ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2011) **9**(2); 155–158
doi:10.1017/S1479262111000098

Genomics of plant genetic resources: past, present and future

Kyujung Van[1], Dong Hyun Kim[1], Jin Hee Shin[1] and Suk-Ha Lee[1,2]*

[1]*Department of Plant Science and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Korea and* [2]*Plant Genomics and Breeding Institute, Seoul National University, Seoul 151-921, Korea*

Abstract

Plant genetic resources (PGR) include cultivars, landraces, wild species closely related to cultivated varieties, breeder's elite lines and mutants. The loss of genetic diversity caused by the practice of agriculture and the availability of genetic information has resulted in a great effort dedicated to the collection of PGR. Prior to the advent of molecular profiling, accessions in germplasm collections were examined based on morphology. The development of molecular techniques now allows a more accurate analysis of large collections. Next-generation sequencing (NGS) with *de novo* assembly and resequencing has already provided a substantial amount of information, which warrants the coordination of existing databases and their integration into genebanks. Thus, the integration and coordination of genomic data into genebanks is very important and requires an international effort. From the determination of phenotypic traits to the application of NGS to whole genomes, every aspect of genomics will have a great impact not only on PGR conservation, but also on plant breeding programmes.

Keywords: genomics; germplasm collection; next-generation sequencing; plant genetic resources

Introduction

Plant genetic resources (PGR) began to establish around 1993 as a consequence of growing concerns about biodiversity, its conservation and genetic erosion. Although the rate of population growth is slowing down, global food production is still a major challenge for the future of mankind (Hoisington *et al.*, 1999; Hammer, 2003; Gepts, 2006). Therefore, securing PGR for future generations has become a priority not only in developing countries but also in the entire world. The development and application of molecular techniques and genomics have dramatically improved the characterization and deployment of PGR. This review surveys the past and current status of the application of genomics to the PGR characterization and discusses future directions.

* Corresponding author. E-mail: sukhalee@snu.ac.kr

Early impact of genomics on PGR

The advent of agriculture made possible by domestication greatly affected the diversity of crops (Gepts, 2006). The voyages of Christopher Columbus marked the earliest recorded acquisition of new plant resources, and, ever since, collected plants have been conserved in botanical gardens and herbaria (Short, 2003). The rediscovery of Mendel's law in the early 20th century helped the dramatic increase in agricultural productivity, although the overall genetic diversity decreased as a result of modern agricultural practices. Fearing genetic erosion, the world community increased the effort to better evaluate PGR in genebanks (Hoisington *et al.*, 1999).

The characterization of PGR by comparisons of plant morphology, such as yield, colour, texture, taste, etc., is the simplest and easiest approach (Gilbert *et al.*, 1999; Hoisington *et al.*, 1999). In addition to these qualitative/quantitative phenotypic traits, pedigree analysis and geographical distribution are also helpful for measuring

genetic diversity (Hammer, 2003). A renewed impetus towards PGR characterization was made possible by the development of modern molecular techniques.

Current status of plant genomics

The genetic diversity of major crops has been declining through domestication and the introduction of modern plant breeding (Tanksley and McCouch, 1997; Hyten *et al.*, 2009). To prevent the genetic vulnerability of crops and to preserve valuable genetic resources, it needs to collect, preserve, examine and utilize germplasm effectively. The concept of the core set was proposed to minimize replicates and ensure the representation of the maximum genetic diversity of the entire germplasm collection (Frankel, 1984; Brown, 1989; van Hintum, 1999). Phenotyping was the traditional criteria for germplasm evaluation; however, currently, these evaluations are changed to genotyping by molecular markers (Tanksley and McCouch, 1997).

Genetic markers are powerful tools for genetic mapping, and molecular markers are highly polymorphic, easily detected and unaffected by the environment (Andersen and Lubberstedt, 2003). Various molecular markers have been developed, such as restriction fragment length polymorphisms, randomly amplified polymorphic DNA, simple sequence repeats, amplified fragment length polymorphisms and single nucleotide polymorphisms (SNPs) (Gupta *et al.*, 2001), which are used for the construction of genetic and physical maps. These markers are applied in plant breeding for quantitative trait loci (QTLs) mapping, map-based cloning, marker-assisted selection, etc. (Moose and Mumm, 2008).

