A common pathway controls cell size in the sepal and leaf epidermis leading to a non-random pattern of giant cells

Frances K. Clark[1,2]*, Gauthier Weissbart[3]*, Xihang Wang[2]*, Kate Harline[2], Chun-Biu Li[4], Pau Formosa-Jordan[3,5,6†], Adrienne H. K. Roeder[2,6†].

1 Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853.
2 Weill Institute for Cell and Molecular Biology and School of Integrative Plant Science, Section of Plant Biology, Cornell University, Ithaca, NY 14853.
3 Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany.
4 Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden.
5 Cluster of Excellence on Plant Sciences (CEPLAS), Max Planck Institute for Plant Breeding Research, Cologne, Germany.
6 Polyploidy Integration and Innovation Institute.

* Authors contributed equally to this work
† Authors for correspondence (PFJ) pformosa@mpipz.mpg.de, (AHKR) ahr75@cornell.edu

Short title (50 characters max): Leaf cell size patterning

Abstract:

Arabidopsis leaf epidermal cells have a wide range of sizes and ploidies, but how large cells are spatially patterned alongside smaller cells remains unclear. Here, we demonstrate that the same genetic pathway that creates giant cells in sepals is also responsible for their formation in the leaf epidermis. In both sepals and leaves, giant cells are scattered among smaller cells; therefore, we asked whether the spatial arrangement of giant cells is random. By comparing sepal and leaf epidermises with computationally generated randomized tissues we show that giant cells are clustered more than is expected by chance. Our cell-autonomous and stochastic computational model recapitulates the observed giant cell clustering, indicating that clustering emerges as a result of the cell division pattern. Overall, cell size patterning is developmentally regulated by common mechanisms in leaves and sepals rather than a simple byproduct of cell growth.

Teaser: The spatial pattern of giant cells becomes non-random as the surrounding cells divide.

Introduction

During development, initially identical cells differentiate into distinct cell types arranged in complex spatial patterns. How these patterns are formed is a central question in developmental biology. The plant epidermis is a suitable system for the study of cellular patterning. Epidermal cells comprise the outermost cell layer and are easy to view under a microscope and image at different developmental stages. Unlike in animals, where most tissue differentiation occurs during embryogenesis, plants continuously form new organs and therefore new epidermises are patterned throughout their life cycles. The *Arabidopsis thaliana* (hereafter Arabidopsis) mature leaf blade

epidermis contains three main cell types: stomatal guard cells, trichomes, and pavement cells (1). Stomatal guard cells surround stomatal pores through which gas exchange occurs, and trichomes are large branched hair cells that serve to discourage herbivory, among other functions (2). All other epidermal cells in the mature leaf blade epidermis (the expanded part of the leaf between the midrib and the margin) are classified as pavement cells. However, pavement cells are not a homogeneous group of cells, but rather exhibit a variety of sizes, ploidies, and shapes (3, 4). Much research has focused on the patterning of stomata (5-7) and trichomes (8, 9), leading to important insights into how the regulation of intercellular signaling, cell fate specification, the cell cycle, and polarized cell division orientation give rise to their spatial arrangement. However, the patterning of pavement cells is understudied. In particular, little is known about how some pavement cells are specified to become larger and more highly polyploid than others.

Pavement cell size patterning has been studied in the Arabidopsis sepal. Pavement cells in the sepal vary in size and ploidy, with some cells reaching up to 800 μm in length (Fig. 1A) and having ploidies up to 32C (10). These very large pavement cells that have a characteristic highly anisotropic shape and bulge out of the epidermis have been named 'giant cells' (10), and these form when a cell endoreduplicates early during growth (10). Endoreduplication occurs when a cell replicates its DNA but does not enter mitosis or divide and instead continues to grow and increases its ploidy. Once a cell enters endoreduplication, it terminally differentiates and almost never reenters the mitotic cycle (10). Similar numbers of giant cells form on sepals within an Arabidopsis plant and among plants, but the precise spatial arrangement of giant cells differs from sepal to sepal.

Forward-genetic screens have identified the genes involved in sepal giant cell patterning, and double mutant analysis has allowed these genes to be ordered within a genetic pathway (10-13) (Fig. 1). The homeodomain leucine zipper (HD-ZIP) Class IV transcription factor Arabidopsis thaliana MERISTEM LAYER1 (ATML1) promotes giant cell specification in a dose-dependent manner (11, 12). Loss of *ATML1* function in sepals greatly reduces giant cell number, and overexpression of *ATML1* leads to ectopic giant cell formation (Fig. 1A,C,G) (11, 12). ATML1 protein concentration fluctuates in the protodermal nuclei of developing sepals (12). High concentrations of ATML1 reached during the G2 phase of the cell cycle are strongly correlated with giant cell differentiation, consistent with a model in which an ATML1 concentration that surpasses a threshold in G2 results in giant cell specification, early endoreduplication, and giant cell differentiation (12). The receptor-like kinase ARABIDOPSIS CRINKLY 4 (ACR4) functions upstream of ATML1 to promote giant cell formation (11, 12, 14-16) (Fig. 1B,H). Loss of function of *ACR4* leads to a modest reduction in the number of giant cells (11) (Fig. 1A, B). The calpain protease DEFECTIVE KERNEL (DEK1) and the CDK inhibitor LOSS OF GIANT CELLS FROM ORGANS (LGO; also known as SIAMESE-RELATED 1, SMR1) function genetically downstream of ATML1 to promote giant cell formation (12) (Fig. 1H). A hypomorphic mutant *dek1* allele (*dek1-4*) results in the complete loss of giant cells from sepals (11) (Fig. 1D). Similarly, sepals from plants homozygous for a loss-of-function mutation in LGO have no giant cells (10, 11) (Fig. 1E), and overexpression of *LGO* increases giant cell number (11) (Fig. 1F). It is unknown whether this genetic pathway affects cell size only in the sepal or whether it is also a more general mechanism of epidermal cell size patterning in other organs.

Leaf pavement cell size is affected by the family of CDK inhibitors that includes LGO, known as the SIAMESE/SIAMESE-RELATED (SIM/SMR) family (17, 18). SMR proteins bind to cyclin CDK complexes and inhibit their phosphorylation of downstream targets (18). *lgo-1* mutants lack large pavement cells and have a reduction in endoreduplication of the leaf cells as compared with those of wild type (17, 18). In *lgo* mutants, pavement cells that should be mature continue to divide (19). Furthermore, overexpression of the closely related paralog of *LGO*, *SIM*, results in larger and more highly endoreduplicated leaf epidermal pavement cells (17). In sepals, LGO upregulates defense response gene expression, including glucosinolate biosynthesis genes (20), whereas in leaves, ATML1 promotes the formation of ER bodies, which contain components of the glucosinolate system, in large pavement cells (21), suggesting a common role of large cells in defense response. Whether the same upstream components of the sepal giant cell pathway also function in leaf cell size patterning has not been thoroughly investigated. One study did compare pavement cell size in *dek1-4* and wild-type cotyledons and found no evidence that the cells differed in ploidy (22). However, true leaves were not examined.

In leaves and sepals, it is unknown whether giant cells exhibit a spatially ordered pattern across the organ, or if instead their spatial arrangement is random. Other epidermal cell types are non-randomly distributed across the leaf tissue. For instance, trichomes do not form in adjacent cells due to lateral inhibition (via both activator–inhibitor and activator–depletion systems) (9), and stomata rarely differentiate in adjacent cells due to both lineage-specific division orientation and intercellular signaling (23). In contrast to stomata and trichomes, sepal giant cells can be in contact with one another. However, it is unknown whether giant cell contacts are likely to be formed by chance. Due to their large shapes, quantifying the spatial arrangement of giant cells has remained challenging, and standard methods for assessing point pattern randomness are not applicable (24-26).

When a leaf develops, epidermal and mesophyll cell layers grow and differentiate simultaneously. Coordination between epidermal and mesophyll layer development is apparent because stomata are positioned preferentially above mesophyll air spaces (27-29) and, in grasses, stomata are organized in rows along the sides of underlying veins (30, 31). For instance, the peptide STOMAGEN is produced in the developing mesophyll cell layer and moves to the developing epidermis to promote specification of stomatal progenitor cells (32). Whether the positioning of larger pavement cells in Arabidopsis leaves is correlated with specific features of underlying cell layers, such as vascular bundles, has yet to be investigated.

Here, we imaged and analyzed large areas of leaves to obtain a holistic understanding of both the size distributions and the spatial arrangements of epidermal pavement cells in the leaf blade (excluding midrib and margin cells). We compared pavement cell sizes between wild-type leaves and leaves of mutants in genes involved in sepal giant cell formation. We discovered that the genetic pathway that controls sepal giant cell formation also has a broader role in patterning epidermal pavement cell size in leaves. We quantified the spatial organization of large cells using simulated randomized tissues and found that large cells tend to cluster together in both mature leaves and sepals. Using modeling and data analysis, we found that giant cells emerge randomly in space at early stages of development, but their spatial arrangement becomes non-random over time due to divisions of the surrounding small cells. Our computational modeling supports the

notion that a non-random clustered pattern can emerge in a cell-autonomous and stochastic manner.

Results

Arabidopsis leaves exhibit a large range of cell sizes, similar to sepals

In sepals, giant cells are easily visible because they are highly elongated (Fig. S1A). Similarly, we observe large and highly anisotropic cells in cauline leaves that appear similar to sepal giant cells (Fig. S1B). In rosette leaves, pavement cells of the epidermis are puzzle shaped with lobes and necks, such that cell size is not readily apparent by eye (Fig. S1C). However, there is heterogeneity in pavement cell sizes in leaf epidermal tissues (4, 33). Therefore, we wondered to what extent the distribution of cell sizes observed in sepals, ranging from giant cells to small cells, also occurs in rosette leaves. We imaged large sections of the blade (excluding midrib and margin cells) of leaf 1 or 2 from wild-type plants expressing a plasma membrane marker (*p35S::mCitrine-RCI2A*) and a nuclear marker (*pUBQ::H2B-TFP*) at 25 days post germination (dpg). At 25 dpg, leaves 1 and 2 of the rosette are fully expanded and mature. Leaves 1 and 2 initiate simultaneously and are indistinguishable; therefore, we refer to them interchangeably as leaf 1 or 2. We segmented the epidermal cells of leaves 1 or 2 and sepals on both abaxial (bottom) and adaxial (top) sides using MorphoGraphX (34, 35) and computed their area (Fig. 2A–D). We observed that on the abaxial side, the cell size distributions for both sepals and leaves are asymmetric, with long tails representing large cells (Fig. 2E). However, the larger cells on the abaxial side of the sepals exceed the average cell size to a greater extent than those in the leaf, resulting in a more extended tail in the distribution (Fig. 2F). Still, we observed that the cell size range in the leaf and sepal are similar and the largest cells of the sepal are about the same size as the largest cells of the leaf (Fig. 2A–E). We conclude that Arabidopsis leaves have a diverse range of cell sizes characterized by a long-tailed distribution, similar to the abaxial side of sepals.

Large cells are formed on the adaxial side as well as the abaxial side of the leaf

In sepals, giant cells are restricted to the abaxial (outer) surface (Fig. 2A–B and Fig. S2A–D). We asked whether there was a difference in cell size between adaxial (top) and abaxial (bottom) surfaces of the leaf. Large cells of similar size are formed on both the adaxial and abaxial surfaces, in contrast to the sepal (Fig. 2A–F and Fig. S2A–F). However, we found that the leaf adaxial side has fewer cells per unit area (leaf replicate 1: 234 cells mm^{-2} on the abaxial side and 156 cells mm^{-2} on the adaxial side, leaf replicate 2: 284 cells mm^{-2} on the abaxial side and 177 cells mm^{-2} on the adaxial side) (Fig. 2C–D and Fig. S2E–F); therefore, many cells are slightly more expanded on the adaxial side (Fig. 2D and Fig. S2F). This difference in cell density is likely attributable to the greater number of stomata and stomatal lineage cells we observed on the abaxial side compared with the adaxial side (Fig. 2C–D and Fig. S2E–F). The greater number of stomata and stomatal lineage cells on the abaxial side is reflected in the increased proportion of very small cells in the abaxial cell size distribution compared with the adaxial cell size distribution (Fig. 2E). We also observed that the abaxial cells are more lobed than the adaxial cells (Fig. 2C–D and Fig. S2E–F). Despite slight differences, the cell size distributions of the abaxial and adaxial sides of the leaf are

quite similar, particularly in the tails, where both sides exhibit a similar range of larger cells, in contrast to the sepal, where only the abaxial side has very large cells.

Cell area correlates with DNA content

Cell area and ploidy are positively correlated in leaf epidermal cells (4). To validate this correlation here, we measured DNA content by quantifying total fluorescence of Histone 2B-TFP (*pUBQ::H2B-TFP*) within each cell nucleus of the 25-dpg leaf images, which approximates to cell ploidy. Each nucleus was matched to its corresponding cell by visual inspection. As expected, a strong linear correlation between DNA content and cell area was observed for both the abaxial surfaces (R^2 = 0.85 and 0.91; n = 2) and the adaxial surfaces (R^2 = 0.79 and 0.82; n = 2) (Fig. 2G and Fig. S2G–H). Therefore, we focus on analyzing cell size, and infer that large cell size indicates high ploidy.

We wondered whether cells of similar size on the abaxial and adaxial side of the same leaf also have a similar DNA content. We found that cells of similar DNA content are larger on the adaxial side than on the abaxial side (Fig. 2G and Fig. S2G–H), suggesting that adaxial cells have expanded more than abaxial cells.

Because the largest sepal cells and the largest leaf cells had approximately the same areas, we asked whether the DNA content of these cells was also similar. We plotted the total fluorescence of Histone 2B-TFP of the cells with largest area in both the leaf and sepal (largest cells were defined as cells with areas exceeding 4308 μm^2, which is the average of the 98th percentile cell areas of the three sepal replicates). We found that the total fluorescence values were very similar between sepal and leaf, suggesting that these largest cells are similar in ploidy (Fig. 2H).

Cell size patterning emerges at the tip and progresses basipetally as the leaf differentiates

To determine how the cell size pattern emerges in the leaf during development, we imaged both the adaxial and abaxial surfaces of each leaf at different stages of development from 5 dpg to 9 dpg. After quantifying cell size (Fig. 3A–B and Fig. S3A–B), we observed that from 5 dpg to 9 dpg, cell size increases greatly (Fig. S3), as expected. At day 5, cells throughout the blade are fairly homogeneous in size, with a few cells starting to expand near the distal tip, and the large cells of the margin and overlying midrib already apparent (Fig. 3A). Excluding the large margin cells and cells overlying the midrib, the cell size pattern consisting of large cells interspersed between small cells progressively develops basipetally from the tip (Fig. 3A–C), whereas at the base the cells remain uniformly small. The progression of cell size patterning down the leaf is consistent with the well-established basipetal wavefront of differentiation and cessation of cell division (36). The cell area distributions (Fig. 3D), excluding margin cells and cells overlying the midrib, showed that more large cells appear throughout development and the maximal cell size increases (Fig. 3A–B, D) more than the median cell size due to the large number of stomata and small pavement cells (Fig. 3D). By 9 dpg, cell size has been patterned almost to the base of the leaf (Fig. 3A–B).

227 We next asked whether the wavefront of cell size patterning progresses basipetally at the same rate
228 on the abaxial and adaxial sides of the leaf. Using images of both the abaxial and adaxial sides of
229 the same leaf, we plotted the positions of the centers of the largest cells on both sides to
230 qualitatively study the spatial locations of large cells. For each leaf, the area threshold determined
231 for large cells was the same for both abaxial and adaxial sides and was taken to be the cell area of
232 the 98th percentile on the abaxial side. We found that large cells (excluding the cells of the margin
233 and overlying the midrib) are at the same proximal–distal position on abaxial and adaxial sides
234 during development (Fig. 3C). These results suggest that the wavefront of patterning and
235 differentiation is coordinated across the abaxial/adaxial axis of the leaf.
236
237 Although the differentiation wavefront is coordinated across the abaxial/adaxial sides, we
238 observed that the large cells on the blade on one side frequently did not form directly opposite the
239 large cells on the blade on the other side (Fig. 3C), with the exception of the large cells of the
240 margin and those overlying the midrib. This finding suggests that the cell size patterning on each
241 side is established independently.
242
243
244 <u>The sepal giant cell specification pathway also patterns giant cells in leaves</u>
245
246 Because the cell size distributions have similarities in leaves and sepals, we tested whether the
247 giant cell specification pathway in sepals (Fig. 1H) also functions in the leaf to pattern cell size.
248 We imaged leaf 1 or 2 at both 9 dpg and 25 dpg from wild type and giant cell pathway mutants.
249 At 9 dpg, patterning has just extended to the base of the leaf, and the leaf is still small enough that
250 we could image the whole upper abaxial quadrant to determine the pattern over a large fraction of
251 the leaf blade (Fig. 4 and Figs. S4, S5). At 25 dpg, the leaf is fully differentiated, fully expanded,
252 and the pattern is established (Fig. 5 and Figs. S6, S7). We computed cell areas and compared
253 them across genotypes at each stage. We found that cell size patterning in the leaf is similarly
254 affected in the mutants at both 9 dpg and 25 dpg as in the mature sepal. Notably, the largest cells
255 show similar variations in their quantities across genotypes. Similar to the sepal, the size of the
256 largest cells is moderately reduced in *acr4-2* mutants (Figs. 1B, 4B, 4H–J, 5B, 5H–I, K), and more
257 greatly reduced in *atml1-3* mutants (Figs. 1C, 4C, 4H–J, 5C, 5H–I, K). The reduction in large cells
258 is drastic in *dek1-4* and *lgo-2* mutant sepals and leaves, resulting in the absence of a long tail in
259 the cell size distribution (Figs. 1D–E, 4D–E, H–J and 5D–E, H–I, K). For these genotypes, the
260 number of cells of medium size is also substantially decreased (Figs. 1D–E, 4D–E, I , 5D–E, I).
261 Conversely, the overexpression of *ATML1* (*ATML1-OX*) or *LGO* (*LGO-OX*) leads to an increase
262 in the size of large cells and in fewer small cells compared to wild type, as in the sepal (Figs. 1F–
263 G, 4F–J, 5F–I, K).
264
265 To quantify the variations in the number of large cells precisely, we quantitatively defined leaf
266 giant cells on the basis of a cell area threshold. Specifically, we first classified pavement cells and
267 stomata using a Support Vector Machine (SVM) classifier based on features of cell shape
268 (Materials and Methods, Fig. S8). Next, a cell size threshold was established in the mature sepal
269 and in the leaf, at both 9 dpg and 25 dpg, using the *atml1-3* mutants, which are known to have very
270 few giant cells in sepals (Materials and Methods, Fig. 1C and Fig. S8). Those cells in the 9-dpg
271 and 25-dpg leaves as well as in the sepal that exceeded their associated threshold were categorized
272 as giant cells (see cell-type classification outcomes in Fig. 6 and Fig. S9). On the basis of this

