Hitchhiking mapping in soybean: Perspectives for more targeted future breeding

by Marina TOMIČIĆ*, Vuk ĐORĐEVIĆ, Jegor MILADINOVIĆ, Ljiljana BRBAKLIĆ, Sanja MIKIĆ and Dragana TRKULJA

Abstract: The identification of genomic regions that have retained a signature of selection during breeding of soybean (*Glycine max* (L.) Merrill) may indicate the positions of genes related to important agronomic traits or underlying adaptation to a specific environment. This can be achieved by analysing populations comprised of ancestral and elite varieties, applying multiple outlier detection tests and population-genetic principles. The hitchhiking mapping has some advantages over other approaches, as well as several limitations. The knowledge of the consequences of selection on the soybean genome in a specific breeding program could have practical importance for future breeding and yield improvement.

Key words: genetic hitchhiking, selection, soybean, yield

Soybean (*Glycine max* (L.) Merrill) presents highly important grain legume in the world, primarily because of the seed quality, rich in protein (40%) and oil (20%). The constant expanding of soybean acreage in the last 60 years was caused by its multiple uses, mainly as human food and animal feed, as well as source of edible oil and biofuel. In order to meet the growing needs for soybean, conventional breeding is performed through multiple cycles of selection, with the main objective to develop elite varieties, superior to their ancestors, adapted to specific growing conditions, with stable and high yield.

Institute of Field and Vegetable Crops, NoviSad, Serbia (marina.tomicic@ifvcns.ns.ac.rs)

Long-term selection for phenotypic traits during breeding has indirectly resulted insignificant loss of genetic diversity of agronomically important genes (10). The process by which a positive selection eliminates or reduces variation of the selectively important loci in the population is known as selective sweep (4). Except reduced variability in target genes, signatures of selection might include shift in allele frequency spectrum, increased linkage disequilibrium and increased levels of population differentiation (Fig. 1) (7). Due to the genetic linkage and consequently low recombination rates, selectively neutral loci that are linked to the targets of selection are also affected by a selective sweep and this phenomenon is marked as genetic hitchhiking (5). A population-genomic approach which utilizes a genome scan of neutral molecular markers to detect regions with signatures of selection forms the basis of the hitchhiking mapping (2, 8). The highest number of tests for the identification of loci favoured during selection compares levels of differentiation and levels of diversity between populations. The power of differentiation based tests does not decline over time as long as different alleles predominate in compared populations. On the contrary, the power of variability based tests is the highest immediately following selection, when the number of alleles is drastically reduced, and declines over time as population increases diversity. Methods for detecting selection can also be divided into model-based, that rely on assumptions about population structure and model-free, where candidate loci are identified as outliers in an empirical distribution of a summary statistic (9).

The application of the hitchhiking mapping started at the beginning of this century, aiming to determinate the genetic basis of local adaptation of natural populations. Soon after, a number of studies have been performed to identify genomic regions responsible for the crop domestication and improvement, by comparing wild relatives and modern varieties (1). This approach has been successfully conducted in a number of major crops including wheat (*Triticum aestivum* L.), maize (*Zea mays* L.), sunflower (*Helianthus annuus* L.), rice (*Oryza sativa* L.) and sorghum (*Sorghum bicolor* (L.) Moench). For soybean, there have been some recent studies on the effects of artificial selection, during domestication and improvement. A resequencing study of soybean indicated that 1,52% of the whole genome and 4-5% of the annotated genes were associated with significant signs of artificial selection by the modern breeding (3), confirming the hypothesis that selection affected a limited number of loci, influencing small reduction in diversity during modern soybean breeding. If a similar genome scan would be applied in breeding programmes to compare ancestral and elite soybean varieties, genomic regions that retained signatures of selection during soybean improvement in specific environments could be detected. Since the pedigrees of modern soybean varieties mainly include elite parents, it can be speculated that high and stable yield, is the only constant criteria during breeding. Considering that the plants with higher yield are better adapted to certain environments, the alleles and haplotypes that confer high yield over different growing conditions were favoured during the breeding. It is of great interest for soybean breeders to identify genomic regions that exhibit signatures of selection through the development of elite varieties, as they may represent important regions for agronomic improvement or crop adaptation, and may have implications in more targeted subsequent breeding (6).
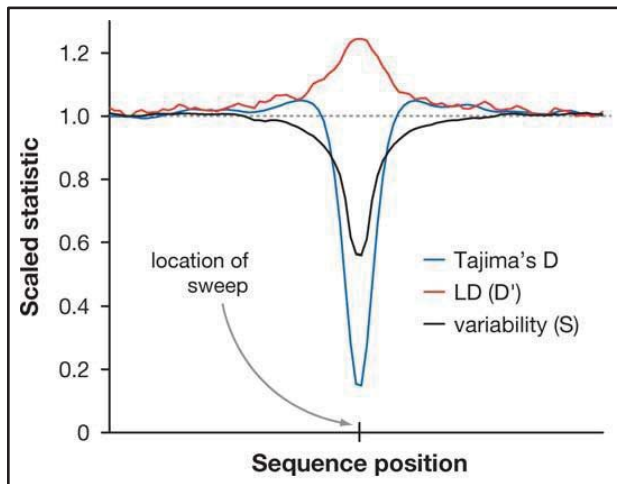
Figure 1. The effect of a selective sweep on genetic variability, linkage disequilibrium (LD) and the frequency spectrum (Tajima's D) in the surrounding region (7)

