

# Challenges of integrating conventional breeding and biotechnology: a personal view!

John W Snape

John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UK. [www.jic.bbsrc.ac.uk](http://www.jic.bbsrc.ac.uk); [john.snape@bbsrc.ac.uk](mailto:john.snape@bbsrc.ac.uk)

## Abstract

The development of 'scientific' breeding methodologies in the first half of the 20<sup>th</sup> Century was based partly on the application of genetic theory following the rediscovery of Mendel's Laws, but mostly on developing procedures for systematic phenotypic selection. Although, new technologies in tissue culture, pathology, statistics and quantitative genetics contributed to advances in the second half of the 20<sup>th</sup> Century, these were limited. The last decade has seen an explosion of genetics and genomics information, yet this also has yet to make a major impact, and there are major challenges to be overcome in translating and integrating this information into plant breeding practice. A foremost challenge is exploiting the detailed molecular information from cloning genes and sequencing model genomes. We need methodologies for utilizing information from *Arabidopsis* and rice in plant breeding, particularly in cloning useful agronomic genes, and allele mining in germplasm collections. At the administrative level, we face challenges in creating the right balance of investment in model systems and crops, and in bridging the interface between them. There are, however, already significant successes emerging in the application of genetic information in plant breeding, particularly through the use of marker-assisted selection. At present this is almost exclusively used for major genes, and this needs to be extended into QTL controlling complex traits to greatly increase the impact. Coupled with this are methodologies for high-through-put genotyping, and the development of array of 'functional' markers. It should also not be forgotten that 'low tech' approaches can have major impacts on plant breeding practice, as exemplified by the application of doubled haploid systems and alien gene transfers.

## Media Summary

Although huge progress has been made in accumulating genetic and genomics technologies and insights into the biology of 'model' plants, a 'technology gap' exists in translating this knowledge into tools that plant breeders can use to produce improved crop varieties.

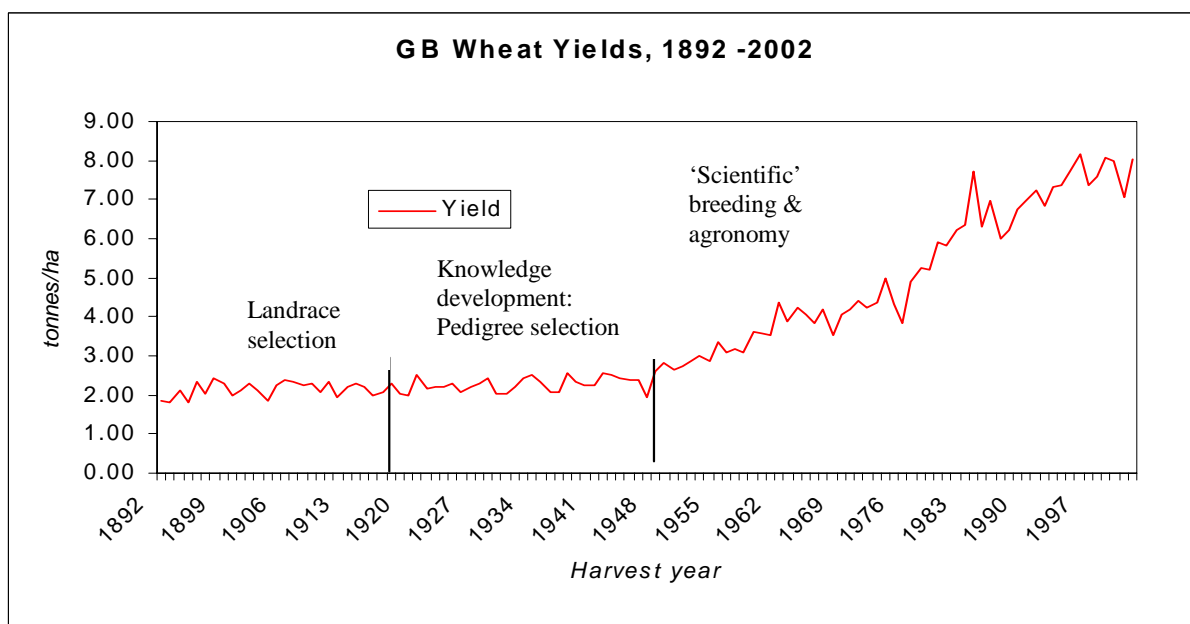
## Keywords

Genomics, marker-assisted selection, crop-model transitions, technology gap

## Introduction :

### *The historic background to plant breeding and biotechnology*

The domestication of crops was, of course, the earliest form of plant breeding, and this produced crops that were suitable for human and animal consumption and the practices of early agriculture. However, arguably, for several hundred years until the beginning of the 20<sup>th</sup> century, crop improvement did not make great advances in terms of improvements in yields (in particular) and crop quality, although sustainable levels of production were generally achievable. However, famine and economic ruin was always around the corner, due particularly to disease pressure, as witnessed by the Irish potato famine in the 1840s and the downy mildew of vines in France in the 1870s. Historical records would suggest that variety turnover was rather static and relied on population selection within landraces, and that dramatic advances in productivity or quality were not obtained. For example, records of wheat yields in the UK (Figure 1) go back to the 1890s and show that up until the 1920's wheat yields in farmer's fields were static at about two tonnes per hectare.



