i.SSS

*international society
for seed science*

10th ISSS
Conference of the International
Society for Seed Science - ISSS
"Seed Science in the 21st Century"

REVIEW

Unravelling the complex trait of seed quality: using natural variation through a combination of physiology, genetics and -omics technologies

Wilco Ligterink*, Ronny V.L. Joosen and Henk W.M. Hilhorst
Wageningen Seed Lab, Lab of Plant Physiology, Wageningen University, Wageningen, The Netherlands

Abstract

Seed quality is a complex trait that is the result of a large variety of developmental processes. The molecular-genetic dissection of these seed processes and their relationship with seed and seedling phenotypes will allow the identification of the regulatory genes and signalling pathways involved and, thus, provide the means to predict and enhance seed quality. Natural variation for seed-quality aspects found in recombinant inbred line (RIL) populations is a great resource to help unravel the complex networks involved in the acquisition of seed quality. Besides extensive phenotyping, RILs can also be profiled by -omics technologies, such as transcriptomics, proteomics and metabolomics in a sophisticated so-called generalized genetical genomics approach. This combined use of physiology, genetics and several -omics technologies, followed by advanced data analysis, allows the construction of regulatory networks involved in the various attributes of seed and seedling quality. This type of analysis of the genetic variation in RIL populations in combination with genome-wide association (GWA) studies will allow a relatively rapid identification of genes that are responsible for quality-related traits of seeds and seedlings. New developments in several -omics technologies, especially the fast-evolving next-generation sequencing techniques, will make a similar system-wide approach more applicable to non-model species in the near future and this will be a huge boost for the potential to breed for seed quality.

Seed quality

Seed quality is a complex trait and comprises many different attributes describing the condition of a seed batch. These attributes include germination characteristics, dormancy, seed and seedling vigour, uniformity in seed size, normal embryo and seedling morphology, storability, absence of mechanical damage, as well as the ability to develop into a normal plant (Dickson, 1980; Hilhorst and Toorop, 1997). Seed quality is largely established during seed development and maturation, as a result of, often complex, interactions between the genome and the environment. This mechanism is part of the normal adaptation of plants to a varying environment and is aimed at maximizing the probability of successful offspring (Huang *et al.*, 2010a).

The practical definition of seed quality is determined by the end user and, therefore, will differ substantially, depending on the use of seeds as propagule or commodity. For farmers or plant growers, high-quality seeds are those seeds that germinate and produce seedlings to a high percentage

*Correspondence
Email: Wilco.Ligterink@wur.nl

under a wide range of field conditions. On the other hand, high-quality seeds for use in the food industry may be seeds with a high starch or oil content or oil seeds with a specific fatty acid composition (Nesi *et al.*, 2008). As a result of the complexity of seed quality, testing for seed quality in order to predict subsequent behaviour in the field is troublesome and at best an 'educated guess' (Powell, 2006). Therefore, seed producers have included additional attributes to the term 'seed quality', such as usable plants and seedling and crop establishment. The trait 'usable plants' is one of the most important attributes of seed quality used by seed producers and plant growers.

Seed companies may enhance seed quality at all the different steps of the production process. At present, seed companies try to obtain the best possible seeds mainly by varying the time and method of harvest, but especially by post-harvest treatments such as cleaning, sorting, coating and priming, and controlling the storage conditions. Besides these methods, seed quality can also be improved by controlling the production environment. It is known that seed quality is largely acquired during seed development and particularly during the maturation phase, by the successive acquisition of seed-quality attributes such as germinability, desiccation tolerance, dormancy, vigour and longevity (Harada, 1997), and that the environmental conditions during development have a huge impact on these different seed-quality aspects. As a result, the quality of different seed lots that are produced in different seasons and locations will vary. Nevertheless, influencing production environments is difficult, even under greenhouse conditions. Furthermore, since there is a complex interaction between the genome and the environment during development, the final effect of the environment on seed quality is difficult to determine and still largely unknown. However, the genetic component of the interaction between the genome and the environment can be investigated and this variation in genetic adaptation provides great opportunities for seed companies to breed for seed quality.

