Published in "The Plant Journal 95 (6): 1084–1101, 2018"
which should be cited to refer to this work.

Wild tomato endosperm transcriptomes reveal common roles of genomic imprinting in both nuclear and cellular endosperm

Morgane Roth[†], Ana M. Florez-Rueda[‡], Margot Paris[§] and Thomas Städler*
Plant Ecological Genetics, Institute of Integrative Biology & Zurich–Basel Plant Science Center, ETH Zurich, 8092 Zurich, Switzerland

*For correspondence (e-mail thomas.staedler@env.ethz.ch).
[†]Present address: Agroscope, Research Division Plant Breeding, 8820 Wädenswil, Switzerland.
[‡]Present address: Plant Developmental Genetics, Department of Plant and Microbial Biology & Zurich–Basel Plant Science Center, University of Zurich, 8008 Zurich, Switzerland.
[§]Present address: Unit of Ecology & Evolution, Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland.

SUMMARY

Genomic imprinting is a conspicuous feature of the endosperm, a triploid tissue nurturing the embryo and synchronizing angiosperm seed development. An unknown subset of imprinted genes (IGs) is critical for successful seed development and should have highly conserved functions. Recent genome-wide studies have found limited conservation of IGs among distantly related species, but there is a paucity of data from closely related lineages. Moreover, most studies focused on model plants with nuclear endosperm development, and comparisons with properties of IGs in cellular-type endosperm development are lacking. Using laser-assisted microdissection, we characterized parent-specific expression in the cellular endosperm of three wild tomato lineages (Solanum section Lycopersicon). We identified 1025 candidate IGs and 167 with putative homologs previously identified as imprinted in distantly related taxa with nuclear-type endosperm. Forty-two maternally expressed genes (MEGs) and 17 paternally expressed genes (PEGs) exhibited conserved imprinting status across all three lineages, but differences in power to assess imprinted expression imply that the actual degree of conservation might be higher than that directly estimated (20.7% for PEGs and 10.4% for MEGs). Regardless, the level of shared imprinting status was higher for PEGs than for MEGs, indicating dissimilar evolutionary trajectories. Expression-level data suggest distinct epigenetic modulation of MEGs and PEGs, and gene ontology analyses revealed MEGs and PEGs to be enriched for different functions. Importantly, our data provide evidence that MEGs and PEGs interact in modulating both gene expression and the endosperm cell cycle, and uncovered conserved cellular functions of IGs uniting taxa with cellular- and nuclear-type endosperm.

Keywords: genomic imprinting, endosperm, cellular-type endosperm, seed development, epigenetic, laser-assisted dissection, transcriptomics, Solanum, wild tomatoes.

INTRODUCTION

Parent-of-origin-dependent gene expression (i.e. genomic imprinting) is an epigenetic phenomenon with important functions for developmental programs in mammals and flowering plants (Feng *et al.*, 2010). Pioneering experiments on nuclear transplantation in mouse embryos demonstrated that the two parental genomes are not equivalent (McGrath and Solter, 1984; Surani *et al.*, 1984). Disrupted imprinting is responsible for mouse embryo abortion and several human diseases collectively called imprinting disorders (Buiting *et al.*, 1995; Eggermann *et al.*, 2015). Imprinting has also been identified in the mammalian placenta, with mis-imprinting resulting in serious developmental abnormalities (Fowden *et al.*, 2006; McMinn *et al.*, 2006). As a compartment of angiosperm seeds, the endosperm is a tissue functionally equivalent to the mammalian placenta (Scott and Spielman, 2006). This seed compartment results from double fertilization and has a triploid genome constitution (2m:1p); it nurtures the embryo and coordinates development between seed compartments (Berger *et al.*, 2006; Lafon-Placette and Köhler, 2014). Imprinting is restricted to a few genes in angiosperm embryos, but widespread among endosperm-

expressed genes (Luo *et al.*, 2011; García-Aguilar and Gillmor, 2015). Experiments in *Arabidopsis thaliana* have demonstrated that imprinting defects in the endosperm may cause seeds to abort, yet the main functions of imprinting in the endosperm are not fully understood (Gehring and Satyaki, 2017).

Several costs and benefits potentially driving the evolution of genomic imprinting have been formulated (Hurst and McVean, 1998; Haig, 2000; Spencer and Clark, 2014), and can be interpreted in the specific context of flowering plants (Ko *et al.*, 2010). The leading explanatory framework for the evolution of imprinting is the kin conflict theory championed by Haig and Westoby (1989, 1991; Haig, 2013). Their theory posits that imprinting is evolutionarily fuelled by diverging parental interests over resource allocation to offspring. Under conditions of parental conflict, imprinted expression might be a means to balance divergent parental interests and be necessary for proper endosperm (and seed) development. It has been proposed that endosperm growth is optimized when paternally expressed genes (PEGs; supposedly promoting growth) and maternally expressed genes (MEGs; supposedly repressing growth) are jointly expressed with suitable dosage (Leblanc *et al.*, 2002). This would imply that imprinted genes (IGs) are likely to functionally interact. The coordinated expression of MEGs and PEGs might require, or be facilitated by, their organization in functional, regulatory and physical networks likely involving co-adaptation (Holman and Kokko, 2014; Patten *et al.*, 2016). However, the putative functional and physical linkage between IGs has not been extensively studied in flowering plants; such analyses were mostly limited to a low number of candidate genes (Wolff *et al.*, 2011; Zhang *et al.*, 2011).

The kin conflict theory further predicts that imprinting will evolve under strong selective constraints in outcrossing lineages with high levels of multiple paternity, while such constraints might be relaxed under long-term inbreeding (Brandvain and Haig, 2005). To our knowledge, genomic imprinting in the endosperm has been studied in representatives of four angiosperm families (Brassicaceae, Poaceae, Solanaceae and Euphorbiaceae), including: (i) (mainly) inbreeding species such as *A. thaliana* (Wolff *et al.*, 2011; Pignatta *et al.*, 2014), *Capsella rubella* (Hatorangan *et al.*, 2016) and rice (Yuan *et al.*, 2017); and (ii) (mainly) outcrossing species such as *Arabidopsis lyrata* (Klosinska *et al.*, 2016), maize (Zhang *et al.*, 2014), *Sorghum bicolor* (Zhang *et al.*, 2016), *Solanum peruvianum* (Florez-Rueda *et al.*, 2016) and castor bean (Xu *et al.*, 2014). While seed phenotypes of homoploid crosses between closely-related species with different mating systems are consistent with predictions of the kin conflict theory, they have hitherto not been associated with specific imprinting patterns (Rebernig *et al.*, 2015; Lafon-Placette *et al.*, 2017).

Moreover, the extent of imprinting conservation between taxa has proved difficult to quantify and interpret, and is often referred to as being quite limited (Zhang *et al.*, 2011; Pignatta *et al.*, 2014). Some of the suggested explanations for low levels of imprinting conservation are mainly technical. Studies on genomic imprinting involved various dissection methods and developmental time points, as well as different bioinformatics pipelines, likely impacting the number and identity of 'imprinted' genes. Moreover, serious concerns have been raised about maternal contamination of endosperm transcriptomes, found to be widespread in manually dissected *A. thaliana* endosperm samples (Schon and Nodine, 2017). A second, not mutually exclusive, explanation for varying levels of imprinting conservation rests on the potentially diverse evolutionary trajectories of IGs. Specifically, imprinting conservation should only be expected for genes whose imprinted expression is maintained by natural selection for its functional consequences, leaving large scope for evolutionarily ephemeral IGs, for example those whose parentally biased expression may be incidentally caused by nearby insertions of transposable elements (TEs; Gehring *et al.*, 2009; Pignatta *et al.*, 2014). Importantly, the possibility that imprinting mismatch between lineages may be causal for (or contribute to) hybrid seed failure (and thus might promote reproductive isolation and speciation) is increasingly appreciated (Florez-Rueda *et al.*, 2016; Lafon-Placette and Köhler, 2016). Despite these conceptual and empirical leads, qualitative and quantitative comparisons of genomic imprinting between closely related lineages have only just started to emerge (Hatorangan *et al.*, 2016; Klosinska *et al.*, 2016).

Wild tomatoes (Solanum section Lycopersicon) provide an excellent model system to study the incidence, evolution and consequences of imprinting in the endosperm; they diverged fairly recently (~2.5 Mya; Pease *et al.*, 2016), harbor a diversity of mating systems and exhibit various degrees of post-zygotic isolation mediated by endosperm-based hybrid seed failure (Baek *et al.*, 2016; Roth *et al.*, 2018). Unlike all other taxa that have been the focus of studies on endosperm genomic imprinting, Solanum is characterized by a cellular- rather than a nuclear-type endosperm during early development. The cellular-type endosperm development implies that there is no transition from an initial syncytial to a later cellular stage (Vijayaraghavan and Prabhakar, 1984), and is thought to be ancestral to other endosperm types (Floyd and Friedman, 2000). To date, no study has probed the conservation of imprinting across lineages with different developmental types of endosperm.

Using laser-assisted microdissection of developing endosperm from three reciprocal crosses and subsequent transcriptome sequencing, we have characterized and compared imprinting landscapes in three self-incompatible

wild tomato lineages: *Solanum arcanum* var. marañón (*A*), *Solanum chilense* (*C*) and *S. peruvianum* (*P*). These lineages are morphologically and ecologically well differentiated (Rick and Lamm, 1955; Rick, 1986; Peralta *et al.*, 2008), and are reproductively isolated from each other to varying degrees by endosperm-based hybrid seed failure, causing partial to near-complete seed abortion (Roth *et al.*, 2018). Using estimates of maternal expression proportions for between 7730 and 13 198 genes per lineage, our study characterizes genomic imprinting and probes the level of imprinting divergence between the three closely related wild tomato lineages. In addition, we have assessed the putative genomic localization and functions of IGs as well as their predicted protein interactions. Finally, we also provide insights into the potential mechanisms underlying parent-specific expression (PSE) in the developing Solanum endosperm.

RESULTS

Identification of IGs and their degree of conservation

By generating RNA-Seq data from developing endosperm in replicated reciprocal crosses, we assessed genomic imprinting in the three wild tomato lineages *A*, *C* and *P*. To maximize power to assess PSE, the parental combinations were chosen based on high inter-individual molecular polymorphism as well as high seed viability (Roth *et al.*, 2018). Biological variation was taken into consideration by using three clonal replicates for each cross, resulting in a total of 18 endosperm samples. Requiring the use of laser microdissection, we sampled developing endosperm tissue at 12 days after pollination (DAP), corresponding to the early globular embryo stage in all three crosses. Mapping against the SL2.50 tomato reference generated an average of 83.9% mapped reads with mapping quality >20. Using the ITAG2.4 tomato genome annotation, this allowed us to quantify PSE for 7730 (*A*), 9623 (*C*) and 13 198 (*P*) nuclear genes, often based on multiple single nucleotide polymorphisms (SNPs) per gene (Table 1). In this set of genes, 6169 were found commonly expressed and polymorphic in all three reciprocal crosses. As pioneered by Florez-Rueda *et al.* (2016), we inferred maternal proportions from both homozygous and heterozygous parental SNPs (Figure 1; Tables 1 and S1). Using empirically derived cross-specific thresholds of maternal proportion, we identified a total of 812 candidate MEGs and 213 candidate PEGs among the three reciprocal crosses (Figure 2; Table 1). Importantly, we did not observe any reversal in imprinting status between lineages, i.e. no changes from candidate MEG to candidate PEG or vice versa.

Following Schon and Nodine (2017), we probed for potential seed-coat and embryo contamination of our endosperm transcriptomes. Using tissue-specific transcriptome data obtained via laser-capture microdissection from early developing wild tomato seeds as reference data (Pattison *et al.*, 2015), there was no discernible signal of seed-coat or embryo contamination in any of our endosperm samples (Figure S1; all *P*-values = 1).

As imprinting assessment relies on parental polymorphisms that are specific to each cross, quantifying imprinting conservation across species faces uncertainties that deserve particular attention. To evaluate the proportion of 'conserved' IGs between two or three taxa, one has to consider the common gene universe for which PSE can be assessed. In our data, this set is represented by the 6169 genes with parental SNPs in each of the three reciprocal crosses that are expressed in developing endosperm at 12 DAP. Within this set of genes, we found 485 to be candidate imprinted in at least one lineage (47.3% of all candidate IGs; Figure 3; Table S2). Among them, 59 (12.2%) are candidate imprinted in all three crosses; this conserved status is shown by 42 MEGs (10.4% of all assessable candidate MEGs) and 17 PEGs (20.7% of all assessable candidate PEGs; Figure 3; Table S3). As the probability of randomly sampling the same genes in the *A*, *C* and *P* gene universes is very low, this degree of overlap is highly significant (10 000 iterations, *t*-test, $P < 2.2e\text{-}16$ for PEGs and MEGs). In pairwise comparisons among lineages, conservation in imprinting status ranged between 19.9 and 24.6% for candidate MEGs, and between 31.3 and 39.3% for candidate PEGs. Overall, these data indicate a higher conservation of PEGs compared with MEGs in these wild tomato lineages. Considering candidate MEGs and PEGs together, we found the highest imprinting overlap between *C* and *P* (26.7%), closely followed by *A* and *P* (25.2%), and *A* and *C* (21.9%).

