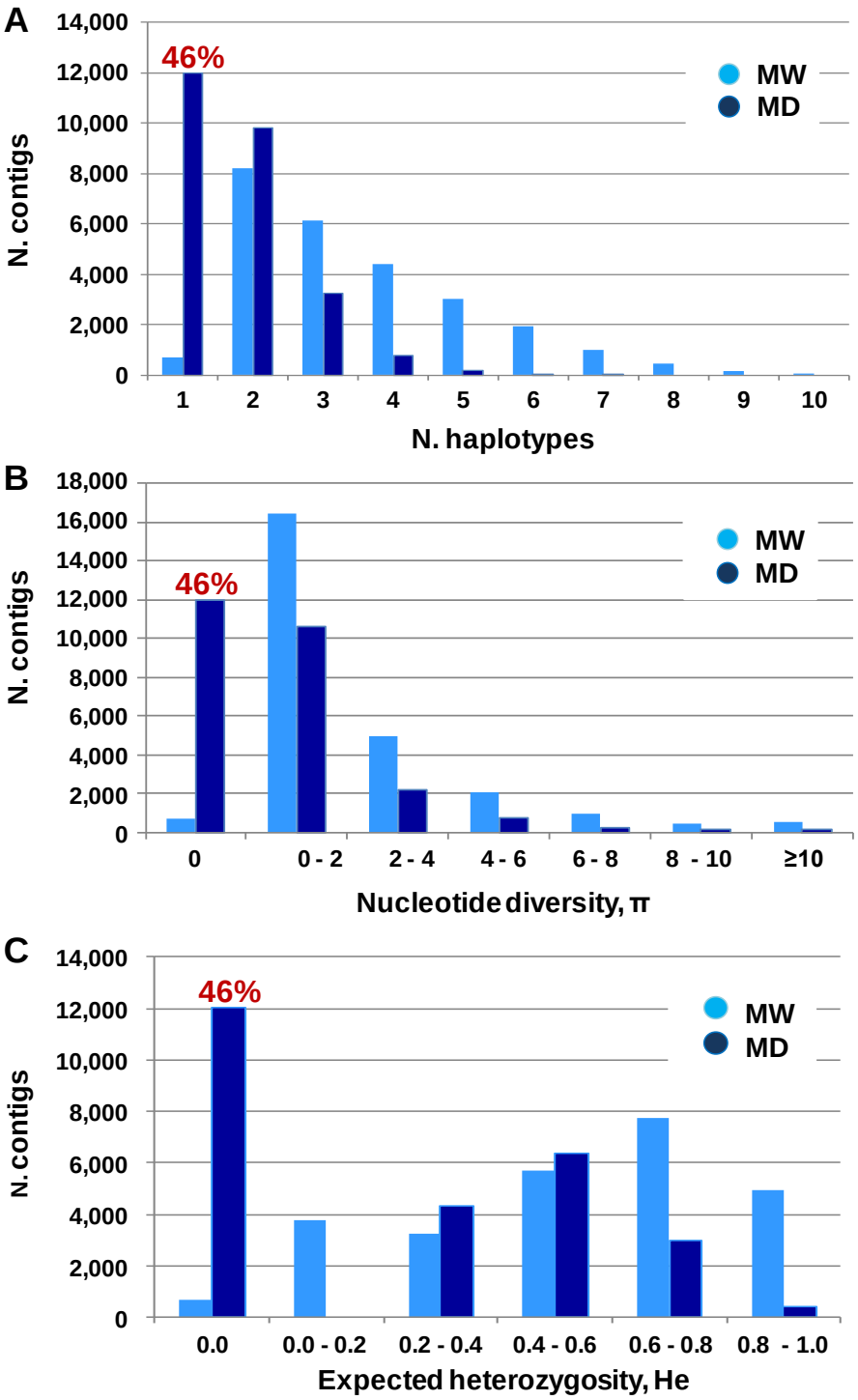


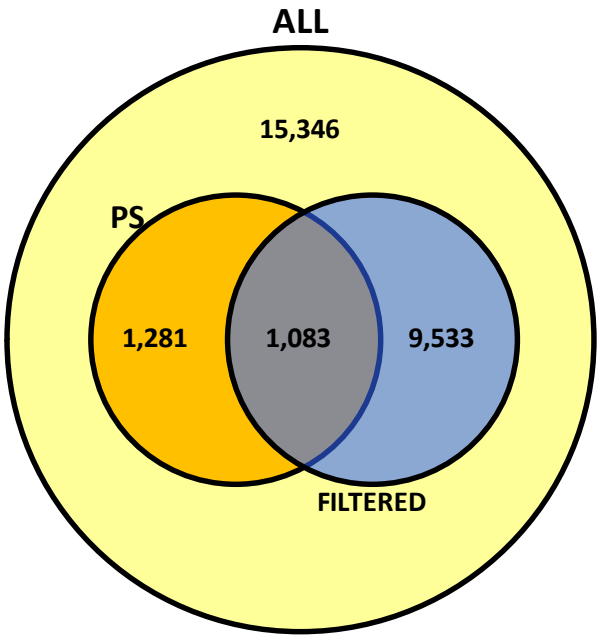
Supplemental Figure 1. Diversity estimates in MW and MD populations.

Number of contigs for the different classes.

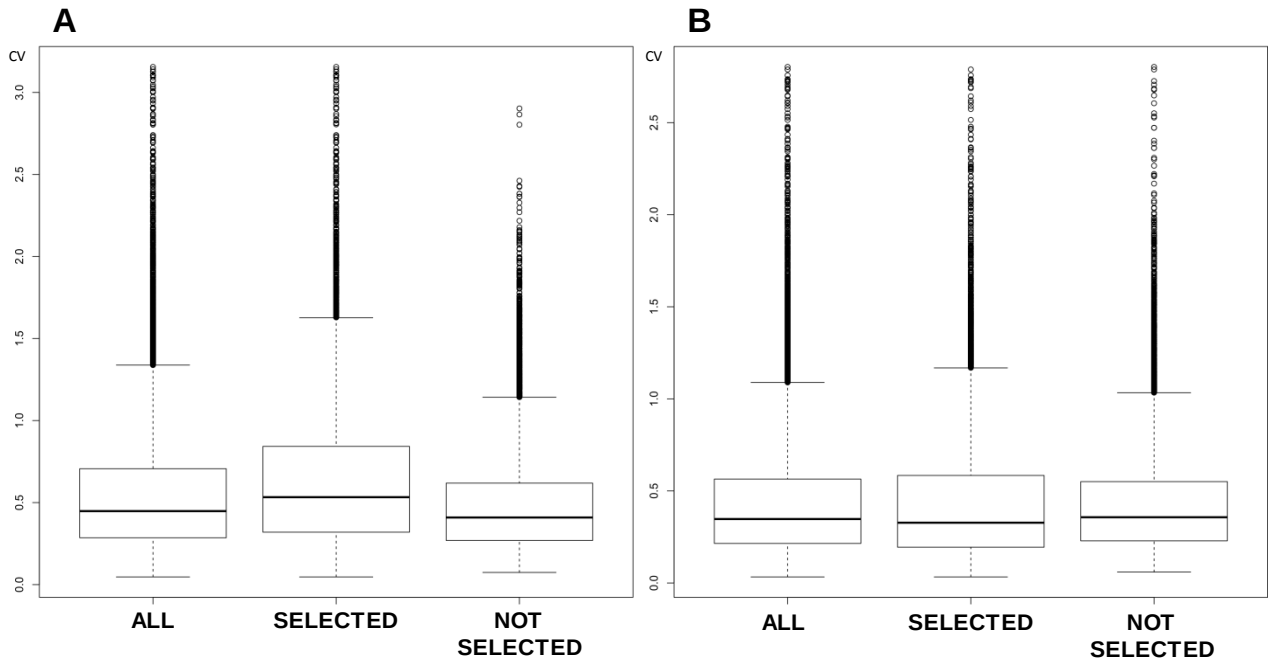
(A) Number of haplotypes (nH).

(B) Nucleotide diversity (π).

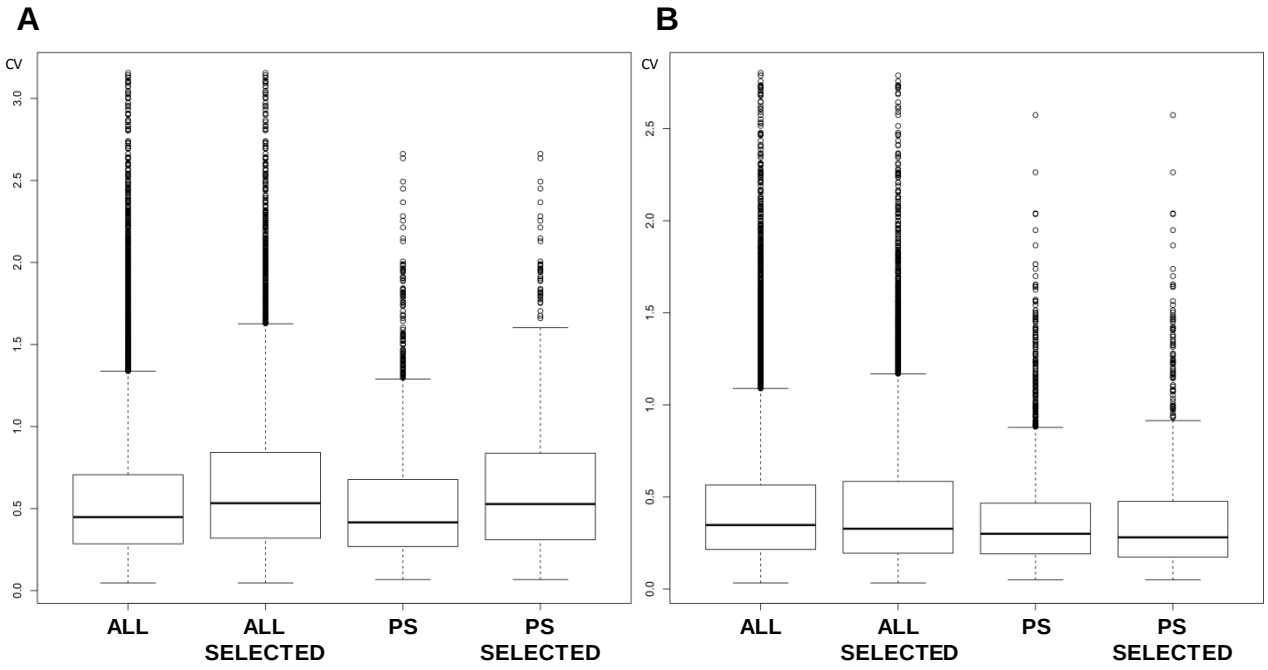
(C) Expected heterozygosity (He).



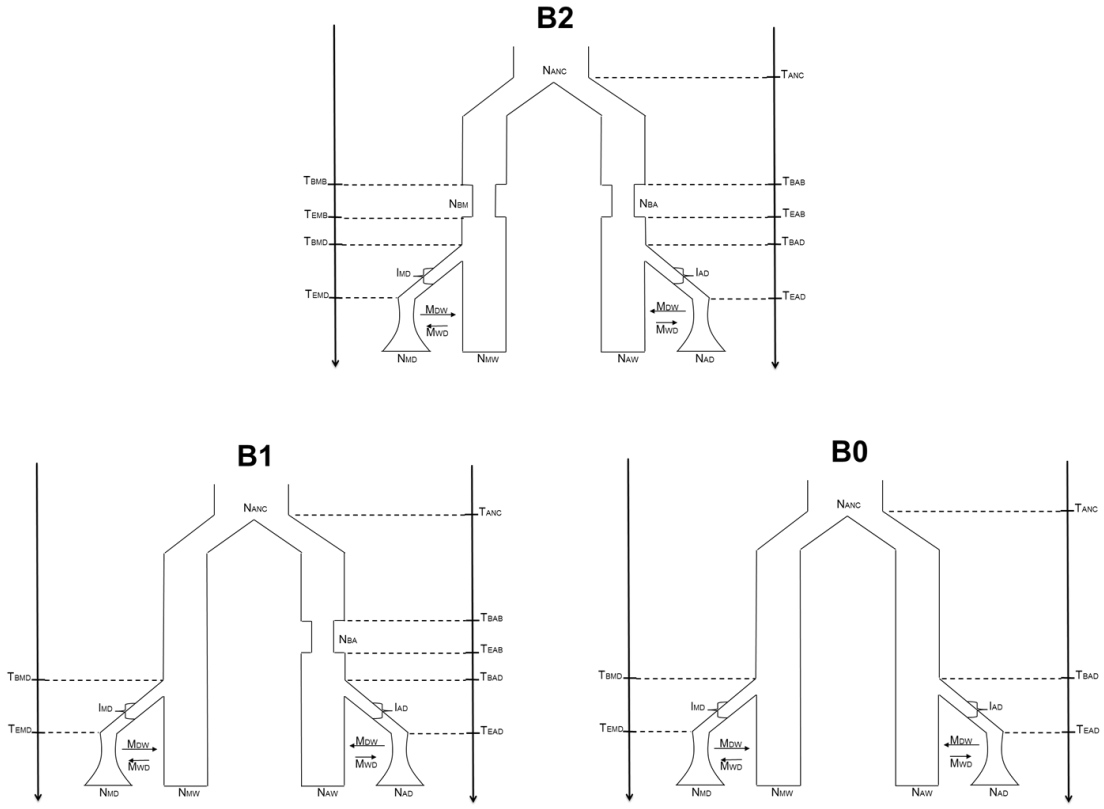
Supplemental Figure 2. Network-Based Analysis: Results of the selection strategy. The Venn diagram illustrates the overlap between the entire set of contigs (yellow), the putatively under selection (PS) contigs (orange), and the contigs selected using the outlined strategy for subsequent network-based analysis (blue). Altogether 1,083 of 2,364 PS contigs were retained for the network-based analysis.



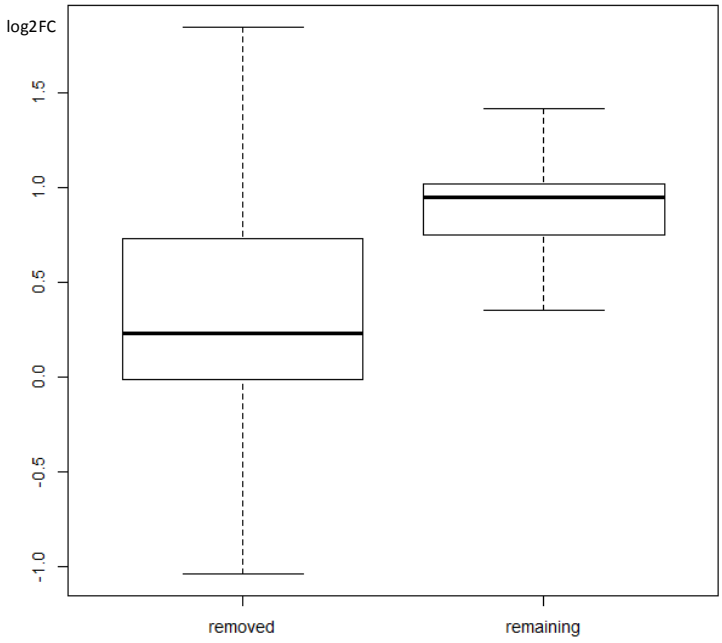
Supplemental Figure 3. Network-Based Analysis: Selection strategy and bias of the CVs. Boxplots of the CVs over all of the contigs (ALL), those selected for the subsequent network-based analysis (SELECTED), and those not selected by the strategy used (NOT SELECTED), for the MW **(A)** and MD **(B)** populations. The boxplots support the absence of systemic bias.