273 definition, we performed a quantitative comparison and statistically compared the number of giant
274 cells per unit area among genotypes in leaves. Two-sample, two-tailed *t*-tests showed that in the
275 9-dpg leaf and the mature leaf, wild-type had significantly more giant cells than *lgo-2* (9 dpg: $p =$
276 0.002, 25 dpg: $p = 0.003$), *dek1-4* (9 dpg: $p = 0.002$, 25 dpg: $p = 0.002$), *atml1-3* (9 dpg: $p = 0.002$,
277 25 dpg: $p = 0.005$), and *acr4-2* (9 dpg: $p = 0.010$, 25 dpg: $p = 0.044$). Conversely, *LGO-OX* had
278 significantly more giant cells than wild type (9 dpg: $p = 0.001$, 25 dpg: $p = 0.003$). Although no
279 difference in the number of giant cells per unit area was observed between wild type and *ATML1-*
280 *OX* (9 dpg: $p = 0.213$, 25 dpg: $p = 0.75$), the fractional area occupied by giant cells was significantly
281 higher in *ATML1-OX* (9 dpg: $p < 0.005$, 25 dpg: $p < 0.005$).
282
283 Collectively, the similarities in the variation between the number of giant cells in the leaf and the
284 sepal indicates that the sepal giant cell specification pathway also regulates the formation of giant
285 cells in leaves.
286
287
288 Giant cell mutants affect the entire cell size distribution
289
290 We observed that not only are giant cells affected in these mutants, but the entire cell-size
291 distribution is also affected. For example, the number of medium-sized cells in *lgo-2* and *dek1-4*
292 is reduced in addition to the number of giant cells (Figs. 4H–J and 5H–I, K) and, correspondingly,
293 the number of small cells is increased in these mutants. To statistically analyze the difference in
294 cell size distributions, we conducted a principal coordinate analysis based on the Wasserstein
295 distances between cell size distributions (termed Wasserstein distance plot in this study), which
296 showed the difference between leaf samples according to their cell size distributions on a 2-
297 dimensional plane (Figs. 4K, 5J and S10, see Materials and Methods). In this plot, samples
298 clustered according to genotype, indicating that genotype controls cell size distribution. We
299 observed a progressive increase in the number of giant cells along the first principal coordinate V1
300 from *lgo-2* mutants to *ATML1-OX* and *LGO-OX* (Fig. 4K and 5J). *ATML1-OX* and *LGO-OX* were
301 distant from each other in this plot, which might partly reflect the fact that *LGO-OX* has more giant
302 cells, whereas *ATML1-OX* has fewer but larger giant cells. When we created the combined
303 Wasserstein distance plot with both the 9-dpg and 25-dpg leaves (normalized to an average cell
304 size of 1), the samples continued to group according to genotype rather than developmental stage,
305 further supporting that these genes affected the cell size distribution by 9 dpg (Fig. 5L). Thus, we
306 conclude that these genes affect the entire cell size distribution.
307
308 However, some differences in the cell size distribution are apparent between 9-dpg and mature 25-
309 dpg leaves. Firstly, at 9 dpg, *dek1-4* and *lgo-2* mutants are very similar; however, in the fully
310 mature 25-dpg leaves, the *lgo*-2 cell size range is notably smaller than that in the *dek1-4* mutant
311 (Figs. 4D–E, H–I and 5D–E, H–I), suggesting that *lgo-2* cells continue to divide after 9 dpg. In
312 addition, the small cells in *lgo-2* mutants were more uniform in size than all of the other genotypes
313 because the typical small stomatal lineage cells that encircle the stomata in mature leaves were
314 fewer in *lgo-2* (Figs. 4E, H–I, and 5E, H–I). This altered cell size distribution relates to the previous
315 finding that *LGO* affects pavement cell differentiation in these stomatal lineage ground cells and
316 that cells undergo division for a longer time in the absence of LGO (19). Secondly, although at 9
317 dpg the *LGO-OX* giant cells were slightly smaller than the *ATML1-OX* giant cells, at 25 dpg, the
318 *LGO-OX* giant cells were nearly equivalent in size to *ATML1-OX* giant cells (Figs. 4F–I and 5F–

319 I). In addition, we observed that more pavement cells were larger in *LGO-OX*, whereas only a few
320 cells became giant in *ATML1-OX* (Figs. 4F–G, I, 5F–G, I, 6B and Fig. S9). *ATML1-OX* leaves had
321 a few connected giant cells separating large islands of small cells, whereas *LGO-OX* leaves showed
322 more giant cells interspersed among smaller clusters of small cells (Figs. 4F–G, 5F–G, 6B and Fig.
323 S9). These phenotypic differences might reflect inherent differences in ATML1 and LGO activities
324 or the fact that *ATML1* and *LGO* overexpression transgenes are under the control of different
325 promoters that might have differences in activity at different developmental stages.
326
327
328 Relationship between the size and shape of cells and organs
329
330 In plants, compensation is the process by which the presence of fewer or more cells is accompanied
331 by a change in cell size, which maintains a constant organ size (37). Likewise, we observed
332 compensation in our leaf giant cell mutants (Fig. S11). Mature leaves of the mutants *acr4-2*, *atml1-*
333 *3*, *dek1-4* and *lgo-2*, which have fewer giant cells and more small pavement cells, are similar in
334 size to wild-type leaves (Fig. S11I–M, P). However, *ATML1-OX* and *LGO-OX* mature leaves,
335 which have much larger cells (see e.g. Fig. 5F–G), are smaller than wild-type (Fig. S11N–P).
336 Therefore, only partial compensation for having fewer cells by having larger cells is observed in
337 *ATML1-OX* and *LGO-OX* plants..
338
339 Additionally, *ATML1-OX* leaves are narrower than those of wild type and *LGO-OX* (Fig. S11A,
340 F, G, I, N, O). We also observed that giant cells are more directionally elongated in *ATML1-OX*
341 than in other genotypes (Figs. 4F–G, 5F–G and Figs. S4F–G, S5F–G, S6F–G, S7G–H), reflecting
342 the elongated shape of the leaf. This suggests the existence of a relationship between giant cell
343 shape and leaf morphology. Likewise, wild-type cauline leaves are both narrower and more
344 elongated than wild-type rosette leaves, and also have more anisotropic elongated giant cells than
345 in rosette leaves (Fig. S1). This observation supports the idea that cell shape reflects the anisotropy
346 of the growing tissue (33).
347
348
349 Spatial patterning of giant cells within the leaf blade
350
351 In wild-type plants, giant cells vary in position from sepal to sepal and from leaf to leaf (10-12).
352 An open question has been whether the spatial organization of giant cells is random, or whether
353 there is an underlying order. Classically, many specialized cell types such as stomata and trichomes
354 are spaced such that they are not in direct contact to one another (23, 38). Giant cells are frequently
355 adjacent to each other and, therefore, it is clear that there is not a strong lateral inhibition between
356 them. We set out to determine firstly whether giant cell position is correlated with underlying
357 vasculature and secondly, how giant cells are spatially positioned relative to one another.
358
359 Giant cells are not preferentially positioned overlying the vasculature
360
361 We wondered whether giant cell positioning was correlated with the position of leaf vasculature
362 for two reasons. Firstly, we observed that large, highly endoreduplicated cells overlie the midrib
363 of the leaf, extending all the way to the leaf tip (Fig. S12A). We wondered whether giant cells
364 might be similarly preferentially located over the other veins. Secondly, we observed that large,

365 highly endoreduplicated cells often appear to "peel" away from the midrib, as if following vascular
366 branches (Fig. S12A). This phenomenon is most common in *ATML1-OX* leaves (Fig. S12C–F).
367 To investigate whether giant cells overlie veins, we traced the veins from the original confocal
368 image onto the heat map of cell area for a 9-dpg wild-type half leaf and four *ATML1-OX* half
369 leaves. We found that many giant cells did not overlie the vasculature (Fig. S12B–F). Specifically,
370 we noted that the points where giant cells peel off the midrib often do not align with where veins
371 extend from the midrib. Furthermore, the orientation of giant cells do not follow the direction of
372 the veins (Fig. S12B–F). Instead, veins in *ATML1-OX* plants frequently pass through patches of
373 small cells (Fig. S12C–F). We conclude that vascular and giant cell patterns are not obviously
374 correlated.
375
376 Giant cells are clustered more often than expected by chance
377

378 A cell-autonomous and stochastic mechanism has been proposed to explain giant cell formation in
379 the sepal (12). However, it remains unknown whether giant cells are randomly arranged within the
380 tissue. To statistically assess the randomness of the pattern, we needed a random reference (or null
381 model) to compare with our experimental replicates. Previous studies addressing this problem
382 considered cells as points (26, 39), or used a regular hexagonal grid to build a null model (40). In
383 our case, these assumptions are not applicable due to the complexity of giant cell shapes and the
384 heterogeneity of cell shapes and sizes that affect cellular arrangements (41). Therefore, we used
385 the dmSET image-based method (41, 42) to generate randomized tissues from the segmented
386 images (Fig. 7A–B and Fig. S13). In the randomized images, cell positions were randomly
387 shuffled, but cell sizes and shapes of the original tissues were preserved (Fig. S14, Materials and
388 Methods). We generated 400 randomized tissues for each biological replicate segmentation of both
389 the wild-type sepal and 25-dpg leaf. Subsequently, observables, such as the mean number of giant
390 cell neighbors per giant cell, which captures the amount of contacts between giant cells, were
391 computed in the experimental data (i.e., segmentation data in Fig. 7B) and compared with those
392 computed in the corresponding randomized tissues (Fig. 7A). The histogram of values calculated
393 from the randomized tissue replicates formed a null distribution, which indicates the values
394 expected by chance if giant cells were placed randomly. Comparison of the real biological data
395 with this null distribution allowed us to statistically test the randomness of the observable (Fig.
396 7A).

397 We performed this analysis on our experimental data of wild-type 25-dpg leaves and mature sepals
398 (Materials and Methods). When considering the six pooled replicates of leaves or sepals, the mean
399 number of giant cell neighbors per giant cell was greater than in the randomized tissues, and the
400 null hypothesis could be rejected ($p < 0.05$) (Fig. 7C). This result shows the presence of clustering
401 among giant cells both in the leaf and the sepal. In addition, the analysis of the distribution of the
402 number of giant cell neighbors for all giant cells (Fig. 7D) revealed that it was less probable to find
403 isolated giant cells, and more probable to find giant cells in contact with two or more other giant
404 cells compared with what was expected by chance. Notably, the distributions of the number of
405 giant cell neighbors in the leaf and in the sepal (Fig. 7D) presented a similar shape, highlighting
406 similarities in the spatial patterns of giant cells between both tissues. The non-random pattern of
407 giant cells was also supported by the analysis of other observables (Fig. S15). In summary, these
408 results suggest the existence of a non-random mechanism that favors contacts between giant cells
409 during the patterning process.

410

411 Different cell sizes are organized into different spatial patterns

412 To investigate whether the clustered pattern is exclusive to giant cells, we applied the same analysis
413 to distinct sub-populations of pavement cells in the leaf tissues. Four populations of pavement cells
414 were defined: giant cells (Fig. 8A–C), middle-sized cells (Fig. 8D–F), small cells (Fig. 8G–I), and
415 a population of randomly selected pavement cells of any size (Fig. 8J–L). The number of cells in
416 each category was determined such that the total cell area of the cell population was approximately
417 equal to the area occupied by the giant cells. In contrast to the clustered pattern of giant cells (Fig.
418 8A–C), middle-sized pavement cells exhibited a more random organization (Fig. 8D–F; the null
419 hypothesis could not be rejected, with $p = 0.195$), closer to the random arrangement found in
420 randomly selected pavement cells (Fig. 8J–L). Conversely, small pavement cells showed a
421 clustered organization (Fig. 8G–I), because the mean number of neighbors between small
422 pavement cells significantly exceeded the value observed in the randomized tissues. Notably, these
423 small cells were clustered around the stomata, and their spatial arrangement is probably a
424 consequence of the stomatal patterning process. Overall, these analyses highlight a relationship
425 between pavement cell size and cell spatial organization within the tissue. Furthermore, these
426 findings underscore the distinctive clustered arrangement of giant cells in comparison to middle-
427 sized and randomly selected pavement cells.

428

429 A cell-autonomous stochastic model can recapitulate giant cell clustering

430 To investigate how giant cell clustering emerges during leaf and sepal epidermal development, we
431 wondered whether the existing cell-autonomous and stochastic model for giant cell specification
432 in sepals (12) could also recapitulate the clustered feature of the giant cell pattern. In this
433 multicellular computational model, the concentration of ATML1 stochastically fluctuates, is
434 regulated by a self-catalytic feedback loop, and in turn, ATML1 regulates the expression of a
435 downstream cell-cycle regulator target (Fig. 9A). At the end of a cell cycle, a cell either divides or
436 endoreduplicates if the ATML1 target exceeds a specific threshold during the G2 phase. We used
437 the proposed model (12) to investigate the resulting spatial organization of giant cells in simulated
438 tissues (Fig. 9A–B; see Materials and Methods for further details).

439 To assess the randomness of the simulated giant cell pattern, we applied the same method as in the
440 experimental images (Fig. 7A) to images of the final simulation time point (Fig. 9B–C). Giant cells
441 were also defined by a size threshold, which was established such that all cells of ploidy 16C or
442 above were considered to be giant (Materials and Methods). The analysis was repeated for several
443 initial conditions to increase the statistical robustness of the analysis. We observed that the mean
444 number of giant cell neighbors per giant cell was greater than expected if giant cells were randomly
445 distributed ($p < 0.05$, see Fig. 9D), showing that the current cell-autonomous model can also
446 produce a clustered giant cell pattern. Furthermore, the distribution of the number of giant cell
447 neighbors per giant cell (Fig. 9E) was similar to the distribution observed in the experimental
448 sepals (Fig. 7D, bottom). This raises the question of what mechanisms are responsible for cell
449 clustering in a cell-autonomous, multicellular model of dividing cells.

450

451 Cell division contributes to the clustering of giant cells

452

To understand how the giant cell clustering behavior emerges in our computational model, we analyzed how the cellular spatial pattern changes over time. We hypothesize that the initial giant cell pattern arises randomly in space, due to the stochastic nature of ATML1 concentration fluctuations that trigger endoreduplication, but as non-giant cells continue to divide, giant cells become more clustered in fully grown tissue. To test this hypothesis in our simulations, we selected the first-arising giant cells and quantified their spatial organization both at an early time point and at the end of the simulation (Fig. 10A and Fig. S16). We found that the first giant cells to appear were indeed more randomly distributed at the initial time point, where the null hypothesis could not be rejected ($p = 0.185$, Fig. 10B), whereas they were clustered at the final time point ($p < 0.05$, Fig. 10B). Indeed, although the giant cell contacts were preserved over time in the segmentation (red bar in Fig. 10B), we observed a shift in the null distribution of the mean number of giant neighbors per giant cell between the initial and the final time point (Fig. 10B). This reflects the fact that as new cells arise from division, the number of potential cellular configurations (i.e., the number of possible spatial cellular arrangements) increases, which decreases the probability of observing giant cell clusters by chance.

To investigate the emergence of the giant cell spatial pattern over time in real tissues, we used time-lapse data of developing sepals (43), where cells were tracked over time, and we similarly quantified the patterns of the first-arising giant cells at the first available time point (sepal at stage 4, 24 h time point) and a later one (sepal at stage 9, 120-h time point; see Materials and Methods) (Fig. 10C and Fig. S16). Similar to the simulations, we observed that giant cells were more randomly distributed in younger sepals and were more clustered in the more developed sepals (Fig. 10D). This analysis indicates that the stochastic and cell-autonomous model is a plausible model to explain the spatial organization of giant cells. Specifically, it shows that cell clustering can emerge in a growing tissue without the need for cell–cell communication but instead as a result of cell divisions of non-giant cells, which alter the spatial pattern of giant cells over time.

Discussion

We investigated pavement cell size patterning in the Arabidopsis leaf epidermis. We found that the same genetic pathway that controls giant cell formation in sepals also controls cell size and giant cell formation in the leaf. Specifically, the receptor-like kinase ACR4, the transcription factor ATML1, the calpain protease DEK1, and the CDK inhibitor LGO are important for the formation of leaf giant cells. Just as in the sepal, overexpression of *LGO* results in an increased number of giant cells and overexpression of *ATML1* leads to a larger area occupied by giant cells. Although giant cells are only present on the abaxial epidermis of sepals, they are present on both the abaxial and adaxial surfaces of leaves. We found that during leaf development, the basipetal wavefront of cell expansion is coordinated between abaxial and adaxial sides, with giant cells present at the same distance from the tip on both sides of the developing leaf. We observed that giant cells are scattered across the surface, sometimes in contact with one another, in both leaves and sepals. Our analysis demonstrated that giant cells are more likely to be in contact than expected by chance in both organs. Furthermore, their spatial arrangement exhibited similarities between the leaf and the sepal. The division and endoreduplication pattern of cells in a stochastic model of ATML1 fluctuations was sufficient to produce this clustered pattern. Thus, we have demonstrated that the same cell size patterning mechanism is present in sepals and leaves, and the division pattern itself

contributes to a final non-random spatial pattern. We conclude that the process of endoreduplication is developmentally regulated rather than simply a byproduct of cell growth.

