LETTER

REPLY TO HUANG ET AL.: Avoiding "one-size-fits-all" approaches to variant discovery

Michael A. Hardigan[a], F. Parker E. Laimbeer[b], John P. Hamilton[a], Brieanne Vaillancourt[a], David S. Douches[c], Eva M. Farré[a], Richard E. Veilleux[b], and C. Robin Buell[a,1]

Huang et al. (1) argue that variant calling methods less conservative than GATK's Best Practices workflow (2) increased false-positive variant discovery in our study of wild and cultivated potatoes (3), impacting diversity estimates. We disagree with their conclusion and highlight the rationale for the variant calling methods used in our study.

First, GATK Best Practices, developed at the Broad Institute, were specifically designed and optimized for human genomics and medical research. Variant calling methodology for human medical research utilizes parameters and validation thresholds not intended for universal application across genomic studies. GATK's hard filter is openly presented as a bias-prone substitute to their preferred (human-specific) variant quality score recalibration (2). GATK's Best Practices webpage explicitly addresses applications of their workflow to different organisms: "They can be *adapted* for analysis of non-human organisms of all kinds, including non-diploids, and of different data types, with varying degrees of effort depending on how divergent the use case and data type are" (https://software.broadinstitute.org/gatk/best-practices/). We contend that arbitrary extension of filtering parameters established for humans to plants in which heterozygosity, repetitive sequence, structural variation, and divergence from reference genomes are several orders higher (4, 5) demonstrates a reductive approach to genomic research failing to account for studies involving more diverse species or, in our study, numerous species.

Second, Huang et al. (1) falsely report that our study utilized GATK for variant calling and lacked filtering for strand and read depth. We utilized FreeBayes, a tool widely employed for genomic studies (6), as GATK was shown to perform poorly for detecting variants with low allelic frequencies (<5%) (7). These comprise a majority of calls in our study due to inclusion of >20 species. Due to high sequence divergence among wild *Solanum* taxa, a MapQ alignment threshold of 20 was chosen to avoid biasing variant discovery and domestication scans toward gene functions based on sequence conservation with the *Solanum tuberosum* reference genome. Furthermore, we included filters for strand bias (<80%), and minimum read depth, specifically reporting nucleotide diversity estimates for 390 Mb of highly conserved sequence within the 844-Mb potato genome.

Estimates of nucleotide diversity in crop species depend not only on variant calling methods but also genetic bottlenecks, ploidy, and reproductive mode. Thus, it is not surprising that outcrossing maize, watermelon, and potato report higher nucleotide diversity (figure 2C in ref. 3) than inbreeding species with strong genetic bottlenecks such as tomato and soybean (8, 9, 10). However, variant counts generated for potato using GATK hard filtering (1) were lower than those for both tomato and soybean, an unlikely outcome considering the biology of these species.

Huang et al. (1) provide no data supporting their assertion that lower SNP counts were driven by exclusion of false positives, or conversely, address exclusion of true polymorphisms by employing conservative thresholds developed for human genomics research. While their arguments contesting variant numbers in tuber-bearing *Solanum* are not well supported, we agree with the authors' opinion that genetic diversity alone cannot predict phenotypic potential for tuber traits. We reported no such conclusion in our study.

1 Huang B, Spooner DM, Liang Q (2018) Genome diversity of the potato. *Proc Natl Acad Sci USA* 115:E6392–E6393.
2 Van der Auwera GA, et al. (2013) From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43:1–33.

[a]Department of Plant Biology, Michigan State University, East Lansing, MI 48824; [b]Department of Horticulture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061; and [c]Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI 48824

Author contributions: M.A.H., F.P.E.L., J.P.H., B.V., D.S.D., E.M.F., R.E.V., and C.R.B. wrote the paper.
The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: buell@msu.edu.

Published online June 22, 2018.

3 Hardigan MA, et al. (2017) Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc Natl Acad Sci USA* 114:E9999–E10008.
4 Rafalski A, Morgante M (2004) Corn and humans: Recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet* 20:103–111.
5 Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. *Science* 320:481–483.
6 Hwang S, Kim E, Lee I, Marcotte EM (2015) Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci Rep* 5:17875.
7 Sandmann S, et al. (2017) Evaluating variant calling tools for non-matched next-generation sequencing data. *Sci Rep* 7:43169.
8 Bai Y, Lindhout P (2007) Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Ann Bot* 100:1085–1094.
9 Lin T, et al. (2014) Genomic analyses provide insights into the history of tomato breeding. *Nat Genet* 46:1220–1226.
10 Hyten DL, et al. (2006) Impacts of genetic bottlenecks on soybean genome diversity. *Proc Natl Acad Sci USA* 103:16666–16671.