We also assessed the average maternal proportion of genes that failed to reach our imprinting thresholds in one or two species, conditional on those genes being candidate imprinted in one or two of the three crosses. Such non-conserved IGs exhibited average maternal proportions of 78.7% (SD 8.7%) for candidate MEGs and 41.6% (SD 12.7%) for candidate PEGs when they failed to reach imprinting thresholds. This indicates that many of the non-conserved PEGs also showed paternal expression bias and were close to reaching imprinting thresholds; thus, the 'true' conservation among PEGs might be higher than estimated above.

Finally, using a different rationale, we estimated the potential maximum and minimum levels of imprinting conservation from our data. For this, we also took all genes into account whose imprinting conservation could not be assessed empirically due to the absence of sequence polymorphism in one or two of the three crosses. We considered two extreme scenarios where: (i) all genes lacking parental polymorphism are also imprinted like in the informative cross(es); and (ii) all such genes are not imprinted (or have a different imprinting status). This yielded estimates of imprinting conservation across all three lineages

Table 1 Summary of data underlying the estimation of maternal proportions (and numbers of candidate IGs) in endosperm-expressed genes in three wild tomato crosses (lineages)

Statistic	*A* × *A* 2185A → 1626B	*C* × *C* 4329B → 2748B	*P* × *P* 2744B → 2964A
Endosperm-expressed genes with SNPs in reciprocal cross	7730	9623	13 198
Alternative homozygous sites	19 274	19 053	61 552
Heterozygous sites (e.g. CC:AC)	17 195	56 092	127 036
Mean no. of SNPs per gene	4.7	7.8	14.3
Candidate MEGs	356	316	365
Candidate PEGs	66	94	133
Median maternal proportion (across all genes)	0.722 → 0.758	0.693 → 0.757	0.728 → 0.731

MEG, maternally expressed gene; PEG, paternally expressed gene; SNP, single nucleotide polymorphism.
A, Solanum arcanum var. marañón; *C, Solanum chilense*; *P, Solanum peruvianum*. Individual parental genotypes are abbreviated with their four-digit TGRC accession number (e.g. 2185, accession LA2185), followed by a capital letter.

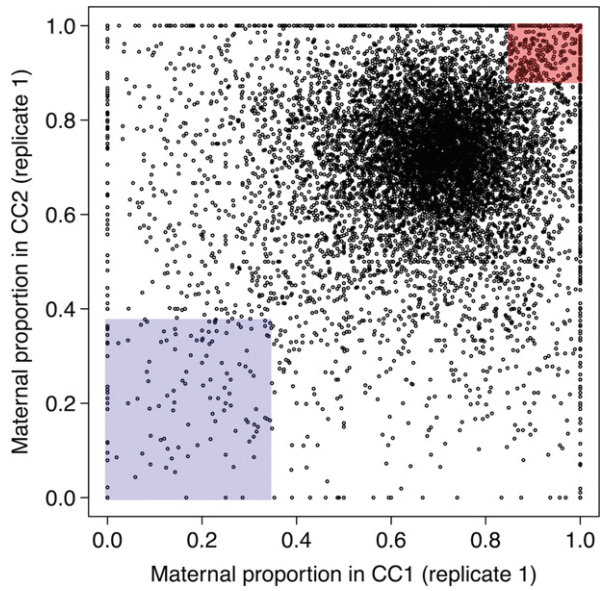

Figure 1. XY-plot representing the distribution of maternal proportion per gene (MAT/TOT allelic counts per gene; 9593 genes) in one biological replicate of the reciprocal cross *CC*.
The first replicates of each cross direction (*CC1* and *CC2*) are shown. Red and blue areas: maternally and paternally biased genes in *CC1* (replicate 1) and *CC2* (replicate 1). Genes located in these highlighted areas may not have been retained as candidate imprinted genes (IGs) because to be considered imprinted, parental bias had to be evident in at least two replicates per cross direction (full data in Table S1).

of 5.2–30.0% (min–max) for MEGs and 8.0–45.5% for PEGs; the maximum overlap could result in 244 MEGs and 97 PEGs (Figure S2; Table S2).

Conserved imprinting among distantly related species is moderate but highlights important functions

Based on previous studies, we screened our set of Solanum candidate IGs for sequence similarity with genes (whether or not imprinted) in *A. thaliana* (*At*), *A. lyrata* (*Al*), *C. rubella* (*Cr*), maize (*Zm*), rice (*Os*), *S. bicolor* (*Sb*) and castor bean (*Rc*). Using BLASTx and stringent filtering criteria, we identified potential homologs to our candidate IGs in these seven model species (Table S4). Specifically, we found that 127 Solanum MEGs and 35 PEGs had the best match (lowest *e*-value) with a gene identified as imprinted in at least one other study (15.6% of candidate MEGs and 16.4% of candidate PEGs). Among the 178 pairs of Solanum IGs and their putative homologs from distantly related species, we identified 'conserved' imprinting status for 147 pairs (82.6%). The other 31 gene pairs (17.4%) showed 'reversed' imprinting status between Solanum and the distantly related taxa (Table S4).

We also used the program OMA to identify IGs in model species that are putative orthologs to Solanum candidate IGs. We thus found 34 Solanum genes, mainly transcription factors (TFs), with putative one-to-one orthologs in one of the model species (Table S4). Among them, 29 gene pairs coincide with the best hits found with the Blastx searches. In most cases, the direction of parental bias was conserved, except for one gene each in *At*, *Al* and rice. This set of putatively orthologous genes represents strong candidates for conserved imprinting across distantly related angiosperm species.

Moreover, among the 167 Solanum candidate IGs with evidence for conserved imprinted expression among distantly related species, we found 11 MEGs and five PEGs with evidence for imprinting in two or three non-Solanum model species, including some with reversed imprinting status (Table 2). These 16 genes appear to primarily code for regulatory elements, such as two RING-domain proteins with conserved MEG status across the tomato lineage *A, Ricinus communis* and *A. thaliana* (the imprinting status in lineages *C* and *P* could not be assessed due to lack of parental sequence differences). These RING-domain

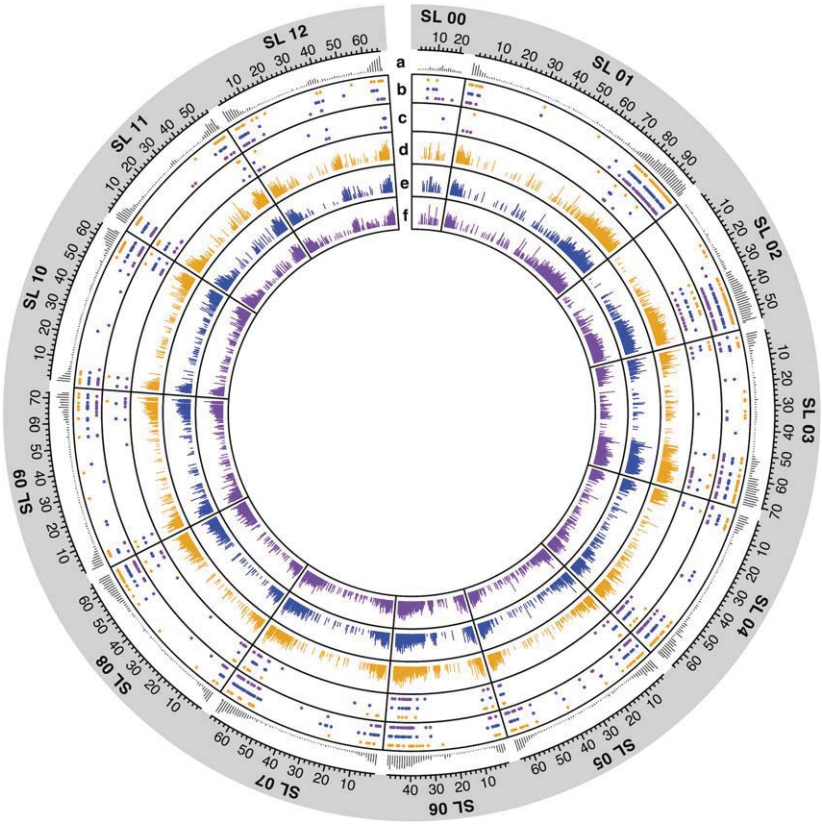

Figure 2. Circos graph representing gene expression and candidate imprinted genes (IGs) in each wild tomato lineage (14 021 genes represented in total). Orange, *Solanum arcanum* var. marañón; blue, *Solanum chilense*; purple, *Solanum peruvianum*. SL00-SL12, linkage groups as given in the tomato genome reference SL2.50. Circles (a–f): (a) gene density along the cultivated tomato reference genome (The Tomato Genome Consortium, 2012); gene density is summarized as the number of genes in consecutive 1-Mb bins; (b) mapping of candidate maternally expressed genes (MEGs); (c) mapping of candidate paternally expressed genes (PEGs); expression levels in *S. arcanum* var. marañón (d), *S. chilense* (e) and *S. peruvianum* (f) intraspecific crosses [transcripts pre million (TPM) averaged across biological replicates].

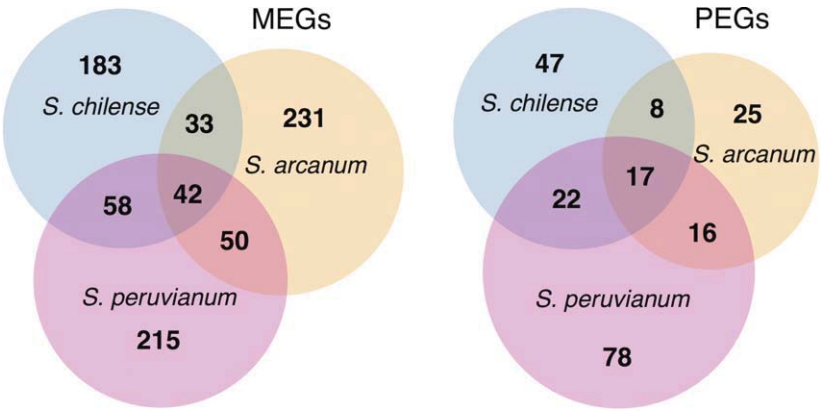

Figure 3. Venn diagrams displaying the overlap of candidate imprinted genes (IGs) assessed in three tomato lineages. Maternally expressed genes (MEGs; $n = 812$); paternally expressed genes (PEGs; $n = 213$).

proteins might be components of the Skp1–Cullin–F-box (SCF) E3 ubiquitin ligase, a protein complex involved in gene repression that might also control cell-cycle switches in the endosperm (Kim *et al.*, 2008; Dumbliauskas *et al.*, 2011).

Levels of gene expression are moderately impacted by imprinting status

Expression levels were assessed for each library by counting transcripts per gene with HTseq and transforming

these counts to transcripts per million (TPM). While 24.8% of all expressed genes were expressed in only one or two lineages, genes with variable imprinting status were invariably expressed across all three taxa. This indicates that in our study, imprinting status of a given gene is not likely to be lost due to missing data (e.g. very low expression) but rather by expression modulation. To study potential co-variation of imprinting status and expression level, we considered expression levels only for genes for which PSE could also be assessed (overall, $n = 30\,537$ observations; Table S5). Variance analysis (ANOVA) revealed that expression levels were influenced by species ($P < 2.2e\text{-}16$), imprinting status ($P = 1.17e\text{-}06$) and by their interaction ($P = 9.1e\text{-}04$). However, comparisons of means did not discriminate between interaction levels (species × imprinting status; Tukey's test, all $P > 0.05$; Figure 4).

