Journal of Experimental Botany
doi:10.1093/jxb/erw348

REVIEW PAPER

Using genomic information to improve soybean adaptability to climate change

Man-Wah Li[1], Dawei Xin[1,2], Yishu Gao[1], Kwan-Pok Li[1], Kejing Fan[1], Nacira Belen Muñoz[1,3,4], Wai-Shing Yung[1] and Hon-Ming Lam[1,*]

[1] Centre for Soybean Research, Partner State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR
[2] Key Laboratory of Soybean Biology of Chinese Ministry of Education, Key Laboratory of Soybean Biology and Breeding/Genetics of Chinese Agriculture Ministry, College of Science, Northeast Agricultural University, Harbin, Heilongjiang Province, People's Republic of China
[3] Instituto de Fisiología y Recursos Genéticos Vegetales, Centro de Investigaciones Agropecuarias-INTA, Córdoba, Argentina
[4] Cátedra de Fisiología Vegetal, Facultad de Ciencias Exactas Físicas y Naturales, Universidad Nacional de Córdoba, Córdoba, Argentina

* Correspondence: honming@cuhk.edu.hk

Received 24 June 2016; Accepted 2 September 2016

Editor: Kadambot Siddique, The University of Western Australia

Abstract

Climate change has brought severe challenges to agriculture. It is anticipated that there will be a drop in crop yield – including that of soybean – due to climatic stress factors that include drastic fluctuations in temperature, drought, flooding and high salinity. Genomic information on soybean has been accumulating rapidly since initial publication of its reference genome, providing a valuable tool for the improvement of cultivated soybean. Not only are many molecular markers that are associated with important quantitative trait loci now identified, but we also have a more detailed picture of the genomic variations among soybean germplasms, enabling us to utilize these as tools to assist crop breeding. In this review, we will summarize and discuss the currently available soybean genomic approaches, including whole-genome sequencing, sequencing-based genotyping, functional genomics, proteomics, and epigenomics. The information uncovered through these techniques will help further pinpoint important gene candidates and genomic loci associated with adaptive traits, as well as achieving a better understanding of how soybeans cope with the changing climate.

Key words: Climate change, epigenomics, genome, genome editing, genome-wide selection, nutrient stress, proteomics, soybean, temperature stress, transcriptomics, water stress, whole-genome sequencing.

Introduction

Soybean is cultivated mainly at latitudes 20–50° N and 10–40° S (Leff *et al.*, 2004). The wide distribution of its cultivation suggests that soybean is highly adaptable towards different environments and climates. However, that degree of adaptability may not be enough under the influence of climate change. In general, soybean can survive in a wide range of temperatures: from 10 to 40 °C, depending on genotype. However, extreme temperatures (below 12 °C or above 36 °C) can lead to reduced soybean germination (Tyagi and Tripathi, 1983) and the abolishment of pollen tube germination and

elongation (Luo, 2011), resulting in yield loss. At the same time, commonly used rhizobium strains are sensitive to both high and low temperatures, and thus extreme temperatures could affect nodule formation and nitrogen fixation (Lynch and Smith, 1993; Rahmani *et al.*, 2009). Climate change also leads to an observable and foreseeable increase in the chance of flooding (Muis *et al.*, 2015), a change in extreme precipitation patterns (Donat *et al.*, 2016), the intrusion of sea water into aquifers (Dasgupta *et al.*, 2015), a change in soil composition (Brevik, 2013) and also atmospheric composition such as the ozone concentration (Dentener *et al.*, 2005). All these can contribute to a reduction in soybean yield.

In 2015, the World Bank stated that climate change could lead to a 5% yield reduction from the current food production by 2030 and a 30% reduction by 2080 (Havlík *et al.*, 2015). Like that of other staple crops, soybean production potential was predicted to be supressed by climate change. A simulation model predicted that production would reduce by 10–20% in India due to global warming when CO_2 concentration doubled (Mall *et al.*, 2004). Another report also predicted that soybean production will reduce by at least 30% by 2099 because of climate change (Schlenker and Roberts, 2009), although this prediction could be overstated according to other researchers (Meerburg *et al.*, 2009). A recent statistical model based on more than 18 years of soybean yield records from 12 states in the USA predicted that on average there will be a 2.4% yield reduction for every 1 °C rise in temperature (Mourtzinis *et al.*, 2015). Although an extension of the growing season due to global warming and the slight increase in carbon fertilization (Norby and Zak, 2011; Reyes-Fox *et al.*, 2014) might boost the production of soybean to some degree, it could not completely mitigate the negative effects of climate change.

There is a demand for improved soybean adaptability to climate change. Wild soybeans may harbor a high diversity of adaptive traits against adverse environments. More in-depth research is needed to discover and make good use of these valuable genetic resources. The publishing of the reference genome in 2010 (Schmutz *et al.*, 2010) has accelerated soybean research in various ways. Here in the subsequent sections, we will focus our discussions on genomic researches on the mechanisms by which soybean can adapt to abiotic stresses, which are increasingly exacerbated by climate change.

Current availability of soybean genomic information

A well-annotated reference genome can speed up subsequent analyses such as the detection of single nucleotide polymorphisms, copy number variation and structural variation among germplasms. It can also serve as the foundation for reference-based functional genomic analyses such as transcriptomic, proteomic, epigenomic and non-coding RNA analyses. Currently available soybean genomic information is given in Table 1.

Thus the official release of the soybean cultivar Williams 82 reference genome in 2010 (Schmutz *et al.*, 2010) has marked the new era of soybean research. Since then, the genome has been further refined. The second assembly of the reference genome was released in 2013. In this version, the predicted size of the soybean genome is 1.1 Gb and the assembled size was 978 Mb. There are altogether 56 044 gene models in this assembly.

On the basis of the reference genome mentioned above (Schmutz *et al.*, 2010), Lam *et al.* (2010) reported the re-sequencing of 14 cultivated and 17 wild soybean germplasms in the same year. It was the first definitive work using whole-genome sequencing to show that wild soybeans have higher genetic diversities than cultivated soybeans. This demonstrates that the wild soybean is an important genetic resource for crop improvement. In the same year, a Korean wild soybean IT182932 was re-sequenced to a high depth (52.07×) (Kim *et al.*, 2010). By mapping the reads onto the Williams 82

Table 1. *Currently available whole-genome sequencing information on soybean*

Variety	Method	Accession number	References
Williams 82 (cultivated)	*De novo* sequencing and assembly	GCA_000004515.3	(Schmutz *et al.*, 2010)
14 cultivated 17 wild	Re-sequencing	SRA020131	(Lam *et al.*, 2010)
IT182932 (wild)	Re-sequencing *De novo* assembly of unmapped reads	SRA009252	(Kim *et al.*, 2010)
10 cultivated 5 wild	Re-sequencing	ERP002622	(Chung *et al.*, 2014)
7 wild	*De novo* sequencing and assembly	PRJNA195632	(Li *et al.*, 2014*d*)
9 semi-wild	Re-sequencing	PRJNA227063	(Qiu *et al.*, 2014)
Maliaodou (semi-wild)	*De novo* sequencing and assembly	PRJNA227063 RX375213	(Qiu *et al.*, 2014)
Lanxi1 (wild)	*De novo* sequencing and assembly	PRJNA227063 SRX375212	(Qiu *et al.*, 2014)
W05 (wild)	*De novo* sequencing and assembly	GCA_000722935.2	(Qi *et al.*, 2014*a*)
240 cultivated 62 wild	Re-sequencing	SRP045129	(Zhou *et al.*, 2015)
Enrei (cultivated)	Reference-based assembly	GCA_001269945.2	(Shimomura *et al.*, 2015)

genome along with the *de novo* assembly of unmapped reads, it was found that there was a 3.76% difference between the wild and cultivated soybean genomes, including large structural variations (Kim *et al.*, 2010). These *de novo* assembled scaffolds unique to wild soybean also acted as the framework for subsequent research on wild soybean. Furthermore, a re-sequencing of 302 soybean accessions identified 121 domestication-selective sweeps and 109 improvement-selective sweeps (Zhou *et al.*, 2015). Some of these sweeps were novel while some of them overlapped with and were narrower than previously known domestication-related QTL regions (Zhou *et al.*, 2015). This provided an important foundation for breeding and gene discovery and further demonstrated the robustness of reference-based re-sequencing.

