The contribution of genomic approaches to resolving genotype x environment interactions in *Brassica* crops

Graham J. King
Southern Cross Plant Science, Southern Cross University, Lismore, NSW 2480, Australia,
Email: graham.king@scu.edu.au

ABSTRACT

Geneticists and breeders have conventionally relied on multi-site experiments and structured populations to partition phenotypic variation into component traits, where variation can then be assigned to genotype and environment, with recent efforts focused on resolution of quantitative trait loci (QTL) and identification of underlying genes. For *Brassica* genomes, which are relatively large and complex, this process requires access to interconnected sets of platform experimental resources, including ongoing determination of the complete chromosomal DNA sequences. The *B. rapa* genome sequence has recently been used to provide valuable information for a number of studies. In coming years, the ability to navigate between the various genome sequences and trait genetics will contribute to understanding Genotype x Environment (GxE) interactions at the molecular level and benefit pre-breeding activities. Exploitation of the genome sequence for trait resolution is benefiting from recent integration of genetic maps for *Brassica napus*, as well as development of advanced recombinant and diversity populations. There is particular value in the combined screening of diversity collections and QTL analysis which can demonstrate the extent of variation and genetic basis for traits as diverse as oil composition and mineral nutrition. The availability of genome sequences provide the opportunity to understand GxE interactions at the molecular level, using genetical genomic techniques such as expression QTL (eQTL) analysis. Although the role that epigenetic regulation plays in GxE interactions is currently poorly understood, we are now starting to determine the distribution of heritable epigenetic marks superimposed on *Brassica* genomes, and to explore the scope for epigenetic intervention as contribution to crop improvement. The need for integrated data management that allows navigation from trait to underlying genes is discussed.

Key words: Brassica, GxE, Genomics

INTRODUCTION

Regional consequences of global climate change increase the need for crops that are resilient and productive in different environments. Reduced land and water availability also places a premium on understanding the relationships between variation in soil composition, crop yield and end-product quality. Geneticists and breeders have conventionally relied on multi-site experiments and structured populations to partition phenotypic variation into component traits, where variation can then be assigned to genotype and environment, with recent efforts focused on resolution of quantitative trait loci (QTL) and identification of underlying genes.

Progress in characterising genes which underlie many agronomic traits has been limited by the relatively large size and complexity of *Brassica* crop genomes. However, whole genome sequencing and anchoring to existing genetic maps is enhancing our ability to resolve trait loci, with recent advances such as expression QTL (eQTL) analysis providing greater insight into the functions and relationships between genes underlying specific traits.

Here I present an outline of recent studies focused on increasing genetic and genomic resolution of chromosomal regions affecting specific traits, and place this in the context of available or induced genetic diversity. The need for re-usable experimental resource platforms is discussed, together with the need for more concerted efforts to understand Genotype x Environment interactions, and the benefits of integrating accumulated and expensive datasets.

RESULTS

The genome of canola is complex, resulting from a series of ancient genome duplications and a more recent inter-specific hybridisation that gave rise to the 19 chromosomes of the

amphidiploid *Brassica napus*. The pattern and relationships between the triplicated segments of each constituent diploid (the 'A' genome of *B. rapa* and the 'C' genome of *B. oleracea*) has now been resolved, firstly at genetic level, and more recently in more detail at the level of genomic DNA sequence. We have recently generated a high resolution integrated genetic map for *B. napus* based on sequence-tagged DNA markers, including many microsatellites, that aligns the maps for reference segregating populations generated in the UK, Canada and France (Wang et al., 2011). This confirmed the pattern of conserved collinearity blocks and also enabled direct comparison with the chromosomal sequences of the model experimental plant Arabidopsis and the *B. rapa* 'A' genome. Information on a wider range of genetic maps and associated molecular markers has also been collated and is available online at http://www.brassica.info/resource/maps.php.

The multinational effort to sequence *Brassica* genomes has benefited from the rapid advances in high throughput sequencing technologies over the past 5 years. This has resulted firstly in the completion and publication of the *B. rapa* genome sequence (The *Brassica rapa* Genome Sequencing Project Consortium, 2011). Sequence data for additional genotypes of *B. rapa* have been released in the UK and USA. At present there are efforts to sequence the C genome of *B. oleracea* (two projects), and several 'reference' sequencing projects focused on *B. napus*. A summary of these projects is available at http://www.brassica.info/resource/sequencing.php. As with many 'published genomes' the available *Brassica* sequences are not complete, but do form a very valuable draft platform that is available to the R&D community for ongoing validation, comparison, improvement and detailed functional annotation. The latter effort is greatly enhanced by the close relationship of Arabidopsis within the same plant family.

The *B. rapa* genome sequence has already been used to provide valuable information for a number of studies (Wang et al., 2011; Hammond et al., 2011). In coming years, the ability to navigate between the various genome sequences and trait genetics will contribute to understanding Genotype x Environment interactions and benefit pre-breeding activities. This will require access to a range of persistent and well described segregating and diversity populations.