Natural variation of seed quality

Although abundant natural variation for seed quality exists, genetic components of seed quality have hardly been used in breeding programmes. Exploiting natural variation is a powerful way to find the genes influencing important physiological processes. There are several ways to exploit natural variation but, in plants, quantitative trait locus (QTL) analysis of recombinant inbred line (RIL) populations have been widely used. In this type of analysis, linkage is sought between the genetic variety and the variation of phenotypic traits in the different RILs (Alonso-Blanco and Koornneef, 2000), whereby the QTLs represent the genomic regions explaining the phenotypic variation that is identified in this way. QTL analysis in plants has revealed a long list of genomic regions with variation for a broad variety of phenotypes, and several of the genes underlying these QTLs have been cloned (reviewed in Salvi and Tuberosa, 2005; Gupta *et al.*, 2009). The complex nature of 'seed quality' makes it a perfect trait to decipher with a QTL approach, particularly because different aspects of seed quality have been proven to have sufficient natural variation to tackle this subject. In *Arabidopsis thaliana*, different QTLs were found for dormancy (Bentsink *et al.*, 2010) and several germination characteristics (Clerkx *et al.*, 2004; Galpaz and Reymond, 2010; Joosen *et al.*, 2010). In tomato, different QTLs for germination character-istics under stress (Foolad *et al.*, 2003, 2007) and for seed size (Doganlar *et al.*, 2000) have been identified. In *Medicago truncatula* several QTLs were identified for germination at extreme temperatures (Dias *et al.*, 2011) and germination and seedling growth under osmotic stress (Vandecasteele *et al.*, 2011). Zeng *et al.* (2006) have identified QTLs for seed storability in rice, and in lettuce QTLs have been identified for several germination characteristics, including thermoinhibition (Argyris *et al.*, 2005, 2008).

In spite of these and other studies on specific aspects of seed quality, an integrated study of the genetics of seed quality is still lacking. A more systematic approach, studying genetic populations differing in seed- and seedling-quality parameters, will provide valuable insight into the involvement of genes, and the processes they control, in the acquisition of seed quality. Until now, only a few QTL positions have been cloned and characterized in detail, but if genes or gene sets associated with seed-quality parameters become available, they may be used as diagnostic tools to assess seed quality, in marker-assisted breeding, or in genetic modification to enhance seed quality.

High-throughput phenotyping

With the fast developments in sequencing technologies that enable fast and relatively inexpensive genotyping and expression analysis, accurate phenotyping is now becoming the limiting step in studying large genetic populations. To overcome this problem several initiatives have been taken to enhance phenotyping, mainly by implementing high-throughput phenotyp-ing platforms for analysing plant morphology, as in the Australian 'High Resolution Plant Phenomics Centre' (HRPPC) (http://www.plantphenomics.org/hrppc) and the Lemnatec systems (www.lemnatec.de) that perform fully automated imaging and subsequent

data extraction of growing plants. For the systemic analysis of the different aspects of seed quality, several (semi-)automatic phenotyping systems can be used. One of the most important aspects is the (semi-) automatic scoring of germination. Several methods to achieve this have been reported by Dell'Aquila (2009) and, more recently, by Joosen *et al.* (2010), who introduced the GERMINATOR package. Furthermore, analysis of seedling shape and growth with systems like that of the previously mentioned HRPPC and Lemnatec, and analysis of the root architecture of seedlings with programs such as EZ-Rhizo (Armengaud *et al.*, 2009) and Roottrace (French *et al.*, 2009), will become important for the in-depth analysis of seed quality.

Genetical genomics: -omics QTL analysis

Fine mapping of QTL is a crucial step for plant breeding as genetic drag should be minimized in every step during the breeding process. Furthermore, cloning of genes responsible for the QTL can provide great insight into the molecular mechanism underlying the adaptation. Although the causal genes for several seed-quality QTLs have been cloned and more are under way (Salvi and Tuberosa, 2005), fine-mapping and ultimate cloning of these genes is very labour-intensive and time-consuming. Therefore classical QTL analysis can be considered as a low-throughput technique. To help in candidate gene selection the concept of genetical genomics was developed (Jansen and Nap, 2001). In genetical genomics the traditional QTL analysis is combined with genome-wide expression profiling for all the lines of a RIL population. With these data, a QTL profile of the expression of every gene can be calculated, just like those for traditional physiological traits. The derived QTLs are termed 'expression QTLs' (eQTLs). When performed in organisms with a sequenced genome, the combination of the eQTL together with the known physical position of the genes provides great opportunities for dissecting molecular regulation. eQTLs are divided into two groups: *cis*- and *trans*-eQTLs. *Cis*-eQTLs are those eQTLs of which the causal polymorphism is inside the gene for which expression differences are measured. In contrast, *trans*-eQTLs are eQTLs of which the causal polymorphism is not in the gene for which expression differences are measured, but, for example, in a transcription factor causing these expression differences (West *et al.*, 2007) (Fig. 1). Although the expression is measured at the gene level, QTL mapping remains dependent on the recombination frequency in the population, resulting in a confidence interval for each QTL that often comprises a large genomic region. If a transcription factor, causing expression differences for a specific gene, is located inside the confidence interval of the eQTL for this particular gene, this *trans*-eQTL cannot be distinguished from a *cis*-eQTL for the same gene. Therefore, in the absence of allele-specific expression data (see elsewhere in this review), it is better to use the terms 'local' and 'distant' eQTL (Rockman and Kruglyak, 2006) (Fig. 1).

