

ORIGINAL ARTICLE

Proteomic divergence in *Arabidopsis* autopolyploids and allopolyploids and their progenitors

DW-K Ng¹, C Zhang¹, M Miller¹, Z Shen², SP Briggs² and ZJ Chen¹

Autopolyploidy and allopolyploidy are common in many plants and some animals. Rapid changes in genomic composition and gene expression have been observed in both autopolyploids and allopolyploids, but the effects of polyploidy on proteomic divergence are poorly understood. Here, we report quantitative analysis of protein changes in leaves of *Arabidopsis* autopolyploids and allotetraploids and their progenitors using isobaric tags for relative and absolute quantitation (iTRAQ) coupled with mass spectrometry. In more than 1000 proteins analyzed, the levels of protein divergence were relatively high (~18%) between *Arabidopsis thaliana* and *Arabidopsis arenosa*, relatively low (~6.8%) between an *A. thaliana* diploid and autotetraploid and intermediate (~8.3 and 8.2%) in F₁- and F₈-resynthesized allotetraploids relative to mid-parent values, respectively. This pattern of proteomic divergence was consistent with the previously reported gene expression data. In particular, many non-additively accumulated proteins (61–62%) in the F₁ and F₈ allotetraploids were also differentially expressed between the parents. The differentially accumulated proteins in functional categories of abiotic and biotic stresses were overrepresented between an *A. thaliana* autotetraploid and diploid and between two *Arabidopsis* species, but not significantly different between allotetraploids and their progenitors. Although the trend of changes is similar, the percentage of differentially accumulated proteins that matched previously reported differentially expressed genes was relatively low. Western blot analysis confirmed several selected proteins with isoforms the cumulative levels of which were differentially expressed. These data suggest high protein divergence between species and rapid changes in post-transcriptional regulation and translational modifications of proteins during polyploidization.

Heredity (2012) **108**, 419–430; doi:10.1038/hdy.2011.92; published online 19 October 2011

Keywords: polyploidy; proteome; iTRAQ; gene expression; hybrids; *Arabidopsis*

INTRODUCTION

Polyploidy has been recognized as a powerful mechanism for evolutionary innovation in many plants and some animals (Wendel, 2000; Doyle *et al.*, 2008; Leitch and Leitch, 2008; Soltis and Soltis, 2009; Wood *et al.*, 2009; Jackson and Chen, 2010). The common occurrence of plant polyploids (including many crops such as wheat, cotton and canola) suggests an evolutionary advantage of having raw genetic materials for selection and adaptation. The molecular basis for this is poorly understood and is probably related to genomic and gene expression changes, which have been demonstrated in many resynthesized and natural polyploid plants, including *Arabidopsis* (Wang *et al.*, 2004, 2006b), *Brassica* (Gaeta *et al.*, 2009), *Gossypium* (Flagel *et al.*, 2008; Hovav *et al.*, 2008), *Senecio* (Hegarty *et al.*, 2006), *Spartina* (Salmon *et al.*, 2005) and *Tragopogon* (Tate *et al.*, 2006; Buggs *et al.*, 2009). These studies have documented several common features of gene expression changes, including non-additive expression, tissue-specific expression of homeologous loci and parental-dependent repression or activation (Jackson and Chen, 2010). In *Arabidopsis*, the natural allotetraploid *Arabidopsis suecica* was formed between extant progenitors *Arabidopsis thaliana* and *Arabidopsis arenosa* (O’Kane *et al.*, 1995). A previous microarray study revealed a 15% transcriptome difference between *A. thaliana* and *A. arenosa* autotetraploids (Wang *et al.*, 2006b). Genome hybridization between these

two autotetraploids resulted in 5–6% transcriptome changes between the resynthesized allotetraploids and their progenitors. In contrast, a very low percentage of transcriptome divergence was observed between an *A. thaliana* diploid and an isogenic autotetraploid line.

Compared with genomic and gene expression changes, changes in proteins and gene products in polyploids and their progenitors are rarely examined. Using protein gel analysis, Birchler and Newton (1981) reported positive and negative effects on protein patterns in maize autopolyploidy series and aneuploids, with greater effects caused by aneuploidy than by ploidy variation (Birchler and Newton, 1981). Protein two-dimensional electrophoresis (2-DE) analysis in diploid, tetraploid and hexaploid wheat showed that the expression of homeologous proteins in hexaploid wheat depended on interactions among the parental A, B and D genomes (Islam *et al.*, 2003). Other recent studies using 2-DE indicated numerous and non-stochastic modifications of proteins in resynthesized *Brassica napus* (Albertin *et al.*, 2006, 2009) and wheat hybrids (Song *et al.*, 2007). Using protein 2-DE coupled with mass spectrometry (MS) assays, a recent study showed a positive correlation of differentially expressed proteins with ploidy levels in a maize ploidy series (diploid, tetraploid and hexaploid) (Yao *et al.*, 2011). However, the highest correlations were found in diploid–hexaploid and tetraploid–hexaploid comparisons rather than in the diploid–tetraploid comparison. Notably, differential protein

¹Section of Molecular Cell and Developmental Biology, Center for Computational Biology and Bioinformatics, and Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX, USA and ²Division of Biology, University of California, San Diego, La Jolla, CA, USA

Correspondence: Professor ZJ Chen, Institute for Cellular and Molecular Biology, The University of Texas at Austin, One University Station, A-4800, Austin, TX 78712, USA.

E-mail: zjchen@mail.utexas.edu

Received 9 May 2011; revised 24 August 2011; accepted 31 August 2011; published online 19 October 2011

regulation did not necessarily correlate with changes in gene expression (Marmagne *et al.*, 2010).

The traditional protein 2-DE has a limited power of detecting the number of protein changes. Isobaric tags for relative and absolute quantitation (iTRAQ) is a suitable method for the study of proteomic changes in polyploids because it is not only capable of simultaneous examination of many more proteins but it also has a high sensitivity for measuring protein abundance in related species when coupled with MS analysis (Ross *et al.*, 2004). iTRAQ-based proteome analysis has been used to identify protein-level differences as small as 50%, and multiple reaction monitoring has been used to confirm these measurements (O'Brien *et al.*, 2010).

Using iTRAQ-based proteomic analysis, in this study, we investigated the effects of genome duplication, divergence and hybridization on proteomic changes in an *A. thaliana* diploid and an isogenic autotetraploid, between *A. thaliana* and *A. arenosa* autotetraploids, and in *A. suecica*-like allotetraploids (F₁ and F₈ generations) relative to the progenitors. These data provide a general survey of protein regulation in response to changes in genome dosage (diploid vs autotetraploid), genome divergence (*A. thaliana* vs *A. arenosa*), genome hybridization (allotetraploids relative to the parents) and genetic stability (F₁ vs F₈ allotetraploids). In addition, the proteomic data were comparatively analyzed with previous microarray data, and gene ontology (GO) analysis was performed to predict the biological consequences of proteomic changes in these comparisons.

MATERIALS AND METHODS

Plant materials and growth conditions

Plant materials included an *A. thaliana* Landsberg *erecta* (*Ler*) diploid (At2), an induced isogenic *Ler* autotetraploid (At4; ABRC, CS3900), one accession of outcrossing tetraploid *A. arenosa* (Aa; ABRC, CS3901) (Comai *et al.*, 2000), multiple F₁-resynthesized *A. suecica*-like allotetraploids (AlloF₁) and one accession (Allo733) of F₈ allotetraploids (Allo733F₈) (Wang *et al.*, 2006b). The *A. thaliana* diploid and autotetraploid lines used in this study were the same as those used in previous experiments (Wang *et al.*, 2004, 2006b). *A. arenosa* is an outcrossing species and is maintained by manual cross-pollination within a small population. Both *A. arenosa* diploids and tetraploids and other hybrid species exist in nature (Koch *et al.*, 2008; Schmickl and Koch, 2011). In the laboratory, allotetraploids could only be resynthesized by pollinating *A. thaliana* autotetraploids with pollen from *A. arenosa* autotetraploids (Comai *et al.*, 2000). Multiple AlloF₁ lines were generated independently between At4 and Aa. These lines were self-pollinated to generate stable allotetraploids Allo733 (F₆), Allo733F₇ and Allo733F₈.

Seeds were germinated in Murashige and Skoog agar media with 3% sucrose (Sigma-Aldrich, St Louis, MO, USA), stratified for 2 days at 4 °C in the dark and grown at 22 °C with a 16/8-h (light/dark) cycle in a growth chamber. Two-week-old seedlings were transferred to fresh Murashige and Skoog agar media and grown under the same condition for 6 days before transferring to Pro-mix BX potting soil (Premier Tech Horticulture, Quakertown, PA, USA). For each genotype, mature leaves from a pool of ≥ 30 plants of At2 and At4 and 7–8 plants of Aa, F₁ and F₈ allotetraploids were harvested as one biological replicate. To minimize developmental variation, mature leaves before bolting (from two biological replications) were harvested for protein preparation. This stage was ~ 4 weeks (including the time for stratification) in *A. thaliana* diploids and tetraploids (*Ler*) under our growth conditions, which is comparable to the growth stage 3.50 in *A. thaliana* (Col-0), as described previously (Boyes *et al.*, 2001). A similar stage in *A. arenosa* and the allotetraploids was ~ 7 weeks (Wang *et al.*, 2004, 2006b).

Fluorescence *in situ* hybridization

To validate the ploidy levels of various plant materials used in this study, fluorescence *in situ* hybridization was performed according to published protocols (Kato *et al.*, 2004; Lysak *et al.*, 2006a). Flower buds (At2, At4, Aa, Allo733F₈) were fixed in ice-cold Carnoy's fixative for at least 1.5 h. Selected flower buds (< 1 mm) were digested in a pectolytic enzyme mixture (0.3% cellulase, 0.3% pectolyase and 0.3% cytohellicase) for 2 h. Digested cells were suspended in 9:1 glacial acetic acid:methanol and mounted on glass slides. Chromosomal DNA was cross-linked to slides by exposure to ultraviolet light (120 mJ cm⁻²).

