screening projects.

These approaches came under consideration following input from Sir Otto Frankel during the New Zealand Department of Scientific and Industrial Research (DSIR) Plant Breeding Symposium (Lincoln, New Zealand, 17-20 February 1986) and a subsequent visit by Sir Otto, AHD Brown and DR Marshall to the AWCC in Tamworth, Australia. Methods to make the AWCC 'more usable' were considered an 'ongoing process' at the time.

The 1986 review (Ibid.) of how a genebank could be more effectively utilized concluded that "*By using these concepts, together with other methods, it's anticipated that the AWCC will be able to select smaller groups of accessions for testing by breeders. The chances of identifying desirable attributes should be significantly improved, while the numbers of accessions to test are reduced*". The underlying objective was effectiveness and efficiency in utilizing PGR; effectiveness through focusing on 'best bet' accessions for the particular adaptive trait under consideration and efficiency through reducing the number of accessions (and therefore the cost) to be evaluated.

# 4 The focused identification of germplasm strategy – FIGS

> *The answer is never the answer. What's really interesting is the mystery. If you seek the mystery instead of the answer, you'll always be seeking. I've never seen anybody really find the answer - they think they have, so they stop thinking. But the job is to seek mystery, evoke mystery, plant a garden in which strange plants grow and mysteries bloom. The need for mystery is greater than the need for an answer*. - Ken Kesey

Since Vavilov's agroecological survey of several crops (1957) there has been little attention paid to how to select a parent for use in breeding programs. The core collection concept was developed not as a way to choose parents for breeding purposes, although there are ample examples where it is stated that the core collection might be helpful for this purpose (for example, Upadhyaya *et al.*, 2006), but more as an approach to concentrate genetic diversity in a smaller subsample of a larger collection. A study of how the AWCC was used (Mackay, 1986) concluded that additional research was needed to more closely meet breeders' needs.

## 4.1 Towards a strategic approach to utilization

The goal to improve utilization of the AWCC by breeders had, by 1989, progressed to an extent where a general approach could be described more completely. The International Center for Agricultural Research in the Dry Areas (ICARDA) held a symposium in conjunction with the official opening of their genebank in 1989. It was at this symposium that Mackay (I) described the need for "*strategic planning for effective evaluation of plant germplasm*". An important point raised was that it is more important to place the emphasis on 'how to select accessions for evaluation' rather than on 'utilization of PGR' in the broader sense. This may, at first sight, seem an insignificant shift in emphasis but this is an imperative if genebanks are to be more effectively and efficiently utilized for breeding purposes. To highlight this shift in emphasis, it can be reworded as "***the necessary shift in thinking from utilization, in general, to targeting subsets of accessions to evaluate in order to find novel genetic variation for specific adaptive traits not already available from material adapted to the target breeding region***". Furthermore, concentrating on the processes involved in selecting germplasm for evaluation was expected to lead towards viable solutions to more effectively identify novel variation in less adapted material such as landraces and CWR. Tanksley & McCouch (1997) recognized that wild relatives were a good source of simply inherited resistance to biotic stresses but saw molecular approaches as the way forward for identifying further agronomically important traits.

A number of issues were addressed in planning for more effective evaluation of the AWCC genebank accessions. These included:

3. **Collaboration:** It is essential for cooperation between plant improvement (breeders) and germplasm scientists (genebank curators) in selecting material for evaluation. Most breeding programs, especially those that involve the more widely cultivated crops, only turn to a genebank for new material when a new challenge emerges that cannot be overcome with more adapted material. Peace (2011) advocates that, in order to facilitate the better use of tree fruit genebanks, one must 'think like a breeder' rather than like a curator, and describes how genebanks can be conduits to utilization.
4. **Address the trait of concern:** The trait for which new genetic variation is being sought should be paramount in the planning process. Traits important to breeders are generally adaptive traits that address economic and productivity challenges in the region targeted by a breeding program. We are largely seeking novel genetic variation, not alleles already present in more agronomically adapted genotypes. Traits that are not considered adaptive in the target region, and do not have any positive influence on a plant's ability to survive and reproduce in this region, can be largely ignored unless they have some deleterious effect in which case they could be selected against.
5. **Evaluation method:** The method of evaluation, which is critical to the design of experiments, is of importance in designing strategic evaluation experiments. The method is also important from the data standards viewpoint; potential users of the information must understand how the observations were derived in order to interpret them appropriately.

6. Methodically **evaluating all accessions** was another option. Whilst not rejected outright, such an approach should be considered only when other approaches have not been successful. Take, for example, the situation of newly emerging pathotypes of fungal diseases. Evaluating all accessions in a genebank today is unlikely to provide accurate information regarding the reaction of the same accessions to the pathotypes of the future. Furthermore, the cost of such an approach is inefficient in contrast to alternatives that can identify the necessary variation by evaluating half, or less, of the total number of accessions. This approach could, however, be considered if all other approaches were unsuccessful in identifying the required genetic variation.

Recognizing and better understanding the relationship between a particular adaptive trait and the environmental factors that shaped its expression is a vital key to more effectively discovering novel genetic variation. Breeders will always seek required variation in germplasm that is agronomically adapted to their target region. If this is not available they then search less adapted material such as landraces and CWR for the necessary variation. Finding novel variation in these categories of germplasm can make use of Vavilov's findings in his agroecological survey of the main field crops (Vavilov, 1957) and even extend it with assistance of new technologies. Mackay (I) used the example of resistance to pre-harvest sprouting in wheat and it was suggested that the possibility of finding such resistance would be more likely in landrace germplasm that evolved in an environment where the incidence of precipitation during grain maturation was an intermittent occurrence compared to other environments. There will always be exceptions, but nature is not expressed in absolutes (black and white) and genetic variation is not always expressed as simply dominant and recessive; there is a continuum of expression that is further influenced by environmental factors.

The full potential of 'predictive' attributes "*will not be fully realized until more suitable tools for linking environments to accessions are developed, such as a database which cross-references soil types and climate with geographic regions*" (Ibid.).

It is possible to plan more strategically in order to choose the non-agronomically adapted genebank accessions most likely to possess variation for the particular trait of interest. At the time three approaches to make evaluation more effective were proposed:

- Use the core collection concept to reduce the number of accessions to evaluate for a specific trait;

- Select economically important traits for evaluation;
- Make the resulting information widely available.

## 4.2 General vs targeted approach

> We cannot do everything at once, but we can do something at once - Calvin Coolidge

For breeders to use a core collection, an advantage in doing so must be apparent. The use of unadapted genebank accessions by breeders is sporadic and only when the required variation cannot be found in adapted material. To engage breeders in the use of a core collection there is the need for it to be evaluated for traits that are useful for breeding and generally such information is not easily accessible (Spagnoletti Zeuli and Qualset, 1995). Marshall and Brown (1981) observe that while evaluation is an initial step in utilization, insufficient information has not prevented use in the past. The question now becomes, does the core collection best meet the needs of the breeder in finding novel genetic variation, or are there more effective and efficient alternatives?

Mackay (II) suggests that the core collection principle can be developed into a more dynamic approach to meet the changing needs of breeders. This is not intended to question the use of the core collection as a way to concentrate genetic variation within a smaller subset of a wider collection, but to extend the concept to include how breeders actually go about finding new genetic variation for deployment in plant improvement programmes. It is less complex, when identifying a small subset of accessions to evaluate, to do so for a single trait than it is to identify a subset that represents virtually all the available genetic variation for all traits in the wider collection.

Another argument for focusing on a single trait at a time is that to completely evaluate a core collection for all traits of contemporary (and future) interest to breeders would require enormous resources that are not always readily available to either genebanks or breeders. For example, the relationship between the number of accessions to evaluate and the level of success in finding useful genetic variation is not linear. Vaughan and Jackson (1995) provide information about the number of rice accessions that were evaluated for ten biotic and abiotic stresses and how many were found to have a resistance or moderate resistance to each stress (Table 1). These are the results from large-scale evaluation of rice genetic resources at the International Rice Research Institute (IRRI). The average success rate for these stresses was about

14%, but it ranges from 1.5% to nearly 62%. Moss *et al.* (1989) and Lehmann (1984) found one desirable *Arachis* accession in every 144 evaluated for 16 stresses, and 11 resistant barley accessions in 6,000 evaluated against six isolates of leaf rust (*Puccinia hordei* Otth.), about 0.7% and 0.2% success rate, respectively.

Table 2. The effectiveness of systematic evaluation of abiotic and biotic stresses in rice - after Vaughan and Jackson (1995).

| **Stress** | **Total number of accessions evaluated** | **Resistant to moderate resistance accessions (number)** | **% success**[1] |
|---|---|---|---|
| Brown planthopper | 44,335 | 682 | 1.5 |
| Green leafhopper | 50,137 | 1,403 | 2.8 |
| Rice whorl maggot | 22,949 | 697 | 3.0 |
| White-backed planthopper | 52,042 | 871 | 1.7 |
| Bacterial blight | 49,752 | 5,512 | 11.1 |
| Blast | 36,634 | 9,616 | 26.2 |
| Sheath blight | 23,088 | 2,153 | 9.3 |
| Drought resistance at early vegetative stage | 28,319 | 4,288 | 15.1 |
| Drought resistance at late vegetative stage | 22,873 | 1,826 | 8.0 |
| Recovery from drought stress | 24,432 | 15,115 | 61.9 |

[1] The percentage of accessions found to be either resistant or moderately resistant to each stress.

Mackay's proposal to develop trait-specific subsets – in contrast to capturing all genetic variation with a minimum of repetitiveness (the core collection concept) – in order to enhance utilization was later recognized by Rana and Kochhar (1996) and was subsequently recommended for assessment by Allem (2001). Trait-specific subsets should not, of course, be considered as a form of core collection because they are selected for a specific use, which is considered outside the scope of core collections by Brown and Spillane (1999).

