COMMENT / COMMENTAIRE

Prospects and challenges for genome-wide association and genomic selection in oilseed *Brassica* species[1]

W.A. Cowling and E. Balázs

Introduction

Papers in this special issue of *Genome* were submitted for publication following the conference "Exploiting Genome-wide Association in Oilseed Brassicas: a model for genetic improvement of major OECD crops for sustainable future farming", held at The University of Western Australia, 9–12 November 2009, which was sponsored by the OECD Co-operative Research Programme on Biological Resource Management for Sustainable Agricultural Systems[3]. This paper summarizes formal and informal discussions among the attendees on prospects and challenges for genome-wide association studies and genomic selection in oilseed *Brassica* species, with subsequent updating before publication.

Genome-wide association and genomic selection in animals and plants

The outcomes sought from genome-wide association studies (GWAS) differ among human geneticists, plant molecular geneticists, and animal and plant breeders. Thousands and often millions of single nucleotide polyporphism (SNP) markers are available in public databases for GWAS in humans and other animal species, *Arabidopsis*, rice, maize, barley and *Brassica* (Weir 2010), and domesticated livestock (Hayes and Goddard 2010). In humans, the major role of GWAS is the discovery of genes and pathways ("candidate" genes), whereas in animal (and presumably plant) breeding there is more emphasis on predicting genetic merit of the phenotype, with the aim of accelerating genetic improvement (Goddard and Hayes 2009). Marker assisted selection is of greatest value for genetic improvement of complex traits in elite breeding programs. In animal breeds, the proportion of variance explained by individual SNPs is small; therefore, it is necessary to use all the SNP information simultaneously and capture all the quantitative trait loci (QTL) in a process called "genomic selection" (Hayes and Goddard 2010).

When applied to breeding programs, SNP genome-wide markers can have the dual benefits of gene identification for complex traits and more rapid genetic improvement. The challenge for plant breeders is how best to achieve these positive outcomes of GWAS in plant breeding. The genus *Brassica* is a good model to test the value of genomic selection inside plant breeding programs, since *Brassica* is related to the model plant *Arabidopsis* (and can benefit from sequence and functional genomics information developed in *Arabidopsis*). Good progress towards SNP whole-genome markers has been made in *B. rapa* and *B. napus*, and oilseed *Brassica* species are amenable to tissue–culture techniques that enable generation of doubled haploids.

However, plants are subject to large genotype × environment (G × E) interactions and the estimation of genetic merit depends greatly on the environment in which complex traits, such as yield, are measured (Cullis et al. 2010). The issue of G × E was considered in a recent review of genomic selection in plant breeding, but no biometrical solutions were presented to help with genomic selection in the presence of G × E (Heffner et al. 2009). In this conference, G × E was a major topic for discussion in the preconference workshop and in papers where genetic information from relatives was used in factor analytic (FA) modelling of multi-

Received 31 August 2010. Accepted 3 September 2010. Published on the NRC Research Press Web site at genome.nrc.ca on 8 November 2010.

W.A. Cowling.[2] Deputy Director, International Centre for Plant Breeding Education and Research, The UWA Institute of Agriculture, c/o School of Plant Biology M084, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia, Australia 6009. **E. Balázs.** Department of Applied Genomics, H-2462 Martonvásár, Agricultural Research Institute, Brunszvik. u. 2, Hungary (previously Theme Coordinator Research Theme 3: The Food Chain [Plants], OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agricultural Systems).

[1]This article is one of a selection of papers from the conference "Exploiting Genome-wide Association in Oilseed Brassicas: a model for genetic improvement of major OECD crops for sustainable farming"

[2]Corresponding author (e-mail: wallace.cowling@uwa.edu.au).

[3]The original presentations at the conference are available at www.oecd-genomeassociation-oz09.com or from the authors in this special issue of *Genome*.

environment trial (MET) data in a canola breeding program (Beeck et al. 2010; Cullis et al. 2010). Pedigree information improved the modelling of G × E and revealed distinct patterns of additive and nonadditive genetic variation across environments. Genome-wide markers integrated into MET and FA analysis should improve further on these analyses, and ultimately lead to genomic selection in plant breeding programs.