Several eQTL analyses have been conducted so far for Arabidopsis (Keurentjes *et al.*, 2007; West *et al.*, 2007), but also for different crop species such as maize (Shi *et al.*, 2007), wheat (Jordan *et al.*, 2007) and barley (Potokina *et al.*, 2008). Besides transcriptomic data, the data of other -omics technologies can also be used for the genetical genomics approach, for example proteomics (pQTL) and metabolomics (mQTL). The power and possibilities of large-scale untargeted metabolomics analysis of genetic populations to reveal mQTLs are reviewed by Keurentjes (2009) and pQTL studies are described by Bourgeois *et al.* (2011) and references therein. An overview of the different aspects of genetical genomics in more depth is given by Joosen *et al.* (2009) and Kliebenstein (2009).

A good example of the power of genetical genomics is described by Jiménez-Gómez *et al.* (2010) who identified *EARLY FLOWERING 3* (*ELF3*) as the most
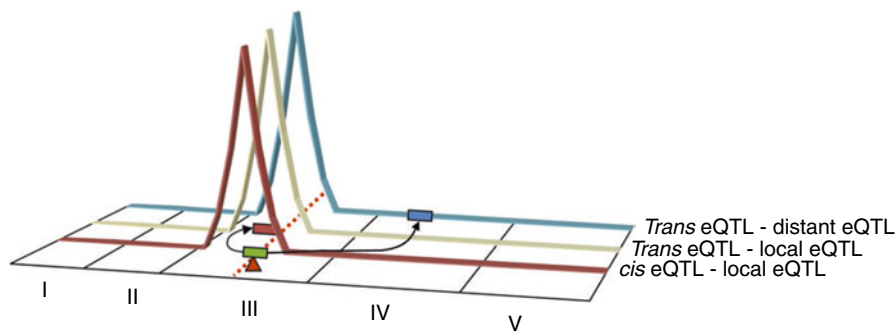

Figure 1. Classification of eQTLs. Allelic variation of the gene under study results in a *cis*-eQTL. All other eQTLs are *trans*, but *trans*-eQTLs of genes located in the confidence interval of the eQTL cannot be distinguished from *cis*-eQTLs in the absence of allele-specific expression information and therefore both should be termed 'local eQTLs'. (A colour version of this figure can be found online at http://journals.cambridge.org/ssr).

likely candidate gene affecting the shade avoidance response of Arabidopsis in a Bayreuth-0 × Shahdara population. To narrow down to *ELF3* as the only candidate causal gene for a shade avoidance QTL identified in this population, they combined publicly available datasets to perform network analysis with eQTL data (West *et al.*, 2007), co-expression analysis (Winter *et al.*, 2007) and functional classification (Ashburner *et al.*, 2000). Drastically narrowing down the number of candidate genes with this kind of approach is feasible for all QTLs where the polymorphism(s) in the causal gene result in differential gene expression of the same gene. However, this approach will not be applicable to the cases where the alleles causal for a QTL do not have an effect on gene expression, but on activity or stability of the encoded protein. In these cases, other levels, such as pQTL or mQTL, and other data types, including protein–protein interactions and metabolic pathways, can help to narrow down to the causal genes.

One of the limitations of a standard genetical genomics approach is that it is only performed for a single developmental stage or environment. Since most phenotypes are not solely the result of the status of a transcriptome, proteome or metabolome at a single stage, it is difficult to choose the most suitable developmental stage. Li *et al.* (2008) have proposed a generalized genetical genomics approach, which enables the analysis of several environments or developmental stages in a single -omics QTL approach. This enriches the genetical genomics approach with the potential to study the dynamics of molecular networks. This type of molecular network is complementary to co-expression networks (Usadel *et al.*, 2009) that are based on the correlation of gene expression. Co-expressed genes over a wide range of developmental stages and environments have a likelihood of being involved in the same biochemical/developmental pathways, as was shown elegantly for the co-expression network built from microarray data of 138 seed-related samples (Bassel *et al.*, 2011). In addition to the information gained about genetic mechanisms underlying natural variation in gene expression, eQTL studies also provide additional genotypic marker information of every used line by the detection of transcript-derived markers (TDMs) (Potokina *et al.*, 2009) in the form of single-feature polymorphisms (SFPs) (Borevitz *et al.*, 2003) or gene expression markers (West *et al.*, 2006) without the need for additional experiments.