## 4.3 A project to explore the hypothesis

Although the general principles of targeting germplasm accessions to evaluate for specific adaptive traits had been described by the early 1990's, little interest could be found in resourcing research to test and further develop this approach. However, by 2000 the Grains Research and Development Corporation (GRDC, based in Canberra, Australia) was looking to expand its plant genetic resources research and development portfolio through collaborative research. An initial proposal was developed in 2001 by Paul Brennan (GRDC Consultant at the time), Kenneth Street (ICARDA) and Michael Mackay (AWCC) during a meeting in Coffs Harbour, NSW. Later the same year the GRDC and the Australian Centre for International Agricultural Research (ACIAR) sponsored the travel of a number of scientists (including Street and Mackay) to St Petersburg with the view to developing the earlier proposal into a full project specification in collaboration with VIR. The resulting project, entitled "*Technologies for the targeted exploitation of the N I Vavilov Institute of Plant Industry (VIR), ICARDA and Australian bread wheat landrace germplasm for the benefit of the wheat breeding programs of the partners*" was resourced for an initial three year period.

The objective of this project was to exchange germplasm and investigate the application of the principles previously enunciated by Mackay (I, II) using the bread wheat landraces held by the collections of the International Center for Agricultural Research in the Dry Areas (ICARDA), VIR and the AWCC. By this time (2002) the development of the geographic information systems (GIS) had advanced significantly and this provided one of the key components that Mackay (I) had suggested was necessary for strategic evaluation – *"The value of these 'predictive' attributes will not be fully realized until more suitable tools for linking environments to accessions are developed…"*.

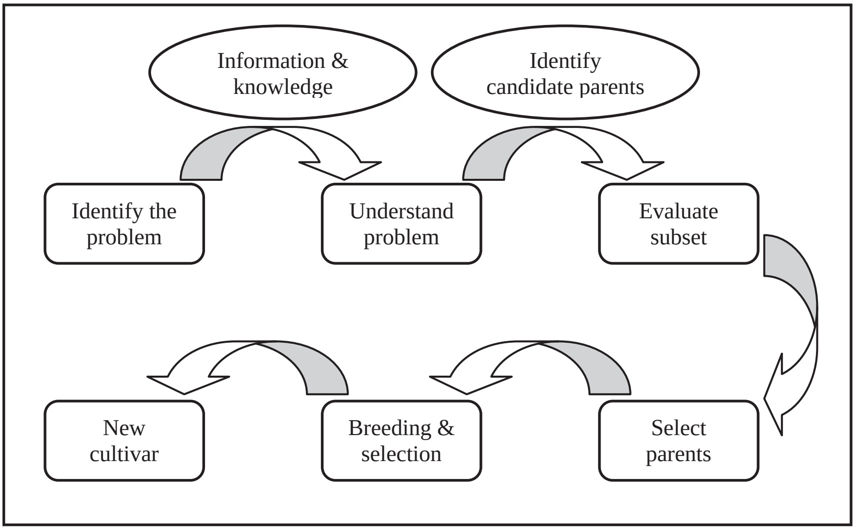

Figure 2. FIGS approach to rational use of genebank accessions for plant improvement (Mackay et al., III).

## 4.4 Linking accessions to environments

The GIS Unit at ICARDA provided the necessary expertise to integrate information about plant genetic resources with geospatial information. The initial work of the project partners involved the painstaking geo-referencing of the bread wheat landraces held in each partner's collection to create a virtual collection of some 16,500 accessions.

Environmental data for each collecting site was then obtained by creating continuous surface digital maps for the whole of Eurasia for environmental parameters from point data using the 'thin-plate smoothing spline' method (Hutchinson and Corbett, 1995) implemented in the ANUSPLIN software (Hutchinson 2000), with modifications by De Pauw (2002). The value of each parameter was then extracted from the digital maps for all collecting sites.

GTOPO30 is a global digital elevation model (DEM) with a horizontal grid spacing of 30 arc seconds (approximately 1 kilometer) (http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30_info. Verified 03/10/2011) that was used to derive elevation and day length; precipitation and temperature were based on different sources. Initially the surfaces were derived using a World Meteorological Organization (WMO) climatic data set (3065 temperature and 2061 precipitation stations). The WMO-based climate attributes were then overwritten by those from the compact disk *Land Resources of Russia*, which had a higher density of sites for Russia than the WMO data set, and those from an ICARDA dataset with a denser place-specific net of stations for Syria (87 temperature, 136 precipitation), the Caucasus (190 temperature, 189 precipitation) and Central Asia (396 temperature, 378 precipitation).

The following primary parameters were derived from the above models and datasets:

Table 3. Primary environmental parameters derived from the models and datasets described here. All, with the exception of elevation, are long term average figures.

| | | |
|---|---|---|
| annual precipitation | annual maximum temperature | annual minimum temperature |
| monthly precipitation | monthly maximum temperatures | monthly minimum temperatures |
| elevation | | |

The parameters in Table 3 were calculated using those listed in Table 2.

Table 4. Environmental parameters derived from the primary parameters in Table 3.

| | |
|---|---|
| long-term average monthly potential evapotranspiration | long-term average monthly aridity indices |
| long-term average annual potential evapotranspiration | long-term average annual aridity indices |
| length of growing period[1] | |

[1] Method of De Pauw (2002).

Table 5. The continuous surface parameters listed were captured from the WorldClim global climatic layers (Hijmans *et al*., 2005). All parameters were based on long-term average data.

| | | |
|---|---|---|
| long-term average isothermality (mean monthly/annual temperature range) | temperature seasonality (standard deviation of monthly temperature x 100) | maximum temperature of warmest month |
| minimum temperature of coldest month | annual temperature range | mean temperature of wettest quarter |
| mean temperature of driest quarter | mean temperature of warmest quarter | mean temperature of coldest quarter |
| precipitation of wettest month | precipitation of driest month | precipitation seasonality (coefficient of variation [cv]) |
| precipitation of wettest quarter | precipitation of driest quarter | precipitation of coldest quarter. |

The collecting-site environments were further characterized using an agro-climatic zoning (ACZ) method developed for this study (Pertziger, personal communication, 2004). This method is based on a simplified length of growing period concept that used threshold values for average air temperature, precipitation and day length to cluster the landrace collection sites into six growing period intervals. The sites were further classified based on whether they were limited by temperature, moisture, both or neither. The resulting zones are detailed in Table 6.

## 4.5 FIGS Drought Set

Two adaptive traits of common interest to the partners were tolerance to drought and salinity. The method used to develop the FIGS ‘drought set’ will be explained in some detail here to demonstrate the progression from the earlier and simpler ways to choose a ‘targeted’ subset, for example for soil boron toxicity (I), to more statistical methodology and onwards to the use of modeling techniques.

Table 6. The ACZ method, based on length of growing period and growth-limiting factors of temperature (T) and moisture (M), either individually, combined (B) or neither (N). Only accessions from the ACZs in bold were included in the FIGS drought set. (LGP = length of growing period.)

| **Climatic Class** | | **Limiting Factor – Class name suffix** | | | |
|---|---|---|---|---|---|
| **LGP (months)** | **Class name Prefix** | **T (temperature)** | **M (moisture)** | **B (both)** | **N (neither)** |
| 1-3 | -2 | -2T | -2M | -2B | |
| 4-5 | -1 | -1T | **-1M** | **-1B** | |
| 6-7 | 0 | 0T | **0M** | **0B** | |
| 8-10 | 1 | 1T | 1M | | 1N |
| 11-12 | 2 | 2T | 2M | | 2N |

The FIGS drought set was chosen using all the agro-climatic data derived for the collecting sites. The set was designed to identify the accessions that are most likely to provide genetic variation for traits that contribute to drought tolerance, resistance and/or avoidance.

The following method was employed in a stepwise fashion: accessions from known irrigated sites were excluded. Next, accessions from ACZs with length of growing period between four and seven months, and from zones where moisture or both moisture and temperature were limiting factors, were retained. Of these accessions, only those from sites where the annual precipitation was in the range of 180-300 mm were kept. Accessions identified in the first three steps were then subjected to a hierarchical cluster analysis using all-site agro-climatic parameters with a pre-set cluster number of 750. One accession from each of these clusters was randomly chosen to make up the FIGS drought set.

As a check set, a core collection – a representative of the virtual bread wheat landrace collection – was developed for comparison with the FIGS drought set. The landrace core collection was developed using techniques similar to those used by others to create core collections based on geographic and ecological information. The core collection was derived by first grouping all accessions into the FAO agro-ecological zones (AEZ) (Fisher *et al.*, 1999) from which they were collected. Sub-groups within the AEZ groups were then formed based on country of origin. The number of accessions within each of these country sub-groups determined acceptance into the core collection, which was designed to contain approximately 5% of the accessions used in the study. Where there were 1-20 accessions in a cluster, one accession was randomly placed in the core collection. Where there were 21-60 accessions in a cluster, 5% of the accessions were chosen at random for the core collection. Where there were more than 61 accessions in the cluster, an additional hierarchical cluster analysis using site data was performed with the number of clusters preset at 5% of the total number of accessions in the sub-group. One accession from the resulting clusters was randomly selected for the core collection.

Principal component analysis (PCA) was conducted to gain a comprehensive picture of climatic influences in the environments from which the accessions in this study were collected. In this way, a comparison could be made between the FIGS drought set and the core collection sites for predominately agro-climatic selection pressures.

The representative core collection contained 711 landrace accessions, collected during 130 missions between 1921 and 2002, representing 51 countries across most regions in North and East Africa, Asia and Europe. By contrast, the FIGS drought set contained 716 accessions, collected during 66 missions between 1924 and 2000, covering 19 countries in North Africa and most Asian regions. There were 31 accessions that were common to both the core collection and the FIGS drought set. The distribution of the collection sites for both subsets is illustrated in Figure 3.