Clearly, major challenges remain to be solved before genomic selection is adopted in commercial plant breeding programs. In *Brassica*, high density genome-wide SNP arrays do not yet exist, and the cost of genotyping with SNP chips may be beyond the commercial reach of most breeding programs. To be effective, every individual in the breeding program that is phenotyped in the field should also be genotyped by SNP arrays. In the example presented by Beeck et al. (2010), 332 *B. napus* genotypes were assessed at 19 sites over 2 years in highly unbalanced designs in MET and FA analysis. The cost of genotyping 332 individuals will not be trivial (332 is a small number of genotypes compared with those tested in most plant breeding programs); missing molecular marker values may cause problems in the analysis (Schrag et al. 2010); and the time and cost involved in "cleaning" data will be substantial (Weir 2010). The process of cleaning SNP data from many individuals for quality assurance and quality control is very demanding, but it is important to detect and eliminate false positive associations by eliminating spurious effects (Weir 2010).

Human studies have demonstrated the value of large sample sizes — tens of thousands are necessary to find subtle single genetic effects, although epistatic interactions have not yet been revealed by GWAS (Weir 2010). A major benefit of plant breeding programs for GWAS is that large numbers of plant progenies can be generated and tested in replicated trials, with control over the population structure. Inbreeding crops such as *B. napus* are interesting models to find epistatic effects through GWAS — homozygous pure lines should have less complex epistasis than outbreeding organisms.

Inbreeding at the level of the population or the individual will require special statistical analysis in GWAS (Weir 2010). Selfing to homozygosity (as occurs in most breeding programs of inbreeding crops) is clearly not essential to achieve benefits from genomic selection; the rate of genetic improvement per year in heterozygous livestock species may be doubled through genomic selection compared with pedigree selection only (Hayes and Goddard 2010). Several differences exist between animal and plant breeding programs. At the population level, many crop breeding programs have a small effective population size (Cowling et al. 2009). The selection unit in animal breeding is the individual, which has a unique genome. In plant breeding, the selection unit is often heterogeneous, there are multiple "genomes" within a plant variety, and these are subject to segregation in each generation of inbreeding. Plant breeders face many challenges in modifying their breeding programs to capitalize on the benefits of genomic selection.

While the power to accurately estimate SNP effects is greatest when the allele frequency is 0.5, targeting intermediate allele frequencies should not become an obsession. There is a strong argument that phenetic or numerical taxonomic studies would be more robust if they included loci with a range of allelic variation (Sneath and Sokal 1973), as this is closer to the expected distribution of QTL allele frequencies.

However, QTL allele frequencies do become important at a low effective population size. In animal populations with an effective population size of 20, alleles with frequency <0.1 were shown to be at risk of loss through genetic drift (Luikart et al. 1998). In practice, it is difficult to introgress "new" positive alleles for quantitative traits in plant breeding programs if their frequency is low (Cowling et al. 2009). Nevertheless, introgression of positive alleles from wide or distant relatives is a major justification for development of genomic selection in crop plant breeding. Potentially, it will be beneficial to combine the population breeding approach (Falk 2010) with genomic selection, and to trace the movement of positive alleles from wild into elite crop varieties as done in rice (McCouch et al. 2007).

Traditional marker assisted selection is based on diagnostic markers for genes or loci that may represent clusters of linked positive allelic forms, normally developed by "prebreeders". The next generation of marker-assisted selection will encompass genomic selection inside breeding programs. Of course, knowledge of candidate genes and QTL identified by linkage analysis will benefit future developments in genomic selection (Zou et al. 2010).

New technologies for GWAS and candidate gene discovery

Second generation sequencing is having a huge impact on candidate gene discovery (Imelfort et al. 2009*a*; Imelfort and Edwards 2009) with large implications for crop improvement (Edwards and Batley 2010). It is now possible to discover genetic variation on a whole genome level, something which was impossible just a few short years ago (Chagné et al. 2007; Edwards et al. 2007; Batley and Edwards 2009*b*; Duran et al. 2009*a*; Duran et al. 2009*b*; Duran et al. 2009*c*; Imelfort et al. 2009*b*). These discoveries when combined with massive throughput SNP genotyping as developed by Illumina, Inc. (San Diego, California) offer whole new opportunities for genome wide association and selection (Appleby et al. 2009; Batley and Edwards 2007; Edwards and Batley 2010; Marshall et al. 2010).

