Next generation sequencing: Gold mine or tsunami for breeding science

In general, conventional breeding for new improved cultivars requires continuous crossing and backcrossing that usually takes several years in order to introduce a desirable trait or gene from one cultivar to another. At present, however, breeders are under constant pressure to release new varieties at the shortest possible time due to global challenges of sustaining an ever-growing population and a highly competitive world market. This is further aggravated by constantly changing environment and rapid depletion of available resources for sustaining agriculture. In the last 20 years, two major advances in biotechnology have made great impact on conventional breeding. The DNA recombination technology facilitated the introduction of alien genes to generate genetically modified (GM) crops and led to the development of many cultivars with improved yield and resistance to abiotic stress. At present, more than 110 million hectares of agricultural lands all over the world are being used for cultivation of GM crops, although here in Japan, growing of GM crops has remained a controversial issue due to strong opposition of the public. Large-scale sequencing also changed the way breeders approach many breeding objectives. The development of molecular markers paved the way for DNA marker assisted selection (MAS) technology and accelerated the selection of the most favorable introgression lines for a target gene. This is further enhanced by genetic analysis for screening available resources carrying a favorable trait, construction of fine genetic maps, and molecular biology techniques for cloning genes corresponding to the target trait. The pioneering works on the application of molecular markers in breeding have been chronicled in the book “DNA-Based Markers in Plants” (edited by Ronald L. Philips and Indra K. Vasil) focusing on the implications in improvement of major cereal crops. Phenotyping strategies that rely on polymorphism of nucleotide sequences (RFLP) or simple sequence repeats (SSR) are now widely incorporated in many breeding programs. Furthermore, the sequencing of entire genomes of rice, maize, wheat etc. produced a large amount of sequence data that enhances the efficiency of MAS and pave the way for more target oriented strategies in breeding.

In the last 2–3 years, with the proliferation of next generation sequencers capable of producing millions of DNA sequence reads in a single run, genome analysis is further raised to a higher level. These advances facilitate not only the re-sequencing of model plants to improve the quality of sequence but also the sequencing of entirely new germplasms, wild species, parental and progeny lines of mapping populations, and a wide range of cultivars for major crop species. With these efforts of sequencing of a large number of germplasm for a particular crop, finding new alleles of agronomically useful genes can be accelerated. As an example, in the case of rice, approximately 3.6 million SNPs have been identified by sequencing 517 Chinese landraces which could be used to detect and measure the degree of association between molecular markers and traits of interest (Huang, X. et al., (2010) Nature Genetics, 42: 961–967). In the case of rice and other cultivated crops, genetic structures are different from species to species reflecting the history of domestication so that there is a real potential in exploiting association genetics. A genomewide collection of SNPs among landraces and modern varieties is undoubtedly useful for further analysis of pedigrees or designing DNA markers within a short genomic region for map-based cloning of important genes or for MAS based breeding strategies. These advances could also facilitate our understanding of many complex phenomena, such as heterosis and epigenetics.

As it is, the genome sequence data deposited in public repositories is no doubt a gold mine for breeders being the source of invaluable information that can be used for elucidating allelic variations to facilitate genomics assisted breeding and accelerate the development of transformation technologies for different crops. However, the explosive growth in sequence data may also lead to uncontrollable outflow of information that could have hazardous implications to breeding similar to a tsunami that could wreak unimaginable havoc along its path. The challenge for breeders is to take advantage of current advances in genomics to develop novel breeding strategies and ultimately generate high yielding cultivars of the major crops that support the foundation of life without totally relegating traditional breeding.

Takuji Sasaki
Editor-in-chief
National Institute of Agrobiological Sciences