nature genetics

Article

https://doi.org/10.1038/s41588-023-01546-0

# Chromosome-level genomes of three key *Allium* crops and their trait evolution

Received: 13 October 2021

Accepted: 20 September 2023

Published online: 06 November 2023

Check for updates

Fei Hao[1,2,13], Xue Liu[1,13], Botong Zhou[1,13], Zunzhe Tian[1,13], Lina Zhou[1,13], Hang Zong[1,13], Jiyan Qi[1,13], Juan He[1,13], Yongting Zhang[1], Peng Zeng[1], Qiong Li[1], Kai Wang[1], Keke Xia[3], Xing Guo[3,4], Li Li[3], Wenwen Shao[3], Bohan Zhang[5], Shengkang Li[3], Haifeng Yang[6], Linchong Hui[6], Wei Chen[6], Lixin Peng[7], Feipeng Liu[8], Zi-Qiang Rong[8], Yingmei Peng[1], Wenbo Zhu[1], John A. McCallum[9], Zhen Li[10], Xun Xu[3,11,14] ✉, Hui Yang[2,14] ✉, Richard C. Macknight[12,14] ✉, Wen Wang[1,14] ✉ & Jing Cai[1,14] ✉

*Allium* crop breeding remains severely hindered due to the lack of high-quality reference genomes. Here we report high-quality chromosome-level genome assemblies for three key *Allium* crops (Welsh onion, garlic and onion), which are 11.17 Gb, 15.52 Gb and 15.78 Gb in size with the highest recorded contig N50 of 507.27 Mb, 109.82 Mb and 81.66 Mb, respectively. Beyond revealing the genome evolutionary process of *Allium* species, our pathogen infection experiments and comparative metabolomic and genomic analyses showed that genes encoding enzymes involved in the metabolic pathway of *Allium*-specific flavor compounds may have evolved from an ancient uncharacterized plant defense system widely existing in many plant lineages but extensively boosted in alliums. Using in situ hybridization and spatial RNA sequencing, we obtained an overview of cell-type categorization and gene expression changes associated with spongy mesophyll cell expansion during onion bulb formation, thus indicating the functional roles of bulb formation genes.

*Allium* species (family—Amaryllidaceae, order—Asparagales), including onion (*Allium cepa* L.), garlic (*Allium sativum* L.) and Welsh onion (*Allium fistulosum* L.), are a group of plants with tremendous value as crops, spices, medicinal materials and ornamentals[1–3]. African lily (*Agapanthus africanus*), which was used as the outgroup in this study, also belongs to the family Amaryllidaceae of the order Asparagales. African lily, an herbaceous and perennial monocotyledon endemic to southern Africa[4,5], is a popular ornamental and medicinal plant with rhizomes, linear strap-like dark green shiny leaves and blue to violet flowers clustered in an umbellate inflorescence[4,5].

[1]School of Ecology and Environment, Northwestern Polytechnical University, Xi'an, China. [2]Center of Special Environmental Biomechanics & Biomedical Engineering, School of Life Sciences, Northwestern Polytechnical University, Xi'an, China. [3]State Key Laboratory of Agricultural Genomics, BGI, Shenzhen, China. [4]BGI Research, Wuhan, China. [5]BGI, Beijing/Shenzhen, China. [6]Lianyungang Academy of Agricultural Sciences, Lianyungang, China. [7]National Engineering Research Center for Non-Food Biorefinery, Guangxi Academy of Sciences, Nanning, China. [8]Frontiers Science Center for Flexible Electronics (FSCFE), Shaanxi Institute of Flexible Electronics (SIFE) & Shaanxi Institute of Biomedical Materials and Engineering (SIBME), Northwestern Polytechnical University, Xi'an, China. [9]The New Zealand Institute for Plant and Food Research, Christchurch, New Zealand. [10]Department of Plant Biotechnology and Bioinformatics, Ghent University and VIB Center for Plant Systems Biology, Ghent, Belgium. [11]Guangdong Provincial Key Laboratory of Genome Read and Write, Shenzhen, China. [12]Department of Biochemistry, University of Otago, Dunedin, New Zealand. [13]These authors contributed equally: Fei Hao, Xue Liu, Botong Zhou, Zunzhe Tian, Lina Zhou, Hang Zong, Jiyan Qi, Juan He. [14]These authors jointly supervised this work: Jing Cai, Wen Wang, Richard C. Macknight, Hui Yang, Xun Xu. ✉e-mail: xuxun@genomics.cn; kittyyh@nwpu.edu.cn; richard.macknight@otago.ac.nz; wenwang@nwpu.edu.cn; jingcai@nwpu.edu.cn

The use of *Allium* species is documented in many historical and classic texts of the ancient world, including the 'Shi Jing,' the Bible and the Quran[6]. The earliest records of onions as human food are found in Egyptian tombs dated to 3,000 BC[7]. Today, onions are one of the most produced vegetables globally, with 100 million tons of dried onions produced annually, ranking second only to tomatoes (http://www.fao.org/faostat/en/#data/QC). The hallmark flavor and beneficial health properties of most *Allium* species are attributed to sulfur-containing compounds, such as allicin and isoallicin. These cleaved products of S-Alk(en)ylcysteine sulfoxide precursors (for example, alliin, isoalliin, methiin and propiin) are catalyzed by hydrolyzing enzymes, such as alliinase and lacrimal factor synthase (LFS), when subjected to mechanical injury or herbivore attack[3,8–10]. However, how these closely related *Allium* species with diverse flavor compounds evolved remains a fundamental unresolved question in plant biology and horticulture. To address this question, comparative genomic studies of *Allium* species have been conducted. Given the low coverage of short-read sequencing, repetitive sequences, which constitute most of the *Allium* genome[11,12], are likely the main reason for genome expansion[13–15], with Gypsy-type long terminal repeats (LTRs) being the most prevalent repetitive sequences[16–19]. Analysis of mid-phase chromosomes and karyotypes suggests that the abundant repetitive sequences in the *Allium* genome may be the result of hybridization and/or polyploid adaptation to different environmental conditions[15,17,20,21]. Comparative analysis of several genes using fluorescence in situ hybridization identified chromosomal inversion between onion and Welsh onion[22].