The use of microarrays and next-generation sequencing in genetical genomics

All performed eQTL studies in plants so far have used microarray analysis or, in one case, cDNA-amplified fragment length polymorphism (AFLP) mapping (Vuylsteke *et al.*, 2006). Depending on the type of array used, this allows the determination of the expression of most of the genes expressed for the studied organism. Most microarrays will give an expression value per gene, providing the basic information needed for an eQTL study. However, more information can be obtained when using whole genome tiling arrays (Mockler and Ecker, 2005). Since these cover the whole genome, independent of any prior annotation of genes, they will be able to analyse expression of genes independent of their annotation. This will not only provide additional information about the expression of unannotated genes (Laubinger *et al.*, 2008; Matsui *et al.*, 2008), but also about alternative splicing (Zhang *et al.*, 2008), as was shown by an eQTL study for *Caenorhabditis elegans* which revealed heritable variation in alternative splicing (Li *et al.*, 2010a). For *Arabidopsis thaliana* a so-called SNPtile microarray was developed (Zhang *et al.*, 2007). Besides tiling of the whole genome, this array also harbours probes for 250,000 SNPs (single nucleotide polymorphisms) and 130,000 CCGG sites for methylation analysis. A genetical genomics study using these arrays will reveal the genetic variation for gene expression and alternative splicing, but with a few additional hybridizations this array will also provide data about allele-specific expression (ASE) and epigenetic polymorphisms. ASE studies help in distinguishing *cis*-eQTLs from local *trans*-eQTLs, where the physical position of the gene under study is within its eQTL confidence interval (see Fig. 1). For this purpose, genome-wide allele-specific expression is measured in an F1 hybrid of the two parents of the RIL population under study. Since both parental alleles share the same genetic background in F1 hybrids and are therefore equally exposed to *trans*-acting factors, any difference in expression from the two different alleles will be the result of a *cis*-eQTL (Zhang and Borevitz, 2009).

Although eQTL studies using microarrays give a wealth of information, RNA sequencing for eQTL studies will increase even further the information that can be gained from this type of study. The first eQTL studies using RNA sequencing have already been performed to study gene expression in *Drosophila* and humans (McManus *et al.*, 2010; Montgomery *et al.*, 2010; Pickrell *et al.*, 2010). These studies show the power of this approach for the analysis of variation in transcription, and a more detailed analysis of variation in splicing and allele-specific expression in comparison to whole-genome tiling arrays. A further advantage of using sequence-based techniques for genetical genomics studies is that they do not rely on the availability of microarrays for the species under study. In fact, one doesn't even need any prior sequence knowledge.

In conclusion, genetical genomic approaches will prove to be especially powerful for model species with a known genome, such as Arabidopsis and tomato, but recent and future developments in second- and third-generation sequencing technologies (Metzker, 2010; Zhang *et al.*, 2011) will open the path towards a successful implementation of genetical genomics approaches for non-model organisms (Varshney *et al.*, 2009).

Integrating genetical genomics with genome-wide association studies

An attractive complement to QTL mapping with the use of RIL populations is linkage disequilibrium (LD) or genome-wide association (GWA) mapping. GWA mapping connects particular ancestral haplotypes to variations in quantitative traits (Hamblin *et al.*, 2011; Ingvarsson and Street, 2011). In GWA studies, mapping populations are used that consist of several hundred to several thousand (wild) accessions or breeding lines. Compared to mapping with the help of RIL populations, where the variation is confined to the two parents of the population, the variation found in a GWA population is much higher. Furthermore, because of the large number of meiosis occurrences in the history of a GWA population, the resolution

