The Plant Cell, Vol. 28: 603–605, March 2016, www.plantcell.org

LETTER TO THE EDITOR

Ws-2 Introgression in a Proportion of *Arabidopsis thaliana Col-0* Stock Seed Produces Specific Phenotypes and Highlights the Importance of Routine Genetic Verification[OPEN]

Arabidopsis thaliana is an important model organism with a robust network of resources that has been of enormous value to the plant science research community. The use of isogenic material as a reference point or control is critical for many types of experiments in plant molecular biology and genetics. Recently, we noticed that some seed from a common source of the widely used *Columbia-0* (*Col-0*) strain gave rise to plants showing features atypical for this strain. Whole-genome DNA-sequencing and allele-specific PCR assays confirmed that the abnormal individuals contain multiple introgressions from the ecotype *Wassilewskija-2* (*Ws-2*), as described below. This emphasizes the importance of practices necessary to maintain the integrity of seed stocks and other biological collections. We urge research groups to evaluate whether they may have been affected and to revisit their materials if needed.

PHENOTYPIC VARIANTS WITHIN A COMMON *Col-0* SEED STOCK CONTAIN CHROMOSOMAL SEGMENTS FROM *Ws-2*

Relative to other Arabidopsis ecotypes, *Col-0* is characterized by a medium rosette size, slightly serrated leaf margins, intermediate height, and an intermediate flowering time (TAIR, www.arabidopsis.org). However, in *Col-0* seed (lot #214-509) obtained from LEHLE SEEDS Company, we observed that some plants were larger than others grown in the same tray. The abnormal plants showed an increase in leaf area, displayed broader, flatter, and more serrated leaves, and tended to flower earlier than other individuals (particularly when grown under long-day cycles; Supplemental Figure 1). Based on phenotypic scoring, we observed that ~6 to 10% (Table 1) of plants from the indicated lot were phenotypically abnormal compared with the majority of the *Col-0* plants.

We suspected that the abnormal individuals could be the result of seed contamination and chose six plants (four abnormal and two typical) for whole-genome DNA-sequencing. Relative to the reference *Col-0* sequence, the two typical plants each had fewer than 700 homozygous single nucleotide polymorphisms (SNPs) detected by the software SHORE (Schneeberger et al., 2009). The four abnormal individuals, however, each had over 131,000 homozygous SNPs (Supplemental Figure 2A), concentrated on Chromosomes 1, 3, and 5. Except for a comparatively small segment on Chromosome 5, the distributions of SNPs were essentially identical across all four abnormal individuals (Supplemental Figure 2B), with an introgression pattern and number consistent with expectations if outcrossing followed by recombination occurred (Giraut et al., 2011). The apparent lack of segregation among the four abnormal samples indicates that the genetic contamination in this seed lot likely arose from a single hybridization event followed by self-pollination for several additional generations, leading to genetic fixation in these individuals.

A comparison of the SNPs in abnormal plants to variant files generated by the Arabidopsis 1001 Genomes Project revealed that ~84% (or 115,000) of SNPs called from each abnormal plant were also found in the *Ws-2* ecotype. The remaining 16% of abnormal sample SNPs unmatched to *Ws-2* follow a similar chromosomal distribution to the SNPs matching *Ws-2* (Supplemental Figure 2B), suggesting that they may be the result of technical differences in software and parameter settings, as opposed to an additional genetic source. The abnormal sample SNPs that matched *Ws-2* also had nearly an equally strong match to the ecotype *Ragl-1*. Indeed, the publicly available *Ws-2* and *Ragl-1* variant files share over 91% SNPs in common, an extremely high proportion for two Arabidopsis ecotypes purportedly from geographically distant regions (*Ws-2* from Russia and *Ragl-1* from the UK). A comparison between 5965 accessions by Anastasio et al. (2011) placed *Ragl-1* within the same haplogroup as *Ws-2* and identified it as one of the hundreds of accessions having potentially misidentified geographic origins. No other ecotype from the 838 accessions we tested accounted for more than 65% of the abnormal sample SNPs, and most accounted for only 18.78 to 50.81% of the abnormal sample SNPs, which is consistent with comparisons between different ecotypes (Salomé and Weigel, 2014).

SNPhylo (Lee et al., 2014) was used to perform SNP-based phylogenetic analysis with variant files of *Ws-2*, additional ecotypes (primarily from the same sequencing project as *Ws-2*), and our samples. For these samples and ecotypes, SNPs within a subset of coordinates contained in the large introgressions on Chromosomes 1, 3, and 5 were included. Corroborating the prior SNP matching approach, these results indicated that the genomic blocks within the abnormal samples are similar to *Ws-2* (Supplemental Figure 2C), while the typical-appearing *Col-0* samples are likely pure *Col-0*. Based on these results, we suggest that *Ws-2*, a relatively common laboratory ecotype provided by several seed distributors, is the donor of the observed genetic variation. We found no compelling evidence of additional genetic history involving artificial mutagenesis in the abnormal samples.