Despite progress, high-quality reference genomes for this important group of plants are still lacking. Mapping phenotype variation to genes is urgently needed to accelerate *Allium* crop breeding given the long generation times due to bulb vernalization and inefficient hand pollination[23]. At present, however, only a few variations (for example, bulb color) have been genetically mapped due to the lack of high-quality reference genomes. Furthermore, the extreme size of *Allium* genomes (from 7.4 Gb to 72.9 Gb)[12,20,24] severely hinders genomic decoding, biological research and breeding. Only two chromosome-level genome has been reported for *Allium* (garlic and Welsh onion)[19,25], but it does not reach the standard required for a reference genome given the relatively low base accuracy (the consensus quality value (QV) 37.06 for Welsh onion), low contig N50 (194 kb for garlic and 7.34 Mb for Welsh onion), low BUSCO score (88.4% for garlic and 91.0% for Welsh onion) and short average coding sequence (CDS) length (797 bp for garlic and 820 bp for Welsh onion), thereby limiting scientific conclusions.

Here we presented high-quality chromosome-level genome assemblies of three key *Allium* crops (onion ($2n = 2x = 16$), garlic ($2n = 2x = 16$) and Welsh onion ($2n = 2x = 16$)), as well as African lily ($2n = 4x = 30$) as the outgroup[4,20,24]. Our results not only provide an important resource for further study but also offer insight into the biology and breeding of these important crops.

## Results

### Chromosome-level genome assembly

We used PacBio sequencing in high fidelity (HiFi) mode with high-throughput chromosome conformation capture (Hi-C) data to assemble the four genomes (Fig. 1a,b, Table 1, Extended Data Fig. 1a and Supplementary Tables 1–6). The final genome assemblies are 15.78 Gb with a contig N50 of 81.66 Mb for onion, 15.52 Gb with a contig N50 of 109.82 Mb for garlic, 11.17 Gb with a contig N50 of 507.27 Mb for Welsh onion and 10.48 Gb with a contig N50 of 7.32 Mb for African lily. The genome sizes of the four assemblies were comparable to flow cytometry estimations reported in previous studies[4,20,24] (Supplementary Table 7). Furthermore, the contig N50 sizes of the three *Allium* species were not only the highest among recently published *Allium* assemblies[19,25,26], but also the highest recorded for plants with large genomes[27–30]. The QV of the genome assemblies of the three *Allium* species and African lily (55.38–69.61) indicate very high base accuracy (Supplementary Table 8 and Supplementary Result 1). Annotation of the assemblies found telomeric repeats at many chromosomal ends, suggesting the chromosome assemblies are of high quality (Supplementary Tables 9–12 and Supplementary Result 1). Annotation of protein-coding genes and BUSCO[31] assessment showed high gene completeness and accuracy (Table 1, Supplementary Tables 13–20 and Supplementary Result 1).

Based on genome annotation, further phylogenetic analysis and divergence time estimation showed that the three *Allium* species clustered together to form a monophyletic group, which diverged from its closest outgroup (African lily) ~69.2 million years ago (Mya; Fig. 1c, Extended Data Fig. 1b–f, Supplementary Table 21 and Supplementary Result 2). Chromosome comparisons of the *Allium* species identified several large-scale reciprocal inversions in Welsh onion–garlic and onion–Welsh onion species pairs (Fig. 1a,d, Supplementary Table 22 and Supplementary Result 3). We also included the previously published garlic genome[19] in chromosome comparison and found that our garlic assembly exhibited better synteny with the other *Allium* species, thereby suggesting higher quality (Fig. 1e, Supplementary Table 22 and Supplementary Result 3). Categorization of repetitive elements showed LTRs were dominant in the four genomes, with Copia (8.90–52.16%) and Gypsy (40.26–54.84%) being the two largest LTR families (Extended Data Fig. 2a, Supplementary Tables 23 and 24 and Supplementary Result 4). Furthermore, divergence analysis of LTR showed that African lily has a sharp concentrated Copia peak at ~10% divergence, whereas the *Allium* species show flatter divergence distribution (Extended Data Fig. 2b,d). For Gypsy, the four species show a more similar distribution (Extended Data Fig. 2c,d and Supplementary Result 4). Gene family expansion and gene evolutionary analysis identified many expanded gene families and rapidly evolving or positively selected genes that were related to defense function and alliin biosynthesis in the *Allium* lineage (Extended Data Fig. 3, Supplementary Tables 25–43 and Supplementary Result 5).

### *Allium*-specific polyploidization events

Whole-genome duplications (WGDs) have an important role in the evolution of angiosperm genomes. In the African lily genome, we discovered numerous groups of eight homologous genomic regions through both intragenomic (Fig. 2a) and intergenomic comparisons (dot-plot analyses) with *Ananas comosus*[32], *Asparagus officinalis*[33] and *Dendrobium chrysotoxum*[34] (Fig. 2b and Supplementary Fig. 1). Based on *Ks* distribution in the African lily genome and phylogenetic analysis of African lily and *A. officinalis* (Fig. 2c, Supplementary Figs. 2–4 and Supplementary Result 6), we found that African lily probably experienced three WGD events, with the most ancient $\tau$ WGD event shared by most monocots[35,36]. Using the $\tau$ WGD event as a time calibration point[37], the other two WGDs were estimated to have occurred at ~70 to 75 Mya (denoted as $\rho$) and 20–22 Mya (denoted as $\omega$), respectively (Supplementary Fig. 3). Further comparison with divergence time corrected by ksrates[38] showed that the second WGD event $\rho$ likely occurred before the divergence of African lily and alliums, whereas the third WGD event $\omega$ likely occurred after their divergence (Supplementary Fig. 5 and Supplementary Result 6).