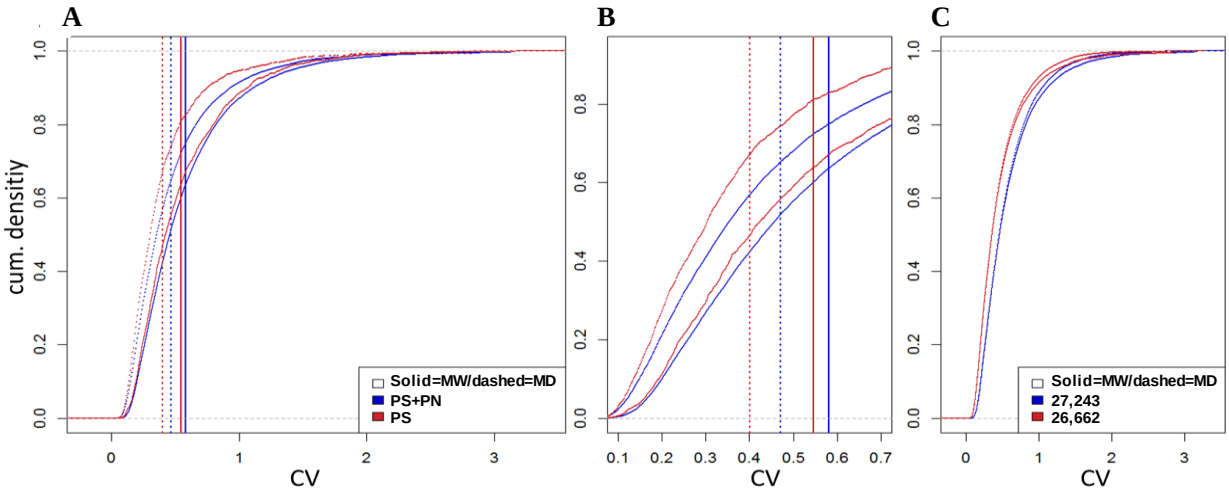
Supplemental Figure 4. Network-Based Analysis: Selection strategy and bias of CVs in the PS contigs. Boxplots of the CVs for all of the contigs (ALL), and those selected for the network-based analysis (ALL SELECTED), and of all the PS contigs (PS) and those retained (PS SELECTED) for the network-based analysis, in the MW **(A)** and MD **(B)** populations. The boxplots support the absence of systematic bias; i.e., no shifts towards higher/ lower values in the CVs for the PS contigs that were retained for further analysis.



Supplemental Figure 5. Demographic models used in this study. Demographic scenarios for the Mesoamerican and the Andean populations. See Supplemental Table 7 online for the details of the parameters.



Supplemental Figure 6. Boxplot illustrating the differences in the log2FC for the 581 removed contigs (removed) and the remaining 26,662 contigs (remaining). Note that negative values indicate higher expression levels in the MD forms.

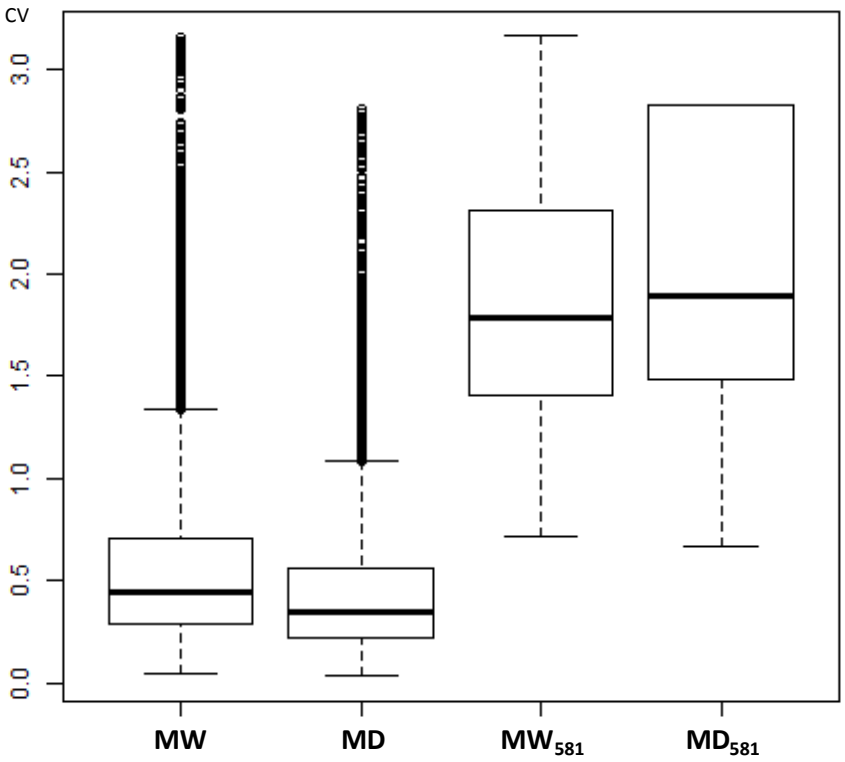


Supplemental Figure 7. Empirical cumulative density functions for the CVs considering all of the 27,243 contigs as reference.

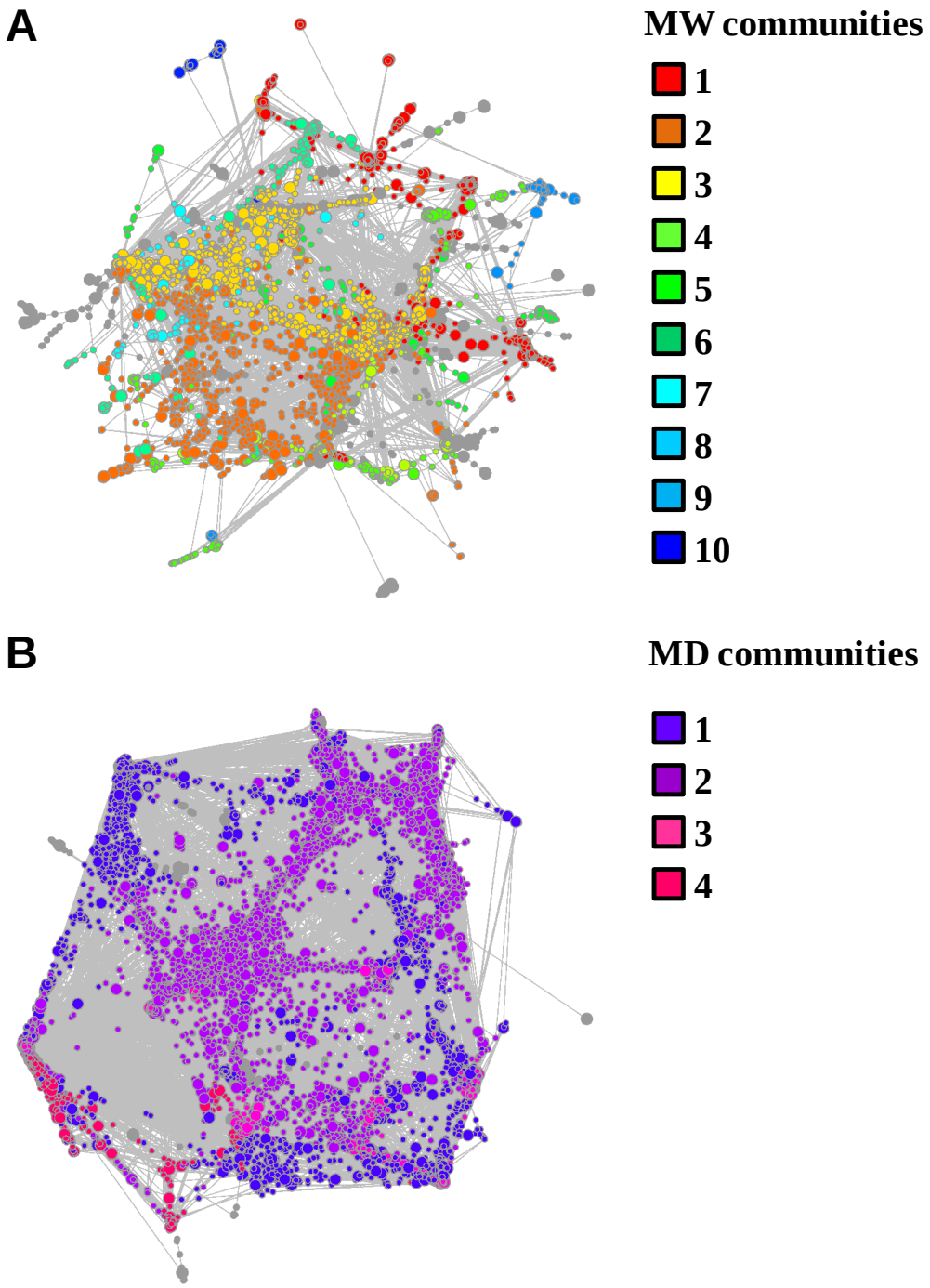
(A) Comparison of the empirical cumulative density functions of the CVs considering all of the contigs and PS.

(B) Detail of the panel in **(A)**, to show the range of the CVs between 0.1 and 0.7.

(C) Effect on the CVs with the removal of the 581 contigs containing more than nine individuals with zero expression levels.



Supplemental Figure 8. Boxplot illustrating the differences in the CVs for the 26,662 contigs (MW, MD) after the removal of the 581 contigs (MW_{581}, MD_{581}) in the MW and MD populations. Note: the majority of high CVs observed for the 581 contigs arise from the very small mean expression values.



Supplemental Figure 9. Relevance network from the MW (A) and MD (B) populations. The relevance network was extracted to match the density of the proximity networks. The different colors indicate the different communities of size ≥40 in the network; the nodes of larger sizes correspond to contigs under selective pressure; isolated nodes are not visualized.

SUPPLEMENTAL TABLES

Supplemental Table 1. Transcriptome assembly statistics. Assembly statistics for the four *P. vulgaris* reference genotypes and for the final nonredundant dataset.

	Assembly statistics			
	G24378	**W617475**	**G12873**	**G12979**
Number of sequences	55,069	65,273	70,826	61,141
Assembled bases	46,396,157	52,104,578	59,851,108	51,083,455
Average length	842.51	798.26	845.04	835.50
Median length	448	431	458	447
N50	1,515	1,393	1,486	1,491
	Nonredundant dataset statistics			
Number of sequences	124,166			
Assembled bases	104,901,858			
Maximum length	16,891			
Minimum length	201			
Average length	844.85			
Median length	428			
N50	1,595			

Supplemental Table 2. Jaccard similarity of the community structure in the MW and MD proximity networks. The Jaccard index was determined for each pair of communities for the extracted MW and MD proximity networks. The low similarity supports the comparison of the corresponding node partitions based on the adjusted Rand index.

		MD network				
		1	**2**	**3**	**4**	**5**
MW network	**1**	0.188	0.214	0.122	0.004	0.066
	2	0.202	0.245	0.133	0.003	0.042
	3	0.066	0.037	0.209	0.002	0.024
	4	0.133	0.052	0.092	0.004	0.054
	5	0.095	0.125	0.104	0.004	0.052
	6	0.001	0.001	0	0	0
	7	0.002	0	0.001	0	0

MW, Mesoamerican wild accessions; MD, Mesoamerican domesticated accessions.

Supplemental Table 3. Community gene function enrichment: List of selected enriched gene functions for each of the communities determined for the MW and MD networks (the full list of enriched terms and corresponding p-values is given in Supplemental Data Sets 1B and 1C online for MW and MD, respectively).

Network	Function
MW	
Community 1	Cell wall (hemicellulose synthesis, glucuronoxylan, cellulose synthesis)
	OPP oxidative PP.6-phosphogluconate dehydrogenase
	Amino acid metabolism (aromatic), hormone metabolism
	RNA regulation
Community 2	Protein synthesis (ribosomal proteins)
	RNA regulation and transcription (HSF, HAD, and AS2-lateral organ boundaries)
	Secondary metabolism (flavanoids and flavanols)
	N-metabolism (ammonia)
	Hormone metabolism (jasmonate and absisic acid synthesis-degradation)
	Signaling
Community 3	DNA synthesis (chromatin structure); lipid metabolism
Community 4	Hormone synthesis (jasmonate synthesis-degradation)
	PS lightreaction (photosystem II)
	Minor CHO metabolism
Community 5	Major CHO biosynthesis (starch)
	Redox (thioredoxin)
	Protein degradation (AAA type)
Community 6	Not enriched for any MapMan bins at significance level of α = 0.05
Community 7	Not enriched for any MapMan bins at significance level of α = 0.05
MD	
Community 1	PS lightreaction (photosystem II and other electron carrier)
	Cell wall (hemicellulose and cellulose synthesis, precursor synthesis (UXS), and glucuronoxylan)
	Minor CHO metabolism (raffinose family)
	RNA regulation of transcription (MYB domain and MADS box transcription factor family, Aux/IAA family, and Trihelix and Triple-Helix transcription factor family),
	N-metabolism (ammonia)
	Stress (abiotic)
	Hormone metabolism (slicylic and abscisic acid)
Community 2	Protein synthesis (ribosomal protein)
	Protein degradation (ubiquitin)
	Protein targeting (secretory pathway)
	Post-transcriptional modification (kinase)
	Signaling (receptor kinase)
	RNA regulation of transcription (bZIP and WRKY domain transcription factor family)
	Fermentation (aldehyde dehydrogenase)
Community 3	DNA synthesis (chromatin structure)
	RNA regulation of transcription (zf-HD)
	Cell wall degradation (mannan-xylose-arabinose-fucose)
Community 4	Not enriched for any MapMan bins at significance level of α = 0.05
Community 5	Not enriched for any MapMan bins at significance level of α = 0.05

MW, Mesoamerican wild accessions; MD, Mesoamerican domesticated accessions.

Supplemental Table 4. Difference in mean node centralities of contigs under selective pressure, and the rest of the nodes in the MW and MD networks. The centrality of each node was estimated for the MW and MD proximity networks. The differences in means between the contigs under selective pressure and the rest of the nodes in the networks were tested. Negligible, although statistically significant, differences were observed for closeness centrality in both the MW and MD proximity networks.

Difference factor	MW proximity network			MD proximity network		
	Mean value of contigs		p value	Mean value of contigs		p value
	Under selective pressure	Remaining		Under selective pressure	Remaining	
Betweenness	26,890	25,470	**0.0063**	28,220	27,920	0.6068
Closeness	1.632E-05	1.621E-05	**3.96E-05**	1.518E-05	1.511E-05	**0.0167**
Degree	12.69	12.88	0.1059	13.69	13.64	0.5963
Ev centr.	0.00217	0.002651	0.7984	0.001401	0.003122	0.025
Page rank	9.356E-05	9.427E-05	0.2401	9.44E-05	9.417E-05	0.6278
Eccentricity	9.208	9.258	0.0031	10.42	10.44	0.6459
Burt's constraint	0.1244	0.1245	0.9176	0.1204	0.1213	0.1221
Transitivity	0.2468	0.2533	0.033	0.2832	0.2886	0.01287

Supplemental Table 5. Difference in assortativity, both nominal and based on the 'selection index', between the MW and MD networks. Significance obtained with permutation tests (permutation of node labels) for the observed assortativity values for the MW and MD networks, and their difference.

