Opportunities and Regulatory Challenges for Genome Engineering in Agriculture

Daniel F. Voytas
University of Minnesota
Minneapolis, Minnesota

Voytas@umn.edu

Plant agriculture faces numerous challenges in the coming decades. The burgeoning world population demands that more food be produced on less, and often increasingly marginal, land. Climate change and diminishing resources, such as water and fertilizer, will make it difficult to achieve the needed increase in productivity. Whereas multiple strategies must be deployed to achieve food security, it is clear that amongst these is the need to accelerate the rate of crop improvement. Recent advances in genome engineering promise to make this possible. From targeted mutagenesis to targeted gene insertion, genome engineering is transforming plant science, making it possible to create genetic diversity with precision, efficiency and control. For the basic plant biologist, genome engineering helps dissect gene function by linking genotype to phenotype. Information garnered about plant-gene function can then be harnessed to create genetic variation relevant to agriculture to achieve increased productivity.

Genome Engineering

Genome engineering, as typically practiced, uses sequence-specific nucleases that recognize unique sites in the plant genome and introduce targeted DNA double-strand breaks (DSBs) (Voytas, 2013). The repair of the DSB can be controlled to achieve the desired DNA-sequence modification at or near the break site (Figure 1). One repair pathway that cells use is non-homologous end-joining (NHEJ), wherein the broken chromosome is simply rejoined (Gorbunova and Levy, 1997; Salomon and Puchta, 1998). Oftentimes, a few to several nucleotides are gained or lost at the break site, creating a targeted mutation. If the mutation occurs in a coding region, it can alter a protein's amino acid sequence or cause a frameshift mutation that destroys or knocks out gene function. A second DNA-repair pathway is homologous recombination (HR) (Puchta *et al.*, 1993, 1996). Through HR, the broken chromosome uses a homologous template to copy information to the break site. The template is most often a sister chromatid or homologous chromosome;

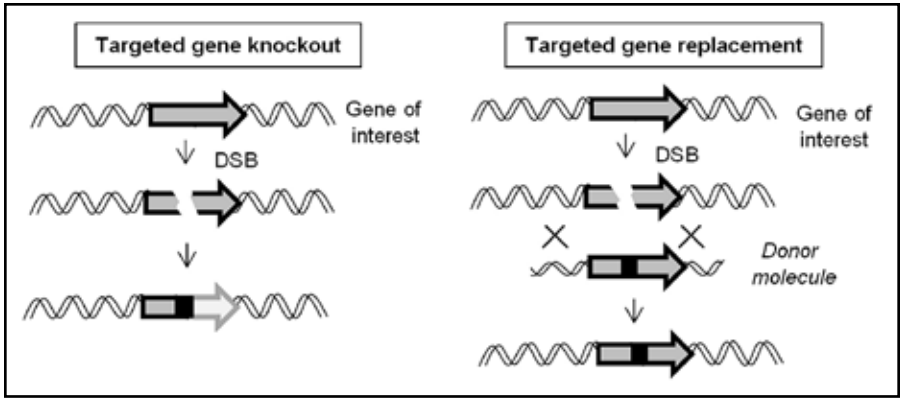


Figure 1. Plant-genome engineering using DNA double-strand breaks (DSBs). DNA-repair pathways can be exploited to introduce desired sequence changes to a plant's genome. Repair of DSBs by non-homologous end joining can result in deletions or insertions at the break-site. Therefore, targeting DNA breaks to a locus or gene of interest can achieve targeted mutagenesis (left panel). Alternatively, DSBs can stimulate homologous recombination with a user-supplied donor molecule. Donor molecules can be designed to contain small point mutations for the purpose of making small changes within genes (targeted gene replacement, illustrated in the right panel) or larger changes, including full genes or gene-regulatory elements (targeted gene insertion). (Courtesy of Nick Baltes.)

however, exogenous templates can be delivered to a cell, and, because they are user-specified, the templates can have specific DNA-sequence alterations that become incorporated at the break site. HR is, therefore, a powerful means to achieve precise alterations to the plant genetic code.

