Family quarrels in seeds and rapid adaptive evolution in *Arabidopsis*

Katherine S. Geist[a], Joan E. Strassmann[a,b], and David C. Queller[a,b,1]

[a]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130; and [b]Wissenschaftskolleg zu Berlin–Institute for Advanced Study, 14193 Berlin, Germany

Edited by Raghavendra Gadagkar, Indian Institute of Science, Bangalore, India, and approved March 29, 2019 (received for review October 15, 2018)

Evolutionary conflict can drive rapid adaptive evolution, sometimes called an arms race, because each party needs to respond continually to the adaptations of the other. Evidence for such arms races can sometimes be seen in morphology, in behavior, or in the genes underlying sexual interactions of host–pathogen interactions, but is rarely predicted a priori. Kin selection theory predicts that conflicts of interest should usually be reduced but not eliminated among genetic relatives, but there is little evidence as to whether conflict within families can drive rapid adaptation. Here we test multiple predictions about how conflict over the amount of resources an offspring receives from its parent would drive rapid molecular evolution in seed tissues of the flowering plant *Arabidopsis*. As predicted, there is more adaptive evolution in genes expressed in *Arabidopsis* seeds than in other specialized organs, more in endosperms and maternal tissues than in embryos, and more in the specific subtissues involved in nutrient transfer. In the absence of credible alternative hypotheses, these results suggest that kin selection and conflict are important in plants, that the conflict includes not just the mother and offspring but also the triploid endosperm, and that, despite the conflict-reducing role of kinship, family members can engage in slow but steady tortoise-like arms races.

kin selection | arms race | molecular evolution | parent–offspring conflict | endosperm

Evolutionary arms races (1, 2) have been documented for strong conflicts between hosts and pathogens (3, 4) and between males and females (5). The mother–offspring relationship is largely amicable, with the mother ensuring the success of her own genes by helping her offspring. However, some conflict is predicted (6, 7), although the conflict is reduced by kinship, so it might be weaker and harder to detect. Mothers are equally related to all their offspring and should help one of them only when the benefit to it exceeds the cost to other offspring. However, each offspring is more related to itself than to its siblings, so it should therefore try to acquire resources in excess of the maternal optimum. There is some evidence that genes expressed in mammalian placentas, which function to provision embryos and are genetically identical to them, evolve rapidly (8). Seeds offer a special opportunity to test within-family conflict theory (9–14). In flowering plants, seeds contain the embryo, a covering of maternal tissue, and the endosperm (Fig. 1). The endosperm does most of the acquisition of resources from the mother and sometimes also stores the resources (15, 16), presumably allowing the embryo to specialize more in developing properly. In most angiosperms, the endosperm is triploid, identical to the embryo but with an extra dose of the maternal alleles, which gives it its own peculiar relatedness patterns (9–14).

Fig. 2 shows how kin selection is predicted to operate on mothers, endosperms, and embryo with respect to transfer of resources to this embryo instead of to other embryos on the same maternal plant (9, 10). There is a large zone of potential conflict where an embryo and its endosperm favor this transfer but the mother does better to provision her other embryos. There is a much smaller zone of conflict where the endosperm is predicted to side with the mother against the embryo. Note also that, if two nonrelatives were selected with respect to providing a benefit to one at a cost to the other, they would be in conflict over the entire positive benefit–cost space in Fig. 2, so relatedness is a moderating factor that reduces conflicts within families.

However, evidence for seed conflict has been indirect. For example, there are possible morphological features consistent with conflict, such as invasive haustoria of endosperms and maternal integumental barriers (9, 11). There is also evidence for paternal effects on seed size (17). Here we present strong tests of the prediction that conflict among these tissues in *Arabidopsis* will lead to high rates of adaptive evolution.

Conflict-based arms races are most likely to be found in genes that are specialized for in tissues engaged in conflict. We therefore first identified sets of *Arabidopsis* genes specialized for a focal organ or tissue as those genes that show significantly higher expression in that organ or tissue compared with other organs or tissues, using published microarray expression datasets (18, 19). We then compared rates of adaptive evolution (α) of these gene sets using $\alpha = 1 - \left(\frac{Pn}{Ps}/\frac{Dn}{Ds}\right)$, where Dn/Ds is the ratio of nonsynonymous to synonymous substitutions between species and Pn/Ps is the corresponding ratio for polymorphisms within species (20–22). This statistic is based on the logic of the McDonald–Kreitman test (23): The two ratios should be the same under a combination of neutrality and purifying selection, but positive selection will elevate Dn/Ds, and therefore α; α provides a

Significance

Evolutionary conflict, such as between pathogens and hosts, can lead to arms races in which each party evolves rapidly in response to the harm inflicted by the other. Kin selection makes relatives much more cooperative, but some conflict is usually still expected. We show that even this reduced conflict appears to drive arms races in seeds of the plant *Arabidopsis*, which contain three genetic relatives: maternal tissue, the embryo, and the triploid endosperm. As expected from potential conflict over how much nutrition the embryo should receive from the mother, genes expressed in seed tissues evolve rapidly, particularly the parts directly involved in nutrient transfer. Moreover, the endosperm appears to have largely taken over the embryo's role in this parent–offspring conflict.