For the three *Allium* species, mapping of the intraspecies syntenic blocks found within African lily and *A. officinalis* to the *Allium* genomes identified up to 16 orthologous syntenic blocks (Fig. 2a,b, Supplementary Figs. 6–12 and 13a,b and Supplementary Result 6), suggesting that the *Allium* species may have undergone four WGD events ($2 \times 2 \times 2 \times 2 = 16$), including two in the lineage leading to *Allium*, one before the divergence between African lily and the lineage leading to *Allium* and one (most ancient $\tau$ WGD) shared by most monocots[35,36]. Clustering based on similarity matrices among homologous syntenic regions between Welsh onion and *A. officinalis* and African lily pairs (Supplementary Figs. 13c and 14 and Supplementary Result 6) suggested that the second oldest WGD was likely the $\rho$ WGD event shared with African lily, as supported by gene tree discordance analysis

https://doi.org/10.1038/s41588-023-01546-0

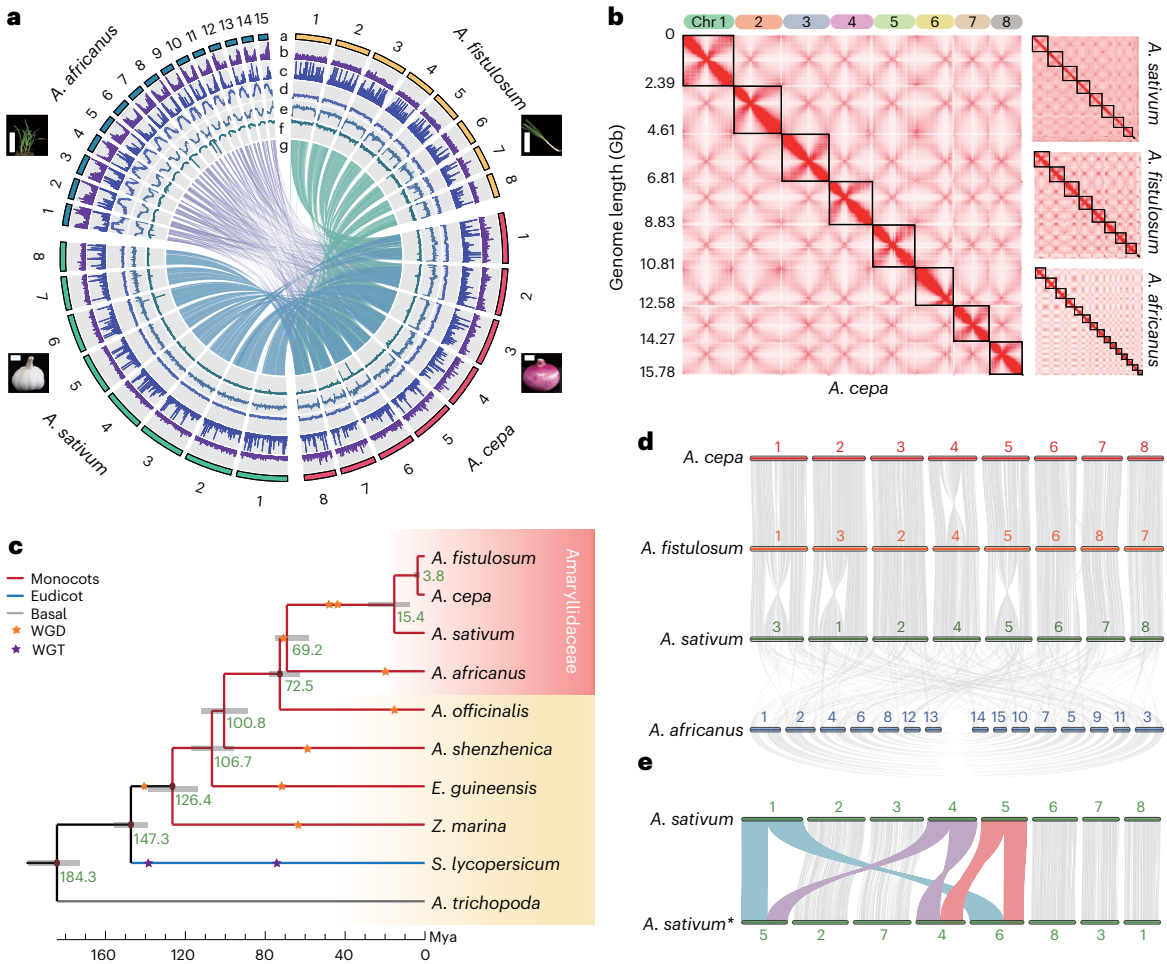



**Fig. 1 | Evolution of *Allium* and African lily genomes. a**, Overview of *Allium* species and African lily genomes. Track a corresponds to chromosome length. Tracks b–f, five rings represent the density distribution of protein-coding genes, gene expression, LTR/Copia, LTR/Gypsy and repeat sequences, respectively. Track g corresponds to syntenic blocks between onion and other three species. Scale bars equal to 5.5 cm for onion, 2.5 cm for garlic, 22.1 cm for Welsh onion and 17.4 cm for African lily. **b**, Interaction map of Hi-C data of four species. Each black frame represents one chromosome. With numbered chromosomes as coordinates, color of each dot represents log value of interaction intensity of the corresponding bin pair of genomes, and interaction intensity increases from white to red. Chr, chromosomes. **c**, Maximum-likelihood phylogenetic tree from single-copy ortholog genes of ten plant species dated with fossil calibrations. To compute node support, 1,000 bootstraps were used, and all nodes showed 100% support. Green numbers denote the divergence time of each node (Mya), and gray bars are 95% confidence intervals for the time of divergence between different clades. WGD or whole-genome triplication (WGT) events[33,36,65,78–80] are shown as asterisks. **d**, Synteny patterns between genomic regions from three *Allium* plants and African lily. Numbers represent individual chromosomes. **e**, Whole-genome alignment between two garlic genome assemblies. Upper one is an assembly from this study, and lower one with an asterisk represents an assembly from a previous study[19]. *A. fistulosum*, *Allium fistulosum*; *A. cepa*, *Allium cepa*; *A. sativum*, *Allium sativum*; *A. africanus*, *Agapanthus africanus*; *A. officinalis*, *Asparagus officinalis*; *A. shenzhenica*, *Apostasia shenzhenica*; *E. guineensis*, *Elaeis guineensis*; *Z. marina*, *Zostera marina*; *S. lycopersicum*, *Solanum lycopersicum*; *A. trichopoda*, *Amborella trichopoda*.

(Supplementary Fig. 15). *Ks* distribution (Supplementary Figs. 2 and 3), dot-plot analysis (Supplementary Fig. 6) and gene tree analysis of syntenic regions (Supplementary Fig. 16 and Supplementary Result 6) in Welsh onion indicated that the two most recent WGD events ($\alpha$ and $\beta$) occurred within a short period (46–50 Mya; Supplementary Fig. 3b) almost simultaneously (referred to as $\alpha/\beta$) before divergence of the three *Allium* species.

