



Biotechnology and bioeconomy of complex traits in crop plants

JAN ANTONI RAFALSKI

Rafalski Consulting, Greenville, DE, United States

Abstract

Most important crop productivity traits, such as yield under normal and environmental stress conditions, are determined by a large number of genes, each with a small phenotypic effect. Genetic improvement of these traits through breeding or genetic engineering has been frustrating researchers in academia and industry. The reasons for this include the complexity of the traits, the difficulty of precise phenotyping and the lack of validated candidate genes. Different approaches to the discovery of the genetic architecture of such traits, such as Genetic Association Mapping and Genomic Selection and their engineering, are expected to yield benefits for farmers and consumers.

Key words: plants, GMO, plant breeding, complex traits

The discovery of a practical *Agrobacterium*-based plant genetic transformation system in the early 1980s (Matzke and Chilton, 1981) quickly led to the development of transgenic crops. The first practical and commercially viable transgenic trait was resistance to glyphosate, an active component of Roundup® herbicide. This was based on the introduction of a bacterial variant of 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase, which, unlike its plant homolog, is resistant to glyphosate (Shah et al., 1986). Next came insect pest-resistance mediated by Bt-toxin genes derived from *Bacillus thuringiensis* (Barton et al., 1987). In both cases, a single gene of bacterial origin was sufficient to enable robust resistance. Transgenic crops, including corn, soybean, and cotton, that carry these traits are currently grown on millions of acres, especially in North and South America (Qaim, 2016). Resistance or tolerance to certain diseases has also been successfully engineered and deployed in the field, for example in papaya (Gonsalves, 2014). These traits owe their popularity to their benefits for the farmer, such as labor savings, and reductions in environmental contamination, especially with insecticides (Gruissem, 2015).

Other genetically simple traits are also relatively straightforward to engineer. Already in the 1980s plant biotechnologists realized that some quality traits and

nutritional properties can be improved, in some cases by introducing single genes, especially if the biochemical pathways involved were relatively well understood. Examples include the manipulation of carotenoid biosynthesis pathways to improve vitamin A content – the well-publicized case of “golden rice” (Potrykus, 2012), or amino acid and fatty acid biosynthesis pathways (Napier et al., 2014). Even though the technical successes were many, for example in producing a healthier fatty acid profile for plant oils or increasing the content of essential amino acids in foods and feeds (Yang et al., 2016), progress toward commercialization has been slow. To some extent, this is the result of the difficult regulatory regime, resulting in regulatory costs in the range of millions of dollars (<http://www.agbioworld.org/biotech-info/articles/biotech-art/crushingcost.html>), making commercialization of products for relatively small market segments economically unfeasible. This has been exacerbated by customer reluctance to accept GMO technology. As a result, many potential products with clear nutritional and health benefits have never been introduced. Thus, only “blockbuster” products with appeal across product lines, such as herbicide and insect tolerance, are broadly available. Also, because of the high costs of research and commercialization, only those companies

with the strongest finances have been able to successfully introduce and market GMO products, and even then only the most profitable ones.

Agricultural production is severely limited both by abiotic environmental stresses such as drought and flooding resistance which affect productivity (yield) and by nutrient utilization efficiency (Mickelbart et al., 2015; Moshelion and Altman, 2016). The importance of such traits is likely to increase with increasing human populations and resulting competition for resources, including cultivated land, as global temperatures increase (Pilbeam, 2015). Both academic and industrial scientists are strongly motivated to identify genes for these quantitative traits. Efficient nutrient extraction and utilization is also of great importance, as it would reduce the amount of both fertilizer needed and ground water contamination from fertilizer run-off (Baligar and Fageria, 2015).