Whether created through NHEJ or HR, the key to achieving a targeted DNA-sequence modification is the DNA DSB. Much effort in the past 15 years has focused on creating reagents capable of recognizing specific DNA sequences in complex genomes to introduce targeted breaks at high efficiency. Four classes of sequence-specific nucleases have been widely deployed (Figure 2). One class is the meganucleases or homing endonucleases—enzymes that naturally recognize and cleave large DNA-sequence signatures (typically >30 bp). The DNA specificity of meganucleases can be altered such that they recognize and cleave novel DNA targets (Smith *et al.*, 2006; Pâques and Duchateau, 2007). Two classes of sequence-specific nucleases use engineered DNA-binding domains fused to the catalytic domain of the type II restriction endonuclease, FokI. These are the zinc finger nucleases (ZFNs) and the transcription activator-like effector nucleases (TALENs). For ZFNs, DNA targeting is achieved by custom arrays of zinc fingers, each of which typically recognizes three base pairs (Bibikova *et al.*, 2003; Carroll, 2011); for the TALENs, custom arrays of TAL effector repeats are assembled, with each repeat recognizing one base pair (Christian *et al.*, 2010; Bogdanove and Voytas, 2011). Both ZFNs and TALENs function as dimers: two DNA-binding domains are engineered to bring the FokI monomers into

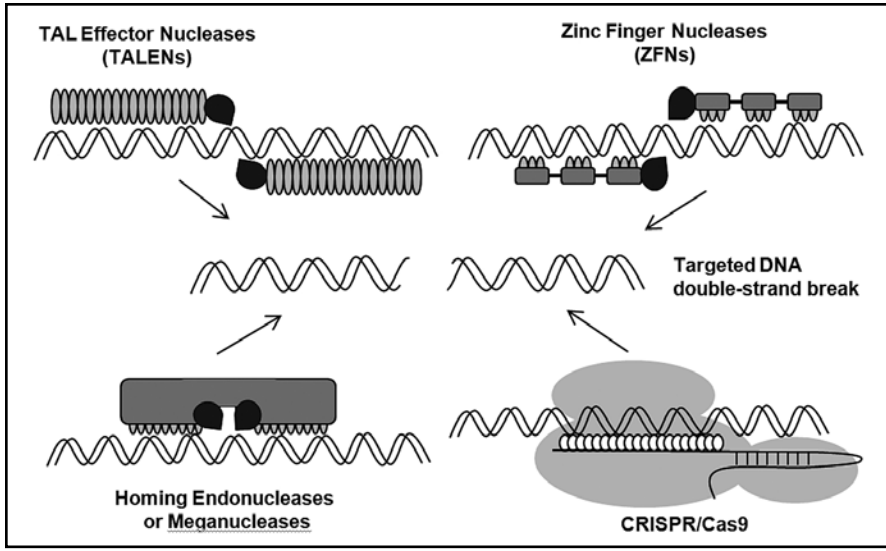


Figure 2. Illustration of the four classes of sequence-specific nucleases: Transcription activator-like (TAL) effector nucleases, zinc-finger nucleases, homing endonucleases or meganucleases and CRISPR/Cas9. All classes of proteins can be "reprogrammed" to recognize and cleave desired DNA sequences. (Courtesy of Nick Baltes.)

proximity on the DNA target. There they dimerize and create the DSB. More recently, CRISPR/Cas9 reagents have emerged as powerful and highly efficient tools for making targeted DSBs (Jinek *et al.*, 2012; Doudna and Charpentier, 2014). For CRISPR/Cas9, targeting is achieved when a guide RNA base pairs with its DNA target. The Cas9 nuclease then introduces the DSB. Because no protein engineering is required and targeting is achieved simply through base pairing, CRISPR/Cas9 has emerged as the reagent of choice for making targeted chromosome breaks.