**Figure 1. Historic average wheat yields in the UK, 1892-2002**

The re-discovery of Mendel's Laws, and the realization that directed plant breeding could bring about systematic improvement, accelerated progress in many crops, particular the main arable crops. For example, again in wheat (Figure 1), the work of Biffin at the newly founded Plant Breeding Institute in Cambridge (founded in 1912), exemplified the start of the process of systematic yield improvement in released UK wheat varieties, although up until the early 1950's yields had only increased up to around three tonnes. The post-war years brought about an accelerated rate of improvement in many arable crops, especially in Developed Countries, due to the enhanced profitability to farmers brought about by a subsidy system, with corresponding research into improvements in agronomy and the application of the scientific method to breeding. In particular, this included the introduction of new genetic variation from a wide variety of sources, caused by the Worldwide exchange of germplasm from the systematic development and exploitation of germplasm collections. It is arguable if these advances can be described as the application of 'biotechnology', although it can be said that this was the start of the impact of 'biotechnology' in its broadest terms in plant breeding.

These advances in crop productivity, and also quality, were particularly prevalent in Europe with the development of the European Union. However, the development of the CGIAR system of institutions following the 'Green Revolution' in wheat and rice production also indicated that advances could be achieved in many different crops adapted to many different agro-ecosystems. I also believe that these advances created a psychological climate amongst plant breeders and geneticists that 'the sky's the limit' in improving all of our major crops, and this provided the confidence and research impetus to justify research and its applications. This was also stimulated by the economic returns that could be obtained from systematic improvement, and this created an acceleration of interest by Industry, often taking plant breeding out of the public domain into the private domain, as exemplified by the privatisation of the Plant Breeding Institute in Cambridge, UK, in 1988. This raises the point that the impact of biotechnology in plant breeding also cuts across the socio-economic climate of varietal production. This is well illustrated by the 'GM debate'. For most of the 20<sup>th</sup> Century, in both Developed and Developing Countries both private and public institutions participated and co-existed in varietal production and variety release. More recently, the balance has turned in favour of large industry controlling most of the varietal production for major crops, which has raised new serious issues concerning intellectual property, germplasm exchange, etc, although these are out of the scope of this paper.

What was the role of biotechnology in these dramatic changes in crop productivity? This, of course, depends on the definition of 'biotechnology' used, but it has to be said, that, probably, it was little! Throughout the last century and into the present, most of the selection for improved productivity, at least, has been based on empirical selection. The whole process is often described as being the plant breeders

'art' because, essentially most of the selection for improved genetic type has been based on empirical selection of observed phenotypes and statistical analysis of observational plot data, rather than for the specific, desirable underlying genetic factors, the genes known to improve those traits. Thus, it has been a subjective exercise, based on the experience and subjective skill of a plant breeder to choose parents for designed crosses or populations, and to select out improved individuals or populations in the progenies of artificial or controlled natural crosses. Indeed, to date, in my experience, very few breeders can quantify the genetic advances they have made in terms of known genes for any **complex** trait, although there are obvious major genes for crop architecture, disease resistance or quality that have been systematically used. However, the real prizes are to be obtained in the systematic improvement of complex traits involved in productivity, quality and biotic and abiotic stresses.

*What are the challenges of integrating biotechnology in plant breeding?*

So in addressing the challenges of applying biotechnology to plant breeding, a major question to be considered (and one that is repeatedly asked by many plant breeders that I know!) is 'I am still making genetic advance – why do I need biotechnology?' The answer is fairly obvious; the great advances currently being made in our genetic and molecular understanding of the biology of our major crops in particular, and through synteny studies into minor crops – in genetics, genomics, proteomics and metabolomics – all the 'omics', suggest that these technologies can and should make an impact! Plant breeders do and will perceive the benefits when they are tangible. The major challenge is to translate the enormous advances in understanding into technologies that plant breeders can apply routinely. This paper will try to discuss the challenges of delivering these! However, the paper will stay away from 'biotechnology' as concerns the technologies and the applications of genetic modification in plant breeding, and focus on the ways in which current developments in plant genetics and genomics can influence 'conventional' plant breeding.

What do plant breeders need? Plant breeding, is, in its essence, a simple process that can be described euphemistically as 'Cross the best with the best, select the best, and hope for the best'! Thus, plant breeders need sources of genetic variation; tools for its manipulation; and tools for validating that they have achieved their objectives in putting together and identifying new adapted gene complexes. Added to this are techniques to speed up the whole process.

Previous to the development of the 'genomics age' there were major gaps in our understanding of the genetical basis of phenotypic variation in crops, and we did not know how to move forward to fill these gaps. We lacked the genetical tools to understand complex traits other than at an overall, quantitative genetics description of the presence of different types of genetical variation – additive, dominance, epistasis, and their relative proportions. We lacked any coherent information on the underlying, individual genetic factors and their structure and modes of actions. This is now changing and for most (all?) of the major crops, and many of the minor crops, we have the genetic markers, maps and genomic tools to enable genetical analysis with a precision never previously possible. Although we still have some way to go for a complete tool-kit, these tools can already be applied for large-scale QTL analysis of yield and yield components to provide targets for marker-assisted selection and gene cloning for transgenic modification. In parallel, there are studies of the underlying physiological consequences of genetic variation to define specific and general adaptation, and yield potential '*per se*', so that varieties can be 'fine-tuned' to their target environment. However, we are still a long way off a mechanistic description in terms of gene action at the proteomics and metabolomics levels, although this may not be necessary for practical plant breeding. But the first major challenge is still translating genetics and genomics information into practice.