Permutation test	**Network**	**Node (n = 10,000)**	
		Observed	**p value**
Assortativity (nominal)	MW	0.005828	0.0608
	MD	0.011888	9.00E-04
	MW-MD	-0.00606	0.1246667
Assortativity (S-Index)	MW	0.008214	0.0154
	MD	0.014609	1.00E-04
	MW-MD	-0.00639	0.1135

MW, Mesoamerican wild accessions; MD, Mesoamerican domesticated accessions.

Supplemental Table 6. Accessions used in this study.

	Accession number/name	Population code[1]	Source[2]	Country	Department, State or Province	Biological status	Altitude (m s.l.m.)	Latitude (°N)	Longitude (°E)
1	G12873*	MW	CIAT	Mexico	Morelos	Wild	1981	19.000	-99.250
2	G9989	MW	CIAT	Mexico	Jalisco	Wild	1400	20.500	-104.817
3	G11050	MW	CIAT	Mexico	Michoacan	Wild	2040	19.683	-101.267
4	G12979*	MW	CIAT	Mexico	Jalisco	Wild	/	20.117	-104.367
5	G22837	MW	CIAT	Mexico	Chihuahua	Wild	1750	26.933	-106.417
6	G24378*	MW	CIAT	Mexico	Oaxaca	Wild	1250	16.040	-97.083
7	PI325677	MW	USDA	Mexico	Morelos	Wild	1828	18.967	-99.100
8	PI417770	MW	USDA	Mexico	Nayarit	Wild	2000	20.667	-102.383
9	G20515	MW	CIAT	Mexico	Puebla	Wild	/	19.800	-97.783
10	G12922	MW	CIAT	Mexico	Jalisco	Wild	1829	20.700	-102.350
11	G5191	MD	CIAT	Venezuela	Distrito Federal	Landrace	/	10.500	-66.917
12	PI151017	MD	USDA	Chile	/	Landrace	660	-33.450	-70.667
13	PI201349	MD	USDA	Mexico	Puebla	Landrace	1000	20.183	-98.050
14	PI281981	MD	USDA	Mexico	Jalisco	Landrace	/	20.417	-103.667
15	PI300668	MD	USDA	Chile	/	Landrace	/	/	/
16	PI309831	MD	USDA	Costa Rica	Cartago	Landrace	1200	9.867	-83.783
17	PI310660	MD	USDA	Guatemala	Chimaltenango	Landrace	2000	14.667	-90.817
18	PI311794	MD	USDA	El Salvador	Santa Ana	Landrace	660	13.983	-89.567
19	W617475*	AW	USDA	Argentina	Salta	Wild	1470	-25.167	-65.617
20	Midas	AD	P.Gepts	Argentina	/	Cultivar	/	/	/
21	PI298109	AD	USDA	Brazil	/	Landrace	/	/	/

[1]MW, Mesoamerican wild; MD, Mesoamerican domesticated; AW, Andean wild; AD, Andean domesticated;
[2]CIAT, International Centre for Tropical Agriculture; USDA, United States Department of Agriculture. Seeds of Midas accession were kindly provided by Prof. Paul Gepts, Department of Plant Sciences, Section of Crop and Ecosystem Sciences, UC Davis, CA, USA.
*Accessions of the hyper-core collection used for the *de novo* assembly.

Supplemental Table 7. Demographic parameters in the B0, B1 and B2 models.

Model	Parameter	Description	Units	Distribution	Mean	S.D.	Min	Max
B0,B1,B2	T_{ANC}	Divergence time between Andean and Mesoamerican gene pools	Generations	Normal	111,000	40,000	67,330	192,835
B2	T_{BMB}	Time of the beginning of the Mesoamerican founder bottleneck	Generations	Normal	99,833	750	98,375	99,833
B1,B2	T_{BAB}	Time of the beginning of the Andean founder bottleneck	Generations	Normal	98,845	3,000	94,051	104,027
B2	T_{EMB}	Time of the ending of the Mesoamerican founder bottleneck	Generations	Normal	67,341	1,400	64,568	67,341
B1,B2	T_{EAB}	Time of the ending of the Andean founder bottleneck	Generations	Normal	67,858	1,500	64,655	70,865
B0,B1,B2	T_{BMD}	Time of the beginning of the Mesoamerican domestication	Generations	Normal	8,160	133	7,922	8,426
B0,B1,B2	T_{BAD}	Time of the beginning of the Andean domestication	Generations	Normal	8,500	8	8,495	8,517
B0,B1,B2	T_{EMD}	Time of the ending of the Mesoamerican domestication	Generations	Normal	6,260	150	5,971	6,567
B0,B1,B2	T_{EAD}	Time of the ending of the Andean domestication	Generations	Normal	7,012	35	6,945	7,075
B0,B1,B2	N_{ANC}	Ancestral effective population size	Haploid individuals	Normal	418,000	105,000	266,000	628,000
B2	N_{BM}	Mesoamerican effective population size during the founder bottleneck	Haploid individuals	Normal	168,000	7,500	153,000	170,000
B1,B2	N_{BA}	Andean effective population size during the founder bottleneck	Haploid individuals	Normal	105,000	20,000	65,000	142,000
B0,B1,B2	N_{MD}	Effective population size of the domesticated Mesoamerican population	Haploid individuals	Uniform			100,000	100,000
B0,B1,B2	N_{MW}	Effective population size of the wild Mesoamerican population	Haploid individuals	Normal	292,000	240,000	125,000	773,000
B0,B1,B2	N_{AW}	Effective population size of the wild Andean population	Haploid individuals	Normal	137,000	182,000	70,000	502,000
B0,B1,B2	N_{AD}	Effective population size of the domesticated Andean population	Haploid individuals	Uniform			100,000	100,000
B0,B1,B2	I_{MD}	Intensity of the domestication bottleneck in Mesoamerica	Ratio	Normal	47.65	3	41.66	52.13
B0,B1,B2	I_{AD}	Intensity of the domestication bottleneck in the Andes	Ratio	Normal	47.26	0.5	46.25	48.59
B0,B1,B2	M_{WD}	Migration rate from wild to domesticated population	Rate	Uniform			0.000001	0.00001
B0,B1,B2	XM	Asymmetric migration factor	Rate	Uniform			2	6
B0,B1,B2	M_{DW}	Migration rate from domesticated to wild population	Rate	XM*MWD				
B0,B1,B2	μ	Mutation rate (per site per generation)	Rate	Normal	$1E^{-9}$	$5E^{-9}$	$1E^{-10}$	$1E^{-8}$
B0,B1,B2	L	Length of the simulated sequences	Base pairs	Normal	1,300	1,500	250	5,000

Supplemental Table 8. Enriched MapMan bins for 581 (140 annotated) contigs removed prior to the network analysis. These 581 contigs were removed from the set of contigs subjected to the network-based analysis due to zero expression estimates in at least nine individuals. These contigs are enriched for the MapMan bins included in the first two columns at a significance level of 0.05.

MapMan bin	Description	p value
28.1.3.2	DNA.synthesis/chromatin structure.histone.core	1.94e-08
28.1.3	DNA.synthesis/chromatin structure.histone	5.01e-08
28.1.3.2.4	DNA.synthesis/chromatin structure.histone.core.H4	0.000118
30.8	Signaling.misc	0.000153
28.1.3.2.3	DNA.synthesis/chromatin structure.histone.core.H3	0.000294
28.1	DNA.synthesis/chromatin structure	0.000978
27.3.24	RNA.regulation of transcription.MADS box transcription factor family	0.002902
30.2.8.1	Signaling.receptor kinases.leucine rich repeat VIII.VIII-1	0.00347
10.2.1	Cell wall.cellulose synthesis.cellulose synthase	0.004913
33	Development	0.005296
28.1.3.2.1	DNA.synthesis/chromatin structure.histone.core.H2A	0.00622
33.99	Development.unspecified	0.006365
30	Signaling	0.01005
10.5.1	Cell wall.cell wall proteins.AGPs	0.01069
10.5.1.1	Cell wall.cell wall proteins.AGPs.AGP	0.01069
28	DNA	0.0132
10.2	Cell wall.cellulose synthesis	0.01921
10	Cell wall	0.02384
30.2.8	Signaling.receptor kinases.leucine rich repeat VIII	0.02721
24	Biodegradation of Xenobiotics	0.03194
27.3.67	RNA.regulation of transcription.putative transcription regulator	0.04299
10.5	Cell wall.cell wall proteins	0.04417
29.7	Protein.glycosylation	0.0499

GLOSSARY

MD: Mesoamerican domesticated accessions.

MW: Mesoamerican wild accessions.

S, ***nH***, ***π***, ***Θ***, ***He***: statistics for the estimation of the levels of molecular diversity.

L_{π}, L_{Θ}, L_{He}: loss in molecular diversity (in MD compared to MW).

Lcv: loss in gene expression diversity (in MD compared to MW).

Log2FC: when significantly shifted, indicates down-regulation or up-regulation of gene expression.

CV: coefficient of variation, estimated for gene expression.

Selection indices: based on two between-groups and one within-groups genetic variation statistics that were likely to have been affected by differential selection for MD compared to MW (see Online Methods for details). Used to identify PS contigs.

PS contigs: contigs putatively under selection in MD compared to MW, identified by computing two selection indices and testing their significance with a coalescent simulation (see Online Methods for details). The opposite is PN contigs (i.e., putatively neutral contigs).

Coalescent simulations of domestication models: used to identify PS contigs by computing two selection indices and testing their significance. Coalescent simulations are used to generate neutral distributions of summary statistics, assuming three likely domestication scenarios (models) that are reconstructed and based on the population histories and demographic parameters estimated in previous studies.

PCCs: Pearson correlation coefficients among gene expression profiles; a wider distribution of PCCs supports stronger correlations.

Network analysis (proximity networks, relevance networks): concerns gene co-expression. In a network, a community is a group of genes with correlated expression (see Online Methods for details).

Rank index, Jaccard similarity: allow the comparison of networks.

Enrichment gene-function analysis and gene-set enrichment analysis: comparative degree of enrichment in gene functions in a comparison among communities and/or groups (e.g., MW *versus* MD; PS *versus* PN), and used to determine biological signals (in the communities) or to examine the functional characterization (PS contigs).

MapMan ontology and test on hypergeometric function: allows the functional characterization of contigs, and to derive which MapMan bins are statistically significantly over-represented.

SUPPLEMENTAL METHODS 1

1. Network-based analysis

1.1 Contig selection strategy and investigation of bias

The contigs for the network analyses were selected to avoid the inclusion of potentially noisy or invariant gene expression profiles, which might lead to the inclusion of spurious edges in the extracted networks.

We removed 581 contigs that showed zero expression levels in at least nine genotypes from the analysis. Of these, 40 were determined to be under selective pressure (PS), with 17 of these contigs having expression values of zero in all of the investigated samples. The removal of these contigs from the total of 27,243 did not introduce bias with respect to the contigs that were determined as PS, as their proportion was within the expected range (binomial test, p-value >0.2). In addition, 140 of these 581 contigs were annotated by MapMan bins. The results of the enrichment analysis are presented in Supplemental Table 8 online.

The log2FC between the MW and MD populations for the 581 contigs show negative values due to the larger expression values in the MD samples (see Supplemental Figure 6 online). There is a shift towards smaller CVs between MD and MW (Supplemental Figures 7A and 7B online), and this also holds when focusing on the PS contigs. The effect of removing the 581 contigs on the distribution of CVs was not pronounced (Supplemental Figure 7C online). However, there were statistically significant differences between the CVs of the 26,662 contigs and the CVs for the 581 removed contigs; these are, however, artefacts due to the high CVs for the removed contigs as a result of their small mean expression levels; this is a common drawback of the CV statistic (Supplemental Figure 8 online).

To select a subset of contigs to use for the network analysis, two statistical tests were performed for each of the remaining 26,662 contigs: the differences in the means and variance of the expression levels between the MW and MD populations, based on ANOVA and on the F-test, respectively (Ho et al., 2008). A very loose level of significance ($\alpha = 0.1$) was considered for both of these tests. The Wilcoxon rank sum test with continuity correction (Hollander and Wolf, 1973) was applied to the CVs computed for MW and MD for each chosen contig, to determine whether the strategy applied for the contig selection had introduced any systematic bias with respect to favoring contigs that vary strongly in both the MW and MD populations, in comparison to the entire set of contigs that were considered. The possibility of a shift in the CV towards higher/ lower values for the PS contigs retained and those excluded from subsequent analysis was also tested.

These tests resulted in the selection of 10,616 contigs, the profiles of which were further subjected to network-based analysis (Supplemental Figure 2 online). There was no bias associated with the contigs variation of gene expression (as shown in Supplemental Figure 3 online). From the remaining 2,324 PS contigs, about 50% were included in the network analysis (Supplemental Figure 2 online). However, we derived two much larger networks using cut-offs of 0.2 and 0.3 for ANOVA and homogeneity of variance (the F-test). We found that the results remained qualitatively unchanged.

1.2 Proximity and relevance networks

The MW and MD expression profiles for the 10,616 contigs were subjected to network-based analysis following two procedures, both of which were based on the Pearson correlation coefficients (PCCs): (i) generation of relevance networks; and (ii) extraction of proximity networks (Klie et al., 2012; Klessen et al., 2013). With the first procedure, an edge is established between two nodes (which represent genes) if the PCC between the corresponding expression profiles is higher than a threshold τ that ensures a given level of statistical significance. According to the second procedure, for a fixed k, an edge is created between two nodes, *u* and *v*, if *u* appears in the list of the k genes with the highest PCC to *v*, and if *v* is in the list of the k genes with the highest PCC to *u*. As a result, the second procedure, which generates a proximity network, takes into consideration the observation that the genes can often be activated as modules of a program to fulfill a particular function.

To compare the networks generated by these two procedures, we required that they were equally informative, as quantified by their density (i.e., the total number of edges divided by the number of edges in a complete network on the same number of nodes). This also implies that the selection of the threshold values in the procedures for network generation has to guarantee an (almost) equal density. The density is a number ranging from 0 to 1, with 0 where there are no edges between any two nodes, and 1 when any two nodes are connected by an edge.

It can be shown that the null distribution for k is almost uniform; therefore, with 10,000 genes, k = 100 will ensure a significance level of $\alpha = 0.01$. Here, we present and discuss the results for k = 15. To examine the robustness of the findings, we also repeated the analysis for $k \in \{10,20\}$ (data not shown). To specify the threshold value τ, we determined the value of the PCC that ensured that the generated network was of (almost) equal density as the proximity network (as established based on the same dataset). For instance, $\tau = 0.91$ in the procedure for extraction of the relevance network corresponds to k = 15 in the procedure for generating proximity networks. These thresholds were applied to both the MW and MD gene expression profiles, and the resulting networks were compared and contrasted based on their network properties.

1.3 Analysis based on relevance networks

The proximity network generation and properties are described in the main text (Online Methods and Results sections). As already indicated above, the relevance networks were extracted so that their densities matched those of the proximity networks. This facilitates further comparative analysis between the two procedures for network generation. With the corresponding threshold value of $\tau = 0.92$, the MW relevance network is again sparser than the MD relevance network: 1,661 isolated nodes (i.e., nodes without any edge) in MW, and 105 isolated nodes in MD. The largest connected component in the MW network included 8,257 nodes, while in the MD network, this contained 10,439 nodes. The relevance networks have a characteristic heavy-tailed degree distribution. As a result, there was a finer community structure in both of the relevance networks (Supplemental Figures 9A and 9B online), with 152 communities with at least two nodes in the MW network and 27 in the MD network. In the MW relevance network, the largest community contained 2,662 nodes, while in the MD relevance network contained 4,665 nodes. Altogether, nine and four communities included more than 100 nodes in the MW and MD networks, respectively. The enrichment analysis largely matched that for the proximity networks.