Many patterning systems rely on cell–cell communication to generate proper spacing (7, 44), and the emergence of clustered patterns in certain cell types is often attributed to cell–cell communication mechanisms in static tissues (41, 45, 46). However, giant cell specification occurs within the context of tissue growth and cell division. Therefore, it is important to consider the influence of these dynamic factors as well. We revisited our previous cell-autonomous model for giant cell specification in which ATML1 stochastically fluctuates, and confirmed that giant cell clustering could arise in that model as a result of cell division, without the need for cell–cell communication. To understand how clustering emerges, we tracked giant cells from their initial emergence both in our modeled tissue and in published experimental time-lapse data of growing sepals (43), and analyzed the evolution of their spatial pattern over time. This analysis suggested that the giant cell pattern initially arises randomly in space in the primordium and becomes more clustered in the fully grown tissue. Therefore, the decrease in the randomness of the giant cell pattern over time appears to be caused by the division of surrounding cells, including dividing stomata lineage cells. Nevertheless, in a proliferating tissue, other mechanisms might operate at the same time that result in giant cell clustering. For instance, correlative effects on cells belonging to the same lineage (e.g., sister or even cousin cells (47)) might influence cell fate decisions. Experimental data at a higher temporal resolution will be necessary to provide evidence to test this hypothesis. Ultimately, further experimental and theoretical studies are needed to determine the relative importance of cell–cell communication and cell proliferation in leading to giant cell clustering.

In the past, researchers have attempted to increase organ size by increasing cell size by promoting endoreduplication, but these efforts have not been successful (48). This is because compensation occurs, in which smaller cell size is accompanied by an increase in cell number, so that organ size is relatively conserved (49, 50). Consistent with this, we observed that leaf 1 or 2 of wild type, *atml1-3*, and *lgo-2* plants are approximately the same size at maturity. Furthermore, instead of having larger leaves, the *ATML1-OX* and *LGO-OX* genotypes that have larger cells actually have slightly smaller leaves than the wild type at maturity. These observations are consistent with what is observed for sepals, because *ATML1-OX* and *LGO-OX* sepals are slightly smaller than wild-type sepals (10, 49). We have previously shown that mitotic division substitutes for endoreduplication to compensate and maintain organ size in mutants lacking giant cells (10). Our images suggest that this mechanism also operates in leaves.

Although giant cell number does not greatly influence organ size, organ shape is altered in sepals and leaves. *ATML1-OX* and *LGO-OX* sepals are narrower than those of wild type and curve outward, so that the bud opens prematurely (11). We speculate that the anisotropy of sepal giant cells drives the change in sepal shape. In *ATML1-OX* leaves, where giant cells are highly anisotropic, we observe a similar change in leaf shape, in which *ATML1-OX* leaf 1 and 2 are more pointed and oblong compared with the rounded wild-type leaf 1 and 2. By contrast, giant cells in *LGO-OX* leaves are isotropic and are more similar to wild-type giant cells, and *LGO-OX* leaves are more rounded. Our results suggest that ATML1 is sufficient to induce anisotropic cell growth, whereas LGO is not. Tang et al. (2023) have shown that the change in shape between rounded juvenile rosette leaves and more elongated adult rosette leaves is accompanied by the appearance

of highly anisotropic giant cells at the leaf base (51). However, they showed that loss of these directional, elongated giant cells does not change adult rosette leaf shape; the adult rosette leaf 7 remains elongated in *lgo-2* leaves where giant cells are not present (51). Thus, the relationship between giant cell shape, anisotropic growth, and organ shape is complex. Further work at the single cell level will be needed to elucidate the influence of giant cells on the shapes of different tissues.

Despite the similarities between cell size patterning in leaves and sepals, subtle differences also exist. Firstly, the distribution of epidermal cell sizes in the leaf is broader than in the sepal where cells are fairly uniformly small except for a scattering of giant cells (Fig. 2E). Secondly, leaves have giant cells on both abaxial and adaxial blades, whereas sepals have giant cells only on the abaxial side and not on the adaxial side that faces the petals. The petal blade does not have giant cells on either abaxial or adaxial sides (52); thus, sepals might be an organ whose identity is transitional between vegetative and floral organs (53). We observe a similar phase change in the anisotropy of giant cells. Rosette leaf giant cells are puzzle shaped and relatively isotropic. Later in the plant life cycle, giant cells in cauline leaves begin to be more anisotropic along the proximal–distal axis and start to resemble sepal giant cells. This supports the hypothesis that cauline leaves represent an intermediate state between rosette leaves and sepals (54). Finally, sepal giant cells are highly anisotropic along the proximal–distal axis. Although sepals and leaves have notable yet subtle differences in cell size, cell size patterning is regulated by the same developmental pathway in both organs.

The genetic pathway that regulates giant cell specification has been co-opted from the epidermal specification pathway, which is a developmental pathway necessary for epidermal and thus plant development (14, 55, 56). Without proper epidermal specification, the plant embryo will not progress past the globular stage of development (55-57). The fact that this fundamental epidermal developmental pathway also patterns giant cells illustrates a common theme in development, namely, that regulatory proteins are commonly reused for more than one developmental process (58).

Taken together, our analysis and theoretical work on patterning during tissue growth highlights that unexpected effects can occur and that these are difficult to infer from the canonical view of pattern formation arising in a static tissue. In this instance, an initially random pattern of giant cells becomes non-random as the surrounding cells divide. Thus, the effects of cell proliferation might also be important to determine the spatial distribution of specialized cell types in other tissues.

Materials and Methods

<u>Plant growth conditions</u>

All seeds were sown on LMlll soil in pots and were stratified in the dark for 3 days at 4°C. The pots were then transferred to Percival plant growth chambers set to 60% humidity, 22°C temperature, and 24-h light provided by Philips 800 Series 32-Watt fluorescent bulbs (f32t8/tl841) (~100 µmol m^{-2} s^{-1}). Days post germination (dpg) were counted from the time the pots were transferred to plant growth chambers.

591
592 Cloning fluorescent nuclear markers
593
594 To create a teal fluorescent nuclear (TFP) marker ubiquitously expressed under the UBIQUITIN
595 10 promoter (*pUBQ10::H2B-TFP*; pAR393), an H2B-TFP fusion with an AAAPAAAAA linker
596 was generated by PCR. TFP was amplified by PCR with primers oAR440 (5′-gct gcc gct cca gct
597 gca gct gcc gct ATG GTT TCT AAG GGA GAA GAA ACT ACT ATG-3′) and oAR438 (5′-cct
598 cga gtc aCT TAT AAA GTT CAT CCA TAC CAT CAG TAG-3′). The lower-case letters in the
599 primer sequences represent linkers, restriction sites, and cloning sequences that were added to the
600 gene sequences. H2B was PCR amplified with oAR369 (5′-CAC CGG ATC CAC AAT GGC
601 GAA GGC AGA TAA G-3′) and oAR439 (5′-agc ggc agc tgc agc tgg agc ggc agc AGA ACT
602 CGT AAA CTT CGT AAC CGC CTT AG-3′). Sequences encoding H2B-TFP were fused via
603 overlapping PCR with oAR369 and oAR438 primers. The H2B-TFP PCR product was cloned into
604 pENTR D TOPO to create the pAR198 entry clone. H2B-TFP was recombined into pUB-Dest
605 with LR Clonase II according to the manufacturer's instructions to generate *pUBQ10::H2B-TFP*
606 (pAR393). pAR393 was transformed into Arabidopsis Col-0 plants expressing the pLH13
607 *p35S::mCitrine-RCI2A* yellow fluorescent plasma membrane marker (49) via Agrobacterium
608 tumefaciens (strain GV3101)-mediated floral dipping (59) and selection with glufosinate-
609 ammonium ("Basta" Neta Scientific OAK-044851-25g).
610
611 Generation of mutant plant lines containing fluorescent plasma membrane and nuclear markers for
612 imaging
613
614 Mutant alleles and overexpression transgenes were crossed to plants expressing fluorescent cell-
615 membrane and nuclear markers to obtain plants for imaging. The following mutant alleles were
616 used: *acr4-2*, *atml1-3*, *dek1-4*, and *lgo-2*. In addition, two lines overexpressing either *ATML1*
617 (*pPDF1::FLAG-ATML1*) or *LGO* (*pATML1::LGO*) in a Col-0 background were used. All of these
618 alleles/transgenes are in the Columbia-0 (Col-0) accession. *acr4-2* (SAIL_240_B04) contains a T-
619 DNA insertion in the codon of the second of seven 39-amino acid repeats of the beta propeller
620 extracellular domain, which is upstream of the transmembrane domain and the kinase domain and
621 is therefore presumed to be loss-of-function allele. The *acr4-2* mutant was obtained from Gwyneth
622 Ingram (14), who obtained it from Syngenta (60). The *atml1-3* allele is a T-DNA insertion in the
623 homeodomain and is a loss-of-function mutant (11). The *atml1-3* was obtained from the
624 Arabidopsis Biological Resource Center (ABRC; accession number SALK_033408) (11). The
625 *dek1-4* allele contains a point mutation that changes a conserved arginine to a cysteine within
626 calpain domain III (ABRC accession CS68904) (11). Complete loss of function of *DEK1* is lethal
627 (55); therefore, *dek1-4* must retain some function. The *dek1-4* phenotype is recessive and therefore,
628 *dek1-4* is likely hypomorphic (11). The *dek1-4* mutant was originally isolated in the Landsberg
629 erecta accession and was subsequently back-crossed twice into Col-0 (12). *lgo-2* contains a T-
630 DNA insertion within the coding sequence of the gene and is a loss-of-function allele (10). *lgo-2*
631 was obtained from the ABRC (accession number SALK_033905 and is available as a homozygous
632 mutant as accession CS69160). *pPDF1::FLAG-ATML1* (*ATML1-OX*) was obtained from Gwyneth
633 Ingram (61). *pATML1::LGO* (*LGO-OX*) has been deposited for distribution at ABRC under
634 accession CS69162 (11, 49). *acr4-2*, *atml1-3*, *dek1-4*, *lgo-2*, *pPDF1::FLAG-ATML1*, and
635 *pATML1::LGO* were each crossed to plants expressing both a *p35S::mCitrine-RCI2A* fluorescent
636 plasma membrane marker (pLH13) and a *pUBQ::H2B-TFP* fluorescent nuclear marker (pAR393).

637 The F2 progeny were genotyped for *acr4-2*, *atml1-3*, *dek1-4*, and *lgo-2* (primer sequences in Table
638 S1) and lines were isolated that were homozygous for these alleles and that also expressed both
639 *pUBQ::H2B-TFP* and *p35S::mCitrine-RCI2A*. We could not obtain *atml1-3* homozygous plants
640 that also contained the *p35S::mCitrine-RCI2A* transgene after crossing, which was probably
641 because the *ATML1* gene was linked to the insertion site of the *p35S::mCitrine-RCI2A* transgene.
642 To obtain plants expressing *p35S::mCitrine-RCI2A* in a homozygous *atml1-3* background, the
643 plasmid containing *p35S::mCitrine-RCI2A* was transformed into *atml1-3* homozygous plants with
644 *pUBQ::H2B-TFP* through Agrobacterium tumefaciens (strain GV3101)-mediated floral dipping
645 (59). A T1 line was chosen that strongly expressed the mCitrine membrane signal and this line
646 was used for future experiments. For the overexpression transgenes *pPDF1::FLAG-ATML1* and
647 *pATML1::LGO*, seeds were collected from F2 plants and F3 plants were genotyped for
648 *pPDF1::FLAG-ATML1* or *pATML1::LGO* (Table S1). Those F2 plants that produced only F3
649 plants having *pPDF1::FLAG-ATML*1 or *pATML1::LGO* were isolated as homozygous for
650 *pPDF1::FLAG-ATML1* or *pATML1::LGO*, respectively. Those lines with *pPDF1::FLAG-ATML1*
651 or *pATML1::LGO* homozygous and that expressed both *pUBQ::H2B-TFP* and *p35S::mCitrine-*
652 *RCI2A* were used for imaging.
653
654 Sample preparation for imaging
655
656 Leaves and sepals were mounted in 0.001% (v/v) Triton for imaging. Leaves were imaged between
657 two coverslips, and sepals were imaged on a slide with a coverslip. Curvy leaves were cut with a
658 razor blade to ensure they could be placed flat under the coverslip. Samples were imaged
659 immediately after preparation. Sepals were imaged at stage 14 (62).
660
661 Imaging with confocal microscopy
662
663 A ZEISS LSM 710 Axio Examiner confocal microscope with a W Plan-Apochromat 20×/1.0 DIC
664 D-0.17 M27 75 mm water-immersion objective lens was used to image leaf 1 or 2 of the
665 Arabidopsis rosette and mature (stage 14) sepals. A 458 nm laser was used to excite *pUBQ::H2B-*
666 *TFP* (collection range 463–600 nm) and a 514 nm laser was used to excite *p35S::mCitrine-RCI2A*
667 (collection range 519–645 nm). Images were captured with a 1× zoom. The gain and laser power
668 varied slightly between images to accommodate slight differences in signal intensity between
669 samples. Each image was composed of several tiles. The dimensions of each voxel were 0.415 µm
670 (x) by 0.415 µm (y) by 1 µm (z).
671
672 For the images of abaxial and adaxial faces of the same organ (two leaf replicates and three sepal
673 replicates; Fig. 2 and Fig. S3), the 458-nm laser power and gain used for imaging *pUBQ::H2B-*
674 *TFP* were adjusted so that the TFP signal was below saturation and was then held constant for all
675 images.
676
677 Leaf areas were calculated from confocal images of entire leaves taken using a 2.5× objective for
678 each 9-dpg and 25-dpg leaf replicate.
679
680 Image processing
681

682 Tiles were stitched in the horizontal direction by ZEISS stitching software (overlap of 5% and
683 threshold of 0.7) and in the vertical direction with MorphoGraphX (34, 35) using the process
684 "Stacks/Multistack/Merge Stacks" (parameters: method = max; interpolation = linear). Assembled
685 images were saved as MorphoGraphX stack files. A surface mesh was created in MorphoGraphX
686 from each image to perform segmentation and analysis on the epidermis. First, extraneous parts of
687 the image were removed with the Voxel Edit tool. (Such extraneous parts of the image include
688 trichomes on the adaxial images and pollen grains/nematode eggs on some leaf images.) Then, an
689 image was subjected to Gaussian Blur using the process "Stack/Filter/Gaussian Blur Stack"
690 (parameters: x = 2; y = 2; z = 2). Next, the tissue surface was identified with the process
691 "Stack/Morphology/Edge Detect" (parameters: threshold varied between 2,300–7,000 according
692 to individual image brightness; multiplier = 2.0; adapt factor = 0.3; fill value = 30,000). These
693 steps extracted a surface of the leaf. The process "Stack/Morphology/Fill Holes" was applied to
694 some images when holes were apparent in the surface (parameters: x-radius = 20; y-radius = 20;
695 threshold = 10,000; depth = 0; fill value = 30,000). This surface was then used to generate a mesh
696 with the process "Mesh/Creation/Marching Cube Surface" (parameters: cube size = 5 µm;
697 threshold = 20,000). The mesh was smoothed with "Mesh/Structure/Smooth mesh" (parameters:
698 number of passes varied between 20–45; Walls Only = no). The mesh obtained was then
699 subdivided with the process "Mesh/Structure/Subdivide" either once or twice depending on its
700 size. In order to obtain the cell membrane signal on the surface, the process "Meshes/Signal/Project
701 Signal" (parameters: min/max distances ranged between 5–15 µm; MinSig = 0.0; MaxSig =
702 60,000) was used to project the *p35S::mCitrine-RCI2A* plasma membrane marker original signal
703 onto the mesh at an optimal depth. The depth range yielding the clearest cell membrane signals
704 with minimal distortion was selected. To perform cell segmentation, each individual cell in the
705 leaf was first manually identified with a cell label marking (seed). Using these seeds, watershed
706 segmentation was performed using the process "Meshes/Segmentation/Watershed Segmentation".
707 Adjacent pairs of stomatal guard cells were segmented together to form a single cell, which was
708 called stomata. Errors in segmentation were identified and corrected by removing the label for
709 those cells and reseeding. The Heat Map processes computed the cell area and other morphological
710 cell features as well as the position of every cell. The cell area data was exported into a data table
711 file for each image. In addition, other cellular shape features were computed and exported into a
712 data table for the cell type classification (Materials and Methods: cell type classification).
713
714 To analyze the nuclear signal from images of leaves expressing *pUBQ::H2B-TFP*, vertical
715 stitching of tiles (already horizontally stitched with Zeiss stitching software) was performed in
716 MorphoGraphX (max method and linear interpolation). Because the *TFP* reporter was expressed
717 under the *UBIQUITIN 10* promoter, TFP was localized in nuclei of the mesophyll cells in addition
718 to cells of the epidermis. Mesophyll nuclei were removed with the Voxel Edit tool. Nuclei were
719 identified as being from the mesophyll by lining up the nuclear signal images with their
720 corresponding membrane signal images and comparing the nuclei within the bounds of each
721 epidermal cell membrane. When compared with an epidermal cell nucleus, mesophyll cell nuclei
722 were often dimmer and lower down and therefore, excess nuclei were removed according to these
723 criteria so that each epidermal cell had one nucleus. When it was ambiguous which of two nuclei
724 in a single cell was from the mesophyll or epidermis, both nuclei were removed and excluded from
725 the analysis. Segmentation of the nuclei was performed in MorphoGraphX so that the total signal
726 could be calculated for each nucleus. To do so, the confocal image was first subjected to
727 "Stack/Filters/Brighten Darken" (parameter: 1). Next, a gaussian blur was performed using

728 "Stack/Filters/Gaussian Blur" (parameters: x = 1, y = 1, z = 1), followed by a binarization with the
729 process "Stack/Filters/Binarize" (parameters: threshold = 2,000), which functioned to select pixels
730 above a threshold value to identify the edges of each nucleus. A lower threshold value was chosen
731 so that we could identify the entire nuclei even for dim nuclei. The Voxel edit tool was used to
732 separate nuclei that inflated into one another. We then created a mesh from the binarized image
733 using the process "Mesh/Creation. Marching Cubes 3D" (parameters: cube size = 1, min voxel =
734 0, smooth passes = 3, label = 0). To ensure that the mesh covered all fluorescence of each nucleus,
735 we expanded the mesh using "Mesh/Structure/Shrink Mesh" with a negative value (parameter:
736 distance = -1). Individual nuclei were manually seeded and then the watershed segmentation was
737 performed with the process "Mesh/Segmentation/Watershed Segmentation" to identify each
738 nucleus. The Heat Map function calculated the total H2B-TFP fluorescence within each nucleus,
739 as a representation of DNA content. To study correlations between total nuclear H2B-TFP signal
740 and cell size, individual cells from cell area meshes were matched with their constituent nuclei
741 from nuclear signal meshes using MorphoGraphX parent tracking. For the leaf replicates, total
742 nuclear H2B-TFP signal was calculated for as many cells as possible from both the abaxial and
743 adaxial sides. For the sepal replicates, total nuclear H2B-TFP signal was calculated only on the
744 abaxial side and only for the largest cells.
745
746 To create the heat maps overlaid with vasculature in Fig. S12, confocal images of leaves expressing
747 *p35S::mCitrine-RCI2A* were used to create surfaces and were segmented as described above to
748 create cell area heat maps. The mCitrine-RCI2A confocal images were found to have signal in the
749 vasculature, so that the trajectories of veins could be traced in images from the abaxial surface of
750 the image, one can see. The mCitrine-RCI2A confocal images were transformed around the *z*-axis
751 in MorphoGraphX. For each leaf, the cell area heat map and the mCitrine-RCI2A confocal image
752 transformed around the *z*-axis were aligned in MorphoGraphX and PNG screen captures were
753 taken of each. These PNGs were then loaded into Adobe Illustrator and the veins were traced in
754 white onto the heat maps.
755
756 Please note that wild-type 25 dpg leaf replicates 1, 3, and 4, *lgo-2* 25 dpg leaf replicates 1 and 2,
757 and *LGO-OX* replicates 1, 2, and 3 were used in for an independent analysis of cell shape
758 (specifically lobeyness) in (63).
759
760
761 <u>Statistical Analysis</u>
762
763 To analyze the relationship between total nuclear H2B-TFP signal (DNA content) and cell area for
764 the leaves in Fig. 2 and Fig. S2, linear regressions were performed on R statistical software
765 (https://www.r-project.org/). To compare the total nuclear H2B-TFP signal (DNA content) of the
766 cells of largest area between sepals and leaves, the cell area at the 98th percentile was calculated
767 for each of the three abaxial sepal replicates and these three cell areas were averaged for an area
768 threshold of 4,308 μm^2. Cell area versus total nuclear H2B-TFP signal (DNA content) was plotted
769 for cells above this 4,308 μm^2 area threshold for the abaxial sepals and the abaxial and adaxial
770 leaves.
771
772 To compare positions of the largest cells on the abaxial and adaxial sides of each leaf at different
773 stages of development (Fig. 3), the abaxial and adaxial images were aligned in MorphoGraphX.