*The technology gap between genetical understanding and its exploitation*

Did biotechnology develop because of plant breeding needs or independent of it? A bit of both, although curiosity driven scientific advance has always been way ahead of practical plant breeding, and, in my view, there has always been a consistent, and quite large time gap before discovery and the realization of its application. There is a continuing argument in many research institutions, including my own, of whether plant research, including that on crops, should be curiosity driven, or driven by the needs of the market. What should the balance of investment be between research on model organisms, such as *Arabidopsis*, and crop plants critical to the economy? How far is research on model crops directly applicable to plant breeding? So, we have an administrative challenge in creating the right mix and

balance between investment into fundamental discovery into plant processes and the market driven needs of crop improvement. However, the ‘success percentage’ of scientific discovery in models, ‘faster, easier, tools there’ attitude, often drives awards from peer-review granting bodies, rather than drivers for solving a plant breeding problem or developing new paradigms for plant breeding.

Thus, we have created a major issue in the context of accelerated scientific advances in plant science over the last ten years and its application in crop improvement. We have created a ‘technology gap’ where our understanding and resource development capacity in plant genomics and biological understanding has exceeded our ability to apply it to practical plant breeding problems and situations. Also, arguably, there is much less investment in making the progress necessary in technology transfer, such that the gap is opening wider, rather than closing.

The last ten years, and particularly the last five years, have seen an explosion in the amount of research in plant genetics and genomics and consequently the information concerning the fundamental structure of plant genomes. This is continuing. For example, it is estimated that the US National Science Foundation has spent US\$350 million over the last five years and intends to spend about US\$1300 million over the next five. The EU as a whole probably spends at least 80 million Euro on plant genomics. Other programmes are in action in India, Japan, China, Australia and other Asia-Pacific countries. Combined with this, there has been large investments by Industry which parallel (and probably duplicate) much of the investment in the public sector. However, much of this investment, to date, has been based on plant models, notably *Arabidopsis* and rice. We now have complete genome sequences for these plants, and large collections of genomics resources – ESTs, insertion libraries, mutant populations, mapping populations, germplasm resources. Combined with this there are the development of technologies and resources in proteomics, metabolomics, and importantly, bioinformatics, to pull all of the information together. The ultimate objective has been very nicely encapsulated in the US programme as ‘*to understand structure and function of all plant genes at levels from molecular to the organismal and to interactions within ecosystems*’ (National Science Foundation 2003). It is obvious when talking to plant breeders that they currently regard this information as being hugely distanced from the practicalities of plant breeding. This is a gap that has to be filled.

On the positive side, recently, there has been a notable change in the balance of funding between models and crop research in several countries, driven by three things. First, the maturing of the phases of the development of genomic resources in models, so that the efforts can be focused on understanding plant processes using the available tools. Secondly, the realization that crop systems are now tractable to in-depth genomic and genetic analysis. It is now appreciated that it is possible to clone genes, for example in the large genome cereals, through a variety of approaches, including the development of large insert libraries, and even to move to whole genome sequencing as the costs of this come down and are practicable. However, this is still no a trivial or cheap exercise. Finally, and importantly, is the realization from comparative studies to date, that models are not going to provide a ‘one size fits all’ solution to getting at orthologous genes of interest in crops. This has to be combined with a realization by the ‘academic’ plant science community that there is a real intellectual challenge in genetic and genomics research in crops, and a need for funding initiatives to drive research in crop-model transitions.