In summary, the African lily and common ancestor of *Allium* species first shared ancient $\tau$ and $\rho$ WGD events, then independently experienced more recent WGD events. African lily experienced a WGD event ($\omega$) around 20–22 Mya, whereas *Allium* species experienced two WGD events ($\alpha$ and $\beta$) over a short period around 46–50 Mya (Figs. 1c and 2c).

## Evolution of allicin/isoallicin pathway in *Allium*

Although the putative biosynthetic pathway for allicin and isoallicin was recently reported[3,39] (Fig. 3a), the evolutionary history of this pathway in *Allium* remains to be elucidated. Here we searched for homologous genes involved in the biosynthetic pathway of allicin and isoallicin in 89 representative plant species. We found that genes responsible for the biosynthesis of precursors alliin/isoalliin and genes encoding hydrolyzing enzymes were present in most plant genomes. However, the copy numbers of the alliinase and *LFS* genes, which encode key enzymes of the alliin/isoalliin cleavage reaction, underwent tremendous *Allium*-specific expansion. While most plant species contained only a few copies, the *Allium* plants contained as many as 60 copies of the alliinase gene and 30 copies of the *LFS* gene (Fig. 3b and Supplementary Table 44). Metabolomic analysis also showed that alliin content, used as a precursor for allicin biosynthesis, was 1,000 times lower in the African lily than in the tissues of garlic, onion and Welsh onion (Supplementary Table 45), as supported in previous studies showing that allicin only exists in *Allium*[3,10,12]. Phylogenetic analysis of the alliinase and *LFS* gene families showed that the alliinase and *LFS* genes

https://doi.org/10.1038/s41588-023-01546-0

**Table 1 | Assembly statistics of four genomes**

| Parameters | Onion | Garlic | Welsh onion | African lily |
|---|---|---|---|---|
| Contig N50 (Mb) | 81.66 | 109.82 | 507.27 | 7.32 |
| Number of contigs | 2,984 | 8,155 | 605 | 4,914 |
| Total length of contigs (bp) | 15,941,260,873 | 16,462,710,162 | 11,674,854,025 | 20,167,188,658[a] |
| Hi-C mounting ratio (%) | 99.01 | 94.30 | 95.69 | 99.25 |
| Total size of assembled chromosomes (bp) | 15,782,911,972 | 15,523,730,015 | 11,171,194,818 | 10,477,061,407 |
| GC content (%) | 33.67 | 35.87 | 35.45 | 42.39 |
| Complete ratio of BUSCO (%) | 96.40 | 92.60 | 96.60 | 94.40 |
| Repeat content (%) | 93.79 | 92.75 | 92.10 | 94.36 |
| Gene number | 61,619 | 57,183 | 50,234 | 87,418 |
| Average gene length (bp) | 3,552.06 | 5,228.54 | 6,110.15 | 7,154.27 |
| Average coding sequence length (bp) | 1,157.08 | 1,190.76 | 1,021.84 | 1,282.78 |

[a]The total size of all contigs from both haploid genomes.

formed 15 and 9 distinct groups, respectively (Fig. 3c and Extended Data Figs. 4–5), many of which were shared by the three *Allium* species. We also identified species-specific groups with many gene copies, suggesting these genes have experienced continuous expansion and may be responsible for the distinct flavors of these alliums.

Our results showed that neither alliinase nor *LFS* genes were found in the syntenic blocks derived from WGD events; rather, most formed clusters on chromosomes, indicating that they arose from local gene duplication. Furthermore, the allicin/isoallicin synthesis pathway genes were mainly distributed on chromosome 2 in Welsh onion and onion, and on chromosome 1 in garlic (Extended Data Fig. 6). This clustering pattern may facilitate efficient transcriptional coregulation, as observed in the synthesis of other secondary metabolites[40,41].