Advances in the performance and adaptability of *Brassica* crops relies in part on understanding and harnessing the allelic diversity available within the relevant genepools. A number of research groups around the world have made use of long-term *ex situ* genetic resource collections to compile core collections representing diversity within *B. napus*. In the UK we pioneered the development of Diversity Fixed Foundation Sets (DFFS), defined as "an informative set of genetically fixed lines representing a structured sampling of diversity across a genepool". The advantage of generating fixed homozygous lines (either doubled haploid or recurrent inbred) is that they may be regarded as being immortal and therefore have great value for assessing traits in replicated trials across different environments and occasions. The persistence of such populations also contributes to the cumulative acquisition of data by different research groups over time, and allows a greater understanding of trait interactions. This approach has been taken in the UK within the Oilseed Rape Genetic Improvement Network (OREGIN, http://www.oregin.info/resources/trials.php), where the research community and breeders have been assessing diversity trials for a wide range of *B. napus* germplasm. There is also scope for canola breeding and pre-breeding studies to harness the variation present within the constituent diploid genepools, again making use of molecular and phenotypic assessment of DFFS or core collections. This can be informed by the continued interest in understanding the polyphyletic origins of the amphidploid *B. napus*. To date we have characterised the distribution of chloroplast haplotypes and nuclear DNA markers within diploid and amphidiploid diversity sets to provide some clarification of the relationships and geographic distribution of potential progenitors in different centres of genetic diversity (Allender & King, 2010). Although this endeavour requires further effort and integration of data, it has the potential to identify additional diploid material harnessing alleles that contribute to stable amphidiploid *B. napus*.

Earlier studies that made use of diversity collections have demonstrated the considerable variation available for traits as diverse as oil composition (Barker et al., 2007) and mineral uptake efficiency (Broadley et al., 2008; Hammond et al., 2009). Both these studies combined diversity studies with identification of QTL associated with the same traits, detected from segregation in reference bi-parental mapping populations. This approach is of particular value for placing trait variation available within the genepool in the context of the underlying genetic

complexity and GxE interactions. In many situations we have found much of the trait variation observed across diversity collections can also be 'unlocked' through transgressive segregation in segregating populations. In the latter situation the variation arises from the combinatorial effects of alternative alleles at each locus present within the homozygous parent lines.

In order to make full use of available genome sequence for trait characterisation, and so identify underlying genes, it is essential to resolve QTL to short well-defined chromosomal regions. This requires moving beyond conventional segregating populations of homozygous lines (doubled haploid or recombinant inbred), preferably by developing series of near-isogenic or substitution lines that enable either genome-wide or targeted locus resolution. In many cases such populations have a valuable life-time for many years or decades, as they may be used for analysis of a wide range of traits that often display sufficient transgressive segregation compared with parental values. The availability of genome sequences also add considerable value to other platform resources such as EMS-mutated populations that may be used for reverse-genetic screening (Stephenson et al., 2010).

Further information contributing to understanding the functional basis of QTL variation associated with crop phenotypes in different environments may be obtained from advanced genetical genomics approaches. Expression QTL (eQTL) analysis makes use of whole-genome transcriptional data, either using transcriptional micro-arrays or high throughput sequencing. We developed a comprehensive design for an Affymetrix GeneChip™ Exon array based on 135,000 *Brassica* unigene sequences (Love et al., 2010), and have recently used an earlier transcriptional array platform for a successful eQTL analysis in *Brassica rapa* (Hammond et al., 2011). This analysis involved comparing the level of transcription for individual genes across a segregating population in contrasting environments. By screening each gene for an association of its pattern of expression across all chromosomes, it is possible to identify significant eQTL. The resultant analysis provides information on the contribution of genes in response to the environmental variable (in this case phosphate). Comparison of the position of significant eQTL with the reference genome sequence allows categorisation of loci into those that map to the same position as the underlying gene (*cis*-eQTL) and those that affect the transcription of genes at other positions in the genome (*trans*-eQTL). Where a single *trans*-eQTL in the genome affects the transcription of genes at many other loci , this indicates the potential presence of a major regulator having large-scale pleiotropic effects.

Epigenetic mechanisms have been defined as those that involve heritable changes in DNA other than changes in nucleotide sequence. Superimposed on eukaryote chromosomal DNA is a series of epigenetic marks that can provide considerable agility in terms of modulating gene expression, ontogeny, and response to the environment. These 'epigenetic' marks affect chromatin structure, and include cytosine methylation of DNA and/or modifications to histone proteins. In plants the pattern of epigenetic marks are typically, but not always reset through meiosis. It is now apparent that a wide range of plant developmental processes are affected to some extent by stably inherited epigenetic modifications.

