Associative expression and systems analysis of complex traits in oilseed rape / canola

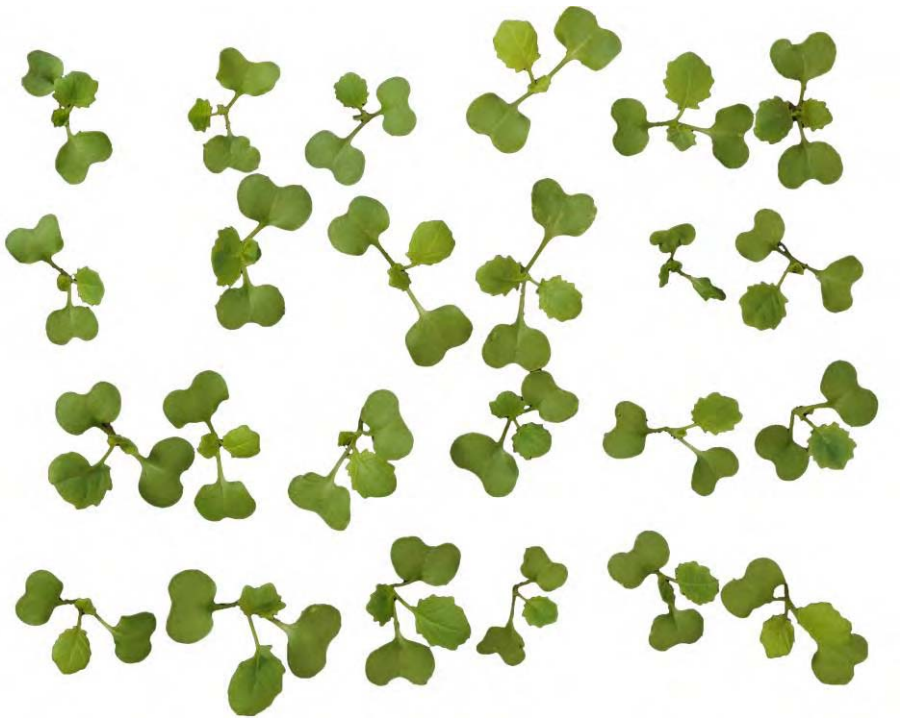
Rod Snowdon[1], Bertha Salazar[1], Aysha Kiran[1], Andrew Sharpe[2], Christine Sidebottom[2], Sue Abrams[2], Irina Zaharia[2], Pierre Fobert[2], Kerry Boyle[2], Isobel Parkin[3], Wentao Zhang[3], Erin Higgins[3], Benjamin Stich[4], Anja Bus[4], Niklas Körber[4], Janet Higgins[5], Ian Bancroft[5]

[1] Department of Plant Breeding, Justus Liebig University, Heinrich-Buff-Ring 26-31, 35392 Giessen, Germany
[2] National Research Council of Canada Plant Biotechnology Institute, 110 Gymnasium Place, Saskatoon, S7N 0W9, Canada
[3] Agriculture and AgResearch Canada, Saskatoon Research Centre, 107 Science Place, Saskatoon, S7N 0X2, Canada
[4] Max Planck Institute for Plant Breeding Research, Carl von Linné Weg 10, 50829 Cologne, Germany
[5] John Innes Centre, Norwich Research Park, Colney Lane, Norwich, NR4 7UH, UK

Oilseed rape/canola (*Brassica napus*) has a large and complex, highly duplicated, polyploid genome that presents problems for marker development, accurate genetic mapping and map-based gene cloning. As yet there is no completed public reference genome sequence, and extensive public-domain genomics resources like single-nucleotide polymorphism (SNP) markers or physical maps are only just starting to become accessible. On the other hand *B. napus* is very closely related to *Arabidopsis thaliana* and via comparative genomics approaches it is possible, when sufficient sequence data can be generated in well-defined plant material, to exploit the vast array of genome resources that are available for the model crucifer.

