The Plant Cell, Vol. 30: 2330–2351, October 2018, www.plantcell.org © 2018 ASPB.

Check for updates

A Spatiotemporal DNA Endoploidy Map of the Arabidopsis Root Reveals Roles for the Endocycle in Root Development and Stress Adaptation

Rahul Bhosale,[a,b,c,d,1] Veronique Boudolf,[a,b,1] Fabiola Cuevas,[a,b,1] Ran Lu,[a,b] Thomas Eekhout,[a,b] Zhubing Hu,[a,b,e] Gert Van Isterdael,[a,b,f] Georgina M. Lambert,[g] Fan Xu,[h] Moritz K. Nowack,[a,b] Richard S. Smith,[i] Ilse Vercauteren,[a,b] Riet De Rycke,[a,b,j] Veronique Storme,[a,b] Tom Beeckman,[a,b] John C. Larkin,[k] Anna Kremer,[d,j] Herman Höfte,[h] David W. Galbraith,[g,l] Robert P. Kumpf,[a,b] Steven Maere,[a,b,c,2] and Lieven De Veylder[a,b,2]

[a] Department of Plant Biotechnology and Bioinformatics, Ghent University, B-9052 Ghent, Belgium
[b] VIB Center for Plant Systems Biology, B-9052 Ghent, Belgium
[c] Bioinformatics Institute Ghent, Ghent University, B-9052 Ghent, Belgium
[d] Center for Plant Integrative Biology, University of Nottingham, Sutton Bonington LE12 5RD, United Kingdom
[e] Institute of Plant Stress Biology, State Key Laboratory of Cotton Biology, Department of Biology, Henan University, 475004 Kaifeng, China
[f] VIB Flow Core, B-9052 Ghent, Belgium
[g] School of Plant Sciences, University of Arizona, Tucson, Arizona 85721
[h] Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, 78000 Versailles, France
[i] Department of Comparative Development and Genetics, Max Planck Institute for Plant Breeding Research, D-50829 Cologne, Germany
[j] VIB Bio Imaging Core, B-9052 Ghent, Belgium
[k] Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803
[l] Key Laboratory of Plant Stress Biology, State Key Laboratory of Cotton Biology, School of Life Sciences, Henan University, Kaifeng 475004, China

ORCID IDs: 0000-0001-6515-4922 (R.B.); 0000-0001-6771-142X (V.B.); 0000-0002-6523-8584 (F.C.); 0000-0002-2536-1692 (R.L.); 0000-0002-2878-1553 (T.E.); 0000-0002-5394-9864 (Z.H.); 0000-0001-6626-1316 (G.V.I.); 0000-0002-5000-9666 (G.M.L.); 0000-0002-5740-3156 (F.X.); 0000-0001-8918-7577 (M.K.N.); 0000-0001-9220-0787 (R.S.S.); 0000-0001-9561-8468 (I.V.); 0000-0001-8270-7015 (R.D.R.); 0000-0003-4762-6580 (V.S.); 0000-0001-8656-2060 (T.B.); 0000-0002-7156-508X (J.C.L.); 0000-0002-4613-1060 (A.K.); 0000-0002-5728-146X (H.H.); 0000-0003-4020-1635 (D.W.G.); 0000-0002-4046-6935 (R.P.K.); 0000-0002-5341-136X (S.M.); 0000-0003-1150-4426 (L.D.V.)

Somatic polyploidy caused by endoreplication is observed in arthropods, molluscs, and vertebrates but is especially prominent in higher plants, where it has been postulated to be essential for cell growth and fate maintenance. However, a comprehensive understanding of the physiological significance of plant endopolyploidy has remained elusive. Here, we modeled and experimentally verified a high-resolution DNA endoploidy map of the developing *Arabidopsis thaliana* root, revealing a remarkable spatiotemporal control of DNA endoploidy levels across tissues. Fitting of a simplified model to publicly available data sets profiling root gene expression under various environmental stress conditions suggested that this root endoploidy patterning may be stress-responsive. Furthermore, cellular and transcriptomic analyses revealed that inhibition of endoreplication onset alters the nuclear-to-cellular volume ratio and the expression of cell wall-modifying genes, in correlation with the appearance of cell structural changes. Our data indicate that endopolyploidy might serve to coordinate cell expansion with structural stability and that spatiotemporal endoreplication pattern changes may buffer for stress conditions, which may explain the widespread occurrence of the endocycle in plant species growing in extreme or variable environments.

INTRODUCTION

Organismal development involves continuous and reiterative organogenesis, during which complex molecular and developmental programs maintain the production of new cells and their subsequent differentiation. In plants, this process mainly occurs at the root and shoot apical meristems, the focal points where cells proliferate through mitotic cell divisions. Upon leaving these meristems, the cells differentiate and simultaneously increase their cell size through postmitotic expansion. This switch from proliferation to differentiation is accompanied in some plant

[1] These authors contributed equally to this work.
[2] Address correspondence to steven.maere@ugent.vib.be or lieven.deveylder@ugent.vib.be.
The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: Steven Maere (steven.maere@ugent.vib.be) and Lieven De Veylder (lieven.deveylder@ugent.vib.be).
www.plantcell.org/cgi/doi/10.1105/tpc.17.00983

IN A NUTSHELL

Background: Every new cell arises from division of an existing cell. In plants, these cell divisions mainly occur at the tip of shoots and roots. After a few rounds of cell division, cells stop dividing and differentiate, while simultaneously increasing their size. In many plant species, the start of differentiation is accompanied by the onset of endoreplication, an alternate cell cycle during which cells duplicate their genome without cell division, resulting in a duplication of their DNA content, i.e., an increase in ploidy level. In the root tip of the model plant species *Arabidopsis thaliana*, up to three endocycles can be observed, resulting in cells with a 2C, 4C, 8C, or 16C DNA content. The biological significance of these endoploidy increases is still under debate.

Question: Although we know that many plant species undergo endoreplication, our knowledge of which cell types become polyploid and how cells with different endoploidy levels are integrated into a developing organ is limited. Current protocols that quantify DNA ploidy are either time-consuming or lack tissue resolution. We addressed this problem by combining biological data with computational modeling.

Findings: Through the integration of different experimental data sets into a mathematical model, we pinpointed a set of 332 marker genes that reliably predict a three-dimensional DNA ploidy map for the complete root tip of Arabidopsis. The map obtained revealed that the level of endoreplication is to a remarkable extent differentially controlled across the distinct root tissues and their developmental stages. Moreover, a simplified version of the mathematical model allowed us to predict the effect of 149 different stress treatments on the root's endoploidy distribution, and our results suggest that the root's DNA ploidy pattern is highly stress-responsive. Furthermore, we demonstrated that endoreplication triggers the expression of cell wall-modifying genes, which might function to safeguard the cell's structural stability upon turgor-driven cell expansion.

Next steps: We are currently aiming to understand exactly how the spatiotemporal root endoreplication pattern changes under various stresses and whether these changes are linked to stress adaptation. We are also studying how the endoploidy level of a cell controls expression of cell wall biosynthesis and cell wall-modifying genes.

species by a transition from the mitotic cell cycle to the endocycle, an alternative cell cycle during which chromosomes are replicated but cells do not divide (De Veylder et al., 2011; Breuer et al., 2014). During such endocycles, also known as endoreplication, every round of full-genome DNA replication results in a doubling of the endoploidy level of the cell.

Endoreplication is not restricted to higher plants but is observed also across a wide variety of cell types in lower invertebrates, arthropods, and mammals (Fox and Duronio, 2013). In lower invertebrates, endoreplication is most often associated with increased cell size, and it is believed to be a crucial determinant of adult body size (Flemming et al., 2000). In arthropods such as the fruit fly *Drosophila melanogaster*, endoreplication is associated with growth and increased metabolic activity of different cell types, including the salivary glands, gut epidermis, and follicle and nurse cells of the adult fly (Edgar and Orr-Weaver, 2001). Mammalian cells that endoreplicate include placental trophoblast giant cells, hepatocytes, blood megakaryocytes, and epithelial keratinocytes (Fox and Duronio, 2013).

Across flowering plants, endoreplication appears to be particularly common and it often occurs in economically important tissues, including cereal endosperm (Sabelli and Larkins, 2009), cotton (*Gossypium hirsutum*) fibers (Van't Hof, 1999), and tomato (*Solanum lycopersicum*) fruits (Chevalier et al., 2011). Endoreplication typically occurs in plant tissues that develop mass very quickly or that possess a high metabolic activity. It is thought that the extra gene copies generated by endoreplication support transcription of sufficient RNA to sustain metabolic processes. Correspondingly, Bourdon et al. (2012) demonstrated that rRNA, RNA polymerase II abundance, and gene transcript levels in tomato fruits increase with endoploidy level. Next to a role in driving metabolism, a correlation is frequently observed between the endoploidy level and the size of a cell. This is most clearly illustrated by the epidermal pavement and hair cells within *Arabidopsis thaliana* leaves, where the largest cells possess the highest endoploidy level (Melaragno et al., 1993; Roeder et al., 2010). However, many experiments assessing the relationship between cell growth and endoploidy have revealed a lack of absolute rules; for instance, there are cases in which trichome cell size does not correlate with endoploidy (Schnittger et al., 1998, 2003). Similarly, plants that ectopically express the cyclin-dependent kinase inhibitor *KIP-RELATED2* display a strong inhibition of their endocycle, but enlarged pavement cell size in comparison to control plants (De Veylder et al., 2001), and it has been suggested that the relationship between cell volume and endoploidy depends on cell identity (Katagiri et al., 2016). Finally, endoreplication has been reported to play a role in sustaining cell fate (Bramsiepe et al., 2010) and cell size patterning (Roeder et al., 2010).

Different environmental factors have an effect on the endoreplication level of cells and tissues (De Veylder et al., 2011; Scholes and Paige, 2015). Among these, absence of light triggers an extra endoreplication cycle in Arabidopsis hypocotyls (Gendreau et al., 1997). Partial shading also affects the endoploidy level, as exemplified by the reduced DNA content in leaves of Arabidopsis plants grown under reduced light intensity, whereas an increased endoreplication in leaves has been observed under water deficit conditions (Cookson and Granier, 2006; Cookson et al., 2006). Endoreplication can also be triggered at biotic interaction sites, as observed upon symbiotic interactions with mycorrhizal fungi (Lingua et al., 2001) and nitrogen-fixing bacteria (Cebolla et al., 1999), and interactions with pathogens such as powdery mildew (Chandran et al., 2010) and nematodes

(de Almeida Engler et al., 2012). In such cases, endoploidy changes are probably triggered by an alteration of phytohormone balances, with auxin and jasmonate known to inhibit the mitosis-to-endocycle transition, and cytokinin promoting it (Ishida et al., 2010; Noir et al., 2013; Takahashi et al., 2013).

Although in recent years, many genes have been identified that control endoreplication onset and progression in plants, lack of a detailed knowledge of the temporal and spatial occurrence of endopolyploidy in an endoreplicating species has hampered the study of the physiological roles of the endocycle. In Arabidopsis, the endocycle is very common and endopolyploidization is seen during development of organs throughout its life cycle (Galbraith et al., 1991). However, in contrast to germline polyploidy, in which all cells within the organism possess the same DNA ploidy level, endopolyploidization does not occur in all cells equally, resulting in subpopulations of cells with different DNA content. A major open question is how cells and tissues with different endoploidy levels are integrated into a developing organ and how this organization contributes to plant growth under different environmental conditions. Protocols to quantify endoploidy include flow cytometry, DNA densitometry, and fluorescent in situ hybridization (Bourdon et al., 2011; Katagiri et al., 2016; Melaragno et al., 1993). Whereas flow cytometry allows rapid measurement of the ploidy level of large numbers of cells, it falls short in terms of spatial resolution. DNA densitometry and fluorescent in situ hybridization, encompassing microscopy and image analysis of fluorescent-stained nuclei through quantitative DNA dyes, provide detailed endoploidy maps of individual tissues but are rather laborious and low throughput.

In this work, we generated an Arabidopsis tissue- and endoploidy-specific transcriptome data set that was combined with previously published spatiotemporal root expression data to computationally predict a spatiotemporal endoploidy map for the complete root tip of Arabidopsis, which was subsequently validated experimentally. This map reveals that the endocycle is, to a remarkable extent, differentially controlled across root tissues and developmental stages, with a putative role in stress responses and the coordination of cell expansion.

RESULTS

Transcriptome Profiles of Arabidopsis Root Cortex Cells of Different Endoploidy

To assess how gene expression varies with the endoploidy level of cells, Arabidopsis root cortical nuclei were collected by

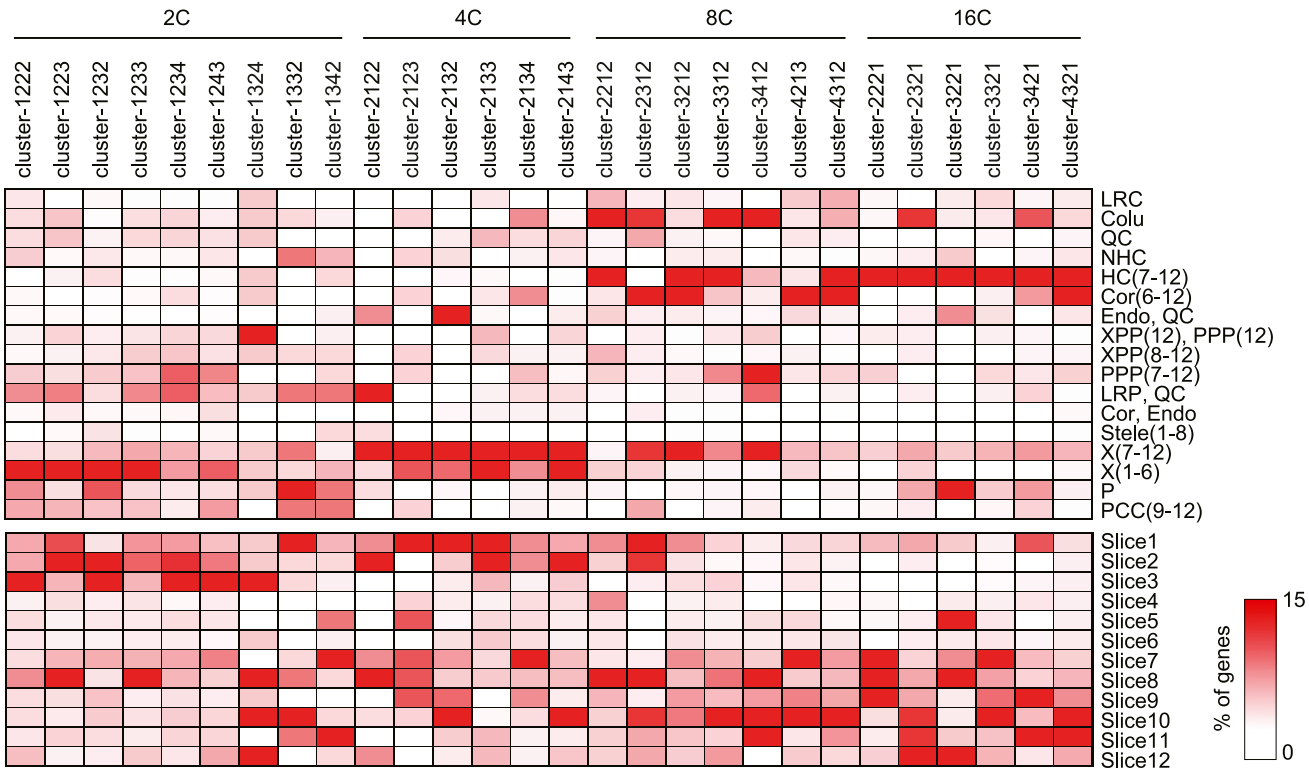

Figure 1. Peak Expression Distribution of Genes in 28 Clusters across Root Tissues and Sections.