Interestingly, IGs conserved across all three lineages showed higher expression levels than candidate IGs with variable imprinting status [Wilcoxon rank sum test (WRST), PEGs, $P = 1.3e\text{-}03$; MEGs, $P = 9.2e\text{-}08$]. Among these 59 conserved IGs, MEGs revealed higher expression in *P* compared with both *C* and *A* (paired WRST *P–C*, $P = 0.04$; *P–A*, $P = 0.04$; Figure 5), and PEGs higher expression in *P* compared with *C* but similar expression levels to *A* (paired WRST *P–C*, $P = 0.016$; *P–A*, $P = 0.14$; Figure 5). No significant differences in expression levels of conserved IGs were found between *A* and *C* (paired WRST, MEGs, $P = 0.94$; PEGs, $P = 0.37$; Figure 5). These differences in expression levels might reflect higher levels of parental conflict in lineage *P* compared with both *A* and *C* (see Discussion).

Next, we examined whether expression level changes with imprinting status for a given gene. We restricted this analysis to the 426 genes exhibiting variable imprinting status across the three lineages [referred to hereafter as NI/MEGs ($n = 361$) and NI/PEGs ($n = 65$)]. In the context of the triploid endosperm, an additive model of gene expression predicts that silencing the paternal or maternal gene copies reduces gene expression by 1/3 and 2/3, respectively, compared with biparentally expressed genes. In both cases, the expectation under this scenario is lower gene expression when imprinted than when not imprinted. We found no evidence for reduced expression correlating with imprinting status. Instead, expression of NI/MEGs was higher when genes were imprinted (paired WRST, $P = 0.017$), and expression of NI/PEGs was not significantly different between imprinted and non-imprinted state (paired WRST, $P = 0.34$). This suggests that distinct mechanisms regulate the expression of PEGs and MEGs with variable imprinting status, MEGs being specifically upregulated when imprinted.

IGs are involved in metabolism, transcription regulation and cell cycle

Biological functions of endosperm-IGs have been characterized in relatively few species and appear to be quite diverse. It is thus still unclear whether some functions are key to IGs and whether these functions are conserved across species. In our study, IGs were found to be mainly involved in biosynthetic and catabolic processes, transcription regulation, signaling and development (Figure 6). There is evidence that IGs are strongly involved in gene expression regulation, as the set of candidate genes contains 38 TFs (29 MEGs and 9 PEGs; Table S2), the main gene families being BZIP, MYB, GATA and MADS-box. We also found 17 imprinted methyltransferases (11 MEGs and 6 PEGs; Table S2), which are key mediators of gene silencing. Three helicases have also been identified; one is a SNF2-related helicase (PEG), a family deemed crucial for rice endosperm development (Hara *et al.*, 2015). One MEG and four PEGs are mRNA splicing factors that may be essential for the correct production of isoforms specific to the endosperm (Lu *et al.*, 2013).

Because MEGs and PEGs might comprise different functions, we analyzed their associated gene ontology (GO) terms separately (Table S6). MEGs and PEGs were significantly enriched for many biological processes, molecular functions and cellular components. MEGs were mainly enriched for metabolic processes (GO:0008152), such as carbohydrate transport and metabolism (GO:0004556, 0000271), cell wall (GO:0042546, 0048046), and in signaling via hormone regulation (GO:0010817) such as auxin (GO:0009734) and stress response (GO:0034599). PEGs were enriched for several chromatin-binding functions, such as histones (GO:0044154, 0043972, 0043971) and euchromatin binding (GO:1990188). They were also involved in amino acid and nucleotide transport (GO:0015802, 0043090, 0006862) and signaling (GO:0001789, 0007186). Lastly, they were also enriched for ubiquitination (GO:0036459) and post-transcriptional silencing by RNA activities (GO:0035194). Jointly, these results may indicate that PEGs contribute more specifically to cell-cycle control and gene regulation. Interestingly, 53 MEGs and five PEGs are nuclear-encoded chloroplastic genes, while 19 MEGs and two PEGs are related to mitochondrial function (ITAG2.4 and Panther database; Table S2).

To assess how IGs' encoded proteins may interact, we utilized the available information from several databases using the software STRING (see Experimental Procedures). The majority of interaction networks involved both MEGs and PEGs (Figure 7; Table S7). We found 179 significant pairwise interactions between IGs, 33 of them between MEGs and PEGs, 140 between MEGs and six between PEGs. In particular, the MEG Solyc05g018300.2, a

Table 2 Candidate IGs in Solanum with potential orthologs in two or three distantly related model species

Gene ID tomato	Imprinting in Solanum	Annotation in tomato (ITAG 2.4)	*Arabidopsis lyrata*	*Arabidopsis thaliana*	*Capsella rubella*	*Oryza sativa*	*Sorghum bicolor*	*Zea mays*	*Ricinus communis*
Solyc01g080070.2	MEG in AP	Copper chaperone		AT3G06130.2		Os12g05040.1			
Solyc01g099160.2	MEG in P	Lipoxygenase		AT1G55020.1			Sobic.001G125900.1		
Solyc02g083400.2	MEG in A(CP)	RING finger protein 38		AT2G17730.1					29784.m000358
Solyc02g085180.2	MEG in AC	Catalytic hydrolase		AT2G18360.1	Carub.0007s0372.1				
Solyc02g089620.2	MEG in ACP	Proline dehydrogenase					Sobic.001G304700.2	Zm00001d029853	
Solyc03g033330.2	MEG in A(CP)	RING finger protein 44		AT4G35840.1					29784.m000358
Solyc03g115860.2	MEG in AC	Endoplasmic reticulum membrane protein YGL010W		AT1G18720.2			Sobic.004G191500.1		
Solyc06g005500.2	MEG in CP(A)	ATP binding/serine-threonine kinase				Os03g60710.1		Zm00001d034662	
Solyc06g076770.2	MEG in P	Myb family transcription factor		AT5G58900.1	Carub.0002s1670.1				
Solyc07g054270.2	MEG in C	Os08g0100600 protein (fragment)–unknown function		AT4G03420.1			Sobic.001G207700.1		
Solyc10g007410.2	MEG in A(C)	Gamma-glutamyl hydrolase 2		AT1G78680.1			Sobic.009G199200.1		
Solyc02g084510.2	PEG in CP(A)	Novel protein containing a PHD-finger domain (fragment)	AL937974	AT3G51120.1				Zm00001d027334	
Solyc07g006760.2	PEG in A(C)	Serine/threonine protein kinase				Os02g54510.3			30179.m000563
Solyc08g080110.2	PEG in C	Fam63a protein					Sobic.010G261600.1	Zm00001d014761	
Solyc10g074870.1	PEG in P(A)	KH domain cont. RNA binding signal transduction associated 1				Os07g12490.1		Zm00001d019215	
Solyc12g049540.1	PEG in P(A)	Polyphosphoinositide phosphatase		AT3G14205.1	Carub.0003s1388.1				

MEG, maternally expressed gene; PEG, paternally expressed gene.

Inference on potential orthology is based on Blastx analyses and/or OMA 'one-to-one' function (see Experimental procedures and Table S4). Colors indicate the imprinting status referenced for the given species (red, MEGs; blue, PEGs). Wild tomato lineages (*A*, *C*, *P*) listed in parentheses in the second column could not be assessed for imprinting status due to lack of parental sequence polymorphism.

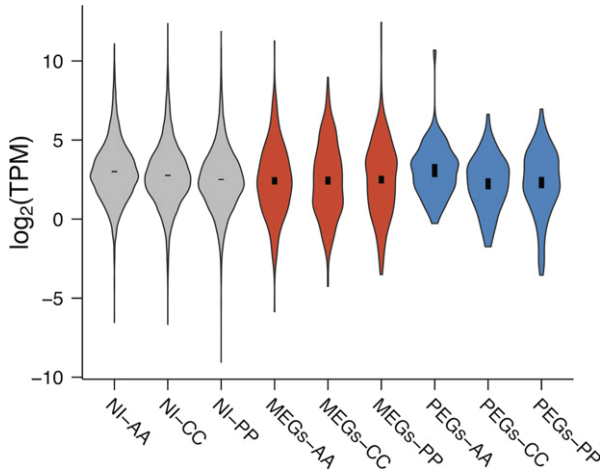

Figure 4. Violin plots representing the distribution of expression levels in the endosperm across all 'imprinting by species' categories [log_2(TPM) values, 14 021 genes in total].
Expression values are averaged across replicates. TPM, transcripts per million. Gray, non-imprinted genes (NI); red, maternally expressed genes (MEGs); blue, paternally expressed genes (PEGs); intraspecific crosses within *Solanum arcanum* var. marañón (*AA*), *Solanum chilense* (*CC*) and *Solanum peruvianum* (*PP*). NI-*AA*, 7304 genes; NI-*CC*, 9208 genes; NI-*PP*, 12 695 genes; MEGs-*AA*, 356 genes; MEGs-*CC*, 316 genes; MEGs-*PP*, 365 genes; PEGs-*AA*, 66 genes; PEGs-*CC*, 94 genes; PEGs-*PP*, 133 genes.

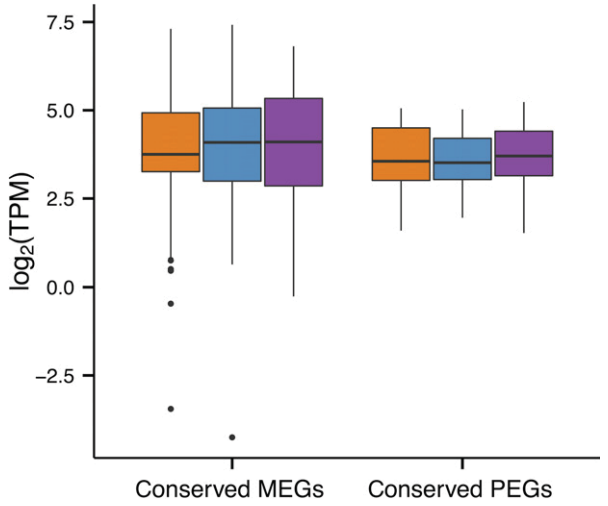

Figure 5. Box plots representing the distribution of expression levels [log_2(TPM) values] across lineages *A* (orange), *C* (blue) and *P* (purple) for conserved maternally expressed genes (MEGs; 42 genes) and paternally expressed genes (PEGs; 17 genes). TPM, transcripts per million.

phosphatase 2C, interacts with 22 other IGs (10 PEGs and 12 MEGs); it is a hub between DNA-binding elements, ribosomal proteins and amylases (Table S7). We further explored whether additional genes might potentially interact with our candidate IGs, and thus highlight key networks or pathways. Among the 10 inferred interactors, three are interconnected hsp40 chaperone proteins (Solyc04g081570.2, Solyc04g081630.1, Solyc04g081640.1), interacting mainly with DNA-binding elements, other chaperones and serine/threonine kinases (Table S7). The Solyc12g044540.1 protein is at the center of a large cluster of IGs, mainly serine/threonine kinases represented by 26 MEGs and six PEGs. It has no functional annotation but is predicted to be a phosphatase 2C in *Solanum lycopersicum* (Blastx, *e* = 2.0e-24). Due to sequence identity in all three crosses, we were unable to assess the imprinting status of this gene.

IGs tend to be physically clustered

Candidate IGs were located on all chromosomes; however, MEGs were particularly abundant on chromosome 1 (*n* = 127) and PEGs on chromosome 3 (*n* = 28; Table S8). We tested for clustering in non-overlapping 100-kb bins using the full set of 1025 candidate IGs across all three lineages. We thus found 164 significant clusters or physically close pairs of IGs representing 373 genes [one-sample *t*-test, $P < 0.05$ with false discovery rate (FDR) correction 0.05]. They either contained only MEGs (*n* = 104, 2–10 genes), only PEGs (*n* = 15, 2–4 genes) or both classes of IGs (*n* = 45, 2–5 genes). The densest cluster was located on chromosome 1, where 10 out of 13 genes (expressed and polymorphic) are candidate MEGs (SL2.50ch01_bin19; Table S8). Among them, nine genes have a chloroplastic function and are involved in translation (five are ribosomal proteins).

The next-largest cluster of MEGs contained four genes exclusively imprinted in lineage *A* (SL2.50ch05_bin17). The largest cluster of PEGs contained regulatory elements with a RING-domain and an F-box, putative elements from the SCF complex (SL2.50ch08_bin16). MEGs and PEGs also clustered jointly within each of 45 clusters of 2–5 genes. In cluster SL2.50ch10_bin610, the two PEGs and two MEGs are involved in transcription regulation. Interestingly, we observed 10 bins each comprising two IGs with identical annotation. In all cases, these tandemly arranged genes were imprinted for the same parent. These observations suggest that IG clustering is extensive in wild tomatoes, and could potentially result from gene duplication of IGs together with shared *cis*-regulatory elements. Given that we could only analyze polymorphic genes, we may expect this physical clustering of IGs to be even more pronounced than could be assessed with the data at hand.