For probe preparation, centromere repetitive sequences of At (AtCEN) and Aa (AaCEN) were amplified by PCR using genomic DNA and the primer pairs as described previously (Comai *et al.*, 2003). PCR products were purified using QIAquick PCR purification kit (Qiagen, Valencia, CA, USA). In all, 1 μ g of purified AtCEN and AaCEN PCR products were labeled with fluorescein-12-dUTP (Fermentas, Glen Burnie, MD, USA) and Texas red-5-dCTP (Perkin-Elmer, Waltham, MA, USA), respectively, using the nick-translation method. Labeled probes were purified using MinElute PCR purification kit (Qiagen).

Hybridization of chromosome mounts was performed at 55 °C for 4–6 h in the presence of labeled probes. Vectashield mounting medium with DAPI (Vector Laboratories Inc., Burlingame, CA, USA) was applied to hybridized slides before observation. Images were taken using a Zeiss Axiovert 200M microscope equipped with a Zeiss Axiocam fluorescence camera (Carl Zeiss Microscopy, LLC, Thornwood, NY, USA).

Flow cytometry

Endoreduplication is common in many plants, especially in *A. thaliana* (Galbraith *et al.*, 1983), and the difference in endopolyploidy in leaves of different *Arabidopsis* lines may contribute to protein expression changes. To determine whether endopolyploidy contributes to proteomic changes among different *Arabidopsis* lines, flow-cytometric analysis was used to evaluate relative levels of endoreduplication in mature leaves of At2, At4, Aa and Allo733F₈. Mature leaf tissues (~ 150 mg; 4–5 leaves) were collected and placed in a petri dish on ice. Suspension of nuclei was prepared by chopping the plant tissues in a 1-ml buffer (pH 7) containing 45 mM MgCl₂, 30 mM sodium citrate, 20 mM 4-morpholine propane sulfonate and 0.1% Triton X-100 using a razor blade (Galbraith *et al.*, 1983). Finely chopped leaf tissues were then filtered through a 30- μ m Partec CellTrics filter (Partec North America, Swedesboro, NJ, USA) into a 12 \times 75 mm² polystyrene culture tube. Nucleus suspension was stained by adding 5 mg ml⁻¹ propidium iodide (Sigma-Aldrich) to a final concentration of 100 μ g ml⁻¹. The nuclei were stained for 40–60 min before flow cytometry using a BD Biosciences FACSCalibur flow cytometer (BD Biosciences, San Jose, CA, USA). Flow-cytometry data were collected using CellQuestPro software (BD Biosciences). Samples were run under low pressure using the FL2 detector set at 480 nm. A total of ~ 80 000 events (nuclei) were counted for each sample. Three biological replicates were performed for each line.

Data analysis was performed using the CellQuestPro software, and a scatter plot was generated from the collected data. Populations of nuclei were then identified and gated. Relative proportion of nuclei (from 0, 1 and 2 endoreduplication events) within these gated peak regions (FL2-A) were measured and calculated. To determine the degree of endoreduplication in these lines, the average number of endocycles per 100 cells was estimated using the endoreduplication factor (EF) (Cookson *et al.*, 2006).

Protein extraction and digestion

For proteomic comparison among the five *Arabidopsis* lines (At2, At4, Aa, AlloF₁, and Allo733F₈), total proteins extracted from mature rosette leaves were used to test the effects of genome duplication (At2 vs At4), genotype difference (At4 vs Aa) and genome hybridization (F₁ or F₈ vs progenitors) on proteome changes. To prepare peptides for iTRAQ reagent labeling, mature leaves (3 g) of multiple plants from each plant genotype (see above) were ground in liquid nitrogen and washed three times with 50 ml -20°C methanol containing 0.2 mM Na₃VO₄ in a 50-ml tube, followed by three washes and precipitation with 50 ml -20°C acetone. Protein pellets were dried in a SpeedVac at 4°C and resuspended in 1 ml solution containing 0.2 mM Na₃VO₄ and 0.2% RapiGest (Waters Corporation, Milford, MA, USA). Supernatants for protein digestion were collected by centrifugation (16 100 g) of samples at 4°C for 15 min. Cysteines were reduced and alkylated using 1 mM Tris (2-carboxyethyl) phosphine (Fisher Scientific, Waltham, MA, USA) at 95°C for 5 min and then 2.5 mM iodoacetamide (Fisher Scientific) at 37°C in the dark for 15 min. Proteins (100 μg) were digested with trypsin (Roche, Indianapolis, IN, USA) with an enzyme to substrate ratio (w:w) of 1:50 at 37°C overnight.

iTRAQ labeling

A multiplex iTRAQ shotgun proteomic approach was used to perform quantitative analysis of four samples in a single MS experiment (Ross *et al.*, 2004). The effects of genome duplication on proteomic changes were compared between At2 and At4. In another multiplexed iTRAQ analysis, proteomes of At4, Aa and two allotetraploids (AlloF₁ and Allo733F₈) were compared. The genotypic or species effects on proteomes were analyzed between an induced *A. thaliana* autotetraploid (At4) and natural *A. arenosa* tetraploids (Aa). For simplicity, we use the terms 'genotype' and 'species' interchangeably between these comparisons. Finally, genome hybridization effects were compared between either AlloF₁ or Allo733F₈ and their progenitors (At4 and Aa). In a single MS experiment, proteomes from four samples were analyzed simultaneously using four iTRAQ reagents (AB SCIEX, Foster City, CA, USA) containing different reporter groups of mass to charge ratio (m/z) of 114, 115, 116 or 117. In comparing diploid and autotetraploid *A. thaliana*, iTRAQ 114 and 115 reagents were used to label tryptic-digested peptides from At2, and iTRAQ 116 and 117 reagents were used for At4 tryptic-digested peptides labeling. In another multiplexed iTRAQ analysis, tryptic-digested peptides from At4, Aa, AlloF₁ and Allo733F₈ allotetraploids were labeled with iTRAQ reagents with m/z of 114, 115, 116 and 117, respectively. Each peptide solution was reacted with one-half tube of iTRAQ reagent in 90% isopropyl alcohol and 250 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) at a pH of 7.2 for 15 min at room temperature. Water was then added to a final isopropyl alcohol concentration of 60%. The second half tube of the iTRAQ reagent was added and samples were incubated at room temperature for 2 h. Peptides were dried in a vacuum concentrator and resuspended in water (100 μl). Samples tagged with different iTRAQ reagents were pooled together. Thereafter, 1% trifluoroacetic acid (final pH < 2) was added to precipitate RapiGest. Samples were incubated at 4°C overnight and then centrifuged at 16 100 g for 15 min. The supernatants were collected and centrifuged through a 0.22- μm filter for subsequent liquid chromatography (LC)-MS/MS analysis.

Three-phase online nano-LC

The iTRAQ-labeled peptides were separated into fractions using nano-LC for subsequent MS analysis. Nano-LC was performed with

an Agilent 1100 HPLC system (Agilent Technologies, Santa Clara, CA, USA) using the exact conditions and columns as described previously (O'Brien *et al.*, 2010). A total of 28 salt fractions were collected for each full proteome analysis.

Tandem MS

The collected peptide fractions from nano-LC were subjected to tandem MS for peptide identification and quantitation. A LTQ linear ion trap tandem mass spectrometer (Thermo Electron Corporation, San Jose, CA, USA) was used to acquire spectra under automated, data-dependent acquisition. MS was performed according to the conditions as described previously (O'Brien *et al.*, 2010). iTRAQ reagents were originally not usable in ion trap instruments because of the 'one-third rule'. To illustrate, fragment ions of a m/z 900 parent will not be detected below m/z 300, and this would normally prevent the detection of iTRAQ reporter ions. To overcome this limit, a new collision-activated fragmentation technique, called pulsed-Q dissociation (PQD), was used to enable routine and reliable measurement of ions down to 50 m/z (Schwartz *et al.*, 2005; Griffin *et al.*, 2007). For peptide identification and quantitation, collision-induced dissociation and PQD scans were collected for the same parent ion (Andreyev *et al.*, 2010; O'Brien *et al.*, 2010). For each MS scan, four pairs of collision-induced dissociation-PQD MS/MS scans were performed for the most intense ions from the parent MS scan. Dynamic exclusion was set at 1 min to improve the duty cycle of MS/MS scans. Approximately 20 000 MS/MS spectra were generated and collected for each salt step fractionation. We have developed and successfully applied a combined collision-induced dissociation/PQD scan approach for peptide identification and quantitation (Andreyev *et al.*, 2010; O'Brien *et al.*, 2010).

Protein database search

To infer protein identities from the identified peptides from MS/MS spectra, the raw data were extracted and searched using Spectrum Mill v3.03 (Agilent Technologies). Both collision-induced dissociation and PQD scans from the same parent ion were merged together. MS/MS spectra with a sequence tag length > 1 were searched against a concatenated forward-decoy database (71 238 protein sequences). The concatenated forward-decoy database was constructed by combining the International Protein Index *Arabidopsis* database (v3.29, 35 619 protein sequences) and a 1:1 decoy database. The decoy database was constructed by first reversing the International Protein Index protein sequences. The tryptic peptides (peptide length > 6) shared with the forward International Protein Index database were then shuffled to ensure that there were no shared tryptic peptides between the forward and decoy databases. Fully tryptic peptides with a maximum miscleavage of one setting were used as the enzyme parameter for the database search. All other search parameters were set to default settings of Spectrum Mill (carbamidomethylation of cysteines, iTRAQ modification, ± 2.5 kDa for precursor ions, ± 0.7 kDa for fragment ions and a minimum matched percentage score peak intensity of 50%). The *in situ* false discovery rate was derived with the constructed concatenated forward-decoy database. To maintain the false discovery rate of protein levels at < 1%, a MS/MS spectrum > 50% score peak intensity and cutoff scores > 15.7, > 13.7 and > 15.0 were dynamically assigned to the data set with 1+, 2+ and 3+ peptides, respectively. Proteins with two or more unique peptides were validated and used for quantitative analysis. To address the issue of database redundancy resulting from proteins with high sequence homology or alternative splice variants, common peptides shared between proteins were grouped such that proteins within the same

group shared the same set or subset of unique peptides. However, if a protein with shared peptides also contained other unique peptides, it was assigned to a different protein group. Within a protein group, the protein with the highest total identification score of all peptides was chosen as the group leader for proteomic comparisons among all *Arabidopsis* lines used in this study.