Recently, phenotypes thought to be caused by a mutant allele for the auxin binding protein ABP1 were called into doubt when it was discovered that the *abp1-5* mutant line harbored multiple second-site mutations, as well as a large *Ws-2* introgression (Enders et al., 2015) that may have resulted from backcrossing to a contaminated *Col-0* plant.

[OPEN]Articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.16.00053

Table 1. Estimated Percentage of Affected Plants in Indicated *Col-0* Seed Lots Based on Phenotypic Scoring or PCR Genotyping

Lot	Year	Method	Plants Grown	Affected Plants	Frequency
#197-089	1998	PCR	23	0	0%
#203-280	2003	PCR	84	8	9.5%
#206-440	2007	PCR	84	5	6.0%
#210-485	2010	Phenotype	275	17	6.2%
#214-509	2014	Phenotype	213	22	10.3%
#215-511	2015	PCR	96	15	15.6%

Although both have introgressions on Chromosome 3, the overall distribution of *Ws-2* SNPs in our abnormal samples is not the same as the *abp1-5* line, suggesting that these events occurred independently.

To validate the DNA sequencing results, single-nucleotide amplified polymorphism (SNAP) primers were designed based on known *Ws-2* SNPs that were also identified in our abnormal samples by SHORE. All the plants that were visually scored as abnormal harbored *Ws-2*-specific SNPs in the introgressed regions and *Col-0*-specific SNPs in the introgression-free region, while all the plants that were visually scored as normal harbored *Col-0*-specific SNPs in all the regions tested (Supplemental Figure 3). Using two PCR markers, random sampling of 96 seedlings from lot #215-511 found that 15 seedlings (15.6% of total) were positive for *Ws-2* SNPs (Table 1), a proportion generally consistent with phenotypic scoring. PCR assays of additional earlier lots indicate that seed lots going back to at least 2003 are affected.

Thus, whole-genome DNA-sequencing and allele-specific PCR assays confirm that a genetic mixture between *Col-0* and *Ws-2* is present in a proportion of this *Col-0* seed stock. Groszmann et al. (2014) found that *Col-0* × *Ws* hybrids show modest (~15%) heterosis in rosette diameter up to the first 28 d of growth. The phenotypic changes in our abnormal samples may reflect *Col-0*/*Ws-2* heterosis.

LEHLE *Col-0* PEDIGREE AND PROTOCOLS

LEHLE SEEDS propagates seed for 21 Arabidopsis ecotypes. Historically, different ecotypes have been planted at different times at four locations in Tucson, AZ and four locations in Round Rock, TX. *Col-0* bulk seed has been grown continuously since 1985, with 48 generations as of the most recent 2015 bulk. *Ws-2* bulk seed has been grown every 1 to 2 years since 1989, with the last bulk in 2003, for a total of 13 generations. In 1995 to 1998, 2000, and 2003, *Col-0* and *Ws-2* bulks were grown in the same location, providing the opportunity for possible cross-pollination.

The standard practice by LEHLE SEEDS is to perform simple sequencing length polymorphism analysis using five separate markers on DNA preparations from several pooled plants and from 20 individual plants. Two of these markers appear to fall within the *Ws-2* introgressions, suggesting that the contamination was at sufficiently low frequency (~6%) when last tested.

From now on, LEHLE SEEDS will abandon the practice of using raw seed from a previous propagation for a high-density planting of the next propagation. Rather, ecotypes will be bulk propagated only from selfed progeny of individual plants, all of which will have passed a growth-stage phenotypic analysis for conformity to ecotype under low-density planting and easily reproducible conditions. A recent hydroponic platform for Arabidopsis looks promising for this purpose (Conn et al., 2013). As there seems to be no solution for eliminating outcrossing completely in confined spaces, LEHLE SEEDS will conduct future bulk propagations of *Col-0* in complete physical isolation from other ecotypes.

RECOMMENDATIONS

Given typical high-density growth conditions of Arabidopsis, along with its branching habit, numerous flowers, fecundity, and small seed size, accidental outcrossing or seed mixtures can sometimes occur despite good practices. Thus, careful observation and molecular characterization are recommended, particularly when a mutation of interest could conceal other phenotypes epistatically or when conducting experiments such as genetic screens where the large number of plants used increases the chances of inadvertently including a contaminant (Greene et al., 2003). If multiple Arabidopsis ecotypes are grown at a single facility, steps taken to ensure genetic purity can include staggered planting, physical separation, staking, floral sleeves, careful seed collection habits, secure seed drying, organized storage, and periodic genotyping.