Great efforts have been made to identify genes that, when introduced into crop plants, would improve these traits. However, due to the genetic complexity of traits, this task was much more challenging than expected. Some of the approaches have relied on forward genetics: identifying genetic loci, and ultimately the genes themselves, by genetic mapping in crosses between individuals differing dramatically in terms of the trait of interest (e.g., drought-tolerant plants crossed with drought-sensitive plants; Jankowicz-Cieslak and Till, 2015). The efforts to map quantitative trait loci ("QTL mapping") have led to the understanding that continuously variable traits such as yield and drought tolerance depend on the action and interactions of multiple genes. Each of these genes contributes relatively little to the phenotype ("minor genes"; Alonso-Blanco and Méndez-Vigo, 2014). This result was predicted by classical quantitative genetics, although biotechnologists were initially very optimistic that a solution would be found. Some suspected that the lack of success in identifying major genes for quantitative traits was due to methodological difficulties, so a reverse genetics approach became popular. This relied on the identification of likely candidates from among the known genes, based on prior biological insights, and their over-expression or elimination in plants, followed by careful determination of the phenotype of interest (Sessions et al., 2002; McCarty and Meeley, 2009). A more extreme approach relied on activation tagging. In this case, a promoter, perhaps an inducible promoter or an enhancer and a color marker gene, are introduced randomly

into the genome by genetic transformation (Weigel et al., 2000). In this approach, tens of thousands of transgenic plants are recovered, each with a promoter in a different genomic location, frequently next to a gene, the expression of which increases. After a massive phenotyping effort, rare plants showing improvement of the targeted traits were identified and the responsible gene cloned based on its proximity to the introduced promoter. This method is practical in model species, such as *Arabidopsis*, and any genes of interest have to be introduced into relevant crop species and re-tested. Several large agricultural corporations, newly formed companies and academic scientists initiated searches for the next generation of GMO products, using the methods discussed above. Major challenges included the difficulty of field phenotyping, due to vagaries of weather, soil non-uniformity within fields, the need for multiple replications and geographical locations needed to obtain sufficient confidence in the results. Many candidate genes for productivity traits were thus identified; yet few if any, when tested in crop plants, exhibited commercial-level improvements.

Classical plant breeding has led to the improvement of yield by 1-3% per year (Duvick, 1984). However, before economic viability can be achieved, the long development cycle and search for successful answers to the challenges of the commercialization of GMOs, as well as addressing the issues of the high cost and time required for regulatory acceptance (Vigani and Olper, 2015) must be considered. An improvement close to 5-10% is required for economic viability. Despite the more than one billion dollars spent on gene discovery, and the further billions spent on the development of transgenic products, results have been disappointing. Early expectations were that at least a few genes with major trait-specific effects would be found, even for complex traits. However, in most cases this has not happened. Although a corn hybrid with improved drought tolerance has been commercialized by Monsanto (Sammons et al., 2014), its performance remains to be fully evaluated, and competitors have released similar non-GMO products, bred by traditional breeding enhanced by genomic technologies (Cooper et al., 2014).

Let us now examine briefly the technological enhancements to plant breeding for these traits. Classically, breeding consists of making genetic crosses between parents differing in desirable traits. The advanced pro-

geny of such crosses (e.g., F2, F3 families) are then field-tested and evaluated for multiple traits, selecting superior individuals for re-testing of their progeny in subsequent years. Ultimately, after several years of testing in multiple environments, a few lines superior to current commercial reference lines may be registered and commercialized. Predictive genetic tests could improve the efficiency of this process (Lübberstedt, 2013). Early on, a gel analysis of enzyme isoforms was used (Karaca 2013). Only with the development of molecular markers, initially restriction fragment length polymorphisms (RFLPs) and later polymerase chain reaction (PCR) – based markers such as simple sequence repeats (SSR) and currently single nucleotide polymorphisms (SNPs), has this approach become practicable and spread through the whole plant-breeding community (Phillips and Vasil, 2013). DNA-based genetic marker assisted breeding has been facilitated by rapid progress in genetic mapping and, in many cases, the cloning of genes for monogenic traits, such as race-specific disease resistance genes or genes coding for enzymes in certain biochemical pathways. However, the application of marker-assisted breeding to complex traits has proved to be challenging. On the one hand, many genes are involved, each usually contributing a modest fraction of the observed genetic variation. On the other hand, different sets of genes are relevant in different populations. Various schemes of “pyramiding” multiple genes into a single individual have been developed (for example, see (Tyagi et al., 2014)). However, in most cases the progress has been slow.