Protein quantification by iTRAQ mass tagging reagent

For quantitative analysis of proteins, shared peptides between protein groups were removed, and unique peptides from a protein group were used for iTRAQ quantitation. Relative protein quantitation between different *Arabidopsis* lines was performed by calculating the iTRAQ reporter ion intensity ratios of respective lines. Protein iTRAQ intensities were derived by summing the peptide iTRAQ intensities from each protein group. Low-intensity spectra with total iTRAQ reporter intensity (sum of all 4 channels) < 100 were filtered. Isotope impurities of iTRAQ reagents were corrected according to the manufacturer's instructions (AB SCIEX). For protein iTRAQ reporter intensities normalization, log ratios between different iTRAQ tags (115/114, 116/114, 117/114) were adjusted globally to attain a median log ratio of zero. The normalized protein iTRAQ intensities were used for quantitative analysis.

Quantitative proteome comparison between different *Arabidopsis* lines

Quantitative protein differences were estimated between autotetraploid and diploid *A. thaliana* (At4/At2), *A. arenosa* and *A. thaliana* (Aa/At4) autotetraploids, the allotetraploids AlloF₁ or Allo733F₈ and the mid-parent value (MPV; ½ At4 + ½ Aa) (F₁/MPV or F₈/MPV). For each comparison, the normalized protein iTRAQ intensities from two biological replicates were used. Only proteins detected in both MS experiments were used to calculate the protein fold change. Student's *t*-tests were performed to determine protein expression differences between each comparison in two replications ($P \leq 0.05$). *P*-value corrections for multiple testing were performed using the SGoF software (Carvajal-Rodriguez *et al.*, 2009). In addition to Student's *t*-test *P*-values, we adjusted *P*-values from various multiple testing methods, including SGoF (Carvajal-Rodriguez *et al.*, 2009), Benjamini and Hochberg (Benjamini and Hochberg, 1995), SGoF plus (de Unal-Alvarez and Carvajal-Rodriguez, 2010), SFisher (Diz *et al.*, 2011) and Bonferroni (Holm, 1979). In addition, a complete list of protein quantification data is provided in Supplementary Data Set 1 (<http://datadryad.org>), in which independent *P*-value adjustments could be performed. The multiple testing results are provided in Supplementary Tables S2–S5. In Supplementary Table S6, we also provide a comparison of the numbers of differentially expressed proteins under each of the correction methods used. All proteins were sorted according to their significance levels. For each comparison (At4/At2, Aa/At4, F₁/MPV or F₈/MPV), preference in upregulation or downregulation of proteins was evaluated using Wilcoxon's signed-rank tests.

Although *P*-value corrections for multiple comparisons would decrease the probability of false positives (type I errors), the probability of false negatives (type II errors) would increase depending on which multiple hypothesis testing methods were used (Diz *et al.*, 2011). In this study, many proteins including two that were validated by western blots were excluded in several corrections for multiple comparisons. In some tests, the number of significant proteins was zero. As the major goal of this study is to survey the overall proteomic changes among different *Arabidopsis* lines, proteins in comparison with the uncorrected *P*-value (≤ 0.05) were used for further analysis. Although this would potentially increase false positives, such

occurrence could be corrected by future experimental validation for the proteins of interest. Moreover, the proteomic data in this study were comparatively analyzed using the previously reported transcriptome (microarray) data (Wang *et al.*, 2006b).

GO analysis

To determine whether the differentially accumulated proteins among different *Arabidopsis* lines were associated with particular biological functions, GO analysis was performed using the online database at The Arabidopsis Information Resource website (Berardini *et al.*, 2004), and proteins were grouped into 14 biological process categories. GO enrichment of all proteins detected from the two iTRAQ experiments were compared against the whole *Arabidopsis* genome categorization in The Arabidopsis Information Resource website. χ^2 Tests were used to determine the significance of GO enrichment for proteins detected in the iTRAQ experiments. GO enrichments of differentially expressed proteins resulting from genome duplication (At4 vs At2), genotype difference (Aa vs At4) and genome hybridization (AlloF₁ or Allo733F₈ vs MPV) were also compared and the significance of enrichment was determined using χ^2 tests.

Extraction of nuclei and thylakoid proteins

To validate histone protein accumulation in various *Arabidopsis* lines, crude nuclei were prepared from the same plant materials for protein extraction and western blots. Rosette leaves from multiple plants of a single line (400 mg) were ground in the presence of liquid nitrogen, and the frozen powder was transferred to a test tube containing 1 ml of nuclei isolation buffer with 0.25 M sucrose, 15 mM Piperazine-*N,N'*-bis (2-ethane-sulfonic acid) (pH 6.8), 5 mM MgCl₂, 60 mM KCl, 15 mM NaCl, 1 mM CaCl₂, 0.9% Triton X-100, 2 $\mu\text{g ml}^{-1}$ Pepstatin A, 1 mM PMSF (phenylmethanesulfonyl fluoride) and 1 \times protease inhibitor cocktail (Sigma-Aldrich) (Saleh *et al.*, 2008). The samples were incubated on ice for 15–30 min, filtered through one layer of Miracloth and subjected to centrifugation at 11 000 g for 20 min at 4 °C. The pellet was resuspended in 100 μl of cold nucleus lysis buffer containing 50 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM PMSF, 1% SDS, 1% Triton X-100, 1 $\mu\text{g ml}^{-1}$ pepstatin A and 1 \times protease inhibitor cocktail (Sigma-Aldrich).

To validate an *Arabidopsis* leaf-type ferredoxin-NADP oxidoreductase (AtLFNR1; At5g66190) accumulation in different *Arabidopsis* lines, thylakoid proteins from rosette leaves (400 mg) were prepared according to a published protocol (Lintala *et al.*, 2007). In brief, frozen tissues were ground and incubated for 30 min in 500 μl of grinding buffer containing 10 mM HEPES (pH 7.6), 5 mM sucrose, 5 mM MgCl₂, 2 mM PMSF and 1 \times protease inhibitor cocktail (Sigma-Aldrich). The samples were filtered through one layer of Miracloth and subjected to centrifugation (2500 g) at 4 °C for 4 min. The pellet (thylakoids) was resuspended in 100 μl of storage buffer containing 10 mM HEPES (pH 7.5), 100 mM sucrose, 5 mM NaCl, 10 mM MgCl₂, 2 mM PMSF and 1 \times protease inhibitor cocktail (Sigma-Aldrich).

SDS-PAGE and western blot analysis

We validated two proteins of the proteomic data using western blot analysis because protein antibodies are very limited. For western analysis, an aliquot (20 μg) of nuclei or thylakoid protein was subjected to 15% SDS-PAGE (6 M urea) and transferred onto a polyvinylidene fluoride membrane (GE Healthcare, Piscataway, NJ, USA). An aliquot (1 μg) of core histones (Millipore, Billerica, MA, USA) was included as a positive control in the western blot analysis of histone H3. For western blotting, the membrane was incubated at room temperature for 1.5 h in a protein-free blocking

buffer (Thermo Scientific, Waltham, MA, USA) containing 0.05% Tween-20. The primary antibodies, anti-H3Cter (1:1000) (Abcam, Cambridge, MA, USA) and anti-AtLFNR1 (1:3000) (Abcam), were used to probe the immunoblots containing nuclei and thylakoid protein extracts, respectively. The goat anti-rabbit IgG, horseradish peroxidase conjugate (1:10 000) (Millipore) was used as a secondary antibody for immunodetection, which was performed using a Super-Signal West Pico kit (Thermo Scientific) and quantified using ImageJ software (National Institutes of Health, Bethesda, MD, USA).

RESULTS

Chromosomal and endopolyploidy analyses in *A. thaliana* diploid (At2), tetraploid (At4), *A. arenosa* autotetraploid (Aa) and allotetraploid (Allo733F₈)

The iTRAQ-based proteomic analysis was performed using rosette leaves (from multiple plants) of various *Arabidopsis* lines: (1) *A. thaliana* (*Ler* ecotype) diploid (At2); (2) an isogenic *A. thaliana* autotetraploid (At4); (3) natural *A. arenosa* autotetraploids (Aa); (4) newly resynthesized *A. suecica*-like F₁ (AlloF₁) lines and (5) a stable *A. suecica*-like allotetraploid (Allo733F₈) (Figures 1a–d). Fluorescence *in situ* hybridization confirmed the ploidy levels of these lines with 10, 20 and 32 chromosomes in At2, At4 and Aa, respectively (Figures 1e–g). Allo733F₈ contained 26 chromosomes, including 10 and 16 from *A. thaliana* and *A. arenosa*, respectively (Figure 1h) (Comai *et al.*, 2000; Wang *et al.*, 2006b).

To address possible complicating effects of endopolyploidy (Galbraith *et al.*, 1983) on interpreting patterns of proteomic divergence, flow-cytometric analysis was used to evaluate relative levels of endoreduplication in At2, At4, Aa and Allo733F₈ (Figures 2a–d). The distribution of 2C, 4C and 8C nuclei in the diploid (At2) was shifted proportionally to the corresponding 4C, 8C and 16C nuclei in tetraploids (At4) (Figure 2e; Supplementary Table S1). At4 showed

an overall higher EF (more cycles of endoreduplication) than did At2 in leaves (Figure 2f; $P=0.01$), indicating that endopolyploidy levels increase proportionally in response to ploidy changes. Among the various *Arabidopsis* tetraploids, At4, Aa and Allo733F₈ had different endopolyploidy distribution patterns (Figure 2e). Overall, Aa showed a lower EF than did At4 ($P=0.001$) or Allo733F₈ (F8) ($P=0.011$) (Figure 2f), and no significant difference was observed between At4 and Allo733F₈ allotetraploids ($P=0.528$). These data indicate that endopolyploidy levels in leaves were low in *A. arenosa* but similar between an *A. thaliana* autotetraploid and F₈ allotetraploid and increased proportionally between an *A. thaliana* diploid and autotetraploid.