We also detected positive selection and rapid evolution of several genes in the alliin biosynthetic pathway in *Allium* (Fig. 3a and Supplementary Table 46). Specifically, two key enzyme genes, *GGT* (encoding γ-glutamyl transpeptidases) and *PCS1* (encoding phytochelatin synthase), experienced strong positive selection on the ancestral branch of the three *Allium* species and contained *Allium*-specific changes in their amino acid sequences (Fig. 3d).

## Evolution of immune system in *Allium* species

Generally, plants recognize pathogen infection through a two-tiered innate immune system composed of pattern-triggered immunity (PTI) and effector-triggered immunity. Here transcriptome analysis revealed that six of the seven genes involved in the alliin biosynthetic pathway (that is, *AcpGSH1* (encoding glutamate cysteine ligase), *AcpGSH2* (encoding glutathione synthetase), *AcpPCS1*, *AcpGGT1*, *AcpGGT3* and *AcpFMO1* (encoding flavin-containing monooxygenase) were upregulated in onion leaf blades infected with a pathogenic fungus (*Phytophthora*) compared with healthy leaf blades (Extended Data Fig. 7a). The upregulated genes were significantly ($P < 0.05$) enriched in the PTI-mitogen-activated protein kinase (MAPK) signaling pathway (Extended Data Fig. 7b). Quantitative real-time reverse transcription polymerase chain reaction showed that the expression levels of *AcpGSH1*, *AcpGGT3* and *AcpFMO1* were also upregulated in onion leaf blades artificially injured and exposed to the *Pseudomonas syringae* pv tomato (Pst) strain DC3000 compared to healthy leaf blades (Extended Data Fig. 7c). Previous research has reported that the plant immune system initiates downstream gene responses by activating signaling pathways such as MAPK[42]. Two important transcription factors activated by the MAPK pathway, WRKY22/WRKY29 and MYB, which specifically interact with the W box (5′-(T)TGAC[CT]-3′) and Myb recognition elements, turn on the transcription of many immune response genes in plants[42]. In this study, by searching the *Allium* genomes for these *cis*-regulatory elements[43] in the upstream sequences of genes in the allicin/isoallicin biosynthetic pathway, we identified W box elements upstream of three genes encoding enzymes involved in allicin/isoallicin biosynthesis, that is, *GSH1*, *GSH2* and *FMO1* (Supplementary Table 47). These findings indicate that the plant immune system may regulate allicin/isoallicin biosynthesis through WRKY22/29 or MYB transcriptional control. Thus, our results suggest that the alliin biosynthetic pathway is upregulated by fungal and bacterial infections and may contribute to downstream immune responses to general microbial infection.

## Spatial pattern of secondary metabolism in onion bulbs

Onion is a biennial plant with a bulb storage organ for overwintering. These bulbs contain a variety of secondary metabolites, including flavonoids such as flavone, flavanol and anthocyanidin. Spatial RNA sequencing (RNA-seq) provides insights into the cellular-level expression patterns of multiple genes in the same metabolic pathway. Here we selected six onions at three developmental stages and performed Spatiotemporal Enhanced REsolution Omics-sequencing (Stereo-seq)[44–46] on the bulb bases. Stereo-seq combined DNA nanoball (DNB) patterned array chips, in situ RNA capture and sequencing. We performed unsupervised clustering of the spatial expression matrix and categorized clusters into six cell types, including spongy mesophyll cells, palisade cells, basal parenchyma cells, epidermal cells, leaf vascular cells and xylem cells based on cell morphology, spatial location and in situ hybridization of cell marker genes (Fig. 4a–c and Extended Data Fig. 8). Previous study has shown that, in the onion bulbs, the *CHS* (encoding chalcone synthase), *CHI* (encoding chalcone isomerase), *F3H* (encoding flavanone 3-hydroxylase), *F3′H* (encoding flavonoid 3′ hydroxylase), *F3′5′H* (encoding flavonoid 3′5′ hydroxylase) and *FLS* (encoding flavonol synthase) genes are involved in flavonoid synthesis and *FLS* converts dihydroflavonols to flavonols, with quercetin being a major flavonoid in onion[47] (Fig. 4d). Quercetin is reported to confer antioxidant activity and disease-resistance in onion[48,49]. Except for *F3′5′H*, these genes showed differential expression among clusters and co-expression in epidermal cells (Fig. 4e, Extended Data Fig. 9a and Supplementary Table 48), indicating that epidermal cells in onion bulbs are the center of flavonoid synthesis.

We also found that the lipid transfer protein gene family (*LTP*) is specifically expressed in epidermal cells based on spatial RNA-seq and RNA in situ hybridization (Fig. 4f and Extended Data Fig. 9), consistent with the expression pattern found in maize leaves by RNA in situ hybridization and in *Arabidopsis* stem epidermis by RNA-seq[50,51]. The LTP proteins are involved in the assembly of cuticular wax[52,53], a hydrophobic component that prevents nonstomatal water loss and provides