With the advent of genome sequencing, millions more genetic markers have become available, leading to the development of several new methodologies. Genome-wide association mapping (GWAS) was first applied to human populations, in which genotypes at millions of SNP markers became available (Visscher et al., 2012). In GWAS, at each of these SNP loci, the distribution of phenotypes, such as the disease status, is compared for each of the two SNP alleles. Simply put, the loci at which individuals carrying one of the two alleles show a much higher occurrence of a disease than those carrying the other allele are declared candidates. This approach allows for the identification of hundreds of candidate loci for complex, quantitative phenotypes (Wallace et al., 2014). The quality of the results depends on the accuracy of the phenotype data, the frequency of alleles, and

the artifacts associated with the population structure (Korte and Farlow, 2013). Plant populations offer some advantages for GWAS mapping. Specifically, population structures can be more easily controlled and phenotyping can be conducted in controlled environments. Thus, GWAS has become widely used, although in some cases the populations are too small or structured, the marker numbers limited, and field phenotyping presents challenges (Rafalski, 2010). Populations designed especially for GWAS have been developed and made broadly available (McMullen et al., 2009). Special software tools are also available (Bradbury et al., 2007).

Availability of very high-density genotypes has led to the development of a new approach to breeding for complex quantitative traits: genomic assisted selection. This was first successfully applied to animal breeding in defined crosses. Success was facilitated by rearing animals in well-controlled conditions and studying precisely measured phenotypes, such as the fat content of meat (Hayes et al., 2009). In this approach, progeny of a cross are genotyped and phenotyped and a computer algorithm is used to identify the best combination of loci and alleles to predict the phenotype. The prediction is used in the next selection cycle. It is possible to skip the laborious and time-consuming phenotyping step for some generations, allowing major savings in time and money. In some cases, breeders have identified some regions in the genome based on longer term studies and have used them as background selection in early stages of testing. For example, DH lines are tested with a small set of markers and some outliers are discarded to allow the testing of more lines in the same season, thus improving the efficiency of the breeding programs. The results of using genomic-assisted selection in plant breeding are still being evaluated (Heffner et al., 2009), but some successful commercially introduced varieties are reportedly being produced using a similar approach (Cooper et al., 2014). Given the variability inherent in field testing across multiple environments and seasons, the accuracy of phenotyping is the most important determinant, and is harder to control in crops relative to animal breeding. In some crop species, this has been facilitated by the development of double haploid technology. Instead of phenotyping F3 families produced by selfing F2 progeny from a cross, each F2 individual is used to produce a genotypically and phenotypically uniform set of double haploid plants, which allow more accurate phenotyping

than segregating populations. This methodology is now widely used in corn (Strigens et al., 2013), and is being developed in other species.

While the above approaches do not require explicit knowledge of the genes behind complex traits, rapid development of a range of high throughput methods of genotyping, expression analysis, and mapping (eQTL), in combination with computational tools, and especially new very specific gene disruption and modification techniques, for example (Feng et al., 2016), provide means of rapid candidate gene identification and validation. These methods, while not a panacea, are expected to advance the description of the architecture of quantitative traits at the gene and haplotype level.