(linkage decay) of the mapping can be as small as 1–300 kb (Buckler and Gore, 2007) which is in huge contrast with the sometimes 10–30 cM confidence intervals for RIL populations, possibly harbouring thousands of genes. As a result of the increased resolution in GWA studies, the number of markers needed in these studies also increases dramatically. This number varies per species and/or population, but can rise up to 750,000 for various maize land races (Sorkheh *et al.*, 2008). Besides the obvious advantages of GWA studies, they have the problem that they generate false positives due to the often complex population structure that can strongly influence the estimation of linkage disequilibrium. Although several approaches have been developed for taking the population structure into account, it is still under debate how to distinguish between false and true positives (Shriner *et al.*, 2007). Furthermore, GWA studies have a reduced statistical power of finding associations as compared to RIL populations, especially for rare alleles that are only found in a few accessions. Besides this reduced power it is often also difficult to detect non-functional alleles in GWA studies (Atwell *et al.*, 2010; Brachi *et al.*, 2010), caused by the fact that genes can become inactive through many independent deletions, insertions or other kinds of null mutations. Despite these disadvantages and the problems described in the literature for GWA
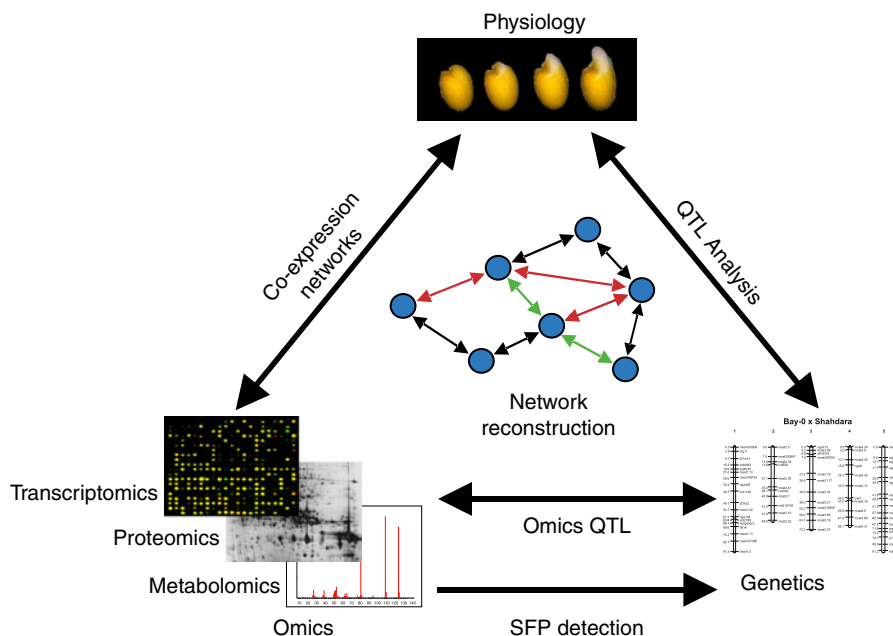

Figure 2. The power of genetical genomics. Phenotypic data of experimental populations can be linked with genotype information to perform quantitative trait locus (QTL) analysis. Omics data linked with phenotypic data can be used to build phenotype related co-expression networks. Omics data can be used to extract genotypic data [single-feature polymorphism (SFP) detection]. The combination of -omics data with genotypic data results in -omics QTL and, finally, all the information together can be used for the reconstruction of molecular networks involved in the physiological phenomenon under study. (A colour version of this figure can be found online at http://journals.cambridge.org/ssr).

studies in humans, GWA studies in plants have resulted in promising results, as shown in Arabidopsis (Atwell *et al.*, 2010; Baxter *et al.*, 2010; Brachi *et al.*, 2010; Li *et al.*, 2010b) and rice (Huang *et al.*, 2010b).

Concluding remarks

Since the introduction of the concept of genetical genomics, it has proved to be a powerful approach to dissect genetic variation. The genetical genomics studies in model species help us to understand the extent of genetic variation and support the development of tools for analysis. This information may then be applied to studies in crop species. The integration of extensive phenotyping with detailed genetic maps and -omics tools, such as transcriptomics, proteomics and metabolomics, will enable accurate and detailed network reconstruction and subsequent unravelling of the genetic and molecular mechanisms underlying complex physiological traits (Fig. 2).

Recent developments in inexpensive high-throughput sequencing and development of tiling microarrays combined with SNP probes and CCGG sites for methylation will soon create opportunities to extend genetical genomics to unravel the genetic variation for gene expression, alternative splicing, allele-specific expression and epigenetic polymorphisms, and allow eQTL mapping in GWA studies. We believe that the combination of global analysis of phenotypic variation and its associated alleles in GWA studies, with a more detailed and in-depth study in populations obtained from experimental crosses, such as RIL populations, will be of tremendous value for unravelling the molecular mechanisms underlying complex traits such as seed quality. Ultimately, this increased knowledge about the factors influencing seed quality will open new possibilities for the breeding industry to understand and control the effects of the maternal environment on seed quality and, above all, allow breeding for high-quality seeds.

Acknowledgements

References

Alonso-Blanco, C. and Koornneef, M. (2000) Naturally occurring variation in Arabidopsis: An underexploited resource for plant genetics. *Trends in Plant Science* **5**, 22–29.

Argyris, J., Truco, M., Ochoa, O., Knapp, S., Still, D., Lenssen, G., Schut, J., Michelmore, R. and Bradford, K. (2005) Quantitative trait loci associated with seed and seedling traits in Lactuca. *TAG Theoretical and Applied Genetics* **111**, 1365–1376.

Argyris, J., Dahal, P., Hayashi, E., Still, D.W. and Bradford, K.J. (2008) Genetic variation for lettuce seed thermoinhibition is associated with temperature-sensitive expression of abscisic acid, gibberellin, and ethylene biosynthesis, metabolism, and response genes. *Plant Physiology* **148**, 926–947.

