Adaptation in plant genomes: bigger is different

Wenbin Mei[1], Markus G Stetter[1], Daniel J Gates[1], Michelle C Stitzer[1,2], and Jeffrey Ross-Ibarra*[1,2,3]

[1]Dept. of Plant Sciences, University of California, Davis, CA 95616
[2]Center for Population Biology, University of California, Davis, CA 95616
[3]Genome Center, University of California, Davis, CA 95616

October 10, 2017

Since their origin 160 million years ago, flowering plants have rapidly diversified into more than 300,000 species, adapting to a striking array of habitats and conditions. In this time, flowering plants have dominated some of the most diverse and extreme environments, harnessed a number of specialized biotic interactions in order to ensure successful pollen and seed dispersal, and adapted to meet the demands of agriculture. Given their diversity and importance, a considerable body of research has been devoted to understanding plant adaptation (Tiffin and Ross-Ibarra, 2014), but the relative importance of the various factors that may impact the process of adaptation are still not well understood. For instance, transitions in polyploidy and mating system have long been considered plausible drivers of flowering plant adaptation (Soltis et al., 2009; Goldberg and Igić, 2012), but both are also associated with evolutionary dead ends (Mayrose et al., 2011; Igic and Busch, 2013). As another example, while the effective size of plant populations is expected to correlate with estimates of the efficiency of natural selection, empirical support for this prediction is mixed (Strasburg et al., 2010; Gossmann et al., 2010). A number of aspects of adaptation have received a degree of theoretical (Hermisson and Pennings, 2017; Ralph and Coop, 2010) or empirical (Strasburg et al., 2012; Anderson et al., 2013; Ågren et al., 2013; Yeaman et al., 2016) support, but it is clear we are far from understanding all of the factors underlying the process of adaptation in flowering plants.

Here we propose that genome size may play a previously under-appreciated role in determining how plants adapt. Rather than focus on the mechanisms of genome size variation (Lynch et al., 2016; Lynch and Conery, 2003) or the adaptive significance of genome size itself (Grime and Mowforth, 1982; Bilinski et al., 2017), our **functional space hypothesis** predicts that interspecific differences in genome size may affect the process of adaptation by changing the number and location of potentially functional mutations. Below, we outline an argument for why existing differences in genome size — spanning more than three orders of magnitude across flowering plants (Gaut and Ross-Ibarra, 2008) — may lead to differences in "functional space", and what this implies about adaptation in plant genomes.

---

*rossibarra@ucdavis.edu

Larger genomes have more functional space

Mutation rates are typically reported at the level of individual nucleotides, but unless these rates change in larger genomes, plants with more nucleotides will inevitably be subject to more mutations *per genome* in each generation. And because variation in gene number across plants is not substantial (Bennetzen et al., 2005), most of the additional mutational input in larger genomes occurs outside of coding sequence.

But not only nucleotides mutate. By definition, large genomes are a consequence of insertion of additional base pairs. In plants, diploid genome size expansion is often the result of amplification of transposable elements (TEs). In addition to the nucleotides contributed by their insertion, the presence of TEs can generate variation in gene expression and regulation and even generate genome rearrangements (Chuong et al., 2016; Lisch, 2013). Beyond TEs, structural variation such as inversions are likely to be abundant in large genomes due to their association with unstable intergenic DNA. Inversions have often been found to play an oversized role in plant adaptation, from perenniality in monkeyflower (Lowry and Willis, 2010) to flowering time in maize (Navarro et al., 2017) and *Boechera* (Lee et al., 2017) to fecundity under drought in *Arabidopsis* (Fransz et al., 2016). Other structural variation is likely more common in larger genomes as well. Gene movement is frequent in large genomes such as wheat and barley (Wicker et al., 2011), frequently due to the action of transposable elements (Morgante et al., 2005). Variation in gene copy number is also common (Żmieńko et al., 2014), and has been identified as the source of a number of different adaptations (Pham et al., 2017; Prunier et al., 2017).

