Precise Genome Modification in Maize Using Zinc-Finger Nucleases

Vipula K. Shukla[a#], Yannick Doyon[b], Jeffrey C. Miller[b], Russell C. DeKelver[b], Erica A. Moehle[b], Sarah E. Worden[a], Jon C. Mitchell[a], Nicole L. Arnold[a], Jeremy M. Rock[b], Gao Zhifang[a], David McCaskill[a], Matthew A. Simpson[a], Beth Blakeslee[a], Scott A. Greenwalt[a], Sarah J. Hinkley[b], Lei Zhang[b], Edward J. Rebar[b], Philip D. Gregory[b], and Fyodor D. Urnov[b]

[a] *Dow AgroSciences, 9330 Zionsville Rd., Indianapolis, IN 46268, USA*
[b] *Sangamo BioSciences, Point Richmond Tech Center, 501 Canal Blvd., Suite A100, Richmond, CA 94804, USA*
[#] *Author to whom correspondence should be addressed: vkshukla@dow.com*

Summary

Agricultural biotechnology is limited by the inefficiency and unpredictability of conventional random mutagenesis and transgenesis. We have developed a broadly applicable, versatile alternative: the use of designed zinc-finger nucleases (ZFNs) that induce double-stranded breaks (DSB) at their target locus. We show that ZFNs designed to cleave an endogenous maize gene disrupt their target via DSB repair by error-prone non-homologous end-joining. We also show that simultaneous expression of ZFNs and delivery of a simple heterologous donor molecule leads to high frequency targeted gene addition at the intended locus. Modified maize plants faithfully transmit these genetic changes to the next generation. ZFNs can be utilized in any plant species amenable to DNA delivery; therefore our results establish a new strategy for plant genetic manipulation for basic science and agricultural applications.

Introduction

Biotechnology is an essential tool in efforts to meet the challenge of increasing global demand for food production. Current approaches to improving agricultural productivity rely on either mutation breeding or transformation of novel genes into the genome; both these approaches are inherently non-specific and inefficient. Targeted genome modification in plants overcomes many of the logistical challenges associated with conventional biotechnology. For example, approaches to drive targeted gene addition in plants have relied on the use of either positive-negative drug selection (*e.g.* rice[1]) or pre-engineered restriction sites (*e.g.* maize[2]). Despite these efforts, efficient targeted genome modification in all plant species, both model and crop, has remained intractable[3].

In this report we describe plant genome editing via engineered zinc finger nucleases (ZFNs). ZFNs are a fusion of a zinc finger-based DNA recognition module and an endonuclease domain[4] that induces a targeted double-strand break (DSB) in the DNA of living cells, stimulating repair mechanisms including homologous

recombination[5]. ZFNs have been shown to be capable of efficiently modifying engineered loci both in human cells[6] and in model plants[7,8]. Because the zinc finger domain[9] can be engineered to recognize novel DNA sequences [10,11], ZFNs have been widely exploited for rational genome engineering at endogenous loci in many eukaryotic systems[12-14].

In this study, we show that engineered ZFNs designed and screened using a proxy system assay[14] can drive disruption of endogenous maize genes via induction of DSBs. We also show that simultaneous delivery of a ZFN expression cassette and a targeting donor DNA construct to maize cells results in homology-directed addition of an herbicide-tolerance gene cassette. This modification is stable, heritable and confers a phenotype on progeny plants.

Results and Discussion

We selected *ZmIPK1*[15] as a gene target for ZFN-mediated genome modification; this gene encodes an enzyme that catalyzes the ultimate step in phytate biosynthesis in maize seeds. Our analysis of the IPK genes in *Z. mays* indicated that two closely related, paralogous gene loci, here referred to as *IPK1* and *IPK2*, are found in the maize genome[15]; based on the published sequence of *ZmIPK1* we isolated and sequenced the orthologous genes from maize variety HiII [16]. We focused our *ZmIPK1* ZFN design and screening efforts on a 1.2 kbp region spanning exon 1 through exon 3 and used an archive of validated two-zinc-finger modules[11,14] to design and assemble multiple pairs of ZFNs for these regions.

