doi: 10.17221/192/2016-CJGPB

Application of Next-Generation Sequencing in Plant Breeding

*David VLK and Jana ŘEPKOVÁ**

Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czech Republic
**Corresponding author: repkova@sci.muni.cz*

Abstract

Vlk D., Řepková J. (2017): Application of next-generation sequencing in plant breeding. Czech J. Genet. Plant Breed., 53: 89–96.

In the past decade, next-generation sequencing (NGS) platforms have changed the impact of sequencing on our knowledge of crop genomes and gene regulation. These techniques are today acquiring a great potential in metagenomic and agrigenomic research while showing prospects for their utilization in plant breeding. We can now obtain new and beneficial information about gene regulation on the cellular as well as whole-plant level through RNA-sequencing and subsequent expression analyses of genes participating in plant defence reactions to pathogens and in abiotic stress tolerance. NGS has facilitated the development of methods to genotype very large numbers of single-nucleotide polymorphisms. Genotyping- by-sequencing and whole-genome resequencing can lead to the development of molecular markers suited to studies of genetic relationships among breeding materials, creation of detailed genetic mapping of targeted genes and genome-wide association studies. Plant genotyping can benefit plant breeding through selection of individuals resistant to climatic stress and to pathogens causing substantial losses in agriculture.

Keywords: gene expression; marker-assisted selection; molecular marker; RNA-sequencing; SNP

Next-generation sequencing techniques began to develop in response to a need for capability to sequence larger numbers of samples at lower cost. The followed Sanger-based methods have not been suitable for processing huge numbers of samples, and massive parallel signature sequencing (MPSS; Brenner *et al.* 2000) also has limitations. At the beginning of the 21st century, the Roche Company introduced its Genome Sequencer. This was the first commercially available next-generation instrument, performing what was known as 454 sequencing. The following year, the Solexa Company developed its Genome Analyzer. A short time later Solexa was acquired by Illumina. The third next-generation device was created by Applied Biosystems and was called SOLiD (for Sequencing by Oligo Ligation and Detection). Since 2007, the market has been enriched by devices known as "third-generation" sequencing. Unlike next-generation sequencing, these devices do not require the amplification of a sample and enable sequencing of a single DNA molecule. While next-generation sequencing generates short reads a few hundred base pairs long, the third-generation technologies produce over 10 000 bp reads, allowing production of highly accurate *de novo* assemblies and generation of more contiguous reconstruction of the genomes with a high content of repetitive elements. The first third-generation sequencer was HeliScope by Helicos Biosciences. Then followed a PacBio RS sequencer from Pacific Biosciences. Today, the most advanced sequencing technologies include semiconductor sequencing by Ion Torrent and nanopore sequencing by Oxford Nanopore Technologies and Nabsys. The history of sequencing technologies was reviewed by Kumar *et al.* (2012). These technologies have been directed to clinical diagnosis of genes conferring human diseases, forensic genomics, metagenomics, epigenetics, and expression analyses. In the plant research area, next-generation sequencing (NGS) technologies have become important tools

doi: 10.17221/192/2016-CJGPB

for assembly of crop reference genomes, transcriptome sequencing for the study of gene expression, whole-genome molecular marker development, and identification of markers in known-function genes. Some of these have become useful in the breeding of various crops.

Reference genomes

Recent sequencing projects have expended a tremendous amount of effort to sequence more complex genomes. The plant genomes consist of high content of repetitive elements due to the high copy number and amplifying nature of transposable elements with frequent segmental or tandem duplication. Ploidy is another challenge for sequencing projects, and results are dependent on many aspects, such as autopolyploid or allopolyploid character of the genome or the age of ploidization event. This complexity of the genomes had been a problem for a long time and it needed to be reduced involving sequencing library with partial representation of the genome using restriction enzymes or capturing sequences without enzyme digestion (Ray & Satya 2014). Many projects aimed to generate reference genome sequences for the species of interest. A reference genome sequence is an important tool for exploring genome structure and function, as well as to guide the genome assembly of closely related species. Moreover, the availability of reference genome sequences enables the mining of large amounts of molecular markers and candidate genes. Resequencing projects are more suited to pre-breeding activities and are directed to identifying genomic variations while inferring information about useful polymorphisms. To date, approximately 100 plant species have been sequenced into draft genome sequences.