References

- Alonso-Blanco C., Belén M.-V. (2014) *Genetic architecture of naturally occurring quantitative traits in plants: an updated synthesis*. *Curr. Opin. Plant Biol.* 18: 37-43.
- Baligar V.C., Fageria N.K. (2015) *Nutrient use efficiency in plants: an overview*. [in:] *Nutrient use efficiency: from basics to advances*. Ed. Rakshit A., Singh B.H., Sen A. New Delhi, India. Springer: 1-14. doi: 10.1007/978-81-322-2169-2_1.
- Barton K.A., Whiteley H.R., Yang N.S. (1987) *Bacillus-thuringiensis delta-endotoxin expressed in transgenic Nicotiana tabacum provides resistance to lepidopteran insects*. *Plant Physiol.* 85(4): 1103-1109. doi: 10.1104/pp.85.4.1103.
- Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. (2007) *TASSEL: Software for association mapping of complex traits in diverse samples*. *Bioinformatics* 23(19): 2633-2635. doi: 10.1093/bioinformatics/btm308.
- Cooper M., Gho C., Leafgren R., Tang T., Messina C. (2014) *Breeding drought-tolerant maize hybrids for the US cornbelt: discovery to product*. *J. Exp. Bot.* 65(21): 6191-6204. doi: 10.1093/jxb/eru064.
- Duvick D.N. (1984) *Genetic contributions to yield gains of U.S. hybrid maize, 1930 to 1980*. [in:] *Genetic contributions to yield gains of five major crop plants*, 15-47. CSSA Special Publication SV - 7. Madison, WI: Crop Science Society of America and American Society of Agronomy. doi:10.2135/cssaspecpub7.c2.
- Feng C., Yuan J., Wang R., Liu Y., Birchler J.A., Han F. (2016) *Efficient targeted genome modification in maize using CRISPR/Cas9 system*. *J. Genet. Gen.* 43(1): 37-43. doi: http://dx.doi.org/10.1016/j.jgg.2015.10.002.
- Gonsalves D. (2014) *Hawaii's transgenic papaya story 1978-2012: a personal account*. [in:] *Genetics and Genomics of Papaya*. Ed. Ming R. and Moore P.H., 115-142. New York, NY: Springer New York. doi: 10.1007/978-1-4614-8087-7_7.
- Gruissem W. (2015) *Genetically modified crops: the truth unveiled*. *Agricult. Food Secur.* 4(1): 1-2. doi: 10.1186/s40066-015-0022-8.
- Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. (2009) *Invited review: genomic selection in dairy cattle: progress and challenges*. *J. Dairy Sci.* 92(2): 433-443. doi: http://dx.doi.org/10.3168/jds.2008-1646.
- Heffner E.L., Sorrells M.E., Jannink J.-L. (2009) *Genomic selection for crop improvement*. *Crop Sci.* 49: 1-12. doi: 10.2135/cropsci2008.08.0512.
- Jankowicz-Cieslak J., Till B.J. (2015) *Forward and reverse genetics in crop breeding*. [in:] *Advances in plant breeding strategies: breeding, biotechnology and molecular tools*. Ed. Al-Khayri M., Jameel J.M.S., Johnson V.D. Springer International Publishing. Cham: 215-240. doi: 10.1007/978-3-319-22521-0_8.
- Karaca M. (2013) *Isozymes as biochemical markers in plant genetics*. *Inter. J. AgriSci.* 3(11): 851-861.
- Korte A., Farlow A. (2013) *The advantages and limitations of trait analysis with GWAS: a review*. *Plant Meth.* 9(1): 1-9. doi: 10.1186/1746-4811-9-29.
- Lübberstedt T. (2013) *Diagnostics in plant breeding*. [in:] *Diagnostics in plant breeding*. Ed. Thomas Lübberstedt T., Rajeev Varshney K., Springer Netherlands. Dordrecht: 3-9. doi: 10.1007/978-94-007-5687-8_1.
- Matzke A.J., Chilton M.D. (1981) *Site-specific insertion of genes into T-DNA of the agrobacterium tumor-inducing plasmid: an approach to genetic engineering of higher plant cells*. *J. Mol. Appl. Genet.* 1(1): 39-49. http://europepmc.org/abstract/MED/6955419.
- McCarty D.R., Meeley R.B. (2009) *Transposon resources for forward and reverse genetics in maize*. [in:] *Handbook of maize: genetics and genomics*. Ed. Bennetzen J.L., Hake S. Springer New York. New York, NY: 561-584. doi: 10.1007/978-0-387-77863-1_28.
- McMullen M.D., Kresovich S., Villeda H.S., Bradbury P., Li H., Sun Q., Flint-Garcia S. et al. (2009) *Genetic properties of the maize nested association mapping population*. *Science* 325(5941): 737-740. doi: 10.1126/science.1174320.
- Mickelbart M.V., Hasegawa P.M., Bailey-Serres J. (2015) *Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability*. *Nature Rev. Genet.* 16(4): 237-251.
- Moshelion M., Altman A. (2016) *Current challenges and future perspectives of plant and agricultural biotechnology*. *Trends Biotech.* 33(6): 337-342. doi: 10.1016/j.tibtech.2015.03.001.
- Napier J.A., Haslam R.P., Beaudoin F., Cahoon E.B. (2014) *Understanding and manipulating plant lipid composition: metabolic engineering leads the way*. *Curr. Opin. Plant Biol.* 19: 68-75. doi: http://dx.doi.org/10.1016/j.pbi.2014.04.001.
- Phillips R.L., Vasil I.K. (2013) *DNA-Based markers in plants*. Vol. 6. Springer Science & Business Media.
- Pilbeam D.J. (2015) *Breeding crops for improved mineral nutrition under climate change conditions*. *J. Exp. Bot.* 66(12): 3511-3521. doi: 10.1093/jxb/eru539.