Previously, genome-sequencing projects depended on Sanger sequencing methods. Recently, introduction of next-generation sequencing (NGS) technologies into plant breeding programmes has enabled the acquisition of high-throughput sequence data inexpensively in a short time (Morozova and Marra, 2008). However, the *de novo* assembly of plant genomes using NGS with short-read length is not yet adequate because most plant genomes are large and harbour long repeat sequences (Varshney *et al.*, 2009). Thus, NGS technologies are applied for the resequencing of species for which a complete reference genome sequence exists and are actively used for high-throughput genotyping of up to a million SNP markers in *Arabidopsis* and several polyploidy crops (Rostoks *et al.*, 2006; Weber *et al.*, 2007; Hyten *et al.*, 2008; Akhunov *et al.*, 2009; Yan *et al.*, 2010). Genome-wide SNP genotyping is a powerful tool for association mapping and evolutionary studies (Akhunov *et al.*, 2009). Furthermore, SNP markers can be used more effectively when combined with genotypes and haplotypes (Hamblin *et al.*, 2007; Yan *et al.*, 2010). This multiplexed genotyping technology facilitates the effective examination and selection of germplasms by unravelling novel and potentially agronomically useful alleles (Tanksley and McCouch, 1997). The QTL mapping of soybean rust was successfully conducted by SNP genotyping using the GoldenGate assay (Hyten *et al.*, 2009). These NGS technologies and the massively developed genome-wide markers are also applied for the construction of high-density maps and genetic diversity analysis (Gupta *et al.*, 2008).

Future directions

A wealth of genetic resources in *Arabidopsis* and other model species have promoted great advances in plant science. Furthermore, whole genome sequencing projects involving more than 20 plants will be completed in the near future (Gupta *et al.*, 2008). With the improvement in sequencing techniques, more genetic resources, including the sequences, will be available in the future. Second (next) generation sequencers – Illumina's GA, Roche's 454 and Applied Biosystems' SOLiD – have generated large amounts of short DNA sequence reads. These have been updated to produce longer read lengths and greater amounts of sequence reads. Currently, several companies are attempting to introduce a new sequencing machine, which will be called third generation sequencing (Rusk, 2009). Helicos Biosciences developed a true single molecule sequencer that sequenced the virus M13 genome by an amplification-free method (Harris *et al.*, 2008). Pacific Biosciences developed a single molecular real-time sequencing machine, based on an assessment of the temporal order of incorporation of fluorescently labelled nucleotides, which can produce reads longer than 1 kb (Eid *et al.*, 2009). Oxford Nanopore's sequencer is designed to avoid amplification or labelling by detecting a direct electrical signal (Clarke *et al.*, 2009). Despite dramatic improvements in sequencing speed and capacity, third generation sequencers will not completely replace the previous sequencing methods. Frequent use by researchers will likely reveal not only the benefits but also the limitations of these new techniques. Similar to the use of second generation sequencers together with ABI 3730, new sequencers will also be used with earlier technologies.

Until several years ago, whole genome plant sequencing projects were limited to model species. However, *de novo* sequencing and assembly are now easier due to longer reads and lower costs, which in the past few years has allowed for much greater sequencing depth. In addition to *de novo* genome sequencing, the whole

https://doi.org/10.1017/S1479262111000098 Published online by Cambridge University Press

genome sequence variations in 1001 accessions of *Arabidopsis* were analyzed in 2008 (Weigel and Mott, 2009). Furthermore, in rice, a high-throughput method for genotyping recombinant populations was developed (Huang *et al*., 2009). Third generation sequencers could be also used for detection of sequence variation, associations between important agronomic traits and gene identification in regulatory networks by ChIP-chip and ChIP-seq protocols.

Presently, bioinformatics is the major bottleneck for a more complete exploitation of the information of genetic resources that is rapidly accumulating. The integration and organization of the available genomic resources to facilitate their use by researchers are therefore important. It could be a similar concept to that of an 'omic space' comprising a comprehensive omic planes (Toyoda and Wada, 2004). Several integrated databases, such as the arabidopsis information resource (*Arabidopsis*), Gramene (rice) and SoyBase (soybean), provide genetic maps, genomic sequences, gene predictions, expressed sequence tags, marker data, QTLs, repetitive sequences, etc. One of the most significant contributions of the comprehensive genomic resources is that it provides a benefit to researchers who want to start new experiments or compare related information.