Transcriptome research

RNA-sequencing (RNA-seq) is a relatively new method for both quantifying and mapping transcriptomes, which uses recently developed deep-sequencing technologies. This approach consists of converting RNA molecules to a library of cDNA fragments with adaptors, these fragments are sequenced, and the resulting reads are either aligned to a reference genome, or assembled *de novo* (Wang *et al.* 2009b). RNA-seq is used to obtain expressed sequence data in a specific tissue within a defined time. Moreover, this is possible for species even without a reference genome (Novaes *et al.* 2008). *De novo* transcriptome assembly using NGS data is an attractive option for the study of large and complex genomes. Roche technology was successfully used in sequencing a series of non-model plants, for instance in comparative sequencing of transcripts from two olive trees during fruit development (*Olea europaea* L.; Alagna *et al.* 2008) and in transcriptome analysis of the bread wheat cultivar Yunong 201 (*Triticum aestivum* L.; Zhang *et al.* 2016). In addition to Sanger sequencing technique associated with the identification of expressed sequence tags (EST; Swarbreck *et al.* 2011), Illumina technology is generally useful for its better coverage of plant transcriptomes.

Data acquired by RNA-seq are universal. Furthermore, they can be used in gene characterization (Dassanayake *et al.* 2009) and molecular marker development (Trick *et al.* 2009).

There are tools that provide user-friendly interfaces for gene discovery in *de novo* transcriptomes, such as Trapid (Van Bel *et al.* 2013) and TrinotateWeb (http://trinotate.github.io). Kamei *et al.* (2016) developed a tool that enables molecular breeders even without extensive bioinformatics knowledge to efficiently study *de novo* transcriptome data from any crop (Orphan Crop Browser; http://www.bioinformatics.nl) with a large and complex genome. They used that tool to identify the putative orthologues of 17 known lignin biosynthetic genes from maize and sugarcane in the orphan crop *Miscanthus sinensis* Andersson.

Identification of expressed genes

NGS technologies and RNA-seq enable the study of gene expression, which is becoming an important tool for plant breeding and identification of genes of interest conferring defence mechanisms against biotic and abiotic stresses. A study focused upon the pathogen *Puccinia striiformis* f.sp. *tritici* Erikss., which causes extensive damage in wheat, used RNA-seq to find genes encoding effector proteins and which may be useful in breeding wheat varieties resistant to this pathogen (Garnica *et al.* 2013). A number of genes associated with phases of development were identified in the transcriptome analyses of cucumber (*Cucumis sativus* L.) using 454 sequencing (Ando *et al.* 2012). Nigam *et al.* (2014) used a combination of microarray and Roche technology to identify genes and their products associated with the quality of cotton fibre.

doi: 10.17221/192/2016-CJGPB

To identify genes responsible for drought tolerance, Tang *et al.* (2013) used Roche's 454-GS FLX System in an RNA-seq analysis of *Populus euphratica* Oliv., which grows in arid or semiarid regions. Similarly, a transcriptome of red clover (*Trifolium pratense* L.) was sequenced using Illumina technology and genes responsible for drought tolerance were discovered. Thus, three metabolites (pinitol, proline, and malate) whose concentrations increased in leaves as an impact of drought stress were identified (Yates *et al.* 2014). Soil salinity is becoming a major problem in many regions and, therefore, several studies have been directed to discovering a molecular mechanism of salt tolerance in plants. Such a mechanism was ascertained for example in soybean (*Glycine max* /L./ Merr.; Fan *et al.* 2012), cotton (*Gossypium aridum* Rose & Standl; Xu *et al.* 2013a), and the halophyte turf grass *Sporobolus virginicus* (L.) Kunth (Yamamoto *et al.* 2015). In addition, Illumina technology was used to identify genes responsible for copper tolerance (Wang *et al.* 2015) and for metabolism-based herbicide resistance in *Lolium rigidum* Gaudin (Gaines *et al.* 2014). In the context of studying plant development, Illumina technology was used in a whole-genome study exploring the function of plant-specific NAC transcription factor family during development and dehydration stress in soybean (Le *et al.* 2011).

In addition to Roche and Illumina technology, Ion Torrent was used in transcriptome analysis of finger millet (*Eleusine coracana* L.), which is a hardy cereal known for its tolerance to salinity, drought, and diseases (Rahman *et al.* 2014). This technology was further utilized in transcriptome profiling of Jatropha roots (*Jatropha curcas* L.) to elucidate molecular responses to waterlogging (Juntawong *et al.* 2014). Finally, SMRT technology by Pacific Biosciences was used in studying the interaction of the bacterial pathogen *Xanthomonas oryzae* pv. *oryzicola* and its host, *Oryza sativa* L., using whole-genome sequencing of the pathogen and RNA-sequencing of the attacked host (Wilkins *et al.* 2015).