Using a "proxy" system for ZFN screening in budding yeast[14], we were able to (i) directly compare ZFN activity at different binding sites within the target gene and (ii) compare the sequence preference of representative ZFNs for *IPK1* and *IPK2* (Fig. 1, left and right panels, respectively). In all cases, a marked preference for the intended paralog was observed (Fig. 1, right). These results confirmed that the ZFN pairs tested were capable of inducing DSBs within *ZmIPK1* and discriminating between *ZmIPK1* and *ZmIPK2*.

To determine whether the selected ZFN pairs were capable of inducing a DSB in maize cells at *ZmIPK1*, we exploited the error-prone nature of DSB repair by non-homologous end joining (NHEJ). NHEJ is known to generate small DNA deletions / insertions at the site of a ZFN-induced break[8,12]. We therefore utilized maize suspension cultures and transient expression of ZFNs to test for the generation of such mutations. In the example illustrated in Fig. 2, analysis of 6.5×10^4 sequence reads revealed 28 deletions and 2 insertions aligning to the ZFN target/cleavage site in cells expressing ZFN 12 but not control samples. When the experiment was repeated to interrogate the sequence of the *ZmIPK2* paralog, in which the binding site for this particular ZFN pair differs by 1 nucleotide, no such mutations were detected in a set of 6.2×10^4 sequence reads (data not shown). Comparable results were obtained with the other ZFN pairs analyzed, indicating that a short, transient burst of ZFN activity is sufficient to induce a targeted DSB at the ZFN-specified genomic location in cultured maize cells.

Having demonstrated that the ZFNs designed against the *ZmIPK1* gene can introduce DSBs in a sequence-specific manner *in vivo*, we set out to investigate if they could drive targeted gene addition to this locus. Donor constructs carrying short homology arms (Fig. 3a) encompassing the phosphinothricin acetyl transferase (PAT) gene open reading frame, which confers tolerance to the herbicide bialaphos[17], were constructed. One donor construct carried an autonomous PAT gene driven by a heterologous promoter, while the second contained a non-autonomous PAT configured to be under the control of the endogenous *ZmIPK1* promoter, following a ribosome stuttering signal (2A peptide[18]).

Four ZFN pairs designed to cleave *ZmIPK1* at two positions in exon 2 were independently delivered to maize cells, along with either autonomous or non-autonomous donor plasmids, using direct DNA methods. Herbicide-tolerant calli were PCR-genotyped at the *IPK1* locus (Fig. 3b); results revealed both monoallelic insertions of PAT at *ZmIPK1*, as well as events containing only a transgenic chromatid (Fig. 3b, lanes indicated as "TI/-" and "TI/TI," respectively). Cloning and sequencing of PCR products confirmed that gene addition had occurred in a precise, homology-directed manner (Fig. 3c). A summary of PCR-genotyping and subsequent DNA sequencing of all (~600) individual, isolated herbicide-tolerant callus events is shown in Table 1. All 4 ZFN pairs validated in the yeast proxy system drove targeted gene addition into their cognate loci, albeit with different efficiencies; as expected, the non-autonomous donor strategy yielded fewer herbicide tolerant events (lower half, Table 1), but significantly enriched this population for targeted versus random integration.

Table I. Frequency of Targeted Integration of PAT into *ZmIPK1*

ZFN design #	*ZF-binding position*	*Donor Configuration*	*Total # HT* Events Recovered*	*# Targeted Integration Events*	*Percent Targeted Integration*
8	exon 2-1	autonomous PAT	29	1	3.4
12	exon 2-1	autonomous PAT	31	4	12.9
15	exon 2-2	autonomous PAT	195	43	22.1
16	exon 2-2	autonomous PAT	216	46	21.3
12	exon 2-1	combo aPAT + ZFN	39	6	15.4
n/a	n/a	autonomous PAT	25	0	0.0
			535	**100**	**sum**
8	exon 2-1	non-autonomous PAT	1	1	100.0
12	exon 2-1	non-autonomous PAT	5	3	60.0
15	exon 2-2	non-autonomous PAT	16	11	68.8
16	exon 2-2	non-autonomous PAT	8	4	50.0
12	exon 2-1	combo naPAT + ZFN	30	5	16.7
			60	**24**	**sum**

**HT: herbicide tolerant*

Southern blot analysis of a representative panel of herbicide-tolerant events confirmed observations based on PCR. Using an *IPK* gene probe (Fig. 4, top), control samples and herbicide-tolerant events designated as non-targeted based on PCR data (*e.g.*