Regarding the assortativity with respect to the contigs under selection pressure, our findings were similar to those from proximity networks: 0.010 for MW and 0.008 for MD. For assortativity with respect to the CVs, as in the case of proximity networks, we found that MW showed higher clustering of similar contigs (0.44) in comparison to MD (0.35). The findings regarding the centrality of the PS contigs were in line with those obtained from the proximity networks.

2. Gene-set enrichment analysis using MapMan

To characterize the molecular functions and biological processes of the *Phaseolus* contigs under study, the latest version that was available for MapMan and *Arabidopsis* (version 1.1 from January 2010, http://mapman.gabipd.org/) was used (Thimm et al., 2004). By using the sequence homology of *Phaseolus* contigs and *Arabidopsis* genes, a total of 13,228 contigs were assigned with MapMan bins, disregarding the 35 MapMan (sub-)bins that corresponded to '*not assigned*'. The *Phaseolus* contig-to-MapMan mapping that was obtained was further processed to include the parent bins and to remove inconsistencies, as described in Klie and Nikoloski (2012).

To further assess whether a particular set of contigs (i.e., members of network communities, PS contigs, or contigs up-regulated in the wild forms) is predominantly involved in similar molecular functions or biochemical processes in MW and MD, gene-set enrichment analysis (GSEA) was performed (Subramanian et al., 2005). Briefly, for a given set of contigs, it was statistically tested whether certain MapMan bins are over-represented within this particular group of contigs; i.e., over-enriched. Classically, hypergeometric tests are used to derive which MapMan bins are statistically significantly over-represented (Rivals et al., 2007). The enrichment results presented throughout this manuscript were obtained using the previously mentioned hypergeometric test and the Benjamini-Hochberg false discovery rate correction to obtain *p* values

(Benjamini and Hochberg, 1995), as implemented in the statistical environment R. For all of the enrichment tests, the significance level was set to 5%.

SUPPLEMENTAL METHODS 2

3. Survey of enriched gene functions for the MW and MD communities

To determine whether the network communities carry biological signals, the enriched gene functions for each of the communities determined were investigated (at a significance level of $\alpha = 0.01$) using the MapMan ontology. The communities corresponded to modular structures of largely different functions (Supplemental Table 3 online; Supplemental Data Sets 1B and 1C online). We investigated whether the different classes of genes that were over-represented in either MW or MD were shown to be involved in the domestication process, either through direct experiment or by implication on the basis of function. The results of this survey are reported below.

- MADS-BOX. The MADS-BOX class of genes, which includes *APETALA1* (AP1; Mandel et al., 1992; Gustafson-Brown et al., 1994) and *CAULIFLOWER* (*CAL*; Kempin et al., 1995), have been shown to control the specification of floral meristem identity (Alvarez-Buylla et al., 2006). In maize, a domestication candidate gene, *GRMZM2G448355* (Hufford et al., 2012), is an ortholog of the Os*MADS56* gene, which delays flowering under long-day conditions in rice (Ryu et al., 2009). Interestingly, the evolution of the domesticated cauliflower (*B. oleracea* spp. *botrytis*) appears to be associated with mutations in the MADS-BOX floral meristem identity genes *CAL* and *AP1* (Kempin et al., 1995; Lowman and Purugganan, 1999).

- bZIP, WRKY. Drought responses include the induction of proteins belonging to several transcription factor and regulator families used in the reprogramming of the transcriptome to initiate adaptation strategies (Ramanjulu and Bartels, 2002), such as the bZIP, the MYB transcription factors, WRKY, and zinc-finger proteins. The bZIP transcription factor family induces transcription by binding to upstream abscisic-acid-responsive elements (Ramanjulu and Bartels, 2002), and differential expression of a member of this transcription-factor family was observed in a leaf subtractive cDNA library of a drought-sensitive genotype of wheat (*TTD-22*; Ergen and Budak, 2008). It is worth noting that several bZIP genes have been postulated to be involved in domestication. In maize, the *OPAQUE 2* (*O2*) gene, for example, was shown to encode a transcriptional activator that has a role in alterations to the grain phenotype during domestication (Henry et al., 2005). In addition, studies of natural variations in the seed dormancy in *Arabidopsis* led to the cloning of a QTL, *DELAY OF GERMINATION 1* (*DOG1*), which affects embryonic dormancy (Bentsink et al., 2006). *DOG1* encodes a member of a plant-specific protein family with a domain shared by the D class bZIP DNA-binding proteins (Sugimoto et al., 2010).

Interestingly, members of the WRKY family of proteins (Eulgem and Somssich, 2007) that are well-known for their roles in response to several abiotic stresses (Ramanjulu and Bartels, 2002) were almost exclusively induced in a drought-tolerant emmer wheat genotype (TR39477; Ergen and Budak, 2008). A number of transcription factors including many from the WRKY, AP2/EREBP, and MYB families, are also involved in leaf senescence (Eulgem et al., 2000; Chen et al., 2002).

The maternally expressed *Arabidopsis* WRKY transcription factor *TRANSPARENT TESTA GLABRA 2* (*TTG2*) appears to have a controlling role in seed size via the integument growth pathway; the seed length of a *ttg2* mutant is much smaller than the wild-type (Garcia et al., 2005).

Duplicated WRKY genes have been maintained in wild and cultivated plant species in the course of selection during domestication and polyploidization (Petitot et al., 2008). However, the majority of the WRKY genes analyzed respond to pathogen attack and to the endogenous signaling molecule salicylic acid (Eulgem and Somssich, 2007).

- Trihelix. The rice *Shattering 1* (*SHA1*) gene that encodes a GT-1-type factor has an important role in the activation of cell separation, and a mutation in the trihelix domain resulted in the elimination of seed shattering in cultivated rice (Lin et al., 2007). All of the the domesticated rice cultivars analysed harbor the mutant *sha1* gene, and have thus lost the ability to shed their seeds at maturity.

- Xylan. The cell wall mediates growth responses and often acts as the first barrier in interactions with a wide range of environmental factors (Stebbins, 1992; Somerville et al., 2004). Selection during the domestication process produced changes in the cell wall structure and composition; in this regard, in the study of Zhang et al. (2012), it was shown that the wild rice Yuanj has a reduced level of xylan associated with a less upright growth habit. Domestication appears to have conferred an increased level of xylan backbone to cultivated rice plants, along with an erect growth habit.

- Raffinose family. The most common of the raffinose family oligosaccharides (RFOs) are raffinose (RFO-trisaccharide) and stachyose (RFO-tetrasaccharide). The RFOs have diverse roles in plants, as they are used for transport and storage of carbon, and as solutes for the protection against abiotic stress (Bachmann et al., 1994; Haritatos et al., 1996; Taji et al., 2002). In many plant species, raffinose is also stored in high amounts in seeds, where it is thought to have an additional role in desiccation tolerance (Obendorf, 1997). The sucrose and RFO content in seeds might determine and indicate their storing capacity. Tahir et al. (2012) reported variations in relative concentrations of stachyose and raffinose between wild and domesticated lentil genotypes.

- MYB. Anthocyanin accumulation is controlled through the coordinated expression of genes encoding the anthocyanin biosynthetic pathway enzymes. From studies in a diverse array of plant species, it is apparent that this coordinated expression is controlled at the transcriptional level, usually by an R2R3 MYB and/or a basic helix-loop-helix transcription factor (Mol et al., 1996; Winkel-Shirley, 2001). More recently, advances have been achieved in apple (*Malus* × *domestica*) where the R2R3 MYB tanscriptional factors responsible for apple anthocyanin accumulation were identified (Espley et al., 2007). In a study seeking to examine the control of apple fruit skin anthocyanins (Ban et al., 2007), Md-MYBA was isolated from a pale red-skinned apple cultivar and characterized in a deep red-skinned cultivar.

A member of the MYB transcription factor family proteins that is involved in the control of several biological functions upon biotic and abiotic stresses (Lee et al., 2007) was differentially induced in a drought-sensitive genotype of wheat (TTD-22; Ergen and Budak 2008).

A gene identified in rice as producing a non-shattering phenotype (*sh4*) has a MYB3 DNA binding domain, which suggests that it is a transcription factor. A single SNP that leads to a single conserved amino-acid change has resulted in a change in function, which appears to result in either the incomplete formation of the abscission zone (Li et al., 2006) or the failure to initiate cell degradation (Lin et al., 2007). In the case of rice, two independent genetic pathways to nonshattering have occurred. The nonshattering phenotype probably came after other mutations that were associated with an increase in grain size and the waxy phenotype (Shomura et al., 2008), which supports the idea that the mutation arose in the domesticated population of rice.

- ZF-HD. Abscission is a process that is often implicated in domestication. Nakano et al. (2013) examined the transcriptomes of three tomato flower pedicel regions, the abscission zone and the flanking proximal and distal regions. They identified 89 genes that were preferentially expressed in the abscission zone compared to both the proximal and distal regions. These genes included several transcription factors that regulate apical or axillary shoot meristem activity. Nine genes belonging to seven transcription factor families were identified, among which there was a zinc-finger homeodomain (ZF-HD) family gene (*CK715116*). In particular, this gene encodes a protein with sequence similarity to the *Arabidopsis* ZF-HD family gene *HOMEOBOX PROTEIN33* product, which is involved in the abscisic acid response pathway (Wang et al., 2011).

4. Gene-function investigation

A survey of the function of a subset of PS contigs was carried out to investigate whether they are known to be associated to the domestication process in other species, using either direct experimentation or because of their function. In particular, we focused our survey on the 380 contigs that had the highest selection index (290), and considering also different patterns of polymorphism, all of the 23 PS contigs with alternative allelic states in MW and MD, and all of the 67 PS contigs that are monomorphic in MW and polymorphic in MD. Almost all of these contigs investigated had a BLASTX e-value ≤$10e^{-4}$, with few exceptions (six contigs had an e-value ≤$10e^{-2}$ and their functions were not relevant to domestication).

4.1 Contigs with the highest selection indices

The survey of the function of 221 annotated contigs among the 290 with the highest selection indices resulted in the following main points:

- Genes involved in "light" response pathway: Under directional selection, several genes were identified as involved in light perception or signaling and circadian and rhythm determination (*PHYA*, *EARLY FLOWERING4-Like 3*), photoperiod response (*GIGANTEA*), photoperiod adaptation (*DWARF IN LIGHT*), and timing of bolting and flowering (*VERNALIZATION INDEPENDENCE 5*, *AROGENATE DEHYDRATASE 1*, *AGAMOUS-LIKE8*). Another gene that was under positive selection, *KANADI2* (that together with *K1* modulates *PIN1*, a gene that, in turn, encodes an auxin transport protein), is a target (as also for *AGAMOUS* [see below] and other genes) of *LEAFY*, a master gene that has a central role in the regulation of the transition from vegetative growth to flower formation (Winter et al., 2011). *BRIZ1* is another interesting gene here, which has been shown to be comprised in the interval of Flt-2L QTL for flowering time in barley (Chen et al., 2009a). This QTL comprises Hv-*AP2*, which is highly similar to the wheat domestication gene Q (Förster et al., 2013). Finally, there is the *HIGH MOBILITY GROUP A* (*HMGA*) gene that encodes a DNA binding protein that interacts with A/T-rich stretches of DNA. Although this specific gene has an unknown function, it should be noted that other proteins, such as AHL22, that contain an AT-hook motif in *Arabidopsis* regulate flowering initiation by modifying the *FLOWERING LOCUS T* chromatin (Yun et al., 2012).

Among these genes, *GIGANTEA* (*GI*) has a pivotal role in the photoperiod response, as it controls flowering under long days in a circadian clock-controlled flowering pathway. In *Arabidopsis*, *GI* acts earlier than *CONSTANS* (*CO*) and *FLOWERING TIME* (*FT*) in the pathway, by increasing the levels of *CO* and *FT* mRNA (Mizoguchi et al., 2005). Literature surveys have shown that the genes with which *GI* interacts directly (i.e., *FT*, *CO*) were targets of selection during the domestication in crops like rice (*DTH CONSTANS*-like, Wu et al., 2013; and *Hd1 CONSTANS* homolog, Takahashi and Shimamoto, 2011) and sunflower (*HaFT1* and paralogs, Blackman et al., 2011). In pea, Hecht et al. (2007) identified *LATE BLOOMER 1* (*LATE1*) as the pea ortholog of *Arabidopsis GI*, and showed that it is necessary for promotion of flowering, production of a mobile flowering stimulus, and induction of an *FT* homolog under long-day conditions. Moreover, several other genes involved in flowering time were selected under domestication, such as in rice, wheat, barley, maize, rapeseed, sunflower, lentil, pea and sorghum (see Lenser and Theißen, 2013; Olsen and Wendel, 2013, for reviews). Interestingly, in *Arabidopsis*, it has been shown that *GIGANTEA* and *EARLY FLOWERING 4* (*ELF4*) show differential phase-specific genetic influences over a diurnal cycle (Kim et al., 2012), and that *GI* is epistatic to *ELF4* in flowering time determination, while *ELF4* is epistatic to *GI* in hypocotyl growth regulation. Moreover, *GI* and *ELF4* have a synergistic or additive effect on endogenous clock regulation. Kim et al. (2012), which suggests that this might allow diversity in the regulation of circadian physiological outputs to be achieved, including flowering time and hypocotyl growth. However, the present study is the first that reports that *GI* was selected during the process of domestication, while a putative transcriptional regulator that potentially functions in the Phytochrome A light signaling pathway was shown to be under selection during domestication of soybean (*E1* gene; Xia et al., 2012). This congruence is relevant, as the recent divergence of flowering-related genes in three legume species has been documented (*Glycine soya, Lotus japonicus*, *Medicago truncatula*; Kim et al., 2013).

AGL8 is another gene of outstanding interest. In *Arabidopsis*, it is negatively regulated by *APETALA 1*. The homolog of *APETALA 1* in wheat has been found to be selected under domestication (Yan et al., 2003; Golovnina et al., 2010). Moreover, *AGAMOUS* occupies a nodal position between the "floral induction" pathway and the "floral identity genes and meristem maintenance" pathway that control the formation of stamens, carpels, ovule and fruit development, and that determine the flowers across all angiosperms (Lenser and Theißen, 2013). Mutations in this "input-output" gene might alter a trait (or a series of traits) in a way that might otherwise only be achieved by concerted mutations within several upstream or downstream genes simultaneously, and it is for this reason that it is regarded as a good example to explain the molecular basis of convergent evolution during domestication (Lenser and Theißen, 2013).

Furthermore, the homolog of *AGL8* in soybean is considered as one of the three candidates for a major QTL for flowering time on chromosome 6 (Zhang et al., 2013).