Study of epigenetic regulation

Epigenetic changes are responsible for alternations in gene regulation. Epigenetics includes some stable changes in the structure of proteins (prions), expression of small RNAs, and modification of chromatins (i.e. DNA methylation and adjustment of histone tails like in the cases of acetylation, methylation, ubiquitination, and phosphorylation).

Traditional methods for studying epigenetics comprise methylation-sensitive restriction enzymes, antibodies specific to 5-methylcytosine, and bisulphite conversion. These have been coupled with microarray-based methods enabling the genome-wide analysis of DNA methylation (ChIP-chip method; Buck & Lieb 2004). The emergence of NGS brought a breakthrough for studying epigenetics, and these technologies have become important tools for ChIP-seq when high coverage of sequence reads is required. Illumina technology was used in an extensive study of organ-specific epigenetic modifications and their impacts on mRNA and sRNA in maize (Wang *et al.* 2009a). An analysis of methylated regions in the tomato genome was performed by the combined technique of bisulphite conversion and Illumina sequencing, whereby it was demonstrated that epigenetic regulation along with hormonal treatment controls the ripening of tomato fruits (Zhong *et al.* 2013).

Small (~25 nt) endogenous RNAs known as micro RNAs (miRNAs) have emerged as key post-transcriptional regulators in eukaryotic gene expression. They appear to be the principal regulators of development and various stress responses. The majority of miRNAs are highly conserved and complementary with the target mRNAs. In plants, these sites are mostly in coding regions and less frequently in the 5' untranslated region. Several approaches can be used for the identification and verification of miRNAs. *In silico* prediction based on conserved sequences and secondary structures is commonly used, and these fast and low-cost methods have been successfully applied in rice (Bonnet *et al.* 2004). Another possibility is to create miRNA libraries and to follow this with cloning and sequencing, but this approach is limited by the low expression of these molecules, which is moreover time- and tissue-specific.

Using Illumina technology, the identification of potential cadmium-responsive miRNAs and their target genes in radish (*Raphanus sativus* L.) roots has been performed (Xu *et al.* 2013b). SMRT technology by Pacific Biosciences was used with the aim of identifying circular RNAs using transcriptome analysis. These molecules play an important role in the function of miRNA and transcriptional control because they act as competitive endogenous RNA and as positive and negative regulators also of their parent genes (Lu *et al.* 2015).

Recent studies have revealed another important regulatory mechanism represented by long non-

doi: 10.17221/192/2016-CJGPB

coding RNAs (lncRNAs). These RNA molecules are longer than 200 bp and do not encode any protein product. The recent studies have linked them to such processes as gene silencing, flowering time regulation, and abiotic stress responses (Wang *et al*. 2014; Zhang *et al*. 2014). Identification of lncRNAs is performed using tiling array, EST analyses, and RNA sequencing. Recently, these molecules were identified in some crops, including wheat (Xin *et al.* 2011), rape mustard (Yu *et al*. 2013), apple (Celton *et al*. 2014), and poplar (Shuai *et al*. 2014).

Mining of molecular markers

There are many types of molecular markers, but single-nucleotide polymorphisms (SNP) and simple sequence repeats (SSR) are the most widely used. Mining of molecular markers through NGS was originally limited to model species of *Arabidopsis* and rice. Molecular markers have been gradually discovered even in species without reference genomes, for example in durum wheat (*Triticum durum* Desf.; Trebbi *et al.* 2011), common bean (*Phaseolus vulgaris* L.; Cortés *et al.* 2011), and red clover (*Trifolium pratense* L.; Ištvánek *et al*. 2014, 2017). Nevertheless, mining of molecular markers in many other economically important species remains limited by the error rate of sequencing techniques due to the incomplete reference genome, content of repetitive elements, or mistakes in sequencing. Therefore, Azam *et al.* (2012) developed a new approach to searching SNPs in chickpea (*Cicer arietinum* L.) known as coverage-based consensus calling (CbCC). It consists of four freely available tools for local alignment: Maq, Bow-Tie, Novoalign, and SOAP2.

NGS offers several approaches that are capable of simultaneously performing genome-wide SNP discovery and genotyping in a single step. The most frequently used methods of genotyping utilize restriction enzymes to capture the reduced representation of a genome (Miller *et al*. 2007). A new approach known as genotyping-by-sequencing (GBS; Poland & Rife 2012) has been developed as a rapid and robust approach to sequencing of multiplexed samples. It combines genome-wide molecular marker discovery and genotyping (Davey *et al.* 2011; Elshire *et al.* 2011). The GBS approach includes the digestion of genomic DNA with restriction enzymes followed by ligation of barcode adapter, PCR amplification, and sequencing of the amplified DNA pool on a single lane of flow cells (He *et al*. 2014).