Proteomic divergence in *Arabidopsis*-related species and allotetraploids

In two repeated iTRAQ LC-MS/MS analyses, a total of 80 753 MS/MS spectra were validated. Among them, 131 spectra were from the decoy database, resulting in a false discovery rate of 0.16% at the spectrum level. In addition, 8945 unique peptides were identified with a false discovery rate of 0.65% at the peptide level (58 peptides from the decoy database). A total of 2083 common proteins were identified by two replications of iTRAQ and MS assays between At2 and At4. After analyzing shared peptides, these proteins were classified into 1400 groups. Among them, 95 proteins (~6.8%) displayed expression differences ($P\leq 0.05$) between At2 and At4 (Figures 3a and c). Between various *Arabidopsis* tetraploids (At4, Aa, AlloF₁ and Allo733F₈), a total of 1160 protein groups were identified from 1730 common proteins that were detected in the 2 replicated iTRAQ and MS assays. The proteomic divergence between At4 and Aa was relatively high, and 209 of the 1160 proteins (~18%) present in both replicates showed expression divergence between *A. thaliana* and *A. arenosa* (Figures 3b and c). In the allotetraploids that were derived

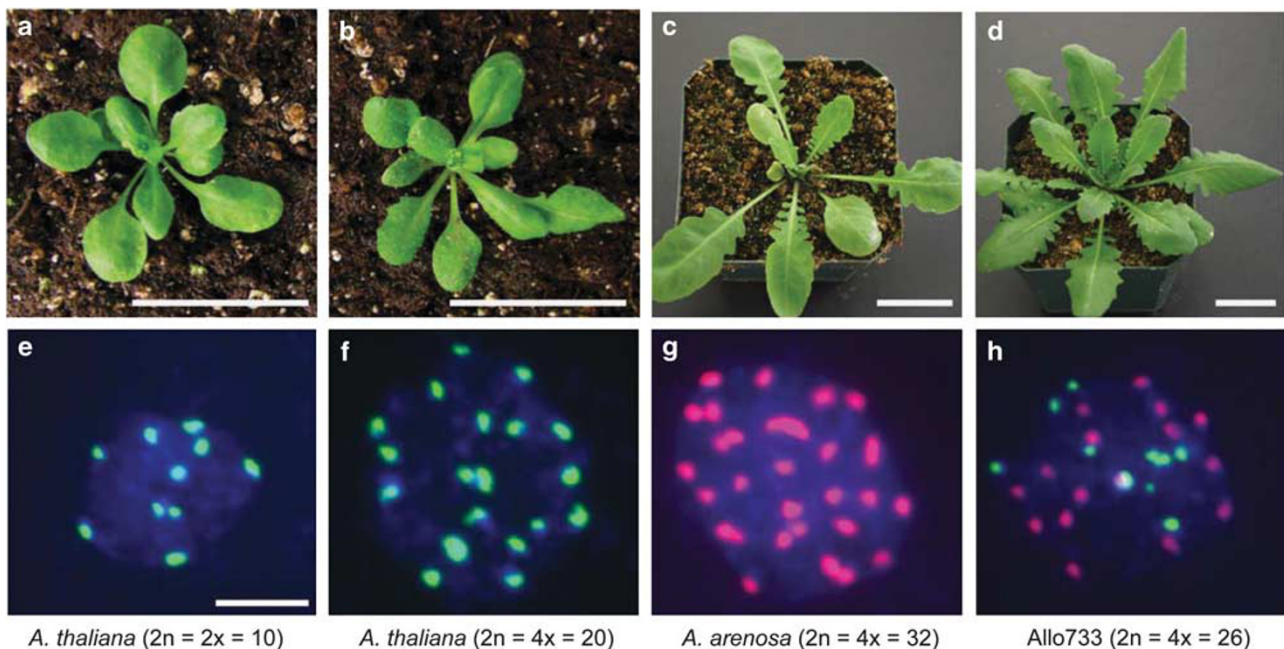


Figure 1 Plant materials and fluorescence *in situ* hybridization (FISH). (a, e) *Arabidopsis thaliana* (*Ler*) diploid (At2; $2n=2x=10$). (b, f) *A. thaliana* autotetraploid (At4; $2n=4x=20$). (c, g) *Arabidopsis arenosa* autotetraploid (Aa; $2n=4x=32$). (d, h) Resynthesized allotetraploid (Allo733F₈; $2n=4x=26$) derived from *A. thaliana* and *A. arenosa*. Fluorescein-labeled (green) and Texas red-labeled (Red) probes were used to detect At and Aa centromeric repeats, respectively. Bars=5 μ m (panels a–d) and 4 cm (panels e–h).

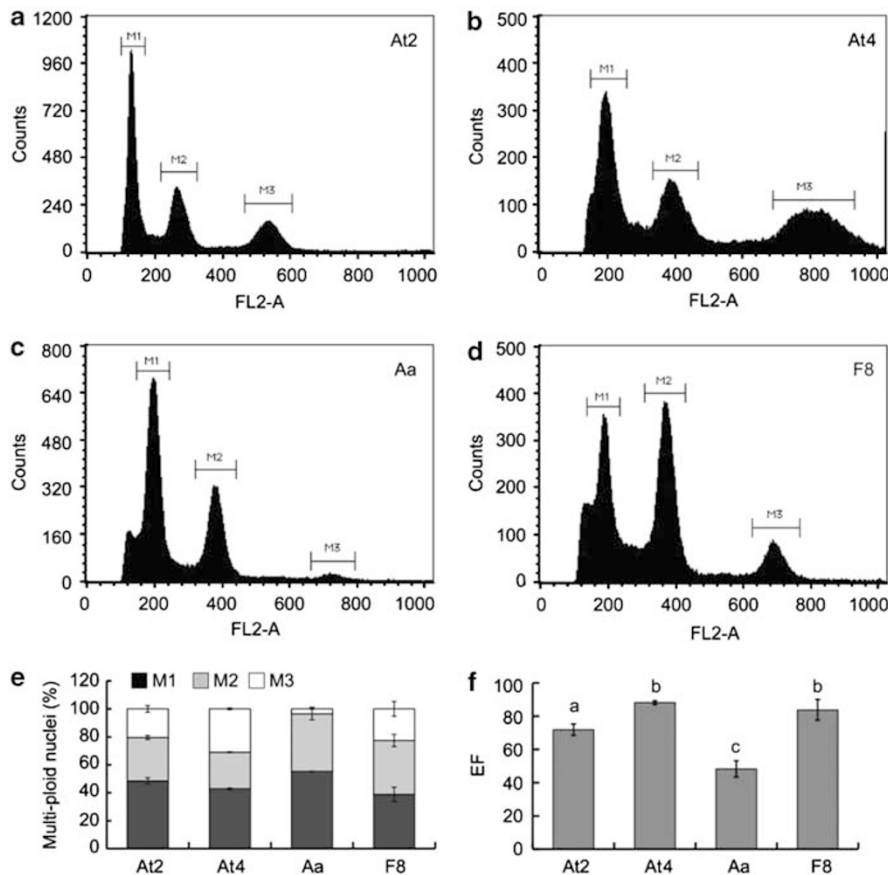


Figure 2 Flow-cytometric analysis of endopolyploidy in mature leaves of different *Arabidopsis* lines. (a–d) Flow-cytometry histograms of gated nuclei fluorescence detected using the FL2 (480 nm) photodetector (FL2-A) using mature leaves of an *A. thaliana* diploid (panel a), *A. thaliana* autotetraploid (panel b), *A. arenosa* tetraploid (panel c) and allotetraploid Allo733 in the F₈ generation (Allo733F₈) (panel d). Markers (M1 to M3) indicate 0, 1 and 2 endoreduplication events, respectively. (e) The proportion of cell ploidy levels ($n=3$). M1, M2 and M3 represent 2C, 4C, 8C nuclei in a diploid, respectively, and 4C, 8C and 16C nuclei in a tetraploid, respectively. (f) The endoreduplication factor (EF) represents the average endocycles per 100 cells ($n=3$) in the 4 lines examined. Error bars indicate s.e. The significance levels ($P \leq 0.05$) are shown as 'a, b and c' in multiple comparison tests.

from *A. thaliana* and *A. arenosa*, 97 (~8.4%) and 95 (~8.2%) proteins in F₁ and F₈ allotetraploids, respectively, showed non-additive abundance relative to the MPV of At4 and Aa, respectively.

Many non-additively expressed proteins in allotetraploids were also differentially expressed between the parents. Among the non-additively accumulated proteins in F₁ and F₈ allotetraploids, approximately 61–62% were also differentially expressed between the progenitors (Figure 3d). A total of 151 proteins were non-additively expressed in F₁ and F₈ allotetraploids, and 41 (~27%) of them were shared between the 2 allotetraploids (Figure 3d).

Effects of genome dosage and hybridization on proteome divergence in *Arabidopsis*

Among the differentially or non-additively expressed proteins, 14 of 95 differentially expressed proteins (~14.7%) between diploid and autotetraploid *A. thaliana* showed ± 2 -fold difference (Figure 4a). Among them, eight proteins were upregulated twofold or more, and six were downregulated twofold or more in the autotetraploid. Overall, the numbers of upregulation or downregulation of proteins were similar between an *A. thaliana* diploid and an isogenic autotetraploid.

Between *A. thaliana* and *A. arenosa* autotetraploids, 158 (~75.6%) of 209 differentially expressed proteins displayed ± 2 -fold expression difference (Figure 4b). Among them, 67 (~42.4%) proteins accumulated at higher levels in *A. arenosa* than in *A. thaliana*, whereas 91

(~57.6%) accumulated at higher levels in *A. thaliana* than in *A. arenosa* (Supplementary Table S3). A slightly larger proportion of non-additively expressed proteins, ~56.7% (55 of 97) and ~53.7% (51 of 95), respectively, were downregulated in F₁ and F₈ allotetraploids (Figures 4c and d). Moreover, ~24.7% (24 of 97) and ~28.4% (27 of 95) of the non-additively accumulated proteins, respectively, were downregulated by ≥ 2 -fold. However, the differences in upregulation or downregulation of proteins within each comparison (Aa/At4, F₁/MPV or F₈/MPV) were not statistically significant.

Finally, among the non-additively accumulated proteins in F₁ and F₈ allotetraploids, expression changes for the majority of proteins were in the same direction; 16 were upregulated, 22 downregulated and 3 in opposing directions (Supplementary Tables S4 and S5). The data suggested that proteomic changes in selfed progeny of polyploids are relatively stable.