- Potrykus I. (2012) "Golden rice", a GMO-product for public good, and the consequences of GE-regulation. *J. Plant Biochem. Biotech.* 21(1): 68-75. doi: 10.1007/s13562-012-0130-5.
- Qaim M. (2016) *Adoption and impacts of GM crops*. [in:] *Genetically modified crops and agricultural development*. Palgrave Macmillan US. New York: 57-84. doi: 10.1057/9781137405722_4.
- Rafalski J.A. (2010) *Association genetics in crop improvement*. *Curr. Opin. Plant Biol.* 13(2): 174-180. doi: 10.1016/j.pbi.2009.12.004.
- Sammons* B., Whitsel J., Stork L.G., Reeves W., Horak M. (2014) *Characterization of drought-tolerant maize MON 87460 for use in environmental risk assessment*. *Crop Sci.* 54: 719-729. doi: 10.2135/cropsci2013.07.0452.
- Shah D.M., Horsch R.B., Klee H.J., Kishore G.M., Winter J.A., Tumer N.E., Hironaka C.M. et al. (1986) *Engineering herbicide tolerance in transgenic plants*. *Science* 233(4762): 478-481. <http://www.jstor.org/stable/1697820>.
- Strigens A., Schipprack W., Reif J.C., Melchinger A.E. (2013) *Unlocking the genetic diversity of maize landraces with doubled haploids opens new avenues for breeding*. *PLoS ONE* 8(2): e57234. <http://dx.doi.org/10.1371/journal.pone.0057234>.
- Tyagi S., Mir R.R., Kaur H., Chhuneja P., Ramesh B., Balyan H.S., Gupta P.K. (2014) *Marker-assisted pyramiding of eight QTLs/genes for seven different traits in common wheat (*Triticum aestivum* L.)*. *Mol. Breed.* 34(1): 167-175. doi: 10.1007/s11032-014-0027-1.
- Vigani M., Olper A. (2015) *Patterns and determinants of GMO regulations: an overview of recent evidence*. *AgBioForum* 18(1): art. 6.
- Visscher P.M., Brown M.A., McCarthy M.I., Yang J. (2012) *Five years of GWAS discovery*. *Amer. J. Human Genet.* 90(1): 7-24. doi: <http://dx.doi.org/10.1016/j.ajhg.2011.11.029>.
- Wallace J.G., Bradbury P.J., Zhang N., Gibon Y., Stitt M., Buckler E.S. (2014) *Association mapping across numerous traits reveals patterns of functional variation in maize*. *PLoS Genet.* 10(12): e1004845. <http://dx.doi.org/10.1371/journal.pgen.1004845>.
- Weigel D., Ahn J.H., Blázquez M.A., Borevitz J.O., Christensen S.K., Fankhauser C., Ferrándiz C. et al. (2000) *Activation tagging in Arabidopsis*. *Plant Physiol.* 122(4): 1003-1014. doi: 10.1104/pp.122.4.1003.
- Yang Q.-Q., Zhang C.-Q., Chan M.-L., Zhao D.-S., Chen J.-Z., Wang Q., Li Q.-F. et al. (2016) *Biofortification of rice with the essential amino acid lysine: molecular characterization, nutritional evaluation, and field performance*. *J. Exp. Bot.* 67(14): 4285-4296. doi:10.1093/jxb/erw209.