Its cost-effective nature makes GBS an excellent tool for many applications in breeding inasmuch as it can genotype thousands and even hundreds of thousands of SNPs in crop genomes and populations and then identify SNPs correlated with traits of interest. Thus, marker-assisted selection (MAS) could be widely applied to enhance crop yield, quality, and tolerance to biotic or abiotic stresses.

Current demands of plant breeding

Genome sequencing of important crops is becoming an initial step for ascertainment of the genome and evolution while ensuing resequencing steps allow elucidating genetic variability among individuals. Determination of the sequence further enables the targeted modification of specific genes using genome editing or identification of appropriate mutations in order to obtain a new allelic form.

Typical plant breeding programmes are mostly based on phenotyping, but, due to the growing knowledge of the genetic background of important agronomic traits, there has recently been an urgent demand for genotype-based selection (Myles *et al*. 2010). A key factor in breeding which uses high-throughput sequencing is to associate a large amount of genomic data with systematic characterization of phenotypes for a wide range of traits and conditions. For this purpose, high-throughput phenotyping platforms allow building a non-destructive record for a wide range of phenotypic traits over time using remote sensing and imaging techniques and specific software applications (Tisné *et al*. 2013; Petrozza *et al*. 2014).

There are two main strategies for identifying marker trait associations (MTA). The one exploits the genotyping of an entire segregating population with markers densely covering the whole genome, and following scrutinizing the associations between phenotypic differences and marker genotypes. This approach is time-consuming and extensive, and therefore the acquisition of precise phenotypic data at this scale may be logistically difficult. The other strategy is based on the genotyping of only that part of the population which manifests extreme phenotypes for target traits. MTA is then derived from allelic frequency differences between the groups of plants with contrasting phenotypes. In the last decade, the emerging high-resolution and cost-effective genotyping platforms have offered the opportunity for performing genome-wide association studies (GWAS). In different plant species, GWAS has been widely adopted to overcome some of the limitation

doi: 10.17221/192/2016-CJGPB

inherent in bi-parental linkage mapping (Lehmensiek *et al.* 2009), and it enables the direct utilization of plenty of MTA in crop design, since they are applicable to a wider germplasm base. In spite of many studies done in crops, the expected effect of a candidate gene has been verified only in a few cases (Chen *et al.* 2014) because several independent studies and pieces of evidence are needed for definitively assigning an SNP association signal to a candidate gene.

The next step, after sequencing and MTA detection, is molecular marker-based selection. There exist two main strategies for molecular selection. The one uses molecular markers which are located inside or nearby a locus with the known phenotypic effect, and this process is known as marker-assisted selection (MAS). This approach started to appear in plant breeding as a result of the fact that some traits are difficult to control through the standard phenotypic selection, and the expression of some traits is dependent on environmental conditions or developmental stages (Xu *et al.* 2003). Another complication is difficult maintenance of recessive alleles during backcrossing or pyramiding multiple monogenic traits. This process is utilized in selecting relatively small numbers of genes with the major phenotypic effect. MAS was usually connected with genetic mapping, and this process comprises multiple consecutive steps from development of mapping populations, genetic mapping, and marker validation to MAS application. The following integration of genetic mapping and MAS relied on combining multiple approaches such as linkage disequilibrium analysis of diverse genotypes or advanced backcross mapping (Xu & Crouch 2008). Two major advantages of the integration consist in the ability to carry out MTA using a breeding population and combining MTA development and validation within a single breeding program. In the last decade, MAS was used in several important crops, such as wheat (Kumar *et al.* 2010), apple (Flachowsky *et al.* 2011) or peanut (Chu *et al.* 2011). The other approach exploits all available markers as predictors of the genomic estimated breeding value (GEBV), and it is known as genomic selection (GS). To calculate GEBV, it is necessary to estimate all locus, haplotype and marker effects through the entire genome (Barabaschi *et al.* 2016). The key point is testing of different statistical models using the genotypic and phenotypic data from control populations to find the one which is able to predict GEBV most accurately, and where the correlation between GEBV and true breeding value reaches the highest levels (Heffner *et al.* 2009). This process has been applied successfully even to crops with large and complex genomes, such as maize (Windhausen *et al.* 2012), wheat (Lado *et al.* 2013), and sugar beet (Würschum *et al.* 2013).