Together, nucleotide changes and structural changes combine to contribute an increased influx of new variation in large plant genomes. Because some portion of these mutations will have phenotypic consequences, this new variation provides greater mutational “space” on which selection can act. While clearly most mutational changes are unlikely to impact function, especially in intergenic regions, if even a small fraction of this mutational input in larger genomes impacts phenotype, selection can act to maintain or remove these mutations in populations. Indeed, the quanitity of expressed intergenic sequence appears to scale with genome size (Lloyd et al., 2017), and machine learning approaches predict a substantial minority of these sequences are likely functional (Lloyd et al., 2017). And though evidence of functionality is difficult to come by for most species, substantially more loci associated with phenotypic variation are found in intergenic regions far from genes (Figure 1) in the large genome of maize than the small *Arabidopsis* genome.

We envision that new mutations in intergenic regions are most likely to impact function if they affect regulatory sequence. Regulatory sequences across the genome can be identified by their signature of open chromatin via nuclease (MNase-seq or DNase-seq) or transposase (ATAC-seq) accessibility (e.g. Oka et al., 2017). Evidence that open chromatin may be a useful proxy for functional sequence can be found in maize, where open chromatin makes up less than 1% of the genome but nucleotide variation in such regions explains more than 40% of phenotypic variation across traits (Rodgers-Melnick et al., 2016). For the additional mutational input in larger genomes to be functional, we expect to find more of the intergenic

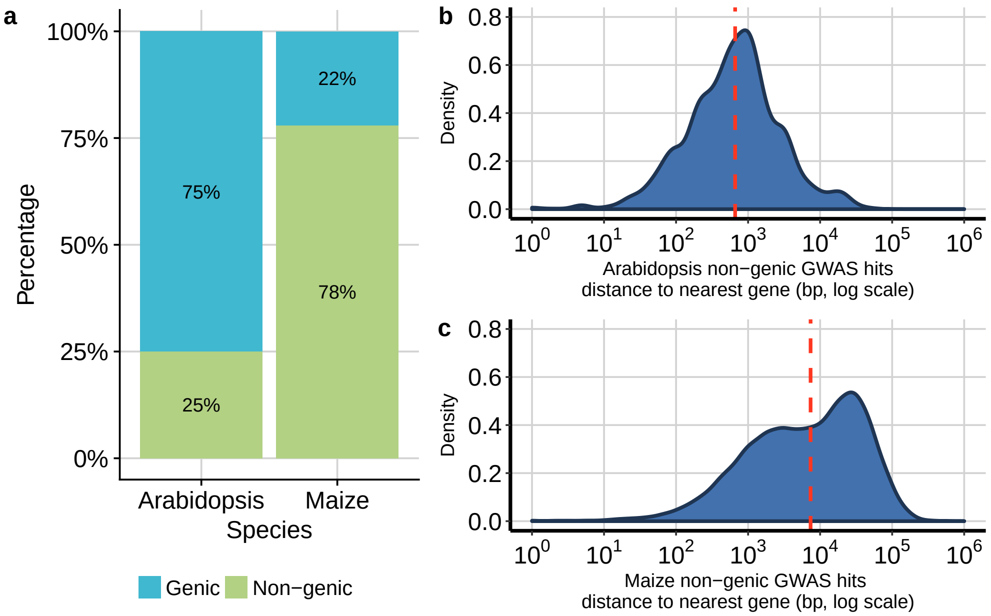

Figure 1: **Genomic distribution of loci identified in genome-wide association studies in *Arabidopsis* and maize.** *Arabidopsis* GWAS hits are top associations from the AraGWAS catalog (`https://aragwas.1001genomes.org`), and maize GWAS hits are the curated results of the nested association mapping population from maize diversity project (`https://www.panzea.org`) (Wallace et al., 2014). (**a**) Proportion of significant GWAS hits in genic (exon and intron) and non-genic regions. (**b-c**) Density plots of non-genic GWAS hits in (**b**) *Arabidopsis* and (**c**) maize. Dotted red lines indicate medians.

sequence in large genomes as open chromatin and ”functional space”. In line with these expectations, Maher et al. (2017) show that total accessible chromatin space increases as genome size increases across four species ranging almost an order of magnitude in genome size (p=0.11, adjusted $R^2 = 0.11$; Figure 2a), and that the largest increase in chromatin between species is in intergenic regions. Although chromatin accessibility data prepared in different labs and from different tissues makes direct comparison difficult, aggregated data from a broad range of species is also suggestive of increases in non-exonic open chromatin as genome size increases (Figure 2b).