774 Then, cell area heat maps were created and the x and y coordinates of the center of each cell were
775 calculated (Fig. 3C). Cell area thresholds for each leaf were determined from the 98th percentile
776 cell area of the abaxial side, and the positions of cells above these area thresholds were plotted for
777 the abaxial and adaxial images of each leaf.
778
779 R statistical software was used to analyze the cell size distributions and create the violin plots and
780 Wasserstein plots. To create the Wasserstein plots, a Wasserstein test was performed between each
781 pair of replicate distributions. A test statistic (also known as Wasserstein distance) and p-values
782 were returned for every test. The p-values are listed in Fig. S10A–B. Classical multidimensional
783 scaling was performed to create a 2D coordinate for each replicate distribution based on the
784 Wasserstein distances, and points from these coordinates were plotted. To ensure that the distances
785 between 2D points adequately reflected the Wasserstein distances among replicate distributions,
786 we plotted the Wasserstein distances against the Euclidean distances between points (Fig. S10C–
787 D). The linear relationships between Wasserstein distances and Euclidean distances showed that
788 the 2D graph accurately represents the differences between distributions.
789
790 To create the Wasserstein plot of the combined 9-dpg and 25-dpg cell area data, cell areas of each
791 replicate were normalized by the mean cell area for that replicate. In this way, each replicate has a
792 mean of 1. This eliminated the difference in the values of the cell size between the 9-dpg and 25-
793 dpg leaves, such that the variances of the cell areas can be compared rather than the absolute sizes.
794
795 To statistically compare the differences in the number of giant cells across genotypes in the leaf at
796 9 dpg and at 25 dpg, two-sample, two-tailed *t*-tests that assumed equal variance were performed
797 on the number of giant cells per segmented area between wild type and the different genotypes.
798 Similarly, two-sample, two-tailed *t*-tests were performed to compare leaf sizes across genotypes
799 (Fig. S11).
800
801 To statistically assess the randomness of the cellular patterns, see section "Statistical analysis of
802 the cellular patterns" below.
803
804 <u>Cell type classification</u>

805 To automatically distinguish stomata from pavement cells, a supervised classification algorithm
806 was used based on cell shape features (Fig. S8). Cell shape features were computed from each
807 2.5D mesh using the MorphoGraphX process "Mesh/Heat Map/Analysis/Cell Analysis 2D" and
808 were extracted with "Mesh/Attributes/Save as CSV" into a data table. Three distinct training
809 datasets were created using a single wild-type replicate – one for the sepal, one for the leaf at 25
810 dpg and one for the leaf at 9 dpg. To get the different training datasets, we manually selected some
811 pavement cells and stomata and labeled them as different cell types, ran the classification processes
812 available within MorphoGraphX, and manually corrected the cells that were wrongly identified.
813 These training datasets were then used to train a supervised learning algorithm (Support Vector
814 Machine quadratic) using the Classification App in MATLAB (R2019b,
815 https://www.mathworks.com). The following cellular shape features were selected to train the
816 classifier in the 25-dpg leaf: area, average radius, length of the major axis, maximum radius,
817 perimeter, circularity, lobeyness (ratio of the cell perimeter over that of its convex hull), and
818 rectangularity (ratio of the cell area over the area of the minimum bounding rectangle in the cell).
819 For the sepal, the aspect ratio and the length of the minor axis were also taken into account. For

820 the 9-dpg leaf, where the variety of cell types was more complex, three cell types were defined
821 (pavement cell, meristemoid and stomata) and meristemoid and stomata were combined in the
822 post-processing script. The shape features used to train the classifier were area, average radius,
823 minimum radius, perimeter, circularity, lobeyness, and visibility stomata (it counts the proportion
824 of straight lines that connects the cell outline without passing through a cell boundary, as described
825 in (64)). To automatically predict cell types in all replicates, a developed Matlab script containing
826 the trained classifier and a post-classifier filter, which corrects for potentially wrong predictions
827 on the basis of known shape criteria, was applied. Manual corrections were finally performed, in
828 which misclassified cells were re-labeled with its correct cell type.

829 Giant cells were defined by a cell-size threshold (Fig. S8). Because a few giant cells were expected
830 in *atml1-3* mutants, *atml1-3* mutants were used as a reference to build this threshold. Fewer than
831 0.7% of the pavement cells were considered to be giant cells in *atml1-3* tissues, which was
832 supported through visual observation in the sepal. Consequently, the giant cell size threshold was
833 set as that corresponding to the average between the 99.3rd percentile cell size value with the cell
834 size value immediately above it in the distribution, taking into account the data of three *atml1-3*
835 pooled replicates. For consistency, the same method was applied to the sepal, and to 9-dpg and 25-
836 dpg leaves, which gave three different threshold values (sepal: 5,290 μm^2, leaf 9 dpg: 2,570 μm^2,
837 leaf 25 dpg: 14,160 μm^2). The percentiles were only calculated on rectangular sections (omitting
838 cells at the outline of the organs) of the sepal to maintain consistency across different organs.
839 Classification output examples in different genotypes are shown in Fig. 6 and Figs. S8, S9.

840 <u>Randomization of the experimental images</u>

841 To assess the randomness of the cellular patterns, it was essential to establish a random reference,
842 or null model, against which the observed pattern could be compared. To produce the required
843 random reference, the image-based method dmSET (41, 42) was applied to generate 400 synthetic
844 random equivalent tissues from each segmented image. Cell positions and orientations were
845 randomly shuffled into new images (named randomizations), while preserving individual
846 approximate cell shapes and sizes (Fig. S14). Only the incomplete cells at the border of the images
847 were fixed. This approach avoids potential biases arising from the heterogeneity of cell sizes and
848 shapes in the tissues, which affect the number of neighboring cells. We ensured that cellular
849 properties, and more specifically cell area and cell circularity ($4\pi\times$ area/(perimeter)^2), were
850 approximately conserved in the randomized tissues (Fig. S14; the Pearson coefficient was > 0.98
851 for cell area correlation and was close to 0.90 for cell circularity correlation). In the sepal
852 randomized tissues, cell orientations were constrained between $-\pi/6$ and $+\pi/6$ compared with their
853 initial orientation, to maintain the anisotropy of the tissue. A custom-made Matlab script was
854 subsequently applied to both original and randomized images to correct errors introduced by the
855 dmSET method and to compute cell shape properties and cellular network information that was
856 used to quantify the cellular pattern.
857

858 Before randomizing the different sepal and leaf replicate images (Fig. 7, and Figs. S13, S17, S18),
859 each 2.5D mesh was first converted into a 2D pixel image using the process "Stack/Mesh
860 Interaction/Mesh To Image" (with a pixel size of 1µm) in MorphoGraphX. Subsequently, a square
861 crop (in the leaf) or rectangular crop (in the sepal) that maximized the tissue section was performed
862 in the segmented images. These 2D segmented images were then randomized using the dmSET

863 method.
864

865 To study the change in the giant cell spatial pattern over time, published time-lapse sepal data were
866 used (43) that were randomized at two different time points (sepal at stage 4: 24 h, and at stage 8:
867 96 h). Cell segmentation and cell lineage tracking were already performed in (43). Using
868 MorphoGraphX, sepal cells were manually selected at the later time point, and the exact
869 corresponding mother cells at the first time point were established using the lineage tracking
870 analysis from (43). In order to quantify the spatial pattern of the same giant cells at two different
871 time points, giant cells at both stages were defined as the pavement cells that did not divide during
872 this period of time. This approach allowed the comparison of the change in tissue organization
873 consistently at two different time points. Then, the 2.5D meshes were projected into 2D images.
874 These images were subsequently randomized using the dmSET method. Here, to facilitate the
875 study of the same giant cells over time, the images were not cropped and the entire studied tissue
876 was randomized, including the cells at the edges. To achieve this, the background region, located
877 outside the tissue of interest, was considered as a single cell that remained fixed in the randomized
878 tissues. Examples of randomizations are shown in Fig. S16C–D. Three different sepal replicates
879 were used for these analyses of the time-lapse data.

880 "Segmentation" and "Randomization" images appearing in figures such as Fig. 7B were generated
881 with Python.
882
883 Statistical analysis of the cellular patterns

884 By comparing a spatial observable in the cellular network of the actual segmentation with the
885 corresponding observable in the cellular networks of the 400 generated randomized tissues,
886 whether the considered observable is likely to be observed by chance can be statistically tested
887 (41). Hence, the use of this method on observables measuring distances or contacts between the
888 studied cells allows the assessment of whether the arrangement of the cells within the tissue is
889 random, clustered or dispersed (Fig. 7A).

890 To quantify the patterns, a custom-made Python script was used to extract pertinent observables
891 (i.e., spatial quantities) from the cellular network, which used the NetworkX Python library
892 (https://networkx.org/). In this manuscript, we mainly focused on the number of giant cell
893 neighbors per giant cell to quantify the number of local contacts between giant cells. Other
894 observables have been quantified, such as the minimum shortest path between giant cells, and the
895 number of giant cells in a cluster (Fig. S15). When dealing with cropped images, giant cells (or
896 any cell population studied, see in Fig. 8) at the border of the image were not considered in the
897 analysis.

898 The number of giant cell neighbors was extracted for every giant cell, and the mean number of
899 giant cell neighbors per giant cell across all giant cells was computed within each experimental
900 replicate. Similarly to the methodology described by the authors of the dmSET method (41), the
901 mean value extracted from the segmentation image was compared with the approximated null
902 distribution formed by the 400 mean values extracted from the randomized images. We first
903 performed the analysis on each replicate independently (Figs. S17 and S18). As the cell size
904 distributions in the different replicates showed similarities across replicates (Figs. 5 and Figs. S6,
905 S7 and S10), replicates were pooled to increase the sample size and statistical power. Six image

replicates were used for both the leaf and sepal wild-type (Fig. 7), and three replicates were used for the wild-type sepal time-lapsed images (Fig. 10). To test the null hypothesis assessing the randomness of the observed metric, a p-value p was obtained as the ratio of the number of random images (defined here as one random image resulting from pooling one random image per replicate) displaying the same or a more extreme value than the one obtained in the segmentation replicates (one-sided test). If the value fell within the null distribution with an associated high p-value ($p > 0.05$), the null hypothesis could not be rejected, indicating that the observed quantity could likely be expected by chance.

In addition, the distribution of the number of giant cell neighbors for all giant cells from the pooled experimental replicates was studied, which provided more insights into their spatial organization. This was compared qualitatively with the distribution expected in a random tissue, extracted from the 400 randomized tissues of all replicates. In addition, the distributions of the minimum shortest path between giant cells and of the number of giant cells in a giant cell cluster were examined (Fig. S15).

All plots derived from these analyses were performed with Python, with the use of the matplotlib (https://matplotlib.org/) and seaborn (https://seaborn.pydata.org/) packages.

Mathematical model for giant cell fate commitment and numerical simulations

To simulate the giant cell fate decisions, our published stochastic and cell-autonomous multicellular model in a growing tissue was used (12). In that model, the transcription factor ATML1 stochastically fluctuates and drives the expression of its target LGO. In the simulated growing tissue, cells divide using a timer with some stochasticity. When the timer of a cell reaches a threshold $\Theta_{C,S}$, cells undergo the S-phase, and therefore cells transition from being diploid (2C) to tetraploid (4C). By default, cells that reach a second and higher timer threshold $\Theta_{C,D}$ will undergo division. However, those cells that have reached a certain LGO concentration threshold Θ_T after undergoing the S-phase will not divide and are maintained in an endoreduplication cycle, which increases their ploidy.

The dynamics of the ATML1, LGO and Timer concentration in cell i, denoted by $[ATML1]_i$, $[Target]_i$ and $[Timer]_i$ respectively, follow the chemical Langevin equations (65) given by

$$\frac{d[ATML1]_i}{dt} = P_A + \frac{V_A[ATML1]_i^{n_A}}{K_A^{n_A}+[ATML1]_i^{n_A}} - G_A[ATML1]_i + \sqrt{\frac{1}{2\varepsilon_i(t)}\left(P_A + \frac{V_A[ATML1]_i^{n_A}}{K_A^{n_A}+[ATML1]_i^{n_A}} + G_A[ATML1]_i\right)}\,\eta_{ATML1,i} \quad (1)$$

$$\frac{d[Target]_i}{dt} = \frac{V_T[ATML1]_i^{n_T}}{K_T^{n_T}+[ATML1]_i^{n_T}} - G_T[Target]_i + \sqrt{\frac{1}{2\varepsilon_i(t)}\left(\frac{V_T[ATML1]_i^{n_T}}{K_T^{n_T}+[ATML1]_i^{n_T}} + G_T[Target]_i\right)}\,\eta_{Target,i} \quad (2)$$

$$\frac{dTimer_i}{dt} = P_C + \sqrt{\frac{1}{2\varepsilon_i(t)}(P_C)}\eta_{Timer,i} \quad , \qquad (3)$$

where P_X is the basal production rate for the X variable (where X is either A for ATML1, T for Target concentration or C for Timer concentration), V_X is the prefactor of the ATML1-dependent production rate for the X variable, K_X is the ATML1 concentration at which the ATML1-dependent production rate has its half-maximal value, n_X is the Hill coefficient, and G_X is the linear degradation rate for the X variable. $\varepsilon_i(t)$ is a normalized cell area, $\varepsilon_i(t) = E_0E_i(t)$, where E_0 is an effective cell area, and $E_i(t)$ is the area of cell i in arbitrary units. η_{Xi} is a random Gaussian variable with zero mean that fulfills $\langle \eta_{Xi}(t)\eta_{X'j}(t')\rangle = \delta(t-t')\delta_{XX'}\delta_{ij}$, where i and j are cell indices, X and X' the modeled variables, $\delta_{XX'}$ and δ_{ij} are Kronecker deltas and $\delta(t-t')$ is the Dirac delta function.

Upon cell division, the Timer was reset. To implement the resetting, the following rule was applied at each time step:

$$Timer_i(t) \rightarrow \{U_i \; if \;\; Timer_i(t) \geq \Theta_{C,D};\; Timer_i(t) \;\; \text{otherwise}\}, \qquad (4)$$

where U_i is a uniform randomly distributed number in the interval [0, 0.5) and $\Theta_{C,D}$ is the cell division threshold for the Timer.

The multicellular template on which the simulations were run and initial conditions were the same as in (12). Initial conditions for ATML1 and Target were randomly uniformly distributed in the interval of [0,1) and [0,0.1), respectively. The Timer initial conditions were set in correlation with the cell areas in the initial template with some stochasticity, as performed in (12). The differences between the used initial conditions were just in the ATML1, Target and Timer initial cellular values, determined by different random numbers.

Tissue growth and division were also implemented as in (12). The multicellular tissue grows anisotropically, to emulate the patterning process in the sepal. After each simulation step, dilution effects due to growth in the modeled variables were taken into account. Cells divided using the shortest path rule together with the constraint of having the division plane through the center of mass of the cell.

Numerical simulations were performed with Tissue software (13, 66), and the integration was performed using an Îto interpretation of the Langevin equations with a Heun algorithm (67). Integration was performed with a time step dt = 0.1, and simulations were stopped at time 135. Parameter values for the simulations are given in Table S2. The outcome of the simulation in Fig. 9B was displayed using Paraview software (https://www.paraview.org/).

We recently proposed a more detailed model of the ATML1 regulatory network to study how giant cell specification and cell fate maintenance depends on VLCFA (13), which is still a stochastic and cell-autonomous model. Here, however, for the sake of simplicity, and the intention of using a minimal, stochastic and cell-autonomous phenomenological model, the former ATML1 model was used (12).