Future perspectives in plant breeding

Another possibility for obtaining new allelic forms is genome editing, which is targeted gene modification to obtain a generation of new allelic variants in the genomes of cultivated individuals. It is supposed that the availability of genome sequences for many important crops, because of using NGS technologies, will facilitate genome editing approaches, because this technology depends on accurate sequence information for precise determination of the target position. Genome editing is based on the induction of double-strand breaks in a targeted locus using sequence-specific nucleases, such as zinc finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN; Ainley *et al.* 2013). Nuclease-induced breaks result in arrays of mutations (e.g. small insertions or deletions) at specific DNA sites. DNA breaks at multiple sites are also utilized in homologous recombination between chromosomal DNA and foreign donor DNA through the homologous recombination pathway. ZNF-induced mutagenesis for acetolactate synthase genes resulted in herbicide resistance in transformed tobacco plants (Townsend *et al.* 2009), and TALEN-mediated mutagenesis was employed to engineer tomato at genes related to gibberellin signalling (Lor *et al.* 2014).

An alternative genome editing strategy is executed using the CRISPR/Cas9 system. This technology, based on archaeal and bacterial clustered regularly interspaced short palindromic repeats (CRISPR) within the adaptive immune system, utilizes CRISPR-associated (Cas) proteins with endonuclease activity and CRISPR RNAs (crRNAs) with sequence specificity. The key point is to thoroughly select the guide RNA in order to eliminate off-target activity and to ensure sequence specificity. To avoid an increased rate of non-target mutations, it is crucial to regulate Cas9 and gRNA expression. This technology is today in broad use because of its relative simplicity, versatility, and efficiency. In plants, the CRISPR/Cas9 system has been used for example in rice to direct the mutagenesis of genes associated with morphological and quality traits (Shan *et al.* 2013) and in cucumber to develop resistance to *Cucumber vein yellowing virus, Zucchini yellow mosaic virus,* and *Papaya ringspot mosaic virus-W* (Chandrasekaran *et al.* 2016).

doi: 10.17221/192/2016-CJGPB

Acknowledgements. Financial support for this work from Ministry of Agriculture of the Czech Republic (Project No. QI111A019) is gratefully acknowledged.

References

Ainley W.M., Sastry-Dent L., Welter M.E., Murray M.G., Zeitler B., Amora R., Corbin D.R., Miles R.R., Arnold N.L., Strange T.L., Simpson M.A., Cao Z., Carroll C., Pawelczak K.S., Blue R., West K., Rowland L.M., Perkins D., Samuel P., Dewes C.M., Shen L., Sriram S., Evans S.L., Rebar E.J., Zhang L., Gregory P.D., Urnov D., Webb S.R., Petolino J.F. (2013): Trait stacking via targeted genome editing. Plant Biotechnology Journal, 11: 1126–1134.

Alagna F., Agostino N.D., Torchia L., Servili M., Rao R., Pietrella M., Giuliano G., Chiusano M.L., Baldoni L., Perrotta G. (2008): Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. BMC Genomics, 10: 399.

Ando K., Carr K.M., Grumet R. (2012): Transcriptome analyses of early cucumber fruit growth identifies distinct gene modules associated with phases of development. BMC Genomics, 13: 518.

Azam S., Thakur V., Ruperao P., Shah T., Balaji J., Amindala B., Farmer A.D., Studholme D.J., May G.D., Edwards D., Jones J.D.G., Varshney R.K. (2012): Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results to identify SNPs in chickpea (*Cicer arietinum*; Fabaceae), a crop species without a reference genome. American Journal of Botany, 99: 186–192.

Barabaschi D., Tondelli A., Desiderio F., Volante A., Vaccino P., Valè G., Cattivelli L. (2016): Next generation breeding. Plant Science, 242: 3–13.

Bonnet E., Wuyts J., Rouzé P., Van de Peer Y. (2004): Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. Proceedings of the National Academy of Sciences of the United States of America, 101: 11511–11516.

Brenner S., Johnson M., Bridgham J., Golda G., Lloyd D. H., Johnson D., Luo S., McCurdy S., Foy M., Ewan M., Roth R., George D., Eletr S., Albrecht G., Vermaas E., Williams S.R., Moon K., Burcham T., Pallas M., DuBridge R.B., Kirchner J., Fearon K., Mao J., Corcoran K. (2000): Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nature Biotechnology, 18: 630–634.

Buck M.J., Lieb J.D. (2004): ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics, 84: 349–360.

Celton J.M., Gaillard S., Bruneau M., Pelletier S., Aubourg S., Martin-Magniette M.L., Navarro L., Laurens F., Renou J.P. (2014): Widespread anti-sense transcription in apple is correlated with siRNA production and indicates a large potential for transcriptional and/or posttranscriptional control. The New Phytologist, 203: 287–299.