DISCUSSION

To characterize and compare patterns of genomic imprinting in three closely-related wild tomato lineages, we conducted a large RNA-Seq experiment on laser-microdissected endosperms at the early-globular stage. Importantly, we found no evidence of RNA contamination from maternal seed-coat or embryo tissues in our endosperm samples, confirming that laser microdissection can achieve high tissue specificity (Schon and Nodine, 2017). Moreover, our experimental design allowed to take biological

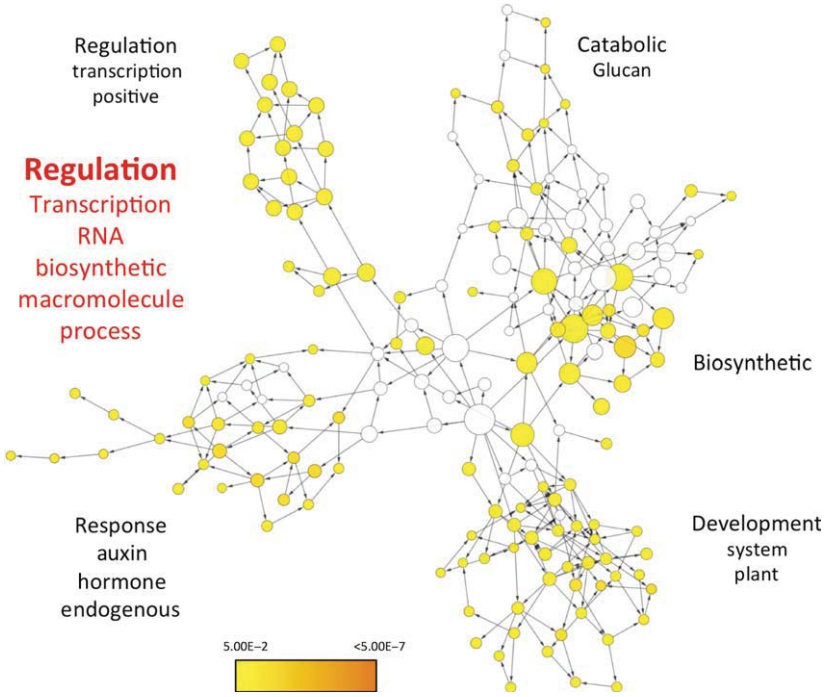

Figure 6. Over-represented gene ontology (GO) terms among candidate imprinted genes (IGs) and interactions between GO terms (biological process terms only).
Each polygon represents a GO term and its size is proportional to the number of genes associated with this term. Color scale: level of significance for each over-represented GO term (*P*-value); white: non-significant. Main functions found with WordCloud (word enrichment) are annotated: black, word enrichment on locally nested GO terms; red, word enrichment on all enriched GO terms.

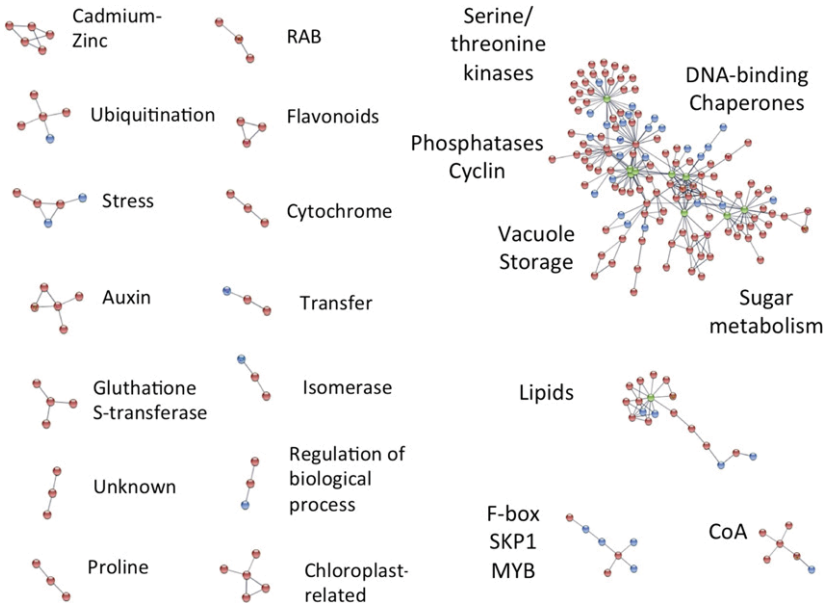

Figure 7. Protein–protein associations among candidate imprinted genes (IGs) identified in *Solanum arcanum* var. marañón, *Solanum chilense* and *Solanum peruvianum* intraspecific endosperms, and significant interacting genes found by the STRING database (348 interactions in total).
Red, maternally expressed genes (MEGs); blue, paternally expressed genes (PEGs); green, genes significantly interacting with several candidate IGs.

variability of PSE into account; this is of particular importance when measuring expression levels of lowly expressed genes from partially degraded tissues (Sims *et al.*, 2014; Anjam *et al.*, 2016). For each of the three species crosses, we identified several hundreds of genes with stable imprinting signatures across replicates.

PEGs and MEGs exhibit different degrees of conservation

In common with previous studies on plants (Waters *et al.*, 2013; Hatorangan *et al.*, 2016; Klosinska *et al.*, 2016), we quantified the observed proportion of conserved imprinting status among lineages *A*, *C* and *P*, and found PEGs to be more conserved than MEGs (ranges for pairwise comparisons: PEGs 31.3–39.3%, MEGs 19.9–24.6%). We also estimated the considerable biological uncertainties associated with measures of imprinting conservation; these uncertainties partly stem from many endosperm-expressed genes with no parental sequence polymorphisms in our crosses. This revealed that imprinting overlap among all three lineages could vary between 5.2 and 30.0% (min–max) for MEGs, and between 8.0 and 45.5% for PEGs. To our knowledge, such uncertainties have not been estimated in other studies. Moreover, the group of non-conserved PEGs showed marked expression bias toward paternal expression (while not reaching imprinting thresholds), implying that the actual degree of imprinting conservation for PEGs is likely to be somewhat higher than estimated here. The apparent stronger conservation of PEGs in wild tomatoes agrees with comparative imprinting analyses in the Brassicaceae (*A. thaliana*–*C. rubella*: 28.6% for PEGs and 14.3% for MEGs; Hatorangan *et al.*, 2016; *A. thaliana*–*A. lyrata*: 50% for PEGs and 35% for MEGs; Klosinska *et al.*, 2016).

These replicated patterns suggest differences in selective pressures impacting the conservation of PEG- versus MEG-status over evolutionary time. Although a trend for faster evolution of MEGs compared with PEGs has been observed in *A. thaliana* (Wolff *et al.*, 2011), no significant differences in sequence evolution between PEGs and MEGs were evident using d_N/d_S ratio comparisons in *A. thaliana* and maize (Wolff *et al.*, 2011; Waters *et al.*, 2013), or by comparisons of rates of adaptive relative to neutral substitutions in *C. rubella* (ω_a; Hatorangan *et al.*, 2016). Thus, further investigations are needed to explore patterns of sequence evolution in MEGs and PEGs, for example by repeating such analyses with more individuals and more IGs, preferentially in outcrossing species where IGs are expected to evolve under stronger parental conflict (Brandvain and Haig, 2005; Haig, 2013). Data on endosperm growth and seed size in interspecific crosses between these three lineages suggest that lineage *P* experienced higher levels of parental conflict compared with both *A* and *C* (Roth *et al.*, 2018), consistent with its higher level of nucleotide diversity (Städler *et al.*, 2008; Tellier *et al.*, 2011; Beddows *et al.*, 2017). Interestingly, we found that conserved IGs tend to have a higher expression in lineage *P* compared with both *A* and *C* (Figure 5), possibly reflecting higher levels of parental conflict and mirroring observations in *A. lyrata*–*A. thaliana* comparisons (Klosinska *et al.*, 2016). Whether directly or indirectly, such expression differences might be involved in the molecular underpinnings of hybrid seed failure observed between these wild tomato lineages (Florez-Rueda *et al.*, 2016; Roth *et al.*, 2018).

Imprinting is moderately conserved across closely and distantly related taxa

As the knockout of single IGs has been shown to be sufficient to induce seed abortion, these appear to be essential for correct seed development and thus potentially conserved across angiosperms (Grossniklaus *et al.*, 2001; Köhler *et al.*, 2003). The 59 candidate IGs identified as being conserved across all three wild tomato lineages represent a non-random overlap, and provide insights into key functions borne by PEGs and MEGs. Moreover, a moderate proportion of our candidate IGs (15.8%) revealed potentially orthologous ('best hit') IGs identified in distantly related plant species, possibly highlighting major functions of a subset of plant IGs. Similarly, Zhang *et al.* (2016) recently reported a potential 30% overlap between their candidate IGs in *S. bicolor* and imprinted potential homologs in maize, rice and *A. thaliana*. Importantly, common functions specific to MEGs and PEGs emerged from our and previous studies, suggesting that they are not tied to the mode of endosperm development. Across Solanum, maize, Sorghum, Arabidopsis and Capsella, PEGs and MEGs are often TFs, but MEGs are primarily involved in transcription regulation, signaling and biosynthetic activities while PEGs are more specifically enriched for chromatin binding and chromatin modifications (Waters *et al.*, 2013; Pignatta *et al.*, 2014; Hatorangan *et al.*, 2016; Zhang *et al.*, 2016).

Any differences in suites of IGs among taxa likely reflect underlying genetic, epigenetic and regulatory divergence arising during and after speciation (Josefsson *et al.*, 2006; Birchler, 2014). In our study system, the closely related taxa *C* and *P* (estimated divergence time 1.25 Mya; Beddows *et al.*, 2017) reveal only slightly higher levels of imprinting conservation (26.7% of all candidate IGs) than the more distantly related pairs *A*–*P* (25.2%) and *A*–*C* (21.9%) that both should be equally diverged (approximately 2 Mya; Pease *et al.*, 2016). Our estimates of IG conservation are overall higher than those obtained for comparisons between the much more diverged *A. thaliana* and *C. rubella* (10–14 Mya; Mitchell-Olds, 2001; Koch and Kiefer, 2005), sharing 20% of their IGs (Hatorangan *et al.*, 2016). However, they are lower than those conservatively estimated between *A. thaliana* and *A. lyrata* (45.3%; Klosinska *et al.*, 2016), which are also markedly more diverged than wild tomatoes (13 Mya; Beilstein *et al.*, 2010). Disregarding potential contributions of maternal RNA contamination in hand-dissected endosperm datasets (Schon and Nodine, 2017) and differences in statistical criteria and power to call candidate IGs between studies, the

proportion of overlapping IGs does not linearly reflect estimates of divergence time between pairs of lineages. However, revealing no instances of reversal between MEG and PEG status might better reflect the fairly recent divergence time of wild tomato lineages. Analyses in Arabidopsis have provided evidence that gene duplication events impacted the incidence and expression profiles of IGs, and that the diverging fates of paralogs have likely fuelled cases of neofunctionalization (Qiu *et al.*, 2014); such processes might largely account for the cases of reversed imprinting status we have uncovered between Solanum and distantly related species (Tables 2 and S4).

The majority of tomato candidate IGs were specific to one or two wild tomato lineages, suggesting their relatively fast turnover over evolutionary timescales. This is consistent with evidence for imprinting polymorphism among genotypes in maize (Waters *et al.*, 2013) and *A. thaliana* (Pignatta *et al.*, 2014), thought to be driven mostly by insertion and turnover of TEs and TE-specific silencing machinery (Gehring *et al.*, 2009; Pignatta *et al.*, 2014; Klosinska *et al.*, 2016). Polymorphism of imprinting status has recently been proposed as a mechanism generating phenotypic polymorphism and promoting efficient diversifying selection on seed traits (Bai and Settles, 2015). This might be especially the case for the less-conserved MEGs; their higher rate of evolutionary turnover might reflect processes of co-adaptation that could contribute to local adaptation of the growing seed (Wolf and Hager, 2006; Holman and Kokko, 2014).

PEGs and MEGs appear to be regulated by distinct mechanisms

In our overall data, about 25% of all genes were not expressed in all three lineages (i.e. they were 'repressed' in one or two lineages). In contrast, non-conserved candidate IGs were never found completely repressed in any of the lineages. This might imply that their biological functions are essential for seed development, irrespective of their imprinting status. It certainly seems plausible that important functions of IGs are conserved across species, rather than their individual identities (Gehring and Satyaki, 2017).