Enrichment of protein functional groups in response to genome duplication, divergence and hybridization

Total proteins identified by the iTRAQ LC-MS/MS analysis were grouped into 14 GO biological process categories based on The Arabidopsis Information Resource GO annotation (Berardini *et al.*, 2004). Compared with the proportion of *A. thaliana* whole-genome proteins, proteins related to stress, abiotic or biotic stimulus, electron transport and energy pathways were significantly enriched (>300%)

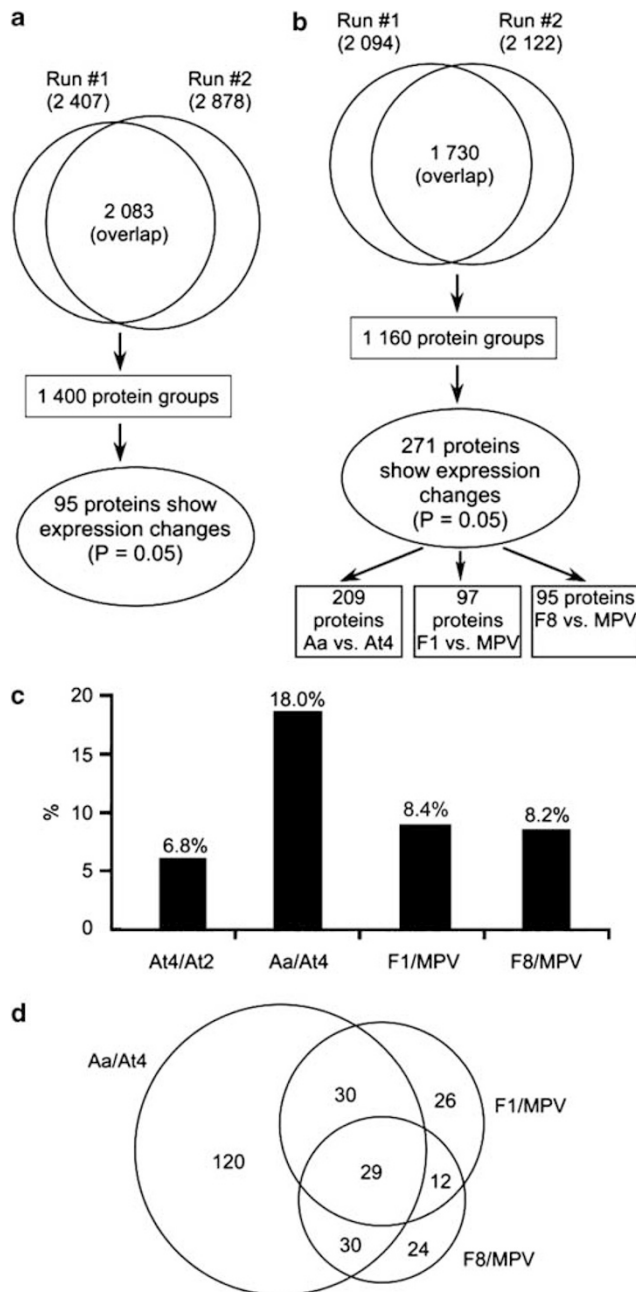


Figure 3 Use of iTRAQ strategy for analysis of proteome divergence and non-additive accumulation in *Arabidopsis*-related species. (a) Proteome comparison between *A. thaliana* (Ler) diploids (At2) and autotetraploids (At4). (b) Proteome comparison between *A. thaliana* autotetraploids (At4), *A. arenosa* autotetraploids (Aa) and their allotetraploid hybrids in the F₁ and F₈ generations. Two LC-MS/MS runs were performed for each comparison. Proteins identified in both runs (overlap) were used for expression divergence analysis. Proteins sharing the same peptides were assigned to a protein group, and the protein with the highest identification score designated as the group leader for proteomic comparisons among all *Arabidopsis* species. Protein expression differences were determined using a significance level of $P \leq 0.05$ within each comparison. MPV, mid-parent value ($\frac{1}{2}$ At4 + $\frac{1}{2}$ Aa). (c) Percentage of proteins showing expression divergence between *A. thaliana* (Ler) diploids (At2) and tetraploids (At4), *A. thaliana* and *A. arenosa* tetraploids (Aa/At4), and resynthesized F₁ and stable (F₈, Allo733F₈) allotetraploids and the MPV ($\frac{1}{2}$ At4 + $\frac{1}{2}$ Aa). (d) Venn diagram showing the number of differentially accumulated proteins between Aa and At4 and non-additively accumulated proteins between F₁ or F₈ allotetraploids and the MPV.

(Figure 5a). Proteins that belong to cell organization and biogenesis, protein metabolism and other biological processes were also over-represented (150–280%). In contrast, proteins of the transcription and signal transduction categories were significantly under-represented. No significant GO enrichments were found for proteins within the categories of developmental processes and DNA/RNA metabolism.

In comparison with differentially expressed proteins between *A. thaliana* diploids (At2) and autotetraploids (At4), GO categories of stress responses (~175%) and abiotic or biotic stimulus (~178%) were significantly enriched (χ^2 test, $P \leq 0.05$) (Figure 5b). Although proteins within electron transport and energy pathways were enriched (~197%), χ^2 tests revealed no significant difference. Within the stress response category, 25 of 37 proteins accumulated at higher levels in At4 than in At2. A similar trend was observed for proteins in response to the abiotic or biotic stimulus category; 24 of 36 proteins accumulated at higher levels in At4 than in At2 (Supplementary Figure S1; Supplementary Table S2). Although the proteomes between *A. thaliana* and *A. arenosa* showed the highest divergence, enrichment of differentially expressed proteins was modest among GO categories. Significant enrichment was found only for proteins related to the response to abiotic or biotic stimulus (Figure 5b). In resynthesized allotetraploids, non-additively expressed proteins involved in developmental processes, abiotic/biotic stimulus response and electron transport/energy pathways were relatively enriched (approximately 123–141%). Among differentially accumulated proteins, proteins in the transcription (in F₁) and DNA/RNA metabolism (in F₁ or F₈) categories were overrepresented by ~190% or more. However, these categories only accounted for a low proportion of proteins that were identified by iTRAQ (Figure 5a). No significant enrichment was found for any of the GO categories in either F₁/MPV or F₈/MPV comparison.

Comparison of proteomic with transcriptomic data

In comparison with a previously reported microarray analysis (Wang *et al.*, 2006b), 8 (~8.4%), 125 (~59.8%) and 44 (~46.3%) differentially or non-additively expressed proteins in this study also showed transcript level differences in At4/At2, Aa/At4 and F₈/MPV, respectively (Figure 6a). Similar upregulation or downregulation of transcript and protein levels suggests transcriptional regulation of gene expression. Between At4 and At2, five genes were upregulated at both transcript and protein levels (Supplementary Table S2). These included pathogenesis-related protein 5 (At1g75040), β -carbonic anhydrase 2 (At5g14740), TSK-associated protein 1 (At1g52410), mitochondrial alanyl-tRNA synthetase (At1g50200) and ribosome-recycling factor (At3g63190). The transcriptomic and proteomic comparison between At4 and Aa revealed 32 (25.6%) and 34 (27.2%) proteins displaying the same trend of upregulation and downregulation, respectively (Supplementary Table S3). In F₈ allotetraploids, 8 (~18.2%) and 18 (~40.9%) proteins shared the same trend of upregulation and downregulation, respectively (Supplementary Table S5). In the proteomic analysis, a major latex protein-like protein, major latex protein-like 423 (At1g24020), was highly expressed in *A. arenosa* (Aa/At4=4.322 \pm 0.318) but down-regulated in the allotetraploids (F₁/MPV=0.437 \pm 0.074; F₈/MPV=0.306 \pm 0.066), which was consistent with the microarray data of upregulation or downregulation at the transcript levels (Supplementary Tables S3 and S5). Interestingly, similar findings were found in resynthesized *B. napus* allotetraploids (stem tissues) using protein two-dimensional analysis (Albertin *et al.*, 2006) and semi-quantitative reverse transcriptase-PCR assays (Marmagne *et al.*, 2010). Finally, expression of 3 (~37.5%), 59 (~47.2%) and 18

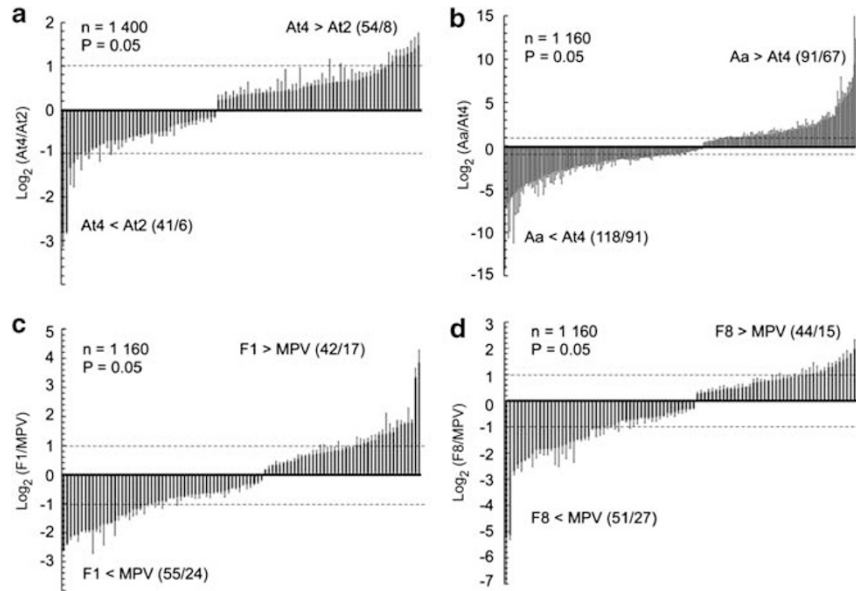


Figure 4 Protein expression changes in *Arabidopsis* autotetraploids and allotetraploids and their progenitors. Histograms showing expression levels (in \log_2 scale) of differentially expressed proteins between (a) *A. thaliana* diploids and autotetraploids; (b) *A. arenosa* and *A. thaliana* tetraploids; (c) F_1 -resynthesized allotetraploids and the MPV; and (d) F_8 -resynthesized allotetraploids (Allo733 F_8) and the MPV. The number of downregulated or upregulated proteins is showed in parenthesis. Error bars showed the s.d. of the log ratio. Dashed lines mark the ± 2 -fold change levels, respectively. MPV: mid-parent value. Total numbers of differentially expressed proteins over those with ± 2 -fold changes are shown in parentheses.