Chandrasekaran J., Brumin M., Wolf D., Leibman D., Klap C., Pearlsman M., Sherman A., Arazi T., Gal-On A. (2016): Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. Molecular Plant Pathology, 17: 1140–1153.

Chen W., Gao Y., Xie W., Gong L., Lu K., Wang W., Li Y., Liu X., Zhang H., Dong H., Zhang W., Zhang L., Yu S., Wang G., Lian X., Luo J. (2014): Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nature Genetics, 46: 714–721.

Chu Y., Wu C.L., Holbrook C.C., Tillman B.L., Person G., Ozias-Akins P. (2011): Marker-assisted selection to pyramid Nematode resistance and the high oleic trait in peanut. Plant Genome, 4: 110–117.

Cortés A.J., Chavarro M.C., Blair M.W. (2011): SNP marker diversity in common bean (*Phaseolus vulgaris* L.). Theoretical and Applied Genetics, 123: 827–845.

Dassanayake M., Haas J.S., Bohnert H.J., Cheeseman J.M. (2009): Shedding light on an extremophile lifestyle through transcriptomics. The New Phytologist, 183: 764–775.

Davey J.W., Hohenlohe P.A., Etter P.D., Boone J.Q., Catchen J.M., Blaxter M.L. (2011): Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature, 12: 499–510.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. (2011): A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE, 6: e19379.

Fan X.D., Wang J.Q., Yang N., Dong Y.Y., Liu L., Wang F.W., Wang N., Chen H., Liu W.C., Sun Y.P., Wu J.Y., Li H.Y. (2012): Gene expression profiling of soybean leaves and roots under salt, saline–alkali and drought stress by high-throughput Illumina sequencing. Gene, 512: 392–402.

Flachowsky H., Le Roux P.M., Peil A., Patocchi A., Richter K., Hanke M.V. (2011): Application of a high-speed breeding technology to apple (*Malus* × *domestica*) based on transgenic early flowering plants and marker-assisted selection. New Phytologist, 192: 364–377.

Gaines T.A., Lorentz L., Figge A., Herrmann J., Maiwald F., Ott M.C., Han H., Busi R., Yu Q., Powles S.B., Beffa R. (2014): RNA-Seq transcriptome analysis to identify genes involved in metabolism-based diclofop resistance in *Lolium rigidum*. Plant Journal, 78: 865–876.

Garnica D.P., Upadhyaya N.M., Dodds P.N., Rathjen J.P. (2013): Strategies for wheat stripe rust pathogenicity identified by transcriptome sequencing. PLoS ONE, 8: e67150.

doi: 10.17221/192/2016-CJGPB

He J., Zhao X., Laroche A., Lu Z.X., Liu H., Li Z. (2014): Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science, 5: 484.

Heffner E.L., Sorrells M.E., Jannink J.-L. (2009): Genomic selection for crop improvement. Crop Science, 49: 1–12.

Ištvánek J., Jaroš M., Křenek A., Řepková J. (2014): Genome assembly and annotation for red clover (*Trifolium pratense*; Fabaceae). American Journal of Botany, 101: 327–337.

Ištvánek J., Dluhošová J., Dluhoš P., Pátková L., Nedělník J., Řepková J. (2017): Gene classification and mining of molecular markers useful in red clover (*Trifolium pratense*) breeding. Frontiers in Plant Science, 8: 367.

Juntawong P., Sirikhachornkit A., Pimjan R., Sonthirod C., Sangsrakru D., Yoocha T., Tangphatsornruang S., Srinives P. (2014): Elucidation of the molecular responses to waterlogging in Jatropha roots by transcriptome profiling. Frontiers in Plant Science, 5: 658.

Kamei C.L.A., Severing E.I., Dechesne A., Furrer H., Dolstra O., Trindade L.M. (2016): Orphan Crops Browser: a bridge between model and orphan crops. Molecular Breeding, 36: 9.

Kumar J., Mir R.R., Kumar N., Kumar A., Mohan A., Prabhu K.V., Balyan H.S., Gupta P.K. (2010): Marker-assisted selection for pre-harvest sprouting tolerance and leaf rust resistance in bread wheat. Plant Breeding, 129: 617–621.

Kumar S., Banks T.W., Cloutier S. (2012): SNP discovery through next-generation sequencing and its applications. International Journal of Plant Genomics: e831460.

Lado B., Matus I., Rodríguez A., Inostroza L., Poland J., Belzile F., del Pozo A., Quincke M., Castro M., von Zitzewitz J. (2013): Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. G3-Genes Genomes Genetics, 3: 2105–2114.