In total, larger genomes have more mutational input, most of which occurs in intergenic regions. Some proportion of this mutational input leads to additional open chromatin, increasing the “functional space” of large genomes.

Large genomes adapt primarily through mutations in regulatory regions

One common approach to study adaptation polarizes synonymous and nonsynonymous substitutions using sequence comparison to an outgroup. Because nonsynonymous mutations are more likely to be functional, an observed excess of nonsynonymous substitions compared to expectations from synonymous sites likely reflects the effects of positive selection.

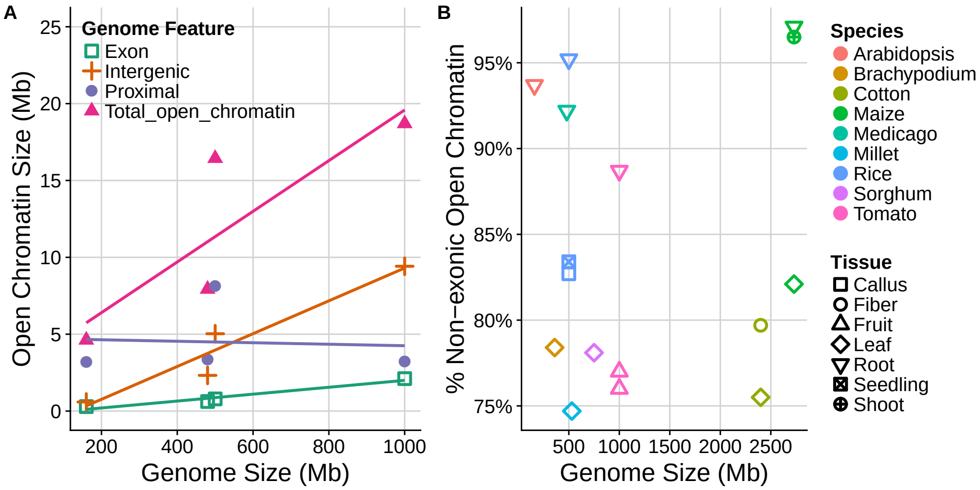

Figure 2: **Open chromatin and genome size.** (**a**) The total amount of open chromatin scales with genome size in four species *Arabidopsis thaliana*, *Medicago truncatula*, *Solanum lycopersicum* and *Oryza sativa* (Maher et al., 2017). Intergenic regions are defined >2kb from the transcription start site and >1kb from the transcription termination site. (**b**) Percentage of open chromatin outside of exons scales with genome size across many species. References and details of analyses for this figure can be found at `https://github.com/RILAB/AJB_MutationalTargetSize_GenomeSize`

The proportion of adaptive nonsynonymous substitutions (α) is used to infer the strength or efficiency of natural selection (Eyre-Walker, 2006). But as most of the additional functional mutations in large genomes occur in regulatory regions outside of genes (Figure 2), as genomes increase in size, a larger proportion of adaptive variation should occur outside genes and enrichment for nonsynonymous substitutions should decrease. Limited evidence for this effect can be found in two studies of environmental adaptation, where putatively selected loci are enriched for nonsynonymous mutations in the small *Arabidopsis* genome (Hancock et al., 2011) but for noncoding sequence near genes in the larger teosinte genome (Pyhäjärvi et al., 2013). We thus predict that, all else being equal, plants with larger genomes should exhibit lower α values than species with smaller genomes. After correcting for phylogeny, we indeed observe such a negative slope (p=0.12; Figure 3), and a negative correlation is also evident within the only two well-sampled genera available (*Helianthus* and *Pinus*). The effect we predict may be relatively weak overall, however, as a number of other factors such as effective population size and population structure (Gossmann et al., 2010) have also been suggested to impact α. Published data of α for species with small genome size are biased towards species with a small effective population size — and thus less efficient natural selection — which might influence the correlation between α and genome size in our meta-analysis (Figure 3).