Nonetheless, a single cultivated soybean reference genome is inadequate, especially when a study involves wild soybean. As a result, a wild soybean draft genome was released in 2014 (Qi *et al.*, 2014*a*). In that study, the authors successfully identified the major salt tolerance gene in soybean by comparing the genomic sequences between the cultivated soybean reference genome and the wild soybean draft genome. After that, a pan-genome representing seven wild soybean accessions from distinct sources was constructed (Li *et al.*, 2014*d*). Approximately 80% of the pan-genome was indispensably shared among the seven accessions, while the remaining dispensable 20% could contain the essential elements for adaptive traits (Li *et al.*, 2014*d*). In addition, some other cultivated soybean genomes were built for specific purposes. For example, the genome of a Japanese cultivar Enrei was built through reference-based assembly (Shimomura *et al.*, 2015), aiming at facilitating the characterization of the soybean cultivar popularized in Japan.

Currently, a large-scale soybean genome project is underway in the USA (http://soybeangenomics.missouri.edu/news/). The genome of the soybean cultivar Lee will be built as the reference genome representing the soybean accessions in the southern USA, complementing the reference genome of Williams 82 that better represents the soybean cultivars in the north of the country. In total, ten reference genomes for cultivated soybean and five for wild soybean will be built. Furthermore, more than 4000 soybean accessions from both public and private depositions in the USA will be re-sequenced. The big dataset generated by this project will eventually bring new insights to both soybean research and production.

Genome-wide studies of gene families

Due to the complexity of the soybean genome, the inefficiency of transformation and thus the generation and maintenance of mutants, the identification and study of genes in soybean are challenging. Assuming that homologous genes perform similar functions across species, studying the homologs of well-characterized genes from the model plant, Arabidopsis, is a common way to identify the adaptability-related genes in soybean. However, owing to whole-genome and segmental duplications (Schmutz *et al.*, 2010), most gene families in soybean have been much expanded compared to Arabidopsis. The diversifications among gene family members as a result of neofunctionalization and subfunctionalization after soybean genome duplication events have further complicated the study of these genes. As a result, the gene homologs between Arabidopsis and soybean do not always form a one-to-one orthologous relationship. A complete annotated soybean genome sequence is therefore needed to allow for ontology curation, blast search, hidden Markov model (HMM)-based search, and so on. Cataloguing the gene families and studying the functions of individual gene members within these families will help identify the candidates for crop improvement.

Since temperature changes are usually considered to be the most important factor affecting yield potential under the influence of climate change, here we have a few examples of heat stress-related gene families.

Heat shock proteins (Hsp) have been shown to confer stress tolerance (Wang *et al.*, 2004). Fifty-one *Hsp20* (Lopes-Caitar *et al.*, 2013), 61 *Hsp70* (Zhang *et al.*, 2015) and 12 *Hsp90* (Xu *et al.*, 2013*a*) genes were identified from the soybean genome through simple BLASTP search, HMMs profile-BLASTP search and keyword curation. The number of genes in these three *Hsp* families is much higher in soybean, compared to just 19 *Hsp20*, 18 *Hsp70* and 7 *Hsp90* genes in Arabidopsis (Krishna and Gloor, 2001; Lin *et al.*, 2001; Waters *et al.*, 2008) and 23 *Hsp20*, 32 *Hsp70* and 8 *Hsp90* genes in rice (Sarkar *et al.*, 2009, 2013; Zhang *et al.*, 2013). Although heat shock proteins have been shown to play important roles under heat stress (Fragkostefanakis *et al.*, 2015), expression data have suggested that they are probably also involved in the response mechanisms for other stresses. At least 40 of the soybean *Hsp*20 genes were induced upon heat stress while five of them were also responsive toward cold stress (four up-regulated and one down-regulated) (Lopes-Caitar *et al.*, 2013). It is interesting that some of the *Hsp20* genes were also responsive to biotic stresses (Lopes-Caitar *et al.*, 2013). On the other hand, 29 soybean *Hsp70* genes were up-regulated by both short-term heat and drought treatments while 27 of the remaining were down-regulated under the same conditions (Zhang *et al.*, 2015). Furthermore, it is interesting that all 12 of the soybean *Hsp90* were highly induced by heat stress, osmotic stress and salinity, and slightly up-regulated under cold stress (Xu *et al.*, 2013*a*). Four out of five soybean *Hsp90* genes could enhance the germination rate under heat stress, osmotic stress and salinity when ectopically expressed in Arabidopsis under a 35S promoter. However, the ectopic expression of these same genes impaired the growth of the transgenic plants such that they perform no better than the wild type control under these stresses in most of the parameters measured except pod setting percentage, chlorophyll content and proline content (Xu *et al.*, 2013*a*). Such observations suggest that low-level expressions of these *Hsp* genes could be beneficial to plant growth under normal conditions while a rapid induction of these genes are essential for the plant to survive under stress.

Heat shock transcription factors (Hsf) regulate gene expression by binding to the heat shock element of target gene promoters upon heat stress (Scharf *et al.*, 2012).

Thirty-eight *Hsf* genes were found in soybean through a genome-wide BLASTP search (Li *et al.*, 2014*b*). Among 19 *Hsf* genes tested, 14 and 13 were induced by drought and heat stress, respectively (Li *et al.*, 2014*b*). The ectopic expression of *GmHsf-34*, which was found to be highly up-regulated by both drought and heat stress, could alleviate the osmotic and heat stress symptoms in the transgenic Arabidopsis (Li *et al.*, 2014*b*).

These examples all illustrate the complex interplay among the genes and gene families involved in abiotic stress adaptations and their effects on growth. This demonstrates that a genome-wide curation of gene family members could yield a clearer overall picture and lead to a more rapid discovery of functional genes that confer stress tolerance.

Impact of sequencing-based genotyping

In the past few decades, efforts have been made to identify the loci involved in the adaptabilities of soybean. Such in-depth studies can help produce soybean crops that are better adapted to climate change.

A list of recent quantitative trait locus (QTL) studies related to the adaptabilities toward temperature stresses, water stresses and mineral stresses is shown in Table 2. Typically, the QTL position is expressed as the genetic distance (cM) between markers. As there has been no available sequence information between markers, the resolutions of the classic genetic markers have been relatively limited (Huang *et al.*, 2009). The loci thus identified usually span large genomic regions and contain a large number of genes. It may be adequate for marker-assisted breeding, but it is difficult to identify the actual phenotype determinants. Thus this has largely hindered the discovery of key adaptation-related genes.

The rise of next-generation sequencing (NGS) has brought about a revolution in crop studies. First of all, the construction of reference genomes allows for the determination of the physical location of each marker in the genome, and therefore sequence and gene information can also be retrieved. Whole-genome sequencing also provides more potential markers, ranging from simple sequence repeats (SSRs) (Hwang *et al.*, 2009), individual single nucleotide polymorphisms (SNPs) (Kim *et al.*, 2010; Lam *et al.*, 2010; Chung *et al.*, 2014; Zhou *et al.*, 2015), bin markers (Qi *et al.*, 2014*a*), insertion/deletion (INDEL) markers (Song *et al.*, 2015*b*), specific-locus amplified fragment (SLAF) markers (Zhang *et al.*, 2016), and so on. As a result, the mapping resolution has been largely

Table 2. *Newly identified soybean QTLs related to adaptations against abiotic stresses associated with climate change*

Traits	Chromosome (Linkage group)	Associated markers	QTL size (cM)	References
Salt stress-related QTLs				
Salt tolerance	3 (N)	SNP14-SNP10	4.7	(Qi *et al.*, 2014*a*)
	3 (N)	SSR03_1335	Not stated	(Ha *et al.*, 2013)
Low temperature stress-related QTLs				
Low temperature-induced seed coat discoloration tolerance	8 (A2)	GmIRCHS	Not stated	(Ohnishi *et al.*, 2011)
	14 (B2)	Sat_342		
Germination stage tolerance	5 (A1)	Sat_271	Not stated	(Zhang *et al.*, 2012)
	5 (A1)	Satt225		
	11 (B1)	Sat_331		
	14 (B2)	Satt168		
	14 (B2)	Satt577		
	04 (C1)	Satt338		
	06 (C2)	Satt640		
	01 (D1b)	Satt041		
	01 (D1b)	Satt271		
	17 (D2)	Satt458		
	17 (D2)	Satt669		
	15 (E)	Satt651		
	12 (H)	Satt142		
	12 (H)	Satt253		
	12 (H)	Satt353		
	20 (I)	Satt440		
	16 (J)	Satt249		
	09 (K)	Sat_126		
	09 (K)	Satt240		
	09 (K)	Satt349		
	19 (L)	Satt513		
	07 (M)	Sat_244		
	07 (M)	Satt323		
	07 (M)	Satt336		
	07 (M)	Satt540		