https://doi.org/10.1038/s41588-023-01546-0

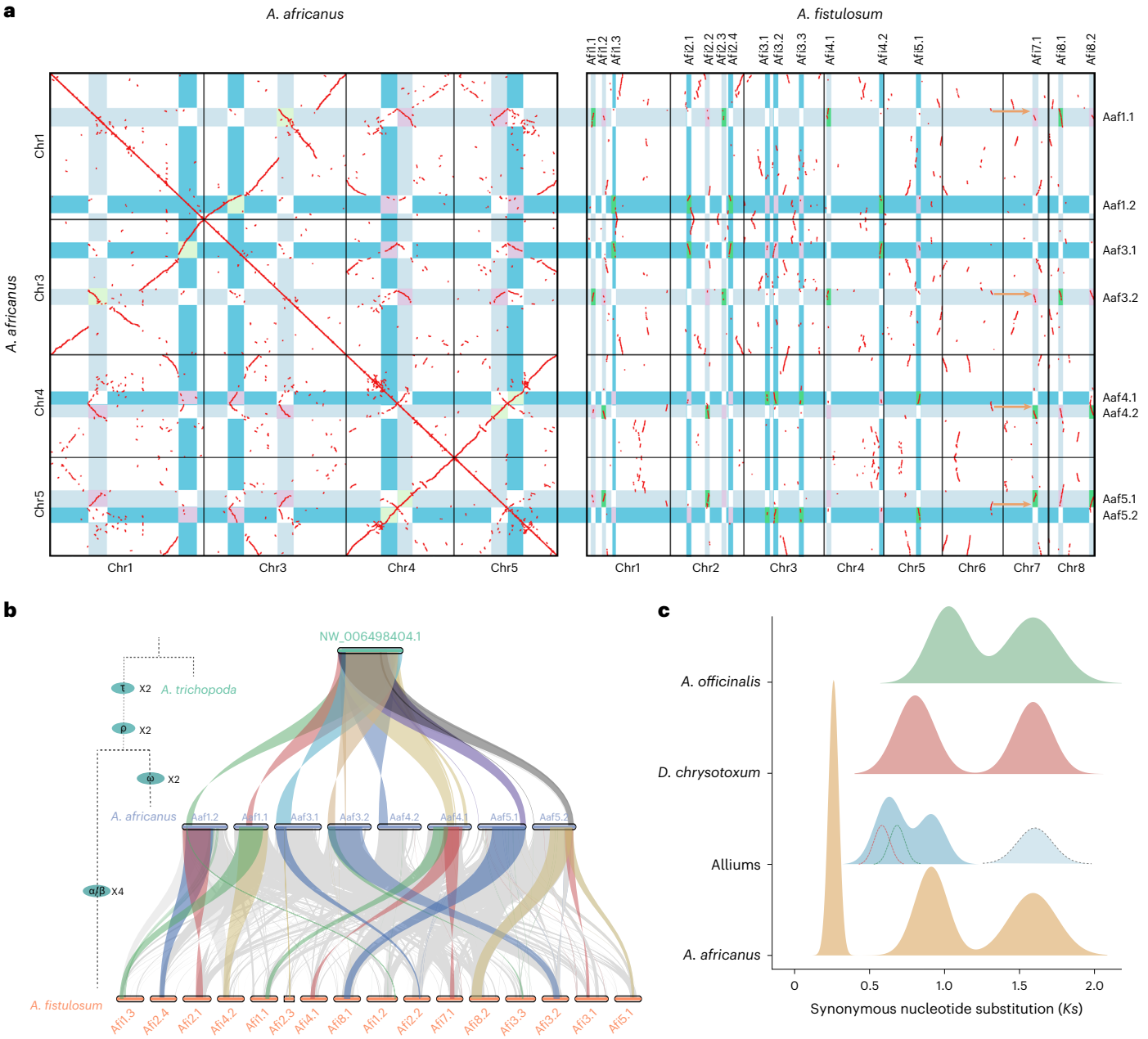



**Fig. 2 | Comparative genomic analyses. a**, Synteny blocks within the African lily genome (left) and between Welsh onion and African lily genomes (right). Dark/light blue rectangles connect synteny blocks with syntenic depth ratios of 8:8 within the African lily and 8:16 between the African lily and Welsh onion. Two colors represent two groups of syntenic blocks originating from the early $\tau$ event, showing better synteny within each group than between two groups. Light green/purple/white boxes on left highlight pairs of synteny blocks tracing to $\omega/\rho/\tau$ event, while green/purple/white boxes on right highlight syntenic pairs tracing to $(\alpha/\beta)/\rho/\tau$ event. Chromosome numbers are shown on left and below. Names of synteny blocks are on top and left, named with three letters for species name (Aaf for African lily and Afi for Welsh onion), followed by chromosome number before the dot and number for block after the dot. **b**, Alignment of homologous genomic fragments among *A. trichopoda*, Welsh onion and African lily. All syntenic blocks identified by WGDI with default parameters were connected to curved lines by MCScan (python version) with default parameters. Large genomic tracts of multiple syntenic blocks are highlighted in different colors. Syntenic depth ratios were 1:8, 8:16, and 1:16 for *A. trichopoda*–African lily, African lily–Welsh onion and *A. trichopoda*–Welsh onion species pairs, respectively. **c**, Gaussian fitting of average synonymous substitution rate (*Ks*) distribution of gene pairs in syntenic blocks in four monocot species. Please see Supplementary Fig. 3a for the original *Ks* distribution and details of approximation. *Y* axis is frequency of genes at corresponding *Ks* in syntenic blocks. Putative peaks for $\alpha/\beta$ and $\tau$ events in *Alliums* are drawn with dotted lines. *Alliums*: onion, garlic and Welsh onion. Chr, chromosome.

protection against harmful elements. Thus, LTP may have an important role in water loss prevention and pest resistance in onion bulbs.