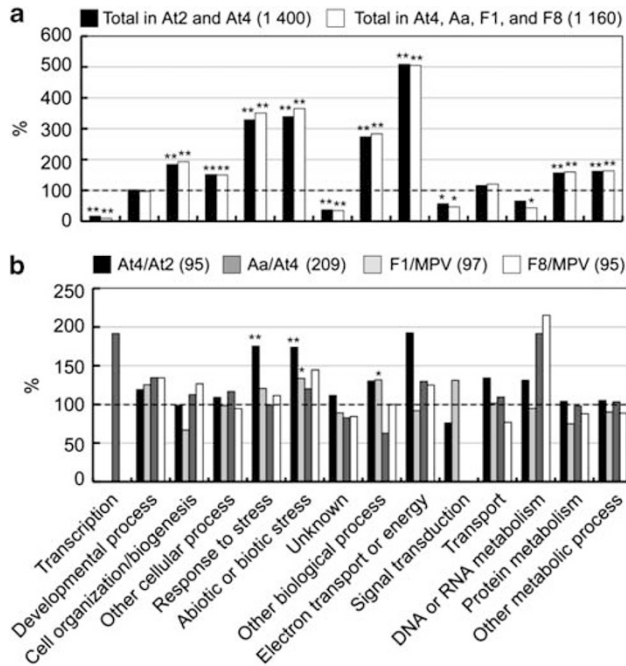


Figure 5 Gene ontology categorization of identified proteins in *Arabidopsis*-related species. (a) Total proteins identified by iTRAQ LC-MS/MS analysis were grouped into 14 biological process categories. The percentage of proteins identified in each category was compared with that in the corresponding category of the entire *A. thaliana* genome (dashed line=average or 100%). (b) The percentage distribution of differentially or non-additively accumulated proteins in *A. thaliana* (At2, At4), *A. arenosa* and resynthesized allotetraploids (F_1 and F_8 generations). The percentage distribution was compared with that of total proteins detected by iTRAQ analysis (dashed line=average or 100%). MPV: mid-parent value. Significant GO enrichments with $P \leq 0.05$ and $P \leq 0.01$ were denoted by '*' and '***', respectively.

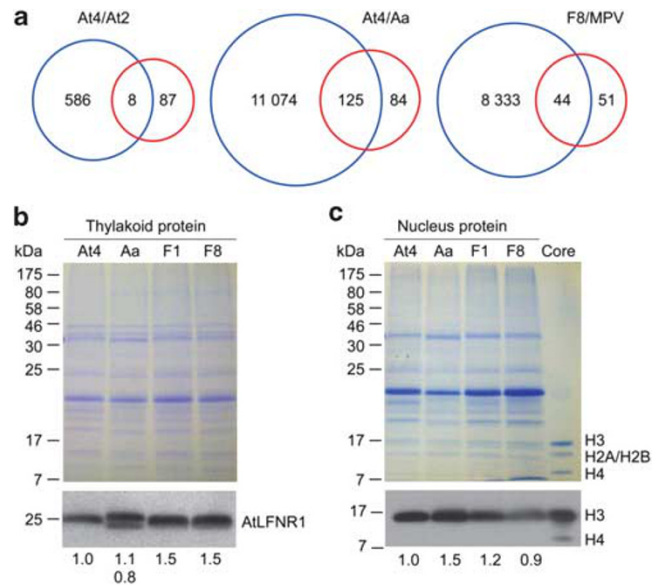


Figure 6 Comparison and validation of proteomic and transcriptomic data. (a) Venn diagrams showing the number of differentially or non-additively expressed proteins in comparison with transcript divergence using per-gene variance in the microarray study (Wang *et al.*, 2006b). (b, c) Validation of protein expression data. (Panel b) Total leaf thylakoid proteins (20 μ g) extracts from At4, Aa, F_1 , F_8 were resolved in a 15% Urea-SDS-PAGE for subsequent western blot analysis using antibodies against AtLFNR1. (Panel c) The nucleus proteins (20 μ g) were prepared from the same plant materials as shown in panel b and used for western blot analysis using the antibodies against histone H3 C terminus. Core histones (1 μ g) were also included as a positive control. Coomassie blue-stained gels were shown as loading controls.

(~40.9%) proteins was inversely correlated with the corresponding microarray data in At4/At2, Aa/At4 and F₈/MPV, respectively (Supplementary Tables S2, S3 and S5).

Validation of iTRAQ proteomic data

In the iTRAQ proteomic analysis, an *Arabidopsis* leaf-type ferredoxin-NADP oxidoreductase (AtLFNR1; At5g66190) was downregulated in *A. arenosa* relative to *A. thaliana* (Aa/At4=0.82 ± 0.03; Supplementary Table S4) and upregulated in F₁ (1.55 ± 0.04) and F₈ (1.38 ± 0.04) allotetraploids relative to the progenitors (Supplementary Tables S5 and S6). Using the antibodies against AtLFNR1, a specific band was detected in thylakoid protein extracts from At4, AlloF₁ and Allo733F₈ leaves (Figure 6b). Interestingly, a doublet with different sizes was detected in Aa. In allotetraploids, the absence of the corresponding signals similar to that detected in Aa suggested that only the *AtLFNR1* homeolog was expressed in allotetraploids. Taking this into consideration, AtLFNR1 intensities in F₁ and F₈ allotetraploids detected by western blot analysis were consistent with the proteomic data showing higher levels in allotetraploids than the MPV. The different LFNR1 bands and higher signal intensities detected in *A. arenosa* than in At4 could result from posttranslational modifications and/or presence of a LFNR1 isoform. In addition, sequence divergence between Aa and AtLFNR1 homologs may lead to underestimation of AaLFNR1 peptide levels in the iTRAQ analysis. These factors may explain why the AtLFNR1 ratio between At4 and Aa was lower in iTRAQ assays (Aa/At4=0.824 ± 0.029) than in the western blot analysis (Figure 6b) or in microarray analysis (Aa/At4=1.52; Supplementary Table S3) (Wang *et al.*, 2006b).

Proteomic data revealed upregulation of histone proteins (H3 and H4) in Aa relative to At4 (Aa/At4=2.83 ± 0.13 for H3 and 1.84 ± 0.18 for H4) (Supplementary Table S3) and downregulation of histones in allotetraploids (AlloF₁ or Allo733F₈; Supplementary Tables S4 and S5) relative to the progenitors. Western blot analysis of nucleus extracts using antibodies against histone H3 indicated the same trend of histone H3 accumulation levels as that in the proteomic data (Figure 6c).

DISCUSSION

Proteomic changes between *A. thaliana* diploid and isogenic autotetraploid are relatively small and enriched in response to stresses

A low level (~6.8%) of iTRAQ proteomic difference between *Arabidopsis* diploids and tetraploids is consistent with the transcriptome data previously reported for *Arabidopsis* (Wang *et al.*, 2006b), potato (Stupar *et al.*, 2007) and maize (Riddle *et al.*, 2010). The leaf or stem proteomes detected by protein 2-DE were also insignificant between *Brassica oleracea* diploids and tetraploids (Albertin *et al.*, 2005). The data collectively suggest relatively small changes in gene and protein expression in response to genome doubling (autopolyploidy). Flow-cytometry analysis revealed a proportionally higher level of endoreduplication in At4 than that in At2. Thus, proteomic changes observed between *A. thaliana* diploid and autotetraploid At4 are not likely caused by somatic endoreduplication.

Differentially expressed proteins related to stimuli or stress responses were more enriched in *A. thaliana* autotetraploids than in diploids. Expression of defense responsive genes often results in a fitness cost to the plants (Zangerl *et al.*, 1997; Heil and Baldwin, 2002; Tian *et al.*, 2003). Constitutive expression of pathogenesis-related genes in *Arabidopsis* mutants has a negative impact on plant size (Bowling *et al.*, 1997; Clarke *et al.*, 1998). Upregulation of stress and defense responsive genes in *A. thaliana* autotetraploids may thus slow

growth. As a result, diploid and autotetraploid plants have a similar size during vegetative growth, despite autotetraploids having generally larger flowers and seeds than diploids (Chen, 2010).

Genome divergence and hybridization have large effects on proteomic changes

Compared with the relatively small proteomic changes between *A. thaliana* diploids and autopolyploids, proteomic divergence was relatively high between the two related species, *A. thaliana* and *A. arenosa*, and their allotetraploids (in F₁ and F₈ generations). The high level of proteome divergence between *A. thaliana* and *A. arenosa* is consistent with the high level of gene expression divergence detected by microarrays (Wang *et al.*, 2006b) between the two species, which diverged ~6 Mya (Koch *et al.*, 2000). *A. thaliana* and *A. arenosa* have five and eight chromosomes in their haploid genomes, respectively. Chromosomal rearrangements such as inversions, translocations and fusions (Kuitinen *et al.*, 2004; Yogeewaran *et al.*, 2005; Lysak *et al.*, 2006b) may lead to some differences in gene expression and proteomes between the two species. Moreover, *cis*- and *trans*-regulatory differences between species (Wang *et al.*, 2006a; Chen, 2007; Wittkopp *et al.*, 2008; Ng *et al.*, 2011) may also lead to proteomic changes.

In resynthesized allotetraploids, a higher percentage of non-additively accumulated proteins showed downregulation in both F₁ and F₈ generations, consistent with the observed repression of non-additively expressed genes (approximately 65–76%) in previous microarray analyses (Wang *et al.*, 2006b). In addition to epigenetic regulation at the transcriptional level, homeologous protein interactions and divergence may also contribute to a high level of divergence in proteomes and regulatory networks between species and in allopolyploids relatively to the parents (Chen, 2007; Doyle *et al.*, 2008; Jackson and Chen, 2010). Moreover, the EF was significantly higher in At4 or Allo733F₈ than in Aa, suggesting that somatic endopolyploidy may contribute to the high proteome divergence between *A. arenosa* tetraploid and *A. thaliana* autotetraploid lines or resynthesized allotetraploids. On the other hand, proteins that are expressed differently between species may be involved in somatic endopolyploidy levels. Therefore, the effect of somatic endopolyploidy on proteome changes may not be easily separated from that of genetic polyploidy. Interestingly, similar somatic endopolyploidy levels between *A. thaliana* autotetraploid and *A. suecica*-like allotetraploid suggests that somatic endopolyploidy levels are related to the maternal progenitor *A. thaliana*. Whether the parent of origin affects somatic endopolyploidy remains to be investigated. In general, the proteomic data suggest larger effects of genome divergence and hybridization on proteomic changes than those of genome doubling (diploid vs tetraploid), consistent with gene expression data (Wang *et al.*, 2006b; Pignatta *et al.*, 2010).