Author contributions: K.S.G., J.E.S., and D.C.Q. designed research; K.S.G. performed research; K.S.G. analyzed data; and K.S.G., J.E.S., and D.C.Q. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Data and authored programs, including gene lists for each category, a measure of each gene's degree of adaptive evolution, statistics computed for each gene set, and the authored Perl wrapper that performs bootstrapping and permutation tests on gene sets using DFE-α, are available at https://github.com/ksgeist/adaptation-in-arabidopsis-seeds.

[1]To whom correspondence should be addressed. Email: queller@wustl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1817733116/-/DCSupplemental.

Published online April 25, 2019.

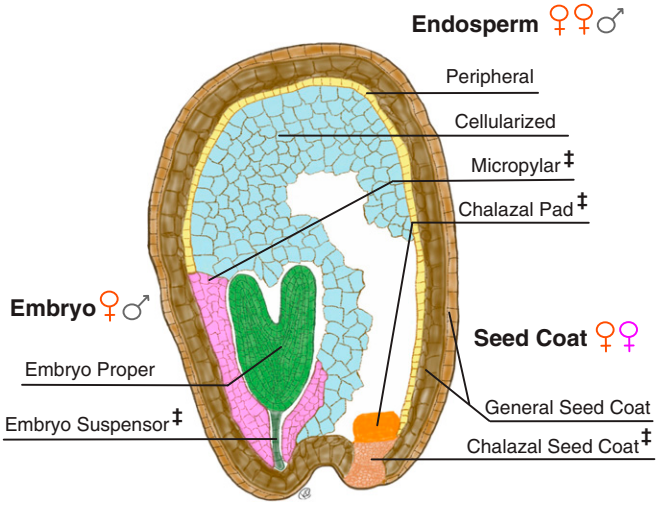

Fig. 1. Components of the angiosperm seed. The developing angiosperm seed, here *Arabidopsis*, has three distinct genetic parties. The mother contributes the seed coat, with the chalazal region being a portal for nutrients. The offspring consists of the embryo proper and a temporary suspensor that may be involved in nutrient acquisition. The endosperm results from a second fertilization and is triploid, genetically identical to the embryo except for an extra set of the maternally derived chromosomes. Nutrients flow to the endosperm through its chalazal region. ‡, Subtissues most involved in transfer between genetically distinct parties.

powerful summary of adaptive evolution averaged over a gene set, and, when two gene sets share the same population history, a higher value of α indicates greater adaptive evolution.

We estimated α in two species pairs, (*i*) *Arabidopsis thaliana* populations with an *Arabidopsis lyrata* outgroup and (*ii*) *A. lyrata* populations with an *Arabidopsis halleri* outgroup (Fig. 3). The first pairing is natural because the expression data come from *A. thaliana*, but there are two complications which we can remove with the second test. First, *A. thaliana* seeds are smaller than those of *A. lyrata* [0.3 mm to 0.5 mm vs. 0.8 mm to 1.2 mm (24)], meaning any excess adaptation observed in seeds could result not just from conflict but from any factor selecting for seed size. However, this is not an issue for *A. lyrata* and *A. halleri*, which have very similar seed sizes [both 0.8 mm to 1.2 mm (24)]. Second, *A. thaliana* is inbred, which changes the relatedness patterns and should reduce conflict (9). This should not be a serious problem, because *A. thaliana* has been outbred for more than 90% of the time since it diverged from *A. lyrata* (25), but the *lyrata–halleri* pairing provides a check with an outbred pair. There are polymorphism data from multiple populations, from which we selected five *A. thaliana* populations and two *A. lyrata* populations for analysis, allowing seven partially independent tests of each prediction (independent polymorphism data but shared outgroup for divergence).

Results

We test conflict predictions at three levels using genes preferentially expressed in organs, in seed tissues, and in seed subtissues (*SI Appendix*, Table S1). First, if there is sufficient family conflict in seeds, then genes with specialized expression in seeds should show more adaptive evolution (higher α) than genes specialized for other organs that are genetically uniform and therefore not subject to conflict. This prediction is successful, with seed genes showing higher α than genes in floral buds, leaf rosettes, stems, and roots in both species pairs; 25 of the 28 comparisons are statistically significant (Fig. 3; all adaptive evolution tests in this paper are permutation tests; see *Methods*).

Second, because the endosperm has taken over the primary nutrient acquisition role for the embryo, we test whether the primary conflict is now between mother and endosperm rather than mother and embryo. This shift would make sense (9) because the endosperm is far less constrained than the embryo, which has to develop in a precise way, and the two tissues differ little in their interests [Fig. 2; even this difference disappears for endosperm genes that are strictly dominant or recessive (10)]. Again, the data support the prediction. Compared with genes with specialized expression in the embryo, we see elevated rates of adaptive evolution in genes specialized for maternal seed coat (four of seven comparisons are significant) and especially for the endosperm (seven of seven comparisons are significant) (Fig. 4, *Left*).

Third, even more informative predictions can be tested within each tissue—maternal, endosperm, and embryo. Genes with specialized expression in subtissues that are most involved in nutrient transfers are predicted to be more engaged in conflict and to have higher α than other subtissues. Most of the actual conflict between mother and endosperm should occur in their chalazal regions where more nutrients are transferred (15, 16) so the conflict hypothesis predicts that these will evolve more rapidly than other subtissues in their respective tissues. This prediction is confirmed both for maternal chalazal seed coat versus the maternal general seed coat (six of seven tests; Fig. 4, *Right*) and for endosperm chalazal pad versus cellularized/peripheral endosperm (seven of seven tests, Fig. 4, *Right*). Two other subtissues are predicted to evolve rapidly only if the embryo still participates in some conflict: the embryo suspensor that is terminally differentiated and participates in nutrient transfer (26) and the micropylar endosperm that surrounds the embryo (15, 16). These predictions are also confirmed, though less strongly, with somewhat higher adaptive evolution in genes with specialized expression in the embryo suspensor versus those in the embryo proper (four of seven tests) and in genes in the micropylar endosperm versus those in the cellularized/peripheral endosperm (five of seven tests) (Fig. 4, *Right*).