Displayed is the proportion (%) of genes in each cluster that are peak-expressed in any given marker line (top) or at any given developmental stage (slices, bottom) of the Arabidopsis root. Only clusters with expression patterns peaking at a single endoploidy level (indicated on top) are shown. For marker lines that do not mark all developmental stages of a particular tissue, the slices marked are given in parentheses. Tissues are abbreviated as follows: lateral root cap (LRC), columella (Colu), quiescent center (QC), atrichoblast (NHC), trichoblast (HC), cortex (Cor), endodermis (Endo), xylem pole pericycle (XPP), phloem pole pericycle (PPP), lateral root primordia (LRP), xylem (X), phloem (P), and phloem companion cells (PCC).

fluorescence-activated nuclear sorting (FANS) based on DNA content (2C, 4C, 8C, and 16C) and subjected to transcriptome analysis (see Methods). In total, 3737 genes were found to be differentially expressed ($P < 0.05$, Benjamini-Hochberg FDR correction) across cortex cells of different endoploidy status (Supplemental Data Set 1). The expression profile of each gene was discretized into a pattern reflecting the expression level ranks across the four endoploidy levels, e.g., the pattern "1 2 3 4" indicates that a given gene is most highly expressed in 2C, then 4C, 8C, and 16C. Equal ranks were allowed, as, e.g., in the pattern "3 2 2 1" for genes that are most highly expressed in 16C, averagely expressed in 4C and 8C, and most lowly expressed in 2C (see Methods). Genes with the same rank pattern were grouped together, resulting in 70 clusters (Supplemental Data Set 2 and Supplemental Figure 1). Downstream analyses focused on the subset of 28 clusters with only one rank-1 expression value (clear peak expression at a single endoploidy level) and with a size of more than 10 genes (Supplemental Figure 2). The spatiotemporal peak expression of these 28 clusters was profiled based on available root cell type- (marker-) and developmental stage-specific gene expression data (Brady et al., 2007) (Figure 1). This analysis revealed that transcripts that exhibit peak expression in 2C or 4C are predominantly found in the meristematic zone (slices 1–3) of the root and in lateral root

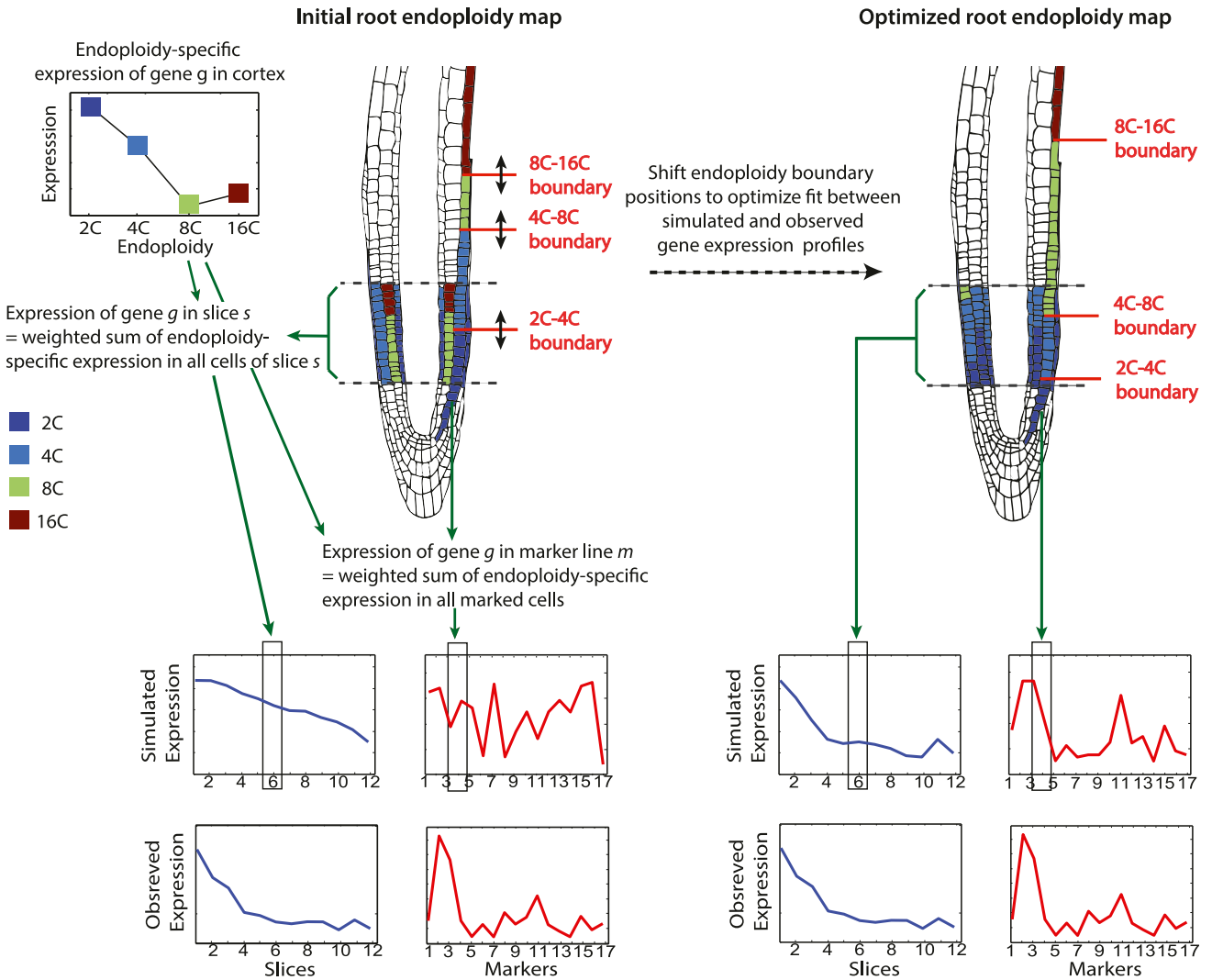

Figure 2. Schematic Representation of the Mathematical Modeling Approach Used to Predict the Root Endoploidy Map.

Shown is an example of how a root endoploidy map is predicted from the endoploidy-specific and spatiotemporal expression patterns of a single gene *g*. Initially, the endoploidy boundaries are randomly positioned on the map. Simulated annealing-based optimization is then used to shift endoploidy boundary positions along the longitudinal root axis and find the root endoploidy map that generates the best possible fit between the simulated and observed spatiotemporal expression profiles for the investigated gene (lower panels). The boxes in the simulated expression panels symbolize the slice *s* and marker line *m* for which root cells in the upper panels have been colored according to their predicted endoploidy levels (no real data). See Supplemental Movie 1 for a time-lapse movie of an example optimization run for a single gene.

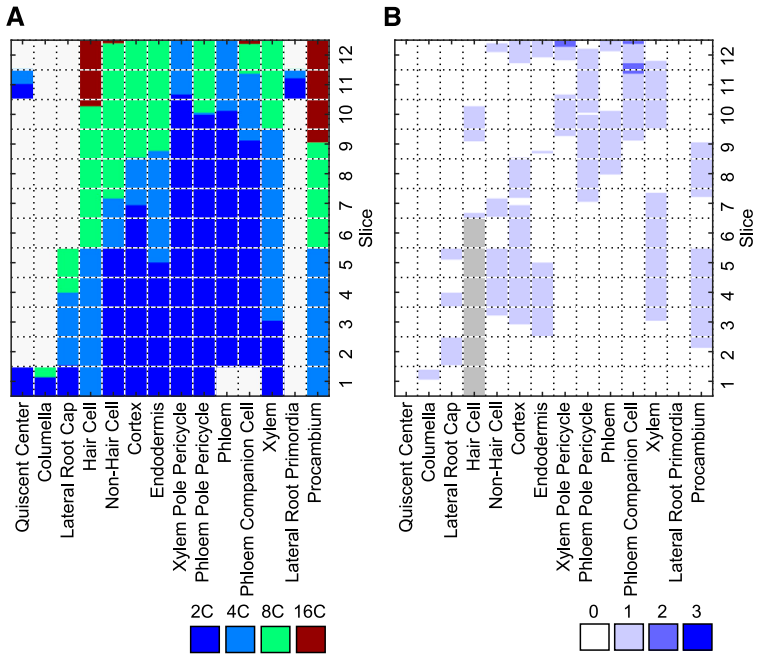

Figure 3. The Predicted Root Endoploidy Map and Experimental Validation.

(A) The endoploidy map of the Arabidopsis root using a balanced set of 332 marker genes.
(B) Difference map between the simulated endoploidy map in **(A)** and the map derived from flow cytometry measurements. Color hues indicate the difference in the number of endoreplication rounds between both maps for each cellular position. Hair cell slices 1 to 6 are unresolved in the flow cytometry map (gray).

primordia, phloem and developing and maturing xylem cells, whereas 8C- and 16C-peaking transcripts are primarily expressed in the transition and maturation zones (slices 7–12), and in columella, epidermal hair cells, cortex, and maturing xylem cells (Figure 1). Gene Ontology (GO) enrichment analyses showed that many of the clusters with peak expression at a particular endoploidy level are enriched in genes of particular function, e.g., several 4C-peaking clusters were found to be enriched in cell wall biogenesis genes, whereas several 8C- and 16C-peaking clusters were found to be enriched in genes involved in transport processes (Supplemental Data Set 3 and Supplemental Figure 3).

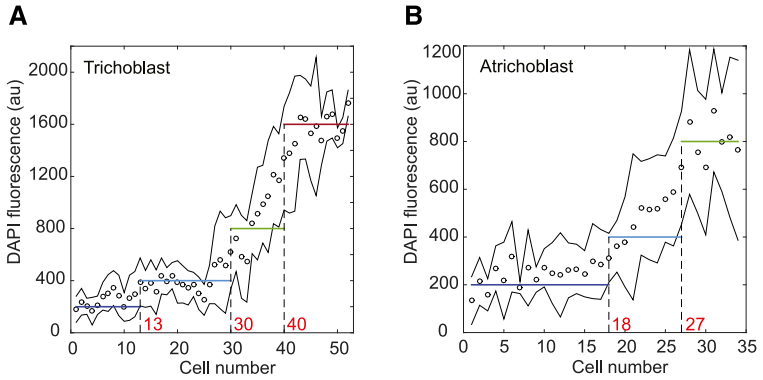

Figure 4. Experimental Mapping of Ploidy Borders in Atrichoblast and Trichoblast Cell Files.

DAPI fluorescence was measured in 12 trichoblast **(A)** and atrichoblast **(B)** cell files. The fluorescence measurements across cell files were normalized so that the average fluorescence of the last three measured cells in each cell file equals 1600 (for trichoblasts) or 800 (for atrichoblasts), and the average of the first three measured cells equals 200, reflecting the expected endoploidy levels at the beginning and end of the measured cell files. A model with 3 (4) endoploidy boundaries and average DAPI fluorescence levels of 200, 400, 800, and 1600 for 2C, 4C, 8C, and 16C cells, respectively, was fit to the atrichoblast (trichoblast) data using a simulated annealing optimization routine in Matlab R2014b. Open circles indicate the average measured fluorescence (after normalization) across all cell files at any given cell number, and the black contours indicate the corresponding standard deviations (standard deviations <10, caused by insufficient numbers of data points at a few early and late cell numbers, were manually set to 100 to avoid technical optimization problems).

A Virtual DNA Endoploidy Map of the Arabidopsis Root Tip

The observation that genes with peak expression at a specific endoploidy level in the cortex are expressed in specific spatiotemporal root zones and that those associations make sense from an endoreplication perspective (genes that are peak-expressed at higher endoploidy levels in the cortex data set are found to be spatiotemporally peak-expressed progressively further away from the root tip) suggests that the gene expression profile of a tissue at a specific developmental stage might be used to predict its endoploidy state. To construct a DNA endoploidy map of the Arabidopsis root tip, we devised a mathematical model that predicts the expression level of genes in 12 different root slices and 17 different tissue marker lines (covering 14 root cell types) as a function of their endoploidy-specific expression levels in the cortex (see Methods). The parameters in this model are the endoploidy boundaries along the longitudinal axis of the different root tissues, i.e., the 2C-4C, 4C-8C, and 8C-16C boundaries, the position of which is optimized to obtain the best possible fit between the predicted and experimentally observed gene expression levels (Figure 2; Supplemental Movie 1). Our model assumes that (1) the endoploidy levels in each cell type (arising through developmentally regulated endoreplication) exhibit a logical increment in DNA content over time (2C→4C→8C→16C from the root tip upwards along the longitudinal root axis), (2) the expression level of a given gene in a particular slice is a weighted sum of the expression levels of the gene in all cells of the different tissue types within that slice, and (3) the measured expression level of a gene in a particular tissue marker line is a weighted sum of the expression levels of the gene in all marked cells, which may have different endoploidy levels (Figure 1). In the model equations, relative gene expression levels in marker lines and slices are modeled as a weighted sum of the relative expression levels of the gene at different endoploidy levels in the cortex (see Methods). The weight of each endoploidy level in the gene expression model for a given slice/marker line is proportional to the number of cells of that endoploidy in the slice/marked tissues concerned, as determined by the tissue-specific endoploidy boundaries to be optimized.

A crucial assumption is that there exists a set of genes for which such a weighting scheme is valid, i.e., genes for which (1) the endoploidy-specific expression levels measured in the cortex can be used as a proxy for the endoploidy-specific expression levels in other tissues and (2) the spatiotemporal expression levels are primarily determined by the endoploidy state of the cell and not by tissue- or developmental stage-related factors that are independent of endoploidy. These genes can then be used as "markers" of the endoploidy state of cells. Two different kinds of genes can reasonably be considered potential markers: genes with a constant relative expression level across tissues at each endoploidy, and genes with a constant absolute expression level across tissues at each endoploidy. Assessing the model equations' validity for these two types of markers predicts that the model should work best for genes with a constant absolute expression level per endoploidy (see Methods).

To identify potential marker genes, the following selection criteria were used. First, genes were selected for high expression levels (>50th quantile, i.e., >73.17) and high endoploidy-specific expression variation in the endoploidy-specific cortex data set (SD/mean expression >50th quantile, i.e., >0.2), since genes that exhibit low endoploidy-specific expression variation (flat profiles in the cortex data set) are uninformative with regard to detecting endoploidy differences. This selection resulted in a reduced list of 4378 genes (Supplemental Data Set 4). Next, genes were selected based on whether or not their spatiotemporal expression pattern could be reliably predicted by the model from their endoploidy-specific expression levels in the cortex, both in terms of quantitative differences (χ^2 statistic R; see Methods) and the Pearson correlation coefficient between simulated and measured expression profiles (Supplemental Data Set 4). For both criteria, we again used the 50th quantile ($R < 372.5350$ and Pearson correlation coefficient > 0.39) as the selection cutoff, resulting in a further reduced set of 954 genes. Among these genes, 407 peak at 2C in the endoploidy-specific cortex data set, 162 at 4C, 302 at 8C, and 83 at 16C. Since employing unequal numbers of genes for the different endoploidy levels leads to endoploidy-specific biases in the model optimization runs and the resulting endoploidy map

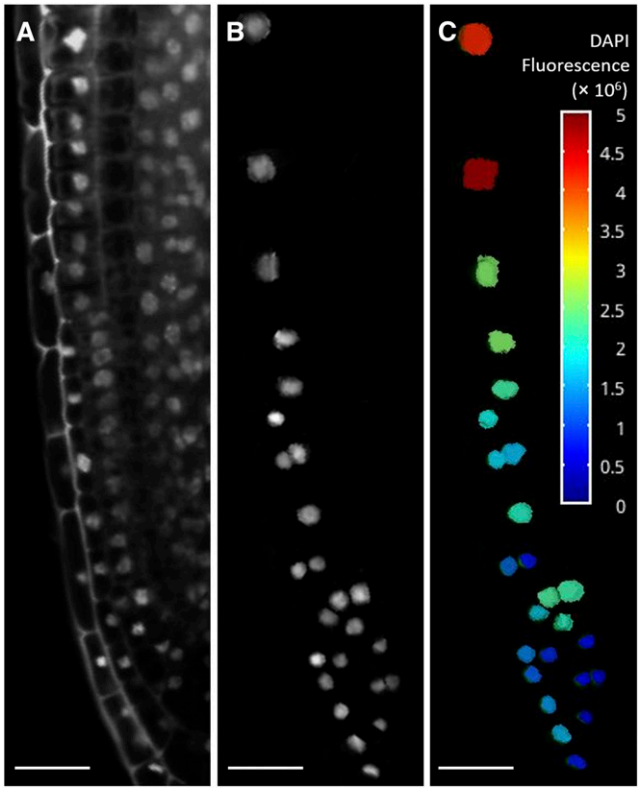

Figure 5. Ploidy Map of Lateral Root Cap Cells.

(A) Representative longitudinal mid-section of a DAPI-stained root meristem (of three independent roots imaged).

(B) Extraction of nuclear outlines of the lateral root cap cells. Note that this image has been processed using MorphoGraphX to highlight the LRC nuclei only (see Methods).

(C) DAPI fluorescence (arbitrary units) heat map. Nuclei fall into three categories likely corresponding to 2C (dark blue, $n = 7$), 4C (light blue-green, $n = 16$), and 8C (red, $n = 2$) nuclei. Bar = 20 μm. Data on independent roots are given in Supplemental Data Set 9.

(Supplemental Figure 4), we selected an equal number of genes from each class (the top 83 in terms of optimized *R* value) to obtain a final balanced set of 332 endoploidy marker genes to be used in the model (Supplemental Data Set 5). In accordance with the prediction that the model should work best for genes with constant absolute expression levels at each endoploidy, this balanced set of 332 genes was found to be significantly enriched in, among others, genes involved in chromatin modification (FDR-corrected P = 7.71E-10), DNA replication (P = 2.66E-09), and epigenetic regulation of gene expression (P = 1.95E-09) (Supplemental Data Set 6), gene categories that are arguably likely to exhibit absolute expression levels linked to DNA content.