### Genes related to evolution and formation of tunicate bulbs

As the main edible part, bulb formation is an important biological process in onion production. Bulb formation begins with cell volume expansion in the sheath of the innermost emerged leaves (usually the fifth or sixth leaves from the outermost leaves), followed by expansion of unemerged leaves from outside to inside[54]. Spongy mesophyll cells are the dominant cell type in bulbs (Supplementary Fig. 17), and their expansion is the morphological basis underlying bulb formation[55]. To clarify their molecular biological processes, we collected bins associated with spongy mesophyll cells from the spatial RNA-seq data matrix and constructed lineage differentiation trajectories using monocle3

https://doi.org/10.1038/s41588-023-01546-0

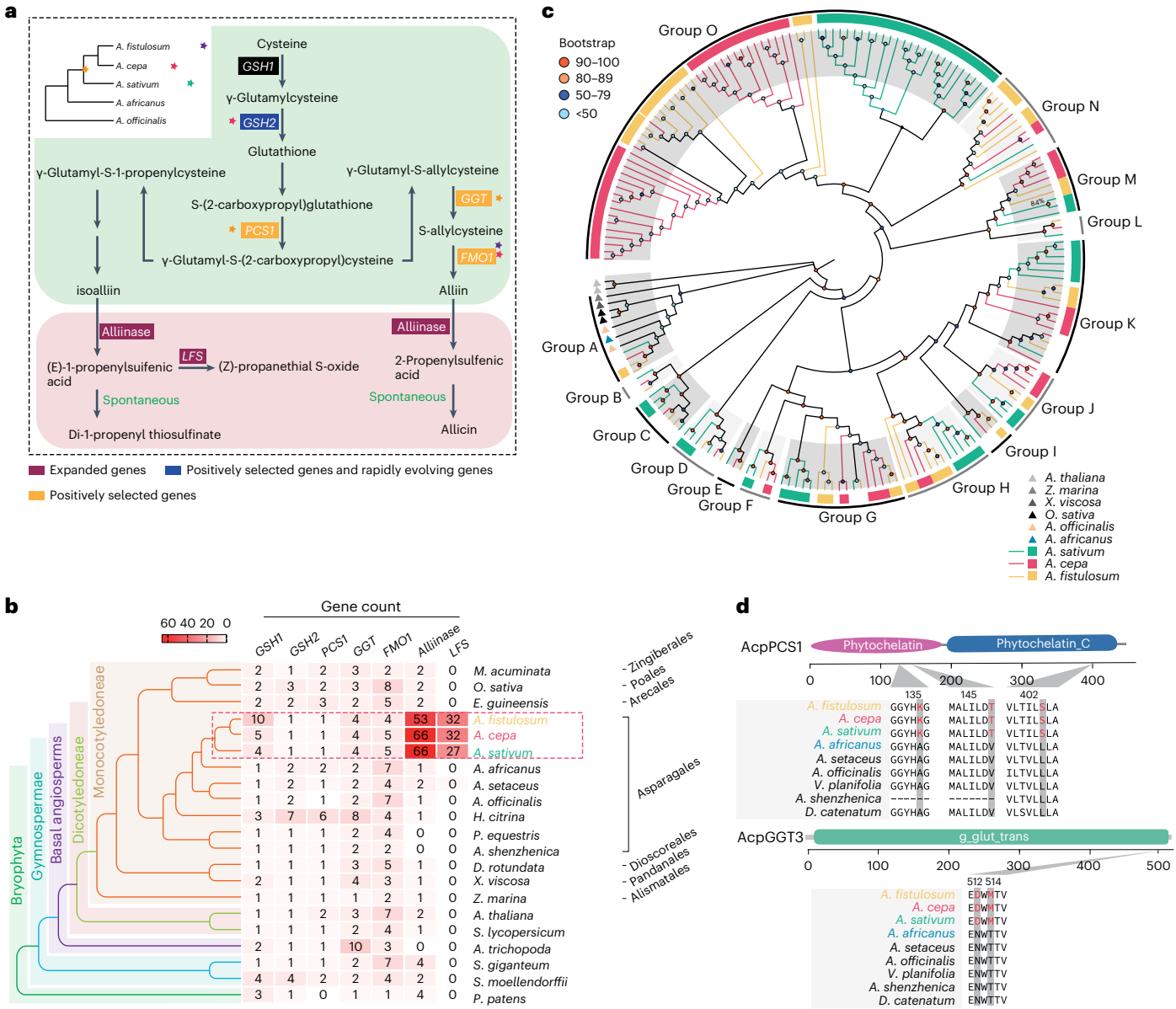



**Fig. 3 | Insights into allicin/isoallicin biosynthesis. a**, Putative biosynthetic pathway of allicin and isoallicin in *Allium* plants[3]. Purple stars indicate positive selection and rapid evolution of genes in Welsh onion. Green stars indicate positive selection and rapid evolution of genes in garlic. Red stars indicate positive selection and rapid evolution of genes in Welsh onion. Orange stars indicate positive selection of genes in ancestors of three *Allium* species (see Supplementary Data 1 for alignments of sequences). **b**, Copy number of homologous genes involved in biosynthesis of allicin and isoallicin in 21 representative species (see Supplementary Table 44 for complete data of 89 species). Phylogenetic tree was drawn according to the Angiosperm Phylogeny Group (APG) IV system. **c**, Characterization of alliinase gene family. All alliinase genes were categorized into 15 groups (see Supplementary Data 2 for alignments of sequences and version with support values on branches). **d**, Specific mutations in *Allium* AcpPCS1 and AcpGGT3. *M. acuminata*, *Musa acuminata*; *A. setaceus*, *Asparagus setaceus*; *H. citrina*, *Hemerocallis citrina*; *D. rotundata*, *Dioscorea rotundata*; *S. giganteum*, *Sequoiadendron giganteum*; *S. moellendorffii*, *Selaginella moellendorffii*; *P. patens*, *Physcomitrella patens*; *V. planifolia*, *Vanilla planifolia*; *D. catenatum*, *Dendrobium catenatum*.