Plant Varieties Created Through Targeted Mutagenesis

One of the simplest means of deploying sequence-specific nucleases is to create mutations through imprecise NHEJ (Voytas, 2013). For targeted mutagenesis, the nuclease is typically delivered to the cell as DNA, either transiently or by stably integrating the nuclease-encoding construct into the genome. If imprecise repair occurs after the break is created, a mutation results. Targeted mutagenesis is particularly valuable for studying gene function. Loss-of-function mutations and their consequential phenotypes are achieved by introducing frameshift mutations near the 5'-end of the gene. Traits of relevance to agriculture can be created through targeted mutagenesis, although the phenotypic variation afforded by loss of gene function is somewhat limited. That said, removing toxins, such as ricin from castor oil, or anti-nutritionals, such as trypsin inhibitors from soybean, are potential traits of value. Similarly, antigenic determinants that cause allergic reactions could be removed from nut or grain proteins.

A recently published example of a trait created through targeted mutagenesis is a soybean variety that produces oil with elevated levels of the monounsaturated fat, oleic acid (Haun *et al.*, 2014). Soybean oil typically has about 20% monounsaturated fats, and, in the past, polyunsaturated fats have been reduced through hydrogenation to improve the oil's storage and frying characteristics (Clemente and Cahoon, 2009). Hydrogenation, however, produces trans-fatty acids, which are unhealthy when consumed. Consequently, there has been a strong push to create soybean varieties that produce oil with elevated levels of monounsaturated fats.

In soybean seeds, the monounsaturated fat, oleic acid, is converted to the polyunsaturated fat, linoleic acid, through the action of fatty acid desaturases (Tang *et al.*, 2005). Soybean has two seed-specific fatty acid desaturase genes, designated *FAD2-1A* and *FAD2-1B* (Schlueter *et al.*, 2007). To mutate these genes, and test whether levels of oleic acid could be increased relative to linoleic acid, TALENs were created that recognize conserved DNA sequences in both *FAD2-1A* and *FAD2-1B* (Haun *et al.*, 2014). Constructs encoding these TALENs were stably transformed into soybean cells and expressed constitutively. As the transformed cells regenerated into soybean plants, the TALENs created mutations, including mutations in cells that gave rise to the germline. Among nineteen transgenic soybean lines that were regenerated, three transmitted to progeny mutations in one or more *FAD2* genes. Consequently, it was possible to recover plants from this population that were homozygous for mutations in *FAD2-1A*, *FAD2-1B* or both genes. In the case of the homozygous double mutant, the desired phenotype was achieved. Oil pressed from this plant's seeds had 80% of the monounsaturated fat, oleic acid, and less than 4% of the polyunsaturated fat, linoleic acid. In contrast, oil from wild-type plants had 20% oleic acid and 50% linoleic acid. This single loss-of-function mutation, therefore, created soybean lines that produce oil with a fatty acid composition that is healthier for human consumption.

In the above example, the construct encoding the TALEN was stably introduced into the soybean genome. The mutations that were created, however, were at *FAD2* genes located at other genomic sites. Consequently, it was possible to segregate away the TALEN transgene and obtain lines of soybean with mutations only in the *FAD2* gene targets (Haun *et al.*, 2014). In contrast to this example, which involves a stable, transgenic intermediate, sequence-specific nucleases can also be introduced into plant cells transiently (Townsend *et al.*, 2009). The nucleases encoded by DNA that enters the plant cell are expressed, and oftentimes the DNA never integrates into the plant genome. This transient expression of the nuclease creates targeted mutations, and transgenic plants are not intermediates in the mutagenesis protocol.

The real advantage of mutagenesis with sequence-specific nucleases is their precision. Traditional methods of mutagenesis that use chemicals, X- or gamma-rays, transposons or T-DNA provide virtually no control over where in the genome mutations are created. Consequently, large populations of mutagenized plants need to be generated and screened to identify those rare individuals with alterations in a particular gene of interest. Oftentimes, mutations that are recovered are not ideal, and perhaps, for example, result in only partial loss of gene function. Sequence-specific nucleases can be used to efficiently create

multiple mutant alleles for study, including complete gene knockouts. It is important to note that all classes of sequence-specific nucleases, when engineered properly, are highly precise (Pauwels *et al.*, 2014). They rarely cleave at unintended or off-target sites and, thus, typically only create mutations at the intended sites. This contrasts with traditional mutagens, which can cause considerable collateral damage to the genome.