Table 2. *Continued*

Traits	Chromosome (Linkage group)	Associated markers	QTL size (cM)	References
Seedling stage tolerant	04 (C1)	Satt338		
	01 (D1b)	Sat_192		
	01 (D1b)	Satt041		
	01 (D1b)	Satt271		
	17 (D2)	Satt669		
	15 (E)	Satt651		
	13 (F)	Satt663		
	12 (H)	Satt142		
	20 (I)	Satt440		
	09 (K)	Sat_020		
	07 (M)	Sat_244		
	07 (M)	Satt336		
	07 (M)	Satt540		
Drought stress-related QTLs				
Canopy-wilting	02 (D1b)	satt296	63.5	(Abdel-Haleem *et al.*, 2012)
	04 (C1)	satt646	36.9	
	05 (A1)	satt276	8	
	12 (H)	satt302	56.8	
	14 (B2)	satt066	74.2	
	17 (D2)	satt135	20.2	
	19 (L)	satt462	55.7	
Fibrous roots	01 (D1a)	satt383-satt580	50.5	(Abdel-Haleem *et al.*, 2011)
	03 (N)	satt339-sat_091	57.3	
	04 (C1)	satt713-sct_191	77.2	
	08 (A2)	satt228-satt429	149.6	
	20 (I)	sat_420-sat_299	77.8	
Flooding stress-related QTLs				
Flooding tolerance score	11 (B1)	BARC-054421-12081	87	(Nguyen *et al.*, 2012)
	13 (F)	BARC-024569-04982	26	
Flooding yield index	11 (B1)	BARC-016279-02316	79	
	13 (F)	sct_033	34	
Ozone				
Ozone tolerance	1	039805–07589		(Burton *et al.*, 2016)
	4	029943–06758		
	6	044133–08626		
	17	017525–03061		
	18	056635–14538		
	19	060295–16596		
	20	029827–06444		

improved due to the increase in marker density (Huang *et al.*, 2009). At the same time, technology-intensive genotyping-by-sequencing (GBS) has also greatly reduced labor time compared to the traditional PCR-based genotyping methods (Huang *et al.*, 2009). Furthermore, compared to hybridization-based SNP genotyping methods, sequencing-based methods also allow for the detection of new variants specific to the population being tested. This illuminates the unique features of the population of interest, which can then be utilized in the future. All these NGS features have shortened the time required and greatly improved precision in identification of key genes for crop improvement.

To date, there are a number of sequencing-based QTL analyses of soybean (Xu *et al.*, 2013*b*; Bastien *et al.*, 2014; Li *et al.*, 2014*a*; Qi *et al.*, 2014*a*; Zhang *et al.*, 2016). Only a limited number of them are related to soybean adaptability and thus useful for developing cultivated soybean that is adaptive to climate change, but they have set good examples for future adaptation studies. For example, a sequencing-based QTL mapping of soybean was done to map the salt tolerance locus in wild soybean (Qi *et al.*, 2014*a*). The locus was mapped to a 978-kb region on chromosome 3 in wild soybean accession W05, using 2757 bin markers (Qi *et al.*, 2014*a*). By saturating the region with SNP markers identified based on re-sequencing information (Lam *et al.*, 2010), the region was further narrowed to a 388-kb region, which is the narrowest region ever to be reported for this trait at that time. By adopting and analysing the re-sequencing consensus from a previous study (Lam *et al.*, 2010), a monovalent cation/proton antiporter gene *GmCHX1* was found to be the common link among different salt-tolerant soybean accessions. As further confirmation of this gene's role, its ectopic expression in the roots of a salt-sensitive soybean accession can alleviate salt stress as well (Qi *et al.*, 2014*a*). This finding has been confirmed by other

independent genomic studies (Guan *et al.*, 2014; Patil *et al.*, 2016), supporting the effectiveness of the sequencing-based study. Furthermore, Kompetitive Allele-Specific Polymerase Chain Reaction (KASP) assays for detecting salt tolerance alleles were designed based on the subsequent whole-genome sequencing analysis of 106 soybean accessions (Patil *et al.*, 2016). The highly precise assay can then help breed salt-tolerant varieties of high economic value. The identification of key salt tolerance genes from soybean has set a good example demonstrating how an integrated genomic approach can be used to identify important adaptation-related genes. Secondly, by improving soybean with these stress tolerance determinants, it allows the crop to be cultivated on barren lands, thus compensating for the loss of arable land due to climate change.

The current overall cost for NGS is still rather high. Hence, more cost-effective improved protocols and alternatives are needed. A new genotyping-by-sequencing (GBS) method called specific-locus amplified fragment sequencing (SLAF-seq) was determined to be cost-effective in genotyping soybean with an even better distribution and coverage of markers than the previously used GBS methods (Sun *et al.*, 2013; Qi *et al.*, 2014*b*). It was used in a low-phosphate stress QTL study on soybean (Zhang *et al.*, 2016). In this study, 6159 SLAF markers were used to generate the genetic map and 85 low-phosphate stress-related QTLs were identified (Zhang *et al.*, 2016). There was a 5-fold increase in resolution in the mapping of a phosphorus-efficiency QTL on chromosome 8, compared to a previous study using the same genetic population and only 306 markers (Zhang *et al.*, 2016).

SLAF-seq, whole-genome resequencing and GBS using reduced-representation sequencing and restriction site-associated sequencing (RAD-seq) normally require relatively high sequencing data throughput for the identification of high-confidence SNPs. The SNPs discovered by these methods are not evenly distributed throughout the genome. By re-sequencing the two parental soybean lines, it was discovered that nearly 90% of the SNPs were clustered in less than 5% of the genome (Li *et al.*, 2014*c*). As a result, the authors suggested that 384 selected markers were good enough to genotype the 254 F8 recombinant inbred lines (Li *et al.*, 2014*c*). The map thus constructed was 2594 cM in length and the average distance between markers was 5.58 cM (Li *et al.*, 2014*c*). Although the distance between markers may not be good enough for fine mapping, this study provided a low-cost alternative for soybean population genotyping.

Microarray technology is another option. Through the reduced-representation sequencing of six cultivated and two wild soybean accessions, a high-density array (SoySNP50K) was produced for soybean genotyping (Song *et al.*, 2013*a*, 2015a). Basically, SNPs were selected so that they were evenly distributed on the reference genome. On average, there were ~111 and 20 SNPs/Mb in euchromatic and heterochromatic regions, respectively (Song *et al.*, 2013*a*). Over 80% of the 52 041 SNPs on the array pool were found to be polymorphic among 20 000 cultivated and wild soybean accessions (Song *et al.*, 2015*a*), which is sufficient for general genotyping purposes. Nevertheless, compared to GBS, the array could not discover novel SNPs, INDELs and other structural variations among different accessions, and these variations may sometimes provide clues for post-mapping gene identification.

Genome selection

Breeding soybean varieties with high yield potential and high adaptability would be important for ensuring stable food production under climate change. High-throughput sequencing and the accumulation of QTL information would be beneficial for marker-assisted selection (MAS) (Collard and Mackill, 2008; He *et al.*, 2014) especially when soybean has a high linkage disequilibrium (Lam *et al.*, 2010). Nevertheless, marker-assisted breeding only favors the stacking of traits controlled by a single locus or by just a few dominant loci, such as in the case of salt tolerance in soybean (Qi *et al.*, 2014*a*; Patil *et al.*, 2016). Unfortunately, traits related to climate change adaptations such as drought tolerance, heat tolerance, cold resistance and nutrient deficiency responses, are usually controlled by multiple small-effect loci. The stacking of traits using MAS could therefore be challenging when yield potential and seed quality are also important considerations. Genomic or genome-wide selection (GS), on the other hand, evaluates the genomic estimated breeding values (GEBVs) based on statistical models that fit together the genotypic and phenotypic data of the training population (reviewed in Nakaya and Isobe, 2012). Thus as long as the trait data have been fitted into the statistical models, their potential can be evaluated in the selection population by selecting individuals or lines in the breeding populations showing high GEBVs. The accuracy of GEBVs varies among plant studies, but it normally achieves a correlation coefficient of over 0.4 with the actual phenotypic values (Nakaya and Isobe, 2012). Although GS studies on soybean lag far behind those on other crops such as maize, barley and wheat, they have been used for studying grain yield (Jarquin *et al.*, 2014), seed weight (Shu *et al.*, 2013) and soybean cyst nematode resistance (Bao *et al.*, 2014). These successful cases further demonstrate the potential of using GS for breeding soybean varieties that are more adaptive to climate change.