Gene ontology (GO) analyses show that extracellular and intracellular communication genes figure prominently, as one might predict under conflict, but a number of other categories are also significantly enriched (*SI Appendix*, Tables S2–S6).

Discussion

The evidence strongly supports multiple predictions of greater adaptive evolution expected from conflict in seeds. Alternative hypotheses cannot account for all of the results. First, the higher α in seeds could reflect an arms race between the maternal seed coat on the outside of the seed against evolving seed predators or soil pathogens and fungi. However, this is not supported by the similarly high α in the endosperm and especially not by the lower α in the general seed coat surrounding the seed versus the chalazal seed coat supplying the nutrients (Fig. 4, *Right*). Second, imprinted genes, which affect seed nutrition, might add still another dimension of kin-selected conflict, that between mother and father (27). We are conducting a separate analysis of imprinted genes, but they constitute small fractions of our gene sets and cannot explain all of the patterns observed. Endosperm tissues show a pattern opposite to an imprinting arms race: Genes preferentially expressed in the slowly evolving cellularized/peripheral region are more often imprinted (8.1%) than those expressed in the rapidly evolving micropylar (0.5%) or chalazal (2.2%) regions [based on 124 stringent, imprinted genes (28)]. Moreover, in maternal tissues, there should be no imprinting conflict, so this hypothesis cannot explain either the high α in the seed coat genes or the higher α in genes in the chalazal region of the seed coat. Finally, the *lyrata–halleri* comparison removes a seed-size selection explanation. However, conflict is always over something—here provisioning—that might be selected for nonconflict reasons, so our results are also consistent with a post hoc hypothesis of nonconflict selection on provisioning. However, some such post hoc explanation could be posited for any