Regulatory Aspects

Regulation is one factor that will determine how broadly and rapidly the products of genome engineering will be deployed in agriculture (Voytas and Gao, 2014). In the United States, plants that have genetic variation created using chemical mutagens or ionizing radiation are not regulated and can be planted directly in the field to test the phenotypic consequence of the induced genetic variation. In contrast, transgenic plants are subjected to exhaustive and costly regulatory scrutiny before they can be planted in the field (Lusser *et al.*, 2012). In many ways, this regulatory burden has restricted the use of transgenic approaches to create genetic diversity to a handful of high-margin row crops. As described above, targeted mutagenesis with sequence-specific nucleases is highly precise, and since the mutant plants often lack foreign DNA, this suggests that they might be treated more like traditional mutants in terms of regulation. In the United States, this appears to be the case. Recent opinion letters from the USDA indicate that plants with targeted mutations made by NHEJ and without transgenes fall outside their regulatory authority (Waltz, 2012; Jones, 2015). Opinion letters were rendered in two cases—for a low phytate line of corn made with ZFNs and a potato variety with improved storage and frying characteristics created by a TALEN-induced mutation. If this trend continues, then many new plant varieties could be made and commercialized without having to accumulate large, costly data packages for regulatory approval. This will likely extend the range of species for which biotechnology is used to create genetic variation of value, and horticultural and vegetable crop varieties will likely enter the marketplace with genomes altered using this technology.

Plant Varieties Created Through Homologous Recombination or Gene Targeting

The repair of DNA breaks by HR, referred to as gene targeting, allows a vast spectrum of DNA-sequence modifications to be introduced into a plant's genome (Voytas, 2013). These modifications can range from single-nucleotide substitutions that alter an amino acid in a coding sequence to the insertion of arrays of transgenes at defined chromosomal sites. The high level of control afforded by DNA repair through HR makes it possible to create plant varieties with complex traits, such as tolerance to biotic or abiotic stress or that more efficiently use inputs such as fertilizer and water. Gene targeting could be used to alter primary metabolism to create varieties that, for example, produce specialty carbohydrates or oils for industrial purposes or for fuel. Plants also produce a remarkable array of complex secondary metabolites, and genome engineering could create plant varieties that overproduce chemicals of pharmaceutical or industrial value. In many cases, achieving such complex traits will require the modification of multiple genes in a pathway.

To illustrate how gene targeting can be used to create a new plant variety, consider an approach to increase disease resistance, specifically resistance of potato to late blight. Late blight is caused by the fungal pathogen *Phytophthora infestans*, and is one of the world's most devastating crop diseases (Kamoun, 2001). If late blight were controlled effectively, potato yields could increase by as much as 50%. Existing methods for combating late blight involve multiple applications of fungicides to potato fields throughout the growing season, which is costly and can have a negative impact on the environment. Resistance can also be achieved genetically. In related species of potato, genes have been identified that confer late-blight resistance (Song *et al.*, 2003; Foster *et al.*, 2009). Traditionally, these genes would be introduced into cultivars of potato through breeding regimes that would take many years to complete. Alternatively, resistance genes could be introduced into the potato genome as transgenes to create resistant, transgenic varieties. This latter approach, however, may be undesirable from a regulatory point of view, in that the resultant resistant varieties carry foreign DNA.

With gene targeting, late-blight resistance in potato can be achieved in a much shorter timeframe than with traditional breeding and with only subtle alterations to the genome. For the potato resistance genes, orthologues that confer susceptibility exist, and they are highly similar to the resistant variant (Song *et al.*, 2003; Foster *et al.*, 2009). Only a handful of base changes distinguish the resistant and susceptible alleles; the DNA-sequence differences confer the ability of the encoded resistant protein to recognize or respond to the pathogen. To confer late-blight resistance through genome engineering, a sequence-specific nuclease, such as a TALEN, ZFN or CRISPR/Cas9 reagent, would be engineered to recognize and cleave the susceptible allele (Figure 3). A construct encoding the nuclease would be introduced into potato cells along with a repair template, which, through HR, would introduce into the susceptible allele the desired DNA-sequence variation from the resistance gene. Potato cells with the desired DNA-sequence modification would then be regenerated into plants, and they should be resistant to late blight. Resistance achieved through genome engineering could be accomplished in as little as a year's time, fast-tracking the production of plants with a trait of commercial value.