Table 7. The analysis was performed on a correlation matrix for the parameters (listed below), based on long-term average data using the SPSS 14.0 (SPSS Inc, Chicago IL, USA) statistical package.

| | | | |
|---|---|---|---|
| altitude (DEM) | monthly precipitation | annual precipitation | monthly maximum temperature |
| yearly maximum temperature | monthly minimum temperature | yearly minimum temperature | monthly potential evapo-transpiration |
| yearly potential evapo-transpiration | monthly aridity index | yearly aridity index | annual mean temperature |
| mean monthly temperature | isothermality (mean monthly/annual temperature range) | temperature seasonality (standard deviation of monthly temperature x 100) | maximum temperature of warmest month |
| minimum temperature of coldest month | annual temperature range | mean temperature of wettest quarter | mean temperature of driest quarter |
| mean temperature of warmest quarter | mean temperature of coldest quarter | precipitation of wettest month | precipitation of driest month |
| precipitation seasonality (cv) | precipitation of wettest quarter | precipitation of driest quarter | precipitation of coldest quarter. |

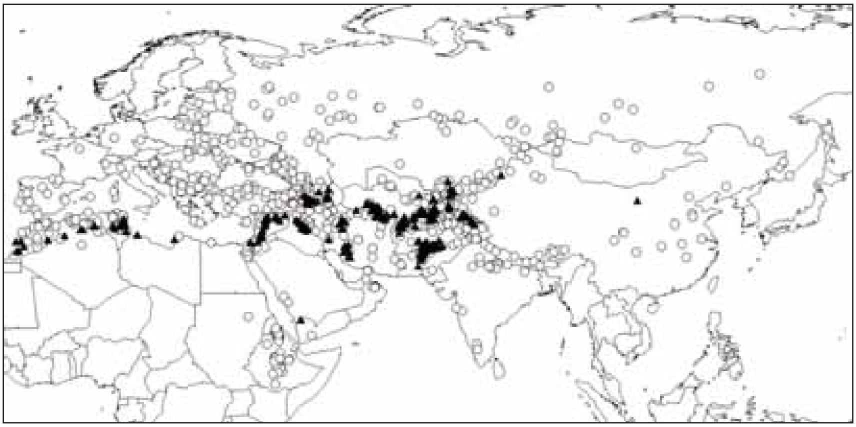

Figure 3. Geographic distribution of representative core collection (open circles) and FIGS drought set (solid triangles) landrace collection sites.

The PCA is a useful tool to summarize key agro-climatic influences in the collecting site environments of each subset using all of the agro-climatic parameters developed for this study. It summarizes the factors having the most impact in differentiating the collecting site environments with the exception of non-environmental factors, such as migration.

Table 8. Frequency of core collection and FIGS drought set landrace accessions by macro-geographical region, country or province and number of collecting sites.

| **FAO Macro Regions** | **Core Collection** | | | **FIGS Drought Set** | | |
|---|---|---|---|---|---|---|
| | **Accs**[1] | **ADM0** | **Sites** | **Accs.** | **ADM0** | **Sites** |
| Eastern Africa | 65 | 7 | 48 | | | |
| Northern Africa | 53 | 31 | 53 | 150 | 13 | 33 |
| Central Asia | 95 | 32 | 88 | 59 | 12 | 46 |
| Eastern Asia | 31 | 14 | 30 | 2 | 2 | 2 |
| Southern Asia | 191 | 56 | 186 | 395 | 21 | 190 |
| Western Asia | 118 | 75 | 118 | 110 | 28 | 72 |
| Eastern Europe | 110 | 66 | 110 | | | |
| Northern Europe | 5 | 4 | 5 | | | |
| Southern Europe | 41 | 22 | 40 | | | |
| Western Europe | 2 | 2 | 2 | | | |
| Totals | 711 | 309 | 680 | 716 | 76 | 343 |

1. Accs = accessions; ADM0 = country or administrative boundary 0; Sites = collection sites.

A plot of the first two principal components is provided in Figure 4; the x axis represents the first component, which explains 47% of the variation between sites, while the y axis represents the second component, which explains a further 21% of the total variation.

The plot clearly illustrates that the core collection is composed of accessions from sites that are influenced by a wide range of climates, from lower temperatures and higher precipitation to more arid environments. By contrast, the drought subset is from environments predominantly experiencing lower precipitation, higher temperatures, higher PET and that are generally more arid, as indicated by Table 8.  About 70% of the FIGS drought subset falls into quadrant 3 with only one outlier in quadrant 1, which represents less arid environments.

Table 9. Summary of distribution of collecting sites as determined by PCA.

| **Quadrant** | **Parameter weightings** | **Core Set** | **Drought Set** | **Both** |
|---|---|---|---|---|
| Quadrant 1 | Lower temps, lower PET, higher precipitation, less arid | 238 | 1 | |
| Quadrant 2 | Decreasing aridity, high temp, high PET | 132 | 96 | 4 |
| Quadrant 3 | Decreasing precipitation, high temp, high PET, arid | 96 | 487 | 17 |
| Quadrant 4 | Lower temps, lower PET, lower precipitation, more arid | 214 | 101 | 10 |

For the FIGS drought set, 86% of the collecting sites are in the driest and warmest quadrants of the PCA plot (quadrants 3 and 4) compared to about half (46%) for the core collection (Figure 4, Table 9). Thus, the chances of finding novel genetic variation for traits conferring adaptation to hot, dry conditions would be higher in the FIGS drought set rather than in a core collection.

The rationale behind the FIGS method is well supported in the literature. That is, the environment strongly influences gene flow, natural selection and thus spatial/geographic differentiation (Lin *et al.*, 1975; Spieth 1979; Epperson 1990). Numerous studies have documented eco-geographic variation for a range of traits that were shown to be linked to environmental parameters (Table 10).

Of course the works of Vavilov (1992) pre-date the examples given in Table 10 and provide an extensive description of traits being in distinct ecological-geographical groupings and how an understanding of this can suggest where to find traits of value to breeding.

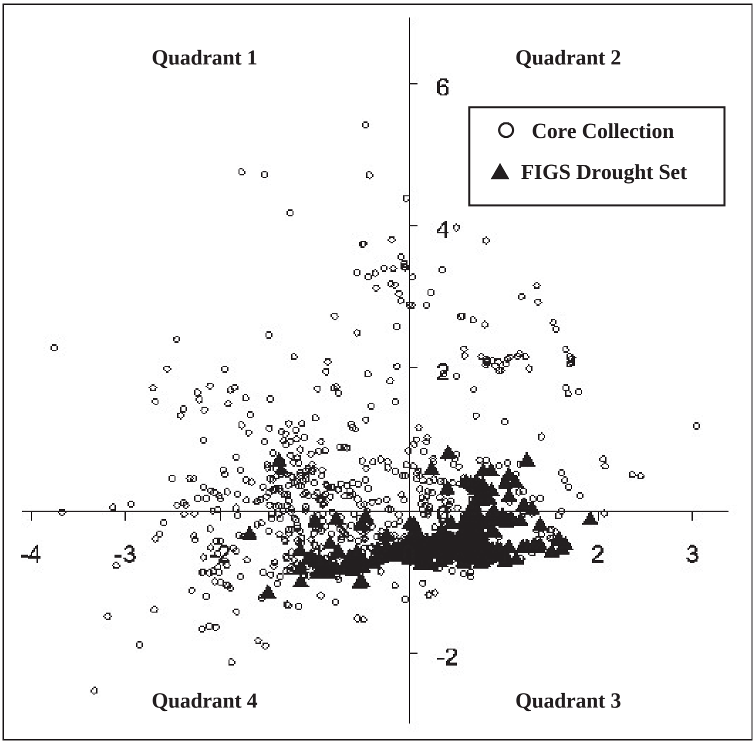

Figure 4. Principal component analysis (PCA) of the two sets of data with environmental parameters.

## 4.6 Powdery mildew

Following a FIGS presentation given at the 10th International Wheat Genetics Symposium (Paestum, Italy, 2003) the University of Zurich requested a FIGS powdery mildew set for bread wheat to be developed for allele-mining research (Kaur *et al.*, 2008; Bhullar *et al.*, IV, V). The approach used to develop a FIGS set was to initially identify some 400 landraces with known resistance to powdery mildew (*Blumeria graminis* f.sp. *tritici*) held by the USDA-ARS National Small Grain Collection in Aberdeen, Idaho. Next, environmental profiles of the sites where these landrace accessions originated were determined and this was used as a reference to identify different collection sites with similar environmental profiles using the virtual collection of nearly 17,000 bread wheat landrace accessions described in the preceding section. The resulting FIGS Powdery Mildew Set (FIGS PM Set) included 1,302 accessions from 10 countries (Table 11).

Table 10. Examples of eco-geographical variation of traits linked to environmental influences

| **Species** | **Trait** | **Environment influence** | **Reference** |
|---|---|---|---|
| *Trifolium repens* | Cyanogenesis | Low altitudes, high winter temperatures, low summer precipitation, greater spring cloudiness | (Pederson *et al.*, 1996) |
| *Medicago* species | Frost tolerance | Low temperature | (Cocks & Ehrman, 1987) |
| Annual legumes | Seed dormancy, early flowering, high seed-to-pod ratio | Aridity | (Ehrman & Cocks 1996) |
| Desert species | Antitelechory | Desert environments | (Ellner & Shmida, 1981) |
| Bread wheat | Tolerance of boron toxicity | Soil type | I |
| Bread wheat | Russian wheat aphid (RWA) resistance | Geographic RWA distribution | Street & Mackay, unpublished data |
| *Triticum dicoccoides* | Drought resistance | Temperature, aridity | (Peleg *et al.*, 2005) |
| Durum wheat | Glume colour and beak length | Altitude | (Bechere *et al.*, 1996) |
| *Triticum dicoccoides* | Heading date, culm length, biomass, grain yield and its components | Climate, soil and water availability | (Beharav & Nevo, 2004) |
| Durum wheat | Glutenin diversity | Precipitation, minimum January temperature, altitude. | (van Hintum & Elings, 1991) |
| Barley | Response to abiotic stresses | Temperature, altitude and rainfall | (Liviero *et al.*, 2002) |
| Barley | Hordatine accumulation (disease defense) | Water relations, temperature and edaphic factors | (Batchu *et al.*, 2006) |

These accessions were subsequently evaluated for their reaction to a number of known powdery mildew isolates based on their pathogenecity to known alleles from the *Pm3* locus. 211 accessions exhibited either resistance or intermediate resistance to at least one of the isolates used. Following molecular characterization a smaller set of 111 candidate accessions were identified with the following characteristics; i) resistant or intermediate resistant to at least one of the isolates used, ii) possessed the *Pm3* haplotype and iii) lacked known *Pm3* alleles (Kaur *et al.*, 2008).