Armengaud, P., Zambaux, K., Hills, A., Sulpice, R., Pattison, R.J., Blatt, M.R. and Amtmann, A. (2009) EZ-Rhizo: integrated software for fast and accurate measurement of root system architecture. *Plant Journal* **57**, 945–956.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nature Genetics* **25**, 25–29.

Atwell, S., Huang, Y.S., Vilhjalmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., Jiang, R., Muliyati, N.W., Zhang, X., Amer, M.A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., de Meaux, J., Ecker, J.R., Faure, N., Kniskern, J.M., Jones, J.D.G., Michael, T., Nemri, A., Roux, F., Salt, D.E., Tang, C., Todesco, M., Traw, M.B., Weigel, D., Marjoram, P., Borevitz, J.O., Bergelson, J. and Nordborg, M. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–631.

Bassel, G.W., Lan, H., Glaab, E., Gibbs, D.J., Gerjets, T., Krasnogor, N., Bonner, A.J., Holdsworth, M.J. and Provart, N.J. (2011) Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. *Proceedings of the National Academy of Sciences, USA* **108**, 9709–9714.

Baxter, I., Brazelton, J.N., Yu, D., Huang, Y.S., Lahner, B., Yakubova, E., Li, Y., Bergelson, J., Borevitz, J.O., Nordborg, M., Vitek, O. and Salt, D.E. (2010) A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter athkt1;1. *PLoS Genetics* **6**, e1001193.

Bentsink, L., Hanson, J., Hanhart, C.J., Blankestijn-de Vries, H., Coltrane, C., Keizer, P., El-Lithy, M., Alonso-Blanco, C., de Andres, M.T., Reymond, M., van Eeuwijk, F., Smeekens, S. and Koornneef, M. (2010) Natural variation for seed dormancy in Arabidopsis is regulated by additive genetic and molecular pathways. *Proceedings of the National Academy of Sciences, USA* **107**, 4264–4269.

Borevitz, J.O., Liang, D., Plouffe, D., Chang, H.S., Zhu, T., Weigel, D., Berry, C.C., Winzeler, E. and Chory, J. (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research* **13**, 513–523.

Bourgeois, M., Jacquin, F., Cassecuelle, F., Savois, V., Belghazi, M., Aubert, G., Quillien, L., Huart, M., Marget, P. and Burstin, J. (2011) A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. *Proteomics* **11**, 1581–1594.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J. and Roux, F. (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics* **6**, e1000940.

Buckler, E. and Gore, M. (2007) An Arabidopsis haplotype map takes root. *Nature Genetics* **39**, 1056–1057.

Clerkx, E.J.M., El-Lithy, M.E., Vierling, E., Ruys, G.J., Blankestijn-De Vries, H., Groot, S.P.C., Vreugdenhil, D. and Koornneef, M. (2004) Analysis of natural allelic variation of Arabidopsis seed germination and seed longevity traits between the accessions Landsberg erecta and Shakdara, using a new recombinant inbred line population. *Plant Physiology* **135**, 432–443.

Dell'Aquila, A. (2009) Digital imaging information technology applied to seed germination testing. A review. *Agronomy for Sustainable Development* **29**, 213–221.

Dias, P., Brunel-Muguet, S., Dürr, C., Huguet, T., Demilly, D., Wagner, M.-H. and Teulat-Merah, B. (2011) QTL analysis of seed germination and pre-emergence growth at extreme temperatures in *Medicago truncatula*. *Theoretical and Applied Genetics* **122**, 429–444.

Dickson, M.H. (1980) Genetic-aspects of seed quality. *Hortscience* **15**, 771–774.

Doganlar, S., Frary, A. and Tanksley, S.D. (2000) The genetic basis of seed-weight variation: tomato as a model system. *Theoretical and Applied Genetics* **100**, 1267–1273.

Foolad, M.R., Zhang, L.P. and Subbiah, P. (2003) Genetics of drought tolerance during seed germination in tomato: inheritance and QTL mapping. *Genome* **46**, 536–545.

Foolad, M.R., Subbiah, P. and Zhang, L. (2007) Common QTL affect the rate of tomato seed germination under different stress and nonstress conditions. *International Journal of Plant Genomics* **2007**, 97386.

French, A., Ubeda-Tomas, S., Holman, T.J., Bennett, M.J. and Pridmore, T. (2009) High-throughput quantification of root growth using a novel image-analysis tool. *Plant Physiology* **150**, 1784–1795.

Galpaz, N. and Reymond, M. (2010) Natural variation in *Arabidopsis thaliana* revealed a genetic network controlling germination under salt stress. *PLoS ONE* **5**, e15198.