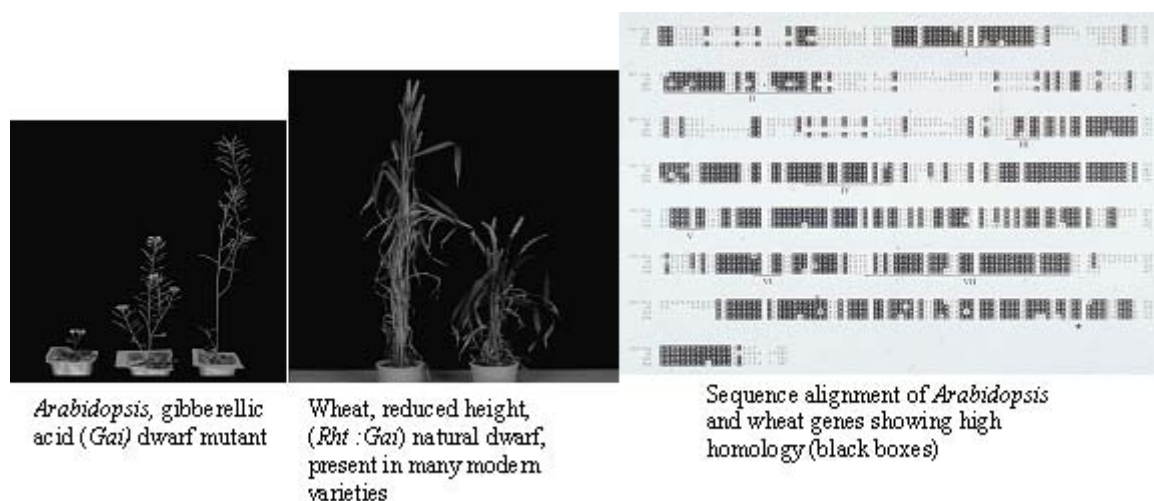
#### *Model to crop translations*

One of the immediate challenges facing the application of biotechnology to plant breeding is to exploit the vast amount of genomics information, such as DNA sequence, that has emerged from the studies of model organisms, not only plants, but also organisms as diverse as bacteria, the nematode worm *Caenorhabditis elegans*, *Drosophila*, and even humans. However, the primary model system for fundamental plant science research is *Arabidopsis*, for the well documented reasons of small genome size, ease of growth, and short generation time. The renaissance of *Arabidopsis* research started at the end of the 1980s (there was a first flurry in the 1970s, and before, for example the mutant work by George Redei at the University of Missouri, USA). The last fifteen years or so has seen incredible advances in the development of resources for the study of *Arabidopsis*, and from these our understanding of plant processes. Intellectually, it has been a hugely rewarding and productive period. The production of the complete genome sequence in 2001 was a historic landmark in plant biology. But can we, and how are we, to use this information in practical plant breeding?

The primary usage, of course, will be to identify plant genes controlling important plant processes so that we can identify novel allelic variation in crop plants for marker-assisted selection (to be discussed below) or for transgenic modification. Both of these processes rely on the ability to understand and to use the model system to understand the basis of genetic variation at the molecular level, and the role of the gene product in a particular plant process.

Initially, this can be done at the *in silico* level, exploiting the existing data on the known structure and function of a particular sequence from previously cloning the gene or discovering homology in *Arabidopsis* with already known function in another species. This information is then combined with the experimental approach to generate homologues of the gene using genomic DNA of the target crop species through the design of degenerate primers. The success rate of this approach will be related to the closeness of the species to *Arabidopsis*, on the one hand, and the degree of conservation of the plant system and its components, across evolutionary time, on the other. For example, this has been very successful in pulling out homologues related to quality characteristics in Brassicas, the nearest crop species group to *Arabidopsis*. As an illustration, Li and Queros (2002) cloned a major aliphatic glucosinolate gene, *BoGSL-ELONG*, essential for manipulation of the aliphatic GSL profile in *Brassica oleraceae* using information from the *Arabidopsis* sequence database. This approach has merit and considerable application, particularly when combined with allele mining for novel sequences in germplasm collections. There are now an accumulating number of examples in Brassicas where this approach can reveal valuable genes for conventional and transgenic approaches to agronomic (eg flowering time) and quality manipulation. However, it is not surprising that this approach works well in Brassicas, as these are of course, the closest crop group to *Arabidopsis*. If it does not work here, it will work much less with other, more distantly related, species!

However, it is a point of intense debate amongst plant geneticists as to how far *Arabidopsis* will contribute to gene discovery and the facile cloning of important agronomic genes in major crops, particularly the cereals. An example, and one of the very few (only?) to date, of how *Arabidopsis* can be used to establish the function of important genes in cereals can be illustrated by the analysis of the famous *Rht* dwarfing genes in wheat (Peng et al. 1999). Most people are aware of the importance of dwarfing genes in wheat and rice, which greatly contributed to the Green Revolution in the 1970s by enabling the development of short statured plants that could tolerate high inputs of fertilizers without lodging. The wheat genes originated as natural mutants from Japan, but despite their widespread usage, nobody had been able to interpret their mechanism of height reduction, although a biochemical observation was made that these plants were insensitive to applied gibberellic acid (GA), which in normal plants, causes etiolation when applied in a liquid feed. Following mutation experiments, a very similar mutant was identified in *Arabidopsis*, and because of the similarities in phenotype (see Figure 2), this became the subject of intensive molecular study



**Figure 2. Cloning *Rht* genes in wheat from sequence homology with *Arabidopsis***

of the mechanism. These studies first led to the gene in *Arabidopsis* being isolated and studied to relate its structure to function. It was found that the gene is a receptor for GA, which acts as a plant hormone to stimulate cell growth and stem elongation. However, the mutated version acts as an inefficient receptor, so that plants cannot respond adequately to the internally produced GA during normal development. The question then arose of whether the mechanism was the same in dwarf phenotypes of crops such as wheat. To understand this necessitated isolating the gene in wheat. To do this, first, the *Arabidopsis* sequence was used to probe DNA libraries in rice and a similar gene sequence was identified. The rice gene was then used to clone the wheat gene from DNA of normal and dwarf wheat plants. These wheat sequences were studied and compared, and the same genetic mutation found in dwarf *Arabidopsis* was found in the DNA from dwarf wheat, suggesting the same mechanism. This was confirmed when the wheat DNA sequence was shown to genetically map to the same location on chromosomes 4B and 4D known to carry the dwarfing genes in wheat. This research now allows the interesting possibility of 'designing' new dwarfing genes for wheat, but also using the *Arabidopsis* and wheat genes to genetically modify other crops to dwarf phenotypes.