Subsequently, the isolation of *Pm3* alleles was done only on the resistant landrace accessions and coding sequences were amplified from 45 of these before being cloned and sequenced. Following further molecular analyses seven novel functional *Pm3* alleles were identified, doubling the number of active *Pm3* alleles known in bread wheat (Bhullar *et al.*, III). Glaszmann *et al.* (2010) described this study as 'of outstanding interest' in assessing genetic diversity for crop improvement.

Further studies of allelic diversity at the *Pm3* locus (Bhullar *et al.*, V) provided insights into the geographic origin of a number of the alleles. For example, *Pm3e* was initially described in an Australian cultivar (W150 also known as Cowra No. 24 and Gullen) but later found to be in the region the *Pm3Go/Jho* allele; the associated proteins only differed by a single amino acid and a Himalayan origin for *Pm3e* was proposed. William Farrer was known to have obtained wheat varieties from India in the 1890s, which might explain the presence of this allele in an Australian cultivar. Evidence obtained also suggests *Pm3b* originating in the vicinity of Afghanistan and *Pm3c* from a geographic area including India and Nepal.

Table 11. Origin of FIGS PM Set accessions.

| **Country** | **No** |
|---|---:|
| Turkey | 419 |
| Iran | 391 |
| Afghanistan | 292 |
| Pakistan | 133 |
| Armenia | 34 |
| Turkmenistan | 16 |
| Russia | 9 |
| India | 6 |
| Azerbaijan | 1 |
| Uzbekistan | 1 |

## 4.7 Russian wheat aphid

Prior to 2006 ICARDA had randomly screened some 5,000 bread wheat landraces from its genebank without identifying any resistance to Russian wheat aphid (RWA), *Diuraphis noxia* (Kurdjumov). The FIGS approach was used to develop a RWA Set from the AWCC, ICARDA and VIR virtual collection of bread wheat landraces (El Bouhssini *et al.*, VI).

The steps involved in developing the FIGS RWA Set included applying progressive filters to the virtual collection. The 1st filter selected 10,200 accessions from countries where RWA had been reported; the 2nd filter chose only those accessions whose collection site was classified as arid, semi-arid and semi-humid according to the ENESCO (United Nations Educational, Scientific and Cultural Organization) agro-climatic zone classification and reduced the number of candidate accessions to 3,338; because temperature and elevation also appear to be related to RWA distribution; the 3rd filter only chose those accessions belonging to three country groups, with specific ranges of elevation, and reduced the candidate accessions to 1,125; of these candidates 510 were stored in the ICARDA genebank and only these were used for the study.

These accessions were grown in the field and infested with RWA collected locally at Tel Hadya, at the 2 leaf stage. Observations were made after symptoms were clearly visible on susceptible check varieties. Each of the twelve (12) accessions identified as promising in the field were subsequently grown in the greenhouse (in a randomized complete block design involving four replications) and infested with ten adult aphids at the 2 leaf stage. Observations were made four weeks after infestation and six of the accessions were found to have a high level of resistance, four a good level of resistance and two only a moderate level of resistance. Because these were the first sources of resistance to the local Syrian strain of the RWA identified in the ICARDA genebank, despite some 5,000 accessions being previously screened, the FIGS approach to identifying accessions likely to possess novel genetic variation for specific adaptive traits was seen to be both effective and efficient. FIGS was also used to successfully identify wheat with resistance to Sunn pest *Eurygaster integriceps* Puton (El Bouhssini *et al*., 2009).

Endresen et al. (VII) further endorsed the utility of the FIGS approach in a study involving two different fungal pathogens in two crops. FIGS subsets of accessions were developed using climatic data to identify accessions with resistance to each crop and associated pathogens. Additionally, a random subset of accessions was developed to compare with the FIGS sets. The results support previous studies mentioned (IV, VI), the difference here being the comparison of FIGS sets with a random sample, thereby providing proof of concept.

# 5 Information - a key to effective utilization

> *If you think in terms of a year, plant a seed; if in terms of ten years, plant trees; if in terms of 100 years, teach the people.* Confucius (551-479 BC)

## 5.1 A global information system for plant genetic resources

The International Treaty on Plant Genetic Resources for Food and Agriculture (IT-PGRFA, or Treaty) entered into force in June 2004. Article 17 of the Treaty describes the need for and purpose of a 'global information system on plant genetic resources for food and agriculture' which includes facilitating '*the exchange of information, based on existing information systems, on scientific, technical and environmental matters related to plant genetic resources for food and agriculture, with the expectation that such exchange of information will contribute to the sharing of benefits by making information on plant genetic resources for food and agriculture available to all Contracting Parties*'.

The Commission on Genetic Resources for Food and Agriculture (Commission), at its Twelfth Regular Session (October 2009, Rome, Italy), requested its Secretary "*to collaborate … in the development of the vision paper … to take stock of existing information systems and to outline a process for the development of the global information system in the context of Article 17 of the International Treaty*." (Report available at http://www.fao.org/nr/cgrfa/cgrfa-meetings/cgrfa-comm/twelfth-reg/en/, verified 24/08/2011). The vision paper was subsequently prepared and presented at the Fourth Session of the Governing Body in March 2011 and is available on the Treaty website (http://www.planttreaty.org).

The Treaty and Commission clearly recognize the need for a global information system that will enable access to PGR and its subsequent sustainable use.

As another link in the ‘information chain’, the Global Crop Diversity Trust (GCDT) initiated a process in 2005 that led to the development of more than 30 global crop and regional strategies for *ex situ* conservation and utilization. Khoury *et al*. (2010) reviewed these strategies and identified the ‘primary constraints affecting conservation, use, and planning are the quality and availability of accession-level information, and availability of resources for regeneration, collecting, and research’. They went on to examine the eight themes (Table 12) that emerged from the strategies as global *ex situ* needs. One of these was information systems.

Table 12. Eight themes examined by Khoury *et al*., 2010.

| | |
|---|---|
| Regeneration | Information systems |
| Crop wild relatives | User priorities |
| Collecting | New technologies and/or research |
| Crop descriptors | Challenges to building a robust strategy for rational conservation |

Information about *ex situ* PGR is collected and aggregated by many different stakeholders. These include plant collectors, genebanks (both *in situ* and *ex situ*), breeders and other researchers. There are several sources of information that can facilitate the use of PGR including genebanks and breeders. Many genebanks have a mandate to fully document accessions and maintain them whilst others have extended mandates that include evaluation and research into genetic diversity. Thus, genebanks require an information system that enables them to fully manage all the information associated with their mandated activities and such systems become a key source of data for utilization.

In terms of utilization by breeders, it is often necessary to have access to germplasm from more than one or two genebanks. There are a number of examples of PGR information portals that aggregate data from numerous genebanks and publish this online. The Genetic Resources Information Network (GRIN) of the United States Department of Agriculture Agricultural Research Service (USDA-ARS) is an example of aggregating data from a number of genebanks in one online system. GRIN also accumulated significant quantities of evaluation data associated with the genebank accessions and made these available through the online interface. Another web based catalogue of *ex situ* germplasm accessions is the European Cooperative Programme for Plant Genetic Resources’ EURISCO providing access to passport data for more than a million accessions held in the national genebanks of over 42 European countries. The System-wide Information Network for Genetic Resources (SINGER) of the CGIAR is a third significant online resource for passport data associated with the germplasm held in the International Center's genebanks, comprising some 600,000 accessions.

These three systems are largely based on passport data. Breeders and others have frequently expressed the need for access to additional information to facilitate the use of PGR. The additional information includes characterization and evaluation (phenotypic) data, molecular data and, more recently, environmental data (Khoury *et al.*, 2010). The crop and regional strategies also identify the "*greatest constraint on utilization*" as "*the lack of accession level information, including passport, characterization (morphological and molecular) and evaluation data, especially for 'useful traits'*" (Ibid.).

## 5.2 Global Information on Germplasm (GIG) project

The Trust partnered with Bioversity International (Bioversity) and the Treaty in 2007 to develop a project proposal entitled '*Global Information on Germplasm (GIG): Information management in support of the global system for the conservation and sustainable use of PGRFA*'. The scope of the GIG project included the CGIAR mandate crops, with a focus on the following 22 crops: banana, barley, beans, breadfruit, cassava, chickpea, coconut, cowpea, fababean, finger millet, grass pea, maize, major aroids, lentil, pearl millet, pigeon pea, potato, rice, sorghum, sweet potato, wheat and yam, and a limited number of other crops of interest to the project collaborators. The project had three components:

- Development of characterization and evaluation data standards for 22 target crops;
- Deployment of GRIN-Global, a new genebank information management system developed through a complementary USDA project; and
- Development of a global accession-level information system.

The project commenced in mid 2008. The data standards component was delivered in May 2010 and the deployment/evaluation of GRIN-Global is due for completion in December 2011.

In May 2011 the first version of a global accession level information system (the 3rd component), called GENESYS, was launched. GENESYS was developed under the guidance of an International Steering Committee (ISC), the outcome of which has significantly shifted the paradigm for access to and use of information about PGR. The first version of GENESYS combined the passport data from GRIN, EURISCO and SINGER in a single portal covering some 2.33 million accessions – about one third of the accessions estimated in the 2010 State of the World report (FAO, 2010). In addition to the passport data over 11 million records of characterization (morphological) and evaluation data, linked to the accessions in the participating genebanks, were included. Furthermore, some 11 million records of climatic information for the 625,290 accessions with collection site geo-references are also included. The *pièce de résistance* that GENESYS introduced was to give users the ability to build completely customizable queries using any of the available data. Thus, a breeder can use any combination of passport, characterization, evaluation and environmental data to identify accessions that best meet the requirements of the breeding program. For example, a wheat breeder can enter the GENESYS portal (http://www.genesys-pgr.org), select wheat from the crop list, and then build a query (mostly using the computer mouse) that, for example, selects accessions from a specific country where the annual precipitation is within a specified range, has red kernels and is resistant to stem rust. The GENESYS portal has a user friendly and intuitive interface that provides access to information in a sophisticated database designed especially for this purpose (Nawar and Mackay, 2009).