The model was jointly optimized for all 332 marker genes to predict a DNA endoploidy map for the complete root tip. In total, 10 optimization runs were performed to optimize the endoploidy boundaries in the model (see Methods), of which four converged on Optimal Map I (Figure 3A; *R* = 63.3373), whereas six runs converged on Optimal Map II (Supplemental Figure 5A; *R* = 62.9712). While exhibiting largely the same endoploidy patterning and having comparable *R* values, the two maps exhibit substantial differences in the 4C-8C boundary position for non-hair cells and in the 2C-4C boundary position for phloem companion cells (Supplemental Figure 5B and Supplemental Data Set 7), next to small differences in a number of 2C-4C boundary positions (lateral root cap, cortex, endodermis, xylem pole pericycle, phloem pole pericycle, phloem, and lateral root primordia), 4C-8C boundary positions (lateral root cap, cortex, endodermis, xylem, phloem pole pericycle, phloem companion cell, and procambium), and 8C-16C boundary positions (non-hair cell and phloem companion cell). A breakdown of the contribution of each marker and slice to the total *R* value of both maps indicates that the observed boundary shifts reflect trade-offs in optimization of the model fit for particular slices and marker lines (Supplemental Data Set 8). In particular, the backward shift of the non-hair cell 4C-8C boundary from slice 7 to slice 9 in Optimal Map II, which is beneficial for the fit of the model to the *GL2* marker data, has a negative effect on the model fit for slices 7 to 9, which is only partially compensated for by (less detrimental) slight forward shifts of the cortex and endodermis 4C-8C boundaries into slice 8, and to a lesser extent of the xylem and procambium 4C-8C boundaries into slice 9. The reason for the phloem companion cell 2C-4C boundary shift between both maps is less obvious, although it also impacts slice 9. Compensatory small 2C-4C boundary shifts can be seen for the phloem and phloem pole pericycle in slice 10 and for the xylem pole pericycle in slice 11. In general, Optimal Map I performs better for seven slices and six markers, while Optimal Map II performs better for only two slices and five markers, despite having an overall lower (better) *R* score. Optimal Map I performs better for the combined set of slices, whereas Optimal Map II performs better for the combined set of marker lines, again indicating a slice/marker optimization trade-off. Experimental validation efforts (see below) suggest that Optimal Map I is closer to the true spatiotemporal endoploidy map, which is why this map is displayed in Figure 3A.

Validation of the Endoploidy Map Predictions

The endoploidy map predictions were experimentally validated in different ways. First, we measured the DNA content of cells of

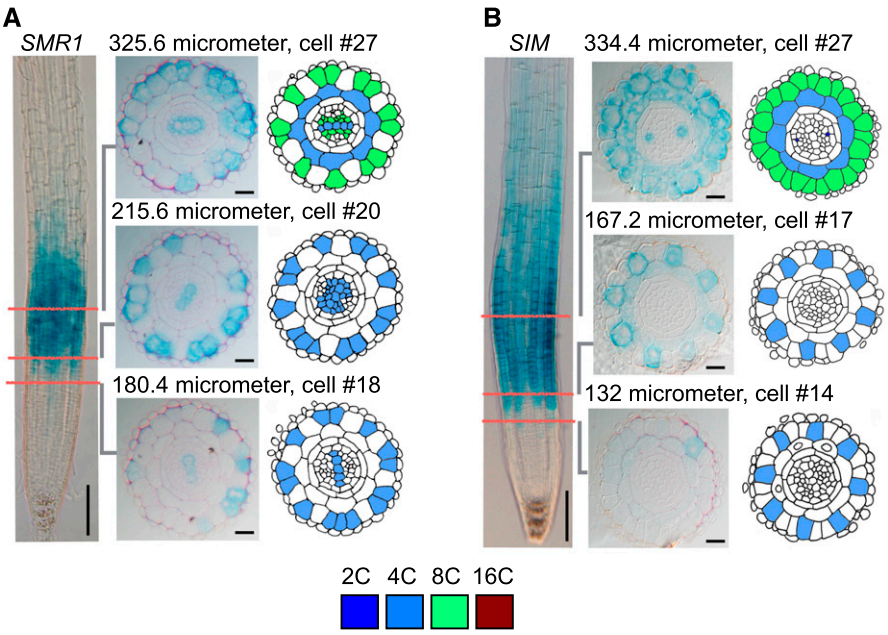

Figure 6. Temporal-Spatial Expression Patterns of the *SIM* and *SMR1* Endocycle Markers.

Cross sections of *SMR1:GFP-GUS* **(A)** and *SIM:GUS* **(B)** marker lines. The positions of the cross sections along the longitudinal root axis are indicated by red lines. The colored cross sections on the right depict the endoploidy levels predicted by the root endoploidy map shown in Figure 3A for the tissues marked by *SMR1* (atrichoblast, cortex, xylem, and procambium; **[A]**) or *SIM* (trichoblast, atrichoblast, cortex, and phloem; **[B]**). Predicted 4C and 8C cells are colored blue and green, respectively. Bars for total root and cross sections are 100 and 20 μm, respectively.

specific tissues in root tips using flow cytometry on cell type-specific GFP marker lines (Supplemental Figure 6), and a cytometry-based endoploidy map was constructed from the endoploidy-specific cell counts across all marker lines (see Methods and Supplemental Figure 7). The experimental and predicted maps present a very similar overall picture of spatiotemporal endoploidy organization, except for the location of the 2C-4C boundaries (Figure 3B), which is likely due to the fact that flow cytometry cannot distinguish mitotic G2 nuclei from G1 endoreplicating nuclei.

For the hair cells, the 2C-4C and 4C-8C boundaries in both the modeled maps and the cytometry-derived map are unreliable due to the late temporal expression of the underlying tissue marker gene (*COBL9*). To circumvent this problem, the DNA content progression in trichoblast cell files was mapped experimentally by 4′,6-diamidino-2-phenylindole (DAPI) staining and fluorescence microscopy. This technique mapped the 2C-4C boundary to the 13th hair cell on average (counted per file starting from the QC) and the 4C-8C transition to the 30th cell (Figure 4A). The 8C-16C boundary was found to be located around the 40th cell, in accordance with the model predictions. DAPI staining and fluorescence microscopy on atrichoblasts mapped the 2C-4C and 4C-8C boundaries to the 18th and 27th cell, respectively (Figure 4B), fitting well with the predicted borders of Optimal Map I (at the 18th and 25th cell, respectively) but less well with the atrichoblast 4C-8C border in Optimal Map II (at the 32nd cell).

In contrast to the majority of root tissues, in which all cells arise linearly from a group of stem cells that surround the quiescent center, the lateral root cap (LRC) consists of distinct layers that repeatedly displace the older LRC layers toward the root periphery. Within these layers, the cells divide and subsequently elongate, followed by their elimination through programmed cell death in the distal LRC (Fendrych et al., 2014). Because of this multilayered structure, there is no strict correspondence in the LRC between cell age and distance from the QC as for the other root tissues, partially violating our modeling assumptions. The spatiotemporal endoploidy structure of the LRC can therefore not directly be inferred from the modeled map, which is expected to only show the average endoploidy trend across LRC layers. To assess whether our model correctly predicts the presence of 2C, 4C, and 8C nuclei in the LRC, we quantified the nuclear content of LRC cells by DAPI staining. Quantification revealed that within a single root, nuclei fell into three distinct classes, likely corresponding to 2C, 4C, and 8C nuclei (Figure 5A; Supplemental Data Set 9). Moreover, the pattern of staining suggests a spatial organization mimicking that of the predicted map, with cells presumably reaching 8C endoploidy at the most distal part of the LRC (Figures 5B and 5C).

SIM and *SMR1* encode direct regulators of endoreplication onset (Churchman et al., 2006; Roeder et al., 2010; Yi et al., 2014). To validate the predicted order of endoreplication onset across different cell types, we examined sequential cross sections of *SMR1:GFP-GUS* and *SIM:GUS* endocycle onset marker lines, confirming that endocycle onset in xylem and atrichoblast cells precedes that of the cortex cells (Figure 6A), whereas trichoblast cells enter the endocycle before atrichoblast, cortex, and phloem tissues (Figure 6B), in accordance with the virtual endoploidy map predictions. The fluorescent marker of the *SMR1:GFP-GUS* line allowed the experimental mapping of endocycle onset within the atrichoblast and cortex cells to the 19th ± 1 and 29th ± 1 cell, respectively, corresponding well to the 2C-4C boundaries predicted by Optimal Maps I and II (at the 18th and 26th atrichoblast and cortex cell, respectively; Supplemental Data Set 10).

Identification of Genes Whose Transcript Level Is Most Strongly Linked to the Endoploidy Level of the Cell

Genes for which the expression levels across various root tissues and developmental stages are primarily a function of the endoreplication state of the cells concerned, rather than of tissue or developmental stage, are interesting candidates for being either tightly regulated by or regulating the endoreplication process. To identify the genes for which the measured spatiotemporal expression profile conforms most to the expression profile predicted by the root endoploidy map, we used our modeling approach in reverse. Given Optimal Map I as input, we predicted the spatiotemporal expression profiles of the 4378 genes in Supplemental Data Set 4 and ranked them in order of best fit between predicted and observed expression profiles (Supplemental Data Set 11). Not surprisingly, many of the top-ranked genes are part of the balanced 332 gene set used to build the virtual root endoploidy map. However, many other genes also show a spatiotemporal expression profile that closely matches the endoploidy-based

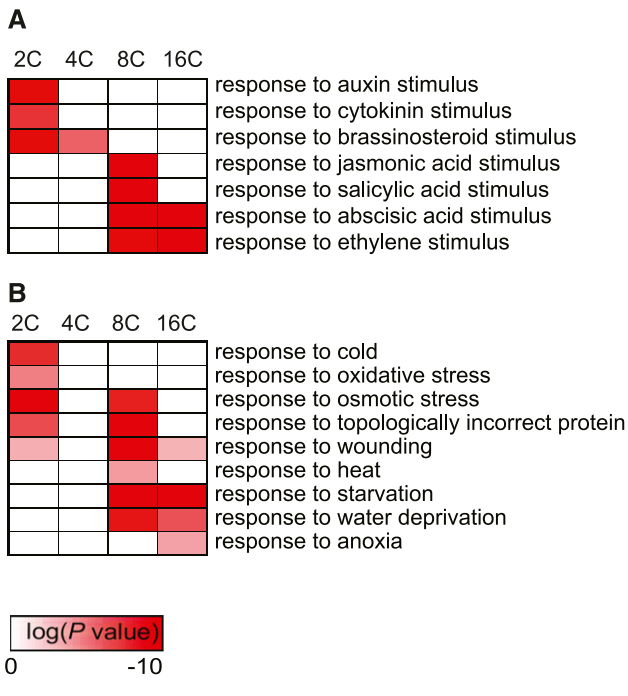

Figure 7. Functional Enrichment of Transcripts Peaking at a Given Endoploidy Level for Hormone and Stress Response GO Categories.

(A) Hormones.

(B) Stress conditions.

Cells are colored according to their enrichment P value after Benjamini and Hochberg false discovery rate correction (only for corrected P ≤ 0.05). Columns, endoploidy level; rows, GO categories.

predictions. GO enrichment analysis of the top 332 genes in the best-fit ranking reveals similar functional enrichments as in the balanced gene set, but generally stronger enrichments (Supplemental Data Set 12; e.g., chromatin modification FDR-corrected P = 1.25E-18, DNA replication P = 3.00E-15, epigenetic regulation of gene expression P = 7.50E-17). Interestingly, about one-third (104) of the top 332 genes are annotated in GO as "developmental process" genes (P = 4.14E-06).

A Simplified Model Predicts the Arabidopsis Root Endoreplication Response to Various Stresses

The extent of endoreplication is controlled by both environmental (Gendreau et al., 1998; Engelen-Eigles et al., 2001; Cookson and Granier, 2006) and endogenous factors, such as phytohormones (Ishida et al., 2010; Myers et al., 1990; Artlip et al., 1995; Valente et al., 1998). Accordingly, gene transcripts peaking at different endoploidy levels show distinct functional enrichment for hormonal and stress responses (Figure 7), suggesting that transcripts whose expression is primarily determined by the endoploidy state of the cell may be useful for predicting the impact of stress and hormonal treatments on the endocycle program during root development. A non-spatiotemporal version of the root endoploidy model (see Methods) was used to predict the impact on the root endoreplication state of 149 treatments for which gene expression data in root tips is publicly available (Supplemental Data Set 13). To accurately predict changes in the endoploidy distribution upon stress treatment, genes that exhibit stress-responsive expression changes that cannot be attributed to endoploidy changes were removed from the balanced set of 332 genes used in the spatiotemporal model version. To this end, we selected 26 out of the 149 treatments (stress versus control; Supplemental Data Set 14) that were generated in a similar experimental setup as used for the spatiotemporal root expression map (except for the stress treatment). Then, we used the non-spatiotemporal model to predict the expression under stress conditions for each of the 332 genes individually, optimizing the fit to the measured expression profiles under stress given the endoploidy-specific expression levels in the cortex data set, and we calculated the sum of squared errors (*SSQ*) between measured and modeled gene

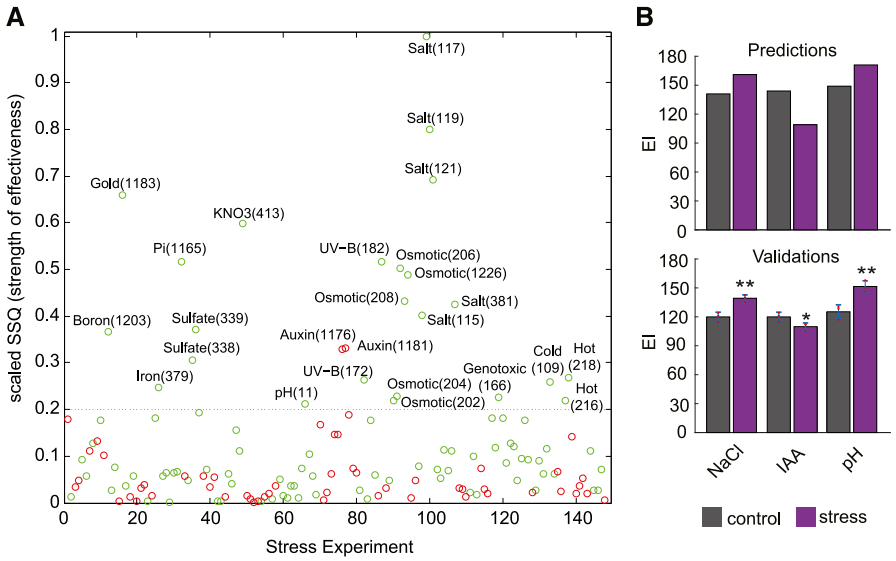

Figure 8. Predicted and Validated Effects of Stress on the Root Endoploidy Content.

(A) The predicted effect of stress conditions on root endoreplication levels. The *x* axis represents 149 publicly available stress/control expression profiling data sets, and the *y* axis depicts the sum of squares (SSQ) of the endoploidy weight differences in the stress and control endoploidy distributions (scaled to the range [0,1]). Red and green circles indicate that endoreplication is suppressed or promoted, respectively. The number between parentheses represents the stress condition index in Supplemental Data Set 12. The dotted line indicates a cutoff below which the conditions are not annotated on the panel. Condition keywords indicate: salt (117) = 150 mM NaCl, 6 h treatment; salt (119) = 150 mM NaCl, 12 h treatment; salt (121) = 150 mM NaCl, 24 h treatment; gold (1183) = 0.125 mM $KAuCl_4$, 6 h treatment; KNO_3 (413) = 5 mM KNO_3, 3.5 h treatment; Pi (1165) = Pi starvation, 6 h treatment; UV-B (182) = 15 min 1.18 W/m^2 Philips TL40W/12, 24 h after treatment; osmotic (206) = 300 mM mannitol, 12 h treatment; osmotic (1226) = 300 mM mannitol, 3 h treatment; osmotic (208) = 300 mM mannitol, 24 h treatment; salt (381) = 140 mM NaCl, 1 h treatment; salt (115) = 150 mM NaCl, 3 h treatment; sulfate (339) = sulfate limitation, 4 h treatment; boron (1203) = 5 mM boric acid, 12 h treatment; auxin (1176) = 1 μM IAA, 12 h treatment; auxin (1181) = 1 μM IAA, 8 h treatment; hot (218) = 38°C, 3 h treatment, 1 h recovery; UV-B (172) = 15 min 1.18 W/m^2 Philips TL40W/12, 0.5 h after treatment; cold (109) = 4°C, 24 h treatment; iron (379) = iron deficiency, 72 h treatment; osmotic (204) = 300 mM mannitol, 6 h treatment; genotoxic (166) = 1.5 μg/ml bleomycin + 22 μg/mL mitomycin, 12 h treatment; osmotic (202) = 300 mM mannitol, 3 h treatment; pH (11) = pH 4.6, 6 h treatment; hot (216) = 38°C, 3 h treatment, 0 h recovery. The dotted line indicates a cutoff below which the conditions are not annotated on the panel.

(B) Predicted and validated endoreplication indices (EI) under stress conditions. Five-day-old seedling were transferred to either control medium or medium holding salt (140 mM NaCl), auxin (1 μM IAA), or low pH (4.6). Ploidy levels were measured in 5-mm root tips 48 h after treatment. Data represent mean ± SD (2 replicates with >200 seedlings per replicate used; *P < 0.05, two-sided Student's *t* test).

expression for each gene across the 26 treatments (Supplemental Figure 8). The nine genes whose stress-responsive expression levels were worst-predicted based on their non-stress endoploidy-specific cortex expression levels were removed from the balanced 332 gene set. These genes (indicated in green in Supplemental Data Set 5) are mainly annotated in the GO database as responsive to jasmonic acid stimulus, wounding, or salt stress. The remaining set of 323 genes was used for predicting endoploidy distributions under all 149 treatments and the corresponding control conditions, and endoploidy distribution changes under treatment versus control conditions were assessed.

The model predicted that the phytohormone auxin negatively controls the extent of endoreplication (Figure 8A), which was confirmed experimentally (Figure 8B) and is in agreement with previous reports (Ishida et al., 2010). Positive effects on endoreplication are predicted for treatments related to macronutrients (such as phosphate and sulfate deficiency), micronutrients (including iron starvation and elevated levels of boron), and environmental stresses (including temperature, DNA stress, and salinity) (Figure 8A). Measurements of the root endoreplication index under salt stress and low pH conditions confirmed the model predictions (Figure 8B).