This avalanche of data creates a huge challenge for bioinformatics (Edwards 2007; Edwards and Batley 2008; Batley and Edwards 2009*a*); however, there are an increasing number of tools available to assist with the storage, comparison, and visualization of *Brassica* genetic markers (Love et al. 2004; Love et al. 2006; Lim et al. 2007; Duran et al. 2010). There remains a gap between the genome and phenome that will require a substantial investment in the collation of phenomic data (Edwards and Batley 2004).

Next generation sequencing will enable epigenetic variation (the "epi-genome") to be characterized and quantified (Joosen et al. 2009; Lister et al. 2008). The adoption of eQTL and genetical genomics approaches may provide more insights as large interconnected datasets become available. This will determine the contribution of the epi-genome to phenotypic variation that mediates G × E interactions.

What impact will such knowledge have on plant breeding? Potentially, there is scope to manipulate epi-allelic variation in relation to providing plants with greater adaptation to variable growing environments (King et al. 2010).

Another prospective use of the sequence data is for predicting genomic estimated breeding values for genomic selection. Meuwissen and Goddard (2010) demonstrated that using sequence data, rather than SNP data, could give further improvements in rate of genetic gain.

Technologies for genome-wide association studies in *Brassica*

SSR markers are valuable to identify diversity among genotypes, as few markers are required and they are multiallelic. For example, up to 10 alleles per locus and 4 or 5 loci per SSR primer pair were detected in a diversity analysis of a diverse population of *B. napus* (Chen et al. 2008). This type of diversity analysis provides useful information for the selection of lines for SNP discovery. SSR markers may be helpful to align maps of *Brassica*, but caution is necessary because of frequent SSR paralogues in replicated regions of the genome. Many sources of *Brassica* SSR markers are available (Burgess et al. 2006; Batley et al. 2007; Hopkins et al. 2007; Ling et al. 2007; Lowe et al. 2004; Piquemal et al. 2005; Suwabe et al. 2002; Suwabe et al. 2006) and additional resources for their discovery (Robinson et al. 2004; Jewell et al. 2006).

SSR markers are problematic for GWAS in polyploid species such as *B. napus* because of the difficulty in distinguishing between homoeology (divergent copies of genes in different genomes) and paralogy (within-genome gene duplication events) and the occurrence of nulls (Mayerhofer et al. 2005; Parkin et al. 2003; H. Raman, personal communication). Also, the mutation rate in and around SSRs is generally higher than in coding regions of neighbouring genes (Mogg et al. 2002; Duran et al. 2009*b*); therefore, polymorphism in SSRs may not be representative of normal evolution, which occurs largely through single base mutations in genes.

SNPs are the most powerful markers for association studies because of their abundance, low rate of reversion to the ancestral state, and relatively low cost of high throughput assay. Up to 20 000 SNPs across the genome have been identified for parents of the *B. napus* mapping population BnaTNDH (Trick et al. 2009) (data are available from www.brassica.bbsrc.ac.uk/). In the UK, the BBSRC sequencing project is currently screening the "digital transcriptome" across the BnaTNDH mapping population (provided by J. Meng, Huazhong Agricultural University, China), therefore, these SNP markers will be mapped at a high density. Other projects are generating gene expression markers (GEMS) based on array data from eQTL experiments.

A collaboration between China and Australia will map 384 SNPs, and a project is underway in Australia to develop and map >3000 SNPs in *B. napus* and across the family Brassicaceae (J. Batley, personal communication). These SNPs require manual input to evaluate and score accurately.

There is a strong interest in making a high-density SNP array publicly available as a tool for the *Brassica* breeding and research communities. A number of participants in the Multinational *Brassica* Genome Project (available from www.brassica.info) have expressed interest in developing a public 60 kb SNP array, and discussions are underway among potential contributors of SNP data from public and commercial *Brassica* researchers. A key to industry involvement is the growing acknowledgement within the breeding industry that GWAS and genomic selection can be extremely useful for identification, characterization, and utilization of novel genetic diversity in practical breeding programs.