## 5.3 GENESYS and its role as part of a global information system

The GENESYS portal provides a solid foundation that can be built upon to further improve the access to and use of PGR. Information systems about germplasm held in genebanks have largely, for the past twenty or more years, only provided access to limited or primary PGR data. Some genebank documentation systems still consist of hard copy data, some use computerized spreadsheets and others use in-house designed relational databases – many of which are not accessible online. Even the larger regional, international or national systems only provide limited data, often not extending beyond some critical passport data and/or being restricted to a specific number of crops or genera.

GENESYS has been built upon existing systems and so does not represent a completely new investment. One driving force behind the development of GENESYS was to provide a ‘one-stop-shop’ where breeders and other germplasm users can have access to, and request, germplasm that is stored in many different genebanks across the globe. GENESYS meets the requirements of the Treaty in that it provides access to the conditions under which material is distributed within the Multi-lateral System (MLS) using the Standard Material Transfer Agreement (SMTA), it allows requests to be sent directly to the genebank(s) holding requested germplasm, and it will continue to develop in compliance with Treaty requirements.

GENESYS has also been developed with the need to be interoperable with other information systems always kept in mind. No one information system can cope with all the requirements of genebanks, breeders, scientists along with all the other users of PGR information. Currently there are numerous initiatives to build information systems that can capture and store data related to many aspects of PGR. One example is the Seeds of Discovery (SeeD) initiative being funded by the Mexican Government and coordinated by CIMMYT. This project will genetically characterize maize and wheat and make this information available via a web based information system. There are numerous facilities developed for the rapid and largely automated phenotyping of germplasm, one example being the Plant Accelerator at The University of Adelaide in Australia (http://www.plantaccelerator.org.au/. Verified 26/09/2011). Currently many CGIAR Centers conduct widespread evaluation of advanced breeding material via international nurseries and the resulting data is often difficult to find and/or use. GENESYS could become a pivotal link between such information systems and provide access to the actual genetic material from which the information has been obtained. Figure 5 illustrates one viewpoint as to how this might evolve over the next 3 to 10 years.

A second phase of developing GENESYS is being planned at this very time. Now that the foundation has been laid and the concept of combining the functionality to build custom queries across numerous data categories has gone beyond ‘proof of concept’, many more stakeholders will participate in the planning, development and implementation of GENESYS II. These will include CGIAR Centers via the Inter-Center Working Group on Genetic Resources (ICWG-GR), the Trust, The Treaty as well as other national and regional partners that might include the USDA-ARS, the European Cooperative Programme on Plant Genetic Resources (ECPGR), India’s National Bureau of Plant Genetic Resources (NBPGR), the Brazilian Agricultural Research Corporation (EMBRAPA) and the newly established Australian Grains Genebank (AGG). To ensure GENESYS provides the functionality required of the users of PGR, especially breeders and others involved in plant improvement, it will be critical for these and other stakeholders to also contribute to the future development. All potential users have the opportunity to contribute to the future development via the GENESYS portal or emailing the Helpdesk (Genesys-Helpdesk@CGIAR.ORG).

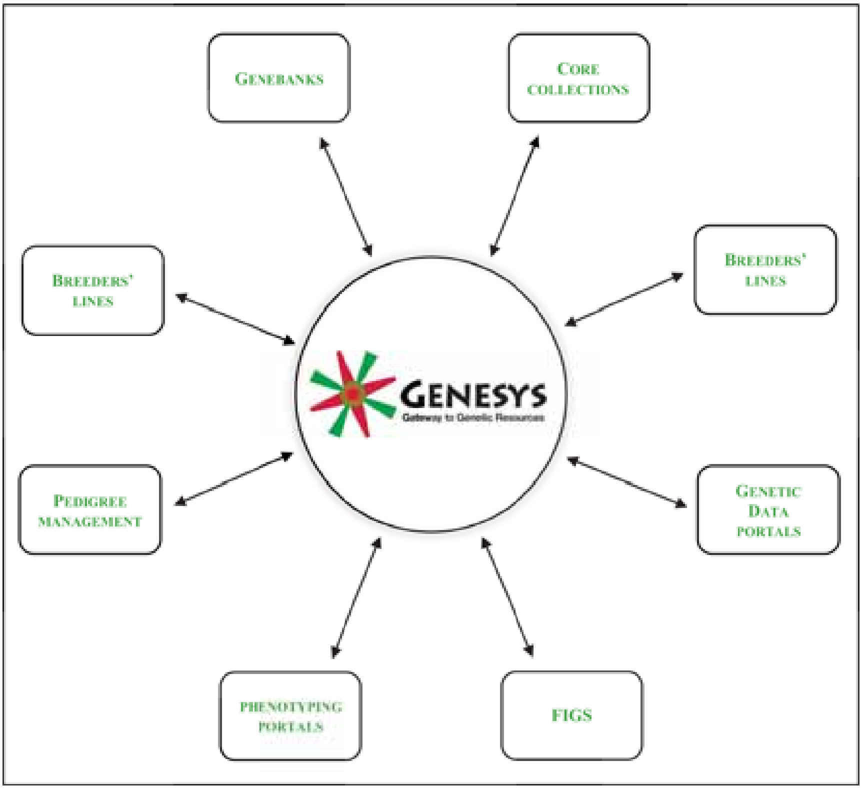

Figure 5. One schema depicting the pivotal role that Genesys could play in linking information from many diverse sources to the actual germplasm which breeders might request to use as a parent in their breeding programmes (FIGS = Focused Identification of Germplasm Strategy).

# 6 Understanding the relationships between the expression of adaptive traits and the environment in which these evolved

*Common sense is the knack of seeing things as they are, and doing things as they ought to be done*. C. E. Stowe.

## 6.1 Review of selecting subsets for discovery of novel genetic variation

There are numerous examples of discovering new sources of variation for adaptive biotic and abiotic traits, and sometimes new alleles, in germplasm chosen for evaluation on the basis of geographic origin.

Stodart *et al*. (2007) evaluated 250 bread wheat landrace accessions, originating from four geographic regions, for tolerance to aluminum under low pH conditions. Of the 35 accessions identified as tolerant, AFLP analysis indicated diverse origin, thus making them potentially useful to expand the diversity of genetic variation in breeding programmes.

In the case of a soil borne biotic stress in rice (*Oryza sativa* L.), Jia *et al*. (2011) found 52 of the 1,794 accessions in the USDA rice core collection (Yan *et al*., 2007) to be significantly more resistant to rice sheath blight (*Rhizoctonia solani)* at the 5% probability than the susceptible check. The USDA rice core collection was developed from the whole collection (ca. 18,500 accessions at the time) using the stratified random sampling method where geographic origin, known genetic characteristics and potential value to breeding are key factors determining inclusion (Brown, 1989a). A similar study was undertaken to identify new sources of resistance to bacterial leaf streak (*Xanthomonas* *translucens* pv. *undulosa*) from the winter wheat accessions of the USDA wheat core subset by Adhikari *et al.* (2011). In this case about 8% of the accessions evaluated were resistant and resistance was found to be significantly more likely in cultivars than landraces or breeding material.

It is apparent that the genetic diversity available for many adaptive traits in genebank accessions is extensive and can be found with relative ease, even when selecting material on the basis of geographic origin. In the cases cited here the germplasm identified to be of potential use in breeding programmes requires further studies to ascertain if the variation is novel or not.

At any one point in time our interest in a given adaptive trait can differ from it being crucial to identify new genetic variation to virtually no interest. It is often the rare alleles that need to be found and deployed quickly in response to new and unpredicted challenges to production. A new question emerges - is it likely that these rare alleles can be found as easily as the ‘new sources’ reported in the studies discussed here? If so, then there is no need to do anything other than select germplasm on the basis of geographic origin and a few other known characteristics before evaluating it for variation of the target trait. If not, then how can we more effectively and efficiently identify the germplasm that is most likely to posses the novel variation required?

This is where the original work of Vavilov and similar minded scientists point in a different direction; better understanding the relationship between environments and adaptive traits to choose candidate accessions for strategic evaluation. Vavilov’s observations gave him the knowledge to link the presence of certain traits with certain regions. Today we could review many studies that link adaptive traits to geographic regions, but do we understand why such relationships exist? Some of the parameters involved might simply be a result of exchange through human migration, dispersal of pollen or seed by insects, or other vectors. However, a significant proportion of the available (possibly unknown) genetic variation is the result of evolution in natural environments where additional factors are likely to have shaped the allelic variation within populations and individuals.

Because we are dealing with almost continuous variation that even the most skilled botanist or breeder is not likely to easily observe, the option of using other means to assist us in understanding these relationships presents itself – mathematics, statistics and associated modeling.

Considerable detail has already been provided on how some of the earlier FIGS subsets were developed. As experience with the FIGS approach has been accumulated, and more researchers have contributed, new methods to concentrate genetic variation (for specific traits) in smaller sized sets of accessions have been examined. The emphasis now moves to a different level where we seek to better understand the link between adaptive trait expression and the eco-geographic profile of the environment from where germplasm was collected.

## 6.2 Understanding the adaptive trait by environment relationship

There are a range of statistical methods that could help to better understand the relationship between the expression of a particular adaptive trait in the environment in which it evolved in tandem with local selection pressure(s).

Endresen *et al*. (2011) undertook a study to verify the FIGS approach to developing targeted subsets of accessions, using four classification methods, for two different biotic stresses in two crops – wheat and barley. In both cases *a priori* data was used to 'train' the statistical methods/models as to which would be the more likely environments in which the requisite genetic variation would evolve.