158, 413, 414, 416, 420) showed only the band corresponding to the non-disrupted allele of *ZmIPK1* (>3.2 kb). In contrast, events designated as carrying a mono- or biallelic targeted-integration (TI) of the PAT cassette based on PCR (Fig. 3b) revealed one transgenic and one wild-type, or a transgenic chromatid only, respectively. Of note, Southern blot analysis using an insert-specific probe revealed multiple random integrants in calli designated as non-targeted (Fig. 4, middle), but only a single band in correctly targeted events. Based on these results, we conclude that in this representative panel of events, primary transformation yielded calli displaying directed transgene insertion only, with no random integration of the heterologous donor DNA. Finally, in leaf tissue of regenerated plants from correctly targeted callus events, no randomly integrated ZFN expression construct was detected (Fig 4, bottom). Taken together, these data show that ZFNs can drive addition of an herbicide resistance marker expression cassette to an investigator-specified locus without additional random integration of either the donor or the ZFN expression cassette.

We regenerated multiple independent plant lines, the majority of which were fertile, from each of several genome-edited callus events. These plants were either self-pollinated or out-crossed to an inbred maize variety (DAS5XH751) lacking herbicide-tolerance genes. Resulting T1/F1 seed were harvested, replanted, and genotyped at the *ZmIPK1* locus in individual plants. Random sampling of 9 T1 seedlings from line 418-8, which was selfed, displayed a 2:6:1 (WT:heterozygotic:homozygotic) segregation pattern for disruption of *ZmIPK* (Fig. 5a, left), while analysis of 10 F1 seedlings from out-crossing of line 418-6 revealed a 1:1 segregation of wild-type *versus* heterozygotic disruption at *ZmIPK1* (Fig. 5a, right). This shows that the ZFN-mediated genome modification observed in primary callus and T0 regenerated plants was transmitted to the next generation in a Mendelian fashion.

Consistent with these data, a proportion of the T1/F1 seedlings displayed an herbicide-tolerant phenotype: F1 progeny of line 418-3 (monoallelic) yielded a 5:4 ratio of tolerant to susceptible plants, while self-pollination plants derived from event #273 (monoallelic) yielded 25 herbicide-tolerant and 16 susceptible progeny. Of note, all outcrossed (33 F1 plants) or self-pollinated (5 T1 plants) progeny of plant lines derived from event #419 survived selection, confirming this event as biallelic.

We assessed the impact of ZFN-mediated insertional mutagenesis at *ZmIPK1* on the accumulation of phytate and its precursors in individual seeds of genome-edited maize plants. Seeds from ZFN-modified, control transgenic (i.e. carrying a random integration of PAT) and wild type maize plants were assayed for the relative ratios of multiple inositol phosphate species. As shown in Fig. 5b, segregating T2 seeds from transgenic plants carrying a randomly inserted PAT gene in the genome show a distribution of high phytate (IP6) and low inorganic phosphate (Pi) accumulation indistinguishable from those obtained with wild type HiII seeds (Fig. 5c left). In contrast, the relative distributions of IP6 and Pi ratios in samples from event 418 show that a significant number of seeds within this segregating population display reduced phytate levels and a concomitant increase in the levels of inorganic phosphate (Fig. 5c, right).

Similar observations were made in lines from each unique event analyzed (data not shown).

In summary, our observations of heritable disruption of *ZmIPK1* expression, coupled to reduced phytate accumulation and increased inorganic phosphate content in seeds, validate our hypothesis that *ZmIPK1* is a suitable gene target for manipulation of phytate reduction and provide the basis for development of a reduced-phytate trait with agronomic and ecological significance. Furthermore, by disrupting *ZmIPK1* via targeted gene addition using a transgene-encoding donor DNA, we have coupled phytate reduction with the introduced herbicide tolerance trait, thereby delivering a dual phenotype, or stacked trait, through precise targeted manipulation of single locus.

The combination of high-fidelity DNA recognition/cleavage by engineered ZFNs plus homology-directed repair at the specified break sites (a widely conserved biological pathway) makes precise modification of native genomes in plants practical and feasible for the first time. The results described here, combined with rapid advances in genome sequencing technologies / bioinformatics, as well as ongoing development of novel DNA delivery methods, establish a new paradigm for the rapid, precise and efficient engineering of plant genomes for development of novel crop traits and basic plant science.