Functional genomics

Climate change-related stresses, such as temperature (heat/cold), water (flooding/drought) and nutritional stresses, could trigger complex responses in plants. A series of transcriptional, translational and post-translational adaptations are employed by plants to cope with these stresses. Instead of studying the effects of a single gene or locus, transcriptomics, proteomics and their derivatives have become popular strategies for identifying gene functions, pathways or systems that are enriched or overrepresented under these stresses. In some occasions, key genes involved in the stress responses can also be identified.

Transcriptomics

Transcriptomics nowadays not only facilitates the analyses of global gene expressions, but sequencing-based transcriptomics

(RNA-seq) also allows the discovery of transcript sequence variants, alternatively spliced transcripts, coding and non-coding RNAs and RNA-editing events under different environmental or treatment conditions. An informative RNA-seq also helps in gene prediction and modelling in the genome annotation pipeline (Thibaud-Nissen *et al.*, 2013). In turn, improved genome annotation could facilitate transcriptomic and proteomic studies. RNA-seq based on NGS generates short reads that require mapping and/or assembly. Some of the information such as alternative splicing and end sequences of the transcripts may be lost during the data manipulation. As a result, single-molecule sequencing has been developed to generate reads up to 20 kb in length (Pan *et al.*, 2008), largely improving the detection of full-length transcripts, though the throughput for cost-effective quantification still needs to be improved.

The tissue-specific expressions of soybean transcripts (Libault *et al.*, 2010; Severin *et al.*, 2010) can be found in publicly available databases such as Phytozome (Goodstein *et al.*, 2012) and eFP Browser (Winter *et al.*, 2007).

Taking carbon fertilization as an example, in a laboratory experiment, an increase in the CO_2 level from 380 to 550 µmol mol^{-1} dry air could increase soybean photosynthesis by ~20% and hence increase the soluble sugar and starch contents by more than 40% and 80%, respectively, without any increase in the number of transcripts of photosynthesis-related genes (Leakey *et al.*, 2009). On the other hand, two studies consistently found that the genes related to starch metabolism, sugar transportation, respiration, glycolysis and the tricarboxylic acid cycle were up-regulated by carbon fertilization (Ainsworth *et al.*, 2006; Leakey *et al.*, 2009). Such observations led to the conclusion that the changes in transcription lead to quicker assimilation of photosynthetic products, followed by faster transportation of carbohydrates across cellular compartments and their subsequent break-down for growth and development. Hence, a higher rate of biomass accumulation is achieved. Therefore, manipulating the photosynthetic reaction at the transcriptional level may not be an effective way to increase yield in response to climate change. Instead, directing the photoassimilates to an effective sink could make better use of the carbon fertilization in crop improvement.

Proteomics

Proteome dynamics are not only determined by the transcript levels, but are also regulated post-transcriptionally and post-translationally. As proteins can have different isoforms, post-translational modifications and subcellular localizations, a proteome is more complex than a transcriptome. Therefore, studying proteomes could provide a deeper understanding of plant stress responses. Nevertheless, the robustness of quantitative proteomics is still limited by the extraction and separation methods, protein compositions, digestive enzymes used and the sensitivity and throughput capacity of mass spectrometry. The study of plant proteomes is also challenged by the presence of large quantities of secondary metabolites and rubisco protein. Therefore, specific protocols for extracting and handling soybean leaf and root protein samples were developed (Mesquita *et al.*, 2012; Rodrigues *et al.*, 2012). Gel-based and quantitative approaches and state-of-the-art mass spectrometry analyses have been summarized in several reviews (Aebersold and Mann, 2003; Bantscheff *et al.*, 2012; Dreger, 2003; Larance and Lomond, 2015; Vadivel, 2015). Although liquid chromatography–tandem mass spectrometry (LC-MS/MS) can identify more proteins, gel-based analyses still cannot be completely replaced under certain circumstances (Komatsu *et al.*, 2009; Nouri and Komatsu, 2010).

The completion of the soybean reference genome also brought about a revolution in proteomic studies. One study compared the proteins identified by searching the NCBI database or the soybean genome sequence using the same set of LC-MS/MS data generated from the plasma membrane proteome of flooding-treated soybean (Komatsu *et al.*, 2009). The search using the NCBI database identified 74 proteins while that using the soybean genome identified 124 proteins including 61 of the proteins from the NCBI search (Komatsu *et al.*, 2009). This shows that a well assembled and annotated genome is important for protein identification in proteomic analyses. In turn, similar to RNA-seq, proteomics studies can provide important information for genome annotation (Thibaud-Nissen *et al.*, 2013), leading to better proteomics studies in the future. In the same study as previously mentioned, in addition to the eight upregulated proteins identified by nano LC-MS/MS, the expressions of yet 12 other proteins were found to be up-regulated and two down-regulated, using two-dimensional polyacrylamide gel electrophoresis followed by mass spectrometry (2-DE-MS) (Komatsu *et al.*, 2009). This confirms that LC-MS still cannot completely replace 2-DE-MS in protein identification. Six of the up-regulated and all of the down-regulated proteins were of unknown functions. Some of the remaining proteins are mainly involved in well-known flooding responses including signal transduction, reactive oxygen species (ROS) scavenging, protein folding and cell wall reinforcement.

Post-translational modifications (PTM) alter protein folding, protein-protein interactions, enzyme activities and protein stability. Enrichment methods for phosphoproteome, acetylome, ubiquitinome, glycoproteome (Mertins *et al.*, 2013; Olsen and Mann, 2013) and redox proteome (Go and Jones, 2013) are available.

An interesting observation is that GAPDH was found to be subjected to glycosylation, phosphorylation and redox modification under different stresses in three previous studies (Galant *et al.*, 2012; Mustafa and Komatsu, 2014; Pi *et al.*, 2016). In general, GAPDH catalyses the conversion of glyceraldehyde 3-phosphate and glycerate into 1,3-bisphosphate in glycolysis. Nonetheless, more and more evidence supports GAPDH playing non-catalytic roles in plants upon stress-induced post-translational modification (Zaffagnini *et al.*, 2013), making GAPDH an important hub for stress-related responses.

Epigenomics: DNA methylation and histone modifications

Epigenomics refers to the study of the genomic-wide reversible modifications of DNA or histones. Such modifications

regulate gene expressions and lead to changes in the cellular performance and phenotypes of the organism. The epigenome is dynamic – it changes according to different developmental stages and environmental influences. Evidence suggests that plants can retain the epigenetic changes from a previous wave of stress and can thereby trigger a rapid transcriptomic response upon the next wave of stress (Kinoshita and Seki, 2014), thus increasing fitness and adaptabilities. Some stress-induced epigenomic modifications were believed to be inheritable though it is still controversial (Schmitz *et al.*, 2013; Tricker, 2015). Epigenomics is such a relatively new topic in soybean genomic research that there are only a limited number of publications on this topic (Schmitz *et al.*, 2013; Song *et al.*, 2013*b*; Kim *et al.*, 2015).

An immunofluorescence study found that heterochromatin DNA methylation increased upon chilling stress but decreased after recovery (Stepinski, 2012). In the same study, the immunofluorescence signal from the dimethylation of Lysine 9 of Histone 3 (H3K9me2) was strongest under chilling stress and it was at its weakest upon recovery (Stepinski, 2012). On the contrary, the signals for H3K4me3, H4K12 acetylation (H4K12ac) and H3K9ac were reduced upon chilling but increased after recovery (Stepinski, 2012). Although these observations did not directly link histone modifications to chilling stress responses in soybean, they implicated that DNA methylation could play a role in regulating gene expressions under chilling stress.

Forty-five out of 1335 soybean transcription factors (TFs) were found to be significantly induced upon salt stress (Song *et al.*, 2012). Among these 45 TFs, the promoters of *Glyma11g02400* (*GmMYB*), *Glyma08g41450* (*Gmb-ZIP*), *Glyma16g27950* (*GmAP2/DREB*) and *Glyma20g30840* (*GmAP2/DREB*) were differentially methylated and enriched with different histone modifications upon salt stress (Song *et al.*, 2012). Specifically, the expressions of *Glyma11g02400, Glyma16g27950* and *Glyma20g30840* were negatively correlated with their DNA methylation levels. On the other hand, the expressions of *Glyma11g02400*, *Glyma20g30840* and *Glyma08g41450* were negatively correlated with the levels of H3K9me2 and positively correlated with those of H3K4me3, while the expression of *Glyma16g27950* was only negatively correlated with DNA methylation (Song *et al.*, 2012). The expressions of *Glyma20g30840* and *Glyma08g41450* were also positively correlated with H3K9ac (Song *et al.*, 2012).