The underlying tenet in such work is that by identifying a mutant for a trait/process in *Arabidopsis*, the same phenotypic mutant in a crop plant will have the same functional lesion. However, this approach does not always work, because although there is sequence similarity and suggested conservation over time, evolution has produced a divergence of function. This is shown by a comparison of genes controlling flowering between wheat and *Arabidopsis*, which appear, for the most part with respect to vernalization and photoperiod, to act differently. For example, at my Institute, the John Innes Centre, Griffiths et al. (2003) isolated eight homologues in barley and identified sixteen sequences in rice of the CONSTNS (CO) gene of *Arabidopsis* using the *Arabidopsis* CO sequence as a probe, or primers derived from it. This gene has an important role in the regulation of photoperiod response in *Arabidopsis*. Although one homologue was a counterpart of *OsA (Hd1)* a major determinant of photoperiod response in rice, none corresponded to major genes already mapped in barley or wheat. This illustrates that although, often, there is homology between sequence from *Arabidopsis* and cereal genes, the function is always not immediately obvious, and does not correspond to genes already identified as controlling the same plant process in the crop plant.

The difficulties of using *Arabidopsis* to determine function of an important plant process can be also illustrated by studies to clone the major vernalization gene *Vrn-A1* in wheat, which controls whether plants will have a winter or spring habit, critically important for eco-geographical adaptation, and this gene has been a major target for many years. The *Vrn-A1* gene was mapped and cloned using an extensive (and expensive) map based-cloning approach in the diploid wheat, *Triticum monococcum*. On cloning, the most likely candidate identified in wheat turned out, in fact, to be a gene homologous to an *Arabidopsis* gene, *Ap1* (Yan et al. 2003). However, *Ap1* homologues had previously been cloned from wheat (Murai et al. 1998; 2003), and suggested as candidates for *Vrn-A1*, but the lack of a functional link, particularly in discovering sequence variation between winter and spring alleles that related to function, did not provide proof that they were indeed the *Vrn* genes. So the unreliability of being able to immediately assume function creates a quandary when trying to interpret whether a cereal homologue of an *Arabidopsis* gene has the same or a very similar function. An extensive series of complementation tests through transformation, or finding a loss of function mutant in the crop species by mutation techniques, or gene silencing by RNA interference, is needed to do so. At present, such experiments are expensive and time consuming.

Another difficulty of model to crop translation using *Arabidopsis* is the present limitation on gene description in *Arabidopsis*. To date, still only about 50% of *Arabidopsis* genes can be ascribed function. Work is continuing apace to discover and 'prove' a function to the unknown sequences. But, this will take a great deal of time and effort over the coming years, and also more resource development including the development of comprehensive libraries of 'gene knockout' lines where specific unknown genes are silenced by mutation, so that the change in phenotype can be related to absence of function.

Nevertheless, *Arabidopsis* is a tool that plant breeding cannot ignore given the vast quality of exciting biological information that is being produced. Studies of *Arabidopsis* are being excellent in highlighting the complexity of plant processes and illustrating the necessity to move from genomic studies into proteomics, so that having identified a gene product, we can relate its structure to its function. Gene

discovery via *Arabidopsis* will allow opportunities for allele mining for conventional breeding strategies and new genes for GM strategies. The challenge is now seeing, and also being honest, as to how far *Arabidopsis* genes will be useful in identifying useful crop plant homologues, and in facilitating the process. So, how are we to facilitate efficient model to crop translations? I believe that one starting point is more effort devoted to doing so in a serious way. A commitment is needed from scientists working on models to bridge the gap to crops, and a realization by crop scientists of the information and tools available in *Arabidopsis*. However, inevitably for a particular trait it will be 'horses for courses' and in some circumstance *Arabidopsis* can short-circuit the process of gene discovery in crops, whilst in others, the mapping-cloning approaches within the crop will be needed. One need to facilitate the process would be a high throughput system for generating and testing crop homologues of known *Arabidopsis* genes, without which, using information for this model system is limited in its scope. Finally, model to crop translations have to be balanced by a distribution of funding to make such transitions possible.

#### *'Horses for courses: 'new' models for crops*

An additional consideration in model to crop translations is the difference in gene number which is being found between *Arabidopsis* and other sequenced plant genomes, such as rice. Whole genome sequence comparison has shown that there are approximately 26K genes in *Arabidopsis* versus 40K genes in rice. Thus cereals, and probably other crops, contain many more genes which *Arabidopsis* does not have, so that the analogy of function must inevitably break down. Additionally, many genes present in *Arabidopsis* are not represented by homologues in rice. Thus, it has been recognized that there is a need for additional crop models. Already, there has been considerable additional investment in rice sequencing and draft sequences of both indica and japonica rice varieties were published in 2002 (Goff et al. 2002; Yu et al. 2002). The systematic sequencing of the rice genome by a worldwide consortium is also well advanced, and much of this has already been published. Studies using this sequence, particularly in Japan, are being highly productive in identifying and cloning genes for specific agronomic traits in rice and identifying homologues in wheat, barley, maize, and minor cereals, and relating these to function. To exploit this investment in cereals will require that many laboratories that are interested in the application of genomics in their mandate crops need to re-divert their investment from *Arabidopsis* into rice and other models. However, in some respects this is like turning around an aircraft carrier! Not easily done in a small time period, particularly since *Arabidopsis* research is still being highly productive in terms of fundamental biological discovery. Also, it will continue to be an excellent model for many dicots, particularly, the Brassicas.