Imprinting has been proposed as a general mechanism to attain specific expression levels of functionally important genes (Haig, 2000; Patten *et al.*, 2014; Wolf *et al.*, 2014). It was recently shown that imprinted *A. lyrata* genes maintain higher expression than non-IGs (Klosinska *et al.*, 2016). Comparing expression levels between non-imprinted and candidate IGs in wild tomatoes, our results show no clear signal for up- or downregulation of candidate IGs at genome-wide scales (Figure 4). This may simply reflect the broad range of expression optima among the suite of endosperm-expressed genes. Nevertheless, conserved IGs showed increased gene expression compared with genes with variable imprinting status, which might be interpreted in different ways. On the one hand, genes with higher expression levels may be more likely to be detected as parentally biased in all lineages. On the other hand, conserved IGs may be more likely to be essential for seed development, thus requiring relatively high expression levels.

To further quantify the co-variation of imprinting and gene expression levels, we selected genes with varying imprinting status (i.e. non-conserved across the three lineages), and compared expression levels between non-imprinted and imprinted states. We found that genes with variable imprinting status tend to be slightly upregulated when maternally imprinted and expressed at similar levels when paternally imprinted, compared with their expression as NIs. This suggests that distinct mechanisms regulate the expression of PEGs versus MEGs. Imprinting is set either by allele-specific derepression of methylated genes, involving DNA glycosylases (Choi *et al.*, 2002; Gehring *et al.*, 2004; Kinoshita *et al.*, 2004; Bauer and Fischer, 2011), and/or by allele-specific silencing via histone modifications through the activity of the PRC2 protein complex (Hsieh *et al.*, 2011; Ikeda, 2012; Raissig *et al.*, 2013). In light of our results, the expression of MEGs might be mostly caused by derepression or upregulation of the maternal allele leading to increased gene expression (whether or not the paternal allele is completely silenced). For PEGs, the paternal allele might be sufficiently upregulated to compensate for the silencing of the maternal allele. This would correspond to a model where MEGs are more likely regulated by allele-specific activation and PEGs by allele-specific silencing of the maternal and upregulation of the paternal allele. The well-characterized imprinted *A. thaliana* genes *MEDEA* (MEG) and *PHERES1* (PEG) illustrate this inference; the maternally derived *MEDEA* allele is activated in the central cell by DEMETER, with *MEDEA* remaining silent in *dme* mutants (Gehring *et al.*, 2006), while the maternally derived *PHERES1* allele is repressed by MEDEA, with *PHERES1* being overexpressed in *mea* mutants (Köhler *et al.*, 2003, 2005). Along the same lines, it was recently found that MEGs are more likely to be endosperm-specific (i.e. repressed in other tissues), implying that their maternal alleles are specifically activated in the endosperm (Klosinska *et al.*, 2016).

PEGs and MEGs may be physically and functionally linked

Mammalian IGs occur essentially in clusters (Brannan and Bartolomei, 1999); some imprinted regions have been well-characterized in mouse and human embryos where the expression of neighboring IGs is regulated by so-called imprinting control regions (Verona *et al.*, 2003; Edwards and Ferguson-Smith, 2007). IGs in mammals not only co-localize and share common *cis*-regulatory elements, but also regulate each other in IG networks (Varrault *et al.*,

2006; Gabory *et al.*, 2009). This physical clustering could emerge from the proliferation of IGs via gene duplication and/or from shared regulatory elements (Walter and Paulsen, 2003; Holman and Kokko, 2014). Physical clusters comprising a few IGs have been identified in *A. thaliana* and maize, but do not tend to be a common feature of plant genomes (Feil and Berger, 2007; Hsieh *et al.*, 2011; Wolff *et al.*, 2011; Zhang *et al.*, 2011); however, this appears to not have been rigorously evaluated in the published literature.

We found many examples of physical linkage between candidate IGs (164 clusters containing 36.4% of total candidate IGs), often involving both PEGs and MEGs (45 clusters; Table S8). As the identification of IGs relies on parental sequence polymorphisms, it is likely that even more extensive genomic clustering is hidden by this technical constraint in detecting neighboring IGs. Among our three within-lineage crosses, 45–70% of all endosperm-expressed genes could not be assessed for PSE, leaving considerable scope for discovering more IGs. We also identified 10 bins containing putative duplicated genes, consistent with clustered *A. thaliana* IGs often being paralogs (five of eight clusters comprise paralogous IGs; Wolff *et al.*, 2011); such observations indicate that gene duplication might be a driver of IG clustering and/or that duplicated genes tend to have the same imprinting status. It is generally acknowledged that tomato chromosomes are highly syntenic with conserved structural features, such as gene-rich euchromatic arms and gene-poor, low-recombination centromeric regions (The Tomato Genome Consortium, 2012). Some small-scale chromosomal rearrangements have been identified, but most of them map to gene-poor heterochromatic regions (Anderson *et al.*, 2010). Given that we probed 100-kb bins containing at least five expressed genes, any such rearrangements among lineages are unlikely to have created false-positives among our inferred clusters.

We were also interested in interactions between PEGs and MEGs at the functional level. Functional cooperation between PEGs and MEGs is expected under the premises of the kin-conflict theory (Haig, 2000). In our study, we found candidate PEGs and MEGs to be enriched for distinct functions, suggesting they are not functionally equivalent. This non-equivalence, however, does not preclude interactions among their gene products. By identifying numerous predicted protein–protein interactions between PEGs and MEGs, we provide evidence that Solanum candidate IGs interact in functional networks, highlighting common biological functions such as expression regulation and cell-cycle control (Figure 7). Although candidate PEGs are greatly outnumbered by candidate MEGs, both MEGs and PEGs are represented in the majority of protein classes identified (Table S2). For example, we found four imprinted ethylene responsive factors (ERFs; two MEGs and two PEGs); this diverse gene family is involved in primary and secondary metabolism, stress response and developmental control (Licausi *et al.*, 2013). Parentally biased expression might achieve optimized stoichiometry of duplicated ERFs that are crucial for the endosperm developmental program. Moreover, the activity, localization and abundance of ERFs is regulated by specific proteins such as phosphorylases, acyl-coA and ubiquitins, some of which were identified as candidate MEGs and PEGs in our data, suggesting functional interactions between IGs (Licausi *et al.*, 2013).

More generally speaking, functional and/or physical units constituted by IGs might reflect many instances of 'innocent bystanders' associating with 'first-order IGs' that happened to be selected for (Varmuza and Mann, 1994; Patten *et al.*, 2016). Moreover, as linkage disequilibrium can facilitate co-adaptation between loci, maternally biased imprinted expression of physically linked genes may enhance the evolution of co-adapted maternal and offspring traits when these are favored by natural selection (Wolf, 2013; Bai and Settles, 2015). Natural selection may have shaped the evolution of genes with functionally important imprinted expression, but where genes have gained an imprinted status as regulatory or physical byproducts of other IGs, natural selection might be relaxed (Spillane *et al.*, 2007; Holman and Kokko, 2014).

PEGs and MEGs as potential agents of cell-cycle transitions

Ubiquitination is a process responsible for targeted protein degradation that is essential for cell-cycle control and involves highly conserved protein complexes (Tyers and Jorgensen, 2000; Nigg, 2001; Reed, 2003; Pines, 2006). We found 16 candidate IGs (nine MEGs and seven PEGs) contributing to ubiquitination (Table S2), and PEGs are significantly enriched for one ubiquitin-related GO term (Table S6). In particular, 12 genes from a major family of E3 ubiquitin ligases, which are part of the SCF complex, were found among our candidate IGs (10 MEGs and two PEGs; Table S2), and two of them encoding RING-domain proteins have conserved MEG status in *A. thaliana* and *R. communis*. We also identified a cluster of SCF-related PEGs on chromosome 8. Via ubiquitination, the SCF complex controls cyclin concentration, which serves as a switch between cell-cycle steps; it has a central role in plant development and appears to be highly conserved across species (Gray *et al.*, 1999; Kim *et al.*, 2008; Jeong *et al.*, 2011; Boycheva *et al.*, 2015). Our results indicate that ubiquitination might be intricately linked to imprinting, and that imprinted expression of specific SCF-complex genes might be required for cell-cycle control in the endosperm.

The timing of cell division and differentiation in different endosperm regions is crucial for seed development and embryo survival (Sabelli and Larkins, 2009; Dante *et al.*, 2014). Parental imprints may dictate the correct pace of cell division after genome and epigenome duplication. These ideas are consistent with the perturbed cell proliferation observed in endosperm of interploidy crosses in maize and *A. thaliana* (Lin, 1982; Scott *et al.*, 1998; Kradolfer *et al.*, 2013). Leblanc *et al.* (2002) proposed that specific IGs regulate mitosis in the developing maize endosperm, hypothesizing that a set of MEGs holds mitosis in check while a set of PEGs controls DNA synthesis. It seems plausible that such parental control occurs, at least partly, in the cellular Solanum endosperm. For example, we identified one cyclin-dependent kinase (CdK) as a candidate MEG and three sister chromatid cohesion proteins, all as candidate PEGs (Table S2). Interestingly, the degradation of sister chromatid cohesion proteins, necessary for the transition from metaphase to anaphase, is carried out by CdK (Murray, 1995; Pines, 2006).

In addition to the potentially critical role of Solanum candidate IGs for cellular switches, they appear to be specifically involved in cell and nuclear organization. As chromosomes in Arabidopsis mainly form biparental homologous pairs in endosperm nuclei (Baroux *et al.*, 2017), IG products such as sister chromatid cohesion proteins might contribute to the correct pairing of chromosomes (Pardo-Manuel de Villena *et al.*, 2000). We also found candidate IGs related to the cytoskeleton (11 MEGs and six PEGs; Table S2), among them actin- and microtubule-binding proteins that may contribute to correct cell organization and mitosis. Moreover, Arabidopsis endosperm chromatin has atypical features compared with other cell types, such as low interphase condensation and enlarged nuclei (Baroux *et al.*, 2007). It was shown that specific heterochromatin domains are established under maternal control and may affect the entire chromatin structure throughout endosperm development, contributing to the characteristic hypomethylation of the endosperm genome (Baroux *et al.*, 2007; Gehring *et al.*, 2009). We found five candidate IGs potentially involved in chromosome condensation, but no evidence for a major maternal role (three PEGs and two MEGs; Table S2). These features suggest that an important function of IGs might be to ensure correct segregation of parental chromosomes during mitosis (Pardo-Manuel de Villena *et al.*, 2000); parental control might be favored in the endosperm because of its atypical, triploid genome.

EXPERIMENTAL PROCEDURES

Plant material and crossing design

Seeds were provided by the Tomato Genetics Resource Center (TGRC, University of California, Davis, USA, https://tgrc.ucdavis.edu). To maximize molecular polymorphism (i.e. power to detect imprinting) for each of the three reciprocal crosses, two genotypes from different source populations were selected. For the *AA* cross, we chose plants from populations LA2185 (Amazonas, Peru) and LA1626 (Ancash, Peru); populations LA4329 (Antofagasta, Chile) and LA2748 (Tarapaca, Chile) were chosen for *CC*; and populations LA2744 (Arica and Parinacota, Chile) and LA2964 (Tacna, Peru) were chosen for *PP*. All selected crosses produced 'normal' quantities of seeds per fruit with high seed viability (Roth *et al.*, 2018). Plants were grown from seed in an insect-free greenhouse at ETHZ (Lindau-Eschikon, canton Zurich, Switzerland). They were regularly repotted in 5-L pots using fresh soil (Ricoter Substrate 214, Ricoter Erdaufbereitung AG, Aarberg, Switzerland, https://www.ricoter.ch) and fertilizing granules (Gartensegen, Hauert HBG Dünger AG, Grossaffoltern, Switzerland, https://www.hauert.com). Additional liquid fertilizer was applied once or twice per month depending on the season (Wuxal® NPK solution, Aglukon Spezialdünger GmbH & Co. KG, Düsseldorf, Germany, https://www.aglukon.com). Plants were watered two-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 whole experiment (12 h light at 18 Klux and 50% relative humidity, 12 h darkness at 0 Klux with 60% relative humidity).