References

1 R. Terada, H. Urawa, Y. Inagaki et al., *Nat Biotechnol* **20** (10), 1030 (2002); R. Terada, Y. Johzuka-Hisatomi, M. Saitoh et al., *Plant Physiol* **144** (2), 846 (2007).

2 K. D'Halluin, C. Vanderstraeten, E. Stals et al., *Plant biotechnology journal* **6** (1), 93 (2008).

3 H. Puchta, *Plant Mol Biol* **48** (1-2), 173 (2002); M. Hanin and J. Paszkowski, *Current opinion in plant biology* **6** (2), 157 (2003).

4 Y. G. Kim, J. Cha, and S. Chandrasegaran, *Proc Natl Acad Sci U S A* **93** (3), 1156 (1996).

5 M. Bibikova, D. Carroll, D. J. Segal et al., *Mol Cell Biol* **21** (1), 289 (2001).

6 M. H. Porteus, T. Cathomen, M. D. Weitzman et al., *Mol Cell Biol* **23** (10), 3558 (2003).

7 D. A. Wright, J. A. Townsend, R. J. Winfrey, Jr. et al., *Plant J* **44** (4), 693 (2005).

8 A. Lloyd, C. L. Plaisier, D. Carroll et al., *Proc Natl Acad Sci U S A* **102** (6), 2232 (2005).

9 J. Miller, A. D. McLachlan, and A. Klug, *Embo J* **4** (6), 1609 (1985).

10 C. O. Pabo, E. Peisach, and R. A. Grant, *Annu Rev Biochem* **70**, 313 (2001).

11 M. Isalan, A. Klug, and Y. Choo, *Nat Biotechnol* **19** (7), 656 (2001).

12 M. Bibikova, M. Golic, K. G. Golic et al., *Genetics* **161** (3), 1169 (2002).

13 M. Bibikova, K. Beumer, J. K. Trautman et al., *Science* **300** (5620), 764. (2003); F. D. Urnov, J. C. Miller, Y. L. Lee et al., *Nature* **435** (7042), 646 (2005); K. Beumer, G. Bhattacharyya, M. Bibikova et al., *Genetics* **172** (4), 2391 (2006); E. A. Moehle, J. M. Rock, Y. L. Lee et al., *Proc Natl Acad Sci U S A* **104** (9), 3055 (2007); A. Lombardo, P. Genovese, C. M. Beausejour et al., *Nat Biotechnol* **25** (11), 1298 (2007); J. C. Miller, M. C. Holmes, J. Wang et al., *Nat Biotechnol* **25**

(7), 778 (2007); Y. Santiago, E. Chan, P. Q. Liu et al., *Proc Natl Acad Sci U S A* **105**, 5809 (2008); X. Meng, M. B. Noyes, L. J. Zhu et al., *Nat Biotechnol* **26** (6), 695 (2008); E. E. Perez, J. Wang, J. C. Miller et al., *Nat Biotechnol* **26** (7), 808 (2008); M. L. Maeder, S. Thibodeau-Beganny, A. Osiak et al., *Mol Cell* **31** (2), 294 (2008).

14 Y. Doyon, J. M. McCammon, J. C. Miller et al., *Nat Biotechnol* **26** (6), 702 (2008).

15 Y. Sun, M. Thompson, G. Lin et al., *Plant Physiol* **144** (3), 1278 (2007).

16 C. Armstrong, C. Green, and R. Phillips, *Maize Genet Coop News Lett* **65**, 92:93 (1991).

17 Arnold W Wohlleben W, Broer I, Hillemann D, Strauch E, Pühler A., *Gene* **70** (1), 25 (1988).

18 N.M. Mattion, Harnish, E.C., Crowley, J.C. & Reilly, P.A. 70, 8124-8127, *J. Virol.* **70** (11), 8124 (1996).

Figure 1. Results of normalized ZFN reporter gene correction activity in a budding yeast proxy screening system. The left panel shows results from testing the indicated ZFNs in a yeast strain that carries the genomic *IPK1* locus. The right panel shows results from testing the same ZFNs in a yeast strain that carries the paralogous *IPK2* locus. For each ZFN pair, activity in the medium of the Mel1 enzyme is shown in mU.