Further Regulatory Aspects

How will plant varieties created through gene targeting be regulated? It is likely that each new variety will be considered on a case-by-case basis. In the above example, the genetic variation that conferred resistance to late blight already existed in nature. Further, the resistant variety created through gene targeting is largely equivalent to a variety derived through traditional breeding—an unregulated process. Since DNA-sequence variation closely linked to the resistance gene would also be introduced through breeding, gene targeting is actually more precise. The potato genome modified through gene targeting has only the desired DNA-sequence alteration, and this could be easily confirmed using approaches such as whole-genome sequencing.

The need for case-by-case evaluation of plants derived from gene targeting is warranted because of the range of modifications that can be created. In the potato example, only a handful of DNA-sequence changes—identified from a wild, resistant relative—were

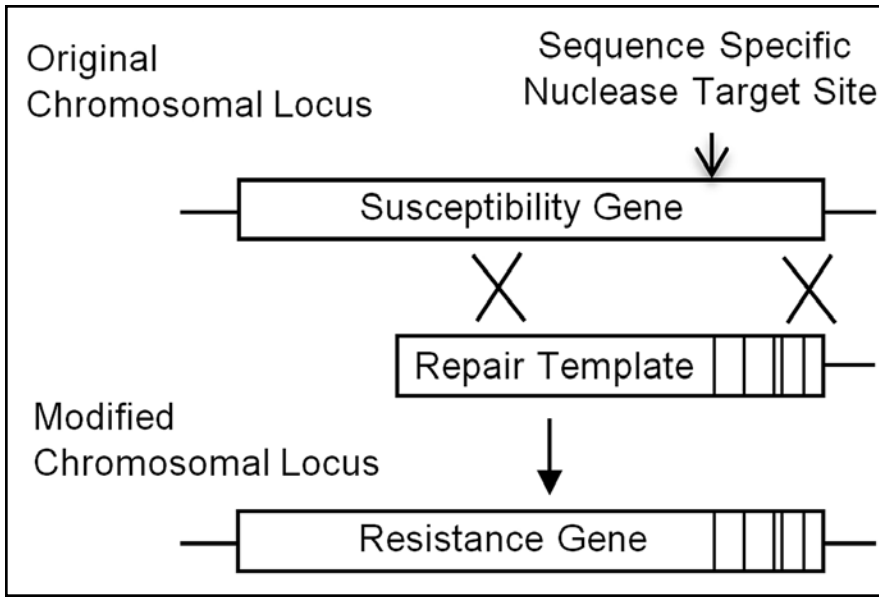


Figure 3. Achieving resistance to late blight through genome engineering. A sequence-specific nuclease is engineered to recognize and cleave the susceptibility gene. A repair template is provided that incorporates mutations (vertical lines) in the susceptibility gene, such that it now confers resistance to late blight.

needed to achieve the desired phenotype. However, variation could also be introduced that is not found in nature. For example, an enzyme's activity could be altered or optimized in the laboratory through *in vitro* evolution or directed mutagenesis to create a novel variant. The genetic changes that underlie the novel activity could then be introduced into the native gene in the plant genome. In considering how to regulate such a plant, a variety of factors will have to be considered, ranging from potential effects if the plant product is consumed (*e.g.* is the variant protein immunogenic?) or its impact on the environment (*e.g.* what are the consequences if the genetic variation moves into weedy relatives of the crop plant?). A clearer picture of how crop varieties created through gene targeting will emerge as new plant varieties are developed and brought to the regulatory authorities for consideration. The guidance provided will be invaluable for those parties using the technology, particularly with respect to estimating the costs needed to pass the regulatory steps prior to field release of a new variety.

Conclusion

Genome engineering has emerged as a powerful means to create genetic variation in plants and is rapidly being deployed for both basic and applied plant biology. The sequence-specific nucleases that enable targeted DNA-sequence modification are precise and accurate, and they alter the genome through well-understood mechanisms of DNA repair. The types of genetic variation that can be created through genome engineering

will contribute to agricultural productivity and help meet the world's burgeoning need for food and other agricultural products. Because genome engineering is a new approach to introduce genetic variation in plants, responsible regulation is required so that the technology can be best deployed for the public good.