This study used geo-referenced landrace evaluation data made available from the USDA-ARS GRIN database for wheat stem rust (*Puccinia graminis* f.sp. *tritici)* and barley net blotch (*Pyrenophora teres* Drechs.). The two data sets were randomly divided into two subsets; two-thirds of the accessions and their observations were used to train the models whilst the other one-third was used to 'test' how precisely the model could classify the collection site environments into categories producing a certain reaction to the biotic stress. The algorithms employed to classify the environments were:

i) Linear discriminant analysis (LDA).
ii) Partial least squares discriminant analysis (PLS-DA).
iii) k-nearest neighbor (kNN).
iv) Soft independent method of class analogy (SIMCA).

Six separate experiments were carried out using different eco-climatic data resolution and different data stratification approaches involving eco-climatic datasets, taxonomy, experimental location, year of experiment and limiting the number of accessions per site.

The results indicated that neither wheat stem rust nor barley net blotch resistance were randomly distributed, but were linked to climatic parameters. Furthermore, it was concluded that the variables used to describe the collection sites can also be used to identify disease resistant landraces more effectively than accessions selected randomly. As would be expected, the year and location of the evaluation experiments, as well as the specific pathogen involved, were seen to influence the results. Some pathogens have become more widespread and important in recent times, possibly because of large scale farming or cultivation practices such as minimum tillage, which can influence pathogen distribution and survival between growing seasons.

The study also identified some limitations to the FIGS approach. These included the need to have a set of *a priori* data to 'train' the model, which might possibly be overcome by expert knowledge to identify the 'training' set. The requirement of having landrace accessions geo-referenced for environmental profiling may be conditional on the specific trait involved, for example, the identification of accessions to evaluate for tolerance to boron toxicity did not require geo-referencing (Mackay, 1986), but access to soil maps to identify other regions where a crop might have evolved with similar edaphic conditions to the target environment.

In another recent study to detect the linkages between a biotic stress (stem rust in wheat) and climatic parameters, Bari *et al*. (VII) look at using such links to concentrate desirable genetic variation from genebanks into small, more effectively exploitable subsets using the FIGS approach as well as investigating five additional modeling approaches to assist in the future deployment of FIGS. The climatic parameters used were:

1. Monthly potential evapo-transpiration
2. Monthly moisture index (ari)
3. Monthly precipitation cm
4. Monthly minimum temperature
5. Monthly maximum temperature

The five modeling techniques (including parametric, non-parametric and machine learning techniques) were investigated to quantify the link between the expression of the trait and a number of climatic parameters:

1. Principal component analysis (PCA)
2. Partial least square (PLS)
3. Random forests (RF)
4. Neural Networks (NN)
5. Support Vector Machines (SVM)

Each of these techniques (different to those used in the study of Endresen *et al.*, 2011) were individually 'tuned' to identify the parameters that provide the best predictive power before being compared. Three of the approaches (PCA, PLS and RF) used to quantify the assumed linkage were able to successfully classify collection sites as having resistant or susceptible accessions in 76% of the cases. The other two techniques (SVM and NN, improved the predictions by 1-2%.

In this case (VII), as was concluded in the study by Endresen *et al.* (2011), the distribution of resistance to stem rust in wheat landraces can be linked to climatic parameters associated with collecting sites. Spooner *et al.* (2009), in a study of potato wild relatives, found that they could only reliably predict resistance to six out of 32 pests and diseases using environmental variables. They did, however, mention factors which might have impeded their results, such as the resolution of the climate grids from which the variables were extracted.

Another observation made was that the predictability of the models increased as the number of variables increased (VII), an assertion supported by Stockwell (2007) in developing efficient ecological models. Furthermore, the choice of variable could influence predictability. It is reasonable to expect that different environmental variables, or combinations thereof, would provide higher predictive performance for different adaptive traits because the factors exerting selection pressure on a given abiotic stress would be quite different to those doing so for a biotic stress. For example variables linked to humidity could increase predictability for diseases such as stem rust because it plays a role in infection by the pathogen whilst edaphic variables could be expected to play a greater role in the case of mineral toxicities.

Modeling approaches, such as the ones discussed here, provide a predictive framework to quantify adaptive trait by environment relationships that can contribute to the more effective and efficient utilization of PGR held in *ex situ* genebanks. This ‘adaptive’ trait by environment relationship (T x E), in contrast to the more widely known genotype by environment (G x E) interaction, is more about how environmental variables (be they climatic, edaphic, biological or of another basis) shape the selection of certain genetic variation in individuals or populations, in preference to other variation, because it bestows upon them ‘better adaptability’ to reproduction and survival in that particular environment.

The underlying assumption is that novel genetic variation in landraces and CWRs will be a consequence of local selection pressures, in some cases including the involvement of man, and such variation is likely to be found where other functional variation has previously been identified.

# 7 Conclusions and recommendations

- The value of plant genetic resources, PGR, has been increasingly recognized over the past 100 plus years. Early researchers, such as N I Vavilov, recognized associations between the various expressions of plant adaptive traits and the eco-geographical regions from where they originated.
- There is still a lot of untapped genetic variation available in *ex situ* genebank accessions. Landraces and crop wild relatives, CWR, are especially anticipated to be extremely rich sources of novel variation.
- Whilst novel alleles have been identified and introgressed into modern cultivars over the past century or so, there has been no generally accepted way to more effectively and efficiently 'mine' *ex situ* genebanks for new genetic variation for targeted traits.
- There has been minimal research into how best to choose candidates for novel genetic variation since the work of Vavilov (1957) who stated "*The basic purpose of this work is to assist the selectionist in getting to know and choose correctly the starting material suitable to different areas …*".
- The core collection concept has value in that it aims to concentrate diverse variation for all traits in a smaller set of accessions.
- The Focused Identification of Germplasm Strategy (FIGS) approach builds on Vavilov's ideas and the core collection concept by using the link between adaptive traits and eco-geographical and/or environmental parameters to focus in on a specific trait and then concentrate genetic variation for this trait in a small subset of accessions.
- Information is a significant key to utilizing PGR more effectively and efficiently. Every situation where a user seeks novel variation for a specific trait should be considered as unique where the data and associated methods employed to identify a set of accessions 'rich' in the target variation will vary accordingly.

- The global plant genetic resources information system, GENESYS, provides ‘proof of concept’ that it is feasible to concentrate access to large quantities of dispersed and disparate data, and build custom queries across all data categories (passport, characterization, evaluation and environmental), via a single portal.
- Both ‘legacy’ and currently generated data can be shared via GENESYS to facilitate access and use of PGR, so a second phase of GENESYS needs to focus on capturing this data.
- Ideally GENESYS should also link to other information portals, for example systems publishing molecular data, breeders’ trial results etc to facilitate interoperable use of all data.
- Tools and utilities to utilize the data in the Genesys portal through a FIGS or similar approach are recommended for future development.
- Similarly, GENESYS data should be harvestable by other initiatives, such as the CGIAR Research Programme on Climate Change, to assist in identifying and deploying traits required for the development of ‘climate ready’ varieties.
- The use of various modeling techniques to identify linkages between evolutionary/domestication environments and the desirable expression of individual adaptive traits, are producing promising results in improving the effectiveness and efficiency with which novel variation can be identified and exploited.
- Further research is required to progress the understanding of such linkages;
- The process should become more effective and efficient as more data associated with *ex situ* genebank accessions becomes available; effective in terms of actually identifying novel genetic variation, and efficient in terms of doing so at a minimal cost.

# References

Adhikari, T.B., Hansen, J. M., Gurung, S., and Bonman, J. M. 2011. Identification of new sources of resistance in winter wheat to multiple strains of Xanthomonas translucens pv. undulosa. *Plant Disease* 95**,** 582-588.

Alercia, A.D., S.; Metz, T. 2001. FAO/IPGRI Multi-crop passport descriptors. Bioversity International.

Allem, A.C. 2001. Managing genebanks: seed base collection examined. *Genetic Resources and Crop Evolution,* 48**,** 321-328.

Bari, A., Martin, A., Boulouha, B., Barranco, D., Gonzalez-Andujar, J.L., Trujillo I., Ayad. G 2003. Image feature extraction combined with a neural networks approach for the identification of olive cultivars. *3rd IASTED International Conference on Visualization, Imaging and Image Processing.* ACTA Press, Benalmádena, Spain.

Bartish, G.I., Jeppsson, N., Bartish, I.V. & Nybom, H. 2000. Assessment of genetic diversity using RAPD analysis in a germplasm collection of sea buckthorn. *Agricultural and Food Science in Finland,* 9**,** 279-289.

Batchu, A.K., Zimmermann, D., Schulze-Lefert, P. & Koprek, T. 2006. Correlation between hordatine accumulation, environmental factors and genetic diversity in wild barley (Hordeum spontaneum C. Koch) accessions from the Near East Fertile Crescent. *Genetica,* 127**,** 87-99.

Bechere, E., Belay, G., Mitiku, D. & Merker, A. 1996. Phenotypic diversity of tetraploid wheat landraces from northern and north-central regions of Ethiopia. *Hereditas (Landskrona),* 124**,** 165-172.

Beharav, A. & Nevo, E. 2004. Variation in agronomically important traits in natural populations of wild emmer wheat, *Triticum dicoccoides,* in Israel. *Plant Genetic Resources: Characterization and Utilization,* 2**,** 81-84.

Bennett, E.E. 1968. Technical Conference on the Exploration, Utilization and Conservation of Plant Genetic Resources. *Technical Conference on the Exploration, Utilization and Conservation of Plant Genetic Resources.* FAO, Rome, Italy.

Bhullar, N., Street, K, Mackay, M, Yahiaoui, N and Keller, B 2009. Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. *Proceedings of the National Academy of Sciences,* 106**,** 9519-9524.