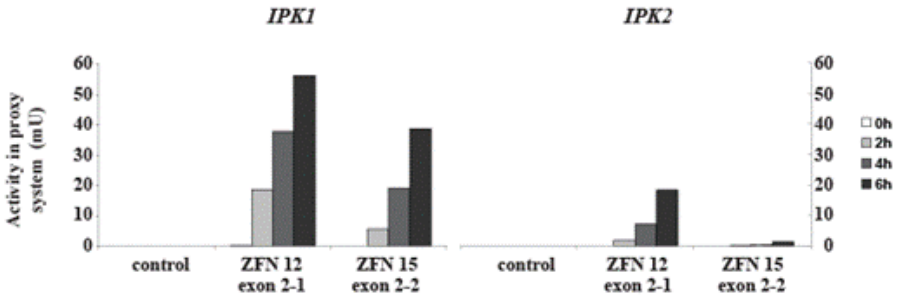


Figure 2. Representative results of assaying for ZFN-induced deletions at *ZmIPK1*. Direct sequence analysis of PCR products via deep "454" pyrosequencing revealed the presence of multiple deletion and insertion mutations induced by NHEJ at the precise predicted ZFN12 cleavage site (box) delimited by the ZFN binding sites (underlined).

Figure 3. Targeted gene addition at *ZmIPK1*.
3a. *ZmIPK1* donor constructs comprised of an autonomous (top) or non-autonomous (middle) PAT gene cassette flanked by short (815 bp) segments of *ZmIPK1* sequence identity (shown as thick black lines). Bottom: schematic representation of the expected edited chromosome following targeted integration of the non-autonomous donor into the endogenous *ZmIPK1* locus.
3b. Representative results of PCR-based genotyping of the *ZmIPK1* locus. Primers hybridizing to *ZmIPK1* gene sequences spanning the predicted integration site (schematic on the right) were used to amplify gDNA from all herbicide tolerant calli. Lanes (left to right): molecular weight markers, wild type HiII, targeted integration events # 273, 278, 417, 418, 419. Lane header TI/- : mono-allelic (heterozygous) insertions. Lane header TI/TI: bi-allelic (homozygous) insertions.
3c. Consensus sequence of target locus analysis of ZmIPK1 PCR-genotyping amplicons. High fidelity integration of the non-autonomous PAT donor was diagnosed based on integrity of the indicated genome/donor boundary sequences, homology arm integration cassette boundaries, preservation of diagnostic sequences in the donor molecule (SpeI, underlined) and PAT gene cassette components.

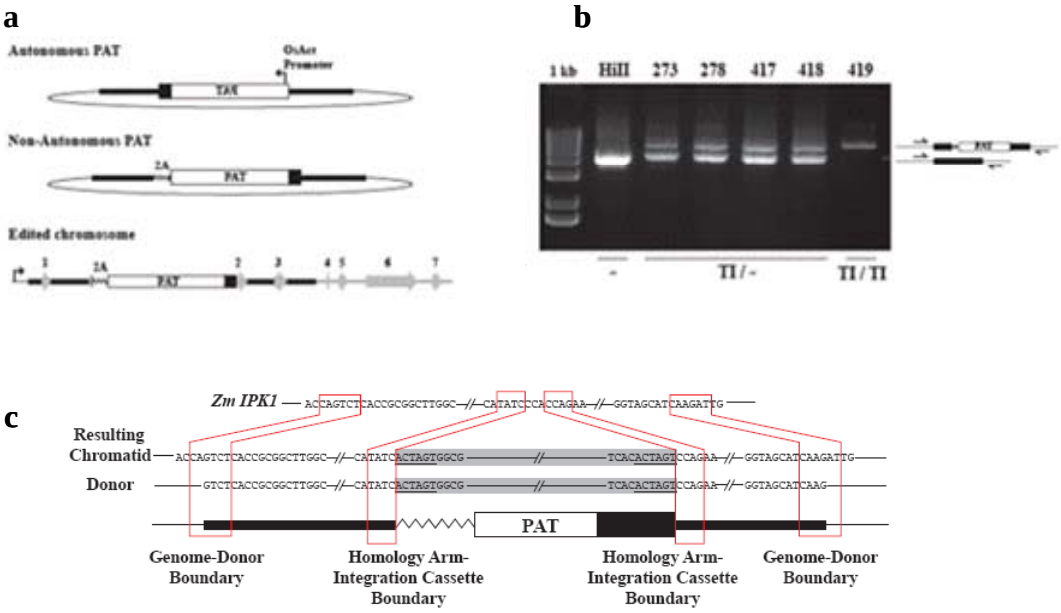
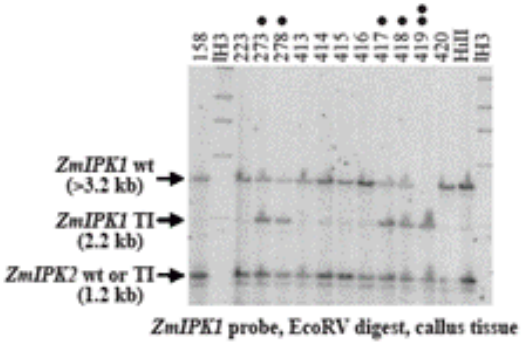
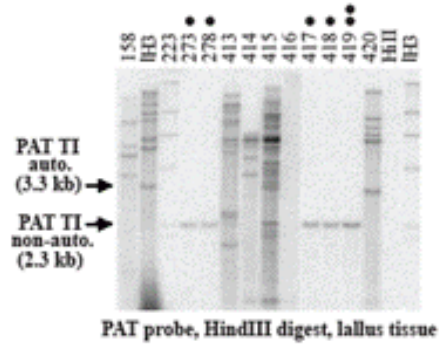