Gupta, V., Mathur, S., Solanke, A.U., Sharma, M.K., Kumar, R., Vyas, S., Khurana, P., Khurana, J.P., Tyagi, A.K. and Sharma, A.K. (2009) Genome analysis and genetic enhancement of tomato. *Critical Reviews in Biotechnology* **29**, 152–181.

Hamblin, M.T., Buckler, E.S. and Jannink, J.-L. (2011) Population genetics of genomics-based crop improvement methods. *Trends in Genetics* **27**, 98–106.

Harada, J.J. (1997) Seed maturation and control of germination. pp. 545–592 *in* Larkins, B.; Vasil, I. (Eds) *Cellular and molecular biology of plant seed development*. Dordrecht, Kluwer Academic Publishers.

Hilhorst, H.W.M. and Toorop, P.E. (1997) Review on dormancy, germinability, and germination in crop and weed seeds. *Advances in Agronomy* **61**, 111–165.

Huang, X., Schmitt, J., Dorn, L., Griffith, C., Effgen, S., Takao, S., Koornneef, M. and Donohue, K. (2010a) The earliest stages of adaptation in an experimental plant population: Strong selection on QTLs for seed dormancy. *Molecular Ecology* **19**, 1335–1351.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q.,

Liu, K., Huang, T., Zhou, T., Jing, Y., Lin, Z., Buckler, E.S., Qian, Q., Zhang, Q.F., Li, J. and Han, B. (2010b) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* **42**, 961–967.

Ingvarsson, P.K. and Street, N.R. (2011) Association genetics of complex traits in plants. *New Phytologist* **189**, 909–922.

Jansen, R.C. and Nap, J.P. (2001) Genetical genomics: the added value from segregation. *Trends in Genetics* **17**, 388–391.

Jiménez-Gómez, J.M., Wallace, A.D. and Maloof, J.N. (2010) Network analysis identifies *ELF3* as a QTL for the shade avoidance response in Arabidopsis. *PLoS Genetics* **6**, e1001100.

Joosen, R.V., Ligterink, W., Hilhorst, H.W. and Keurentjes, J.J. (2009) Advances in genetical genomics of plants. *Current Genomics* **10**, 540–549.

Joosen, R.V., Kodde, J., Willems, L.A., Ligterink, W., van der Plas, L.H. and Hilhorst, H.W. (2010) Germinator: a software package for high-throughput scoring and curve fitting of arabidopsis seed germination. *Plant Journal* **62**, 148–159.

Jordan, M.C., Somers, D.J. and Banks, T.W. (2007) Identifying regions of the wheat genome controlling seed development by mapping expression quantitative trait loci. *Plant Biotechnology Journal* **5**, 442–453.

Keurentjes, J.J.B. (2009) Genetical metabolomics: closing in on phenotypes. *Current Opinion in Plant Biology* **12**, 223–230.

Keurentjes, J.J., Fu, J., Terpstra, I.R., Garcia, J.M., van den Ackerveken, G., Snoek, L.B., Peeters, A.J., Vreugdenhil, D., Koornneef, M. and Jansen, R.C. (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proceedings of the National Academy of Sciences, USA* **104**, 1708–1713.

Kliebenstein, D. (2009) Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. *Annual Review of Plant Biology* **60**, 93–114.

Laubinger, S., Zeller, G., Henz, S., Sachsenberg, T., Widmer, C., Naouar, N., Vuylsteke, M., Scholkopf, B., Ratsch, G. and Weigel, D. (2008) At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in *Arabidopsis thaliana*. *Genome Biology* **9**, R112.

Li, Y., Breitling, R. and Jansen, R.C. (2008) Generalizing genetical genomics: getting added value from environmental perturbation. *Trends in Genetics* **24**, 518–524.

Li, Y., Breitling, R., Snoek, L.B., van der Velde, K.J., Swertz, M.A., Riksen, J., Jansen, R.C. and Kammenga, J.E. (2010a) Global genetic robustness of the alternative splicing machinery in *Caenorhabditis elegans*. *Genetics* **186**, 405–410.

Li, Y., Huang, Y., Bergelson, J., Nordborg, M. and Borevitz, J.O. (2010b) Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **107**, 21199–21204.

Matsui, A., Ishida, J., Morosawa, T., Mochizuki, Y., Kaminuma, E., Endo, T.A., Okamoto, M., Nambara, E., Nakajima, M., Kawashima, M., Satou, M., Kim, J.-M., Kobayashi, N., Toyoda, T., Shinozaki, K. and Seki, M. (2008) Arabidopsis transcriptome analysis under

drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant and Cell Physiology* **49**, 1135–1149.

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R. and Wittkopp, P.J. (2010) Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Research* **20**, 816–825.

Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nature Reviews Genetics* **11**, 31–46.

Mockler, T.C. and Ecker, J.R. (2005) Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**, 1–15.

Montgomery, S.B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R.P., Ingle, C., Nisbett, J., Guigo, R. and Dermitzakis, E.T. (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777.

Nesi, N., Delourme, R., Brégeon, M., Falentin, C. and Renard, M. (2008) Genetic and molecular approaches to improve nutritional value of *Brassica napus* L. seed. *Comptes Rendus Biologies* **331**, 763–771.

Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y. and Pritchard, J.K. (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772.

Potokina, E., Druka, A., Luo, Z., Wise, R., Waugh, R. and Kearsey, M. (2008) Gene expression quantitative trait locus analysis of 16,000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. *Plant Journal* **53**, 90–101.

Potokina, E., Druka, A. and Kearsey, M.J. (2009) Transcript profiling and expression level mapping. pp. 81–92 *in* Somers, D.J.; Langridge, P.; Gustafson, J.P. (Eds) *Plant genomics*. New York, Humana Press.

Powell, A.A. (2006) Seed vigor and its assessment. pp. 603–648 *in* Basra, A.S. (Ed.) *Handbook of seed science and technology*. Binghamton, USA, Food Products Press.

Rockman, M.V. and Kruglyak, L. (2006) Genetics of global gene expression. *Nature Reviews Genetics* **7**, 862–872.

Salvi, S. and Tuberosa, R. (2005) To clone or not to clone plant QTLs: present and future challenges. *Trends in Plant Science* **10**, 297–304.

Shi, C., Uzarowska, A., Ouzunova, M., Landbeck, M., Wenzel, G. and Lübberstedt, T. (2007) Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint × Flint maize recombinant inbred line population. *BMC Genomics* **8**, 22.

Shriner, D., Vaughan, L.K., Padilla, M.A. and Tiwari, H.K. (2007) Problems with genome-wide association studies. *Science* **316**, 1840–1842.

Sorkheh, K., Malysheva-Otto, L.V., Wirthensohn, M.G., Tarkesh-Esfahani, S. and Martínez-Gómez, P. (2008)

Linkage disequilibrium, genetic association mapping and gene localization in crop plants. *Genetics and Molecular Biology* **31**, 805–814.

Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S. and Provart, N.J. (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant, Cell & Environment* **32**, 1633–1651.

Vandecasteele, C., Teulat-Merah, B., Morere-LePaven, M.C., Leprince, O., LyVu, B., Viau, L., Ledroit, L., Pelletier, S., Payet, N., Satour, P., Lebras, C., Gallardo, K., Huguet, T., Limami, A.M., Prosperi, J.M. and Buitink, J. (2011) QTL analysis reveals a correlation between the ratio of sucrose/raffinose family oligosaccharides and seed vigour in *Medicago truncatula*. *Plant, Cell & Environment* **34**, 1473–1487.

Varshney, R.K., Nayak, S.N., May, G.D. and Jackson, S.A. (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* **27**, 522–530.

Vuylsteke, M., Van Den Daele, H., Vercauteren, A., Zabeau, M. and Kuiper, M. (2006) Genetic dissection of transcriptional regulation by cDNA-AFLP. *Plant Journal* **45**, 439–446.

West, M.A.L., Van Leeuwen, H., Kozik, A., Kliebenstein, D.J., Doerge, R.W., St. Clair, D.A. and Michelmore, R.W. (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. *Genome Research* **16**, 787–795.

West, M.A.L., Kim, K., Kliebenstein, D.J., Van Leeuwen, H., Michelmore, R.W., Doerge, R.W. and St. Clair, D.A. (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. *Genetics* **175**, 1441–1450.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J. (2007) An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* **2**, e718.

Zeng, D.L., Guo, L.B., Xu, Y.B., Yasukumi, K., Zhu, L.H. and Qian, Q. (2006) QTL analysis of seed storability in rice. *Plant Breeding* **125**, 57–60.

Zhang, J., Chiodini, R., Badr, A. and Zhang, G. (2011) The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics* **38**, 95–109.

Zhang, X. and Borevitz, J.O. (2009) Global analysis of allele-specific expression in *Arabidopsis thaliana*. *Genetics* **182**, 943–954.

Zhang, X., Richards, E.J. and Borevitz, J.O. (2007) Genetic and epigenetic dissection of cis regulatory variation. *Current Opinion in Plant Biology* **10**, 142–148.

Zhang, X., Byrnes, J., Gal, T., Li, W.-H. and Borevitz, J. (2008) Whole genome transcriptome polymorphisms in *Arabidopsis thaliana*. *Genome Biology* **9**, R165.