The alternative strategy to gene discovery and gene isolation in crops is, of course, to produce the resources and approaches need to identify genes and clone genes in the crops of interest themselves. This is starting to be pursued. Already in the USA there are projects to sequence the maize genome and, recently, an international consortium was established to start sequencing of the wheat genome. None of this is likely to be cheap, and it has been estimated that it will take \$70 million US to fully sequence the wheat genome. For these large genomes, new and novel strategies need to be developed, such as sample sequencing only the 'gene rich' regions of the genome using various methodologies. Pilot projects suggest that such enrichment strategies can work and enable the sequencing of 70%+ of the genes. Most strategies for genome sequencing take one variety of the species, usually the one most commonly used in genetical studies, Nipponbare in rice, B73 in maize, Chinese Spring in wheat, etc. However, interesting studies are now showing that there may be sequence variation between individual genotypes, as has been recently demonstrated already in maize (Fu and Dooner 2002). Thus, to discover the totality of gene content in a species, it may be necessary to sequence several different, possibly diverse, genotypes. Obviously, unless the cost of sequencing genomes comes down dramatically, this is a pipe-dream! However, there are claims that future technologies will be able to sequence a genome for a \$1K US!

It is apparent from the above discussion that we are in the middle of an exciting time for genomics research and its application in plant breeding. However, the speed and ease of translation of advances through to application have clearly been over emphasized, although the opportunities are surely to be found. So, the challenge is also one of restraint in not over-exaggerating the promise and the speed of living up to that promise. The history of biotechnology in plant breeding research is arguably littered with 'over-egged' research (Simmonds 1999)!

### *The promise and realities of marker-assisted selection*

The current most obvious and pertinent use of genetics and genomics information in conventional plant breeding is its application for marker-assisted selection. This is an ongoing success of biotechnology in plant breeding, where plant selection is done on the basis of genotype in the laboratory, rather than phenotype in the glasshouse or field. The advantages have been extensively discussed (eg Koebner and Summers 2003). They include speeding up generation time to release, the obviation of a selection environment, particularly in cases where the environment may not be expressed in every field season, for example, with biotic and abiotic stresses, and efficient selection for complex traits of low heritability. MAS holds huge promise for selection of advanced lines in the laboratory on the basis of DNA profile, rather than on phenotype in the glasshouse or field. We are, undoubtedly, only at the start of MAS in most plant breeding programmes, and its current applications are limited in extent. Even in the major crops a major limitation is the shortage of target traits and genes for these. Nearly all of the MAS presently carried out is for major genes for biotic and abiotic stresses, and a few for quality attributes. For example, in barley (Langridge and Barr 2003) only 27 loci covering 18 traits, mostly biotic and abiotic stress targets, are being subjected to MAS. Very few QTL for complex traits are tagged and being applied for MAS at the present time.

Successful MAS requires three components for successful implementation, gene discovery, marker implementation, and validation. The gene discovery phase is probably the easiest phase. In virtually all major and many minor crops we now have the molecular marker technologies, the genetic maps, and the statistical methodologies, for very successful major gene and QTL discovery for any trait worthy of study. Possibly, the only limitation of this stage is the 'phenomics', the ability to identify and score pertinent traits. At present there is only limited engagement of plant physiologists, for example, in defining the component traits for complex phenotypes. For example, a yield QTL may relate to a photosynthetic characteristic, rather than a yield component *per se*, and we need to dissect out these subtleties. There is also a recognized problem in this respect in having trained personnel able to recognize phenotypes! This may sound trivial, but is a real problem in European research Institutes which have a surfeit of molecular biologists, but a huge deficit of phenotypers! My Institute is a case in point, where people trained in the science of whole plant physiology and plant breeding research are difficult to find and attract!

The next, more difficult, phases are marker implementation and validation. Discovering a major gene or QTL and an associated marker is still a long way away from its efficient selection, and this is often where the greatest technology gap exists in applying MAS. It is not an easy task to find closely linked and diagnostic markers for traits of interest. It sometimes occurs by chance that markers associated with a particular trait are diagnostic for that trait in different crosses of the same species. This requires, of course, strict linkage disequilibrium between the marker polymorphism and 'useful' alleles controlling the trait. An excellent example of this over the last two decades and currently, and one of the first uses of marker-assisted selection in plant breeding, is that of using polymorphism identified by SDS-PAGE to select the high-molecular weight glutenins in wheat. Work by Payne et al. (1983) showed a functional relationship between allele polymorphism for the storage proteins, and variation in bread-making quality that they impart. Through evaluating the extensive allelic variation at these loci, and showing its association with bread-making quality, it was possible to develop an 'allelic score' and hence predict and assemble the bread-making quality of a variety from its allele profile on SDS-PAGE. This has revolutionized selection for bread-making quality in many breeding programmes around the World. This same example also illustrates the direction in which marker-assisted selection must now move to be more efficient – to the development of 'functional' markers, where allelic variation for the marker is directly associated with phenotype. We now have the opportunity, for example, to use the extensive EST information becoming available in crop species to associate candidate genes with function, and then allelic variation at these candidate genes with function. The most suitable marker system to enable such associations has to be single nucleotide polymorphisms, SNPs. The discovery of SNPs using EST information from a range of varieties can allow haplotypes to be assembled which can become diagnostic for trait variation. One such method for the rapid discovery of SNPs is 'genotype partitioning' developed at my Institute (Zhu 2002). Although SNP discovery is presently laborious and expensive, it must be the way forward to combine extensive phenotyping with diagnostic markers. Once we have widespread SNPs, it will also enable the more efficient use of the World's extensive crop germplasm collections through 'allele mining'. This will enable identified alleles, which have been shown to be 'useful', to be used for indirect selection. Even in cases where candidate haplotypes are not the genes controlling the



trait, if they are closely linked and exhibit strong linkage disequilibrium with the trait, the haplotype can be diagnostic, since the chromosome segment involved should be identical by descent.