The above examples illustrate that the epigenome plays vital roles in stress adaptation. To improve the fitness of soybean under stresses due to climate change, manipulating the epigenome through genetic engineering may be one solution. Alternatively, pre-treating plants with mild stresses or chemicals to trigger epigenomic changes prior to the predicted severe stresses could theoretically be a way to prepare the plant for the bigger challenge.

Epigenomics: non-coding RNA

Small non-coding RNAs have played vital roles in cellular functions including stress responses (Guleria *et al.*, 2011). Thus far, small RNAs in plants have been found to regulate RNA degradation, translation, DNA methylation and histone modifications (reviewed in Guleria *et al.*, 2011). The classification and biogenesis of small non-coding RNA in plants have been intensively reviewed (Guleria *et al.*, 2011; Borges and Martienssen, 2015). Published miRNA sequences and annotations, including those from soybean, can be found in miRBase (Griffiths-Jones *et al.*, 2008). Currently, it contains the sequences of 573 miRNA precursors and 639 mature miRNAs of *Glycine max* and 13 miRNA precursors and 13 mature miRNAs of *Glycine soja* (Release 21, retrieved May 2016). A functional network of soybean miRNAs was built based on the information retrieved from miRBase (Xu *et al.*, 2014*a*) and is available online (Xu *et al.*, 2014*b*).

NGS has also sped up the identification of stress- or adaptation-related small RNAs in soybean. Current studies have focused on tissue-specific miRNAs in the soybean plant (Joshi *et al.*, 2010), small RNAs in root nodules (Turner *et al.*, 2012), stress- and disease-related miRNAs (Kulcheski *et al.*, 2011), phasiRNAs (Arikit *et al.*, 2014) and aluminum-induced miRNAs from wild soybean (Zeng *et al.*, 2012). Furthermore, parallel analyses of RNA ends (PARE) sequencing has also been used for identifying miRNA cleavage sites on mRNA targets during seed development (Shamimuzzaman and Vodkin, 2012).

The levels of miRNAs that target multiple stress-related TFs were found to be significantly correlated with chilling stress in soybean root nodules (Zhang *et al.*, 2014). Such a discovery supports the idea that miRNAs play important roles in stress adaptations by indirectly modulating the expressions of a bundle of genes, through targeting TFs. On the other hand, a miRNAome of the soybean root apex identified miRNAs that responded to both salt stress and exogenous auxin treatments (Sun *et al.*, 2016), implying that miRNAs from soybean also mediate the cross-talk between hormonal pathways and stress responses. Furthermore, the ectopic expression of one of the salt- and auxin-responsive miRNA, *miR399a* in soybean hair roots, reduced root growth by 40% under salt treatment (Sun *et al.*, 2016). This demonstrated that the manipulation of miRNAs could bring about huge effects on soybean adaptation.

Long non-coding RNAs (lncRNAs) are another class of non-coding transcripts, which are normally longer than 200 nt. Although RNA sequencing has sped up the pace of lncRNA research, thus far, the functions and working mechanisms of these non-coding RNAs are still largely unclear. More than 13 000 lncRNAs were identified, out of a total of 103 106 that expressed transcripts in the developing soybean embryo (Aghamirzaie *et al.*, 2015). Over 70% of them were transcribed from protein-encoding gene regions and have expression levels lower than those of the protein-coding transcripts (Aghamirzaie *et al.*, 2015). The remaining lncRNAs were long inter-genic non-coding RNAs and non-coding antisense transcripts. Little has been done to study the lncRNAs in soybean, but the high proportion of lncRNAs in the transcriptome may infer their importance. As more and more soybean transcriptome analyses are being published, more adaptation-related lncRNAs will be identified.

Genome editing

Since soybean is directly consumed by humans, the public concern over food and environmental safety should be taken into account seriously. Genome editing, whether using the zinc finger nuclease (ZFN), transcription activator-like effector-based nucleases (TALEN) or the clustered regulatory interspaced short palindromic repeat (CRISPR)/cas9, is still one of the more acceptable ways to manipulate crop genomes for improving their adaptabilities (Curtin *et al.*, 2011; Chen and Gao, 2014; Du *et al.*, 2016). A 'DNA-free' protocol has also been developed to ease the public concern (Woo *et al.*, 2015). These systems generate cleavage sites on desired genomic regions. Non-homologous end-joining of the cleavage sites by the endogenous DNA repairing machinery creates an insertion or a deletion and thus leads to a mutation in the target genome. Furthermore, replacement of the target sequence with a homologous sequence can also be achieved through homology-directed DNA repairs. The advantage of genome editing over traditional transgenic plants is that genome editing practically leaves no trace of foreign DNA in the edited genome.

Among these three methods, CRISPR/cas9 is considered the best, the main reason being that single-guide RNAs (sgRNAs), consisting of the transacting CRISPR RNA (tracRNA) and the CRISPR RNA (crRNA), are used for sequence targeting in the CRISPR/cas9 system instead of DNA-binding proteins (Gaj *et al.*, 2013). SgRNAs are easier to design compared to DNA-binding proteins and can target the desired genomic regions with high precision. The CRISPR/cas9 system is well established in mammalian research, but more efforts have to be made to improve its accuracy and efficiency in plants, especially in soybean. Based on the whole-genome sequence information, bioinformatic tools are now available for designing sgRNAs with off-target predictions (Brazelton *et al.*, 2015), making this technology more accessible to researchers. Web-based tools tailored for identifying CRISPR targets in the soybean genome are also available (Michno *et al.*, 2015).

Great efforts have been made to optimize the CRISPR system. RNA polymerase III *U6* promoters are usually used to drive the expression of the sgRNAs. It has been discovered that driving the sgRNA by the native soybean *U6* promoter can bring about better mutation efficiency. Compared to the Arabidopsis *U6-26* promoter, the soybean *U6-10* promoter can increase the mutation efficiency from 3.2–9.7% to 14.7–20.2% (Sun *et al.*, 2015). In another study, the mutation rate of the target is 43.4–48.1% using the soybean *U6-16g-1* (*GmU6-16g-1*) promoter-driven sgRNA, compared to 11.7–18.1% using the *AtU6-26* promoter-driven sgRNA (Du *et al.*, 2016). On the other hand, a report also suggested that the use of meristemic or germline-specific promotors can increase the inheritability of the mutation (Osakabe *et al.*, 2016). Furthermore, there is a synthetic *cas9* gene (*GmCas9*) that is optimized for soybean based on the soybean preferred codon usage to increase the expression efficiency and thus the genome editing rate (Michno *et al.*, 2015).

Off-target mutations are occasionally detected when using the CRISPR/cas9 system, making it a serious concern especially when the soybean genome contains many highly similar duplicated genes (Du *et al.*, 2016; Sun *et al.*, 2015). To test the efficacy of new CRISPR constructs, one may transform these constructs into protoplasts (single cells) or soybean hairy roots (multicellular organs) to assess the mutation rate and off-target rate before making whole-plant transgenics. To date, since CRISPR/cas9 is a relatively new technology in the field of soybean research, for which there are only a handful of published reports (Cai *et al.*, 2015; Jacobs *et al.*, 2015; Michno *et al.*, 2015; Sun *et al.*, 2015; Du *et al.*, 2016). Overall, there has been little focus on adaptations. To improve soybean adaptations toward climate change, researchers will have to focus on mutating and knocking down the stress-sensitive genes or replacing the non-functional/sensitive allele in an elite germplasm with a functional/tolerant one from a non-commercial germplasm.

Conclusion and perspectives

Current genomic studies have the potential to identify candidate loci or genes for soybean improvement, a number of which have been identified through curation, mapping and studies on differential transcription, translation and post-translational modifications. Stress-related gene candidates can be used in soybean improvement for adapting to climatic changes. Nevertheless, most of the soybean genomic studies remain at the gene discovery stage. Candidate genomic loci are waiting for fine mapping and causal gene identification. Thousands of gene candidates from genomic studies have not been functionally tested. Efforts have to be made to fill these gaps which are essential for strategic breeding programs including gene stacking and field application.

Current studies are usually focused on the effects of short periods of stress treatment (lasting from hours to less than a month) at the seedling stages of soybean. However, the effects of climate change are life-long and can persist for generations. Therefore, more intensive studies should be carried out by comparing the adaptability of soybean at different stages of development. Furthermore, the study of inheritable epigenetic imprints will also be useful for understanding how plants inherit acquired adaptations from previous generations.