Figure 4. Southern Blot Analysis of Selected Herbicide Tolerant Events
Top: Interrogation of *ZmIPK1* locus size. Lane headers: event numbers, wild type (HiII), molecular weight markers (λH3). Hybridization bands corresponding to wild type *ZmIPK1* (*ZmIPK1* WT; >3.2 kb), disrupted *ZmIPK1* (*ZmIPK1* TI; 2.2 kb) and the paralogous *ZmIPK2* gene (1.2 kb) are indicated by arrows to the left. Events designated as targeted integration (273, 278, 417, 418, 419) are marked with dots.

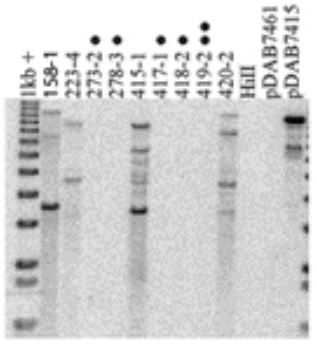
Middle: Interrogation of PAT transgene copy number. Lane headers: event numbers, wild type (HiII), molecular weight markers (λH3). Hybridization bands corresponding to targeted integration (TI) of PAT in either the autonomous (3.3 kb) or non-autonomous (2.3 kb) configuration at *ZmIPK1* are indicated by arrows to the left. Events designated as targeted integration only, with no concurrent random integration of PAT donor (273, 278, 417, 418, 419) are marked with dots.
Bottom: Interrogation of ZFN transgene copy number. Lane headers: event numbers, wild type (HiII), molecular weight markers (1kb+), plasmid controls (pDABX). Events designated as targeted integration only (273, 278, 417, 418, 419) are marked with dots.



***ZmIPK1* probe, EcoRV digest, callus tissue**



PAT probe, HindIII digest, lallus tissue



FokI probe, EcoRV digest, leaf tissue

Figure 5. Stability and Heritability of ZFN-mediated Gene Disruption at *ZmIPK1*
5a. PCR genotyping of T1/F1 plants from event 418. Bands corresponding to wild type (wt) or targeted integration (TI) alleles of *ZmIPK1* are indicated by arrows.
5b. Quantitation of *ZmIPK1* and PAT mRNA in T2/F2 seeds of event 418. X axis: groupings of samples according to detection of *PAT* mRNA; Y axis: relative expression of *ZmIPK1* or PAT mRNA normalized to *EF1α* mRNA. (n): number of individual seeds in each grouping; each seed was subjected to 4-6 independent measurements. Error bars indicate standard deviations of replicate measurements of the same mRNA sample.
5c. Normalized ratios of phytate (IP6) and Pi in T2/F2 seeds of event 418. Data points indicate the distribution of IP6 and Pi ratios in seeds from a randomly-inserted transgenic PAT line (left) and targeted integration event 418 (right). Y axis: normalized peak area for each component measured. (n): number of individual seeds in each type of sample set; each data point represents one of three independent measurements taken per seed.