- Genes that are pivotal to ensure correct hormonal perception, transport or biosynthesis: *TINY2* is an AP2/ERF (APETALA/Ethylene-responsive factor). In *Glycine max*, this gene regulates the Gm-*GA2ox4* gene, causing dwarfism (Suo et al., 2012). Another gene, At-*GRXS17*, has a critical role in redox homeostasis and hormone perception, to mediate temperature-dependent post-embryonic growth (Cheng et al., 2011); its mutants are altered in their perception of the phytohormone auxin and its polar transport. The *PIN3* gene encodes an auxin transport protein that is crucial for correct cellular coordination, and has a direct involvement in gravitropism and phototropism. It is implicated in the shade avoidance response regulated by phyA (here also found under selection) and phyB (Devlin et al., 2003). Moreover, *PIN* interacts with *INDEHISCENT* (*IND*) in regulation of the separation layer cells. The gene *WALL ARE THIN 1* is involved in numerous processes, such as auxin polar transport, positive regulation of auxins, regulation of meristem growth, anthocyanin accumulation in tissues in response to UV light, cell-wall organization and secondary cell-wall biogenesis in fibers, cell growth, regulation of cell size, root hair elongation, root morphogenesis, water transport, response to salt stress, and stem elongation (Ranocha et al., 2010). The gene *FRAGILE FIBER 1* (*FRA1*) is also an interesting candidate, as its mutants show altered orientation of cellulose microfibrils and reduced mechanical strength of fibers. Moreover, phylogenetic studies have shown that the rice gene *GIBBERELLIN INSENSITIVE DWARF1* (*GDD1*) is most similar to *At5g47820* (*FRA1*) of *Arabidopsis*. Mutation of rice *GDD1*, which encodes a kinesin-like protein that binds to a gibberellic acid biosynthesis gene promoter, leads to dwarfism with impaired cell elongation (Li et al., 2011). *FRA1* appears to be located 0.4 cM downstream of the *DINAGZ* QTL (a QTL for fiber digestibility) of chromosome 5 (Barriere et al., 2005). Finally, we found that apyrase 2 (*APY2*) is under selection. At-*APY1* and At-*APY2* have key roles in pollen germination and vegetative growth (Wu et al., 2007; Wolf et al., 2007). In *Arabidopsis*, the *apy1* and *apy2* double knock-out showed failed pollen germination. Moreover, lines silenced under RNA interference showed a dwarf phenotype in overall vegetative growth, and had dramatically reduced growth of the primary root and etiolated hypocotyls (Wu et al., 2007), and abnormal cotyledons (Wolf et al., 2007). Recent studies have indicated that a critical step connecting apyrase suppression to growth suppression is the inhibition of polar auxin transport (Liu et al., 2012). *AHP1* has a role in cytokinin signaling and affects multiple aspects of plant development. Cytokinins have diverse functions that include control of meristem activity, hormonal cross-talk, nutrient acquisition, responses to light (interactions with *PHYB*) and various stress responses (Jones et al., 2010; Brenner et al., 2012). Together with four other *AHP* genes and a histidine kinase, *CYTOKININ-INDEPENDENT1* (*CKI1*) gene, *AHP1* is implicated in the regulation of female gametophyte development and vegetative growth (Deng et al., 2010). Furthermore, a quintuple *ahp* mutant showed various abnormalities in growth and development, including reduced fertility, increased seed size, reduced vascular development, and a shortened primary root (Deng et al., 2010).

As indicated by Lenser and Theißen (2013), despite this seeming plurality of the domestication targets, two strategies have been repeatedly applied to promote the dwarfing habit in different species: alteration of the meristem identity to cause determinate growth (Salamini, 2003), or as the above-mentioned genes also suggest for *P. vulgaris*, interference with hormone metabolism or signaling (McGarry et al., 2012). In addition to the modulation of gibberellic acid signaling that is mainly seen in rice (Itoh et al., 2004; Asano et al., 2007, 2011), barley (Jia et al., 2009) and wheat (Doebley et al., 2006), interference with brassinosteroid signal transduction and with polar auxin transport has been shown to cause a semi-dwarfing growth habit in agronomically important varieties of barley (Chono et al., 2003) sorghum (Multani et al., 2003), and pearl millet (Parvathaeni et al., 2013). However, although in soybean and the common bean, *TERMINAL FLOWER* determines the growth habit and has been shown to be under selection during domestication (Tian et al., 2010; Kwak et al., 2012), it was not on our list of candidates.

- Genes involved in seed development and traits: Some of these genes are important for correct embryo development (*YODA*, Lukowitz et al., 2004; *POR*, Mathur, 2004; *AKRP*, Tzafrir et al., 2004; *BRIZ1*, Hsia and Callis, 2010; *EMB1211*, Liang et al., 2010; *IAA12*, Bureau et al., 2010), and others can affect the seed chemical composition, like for seed protein (*IAA12*, Yonekura et al., 2010), oil (*IAA12*, Kondo et al., 2010; and *LRR-RLK*) and fat (*FATB*, Molina et al., 2008) content,

and for seed maturation (*SEC14p*, Huang et al., 2013) and germination (*DOF6,* Rueda-Romero et al., 2012; *TKL,* Arc et al., 2012; *BGAL3*, Bui et al., 2011*; RBOHB*, Müller et al., 2009). Effects of domestication on these traits are inconceivable, as human selection might have led to seeds with development, maturation and germination that are more adapted to cultivation and field conditions. Moreover, during domestication, variations in seed chemical characteristics were also observed in several cases, such as, for example, the case of the *WAXY* gene in rice, maize, barley, sorghum, millet, and amaranth (Olsen and Wendel, 2013).

- Three genes have roles in the carbon/nitrogen (C/N) balance response: The first of these is a *trans*-membrane gamma-aminobutyrate (GABA) transporter. There is growing evidence that in plants the GABA shunt has a major role in primary C/N metabolism (Fait et al., 2008; Masclaux-Daubresse et al., 2008). The second is *TSD2* (*TUMOROUS SHOOT DEVELOPMENT2*), which encodes a putative methyltransferase with an essential role in cell adhesion that coordinates plant development (Krupková et al., 2007) and appears to be a critical modulator of the C/N nutrient balance response in *Arabidopsis* (Gao et al., 2008). Finally, a transketolase (AT3G60750) is among 16 genes involved in N nutrition and C metabolism up-regulated by NO_3^- (Girin et al., 2007). It has been shown that the transcribed segments of undomesticated plant genomes are the most N poor, with genomes and proteomes with signatures of N limitation (Acquisti et al., 2009). In this context, adaptation to N availability at the whole-plant level is highly conceivable, which would justify a role also for these genes during domestication.

- Genes involved in responses to environmental stresses: Several genes have homologs that encode proteins that are involved in biotic stresses (Ribosomal protein S14p/S29e family protein, Mitsuya et al., 2009; a tautomerase/MIF superfamily protein, Bricchi et al., 2012; HSF A4A, Kuśnierczyk et al., 2007; SIZ1, Lee et al., 2007; G6PDH1, Asai et al., 2011; Siddappaji et al., 2013). In this case, there are genes that control programmed cell death and are involved in innate immunity responses (*MKS1*, Andreasson et al., 2005; *WRKY4*, Lai et al., 2008; *ERD2*, Xu et al., 2012), and also in systemic acquired resistance (*ROF2*, Lin et al., 2012a; Copper transport protein family [*AT5G52740*], Ascencio-Ibáñez et al., 2008). Among the genes that are involved in biotic stress responses, *MKS1* and *WRKY4* are involved in the mitogen-activated protein kinase (MAPK) pathway. This is one of the main phosphorylation pathways that plants use in biotic and abiotic stress resistance. Specifically, the MAPK cascade and WRKY transcription factor functions act in response to both fungal and bacterial pathogens (Asai et al., 2002). In the case of abiotic stress, there are common genes involved in: drought tolerance or osmotic stress (RING E3 Ub ligase, Ryu et al., 2010; Cho et al., 2011; *CYCLIN H*, Zhou et al., 2013; *ERD2*, Taji et al., 1999; *SIZ1*, Catala et al., 2007), in stomata aperture/ closure (*SPHK1*, Worral et al., 2008; Guo et al., 2012), and in the control of abscissic acid synthesis (*AT5G04250*, encoding a cysteine proteinase superfamily protein, Jensen et al., 2013), response to heat (*Hsp70b*, Sung et al., 2001; *CPN10*, Kant et al., 2008; *HSFA4A*, Riechmann et al., 2000), cold (*STA1*, Lee et al., 2006), and salt (*ERD2*, Taji et al., 1999; *CYCLIN H*, Zhou et al., 2013) Stresses. Some other genes are also involved in the response to nutrient starvation (*SIZ1*, Miura et al., 2005, 2011) and in the tolerance to toxic metals in the soil, in particular to aluminum (*BCB*, Ezaki et al., 2000, 2001). The involvement of fast-evolving genes implicated in biotic and abiotic stress has recently been documented in tomato (Koenig et al., 2013). Moreover, metal tolerance is a trait found under domestication in different crops like wheat, rye, sorghum and corn, and constitutes an evident convergent selection during crop domestication (Lenser and Theißen, 2013).

Other genes that do not fall into these categories:

- *RKF1*; This RECEPTOR-LIKE KINASE IN FLOWERS gene showed the highest signal for directional selection. Twenty-five percent of the selectively expressed genes in pollen are included in the signal-transduction category. Most of these genes (26 of 37) encode putative protein kinases. *RKF1* is an example of the pollen-specific receptor kinases that are possibly involved in signaling events (Takahashi et al., 1998).

- *RPF1*; this gene encodes for RNA PROCESSING FACTOR 1 (RPF1), which is implicated in the restoration of cytoplasmic male sterility. This nuclear gene goes through mitochondrial RNA processing that leads to a protein that prevents pollen abortion. This trait is mitochondrially inherited, and it has been observed in various plant species (Holzle et al., 2011). Another gene that

is essential for pollen development and that has a crucial role in male gametophyte development and male fertility is *ATMGT9* (Chen et al., 2009b), which encodes a magnesium transporter that has been shown to have a role in local adaptation to serpentine soils (high heavy-metal content, low calcium-to-magnesium ratios) in *Arabidopsis lyrata* (Turner et al., 2010).

- The homolog of *YABBY5* (*YAB5*); *YAB5* is involved in the molecular determination of abscission in tomato fruit, a trait that is also under domestication (Nakano et al., 2013). Moreover, recently, a *YAB* homologue gene, *SH1*, and its homologs, was identified in the regulation of seed shattering in cereal species, including sorghum, rice and maize (Lin et al., 2012b). In sorghum, the shattering is determined by a single gene (*SH1*) that encodes a YABBY transcription factor. In maize and rice, these encode YABBY-LIKE transcription factors (Lin et al., 2012b).

- A gene encoding proteins that belong to the transducin/WD40 repeat-like superfamily. This gene is involved in the responses to the auxin stimulus and has also been implicated in defense responses (Rodrigues et al., 2013). In *Arabidopsis*, it is involved in several processes, including cell-chromatin silencing by RNA interference, glucuronoxylan metabolic processes, gravitropism, hydrogen peroxide biosynthetic processes, meiotic DNA double-strand break formation, meiotic chromosome segregation, organ morphogenesis, positive regulation of organelle organization, protein desumoylation, reciprocal meiotic recombination, regulation of anion channel activity, regulation of chromosome organization, sister chromatid cohesion, tissue development, vegetative-to-reproductive phase transition of the meristem, and xylan biosynthetic processes.

Interestingly, a WD40 protein in *G. max* is involved in seed development (Joshi et al., 2013). Moreover, a *WD40* (or *WUSCHEL*, as this has not been definitely identified) gene acts to determine the number of locules (fruit size) in tomato, a trait that was under selection in this crop during domestication (Muños et al., 2011; Ranc et al., 2012). Moreover, a WD40 protein is involved in pigmentation in *Sorghum bicolor* (Wu et al., 2012).

- Two genes are homologous of genes that control the color of the “edible part”, a trait that has often been affected by domestication in several herbaceous (potato, rice) and tree crops (orange and grapevine). Again, it is also considered an example of molecular convergence underlying domestication-related phenotypic changes (Lenser and Theißen, 2013; Olsen and Wendel, 2013). The first is *BCH1* that is involved in hydroxylation of beta-carotene (Tian et al., 2003), and is implicated in the pathway that confers color to tomato and pepper (Paran and van der Knaap, 2007). A major QTL for carrot color has been found for the pigment content that is mainly determined by two genes (*Y* and *Y2*) that are associated with domestication in carrot (wild, white; domesticated, orange). The other gene, *ATCDC5*, has been shown to have a key role in plant development, secondary metabolism, hormone signal transduction, disease resistance, and abiotic stress tolerance. However, it has also been observed that the CDC5 transcription factor binds a deleted region of the *DFR-A* onion gene. Interestingly, mutations in this gene were associated with the changed colors of onion after the domestication process (Song et al., 2014).

- A gene related to yield. This encodes a leucine-rich repeat protein kinase family protein (*AT2G26730*). Wild rice (*Oryza rufipogon*) has an important role by contributing to modern rice breeding. Song et al. (2009) sequenced and analyzed a 172-kb genomic DNA region of wild rice around the RM5 locus, which is associated with the yield QTL *yld1.1*. As suggested previously, two putative receptor-like protein kinase genes (one of which is *AT2G26730*) were key suspects for *yld1.1* (Song et al., 2009).

4.2 Contigs monomorphic in MW and polymorphic in MD

We also investigated the 47 annotated contigs among the 67 PS contigs that were polymorphic in MD and monomorphic in MW:

- *ELF4-L2* (*EARLY FLOWERING4-Like 2*); This gene is involved in the control of the circadian rhythm, and also functions as part of the light-input pathway. In *Pisum sativum*, Liew et al. (2009) isolated an ortholog of the *Arabidopsis ELF4-L2* (*DIE NEUTRALIS*, *DNE*) locus, and confirmed that this gene inhibits flowering under noninductive short-day conditions. *DNE* has a role in diurnal and/or circadian regulation of several clock genes, including the pea *GI* ortholog *LATE BLOOMER 1(LATE1)*.

- *KAN1* (*KANADI1*) (that together with *K2* modulates, in turn, the *PIN1* gene, which encodes an auxin transport protein); This gene is a target of *LEAFY*, a master gene that has a central role in the regulation of the transition from vegetative growth to flower formation. Interestingly, while *ELF4-L3* and *KAN2* are under selection toward a reduction in diversity in the domesticated form, for *ELF4-L2* and *KAN1* the opposite is true; i.e., they have evolved toward an increase in diversity in the domesticated forms.

- *MYB102*; De Vos et al. (2006) showed that At-*MYB102* has a role in defense against caterpillars. This genes is also involved in wounding and osmotic stress responses; interestingly, herbivore-damaged plants also suffer from water loss (De Vos et al., 2006).

- *KUP6* (*K^+ uptake transporter 6*); Osakabe et al. (2013) demonstrated that the KUP potassium transporter family has important roles in water-stress responses and growth in plants; moreover, KUP-type K^+ transporters are induced by different stresses with an osmotic component, and specifically inhibit cell expansion while enhancing drought tolerance.

- *APC*; In rice, *APC* controls *RSS1*, which is important for maintenance of the shoot meristem under abiotic stress conditions, and is also thought to control tolerance responses to salt, drought, and cold (Ogawa et al., 2011).

- *ETT* (*ETTIN MUTATIONS*); other names *AUXIN RESPONSE TRANSCRIPTION FACTOR3* (*ARF3*); This gene positively regulates *FIL* (required for the “super-replum” phenotype) activity during leaf development (Garcia et al., 2006; Sorefan et al., 2009). Moreover, in *Arabidopsis*, González-Reig et al. (2012) hypothesized that subtle changes in the expression of the antagonistic factors involved in auxin efflux can produce drastic changes in the size of the different tissue types, including the variability of fruit shape in Brassicaceae and other related species. Another example is in rice, where at least two *ARF*s have been detected as selected genes, and where it has been suggested that these genes can be targeted by domestication selection for enhanced growth responses and productivity (Wang et al., 2010).