Global transcriptome analyses have been applied in recent years for high-density mapping of gene expression markers, genetical genomics studies of gene expression QTL and gene expression association analysis. Although microarray technologies are still commonly used for these applications, ultra-deep transcriptional profiling via next-generation DNA sequencing (NGS) has the potential to be considerably more powerful in terms of detection thresholds and quantification accuracy, particularly for low-abundance transcripts. Since transcription factors and other regulatory RNA can play a key regulatory role in biological systems even at very low transcript levels, an exact quantification of differential levels of such transcripts can give vital information for systems genetics approaches. The exploitation of the huge depth of genome and transcriptome data offered by NGS technologies could be considerably improved by systems genetics approaches that combine different analytical procedures. For example, weighted gene co-expression network analysis approaches (Zhang & Horvarth 2005, Langfelder & Horvarth 2007, Zhu et al. 2007) are being used increasingly in human genetics, where they have been successfully applied to define gene expression networks and identify key regulatory factors involved in complex diseases.

In winter rapeseed, seedling establishment plays an extremely important role in optimisation of nutrient uptake and post-winter plant development prior to flowering. Seedling root and shoot development traits in *B. napus* (e.g. Figure 1) also exhibit notable heterosis (Basunanda et al. 2010) and this may contribute to the general improvement in vigour shown by both spring-sown canola and winter rapeseed hybrids in comparison to inbred varieties. There is also a large variation in seedling vigour among inbred oilseed rape/canola varieties, which may reflect different expression of "fixed heterosis" in the polyploid *B. napus* genome. Seedling development involves extremely complex interactions amongst different hormone signalling pathways, whereby the major regulatory factors governing these interactions are to date largely unknown. On the other hand, seedlings represent a relatively simple biological object from which to obtain well-defined material for high-throughput transcriptome and hormone profiling. This offers the opportunity to use a simple biological system to gather extensive transcript and metabolite data. Hence seedlings can be used as an ideal case study for systems genetic analysis of a complex set of interacting traits that can have a high relevance for crop yield and yield stability, but whose regulation is still very poorly understood. Specific aspects of growth and development involve complex interactions of signalling molecules that can regulate gene expression, therefore hormone profiling has proven to be a useful tool for functional genomics studies. Analytical methods for simultaneous, high-throughput metabolite profiling of more than 40 phytohormones by mass spectrometry have been developed in recent years that are also successful