No functional enrichment of non-additively expressed proteins in allotetraploids

In contrast to autopolyploidization, non-additively expressed proteins in resynthesized allotetraploids (F₁ or F₈ generation) displayed no significant enrichment of particular GO functions. This is similar to the results from synthetic *B. napus*; non-additively expressed proteins were not found to be related to any functional categories or metabolic pathways (Albertin *et al.*, 2007). However, in the transcriptome study, non-additively expressed genes involved in defense and plant hormonal regulation were enriched in *Arabidopsis* allotetraploids (Wang *et al.*, 2006b). This suggests that interactions between *trans*-acting homeologous proteins in the same nuclei of allopolyploids may

provide a mechanism for maintaining the proteomic balance between *A. thaliana* and *A. arenosa* genomes. Such regulation at the protein level may compensate changes in transcript levels that are likely affected by *cis*-regulatory modulation, leading to balanced changes in all categories of proteins. Alternatively, interactions between homologous proteins could also create new regulatory networks such that functional enrichment of non-additively expressed proteins could only be found under certain growth and environmental conditions. This hypothesis is testable in resynthesized allotetraploids and their progenitors in response to external stimuli.

The degree of changes between transcriptomes and proteomes is similar, but the content of shared changes is different

Although the levels of transcriptome and proteome divergence were similar in response to genome doubling, divergence or hybridization, the majority of differentially or non-additively expressed proteins were not significantly expressed in microarray analyses (Wang *et al.*, 2006b). A limited correlation between mRNA and protein levels has been reported in yeast (Gygi *et al.*, 1999). In yeast, the aneuploidy effects on transcript and protein levels have also been inconsistent (Torres *et al.*, 2007). The lack of compensation for gene dosage and aneuploidy in yeast may suggest larger effects of aneuploidy than euploidy on transcriptome and proteome, as well as phenotypic variation in response to growth conditions and fitness tests (Pavelka *et al.*, 2010; Springer *et al.*, 2010). A recent study using proteome and transcriptome analysis in single *Escherichia coli* cells suggested that protein and mRNA copy numbers in single cells for any given gene are uncorrelated, displaying a substantial amount of intrinsic and extrinsic noises (Taniguchi *et al.*, 2010). Comparison of transcriptomes using semi-quantitative and quantitative PCR assays and two-dimensional protein gel electrophoresis in *B. napus* also revealed that the majority of transcript changes (~62%) did not reflect the differential protein accumulation in the resynthesized allotetraploids compared with the diploid progenitors, *B. oleracea* and *B. rapa* (Marmagne *et al.*, 2010). Changes in genome structure, gene expression and interphase organization have also been documented in *A. thaliana* trisomics (Huettel *et al.*, 2008).

The discordance between transcriptomic and proteomic data is probably related to several factors. First, some gene expression changes are associated with small RNAs, including microRNAs and small-interfering RNAs (Ha *et al.*, 2009). Moreover, reactivation of retrotransposons in newly synthesized wheat allopolyploids was found to be responsible for sense and antisense transcript production from adjacent genes, leading to subsequent alteration of gene expression in amphiploids (Kashkush *et al.*, 2003). Second, in addition to transcriptional regulation, discordance in gene expression and proteomic changes may reflect transcript stability, post-translational modifications and/or protein stabilities. Third, several experimental factors could contribute to transcriptome and proteome discordance. As examples, the divergence between *A. arenosa* and *A. thaliana* homologous protein sequences could lead to addition/deletion of either lysine or arginine residues. Therefore, homologous proteins in the *A. arenosa* genome may yield a subset of tryptic-digested peptides that are different from those in the *A. thaliana* genome. As the International Protein Index *Arabidopsis* database is built on protein sequences from *A. thaliana*, MS/MS spectra of tryptic-digested peptides originating in the *A. arenosa* genome may be underrepresented in the database. In addition, we cannot exclude a possibility that the discordance is caused by different sets of plant materials used in these comparative studies, as well as the limitation of specificity and sensitivity associated with the microarray technology. Next-generation

mRNA sequencing technology is expected to provide a more accurate and sensitive evaluation of transcriptome changes in *A. thaliana*-related species and allopolyploids. A more detailed comparative analysis of proteomic and transcriptomic data may help to evaluate the contributions of transcriptional, translational and post-translational regulation of gene expression in *Arabidopsis*-related species and polyploids. Finally, as mentioned above, expression and proteomic changes in allopolyploids and between species may be confounded by different levels of endoreduplication between *A. thaliana* and *A. arenosa* and in allotetraploids. Although EFs are proportional to ploidy levels (for example, diploid vs tetraploid) within the same species (*A. thaliana*), EF factors are different between *A. arenosa* and other ploidy lines, including allotetraploids. The relative effects of allopolyploidy and endopolyploidy on gene expression and proteomes should be revisited in the future using flow sorting of specific cell types for analyses (Zhang *et al.*, 2005).

Western blot validated proteome data

Differential accumulation of AtLFNR1 and histone H3 among the *Arabidopsis*-related species and polyploids detected by the iTRAQ approach were validated by western blot analysis. The same trend of histone H3 accumulation was detected in At4, Aa, AlloF₁ and Allo733F₈ using these two approaches. AtLFNR1 migrated into two distinct spots in two-dimensional protein gel electrophoresis, suggesting possible post-translational modifications of this protein (Lintala *et al.*, 2007), which cannot be detected by microarray analysis in allotetraploids (Wang *et al.*, 2006b). iTRAQ quantitation detected a higher LFNR1 accumulation in *A. thaliana* than that in *A. arenosa* because it could not account for the different-size doublet of LFNR1 found in Aa by western blot analysis. In allotetraploids, only the AtLFNR1 homeolog was detected by western blot analysis, and its accumulation level was higher in Aa than in At4. Interestingly, mutation of *AtLFNR1* led to a reduced rosette size, chlorophyll content and overall biomass in *A. thaliana* (Lintala *et al.*, 2007). This may provide a link between AtLFNR1 accumulation and overall increased growth vigor in allotetraploids, demonstrating the importance of comparing transcriptome with proteomic data in studying the consequences of speciation and ploidy effects. Although data validation was very limited at this stage, our results demonstrate that iTRAQ is a valuable approach to the study of proteomic divergence in *Arabidopsis*-related species and allopolyploids.

CONCLUSIONS

Complementary to previous transcriptome (Wang *et al.*, 2006b) and small RNA data (Ha *et al.*, 2009), our current proteomic analysis provided a broad view of the rapid and dynamic changes in proteomes in closely related *Arabidopsis* species and allopolyploids. Genome duplication within *A. thaliana* resulted in limited changes in proteomes. In contrast but consistent with their evolutionary divergence, autotetraploid *A. thaliana* and outcrossing *A. arenosa* had a relatively high level of proteomic divergence. Differentially expressed proteins between *A. thaliana* diploid and autotetraploid or between *A. thaliana* and *A. arenosa* tetraploids were enriched in functional groups in response to abiotic or biotic stimuli. In contrast, proteomic divergence in allotetraploids relative to the parents was intermediate and occurred over all biological processes. In the future, comparative analyses of these data sets and specific changes in some genes and proteins, as well as genome-wide chromatin modifications (Ha *et al.*, 2011), will allow us to identify additional protein factors and genes that are potentially responsible for altered growth vigor, morphology and adaptation in *Arabidopsis* species and resynthesized allopolyploids.

DATA ARCHIVING

All mass spectra used in this study are publicly available for download at the Proteome Commons Tranche network (<https://proteomecommons.org/tranche/>) using the following hash: FZ0WuK7vW-JYYFb2qLL7vCZGGV5l/YesB2So0SRxp16o34iXM29uQRhAnXu28HaX9MWEOb9ZSZf5heeYBQ3UIUwL7zjkAAAAAABZSw==.

A complete listing of protein quantification data is provided as Supplementary data set 1 and it is available for download at the Dryad repository (<http://datadryad.org>): doi:10.5061/dryad.sk318rd3.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank James A Birchler for the fluorescence *in situ* hybridization (FISH) cytogenetics workshop where CZ adapted the FISH methodology for this study and Angela Bardo in the Core Facility of Institute for Cellular and Molecular Biology at The University of Texas at Austin for assistance with fluorescence image analysis. We also thank Xiaoli Shi for helpful discussions to improve statistical analysis of the proteomic data. This study was supported by grants from the Plant Genome Research Program of the National Science Foundation (IOS0733857 to ZJC and ISO0924023 to SPB).

Author Contributions: ZJC, DW-KN and CZ designed experiments. DW-KN and CZ prepared the biological materials and validated the data. ZS performed iTRAQ proteomic experiments. DW-KN, C Z, ZS, SPB and ZJC analyzed the data. DW-KN and ZJC wrote the paper, and ZS, SPB, DW-KN and ZJC edited the paper.