- Transducin/WD40 repeat-like superfamily protein; This is involved in the response to the auxin stimulus and it is also implicated in defense responses (Rodrigues et al., 2013). A WD40 protein in *G. max* is involved in seed development (Joshi et al., 2013). Moreover, a WD40 protein probably acts together with the *WUSCHEL* gene to determine the number of locules (lc) in tomato, a trait that was under selection in this crop during domestication (Muños et al., 2011; Ranc et al., 2012). Also in this case, it must be noted that members of this family protein have also been selected in the opposite direction; i.e., towards a reduction of diversity in the domesticated form (see above).

- Squamosa BP-like (*SPL*); So far, *SPL* has also been functionally linked to diverse developmental processes, such as seed germination and seedling development (Martin et al., 2010), leaf and plastocron development (Moreno et al., 1997), juvenile-to-adult and floral phase transitions (Cardon et al., 1997; Wu and Poethig, 2006; Gandikota et al., 2007; Wang et al., 2009), fruit ripening (Manning et al., 2006), copper homeostasis (Kropat et al., 2005; Yamasaki et al., 2009), programmed cell death (Stone et al., 2005), domestication (Wang et al., 2005; Preston and Hileman, 2013), and grain yield (Jiao et al., 2010; Miura et al., 2010).

- GRAS transcription factor family; This family includes monoculm1 (*moc1*), which was characterized from a rice mutant that had almost no branching (Li et al., 2003). In rice, *moc1* affects all branching meristems, so that both vegetative and inflorescence branching are severely curtailed (Doust, 2007). This gene is necessary for axillary meristem initiation. It is related to the *LATERAL SUPPRESSOR* gene (*LAve S* and *ls* in the model dicot plant systems of *Arabidopsis* and tomato, respectively), which controls aspects of branching.

- Other genes *ALEURAIN-LIKE PROTEASE* (*AALP*) and *SENESCENCE ASSOCIATED GENE2* (*SAG2*); In *Arabidopsis*, these are involved in senescence (Grbić, 2003).

4.3 Contigs fixed for alternative alleles in MW and MD

Finally, we specifically focused on the 19 annotated contigs among the 23 PS contigs that were fixed for alternative alleles in MW and MD. Some of these genes are involved in responses to biotic stress, such as insect responses (*JAZ*, also known as *TIFY*, Kazan and Manners, 2012; Schweizer et al., 2013) and responses to bacteria (*LOL2*, Epple et al., 2003). Another gene, which encodes a

GDSL-like lipase, is mainly involved in the response to water deprivation (Boyce et al., 2003). Two other genes are also interesting: the ribosomal protein L36 in *Lycopersicon pimpillinelifolium* co-maps with a QTL for wall thickness, and in *L. esculentum* for flavonoid content (Okmen et al., 2011). Finally, *CHC1* is predicted to encode a protein that belongs to the chromodomain remodeling complex. In *Arabidopsis*, two RNA-interference knock-down lines have a dwarf phenotype and reduced rates of *Agrobacterium*-mediated transformation. The low rate of root-mediated transformation might result from altered root morphology or reduced root growth rates (Crane and Gelvin, 2007; Campi et al., 2012). This genes has been suggested to enhance plant growth (De et al., 2012). Other details of these genes are reported below:

- *JAZ12* (also known as *TIFY3B*); In *Arabidospis*, this gene is involved in the jasmonic-acid-mediated signaling pathway, positive regulation of the flavonoid biosynthetic process, regulation of the plant-type hypersensitive response and the response to wounding. The JAZ/TIFY protein has MYC2, MYC3, MYC4 and NINJA as its interacting co-repressors and transcription factors (Kazan and Manners, 2012). Recently, it was shown that *Arabidopsis* basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, and insect performance and feeding behavior (Schweizer et al., 2013).

- Ribosomal protein L36; This is involved in DNA-dependent transcription and elongation, and ribosome biogenesis and translation. In *Arabidopsis*, it is ethylene regulated (De Paepe et al., 2004). This gene co-maps with a QTL for wall thickness in *L. pimpillinelifolium* and flavonoid content in *L. esculentum* (Okmen et al., 2011).

- *CHC1* (*CLATHRIN HEAVY CHAIN1*); In *Arabidopsis*, *CHC1* is predicted to encode a protein that belongs to the chromodomain remodeling complex (Crane and Gelvin, 2007; Campi et al., 2012). Two RNA-interference knock-down lines have a dwarf phenotype and reduced rates of *Agrobacterium*-mediated transformation. The low rate of root-mediated transformation might result from altered root morphology or reduced root growth rates. This gene is among those for complexes of AN3-interacting proteins. *AN3* is an “intrinsic yield gene” in *Arabidopsis*. Their use has been suggested for plant growth (De et al., 2012).

SUPPLEMENTAL REFERENCES

Acquisti, C., Elser J.J., and Kumar, S. (2009). Ecological nitrogen limitation shapes the DNA composition of plant genomes. Mol. Biol. Evol. **26:** 953-956.

Alvarez-Buylla, R., Garcia-Ponce, B., and Garay-Arroyo, A. (2006). Unique and redundant functional domains of APETALA1 and CAULIFLOWER, two recently duplicated *Arabidopsis thaliana* floral MADS-box genes. J. Exp. Bot. **57:** 3099-3107.

Andreasson, E., et al. (2005). The MAP kinase substrate MKS1 is a regulator of plant defense responses. EMBO J. **24:** 2579-2589.

Arc, E., Chibani, K., Grappin, P., Jullien, M., Godin, B., Cueff, G., Valot, B., Balliau, T., Job, D., and Rajjou, L. (2012). Cold stratification and exogenous nitrates entail similar functional proteome adjustments during *Arabidopsis* seed dormancy release. J. Proteome Res. **11:** 5418-5432.

Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.L., Gomez-Gomez, L., Boller, T., Ausubel, F.M., and Sheen, J. (2002). MAP kinase signalling cascade in *Arabidopsis* innate immunity. Nature **415:** 977-983.

Asai, S., Yoshioka, M., Nomura, H., Tone, C., Nakajima, K., Nakane, E., Doke, N., and Yoshioka, H. (2011). A plastidic glucose-6-phosphate dehydrogenase is responsible for hypersensitive response cell death and reactive oxygen species production. J. Gen. Plant Pathol. **77:** 152-162.

Asano, K., Takashi, T., Miura, K., Qian, Q., Kitano, H., Matsuoka, M., and Ashikari, M. (2007). Genetic and molecular analysis of utility of sd1 alleles in rice breeding. Breed. Sci. **57:** 53-58.

Asano, K., et al. (2011). Artificial selection for a green revolution gene during japonica rice domestication. Proc. Natl. Acad. Sci. USA **108:** 11034-11039.

Ascencio-Ibáñez, J.T., Sozzani, R., Lee, T.J., Chu, T.M., Wolfinger, R.D., Cella, R., and Hanley-Bowdoin, L. (2008). Global analysis of *Arabidopsis* gene expression uncovers a

complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol. **148:** 436-454.

Bachmann, M., Matile, P., and Keller, F. (1994). Metabolism of the raffinose family oligosaccharides in leaves of *Ajuga reptans* L. Cold acclimation, translocation, and sink to source transition: discovery of a chain elongation enzyme. Plant Physiol. **105:** 1335-1345.

Ban, Y., Honda, C., Hatsuyama, Y., Igarashi, M., Bessho, H., and Moriguchi, T. (2007). Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. Plant Cell Physiol. **48:** 958-970.

Barrière, Y., Laperche, A., Barrot, L., Aurel, G., Briand, M., and Jouanin, L. (2005). QTL analysis of lignification and cell wall digestibility in the Bay-0× Shahdara RIL progeny of *Arabidopsis thaliana* as a model system for forage plant. Plant Sci. **168:** 1235-1245.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B **57:** 289-300.

Bentsink, L., Jowett, J., Hanhart, C.J., and Koornneef, M. (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis.* Proc. Natl. Acad. Sci. USA **103:** 17042-17047.

Blackman, B.K., Rasmussen, D.A., Strasburg, J.L., Raduski, A.R., Burke, J.M., Knapp, S.J., Michaels, S.D., and Rieseberg, L.H (2011). Contributions of flowering time genes to sunflower domestication and improvement. Genetics **187:** 271-287.

Boyce, J.M., Knight, H., Deyholos, M., Openshaw, M.R., Galbraith, D.W., Warren, G., and Knight, M.R. (2003). The sfr6 mutant of *Arabidopsis* is defective in transcriptional activation via CBF/DREB1 and DREB2 and shows sensitivity to osmotic stress. Plant J. **34:** 395-406.

Brenner, W.G., Ramireddy, E., Heyl, A., and Schmülling, T. (2012). Gene regulation by cytokinin in *Arabidopsis.* Front. Plant Sci. **3:** 8.

Bricchi, I., Bertea, C.M., Occhipinti, A., Paponov, I.A., and Maffei, M.E. (2012). Dynamics of membrane potential variation and gene expression induced by *Spodoptera littoralis*, *Myzus persicae*, and *Pseudomonas syringae* in *Arabidopsis.* PLoS one **7:** e46673.

Bui, M., Lim, N., Sijacic, P., and Liu, Z. (2011). LEUNIG HOMOLOG and LEUNIG regulate seed mucilage extrusion in *Arabidopsis.* J. Integr. Plant. Biol. **53:** 399-408.

Bureau, M., Rast, M.I., Illmer, J., and Simon, R. (2010). JAGGED LATERAL ORGAN (JLO) controls auxin dependent patterning during development of the *Arabidopsis* embryo and root. Plant Mol. Biol. **74:** 479-491.

Campi, M., D'Andrea, L., Emiliani, J., and Casati, P. (2012). Participation of chromatin-remodeling proteins in the repair of ultraviolet-B-damaged DNA. Plant Physiol. **158:** 981-995.

Cardon, G.H., Hohmann, S., Nettesheim, K., Saedler, H., and Huijser, P. (1997). Functional analysis of the *Arabidopsis thaliana* SBP-box gene SPL3: a novel gene involved in the floral transition. Plant J. **12:** 367-377.

Catala, R., Ouyang, J., Abreu, I.A., Hu, Y., Seo, H., Zhang, X., and Chua, N.H. (2007). The *Arabidopsis* E3 SUMO ligase SIZ1 regulates plant growth and drought responses. Plant Cell **19:** 2952-2966.

Chen, A., Baumann, U., Fincher, G.B., and Collins, N.C. (2009a). Flt-2L, a locus in barley controlling flowering time, spike density, and plant height. Funct. Integr. Genomics **9:** 243-254.

Chen, J., Li, L.G., Liu, Z.H., Yuan, Y.J., Guo, L.L., Mao, D.D., Tian, L.F., Chen, L.B., Luan, S., and Li, D.P. (2009b). Magnesium transporter AtMGT9 is essential for pollen development in *Arabidopsis*. Cell Res. **19:** 887-898.

Chen, W., et al. (2002). Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell **14:** 559-574.

Cheng, N.H., Liu, J.Z., Liu, X., Wu, K., Thompson, S.M., Lin, J., Chang, J., Whitham, S.A., Park, S., Cohen, J.D., and Hirschi, K.D. (2011). *Arabidopsis* monothiol glutaredoxin, AtGRXS17, is critical for temperature-dependent postembryonic growth and development via modulating auxin response. J. Biol. Chem. **286:** 20398-20406.

Cho, S.K., Ryu, M.Y., Seo, D.H., Kang, B.G., and Kim, W.T. (2011). The *Arabidopsis* RING E3 ubiquitin ligase AtAIRP2 plays combinatory roles with AtAIRP1 in abscisic acid-mediated drought stress responses. Plant Physiol. **157:** 2240-2257.

Chono, M., Honda, I., Zeniya, H., Yoneyama, K., Saisho, D., Takeda, K., Takatsuto, S., Hoshino, T., and Watanabe, Y. (2003). A semidwarf phenotype of barley uzu results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. Plant Physiol. **133:** 1209-1219.

Crane, Y.M., and Gelvin, S.B. (2007). RNAi-mediated gene silencing reveals involvement of *Arabidopsis* chromatin-related genes in *Agrobacterium*-mediated root transformation. Proc. Natl. Acad. Sci. USA **104:** 15156-15161.

De, J.G., Inzé, D., and Verkest, A. (2012). U.S. Patent No. 20,120,324,602. Washington, DC: U.S. Patent and Trademark Office.

De Paepe, A., Vuylsteke, M., Van Hummelen, P., Zabeau, M., and Van Der Straeten, D. (2004). Transcriptional profiling by cDNA-AFLP and microarray analysis reveals novel insights into the early response to ethylene in *Arabidopsis.* Plant J. **39:** 537-559.

De Vos, M., Van Zaanen, W., Koornneef, A., Korzelius, J.P., Dicke, M., Van Loon, L.C., and Pieterse, C.M. (2006). Herbivore-induced resistance against microbial pathogens in *Arabidopsis.* Plant Physiol. **142:** 352-363.

Deng, Y., Dong, H., Mu, J., Ren, B., Zheng, B., Ji, Z., Yanga, W.C., Lianga, Y., and Zuo, J. (2010). *Arabidopsis* histidine kinase CKI1 acts upstream of histidine phosphotransfer proteins to regulate female gametophyte development and vegetative growth. Plant Cell **22:** 1232-1248.

Devlin, P.F., Yanovsky, M.J., and Kay, S.A. (2003). A genomic analysis of the shade avoidance response in *Arabidopsis.* Plant Physiol. **133:** 1617-1629.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell **127:** 1309-1321.

Doust, A. (2007). Architectural evolution and its implications for domestication in grasses. Ann. Bot. **100:** 941-950.

Epple, P., Mack, A.A., Morris, V.R., and Dangl, J.L. (2003). Antagonistic control of oxidative stress-induced cell death in *Arabidopsis* by two related, plant-specific zinc finger proteins. Proc. Natl. Acad. Sci. USA **100:** 6831-6836.

Ergen, N.Z., and Budak, H. (2008). Sequencing over 13,000 expressed sequence tags from six subtractive cDNA libraries of wild and modern wheats following slow drought stress. Plant Cell Environ. **32:** 220-236.

Espley, R.V., Hellens, R.P., Putterill, J., Stevenson, D.E., Kutty-Amma, S., and Allan, A.C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. Plant J. **49:** 414-427.

Eulgem, T., Rushton, P.J., Robatzek, S., and Somssich, I.E. (2000). The WRKY superfamily of plant transcription factors. Trends Plant Sci. **5:** 199-206.

Eulgem, T., and Somssich, I.E. (2007). Networks of WRKY transcription factors in defense signaling. Curr. Opin. Plant Biol. **10:** 366-371.

Ezaki, B., Gardner, R.C., Ezaki, Y., and Matsumoto, H. (2000). Expression of aluminum-induced genes in transgenic *Arabidopsis* plants can ameliorate aluminum stress and/or oxidative stress. Plant Physiol. **122:** 657-666.

Ezaki, B., Katsuhara, M., Kawamura, M., and Matsumoto, H. (2001). Different mechanisms of four aluminum (Al)-resistant transgenes for Al toxicity in *Arabidopsis.* Plant Physiol. **127:** 918-927.

Fait, A., Fromm, H., Walter, D., Galili, G., and Fernie, A.R. (2008). Highway or byway: the metabolic role of the GABA shunt in plants. Trends Plant Sci. **13:** 14-19.