The current estimated cost (August 2010) for a 60 kb SNP Illumina BeadArray lies between US$800 and $1000 per sample (including analysis and chip) depending on the production run for the chips (R. Snowdon, personal communication), although the price should be less for a lower density chip (G. Durstewitz, personal communication). This may limit the work to (perhaps) founders of breeding populations, as done by maize breeders. Based on this estimate, for the canola breeding program described in Beeck et al. (2010), genotyping alone would cost >$250 000 for the 332 genotypes, and this does not include the cost of the cleaning of data or the time taken before genomic selection could proceed. It is not yet clear what price would be "break-even" for the application of genomic selection via SNP chips in *Brassica* breeding. Many questions remain for the commercial breeder such as how should the breeder handle heterogeneity within varieties, which may not be completely pure breeding? Should the breeder sample 1 plant per variety, 10 plants, or even more for DNA extraction and genomic selection?

Generating and measuring genetic and phenotypic diversity

The importance of precision and accuracy in measuring phenotype cannot be overstated for complex traits. Accurate assessment of phenotype such as yield or oil in canola as described in Beeck et al. (2010) and Cullis et al. (2010) is just as important as accurate assessment of genotype for positive outcomes from GWAS and genomic selection. Accounting for nongenetic variation such as spatial variation in field trials and adding genetic information from pedigrees, MET and FA modelling of canola yield and oil was greatly improved (Beeck et al. 2010; Cullis et al. 2010), but more work is required to integrate genomic selection in the presence of G × E. Plant breeders may have to consider modifying their breeding program to exploit the benefits of genomic selection such as increasing the number of genotypes tested in the MET trial series through the use of *p*-rep designs (Beeck et al. 2010). Breeders and molecular geneticists will also have to resolve the question of sampling within heterogeneous varieties for determination of the variety's "genome".

The research and (or) breeding community also requires more investment to ensure ongoing support for integrated curation and data repositories of large phenotypic and (or) genotypic datasets. One framework for this is being generated (available from www.brassica.info/resource/databases/cropstore.php).

Exploiting association mapping in Brassica and other crops

GWAS will lead to the identification of genomic regions

related to yield and other important traits in crop plants — will it be possible to integrate these data from public and (or) private sources? A canola breeding company donated data for analysis and publication in these proceedings (Beeck et al. 2010; Cullis et al. 2010). Collaboration is possible between public and private researchers to develop technology for genomic selection in the precompetitive area. This is particularly the case for nested association mapping (NAM) populations (Yu et al. 2008), which generally include predominantly prebreeding materials from exotic founders, but are nevertheless of great interest for commercial breeders. Initiatives are now underway to develop, genotype, and phenotype immortal NAM populations for *Brassica* crops. Such investments will considerably increase the power of genome-wide association analysis for complex traits.

Acknowledgements

We thank the many conference participants and authors from this special issue of *Genome* who contributed to this discussion paper at the conference and during the editing phase.

References

Appleby, N., Edwards, D., and Batley, J. 2009. New technologies for ultra-high throughput genotyping in plants. *In* Plant genomics. *Edited by* D. Somers, P. Langridge, and J.P. Gustafson. Humana Press, New York. pp. 19–40.

Batley, J., and Edwards, D. 2007. SNP applications in plants. *In* Association mapping in plants. *Edited by* N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner, and H.N. De Silva. Springer, New York. pp. 95–102.

Batley, J., and Edwards, D. 2009*a*. Genome sequence data: management, storage, and visualization. Biotechniques, **46**(5): 333–336. doi:10.2144/000113134. PMID:19480628.

Batley, J., and Edwards, D. 2009*b*. Mining for single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) molecular genetic markers. *In* Bioinformatics for DNA sequence analysis. *Edited by* D. Posada. Humana Press, New York. pp. 303–322.

Batley, J., Hopkins, C.J., Cogan, N.O.I., Hand, M., Jewell, E., Kaur, J., et al. 2007. Identification and characterization of simple sequence repeat markers from *Brassica napus* expressed sequences. Mol. Ecol. Notes, **7**(5): 886–889. doi:10.1111/j.1471-8286.2007.01738.x.