989 Randomizations of the outcomes from the numerical simulations
990
991 To assess the randomness of the giant cell pattern in the numerical simulations (Fig. 9), the same
992 method was employed as that used for the experimental images. Although randomizations of the
993 tissues were performed similarly (see the "Randomization of the experimental images" section
994 above), a Python script was developed to display the output of the simulation as a multi-labeled
995 image, where each cell was colored with a different label. These images could therefore be
996 randomized using the dmSET method (42). To compare the simulated giant cell pattern (Fig. 9B)
997 to the giant cell pattern found in the experimental mature sepals (Fig. 7B), the output image was
998 cropped using the maximal rectangle in the tissue, and giant cells were also defined by a size
999 threshold, ensuring that all cells with a ploidy of 16C or higher were categorized as giant cells
1000 (Fig. 9C). The few 8C cells that exceeded this threshold were also considered as giant cells.
1001
1002 To study the change in the giant cell spatial pattern over time (Fig. 10), the same simulations were
1003 used, but only the first-arising giant cells (cells that stopped dividing after time t = 55 of the
1004 simulations) were studied. The same method was used to assess the randomness of the cellular
1005 patterns on these giant cells both at time t = 55 and time t = 135. Here, instead of cropping the
1006 image, the whole tissue was randomized (using the dmSET method), including the cells at the
1007 edges, such that exactly the same giant cells were considered at both time points. Examples of
1008 randomizations are shown in Fig. S16A–B. The analysis was performed over five simulation
1009 replicates, with different cellular random initial conditions.
1010
1011 Related "Segmentation" and "Randomization" images appearing in figures such as Fig. 9B were
1012 generated with Python.
1013

1014 ## Acknowledgements

1015
1016 We thank Nicholas J. Russell, John Chandler, Josep Mercadal, Elise Laurelle and Philippe Andrey
1017 for critical comments on the manuscript. We thank Stephen Starkman, Emily Phung, and Isabel
1018 Delo for assistance with segmentation, and Violeta Gibelli for assistance with cell classification
1019 corrections. We thank Richard Smith and Soeren Strauss for their help in using MorphoGraphX;
1020 we are grateful that Richard Smith created a process to generate 2D images from meshes. We also
1021 thank Philippe Andrey for his insights on image analysis and statistics and Elise Laruelle for her
1022 help in using the dmSET randomization code, as well as the Cornell Statistics Consulting Unit,
1023 specifically May Boggess, for help with statistics and coding.
1024

1025 ## Funding

1026
1027 This work was funded by NSF IOS-1553030 (AHKR), NSF DBI-232051 (AHKR and PF-J), the
1028 International Max Planck Research School on "Understanding Complex Plant Traits Using
1029 Computational and Evolutionary Approaches" (GW), a core grant from the Max Planck Society
1030 (GW, PF-J), and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)
1031 under Germany's Excellence Strategy – EXC-2048/1 – project ID 390686111 (PF-J). This research
1032 was advanced through time spent collaborating at the KITP program on Dynamics of Self-
1033 Organization in Animal and Plant Development supported in part by the National Science

Foundation under Grant No. NSF PHY-1748958 and Gordon and Betty Moore Foundation Grant No. 2919.02.

Competing interests:

The authors declare no competing interests.

Author Contributions:

Conception and Design of Experiments: FC, GW, PF-J, AHKR
Imaging: FC
Image analysis: FC, GW, XW
Data analysis: FC, GW, CB, XW,
Analysis of pattern randomness: GW
Modeling: GW, PF-J
Interpretation of results: FC, GW, PF-J, AHKR
Writing manuscript: FC, GW, PF-J, AHKR
Editing manuscript: FC, MB, CB, XW, GW, PF-J, AHKR

Data availability

Microscopy data, MorphoGraphX meshes for the images, as well as material used for data analysis and modeling are available at Open Science Framework (osf.io), DOI:10.17605/OSF.IO/RFCWS.

Reference list

1. D. T. Zuch, S. M. Doyle, M. Majda, R. S. Smith, S. Robert, K. U. Torii, Cell biology of the leaf epidermis: Fate specification, morphogenesis, and coordination. *The Plant Cell* **34**, 209-227 (2021).
2. M. Hulskamp, Trichomes. *Curr Biol* **29**, R273-R274 (2019).
3. Y. Katagiri, J. Hasegawa, U. Fujikura, R. Hoshino, S. Matsunaga, H. Tsukaya, The coordination of ploidy and cell size differs between cell layers in leaves. *Development* **143**, 1120-1125 (2016).
4. J. E. Melaragno, B. Mehrotra, A. W. Coleman, Relationship between Endopolyploidy and Cell Size in Epidermal Tissue of Arabidopsis. *Plant Cell* **5**, 1661-1668 (1993).
5. D. C. Bergmann, F. D. Sack, Stomatal development. *Annu Rev Plant Biol* **58**, 163-181 (2007).
6. R. J. Horst, H. Fujita, J. S. Lee, A. L. Rychel, J. M. Garrick, M. Kawaguchi, K. M. Peterson, K. U. Torii, Molecular Framework of a Regulatory Circuit Initiating Two-Dimensional Spatial Patterning of Stomatal Lineage. *PLoS Genet* **11**, e1005374 (2015).
7. K. U. Torii, Stomatal development in the context of epidermal tissues. *Ann Bot* **128**, 137-148 (2021).
8. M. Hulskamp, Plant trichomes: a model for cell differentiation. *Nat Rev Mol Cell Biol* **5**, 471-480 (2004).

1078 9. R. Balkunde, M. Pesch, M. Hulskamp, Trichome patterning in Arabidopsis thaliana from
1079 genetic to molecular models. *Curr Top Dev Biol* **91**, 299-321 (2010).
1080 10. A. H. Roeder, V. Chickarmane, A. Cunha, B. Obara, B. S. Manjunath, E. M. Meyerowitz,
1081 Variability in the control of cell division underlies sepal epidermal patterning in
1082 Arabidopsis thaliana. *PLoS Biol* **8**, e1000367 (2010).
1083 11. A. H. Roeder, A. Cunha, C. K. Ohno, E. M. Meyerowitz, Cell cycle regulates cell type in
1084 the Arabidopsis sepal. *Development* **139**, 4416-4427 (2012).
1085 12. H. M. Meyer, J. Teles, P. Formosa-Jordan, Y. Refahi, R. San-Bento, G. Ingram, H.
1086 Jonsson, J. C. Locke, A. H. Roeder, Fluctuations of the transcription factor ATML1
1087 generate the pattern of giant cells in the Arabidopsis sepal. *Elife* **6**, (2017).
1088 13. B. V. L. Vadde, N. J. Russell, S. R. Bagde, B. Askey, M. Saint-Antoine, B. Brownfield,
1089 S. Mughal, L. E. Apprill, A. Khosla, F. K. Clark, E. M. Schwarz, S. Alseekh, A. R.
1090 Fernie, A. Singh, K. Schrick, J. C. Fromme, A. Skirycz, P. Formosa-Jordan, A. H. K.
1091 Roeder, The transcription factor ATML1 maintains giant cell identity by inducing
1092 synthesis of its own long-chain fatty acid-containing ligands. *bioRXiv*, (2024).
1093 14. M. L. Gifford, S. Dean, G. C. Ingram, The Arabidopsis ACR4 gene plays a role in cell
1094 layer organisation during ovule integument and sepal margin development. *Development*
1095 **130**, 4249-4258 (2003).
1096 15. H. Tanaka, M. Watanabe, D. Watanabe, T. Tanaka, C. Machida, Y. Machida, ACR4, a
1097 putative receptor kinase gene of Arabidopsis thaliana, that is expressed in the outer cell
1098 layers of embryos and plants, is involved in proper embryogenesis. *Plant Cell Physiol* **43**,
1099 419-428 (2002).
1100 16. I. De Smet, V. Vassileva, B. De Rybel, M. P. Levesque, W. Grunewald, D. Van Damme,
1101 G. Van Noorden, M. Naudts, G. Van Isterdael, R. De Clercq, J. Y. Wang, N. Meuli, S.
1102 Vanneste, J. Friml, P. Hilson, G. Jürgens, G. C. Ingram, D. Inzé, P. N. Benfey, T.
1103 Beeckman, Receptor-Like Kinase ACR4 Restricts Formative Cell Divisions in the
1104 Arabidopsis Root. *Science* **322**, 594-597 (2008).
1105 17. M. L. Churchman, M. L. Brown, N. Kato, V. Kirik, M. Hulskamp, D. Inze, L. De
1106 Veylder, J. D. Walker, Z. Zheng, D. G. Oppenheimer, T. Gwin, J. Churchman, J. C.
1107 Larkin, SIAMESE, a plant-specific cell cycle regulator, controls endoreplication onset in
1108 Arabidopsis thaliana. *Plant Cell* **18**, 3145-3157 (2006).
1109 18. N. Kumar, H. Harashima, S. Kalve, J. Bramsiepe, K. Wang, B. L. Sizani, L. L. Bertrand,
1110 M. C. Johnson, C. Faulk, R. Dale, L. A. Simmons, M. L. Churchman, K. Sugimoto, N.
1111 Kato, M. Dasanayake, G. Beemster, A. Schnittger, J. C. Larkin, Functional Conservation
1112 in the SIAMESE-RELATED Family of Cyclin-Dependent Kinase Inhibitors in Land
1113 Plants. *Plant Cell* **27**, 3065-3080 (2015).
1114 19. M. Dubois, I. Achon, R. A. Brench, S. Polyn, R. Tenorio Berrío, I. Vercauteren, J. E.
1115 Gray, D. Inzé, L. De Veylder, SIAMESE-RELATED1 imposes differentiation of
1116 stomatal lineage ground cells into pavement cells. *Nature Plants* **9**, 1143-1153 (2023).
1117 20. A. H. R. E. M. Schwarz, Transcriptomic Effects of the Cell Cycle Regulator LGO in
1118 Arabidopsis Sepals. *Frontiers in Plant Science* **7**, (2016).
1119 21. A. Wilkens, P. Czerniawski, P. Bednarek, M. Libik-Konieczny, K. Yamada, ATML1
1120 Regulates the Differentiation of ER Body–Containing Large Pavement Cells in Rosette
1121 Leaves of Brassicaceae Plants. *Plant and Cell Physiology*, (2024).

1122 22. R. Galletti, K. L. Johnson, S. Scofield, R. San-Bento, A. M. Watt, J. A. Murray, G. C.
1123 Ingram, DEFECTIVE KERNEL 1 promotes and maintains plant epidermal
1124 differentiation. *Development* **142**, 1978-1983 (2015).
1125 23. J. Dong, D. C. Bergmann, Stomatal patterning and development. *Curr Top Dev Biol* **91**,
1126 267-297 (2010).
1127 24. B. Greese, M. Hulskamp, C. Fleck, Quantification of variability in trichome patterns.
1128 *Front Plant Sci* **5**, 596 (2014).
1129 25. C. Liu, Y. Li, L. Xu, M. Li, J. Wang, P. Yan, N. He, Stomatal Arrangement Pattern: A
1130 New Direction to Explore Plant Adaptation and Evolution. *Front Plant Sci* **12**, 655255
1131 (2021).
1132 26. S. M. Zeng, E. K. W. Lo, B. J. Hazelton, M. F. Morales, K. U. Torii, Effective range of
1133 non-cell autonomous activator and inhibitor peptides specifying plant stomatal patterning.
1134 *Development* **147**, (2020).
1135 27. D. C. Bergmann, Integrating signals in stomatal development. *Curr Opin Plant Biol* **7**,
1136 26-32 (2004).
1137 28. M. R. Lundgren, A. Mathers, A. L. Baillie, J. Dunn, M. J. Wilson, L. Hunt, R. Pajor, M.
1138 Fradera-Soler, S. Rolfe, C. P. Osborne, C. J. Sturrock, J. E. Gray, S. J. Mooney, A. J.
1139 Fleming, Mesophyll porosity is modulated by the presence of functional stomata. *Nature*
1140 *Communications* **10**, 2825 (2019).
1141 29. G. J. Dow, J. A. Berry, D. C. Bergmann, Disruption of stomatal lineage signaling or
1142 transcriptional regulators has differential effects on mesophyll development, but
1143 maintains coordination of gas exchange. *New Phytol* **216**, (2017).
1144 30. K. H. McKown, D. C. Bergmann, Stomatal development in the grasses: lessons from
1145 models and crops (and crop models). *New Phytol* **227**, 1636-1648 (2020).
1146 31. G. L. Stebbins, S. S. Shah, Developmental studies of cell differentiation in the epidermis
1147 of monocotyledons: II. Cytological features of stomatal development in the Gramineae.
1148 *Developmental Biology* **2**, 477-500 (1960).
1149 32. S. S. Sugano, T. Shimada, Y. Imai, K. Okawa, A. Tamai, M. Mori, I. Hara-Nishimura,
1150 Stomagen positively regulates stomatal density in Arabidopsis. *Nature* **463**, 241-244
1151 (2010).
1152 33. K. Kawade, H. Tsukaya, Probing the stochastic property of endoreduplication in cell size
1153 determination of Arabidopsis thaliana leaf epidermal tissue. *PLoS One* **12**, e0185050
1154 (2017).
1155 34. P. Barbier de Reuille, A.-L. Routier-Kierzkowska, D. Kierzkowski, G. W. Bassel, T.
1156 Schüpbach, G. Tauriello, N. Bajpai, S. Strauss, A. Weber, A. Kiss, A. Burian, H.
1157 Hofhuis, A. Sapala, M. Lipowczan, M. B. Heimlicher, S. Robinson, E. M. Bayer, K.
1158 Basler, P. Koumoutsakos, A. H. K. Roeder, T. Aegerter-Wilmsen, N. Nakayama, M.
1159 Tsiantis, A. Hay, D. Kwiatkowska, I. Xenarios, C. Kuhlemeier, R. S. Smith,
1160 MorphoGraphX: A platform for quantifying morphogenesis in 4D. *eLife* **4**, e05864
1161 (2015).
1162 35. S. Strauss, A. Runions, B. Lane, D. Eschweiler, N. Bajpai, N. Trozzi, A. L. Routier-
1163 Kierzkowska, S. Yoshida, S. Rodrigues da Silveira, A. Vijayan, R. Tofanelli, M. Majda,

E. Echevin, C. Le Gloanec, H. Bertrand-Rakusova, M. Adibi, K. Schneitz, G. W. Bassel, D. Kierzkowski, J. Stegmaier, M. Tsiantis, R. S. Smith, Using positional information to provide context for biological image analysis with MorphoGraphX 2.0. *Elife* **11**, (2022).

36. P. M. Donnelly, D. Bonetta, H. Tsukaya, R. E. Dengler, N. G. Dengler, Cell cycling and cell enlargement in developing leaves of Arabidopsis. *Dev Biol* **215**, 407-419 (1999).

37. A. H. Haber, Nonessentiality of concurrent cell divisions for degree of polarization of leaf growth I. studies with radiation-induced mitotic inhibition. *American Journal of Botany* **49**, 583–589 (1962).

38. B. V. L. Vadde, A. H. K. Roeder, Can the French flag and reaction-diffusion models explain flower patterning? Celebrating the 50th anniversary of the French flag model. *J Exp Bot* **71**, 2886-2897 (2020).

39. R. Balkunde, A. Deneer, H. Bechtel, B. Zhang, S. Herberth, M. Pesch, B. Jaegle, C. Fleck, M. Hulskamp, Identification of the Trichome Patterning Core Network Using Data from Weak ttg1 Alleles to Constrain the Model Space. *Cell Rep* **33**, 108497 (2020).

40. J. Mercadal, A diffusion-driven switch specifies rhizoid precursor cells in Marchantia polymorpha. *bioRXiv*, (2023).

41. E. Laruelle, N. Spassky, A. Genovesio, Unraveling spatial cellular pattern by computational tissue shuffling. *Commun Biol* **3**, 605 (2020).

42. E. Laruelle, A. Genovesio, paper presented at the 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), Kolkata, India, 2022 2022.

43. N. Hervieux, M. Dumond, A. Sapala, A. L. Routier-Kierzkowska, D. Kierzkowski, A. H. Roeder, R. S. Smith, A. Boudaoud, O. Hamant, A Mechanical Feedback Restricts Sepal Growth and Shape in Arabidopsis. *Curr Biol* **26**, (2016).

44. M. Grebe, The patterning of epidermal hairs in Arabidopsis--updated - PubMed. *Current opinion in plant biology* **15**, (2012).

45. Spatial statistics is a comprehensive tool for quantifying cell neighbor relationships and biological processes via tissue image analysis. *Cell Reports Methods* **2**, (2022).

46. G. Kumar, R. K. Pandurengan, E. R. Parra, K. Kannan, C. Haymaker, Frontiers | Spatial modelling of the tumor microenvironment from multiplex immunofluorescence images: methods and applications. *Frontiers in Immunology* **14**, (2023).

47. O. Sandler, S. P. Mizrahi, N. Weiss, O. Agam, I. Simon, N. Q. Balaban, O. Sandler, S. P. Mizrahi, N. Weiss, O. Agam, I. Simon, N. Q. Balaban, Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature* **519**, (2015).

48. H. Kobayashi, Variations of endoreduplication and its potential contribution to endosperm development in rice. *Plant Production Science* **22**, 227-241 (2019).

49. D. O. Robinson, J. E. Coate, A. Singh, L. Hong, M. Bush, J. J. Doyle, A. H. K. Roeder, Ploidy and Size at Multiple Scales in the Arabidopsis Sepal. *Plant Cell* **30**, 2308-2329 (2018).

50. G. Horiguchi, H. Tsukaya, Organ size regulation in plants: insights from compensation. *Front Plant Sci* **2**, 24 (2011).

51. H. B. Tang, J. Wang, L. Wang, G. D. Shang, Z. G. Xu, Y. X. Mai, Y. T. Liu, T. Q. Zhang, J. W. Wang, Anisotropic cell growth at the leaf base promotes age-related changes in leaf shape in Arabidopsis thaliana. *Plant Cell* **35**, 1386-1407 (2023).