A Link between Endoreplication and Cell Volume (or Wall Structure)

Endoreplication is generally thought to drive growth. Many cell wall biogenesis genes were found to be peak-expressed in 4C cortex cells in wild-type Arabidopsis plants (Supplemental Figure 3), suggesting that endoreplication-driven growth might be associated with cell wall modifications. To study the role of endoreplication in cell growth and cell wall processes in more detail, we introgressed the *SMR1:GFP-GUS* marker in the *smr1* mutant background. Subsequently, we first mapped the nuclear and cellular size of atrichoblast cells in both genotypes in relation to the expression of the *SMR1:GFP-GUS* marker. In the control plants, *SMR1* expression appeared before the onset of cell expansion (Figure 9; Supplemental Figure 9A). Likewise, expression preceded the increase in nuclear size (Supplemental Figure 9B). By contrast, *smr1* cells displayed no significant increase in nuclear size, corresponding with their endoploidy phenotype, but still underwent rapid cell expansion (Supplemental Figure 9), likely driven by turgor generated by the expanding vacuole. To map the covariates of this altered nuclear-to-cellular volume ratio at the transcript level, we compared the transcriptome of SMR1:GFP-GUS fluorescent protoplasts isolated from wild-type versus *smr1* mutant roots. In total, 40 genes were found to be differentially expressed (FDR < 0.05) (Figure 10; Supplemental Data Set 15). Among the 29 genes downregulated in wild-type versus *smr1* plants, six histone and two histone modifying genes were found, which might reflect extensive chromatin reorganization coinciding with endocycle onset, as reported recently (Otero et al., 2016). Also, *CDKB1;1* was found to be downregulated in the wild type, fitting with its role as suppressor of the endocycle onset (Boudolf et al., 2004). At a higher FDR, another five chromatin genes and *CDKB1;2* were found to be differentially expressed as well (Figure 10).

Among the 11 genes upregulated in wild-type plants versus *smr1*, three cell wall biosynthesis genes were identified, including *GUX5* and *TBL41*. Although neither have yet been characterized yet, they are homologous to genes that add glucuronic acid (*GUX1*

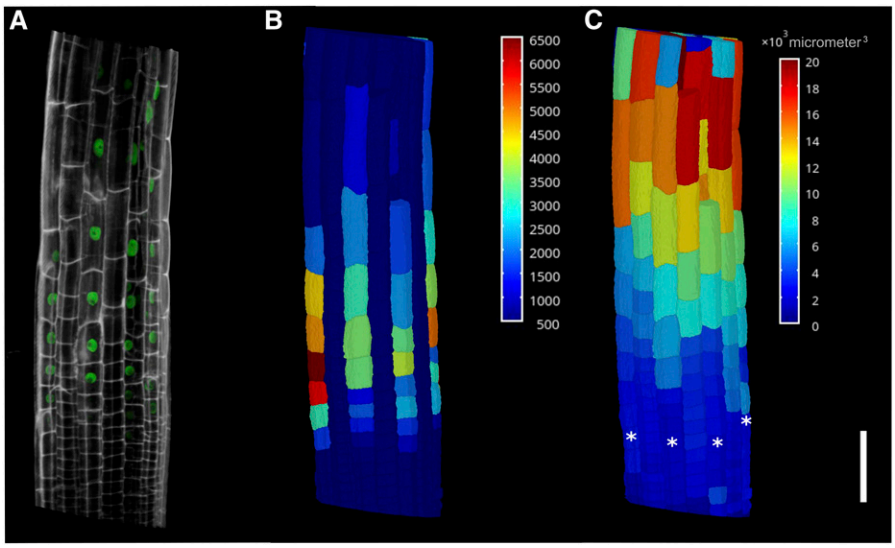

Figure 9. *SMR1* Expression Precedes Cell Expansion.

(A) Representative confocal image of a *SMR1:GFP-GUS* root counterstained with propidium iodide, displaying atrichoblast expression.
(B) *GFP* expression heat map.
(C) Cell volume heat map. The asterisks indicate the position of the first GFP signal. Note that for the purpose of visualization, MorphographX was used to digitalize the images. The confocal image was converted into a working stack to edit the stacked images, deleting the layers over and under the epidermis to enable visualization of the boundaries between epidermal cells and the presence of GFP. Bar = 50 μm.

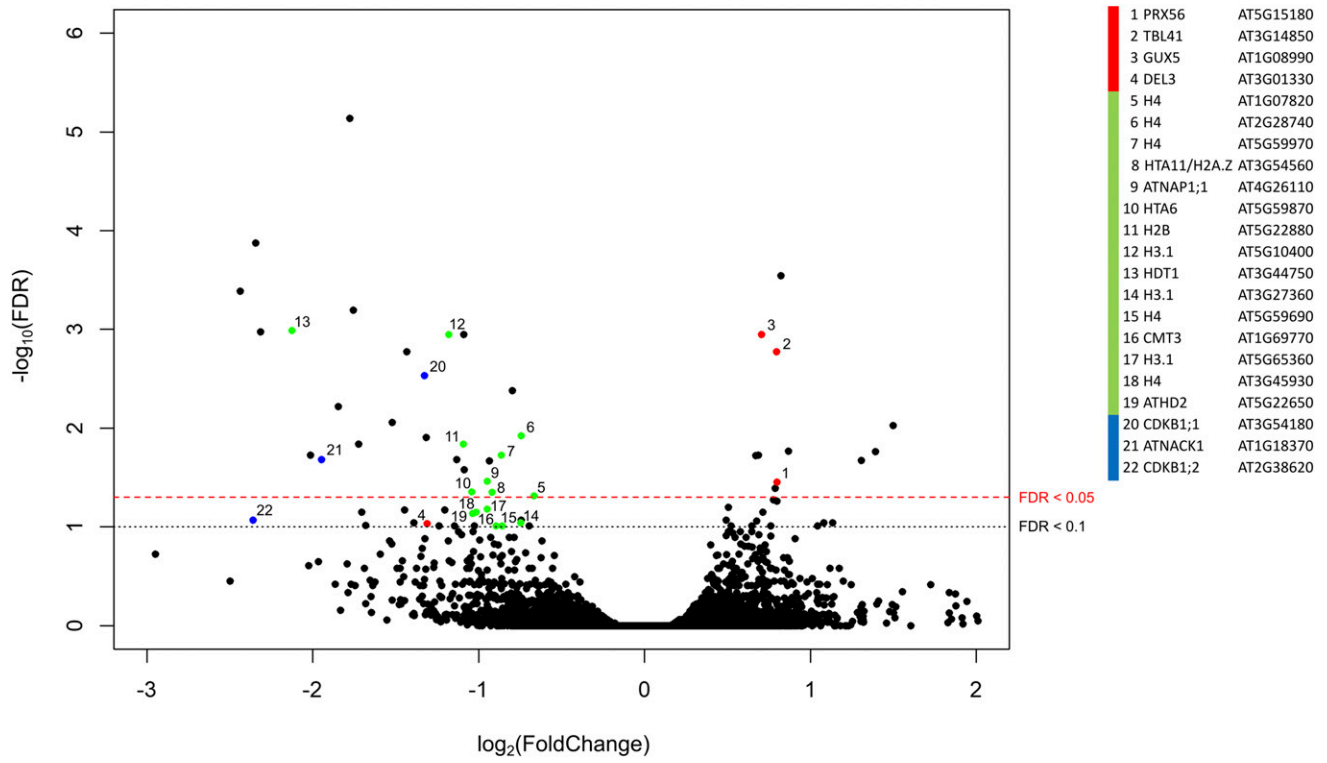

Figure 10. Volcano Plot Showing Transcripts That Are Up- and Downregulated in SMR1:GFP-GUS Fluorescent Protoplasts of Control (Col-0) versus *smr1* (Col-0) Plants.

Red, green, and blue dots correspond to cell wall modifying, chromatin-linked, and cell cycle genes, respectively. Gene IDs are given at the right.

and *GUX2*) or acetyl groups (*TBL27* and *ESKIMO1*) to xylans, respectively (Bromley et al., 2013; Gille et al., 2011; Xiong et al., 2013; Yuan et al., 2013). The pattern of xylan substitution determines whether or not xylan interacts with cellulose (Grantham et al., 2017) and hence determines the cell wall flexibility and strength. Strikingly, we also identified the atypical E2F transcription factor gene *DEL3*/*E2Ff* as downregulated in wild-type plants, although at a higher FDR threshold (Figure 10). This transcriptional repressor has been demonstrated to restrain the growth of postmitotic cells by inhibiting the expression of cell wall-modifying genes, including expansins (*EXP3*, *EXP7*, and *EXP9*) and an UDP-glucose-glycosyl transferase (Ramirez-Parra et al., 2004)

Because xylans play a major role in strengthening cell walls (Scheller and Ulvskov, 2010), we took a closer look at cell wall phenotypes in Col-0 and *smr1* roots, comparing radial sections through the root elongation zone of wild-type and mutant plants. Although no striking differences in tissue organization or cell wall thickness were observed, *smr1* trichoblast cells were found to be more radial in shape (Figures 11). A similar phenotype was observed in the independent *lgo-1* mutant (Supplemental Figure 10), although less strongly, which may be due to *lgo-1* being a weaker allele in a different ecotype (Roeder et al., 2010). Together with the cell volume effect, these data suggest that the endoreplication cycle may play a role in coordinating turgor-driven rapid cell expansion with adequate production of specific cell wall compounds. Given the substantially thicker exterior cell wall of the outer tissues (Dyson et al., 2014), a link between endoreplication and cell wall biogenesis may also help explain the overall outwardly increasing endoploidy levels observed in the Arabidopsis root tip (Figure 3A).

DISCUSSION

In this study, we predicted an endoploidy map of the developing Arabidopsis root tip from endoploidy-specific gene expression profiles in the cortex and existing spatiotemporal gene expression data sets of the Arabidopsis root (Brady et al., 2007). Flow cytometric experiments on GFP marker lines (Supplemental Figure 6) produced an overall similar endoploidy map of the root tip, except that the 2C-4C boundaries could not be resolved because cytometry cannot distinguish G2 phase nuclei from cells that underwent one endocycle. Fluorescence microscopy on the DAPI-stained wild-type root epidermis and on cortex and atrichoblast cell files in a *SMR1:GFP-GUS* marker line confirmed that the modeled endoploidy map predicts 2C-4C boundaries to within a few cells from their true location in atrichoblasts and cortex (Supplemental Data Set 10). Interestingly, not only 2C-4C boundaries but also higher-level endoploidy boundaries appear to be more accurately predicted by the model than by the cytometry-derived map. Fluorescence microscopy on DAPI-stained roots revealed that both the 4C-8C boundary in atrichoblasts and the

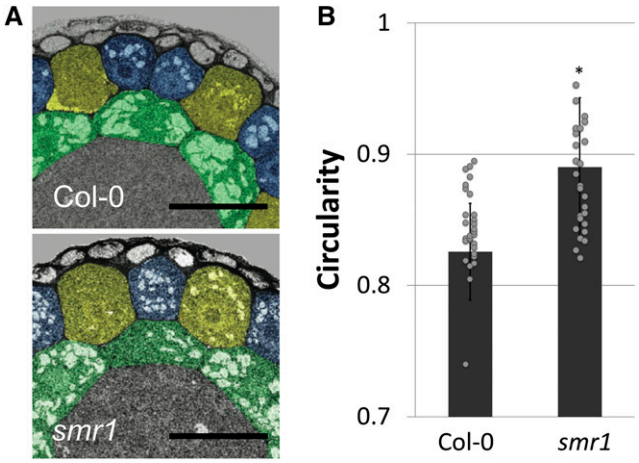

Figure 11. *smr1* Mutant Cells Display Altered Shape.

(A) Serial block-face scanning electron microscopy radial sections through the elongation zone of Col-0 and *smr1* mutant plants. Trichoblasts, atrichoblasts, and cortex cells are false-colored in yellow, blue, and green, respectively. Bar = 25 μm.

(B) Circularity of control (Col-0) and *smr1* trichoblast cells within the transition zone. Data represent mean ± SE ($n > 20$, *P ≤ 0.05, two-sided Student's *t* test).

8C-16C boundary in trichoblasts align more closely with the model predictions than with the cytometry-derived map (Supplemental Data Set 10). This may point to the existence of certain biases in the FANS-sorting efficacy for cells of different endoploidy levels within a given tissue. Although sorted cells of comparable lines were also used to generate the tissue expression data (Brady et al., 2007) used in the model, the inclusion of root slice expression data from sorted protoplasts (Brady et al., 2007) in the model appears to compensate for such biases.

Next to the fact that the model outperforms experimental approaches such as flow cytometry (in terms of accuracy) and fluorescence microscopy (in terms of feasibility for inner tissues) for mapping the spatiotemporal distribution of root endoploidy levels, the modeling approach also offers other advantages. Given the established root endoploidy map as input, we used the model in reverse to assess which genes exhibit spatiotemporal root expression patterns that can be predicted accurately by the root endoploidy patterning. In line with the prediction that the model should primarily identify genes that have a constant absolute expression level per cell at each endoploidy, the top of the list contains many chromatin modifiers and cell cycle genes (Supplemental Data Sets 11 and 12). Other than these, many regulatory and developmental genes are also found to exhibit expression patterns that are linked to the endoreplication state of the cell, offering inroads to study how endoreplication patterns are linked to root physiology.

The model also allowed us to predict the endoreplication response to stress treatments in Arabidopsis roots (Figure 7) from publicly available expression data. Interestingly, some degree of endoreplication response was predicted for the majority of treatments in the expression compendium that we analyzed. The strongest response was predicted for salt, osmotic, and temperature stress, various nutrient and DNA stresses, and auxin treatment. Several of these predictions (for salt, pH, and IAA treatments) were qualitatively validated, but more work is needed to unravel exactly how spatiotemporal endoreplication patterns change under various stresses and whether and how these pattern changes are linked to stress adaptation. Endoploidy pattern changes under stress might in some cases rather be associated with developmental and root morphology changes under stress than be a direct response to the stress treatment. On the other hand, in some cases, developmental and endoploidy changes might be inextricably linked. Regardless, our model can be considered a tool to scout gene expression databases for treatments that might provoke an interesting endoreplication response, which can then be studied in more detail to unravel the physiological roles of endoreplication under stress.

Most importantly, the obtained endoploidy map details the extent and order of endoreplication among different root tissue types, revealing that endoreplication is under strict spatiotemporal control. The predicted endoploidy levels appear to be associated primarily with the radial organization of the root, with the outer tissue layers displaying higher endoploidy levels than the inner tissues. Xylem pole pericycle cells, but not phloem pole pericycle cells, are predicted to undergo endoreplication only after lateral root primordia have been established, in accordance with their role as an "extended meristem" from which lateral roots emerge (Casimiro et al., 2003). Remarkably, epidermal hair cells undergo one more endocycle than the non-hair epidermal cells in the root tip. The predicted 16C value for the hair cells in the maturation zone confirms what has been experimentally detected before (Sliwinska et al., 2012; Sugimoto-Shirasu et al., 2005) and suggests the need of the third endocycle to support hair tip growth. However, when studying the endoploidy level in a number of root hair mutants, no direct correlation between the hair tip size and endoploidy level was observed (Sliwinska et al., 2015); therefore, the exact role of the third endocycle in the process of hair development still needs to be unraveled. A putative scenario might be that the high ploidy level is required for an increased metabolic output required for tip outgrowth. Root hairs have been found to express unique cell expansion and cell wall-modifying enzymes, including specific expansins and pectin-interacting proline-rich proteins (Bernhardt and Tierney, 2000; Boron et al., 2014; Cho and Cosgrove, 2002; Vissenberg et al., 2001).

Among vasculature tissues, xylem is predicted to undergo endoreplication much earlier than other tissues, resulting in a deviation from the radial endoploidy pattern. Given that the mature xylem cells possess a thick secondary cell wall providing mechanical strength for upward transport of water and nutrients, the early transition of xylem cells to 4C suggests that there may be a functional link between endoreplication and cell wall architecture. Strikingly, genes with peak expression in 4C cortex cells display a functional enrichment for cell wall-related processes, including "cell wall organization or biogenesis" and "xylem development" (Supplemental Figure 3), whereas among the genes being differentially expressed in *SMR1:GFP-GUS* positive cells several cell wall-modifying genes were found to be upregulated in wild-type versus *smr1* mutant roots (Figure 10). One of these (*TBL41*) is homologous to *ESKIMO* (*ESK1*). ESK1 possesses xylan *O*-acetyltransferase activity and its mutation results in collapsed

xylem vessels (Lefebvre et al., 2011). Recently, a functional link between endopolyploidy and ESK1 activity has been suggested, as a suppressor mutant screen using *esk1-5* dwarf plants resulted in the isolation of a mutant *KAKTUS* allele, fully recovering the xylem morphology and its hydraulic conductivity (Bensussan et al., 2015). *KAKTUS* encodes a suppressor of endoreplication (El Refy et al., 2003; Perazza et al., 1999), and its mutation results in increased endoploidy levels.

A putative role for endocycle onset in control of cell wall modifications might also explain reports on the observed pathogen susceptibility of endocycle mutants, such as *smr1* and *sim smr1* (Hamdoun et al., 2016; Wang et al., 2014). On the other hand, ectopic overexpression of the endocycle-promoting *UVI4* and *OSD1* genes results in enhanced disease resistance (Bao and Hua, 2014; Bao et al., 2013). We speculate that endocycle-driven cell wall modifications (or lack of such modifications) might impact the success of pathogen penetration. A link between endoploidy and cell wall composition is also suggested by the phenotypes of plants mutated in the *CONSTITUTIVE EXPRESSION OF PATHOGENESIS-RELATED GENES5*. These mutants display trichomes with reduced DNA content, linked to the presence of a thinner cell wall and reduced cellulose content (Brininstool et al., 2008; Kirik et al., 2001).

