SCIENCE CHINA
Life Sciences

• INSIGHT •

August 2015 Vol.58 No.8: 813–815
doi: 10.1007/s11427-015-4908-2

Genomics progress will facilitate molecular breeding in soybean

WANG Zheng & TIAN ZhiXi*

State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

Received June 23, 2015; accepted July 8, 2015; published online July 17, 2015

Citation: Wang Z, Tian ZX. Genomics progress will facilitate molecular breeding in soybean. Sci China Life Sci, 2015, 58: 813−815, doi: 10.1007/s11427-015-4908-2

It has been suggested that the soybean (*Glycine max* [L.] Merr.) currently cultivated in commercial agriculture was domesticated from the wild soybean (*G. soja* Sieb. & Zucc.) in China approximately 5,000 years ago. Because of its high protein and oil content, cultivated soybean has become a major economic crop by providing 69% and 30% of the world's plant protein and oil. Soybean breeders have made considerable efforts to develop elite varieties that can meet this ever-increasing demand. However, over the last century, advances in soybean breeding have progressed slowly. One of the major reasons is the genetic bottlenecks caused by the domestication practice of using seeds from only a small number of plants with desirable traits to propagate each new generation during introduction and improvement [1]. Well-established genome sequences and a better understanding of the underlying genetic bases of agronomically important traits will expedite the progress of marker-assisted breeding programs for soybean.

At the beginning of 2010, the first assembled reference genome of a soybean cultivar, Williams 82, was released [2]. This draft genome sequence demonstrated that soybean contains 20 chromosomes, and approximately 60% of the genome was occupied by repetitive DNA. The genome consisted of 46,430 predicted protein-coding genes, 78% of which were located at the end of a chromosome. Moreover, the genome sequence revealed that this genome experienced two rounds of whole-genome duplication, one 59 million years ago and the other 13 million years ago, which resulted in the presence of nearly 75% of genes in multiple copies. This made the diploid soybean more like an ancient palaeopolyploid. This reference genome is a milestone of soybean genomics. It ushered in the era of soybean functional genomics.

Soon after the publication of the Williams 82 reference genome, Kim *et al*. [3] published the genome sequence of a wild soybean (*G. soja* var. IT182932), generated using two MPS platforms: Illumina-GA and GS-FLX. A difference of approximately 0.31% was found between the genomes of *G. max* and *G. soja*, which included 2.5 Mb of substituted bases and 406 kb of small insertions/deletions. In addition, they detected 32.4 Mb of large deletions and 8.3 Mb of novel sequence contigs in the *G. soja* genome. Recently, significant work on the pan-genome of *G. soja* was carried out by sequencing and *de novo* assembling seven phylogenetically and geographically representative accessions [4]. This pan-genome covered 94% of the genes of *Glycine max*. Nearly half of the gene families and 80% of the pan-genome sequence were present in all seven *Glycine soja* accessions. The other gene families were dispensable and exhibited greater variation, which may be responsible for the variation of agronomic traits. This study also revealed that, comparing to the cultivar accessions, more genes associated with abiotic and biotic tolerance and mitigation were selected during the course of evolution. These additional genomes provide valuable information about lineage-specific sequences, which will be of great benefit to future gene identification studies.

Compared to wild soybean, modern cultivated soybean exhibited significant changes in both morphological characteristics and seed quality, suggesting that strong genetic

*Corresponding author (email: zxtian@genetics.ac.cn)

selection has been experienced during soybean breeding. Prior to the publication of the reference genome, an analysis of 111 fragments from 102 genes indicated that genetic bottlenecks have occurred during soybean domestication and improvement [1]. To obtain a comprehensive overview of the genome-wide sequence variation, Lam *et al*. [5] resequenced the genomes of a diverse group of 17 wild and 14 cultivated soybean accessions, identifying a total of 6,318,109 SNPs and 186,177 present and absent variations. Genomic analysis suggested that, unlike other crops, the soybean genomes had higher levels of linkage disequilibrium (LD) and a high ratio of average nonsynonymous to synonymous nucleotides. In addition, cultivated accessions exhibited lower genetic diversity than wild soybeans. Li *et al*. [6] then published the resequenced genomes of 8 wild soybeans, 8 landraces, and 8 elite soybeans. Combined with the sequence data of the 31 accessions, this study provided a clear molecular footprinting picture of soybean domestication and improvement. They found that genetic diversity was reduced more significantly during domestication than during the switch from landraces to elite cultivars. They also detected that approximately 2.99% of all genomic regions were affected by artificial selection. Moreover, selection regions were distributed unequally throughout the genome, with some selection hotspots existing in certain genomic regions.