The role that epigenetic regulation plays in Genotype by Environment (GxE) interactions is poorly understood, although it appears to be a normal and essential mechanism for co-ordinating eukaryote genome activity. It is currently unknown how much inadvertent selection of epiallelic variation has contributed to yield increases. We have started to determine the extent and distribution of epiallelic variation in crop genomes, and demonstrated stable inheritance through multiple meioses. There is scope to harness epigenetic variation for crop improvement (King et al., 2010), particularly within complex crop genomes such as *Brassica* spp. Where it is clear that specific epialleles contribute to enhanced phenotypic traits, this information will add value to existing marker-assisted selection. A number of approaches are available for epigenetic intervention, including stochastic modulation of DNA methylation, followed by forward or reverse selection of epialleles. In the longer term there will be a requirement to establish optimal strategies to ensure stable retention of desirable epi-alleles within breeding material, and also to develop new techniques for targeted epigenetic modification.

The rapid increase in available data arising from combined genome and trait studies, and the accumulated long-term investments in genetics R&D requires considerable effort to be devoted to integration and sharing of data. We have taken a step-wise approach to accumulating reference information for *Brassica* species and storing this in data registries that clearly define plant genetic resources such as diversity and segregating populations, experimental lines and accessions, detailed information relating to phenotypic traits and

experimental trials, together with genetic maps and markers. The CropStoreDB database (www.cropstoredb.org) addresses the first stage in dataset curation and dissemination, by providing a data model and tools that reflect the relationships understood and manipulated by researchers and breeders. The design philosophy for CropStoreDB took into account the requirement to provide an explicit relationship between database tables and the entities that are commonly manipulated by plant geneticists. Providing a means to explain such relationships visually is proving to be beneficial in encouraging those who generate datasets to be responsible for, and involved in, the curation and data validation process. This approach is appropriate in many institutions or projects where there is insufficient investment to carry the overhead of developing and modifying sophisticated data-checking software, and where reference datasets are generated over long periods of time. These data are progressively being integrated with the reference genomic sequence, and associated datasets and bioinformatic tools and interfaces have been developed to aid navigation from trait to gene and *vice versa*.

ACKNOWLEDGEMENTS

Much of the work reported here was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) and Department for Environment Food and Rural Affairs (Defra). Colleagues who contributed to this work are co-authors of the papers and websites cited.

REFERENCES

Allender C, King GJ (2010) Clarifying the Origins of the Amphiploid Species *Brassica napus* L. using Chloroplast and Nuclear Molecular Markers. BMC Plant Biology 10:54.

Barker GC, Larson TR, Graham IA, Lynn JR, King GJ (2007) Novel Insights into Seed Fatty Acid Synthesis and Modification Pathways from Genetic Diversity and QTL Analysis of the *Brassica* C genome. Plant Physiology. 144: 1827-1842.

Broadley MR, Hammond JP, King GJ, Astley D, Bowen HC, Meacham MC, Mead A, Pink DAC, Teakle GR, Hayden RM, Spracklen WP, White PJ (2008) Leaf-calcium (Ca) and magnesium (Mg) concentrations differ between subtaxa, are highly heritable, and associate with potentially pleiotropic loci in *Brassica oleracea.* Plant Physiology 146:1707-1720.

Hammond JP, Broadley MR, White PJ, King GJ, Bowen HC, Hayden R, Meacham MC, Mead A, Overs T, Spracklen WP, Greenwood DJ (2009) Shoot yield drives phosphorus use efficiency in *Brassica oleracea* and correlates with root architecture traits**.** J Exp. Bot. 60:1953-1968

Hammond JP, Mayes S, Bowen HC, Graham NS, Hayden RMm Love CG, Spracklen WP, Wang J, Welham SJ, White PJ, King GJ, Broadley MR. (2011) Novel regulatory hotspots control plant gene expression under varying soil phosphorus (P) supply. Plant Physiology.

King GJ, Amoah S, Kurup S (2010) Exploring and Exploiting Epigenetic Variation in Crop Plants. Genome 53:856-868.

Love CG, Graham NS, Ó Lochlainn S, Bowen HC, May ST, White PJ, Broadley MR, Hammond JP, King GJ (2010) A *Brassica* Exon Array for Whole-Transcript Gene Expression Profiling. PLOS One 5(9): e12812.

Stephenson P, Baker D, Girin T, Perez A, Amoah S, King GJ, Østergaard L (2010) A rich TILLING resource for studying gene function in *Brassica rapa*. BMC Plant Biology 10:62

The *Brassica rapa* Genome Sequencing Project Consortium (2011) The genome of the mesopolyploid crop species *Brassica rapa.* Nature Genetics (in press).

Wang J, Lydiate DJ, Parkin IAP, Falentin C, Delourme R, Carion P, King GJ (2011) Integration of Linkage Maps for the Amphidiploid *Brassica napus* and Comparative Mapping with *Arabidopsis* and *Brassica rapa.* BMC Genomics 12:101. doi:10.1186/1471-2164-12-101.