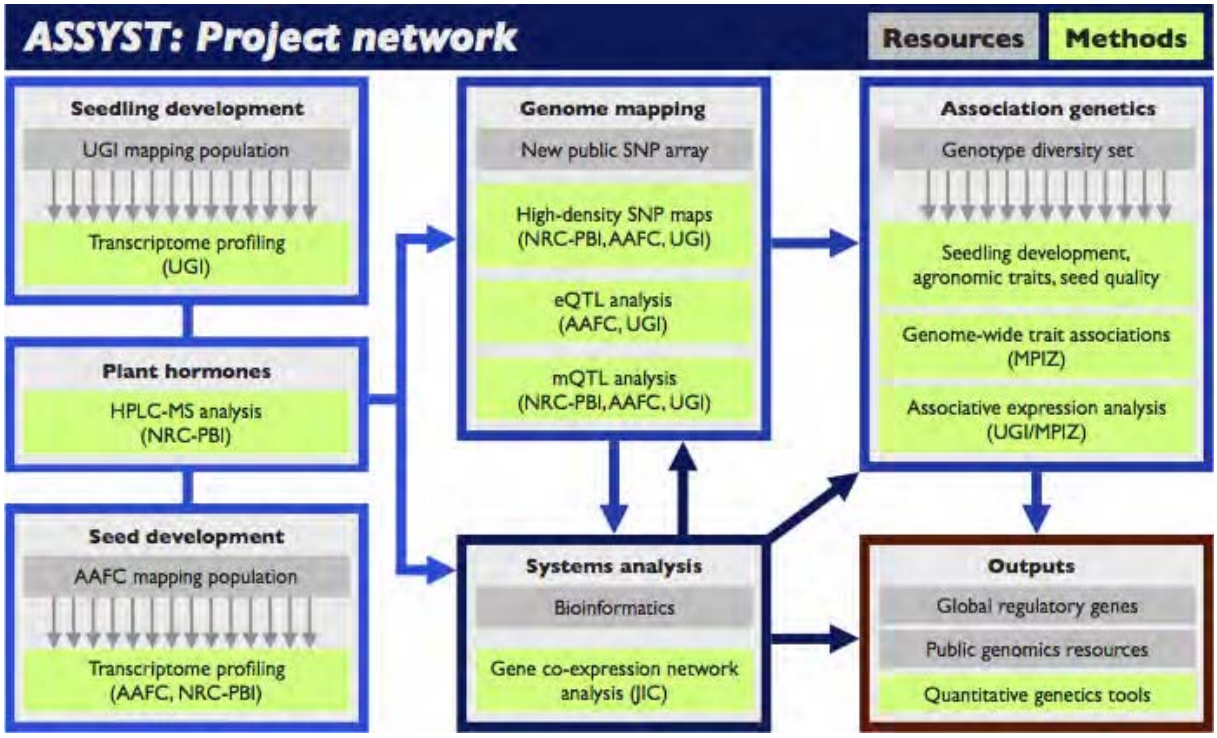


when applied in *B. napus* seeds and seedlings (Chiwocha et al 2003, 2005; Ross et al 2004).

Figure 1. Digital phenotyping of seedling vigour in the ERANET ASSYST *B. napus* diversity set.

The tri-national ERANET Plant Genomics consortium ASSYST (Figure 2) aims at identifying regulatory genes involved in expression of complex traits in *B. napus*. Quantitative gene expression

data from well-defined populations of segregating winter oilseed rape and spring canola populations are being integrated with segregating marker data, quantitative hormone metabolite profiles and phenotype data from greenhouse, growth chamber and field trials, using a systems genetics approach that combines an analysis of gene co-expression networks with expression QTL approaches. Furthermore, genome-wide marker data from a *B. napus* 1536 SNP array is being implemented for association analyses in a set of over 500 genetically diverse *B. napus* inbred lines, and quantitative expression levels of interesting regulatory candidate genes will be investigated in the diversity collection for associations with traits of interest. One focus of the work is on seedling development traits, as a case study for a complex interactive system that is genetically very poorly understood, but is agronomically extremely important. The network analysis tools are also being applied for a systems analysis of important seed quality characters, to identify key regulatory factors involved in biosynthesis of oil, protein and fibre components. The project incorporates the most recent technological developments in the field of next-generation sequencing for ultra-deep transcription profiling and SNP discovery. Gene co-expression network analysis, classical QTL analysis, genetical genomics and association genetics concepts are being integrated in a manner that until now has not been used for functional genomics of complex traits in crop plants. We intend to make the genotype diversity set and all marker data publicly available as a resource for genome-wide association studies in *B. napus*.



UGI: Department of Plant Breeding, University of Giessen, Germany
NRC-PBI: Plant Biotechnology Institute, Saskatoon, Canada
AAFC: Agriculture and AgResearch Canada, Saskatoon, Canada
MPIZ: Max Planck Institute for Plant Breeding Research, Cologne, Germany
JIC: John Innes Centre, Norwich, UK

Figure 2. Overview of the ERANET ASSYST consortium and project aims.

References

Basunanda et al. 2010. *Theor Appl Genet* 120: 271-281

Chiwocha SDS, et al. 2003: *Plant J* 35: 405-417.

Langfelder P, Horvath S, 2007: *BMC Systems Biology* 2007-1: 54

Ross ARS, et al. 2004. *Anal Biochem* 392: 324-333

Zhang B & Horvath S, 2005: *Stat Appl Genetics Molecular Biology* 4: No. 1, Article 17

Zhu J, et al. 2007: *PLoS Comput Biol* 3: e69

Acknowledgements

ERANET-ASSYST is funded by the contributing ERANET partner organisations DFG (Germany), BBSRC (UK), Agriculture and Agri-Food Canada, National Research Council of Canada. The German breeding companies NPZ Lembke, Deutsche Saatveredelung, KWS Saat, Limagrain are kindly contributing to field evaluations of the ASSYST *B. napus* diversity set.