52. R. G. Franks, Z. Liu, R. L. Fischer, R. G. Franks, Z. Liu, R. L. Fischer, SEUSS and LEUNIG regulate cell proliferation, vascular development and organ polarity in Arabidopsis petals. *Planta* **224**, (2006).
53. V. Monniaux, How to Evolve a Perianth: A Review of Cadastral Mechanisms for Perianth Identity. *Frontiers in Plant Science* **9**, (2018).
54. C. Le Gloanec, A. Gómez-Felipe, V. Alimchandani, E. Branchini, A. L. Routier-Kierzkowska, D. Kierzkowski, Modulation of cell differentiation and growth dynamics underlie the shift from bud protection to light capture in cauline leaves. *bioRxiv*, (2023).
55. K. L. Johnson, K. A. Degnan, J. Ross Walker, G. C. Ingram, AtDEK1 is essential for specification of embryonic epidermal cell fate. *Plant J* **44**, 114-127 (2005).
56. M. Abe, H. Katsumata, Y. Komeda, T. Takahashi, Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. *Development* **130**, (2003).
57. E. Ogawa, Y. Yamada, N. Sezaki, S. Kosaka, H. Kondo, N. Kamata, M. Abe, Y. Komeda, T. Takahashi, ATML1 and PDF2 Play a Redundant and Essential Role in Arabidopsis Embryo Development. *Plant and Cell Physiology* **56**, (2015).
58. P. Li, M. B. Elowitz, A. Klein, B. Treutlein, Communication codes in developmental signaling pathways. *Development* **146**, (2019).
59. S. J. Clough, A. F. Bent, Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* **16**, 735-743 (1998).
60. A. Sessions, E. Burke, G. Presting, G. Aux, J. McElver, D. Patton, B. Dietrich, P. Ho, J. Bacwaden, C. Ko, J. D. Clarke, D. Cotton, D. Bullis, J. Snell, T. Miguel, D. Hutchison, B. Kimmerly, T. Mitzel, F. Katagiri, J. Glazebrook, M. Law, S. A. Goff, A high-throughput Arabidopsis reverse genetics system. *Plant Cell* **14**, 2985-2994 (2002).
61. R. San-Bento, E. Farcot, R. Galletti, A. Creff, G. Ingram, Epidermal identity is maintained by cell-cell communication via a universally active feedback loop in Arabidopsis thaliana. *Plant J* **77**, 46-58 (2014).
62. D. R. Smyth, J. L. Bowman, E. M. Meyerowitz, Early flower development in Arabidopsis. *Plant Cell* **2**, 755-767 (1990).
63. N. Trozzi, B. Lane, A. Perruchoud, Y. Wang, L. Hörmayer, M. Ansel, C. Mollier, A. Malivert, F. Clark, T. Reichgelt, A. H. K. Roeder, O. Hamant, A. Boudaoud, D. Kwiatkowska, A. Runions, R. S. Smith, M. Majda, Puzzle cell shape emerges from the interaction of growth with mechanical constraints. *bioRxiv*, (2023).
64. R. v. Spoordonk, R. Schneider, A. Sampathkumar, Mechano-chemical regulation of complex cell shape formation: Epidermal pavement cells—A case study. *Quantitative Plant Biology* **4**, (2023).
65. D. T. Gillespie, The chemical Langevin equation. *Journal of Chemical Physics* **113**, 297-306 (2000).
66. O. Hamant, H. Jönsson, P. Krupinski, M. Uyttewaal, P. Bokov, F. Corson, P. Sahlin, A. Boudaoud, E. M. Meyerowitz, Y. Couder and J. Traas, Developmental Patterning by Mechanical Signals in Arabidopsis. Science 322, 1650 - 1655 (2008).

67. O. Carrillo, M. Ibanes, J. Garcia-Ojalvo, J. Casademunt, J. M. Sancho, Intrinsic noise-induced phase transitions: beyond the noise interpretation. *Phys Rev E Stat Nonlin Soft Matter Phys* **67**, 046110 (2003).

Figure 1

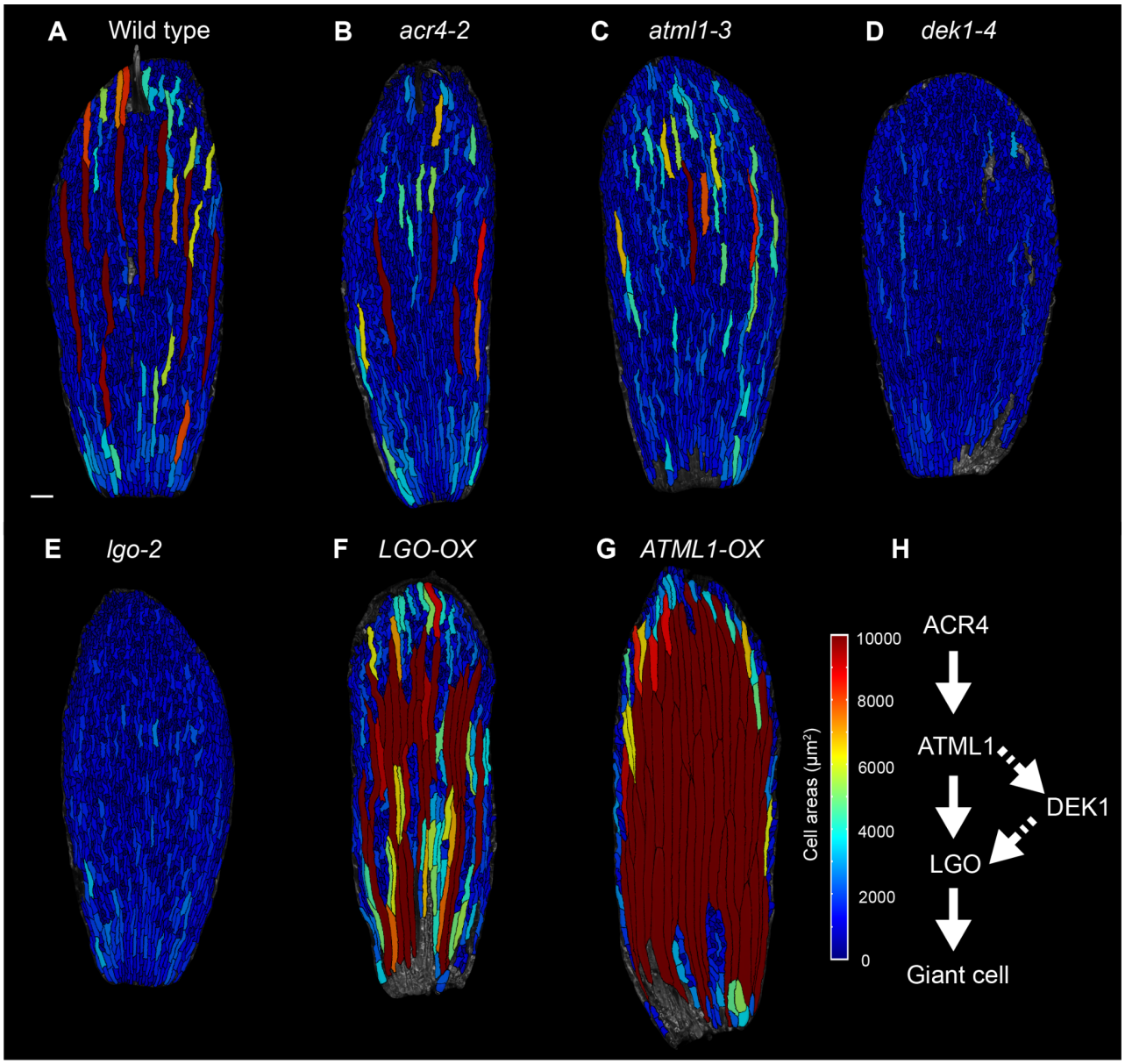

Fig. 1. The genetic pathway that regulates giant cell development in sepals. (**A**–**G**) Cell area heat maps in μm^2 of the abaxial (outer) surface of a stage 14 adult sepal of (**A**) wild type, (**B**) *acr4-2*, (**C**) *atml1-3*, (**D**) *dek1-4*, (**E**) *lgo-2*, (**F**) *LGO-OX* (*pATML1::LGO*) and (**G**) *ATML1-OX* (*pPDF1::FLAG-ATML1*). Scale bar represents 100 µm. (**H**) The ordering of genes into a genetic pathway according to double-mutant phenotypic analysis.

Figure 2

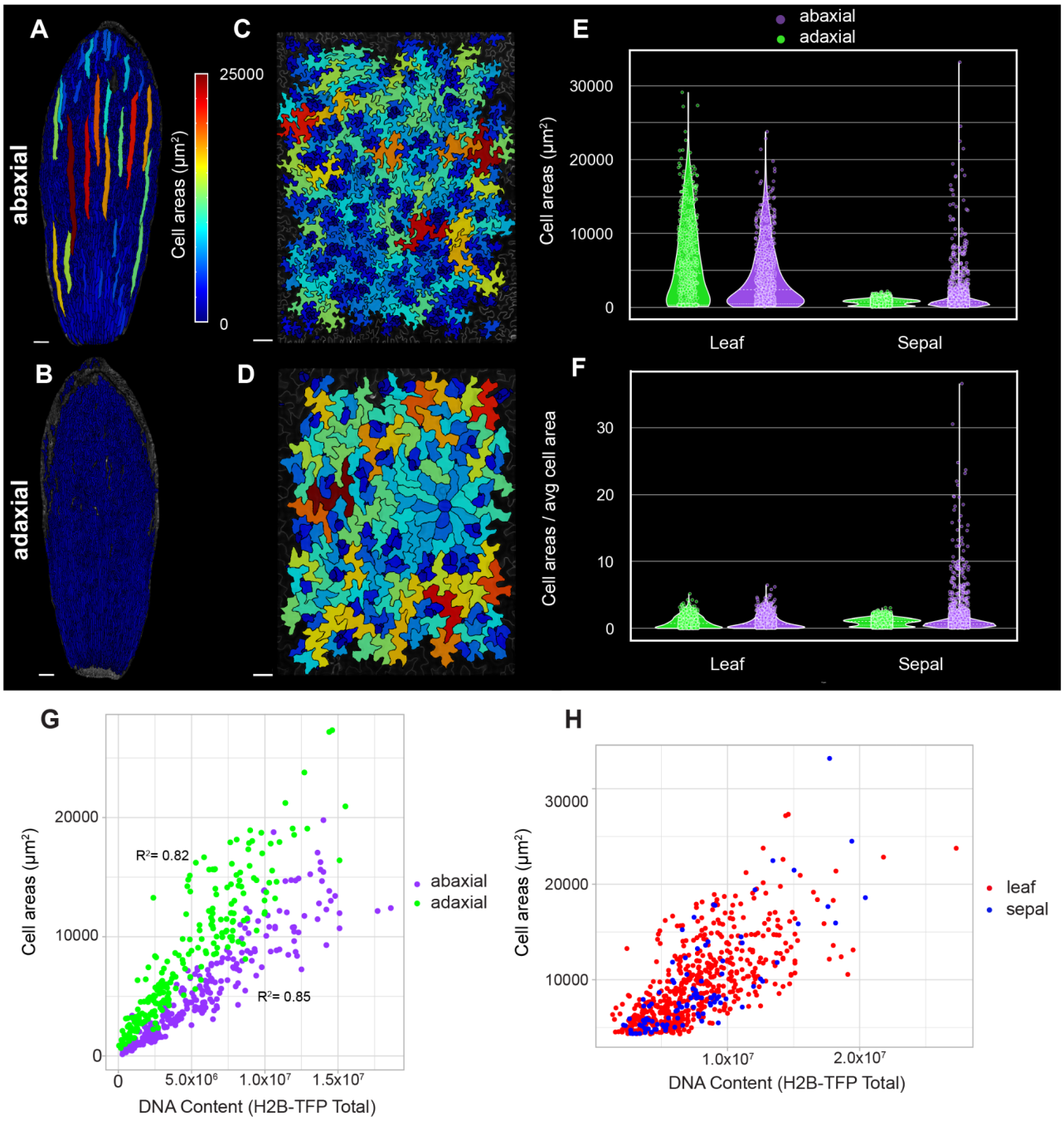

Fig. 2. Abaxial and adaxial cell size distribution in the wild-type leaf and sepal epidermis; size correlates with DNA content. (**A–D**) Cell area heat maps in µm^2 of (**A**) abaxial surface of wild-type sepal, (**B**) adaxial surface of wild-type sepal, (**C**) abaxial surface of 25-dpg wild-type leaf 1 or 2 (**D**) adaxial surface of 25-dpg wild-type leaf 1 or 2. Scale bars represent 100 µm. (**E–F**) Violin and strip plots of (**E**) cell areas and (**F**) cell areas normalized by the average cell area of abaxial and adaxial sides of 25-dpg wild-type leaves (two pooled replicates) and adult wild-type sepals (three pooled replicates). (**G**) Adaxial side (green) and abaxial side (purple) of 25 dpg-leaf cell area versus DNA content as measured by H2B-TFP total nuclear fluorescence, with $R^2 = 0.85$ for the abaxial side and $R^2 = 0.82$ for the adaxial side (one of two replicates). (**H**) Cell area of the largest cells (area > 4,308 µm^2) versus DNA content as

measured by H2B-TFP total nuclear fluorescence in both the abaxial and adaxial side of the 25-dpg leaf (red) and in the abaxial side of the adult sepal (blue). See Fig. S2 for replicates.

Figure 3

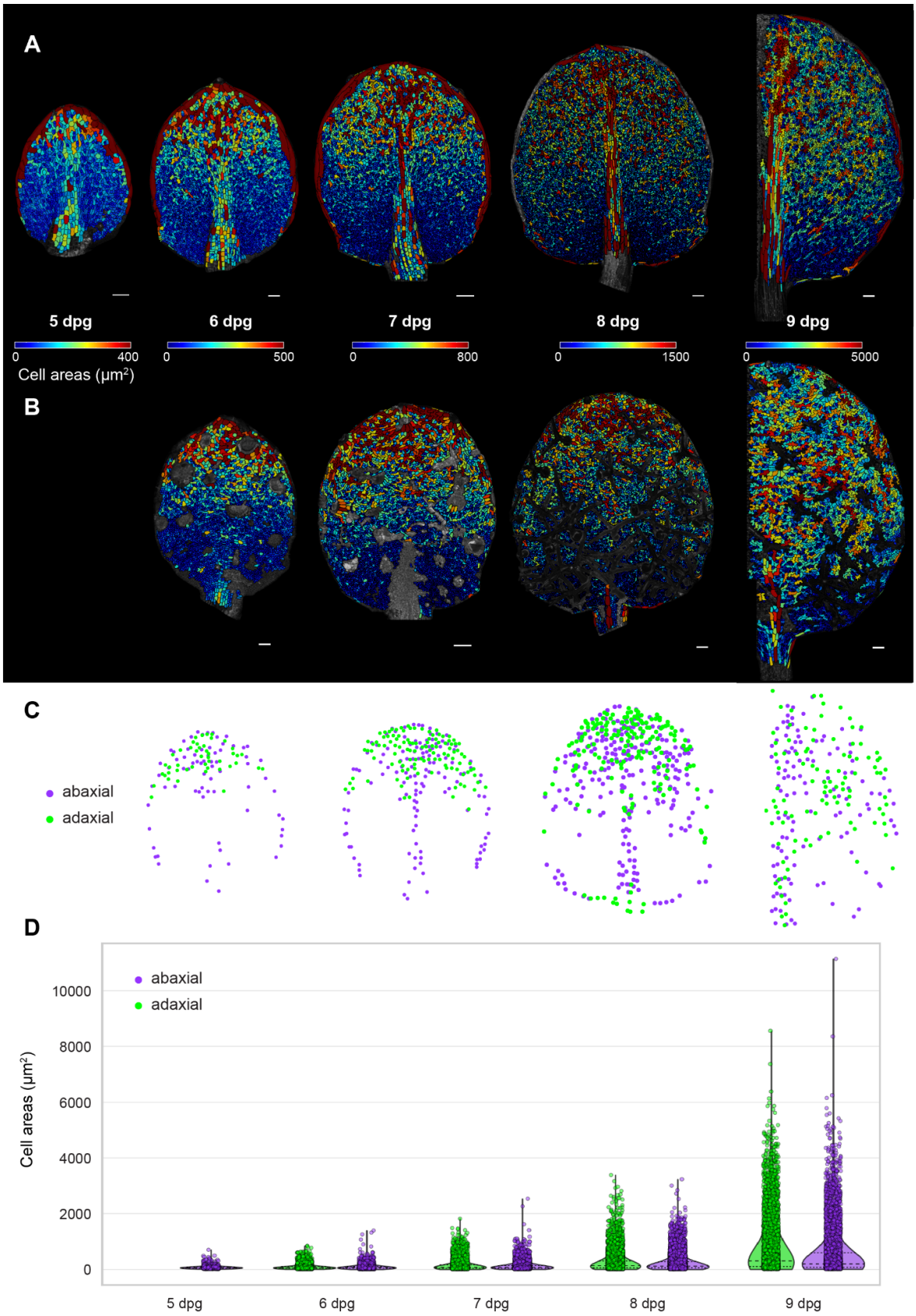
A
5 dpg
6 dpg
7 dpg
8 dpg
9 dpg
0
400
0
500
0
800
0
1500
0
5000
Cell areas (μm²)
B
C
abaxial
adaxial
D
abaxial
adaxial
Cell areas (μm²)
10000
8000
6000
4000
2000
0
5 dpg
6 dpg
7 dpg
8 dpg
9 dpg

Fig. 3. Cell size patterning occurs as a basipetal wave simultaneously on the adaxial and abaxial sides of the leaf. (**A–B**) Cell area heat maps in μm^2 of leaf 1 or 2 at different stages of development on (**A**) the abaxial side and on (**B**) the adaxial side at 5 dpg, 6 dpg, 7 dpg, 8 dpg and 9 dpg (half leaf). Unsegmented regions on adaxial leaves correspond to trichomes, which were not considered in this analysis. Each stage is associated with a distinct heat map color range. Scale bars represent 50 µm at 5 dpg and 6 dpg, and 100 µm at 7 dpg, 8 dpg and 9 dpg. (**C**) Spatial positions of large cells (those above an area threshold, see below) on the abaxial (purple points) and adaxial (green points) sides of the same leaf at 6 dpg, 7 dpg, 8 dpg and 9 dpg. Area thresholds for each leaf were determined from the 98th percentile cell area of the abaxial side. (**D**) Violin and strip plots of cell areas in μm^2 on abaxial and adaxial sides of leaves at different developmental stages, excluding margin and midrib cells. Abaxial and adaxial sides are from the same leaf. See also Fig. S3 for the leaves shown to scale.