Förster, S., Schumann, E., Baumann, M., Weber, W.E., and Pillen, K. (2013). Copy number variation of chromosome 5A and its association with Q gene expression, morphological aberrations, and agronomic performance of winter wheat cultivars. Theor. Appl. Genet. **126:** 3049-3063.

Gandikota, M., Birkenbihl, R.P., Höhmann, S., Cardon, G.H., Saedler, H., and Huijser, P. (2007). The miRNA156/157 recognition element in the 3' UTR of the *Arabidopsis* SBP box gene SPL3 prevents early flowering by translational inhibition of seedlings. Plant J. **49:** 683-693.

Gao, P., Xin, Z., and Zheng, Z.L. (2008). The OSU1/QUA2/TSD2-encoded putative methyltransferase is a critical modulator of carbon and nitrogen nutrient balance response in *Arabidopsis*. PLoS one **3:** e1387.

Garcia, D., Gerald, J.N.F., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis.* Plant Cell **17:** 52-60.

Garcia, D., Collier, S.A., Byrne, M.E., and Martienssen, R.A. (2006). Specification of leaf polarity in *Arabidopsis* via the trans-acting siRNA pathway. Curr. Biol. **16:** 933-938.

Girin, T., Lejay, L., Wirth, J., Widiez, T., Palenchar, P.M., Nazoa, P., Touraine, B., Gojon, A., and Lepetit, M. (2007). Identification of a 150 bp cis-acting element of the AtNRT2. 1 promoter involved in the regulation of gene expression by the N and C status of the plant. Plant Cell Environ. **30:** 1366-1380.

Golovnina, K., Kondratenko, E.Y., Blinov, A.G., and Goncharov, N.P. (2010). Molecular characterization of vernalization loci VRN1 in wild and cultivated wheats. BMC Plant Biol. **10:** 168.

González-Reig, S., Ripoll, J.J., Vera, A., Yanofsky, M.F., and Martínez-Laborda, A. (2012). Antagonistic gene activities determine the formation of pattern elements along the mediolateral axis of the *Arabidopsis* fruit. PLoS Genet. **8:** e1003020.

Grbić, V. (2003). SAG2 and SAG12 protein expression in senescing *Arabidopsis* plants. Physiol. Plant **119:** 263-269.

Guo, L., Mishra, G., Markham, J.E., Li, M., Tawfall, A., Welti, R., and Wang, X. (2012). Connections between sphingosine kinase and phospholipase D in the abscisic acid signaling pathway in *Arabidopsis*. J. Biol. Chem. **287:** 8286-8296.

Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1994). Regulation of the *Arabidopsis* floral homeotic gene APETALA1. Cell **76:** 131-143.

Haritatos, E., Keller, F., and Turgeon, R. (1996). Raffinose oligosaccharide concentrations measured in individual cell and tissue types in *Cucumis melo* L. leaves: implications for phloem loading. Planta **198:** 614-622.

Hecht, V., Knowles, C.L., Vander Schoor, J.K., Liew, L.C., Jones, S.E., Lambert, M.J., and Weller, J.L. (2007). Pea LATE BLOOMER1 is a GIGANTEA ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. Plant Physiol. **144:** 648-661.

Henry, A.M., Manicacci, D., Falque, M., and Damerval, C. (2005). Molecular evolution of the Opaque-2 gene in *Zea mays* L. J. Mol. Evol. **61:** 551-558.

Ho, J.W.K., Stefani, M., dos Remedios, C.G., and Charleston, M.A. (2008). Differential variability analysis of gene expression and its application to human diseases. Bioinformatics **24:** i390-i398.

Hollander, M., and Wolfe, D.A. (1973). Nonparametric Statistical Methods. New York: John Wiley and Sons. pp. 68-75.

Hölzle, A., Jonietz, C., Törjek, O., Altmann, T., Binder, S., and Forner, J. (2011). A RESTORER OF FERTILITY-like PPR gene is required for 5′-end processing of the nad4 mRNA in mitochondria of *Arabidopsis thaliana.* Plant J. **65:** 737-744.

Hsia, M.M., and Callis, J. (2010). BRIZ1 and BRIZ2 proteins form a heteromeric E3 ligase complex required for seed germination and post-germination growth in *Arabidopsis thaliana.* J. Biol. Chem. **285:** 37070-37081.

Huang, D., Koh, C., Feurtado, J.A., Tsang, E.W., and Cutler, A.J. (2013). MicroRNAs and their putative targets in *Brassica napus* seed maturation. BMC Genomics **14:** 140.

Hufford, M.B., et al. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. **44:** 808-811.

Itoh, H., Tatsumi, T., Sakamoto, T., Otomo, K., Toyomasu, T., Kitano, H., Ashikari, M., Ichihara, S., and Matsuoka, M. (2004). A rice semi-dwarf gene, Tan-Ginbozu (D35), encodes the gibberellin biosynthesis enzyme, ent-kaurene oxidase. Plant Mol. Biol. **54:** 533-547.

Jensen, M.K., Lindemose, S., Masi, F.D., Reimer, J.J., Nielsen, M., Perera, V., Workman, C.T., Turck, F., Grant, M.R., Mundy, J., Petersen M., and Skriver, K. (2013). ATAF1 transcription factor directly regulates abscisic acid biosynthetic gene NCED3 in *Arabidopsis thaliana.* FEBS open bio **3:** 321-327.

Jia, Q., Zhang, J., Westcott, S., Zhang, X. Q., Bellgard, M., Lance, R., and Li, C. (2009). GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. Funct. Integr. Genomics **9:** 255-262.

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., Dong, G., Zeng, D., Lu, Z., Zhu, X., Qian, Q., and Li, J. (2010). Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. Nat. Genet. **42:** 541-545.

Jones, B., Gunnerås, S.A., Petersson, S.V., Tarkowski, P., Graham, N., May, S., Dolezal, K., Sandberga, G., and Ljung, K. (2010). Cytokinin regulation of auxin synthesis in *Arabidopsis* involves a homeostatic feedback loop regulated via auxin and cytokinin signal transduction. Plant Cell **22:** 2956-2969.

Joshi, T., Valliyodan, B., Wu, J.H., Lee, S. H., Xu, D., and Nguyen, H.T. (2013). Genomic differences between cultivated soybean, *G. max* and its wild relative *G. soja.* BMC Genomics **14**: S5.

Kant, P., Gordon, M., Kant, S., Zolla, G., Davydov, O., Heimer, Y. M., Chalifa-Caspi, V., Shaked, R., and Barak, S. (2008). Functional-genomics-based identification of genes that regulate *Arabidopsis* responses to multiple abiotic stresses. Plant Cell Environ. **31:** 697-714.

Kazan, K., and Manners, J.M. (2012). JAZ repressors and the orchestration of phytohormone crosstalk. Trends Plant Sci. **17:** 22-31.

Kempin, S.A., Savidge, B., and Yanofsky, M.F. (1995). Molecular basis of the cauliflower phenotype in *Arabidopsis.* Science **267:** 522-525.

Kim, Y., Yeom, M., Kim, H., Lim, J., Koo, H. J., Hwang, D., Somers, D., Nam, H.G (2012). GIGANTEA and EARLY FLOWERING 4 in *Arabidopsis* exhibit differential phase-specific genetic influences over a diurnal cycle. Mol. Plant **5:** 152-161.

Kim, M.Y., Kang, Y.J., Lee, T., and Lee, S.H. (2013). Divergence of flowering-related genes in three legume species. Plant Genome **6:** 3.

Kleessen, S., Klie, S., and Nikoloski, Z. (2013). Data integration through proximity-based networks provides biological principles of organization across scales. Plant Cell **25:** 1917-27.

Klie, S., and Nikoloski, Z. (2012). The choice between MapMan and Gene Ontology for automated gene function prediction in plant science. Front. Genet. **3:** 115.

Klie, S., Mutwil, M., Persson, S., and Nikoloski, Z. (2012). Inferring gene functions through dissection of relevance networks: interleaving the intra- and inter-species views. Mol. Bio. Syst. **8:** 2233-2241.

Koenig, D., et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc. Natl. Acad. Sci. USA **110:** E2655-E2662.

Kondo, S., Ohto, C., Takagi, M., Matsui, K., Koyama, T., Mitsuda, N., Muramoto, N., Mitsukawa, N., and Tanaka, T. (2010). U.S. Patent Application 13/376,138.

Kropat, J., Tottey, S., Birkenbihl, R.P., Depege, N., Huijser, P., and Merchant, S. (2005). A regulator of nutritional copper signaling in *Chlamydomonas* is an SBP domain protein that recognizes the GTAC core of copper response element. Proc. Natl. Acad. Sci. USA **102:** 18730-18735.

Krupková, E., Immerzeel, P., Pauly, M., and Schmülling, T. (2007). The TUMOROUS SHOOT DEVELOPMENT2 gene of *Arabidopsis* encoding a putative methyltransferase is required for cell adhesion and co-ordinated plant development. Plant J. **50:** 735-750.

Kuśnierczyk, A., Winge, P., Midelfart, H., Armbruster, W.S., Rossiter, J.T., and Bones, A.M. (2007). Transcriptional responses of *Arabidopsis thaliana* ecotypes with different glucosinolate profiles after attack by polyphagous *Myzus persicae* and oligophagous *Brevicoryne brassicae.* J. Exp. Bot. **58:** 2537-2552.

Kwak, M., Toro, O., Debouck, D.G., and Gepts, P. (2012). Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). Ann. Bot. **110:** 1573-1580.

Lai, Z., Vinod, K.M., Zheng, Z., Fan, B., and Chen, Z. (2008). Roles of *Arabidopsis* WRKY3 and WRKY4 transcription factors in plant responses to pathogens. BMC Plant biol. **8**: 68.

Lee, B.H., Kapoor, A., Zhu, J., and Zhu, J.K. (2006). STABILIZED1, a stress-upregulated nuclear protein, is required for pre-mRNA splicing, mRNA turnover, and stress tolerance in *Arabidopsis.* Plant Cell **18**: 1736-1749.

Lee, J., et al. (2007). Salicylic acid-mediated innate immunity in *Arabidopsis* is regulated by SIZ1 SUMO E3 ligase. Plant J. **49**: 79-90.

Lenser, T., and Theißen, G. (2013). Molecular mechanisms involved in convergent crop domestication. Trends Plant Sci. **18**: 704-714.

Li, X., et al. (2003). Control of tillering in rice. Nature **422**: 618-21.

Li, C., Zhou, A., and Sang, T. (2006). Rice domestication by reducing shattering. Science **311**: 1936-1939.

Li, J., et al. (2011). Mutation of rice BC12/GDD1, which encodes a kinesin-like protein that binds to a GA biosynthesis gene promoter, leads to dwarfism with impaired cell elongation. Plant Cell **23**: 628-640.

Liang, Q., Lu, X., Jiang, L., Wang, C., Fan, Y., and Zhang, C. (2010). EMB1211 is required for normal embryo development and influences chloroplast biogenesis in *Arabidopsis.* Physiol. Plant **140**: 380-394.

Liew, L.C., Hecht, V., Laurie, R.E., Knowles, C.L., Vander Schoor, J.K., Macknight, R.C., and Weller, J.L. (2009). DIE NEUTRALIS and LATE BLOOMER 1 contribute to regulation of the pea circadian clock. Plant Cell **21**: 3198-3211.

Lin, Z., Griffith, M.E., Li, X., Zhu, Z., Tan, L., Fu, Y., Zhang, W., Wang, X., Xie, D., and Sun, C. (2007). Origin of seed shattering in rice (*Oryza sativa* L.). Planta **226**: 11-20.

Lin, J.Y., Mendu, V., Pogany, J., Qin, J., and Nagy, P.D. (2012a). The TPR domain in the host Cyp40-like cyclophilin binds to the viral replication protein and inhibits the assembly of the tombusviral replicase. PLoS Pathog. **8**: e1002491.

Lin, Z., et al. (2012b). Parallel domestication of the Shattering1 gene in cereals. Nat. Genet. **44**: 720-724.

Liu, X., Wu, J., Clark, G., Lundy, S., Lim, M., Arnold, D., Chan, J., Tang, W., Muday, G.K., Gardner, G., and Roux, S.J. (2012). Role for apyrases in polar auxin transport in *Arabidopsis.* Plant Physiol. **160**: 1985-1995.

Lowman, A.C., and Purugganan, M.D. (1999). Duplication of the *Brassica oleracea* APETALA1floral homeotic gene and the evolution of domesticated cauliflower. J. Hered. **90**: 514-520.

Lukowitz, W., Roeder, A., Parmenter, D., and Somerville, C. (2004). A MAPKK Kinase gene regulates extra-embryonic cell fate in *Arabidopsis.* Cell **116**: 109-119.

Mandel, M.A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene APETALA7. Nature **360**: 273-277.

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A.J., King, G.J., Giovannoni, J.J., and Seymour, G.B. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat. Genet. **38**: 948-952.

Martin, R.C., et al. (2010). The regulation of post-germination transition from the cotyledon- to vegetative-leaf stages by microRNA-targeted SQUAMOSA PROMOTER-BINDING PROTEINLIKE13 in *Arabidopsis.* Seed Sci. Res. **20**: 89-96.

Masclaux-Daubresse, C., Reisdorf-Cren, M., and Orsel, M. (2008). Leaf nitrogen remobilisation for plant development and grain filling. Plant Biol. **10**: 23-36.

Mathur, J. (2004). Cell shape development in plants. Trends Plant Sci. **9**: 583-590.

McGarry, R.C., and Ayre, B.G. (2012). Manipulating plant architecture with members of the CETS gene family. Plant Sci. **188**: 71-81.

Mitsuya, Y., Takahashi, Y., Berberich, T., Miyazaki, A., Matsumura, H., Takahashi, H., Terauchi, R., and Kusano, T. (2009). Spermine signaling plays a significant role in the defense response of *Arabidopsis thaliana* to cucumber mosaic virus. *J. Plant Physiol.* **166**: 626-643.

Miura, K., Rus, A., Sharkhuu, A., Yokoi, S., Karthikeyan, A.S., Raghothama, K.G., Baek, D., Koo, Y.D., Jin, J.B., Bressan, R.A., Yun, D.-J., and Hasegawa, P.M. (2005). The *Arabidopsis* SUMO E3 ligase SIZ1 controls phosphate deficiency responses. Proc. Natl. Acad. Sci. USA **102**: 7760-7765 .

Miura, K., Ikeda, M., Matsubara, A., Song, X.J., Ito, M., Asano, K., Matsuoka, M., Kitano, H., and Ashikari, M. (2010). OsSPL14 promotes panicle branching and higher grain productivity in rice. Nat. Genet. **42**: 545-550.

Miura, K., Lee, J., Gong, Q., Ma, S., Jin, J.B., Yoo, C.Y., Miura, T., Sato, A., Bohnert, H.J., and Hasegawa, P.M. (2011). SIZ1 regulation of phosphate starvation-induced root architecture remodeling involves the control of auxin accumulation. Plant Physiol. **155**: 1000-1012.

Mizoguchi, T., Wright, L., Fujiwara, S., Cremer, F., Lee, K., Onouchi, H., Mouradov, A., Fowler, S., Kamada, H., Putterill, J., and Coupland, G. (2005). Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in *Arabidopsis*. Plant Cell **17**: 2255-2270.