In most current situations aimed at gene discovery of targets for MAS, functional markers are not available, and closely linked 'anonymous' markers have to be 'converted' in some way, usually to a PCR format, to be diagnostic. This is usually achieved by designing specific PCR primers from sequence information of the closest convenient marker to create, for example, a sequence tagged site. This is usually effective when the gene-marker combination is introduced from an exotic source into an adapted gene pool, creating quasi linkage disequilibrium. However, it is often quite difficult to extend the diagnostic marker to other allelic variants or to other gene pools. This is when a 'validation' phase has to be introduced into any MAS programme, which greatly slows down the adoption of the technology by different breeding programmes. In the end, there has to be a concerted effort to use sequence information to facilitate the finding of functional markers and hence to obviate the need for extensive testing and validation of every trait-marker combination. Unless this is achieved, MAS will only be useful with characterized major genes and not extend to QTLs for complex traits.

Finally, there are also a number of technical problems that have to be overcome before we enter an era of 'true', directed, 'designer' genotype selection, particularly the costs of high-throughput assays. Although there are high-throughput platforms using microsatellites with fluorescent primers, for example, these are limited by cost for the numbers that plant breeders would like to put through. So we have the problem of reducing the cost per data point, relative to the genetic gains that can be made by conventional selection. We are not yet at the stage of having high-throughput analytical platforms at a price, and throughput, worth the investment in large scale MAS in most crops, unless the targets give a real return on investment. This is why current targets are restricted to abiotic and biotic stresses, which cause large yield or quality reductions, rather than desirable, more marginal traits in terms of the economic returns.

#### *Using 'steam biotechnology' for plant breeding*

Another, but perhaps more obscure, and under-rated, challenge to the application of biotechnology in plant breeding is the concentration of resources in looking for the development of 'high tech' rather than 'low tech' solutions to plant breeding problems. Today, much of the research investment is in large scale genomics and genetic engineering programmes, whilst much less resource is put into the development of 'steam biotechnology'. In this category I place systems that do not require complex equipment, protocols, consumables or highly trained personnel. A good example is tissue culture systems which speed up breeding programmes, a critical component of competitiveness and responding to new biotic and abiotic challenges. The most obvious and successful example in recent years has been the development of doubled haploid systems which can short circuit the pedigree system of breeding in many inbred crops by going from heterozygous F<sub>1</sub>s to completely homozygous progenies in a single generation. The advantages are well catalogued (see Snape 1989, for example). In the UK, most of the high yielding winter wheats currently coming through the testing system are doubled haploids produced using the maize cross system (Laurie and Bennett 1988) pioneered at the Plant Breeding Institute in Cambridge. Canola is another crop where anther culture is used routinely as a breeding tool. Yet, in many crops, there are no adequate doubled haploid systems, and very little investment in developing these. Even with the doubled haploid systems currently being used, neither anther culture nor interspecific hybridisation, are particularly efficient in terms of resources and costs per doubled haploid produced. This could be overcome if more efficient high-throughput systems were developed, particularly workable microspore systems, so that hundreds of plants can be produced with small resources. However, to my knowledge there is little current investment in this area despite the obvious advantages.

Another example of 'steam biotechnology' is the use of cytogenetics methods for alien gene transfer into crops. It is well known that the wild relatives of our cultivated species contain a wealth of genetic diversity for 'useful' genes, particularly for disease and stress resistances. These can be transferred by the creation of interspecific hybrids and subsequent backcrossing with cytogenetical manipulation of recombination or irradiation to induce translocations and to stabilize the chromosome number and structure. This has been well practiced in cereals, particularly wheat, where several of the current disease resistance genes are derived from alien sources. However, current investment in research tends to look for more 'creative' solutions through the use of genetic engineering, rather than exploiting the genetic diversity that already exists in germplasm collections and wild relatives.

In the same vein, the creation of completely new crops through the production of new synthetic allopolyploids is possible by interspecific and intergeneric hybridisation, as has been practiced for over sixty years, for example in the development of triticale, now a common cereal for forage or grain uses. This simple technology offers opportunities to create new and novel crops for particular environments or end-uses. A recent example from work of my colleagues at the John Innes Centre was the creation of a salt tolerant cereal *Tritipyrum*, (Forster et al. 1987), by combining the productivity of durum wheat with the salt tolerance of *Agropyron junceum*. However, there has been very little interest in exploiting this for marginal areas devastated by saline toxicity through continuous irrigation, whilst high tech solutions are undergoing heavy investment. Indeed, the science of cytogenetics is undergoing a decline in investment and skilled personnel, being regarded, at least by funding agencies, as passé, an old fashioned science, despite the obvious utility. There is a serious risk that the skills will be lost to plant breeding as the major practitioners of 'traditional' science' retire and the lab, glasshouse and field skills passed down in laboratories for much of the last century will be lost to the corporate memory.