In plant cells, vacuolar expansion rather than an increase in cytoplasmic volume is believed to be the major force driving cell growth (Sugimoto-Shirasu and Roberts, 2003). Correspondingly, *smr1* mutant cells still expand in the absence of endocycle onset (Supplemental Figure 9). Based on the observations that endoreplication onset precedes cell elongation (Figure 9; Hayashi et al., 2013), that *smr1* trichoblast cells exhibit enhanced radial symmetry (Figure 11), that particular cell wall-related gene classes, including xylan biosynthesis genes, are overrepresented among the genes exhibiting peak expression in 4C cells (Supplemental Figure 3), and that several cell wall-related genes are downregulated in the *smr1* mutant, we put forward the hypothesis that endoreplication might prepare the cell to cope with the turgor-driven rapid cell expansion by inducing appropriate cell wall modifications. This might be of particular importance for tissues holding extremely rapidly expanding cells, where an increase in the gene copy number might be a way to cope with the high demand for new cell wall materials. Correspondingly, Bourdon et al. (2012) have demonstrated a clear link between cellular endoploidy and the level of gene transcription. How the endocycle specifically controls expression of cell wall biosynthesis and cell wall-modifying genes remains to be addressed, although the high number of chromatin-related genes being downregulated in wild-type versus *smr1* plants upon activation of the *SMR1* promoter suggests an epigenetic component.

Besides sensitivity to pathogens, mutant plants with reduced endoreplication levels typically do not display any obvious plant phenotype, suggesting that at least under normal growth conditions a loss of coordination between rapid cell expansion and gene transcription might still be fairly tolerable to the plant. Nevertheless, our stress modulation data suggest that the endocycle is stimulated by many different stresses (Figure 8), including salinity and boron, which are both known to affect the plant cell wall (Tenhaken, 2015). It is therefore conceivable that the changes to the normal endoreplication pattern that are observed under stress conditions may also be related to cell wall adjustments to mitigate the stress. In any case, the observed changes in root tip endoploidy patterning in response to several stresses suggest that endoreplication pattern changes may play a role in adaptive stress responses, which may explain why the endocycle is mainly observed in species growing in variable environments (Barow, 2006; Scholes and Paige, 2015). However, these speculations need to be further substantiated.

METHODS

Plant Lines and Growth Conditions

Endoploidy-specific gene expression profiles were obtained from sorted nuclei of the cortical cells of the *Arabidopsis thaliana pCO2:YFP-H2b* line (Heidstra et al., 2004). Tissue-specific endoploidy measurements were obtained via flow cytometric analysis of the 13 Arabidopsis marker lines listed in Supplemental Data Set 16 (Brady et al., 2007; Dietrich et al., 2017). Endoploidy boundary positions were confirmed using *SIM:GUS* and *SMR1:GFP-GUS* reporter lines (Yi et al., 2014). The *sim*, *smr1*, and *lgo-1* mutant lines have been described by Churchman et al. (2006) and Roeder et al. (2010), respectively.

Seeds were surface-sterilized using a solution of 20 parts by volume of commercial bleach and 80 parts by volume of 100% ethanol and then washed twice with 100% ethanol. The dried seeds were germinated vertically on plates containing half-strength Murashige and Skoog (MS) medium (Murashige and Skoog, 1962), 1% sucrose, and 0.5 g/L MES (pH 5.7) in 1% agar, except for the *SMR:GFP-GUS* sorting experiment in which sucrose was omitted from the medium. Plants were grown under long-day conditions (16 h light, 8 h darkness) at 22°C. For pH, salt, and IAA treatment experiments, plants were grown on a layer of nylon strip embedded on the agar surface. Five days after germination, the nylon filters were transferred for 48 h onto control or treatment media, followed by sampling of 5-mm root tips for flow cytometric analysis. Low pH (4.6), high salt (140 mM NaCl), auxin (1 μM IAA), and respective MS standard media were prepared as described by Dinneny et al. (2008), Iyer-Pascuzzi et al. (2011), and Lewis et al. (2013), respectively.

Endoploidy-Specific Microarray Data Acquisition

Roots of 5-d-old *pCO2:YFP-H2B* (Col-0 ecotype) plants grown under continuous light conditions at 22 to 23°C were excised using a razor blade just below the hypocotyl region. Samples of combined root material (10 g fresh weight) were collected in a glass Petri dish and chopped with slicing action after adding 10 mL nuclear isolation buffer (45 mM $MgCl_2$, 30 mM sodium citrate [trisodium], and 20 mM MOPS, pH adjusted to 7.0). The solution was collected into a 15-mL tube and centrifuged at 2500 rpm in a Sorvall swinging bucket AH-3000 at 4°C for 8 min. The pellet was drained to ~0.5 mL and resuspended in nuclear isolation buffer to 4 mL, then transferred to a 30-μm strainer in a 5-mL tube and DAPI was added (20 μL/mL 0.1 mg/mL stock). Biparametric sorting was then done based on YFP fluorescence and nuclear DNA content, as previously described (Zhang et al., 2005). Sorting of isolated nuclei was done using a Dako-Cytomation MoFlo flow cytometer/cell sorter as described (Zhang et al., 2008). The nuclear RNA was extracted from each nuclear DNA content population (2C, 4C, 8C, and 16C, ~0.2 mL [200,000 nuclei]/0.95 mL RLT) using the Qiagen RNeasy kit according to the manufacturer's instructions. Prior to Affymetrix ATH1 array hybridization, two consecutive rounds of RNA amplification were done in the VIB Nucleomics core, using standard Affymetrix protocols for small samples. The amplified nuclear RNA of each DNA content class was used for microarray analysis.

Microarray Data Normalization and Exploratory Analyses

The raw endoploidy-specific microarray data were preprocessed using the robust multichip average (RMA) normalization approach (background correction, quantile normalization, and summarization) implemented in the Bioconductor R package, version 2.5 (Irizarry et al., 2003; Gentleman et al., 2004). The Bioconductor package Limma (Smyth, 2004) was used to identify differentially expressed genes. Pairwise comparisons between any two endoploidy levels were performed using moderated *t* statistics and the eBayes method as implemented in Limma (Smyth, 2004). P values were corrected for multiple testing using the Benjamini-Hochberg method (Hochberg and Benjamini, 1990) at an FDR threshold of 0.05.

The expression profile of each gene was discretized into a pattern reflecting the expression level ranks across the four endoploidy levels. To allow for patterns with equal ranks among two or more endoploidy levels (e.g., "3 2 2 1" for genes that are most highly expressed in 16C, averagely expressed in 4C and 8C, and most lowly expressed in 2C), a criterion is needed to decide whether two expression levels are the same or different. As differential expression calls proved to be not sensitive enough to discriminate more than a few of the possible rank patterns, we instead discretized expression profiles using the square root of the averaged expression values as a yardstick. The following protocol was followed to discretize expression profiles into rank patterns: (1) Rank endoploidy-specific expression levels from high to low (indexed by *i*). (2) The first (highest) expression level ($i = 1$) gets rank $r = 1$ and becomes the reference level. Augment *i*. (3) The next highest expression level *i* is compared with the reference level. There are two possibilities: (a) The expression difference between both levels is less than the square root of the reference level. Expression level *i* gets the same rank *r* as the reference level, and the reference level remains unchanged. Augment *i* and repeat step 3. (b) The expression difference between both levels is equal or more than the square root of the reference level. Augment *r*. Expression level *i* is assigned the augmented rank *r* and becomes the new reference level. Augment *i* and repeat step 3. (4) Reorder the resulting discretized expression patterns to match the original endoploidy sequence 2C-4C-8C-16C.

A spatiotemporal root gene expression data set (Brady et al., 2007) was used together with the endoploidy-specific cortex data set in mathematical model I (see below). To this end, the raw data from both data sets were jointly RMA normalized as described above. The untransformed (i.e., non-log-scale) expression values were used in the model. For mathematical model II (see below), raw microarray data sets for stress experiments on whole roots were obtained from CORNET (De Bodt et al., 2010, 2012) and the GEO repository (Edgar et al., 2002) (Supplemental Data Set 13). These data were RMA normalized together with the endoploidy-specific cortex data as described above. Again, untransformed expression values were used in the mathematical model.

GO Enrichment Analyses

Functional enrichment results reported in this study were calculated with the BiNGO tool (Maere et al., 2005) using hypergeometric tests and Benjamini-Hochberg (Benjamini and Hochberg, 1995) multiple testing correction at FDR = 0.05. GO information and annotations for Arabidopsis were obtained from the GO database (www.geneontology.org; version downloaded on 4/22/2015). The 19,937 genes in the preprocessed Affymetrix ATH1 microarray data set were used as the reference set for the GO enrichment analyses presented in Supplemental Figures 4 and 10 and Supplemental Data Set 3. The 4378 genes listed in Supplemental Data Set 4 were used as the reference set for all other analyses.

RNA-Seq Analysis of Fluorescence-Sorted Protoplasts

Protoplasts were prepared as described by Birnbaum et al. (2003). Approximately 10,000 seeds (per replicate) of Col-0 and *smr1* plants holding the *SMR1:GUS-GFP* construct were sown on top of nylon mesh on square plates holding 1% agar and 0.5× MS medium. Five days after sowing, 0.5-cm root tips were isolated and placed at room temperature in 30 mL protoplasting solution B [solution B = (solution A + 1.5% cellulase and 0.1% pectolyase), solution A = (600 mM mannitol, 2 mM $MgCl_2$, 0.1% BSA, 2 mM CaCl, 2 mM MES, and 10 mM KCl, pH 5.5)]. After 50 min, the protoplasting solution was filtered using a 40-μm strainer into a 50-mL conical tube. Twenty milliliters of solution B was added to the residual roots to wash out the remaining protoplasts and added to the 50-mL conical tube. Samples were centrifuged at 4°C for 8 min at 200g. The supernatant was discarded and resuspended in 1 mL solution A. Then the samples were sorted with a BD FACSAria II cell sorter (85-μ nozzle, 45 p.s.i.) following a previously described protocol (Brady et al., 2007). The samples were sorted within 15 to 20 min of finishing the protoplasting step, and a period of 30 min was needed for collecting 120,000 GFP-positive protoplasts (per replicate). Samples were collected in 1.5-mL tubes containing 300 μL RLT buffer (Qiagen) supplemented with 3 μL β-mercaptoethanol and immediately frozen in liquid nitrogen to be stored at −70°C. RNA was extracted using the Qiagen RNeasy Micro Kit. RNA concentration and purity were determined spectrophotometrically using a Nanodrop ND-1000 (Nanodrop Technologies), and RNA integrity was assessed using a Bioanalyzer 2100 (Agilent). Per sample, an amount of 250 ng of total RNA was used as input. Using the Illumina TruSeq Stranded mRNA sample prep kit (protocol version: part no. 15031047 rev. E, October, 2013), poly(A)-containing mRNA molecules were purified from the total RNA input using poly-T oligo-attached magnetic beads. In a reverse transcription reaction using random primers, RNA was converted into first strand cDNA and subsequently converted into double-stranded cDNA in a second-strand cDNA synthesis reaction using DNA Polymerase I and RNase H. The cDNA fragments were extended with a single "A" base to the 3′ ends of the blunt-ended cDNA fragments, after which multiple indexing adapters were ligated, introducing different barcodes for each sample. Finally, enrichment PCR was performed to enrich those DNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the library. Sequence libraries of each sample were equimolarly pooled and sequenced on an Illumina NextSeq 500 instrument (medium output, 75 bp, paired end, v2)

Transcript quantification was performed by Salmon version 0.7.2 (Patro et al., 2017), using the TAIR10 list of all coding sequences. The option to correct for sequence-specific bias was turned off and all parameters were set to their default values. Transcript quantification results generated by Salmon were corrected for gene length variations across the samples using the tximport 1.6.0 R package (Soneson et al., 2015) and were further analyzed using the DESeq2 package (Love et al., 2014). Transcripts with a read count of less than 10 in all samples combined were removed from further analyses. Principal component analysis showed one Col-0 sample not clustering with the other two Col-0 samples, and this sample was omitted from further analyses. Normalized read counts were $\log_2$ transformed, and a negative binomial generalized linear model was fitted to the data using the genotype as the factor to identify differential gene expression. The significance of the expression changes was determined using Wald tests and corrected for multiple testing by the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

Mathematical Models

Model I: Predicting the Spatiotemporal Root Endoploidy Map

Model I simulates the expression patterns of genes in root slices and marker lines as a function of their endoploidy-specific expression levels in the cortex (Figure 2). The model equations are as follows:

$$E_S(g,s) = \frac{\sum_{t=1}^{14}\left[\sum_{c\in C(s,t)}\left(\sum_p w_P(t,c,p)\cdot E_P(g,p)\right)\right]}{\sum_{t=1}^{14}|C(s,t)|} \quad (1)$$

$$E_M(g,m) = \frac{\sum_{t\in T(m)}\left\{\sum_{s\in S(m,t)}\left[\sum_{c\in C(s,t)}\left(\sum_p w_P(t,c,p)\cdot E_P(g,p)\right)\right]\right\}}{\sum_{t\in T(m)}\left(\sum_{s\in S(m,t)}|C(s,t)|\right)} \quad (2)$$

In these equations, $E_S(g,s)$ and $E_M(g,m)$ represent the simulated expression of gene g in slice s and marker line m, respectively. t indexes tissues (cell types). The index p represents the endoploidy level (2C, 4C, 8C, or 16C) and the index c indicates the position of a cell in a particular tissue t along the longitudinal axis. $E_P(g,p)$ represents the endoploidy-specific expression level of gene g at endoploidy p in the cortex data set. W_P is the endoploidy matrix, with $w_P(t,c,p) = 1$ if cell c in tissue t has endoploidy level p, and $w_P(t,c,p) = 0$ otherwise. $C(s,t)$ is the set of cell numbers in slice s and tissue t, as derived from the cell count matrix W_C, which incorporates average cell count estimates across slices and tissues obtained from the visual inspection of 10 confocal images of Arabidopsis wild-type (Col-0) roots for the cell types hair cell, cortex, and endodermis in the meristematic and elongation zones (slices 1–8), and for the xylem pole pericycle and phloem pole pericycle in the meristematic zone (slices 1–6) (Supplemental Data Set 17). The W_C cell counts in the non-hair, phloem, phloem companion, xylem, and procambium cell files in slices 1 to 8 were deduced from the measured counts for other cell types. The cell counts in the remaining tissues and slices were based on the cell counts provided by Cartwright et al. (2009). In Equation 2, $T(m)$ represents the set of tissues covered in at least some developmental stages (slices) by a particular marker m, $S(m,t)$ being the set of slices in which marker m covers tissue t (Brady et al., 2007).

Equations 1 and 2 essentially sum up the endoploidy-specific expression levels of gene g in all cells in a slice s and all marked cells in marker line m, where the endoploidy of each cell (and, hence, its contribution to the gene's expression level) is determined from the endoploidy matrix W_P. Simulated slice and marker line expression levels are then compared with the slice and marker line expression levels experimentally determined by Brady et al. (2007). The parameters of the model, the endoploidy boundaries (i.e., the cells at which the endoploidy level changes along the longitudinal root axis, as encoded in the endoploidy matrix W_P), are randomly assigned at the beginning of a simulation and optimized using a Monte Carlo Simulated Annealing strategy to obtain the best possible fit between the simulated and measured expression patterns across all slices and marker lines (see "Simulation and Optimization Strategy").

Two different derivations of the weighting scheme used in Equations 1 and 2, for two different classes of candidate endoploidy marker genes, are described below using a hypothetical example (Supplemental Figure 11).