Bhullar, N., Mackay, M and Keller, B 2010. Genetic Diversity of the Pm3 Powdery Mildew Resistance Alleles in Wheat Gene Bank Accessions as Assessed by Molecular Markers. *Diversity,* 2**,** 768-786.

Bisht, I.S., Mahajan, R.K. & Patel, D.P. 1998. The use of characterisation data to establish the Indian mungbean core collection and assessment of genetic diversity. *Genetic Resources and Crop Evolution,* 45**,** 127-133.

Brown, A.H.D. 1989a. Core Collections - a Practical Approach to Genetic-Resources Management. *Genome,* 31**,** 818-824.

Brown, A.H.D. 1989b. The case for core collections. In Brown, A.H.D., Frankel, O.H., Marshall, D.R. & Williams, J.T. (Eds.) *The use of plant genetic resources.* Cambridge University Press, Cambridge.

Brown, A.H.D. & Spillane, C. 1999. Implementing core collections-principles, procedures, progress, problems and promise. In R.C. Johnson and T. Hodgkin, E. (Ed.) *Core collections for today and tomorrow* International Plant Genetic Resources Institute, Rome.

Brown, J.C., Pds 2008. *An introduction to plant breeding,* Blackwell Publishing, Ames.

Casler, M.D. 1995 Patterns of Variation in a Collection of Perennial Ryegrass Accessions. *Crop Science,* 35**,** 1169-1177.

Cocks, P.S. & Ehrman, T.A.M. 1987. The Geographic Origin of Frost Tolerance in Syrian Pasture Legumes. *Journal of Applied Ecology,* 24**,** 673-683.

De Pauw, E. 2002. *An agroecological exploration of the Arabian Peninsula,* An agroecological exploration of the Arabian Peninsula. International Center for Agricultural Research in the Dry Areas (ICARDA) Aleppo Syria: 2002. x + 77. many ref.

Diwan, N., Bauchan, G.R. & Mcintosh, M.S. 1994. A Core Collection for the United-States Annual Medicago Germplasm Collection. *Crop Science,* 34**,** 279-285.

Ehrman, T. & Cocks, P.S. 1996. Reproductive patterns in annual legume species on an aridity gradient. *Vegetatio,* 122**,** 47-59.

El Bouhssini, M., Street, K., Amri, A., Mackay, M., Ogbonnaya, F.C., Omran, A., Abdalla, O., Baum, M, Dabbous, A. and Rihawi, F. 2010. Sources of resistance in bread wheat to Russian wheat aphid (Diuraphis noxia) in Syria identified using the Focused Identification of Germplasm Strategy (FIGS). *Plant Breeding,* 130**,** 96-97.

El Bouhssini, M., Street, K., Joubi, A., Ibrahim, Z. & Rihawi, F. 2009. Sources of wheat resistance to Sunn pest, Eurygaster integriceps Puton, in Syria. *Genetic Resources and Crop Evolution,* 56**,** 1065-1069.

Ellner, S. & Shmida, A. 1981. Why Are Adaptations for Long-Range Seed Dispersal Rare in Desert Plants. *Oecologia,* 51**,** 133-144.

Endresen, D., Street, K., Mackay, M., Bari, A., De Pauw, E., Amri, A. & Yahyaoui, A. 2011. Sources of resistance to stem rust (Ug99) in bread wheat and durum wheat identified using focused identification of germplasm strategy (FIGS). (Submitted).

Epperson, B.K. 1990. Spatial patterns of genetic variation within plant populations. In Brown, A.H.D., Clegg, M.T., Kahler, A.L. & Weir, B.S. (Eds.) *Plant population genetics, breeding, and genetic resources.* Sinauer Associates Inc., Sunderland, Massachusetts.

FAO 2010. The Second Report on the State of the World’s Plant Genetic Resources for Food and Agriculture. FAO, Rome.

FAO 2011a. The dynamic tension between public and private plant breeding in Thailand. *FAO Plant Production and Protection Paper 208.*

FAO 2011b. The strategic role of plant breeding in Uruguay: analysis through an agricultural innovation system framework. *FAO Plant Production and Protection Paper 209.*

Fisher, G., Makowski, M. & Granat, J. 1999. AEZWIM An interactive multiple-criteria analysis tool for land resources appraisal. World Soil Resources Reports. Food and Agriculture Organization (FAO) Rome Italy: 1999. 87 90. 23 ref.

Frankel, O.H. 1977. Genetic resources. *Annals of the New York Academy of Sciences,* Vol 287**,** 332-344.

Frankel, O.H. 1984 Genetic perspectives of germplasm conservation. In Arber, W., Illmensee, K., Peacock, W.J. & Starlinger, P. (Eds.) *Genetic manipulation: impact on man and society.. Cambridge University Press, for ICSU Press, Cambridge, UK: 1984. 161-170. 19 ref.*

Frankel, O.H. & Bennett, E. 1970. Genetic Resources. In Frankel, O.H.A.B., E. (Ed.) *Genetic Resources in Plants - their Exploration and Conservation.* International Biological Programme, Glascow.

Fu, Y.B., Peterson, G.W., Williams, D., Richards, K.W. & Fetch, J.M. 2005. Patterns of AFLP variation in a core subset of cultivated hexaploid oat germplasm. *Theoretical and Applied Genetics,* 111**,** 530-539.

Furman, B.J. 2006. Methodology to establish a composite collection: case study in lentil. *Plant Genetic Resources: Characterization and Utilization,* 4**,** 2-12.

Glaszmann, J.C., Kilian, B., Upadhyaya, H.D. & Varshney, R.K. 2010. Accessing genetic diversity for crop improvement. *Current Opinion in Plant Biology,* 13**,** 167-173.

Harlan, J.R. 1992. *Crops and Man,* American Society Of Agronomy, Inc, Crop Science Society Of America, Inc, Madison.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* 25**,** 1965-1978.

Hodgkin, T., Brown, A.H.D., Hintum, T.J.L.V. & Morales, E.A.V. (Eds.) 1995. *Core collections of plant genetic resources,* John Wiley & Sons Chichester UK.

Holbrook, C.C., Anderson, W.F. & Pittman, R.N. 1993. Selection of a Core Collection from the United-States Germplasm Collection of Peanut. *Crop Science,* 33**,** 859-861.

Hu, J., Zhu, J. & Xu, H.M. 2000. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theoretical and Applied Genetics,* 101**,** 264-268.

Hutchinson, M. & Corbett, J. 1995. Spatial interpolation of climate data using thin plate smoothing splines. In Fao (Ed.) *Coordination and harmonization of databases and software for agroclimatic applications.* FAO, Rome.

Hutchinson, M.F. 2000. *ANUSPLIN Version 4.1. User Guide,* The Australian National University Centre for Resource and Environmental Studies, Canberra.

Jia, L., Yan, W., Agrama, H.A., Yeater, K., Li, X., Hu, B., Moldenhauer, K., Mcclung, A. & Wu, D. 2011. Searching for Germplasm Resistant to Sheath Blight from the USDA Rice Core Collection. *Crop Science,* 51**,** 1507-1517.

Kaur, N., Street, K., Mackay, M., Yahiaoui, N. & Keller, B. 2008. Allele mining and sequence diversity at the wheat powdery mildew resistance locus Pm3. In Appels, R., Eastwood, R., Lagudah, E., Langridge, P., Mackay, M., Mcintyre, L. & Sharp, P. (Eds.) *11th International Wheat Genetics Symposium.* Sydney University Press, Brisbane.

Khoury, C., Laliberte, B. & Guarino, L. 2010. Trends in ex situ conservation of plant genetic resources: a review of global crop and regional conservation strategies. *Genetic Resources and Crop Evolution,* 57**,** 625-639.

Kingdon, J., Agwanda, B., Kinnaird, M., O'Brien, T., Holland, C., Gheysens, T., Boulet-Audet, M. & Vollrath, F. 2011. A poisonous surprise under the coat of the African crested rat. *Proceedings of the Royal Society B: Biological Sciences.*

Kingsbury, N. 2009. *Hybrid. The History and Science of Plant Breeding,* University of Chicago Press, Chicago

Knott, D. 1961. The inheritance of rust resistance VI. The transfer of stem rust resistance from Agropyron elongatum to common wheat. *Canadian Journal of Plant Science,* 41**,** 109-123.

Lee, R. & Devore, I.E. 1968. *Man the Hunter,* Aldine Publishing, Chicago.

Lehmann, C. 1984. Germplasm evaluation at Gatersleben: the relationship between genebank and breeder. In Holden, J. & Williams, J. (Eds.) *Crop Genetic Resources: Conservation and evaluation.* Allen and Unwin, London.

Lin, W., Bradshaw, A.D. & Thurman, D.A. 1975. The potential for evolution of heavy metal tolerance in plants. 3. The rapid evolution of copper tolerance in Agrostis stolonifera. *Heredity,* 34**,** 165-187.

Liviero, L., Maestri, E., Gulli, M., Nevo, E. & Marmiroli, N. 2002. Ecogeographic adaptation and genetic variation in wild barley, Application of molecular markers targeted to environmentally regulated genes. *Genetic Resources and Crop Evolution,* 49**,** 133-144.

Macindoe, S. & Walkden Brown, C. 1968. *Wheat breeding and varieties in Australia,* Department of Agriculture, Sydney.

Mackay, M. 1986. Utilizing Wheat Genetic Resources in Australia. In Mclean, R. (Ed.) *5th Wheat Breeding Assembly.* Wheat Breeding Society of Australia (Inc.), Perth/Merredin.

Mackay, M.C. 1990. Strategic planning for effective evaluation of plant germplasm. In Srivastava, J.P. & Damania, A.B. (Eds.) *Wheat genetic resources: meeting diverse needs.* John Wiley & Sons, Chichester.

Mackay, M.C. 1995. One core collection or many? In Hodgkin T., Brown, A.H.D., Van Hintum T.J.L., Morales E.A.V. (Eds.) *Core Collections of Plant Genetic Resources.* John Wiley & Sons Ltd., Chichester, U.K. and New York, USA, Brasila, Brazil.