## Conclusions

Enormous progress has been made in the last decade in progressing our fundamental understanding of plant biology through the development of new ideas and technologies in genetics and the 'omics' (genomics, transcriptomics, metabolomics) in model organisms, particularly, but not exclusively, *Arabidopsis*. This is providing enormous insight into how plant processes are controlled, eg flowering, development, disease responses, abiotic stress responses, product formation, so that we are developing expert knowledge as to how plants function. However, our capacity to translate the technologies and knowledge into solutions for practical plant breeding has been, and is, limited. The challenge is to bridge this gap, not only scientifically but administratively, by providing incentives to scientists to devote their intellect to make crop-model transitions. Alongside this, we need also to develop the technological resources within crops themselves to produce high-through-put marker technologies for marker assisted selection, and facile gene cloning. It also needs to be acknowledged that 'low-tech' applications and solutions are available to many plant breeding problems, providing that funding can be carefully channelled to their development without competition from more up-stream research. The resurgence of interest, particularly in developed countries in 'public good plant breeding' can facilitate this. Finally, we have to be careful that many of the scientific skills of use to plant breeders are not lost to the corporate memory. We need to maintain the infrastructures and training to maintain these so that the 'intellectual space' between fundamental plant science and plant breeding application is populated.

## References

- Forster B P, Gorham J, and Miller T E (1987). Salt tolerance of an amphiploid between *Triticum aestivum* and *Agropyron junceum*. *Plant Breeding*, 98,1-8.
- Fu H and Dooner H K (2002). Intraspecific violation of genetic colinearity and its implications in maize. *Proceedings National Academy of Sciences USA* 99, 9573-9578.
- Goff S A, Ricke D, Lang T, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange B M, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun W, Chen L, Cooper B, Park S, Wood T C, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller R M, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T and Oliphant A (2002). A Draft sequence of the Rice Genome (*Oryza sativa* L. ssp, *japonica*). *Science* 296, 92-105.
- Griffiths S, R P Dunford, Coupland G and Laurie D A (2003). The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. *Plant Physiology* 131, 1855-1867.
- Koebner RMDK and Summers RW (2003). 21<sup>st</sup> Century wheat breeding : plot selection or plate detection? *Trends in Biotechnology* 21, 59-63.
- Langridge P and Barr A (2003) Preface to 'Better Barley Faster: the Role of Marker Assisted Selection'. *Australian Journal of Agricultural Research* 54, pp. i-iv.
- Laurie D A and Bennett M D (1988). The production of haploid wheat plants from wheat x maize crosses. *Theor Appl Genet* 73,393-397.

- Li G Y and Queros C F (2002). Genetic analysis, expression and molecular characterization of BoGSL-ELONG, a major gene involved in the aliphatic glucosinolate pathway of Brassica species. *Genetics* 162, 1937-1943.
- Murai K, Muria R, Takumi S and Ogihara Y (1998). Cloning and characterization of cDNAs corresponding to the wheat MADS box genes. *Proceedings of the 9<sup>th</sup> International Wheat Genetics Symposium* pp.89-94. (University of Saskatchewan Press).
- Murai K, Miyamae M, Kato H, Takumi S and Ogihara Y (2003). A MADS-box Gene Controlling Heading Time in Common Wheat. *Proceedings of Tenth International Wheat Genetics Symposium, Paestum, 2003* pp. 189-192. (Istituto Sperimentale per la Cerealicoltura).
- National Science foundation (2003). National Plant Genome Initiative Progress Report ([www.ostp.gov/NSTC/html/npgi2003/executive.htm](http://www.ostp.gov/NSTC/html/npgi2003/executive.htm)).
- Payne P I, Holt L M, Thompson R D, Bartels D, Harberd N P, Harris P A and Law C N (1983). The high molecular weight subunits of glutenin: classical genetics, molecular genetics and the relationship of bread-making quality. *Proceedings of the 6<sup>th</sup> International Wheat Genetics Symposium* pp. 827-834, (Kyoto Japan).
- Peng J, Richards D E, Hartley N M, Murphy G P, Devos K M, Flintham J E, Beales J, Fish L J, Worland A J, Pelica F, Sudhakar D, Christou P, Snape J W, Gale M D, and Harberd N P (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400, 256-261.
- Simmonds N W (1999). Agricultural research revolutionized. *TAA Newsletter* 19(2), 36–39.
- Snape J W (1989). Doubled haploid breeding: Theoretical basis and practical applications. *Review of advances in Plant Biotechnology 1985-88: 2nd International Symposium on Genetic Manipulation in Crops*, pp. 19-30. (CIMMYT, Mexico D F and IRRI, Manila, Philippines).
- Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T and Dubcovsky J (2003). Positional Cloning of the Wheat Vernalization Gene *VRN1*. *PNAS* 100, 6263-6268.
- Yu J, Hu S, Wang J, Wong G K, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu W, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L and Yang H (2002). A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *indica*). *Science* 296, 79-92.
- Zhu J (2002). Genome Partitioning Technology. UK Patent Application: 6092456.