References

(1) Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127:1309-1321
(2) Harr B, Kauer M, Schlötterer C (2002) Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. Proc Natl Acad Sci U S A 99:12949-12954
(3) 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 (2013) Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. BMC Genomics 14:579-591
(4) Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4:981-994
(5) Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23:23-35
(6) Morrell PL, Buckler ES, Ross-Ibarra J (2011) Crop genomics: Advances and applications. Nat Rev Genet 13:85-96
(7) Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genet 39:197-218
(8) Schlötterer C (2003) Hitchhiking mapping-functional genomics from the population genetics perspective. Trends Genet 19:32-38
(9) Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol 14:671-688
(10) Yamasaki M, Wright SI, McMullen MD (2007) Genomic screening for artificial selection during domestication and improvement in maize. Ann Bot 100:967-973

The advantage of hitchhiking mapping within breeding programmes is that it does not require creating neither mapping populations nor collecting replicated phenotypic data. By using molecular markers alone, it can take advantage of yield progress that occurred during the entire period of crop breeding, enabling the isolation of loci of more practical importance for developing varieties in target environment. In addition, there are also several challenges for successful application of this approach. Since the demographic events, such as genetic drift and population bottleneck, cause the reduction of genetic diversity, it can be difficult to separate the signatures of selection from those resulting from demographic history of a population. Further testing is necessary to be conducted, having in mind that selection acts in a locus-specific manner and demographic events have genome wide effects (10). Besides that, considering that hitchhiking mapping faces the problem of identifying false positive results, it is important to apply multiple tests that assume different concepts and hypothesis to raise the candidate status of the identified loci (9). Moreover, different approaches could identify additional regions that may be overlooked by a particular test. However, the strongest validation of positive selection will with no doubt come from functional characterization. Integration of population genomics with available genomic resources would provide information on potential functions of genes under selection. The identified QTLs will likely be in accordance with the criteria used during selection, playing a role in the process of soybean improvement. If these loci are identified in different environments, they probably influence general productivity and could be regarded as universal. Loci identified in single environment could be considered as environment-specific, being influenced by different selective pressures within local conditions and, as a consequence, strong genotype by environment interactions.

In soybean, yield and most agronomic traits, including resistance to abiotic and biotic stress, protein concentration and oil content, are quantitative traits controlled by multiple genes, with low heritability and strong genotype by environment interaction. This greatly affects the identification of genetically superior plants based on phenotype and can be considered as one of the major hindrance during conventional breeding. The specific alleles of the identified selectively important loci may serve in creating favourable haplotypes for a target environment and possibly be used in marker-assisted breeding. Thus, the detection of selectively important genes, as well as the identification of associated molecular markers, presents the prerequisite for successful marker assisted selection that would greatly improve efficiency of breeding.