References

Bibikova M *et al.* (2003) Enhancing gene targeting with designed zinc finger nucleases. Science 300 764.

Bogdanove AJ Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. Science 333 1843–1846.

Carroll D (2011) Genome engineering with zinc-finger nucleases. Genetics 188 773–782.

Christian M *et al.* (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186 757–761.

Clemente TE Cahoon EB (2009) Soybean oil: genetic approaches for modification of functionality and total content. Plant Physiology 151 1030–1040.

Doudna JA Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346 1258096.

Foster SJ *et al.* (2009) Rpi-vnt1.1, a Tm-2(2) homolog from *Solanum venturii*, confers resistance to potato late blight. Molecular Plant-Microbe Interactions 22 589–600.

Gorbunova V Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Research 25 4650–4657.

Haun W *et al.* (2014) Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnology Journal 12 934–940.

Jinek M *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337 816–821.

Jones H (2015) Regulatory uncertainty over genome editing. Nature Plants 1. doi:10.1038/nplants.2014.11

Kamoun S (2001) Nonhost resistance to Phytophthora: novel prospects for a classical problem. Current Opinion in Plant Biology 4 295–300.

Lusser M *et al.* (2012) Deployment of new biotechnologies in plant breeding. Nature Biotechnology 30 231–239.

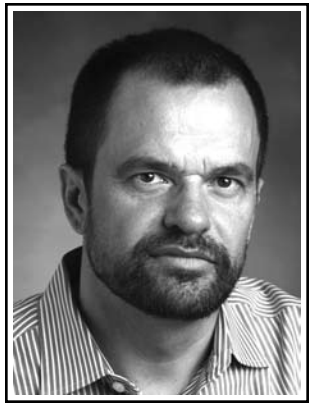
Pâques F Duchateau P (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Current Gene Therapy 7 49–66.

Pauwels K *et al.* (2014) Engineering nucleases for gene targeting: safety and regulatory considerations. New Biotechnology 31 18–27.

Puchta H *et al.* (1993) Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. Nucleic Acids Research 21 5034–5040.

Puchta H *et al.* (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proceedings of the National Academy of Sciences of the USA 93 5055–5060.

Salomon S Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO Journal 17 6086–6095.
Schlueter J *et al*. (2007) The *FAD2* gene family of soybean: insights into the structural and functional divergence of a paloepolyploid genome. Crop Science 47 14–26.
Smith J *et al*. (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Research 34 e149.
Song J *et al*. (2003) Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. Proceedings of the National Academy of Sciences of the USA 100 9128–9133.
Tang GQ *et al*. (2005) Oleate desaturase enzymes of soybean: evidence of regulation through differential stability and phosphorylation. Plant Journal 44 433–446.
Townsend JA *et al*. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459 442–445.
Voytas DF (2013) Plant genome engineering with sequence-specific nucleases. Annual Review of Plant Biology 64 327–350.
Voytas DF Gao C (2014) Precision genome engineering and agriculture: opportunities and regulatory challenges. PLoS Biology 12 e1001877.
Waltz E (2012) Tiptoeing around transgenics. Nature Biotechnology 30 215–217.



DANIEL VOYTAS is a professor in the Department of Genetics, Cell Biology and Development at the University of Minnesota (UMN) and director of the UMN's Center for Genome Engineering.

HE GRADUATED from Harvard College in 1984 and received his PhD in genetics from Harvard Medical School in 1990. He conducted postdoctoral research at Johns Hopkins University School of Medicine where he was a fellow of the Life Science Research Foundation. In 1992, he joined the faculty at Iowa State University where he was promoted to associate professor in 1997 and to professor in 2001. In 2008, he joined the faculty at UMN.

DR. VOYTAS advises agricultural biotechnology companies on the use of new methods of genome engineering for crop improvement and he serves as chief science officer for Cellectis Plant Sciences in New Brighton, MN. He is an elected fellow of the American Association for the Advancement of Science.