

# Soybean diversity map may provide important basis for breeding

In an international project jointly funded by NSFC, a research team composed of scientists from the Chinese University of Hong Kong, BGI-Shenzhen, the Chinese Ministry of Agriculture, the Chinese Academy of Agricultural Sciences and University of Copenhagen, reported a large-scale analysis of the patterns of genome-wide genetic variation in soybeans. The results, published online in *Nature Genetics* on 14th November 2010, provide a valuable resource for the genomic and genetic analysis of soybeans and facilitate future breeding and quantitative trait analysis.

According to a report by EurekAlert, using the next generation sequencing platform, the team re-sequenced the genomes of 17 wild and 14 cultivated soybeans and obtained 900 million 45bp or 76bp paired-end reads which were aligned to the soybean reference genome using SOAP2. From the analysis, the researchers identified a total of 6,318,109 single nucleotide polymorphisms (SNPs), which is the first genome-wide variation map of soybean. Using the program SOAPdenovo, they assembled the wild and cultivated soybeans and identified 186,177 present and absent variations (PAVs), including genes which might be lost in cultivated soybean during domestication. This research established the first comprehensive re-sequencing data of wild and cultivated soybean genomes as well of Fabaceae family members.

The researchers identified two unique features of the soybean genome that are distinct from other crop plants: they have exceptionally high linkage disequilibrium (LD) and a high ratio of average nonsynonymous versus synonymous nucleotide changes (Nonsyn/Syn), which indicates that marker-assisted breeding would be less challenging than map-based cloning for soybean improvement.

The researchers found higher genomic diversity in the wild soybeans than in the cultivated soybeans, indicating that human selection had a strong impact on the genetic diversity in the cultivated soybeans. The team also discovered an unexpected observation that wild soybeans have less low frequency SNPs as compared to cultivated soybean, which would be explained by the reduced population size of wild soybean as a result of the shrinking habitat. The team also reported linkage disequilibrium block location and distribution and identified a set of 205,614 tag SNPs that may be useful for QTL mapping and association studies.

"The study lays the foundation for future large scale population studies, marker-assisted breeding application and gene function identification in soybeans. It provides great benefits to the international soybean research community and breeding groups," said Dr Xun Xu, Project Investigator at BGI.

The project investigator at CUHK, Dr. Hon-Ming Lam said: "This research has generated a huge amount of genomic data to expedite future soybean researches. It also provides important information to facilitate soybean breeding program. It is the first time a large scale soybean genome project can be completed solely by Chinese scientists in the home of soybean, China. It is also an exemplar of deep collaborations between research institutes in Hong Kong and Mainland that will lead to significant breakthrough in science."