Reciprocal crosses, sample collection and preparation

The 18 plants (six genotypes × three clonal replicates) started flowering a few months after propagation such that all crosses could be performed concurrently, always between 08:00 hours and 10:00 hours. For each reciprocal cross, pollen was manually transferred from stamens of the paternal plant to pistils of the maternal plant. We sampled developing fruits 12 DAP, always between 13:00 hours and 15:00 hours. Each sample represents one biological replicate for a given cross. The protocol for fruit fixation, cryoprotection and cryosection was adapted from Nakazono *et al.* (2003). Briefly, fruits were cut into two equal parts and put in 15-ml tubes containing 4°C Farmer's solution (3:1 ethanol:acetic acid). This fixative solution was infiltrated under vacuum on ice for 15 min, and swirled for 1 h on a rocker shaker at 4°C. Samples were vacuumed a second time in the same way and left swirling overnight at 4°C. The next day, samples were infiltrated by three successive pre-chilled cryoprotectant solutions (sucrose grades 10, 20 and 30% in phosphate-buffered saline solution). For each solution, fruits were infiltrated under vacuum for 15 min on ice and shook for 1 h at 4°C. Then, fruits were carefully drained, placed in Tissue-Tek® cryomolds (25 × 20 × 5 mm, Sakura® Finetek, Tokyo, Japan, https://www.sakuraus.com) and filled with Tissue-Tek® O.C.T. medium (Optimal Cutting Temperature, Sakura® Finetek). The obtained embeddings were frozen in liquid nitrogen, stored at −80°C until use and transported on dry ice. For each parental genotype, flower buds were sampled from the original plant (before cuttings) and stored in liquid nitrogen for later RNA sequencing.

Endosperm isolation

While preserving an adequate histological context for lasering out the endosperm, our fruit preparation protocol allowed for cutting fruits at low temperature to minimize RNA degradation and maximize RNA yield. Cryosections were performed with a Thermo Scientific cryostat (Cryostar NX70; Thermo Fisher Scientific, Waltham, MA, USA, https://www.thermofisher.com). Cryoprotected fruits embedded in O.C.T. were kept at −20°C for 10 min in

the cryochamber to equilibrate temperature. Fruits were sliced in 12-μm cuts recovered on LCM tissue tape (Section-Lab, Hiroshima, Japan, www.section-lab.jp) with the slightly modified Kawamoto's film method (Kawamoto, 2003). After cutting, slices stuck to the tape were washed in pre-chilled ethanol grades (50, 75, 95%) and dehydrated in 100% ethanol in the cryochamber at −20°C for 30 min. Tapes were dried for 5 min under a fume hood and transported on dry ice for direct lasering or stored at −80°C for later lasering on the same day. Each tape was mounted on a metal frame allowing for visualizing and lasering endosperms directly on the tape. Dissection was done with a PALM® Microbeam System (Zeiss, Oberkochen, Germany, https://www.zeiss.com) and the assistance of the PALM® Robosoftware (Zeiss). Upon lasering, endosperm slices were catapulted jointly with the underlying tape to the cap of an adhesive tube (Zeiss). For each sample, RNA was extracted immediately after endosperm collection (at least 100 slices). Endosperm samples were immerged in extraction buffer (XB, Arcturus® Picopure RNA isolation kit; Thermo Fisher Scientific) and incubated at 42°C for 1 h. The obtained cell extracts were stored at −80°C for a maximum of 4 days prior to RNA isolation.

RNA isolation, library preparation and sequencing

RNA isolation was performed following the manufacturer's protocol (Arcturus® Picopure RNA isolation kit; Thermo Fisher Scientific), including the application of DNase (RNase-free DNase Set, Qiagen, Hilden, Germany, https://www.qiagen.com/de) for 15 min on the column. RNA was quantified with a Qubit® fluorometer (Invitrogen, Waltham, MA, USA, https://www.thermofisher.com) and its quality assessed with the 2200 TapeStation instrument (Agilent Technologies, Santa Clara, CA, USA, https://www.agilent.com). A minimum of 100 ng of RNA per sample was treated a second time with DNAse prior to library preparation (gDNA Wipeout Buffer, QuantiTect Reverse Transcription Kit, Qiagen). Ribosomal RNA was removed with the Ribo-Zero Plant Library Prep Kit (Illumina, San Diego, CA, USA, https://www.illumina.com) and libraries were prepared with the TruSeq Stranded Total RNA Library Prep Kit following the protocol supplied by Illumina; replicate samples were prepared on different days. Library quality was assessed on a 2100 Bioanalyzer and on TapeStation (Agilent Technologies). The 18 endosperm libraries were paired-end sequenced (2 × 125 bp) on an Illumina HiSeq 2500 v4 at the Functional Genomics Center Zurich (www.fgcz.ch). To obtain satisfactory coverage, we pooled eight samples per lane following the TruSeq pooling protocol, with all cross replicates being sequenced across at least two lanes.

RNA from parental flower buds was extracted with RNAeasy RNA isolation kit (Qiagen) and libraries prepared with the Illumina TruSeq RNA Sample Preparation Kit v2 (Illumina). These libraries were sequenced on a HiSeq2000 to produce 150-bp paired-end reads (ETH Department of Biosystems Science and Engineering, Basel, Switzerland).

Bioinformatics analyses

Filtering and read mapping. Once reads were de-multiplexed, quality assessment of all samples was done with the FastQC program (http://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), and reads with $Q < 25$ and length <30 nucleotides were filtered out. 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 check was done with Qualimap version 2.2 (Okonechnikov *et al.*, 2016) and RseQC (Wang *et al.*, 2012).

Inferring maternal counts from endosperm data. Single nucleotide polymorphisms were called from the parental transcriptomes (flower bud tissue) to infer the expected genotype in the endosperm for each polymorphic site. Variants were identified with the mpileup tool from Samtools (Li *et al.*, 2009), retaining only high-quality, uniquely mapping reads ($Q > 20$). For each sample and each SNP, allele-specific endosperm expression of genes was assessed using the workflow described in Florez-Rueda *et al.* (2016), with slight parameter modifications (https://github.com/MorganeRoth). The main advantage of this approach is the possibility to use both homozygous and heterozygous nucleotide differences between parental individuals for estimating maternal expression proportions. Total and maternal allelic counts were thus quantified for each SNP from the raw allelic counts obtained from the endosperm transcriptomes. Using the genome annotation ITAG2.4, we then summed the computed allelic counts at all SNPs per gene and estimated maternal proportions for each gene as the ratio maternal/total counts.

Detection of candidate IGs. Once total and maternal counts were computed for each sample, we calculated the distribution of maternal proportions across genes. For each of our six (unilateral) crosses, the distribution of maternal proportions across genes was calculated after merging counts from the three replicates. For each cross, we called MEGs and PEGs after calculating thresholds based on each empirically derived distribution. We considered the median observed maternal proportion as the expected maternal proportion of non-IGs. Genes exhibiting <50% paternal expression compared with the expected median were considered as candidate MEGs. Conversely, we defined candidate PEGs as those having their expected maternal expression at least halved. Hence, genes with maternal proportion > [median + ((1 − median)/2)] were called maternally biased (MEGs) and genes with maternal proportion < [median/2] were called paternally biased (PEGs). The inference of parental bias was considered as robust only if it was found in two or three (of three) replicates for a given cross. Moreover, a gene was considered imprinted for the corresponding cross only if the direction of parental bias matched between reciprocals. As an additional criterion, we only retained genes for which total parental counts across replicates were significantly different from the empirical median values in each direction of the cross (χ^2 test with 5% FDR correction). Putative genomic positions of all candidate IGs, gene expression levels and gene density were plotted on a circos plot using the circlize package in R (R Development Core Team, 2014; Gu, 2017). Gene density was estimated by counting annotated *S. lycopersicum* genes in 1-Mb bins using the BEDTools utility (Quinlan and Hall, 2010).

Assessment of maternal contamination. Previous endosperm-based transcriptomic studies in Arabidopsis have been shown to be partly compromised by RNA contamination from the maternal seed-coat, particularly as they relate to the number and identity of MEGs (Schon and Nodine, 2017). To probe for potential maternal contamination of our endosperm data, we used the statistical method for tissue-specific enrichment advocated by Schon and Nodine (2017), using recent transcriptomic data from *Solanum pimpinellifolium* 4-DAP seed tissues (pre-globular embryo

stage) extracted via laser-microdissection as a reference (Pattison *et al.*, 2015). We defined tissue-specific expression as sets of genes having expression levels at least two times higher in a given tissue (endosperm, seed-coat, embryo) compared with the two others. Multiple-test correction was performed by adjusting *P*-values with a 1% FDR.

Quantification of gene expression levels. Read counts per gene were calculated with the HTSeq program (Anders *et al.*, 2015), and these counts were transformed to TPM (Wagner *et al.*, 2012). These analyses were performed in R (R Development Core Team, 2014).

Network and ontology analyses. Gene ontology networks were calculated with BiNGO (Maere *et al.*, 2005), and functional networks were calculated with STRING (80% confidence interval; Szklarczyk *et al.*, 2017), both using the Cytoscape 3.4 interface (Cline *et al.*, 2007; Su *et al.*, 2014). We used the Panther database to obtain additional putative annotation for candidate IGs and to perform GO-term overrepresentation tests (Mi *et al.*, 2017). GO-term enrichment among IGs was tested with the R package TopGO, using the weight01 algorithm (R Development Core Team, 2014; Alexa and Rahnenführer, 2016). The considered gene universe was the union of expressed and polymorphic genes in *A*, *C* and *P*, and *P*-values were corrected with the Benjamini–Hochberg method using a 5% FDR.

We created an inventory of published IGs identified in *A. thaliana* (*At*; Köhler *et al.*, 2005; Spillane *et al.*, 2007; Gehring *et al.*, 2011; Hsieh *et al.*, 2011; McKeown *et al.*, 2011; Wolff *et al.*, 2011; Bratzel *et al.*, 2012; Pignatta *et al.*, 2014; Burkart-Waco *et al.*, 2015; Jeong *et al.*, 2015), *A. lyrata* (*Al*; Klosinska *et al.*, 2016), *C. rubella* (*Cr*; Hatorangan *et al.*, 2016), rice (*Os*; Luo *et al.*, 2011; Yuan *et al.*, 2017), maize (*Zm*; Gutiérrez-Marcos *et al.*, 2004; Zhang *et al.*, 2011, 2014; Waters *et al.*, 2013; Xin *et al.*, 2013), *S. bicolor* (*Sb*; Zhang *et al.*, 2016) and castor bean (*Rc*; Xu *et al.*, 2014). Sequences were retrieved from public databases: *At*, The Arabidopsis Information Resource (TAIR; Swarbreck *et al.*, 2008) at http://www.arabidopsis.org; *Al* and *Cr*, Phytozome v11.0 (Goodstein *et al.*, 2012) at https://phytozome.jgi.doe.gov; maize, MaizeGDB (Lawrence, 2004) at http://www.maizegdb.org; rice and *Sb*, PlantGDB (Duvick *et al.*, 2007) at http://www.plantgdb.org; and *Rc*, Castor Bean Genome Project (Chan *et al.*, 2010) at http://castorbean.jcvi.org. We estimated sequence similarity between protein sequences from genes in these distantly related species and candidate IGs found in our study using BLASTx with *e*-values $\leq$ 1e-10, retaining and reporting the best hit for each Solanum-model species gene pair only when that non-Solanum gene was found to be imprinted in at least one study (Table S4). Putative orthologs between our IGs and IGs found in *At*, *Al*, *Sb*, maize and rice were also inferred with OMA (Dessimoz *et al.*, 2005) using the 'all-against-all' function.

Physical clustering of genes. We assessed whether IGs tend to be located in physical clusters. For this we adapted the method described in Wolff *et al.* (2011) and Zhang *et al.* (2011) by defining non-overlapping bins of 100 kb along the reference genome (The Tomato Genome Consortium, 2012) and counting the number of inferred IGs per bin; this analysis was performed jointly on all IGs inferred among the three reciprocal crosses. We assumed that the IG distribution was independent between chromosomes and performed individual tests on each chromosome, simulating the random sampling of the total number of inferred IGs among the expressed and polymorphic genes (for which we have parental expression data) in each chromosome. We then performed a one-sample *t*-test on each bin to assess whether the observed number of IGs was higher than the number obtained from a random distribution. The obtained *P*-values were corrected with the FDR method (R Development Core Team, 2014).

AVAILABILITY OF DATA AND MATERIAL

Raw sequence data for the RNA-sequencing dataset used in this study are available from the Sequence Read Archive (https://trace.ncbi.nlm.nih.gov/Traces/sra/) with the accession numbers SRP132466 (this study) and SRX1850236 (maternal plant LA4329B; Florez-Rueda *et al.*, 2016).