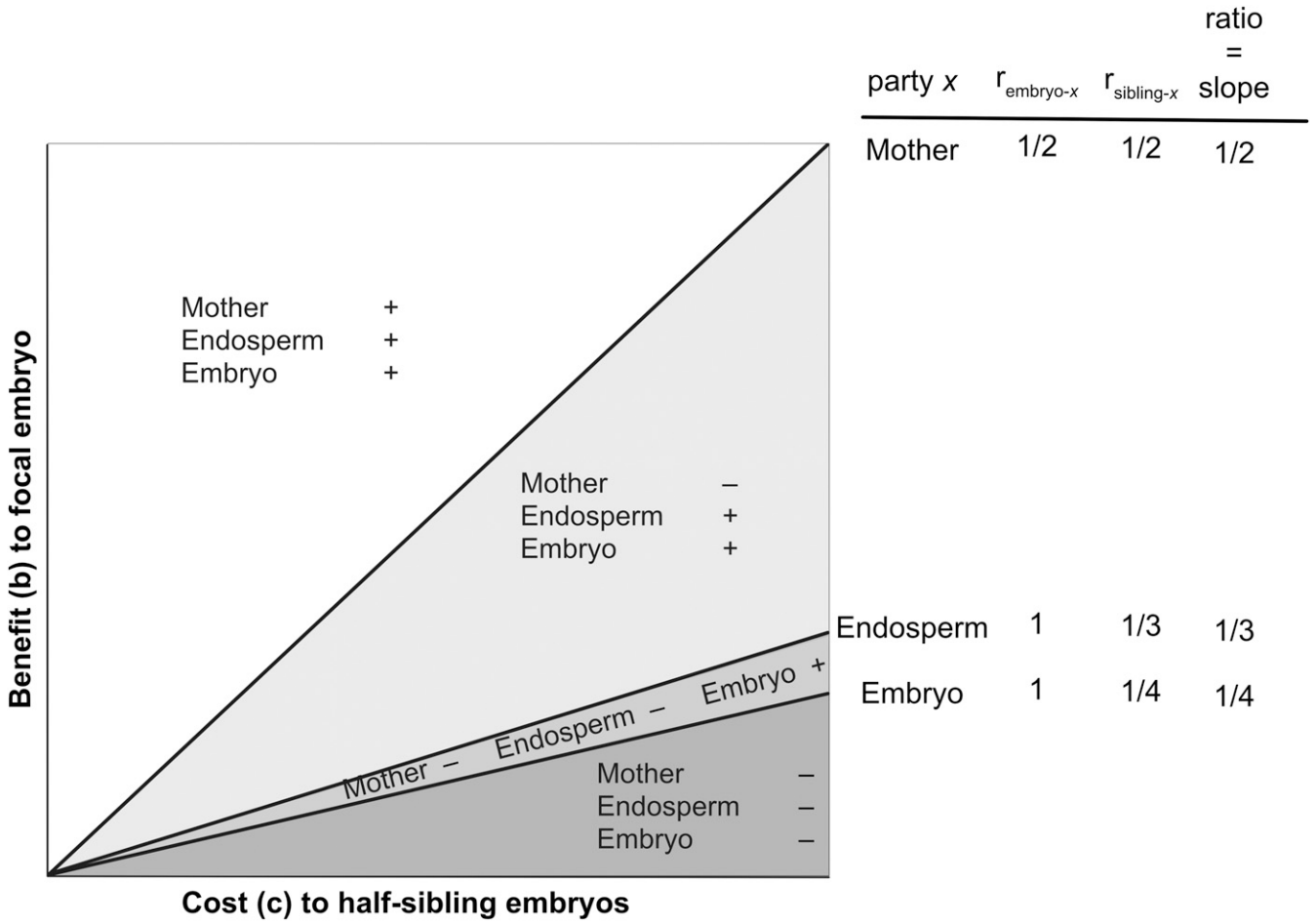

Fig. 2. Conflict in seeds. Suppose an allele is expressed in tissue x (x = embryo, endosperm, mother) of a focal seed. It increases nutrient flow to the embryo of this seed, increasing its fitness by b (y axis) and decreasing the total fitness of its current or future maternal half siblings by c (x axis). Hamilton's kin selection rule (29) states that this allele will be favored when $r_{embryo\text{-}x}b - r_{sibling\text{-}x}c > 0$, where the two rs are the relatednesses of tissue x to the focal embryo and to half-sibling embryos (9). Therefore, each party should favor this transfer (+) when $b > c{*}r_{sibling\text{-}x}/r_{embryo\text{-}x}$, and disfavor it (–) when this inequality is reversed. If b/c is high enough (white), all parties favor the focal embryo, and, if it is low enough, none do (dark gray). In between, there are zones of potential conflict where some tissues would gain from the transfer and others would lose from it.

of the over 600 possible patterns of significance rankings (*SI Appendix*, Table S7) of the five plant organs tested (Fig. 3), and more if we included the patterns of Fig. 4. In contrast, the conflict hypothesis predicted a priori the single pattern actually observed (seeds show more adaptive evolution than the other four), passing a severe attempt at falsification.

In the absence of viable alternative explanations, these results suggest a number of implications. They add support to Hamilton's (29) assertion that kin selection is important far beyond its canonical applications to the evolution of altruism in animals such as social insects, in this case, to plants (8–14, 30). They also add support to the idea that kin selection is relevant not just to driving altruism but also to limiting selfish behavior and conflicts. Our results—together with parallel ones on rodents (8)—also provide support for parent–offspring conflict in general and against the idea (31) that parents should completely win because they have initial control of the contested resources. If that were true, there would be no ongoing conflict and elevated rates of adaptation. Our results suggesting pronounced endosperm conflict with the mother and weaker conflict with the embryo provide support for the idea that the peculiar triploid endosperm, which never lives independently or reproduces directly, evolves according to its own relatedness-based interests. They therefore lend credence to kin-selection theories of the origin and evolution of the endosperm (9–13, 27, 32). Finally, the idea that some parts of the seed have evolved to increase seed size and others have evolved to moderate seed size is likely relevant to strategies for artificially selecting seeds, such as in the cereals that constitute a large part of the human diet.

Relatedness is expected to decrease conflict, so it is interesting that kin interactions nevertheless seem to drive rapid evolution, consistent with an evolutionary arms race. One reason may be the constancy of the conflict. In Aesop's fable, a tortoise raced against a hare, but arms races can pair two hares or two tortoises. Hosts and pathogens may be hares, with selection that is strong but irregular because not every host encounters a pathogen and also because host–pathogen species pairings shift (2). In contrast, family quarrels may resemble races among tortoises. The pace may be slower, but it never wanes, because every offspring has a mother and every evolutionary successful mother has offspring.

Methods

Genes Specialized for Particular Organs and Tissues. From two published *A. thaliana* microarray expression datasets, we identified genes specialized for seed tissues (19) (series GSE12404; *SI Appendix*, Table S1), as well as for seeds and the following nonseed organs: floral bud, leaf rosette, stem, and root (18) (series GSE680; *SI Appendix*, Table S1). The seed expression dataset (19) (series GSE12404) includes time series Affymetrix ATH1 microarray data across six developmental stages from microdissected seed tissues and from

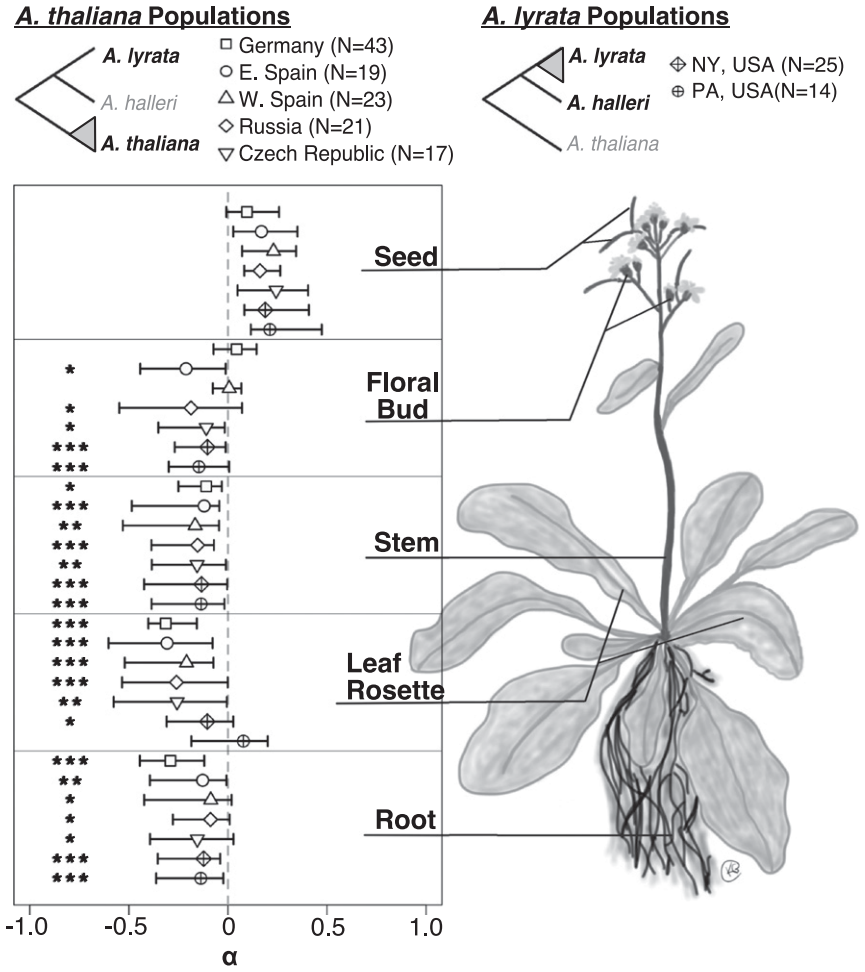

Fig. 3. Adaptive evolution α is higher for genes upregulated in seeds compared with other organs. Each panel shows seven estimates of the rate of adaptive evolution, α, for five *A. thaliana* populations with an *A. lyrata* outgroup, as well as two *A. lyrata* populations with an *A. halleri* outgroup. For floral buds, stems, leaf rosettes, and roots, asterisks show significant differences from the corresponding seed (permutation tests; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$).

two to three biological replicates. The life stages expression dataset [GSE680 (18)] includes time series ATH1 microarray data from seeds across the same six developmental stages, as well as mature plant organs at single time points. For each microarray experiment, we extracted normalized $\log_2$-transformed expression values for each mRNA sequence with the Robust Multiarray Average preprocessing approach (33) using the Affy package (34) in Bioconductor v.2.12 (35) implemented in R v.2.15 (36).

We used the limma package (37) to identify which RNA sequences were significantly enriched in each organ or tissue that we used in subsequent analyses. We performed pairwise contrasts of a focal tissue against other tissue(s), as specified in the next three paragraphs, with *t* tests moderated by an empirical Bayes function, because there were few replicates available for each microarray experiment. The mRNA transcripts with a Benjamini–Hochberg (38) adjusted *P* value less than 0.01 were considered enriched for a focal tissue. We assigned gene identities to mRNA transcripts using a reannotated array based on the TAIR10 *A. thaliana* genome release (https://www.arabidopsis.org). Transcripts that mapped to more than one sequence were excluded from further analyses. When multiple transcripts mapped to the same gene, we required all of them to be significantly enriched in the focal tissue.

We applied these procedures to perform contrasts to identify gene sets specialized for particular organs or tissues at three different levels. First, to identify genes specialized for particular organs, we performed pairwise contrasts to identify genes encoding mRNAs enriched in each of the following organs: seed, floral bud, leaf rosette, stem, and root [data from series GSE680 (18)]. For the seed genes in this analysis, expression data were averaged across ontogeny.

Second, to identify genes specialized for particular seed tissues, we identified genes significantly enriched in mRNA expression, in at least one developmental stage, in each of the three seed tissues (maternal seed coat, endosperm, and embryo) relative to the other two tissues [data from series GSE12404 (19)]. From these gene sets, we deleted any genes previously found to have significantly enriched expression in floral bud, leaf, stem, or root organs to limit any effects of selection on genes during those stages of the plant life cycle.

Third, to identify genes specialized for particular subtissues, we identified genes with significantly enriched mRNA expression in each of the seed subtissues within the seed coat, embryo, and endosperm tissues. We compared subtissues within a given tissue only. For example, genes specialized for chalazal endosperm were those with significantly enriched expression in the chalazal region of the endosperm, relative to the other three endosperm regions. Again, genes were chosen when their expression was enriched in at least one developmental stage of the focal subtissue but not in any other subtissue. We again deleted any genes in the set previously found to have enriched expression in floral bud, leaf, stem, or root organs. Because of small sample sizes, limited presence in stages, shared ontogenetic origins (16), and shared predictions, we combined the gene sets for "cellularized" and "peripheral" endosperms.

Tests for Molecular Signatures of Positive Selection in Tissue-Specific Genes. We used both interspecific and intraspecific sequence comparisons to test for positive selection in plant organs, seed tissues, and subtissues. For the gene sets specialized in each, we estimated the proportion of adaptive substitutions as $\alpha = 1 - \left(\frac{Pn}{Ps}/\frac{Dn}{Ds}\right)$, where *Pn* and *Ps* are the numbers of nonsynonymous and synonymous polymorphisms within species, respectively, and *Dn* and *Ds* are the numbers of nonsynonymous and synonymous differences between species (21, 39). The metric α is an extension of the McDonald–Kreitman test (23), which assumes that, if an adaptive mutation arises, it is swept to fixation quickly, contributing to between-species divergence but not within-species polymorphism. Thus, if $Dn/Ds > Pn/Ps$, the sequence is thought to be under strong positive selection. The McDonald–Kreitman test typically looks at a single gene, limiting sample size and power, but α is calculated cumulatively across a class of genes. Thus, no single gene in the set need be significant under the McDonald–Kreitman criteria to detect adaptation. We first estimated α with polymorphism counts from *A. thaliana* and with divergences

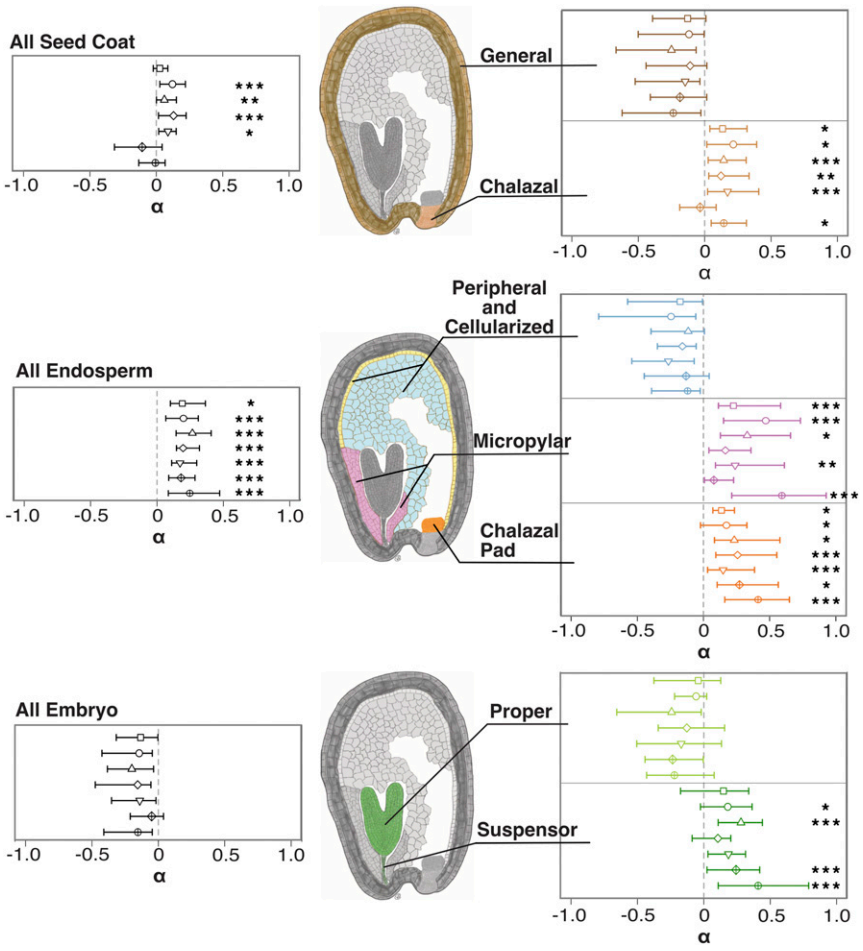

Fig. 4. Adaptive evolution α in genes specialized for different seed tissues and subtissues. (*Left*) Adaptive evolution in the genes specialized for maternal seed coat and endosperm is higher than in those specialized for embryo, supporting the hypothesis that most of the conflict is between mother and endosperm. Asterisks indicate significance relative to the same-population embryo α. (*Right*) Within each of the maternal, endosperm, and embryo tissues, adaptive evolution is higher in genes specialized for subtissues more directly involved in nutrient transfers. Asterisks indicate significant differences relative to same-population α of the subtissue(s) less involved in nutrient transfer (maternal general seed coat, peripheral and cellularized endosperm, embryo proper). The seven populations in each panel are as in Fig. 3 (permutation tests; $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$).

from its sister species, *A. lyrata*. We also estimated α for a pair of two closely related outbred species, *A. lyrata* and *A. halleri*, for which we obtained publicly available polymorphism data for *A. lyrata* (40).

Pairwise divergence estimates. To estimate divergence between *A. thaliana* and *A. lyrata*, or between *A. lyrata* and *A. halleri*, we first identified whole-genome orthologs using our version of standalone InParanoid v.4 (41), which we updated to work with BLAST+ (42). This gave us reciprocal BLAST comparisons of *A. thaliana*, *A. lyrata* (*v*. 1.0) (43), and *A. halleri* (*v*. 1.0) (44) protein sequences. With the 1:1 orthologs, we performed pairwise global alignments of *A. thaliana*, *A. lyrata*, and *A. halleri* proteins using MUSCLE (45), which were again back-translated and trimmed using PAL2NAL v. 14 (46) and trimAl v. 1.2 (47), respectively. We then used the Nei and Gojobori method (48) implemented in the codeml package of PAML 4.0 (runmode = −2, CodonFreq = 2) (49) to estimate the numbers of nonsynonymous and synonymous sites and substitutions per gene between each species pair. We excluded all genes with saturated divergence ($dS \geq 1$; *A. thaliana*–*A. lyrata*, 61 genes; *A. lyrata*–*A. halleri*, 37 genes) from future tests.

***Polymorphism estimates of* A. thaliana *and* A. lyrata.** To reduce the chance of unusual results owing to an unusual recent population history, we estimated within-species polymorphism for five populations of *A. thaliana* and two populations of *A. lyrata*. To obtain *Pn* and *Ps* for *A. thaliana*, we used SNP data from resequenced *A. thaliana* genomes as part of the 1,001 Genomes Project (www.1001genomes.org, *SI Appendix*, Table S2). We chose five populations, each consisting of geographically clustered accessions, to minimize any effects of population structure: Germany ($n = 43$), Czech ($n = 17$), Russia ($n = 21$), E. Spain ($n = 18$), and W. Spain ($n = 23$). We converted the variant call format (VCF) files for each *A. thaliana* individual in each population into a variant FASTA sequence file of the *A. thaliana* reference genome (TAIR10, downloaded May 7, 2012) with a custom Perl script. We used BEDTools (50) to extract the coding sequences for each gene and translated these to amino acids with a custom Perl script. We aligned, back-translated, and trimmed misaligned regions using the same methods described for estimating divergence. We analyzed the coding alignments with PolyMORPHOrama without a minor allele frequency cutoff.

For *A. lyrata*, we obtained pooled resequenced genome data and genotype-by-sequencing (GBS) data for the two available populations, one collected from Erie, PA ($n = 14$), and the other from Jamesville, NY ($n = 25$) (40). To obtain our polymorphism counts for the two available populations of *A. lyrata*, we began with all GBS and pooled-sequencing FASTQ files (European Nucleotide Archive: https://www.ebi.ac.uk/ena, accession PRJEB8335). These FASTQ were deinterleaved and had been demultiplexed (GBS) and trimmed to a minimum PHRED quality score of 20 before they were added to the repository. We merged the FASTQ sequence quality files from multiple lanes of pooled sequence for the New York population to increase coverage. We then mapped all paired end reads to version 1.0 of the *A. lyrata* reference genome (43) with *sampe* in the Burrows–Wheeler Aligner v. 0.7.15 (51). We sorted and indexed the alignment files with SAMTOOLS v. 1.3 (51–53), and then realigned insertions/deletions with the Genome Analysis ToolKit v.3.3.0 (54), removed low-quality reads (<20) and those that failed to map with SAMTOOLS, and removed duplicate reads with Picard (v. 1.128; broadinstitute.github.io/picard/). Because we were only interested in polymorphism in coding sequence, we called variants using the primary coding sequence of *A. lyrata* with Genome Analysis Toolkit HaplotypeCaller (54) with a quality score of >25 and filtered for SNPs only. To ensure optimal coverage across all coding sequence genome-wide, we merged all resulting VCF files by population with the vcf-merge tool of VCFTOOLS v. 1.14 (55). Our same custom Perl script converted the merged VCF files to a variant FASTA, which we then used to extract the coding sequences for each gene and calculate nonsynonymous and synonymous polymorphism counts as described for the *A. thaliana* populations.

***The proportion of adaptive substitutions as estimated by* α.** We estimated the proportion (α) and rate (ω_a) of adaptive substitutions and the proportion of nonsynonymous mutations that are deleterious ($1 - f$) for our different gene sets using the standalone version of the Distribution of Fitness Effects (DFE)-α program v. 2.15 (22). Using a custom Perl wrapper that sums and formats the nonsynonymous and synonymous polymorphism frequency spectra provided

by PolyMORPHOrama creates all necessary run files, incorporates divergence information, and performs either permutations or bootstrapping. In DFE-α, we used default parameters, except that we used a two-epoch model without folded site-frequency spectra and a Jukes–Cantor correction when calculating nucleotide divergence. Results are shown in *SI Appendix*, Table S1.

Statistics. For each focal gene set, we generated confidence intervals for α, ω_a, and $1 - f$ by bootstrapping across loci 1,000 times. For each parameter X, i genes in the focal set were randomly drawn with replacement 1,000 times, from which we reran DFE-α for each resample, recomputed X. These data were used for a 95% confidence interval.

To ask whether focal gene sets differed from each other, we employed permutations to test for differences between two samples. To test for a difference between the statistics of two gene sets, X_1 and X_2 with i and j numbers of genes, we randomly drew without replacement i and j genes from combined sets of genes, reran DFE-α, and recalculated $X_1 - X_2$. We calculated the P value as the proportion of times the permuted difference was greater than zero in the direction predicted. All P values reported are one-tailed.

GO. We performed a GO analysis to rudimentarily examine the functions of the genes in our focal tissues. We used the DAVID Functional Annotation Clustering Tool (56, 57) to obtain clusters of significantly enriched GO terms with a stringency setting of "High" for all gene sets, using the *A. thaliana* genome as background. Clusters for each gene set are given in *SI Appendix*, Tables S2–S6.

Data and Code Availability. Data and authored programs have been archived at https://github.com/ksgeist/adaptation-in-arabidopsis-seeds (58). We include gene lists for each category along with a measure of each gene's degree of adaptive evolution, as well as summary statistics computed for each gene set. We also provide the Perl wrapper we authored that performs bootstrapping and permutation tests on gene sets using DFE-α (21).

ACKNOWLEDGMENTS. We thank Justin Fay, Kenneth Olsen, and Jonathan Myers for comments on this project. This material is based upon work supported by the National Science Foundation under Grants IOS-1656756, IOS-1753743, and DGE-1143954, and by John Templeton Foundation Grant 43667.

1. Dawkins R, Krebs JR (1979) Arms races between and within species. *Proc R Soc Lond B Biol Sci* 205:489–511.
2. Brockhurst MA, et al. (2014) Running with the red queen: The role of biotic conflicts in evolution. *Proc Biol Sci* 281:20141382.
3. Paterson S, et al. (2010) Antagonistic coevolution accelerates molecular evolution. *Nature* 464:275–278.
4. Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-genes. *Science* 292:2281–2285.
5. Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ (2014) Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Mol Biol Evol* 31:574–583.
6. Trivers RL (1974) Parent-offspring conflict. *Am Zool* 14:249–264.
7. Kilner RM, Hinde CA (2012) Parent-offspring conflict. *The Evolution of Parental Care*, eds Royle NJ, Smiseth PT, Kölliker M (Oxford Univ Press, Oxford), pp 119–132.
8. Chuong EB, Tong W, Hoekstra HE (2010) Maternal-fetal conflict: Rapidly evolving proteins in the rodent placenta. *Mol Biol Evol* 27:1221–1225.
9. Queller DC (1983) Kin selection and conflict in seed maturation. *J Theor Biol* 100: 153–172.
10. Queller D (1984) Models of kin selection on seed provisioning. *Heredity* 53:151–165.
11. Westoby M, Rice B (1982) Evolution of the seed plants and inclusive fitness of plant tissues. *Evolution* 36:713–724.
12. Willson MF, Burley N (1983) *Mate Choice in Plants* (Princeton Univ Press, Princeton).
13. Charnov EL (1979) Simultaneous hermaphroditism and sexual selection. *Proc Natl Acad Sci USA* 76:2480–2484.
14. Shaanker RU, Ganeshaiah KN, Bawa KS (1988) Parent-offspring conflict, sibling rivalry, and brood size patterns in plants. *Annu Rev Ecol Syst* 19:177–205.
15. Costa LM, Gutièrrez-Marcos JF, Dickinson HG (2004) More than a yolk: The short life and complex times of the plant endosperm. *Trends Plant Sci* 9:507–514.
16. Li J, Berger F (2012) Endosperm: Food for humankind and fodder for scientific discoveries. *New Phytol* 195:290–305.
17. Cailleau A, Grimanelli D, Blanchet E, Cheptou P-O, Lenormand T (2018) Dividing a maternal pie among half-sibs: Genetic conflicts and the control of resource allocation to seeds in maize. *Am Nat* 192:577–592.
18. Le BH, et al. (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci USA* 107:8063–8070.
19. Belmonte MF, et al. (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. *Proc Natl Acad Sci USA* 110:E435–E444.
20. Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405–1413.
21. Smith NGC, Eyre-Walker A (2002) Adaptive protein evolution in *Drosophila*. *Nature* 415:1022–1024.
22. Eyre-Walker A, Keightley PD (2009) Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol* 26:2097–2108.
23. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
24. Al-Shehbaz IA, O'Kane SL, Jr (2002) Taxonomy and phylogeny of *Arabidopsis* (brassicaceae). *Arabidopsis Book* 1:e0001.
25. Tang C, et al. (2007) The evolution of selfing in *Arabidopsis thaliana*. *Science* 317: 1070–1072.
26. Kawashima T, Goldberg RB (2010) The suspensor: Not just suspending the embryo. *Trends Plant Sci* 15:23–30.
27. Haig D, Westoby M (1989) Parent-specific gene expression and the triploid endosperm. *Am Nat* 134:147–155.
28. Schon MA, Nodine MD (2017) Widespread contamination of *Arabidopsis* embryo and endosperm transcriptome data sets. *Plant Cell* 29:608–617.
29. Hamilton WD (1964) The genetical evolution of social behaviour. I. *J Theor Biol* 7:1–16.
30. Dudley SA (2015) Plant cooperation. *AoB Plants* 7:plv113.
31. Alexander RD (1974) The evolution of social behavior. *Annu Rev Ecol Syst* 4:325–383.
32. Friedman WE (1995) Organismal duplication, inclusive fitness theory, and altruism: Understanding the evolution of endosperm and the angiosperm reproductive syndrome. *Proc Natl Acad Sci USA* 92:3913–3917.
33. Irizarry RA, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249–264.
34. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) Affy–Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20:307–315.
35. Gentleman RC, et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.
36. R Core Team (2014) R: A Language and Environment for Statistical Computing (R Found Stat Comput, Vienna).
37. Ritchie ME, et al. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47.
38. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57:289–300.
39. Fay JC, Wyckoff GJ, Wu C-I (2002) Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* 415:1024–1026.
40. Fracassetti M, Griffin PC, Willi Y (2015) Validation of pooled whole-genome resequencing in *Arabidopsis lyrata*. *PLoS One* 10:e0140462.
41. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and inparalogs from pairwise species comparisons. *J Mol Biol* 314:1041–1052.
42. Camacho C, et al. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421.
43. Hu TT, et al. (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet* 43:476–481.
44. Briskine RV, et al. (2017) Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. *Mol Ecol Resour* 17:1025–1036.
45. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
46. Suyama M, Torrents D, Bork P (2006) PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34: W609-12.
47. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.
48. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426.
49. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
50. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
51. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
52. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993.
53. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
54. McKenna A, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303.
55. Danecek P, et al.; 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
56. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.
57. Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1–13.
58. Geist KS (2019) Adaptation in *Arabidopsis* seeds. Github. Available at https://github.com/ksgeist/adaptation-in-arabidopsis-seeds. Deposited April 10, 2019.