Sweeps are softer in larger genomes

At the onset of a new selective pressure, adaptation can either proceed using genetic variation currently segregating in the population or by acting upon *de novo* beneficial mutations

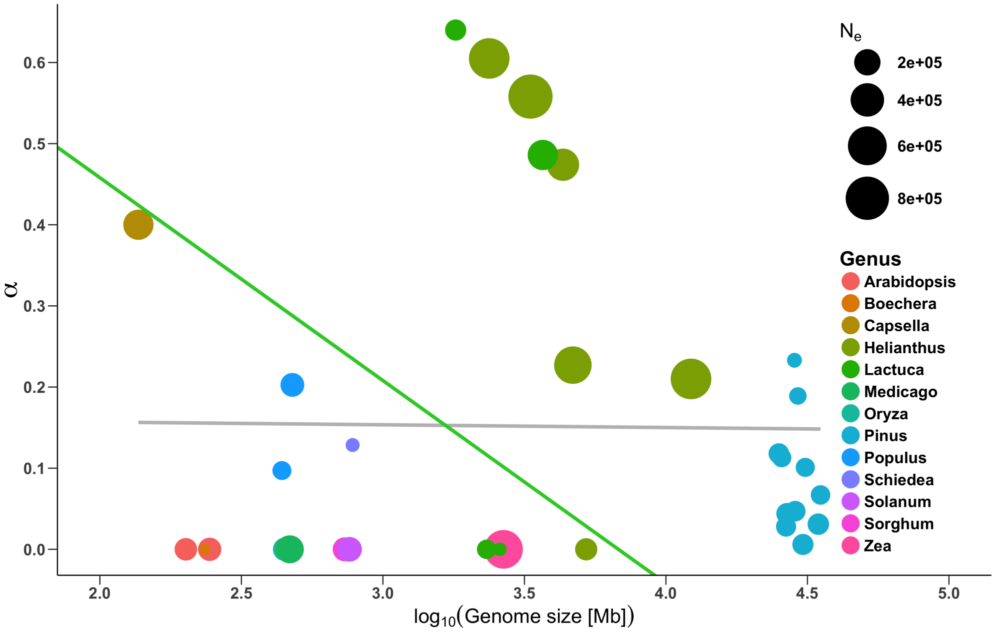

Figure 3: **Pattern of adaptive substitutions and genome size.** The proportion of adaptive substitutions (α) across plants with different genome sizes. Colors represent genera and point sizes show estimates of the effective population size (N_e). Lines represent regression analyses performed with (green) and without (black) phylogenetic correction. References and details of analyses for this figure can be found at `https://github.com/RILAB/AJB_MutationalTargetSize_GenomeSize`

as they enter the population. If mutations are limited, adaptation occurs predominantly by new beneficial mutations, which rapidly reach high frequency and leave a footprint of reduced local genetic diversity known as a "hard sweep" (Smith and Haigh, 1974). A number of other modes of adaptation are possible, however, and often collectively referred to as "soft sweeps". If substantial functional variation currently exists in the population, adaptation can make use of existing variants and affect surrounding genetic diversity to a lesser degree (Hermisson and Pennings, 2017). Similarly, if the influx of new beneficial mutations across multiple loci is sufficiently high, adaptation can occur via partial sweeps that increase the frequency — but not fix — multiple independent beneficial mutations (Hermisson and Pennings, 2017).