Beeck, C.P., Cowling, W.A., Smith, A.B., and Cullis, B.R. 2010. Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. Genome, **53**(11): 992–1001. doi:10.1139/G10-051.

Burgess, B., Mountford, H., Hopkins, C.J., Love, C., Ling, A.E., Spangenberg, G.C., et al. 2006. Identification and characterization of simple sequence repeat (SSR) markers derived in silico from *Brassica oleracea* genome shotgun sequences. Mol. Ecol. Notes, **6**(4): 1191–1194. doi:10.1111/j.1471-8286.2006.01488.x.

Chagné, D., Forster, J.W., Cogan, N.O.I., Batley, J., and Edwards, D. 2007. Single nucleotide polymorphism discovery. *In* Association mapping in plants. *Edited by* N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner, and H.N. De Silva. Springer, New York. pp. 53–76.

Chen, S., Nelson, M.N., Ghamkhar, K., Fu, T., and Cowling, W.A. 2008. Divergent patterns of allelic diversity from similar origins: the case of oilseed rape (*Brassica napus* L.) in China and Australia. Genome, **51**(1): 1–10. doi:10.1139/G07-095. PMID: 18356934.

Cowling, W.A., Buirchell, B.J., and Falk, D.E. 2009. A model for incorporating novel alleles from the primary gene pool into elite crop breeding programs while reselecting major genes for domestication or adaptation. Crop Pasture Sci. **60**(10): 1009–1015. doi:10.1071/CP08223.

Cullis, B.R., Smith, A.B., Beeck, C.P., and Cowling, W.A. 2010. Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. Genome, **53**(11): 1002–1016. doi:10.1139/G10-080.

Duran, C., Appleby, N., Edwards, D., and Batley, J. 2009*a*. Molecular genetic markers: discovery, applications, data storage and visualisation. Curr. Bioinf. **4**(1): 16–27. doi:10.2174/157489309787158198.

Duran, C., Appleby, N., Vardy, M., Imelfort, M., Edwards, D., and Batley, J. 2009*b*. Single nucleotide polymorphism discovery in barley using autoSNPdb. Plant Biotechnol. J. **7**(4): 326–333. doi:10.1111/j.1467-7652.2009.00407.x. PMID:19386041.

Duran, C., Edwards, D., and Batley, J. 2009*c*. Molecular marker discovery and genetic map visualisation. *In* Applied bioinformatics: tools and applications. *Edited by* D. Edwards, D. Hanson, and J. Stajich. Springer, New York. pp. 95–102.

Duran, C., Boskovic, Z., Imelfort, M., Batley, J., Hamilton, N.A., and Edwards, D. 2010. CMap3D: a 3D visualization tool for comparative genetic maps. Bioinformatics, **26**(2): 273–274. doi:10.1093/bioinformatics/btp646. PMID:19942584.

Edwards, D. 2007. Bioinformatics and plant genomics for staple crops improvement. *In* Breeding major food staples. *Edited by* M.S. Kang and P.M. Priyadarshan. Blackwell Publishing, Ames, Iowa. pp. 93–106.

Edwards, D., and Batley, J. 2004. Plant bioinformatics: from genome to phenome. Trends Biotechnol. **22**(5): 232–237. doi:10.1016/j.tibtech.2004.03.002. PMID:15109809.

Edwards, D., and Batley, J. 2008. Bioinformatics: fundamentals and applications in plant genetics, mapping and breeding. *In* Principles and practices of plant genomics. *Edited by* C. Kole and A.G. Abbott. Science Publishers, Inc., Enfield, New Hampshire, USA. pp. 269–302.

Edwards, D., and Batley, J. 2010. Plant genome sequencing: applications for crop improvement. Plant Biotechnol. J. **8**(1): 2–9. doi:DOI: 10.1111/j.1467-7652.2009.00459.x.

Edwards, D., Batley, J., Cogan, N.O.I., Forster, J.W., and Chagné, D. 2007. Single nucleotide polymorphism discovery. *In* Association mapping in plants. *Edited by* N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner, and H.N. De Silva. Springer, New York. pp. 53–76.

Falk, D.E. 2010. Generating and maintaining diversity at the elite level in crop breeding. Genome, **53**(11): 982–991. doi:10.1139/G10-081.