Figure 4

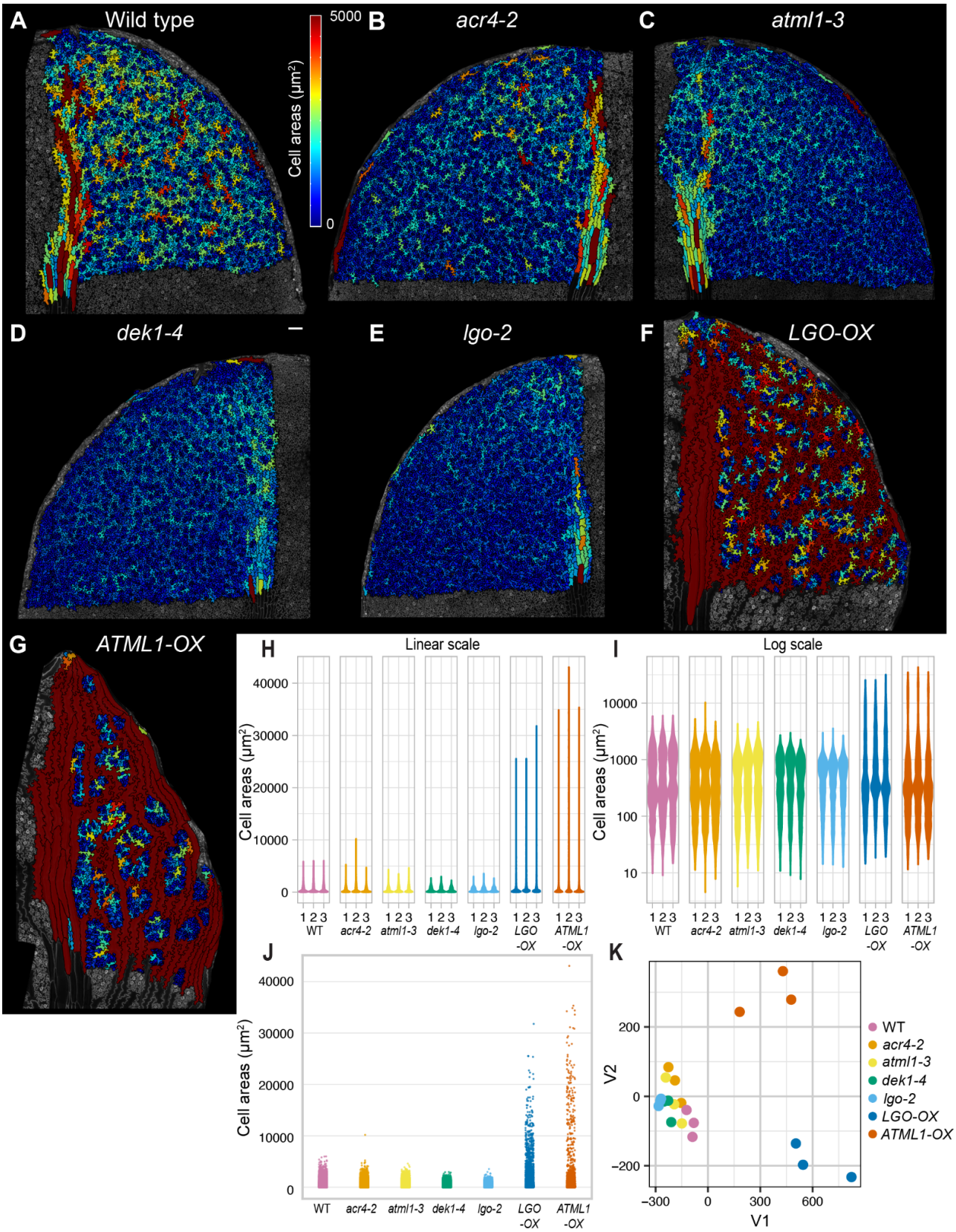
A
Wild type
B
acr4-2
C
atml1-3
D
dek1-4
E
lgo-2
F
LGO-OX
G
ATML1-OX
Cell areas (µm²)
5000
0
H
Linear scale
I
Log scale
Cell areas (µm²)
WT
acr4-2
atml1-3
dek1-4
lgo-2
LGO -OX
ATML1 -OX
J
K
V1
V2
WT
acr4-2
atml1-3
dek1-4
lgo-2
LGO-OX
ATML1-OX

Fig. 4. The sepal giant cell specification pathway also patterns cell size in 9-dpg leaves. (**A–G**) Cell area heat maps in μm^2 of the upper abaxial quadrant of leaf 1 or 2 at 9 dpg for the genotypes: (**A**) wild type, (**B**) *acr4-2*, (**C**) *atml1-3*, (**D**) *dek1-4*, (**E**) *lgo-2*, (**F**) *LGO-OX* (*pATML1::LGO*) and (**G**) *ATML1-OX* (*pPDF1::FLAG-ATML1*). Scale bar represents 100 µm. (**H, I**) Violin plots of cell area densities on a linear scale (**H**) and on a $\log_{10}$ scale (**I**). Cells were in the upper quadrants of leaf 1 or 2 for three replicates of each genotype. (**J**) The same data as in (H) and (I) but displayed as a dot plot, with all replicates per genotype pooled, revealing giant cells as the large cell outliers. (**K**) 2D Wasserstein distance plot for 9-dpg replicates. Cell area heat maps of other replicates are shown in Figs. S4 and S5. The Wasserstein statistical tests among replicates are shown in Fig. S10. Areas of midrib cells (large cells overlying the midrib to the tip of the leaf) and margin cells were removed in (H–K) so that only leaf blade cells were compared.

Figure 5

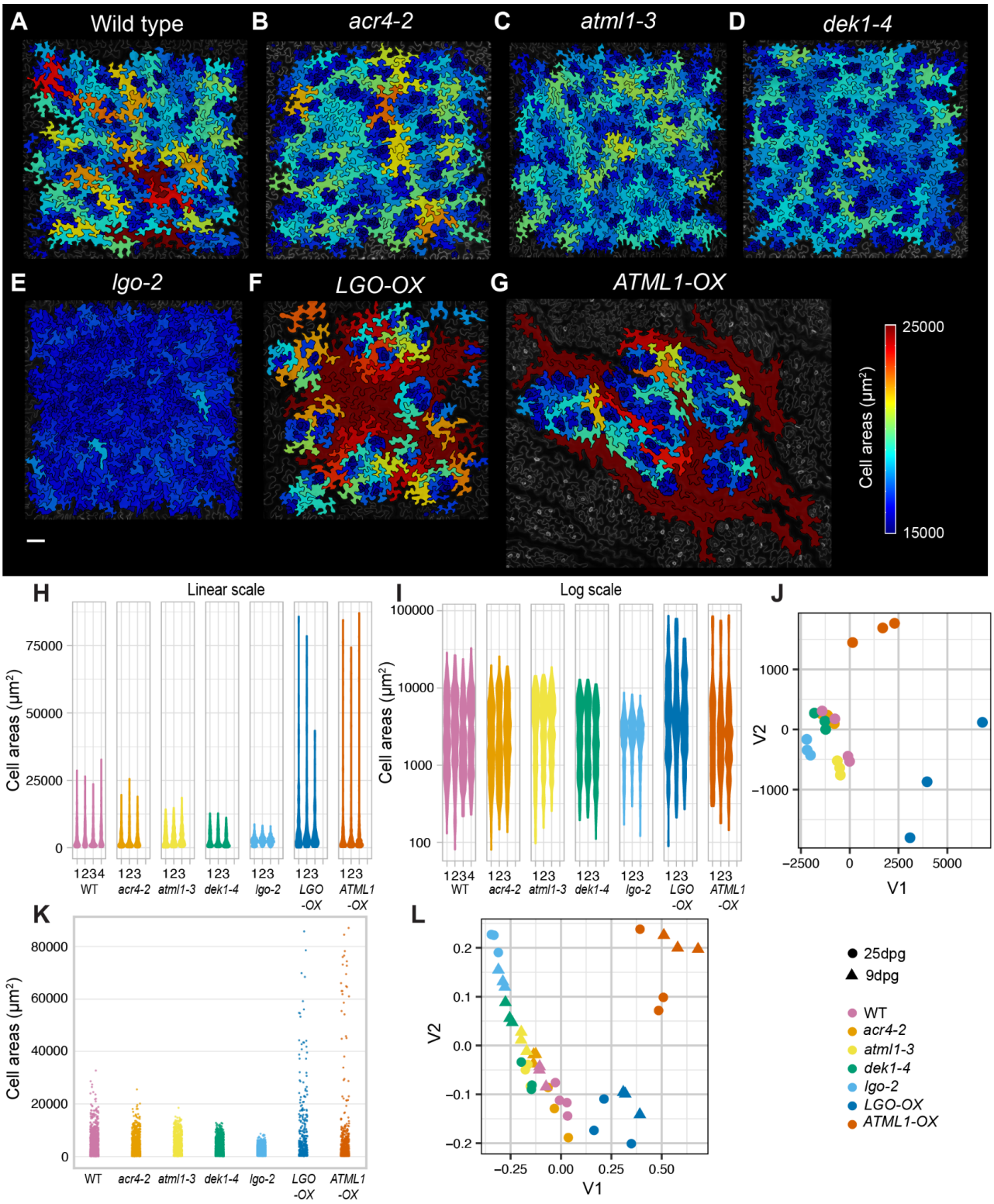

Fig. 5. The sepal giant cell specification pathway also patterns cell size in 25-dpg mature leaves. (**A–G**) Cell area heat maps in µm^2 of a 25-dpg leaf 1 or 2 area approximately midway between the midrib and margin and between tip and base on the abaxial side for the genotypes: (**A**) wild type (**B**) *acr4-2*, (**C**) *atml1-3*, (**D**) *dek1-4*, (**E**) *lgo-2*, (**F**) *LGO-OX* (*pATML1::LGO*), (**G**) *ATML1-OX* (*pPDF1::FLAG-*

ATML1). Scale bar represents 100 µm. (**H–I**) Violin plots of cell area densities on a linear scale (**H**) and on a $\log_{10}$ scale (**I**) for abaxial sections of leaf 1 or 2 for four replicates of each wild type and three replicates of other genotypes. (**J**) Wasserstein distances for 25-dpg replicates displayed as Euclidean distances embedded in 2D. (**K**) The same data as in (H) and (I) but displayed as a dot plot, with all replicates pooled per genotype, revealing giant cells as the large cell outliers. Stomata were removed in (H–K) so that only leaf blade pavement cells were compared. (**L**) Wasserstein tests were performed for all replicates of both 9 dpg (with stomata) and 25 dpg (without stomata) and the Wasserstein distances plot is displayed. For comparison, cell area distributions of 9-dpg leaves and 25-dpg leaves were normalized to have averages of 1. The 25-dpg replicates are indicated by circular dots and 9-dpg replicates by triangular dots. Cell area heat maps of other replicates are shown in Figs. S4–S7. Wasserstein statistical tests among replicates are shown in Fig. S10. Dataset F was also used for an independent analysis in Trozzi et al., 2023.

Figure 6

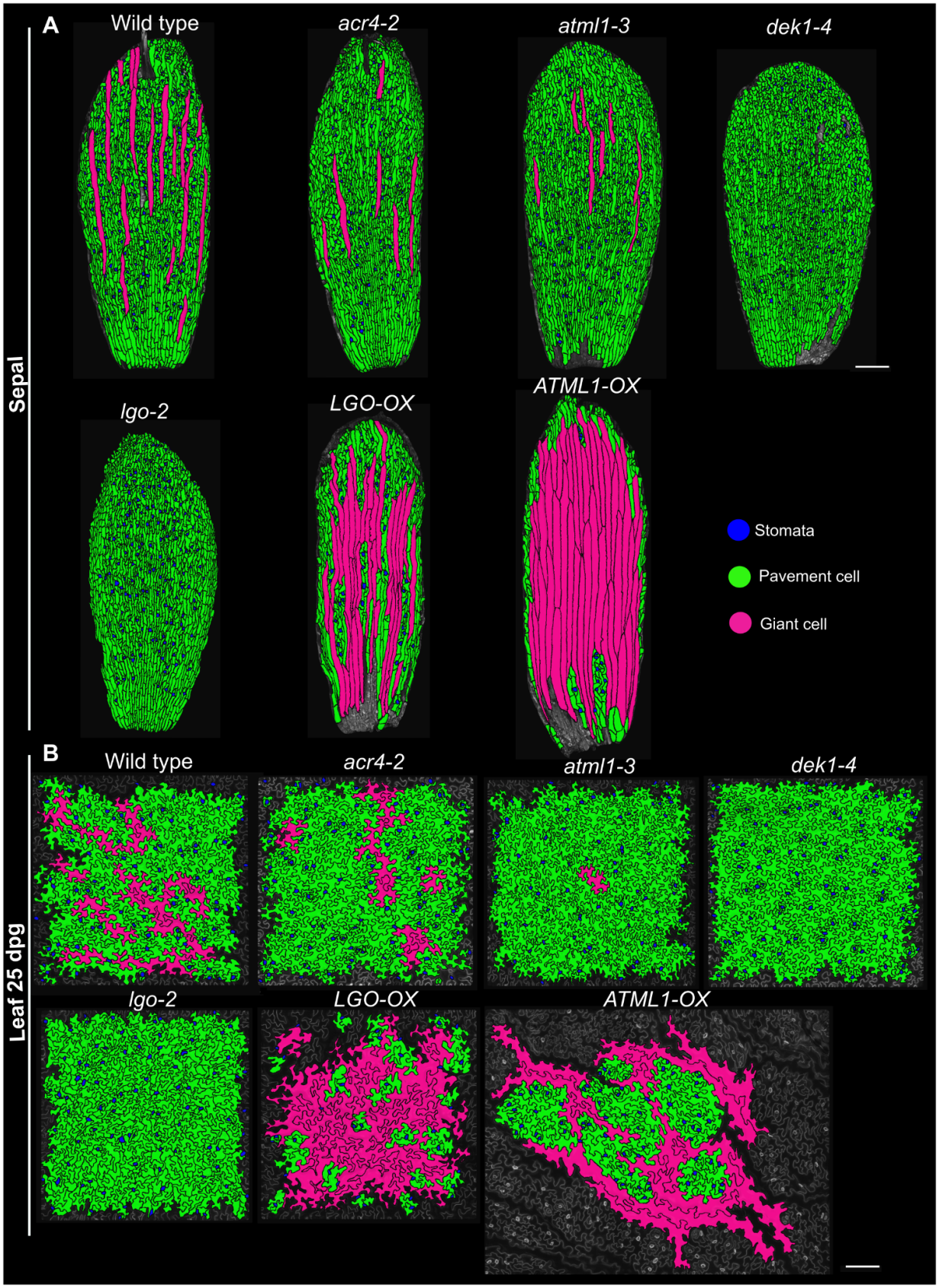
A
Wild type
acr4-2
atml1-3
dek1-4
Sepal
lgo-2
LGO-OX
ATML1-OX
Stomata
Pavement cell
Giant cell
B
Wild type
acr4-2
atml1-3
dek1-4
Leaf 25 dpg
lgo-2
LGO-OX
ATML1-OX

Fig. 6. Cell type classification in the leaf and sepal. (**A–B**) Segmented meshes of one replicate for each genotype after cell type classification in the mature sepal (**A**) and the mature leaf (**B**). Cells are colored with their corresponding cell type: pavement cells (in green), stomata (in blue) and giant cells (in magenta). Stomata and pavement cells were first classified using a trained classification algorithm based on cell shape features. Giant cells were defined as the largest cells, using a size threshold based on *atml1-3* mutants (see Materials and Methods). Scale bars represent 200 µm. See also Figs. S8 and S9.

Figure 7

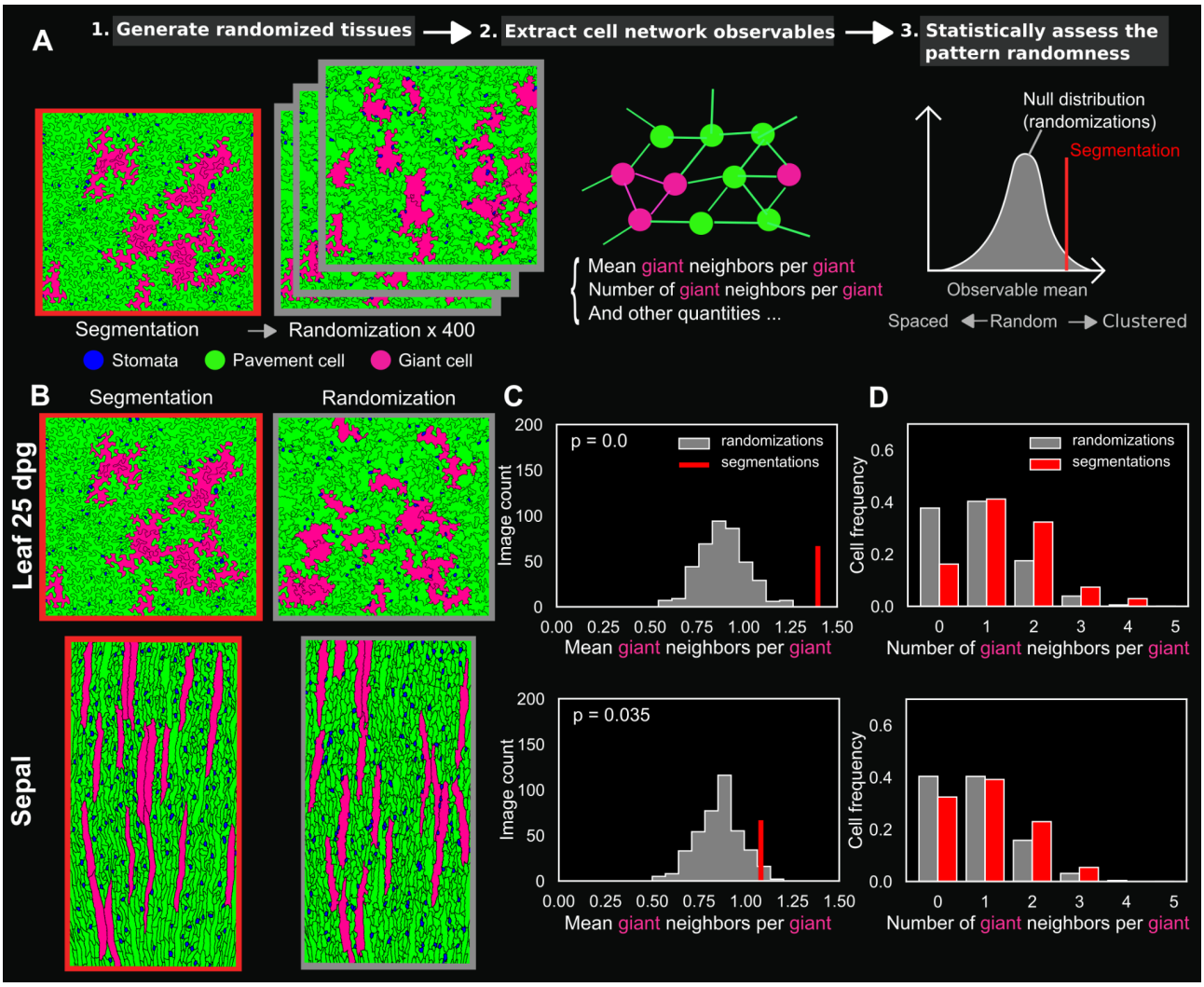

Fig. 7. Giant cells are more clustered than expected by chance both in the wild-type leaf and sepal. (**A**) Scheme summarizing the method used to assess the randomness of the cellular patterns. Each segmentation is computationally randomized using the dmSET method into 400 randomized tissues where cell positions (and orientation in the case of the leaf) have been randomly shuffled (see Materials and Methods). To statistically assess the extent to which the segmented image shows a random giant cell pattern, a quantitative observable (middle) extracted from the segmentation is compared with the same observable computed in all randomized tissues, forming the estimated 'null distribution' (right). (**B**) Example of a representative segmentation of a wild-type leaf 25 dpg (top left) and a wild-type sepal (bottom left) and one of their randomized tissue (randomization) images (right). (**C**) Mean number of giant cell neighbors per giant cell in leaves (top) and sepals (bottom). The value extracted from the segmentations (in red) was statistically tested against all the values extracted from the 400 pooled randomizations (in gray). The mean number of giant cell neighbors per giant cell is higher than expected by chance by the null distribution, and the null hypothesis can be rejected (p-value < 0.05), indicating that giant cells are clustered. (**D**) Distributions of the number of giant cell neighbors for all giant cells found in all replicates of segmentations (in red) and randomizations (in gray) in leaves (top) and sepals (bottom). In the segmentations, fewer isolated giant cells are present than expected by chance, and more giant cells are in contact with more than one other giant cell, highlighting the tendency of giant cells to

form clusters. Total number of giant cells counted (excluding giant cells at the image border) in the analysis: n = 68 (leaf, segmentations), n = 68 × 400 (leaf, randomizations), n = 74 (sepal, segmentations), n = 74 × 400 (sepal, randomizations). See also Figs. S13, S14, S15, S17 and S18.