Derivation 1. For transcripts of which the absolute expression level scales with cytoplasmic volume, as appears to be the case for most transcripts in, e.g., mammals (Padovan-Merhar et al., 2015) and yeast (Marguerat and Bähler, 2012), the relative expression level is the same across cells with different volume. Under the assumption that there exists a set of genes for which the relative expression level at any particular endoploidy is approximately the same across tissues, the relative expression level of any such gene g in a marker line m or slice s should be a volume-weighted sum across the component cells, e.g., for a slice s containing a number of 4C and 8C cells:

$$E_S(g,s) = \frac{V_{tot,4C}\ E_P(g,4C) + V_{tot,8C}\ E_P(g,8C)}{V_{tot,4C} + V_{tot,8C}} \quad (3)$$

with $E_P(g,4C)$ and $E_P(g,8C)$ the relative expression levels of gene g in 4C and 8C cells, respectively, of the cortex (which are assumed to be equal to the relative expression levels in other tissues), and $V_{tot,4C}$ and $V_{tot,\ 8C}$ the total cytoplasmic volume of the 4C and 8C cells in slice s. Conceptually, Equation 3 is the easiest form for describing relative expression levels in an ensemble of cells (be it a slice or a tissue) as a function of the endoploidy-specific relative expression levels measured in the cortex. However, determining the volumes $V_{tot,4C}$ and $V_{tot,8C}$ for any slice or marker line requires a detailed assessment of the cytoplasmic volumes of individual cells across all root tissues and developmental stages, which is very challenging. Note that measuring total root cell volumes instead of cytoplasmic volumes is not sufficient, due to the interfering impact of expanding plant cell vacuoles on the cytoplasmic/cell volume ratio. Equation 3 can be approximated by assuming that the average cytoplasmic volume of cells in a particular endoploidy state is the same across tissues/slices, say $\overline{V_{4C}}$ and $\overline{V_{8C}}$ for 4C and 8C cells, respectively. Plant root cells of different types have indeed been found to exhibit fairly similar cytoplasmic volumes on average (averaged across endoploidies) (Petersson et al., 2009). Equation 3 then becomes:

$$E_S(g,s) = \frac{n_{4C}\overline{V_{4C}}\ E_P(g,4C) + n_{8C}\overline{V_{8C}}\ E_P(g,8C)}{n_{4C}\overline{V_{4C}} + n_{8C}\overline{V_{8C}}} \quad (4)$$

with n_{4C} and n_{8C} the number of 4C and 8C cells, respectively, in slice s. If the volumes of 4C and 8C cells are written as a function of the volume of an average 2C cell ($\overline{V_{2C}}$), with scaling factors f_{4C} and f_{8C}, respectively, this reduces to:

$$\begin{aligned} E_S(g,s) &= \frac{n_{4C} f_{4C}\overline{V_{2C}}\ E_P(g,4C) + n_{8C} f_{8C}\overline{V_{2C}}\ E_P(g,8C)}{n_{4C} f_{4C}\overline{V_{2C}} + n_{8C} f_{8C}\overline{V_{2C}}} \\ &= \frac{n_{4C} f_{4C} E_P(g,4C) + n_{8C} f_{8C}\ E_P(g,8C)}{n_{4C} f_{4C} + n_{8C} f_{8C}} \end{aligned} \quad (5)$$

When assuming $f_{4C} = f_{8C} = 1$ (see below for rationale), Equation 5 becomes the toy example equivalent of Equation 1. Equation 2 can be derived similarly.

Derivation 2. The same equation can be derived for genes for which the absolute (rather than the relative) expression levels remain approximately constant across cells of the same endoploidy, regardless of volume. Although such genes are expected to constitute a minority, some genes, in particular some transcription factor genes, have been shown to have absolute expression levels that do not scale with cell volume (Marguerat and Bähler, 2012; Schmidt and Schibler, 1995). Under the assumption that a gene g has a constant absolute expression level at any given endoploidy level, regardless of cell volume, the relative expression level of g in, e.g., a slice s containing a number of 4C and 8C cells (counts n_{4C} and n_{8C}, respectively) can be written as:

$$\begin{aligned} E_S(g,s) &= \frac{E_{abs,S}(g,s)}{mRNA_{tot}(s)} \\ &= \frac{\sum_{i=1}^{n_{4C}} E_{abs,4C}(g,i) + \sum_{j=1}^{n_{8C}} E_{abs,8C}(g,j)}{\sum_{i=1}^{n_{4C}} mRNA_{4C}(i) + \sum_{j=1}^{n_{8C}} mRNA_{8C}(j)} \\ &\overset{(a)}{=} \frac{n_{4C}\ E_{abs,P}(g,4C) + n_{8C}\ E_{abs,P}(g,8C)}{\sum_{i=1}^{n_{4C}} mRNA_{4C}(i) + \sum_{j=1}^{n_{8C}} mRNA_{8C}(j)} \\ &\overset{(b)}{=} \frac{n_{4C}\, c\overline{V_{4C}}\ E_P(g,4C) + n_{8C}\ c\ \overline{V_{8C}}\ E_P(g,8C)}{\sum_{i=1}^{n_{4C}} cV_{4C}(i) + \sum_{j=1}^{n_{8C}} cV_{8C}(j)} \\ &\overset{(c)}{\cong} \frac{n_{4C}\overline{V_{4C}}\ E_P(g,4C) + n_{8C}\overline{V_{8C}}\ E_P(g,8C)}{n_{4C}\overline{V_{4C}} + n_{8C}\overline{V_{8C}}} \\ &= \frac{n_{4C}\ f_{4C}\ E_P(g,4C) + n_{8C}\ f_{8C}\ E_P(g,8C)}{n_{4C} f_{4C} + n_{8C} f_{8C}} \end{aligned} \quad (6)$$

where $E_{abs,S}(g,s)$ is the absolute expression level of gene g in slice s, $E_{abs,4C}(g,i)$ is the absolute expression level of gene g in 4C cell i, $E_{abs,P}(g,4C)$ is the absolute expression level of gene g in 4C cortex cells (remember that in this derivation, absolute expression levels in 4C cells are assumed to be constant across tissues), $E_P(g,4C)$ is the relative expression of gene g in the average 4C cortex cell, $mRNA_{tot}(s)$ indicates the total mRNA count in slice s, $mRNA_{4C}(i)$ is the total mRNA count in 4C cell i, $V_{4C}(i)$ is the cytoplasmic volume of 4C cell i, $\overline{V_{4C}}$ is the cytoplasmic volume

of an average 4C cortex cell, *c* is a constant total mRNA concentration per unit volume, and f_{4C} and f_{8C} are volume scaling factors, as in Equation 5. The codes (a) to (c) indicate the use of the following assumptions/approximations in the derivation: (a) The absolute expression level $E_{abs,4C}(g,i)$ in any 4C cell is the same as the absolute expression level in cortex cells $E_{abs,P}(g,4C)$, and similarly for 8C cells. (b) The total mRNA content per unit volume *c* is roughly constant across cells of different size and cell type. The assumption of a constant mRNA concentration across different-sized cells is supported in, e.g., mammals (Padovan-Merhar et al., 2015) and yeast (Marguerat and Bähler, 2012), also for cells of different (endo)ploidy (Padovan-Merhar et al., 2015; Schmidt and Schibler, 1995). (c) In any tissue, cells of a particular endoploidy are assumed to have more or less the same cytoplasmic volume on average as cortex cells of that endoploidy (Petersson et al., 2009).

Although the form of Equations 5 and 6 is identical, fewer approximations are made in the case of derivation 2, where the marker genes *g* have a constant absolute expression level across cells of different volume. This is because approximation (c) is only used in the denominator of Equation 6, whereas it is used in both the numerator and denominator of Equation 5, with its use in the numerator directly affecting the balance between 4C and 8C endoploidy contributions to the slice expression level $E_S(g,s)$. An alternative way of appreciating the difference in approximation effects between genes with constant relative versus constant absolute expression levels for any given endoploidy is to consider Equations 5 and 6 as reshuffling the cytoplasmic volumes of all cells in slice *s* to match the cell sizes for any particular endoploidy to the cortex average (Supplemental Figure 11).

A residual problem is the determination of the volume-scaling factors *f* for the different endoploidy levels in Equations 5 and 6. According to the karyoplasmic ratio theory, the cytoplasmic volume of cells should scale with ploidy (Wilson, 1925). Evidence supporting the karyoplasmic ratio theory has been presented previously for several organisms, including plants, but a clear relationship between endoploidy and cell (cytoplasm) size is not always observed (Chevalier et al., 2014). We tested several possible *f* weighting schemes, and surprisingly, the predicted ploidy maps I and II using the weighting scheme $f_{2C} = f_{4C} = f_{8C} = f_{16C} = 1$ (i.e., cells of different endoploidy are assumed to have roughly the same cytoplasmic volume) correspond best to the flow cytometry-derived endoploidy map presented in Supplemental Figure 8 and the endoploidy boundaries observed in other validation experiments (Figure 2; Supplemental Figure 9). This suggests that in Arabidopsis roots, the cytoplasmic volume of a cell may not scale with its endoploidy level. We used the equal *f* weighting assumption in equations 1 and 2 to predict the root endoploidy map presented in Figure 3. Other weighting schemes in which *f* scales with endoploidy lead to qualitatively similar root endoploidy map layouts, but with most endoploidy boundaries shifted to later developmental stages. An example optimized map for $f_{2C} = 1;\ f_{4C} = 2;\ f_{8C} = 4;\ f_{16C} = 8$ (i.e., under the assumption that cytoplasmic volume doubles with every endocycle) is presented in Supplemental Figure 11.

Model II: Prediction of Endoploidy Distribution Changes upon Stress Treatment

Model II is a simplified version of model I that simulates gene expression patterns in a whole root as a function of the endoploidy-specific expression levels in the cortex data set. The expression level of a particular gene in a root is taken to be a weighted sum of the expression levels in cells with different endoploidy levels in that root:

$$E(g) = \sum_p w(p) \cdot E_P(g,p) \tag{7}$$

where $E(g)$ represents the simulated expression level of gene *g* to be compared with the measured expression level in a given publicly available gene expression data set on whole roots subjected to a particular treatment, $w(p)$ represents the weight of endoploidy *p*, i.e., the fraction of root cells having endoploidy *p*, and $E_P(g,p)$ represents the endoploidy-specific expression level of gene *g* at endoploidy level *p* in the cortex data set. Model II uses the same optimization strategy as model I to optimize the endoploidy weights $w(p)$ (see below).

Model Optimization Strategy. We used a Monte Carlo Simulated Annealing strategy with exponential temperature decay to optimize the parameters in both models I and II. The optimization strategy is schematically represented in Supplemental Figure 12. Initially, the parameters, i.e., the endoploidy boundaries in each cell type along the longitudinal axis of the root for model I, or the endoploidy weights for model II, are randomly assigned, and optimization progresses by attempting random steps in parameter space, i.e., by moving a particular endoploidy boundary one cell up or down in a single cell type (model I) or by simulating the effect of a slight endoploidy shift on the endoploidy percentages in a root (model II). A step was accepted if

$$rand < exp(-\Delta R/SA) \tag{8}$$

with *rand* a random number drawn uniformly from the interval [0,1], ΔR the change in optimization potential (see Equation 9) upon taking a step in parameter space, and *SA* the simulated annealing parameter (temperature), which gradually decreases during the course of an optimization run. Endoploidy map optimizations for single genes, as used in the selection of the balanced gene set, were performed with the following parameter settings: start temperature $SA_{init} = 5$, end temperature $SA_{end} = 0.001$, and exponential cooling scheme $SA_i = 0.9999\ SA_{i-1}$ with a single iteration at each SA_i. Endoploidy map optimizations for the full gene sets were performed with the parameter settings $SA_{init} = 10$, $SA_{end} = 0.00001$ and the exponential cooling scheme $SA_i = 0.99999\ SA_{i-1}$ with 1000 iterations at each SA_i. The optimization potential *R* is defined by the χ^2 (goodness-of-fit) statistic between measured and simulated expression levels, averaged per data point, e.g., in the case of Model I:

$$R = \frac{\sum_{g=1}^{nG}\left[\left(\frac{M_g - simM_g}{\sigma(M_g)}\right)^2 + \left(\frac{S_g - simS_g}{\sigma(S_g)}\right)^2\right]}{nG \times 29} \tag{9}$$

where *nG* is the total number of genes included in the simulation, S_g and M_g are the measured expression profiles (vectors) of gene *g* across 12 slices and 17 marker lines, respectively, and $simS_g$ and $simM_g$ are the corresponding simulated expression profiles, i.e., the vectors of predicted $E_S(g,s)$ and $E_M(g,m)$ values across all slices and marker lines, respectively, as derived from Equations 1 and 2 using the endoploidy boundaries of the map that is being optimized. $\sigma(S_g)$ and $\sigma(M_g)$ represent SD vectors approximated by element-wise square roots of the S_g and M_g vectors, respectively [note that the divisions involving $\sigma(S_g)$ and $\sigma(M_g)$ are element-wise divisions]. The calculated *R* values are divided by a normalization factor $nG \times 29$ representing the number of data points fitted, where 29 is the sum of the number of slices (12) and marker lines (17) represented in the spatiotemporal expression data set (Brady et al., 2007). This normalization factor has no effect on the optimal map layout.

Endoploidy Map Validation Experiments

I. Flow Cytometer Experiments

The endoploidy content of cells of different tissue types (Supplemental Figure 6) was measured in 0.5-cm-long Arabidopsis root tips using flow cytometric analysis of cell type-specific GFP marker lines (Brady et al., 2007; Dietrich et al., 2017). For some tissues, multiple marker lines cover different developmental stages (Brady et al., 2007). In this case, we used marker lines that cover the later developmental stages (Supplemental Data Set 16).

Sections from roots of 5-d-old plants were excised with a razor blade ~0.5 cm above the root tip. For sorting of protoplasts, the excised tips were incubated in 8 mL of protoplasting solution (1.25% cellulase [Yakult], 0.3% macerozyme [Yakult], 0.4 M mannitol, 20 mM MES, 20 mM KCl, pH 5.7 adjusted with 1 M Tris/HCl, pH 7.5, activated at 55°C for 10 min and then cooled to room temperature, 0.1% BSA, and 10 mM $CaCl_2$) in 25-mL Erlenmeyer flasks for 2 h on an orbital shaker (100 rpm) under continuous white light. The protoplasts were then filtered through a 40-μm filter and centrifuged at 1000 rpm at 4°C for 10 min. The pellets were resuspended in 1 mL of wash buffer (identical composition to that of the protoplasting solution but lacking the enzymes and activation pretreatment). The *GFP*-expressing protoplasts were flow-sorted and collected in 200 μL CyStain UV Precise nuclei extraction buffer (Partec) and their nuclei were stained by adding 800 μL nuclei staining buffer (Partec). DNA contents of *GFP*-expressing protoplasts were measured with a CyFlow flow cytometer (Partec) excited by illumination at 395 nm and analyzed with the FloMax software (Partec).

For flow cytometry on nuclear GFP lines, cut root tips were chopped with a razor blade in 200 μL of nuclei extraction buffer containing 45 mM $MgCl_2$, 30 mM sodium citrate, and 20 mM 3-morpholinopropane-1-sulfonic acid, pH 7.0 (Galbraith et al., 1991) for 2 min, then filtered through a 50-μm nylon filter. The DNA was stained with 1 mg/mL DAPI (Zhang et al., 2005). Nuclei were measured using a CyFlow flow cytometer excited by illumination at 395 nm and equipped with an additional 488-nm laser to excite and detect GFP-specific fluorescence. The DNA content of cells was derived from DAPI fluorescence measurements using FloMax software (Partec).

To represent the flow cytometry-derived endoploidy distributions in a format comparable to the predicted endoploidy map, the number of cells analyzed by flow cytometry for a given cell type is rescaled to fit the number of cells present on the predicted map in the slices marked by the cell type marker concerned. The rescaled numbers of measured 2C, 4C, 8C, and 16C cells are then distributed in a logical order (2C→4C→8C→16C) over the root slices marked, and endoploidy levels for earlier or later developmental stages are inferred where possible (Supplemental Figure 7). For procambium cells, a special protocol was followed because *WOL* marks not only procambium cells, but the entire stele in slices 1 to 8. The *WOL*-derived endoploidy, cell counts were normalized to the total number of stele cells in slices 1 to 8 of the predicted map, and the numbers of 2C, 4C, 8C, and 16C cells obtained for nonprocambium stele tissues (phloem pole pericycle, xylem pole pericyle, xylem, phloem, and phloem companion cells) in slices 1 to 8 using other marker lines were subtracted from the rescaled *WOL* endoploidy cell counts to obtain an estimate of the number of 2C, 4C, 8C, and 16C procambium cells to be filled in on the validation map. Using this protocol, a negative estimate of the 2C procambium cells was obtained, i.e., the number of 2C cells observed for nonprocambium stele tissues using tissue-specific markers is larger than the number inferred from the *WOL* measurements, indicating that *WOL* may mark different stele tissues with different efficiency. For filling in the validation map, the surplus of nonprocambium 2C cells was treated as 4C cells and subtracted from the total 4C stele cell count. Because of the lack of a procambium-specific marker line, the procambium endoploidy distributions in both the predicted endoploidy map and the flow cytometry-based map are likely less reliable.

II. Endoploidy Border Analysis

The endoploidy boundary positions observed in the flow cytometry-derived map are not always reliable because they are highly dependent on the absolute number of nuclei or protoplasts extracted and measured per tissue in the flow cytometer analysis. The efficiency of protoplasting and nuclear extraction are dependent on the duration of the treatment of the roots with the respective extraction buffers. Such dependency often gives quantitative variations in the endoploidy distributions, and it is not straightforward to compare data obtained from different experiments. Moreover, flow cytometry cannot distinguish mitotically dividing G2 cells from endoreplicating 4C cells, leading to uncertainties in the location of the 2C-4C boundary in any given cell type. To overcome these technical limitations, we used confocal microscopy to locate the first atrichoblast and cortex cells from the QC exhibiting a visible GFP signal in the *SMR1:GFP-GUS* marker line. We performed this analysis on three to five roots and the average first cell number was taken as the endoploidy boundary cell number estimate.

III. Mapping Endoploidy Borders through Histochemical GUS Staining

To validate the order of endoreplication onset across tissues, we studied root cross sections at various distances from the QC in GUS-stained endocycle marker lines (*SMR1* and *SIM*). First, we generated a cell distance matrix of the atrichoblast cell type by counting the cells along the longitudinal axis of the root and measuring their distance from the QC. We repeated this analysis for 30 different roots to generate an average distance map (Supplemental Data Set 18). Next, we used the measured average distance matrix to identify the atrichoblast cell number (counted along the longitudinal axis) visible in a root cross section at a given distance from the QC. The measured endoreplication onset order was compared with the predicted order at the inferred atrichoblast cell number on the virtual endoploidy map.