Goddard, M.E., and Hayes, B.J. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. **10**(6): 381–391. doi:10.1038/nrg2575. PMID: 19448663.

Hayes, B., and Goddard, M. 2010. Genome-wide association and genomic selection in animal breeding. Genome, **53**(11): 876–883. doi:10.1139/G10-076.

Heffner, E.L., Sorrells, M.E., and Jannink, J.-L. 2009. Genomic selection for crop improvement. Crop Sci. **49**(1): 1–12. doi:10.2135/cropsci2008.08.0512.

Hopkins, C.J., Cogan, N.O.I., Hand, M., Jewell, E., Kaur, J., Li, X., et al. 2007. Sixteen new simple sequence repeat markers from *Brassica juncea* expressed sequences and their cross-species am-

plification. Mol. Ecol. Notes, **7**(4): 697–700. doi:10.1111/j.1471-8286.2007.01681.x.

Imelfort, M., and Edwards, D. 2009. De novo sequencing of plant genomes using second-generation technologies. Brief. Bioinform. **10**(6): 609–618. doi:10.1093/bib/bbp039. PMID: 19933209.

Imelfort, M., Batley, J., Grimmond, S., and Edwards, D. 2009*a*. Genome sequencing approaches and successes. *In* Plant genomics. *Edited by* D. Somers, P. Langridge, and J.P. Gustafson. Humana Press, New York. pp. 345–358.

Imelfort, M., Duran, C., Batley, J., and Edwards, D. 2009*b*. Discovering genetic polymorphisms in next-generation sequencing data. Plant Biotechnol. J. **7**(4): 312–317. doi:10.1111/j.1467-7652.2009.00406.x. PMID:19386039.

Jewell, E., Robinson, A., Savage, D., Erwin, T., Love, C.G., Lim, G.A.C., et al. 2006. SSR Primer and SSR Taxonomy Tree: Biome SSR discovery. Nucleic Acids Res. **34:**(4): W656–W659. doi:10.1093/nar/gkl083. PMID:16845092.

Joosen, R.V.L., Ligterink, W., Hilhorst, H.W.M., and Keurentjes, J.J.B. 2009. Advances in genetical genomics of plants. Curr. Genomics, **10**(8): 540–549. doi:10.2174/138920209789503914. PMID:20514216.

King, G.J., Amoah, S., and Kurup, S. 2010. Exploring and exploiting epigenetic variation in crops. Genome, **53**(11): 856–868. doi:10.1139/G10-059.

Lim, G.A.C., Jewell, E.G., Li, X., Erwin, T.A., Love, C., Batley, J., et al. 2007. A comparative map viewer integrating genetic maps for *Brassica* and *Arabidopsis.* BMC Plant Biol. **7**(1): 40. doi:10.1186/1471-2229-7-40. PMID:17645810.

Ling, A.E., Kaur, J., Burgess, B., Hand, M., Hopkins, C.J., Li, X., et al. 2007. Characterization of simple sequence repeat markers derived *in silico* from *Brassica rapa* bacterial artificial chromosome sequences and their application in *Brassica napus.* Mol. Ecol. Notes, **7**(2): 273–277. doi:10.1111/j.1471-8286.2006.01578.x.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis.* Cell, **133**(3): 523–536. doi:10.1016/j.cell.2008.03.029. PMID:18423832.

Love, C.G., Batley, J., Lim, G., Robinson, A.J.R., Savage, D., Singh, D., et al. 2004. New computational tools for *Brassica* genome research. Comp. Funct. Genomics, **5**(3): 276–280. doi:10.1002/cfg.394. PMID:18629157.

Love, C., Logan, E., Erwin, T., Spangenberg, G., and Edwards, D. 2006. Analysis of the *Brassica* A and C genomes and comparison with the genome of *Arabidopsis thaliana.* Acta Hortic. **706**: 99–104.

Lowe, A.J., Moule, C., Trick, M., and Edwards, K.J. 2004. Efficient large-scale development of microsatellites for marker and mapping applications in *Brassica* crop species. Theor. Appl. Genet. **108**(6): 1103–1112. doi:10.1007/s00122-003-1522-7. PMID:15067397.