Although climate change brings about various stresses to crops and hence hampers crop production, it has also created new opportunities for agriculture. Crop production is predicted to move away from the equator as a result of warming climate. High-latitude regions, which were previously not suitable for crop production due to low average annual temperatures, will soon have annual accumulated temperatures high enough for crop growth (Long and Ort, 2010). Nonetheless, other factors such as day length, soil fertility and water availability in these regions are normally suboptimal for plant growth. Consequently, only early-mature soybean varieties are suitable for cultivation in high-latitude regions. In addition to improving the adaptability of crops against stresses on current arable lands, research should also be conducted to enhance the adaptability of soybean in these high-latitude regions.

Acknowledgements

This work was supported by grants from the Hong Kong Research Grants Council (General Research Fund: 14108014) to HML. Ms Jee-Yan Chu copyedited the manuscript. We also thank Miss Fuk-Ling Wong and Mr Ming-Sin Ng for proofreading the manuscript.

References

Abdel-Haleem H, Carter TE, Purcell LC, King CA, Ries LL, Chen PY, Schapaugh W, Sinclair TR, Boerma HR. 2012. Mapping of quantitative trait loci for canopy-wilting trait in soybean (*Glycine max* L. Merr). Theoretical and Applied Genetics **125,** 837–846.

Abdel-Haleem H, Lee GJ, Boerma RH. 2011. Identification of QTL for increased fibrous roots in soybean. Theoretical and Applied Genetics **122,** 935–946.

Aebersold R, Mann M. 2003. Mass spectrometry-based proteomics. Nature **422,** 198–207.

Aghamirzaie D, Batra D, Heath LS, Schneider A, Grene R, Collakova E. 2015. Transcriptome-wide functional characterization reveals novel relationships among differentially expressed transcripts in developing soybean embryos. BMC Genomics **16,** 928.

Ainsworth EA, Rogers A, Vodkin LO, Walter A, Schurr U. 2006. The effects of elevated CO_2 concentration on soybean gene expression. An analysis of growing and mature leaves. Plant Physiology **142,** 135–147.

Arikit S, Xia R, Kakrana A, Huang K, *et al*. 2014. An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. Plant Cell **26,** 4584–4601.

Bantscheff M, Lemeer S, Savitski MM, Kuster B. 2012. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Analytical and Bioanalytical Chemistry **404,** 939–965.

Bao Y, Vuong T, Meinhardt C, Tiffin P, Denny R, Chen SY, Nguyen HT, Orf JH, Young ND. 2014. Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. Plant Genome **7, doi:**10.3835/plantgenome2013.3811.0039.

Bastien M, Sonah H, Belzile F. 2014. Genome wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a genotyping-by-sequencing approach. Plant Genome **7, doi:** 10.3835/plantgenome2013.3810.0030.

Borges F, Martienssen RA. 2015. The expanding world of small RNAs in plants. Nature Reviews Molecular Cell Biology **16,** 727–741.

Brazelton VA, Zarecor S, Wright DA, Wang Y, Liu J, Chen KT, Yang B, Lawrence-Dill CJ. 2015. A quick guide to CRISPR sgRNA design tools. Gm Crops & Food-Biotechnology in Agriculture and the Food Chain **6,** 266–276.

Brevik E. 2013. The potential impact of climate change on soil properties and processes and corresponding influence on food security. Agriculture **3,** 398.

Burton AL, Burkey KO, Carter TE, Orf J, Cregan PB. 2016. Phenotypic variation and identification of quantitative trait loci for ozone tolerance in a Fiskeby III x Mandarin (Ottawa) soybean population. Theoretical and Applied Genetics **129,** 1113–1125.

Cai YP, Chen L, Liu XJ, Sun S, Wu CX, Jiang BJ, Han TF, Hou WS. 2015. CRISPR/Cas9-mediated genome editing in soybean hairy roots. PLoS One **10,** e0136064.

Chen KL, Gao CX. 2014. Targeted genome modification technologies and their applications in crop improvements. Plant Cell Reports **33,** 575–583.

Chung WH, Jeong N, Kim J, *et al*. 2014. Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. DNA Research **21,** 153–167.

Collard BCY, Mackill DJ. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B – Biological Sciences **363,** 557–572.

Curtin SJ, Zhang F, Sander JD, *et al*. 2011. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant Physiology **156,** 466–473.

Dasgupta S, Hossain MM, Huq M, Wheeler D. 2015. Climate change and soil salinity: the case of coastal Bangladesh. Ambio **44,** 815–826.

Dentener F, Stevenson D, Cofala J, Mechler R, Amann M, Bergamaschi P, Raes F, Derwent R. 2005. The impact of air pollutant and methane emission controls on tropospheric ozone and radiative forcing: CTM calculations for the period 1990–2030. Atmospheric Chemistry and Physics **5,** 1731–1755.

Donat MG, Lowry AL, Alexander LV, Ogorman PA, Maher N. 2016. More extreme precipitation in the world's dry and wet regions. Nature Climate Change **6,** 508–513.

Dreger M. 2003. Emerging strategies in mass-spectrometry based proteomics. European Journal of Biochemistry **270,** 569–569.

Du HY, Zeng XR, Zhao M, Cui XP, Wang Q, Yang H, Cheng H, Yu DY. 2016. Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. Journal of Biotechnology **217,** 90–97.

Fragkostefanakis S, Roth S, Schleiff E, Scharf KD. 2015. Prospects of engineering thermotolerance in crops through modulation of heat stress transcription factor and heat shock protein networks. Plant, Cell & Environment **38,** 1881–1895.

Gaj T, Gersbach CA, Barbas CF. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology **31,** 397–405.

Galant A, Koester RP, Ainsworth EA, Hicks LM, Jez JM. 2012. From climate change to molecular response: redox proteomics of ozone-induced responses in soybean. New Phytologist **194,** 220–229.

Go YM, Jones DP. 2013. The redox proteome. Journal of Biological Chemistry **288,** 26512–26520.

Goodstein DM, Shu SQ, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40,** D1178–D1186.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. Nucleic Acids Research **36,** D154–D158.

Guan RX, Qu Y, Guo Y, *et al*. 2014. Salinity tolerance in soybean is modulated by natural variation in GmSALT3. Plant Journal **80,** 937–950.

Guleria P, Mahajan M, Bhardwaj J, Yadav SK. 2011. Plant small RNAs: biogenesis, mode of action and their roles in abiotic stresses. Genomics, Proteomics & Bioinformatics **9,** 183–199.

Ha BK, Vuong TD, Velusamy V, Nguyen HT, Shannon JG, Lee JD. 2013. Genetic mapping of quantitative trait loci conditioning salt tolerance in wild soybean (*Glycine soja*) PI 483463. Euphytica **193,** 79–88.

Havlík P, Valin HJP, Gusti M, Schmid E, Forsell N, Herrero M, Khabarov N, Mosnier A, Cantele M, Obersteiner M. 2015. Climate change impacts and mitigation in the developing world: an integrated assessment of the agriculture and forestry sectors. *Policy Research Working Paper*, **Vol. 1**. Washington, DC: World Bank Group.

He JF, Zhao XQ, Laroche A, Lu ZX, Liu HK, Li ZQ. 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science **5, doi:** 10.3389/Fpls.2014.00484.

Huang XH, Feng Q, Qian Q, *et al*. 2009. High-throughput genotyping by whole-genome resequencing. Genome Research **19,** 1068–1076.

Hwang TY, Sayama T, Takahashi M, *et al*. 2009. High-density integrated linkage map based on SSR markers in soybean. DNA Research **16,** 213–225.

Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA. 2015. Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotechnology **15,** 16.

Jarquin D, Kocak K, Posadas L, Hyma K, Jedlicka J, Graef G, Lorenz A. 2014. Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC Genomics **15,** 740.

Joshi T, Yan Z, Libault M, *et al*. 2010. Prediction of novel miRNAs and associated target genes in *Glycine max*. BMC Bioinformatics **11,** S14.

Kim KD, El Baidouri M, Abernathy B, Iwata-Otsubo A, Chavarro C, Gonzales M, Libault M, Grimwood J, Jackson SA. 2015. A comparative epigenomic analysis of polyploidy-derived genes in soybean and common bean. Plant Physiology **168,** 1433–1447.

Kim MY, Lee S, Van K, *et al*. 2010. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proceedings of the National Academy of Sciences, USA **107,** 22032–22037.

Kinoshita T, Seki M. 2014. Epigenetic memory for stress response and adaptation in plants. Plant and Cell Physiology **55,** 1859–1863.

Komatsu S, Wada T, Abalea Y, Nouri MZ, Nanjo Y, Nakayama N, Shimamura S, Yamamoto R, Nakamura T, Furukawa K. 2009. Analysis of plasma membrane proteome in soybean and application to flooding stress response. Journal of Proteome Research **8,** 4487–4499.