- Albertin W, Alix K, Balliau T, Brabant P, Davanture M, Malosse C *et al.* (2007). Differential regulation of gene products in newly synthesized *Brassica napus* allotetraploids is not related to protein function nor subcellular localization. *BMC Genomics* **8**: 56.
- Albertin W, Balliau T, Brabant P, Chevre AM, Eber F, Malosse C *et al.* (2006). Numerous and rapid nonstochastic modifications of gene products in newly synthesized *Brassica napus* allotetraploids. *Genetics* **173**: 1101–1113.
- Albertin W, Brabant P, Catrice O, Eber F, Jenczewski E, Chevre AM *et al.* (2005). Autopolyploidy in cabbage (*Brassica oleracea* L.) does not alter significantly the proteomes of green tissues. *Proteomics* **5**: 2131–2139.
- Albertin W, Langella O, Joets J, Negroni L, Zivy M, Damerval C *et al.* (2009). Comparative proteomics of leaf, stem, and root tissues of synthetic *Brassica napus*. *Proteomics* **9**: 793–799.
- Andreyev AY, Shen Z, Guan Z, Ryan A, Fahy E, Subramaniam S *et al.* (2010). Application of proteomic marker ensembles to subcellular organelle identification. *Mol Cell Proteomics* **9**: 388–402.
- Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soci (Series B)* **57**: 289–300.
- Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P *et al.* (2004). Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiol* **135**: 745–755.
- Birchler Ja, Newton KJ (1981). Modulation of protein-levels in chromosomal dosage series of maize—the biochemical basis of aneuploid syndromes. *Genetics* **99**: 247–266.
- Bowling SA, Clarke JD, Liu Y, Klessig DF, Dong X (1997). The *cpr5* mutant of *Arabidopsis* expresses both *NPR1*-dependent and *NPR1*-independent resistance. *Plant Cell* **9**: 1573–1584.
- Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR *et al.* (2001). Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* **13**: 1499–1510.
- Buggs RJ, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA *et al.* (2009). Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. *Heredity* **103**: 73–81.
- Carvajal-Rodriguez A, de Una-Alvarez J, Rolan-Alvarez E (2009). A new multitest correction (SGoF) that increases its statistical power when increasing the number of tests. *BMC Bioinformatics* **10**: 209.
- Chen ZJ (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol* **58**: 377–406.
- Chen ZJ (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci* **15**: 57–71.
- Clarke JD, Liu Y, Klessig DF, Dong X (1998). Uncoupling PR gene expression from NPR1 and bacterial resistance: characterization of the dominant *Arabidopsis cpr6-1* mutant. *Plant Cell* **10**: 557–569.
- Comai L, Tyagi AP, Lysak MA (2003). FISH analysis of meiosis in *Arabidopsis* allopolyploids. *Chromosome Res* **11**: 217–226.
- Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y *et al.* (2000). Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* **12**: 1551–1568.
- Cookson SJ, Radziejowski A, Granier C (2006). Cell and leaf size plasticity in *Arabidopsis*: what is the role of endoreduplication? *Plant Cell Environ* **29**: 1273–1283.
- de Una-Alvarez J, Carvajal-Rodriguez A (2010). 'SGoFance Trace': assessing significance in high dimensional testing problems. *PLoS One* **5**: e15930.
- Diz AP, Carvajal-Rodriguez A, Skibinski DO (2011). Multiple hypothesis testing in proteomics: a strategy for experimental work. *Mol Cell Proteomics* **10**: M110.004374.
- Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS *et al.* (2008). Evolutionary genetics of genome merger and doubling in plants. *Annu Rev Genet* **42**: 443–461.
- Flagel L, Udall J, Nettleton D, Wendel J (2008). Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol* **6**: 16.
- Gaeta RT, Yoo SY, Pires JC, Doerge RW, Chen ZJ, Osborn TC (2009). Analysis of gene expression in resynthesized *Brassica napus* Allopolyploids using arabidopsis 70mer oligo microarrays. *PLoS One* **4**: e4760.
- Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **220**: 1049–1051.
- Griffin TJ, Xie H, Bandhakavi S, Popko J, Mohan A, Carlis JV *et al.* (2007). iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer. *J Proteome Res* **6**: 4200–4209.
- Gygi SP, Rochon Y, Franza BR, Aebersold R (1999). Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**: 1720–1730.
- Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ *et al.* (2009). Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proc Natl Acad Sci USA* **106**: 17835–17840.
- Ha M, Ng DW, Li WH, Chen ZJ (2011). Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Res* **21**: 590–598.
- Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ (2006). Transcriptome shock after interspecific hybridization in senescio is ameliorated by genome duplication. *Curr Biol* **16**: 1652–1659.
- Heil M, Baldwin IT (2002). Fitness costs of induced resistance: emerging experimental support for a slippery concept. *Trends Plant Sci* **7**: 61–67.
- Holm S (1979). A simple sequentially rejective multiple test procedure. *Scand J Stat* **6**: 65–70.
- Hovav R, Udall JA, Chaudhary B, Hovav E, Flagel L, Hu G *et al.* (2008). The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. *PLoS Genet* **4**: e25.
- Huettel B, Kreil DP, Matzke M, Matzke AJ (2008). Effects of aneuploidy on genome structure, expression, and interphase organization in *Arabidopsis thaliana*. *PLoS Genet* **4**: e1000226.
- Islam N, Tsujimoto H, Hirano H (2003). Proteome analysis of diploid, tetraploid and hexaploid wheat: towards understanding genome interaction in protein expression. *Proteomics* **3**: 549–557.
- Jackson S, Chen ZJ (2010). Genomic and expression plasticity of polyploidy. *Curr Opin Plant Biol* **13**: 153–159.
- Kashkush K, Feldman M, Levy AA (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat Genet* **33**: 102–106.
- Kato A, Lamb JC, Birchler JA (2004). Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc Natl Acad Sci USA* **101**: 13554–13559.
- Koch MA, Haubold B, Mitchell-Olds T (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol* **17**: 1483–1498.
- Koch MA, Wernisch M, Schmickl R (2008). *Arabidopsis thaliana*'s wild relatives: an updated overview on systematics, taxonomy and evolution. *Taxon* **57**: 933–943.
- Kuittinen H, de Haan AA, Vogl C, Oikarinen S, Leppala J, Koch M *et al.* (2004). Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. *Genetics* **168**: 1575–1584.
- Leitch AR, Leitch IJ (2008). Genomic plasticity and the diversity of polyploid plants. *Science* **320**: 481–483.
- Lintala M, Allahverdiyeva Y, Kidron H, Piippo M, Battchikova N, Suorsa M *et al.* (2007). Structural and functional characterization of ferredoxin-NADP+ oxidoreductase using knock-out mutants of *Arabidopsis*. *Plant J* **49**: 1041–1052.
- Lysak M, Franz P, Schubert I (2006a). Cytogenetic analyses of *Arabidopsis*. *Methods Mol Biol* **323**: 173–186.
- Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I (2006b). Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc Natl Acad Sci USA* **103**: 5224–5229.
- Marmagne A, Brabant P, Thiellement H, Alix K (2010). Analysis of gene expression in resynthesized *Brassica napus* allotetraploids: transcriptional changes do not explain differential protein regulation. *New Phytol* **186**: 216–227.
- Ng DW, Zhang Z, Miller M, Palmer G, Whiteley M, Tholl D *et al.* (2011). *Cis-* and *trans-*regulation of miR163 and target genes confers natural variation of secondary metabolites in *Arabidopsis* related species and allopolyploids. *Plant Cell* **23**: 1729–1740.
- O'Brien RN, Shen Z, Tachikawa K, Lee PA, Briggs SP (2010). Quantitative proteome analysis of pluripotent cells by iTRAQ mass tagging reveals post-transcriptional regulation of proteins required for ES cell self-renewal. *Mol Cell Proteomics* **9**: 2238–2251.
- O'Kane S, Schaal B, Al-Shehbaz I (1995). The origins of *Arabidopsis suecica* (Brassicaceae), as indicated by nuclear rDNA sequences, and implications for rDNA evolution. *System Bot* **21**: 559–566.

- Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L *et al.* (2010). Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468**: 321–325.
- Pignatta D, Dilkes BP, Yoo SY, Henry IM, Madlung A, Doerge RW *et al.* (2010). Differential sensitivity of the *Arabidopsis thaliana* transcriptome and enhancers to the effects of genome doubling. *New Phytologist* **186**: 194–206.
- Riddle NC, Jiang H, An L, Doerge RW, Birchler JA (2010). Gene expression analysis at the intersection of ploidy and hybridity in maize. *Theor Appl Genet* **120**: 341–353.
- Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S *et al.* (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**: 1154–1169.
- Saleh A, Alvarez-Venegas R, Avramova Z (2008). An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in *Arabidopsis* plants. *Nat Protoc* **3**: 1018–1025.
- Salmon A, Ainouche ML, Wendel JF (2005). Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Mol Ecol* **14**: 1163–1175.
- Schmickl R, Koch MA (2011). *Arabidopsis* hybrid speciation processes. *Proc Natl Acad Sci USA* **108**: 14192–14197.
- Schwartz JC, Syka JEP, Quarmby ST (2005). *The 53rd ASMS Conference on Mass Spectrometry and Allied Topics*. San Antonio, Texas.
- Soltis PS, Soltis DE (2009). The role of hybridization in plant speciation. *Annu Rev Plant Biol* **60**: 561–588.
- Song X, Ni Z, Yao Y, Xie C, Li Z, Wu H *et al.* (2007). Wheat (*Triticum aestivum* L.) root proteome and differentially expressed root proteins between hybrid and parents. *Proteomics* **7**: 3538–3557.
- Springer M, Weissman JS, Kirschner MW (2010). A general lack of compensation for gene dosage in yeast. *Mol Syst Biol* **6**: 368.
- Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, Ouyang S *et al.* (2007). Phenotypic and transcriptomic changes associated with potato autopolyploidization. *Genetics* **176**: 2055–2067.
- Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J *et al.* (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**: 533–538.
- Tate JA, Ni Z, Scheen AC, Koh J, Gilbert CA, Lefkowitz D *et al.* (2006). Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* **173**: 1599–1611.
- Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003). Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**: 74–77.
- Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ *et al.* (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**: 916–924.
- Wang J, Tian L, Lee HS, Chen ZJ (2006a). Nonadditive regulation of *FRI* and *FLC* loci mediates flowering-time variation in *Arabidopsis* allopolyploids. *Genetics* **173**: 965–974.
- Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B *et al.* (2006b). Genome-wide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* **172**: 507–517.
- Wang J, Tian L, Madlung A, Lee HS, Chen M, Lee JJ *et al.* (2004). Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* **167**: 1961–1973.
- Wendel JF (2000). Genome evolution in polyploids. *Plant Mol Biol* **42**: 225–249.
- Wittkopp PJ, Haerum BK, Clark AG (2008). Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet* **40**: 346–350.
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009). The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci USA* **106**: 13875–13879.
- Yao H, Kato A, Mooney B, Birchler JA (2011). Phenotypic and gene expression analyses of a ploidy series of maize inbred Oh43. *Plant Mol Biol* **75**: 237–251.
- Yogeewaran K, Frary A, York TL, Amenta A, Lesser AH, Nasrallah JB *et al.* (2005). Comparative genome analyses of *Arabidopsis* spp.: inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. *Genome Res* **15**: 505–515.
- Zangerl AR, Arntz AM, Berenbaum MR (1997). Physiological price of an induced chemical defense: photosynthesis, respiration, biosynthesis, and growth. *Oecologia* **109**: 433–441.
- Zhang C, Gong FC, Lambert GM, Galbraith DW (2005). Cell type-specific characterization of nuclear DNA contents within complex tissues and organs. *Plant Methods* **1**: 7.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)