Luikart, G., Allendorf, F.W., Cornuet, J.-M., and Sherwin, W.B. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J. Hered. **89**(3): 238–247. doi:10.1093/jhered/89.3.238. PMID:9656466.

Marshall, D.J., Hayward, A., Eales, D., Imelfort, M., Stiller, J., Berkman, P.J., et al. 2010. Targeted identification of genomic regions using TAGdb. Plant Methods, **6**(1): 19. doi:10.1186/1746-4811-6-19. PMID:20727174.

Mayerhofer, R., Wilde, K., Mayerhofer, M., Lydiate, D., Bansal, V.K., Good, A.G., and Parkin, I.A.P. 2005. Complexities of chromosome landing in a highly duplicated genome: toward map-based cloning of a gene controlling blackleg resistance in *Brassica napus.* Genetics, **171**(4): 1977–1988. doi:10.1534/genetics.105.049098. PMID:16143600.

McCouch, S.R., Sweeney, M., Li, J., Jiang, H., Thomson, M., Septiningsih, E., et al. 2007. Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa.* Euphytica, **154**(3): 317–339. doi:10.1007/s10681-006-9210-8.

Meuwissen, T., and Goddard, M.E. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics, **185**(2): 623–631. doi:10.1534/genetics.110.116590. PMID:20308278.

Mogg, R., Batley, J., Hanley, S., Edwards, D., O'Sullivan, H., and Edwards, K. 2002. Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. Theor. Appl. Genet. **105**(4): 532–543. doi:10.1007/s00122-002-0897-1. PMID:12582502.

Parkin, I.A., Sharpe, A.G., and Lydiate, D.J. 2003. Patterns of genome duplication within the *Brassica napus* genome. Genome, **46**(2): 291–303. doi:10.1139/g03-006. PMID:12723045.

Piquemal, J., Cinquin, E., Couton, F., Rondeau, C., Seignoret, E., Doucet, I., et al. 2005. Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theor. Appl. Genet. **111**(8): 1514–1523. doi:10.1007/s00122-005-0080-6. PMID: 16187118.

Robinson, A.J., Love, C.G., Batley, J., Barker, G., and Edwards, D. 2004. Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics, **20**(9): 1475–1476. doi:10.1093/bioinformatics/bth104. PMID:14962913.

Schrag, T.A., Möhring, J., Melchinger, A.E., Kusterer, B., Dhillon, B.S., Piepho, H.-P., and Frisch, M. 2010. Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. Theor. Appl. Genet. **120**(2): 451–461. doi:10.1007/s00122-009-1208-x. PMID: 19916002.

Sneath, P.H.A., and Sokal, R.R. 1973. Numerical taxonomy. Freeman, San Francisco.

Suwabe, K., Iketani, H., Nunome, T., Kage, T., and Hirai, M. 2002. Isolation and characterization of microsatellites in *Brassica rapa* L. Theor. Appl. Genet. **104**(6–7): 1092–1098. doi:10.1007/s00122-002-0875-7. PMID:12582617.

Suwabe, K., Tsukazaki, H., Iketani, H., Hatakeyama, K., Kondo, M., Fujimura, M., et al. 2006. Simple sequence repeat-based comparative genomics between *Brassica rapa* and *Arabidopsis thaliana*: the genetic origin of clubroot resistance. Genetics, **173**(1): 309–319. doi:10.1534/genetics.104.038968. PMID: 16723420.

Trick, M., Long, Y., Meng, J., and Bancroft, I. 2009. Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. Plant Biotechnol. J. **7**(4): 334–346. doi:10.1111/j.1467-7652.2008.00396.x. PMID:19207216.

Weir, B.S. 2010. Statistical genetic issues for genome-wide association studies. Genome, **53**(11): 869–875. doi:10.1139/G10-062.

Yu, J., Holland, J.B., McMullen, M.D., and Buckler, E.S. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics, **178**(1): 539–551. doi:10.1534/genetics.107.074245. PMID:18202393.

Zou, J., Jiang, C., Cao, Z., Li, R., Long, Y., Chen, S., and Meng, J. 2010. Association mapping of seed oil content in *Brassica napus* and comparison with quantitative trait loci identified from linkage mapping. Genome, **53**(11): 908–916. doi:10.1139/G10-075.