ACKNOWLEDGEMENTS

The authors are grateful to Maja Frei and Esther Zürcher for taking expert care of the plants, to Beatrice Arnold for preparing RNA-Seq libraries, and to Claudia Michel, Silvia Kobel and Joachim Hehl for further technical help. The authors thank Niklaus Zemp, Stefan Zoller and Mathias Scharmann for their generous bioinformatics advice, and Alex Widmer for his general support of this project. The authors thank the CM Rick Tomato Genetics Resource Center at U.C. Davis for generously supplying seed samples. The authors acknowledge the technical support on 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). The authors thank the Genetic Diversity Center (ETH Zurich, Switzerland) and the Swiss Institute for Bioinformatics (Lausanne, Switzerland) for providing valuable tools and training for bioinformatics analyses. This work was supported by the Swiss National Science Foundation (31003A_130702 to TS) and an ETH Research Grant (ETH-40 13-2 to TS and Alex Widmer).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Heat map representing tissue-specific enrichment for each of our 18 endosperm samples (following Schon and Nodine, 2017; using data from Pattison *et al.*, 2015).

Figure S2. Venn diagrams displaying the potential maximum overlap of IGs assessed in three tomato lineages if parental polymorphisms were present in all relevant genes.

Table S1. Maternal proportions, imprinting status and metrics to assess PSE for all expressed and polymorphic genes.

Table S2. Identity and annotation of the 1025 candidate IGs identified among *AA*, *CC* and *PP* reciprocal intraspecific crosses, with information on conserved imprinting among taxa.

Table S3. List of 59 conserved Solanum candidate IGs with functional annotation.

Table S4. Imprinted putative homologs/orthologs of Solanum candidate IGs in seven other species (178 gene pairs Solanum-model species).

Table S5. Expression levels in the three reciprocal wild tomato crosses (30 537 observations across species), and expression comparisons for 6167 genes polymorphic and expressed in all three crosses, including genes with variable imprinting status.

Table S6. GO-term enrichment analyses of candidate IGs with TopGO in R and BiNGO in Cytoscape; separate tests were run for MEGs, PEGs, and all candidate IGs (MEGs & PEGs).

Table S7. Edges (protein–protein interactions) obtained by network analysis (STRING program) of 1025 candidate IGs and 10 additional, interacting non-IGs.

Table S8. List of 373 Solanum candidate IGs found to be significantly enriched in physical clusters (bin size 100 kb), and properties of the 164 100-kb bins with significant clustering of Solanum candidate IGs.

REFERENCES

Alexa, A. and Rahnenführer, J. (2016) topGO: enrichment analysis for gene ontology. *R Package Version 2.28.0*.

Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

Anderson, L.K., Covey, P.A., Larsen, L.R., Bedinger, P. and Stack, S.M. (2010) Structural differences in chromosomes distinguish species in the tomato clade. *Cytogenet. Genome Res.* **129**, 24–34.

Anjam, M.S., Ludwig, Y., Hochholdinger, F., Miyaura, C., Inada, M., Siddique, S. and Grundler, F.M.W. (2016) An improved procedure for isolation of high-quality RNA from nematode-infected *Arabidopsis* roots through laser capture microdissection. *Plant Methods*, **12**, 25.

Baek, Y.S., Royer, S.M., Broz, A.K. *et al.* (2016) Interspecific reproductive barriers between sympatric populations of wild tomato species (*Solanum* section *Lycopersicon*). *Am. J. Bot.* **103**, 1964–1978.

Bai, F. and Settles, A.M. (2015) Imprinting in plants as a mechanism to generate seed phenotypic diversity. *Front. Plant Sci.* **5**, 780.

Baroux, C., Pecinka, A., Fuchs, J., Schubert, I. and Grossniklaus, U. (2007) The triploid endosperm genome of *Arabidopsis* adopts a peculiar, parental-dosage-dependent chromatin organization. *Plant Cell*, **19**, 1782–1794.

Baroux, C., Pecinka, A., Fuchs, J., Kreth, G., Schubert, I. and Grossniklaus, U. (2017) Non-random chromosome arrangement in triploid endosperm nuclei. *Chromosoma*, **126**, 115–124.

Bauer, M.J. and Fischer, R.L. (2011) Genome demethylation and imprinting in the endosperm. *Curr. Opin. Plant Biol.* **14**, 162–167.

Beddows, I., Reddy, A., Kloesges, T. and Rose, L.E. (2017) Population genomics in wild tomatoes – the interplay of divergence and admixture. *Genome Biol. Evol.* **9**, 3023–3038.

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R. and Mathews, S. (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **107**, 18724–18728.

Berger, F., Grini, P.E. and Schnittger, A. (2006) Endosperm: an integrator of seed growth and development. *Curr. Opin. Plant Biol.* **9**, 664–670.

Birchler, J.A. (2014) Interploidy hybridization barrier of endosperm as a dosage interaction. *Front. Plant Sci.* **5**, 281.

Boycheva, I., Vassileva, V., Revalska, M., Zehirov, G. and Iantcheva, A. (2015) Cyclin-like F-box protein plays a role in growth and development of the three model species *Medicago truncatula*, *Lotus japonicus*, and *Arabidopsis thaliana*. *Protoplasma*, **6**, 117–130.

Brandvain, Y. and Haig, D. (2005) Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. *Am. Nat.* **166**, 330–338.

Brannan, C.I. and Bartolomei, M.S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* **9**, 164–170.

Bratzel, F., Yang, C., Angelova, A., López-Torrejón, G., Koch, M., del Pozo, J.C. and Calonje, M. (2012) Regulation of the new *Arabidopsis* imprinted gene AtBMI1C requires the interplay of different epigenetic mechanisms. *Mol. Plant*, **5**, 260–269.

Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R.D. and Horsthemke, B. (1995) Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat. Genet.* **9**, 395–400.

Burkart-Waco, D., Ngo, K., Lieberman, M. and Comai, L. (2015) Perturbation of parentally biased gene expression during interspecific hybridization. *PLoS ONE*, **10**, e0117293.

Chan, A.P., Crabtree, J., Zhao, Q. *et al.* (2010) Draft genome sequence of the oilseed species *Ricinus communis*. *Nat. Biotechnol.* **28**, 951–956.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E. and Fischer, R.L. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell*, **110**, 33–42.

Cline, M.S., Smoot, M., Cerami, E. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protocols*, **2**, 2366–2382.

Dante, R.A., Larkins, B.A. and Sabelli, P.A. (2014) Cell cycle control and seed development. *Front. Plant Sci.* **5**, 493.

Dessimoz, C., Cannarozzi, G., Gil, M., Margadant, D., Roth, A., Schneider, A. and Gonnet, G.H. (2005) OMA, a comprehensive, automated project for the identification of orthologs from complete genome data: introduction and first achievements. In *Comparative Genomics* (McLysaght, A. and Huson, D.H., eds). Berlin, Germany: Springer, pp. 61–72.

Dumbliauskas, E., Lechner, E., Jaciubek, M. *et al.* (2011) The *Arabidopsis* CUL4-DDB1 complex interacts with MSI1 and is required to maintain MEDEA parental imprinting. *EMBO J.* **30**, 731–743.

Duvick, J., Fu, A., Muppirala, U., Sabharwal, M., Wilkerson, M.D., Lawrence, C.J., Lushbough, C. and Brendel, V. (2007) PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res.* **36**, D959–D965.

Edwards, C.A. and Ferguson-Smith, A.C. (2007) Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.* **19**, 281–289.

Eggermann, T., de Nanclares, G.P., Maher, E.R. *et al.* (2015) Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. *Clin. Epigenet.* **7**, 123.

Feil, R. and Berger, F. (2007) Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet.* **23**, 192–199.

Feng, S., Jacobsen, S.E. and Reik, W. (2010) Epigenetic reprogramming in plant and animal development. *Science*, **330**, 622–627.

Florez-Rueda, A.M., Paris, M., Schmidt, A., Widmer, A., Grossniklaus, U. and Städler, T. (2016) Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Mol. Biol. Evol.* **33**, 2935–2946.

Floyd, S.K. and Friedman, W.E. (2000) Evolution of endosperm developmental patterns among basal flowering plants. *Int. J. Plant Sci.* **161**, S57–S81.

Fowden, A.L., Sibley, C., Reik, W. and Constancia, M. (2006) Imprinted genes, placental development and fetal growth. *Hormone Res.* **65**, 50–58.

Gabory, A., Ripoche, M.-A., Le Digarcher, A. *et al.* (2009) H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development*, **136**, 3413–3421.

García-Aguilar, M. and Gillmor, C.S. (2015) Zygotic genome activation and imprinting: parent-of-origin gene regulation in plant embryogenesis. *Curr. Opin. Plant Biol.* **27**, 29–35.

Gehring, M. and Satyaki, P.R. (2017) Endosperm and imprinting, inextricably linked. *Plant Physiol.* **173**, 143–154.

Gehring, M., Choi, Y. and Fischer, R.L. (2004) Imprinting and seed development. *Plant Cell*, **16**, S203–S213.

Gehring, M., Huh, J.H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B. and Fischer, R.L. (2006) DEMETER DNA glycosylase establishes MEDEA Polycomb gene self-imprinting by allele-specific demethylation. *Cell*, **124**, 495–506.

Gehring, M., Bubb, K.L. and Henikoff, S. (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science*, **324**, 1447–1451.

Gehring, M., Missirian, V. and Henikoff, S. (2011) Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS ONE*, **6**, e23687.

Goodstein, D.M., Shu, S.Q., Howson, R. *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186.

Gray, W.M., Pozo, J.C., Walker, L., Hobbie, L., Risseeuw, E., Banks, T., Crosby, W.L., Yang, M., Ma, H. and Estelle, M. (1999) Identification of an SCF ubiquitin–ligase complex required for auxin response in *Arabidopsis thaliana*. *Genes Dev.* **13**, 1678–1691.

Grossniklaus, U., Spillane, C., Page, D.R. and Köhler, C. (2001) Genomic imprinting and seed development: endosperm formation with and without sex. *Curr. Opin. Plant Biol.* **4**, 21–27.

Gu, Z. (2017) Circlize: circular visualization. *R Package Version 0.4.1*.

Gutiérrez-Marcos, J.F., Costa, L.M., Biderre-Petit, C., Khbaya, B., O'Sullivan, D.M., Wormald, M., Perez, P. and Dickinson, H.G. (2004) *maternally expressed gene1* is a novel maize endosperm transfer cell-specific gene

with a maternal parent-of-origin pattern of expression. *Plant Cell*, **16**, 1288–1301.

Haig, D. (2000) The kinship theory of genomic imprinting. *Annu. Rev. Ecol. Syst.* **31**, 9–32.

Haig, D. (2013) Kin conflict in seed development: an interdependent but fractious collective. *Annu. Rev. Cell Dev. Biol.* **29**, 189–211.

Haig, D. and Westoby, M. (1989) Parent-specific gene expression and the triploid endosperm. *Am. Nat.* **134**, 147–155.

Haig, D. and Westoby, M. (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. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **333**, 1–13.

Hara, T., Katoh, H., Ogawa, D. *et al.* (2015) Rice SNF2 family helicase ENL1 is essential for syncytial endosperm development. *Plant J.* **81**, 1–12.

Hatorangan, M.R., Laenen, B., Steige, K.A., Slotte, T. and Köhler, C. (2016) Rapid evolution of genomic imprinting in two species of the Brassicaceae. *Plant Cell*, **28**, 1815–1827.

Holman, L. and Kokko, H. (2014) The evolution of genomic imprinting: costs, benefits and long-term consequences. *Biol. Rev.* **89**, 568–587.

Hsieh, T.-F., Shin, J., Uzawa, R. *et al.* (2011) Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc. Natl Acad. Sci. USA*, **108**, 1755–1762.

Hurst, L.D. and McVean, G.T. (1998) Do we understand the evolution of genomic imprinting? *Curr. Opin. Genet. Dev.* **8**, 701–708.

Ikeda, Y. (2012) Plant imprinted genes identified by genome-wide approaches and their regulatory mechanisms. *Plant Cell Physiol.* **53**, 809–816.

Jeong, C.W., Roh, H., Dang, T.V., Choi, Y.D., Fischer, R.L., Lee, J.S. and Choi, Y. (2011) An E3 ligase complex regulates SET-domain polycomb group protein activity in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **108**, 8036–8041.

Jeong, C.W., Park, G.T., Yun, H., Hsieh, T.-F., Choi, Y.D., Choi, Y. and Lee, J.S. (2015) Control of paternally expressed imprinted *UPWARD CURLY LEAF1*, a gene encoding an F-Box protein that regulates CURLY LEAF polycomb protein, in the *Arabidopsis* endosperm. *PLoS ONE*, **10**, e0117431.

Josefsson, C., Dilkes, B. and Comai, L. (2006) Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* **16**, 1322–1328.

Kawamoto, T. (2003) Use of a new adhesive film for the preparation of multi-purpose fresh-frozen sections from hard tissues, whole-animals, insects and plants. *Arch. Histol. Cytol.* **66**, 123–143.

Kim, H.J., Oh, S.A., Brownfield, L., Hong, S.H., Ryu, H., Hwang, I., Twell, D. and Nam, H.G. (2008) Control of plant germline proliferation by SCF^{FBL17} degradation of cell cycle inhibitors. *Nature*, **455**, 1134–1137.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L. and Kakutani, T. (2004) One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. *Science*, **303**, 521–523.

Klosinska, M., Picard, C.L. and Gehring, M. (2016) Conserved imprinting associated with unique epigenetic signatures in the *Arabidopsis* genus. *Nat. Plants*, **2**, 16145.

Ko, C., Köhler, C. and Weinhofer-Molisch, I. (2010) Mechanisms and evolution of genomic imprinting in plants. *Heredity*, **105**, 57–63.

Koch, M.A. and Kiefer, M. (2005) Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species – *Capsella rubella*, *Arabidopsis lyrata* subsp. *petraea*, and *A. thaliana*. *Am. J. Bot.* **92**, 761–767.

Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W. and Grossniklaus, U. (2003) The *Polycomb*-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. *Genes Dev.* **17**, 1540–1553.

Köhler, C., Page, D.R., Gagliardini, V. and Grossniklaus, U. (2005) The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nat. Genet.* **37**, 28–30.

Kradolfer, D., Hennig, L. and Köhler, C. (2013) Increased maternal genome dosage bypasses the requirement of the FIS Polycomb Repressive Complex 2 in *Arabidopsis* seed development. *PLoS Genet.* **9**, e1003163.

Lafon-Placette, C. and Köhler, C. (2014) Embryo and endosperm, partners in seed development. *Curr. Opin. Plant Biol.* **17**, 64–69.

Lafon-Placette, C. and Köhler, C. (2016) Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Mol. Ecol.* **25**, 2620–2629.

Lafon-Placette, C., Johannessen, I.M., Hornslien, K.S. *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.

Lawrence, C.J. (2004) MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.* **32**, D393–D397.

Leblanc, O., Pointe, C. and Hernandez, M. (2002) Cell cycle progression during endosperm development in *Zea mays* depends on parental dosage effects. *Plant J.* **32**, 1057–1066.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Licausi, F., Ohme-Takagi, M. and Perata, P. (2013) APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytol.* **199**, 639–649.

Lin, B.Y. (1982) Association of endosperm reduction with parental imprinting in maize. *Genetics*, **100**, 475–486.

Lu, X., Chen, D., Shu, D., Zhang, Z., Wang, W., Klukas, C., Chen, L.-l., Fan, Y., Chen, M. and Zhang, C. (2013) The differential transcription network between embryo and endosperm in the early developing maize seed. *Plant Physiol.* **162**, 440–455.

Luo, M., Taylor, J.M., Spriggs, A., Zhang, H., Wu, X., Russell, S., Singh, M. and Koltunow, A. (2011) A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet.* **7**, e1002125.

Maere, S., Heymans, K. and Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448–3449.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12.

McGrath, J. and Solter, D. (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, **37**, 179–183.

McKeown, P.C., Laouielle-Duprat, S., Prins, P. *et al.* (2011) Identification of imprinted genes subject to parent-of-origin specific expression in *Arabidopsis thaliana* seeds. *BMC Plant Biol.* **11**, 113.

McMinn, J., Wei, M., Schupf, N., Cusmai, J., Johnson, E.B., Smith, A.C., Weksberg, R., Thaker, H.M. and Tycko, B. (2006) Unbalanced placental expression of imprinted genes in human intrauterine growth restriction. *Placenta*, **27**, 540–549.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D. and Thomas, P.D. (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45**, D183–D189.

Mitchell-Olds, T. (2001) *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol. Evol.* **16**, 693–700.

Murray, A. (1995) Cyclin ubiquitination: the destructive end of mitosis. *Cell*, **81**, 149–152.

Nakazono, M., Qiu, F., Borsuk, L.A. and Schnable, P.S. (2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. *Plant Cell*, **15**, 583–596.

Nigg, E.A. (2001) Mitotic kinases as regulators of cell division and its checkpoints. *Nat. Rev. Mol. Cell Biol.* **2**, 21–32.

Okonechnikov, K., Conesa, A. and García-Alcalde, F. (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, **32**, 292–294.

Pardo-Manuel de Villena, F., de la Casa-Esperón, E. and Sapienza, C. (2000) Natural selection and the function of genome imprinting: beyond the silenced minority. *Trends Genet.* **16**, 573–579.

Patel, R.K. and Jain, M. (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE*, **7**, e30619.

Patten, M.M., Ross, L., Curley, J.P., Queller, D.C., Bonduriansky, R. and Wolf, J.B. (2014) The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity*, **113**, 119–128.

Patten, M.M., Cowley, M., Oakey, R.J. and Feil, R. (2016) Regulatory links between imprinted genes: evolutionary predictions and consequences. *Proc. R. Soc. Lond. B Biol. Sci.* **283**, 20152760.

Pattison, R.J., Csukasi, F., Zheng, Y., Fei, Z., van der Knaap, E. and Catalá, C. (2015) Comprehensive tissue-specific transcriptome analysis reveals distinct regulatory programs during early tomato fruit development. *Plant Physiol.* **168**, 1684–1701.

Pease, J.B., Haak, D.C., Hahn, M.W. and Moyle, L.C. (2016) Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol.* **14**, e1002379.

Peralta, I.E., Spooner, D.M. and Knapp, S. (2008) Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*; sect. *Juglandifolia*; sect. *Lycopersicon*; Solanaceae). *Syst. Bot. Monogr.* **84**, 1–186.

Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W. and Gehring, M. (2014) Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. *eLife*, **3**, e03198.

Pines, J. (2006) Mitosis: a matter of getting rid of the right protein at the right time. *Trends Cell Biol.* **16**, 55–63.

Qiu, Y., Liu, S.-L. and Adams, K.L. (2014) Frequent changes in expression profile and accelerated sequence evolution of duplicated imprinted genes in *Arabidopsis*. *Genome Biol. Evol.* **6**, 1830–1842.

Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

R Development Core Team (2014) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raissig, M.T., Berner, M., Baroux, C. and Grossniklaus, U. (2013) Genomic imprinting in the *Arabidopsis* embryo is partly regulated by PRC2. *PLoS Genet.* **9**, e1003862.

Rebernig, C.A., Lafon-Placette, C., Hatorangan, M.R., Slotte, T. and Köhler, C. (2015) Non-reciprocal interspecies hybridization barriers in the *Capsella* genus are established in the endosperm. *PLoS Genet.* **11**, e1005295.

Reed, S.I. (2003) Ratchets and clocks: the cell cycle, ubiquitylation and protein turnover. *Nat. Rev. Mol. Cell Biol.* **4**, 855–864.

Rick, C.M. (1986) Reproductive isolation in the *Lycopersicon peruvianum* complex. In *Solanaceae, Biology and Systematics* (D'Arcy, W., ed). New York, NY, USA: Columbia University Press, pp. 477–495.

Rick, C.M. and Lamm, R. (1955) Biosystematic studies on the status of *Lycopersicon chilense*. *Am. J. Bot.* **42**, 663–675.

Roth, M., Florez-Rueda, A.M., Griesser, S., Paris, M. and Städler, T. (2018) Incidence and developmental timing of endosperm failure in post-zygotic isolation between wild tomato lineages. *Ann. Bot.* **121**, 107–118.

Sabelli, P.A. and Larkins, B.A. (2009) The contribution of cell cycle regulation to endosperm development. *Sex. Plant Reprod.* **22**, 207–219.

Schon, M.A. and Nodine, M.D. (2017) Widespread contamination of *Arabidopsis* embryo and endosperm transcriptome data sets. *Plant Cell*, **29**, 608–617.

Scott, R.J. and Spielman, M. (2006) Genomic imprinting in plants and mammals: how life history constrains convergence. *Cytogenet. Genome Res.* **113**, 53–67.

Scott, R.J., Spielman, M., Bailey, J. and Dickinson, H.G. (1998) Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development*, **125**, 3329–3341.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A. and Ponting, C.P. (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **15**, 121–132.

Spencer, H.G. and Clark, A.G. (2014) Non-conflict theories for the evolution of genomic imprinting. *Heredity*, **113**, 112–118.

Spillane, C., Schmid, K.J., Laoueillé-Duprat, S., Pien, S., Escobar-Restrepo, J.-M., Baroux, C., Gagliardini, V., Page, D.R., Wolfe, K.H. and Grossniklaus, U. (2007) Positive darwinian selection at the imprinted MEDEA locus in plants. *Nature*, **448**, 349–352.

Städler, T., Arunyawat, U. and Stephan, W. (2008) Population genetics of speciation in two closely related wild tomatoes (*Solanum* section *Lycopersicon*). *Genetics*, **178**, 339–350.

Su, G., Morris, J.H., Demchak, B. and Bader, G.D. (2014) Biological network exploration with Cytoscape 3. *Curr. Protoc. Bioinformatics*, **47**, 8.13.1–8.13.24.

Surani, M.A.H., Barton, S.C. and Norris, M.L. (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, **308**, 548–550.

Swarbreck, D., Wilks, C., Lamesch, P. *et al.* (2008) The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* **36**, D1009–D1014.

Szklarczyk, D., Morris, J.H., Cook, H. *et al.* (2017) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368.

Tellier, A., Fischer, I., Merino, C., Xia, H., Camus-Kulandaivelu, L., Städler, T. and Stephan, W. (2011) Fitness effects of derived deleterious mutations in four closely related wild tomato species with spatial structure. *Heredity*, **107**, 189–199.

The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.

Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.

Tyers, M. and Jorgensen, P. (2000) Proteolysis and the cell cycle: with this RING I do thee destroy. *Curr. Opin. Genet. Dev.* **10**, 54–64.

Varmuza, S. and Mann, M. (1994) Genomic imprinting – defusing the ovarian time bomb. *Trends Genet.* **10**, 118–123.

Varrault, A., Gueydan, C., Delalbre, A. *et al.* (2006) Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. *Dev. Cell*, **11**, 711–722.

Verona, R.I., Mann, M.R.W. and Bartolomei, M.S. (2003) Genomic imprinting: intricacies of epigenetic regulation in clusters. *Annu. Rev. Cell Dev. Biol.* **19**, 237–259.

Vijayaraghavan, M.R. and Prabhakar, K. (1984) The endosperm. In *Embryology of Angiosperms* (Johri, B.M., ed). Berlin, Germany: Springer, pp. 319–376.

Wagner, G.P., Kin, K. and Lynch, V.J. (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**, 281–285.

Walter, J. and Paulsen, M. (2003) The potential role of gene duplications in the evolution of imprinting mechanisms. *Hum. Mol. Genet.* **12**, 215–220.

Wang, L., Wang, S. and Li, W. (2012) RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, **28**, 2184–2185.

Waters, A.J., Bilinski, P., Eichten, S.R., Vaughn, M.W., Ross-Ibarra, J., Gehring, M. and Springer, N.M. (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.

Wolf, J.B. (2013) Evolution of genomic imprinting as a coordinator of coadapted gene expression. *Proc. Natl Acad. Sci. USA*, **110**, 5085–5090.

Wolf, J.B. and Hager, R. (2006) A maternal–offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biol.* **4**, e380.

Wolf, J.B., Oakey, R.J. and Feil, R. (2014) Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. *Heredity*, **113**, 167–175.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., Spillane, C., Nordborg, M., Rehmsmeier, M. and Köhler, C. (2011) High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet.* **7**, e1002126.

Xin, M., Yang, R., Li, G. *et al.* (2013) Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development. *Plant Cell*, **25**, 3212–3227.

Xu, W., Dai, M., Li, F. and Liu, A. (2014) Genomic imprinting, methylation and parent-of-origin effects in reciprocal hybrid endosperm of castor bean. *Nucleic Acids Res.* **42**, 6987–6998.

Yuan, J., Chen, S., Jiao, W. *et al.* (2017) Both maternally and paternally imprinted genes regulate seed development in rice. *New Phytol.* **216**, 373–387.

Zhang, M., Zhao, H., Xie, S. *et al.* (2011) Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc. Natl Acad. Sci. USA*, **108**, 20042–20047.

Zhang, M., Xie, S., Dong, X. *et al.* (2014) Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. *Genome Res.* **24**, 167–176.

Zhang, M., Li, N., He, W., Zhang, H., Yang, W. and Liu, B. (2016) Genome-wide screen of genes imprinted in sorghum endosperm, and the roles of allelic differential cytosine methylation. *Plant J.* **85**, 424–436.