Krishna P, Gloor G. 2001. The Hsp90 family of proteins in *Arabidopsis thaliana*. Cell Stress & Chaperones **6,** 238–246.

Kulcheski FR, de Oliveira LFV, Molina LG, *et al*. 2011. Identification of novel soybean microRNAs involved in abiotic and biotic stresses. BMC Genomics **12,** 307.

Lam HM, Xu X, Liu X, *et al*. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nature Genetics **42,** 1053–1059.

Larance M, Lomond AI. 2015. Multidimensional proteomics for cell biology. Nature Reviews Molecular Cell Biology **16,** 269–280.

Leakey ADB, Xu F, Gillespie KM, McGrath JM, Ainsworth EA, Ort DR. 2009. Genomic basis for stimulated respiration by plants growing under elevated carbon dioxide. Proceedings of the National Academy of Sciences, USA **106,** 3597–3602.

Leff B, Ramankutty N, Foley JA. 2004. Geographic distribution of major crops across the world. Global Biogeochemical Cycles **18**.

Li B, Tian L, Zhang J, Huang L, Han F, Yan S, Wang L, Zheng H, Sun J. 2014*a*. Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in *Glycine max*. BMC Genomics **15,** 1086.

Li PS, Yu TF, He GH, Chen M, Zhou YB, Chai SC, Xu ZS, Ma YZ. 2014*b*. Genome-wide analysis of the Hsf family in soybean and functional identification of GmHsf-34 involvement in drought and heat stresses. BMC Genomics **15,** 1009.

Li YH, Liu YL, Reif JC, Liu ZX, Liu B, Mette MF, Chang RZ, Qiu LJ. 2014*c*. Biparental resequencing coupled with SNP genotyping of a segregating population offers insights into the landscape of recombination and fixed genomic regions in elite soybean. G3-Genes Genomes Genetics **4,** 553–560.

Li YH, Zhou G, Ma J *et al*. 2014*d*. *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nature Biotechnology **32,** 1045–1052.

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, Stacey G. 2010. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. Plant Journal **63,** 86–99.

Lin BL, Wang JS, Liu HC, Chen RW, Meyer Y, Barakat A, Delseny M. 2001. Genomic analysis of the Hsp70 superfamily in *Arabidopsis thaliana*. Cell Stress & Chaperones **6,** 201–208.

Long SP, Ort DR. 2010. More than taking the heat: crops and global change. Current Opinion in Plant Biology **13,** 241–248.

Lopes-Caitar VS, de Carvalho MCCG, Darben LM, Kuwahara MK, Nepomuceno AL, Dias WP, Abdelnoor RV, Marcelino-Guimaraes FC. 2013. Genome-wide analysis of the Hsp20 gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. BMC Genomics **14,** 577.

Luo QY. 2011. Temperature thresholds and crop production: a review. Climatic Change **109,** 583–598.

Lynch DH, Smith DL. 1993. Soybean (*Glycine max*) nodulation and N2-fixation as affected by exposure to a low root-zone temperature. Physiologia Plantarum **88,** 212–220.

Mall RK, Lal M, Bhatia VS, Rathore LS, Singh R. 2004. Mitigating climate change impact on soybean productivity in India: a simulation study. Agricultural and Forest Meteorology **121,** 113–125.

Meerburg BG, Verhagen A, Jongschaap RE, Franke AC, Schaap BF, Dueck TA, van der Werf A. 2009. Do nonlinear temperature effects indicate severe damages to US crop yields under climate change? Proceedings of the National Academy of Sciences, USA **106,** E120.

Mertins P, Qiao JW, Patel J, Udeshi ND, Clauser KR, Mani DR, Burgess MW, Gillette MA, Jaffe JD, Carr SA. 2013. Integrated proteomic analysis of post-translational modifications by serial enrichment. Nature Methods **10,** 634–637.

Mesquita RO, Soares ED, de Barros EG, Loureiro ME. 2012. Method optimization for proteomic analysis of soybean leaf: Improvements in identification of new and low-abundance proteins. Genetics and Molecular Biology **35,** 353–361.

Michno JM, Wang XB, Liu JQ, Curtin SJ, Kono TJY, Stupar RM. 2015. CRISPR/Cas mutagenesis of soybean and *Medicago truncatula* using a new web-tool and a modified Cas9 enzyme. Gm Crops & Food-Biotechnology in Agriculture and the Food Chain **6,** 243–252.

Mourtzinis S, Specht JE, Lindsey LE, *et al*. 2015. Climate-induced reduction in US-wide soybean yields underpinned by region- and in-season-specific responses. Nature Plants **1,** 1–4.

Muis S, Guneralp B, Jongman B, Aerts JC, Ward PJ. 2015. Flood risk and adaptation strategies under climate change and urban expansion: a probabilistic analysis using global data. The Science of theTotal Environment **538,** 445–457.

Mustafa G, Komatsu S. 2014. Quantitative proteomics reveals the effect of protein glycosylation in soybean root under flooding stress. Frontiers in Plant Science **5, doi:** 10.3389/Fpls.2014.00627.

Nakaya A, Isobe SN. 2012. Will genomic selection be a practical method for plant breeding? Annals of Botany **110,** 1303–1316.

Nguyen VT, Vuong TD, VanToai T, Lee JD, Wu X, Mian MAR, Dorrance AE, Shannon JG, Nguyen HT. 2012. Mapping of quantitative trait loci associated with resistance to *Phytophthora sojae* and flooding tolerance in soybean. Crop Science **52,** 2481–2493.

Norby RJ, Zak DR. 2011. Ecological lessons from free-air CO2 enrichment (FACE) experiments. Annual Review of Ecology, Evolution, and Systematics **42,** 181–203.

Nouri MZ, Komatsu S. 2010. Comparative analysis of soybean plasma membrane proteins under osmotic stress using gel-based and LC MS/MS-based proteomics approaches. Proteomics **10,** 1930–1945.

Ohnishi S, Funatsuki H, Kasai A, *et al*. 2011. Variation of GmIRCHS (Glycine max inverted-repeat CHS pseudogene) is related to tolerance of low temperature-induced seed coat discoloration in yellow soybean. Theoretical and Applied Genetics **122,** 633–642.

Olsen JV, Mann M. 2013. Status of large-scale analysis of post-translational modifications by mass spectrometry. Molecular & Cellular Proteomics **12,** 3444–3452.

Osakabe Y, Watanabe T, Sugano SS, Ueta R, Ishihara R, Shinozaki K, Osakabe K. 2016. Optimization of CRISPR/Cas9 genome editing to modify abiotic stress responses in plants. Scientific Reports **6,** doi:10.1038/srep26685.

Pan Q, Shai O, Lee LJ, Frey J, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature Genetics **40,** 1413–1415.

Patil G, Do T, Vuong TD, Valliyodan B, Lee JD, Chaudhary J, Shannon JG, Nguyen HT. 2016. Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. Scientific Reports **6, doi:** 10.1038/srep19199.

Pi EX, Qu LQ, Hu JW, Huang YY, Qiu LJ, Lu H, Jiang B, Liu C, Peng TT, Zhao Y, Wang HZ, Tsai SN, Ngai SM, Du LQ. 2016. Mechanisms of soybean roots' tolerances to salinity revealed by proteomic and phosphoproteomic comparisons between two cultivars. Molecular & Cellular Proteomics **15,** 266–288.

Qi X, Li MW, Xie M, *et al*. 2014*a*. Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nature communications **5,** 4340.

Qi ZM, Huang L, Zhu RS, Xin DW, Liu CY, Han X, Jiang HW, Hong WG, Hu GH, Zheng HK, Chen QS. 2014*b*. A high-density genetic map for soybean based on specific length amplified fragment sequencing. PLoS One **9,** e104871.

Qiu J, Wang Y, Wu S, *et al*. 2014. Genome re-sequencing of semi-wild soybean reveals a complex Soja population structure and deep introgression. PLoS One **9,** e108479.

Rahmani H, Saleh-Rastin N, Khavazi K, Asgharzadeh A, Fewer D, Kiani S, Lindstrom K. 2009. Selection of thermotolerant bradyrhizobial strains for nodulation of soybean (*Glycine max* L.) in semi-arid regions of Iran. World Journal of Microbiology & Biotechnology **25,** 591–600.

Reyes-Fox M, Steltzer H, Trlica MJ, McMaster GS, Andales AA, LeCain DR, Morgan JA. 2014. Elevated CO_2 further lengthens growing season under warming conditions. Nature **510,** 259–262.