A key component in determining which of these processes take place as adaptation progresses is the mutational target size — what we term here more generally as "functional space". More formally, the likelihood of sweeps being hard or soft depends on the population mutation rate θ_b, which is the product of twice the effective population size (N_e) and the genome-wide beneficial mutation rate (U_b) (Hermisson and Pennings, 2017). While there is little evidence that the effective size of plant populations scales with genome size (Whitney et al., 2010), as we argue above the genome-wide rate of beneficial mutations should increase as genomes increase in size. Theory suggests that soft sweeps should predominate whenever $\theta_b > 1$ (Hermisson and Pennings, 2017), leading to the prediction that adaptation from extant variation or multiple independent mutations should be more common in larger genomes. Although empirical examinations of the amount of soft and hard sweeps across populations is still scarce, patterns of diversity around nonsyonymous substitutions in the small *Capsella grandiflora* genome suggests a predominance of hard sweeps (Williamson et al., 2014), while similar analyses in the larger maize genome find little evidence for hard sweeps around nonsynonymous substitutions (Beissinger et al., 2016).

Functional space and adaptation in plant genomes

Here we have proposed the **functional space hypothesis**, positing that mutational target size scales with genome size, impacting the number, source, and genomic location of beneficial mutations that contribute to adaptation. Though motivated by preliminary evidence, mostly from *Arabidopsis* and maize, more data are needed before any rigorous assessment of the hypothesis can be made. If correct, the functional space hypothesis suggests that we should expect plants with large genomes to exhibit more functional mutations outside of genes, more regulatory variation, and likely less signal of strong selective sweeps reducing diversity. These differences have implications for how we study the evolution and development of plant genomes, from where we should look for signals of adaptation to what patterns we expect adaptation to leave in genetic diversity or gene expression data. While flowering plant genomes vary across more than three orders of magnitude in size, most studies of both functional and evolutionary genomics have focused on species at the extreme small edge of this scale. Our hypothesis predicts that methods and results from these small genomes may not replicate well as we begin to explore large plant genomes. Finally, while we have focused here on evidence from plant genomes, we see no *a priori* reason why similar arguments might not hold in other taxa as well.

Acknowledgments

The authors wish to acknowledge funding support from the NSF (IOS-1238014) and the USDA (Hatch project CA-D-PLS-2066-H) as well as Roger Deal, Pär K Ingvarsson, Toni I. Gossmann, and Aurélien Tellier for access to raw data. We thank Graham Coop and Stephen Wright for helpful discussion and Peter Tiffin, Vince Buffalo, the Prost journal club, and numerous other colleagues for comments on an early draft of the paper.

References

Jon Ågren, Christopher G Oakley, John K McKay, John T Lovell, and Douglas W Schemske. Genetic mapping of adaptation reveals fitness tradeoffs in arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, 110(52):21077–21082, 2013.

Jill T Anderson, Cheng-Ruei Lee, Catherine A Rushworth, Robert I Colautti, and Thomas Mitchell-Olds. Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, 22(3):699–708, 2013.

Timothy M Beissinger, Li Wang, Kate Crosby, Arun Durvasula, Matthew B Hufford, and

Jeffrey Ross-Ibarra. Recent demography drives changes in linked selection across the maize genome. *Nature plants*, 2:16084, 2016.

Jeffrey L Bennetzen, Jianxin Ma, and Katrien M Devos. Mechanisms of recent genome size variation in flowering plants. *Annals of botany*, 95(1):127–132, 2005.

Paul Bilinski, Patrice S Albert, Jeremy J Berg, James Birchler, Mark Grote, Anne Lorant, Juvenal Quezada, Kelly Swarts, Jinliang Yang, and Jeff Ross-Ibarra. Parallel altitudinal clines reveal adaptive evolution of genome size in zea mays. *bioRxiv*, page 134528, 2017.

Edward B Chuong, Nels C Elde, and Cédric Feschotte. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics*, 2016.

Adam Eyre-Walker. The genomic rate of adaptive evolution. *Trends in ecology & evolution*, 21(10):569–575, 2006.