Recent advancements in bioinformatics tools enable SNP calling from a variety of sequencing applications (Ossowski et al., 2008; Van der Auwera et al., 2013) as a confirmation of genetic background and purity. Since an increasing number of research projects leverage some form of sequencing data, one recommendation is that, whenever appropriate, SNP analysis be routinely included in the bioinformatics pipeline of high-throughput sequencing experiments.

Supplemental Data

Supplemental Figure 1. Phenotype of abnormal plants in LEHLE SEEDS Company *Col-0* lot #214-509.

Supplemental Figure 2. DNA-sequencing of abnormal plants reveals large *Ws-2* introgressions.

Supplemental Figure 3. Allele-specific PCR assays confirm presence of *Ws-2* introgression in abnormal plants.

Supplemental Table 1. SNAP PCR primer sequences used for genotyping shown in Supplemental Figure 3.

Mon-Ray Shao
Department of Agronomy and Horticulture
University of Nebraska
Lincoln, NE 68588

Vikas Shedge
Department of Agronomy and Horticulture
University of Nebraska
Lincoln, NE 68588

Hardik Kundariya
Department of Agronomy and Horticulture
University of Nebraska
Lincoln, NE 68588

Fredric R. Lehle
LEHLE SEEDS
Round Rock, TX 78681-2366
ORCID ID: 0000-0002-9019-6099

Sally A. Mackenzie
Department of Agronomy and Horticulture
University of Nebraska
Lincoln, NE 68588
smackenzie2@unl.edu
ORCID ID: 0000-0003-2077-5607

ACKNOWLEDGMENTS

We thank the University of Nebraska Beadle Center Greenhouse Facility for assistance in tracing the seed contamination and identifying seed lots. Funding for this analysis was provided by a grant from the Bill and Melinda Gates Foundation to S.A.M.

AUTHOR CONTRIBUTIONS

M.-R.S., V.S., and S.A.M. designed the experiments. M.-R.S. performed DNA-sequencing analysis. V.S. performed SNAP PCR assays. H.K. provided phenotypic measurements. M.-R.S., F.L., and S.A.M. wrote the article. M.-R.S and V.S. contributed equally to this work.

Received January 26, 2016; revised March 15, 2016; accepted March 15, 2016; published March 15, 2016.

REFERENCES

Anastasio, A.E., Platt, A., Horton, M., Grotewold, E., Scholl, R., Borevitz, J.O., Nordborg, M., and Bergelson, J. (2011). Source verification of mis-identified *Arabidopsis thaliana* accessions. Plant J. **67:** 554–566.

Conn, S.J., Hocking, B., Dayod, M., Xu, B., Athman, A., Henderson, S., Aukett, L., Conn, V., Shearer, M.K., Fuentes, S., Tyerman, S.D., and Gilliham, M. (2013). Protocol: optimising hydroponic growth systems for nutritional and physiological analysis of *Arabidopsis thaliana* and other plants. Plant Methods **9:** 4.

Enders, T.A., Oh, S., Yang, Z., Montgomery, B.L., and Strader, L.C. (2015). Genome sequencing of Arabidopsis abp1-5 reveals second-site mutations that may affect phenotypes. Plant Cell **27:** 1820–1826.

Giraut, L., Falque, M., Drouaud, J., Pereira, L., Martin, O.C., and Mézard, C. (2011). Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. PLoS Genet. **7:** e1002354.

Greene, E.A., Codomo, C.A., Taylor, N.E., Henikoff, J.G., Till, B.J., Reynolds, S.H., Enns, L.C., Burtner, C., Johnson, J.E., Odden, A.R., Comai, L., and Henikoff, S. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics **164:** 731–740.

Groszmann, M., Gonzalez-Bayon, R., Greaves, I.K., Wang, L., Huen, A.K., Peacock, W.J., and Dennis, E.S. (2014). Intraspecific Arabidopsis hybrids show different patterns of heterosis despite the close relatedness of the parental genomes. Plant Physiol. **166:** 265–280.

Lee, T.H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics **15:** 162.

Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res. **18:** 2024–2033.

Salomé, P.A., and Weigel, D. (2014). Plant genetic archaeology: whole-genome sequencing reveals the pedigree of a classical trisomic line. G3 (Bethesda) **5:** 253–259.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jørgensen, J.E., Weigel, D., and Andersen, S.U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat. Methods **6:** 550–551.

Van der Auwera, G.A., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics **11:** 11.10.1–11.10.33.