Figure 8

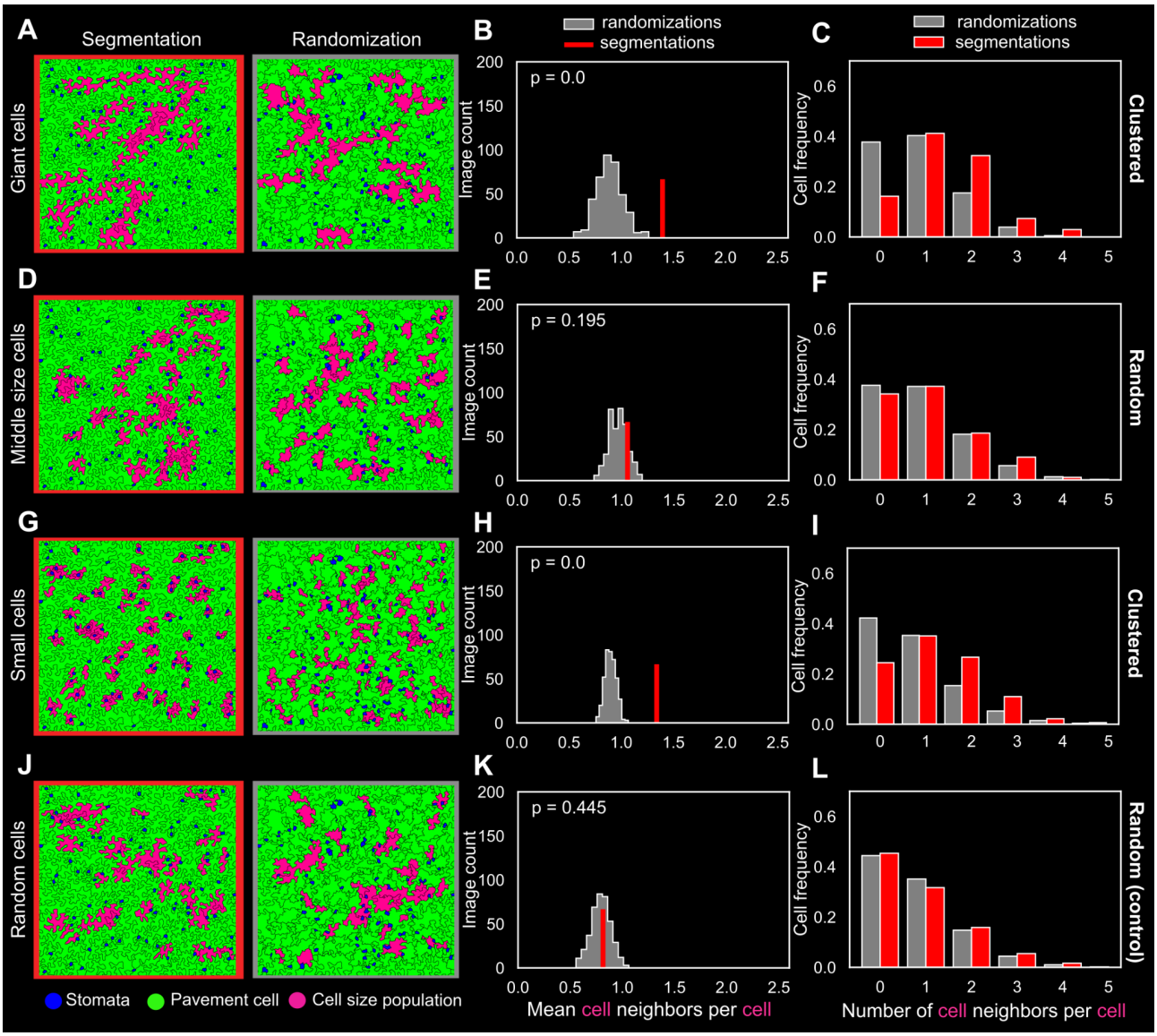

Fig. 8. Different cell sizes display different spatial patterns in the wild-type leaf. The method used to assess the randomness of the giant cell patterns (Fig. 7) was applied here on different pavement cell size populations within the mature 25-dpg leaf: (**A–C**) giant, (**D–F**) mid-size (around 5,000 μm^2), (**G–I**) small (smallest pavement cells), and (**J–L**) random (randomly selected pavement cells). (**A**, **D**, **G**, **J**) Example of representative segmentation of a 25-dpg wild-type leaf (left) and one of its corresponding randomized tissue randomization (right), where cell locations have been computationally shuffled. Cells colored in magenta represent the cells within the studied pavement cell size population. (**B**, **E**, **H**, **K**) Mean number of cell neighbors per cell within the same size population. The value extracted in segmentations (in red) was statistically compared with the 400 values extracted from the randomizations (in gray). (**B**) The mean number of giant cell neighbors per giant cell is higher than expected by chance ($p < 0.05$), indicating that giant cells are clustered. Same data as in Fig. 7C, top. (**E**) Middle-size cells are less clustered than giant cells and more randomly organized (the null hypothesis cannot be rejected, $p = 0.195$). (**H**) The mean number of small cell neighbors per small cell is significantly higher than in the randomized tissues ($p < 0.05$), highlighting that small cells form clusters. (**K**) As expected, the randomly selected pavement cells (with area > 2,000 μm^2) show a value that falls right in the center of

the null distribution ($p = 0.445$). (**C**, **F**, **I**, **L**) Distributions of the number of cell neighbors belonging to the studied cell population per cell of that population in the segmentations (in red) and the randomizations (in gray). All six replicates were pooled to increase the statistical power of the tests. Total number of cells in cell populations counted in the analysis: $n = 68$ (giant cells, segmentations), $n = 68 \times 400$ (giant cells, randomizations), $n = 199$ (middle-size cells, segmentations), $n = 199 \times 400$ (middle-size cells, randomizations), $n = 639$ (small cells, segmentations), $n = 639 \times 400$ (small cells, randomizations), $n = 162$ (random cells, segmentations), $n = 162 \times 400$ (random cells, randomizations).

Figure 9

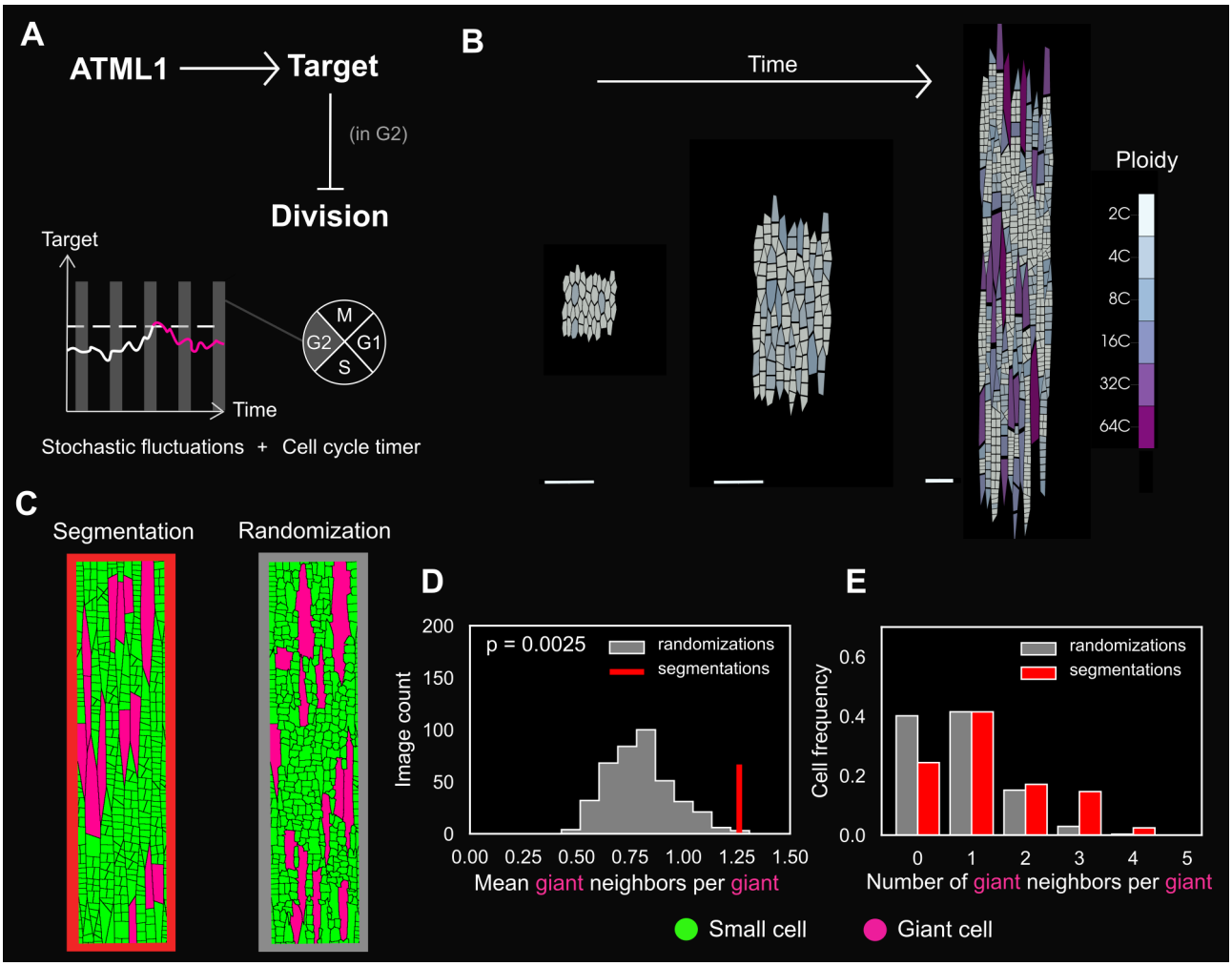

Fig. 9. A cell-autonomous stochastic model can recapitulate giant cells clustering. (**A**) Cartoon of the computational model for giant cell patterning. ATML1 activates a target (LGO), which, if above a certain threshold and during the G2 cell-cycle phase, prevents cell division and instead drives the entry into endoreduplication and giant cell formation. (**B**) Simulation snapshots of the simulated growing sepal, at three different time points. Color codes indicate the cell ploidy levels. Scale bars represent the same size in arbitrary units. (**C**) A rectangular section of the simulation output is then used to quantify the giant cell pattern. "Segmentation" refers to one simulation output (left) and "Randomization" to one randomization of the simulated output (right). Giant cells, labeled in magenta, were defined by a size threshold (see Material and Methods). (**D**) Mean number of giant cell neighbors per giant cell in the simulations (called segmentation in red) and in its randomizations (in gray). The mean number of giant cell neighbors per giant cell is higher than expected by chance ($p < 0.05$), indicating a clustered pattern of giant cells, as in the experimental sepal replicates (Fig. 7C). (**E**) Distribution of the number of giant cell neighbors per giant cell, demonstrating that it is more likely to find giant cells in contact with at least two other giant cells in our simulations compared with the randomized simulated outputs, as in the experimental sepal replicates (Fig. 7D). Five simulation outputs with five different initial conditions were performed and combined for the analysis. Total number of giant cells (excluding giant cells at the image border) counted in the analysis: n = 42 (segmentations), n = 42 × 400 (randomizations).

Figure 10

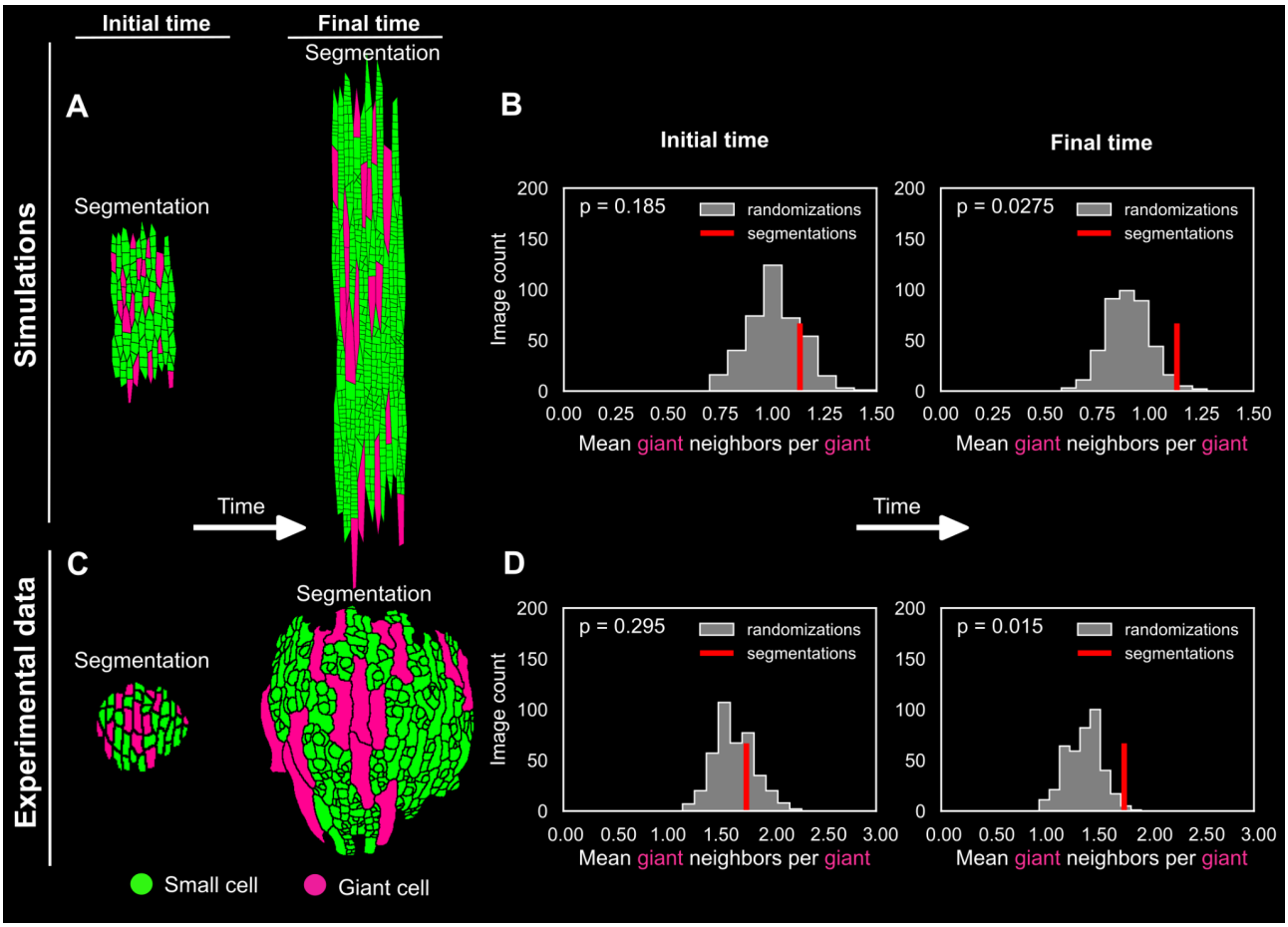

Fig. 10. The giant cell spatial pattern arises randomly and becomes clustered over time. (**A, C**) Segmented images of (**A**) a simulated tissue and (**C**) a real sepal at two different time points, referred to as 'initial time' and 'final time'. Small cells are labeled in green and giant cells in pink. Giant cells are defined as the cells that did not divide from the initial time point. (**A**) Numerical simulation output at t = 55 and t = 135. (**C**) Time-lapse sepal epidermis images at stage 4 and stage 9. (**B**, **D**) Statistical assessment of the randomness of the giant cell pattern (comparing the "segmentations" in red with the randomized tissues in gray) at initial time (left) and final time point (right) in (B) the simulations and (D) the real tissues. The shift in the null distribution over developmental time means that the initially random distribution of giant cells becomes non-random during development. At the initial time point, the null hypothesis could not be rejected (giant cells are randomly distributed, with $p = 0.185$ in simulations and $p = 0.295$ in experimental data). At the final time point, the mean giant neighbors per giant cell became significantly greater than expected by chance ($p < 0.05$). All five replicates (in simulations) and three replicates (in experimental data) were pooled. Total number of giant cells counted in the analysis: n = 83 (simulations, segmentations), n = 83 × 400 (simulations, randomizations), n = 49 (experimental data, segmentations), n = 14 × 400 (experimental data, randomizations). Dataset in C was also used for an independent analysis in Hervieux et al. 2016. See randomization snapshots related to this figure in Fig. S16.