IV. Mapping Endoploidy Borders in Atrichoblast and Trichoblast Cell Files through Densitometry

Whole seedlings were fixed in 1.5% (v/v) formaldehyde and 0.5% (v/v) glutaraldehyde made up in PEMT buffer (50 mM PIPES, 2 mM EGTA, 2 mM $MgSO_4$, and 0.05% [v/v] Triton X-100, pH 7.2) for 40 min and rinsed in PEMT buffer three times for 10 min and then rinsed three more times in PBS buffer. Seedlings were stained for 10 min with DAPI using 25% CyStain B solution (Partec) in PBS buffer, washed with PBS buffer three times, and then mounted in PBS buffer. DAPI signal was visualized by fluorescence microscopy using a Carl Zeiss inverted LSM710 confocal laser microscope, equipped with objectives c-apochromat 40×/1.2 W Korr M27. Fiji was used to export the .lsm z-stack file as a TIFF file. MorphoGraphX (MGX) (Barbier de Reuille et al., 2015) was used to digitalize the images and create 3D meshes of the nuclei that were then used to extract the nuclear volume. The position of the cells/nuclei in relation to the QC was used to assess the changes in nuclear volume.

V. Mapping Nuclear Volumes within the LRC

Seedlings were incubated in water with RNase A (20 μg/mL) for 30 min and subsequently fixed with 2% (v/v) paraformaldehyde and 0.1% (v/v) triton in MTSB buffer (15 g PIPES, 1.90 g EDTA, 1.22 g $MgSO_4 \cdot 7H_2O$, 2.5 g KOH dissolved in a total of 1 liter water at pH 7.0). Vacuum was applied two times for 10 min for air removal, followed by incubation in the dark on a shaker (100 rpm) for 30 min, and seedlings were then rinsed three times with MTSB buffer. Seedlings were transferred to methanol for 5 min at 60°C and the methanol was slowly diluted by adding small volumes of water until a concentration of 50% methanol was obtained. After two washes with water, this procedure was repeated. After a wash with MTSB, the buffer was replaced with MTSB buffer holding 3% (v/v) nonionic detergent IGEPAL CA-630 and 10% (v/v) DMSO for 25 min. Samples were washed twice in

MTSB buffer, followed by removal of MTSB buffer and addition of DAPI (1 mg/mL) for 10 min. After removal of the DAPI solution, glycerin (50%) was added in a small volume. Finally, seedlings were mounted on cover slips in 100% glycerin. The DAPI signal was visualized by fluorescence microscopy using a Carl Zeiss inverted LSM710 confocal laser microscope equipped with objectives c-apochromat 40×/1.2 W Korr M27. Fiji was used to export the .lsm z-stack file as a TIFF file. MGX (Barbier de Reuille et al., 2015) was used to wipe all non-nuclear background signals and to create 3D meshes of the nuclei for DAPI fluorescence measurements.

Mapping of Nuclear and Cellular Volumes

Five-day-old seedlings were incubated in 15 μM (10 μg/mL in distilled water) propidium iodide (Sigma-Aldrich) for 5 min and then mounted in water. GFP and PI signal were visualized by fluorescence microscopy using a Carl Zeiss inverted LSM710 confocal laser microscope, equipped with objectives c-apochromat 40×/1,2 W Korr M27. Fiji was used to split the channels (PI/GFP) and export them as TIFF. If there were shift produced by the movement of the root during the imaging process, we used the Fiji plug-in Registration>StackReg>Rigid Body to correct them. MGX (Barbier de Reuille et al., 2015) was used to digitalize the GFP channel images and create 3D cellular and nuclear meshes that were then used to extract the cellular and nuclear volumes. The analyzed region was determined based on the cell/nucleus position with the first peak expression of *SMR1:GFP-GUS*. Based on this position, we selected two cells bellow and seven cells above this position for the nuclear and cellular volume analysis.

Transmission Electron Microscopy

Root tips of 5-d-old seedlings of Arabidopsis were excised, immersed in 20% (w/v) BSA, and frozen immediately in a high-pressure freezer (Leica EM ICE; Leica Microsystems). Freeze substitution was performed using a Leica EM AFS (Leica Microsystems) in dry acetone containing 1% (w/v) OsO_4 and 0.2% glutaraldehyde over a 4-d period as follows: −90°C for 54 h, 2°C per hour increase for 15 h, −60°C for 8 h, 2°C per hour increase for 15 h, and −30°C for 8 h. Samples were then slowly warmed up to 4°C, infiltrated stepwise over 3 d at 4°C in Spurr's resin and embedded in capsules. The polymerization was performed at 70°C for 16 h. Ultrathin sections were made from the late meristematic zone (200–250 μm from the root tip) up to the mature zone (350–400 μm from the root tip) every 15 μm using an ultramicrotome (Leica EM UC6) and poststained in a Leica EM AC20 for 40 min in uranyl acetate at 20°C and for 10 min in lead stain at 20°C. Sections were collected on formvar-coated copper slot grids. Grids were viewed with a JEM 1400plus transmission electron microscope (JEOL) operating at 60 kV.

Serial Block-Face Scanning Electron Microscopy

Five-day-old seedlings were fixed on a plate for 2 h using 0.1 M phosphate buffer, pH 6.8, 2.5% (v/v) glutaraldehyde, and 0.5% (v/v) paraformaldehyde, after which the root tips were isolated by cutting, transferred to fresh fixative, and kept overnight at 4°C. Samples were washed the next day in cold 0.15 M cacodylate buffer. En bloc contrast staining was performed by consecutive incubations in heavy metal-containing solutions. Between incubation steps, samples were washed in ultrapure water (UPW). The first staining step was in 0.2% (w/v) ruthenium red and 2% (v/v) aqueous osmium tetroxide in 0.15 M cacodylate buffer (pH 7.4). After washing, the samples were incubated in a fresh thiocarbohydrazide (TCH) solution (1% w/v in UPW). The next washing step was followed by incubation in 1% osmium in UPW. After this, an additional incubation in TCH followed by 1% (v/v) osmium was performed. Next, samples were incubated in 1% (w/v) uranyl acetate followed by a Walton's lead aspartate staining. For this, a 30-mM L-aspartic acid solution was used to freshly dissolve lead nitrate (20 mM, pH 5.5). The solution was filtered through a 0.22-μm Millipore syringe filter after a 30-min incubation at 60°C. After the final washing steps, the samples were dehydrated using ice-cold solutions of 2× 30%, 50%, 70%, 90%, 100% ethanol (anhydrous), and 2× 100% acetone. Resin embedding was done by placing the samples in solutions containing 30%, 50%, 70%, 90%, and 100% Spurr's resin (ethyl methyl sulfonate) in acetone. Next, samples were placed in fresh 100% Spurr's resin for polymerization. All staining, washing, dehydration, infiltration, and polymerization steps were performed in a Pelco Biowave Pro microwave (Ted Pella). For serial block-face imaging, the resin-embedded root tips were mounted on an aluminum specimen pin (Gatan), using conductive epoxy (Circuit Works). The specimens were trimmed in a pyramid shape using an ultramicrotome and coated with 5 nm of Pt, in a Quorum sputter coater. The aluminum pins were placed in the Gatan 3View2 in a Zeiss Merlin scanning electron microscope, for imaging at 1.5 kV with a Gatan Digiscan II energy-selective backscattered detector and slicing 80-nm slices. The 3D serial block-face-scanning electron microscopy Z-stacks have been registered with IMOD (http://bio3d.colorado.edu/imod/) to correct for movements of the specimen within the microscope during the imaging process.

Accession Numbers

RNA microarray and RNA-seq data have been deposited in the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE108399 and GSE115001. Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the following accession numbers: *SIM* (AT5G04470) and *SMR1* (AT3G10525).

Supplemental Data

Supplemental Figure 1. Endoploidy-specific expression patterns in the Arabidopsis root cortex.

Supplemental Figure 2. Centroids of endoploidy-specific expression clusters.

Supplemental Figure 3. GO enrichment of endoploidy-specific expression clusters.

Supplemental Figure 4. Optimized endoploidy maps for selected unbalanced gene sets.

Supplemental Figure 5. Comparison between two optimized endoploidy map alternatives (Optimal Maps I and II).

Supplemental Figure 6. Flow cytometry-based tissue-specific endoploidy distributions.

Supplemental Figure 7. Endoploidy map determined through flow cytometry.

Supplemental Figure 8. Expression prediction performance for the 332 genes in the balanced gene set in selected stress conditions.

Supplemental Figure 9. Nuclear and cellular volume of control (black bars) versus *smr1* (gray) atrichoblast cells.

Supplemental Figure 10. Circularity measurements of control (Ler) and *lgo1* trichoblast cells within the transition zone.

Supplemental Figure 11. Endoploidy map modeling assumptions for volume scaling.

Supplemental Figure 12. Schematic representation of the endoploidy boundary optimization approach.

Supplemental Data Set 1. Log_2 expression values and differential expression analysis results for all genes in the cortex endoploidy-specific expression data set.

Supplemental Data Set 2. Cluster membership for all differentially expressed genes in the cortex data set.

Supplemental Data Set 3. GO enrichment results for 28 clusters with clear peak expression at one endoploidy level in the cortex expression data set.

Supplemental Data Set 4. Single-gene Model I optimization R scores and Pearson correlations for the reduced set of 4378 genes.

Supplemental Data Set 5. Annotation of the balanced set of 332 genes used in Model I.

Supplemental Data Set 6. GO enrichment results for the balanced set of 332 genes used in Model I.

Supplemental Data Set 7. Comparison between predicted endoploidy boundaries on Optimal Maps I and II.

Supplemental Data Set 8. Chi-squared contributions of individual markers and slices to the R values of Optimal Maps I and II.

Supplemental Data Set 9. Endoploidy classes of lateral root cap cells as inferred from DAPI staining.

Supplemental Data Set 10. Comparison of predicted and validated endoploidy boundaries for trichoblast, atrichoblast, and cortex tissues.

Supplemental Data Set 11. List of genes ranked in order of best fit between their observed spatiotemporal expression levels and the expression levels predicted from the optimized endoploidy map.

Supplemental Data Set 12. GO enrichment results for top 332 genes whose transcript level is most strongly linked to the endoploidy level of the cell.

Supplemental Data Set 13. Description of publicly available gene expression data sets used for predicting endoploidy distribution shifts under stress.

Supplemental Data Set 14. Subset of stress experiments used for eliminating stress-sensitive genes from the balanced gene set for use in model II.

Supplemental Data Set 15. Transcriptome data of sorted SMR1:GFP-GUS cells in the wild type versus *smr1.*

Supplemental Data Set 16. Cell types and cytoplasmic or nuclear-tagged GFP marker lines used for validating the Arabidopsis root endoploidy map by flow cytometry.

Supplemental Data Set 17. Adapted cell counts of 14 cell types in 12 slices.

Supplemental Data Set 18. Average distance of atrichoblast cells (numbered along the longitudinal axis) from the quiescent center.

Supplemental Movie 1. Model optimization run for a single gene (AT4G22070).

Supplemental Movie Legend.

ACKNOWLEDGMENTS

We thank Sonia Bartunkova for technical support, Annick Bleys for help in preparing the manuscript, and Ingrid Lieten for the special grant to develop correlative microscopy at VIB. R.B. and R.P.K. acknowledge the support of a Newton International and OMICS@vibMarie Curie COFUND fellowship, respectively. J.C.L. acknowledges support from the National Science Foundation (IOS 1146620). D.W.G. acknowledges support from National Science Foundation Grant DBI-052163. The work by F.X. was supported by a CSC fellowship and ANR project "Pectosign." The Institut Jean-Pierre Bourgin benefits from the support of the LabEx Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS). This work was supported by grants of the Research Foundation Flanders (G.002911N) and the Interuniversity Attraction Poles Programme (IUAP P7/29 "MARS"), initiated by the Belgian Science Policy Office.

AUTHOR CONTRIBUTIONS

S.M. and L.D.V. designed and supervised the study. H.H., D.W.G., and L.D.V. designed wet-lab experiments. S.M. designed the virtual root endoploidy map modeling approach. R.B. and S.M. implemented the model and ran the simulations. R.B., V.B., F.C., R.P.K, Z.H, G.V.I., J.C.L., R.D.R., I.V., R.L., F.X., and A.K. performed wet-lab experiments. T.E. processed and analyzed RNA-seq data. G.M.L. assisted with technical aspects of the flow sorting experiments under guidance of D.W.G. V.S. performed statistical analyses. T.B. and M.K.N. contributed resources. R.S. gave prepublication access to software. R.B., S.M., and L.D.V. wrote the manuscript.

Received January 2, 2018; revised June 18, 2018; accepted August 8, 2018; published August 16, 2018.

REFERENCES

Artlip, T.S., Madison, J.T., and Setter, T.L. (1995). Water deficit in developing endosperm of maize: cell division and nuclear DNA endoreduplication. Plant Cell Environ. **18:** 1034–1040.

Bao, Z., and Hua, J. (2014). Interaction of CPR5 with cell cycle regulators UVI4 and OSD1 in Arabidopsis. PLoS One **9:** e100347.

Bao, Z., Yang, H., and Hua, J. (2013). Perturbation of cell cycle regulation triggers plant immune response via activation of disease resistance genes. Proc. Natl. Acad. Sci. USA **110:** 2407–2412.

Barbier de Reuille, P., et al. (2015). MorphoGraphX: A platform for quantifying morphogenesis in 4D. eLife **4:** 05864.

Barow, M. (2006). Endopolyploidy in seed plants. BioEssays **28:** 271–281.

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

Bensussan, M., et al. (2015). Suppression of dwarf and irregular xylem phenotypes generates low-acetylated biomass lines in Arabidopsis. Plant Physiol. **168:** 452–463.

Bernhardt, C., and Tierney, M.L. (2000). Expression of AtPRP3, a proline-rich structural cell wall protein from Arabidopsis, is regulated by cell-type-specific developmental pathways involved in root hair formation. Plant Physiol. **122:** 705–714.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science **302:** 1956–1960.

Boron, A.K., Van Orden, J., Nektarios Markakis, M., Mouille, G., Adriaensen, D., Verbelen, J.-P., Höfte, H., and Vissenberg, K. (2014). Proline-rich protein-like PRPL1 controls elongation of root hairs in *Arabidopsis thaliana*. J. Exp. Bot. **65:** 5485–5495.

Boudolf, V., Vlieghe, K., Beemster, G.T., Magyar, Z., Torres Acosta, J.A., Maes, S., Van Der Schueren, E., Inzé, D., and De Veylder, L. (2004). The plant-specific cyclin-dependent kinase CDKB1;1 and transcription factor E2Fa-DPa control the balance of mitotically dividing and endoreduplicating cells in Arabidopsis. Plant Cell **16:** 2683–2692.

Bourdon, M., Coriton, O., Pirrello, J., Cheniclet, C., Brown, S.C., Poujol, C., Chevalier, C., Renaudin, J.-P., and Frangne, N. (2011). In planta quantification of endoreduplication using fluorescent in situ hybridization (FISH). Plant J. **66:** 1089–1099.

Bourdon, M., Pirrello, J., Cheniclet, C., Coriton, O., Bourge, M., Brown, S., Moïse, A., Peypelut, M., Rouyère, V., Renaudin, J.-P., Chevalier, C., and Frangne, N. (2012). Evidence for karyoplasmic homeostasis during endoreduplication and a ploidy-dependent increase in gene transcription during tomato fruit growth. Development **139:** 3817–3826.

Brady, S.M., Orlando, D.A., Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science **318:** 801–806.

Bramsiepe, J., Wester, K., Weinl, C., Roodbarkelari, F., Kasili, R., Larkin, J.C., Hülskamp, M., and Schnittger, A. (2010). Endoreplication controls cell fate maintenance. PLoS Genet. **6:** e1000996.

Breuer, C., Braidwood, L., and Sugimoto, K. (2014). Endocycling in the path of plant development. Curr. Opin. Plant Biol. **17:** 78–85.

Brininstool, G., Kasili, R., Simmons, L.A., Kirik, V., Hülskamp, M., and Larkin, J.C. (2008). Constitutive Expressor Of Pathogenesis-related Genes5 affects cell wall biogenesis and trichome development. BMC Plant Biol. **8:** 58.

Bromley, J.R., Busse-Wicher, M., Tryfona, T., Mortimer, J.C., Zhang, Z., Brown, D.M., and Dupree, P. (2013). GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns. Plant J. **74:** 423–434.

Cartwright, D.A., Brady, S.M., Orlando, D.A., Sturmfels, B., and Benfey, P.N. (2009). Reconstructing spatiotemporal gene expression data from partial observations. Bioinformatics **25:** 2581–2587.

Casimiro, I., Beeckman, T., Graham, N., Bhalerao, R., Zhang, H., Casero, P., Sandberg, G., and Bennett, M.J. (2003). Dissecting Arabidopsis lateral root development. Trends Plant Sci. **8:** 165–171.

Cebolla, A., Vinardell, J.M., Kiss, E., Oláh, B., Roudier, F., Kondorosi, A., and Kondorosi, E. (1999). The mitotic inhibitor ccs52 is required for endoreduplication and ploidy-dependent cell enlargement in plants. EMBO J. **18:** 4476–4484.

Chandran, D., Inada, N., Hather, G., Kleindt, C.K., and Wildermuth, M.C. (2010). Laser microdissection of Arabidopsis cells at the powdery mildew infection site reveals site-specific processes and regulators. Proc. Natl. Acad. Sci. USA **107:** 460–465.

Chevalier, C., Bourdon, M., Pirrello, J., Cheniclet, C., Gévaudant, F., and Frangne, N. (2014). Endoreduplication and fruit growth in tomato: evidence in favour of the karyoplasmic ratio theory. J. Exp. Bot. **65:** 2731–2746.