Malosetti, M. & Abadie, T. 2001. Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphological traits. *Genetic Resources and Crop Evolution,* 48**,** 381-390.

Mann, C. 1997. Cashing in on Seed Banks' Novel Genes. *Science,* 277**,** 1042.

Marshall, D.R. & Brown, A.H.D. 1981. Wheat genetic resources. Wheat science - today and tomorrow.. Cambridge University Press. Cambridge UK: 1981. 21-40. 42 ref.

Martynov, S. & Dobrotvorskaya, T. 2006. Genealogical Analysis of Diversity of Russian Winter Wheat Cultivars (Triticum aestivum L.). *Genetic Resources and Crop Evolution,* 53**,** 379-386.

McKhann, H.I., Camilleri, C., Berard, A., Bataillon, T., David, J.L., Reboud, X., Le Corre, V., Caloustian, C., Gut, I.G. & Brunel, D. 2004. Nested core collections maximizing genetic diversity in Arabidopsis thaliana. *Plant Journal,* 38**,** 193-202.

Morrell, P. & Clegg, M. 2007. Genetic evidence for a second domestication of Barley (Hordeum vulgare) east of the Fertile Crescent. *Proceedings of the National Academy of Sciences,* 104**,** 3289-3294.

Moss, J., Ramamatha Rao, V. & Gibbons, R. 1989. Evaluatingt he germplasm of groundnut (Arachis hypogaea) and wild Aranchis species at ICRISAT. In Brown, A.H.D., Frankel, O.H., Marshall, D.R. & Williams, J.T. (Eds.) *The use of plant genetic resources.* Cambridge University Press, Cambridge.

Nawar, M. & Mackay, M. 2010. A global information portal to facilitate and promote accessibility and rational utilization of ex situ plant genetic resources for food and agriculture. *IAALD XIIIth World Congress.* Montpellier.

Ortiz, R., Ruiz-Tapia, E.N. & Mujica-Sanchez, A. 1998. Sampling strategy for a core collection of Peruvian quinoa germplasm. *Theoretical and Applied Genetics,* 96**,** 475-483.

Peace, C. 2011. DNA Information Facilitates Utilization Of Tree Fruit Genebanks. W292: Genomics of Gene Banks. *Plant & Animal Genomes XIX Conference.* San Diego.

Pederson, G.A., Fairbrother, T.E. & Greene, S.L. 1996. Cyanogenesis and climatic relationships in US white clover germplasm collection and core subset. *Crop Science,* 36**,** 427-433.

Peeters, J. & Williams, J. 1984. Towards better use of genebanks with special reference to information. *Plant Genetic Resources Newsletter,* 60**,** 22-32.

Peleg, Z., Fahima, T., Abbo, S., Krugman, T., Nevo, E., Yakir, D. & Saranga, Y. 2005. Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. *Plant Cell and Environment,* 28**,** 176-191.

Qualset, C. & Shands, H. 2005. *Safegarding the future of U.S. agriculture: The need to conserve threatened collections of crop diversity worldwide,* University of California, Division of Agriculture and Natural Resources Conservation Program, Davis.

Rana, R.S. & Kochhar, S. 1996. Core subsets of base collections and priorities of national programmes: Indian perspective. *Genetic Resources and Crop Evolution,* 43**,** 423-428.

Spagnoletti Zeuli, P. & Qualset, C. 1995. The durum wheat core collection and the plant breeder. In Hodgkin T., Brown, A.H.D., Van Hintum T.J.L., Morales E.A.V. (Eds.) *Core Collections of Plant Genetic Resources.* John Wiley & Sons Ltd., Chichester, U.K. and New York, USA, Brasila, Brazil.

Spieth, P.T. 1979. Environmental heterogeneity: a problem of contradictory selection pressures, gene flow, and local polymorphism. *American Naturalist,* 113**,** 247-260.

Spooner, D.M., Jansky, S.H. & Simon, R. 2009. Tests of taxonomic and biogeographic predictivity: Resistance to disease and insect pests in wild relatives of cultivated potato. *Crop Science,* 49**,** 1367-1376.

Stockwell, D.R.B. 2007. *Niche Modeling: Predictions from Statistical Distributions,* Chapman & Hall/CRC, Boca Raton.

Stodart, B., Raman, H., Coombes, N. & Mackay, M. 2007. Evaluating landraces of bread wheat Triticum aestivum L. for tolerance to aluminium under low pH conditions. *Genetic Resources and Crop Evolution,* 54**,** 759-766.

Tanksley, S.D. & Mccouch, S.R. 1997. Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. *Science,* 277**,** 1063-1066.

Tohme, J., Gonzalez, D.O., Beebe, S. & Duque, M.C. 1996. AFLP analysis of gene pools of a wild bean core collection. *Crop Science,* 36**,** 1375-1384.

Upadhyaya, H.D. & Ortiz, R. 2001. A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theoretical and Applied Genetics,* 102**,** 1292-1298.

Upadhyaya, H.D., Gowda, C.L.L., Buhariwalla, H.K. & Crouch, J.H. 2006. Efficient use of crop germplasm resources: identifying useful germplasm for crop improvement through core and mini-core collections and molecular marker approaches. *Plant Genetic Resources,* 4(1), 25–35.

Vanhintum, T.J.L. & Elings, A. 1991. Assessment of Glutenin and Phenotypic Diversity of Syrian Durum-Wheat Landraces in Relation to Their Geographical Origin. *Euphytica,* 55**,** 209-215.

Vaughan, D. & Jackson, M. 1995. The core as a guide to the whole collection. In Hodgkin T., Brown, A.H.D., Van Hintum T.J.L., Morales E.A.V. (Eds.) *Core Collections of Plant Genetic Resources.* John Wiley & Sons Ltd., Chichester, U.K. and New York, USA, Brasila, Brazil.

Vavilov, N. 1957. *Agroecological Survey of the Main Field Crops,* The Academy of Sciences of the USSR, Moscow.

Vavilov, N.I. 1926. Tzentry proiskhozhdeniya kulturnykh rastenii. (Studies on the origin of cultivated plants). . *Trudy Byuro prikl. Bot. (Bull. of Applied Botany),* 16**,** 139-248.

Vavilov, N.I. 1992. The Phytogeographical Basis for Plant Breeding. In Dorofeyev, V.F. (Ed.) *Origin and Geography of Cultivated Plants.* Cambridge University Press.

Yan, W.G., Rutger, J.N., Bryant, R.J., Bockelman, H.E., Fjellstrom, R.G., Chen, M.H., Tai, T.H. & Mcclung, A.M. 2007. Development and evaluation of a core subset of the USDA rice (Oryza sativa L.) germplasm collection. *Crop Science,* 47**,** 869–878.

# Acknowledgements

My journey in the world of plant genetic resources began with obtaining an Australian Wheat Industry Research Council scholarship to undertake an intensive post graduate course in plant breeding at the Waite Institute campus of The University of Adelaide some 30 years ago. It was here that I was exposed to evaluating germplasm for tolerance to the cereal cyst nematode, largely under the guidance of wheat breeder Tony Rathjen, whom I acknowledge for ‘planting the seed’. Tony kept nourishing this seed by always bringing students to the Australian Winter Cereals Collection (AWCC) for some hands on exposure to genetic resources until I left in 2008.

My role as curator of the AWCC brought me into contact with many institutions and individuals. Providing breeders, including prebreeders, with the germplasm they required was the baseline for my appointment to this position, in 1984, to succeed Ken Symes. Ken saw genebanking as a service to plant improvement and shared this vision with me for which I am grateful.

Numerous cereal breeders in Australia helped me to develop an interest in trying to identify the ‘best bet’ accessions for their breeding programmes; I thank them all.

Otto Frankel, Tony Brown and Don Marshall visited the AWCC to share their ideas about managing germplasm effectively. They introduced me to the core collection which sparked some ideas for further refinement when considering how to give the breeders the variation they sought in a smaller package.

The Grains Research and Development Corporation (GRDC) provided continual support for the AWCC from 1990 onwards, allowing me to take on work of a more global application as the idea of strategic planning for efficient evaluation took hold.

This brought me into contact with the late Bent Skovmand with whom I continued to ponder how to find that elusive diversity more effectively. Bent was to become my second supervisor until his untimely passing in 2007. In fact, during the 1990s I interacted with many people I like to think of as colleagues and I acknowledge all the contributions they made to a developing germplasm use startegy, based on the convergence of so many individual perspectives.

In 2001 Paul Brennan, ex wheat breeder and then consultant to the GRDC, met with Ken Street and me in Coffs Harbour to draft a collaborative project involving the AWCC, ICARDA and VIR to exploit bread wheat landraces in a targeted way that has lasted until the present time. I thank both Ken and Paul for their contributions to developing those fledgling ideas into something that became know as FIGS. Without Ken's mammoth efforts FIGS would probably still just be an idea – thanks, mate! The many contributions of our other colleagues at the AWCC, ICARDA and VIR (Jan Konopka, Olga Mitrofanova, Evgeny Zuev, Sergey Alexanian and Greg Grimes to name just a few) are also gratefully acknowledged.

In 2005 I was fortunate to obtain the support of the GRDC to spend a six month sabbatical at SLU where this doctoral venture began. Again, having the opportunity to interact with numerous people at the then SLU Crop Science Department, the Nordic Gene Bank and other European institutions sparked further opportunities to build upon shared ideas; Dag Terje Filip Endresen and Helmut Knüpffer deserve my thanks for their various contributions and support.

Special acknowledgement is offered to Roland von Bothmer, my primary supervisor and one of nature's gentlemen. Roland not only provided me with the opportunity to undertake this work, he also provided moral support to an absent student who, during the past six years, was often 'burning the candle at both ends' with challenging workplace demands. Roland was always understanding, tolerant and able to elicit an extra effort when it was required – thank you, Roland!

I'd like to express my gratitude to my family who have all supported me with immeasurable patience, understanding and encouragement, especially to Bernadette my soulmate.

Finally, to those who have contributed to my journey and are not mentioned here, please accept my sincere appreciation and gratitude.