Differences in Effective Ploidy Drive Genome-Wide Endosperm Expression Polarization and Seed Failure in Wild Tomato Hybrids

Morgane Roth,¹ Ana M. Florez-Rueda,² and Thomas Städler³

Plant Ecological Genetics, Institute of Integrative Biology and Zurich–Basel Plant Science Center, ETH Zurich, 8092 Zurich, Switzerland

ORCID IDs: 0000-0002-5244-4215 (M.R.); 0000-0003-2139-2528 (A.M.F.-R.); 0000-0003-4513-8583 (T.S.)

ABSTRACT Parental imbalances in the endosperm leading to impaired development and eventual hybrid seed failure are common causes of postzygotic isolation in flowering plants. Endosperm sensitivity to parental dosage is reflected by canonical phenotypes of “parental excess” in reciprocal interploid crosses. Moreover, parental-excess traits are also evident in many homoploid interspecific crosses, potentially reflecting among-lineage variation in “effective ploidy” driven by endosperm properties. However, the genetic basis of effective ploidy is unknown and genome-wide expression perturbations in parental-excess endosperms from homoploid crosses have yet to be reported. The tomato clade (*Solanum* section *Lycopersicon*), encompassing closely related diploids with partial-to-complete hybrid seed failure, provides outstanding opportunities to study these issues. Here, we compared replicated endosperm transcriptomes from six crosses within and among three wild tomato lineages. Strikingly, strongly inviable hybrid crosses displayed conspicuous, asymmetric expression perturbations that mirror previously characterized parental-excess phenotypes. *Solanum peruvianum*, the species inferred to have evolved higher effective ploidy than the other two, drove expression landscape polarization between maternal and paternal roles. This global expression divergence was mirrored in functionally important gene families such as MADS-box transcription factors and E3 ubiquitin ligases, and revealed differences in cell cycle tuning that match phenotypic differences in developing endosperm and mature seed size between reciprocal crosses. Our work starts to uncover the complex interactions between expression divergence, parental conflict, and hybrid seed failure that likely contributed to plant diversity.

KEYWORDS endosperm; hybrid seed failure; effective ploidy; gene expression; wild tomatoes

HYBRID seed failure (HSF) is a common phenotype mediating early-acting postzygotic reproductive isolation in flowering plants. HSF does not necessarily result from F1 embryo defects as embryos may be rescued from developing seeds and grown to become fertile plants (Sharma *et al.* 1996). Such observations have been widely interpreted as

evidence for hybrid endosperms’ compromised ability to nourish the embryo (Lester and Kang 1998; Sekine *et al.* 2013; Rebernig *et al.* 2015). As the products of double fertilization, the embryo and endosperm are genetically closely related, yet these fertilization products are strongly dissimilar, concomitant with their different genome compositions (embryo diploid, 1m:1p and endosperm triploid, 2m:1p) and methylation profiles (Gehring *et al.* 2009). This original “brotherhood” between endosperm and embryo evolved over long periods of coevolutionary history, which might have contributed to the success of flowering plants (Baroux *et al.* 2002; Becraft and Gutierrez-Marcos 2012).

The frequent occurrence of HSF in interploidy crosses has been interpreted to be a consequence of endosperm sensitivity to parental dosage, establishing a reproductive barrier termed the “triploid block” (Köhler *et al.* 2010; Stoute *et al.* 2012). Well-known features of interploid seed failure are the typically contrasting phenotypes of reciprocal developing and/or

Copyright © 2019 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.119.302056>

Manuscript received February 26, 2019; accepted for publication March 18, 2019; published Early Online March 22, 2019.

Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.7770620>.

¹Present address: Agroscope, Research Division Plant Breeding, 8820 Wädenswil, Switzerland.

²Present address: Plant Developmental Genetics, Department of Plant and Microbial Biology and Zurich–Basel Plant Science Center, University of Zurich, 8008 Zurich, Switzerland.

³Corresponding author: Plant Ecological Genetics, Institute of Integrative Biology and Zurich–Basel Plant Science Center, ETH Zurich, Universitätstrasse 16, 8092 Zurich, Switzerland. E-mail: thomas.staedler@env.ethz.ch

mature hybrid seeds (Cooper and Brink 1945; Valentine and Woodell 1963; Scott *et al.* 1998; Leblanc *et al.* 2002; Pennington *et al.* 2008). Specifically, these asymmetric phenotypes comprise smaller seeds when the ovule parent is of higher ploidy (“maternal-excess phenotype”) and larger seeds when the pollen parent is of higher ploidy (“paternal-excess phenotype”; Haig and Westoby 1991). As endosperm size—which largely determines mature seed size—is affected in corresponding directions in such reciprocal interploidy crosses, parental-excess phenotypes have been regarded as a direct consequence of asymmetric parental dosage in their endosperms (Scott *et al.* 1998; Sabelli and Larkins 2009; Stoute *et al.* 2012).

Importantly, such inferred dosage sensitivity is also suspected to play a role in the developmental trajectory and (often) abortion of homoploid hybrid seeds with similar symptoms of parental excess (Josefsson *et al.* 2006; Rebernig *et al.* 2015; Oneal *et al.* 2016; Lafon-Placette *et al.* 2017, 2018). Phenotypic asymmetries between seeds from reciprocal homoploid crosses indicate that incompatibilities expressed in hybrid endosperms encompass parental effects. These phenomena might be caused by differences in so-called effective ploidy, a compound property thought to determine dosage requirements for specific genes in a given lineage [reviewed in Lafon-Placette and Köhler (2016)], and in classical work on tuber-bearing *Solanum* species proposed as the “endosperm balance number” (EBN; Johnston *et al.* 1980; Ortiz and Ehlenfeldt 1992). In crosses between homoploid species with different effective ploidies, the species with higher effective ploidy would mimic the lineage with higher actual (karyotypic) ploidy in an interploidy cross. Relevant studies have recently been performed in two Brassicaceae genera, *Arabidopsis* and *Capsella*, where it was shown that the parent with an outcrossing breeding system (*Arabidopsis lyrata*, *A. arenosa*, or *Capsella grandiflora*) drove seed phenotypes of maternal and paternal excess (Josefsson *et al.* 2006; Rebernig *et al.* 2015; Lafon-Placette *et al.* 2017, 2018); induced ploidy increases in the studied inbreeding species partly restored seed viability (Josefsson *et al.* 2006; Lafon-Placette *et al.* 2017). Beyond such phenotypic evidence, divergence in dosage between parental species of flowering plants and its consequences for genome-wide expression modulation appear to not have been assessed.

To date, genome-wide studies on endosperm gene expression have mainly focused on characterizing genomic imprinting, *i.e.*, the parent-of-origin-dependent expression of genes. A trend for elevated expression of imprinted genes in species with higher effective ploidy was found between closely related species, but whether this might contribute to HSF is uncertain (Klosinska *et al.* 2016; Roth *et al.* 2018b). Of note, genomic imprinting is extensively perturbed in failing wild tomato hybrid endosperm (Florez-Rueda *et al.* 2016), but it is unclear whether misimprinting *per se* or total expression-level changes of functionally important genes (plausibly including imprinted genes) underpin hybrid seed abortion. We may hypothesize that parental imbalances caused by

divergent effective ploidies affect global expression levels and dosage-sensitive processes such as genomic imprinting. Moreover, we expect such parental imbalances to be reflected in opposite patterns of expression change in reciprocal crosses.

Wild tomatoes (*Solanum* section *Lycopersicon*) provide a well-suited plant system for the study of developmental and evolutionary questions regarding HSF (Florez-Rueda *et al.* 2016; Roth *et al.* 2018a). We have recently shown that crosses between *Solanum arcanum* var. marañón (A), *S. chilense* (C), and *S. peruvianum* (P) result in different degrees of endosperm disruption leading to partial or complete seed inviability (Roth *et al.* 2018a). In particular, crosses between A and C yield variable proportions of viable and inviable seeds (here categorized as “weak-HSF” seeds) whereas crosses between P and either A or C result in near-complete seed failure (termed “strong-HSF” seeds; Supplemental Material, Figure S1). Moreover, marked phenotypic asymmetries are characteristic of seeds from reciprocal crosses of the strong-HSF category, where endosperms fathered by species P (*i.e.*, from crosses AP and CP) correspond to paternal-excess phenotypes and endosperms of P maternal plants (*i.e.*, from crosses PA and PC) correspond to maternal-excess phenotypes (Florez-Rueda 2014; Roth *et al.* 2018a). Thus, we hypothesized an increased effective ploidy in lineage P compared to C and A.

The present study seeks to (i) quantify molecular perturbations of gene expression levels in (partly or entirely) failing wild tomato endosperms, (ii) assess genome-wide expression differences between parental-excess seed types, (iii) identify candidate genes/gene families with potentially important roles underlying parental-excess and effective ploidy, and (iv) discuss the potential molecular drivers of effective ploidy variation among closely related species.

Materials and Methods

Plant material and crosses

Seeds were obtained from the Tomato Genetics Resource Center (University of California, Davis; <https://tgrc.ucdavis.edu>). We crossed three genotypes (one per species) in a full diallele design, with all reciprocal crosses producing seed phenotypes typical for weak or strong seed inviability, respectively (Roth *et al.* 2018a; Figure S1). Genotypes were chosen from population LA2185 (Amazonas, Peru) for A, population LA4329 (Antofagasta, Chile) for C, and population LA2744 (Arica and Parinacota, Chile) for P to be used in hybrid crosses (Figure S2). In addition, we chose three genotypes from additional populations of each species to perform intraspecific reciprocal crosses (referred to as “controls”; Figure S2). The corresponding populations were LA1626 (Ancash, Peru) for A, LA2748 (Tarapaca, Chile) for C, and LA2964 (Tacna, Peru) for P. The latter three populations were not used in hybrid crosses. As detailed in Roth *et al.* (2018a), all crosses produced normal quantities of seeds per fruit. Plants

were grown from seed in an insect-free greenhouse at ETH Zurich (Lindau-Eschikon, canton Zurich, Switzerland). They were regularly repotted in 5-liter pots using fresh soil (Ricoter Substrate 214; Ricoter Erdaufbereitung AG, Aarberg, Switzerland) and fertilizing granules (Gartensegen; Hauert HBG Dünger AG, Grossaffoltern, Switzerland). Additional liquid fertilizer was applied once or twice per month depending on the season (Wuxal NPK solution; Aglukon Spezialdünger GmbH and Co. KG, Düsseldorf, Germany). Plants were watered two-to-four times per week.

Well before the onset of the experiments, cuttings yielded multiple ramets per genotype, from which we chose three to serve as biological replicates. All clones were maintained in a climate chamber for the duration of the experiment (12 hr light at 18 klx and 50% relative humidity, 12 hr darkness at 0 klx with 60% relative humidity). Reciprocal crosses were named with the two initial letters of parental lineages within brackets [all reciprocal crosses were: (AC), (AP), (CP), (AA), (CC), and (PP)], and individual crosses were designated by the initial letters of parental lineages without brackets, indicating the cross direction “mother $\times$ father”: AA1, LA2185A $\times$ LA1626B; AA2, LA1626B $\times$ LA2185A; CC1, LA4329B $\times$ LA2748B; CC2, LA2748B $\times$ LA4329B; PP1, LA2744B $\times$ LA2964A; PP2, LA2964A $\times$ LA2744B; AC, LA2185A $\times$ LA4329B; CA, LA4329B $\times$ LA2185A; AP, LA2185A $\times$ LA2744B; PA, LA2744B $\times$ LA2185A; CP, LA4329B $\times$ LA2744B; and PC, LA2744B $\times$ LA4329B. Each cross was performed three times using clonal replicates of each genotype. Fruits were sampled 12 days after manual pollination (DAP)—corresponding to the early globular embryo stage—then embedded, and endosperms were sampled from fruit cryosections via laser-assisted microdissection. Methods for endosperm sampling, RNA extraction, library preparation, and sequencing are detailed in our previous study (Roth *et al.* 2018b).

Alignment and counting methods

Short-read alignment was done as previously described (Roth *et al.* 2018b). Briefly, RNA sequencing (RNA-Seq) quality assessment of all samples was performed with the FastQC program (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapters were removed with cutadapt (Martin 2011). Trimming and quality filtering were done with the Perl script trimmingreads.pl from the NGSQC Toolkit version 2.3 (Patel and Jain 2012). Read mapping was performed with TopHat version 2.1.0 (Trapnell *et al.* 2009) against the SL2.50 reference genome of the cultivated tomato var. Heinz (The Tomato Genome Consortium 2012) with the corresponding annotation ITAG2.4 (International Tomato Annotation Consortium; <https://solgenomics.net/>). Mapping quality checking was done with Qualimap version 2.2 (Okonechnikov *et al.* 2016) and RSeQC (Wang *et al.* 2012). Total reads per gene were counted from bam files with HTseq (Anders *et al.* 2015) using the gff ITAG2.4 annotation file (The Tomato Genome Consortium 2012). Only reads with mapping quality > 20 were retained.

Statistical analyses

Differential gene expression (DGE) analysis was performed with the R package EdgeR (Robinson *et al.* 2010; R Development Core Team 2014). After filtering our data set for lowly expressed genes (only genes with at least one read count per million in at least two of the 36 libraries were kept), 22,006 genes remained for DGE analysis, indicating that 63.4% of the ITAG2.4-annotated genes were jointly expressed in the 12-DAP endosperms of our various cross types. We used Multidimensional Scaling (MDS) plots to assess variation between biological replicates, using the function plotMDS in EdgeR, and target groups “cross_ID” for intraspecific crosses and “cross_spe” for the whole data set (Table S1). A negative binomial model was fitted to each gene using individual crosses as factors, estimating trended dispersions (variance parameters). Differentially Expressed Genes (DEGs) were identified in the selected pairwise comparisons using different contrasts with a generalized linear model likelihood ratio test [*P*-value correction with the Benjamini–Hochberg method for a false discovery rate (FDR) of 5%].

In each comparison, we used specific contrasts to compare two classes of crosses according to different criteria. First, to identify species-specific expression changes we compared intraspecific crosses in a pairwise manner [e.g., in the (AA)-(CC) comparison, we compared all replicates of AA1 and AA2 to all replicates of CC1 and CC2]. Second, to study the effects of hybridization, we (conservatively) focused on consistent expression changes between hybrids and intraspecific crosses in both their parental species [e.g., in AP-intra we compared AP (all replicates) sequentially to PP1, PP2, AA1, and AA2 (taking all replicates) and kept only genes being differentially expressed in the same direction toward AP in all four comparisons]. Finally, we looked at the effect of cross direction on hybrid endosperm expression by comparing reciprocal crosses (e.g., in the AP-PA comparison, all replicates of the AP cross were compared to all replicates of the PA cross); in total, we report 17 different contrasts (Table S2). Count data used for creating heatmaps were obtained from normalized counts per million, averaged across replicates for each cross and among cross directions for intraspecific crosses [AA = mean of (AA1 and AA2); CC = mean of (CC1 and CC2); and PP = mean of (PP1 and PP2)]. Heatmaps were plotted with the R package gplots using hierarchical clustering (R Development Core Team 2014; Warnes *et al.* 2016). The R package topGO (Alexa and Rahnenführer 2016) was used to identify enriched gene ontology (GO) terms from ITAG 2.4 downloaded from the Plant Ensembl Biomart data mining platform (Kinsella *et al.* 2011), using the set of 22,006 endosperm-expressed genes as the gene universe. Venn diagrams were obtained with the R package venneuler (Wilkinson 2011).

Data availability

Raw sequence data for the RNA-Seq data set used in this study are available from the Sequence Read Archive (<https://trace.ncbi.nlm.nih.gov/Traces/sra/>) with the accession numbers

Figure 1 Multidimensional scaling plot representing the distances between endosperm samples (*i.e.*, sequencing libraries) based on the joint expression levels of 22,006 genes. (A) All 18 samples representing intraspecific, reciprocal crosses (AA), (CC), and (PP). (B) All 36 endosperm samples (*i.e.*, intraspecific as well as hybrid) analyzed jointly. AA1, LA2185A × LA1626B; AA2, LA1626B × LA2185A; CC1, LA4329B × LA2748B; CC2, LA2748B × LA4329B; PP1, LA2744B × LA2964A; PP2, LA2964A × LA2744B; AC, LA2185A × LA4329B; CA, LA4329B × LA2185A; AP, LA2185A × LA2744B; PA, LA2744B × LA2185A; CP, LA4329B × LA2744B; PC, LA2744B × LA4329B. Cross specifications are identical in all other display items. A, *S. arcanum* var. marañón; C, *S. chilense*; log₂(FC), log₂-fold change; P, *S. peruvianum*.

PRJNA427095 (18 hybrid endosperm libraries), SRP132466 (18 within-species endosperm libraries and five parental plants; Roth *et al.* 2018b), and SRX1850236 (parent LA4329B; Florez-Rueda *et al.* 2016). All R scripts developed to perform the analyses and generate figures are available online (<https://github.com/MorganeRoth>). Figure S1 details the distribution of seed viability in all crosses used in this study, a subset of a larger phenotypic study of (hybrid) seed viability (Roth *et al.* 2018a). Figure S2 is a diagram of the crossing design representing the six reciprocal crosses used for our endosperm RNA-Seq experiment. Tables S1–S5 are compiled as separate sheets of one large Excel file: Table S1 represents the target file used in EdgeR listing all samples and sample categories used for DGE analyses; Table S2 lists the specific contrasts used in EdgeR to perform 17 cross comparisons; Table S3 reports all DGE results; Table S4 summarizes GO-term enrichments for DEGs in selected categories; and Table S5 reports the statuses of 58 candidate imprinted genes and their differential expression. Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.7770620>.

Results and Discussion

Interspecific expression differences in the endosperm

The MDS plot using expression data from only the intraspecific crosses revealed that samples broadly group by species and cross direction (Figure 1A). In particular, differences in the overall gene expression landscape between (CC) and (PP) endosperms appear to be fewer than between (CC) and (AA), or (PP) and (AA), endosperms: 1822 DEGs were found between (PP) and (CC), 2781 DEGs between (CC) and (AA),

and 2647 DEGs between (PP) and (AA) (Figure 2A, Table 1, and Table S3). This apparent genome-wide expression divergence broadly reflects the differences in divergence time between A, C, and P (Städler *et al.* 2008; Beddows *et al.* 2017). A positive correlation between sequence and expression divergence is expected from theory (Nuzhdin *et al.* 2004; Renaut *et al.* 2012). However, while our results support this notion, the correlation between expression and sequence divergence appears to be either positive (Nuzhdin *et al.* 2004; Khaitovich *et al.* 2005; Renaut *et al.* 2012) or nonsignificant (Jeukens *et al.* 2010; Wolf *et al.* 2010; Moyers and Rieseberg 2013) in previous empirical studies.

In any case, expression divergence may be regarded as a substrate for interspecific incompatibilities accumulating over evolutionary timescales. In turn, we would expect HSF severity to be correlated with the proportion of DEGs between parental species. In the present study system, this would yield expectations for (AP) and (AC) hybrids to exhibit higher levels of inviability than (CP) hybrids. However, (AP) and (CP) seeds turn out to be highly abortive whereas (AC) seeds are partly viable (Figure S1; Roth *et al.* 2018a). This indicates that HSF in wild tomatoes might be driven by expression variation at specific genes rather than by the overall number of genes differentially expressed between species.

Molecular responses to hybridization reflect distinct seed phenotypes

The global expression landscape represented by the joint analysis of all 36 samples revealed several broad expression profiles corresponding to different seed phenotypes; the y-axis of Figure 1B largely separates the intraspecific crosses (AA), (CC), and (PP), and viable hybrids (AC), from strongly

Figure 2 Overview of numbers of differentially expressed genes (DEGs) in different cross comparisons. The direction of expression change refers to the first term as compared to the second term (e.g., genes overexpressed in PP1-PP2 refer to genes overexpressed in PP1 compared to PP2). Red, overexpressed genes; blue, underexpressed genes. (A) DEGs between parental lineages (pairs of intraspecific crosses). (B) DEGs between individual hybrids and their four corresponding intraspecific crosses (see *Materials and Methods*). (C) DEGs between reciprocal crosses. The relatively high number of DEGs in reciprocal intraspecific crosses may be due to our parental individuals being from distinct populations, but also reflects differences in range-wide nucleotide diversity among lineages (Städler *et al.* 2008; Tellier *et al.* 2011). Note the differences in scales of the y-axes between (A and B) and (C). A, *S. arcanum* var. marañón; C, *S. chilense*; HSF, hybrid seed failure; P, *S. peruvianum*.

abortive crosses [(AP) and (CP)], while the x-axis visualizes the fundamentally dissimilar expression landscapes of hybrid endosperms with P as the ovule parent (PA and PC) on the one hand and those with P as the pollen parent (AP and CP) on the other hand (Figure 1B). This marked expression divergence corresponds to opposite seed phenotypes, comprising larger seeds in AP and CP crosses (paternal-excess-like) and smaller seeds in PA and PC crosses (maternal-excess-like; Roth *et al.* 2018a). Partly viable endosperms [intraspecific and (AC)] fall between these two extreme phenotypes along the x-axis of the MDS plot (Figure 1B).

To identify consistent expression changes in hybrids compared to normally developing seeds, we compared each hybrid class to its four corresponding intraspecific crosses (Tables S1–S3). In each case, we found a higher number of DEGs being underexpressed than overexpressed in hybrids (comparisons hybrid-intra; Figure 2B; Table S3). The most striking expression changes were observed in maternal-excess crosses in comparisons PA-intra and PC-intra, with 2191 and 1676 DEGs in total, respectively. Surprisingly, paternal-excess crosses were characterized by lower numbers of consistent expression changes across comparisons to parental species (871 DEGs in AP-intra and 369 in CP-intra; Table S3).

Among GO terms enriched in DEGs found in parental-excess endosperms, we identified three distinct patterns. First, GO terms relating to transcription regulation were overrepresented among under- and overexpressed genes of both parental-excess cross types, showing a global perturbation of gene expression in abortive hybrids (Table S4). Second,

some functions were concomitantly enriched in genes overexpressed in maternal-excess crosses and in genes underexpressed in paternal-excess crosses, such as auxin signaling (Table S4). Mirroring this latter case, some GO terms were enriched among genes overexpressed in paternal-excess crosses and underexpressed in maternal-excess crosses. Many GO terms were found in this category, such as biosynthesis, nucleosome and nucleus, RNA polymerase II, and some specific enzymes (pectinesterase, polygalacturonase, and dimethyl-allyltransferase; Table S4). Thus, these functions seem to be affected asymmetrically between the two distinct classes of parental-excess endosperms. Third, we identified functions specifically enriched in maternal-excess crosses; genes relating to cell division (replication and microtubules) were underexpressed, while genes relating to stress and nutrient reservoir activity were overexpressed. Together, these results reflect an extensive expression deregulation in parental-excess crosses, associated with contrasting expression patterns between maternal- and paternal-excess endosperms.

In contrast, we found only 118 and 249 DEGs for partially viable hybrids AC and CA (AC-intra and CA-intra; Figure 2B; Table S3) which were enriched for 14 GO terms in CA-intra, mainly relating to enzymatic activity, the cell wall, and lipid binding (Table S4). No significant GO term was found in AC-intra DEGs.

Changes relating to signaling have been reported to be potential contributors to HSF in *Arabidopsis* hybrid endosperm (Burkart-Waco *et al.* 2013), but functions relating to

Table 1 Contingency table of DEGs in among-species vs. reciprocal hybrid comparisons

	Direction of expression change		Total # DEGs
	Up in PA (down in AP)	Down in PA (up in AP)	
Up in (PP) [down in (AA)]	544	385	1471
Down in (PP) [up in (AA)]	98	588	1176
Total # DEGs	3222	4005	
	Up in PC (down in CP)	Down in PC (up in CP)	Total # DEGs
Up in (PP) [down in (CC)]	409	270	971
Down in (PP) [up in (CC)]	79	407	851
Total # DEGs	3354	3799	
	Up in CA (down in AC)	Down in CA (up in AC)	Total # DEGs
Up in (CC) [down in (AA)]	365	132	1491
Down in (CC) [up in (AA)]	70	484	1290
Total # DEGs	1513	1784	

Comparisons among species (first column) each include both reciprocal crosses (e.g., PP1, PP2, AA1, and AA2 were used to compare expression levels of species P and A). P, *S. peruvianum*; A, *S. arcanum* var. marañón; #, number; DEGs, differentially expressed genes; C, *S. chilense*.

global transcriptome changes during endosperm-based HSF remain poorly documented. Interestingly, functions enriched among imprinted genes (whose expression levels may be critical for seed development) seem to overlap with perturbed functions observed in parental-excess endosperms (Roth *et al.* 2018b). These GO terms correspond mainly to transcription factor (TF) activity, metabolic processes, and signaling, and are also found enriched among imprinted genes in other species such as *A. thaliana*, rice, maize, and sorghum (Gehring *et al.* 2011; Luo *et al.* 2011; Waters *et al.* 2013; Zhang *et al.* 2016).

Because transcription regulation appears to be markedly affected in abortive crosses, we scrutinized expression changes of TFs and found extensive expression changes in the MADS-box TF family; in particular, we observed two subsets of over- and underexpressed genes in paternal-excess endosperms AP and CP compared to all other cross categories (Figure 3A). MADS-box genes such as *AGAMOUS-LIKE* (AGL) genes are linked to the Polycomb Repressive Complex and are involved in *A. thaliana* endosperm cellularization during development (Kang *et al.* 2008; Walia *et al.* 2009). The paternal-excess phenotype of *A. thaliana* × *A. arenosa* interspecific seeds has been linked to the overexpression of several AGL genes in the developing endosperm (Walia *et al.* 2009), while downregulation of *AGL62* in *A. thaliana* *osd1* mutants results in a maternal-excess phenotype (Kradolfer *et al.* 2013).

Extensive expression polarization between reciprocal crosses reflects parental-excess signatures

Across our entire transcriptome data set, the strongest expression differences were found between reciprocal crosses of (PA) ($n = 7227$ DEGs) and (PC) ($n = 7153$ DEGs), representing about one-third of all endosperm-expressed genes in both cases (Figure 2C and Table 1). These results highlight the typical expression signatures of parental excess characterized by the above-mentioned strong expression differences between maternal- (PA and PC) and paternal-excess endosperm types (AP and CP), combined with many fewer DEGs

between independent hybrid crosses of the same parental-excess type (AP-CP, $n = 2639$ DEGs; PA-PC, $n = 788$ DEGs; Table S3). Moreover, of these two sets of DEGs, 4477 genes were in common and shared the same direction of expression change in both the AP-PA and CP-PC comparisons (only 127 genes showed opposite gene expression changes between them; Figure 4; Table S3).

An alternative view to invoking parental-excess polarization might posit that fundamentally perturbed seed physiology *per se* (expression chaos) is causal for the high numbers of DEGs in (some) inviable hybrid comparisons. However, this fails to explain the modest number of DEGs in the (also inviable) same-parental-excess-type comparisons that are even within the range of DEGs seen in normally developing, reciprocal intraspecific crosses (Figure 2C).

DEGs between reciprocal crosses may reveal functions preferentially controlled by one parent that are perturbed in hybrid endosperms. For example, transcription and chromatin-related activities were more often—but not exclusively—enriched among genes overexpressed with P as the pollen parent (12 GO terms compared to just three when P is the ovule parent; Table S4). Other functions appeared to be more specifically overexpressed when P was the ovule parent, such as energy metabolism (e.g., starch and lipids; eight GO terms), stress signals, cell cycle control (protein phosphorylation, protein serine/threonine kinase, and auxin-related terms; seven GO terms), and cell architecture (cell wall; 13 GO terms; Table 2 and Table S4). DNA binding was also enriched among DEGs between reciprocal strong-HSF crosses (Table 2; Table S4). Overall, these functions overlap with functional categories enriched among candidate imprinted genes (Waters *et al.* 2011; Roth *et al.* 2018b). We also found that 50–67% of wild tomato conserved imprinted genes were differentially expressed between parental-excess endosperm types (Table S5). In particular, maternally expressed genes (MEGs) were mostly overexpressed in maternal-excess endosperms (32/32 differentially expressed MEGs overexpressed in PC-CP and 20/22 differentially expressed MEGs overexpressed

Figure 3 Heatmaps representing expression variability for selected gene families among intraspecific and hybrid crosses. (A) MADS-box TFs, (B) genes annotated as relating to the cell cycle, and (C) E3 ubiquitin ligases. Color scale according to Z-score (darker colors correspond to stronger expression values); samples and genes ordered by hierarchical clustering. A, *S. arcanum* var. marañón; C, *S. chilense*; P, *S. peruvianum*; TFs, transcription factors.

in PA-AP), whereas paternally expressed genes (PEGs) tended to be overexpressed in paternal-excess endosperms (seven out of seven differentially expressed PEGs overexpressed in CP-PC and five out of seven differentially expressed PEGs overexpressed in AP-PA). Thus, a parental-excess scenario in hybrid seeds might alter specific dosage mechanisms regulating the expression of imprinted genes, which is potentially lethal for the endosperm and thus the developing seed (Lafon-Placette *et al.* 2018).

Higher effective ploidy of P drives expression in parental-excess endosperm types

Due to the asymmetric parental contributions to endosperm genomes (2m:1p), reciprocal crosses qualitatively inherit the same parental genomes but differ in dosage from each parent. As a consequence, increased dosage of one parent will not have the same effects in reciprocal crosses. Thus, the extensive expression differences we observe between reciprocal hybrid classes could be a manifestation of global expression differences between lineages. The high proportion of shared gene identity and parallel expression change in AP-PA and CP-PC comparisons suggests that the strongly abortive endosperms respond in symmetric fashion relative to parent P. The relative dosage of P (two as ovule parent and one as pollen parent) mirrors the polarization of global gene expression between these two reciprocal hybrid classes (Figure 1B).

However, P does not have increased ploidy as all three species studied here are diploid. Although P does not exhibit higher genome-wide expression levels (Roth *et al.* 2018b), our DGE analysis found 14–25% more genes overexpressed than underexpressed in (PP) compared to either (AA) or (CC) endosperm [χ^2 test, comparison (PP)-(AA), $P = 5.50E-05$; comparison (PP)-(CC), $P = 5.00E-02$; Figure 2A, Table 1, and Table S3].

This indicates that, compared to both A and C, lineage P features increased expression in the endosperm that is not observed genome-wide but rather restricted to a subset of genes. Interestingly, among the common set of 390 genes overexpressed in (PP) compared to both (CC) and (AA), a

sizable fraction ($n = 252$, 64.6%) comprises genes either overexpressed in both maternal-excess crosses (PA and PC compared to AP and CP, $n = 129$) or overexpressed in both paternal-excess crosses (AP and CP compared to PA and PC, $n = 123$; Table S3). From these sets of genes, genes overexpressed in maternal-excess crosses are mainly enriched for nutrient reservoir activity and galactose metabolism, and genes overexpressed in paternal-excess crosses are enriched for DNA binding, transcription regulation, and biosynthetic processes (Table S4). These enrichments possibly reflect increased maternal influence on resource allocation in maternal-excess endosperms, and increased paternal influence on the control of gene expression and growth, respectively.

It has recently been shown in *Capsella* that imprinted genes, and especially PEGs, tend to have increased expression in species with higher effective ploidy (Lafon-Placette *et al.* 2018). In our recent study on wild tomatoes, only a small fraction of candidate imprinted genes was significantly differentially expressed between (AA), (CC), and (PP), and these were exclusively MEGs (Roth *et al.* 2018b; Table S5). MEGs overexpressed in (PP) were mostly found to be overexpressed in maternal-excess crosses PA and PC (six of seven in PA and four of five in PC; Table S5). Thus, increased expression of imprinted genes in P might contribute only marginally to the expression polarization observed between reciprocals of strongly abortive crosses. Alternatively, the contribution of imprinted genes might be underestimated because some of them remain to be identified due to technical limitations (e.g., lack of parental polymorphism for many genes in the crosses used; Roth *et al.* 2018b).

Gene families potentially involved in effective ploidy variation have not yet been reported in the literature. As mentioned above, AGL genes are involved in typical responses to parental excess in interploidy crosses (Kang *et al.* 2008; Walia *et al.* 2009; Kradolfer *et al.* 2013). In our study, 28 of 41 putative AGL genes expressed in wild tomato endosperm were jointly overexpressed in both paternal-excess crosses (Table S3). Among them, eight were also overexpressed in (PP) compared to both (CC) and (AA) endosperms (Table 3). This

Figure 4 Venn diagrams representing the overlap between DEGs identified between all reciprocal hybrid crosses. (A) The number of genes overexpressed in paternal-excess crosses and (B) the number of genes underexpressed in paternal excess crosses. For completeness, the partly viable hybrid cross comparison AC-CA (exhibiting evidence of effective ploidy differences yet no clear pattern of parental excess) is included. A, *S. arcanum* var. marañón; C, *S. chilense*; DEGs, differentially expressed genes; P, *S. peruvianum*.

pattern of expression suggests that these eight AGL genes might be paternally expressed, but their imprinting status could not be assessed due to a lack of SNPs between our parental plants (only one AGL gene was polymorphic in P and not imprinted; Roth *et al.* 2018b). Yet, AGL genes interact within a network (Walia *et al.* 2009) and are potentially subject to imprinting in the endosperm, as shown by the first-identified PEG *PHERES1*, and further AGL genes being maternally or paternally expressed in *Arabidopsis* (Köhler *et al.* 2003; Shirzadi *et al.* 2011; Bai and Settles 2015). Overall, our data indicate increased expression levels in species P for genes known to be critical for seed size and seed viability, such as AGL genes, with a potential link to imprinting.

Molecular functions underlying parental excess reveal differences in cell cycle tuning

A heatmap representing the variation of genes related to the cell cycle across the whole data set highlights the specific expression pattern found in maternal-excess crosses PA and PC (Figure 3B). More precisely, GO terms associated with genes differentially expressed between maternal- and paternal-excess crosses (*i.e.*, PA and PC vs. AP and CP, respectively) indicate contrasting cell cycle regimes between these two cross directions (Table S4). DNA replication and chiasma assembly were enriched among genes overexpressed in paternal-excess endosperms, indicating that AP and CP endosperm cells were probably still dividing at 12 DAP, while proliferation had most likely stopped in the reciprocal PA and PC endosperms (Table S4). Our prior morphological measurements of various (CP) seed compartments between 10 and 13 DAP bolster this inference (Roth *et al.* 2018a). Also, the enrichment in cell cycle control- and cell wall-related terms in DEGs between hybrid endosperms with P in the maternal vs. paternal role is plausibly linked to cell

proliferation differences. Related to this, we found striking expression polarization in E3 ubiquitin ligases whose protein products are involved in the control of the cell cycle (Figure 3C; Inzé and De Veylder 2006).

The function “negative regulation of growth” was overexpressed in maternal-excess phenotypes, combined with an increased response to auxin, whose concentrations are known to impact the rate of cell division (Tables S3, S4; John *et al.* 1993; Schruff *et al.* 2006; Orozco-Arroyo *et al.* 2015; Figueiredo *et al.* 2016). *A. thaliana* *arf* mutants bear a nonfunctional *AUXIN RESPONSE FACTOR 2* (ARF2) and a paternal-excess phenotype with enlarged seeds, due to delayed and extended cell divisions in seed tissues (Schruff *et al.* 2006). This indicates that maternal factors control the response to auxin, which is responsible for the control of cell cycle transitions. Interestingly, five ARFs were found to be MEGs in wild tomato endosperm and three of them were overexpressed in maternal-excess phenotypes (Solyc04g081240.2, Solyc07g043610.2, and Solyc11g069500.1; Roth *et al.* 2018b; Table S3). Signals for cell differentiation and responses to hormones involved in cell differentiation and seed maturation, such as brassinosteroids and abscisic acid (Orozco-Arroyo *et al.* 2015), were overrepresented among genes overexpressed in the maternal-excess endosperms (Table S3). Compared to all intra-specific endosperms, genes involved in mitotic chromosome condensation and the regulation of G2/M transition of the mitotic cell cycle were mainly underexpressed in PA and PC, whereas genes involved in seed dormancy were overexpressed (Table S4). These concomitant expression changes probably reflect our histological observations that maternal-excess endosperms have stopped dividing and already started to differentiate at the early globular embryo stage (Roth *et al.* 2018a).

Thus, we suggest that hormone concentrations, regulating the progression through the cell cycle, are mainly under maternal control and perturbed in opposite ways in (PA, PC) vs. (AP, CP) endosperms, contributing to maternal- and paternal-excess endosperm morphologies, and the corresponding seed size differences. As proposed for interploidy maize crosses by Leblanc *et al.* (2002), parental dosage would influence the cell cycle such that: (i) rapid mitotic arrest is due to fast G/M transitions in maternal-excess endosperm, and (ii) a longer phase of cell proliferation is due to facilitated reentry into the S phase (DNA replication phase) and delayed G/M transitions in paternal-excess endosperm.

Evolutionary implications of effective ploidy variation for reproductive isolation

We found that transcriptomic differences were associated with phenotypic differences between intraspecific, partially viable, and completely inviable hybrid seeds. Our study system included two crosses with reciprocal strong-HSF phenotypes [(AP) and (CP)] which also exhibited similar global expression signatures. Thus, the (AP) and (CP) data reflect independently evolved yet similar biological features, suggesting shared molecular and physiological underpinnings of reproductive isolation between closely related lineages.

Table 2 Top 10 GO terms enriched among genes differentially expressed between maternal- and paternal-excess hybrid endosperms

Direction of change	Ontology category	GO term ID	GO term description	Annotated genes (#)	Observed genes (#)	Expected genes (#)	Corrected <i>P</i> -value
Overexpressed in paternal-excess (AP and CP) crosses	MF	4161	Dimethylallyltranstransferase activity	20	20	2	1.75E−17
	MF	3677	DNA binding	1358	247	163	3.25E−15
	MF	8234	Cysteine-type peptidase activity	133	49	16	2.00E−11
	MF	46983	Protein dimerization activity	354	90	43	1.63E−09
	BP	6334	Nucleosome assembly	41	20	5	7.75E−08
	BP	6508	Proteolysis	684	122	78	7.75E−08
	MF	8289	Lipid binding	125	33	15	1.50E−07
	MF	1104	RNA polymerase II transcription Cofactor activity	28	13	3	5.58E−05
	MF	46982	Protein heterodimerization activity	118	31	14	1.14E−04
Overexpressed in maternal-excess (PA and PC) crosses	MF	30599	Pectinesterase activity	49	17	6	1.94E−04
	MF	3700	Transcription factor activity	473	98	42	1.95E−14
	MF	43565	Sequence-specific DNA binding	265	62	23	3.75E−11
	MF	8146	Sulfotransferase activity	18	11	2	7.17E−07
	BP	6355	Regulation of transcription, DNA templated	1086	143	98	2.05E−06
	MF	45735	Nutrient reservoir activity	30	13	3	6.63E−06
	BP	9734	Auxin-activated signaling pathway	48	17	4	4.75E−05
	MF	4722	Protein serine/threonine phosphatase activity	79	20	7	1.20E−04
	MF	8289	Lipid binding	125	25	11	2.08E−04
	MF	4674	Protein serine/threonine kinase activity	413	61	36	4.36E−04
	BP	6468	Protein phosphorylation	934	122	84	7.67E−04

GO, gene ontology; ID, identifier; #, number; A, *S. arcanum* var. *marañón*; P, *S. peruvianum*; C, *S. chilense*; MF, molecular function; BP, biological process.

Moreover, the polarized phenotypes and expression landscapes of strongly abortive hybrid seeds during early development indicate that effective ploidy variation might facilitate the establishment of reproductive isolation. More specifically, species P appears to drive HSF at both the molecular and phenotypic levels upon hybridization with lineages C and A. Thus, we propose that P bears increased effective endosperm dosage which can be interpreted as higher effective ploidy (or EBN; Johnston *et al.* 1980; Lafon-Placette and Köhler 2016). Empirical data in *Capsella* suggest a positive correlation between levels of parental conflict within lineages and effective ploidy (Rebernig *et al.* 2015; Lafon-Placette *et al.* 2018). As levels of parental conflict should negatively correlate with relatedness between parents, such conflict should decrease with more intense inbreeding (Brandvain and Haig 2005); indirect evidence supporting this prediction comes from interpopulation crosses of a mixed-mating plant (Raunsgard *et al.* 2018). Although our study included only obligate outcrossers, lineages A, C, and P harbor different levels of range-wide nucleotide diversity; specifically, P is the most diverse and A the least diverse lineage (Städler *et al.* 2008; Tellier *et al.* 2011; Labate *et al.* 2014). Range-wide nucleotide diversity should reflect long-term effective population size; all other things being equal, one would expect lower parental conflict between two randomly drawn plants from the least polymorphic (A) compared to the more polymorphic lineages

(C and, particularly, P). In summary, we infer the relative effective ploidies between lineages to be $P > C > A$.

Lafon-Placette *et al.* (2018) identified higher numbers and expression levels of PEGs in the obligatory outcrosser *C. grandiflora* (inferred to have the highest effective ploidy), compared to the highly selfing *C. rubella* and the more ancient selfer *C. orientalis*. In contrast, our present and previously reported data (Roth *et al.* 2018b) suggest that PEGs are expressed at similar levels between A, C, and P. We also found no significant differences in the proportions of PEGs between A, C, and P (χ^2 test, $P > 0.05$). This lack of a clear signal regarding numbers and expression levels of PEGs, concomitant with apparent divergence in effective ploidy, can be reconciled due to the presumably closer levels of parental conflict among our wild tomato lineages (with A, C, and P all being obligate outcrossers) compared to the *Capsella* system (Lafon-Placette *et al.* 2018). While genomic imprinting probably plays an important role in the evolution of effective ploidy, our results indicate that the functional drivers might not be restricted to imprinted genes [see also von Wangenheim and Peterson (2004)]. We suggest that regulators of parent-specific expression, rather than strictly imprinted genes, might be responsible for evolutionary changes in effective ploidy.

Given only moderate genome-wide differences in expression levels between lineages and relatively few DEGs between

Table 3 Expression pattern and annotation of eight AGL genes potentially contributing to differences in effective ploidy between species

Gene model	Among-species comparisons			Reciprocal hybrid comparisons			Source	
	(PP)-(AA)	(PP)-(CC)	(CC)-(AA)	AP-PA	CP-PC	AC-CA	TFDB v3.0	Annotation in ITAG2.4
Solyc01g097850	Up	Up	Up	Up	Up	Up	AGAMOUS-like 62	MADS-box TF 31
Solyc01g103870	Up	Up	Up	Up	Up	Ns	AGAMOUS-like 98	SRF-type TF family protein
Solyc03g033570	Up	Up	Up	Up	Up	Up	Not found	Agamous-like MADS-box protein AGL62
Solyc04g025110	Up	Up	Up	Up	Up	Ns	AGAMOUS-like 62	MADS-box TF 8
Solyc04g047870	Up	Up	Up	Up	Up	Ns	AGAMOUS-like 62	MADS box TF 1
Solyc06g054680	Up	Up	Up	Up	Up	Up	AGAMOUS-like 62	MADS-box TF
Solyc11g069770	Up	Up	Up	Up	Up	Up	AGAMOUS-like 62	TF MADS-box
Solyc12g016150	Up	Up	Up	Up	Up	Ns	AGAMOUS-like 96	MADS-box protein (fragment)

P, *S. peruvianum*; A, *S. arcanum* var. marañón; C, *S. chilense*; TFDB, Transcription Factor Database v3.0 (http://planttfdb_v3.cbi.pku.edu.cn/); Up, overexpressed in first cross of each pairwise comparison; TF, transcription factor; Ns, nonsignificant expression change; SRF, serum response factor.

them (Figure 2A, Table 1, and Table S3), we propose that the property effective ploidy manifests as the stronger/weaker expression of a limited number of specific genes controlling dosage-sensitive mechanisms; we have provided a number of candidate mechanisms potentially underlying this feature. In particular, expression levels of AGL genes seem to match the inferred genetic value hierarchy (Table 3). Thus, these are candidates for underpinning different effective ploidies between tomato lineages and, as a consequence, they might be decisive for the occurrence and/or severity of HSF. Knocking out single or multiple AGL genes in parental plants, or modifying their expression levels in the endosperm, as has been done in *Arabidopsis* (Walia *et al.* 2009; Kradolfer *et al.* 2013), would allow the validation of their specific roles (if any) in endosperm development and seed failure in *Solanum*.

It has been shown that *Arabidopsis* AGL genes act within a network and that they can be nonimprinted, maternally expressed, or paternally expressed (Walia *et al.* 2009; Bai and Settles 2015). Thus, any perturbation of expression levels among coadapted AGLs in hybrids might be at the root of the genome-wide perturbations observed in strong-HSF hybrids. Within species, parental conflict might be stabilized by gene expression coadaptation within functional networks, which might also determine the property effective ploidy. When parental species have accrued diverged effective ploidies this equilibrium may be disrupted in their hybrids, acting as a postzygotic reproductive barrier with varying quantitative effects depending on the disparity of effective ploidies as manifested in the endosperm. In this context, our work is the first to explore genome-wide expression correlates of dissimilar effective ploidies in the endosperm, thus enabling the exploration of possible links between parental conflict, expression dosage, and HSF in flowering plants. It may also have practical applications in plant breeding, for example to enhance hybridization success between crops and their wild relatives by compensating effective ploidy differences with targeted, experimental ploidy changes.

Acknowledgments

We are grateful to Maja Frei and Esther Zürcher for taking expert care of the plants; Beatrice Arnold for preparing RNA-Seq

libraries; Claudia Michel, Silvia Kobel, and Joachim Hehl for further technical help; Margot Paris for her advice on experimental design and analyses; Niklaus Zemp, Stefan Zoller, and Mathias Scharmann for their generous bioinformatics advice; Alex Widmer for his general support of this project; the C.M. Rick Tomato Genetics Resource Center at the University of California, Davis for generously supplying seed samples; and the Genetic Diversity Center (ETH Zurich, Switzerland) and the Swiss Institute for Bioinformatics (Lausanne, Switzerland) for providing valuable tools and training for bioinformatics analyses. We acknowledge the technical support regarding histological preparations and laser microdissections provided by SECTION-LAB (Hiroshima, Japan) and ScopeM (ETH Zurich, Switzerland). Sequencing data were produced at the Functional Genomics Center Zurich (University of Zurich, Switzerland). This work was supported by the Swiss National Science Foundation (grant 31003A_130702 to T.S.) and an ETH Zurich Research grant (ETH-40 13-2 to T.S. and Alex Widmer).

Literature Cited

- Alexa, A., and J. Rahnenführer, 2016 topGO: enrichment analysis for gene ontology. R Package Version 2.28.0. <https://doi.org/10.18129/B9.bioc.topGO>
- Anders, S., P. T. Pyl, and W. Huber, 2015 HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Bai, F., and A. M. Settles, 2015 Imprinting in plants as a mechanism to generate seed phenotypic diversity. *Front. Plant Sci.* 5: 780. <https://doi.org/10.3389/fpls.2014.00780>
- Baroux, C., C. Spillane, and U. Grossniklaus, 2002 Evolutionary origins of the endosperm in flowering plants. *Genome Biol.* 3: reviews1026.
- Becraft, P. W., and J. Gutierrez-Marcos, 2012 Endosperm development: dynamic processes and cellular innovations underlying sibling altruism. *WIREs. Dev. Biol.* 1: 579–593.
- Beddows, I., A. Reddy, T. Kloesges, and L. E. Rose, 2017 Population genomics in wild tomatoes—the interplay of divergence and admixture. *Genome Biol. Evol.* 9: 3023–3038. <https://doi.org/10.1093/gbe/evx224>
- Brandvain, Y., and D. Haig, 2005 Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. *Am. Nat.* 166: 330–338. <https://doi.org/10.1086/432036>

- Burkart-Waco, D., K. Ngo, B. Dilkes, C. Josefsson, and L. Comai, 2013 Early disruption of maternal-zygotic interaction and activation of defense-like responses in *Arabidopsis* interspecific crosses. *Plant Cell* 25: 2037–2055. <https://doi.org/10.1105/tpc.112.108258>
- Cooper, D. C., and R. A. Brink, 1945 Seed collapse following matings between diploid and tetraploid races of *Lycopersicon pimpinellifolium*. *Genetics* 30: 376–401.
- Figueiredo, D. D., R. A. Batista, P. J. Roszak, L. Hennig, and C. Köhler, 2016 Auxin production in the endosperm drives seed coat development in *Arabidopsis*. *Elife* 5: e20542. <https://doi.org/10.7554/eLife.20542>
- Florez-Rueda, A. M., 2014 Postzygotic barriers to interbreeding in wild tomatoes: genomic imprinting and transcriptional signatures of hybrid seed failure. Ph.D. Thesis, ETH Zurich, Zurich, Switzerland.
- Florez-Rueda, A. M., M. Paris, A. Schmidt, A. Widmer, U. Grossniklaus *et al.*, 2016 Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Mol. Biol. Evol.* 33: 2935–2946. <https://doi.org/10.1093/molbev/msw175>
- Gehring, M., K. L. Bubb, and S. Henikoff, 2009 Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* 324: 1447–1451. <https://doi.org/10.1126/science.1171609>
- Gehring, M., V. Missirlian, and S. Henikoff, 2011 Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS One* 6: e23687. <https://doi.org/10.1371/journal.pone.0023687>
- Haig, D., and M. Westoby, 1991 Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Phil. Trans. R. Soc. Lond. B* 333: 1–13.
- Inzé, D., and L. De Veylder, 2006 Cell cycle regulation in plant development. *Annu. Rev. Genet.* 40: 77–105. <https://doi.org/10.1146/annurev.genet.40.110405.090431>
- Jeukens, J., S. Renaut, J. St-Cyr, A. W. Nolte, and L. Bernatchez, 2010 The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Mol. Ecol.* 19: 5389–5403. <https://doi.org/10.1111/j.1365-294X.2010.04934.x>
- John, P., K. Zhang, C. Dong, L. Diederich, and F. Wightman, 1993 p34^{cdc2} related proteins in control of cell cycle progression, the switch between division and differentiation in tissue development, and stimulation of division by auxin and cytokinin. *Aust. J. Plant Physiol.* 20: 503–526.
- Johnston, S. A., T. P. M. den Nijs, S. J. Peloquin, and R. E. Hanneman, 1980 The significance of genic balance to endosperm development in interspecific crosses. *Theor. Appl. Genet.* 57: 5–9. <https://doi.org/10.1007/BF00276002>
- Josefsson, C., B. Dilkes, and L. Comai, 2006 Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* 16: 1322–1328. <https://doi.org/10.1016/j.cub.2006.05.045>
- Kang, I.-H., J. G. Steffen, M. F. Portereiko, A. Lloyd, and G. N. Drews, 2008 The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell* 20: 635–647. <https://doi.org/10.1105/tpc.107.055137>
- Khaitovich, P., I. Hellmann, W. Enard, K. Nowick, M. Leinweber *et al.*, 2005 Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* 309: 1850–1854. <https://doi.org/10.1126/science.1108296>
- Kinsella, R. J., A. Kähäri, S. Haider, J. Zamora, G. Proctor *et al.*, 2011 Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database (Oxford)* 2011: bar030. <https://doi.org/10.1093/database/bar030>
- Klosinska, M., C. L. Picard, and M. Gehring, 2016 Conserved imprinting associated with unique epigenetic signatures in the *Arabidopsis* genus. *Nat. Plants* 2: 16145. <https://doi.org/10.1038/nplants.2016.145>
- Köhler, C., L. Hennig, C. Spillane, S. Pien, W. Gruissem *et al.*, 2003 The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev.* 17: 1540–1553. <https://doi.org/10.1101/gad.257403>
- Köhler, C., O. Mittelsten Scheid, and A. Erilova, 2010 The impact of the triploid block on the origin and evolution of polyploid plants. *Trends Genet.* 26: 142–148. <https://doi.org/10.1016/j.tig.2009.12.006>
- Kradolfer, D., L. Hennig, and C. Köhler, 2013 Increased maternal genome dosage bypasses the requirement of the FIS Polycomb Repressive Complex 2 in *Arabidopsis* seed development. *PLoS Genet.* 9: e1003163. <https://doi.org/10.1371/journal.pgen.1003163>
- Labate, J. A., L. D. Robertson, S. R. Strickler, and L. A. Mueller, 2014 Genetic structure of the four wild tomato species in the *Solanum peruvianum* s.l. species complex. *Genome* 57: 169–180. <https://doi.org/10.1139/gen-2014-0003>
- Lafon-Placette, C., and C. Köhler, 2016 Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Mol. Ecol.* 25: 2620–2629. <https://doi.org/10.1111/mec.13552>
- Lafon-Placette, C., I. M. Johannessen, K. S. Hornslien, M. F. Ali, K. N. Bjerkkan *et al.*, 2017 Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in central Europe. *Proc. Natl. Acad. Sci. USA* 114: E1027–E1035. <https://doi.org/10.1073/pnas.1615123114>
- Lafon-Placette, C., M. R. Hatorangan, K. A. Steige, A. Cornille, M. Lascoux *et al.*, 2018 Paternally expressed imprinted genes associate with hybridization barriers in *Capsella*. *Nat. Plants* 4: 352–357. <https://doi.org/10.1038/s41477-018-0161-6>
- Leblanc, O., C. Pointe, and M. Hernandez, 2002 Cell cycle progression during endosperm development in *Zea mays* depends on parental dosage effects. *Plant J.* 32: 1057–1066. <https://doi.org/10.1046/j.1365-3113X.2002.01491.x>
- Lester, R. N., and J. H. Kang, 1998 Embryo and endosperm function and failure in *Solanum* species and hybrids. *Ann. Bot.* 82: 445–453. <https://doi.org/10.1006/anbo.1998.0695>
- Luo, M., J. M. Taylor, A. Spriggs, H. Zhang, X. Wu *et al.*, 2011 A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet.* 7: e1002125. <https://doi.org/10.1371/journal.pgen.1002125>
- Martin, M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17: 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Moyers, B. T., and L. H. Rieseberg, 2013 Divergence in gene expression is uncoupled from divergence in coding sequence in a secondarily woody sunflower. *Int. J. Plant Sci.* 174: 1079–1089. <https://doi.org/10.1086/671197>
- Nuzhdin, S. V., M. L. Wayne, K. L. Harmon, and L. M. McIntyre, 2004 Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol. Biol. Evol.* 21: 1308–1317. <https://doi.org/10.1093/molbev/msh128>
- Okonechnikov, K., A. Conesa, and F. García-Alcalde, 2016 Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32: 292–294.
- Oneal, E., J. H. Willis, and R. G. Franks, 2016 Disruption of endosperm development is a major cause of hybrid seed inviability between *Mimulus guttatus* and *Mimulus nudatus*. *New Phytol.* 210: 1107–1120. <https://doi.org/10.1111/nph.13842>
- Orozco-Arroyo, G., D. Paolo, I. Ezquer, and L. Colombo, 2015 Networks controlling seed size in *Arabidopsis*. *Plant Reprod.* 28: 17–32. <https://doi.org/10.1007/s00497-015-0255-5>

- Ortiz, R., and M. K. Ehlenfeldt, 1992 The importance of endosperm balance number in potato breeding and the evolution of tuber-bearing *Solanum* species. *Euphytica* 60: 105–113.
- Patel, R. K., and M. Jain, 2012 NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7: e30619. <https://doi.org/10.1371/journal.pone.0030619>
- Pennington, P. D., L. M. Costa, J. F. Gutierrez-Marcos, A. J. Greenland, and H. G. Dickinson, 2008 When genomes collide: aberrant seed development following maize interploidy crosses. *Ann. Bot.* 101: 833–843. <https://doi.org/10.1093/aob/mcn017>
- Raunsgard, A., Ø. H. Opedal, R. K. Ekrem, J. Wright, G. H. Bolstad *et al.*, 2018 Intersexual conflict over seed size is stronger in more outcrossed populations of a mixed-mating plant. *Proc. Natl. Acad. Sci. USA* 115: 11561–11566. <https://doi.org/10.1073/pnas.1810979115>
- Rebernig, C. A., C. Lafon-Placette, M. R. Hatorangan, T. Slotte, and C. Köhler, 2015 Non-reciprocal interspecies hybridization barriers in the *Capsella* genus are established in the endosperm. *PLoS Genet.* 11: e1005295. <https://doi.org/10.1371/journal.pgen.1005295>
- Renaud, S., C. J. Grassa, B. T. Moyers, N. C. Kane, and L. H. Rieseberg, 2012 The population genomics of sunflowers and genomic determinants of protein evolution revealed by RNAseq. *Biology (Basel)* 1: 575–596.
- R Development Core Team, 2014 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
- Robinson, M. D., D. J. McCarthy, and G. K. Smyth, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Roth, M., A. M. Florez-Rueda, S. Griesser, M. Paris, and T. Städler, 2018a Incidence and developmental timing of endosperm failure in post-zygotic isolation between wild tomato lineages. *Ann. Bot.* 121: 107–118. <https://doi.org/10.1093/aob/mcx133>
- Roth, M., A. M. Florez-Rueda, M. Paris, and T. Städler, 2018b Wild tomato endosperm transcriptomes reveal common roles of genomic imprinting in both nuclear and cellular endosperm. *Plant J.* 95: 1084–1101. <https://doi.org/10.1111/tpj.14012>
- Sabelli, P. A., and B. A. Larkins, 2009 The contribution of cell cycle regulation to endosperm development. *Sex. Plant Reprod.* 22: 207–219. <https://doi.org/10.1007/s00497-009-0105-4>
- Schruff, M. C., M. Spielman, S. Tiwari, S. Adams, N. Fenby *et al.*, 2006 The AUXIN RESPONSE FACTOR 2 gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development* 133: 251–261. <https://doi.org/10.1242/dev.02194>
- Scott, R. J., M. Spielman, J. Bailey, and H. G. Dickinson, 1998 Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* 125: 3329–3341.
- Sekine, D., T. Ohnishi, H. Furuumi, A. Ono, T. Yamada *et al.*, 2013 Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *Plant J.* 76: 792–799. <https://doi.org/10.1111/tpj.12333>
- Sharma, D. R., R. Kaur, and K. Kumar, 1996 Embryo rescue in plants – a review. *Euphytica* 89: 325–337.
- Shirzadi, R., E. D. Andersen, K. N. Bjerkkan, B. M. Gloeckle, M. Heese *et al.*, 2011 Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of AGAMOUS-LIKE36. *PLoS Genet.* 7: e1001303. <https://doi.org/10.1371/journal.pgen.1001303>
- Städler, T., U. Arunyawat, and W. Stephan, 2008 Population genetics of speciation in two closely related wild tomatoes (*Solanum* section *Lycopersicon*). *Genetics* 178: 339–350. <https://doi.org/10.1534/genetics.107.081810>
- Stoute, A. I., V. Varenko, G. J. King, R. J. Scott, and S. Kurup, 2012 Parental genome imbalance in *Brassica oleracea* causes asymmetric triploid block. *Plant J.* 71: 503–516.
- Tellier, A., I. Fischer, C. Merino, H. Xia, L. Camus-Kulandaivelu *et al.*, 2011 Fitness effects of derived deleterious mutations in four closely related wild tomato species with spatial structure. *Heredity* 107: 189–199. <https://doi.org/10.1038/hdy.2010.175>
- The Tomato Genome Consortium, 2012 The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485: 635–641. <https://doi.org/10.1038/nature11119>
- Trapnell, C., L. Pachter, and S. L. Salzberg, 2009 TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105–1111. <https://doi.org/10.1093/bioinformatics/btp120>
- Valentine, D. H., and S. R. J. Woodell, 1963 Studies in British primulas. X. Seed incompatibility in intraspecific and interspecific crosses at diploid and tetraploid levels. *New Phytol.* 62: 125–143. <https://doi.org/10.1111/j.1469-8137.1963.tb06321.x>
- von Wangenheim, K.-H., and H.-P. Peterson, 2004 Aberrant endosperm development in interploidy crosses reveals a timer of differentiation. *Dev. Biol.* 270: 277–289. <https://doi.org/10.1016/j.ydbio.2004.03.014>
- Walia, H., C. Josefsson, B. Dilkes, R. Kirkbride, J. Harada *et al.*, 2009 Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr. Biol.* 19: 1128–1132. <https://doi.org/10.1016/j.cub.2009.05.068>
- Wang, L., S. Wang, and W. Li, 2012 RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28: 2184–2185. <https://doi.org/10.1093/bioinformatics/bts356>
- Warnes, G. R., B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw *et al.*, 2016 gplots: various R programming tools for plotting data. R Package Version 3.0.1. <https://CRAN.R-project.org/package=gplots> (accessed 17 September 2018)
- Waters, A. J., I. Makarevitch, S. R. Eichten, R. A. Swanson-Wagner, C.-T. Yeh *et al.*, 2011 Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* 23: 4221–4233. <https://doi.org/10.1105/tpc.111.092668>
- Waters, A. J., P. Bilinski, S. R. Eichten, M. W. Vaughn, J. Ross-Ibarra *et al.*, 2013 Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc. Natl. Acad. Sci. USA* 110: 19639–19644. <https://doi.org/10.1073/pnas.1309182110>
- Wilkinson, L., 2011 Venneuler: Euler and Venn diagrams. R Package Version 1.1–0. <https://CRAN.R-project.org/package=venneuler> (accessed 17 September 2018)
- Wolf, J. B. W., T. Bayer, B. Haubold, M. Schilhabel, P. Rosenstiel *et al.*, 2010 Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Mol. Ecol.* 19: 162–175. <https://doi.org/10.1111/j.1365-294X.2009.04471.x>
- Zhang, M., N. Li, W. He, H. Zhang, W. Yang *et al.*, 2016 Genome-wide screen of genes imprinted in sorghum endosperm, and the roles of allelic differential cytosine methylation. *Plant J.* 85: 424–436. <https://doi.org/10.1111/tpj.13116>

Communicating editor: N. Springer