Chevalier, C., Nafati, M., Mathieu-Rivet, E., Bourdon, M., Frangne, N., Cheniclet, C., Renaudin, J.P., Gévaudant, F., and Hernould, M. (2011). Elucidating the functional role of endoreduplication in tomato fruit development. Ann. Bot. **107:** 1159–1169.

Cho, H.-T., and Cosgrove, D.J. (2002). Regulation of root hair initiation and expansin gene expression in Arabidopsis. Plant Cell **14:** 3237–3253.

Churchman, M.L., et al. (2006). SIAMESE, a plant-specific cell cycle regulator, controls endoreplication onset in *Arabidopsis thaliana*. Plant Cell **18:** 3145–3157.

Cookson, S.J., and Granier, C. (2006). A dynamic analysis of the shade-induced plasticity in Arabidopsis thaliana rosette leaf development reveals new components of the shade-adaptative response. Ann. Bot. **97:** 443–452.

Cookson, S.J., Radziejwoski, A., and Granier, C. (2006). Cell and leaf size plasticity in Arabidopsis: what is the role of endoreduplication? Plant Cell Environ. **29:** 1273–1283.

de Almeida Engler, J., Kyndt, T., Vieira, P., Van Cappelle, E., Boudolf, V., Sanchez, V., Escobar, C., De Veylder, L., Engler, G., Abad, P., and Gheysen, G. (2012). CCS52 and DEL1 genes are key components of the endocycle in nematode-induced feeding sites. Plant J. **72:** 185–198.

De Bodt, S., Carvajal, D., Hollunder, J., Van den Cruyce, J., Movahedi, S., and Inzé, D. (2010). CORNET: a user-friendly tool for data mining and integration. Plant Physiol. **152:** 1167–1179.

De Bodt, S., Hollunder, J., Nelissen, H., Meulemeester, N., and Inzé, D. (2012). CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations. New Phytol. **195:** 707–720.

De Veylder, L., Beeckman, T., Beemster, G.T., Krols, L., Terras, F., Landrieu, I., van der Schueren, E., Maes, S., Naudts, M., and Inzé, D. (2001). Functional analysis of cyclin-dependent kinase inhibitors of Arabidopsis. Plant Cell **13:** 1653–1668.

De Veylder, L., Larkin, J.C., and Schnittger, A. (2011). Molecular control and function of endoreplication in development and physiology. Trends Plant Sci. **16:** 624–634.

Dietrich, D., et al. (2017). Root hydrotropism is controlled via a cortex-specific growth mechanism. Nat. Plants **3:** 17057.

Dinneny, J.R., Long, T.A., Wang, J.Y., Jung, J.W., Mace, D., Pointer, S., Barron, C., Brady, S.M., Schiefelbein, J., and Benfey, P.N. (2008). Cell identity mediates the response of Arabidopsis roots to abiotic stress. Science **320:** 942–945.

Dyson, R.J., et al. (2014). Mechanical modelling quantifies the functional importance of outer tissue layers during root elongation and bending. New Phytol. **202:** 1212–1222.

Edgar, B.A., and Orr-Weaver, T.L. (2001). Endoreplication cell cycles: more for less. Cell **105:** 297–306.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. **30:** 207–210.

El Refy, A., Perazza, D., Zekraoui, L., Valay, J.-G., Bechtold, N., Brown, S., Hülskamp, M., Herzog, M., and Bonneville, J.M. (2003). The Arabidopsis KAKTUS gene encodes a HECT protein and controls the number of endoreduplication cycles. Mol. Genet. Genomics **270:** 403–414.

Engelen-Eigles, G., Jones, R.J., and Phillips, R.L. (2001). DNA endoreduplication in maize endosperm cells is reduced by high temperature during the mitotic phase. Crop Sci. **41:** 1114.

Fendrych, M., Van Hautegem, T., Van Durme, M., Olvera-Carrillo, Y., Huysmans, M., Karimi, M., Lippens, S., Guérin, C.J., Krebs, M., Schumacher, K., and Nowack, M.K. (2014). Programmed cell death controlled by ANAC033/SOMBRERO determines root cap organ size in Arabidopsis. Curr. Biol. **24:** 931–940.

Flemming, A.J., Shen, Z.Z., Cunha, A., Emmons, S.W., and Leroi, A.M. (2000). Somatic polyploidization and cellular proliferation drive body size evolution in nematodes. Proc. Natl. Acad. Sci. USA **97:** 5285–5290.

Fox, D.T., and Duronio, R.J. (2013). Endoreplication and polyploidy: insights into development and disease. Development **140:** 3–12.

Galbraith, D.W., Harkins, K.R., and Knapp, S. (1991). Systemic endopolyploidy in *Arabidopsis thaliana*. Plant Physiol. **96:** 985–989.

Gendreau, E., Höfte, H., Grandjean, O., Brown, S., and Traas, J. (1998). Phytochrome controls the number of endoreduplication cycles in the *Arabidopsis thaliana* hypocotyl. Plant J. **13:** 221–230.

Gendreau, E., Traas, J., Desnos, T., Grandjean, O., Caboche, M., and Höfte, H. (1997). Cellular basis of hypocotyl growth in *Arabidopsis thaliana*. Plant Physiol. **114:** 295–305.

Gentleman, R.C., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. **5:** R80.

Gille, S., de Souza, A., Xiong, G., Benz, M., Cheng, K., Schultink, A., Reca, I.-B., and Pauly, M. (2011). O-acetylation of Arabidopsis hemicellulose xyloglucan requires AXY4 or AXY4L, proteins with a TBL and DUF231 domain. Plant Cell **23:** 4041–4053.

Grantham, N.J., Wurman-Rodrich, J., Terrett, O.M., Lyczakowski, J.J., Stott, K., Iuga, D., Simmons, T.J., Durand-Tardif, M., Brown, S.P., Dupree, R., Busse-Wicher, M., and Dupree, P. (2017). An even pattern of xylan substitution is critical for interaction with cellulose in plant cell walls. Nat. Plants **3:** 859–865.

Hamdoun, S., Zhang, C., Gill, M., Kumar, N., Churchman, M., Larkin, J.C., Kwon, A., and Lu, H. (2016). Differential roles of two homologous cyclin-dependent kinase inhibitor genes in regulating cell cycle and innate immunity in Arabidopsis. Plant Physiol. **170:** 515–527.

Hayashi, K., Hasegawa, J., and Matsunaga, S. (2013). The boundary of the meristematic and elongation zones in roots: endoreduplication precedes rapid cell expansion. Sci. Rep. **3:** 2723.

Heidstra, R., Welch, D., and Scheres, B. (2004). Mosaic analyses using marked activation and deletion clones dissect Arabidopsis SCARECROW action in asymmetric cell division. Genes Dev. **18:** 1964–1969.

Hochberg, Y., and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. Stat. Med. **9:** 811–818.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4:** 249–264.

Ishida, T., Adachi, S., Yoshimura, M., Shimizu, K., Umeda, M., and Sugimoto, K. (2010). Auxin modulates the transition from the mitotic cycle to the endocycle in Arabidopsis. Development **137:** 63–71.

Iyer-Pascuzzi, A.S., Jackson, T., Cui, H., Petricka, J.J., Busch, W., Tsukagoshi, H., and Benfey, P.N. (2011). Cell identity regulators link development and stress responses in the Arabidopsis root. Dev. Cell **21:** 770–782.

Katagiri, Y., Hasegawa, J., Fujikura, U., Hoshino, R., Matsunaga, S., and Tsukaya, H. (2016). The coordination of ploidy and cell size differs between cell layers in leaves. Development **143:** 1120–1125.

Kirik, V., Bouyer, D., Schöbinger, U., Bechtold, N., Herzog, M., Bonneville, J.M., and Hülskamp, M. (2001). CPR5 is involved in cell proliferation and cell death control and encodes a novel transmembrane protein. Curr. Biol. **11:** 1891–1895.

Lefebvre, V., Fortabat, M.-N., Ducamp, A., North, H.M., Maia-Grondard, A., Trouverie, J., Boursiac, Y., Mouille, G., and Durand-Tardif, M. (2011). ESKIMO1 disruption in Arabidopsis alters vascular tissue and impairs water transport. PLoS One **6:** e16645.

Lewis, D.R., Olex, A.L., Lundy, S.R., Turkett, W.H., Fetrow, J.S., and Muday, G.K. (2013). A kinetic analysis of the auxin transcriptome reveals cell wall remodeling proteins that modulate lateral root development in Arabidopsis. Plant Cell **25:** 3329–3346.

Lingua, G., Fusconi, A., and Berta, G. (2001). The nucleus of differentiated root plant cells: modifications induced by arbuscular mycorrhizal fungi. Eur. J. Histochem. **45:** 9–20.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15:** 550.

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

Marguerat, S., and Bähler, J. (2012). Coordinating genome expression with cell size. Trends Genet. **28:** 560–565.

Melaragno, J.E., Mehrotra, B., and Coleman, A.W. (1993). Relationship between endopolyploidy and cell size in epidermal tissue of Arabidopsis. Plant Cell **5:** 1661–1668.

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and biological assays with tobacco tissue cultures. Physiol. Plant. **15:** 473–497.

Myers, P.N., Setter, T.L., Madison, J.T., and Thompson, J.F. (1990). Abscisic acid inhibition of endosperm cell division in cultured maize kernels. Plant Physiol. **94:** 1330–1336.

Noir, S., Bömer, M., Takahashi, N., Ishida, T., Tsui, T.-L., Balbi, V., Shanahan, H., Sugimoto, K., and Devoto, A. (2013). Jasmonate controls leaf growth by repressing cell proliferation and the onset of endoreduplication while maintaining a potential stand-by mode. Plant Physiol. **161:** 1930–1951.

Otero, S., Desvoyes, B., Peiró, R., and Gutierrez, C. (2016). Histone H3 dynamics reveal domains with distinct proliferation potential in the Arabidopsis root. Plant Cell **28:** 1361–1371.

Padovan-Merhar, O., Nair, G.P., Biaesch, A.G., Mayer, A., Scarfone, S., Foley, S.W., Wu, A.R., Churchman, L.S., Singh, A., and Raj, A. (2015). Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. Mol. Cell **58:** 339–352.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods **14:** 417–419.

Perazza, D., Herzog, M., Hülskamp, M., Brown, S., Dorne, A.M.A., and Bonneville, J.M.J. (1999). Trichome cell growth in *Arabidopsis thaliana* can be derepressed by mutations in at least five genes. Genetics **152:** 461–476.

Petersson, S.V., Johansson, A.I., Kowalczyk, M., Makoveychuk, A., Wang, J.Y., Moritz, T., Grebe, M., Benfey, P.N., Sandberg, G., and Ljung, K. (2009). An auxin gradient and maximum in the Arabidopsis root apex shown by high-resolution cell-specific analysis of IAA distribution and synthesis. Plant Cell **21:** 1659–1668.

Ramirez-Parra, E., López-Matas, M.A., Fründt, C., and Gutierrez, C. (2004). Role of an atypical E2F transcription factor in the control of Arabidopsis cell growth and differentiation. Plant Cell **16:** 2350–2363.

Roeder, A.H.K., Chickarmane, V., Cunha, A., Obara, B., Manjunath, B.S., and Meyerowitz, E.M. (2010). Variability in the control of cell division underlies sepal epidermal patterning in *Arabidopsis thaliana*. PLoS Biol. **8:** e1000367.

Sabelli, P.A., and Larkins, B.A. (2009). The development of endosperm in grasses. Plant Physiol. **149:** 14–26.

Scheller, H.V., and Ulvskov, P. (2010). Hemicelluloses. **61:** 263–289.

Schmidt, E.E., and Schibler, U. (1995). Cell size regulation, a mechanism that controls cellular RNA accumulation: consequences on regulation of the ubiquitous transcription factors Oct1 and NF-Y and the liver-enriched transcription factor DBP. J. Cell Biol. **128:** 467–483.

Schnittger, A., Weinl, C., Bouyer, D., Schöbinger, U., and Hülskamp, M. (2003). Misexpression of the cyclin-dependent kinase inhibitor ICK1/KRP1 in single-celled Arabidopsis trichomes reduces endoreduplication and cell size and induces cell death. Plant Cell **15:** 303–315.

Schnittger, A., Jürgens, G., and Hülskamp, M. (1998). Tissue layer and organ specificity of trichome formation are regulated by GLABRA1 and TRIPTYCHON in Arabidopsis. Development **125:** 2283–2289.

Scholes, D.R., and Paige, K.N. (2015). Plasticity in ploidy: a generalized response to stress. Trends Plant Sci. **20:** 165–175.

Sliwinska, E., Mathur, J., and Bewley, J.D. (2012). Synchronously developing collet hairs in Arabidopsis thaliana provide an easily

accessible system for studying nuclear movement and endoreduplication. J. Exp. Bot. **63:** 4165–4178.

Sliwinska, E., Mathur, J., and Bewley, J.D. (2015). On the relationship between endoreduplication and collet hair initiation and tip growth, as determined using six *Arabidopsis thaliana* root-hair mutants. J. Exp. Bot. **66:** 3285–3295.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. **3:** 3–25.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000 Res. **4:** 1521.

Sugimoto-Shirasu, K., and Roberts, K. (2003). "Big it up": endoreduplication and cell-size control in plants. Curr. Opin. Plant Biol. **6:** 544–553.

Sugimoto-Shirasu, K., Roberts, G.R., Stacey, N.J., McCann, M.C., Maxwell, A., and Roberts, K. (2005). RHL1 is an essential component of the plant DNA topoisomerase VI complex and is required for ploidy-dependent cell growth. Proc. Natl. Acad. Sci. USA **102:** 18736–18741.

Takahashi, N., Kajihara, T., Okamura, C., Kim, Y., Katagiri, Y., Okushima, Y., Matsunaga, S., Hwang, I., and Umeda, M. (2013). Cytokinins control endocycle onset by promoting the expression of an APC/C activator in Arabidopsis roots. Curr. Biol. **23:** 1812–1817.

Tenhaken, R. (2015). Cell wall remodeling under abiotic stress. Front. Plant Sci. **5:** 771.

Valente, P., Tao, W., and Verbelen, J.-P. (1998). Auxins and cytokinins control DNA endoreduplication and deduplication in single cells of tobacco. Plant Sci. **134:** 207–215.

Van't Hof, J. (1999). Increased nuclear DNA content in developing cotton fiber cells. Am. J. Bot. **86:** 776–779.

Vissenberg, K., Fry, S.C., and Verbelen, J.P. (2001). Root hair initiation is coupled to a highly localized increase of xyloglucan endotransglycosylase action in Arabidopsis roots. Plant Physiol. **127:** 1125–1135.

Wang, S., Gu, Y., Zebell, S.G., Anderson, L.K., Wang, W., Mohan, R., and Dong, X. (2014). A noncanonical role for the CKI-RB-E2F cell-cycle signaling pathway in plant effector-triggered immunity. Cell Host Microbe **16:** 787–794.

Wilson, E.B. (1925). The Cell in Development and Heredity. (New York: Macmillan Company).

Xiong, G., Cheng, K., and Pauly, M. (2013). Xylan O-acetylation impacts xylem development and enzymatic recalcitrance as indicated by the Arabidopsis mutant tbl29. Mol. Plant **6:** 1373–1375.

Yi, D., et al. (2014). The Arabidopsis SIAMESE-RELATED cyclin-dependent kinase inhibitors SMR5 and SMR7 regulate the DNA damage checkpoint in response to reactive oxygen species. Plant Cell **26:** 296–309.

Yuan, Y., Teng, Q., Zhong, R., and Ye, Z.-H. (2013). The Arabidopsis DUF231 domain-containing protein ESK1 mediates 2-O- and 3-O-acetylation of xylosyl residues in xylan. Plant Cell Physiol. **54:** 1186–1199.

Zhang, C., Barthelson, R.A., Lambert, G.M., and Galbraith, D.W. (2008). Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. Plant Physiol. **147:** 30–40.

Zhang, C., Gong, F.C., Lambert, G.M., and Galbraith, D.W. (2005). Cell type-specific characterization of nuclear DNA contents within complex tissues and organs. Plant Methods **1:** 7.

A Spatiotemporal DNA Endoploidy Map of the Arabidopsis Root Reveals Roles for the Endocycle in Root Development and Stress Adaptation
Rahul Bhosale, Veronique Boudolf, Fabiola Cuevas, Ran Lu, Thomas Eekhout, Zhubing Hu, Gert Van Isterdael, Georgina M. Lambert, Fan Xu, Moritz K. Nowack, Richard S. Smith, Ilse Vercauteren, Riet De Rycke, Veronique Storme, Tom Beeckman, John C. Larkin, Anna Kremer, Herman Höfte, David W. Galbraith, Robert P. Kumpf, Steven Maere and Lieven De Veylder
Plant Cell 2018;30;2330-2351; originally published online August 16, 2018;
DOI 10.1105/tpc.17.00983

This information is current as of January 14, 2019

Supplemental Data	**/content/suppl/2018/08/15/tpc.17.00983.DC1.html**
References	This article cites 98 articles, 40 of which can be accessed free at: /content/30/10/2330.full.html#ref-list-1
Permissions	https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&issn=1532298X&WT.mc_id=pd_hw1532298X
eTOCs	Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain
CiteTrack Alerts	Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain
Subscription Information	Subscription Information for *The Plant Cell* and *Plant Physiology* is available at: http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY