Molecular Plant
Spotlight

CellPress
Partner Journal

The power of chromosome-scale, haplotype-resolved genomes

The assembly of chromosome-scale and haplotype-resolved reference genomes is now more easily attainable, largely due to various improvements in both assembly algorithms and long-read sequencing technologies (see recent review by Michael and VanBuren, 2020). Due to these technological advancements, there has been a shift away from using highly inbred accessions, in some instances double haploids, to instead assembling the genomes of important reference genotypes. For example, a haplotype-resolved genome was recently assembled for an important cultivated garden strawberry (*Fragaria x ananassa*) cultivar Royal Royce, which is not only highly heterozygous but also octoploid (8×) (Hardigan et al., 2021). Similarly, haplotype-phased genomes have been assembled for other important crops, including cassava (*Manihot esculenta*) (Hu et al., 2021) and potato (*Solanum tuberosum*) (Hoopes et al., 2022). This has resulted in the development of new genomic resources that can be directly used in guiding molecular breeding efforts, without worrying about the gene presence-absence variation that exists within a crop or species (Golicz et al., 2016; Bayer et al., 2021). An additional benefit of having a reference genome of an important cultivar and/or genotype, versus from a highly inbred line, is that valuable insights that are potentially gained from the analysis of genetic variants present among haplotypes are not lost, allowing for additional insights.

Recently, Hu et al. (2022) published the lychee (*Litchi chinensis* Sonn.) genome and provided one of the best example studies to date of how analyzing and comparing haplotypes can result in exciting new discoveries. The analysis of both haplotypes provided novel insights not only into the domestication history of this important tropical fruit but also into the underlying genetics encoding important traits and heterosis (Hu et al., 2022). These findings would not have been revealed from the analysis of a single haplotype as a result of either assembling a genome of a highly inbred line or from collapsing haplotypes to form a single master reference.

Lychee is a tropical perennial fruit tree species, native to southeast Asia, that has been cultivated for its unique appearance and flavor for millennia, and remains of great economic, historical, and cultural value to various parts of the world. The oldest known lychee tree sprouted over 1250 years ago during the Tang dynasty in China and still produces fruit today (Hu et al., 2022). Cultivars are classified into one of three major groups based solely on their fruit maturity: extremely early-maturing cultivars, early- to intermediate-maturing cultivars, and late-maturing cultivars (Hu et al., 2022). However, the domestication of this important crop has remained poorly understood. Wild lychee still exist in rainforests in various regions of southern China, including Haian, Yunnan, Guangxi, and Guangdong provinces (Figure 1), as well as other countries, including Vietnam.

Hu et al. (2022) assembled a chromosome-scale, haplotype-resolved reference genome of the lychee cultivar Feizixiao and generated resequencing data for an additional 72 wild and cultivated accessions aimed at investigating the domestication history and to explore genetic variation present in this crop species. When these resequencing data were aligned to the reference genome, notable coverage differences were observed between each of the haplotypes, with distinct patterns that were specific to the extremely early- and late-maturing cultivar groups. In other words, reads preferentially mapped to haplotype 1 for extremely early-maturing cultivars, while reads preferentially mapped to haplotype 2 for late-maturing cultivars. Furthermore, the observed coverage differences among haplotypes for extremely early-maturing cultivars mirrored wild species from Yunnan, whereas late-maturing cultivars mirrored wild species from Hainan. This observation, in addition to results from phylogenetic and principal component analyses, revealed that these two distinct cultivar groups have unique geographic origins and may represent independent domestication events (Figure 1). Early- to intermediate-maturing cultivars, including Feizixiao, were discovered to be hybrids of extremely early- and late-maturing cultivars.

In addition to allelic variants, numerous gene content variants were identified that were unique to the two haplotypes in the Feizixiao genome. The observed haplotype variation alone supports the need to develop pangenome resources for a particular crop species and the impact of reference genome bias on downstream analyses and applications (Crysnanto and Pausch, 2020). The authors were also able to show that the number of deleterious alleles from the two founding populations likely decreased during domestication and in breeding programs following the selection of the most vigorous high-yielding cultivars. This observation supports the model that hybrid vigor (heterosis) emerges from the masking and/or purging of deleterious alleles (see review by Birchler et al., 2003).