Rodrigues EP, Torres AR, Batista JSD, Huergo L, Hungria M. 2012. A simple, economical and reproducible protein extraction protocol for proteomics studies of soybean roots. Genetics and Molecular Biology **35,** 348–352.

Sarkar NK, Kim YK, Grover A. 2009. Rice sHsp genes: genomic organization and expression profiling under stress and development. BMC Genomics **10,** 393.

Sarkar NK, Kundnani P, Grover A. 2013. Functional analysis of Hsp70 superfamily proteins of rice (*Oryza sativa*). Cell Stress & Chaperones **18,** 427–437.

Scharf KD, Berberich T, Ebersberger I, Nover L. 2012. The plant heat stress transcription factor (Hsf) family: Structure, function and evolution. Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms **1819,** 104–119.

Schlenker W, Roberts MJ. 2009. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. Proceedings of the National Academy of Sciences, USA **106,** 15594–15598.

Schmitz RJ, He YP, Valdes-Lopez O, *et al*. 2013. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Research **23,** 1663–1674.

Schmutz J, Cannon SB, Schlueter J, *et al*. 2010. Genome sequence of the palaeopolyploid soybean. Nature **463,** 178–183.

Severin AJ, Woody JL, Bolon YT, *et al*. 2010. RNA-Seq atlas of Glycine max: A guide to the soybean transcriptome. BMC Plant Biology **10,** 160.

Shamimuzzaman M, Vodkin L. 2012. Identification of soybean seed developmental stage-specific and tissue-specific miRNA targets by degradome sequencing. BMC Genomics **13,** 310.

Shimomura M, Kanamori H, Komatsu S, *et al*. 2015. The *Glycine max* cv. Enrei genome for improvement of Japanese soybean cultivars. International journal of Genomics 2015, 358127.

Shu YJ, Yu DS, Wang D, Bai X, Zhu YM, Guo CH. 2013. Genomic selection of seed weight based on low-density SCAR markers in soybean. Genetics and Molecular Research **12,** 2178–2188.

Song QJ, Hyten DL, Jia GF, Quigley CV, Fickus EW, Nelson RL, Cregan PB. 2013*a*. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One **8,** e54985.

Song QJ, Hyten DL, Jia GF, Quigley CV, Fickus EW, Nelson RL, Cregan PB. 2015*a*. Fingerprinting soybean germplasm and its utility in genomic research. G3-Genes Genomes Genetics **5,** 1999–2006.

Song QX, Lu X, Li QT, Chen H, Hu XY, Ma B, Zhang WK, Chen SY, Zhang JS. 2013*b*. Genome-wide analysis of DNA methylation in soybean. Molecular Plant **6,** 1961–1974.

Song XF, Wei HC, Cheng W, Yang SX, Zhao YX, Li X, Luo D, Zhang H, Feng XZ. 2015*b*. Development of INDEL markers for genetic mapping based on whole genome resequencing in soybean. G3-Genes Genomes Genetics **5,** 2793–2799.

Song YG, Ji DD, Li S, Wang P, Li Q, Xiang FN. 2012. The dynamic changes of DNA methylation and histone modifications of salt responsive transcription factor genes in soybean. PLoS One **7,** e41274.

Stepinski D. 2012. Levels of DNA methylation and histone methylation and acetylation change in root tip cells of soybean seedlings grown at different temperatures. Plant Physiology and Biochemistry **61,** 9–17.

Sun XJ, Hu Z, Chen R, Jiang QY, Song GH, Zhang H, Xi YJ. 2015. Targeted mutagenesis in soybean using the CRISPR-Cas9 system. Scientific Reports **5, doi:**10.1038/srep10342.

Sun XW, Liu DY, Zhang XF, *et al*. 2013. SLAF-seq: An efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. PLoS One **8,** e58700.

Sun ZX, Wang YN, Mou FP, Tian YP, Chen L, Zhang SL, Jiang Q, Li X. 2016. Genome-side small RNA analysis of soybean reveals auxin-responsive microRNAs that are differentially expressed in response to salt stress in root apex. Frontiers in Plant Science **6, doi:** 10.3389/Fpls.2015.01273.

Thibaud-Nissen F, Souvorov A, Murphy T, DiCuccio M, Kitts P. 2013. Eukaryotic Genome Annotation Pipline. *The NCBI Handbook[Internet]*: National Center for Biotechnology Information.

Tricker PJ. 2015. Transgenerational inheritance or resetting of stress-induced epigenetic modifications: two sides of the same coin. Frontiers in Plant Science **6, doi:** 10.3389/fpls.2015.00699.

Turner M, Yu O, Subramanian S. 2012. Genome organization and characteristics of soybean microRNAs. BMC Genomics **13,** 169.

Tyagi SK, Tripathi RP. 1983. Effect of temperature on soybean germination. Plant and Soil **74,** 273–280.

Vadivel AKA. 2015. Gel-based proteomics in plants: time to move on from the tradition. Frontiers in Plant Science **6, doi:** 10.3389/Fpls.2015.00369.

Wang WX, Vinocur B, Shoseyov O, Altman A. 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in Plant Science **9,** 244–252.

Waters ER, Aevermann BD, Sanders-Reed Z. 2008. Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. Cell Stress & Chaperones **13,** 127–142.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. 2007. An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. PLoS One **2,** e718.

Woo JW, Kim J, Il Kwon S, Corvalan C, Cho SW, Kim H, Kim SG, Kim ST, Choe S, Kim JS. 2015. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nature Biotechnology **33,** 1162–U1156.

Xu JY, Xue CC, Xue D, Zhao JM, Gai JY, Guo N, Xing H. 2013*a*. Overexpression of GmHsp90s, a Heat Shock Protein 90 (Hsp90) gene family cloning from soybean, decrease damage of abiotic stresses in *Arabidopsis thaliana*. PLoS One **8,** e69810.

Xu X, Zeng L, Tao Y, Vuong T, Wan J, Boerma R, Noe J, Li Z, Finnerty S, Pathan SM, Shannon JG, Nguyen HT. 2013*b*. Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. Proceedings of the National Academy of Sciences, USA **110,** 13469–13474.

Xu YG, Guo MZ, Liu XY, Wang CY, Liu Y. 2014*a*. Inferring the soybean (*Glycine max*) microRNA functional network based on target gene network. Bioinformatics **30,** 94–103.

Xu YG, Guo MZ, Liu XY, Wang CY, Liu Y. 2014*b*. SoyFN: a knowledge database of soybean functional networks. Database-the Journal of Biological Databases and Curation, doi: 10.1093/database/bau1019.

Zaffagnini M, Fermani S, Costa A, Lemaire SD, Trost P. 2013. Plant cytoplasmic GAPDH: redox post-translational modifications and moonlighting properties. Frontiers in Plant Science **4,** doi:10.3389/fpls.2013.00450.

Zeng QY, Yang CY, Ma QB, Li XP, Dong WW, Nian H. 2012. Identification of wild soybean miRNAs and their target genes responsive to aluminum stress. BMC Plant Biology **12,** 182.

Zhang D, Li H, Wang J, Zhang H, Hu Z, Chu S, Lv H, Yu D. 2016. High-density genetic mapping identifies new major loci for tolerance to low-phosphorus stress in soybean. Frontiers in Plant Science **7, doi:** 10.3389/fpls.2016.00372.

Zhang J, Li JB, Liu BB, Zhang L, Chen J, Lu MZ. 2013. Genome-wide analysis of the *Populus Hsp90* gene family reveals differential expression patterns, localization, and heat stress responses. BMC Genomics **14,** 532.

Zhang L, Zhao HK, Dong QL, Zhang YY, Wang YM, Li HY, Xing GJ, Li QY, Dong YS. 2015. Genome-wide analysis and expression profiling under heat and drought treatments of *HSP70* gene family in soybean (*Glycine max* L.). Frontiers in Plant Science **6, doi:** 10.3339/Fpls.2015.00773.

Zhang SL, Wang YN, Li KX, Zou YM, Chen L, Li X. 2014. Identification of cold-responsive miRNAs and their target genes in nitrogen-fixing nodules of soybean. International Journal of Molecular Sciences **15,** 13596–13614.

Zhang WB, Jiang HW, Qiu PC, Liu CY, Chen FL, Xin DW, Li CD, Hu GH, Chen QS. 2012. Genetic overlap of QTL associated with low-temperature tolerance at germination and seedling stage using BILs in soybean. Canadian Journal of Plant Science **92,** 1381–1388.

Zhou ZK, Jiang Y, Wang Z, *et al*. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nature Biotechnology **33,** 408–U125.