Previous studies have suggested that large numbers of rare alleles (frequency<0.10) were lost during domestication and improvement, and that introduction and adaptive selection have occurred in different geographic areas [1]. To probe how soybean subpopulations have adapted to different geographic areas, and to identify the genes responsible for domestication traits, Zhou *et al*. performed a large-scale assessment of soybean domestication and improvement by re-sequencing 302 wild, landrace, and improved soybean accessions at >11× depth [7]. The large dataset produced approximately two-fold more SNPs and small indels than those previously reported. In addition, the depth of the sequencing data identified 1,614 copy number variations. By independently analyzing the LD of different genomic regions, the results provided a detailed LD landscape across the genome, and suggested that LD was much higher in pericentromeric regions than in arm regions. Bioinformatics analysis identified a total of 121 and 109 selective sweeps during soybean domestication and improvement, respectively. Moreover, through a genome-wide association study, this study revealed associations between 10 selected regions and 9 domestication or improvement traits, and identified 13 previously uncharacterized loci for agronomic traits. Particularly, an investigation involving the combination of previously generated quantitative trait loci (QTL) information revealed that 96 of the 230 selected regions correlated with reported oil QTLs, and 21 contained fatty acid biosynthesis genes. Further pairwise population differentiation analysis across geographic groups identified large numbers of local differentiation signals that were not detected during previous screenings of domestication and improvement selection. The comprehensive investigations carried out in this study provide a valuable resource for future studies of the allelic variation of relevant traits, thereby facilitating soybean crop improvement.

The soybean reference genome accelerated advances in functional genomics work in soybean, and a set of genes responsible for important traits were cloned in recent years [8, 9]. However, compared to other crops like rice and maize, our understanding of the genetic regulation of complex traits in soybean is limited. We believe that the rapid progress from more *de novo* assembled and resequenced genomes will shed light on the genetic bases of agronomically important traits. In addition, knowledge from other omics studies, such as transcriptomics, epigenomics, proteomics, and metabolomics analyses, will be needed to discover the deeper regulation network governing complex traits. In summary, the accumulation of genomics studies in soybean will bring molecular breeding from theory to practice.

1 Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA, 2006, 103: 16666–16671

2 Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. Genome sequence of the palaeopolyploid soybean. Nature, 2010, 463: 178–183

3 Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, Kim DS, Lee YS, Park D, Ma J, Kim WY, Kim BC, Park S, Lee KA, Kim DH, Kim KH, Shin JH, Jang YE, Kim KD, Liu WX, Chaisan T, Kang YJ, Lee YH, Kim KH, Moon JK, Schmutz J, Jackson SA, Bhak J, Lee SH. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proc Natl Acad Sci USA, 2010, 107: 22032–22037

4 Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang SS, Zuo Q, Shi XH, Li YF, Zhang WK, Hu Y, Kong G, Hong HL, Tan B, Song J, Liu ZX, Wang Y, Ruan H, Yeung CK, Liu J, Wang H, Zhang LJ, Guan RX, Wang KJ, Li WB, Chen SY, Chang RZ, Jiang Z, Jackson SA, Li R, Qiu LJ. *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol, 2014, 32: 1045–1052

5 Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, Li J, Jian M, Wang J, Shao G, Wang J, Sun SS, Zhang G. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet, 2010, 42: 1053–1059

6 Li YH, Zhao SC, Ma JX, Li D, Yan L, Li J, Qi XT, Guo XS, Zhang L, He WM, Chang RZ, Liang QS, Guo Y, Ye C, Wang XB, Tao Y, Guan RX, Wang JY, Liu YL, Jin LG, Zhang XQ, Liu ZX, Zhang LJ, Chen J, Wang KJ, Nielsen R, Li RQ, Chen PY, Li WB, Reif JC, Purugganan M, Wang J, Zhang MC, Wang J, Qiu LJ. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. BMC Genomics, 2013, 14: 579–590

7 Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao

Y, Ma Y, Fang C, Shen Y, Liu T, Li C, Li Q, Wu M, Wang M, Wu Y, Dong Y, Wan W, Wang X, Ding Z, Gao Y, Xiang H, Zhu B, Lee SH, Wang W, Tian Z. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol, 2015, 33: 408–414

8 Chan C, Qi X, Li MW, Wong FL, Lam HM. Recent developments of genomic research in soybean. J Genet Genomics, 2012, 39: 317–324

9 Xia Z, Zhai H, Lu S, Wu H, Zhang Y. Recent achievement in gene cloning and functional genomics in soybean. Scientific World J, 2013, 2013: 281367