As noted above, flowering time and the fruit maturation period are key target traits for lychee.

Hu et al. (2022) identified and analyzed over 500 homologs to known flowering-related genes and gene families previously characterized in other species, and identified many unique gene duplicates with diverged expression patterns in lychee. In

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

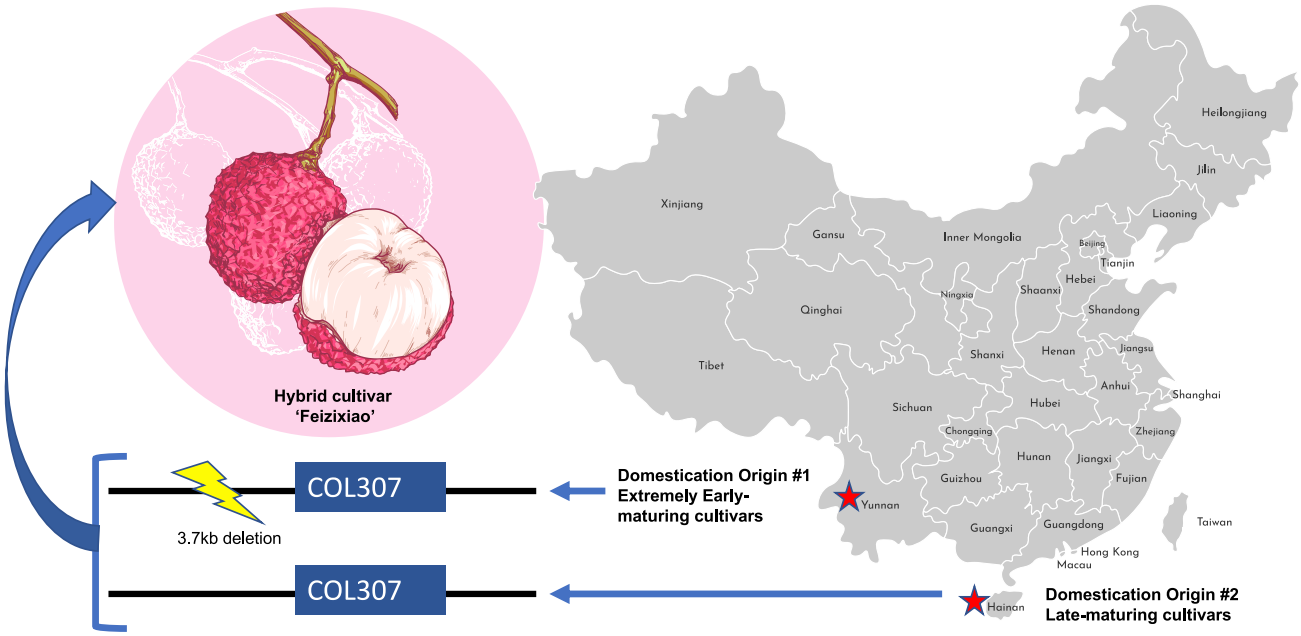

Figure 1. Origin and domestication history of lychee.
Red stars denote the approximate locations of the two independent domestication origins in China; Yunnan for extremely early-maturing cultivars and Hainan for late-maturing cultivars. A 3.7-kb deletion close to a homolog of CONSTANS (COL307), possibly caused by a transposable element, may in large part contribute to the observed differences in flowering-time and fruit maturity dates. The reference genome of the hybrid cultivar Feizixiao contains both variants of this gene and exhibits an intermediate phenotype.

addition, they identified putative flowering time regulators that fell within "selective sweep" regions across the genome. Of particular interest, allelic variants for a gene that encodes a CONSTANS-like (COL) protein, a transcription factor known to regulate flowering time in *Arabidopsis* (Suárez-López et al., 2001), was identified on the two haplotypes. All early-flowering lines were shown to have a 3.7-kb deletion near this gene, while all later-flowering lines lacked this particular deletion (Figure 1). A long-terminal repeat (LTR) retrotransposon was hypothesized to have caused the deletion, resulting in the modified expression of that nearby gene, and ultimately leading to observable flowering time and fruit maturation differences among lychee cultivars (Hu et al., 2022).