Le D.T., Nishiyama R., Watanabe Y., Mochida K., Yamaguchi-Shinozaki K., Shinozaki K., Tran L.S. (2011): Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. DNA Research, 18: 263–276.

Lehmensiek A., Bovill W., Wenzl P., Langridge P., Appels R. (2009): Genetic mapping in the triticeae. In: Feuillet C., Muehlbauer G.J. (eds): Genetics and Genomics of the Triticeae. Heidelberg, Springer: 201–236.

Lor V.S., Starker C.G., Voytas D.F., Weiss D., Olszewski N.E. (2014): Targeted mutagenesis of the tomato PROCERA gene using transcription activator-like effector nucleases. Plant Physiology, 166: 1288–1291.

Lu T., Cui L., Zhou Y., Zhu Q., Fan D., Gong H., Zhao Q., Zhou C., Zhao Y., Lu D., Luo J., Wang Y., Tian Q., Feng Q., Huang T., Han B. (2015): Transcriptome-wide investigation of circular RNAs in rice. RNA, 21: 2076–2087.

Miller M.R., Dunham J.P., Amores A., Cresko W.A., Johnson E.A. (2007): Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research, 17: 240–248.

Myles S., Chia J., Hurwitz B., Simon C., Zhong G.Y., Buckler E., Ware D. (2010): Rapid genomic characterization of the genus *Vitis*. PLoS ONE, 5: e8219.

Nigam D., Kavita P., Tripathi R.K., Ranjan A., Goel R., Asif M., Shukla A., Singh G., Rana D., Sawant S.V. (2014): Transcriptome dynamics during fibre development in contrasting genotypes of *Gossypium hirsutum* L. Plant Biotechnology Journal, 12: 204–18.

Novaes E., Drost D.R., Farmerie W.G., Appas G.J.P., Grattapaglia D., Sederoff R.R., Kirst M. (2008): High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genomics, 9: 312.

Petrozza A., Santaniello A., Summerer S., Di Tommaso G., Di Tommaso D., Paparelli E., Piaggesi E., Perata P., Cellini F. (2014): Physiological responses to Megafol treatments in tomato plants under drought stress: a phenomic and molecular approach. Scientia Horticulturae, 174: 185–192.

Poland J.A., Rife T.W. (2012): Genotyping-by-sequencing for plant breeding and genetics. Plant Genome, 5: 92–102.

Rahman H., Jagadeeshselvam N., Valarmathi R., Sachin B., Sasikala R., Senthil N., Sudhakar D., Robin S., Muthurajan R. (2014): Transcriptome analysis of salinity responsiveness in contrasting genotypes of finger millet (*Eleusine coracana* L.) through RNA-sequencing. Plant Molecular Biology, 85: 485–503.

Ray S., Satya P. (2014): Next generation sequencing technologies for next generation plant breeding. Frontiers in Plant Science, 5: 367.

Shan Q., Wang Y., Li J., Zhang Y., Chen K., Liang Z., Zhang K., Liu J., Xi J.J., Qiu J.L., Gao C. (2013): Targeted genome modification of crop plants using a CRISPR-Cas system. Nature Biotechnology, 7: 686–688.

Shuai P., Liang D., Tang S., Zhang Z., Ye C.Y., Su Y., Xia X., Yin W. (2014): Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. Journal of Experimental Botany, 65: 4975–4983.

Swarbreck S.M., Lindquist E.A., Ackerly D.D., Andersen G.L. (2011): Analysis of leaf and root transcriptomes of soil-grown *Avena barbata* plants. Plant & Cell Physiology, 52: 317–332.

Tang S., Liang H., Yan D., Zhao Y., Han X., Carlson J.E., Xia X., Yin W. (2013): *Populus euphratica*: the transcriptomic response to drought stress. Plant Molecular Biology, 83: 539–557.

Tisné S., Serrand Y., Bach L., Gilbault E., Ben Ameur R., Balasse H., Voisin R., Bouchez D., Durand-Tardif M., Guerche P., Chareyron G., Da Rugna J., Camilleri C., Loudet O. (2013): Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity. Plant Journal, 74: 534–544.

doi: 10.17221/192/2016-CJGPB

Townsend J.A., Wright D.A., Winfrey R.J., Fu F., Maeder M.L., Joung J.K., Voytas D.F. (2009): High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature, 459: 442–445.

Trebbi D., Maccaferri M., de Heer P., Sorensen A., Giuliani S., Salvi S., Sanguineti M.C., Massi A., van der Vossen E.A.G., Tuberosa R. (2011): High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). Theoretical and Applied Genetics, 123: 555–569.