Paul Fransz, Gabriella Linc, Cheng-Ruei Lee, Saulo Alves Aflitos, Jesse R Lasky, Christopher Toomajian, Hoda Ali, Janny Peters, Peter Dam, Xianwen Ji, et al. Molecular, genetic and evolutionary analysis of a paracentric inversion in arabidopsis thaliana. *The Plant Journal*, 88(2):159–178, 2016.

Brandon S Gaut and Jeffrey Ross-Ibarra. Selection on major components of angiosperm genomes. *science*, 320(5875):484–486, 2008.

Emma E Goldberg and Boris Igić. Tempo and mode in plant breeding system evolution. *Evolution*, 66(12):3701–3709, 2012.

Toni I Gossmann, Bao-Hua Song, Aaron J Windsor, Thomas Mitchell-Olds, Christopher J Dixon, Maxim V Kapralov, Dmitry A Filatov, and Adam Eyre-Walker. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Molecular biology and evolution*, 27(8):1822–1832, 2010.

JP Grime and MA Mowforth. Variation in genome sizean ecological interpretation. *Nature*, 299(5879):151–153, 1982.

Angela M Hancock, Benjamin Brachi, Nathalie Faure, Matthew W Horton, Lucien B Jarymowycz, F Gianluca Sperone, Chris Toomajian, Fabrice Roux, and Joy Bergelson. Adaptation to climate across the arabidopsis thaliana genome. *Science*, 334(6052):83–86, 2011.

Joachim Hermisson and Pleuni S Pennings. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, 8(6):700–716, 2017.

Boris Igic and Jeremiah W Busch. Is self-fertilization an evolutionary dead end? *New Phytologist*, 198(2):386–397, 2013.

Cheng-Ruei Lee, Baosheng Wang, Julius P Mojica, Terezie Mandáková, Kasavajhala VSK Prasad, Jose Luis Goicoechea, Nadeesha Perera, Uffe Hellsten, Hope N Hundley, Jenifer Johnson, et al. Young inversion with multiple linked qtls under selection in a hybrid zone. *Nature Ecology & Evolution*, 1:0119, 2017.

Damon Lisch. How important are transposons for plant evolution? *Nature Reviews. Genetics*, 14(1):49, 2013.

John P Lloyd, Zing TY Tsai, Rosalie P Sowers, Nicholas L Panchy, and Shin-Han Shiu. Defining the functional significance of intergenic transcribed regions based on heterogeneous features of phenotype genes and pseudogenes. *bioRxiv*, page 127282, 2017.

David B Lowry and John H Willis. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS biology*, 8(9):e1000500, 2010.

Michael Lynch and John S Conery. The origins of genome complexity. *science*, 302(5649): 1401–1404, 2003.

Michael Lynch, Matthew S Ackerman, Jean-Francois Gout, Hongan Long, Way Sung, W Kelley Thomas, and Patricia L Foster. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, 17(11):704–714, 2016.

Kelsey A Maher, Marko Bajic, Kaisa Kajala, Mauricio Reynoso, Germain Pauluzzi, Donnelly A West, Kristina Zumstein, Margaret Woodhouse, Kerry Bubb, Mike Dorrity, et al. Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *bioRxiv*, page 167932, 2017.

Itay Mayrose, Shing H Zhan, Carl J Rothfels, Karen Magnuson-Ford, Michael S Barker, Loren H Rieseberg, and Sarah P Otto. Recently formed polyploid plants diversify at lower rates. *Science*, 333(6047):1257–1257, 2011.

Michele Morgante, Stephan Brunner, Giorgio Pea, Kevin Fengler, Andrea Zuccolo, and Antoni Rafalski. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nature genetics*, 37(9):997, 2005.

J Alberto Romero Navarro, Martha Willcox, Juan Burgueño, Cinta Romay, Kelly Swarts, Samuel Trachsel, Ernesto Preciado, Arturo Terron, Humberto Vallejo Delgado, Victor Vidal, et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nature genetics*, 49(3):476–480, 2017.