In summary, Hu et al. (2022) provide an excellent example of the power of incorporating haplotype data into genome analyses, which uncovered evidence not only for multiple origins and domestication of lychee but also into the underlying genetics encoding key target traits. Only a partial view of these results would have been revealed with a single-haplotype reference genome. Nearly half of the alleles between the two haplotypes in the Feizixiao genome were differentially expressed, and, given that transposable elements can affect the expression of nearby genes and drive phenotypic variation (see review by Lisch, 2013), future analysis of this and other genomes should incorporate variation of these important genomic features into downstream analyses. An integrated systems biology approach, leveraging high-quality reference genomes such as lychee (Hu et al., 2022), will certainly continue to improve our understanding of how genomes evolve, detect associations between observed genetic and phenotypic variation, and will greatly accelerate future molecular breeding efforts of diverse crops.

FUNDING
This work was supported by Michigan State University AgBioResearch, USDA-HATCH 1009804, USDA-AFRI 2019-51181-30015, and NSF-PGRP 2029959.

ACKNOWLEDGMENTS
We thank MacKenzie Jacobs, Jordan Brock, and the anonymous reviewers for their thoughtful comments. No conflict of interest is declared.

Patrick P. Edger[1,2,*]

[1]Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA
[2]Genetics and Genome Sciences, Michigan State University, East Lansing, MI 48824, USA
***Correspondence: Patrick P. Edger (edgerpat@msu.edu)**
https://doi.org/10.1016/j.molp.2022.02.010

REFERENCES
Bayer, P.E., Scheben, A., Golicz, A.A., Yuan, Y., Faure, S., Lee, H., Chawla, H.S., Anderson, R., Bancroft, I., Raman, H., et al. (2021). Modelling of gene loss propensity in the pangenomes of three Brassica species suggests different mechanisms between polyploids and diploids. Plant Biotechnol. J. **19**:2488–2500.

Birchler, J.A., Auger, D.L., and Riddle, N.C. (2003). In search of the molecular basis of heterosis. Plant Cell **15**:2236–2239.

Crysnanto, D., and Pausch, H. (2020). Bovine breed-specific augmented reference graphs facilitate accurate sequence read mapping and unbiased variant discovery. Genome Biol. **21**:184.

Golicz, A.A., Batley, J., and Edwards, D. (2016). Towards plant pangenomics. Plant Biotechnol. J. **14**:1099–1105.

Hardigan, M.A., Feldmann, M.J., Pincot, D.D.A., Famula, R.A., Vachev, M.V., Madera, M.A., Zerbe, P., Mars, K., Peluso, P., Rank, D., et al. (2021). Blueprint for phasing and assembling the genomes of

heterozygous polyploids: application to the octoploid genome of strawberry. Preprint at bioRxiv. https://doi.org/10.1101/2021.11.03.467115.

Hoopes, G., Meng, X., Hamilton, J.P., Achakkagari, S.R., de Alves Freitas Guesdes, F., Bolger, M.E., Coombs, J.J., Esselink, D., Kaiser, N.R., Kodde, L., et al. (2022). Phased, chromosome-scale genome assemblies of tetraploid potato reveals a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. Mol. Plant, S1674-2052, 00003-X. https://doi.org/10.1016/j.molp.2022.01.003.

Hu, W., Ji, C., Shi, H., Liang, Z., Ding, Z., Ye, J., Ou, W., Zhou, G., Tie, W., Yan, Y., et al. (2021). Allele-defined genome reveals biallelic differentiation during cassava evolution. Mol. Plant **14**:851–854.

Hu, G., Feng, J., Xiang, X., Wang, J., Salojärvi, J., Liu, C., Wu, Z., Zhang, J., Liang, X., Jiang, Z., et al. (2022). Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars. Nat. Genet. **54**:73–83.

Lisch, D. (2013). How important are transposons for plant evolution? Nat. Rev. Genet. **14**:49–61.

Michael, T.P., and VanBuren, R. (2020). Building near-complete plant genomes. Curr. Opin. Plant Biol. **54**:26–33.

Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G. (2001). CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature **410**:1116–1120.