Trick M., Long Y., Meng J., Bancroft I. (2009): Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. Plant Biotechnology Journal, 7: 334–346.

Van Bel M., Proost S., Van Neste C., Deforce D., Van de Peer Y., Vandepoele K. (2013): TRAPID: an efficient online tool for the functional and comparative analysis of de novo RNASeq transcriptomes. Genome Biology, 14: R134.

Wang H., Chung P.J., Liu J., Jang I.C., Kean M.J., Xu J., Chua N.H. (2014): Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. Genome Research, 24: 444–453.

Wang X., Elling A.A., Li X., Li N., Peng Z., He G., Sun H., Qi Y., Liu X.S., Deng X.W. (2009a): Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. Plant Cell, 21: 1053–1069.

Wang Y., Dong C., Xue Z., Jin Q., Xu Y. (2015): *De novo* transcriptome sequencing and discovery of genes related to copper tolerance in *Paeonia* ostii. Gene, 576: 126–135.

Wang Z., Gerstein M., Snyder M. (2009b): RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics, 10: 57–63.

Wilkins K.E., Booher N.J., Wang L., Bogdanove A.J. (2015): TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. oryzicola while strict conservation suggests universal importance of five TAL effectors. Frontiers in Plant Science, 6: 536.

Windhausen V.S., Atlin C.N., Hickey J.M., Crossa J., Jannink J.L., Sorrells M.E., Raman B., Cairns J.E., Tarekegne A., Semagn K., Beyene Y., Grudloyma P., Technow F., Riedelsheimer C., Melchinger A.E. (2012): Effectiveness of genomic predictions of maize hybrid performance in different breeding populations and environments. G3-Genes Genomes Genetics, 2: 1427–1436.

Würschum T., Reif J.C., Kraft T., Janssen G., Zhao Y. (2013): Genomic selection in sugar beet breeding populations. BMC Genetics, 14: 85.

Xin M., Wang Y., Yao Y., Song N., Hu Z., Qin D., Xie C., Peng H., Ni Z., Sun Q. (2011): Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. BMC Plant Biology, 11: 61.

Xu Y. (2003): Developing marker-assisted selection strategies for breeding hybrid rice. Plant Breeding Reviews, 23: 73–174.

Xu Y., Crouch J.H. (2008): Marker-assisted selection in plant breeding: from publications to practice. Crop Science, 48: 391–407.

Xu P., Liu Z., Fan X., Gao J., Zhang X., Zhang X., Shen X. (2013a): De novo transcriptome sequencing and comparative analysis of differentially expressed genes in *Gossypium aridum* under salt stress. Gene, 525: 26–34.

Xu L., Wang Y., Zhai L., Xu Y., Wang L., Zhu X., Gong Y., Yu R., Limera C., Liu L. (2013b): Genome-wide identification and characterization of cadmium-responsive microRNAs and their target genes in radish (*Raphanus sativus* L.) roots. Journal of Experimental Botany, 64: 4271–4287.

Yamamoto N., Takano T., Tanaka K., Ishige T., Terashima S., Endo C., Kurusu T., Yajima S., Yano K., Tada Y. (2015): Comprehensive analysis of transcriptome response to salinity stress in the halophytic turf grass *Sporobolus virginicus*. Frontiers in Plant Science, 6: 241.

Yates S.A., Swain M.T., Hegarty M.J., Chernukin I., Lowe M., Allison G.G., Ruttink T., Abberton M. T., Jenkins G., Skøt L. (2014): *De novo* assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. BMC Genomics, 15: 453.

Yu X., Yang J., Li X., Liu X., Sun C., Wu F., He Y. (2013): Global analysis of cis-natural antisense transcripts and their heat responsive nat-siRNAs in *Brassica rapa*. BMC Plant Biology, 13: 208.

Zhang N., Wang S., Zhang X., Dong Z., Chen F., Cui D. (2016): Transcriptome analysis of the Chinese bread wheat cultivar Yunong 201 and its ethyl methanesulfonate mutant line. Gene, 575: 285–293.

Zhang W., Han Z., Guo Q., Liu Y., Zheng Y., Wu F., Jin W. (2014): Identification of maize long non-coding RNAs responsive to drought stress. PLoS ONE 9: e98958.

Zhong S., Fei Z., Chen Y.R., Vrebalov J., Mcquinn R., Gapper N., Giovannoni J. (2013): Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nature Biotechnology, 31: 154–159.

Received for publication December 13, 2016
Accepted after corrections June 21, 2017
Published online August 8, 2017