Acknowledgements

This work was supported by a grant from the BioGreen 21 Project (code no. 20080401034010), Rural Development Administration, the Republic of Korea. S.-H. Lee is grateful for the Senior Visiting Fellowship provided by the Institute of Advanced Studies at the University of Bologna, Italy.

References

Akhunov E, Nicolet C and Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theoretical and Applied Genetics* 119: 507–517.

Andersen JR and Lubberstedt T (2003) Functional markers in plants. *Trends in Plant Science* 8: 554–560.

Brown AHD (1989) The case for core sets. In: Brown AHD, Frankel OH, Marshall DR and Williams JT (eds) *The Use of Plant Genetic Resources*. Cambridge: Cambridge University Press, pp. 136–155.

Clarke J, Wu H-C, Jayasinghe L, Patel A, Reid S and Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology* 4: 265–270.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J and Turner S (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323: 133–138.

Frankel PH (1984) Genetic perspective of germplasm conservation. In: Arber W, Llimensee K, Peacock WJ and Starlinger P (eds) *Genetic manipulations: Impact on Man and Society*. Cambridge: Cambridge University Press, pp. 161–170.

Gepts P (2006) Plant genetic resources conservation and utilization: the accomplishments and future of a societal insurance policy. *Crop Science* 46: 2278–2292.

Gilbert JE, Lewis RV, Wilkinson MJ and Caligari PDS (1999) Developing an appropriate strategy to assess genetic variability in plant germplasm collections. *Theoretical and Applied Genetics* 98: 1125–1131.

Gupta PK, Roy JK and Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science India* 80: 524–535.

Gupta PK, Rustgi S and Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity* 101: 5–18.

Hamblin MT, Warburton ML and Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *Public Library of Science ONE* 2: e1367.

Hammer K (2003) A paradigm shift in the discipline of plant genetic resources. *Genetic Resources and Crop Evolution* 50: 3–10.

Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, DiMeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H and Xie Z (2008) Single-molecule DNA sequencing of a viral genome. *Science* 320: 106–109.

Hoisington D, Khairallah M, Reeves T, Ribaut J-M, Skovmand B, Taba S and Warburton M (1999) Plant genetic resources: what can they contribute toward increased crop productivity? *Proceeding of the National Academy of Sciences USA* 96: 5937–5943.

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T and Han B (2009) High-throughput genotyping by whole-genome resequencing. *Genome Research* 19: 1068–1076.

Hyten DL, Song Q, Choi IY, Yoon MS, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND and Cregan PB (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theoretical and Applied Genetics* 116: 945–952.

Hyten DL, Smith JR, Frederick RD, Tucker ML, Song Q and Cregan PB (2009) Bulked segregant analysis using the GoldenGate assay to locate the *Rpp3* locus that confers resistance to soybean rust in soybean. *Crop Science* 49: 265–271.

Moose SP and Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology* 147: 969–977.

Morozova O and Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92: 255–264.

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ and Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proceedings of the National Academy of Sciences USA* 103: 18656–18661.

Rusk N (2009) Cheap third-generation sequencing. *Nature Methods* 6: 244–245.

Short P (2003) *In Pursuit of Plants*. Portland: Timber Press.

Tanksley SD and McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277: 1063–1066.

Toyoda T and Wada A (2004) Omic space: coordinate-based integration and analysis of genomic phenomic interactions. *Bioinformatics* 20: 1759–1765.

van Hintum TJL (1999) The general methodology for creating a core collection. In: Johnsons RC and Hodgkin T (eds) *Core Set for Today and Tomorrow*. Rome: International Plant Genetic Resources Institute (IPGRI), pp. 10–17.

Varshney RK, Nayak SN, May GD and Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27: 522–530.

Weber APM, Weber KL, Carr K, Wilkerson C and Ohlrogge JB (2007) Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. *Plant Physiology* 144: 32–42.

Weigel D and Mott R (2009) The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biology* 10: 107.

Yan JB, Yang XH, Shah T, Sanchez-Villeda H, Li JS, Warburton M, Zhou Y, Crouch JH and Xu YB (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. *Molecular Breeding* 25: 441–451.