Mol, J., Jenkins, G., Schäfer, E., Weiss, D., and Walbot, V. (1996). Signal perception, transduction, and gene expression involved in anthocyanin biosynthesis. Crit. Rev. Plant Sci. **15**: 525-557.

Molina, I., Ohlrogge, J.B., and Pollard, M. (2008). Deposition and localization of lipid polyester in developing seeds of *Brassica napus* and *Arabidopsis thaliana.* Plant J. **53**: 437-449.

Moreno, M.A., Harper, L.C., Krueger, R.W., Dellaporta, S.L., and Freeling, M. (1997). Ligueleless1 encodes a nuclear-localized protein required for induction of ligules and auricles during maize leaf organogenesis. Genes Dev. **11**: 616-628.

Müller, K., Carstens, A.C., Linkies, A., Torres, M.A., and Leubner-Metzger, G. (2009). The NADPH-oxidase AtrbohB plays a role in *Arabidopsis* seed after-ripening. New Phytol. **184**: 885-897.

Multani, D.S., Briggs, S.P., Chamberlin, M.A., Blakeslee, J.J., Murphy, A.S., and Johal, G.S. (2003). Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. Science **302**: 81-84.

Muños, S., Ranc, N., Botton, E., Bérard, A., Rolland, S., Duffé, P., Carretero, Y., Le Paslier, M.-C., Delalande, C., Bouzayen, M., Brunel, D., and Causse, M. (2011). Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. Plant Physiol. **156**: 2244-2254.

Nakano, T., Fujisawa, M., Shima, Y., and Ito, Y. (2013). Expression profiling of tomato pre-abscission pedicels provides insights into abscission zone properties including competence to respond to abscission signals. BMC Plant Biol. **13**: 40.

Obendorf, R.L. (1997). Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. Seed Sci. Res. **7**: 63-74.

Ogawa, D., et al. (2011). RSS1 regulates the cell cycle and maintains meristematic activity under stress conditions in rice. Nat. Commun. **2**: 278.

Okmen, B. (2011). Quantitative trait loci (QTL) analysis for antioxidant and agronomically important traits in tomato (*Lycopersicon esculentum*). Turk. J. Agric. For. **35**: 501-514.

Olsen, K.M., and Wendel, J.F. (2013). Crop plants as models for understanding plant adaptation and diversification. Front Plant Sci. **4**: 290.

Osakabe, Y., et al. (2013). Osmotic stress responses and plant growth controlled by potassium transporters in *Arabidopsis*. Plant Cell **25**: 609-624.

Paran, I., and van der Knaap, E. (2007). Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. J. Exp. Bot. **58**: 3841-3852.

Parvathaneni, R.K., Jakkula, V., Padi, F.K., Faure, S., Nagarajappa, N., Pontaroli, A.C., Wu,X., Bennetzen, J.L.,and Devos, K.M. (2013). Fine-mapping and identification of a candidate gene underlying the d2 dwarfing phenotype in pearl millet, *Cenchrus americanus* (L.) Morrone. G3 Genes| Genomes| Genetics **3**: 563-572.

Petitot, A.S., Lecouls, A.C., and Fernandez, D. (2008). Sub-genomic origin and regulation patterns of a duplicated WRKY gene in the allotetraploid species *Coffea arabica.* Tree Genet. Genomes **4**: 379-390.

Preston, J.C., and Hileman, L. (2013). Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. Front. Plant Sci. **4**: 80.

Ramanjulu, S., and Bartels, D. (2002). Drought-and desiccation-induced modulation of gene expression in plants. Plant Cell Environ. **25**: 141-151.

Ranc, N., Muños, S., Xu, J., Le Paslier, M.C., Chauveau, A., Bounon, R., Rolland, S., Bouchet, J.-P., Brunel, D., and Causse, M. (2012). Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *cerasiforme.* G3 Genes| Genomes| Genetics **2**: 853-864.

Ranocha, P., Denancé, N., Vanholme, R., Freydier, A., Martinez, Y., Hoffmann, L., Köhler, L., Pouzet, C., Renou, J.-P., Sundberg, B., Boerjan, W., and Goffner, D. (2010). Walls are thin 1 (WAT1), an *Arabidopsis* homolog of *Medicago truncatula* NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. Plant J. **63**: 469-483.

Riechmann, J.L., et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science **290**: 2105-2110.

Rivals, I., Personnaz, L., Taing, L., and Potier, M.C. (2007). Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics **23**: 401-407.

Rodrigues, C.M., de Souza, A.A., Takita, M.A., Kishi, L.T., and Machado, M.A. (2013). RNA-Seq analysis of *Citrus reticulata* in the early stages of *Xylella fastidiosa* infection reveals auxin-related genes as a defense response. BMC Genomics **14**: 676.

Rueda-Romero, P., Barrero-Sicilia, C., Gómez-Cadenas, A., Carbonero, P., and Oñate-Sánchez, L. (2012). *Arabidopsis thaliana* DOF6 negatively affects germination in non-after-ripened seeds and interacts with TCP14. J. Exp. Bot. **63**: 1937-1949.

Ryu, C.H., Lee, S., Cho, L.H., Kim, S.L., Lee, Y.S., Choi, S.C., Jeong, H.J., Yi, J., Park, S.J., Han, C.-D., and An, G. (2009). OsMADS50 and OsMADS56 function antagonistically in regulating long day (LD)-dependent flowering in rice. Plant Cell Environ. **32**: 1412-1427.

Ryu, M.Y., Cho, S.K., and Kim, W.T. (2010). The *Arabidopsis* C3H2C3-type RING E3 ubiquitin ligase AtAIRP1 is a positive regulator of an abscisic acid-dependent response to drought stress. Plant Physiol. **154**: 1983-1997.

Salamini, F. (2003). Plant biology. Hormones and the green revolution. Science **302**: 71-72.

Schweizer, F., Fernández-Calvo, P., Zander, M., Diez-Diaz, M., Fonseca, S., Glauser, G., Lewsey, M.G., Ecker, J.R., Solano, R., and Reymond, P. (2013). *Arabidopsis* basic Helix-Loop-Helix Transcription Factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. Plant Cell **25**: 3117-3132.

Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., and Yano, M. (2008). Deletion in a gene associated with grain size increased yields during rice domestication. Nat. Genet. **40**: 1023-1028.

Siddappaji, M.H., Scholes, D.R., Bohn, M., and Paige, K.N. (2013). Overcompensation in response to herbivory in *Arabidopsis thaliana*: the role of glucose-6-phosphate dehydrogenase and the oxidative pentose-phosphate pathway. Genetics **195**: 589-598.

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., Vorwerk, S., and Youngs, H. (2004). Toward a systems approach to understanding plant cell walls. Science **306**: 2206-2211.

Song, B.K., Hein, I., Druka, A., Waugh, R., Marshall, D., Nadarajah, K., Yap, S.-J., and Ratnam, W. (2009). The 172-kb genomic DNA region of the *O. rufipogon* yld1. 1 locus: comparative sequence analysis with *O. sativa* ssp*. japonica* and *O. sativa* ssp. *indica*. Funct. Integr. Genomics **9**: 97-108.

Song, S., Kim, C.W., Moon, J.S., and Kim, S. (2014). At least nine independent natural mutations of the DFR-A gene are responsible for appearance of yellow onions (*Allium cepa* L.) from red progenitors. ***Mol. Breed.*** **33**: 173-186.

Sorefan, K., Girin, T., Liljegren, S.J., Ljung, K., Robles, P., Galván-Ampudia, C.S., Offringa, R., Friml, J., Yanofsky, M.F., and Østergaard, L. (2009). A regulated auxin minimum is required for seed dispersal in *Arabidopsis.* Nature **459**: 583-586.

Stebbins, G.L. (1992). Comparative aspects of plant morphogenesis: a cellular, molecular, and evolutionary approach. Am. J. Bot. **79**: 589-598.

Stone, J.M., Liang, X., Nekl, E.R., and Stiers, J.J. (2005). *Arabidopsis* AtSPL14, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1. Plant J. **41**: 744-754.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA **102**: 15545-15550.

Sugimoto, K., Takeuchi, Y., Ebana, K., Miyao, A., Hirochika, H., Hara, N., Ishiyamac, K., Kobayashic, M., Band, Y., Hattori, T., and Yano, M. (2010). Molecular cloning of Sdr4, a regulator involved in seed dormancy and domestication of rice. Proc. Natl. Acad. Sci. USA **107**: 5792-5797.

Sung, D.Y., Vierling, E., and Guy, C.L. (2001). Comprehensive expression profile analysis of the *Arabidopsis* Hsp70 gene family. Plant Physiol. **126**: 789-800.

Suo, H., Ma, Q., Ye, K., Yang, C., Tang, Y., Hao, J., Zhang, Z.J., Chen, M., Feng, Y., and Nian, H. (2012). Overexpression of AtDREB1A causes a severe dwarf phenotype by decreasing endogenous gibberellin levels in soybean [*Glycine max* (L.) Merr.]. PloS one **7**: e45568.

Tahir, M., Båga, M., Vandenberg, A., and Chibbar, R.N. (2012). An Assessment of raffinose family oligosaccharides and sucrose concentration in genus *Lens*. Crop Sci. **52**: 1713-1720.

Taji, T., Seki, M., Yamaguchi-Shinozaki, K., Kamada, H., Giraudat, J., and Shinozaki, K. (1999). Mapping of 25 drought-inducible genes, RD and ERD, in *Arabidopsis thaliana.* Plant Cell Physiol**. 40**: 119-123.

Taji, T., Ohsumi, C., Iuchi, S., Seki, M., Kasuga, M., Kobayashi, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2002). Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. Plant J. **29**: 417-426.

Takahashi, Y., and Shimamoto, K. (2011). Heading date 1 (Hd1), an ortholog of *Arabidopsis* CONSTANS, is a possible target of human selection during domestication to diversify flowering times of cultivated rice. Genes Genet. Syst. **86**: 175-182.

Takahashi, T., Mu, J.H., Gasch, A., and Chua, N.H. (1998). Identification by PCR of receptor-like protein kinases from *Arabidopsis* flowers. Plant Mol Biol **37**: 587-596.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A., Rhee, S.Y., and Stitt, M. (2004). MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes Plant J. **37**: 914-939.

Tian, L., Magallanes-Lundback, M., Musetti, V., and DellaPenna, D. (2003). Functional analysis of β-and ε-ring carotenoid hydroxylases in *Arabidopsis*. Plant Cell **15**: 1320-1332.

Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J.E., Nelson, R.L., McCleanc, P.E., Qiu, L., and Ma, J. (2010). Artificial selection for determinate growth habit in soybean. Proc. Natl. Acad. Sci. USA **107**: 8563-8568.

Turner, T.L., Bourne, E.C., Von Wettberg, E.J., Hu, T.T., and Nuzhdin, S.V. (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat. Genet. **42**: 260-263.

Tzafrir, I., Pena-Muralla, R., Dickerman, A., Berg, M., Rogers, R., Hutchens, S., Sweeney, T.C., McElver, J., Aux, G., Patton, D., and Meinke, D. (2004). Identification of genes required for embryo development in *Arabidopsis*. Plant Physiol. **135**: 1206-1220.

Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L., and Doebley, J.F. (2005). The origin of the naked grains of maize. Nature **436**: 714-719.

Wang, J.W., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana.* Cell **138**: 738-749.

Wang, Y., Shen, D., Bo, S., Chen, H., Zheng, J., Zhu, Q.-H., Cai, D., Helliwell, C., and Fan, L. (2010). Sequence variation and selection of small RNAs in domesticated rice. BMC Evol. Biol. **10**: 119.

Wang, L., Hua, D., He, J., Duan, Y., Chen, Z., Hong, X., and Gong, Z. (2011). Auxin Response Factor2 (ARF2) and its regulated homeodomain gene HB33 mediate abscisic acid response in *Arabidopsis*. PLoS Genet. **7**: e1002172.

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. **126**: 485-93.

Winter, C.M., et al. (2011). LEAFY target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. Dev. Cell **20**: 430-443.

Wolf, C., Hennig, M., Romanovicz, D., and Steinebrunner, I. (2007). Developmental defects and seedling lethality in apyrase AtAPY1 and AtAPY2 double knockout mutants. Plant Mol. Biol. **64**: 657-672.

Worrall, D., Liang, Y.K., Alvarez, S., Holroyd, G.H., Spiegel, S., Panagopulos, M., Gray, J.E., and Hetherington, A. M. (2008). Involvement of sphingosine kinase in plant cell signalling. Plant J. **56**: 64-72.

Wu, G., and Poethig, R.S. (2006). Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. Development **133**: 3539-3547.

Wu, J., Steinebrunner, I., Sun, Y., Butterfield, T., Torres, J., Arnold, D., Gonzalez, A., Jacob, F., Reichler, S., and Roux, S.J. (2007). Apyrases (nucleoside triphosphate-diphosphohydrolases) play a key role in growth control in *Arabidopsis*. Plant Physiol. **144**: 961-975.

Wu, Y., et al. (2012). Presence of tannins in sorghum grains is conditioned by different natural alleles of Tannin1. Proc. Natl. Acad. Sci. USA **109**: 10281-10286.

Wu, W., et al. (2013). Association of functional nucleotide polymorphisms at DTH2 with the northward expansion of rice cultivation in Asia. Proc. Natl. Acad. Sci. USA **110**: 2775-2780.

Xia, Z., et al. (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. Proc. Natl. Acad. Sci. USA **109**: E2155-E2164.

Xu, G., Li, S., Xie, K., Zhang, Q., Wang, Y., Tang, Y., Liu, D., Hong, Y., He, C., and Liu, Y. (2012). Plant ERD2-like proteins function as endoplasmic reticulum luminal protein receptors and participate in programmed cell death during innate immunity. Plant J. **72:** 57-69.

Yamasaki, H., Hayashi, M., Fukazawa, M., Kobayashi, Y., Shikanai, T. (2009). SQUAMOSA promoter binding protein-like7 is a central regulator for copper homeostasis in *Arabidopsis.* Plant Cell **21:** 347-361.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., and Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene VRN1. Proc. Natl. Acad. Sci. USA **100:** 6263-6268.

Yonekura, M., Ohto, C., Muramoto, N., Mitsukawa, N., Takagi, M., and Matsui, K. (2010). U.S. Patent Application 13/376,169.

Yun, J., Kim, Y.S., Jung, J.H., Seo, P.J., and Park, C.M. (2012). The AT-hook motif-containing protein AHL22 regulates flowering initiation by modifying FLOWERING LOCUS T chromatin in *Arabidopsis*. J. Biol. Chem. **287:** 15307-15316.

Zhang, S.J., Song, X.Q., Yu, B.S., Zhang, B.C., Sun, C.Q., Knox, J.P., and Zhou, Y.H. (2012). Identification of quantitative trait loci affecting hemicellulose characteristics based on cell wall composition in a wild and cultivated rice species. Mol. Plant. **5:** 162-175.

Zhang, D., Cheng, H., Hu, Z., Wang, H., Kan, G., Liu, C., and Yu, D. (2013). Fine mapping of a major flowering time QTL on soybean chromosome 6 combining linkage and association analysis. Euphytica **191:** 23-33.

Zhou, X.F., Jin, Y.H., Yoo, C.Y., Lin, X.L., Kim, W.Y., Yun, D.J., Bressan, R.A., Hasegawa, P.M., and Jin, J.B. (2013). CYCLIN H; 1 regulates drought stress responses and blue light-induced stomatal opening by inhibiting reactive oxygen species accumulation in *Arabidopsis*. Plant Physiol. **162:** 1030-1041.