Rurika Oka, Johan Zicola, Blaise Weber, Sarah N Anderson, Charlie Hodgman, Jonathan I Gent, Jan-Jaap Wesselink, Nathan M Springer, Huub CJ Hoefsloot, Franziska Turck, et al. Genome-wide mapping of transcriptional enhancer candidates using dna and chromatin features in maize. *Genome biology*, 18(1):137, 2017.

Gina M Pham, Linsey Newton, Krystle Wiegert-Rininger, Brieanne Vaillancourt, David S Douches, and C Robin Buell. Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *The Plant Journal*, 2017.

Julien Prunier, Sébastien Caron, and John MacKay. Cnvs into the wild: screening the genomes of conifer trees (picea spp.) reveals fewer gene copy number variations in hybrids and links to adaptation. *BMC genomics*, 18(1):97, 2017.

Tanja Pyhäjärvi, Matthew B Hufford, Sofiane Mezmouk, and Jeffrey Ross-Ibarra. Complex patterns of local adaptation in teosinte. *Genome biology and evolution*, 5(9):1594–1609, 2013.

Peter Ralph and Graham Coop. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics*, 186(2):647–668, 2010.

Eli Rodgers-Melnick, Daniel L Vera, Hank W Bass, and Edward S Buckler. Open chromatin reveals the functional maize genome. *Proceedings of the National Academy of Sciences*, 113(22):E3177–E3184, 2016.

John Maynard Smith and John Haigh. The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1):23–35, 1974.

Douglas E Soltis, Victor A Albert, Jim Leebens-Mack, Charles D Bell, Andrew H Paterson, Chunfang Zheng, David Sankoff, P Kerr Wall, Pamela S Soltis, et al. Polyploidy and angiosperm diversification. *American journal of botany*, 96(1):336–348, 2009.

Jared L Strasburg, Nolan C Kane, Andrew R Raduski, Aurélie Bonin, Richard Michelmore, and Loren H Rieseberg. Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. *Molecular biology and evolution*, 28(5): 1569–1580, 2010.

Jared L Strasburg, Natasha A Sherman, Kevin M Wright, Leonie C Moyle, John H Willis, and Loren H Rieseberg. What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587):364–373, 2012.

Peter Tiffin and Jeffrey Ross-Ibarra. Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, 29(12):673–680, 2014.

Jason G Wallace, Peter J Bradbury, Nengyi Zhang, Yves Gibon, Mark Stitt, and Edward S Buckler. Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet*, 10(12):e1004845, 2014.

Kenneth D Whitney, Eric J Baack, James L Hamrick, Mary Jo W Godt, Brian C Barringer, Michael D Bennett, Christopher G Eckert, Carol Goodwillie, Susan Kalisz, Ilia J Leitch, et al. A role for nonadaptive processes in plant genome size evolution? *Evolution*, 64(7): 2097–2109, 2010.

Thomas Wicker, Klaus FX Mayer, Heidrun Gundlach, Mihaela Martis, Burkhard Steuernagel, Uwe Scholz, Hana Šimková, Marie Kubaláková, Frédéric Choulet, Stefan Taudien, et al. Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *The Plant Cell Online*, 23(5): 1706–1718, 2011.

Robert J Williamson, Emily B Josephs, Adrian E Platts, Khaled M Hazzouri, Annabelle Haudry, Mathieu Blanchette, and Stephen I Wright. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of capsella grandiflora. *PLoS genetics*, 10(9):e1004622, 2014.

Sam Yeaman, Kathryn A Hodgins, Katie E Lotterhos, Haktan Suren, Simon Nadeau, Jon C Degner, Kristin A Nurkowski, Pia Smets, Tongli Wang, Laura K Gray, et al. Convergent local adaptation to climate in distantly related conifers. *Science*, 353(6306):1431–1433, 2016.

Agnieszka Żmieńko, Anna Samelak, Piotr Kozłowski, and Marek Figlerowicz. Copy number polymorphism in plant genomes. *Theoretical and applied genetics*, 127(1):1–18, 2014.