(refs. 56,57; Fig. 4g and Extended Data Fig. 10a–c). Pseudotime analysis showed that accompanying the development of onion bulbs, spongy mesophyll cells displayed a continuous cell trajectory from the inner to outer layer and from the base to the top (Fig. 4g and Extended Data Fig. 10a–c). This suggests that cell expansion first occurs in the outer bulb scale away from the basal stem, consistent with previous descriptions of cell expansion and microtubule disruption processes[55]. For each layer of bulb, the order of sponge mesophyll cells expansion was from the outer edge to the inner edge, indicating that the inner edge cells start to expand earlier and are pushed further inward by later expansion of cells near the outer edge. Bulb formation is a process of well-organized spatiotemporal changes in cell morphology. To further investigate the molecular processes underlying cell expansion trajectories, we performed differential analysis of the spatial expression matrix of the spongy mesophyll cell lineage and enrichment analysis of the top 1,000 genes with the lowest $q$ value against all genes (background; Supplementary Table 48). The Gene Ontology (GO) terms significantly enriched ($P < 0.05$) in the genes with the most noticeable change in expression during cell expansion were 'microtubule-based process,' 'structural constituent of cytoskeleton,' 'abscisic acid (ABA) binding' and 'ABA-activated signaling' and Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) term 'xyloglucan:xyloglucosyl transferase,' suggesting these processes have important roles in cell expansion (Fig. 4h and Supplementary Fig. 18).

https://doi.org/10.1038/s41588-023-01546-0

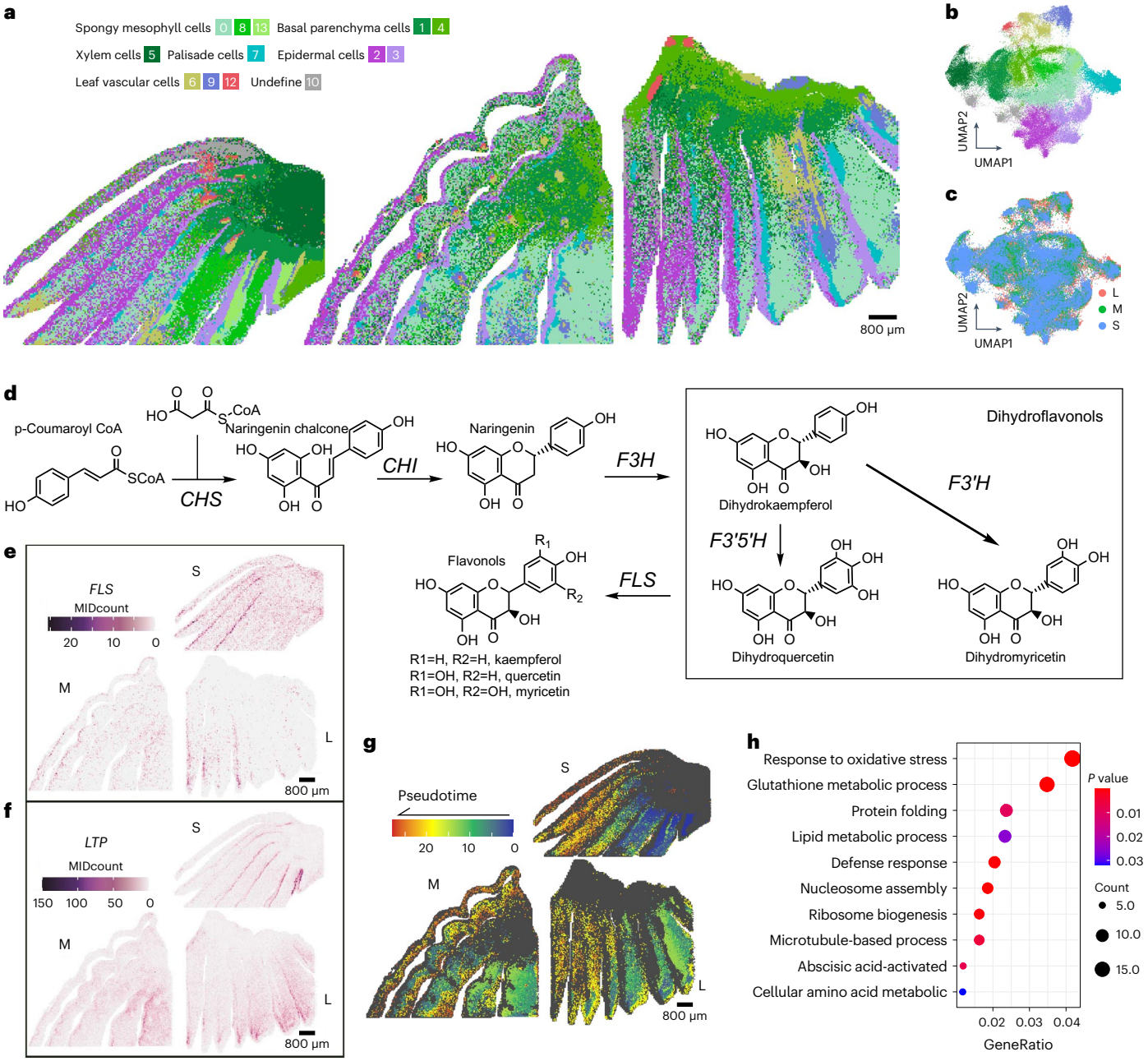



**Fig. 4 | Spatial RNA-seq reveals spatial distribution of gene expression in metabolic process and onion bulb formation. a**, Unsupervised clustering of three consecutive period onion bulbs. We used a bin size of 80 capture sites to approximate cell size. Bins are colored according to cell types coded on top-left. Same color code is also used in **b**. Left, early bulb (S); center, midterm bulb (M); right, mature bulb (L). **b**,**c**, UMAP dimensionality reduction projection of onion bulb cells, grouped by clusters and developmental stages, respectively. UMAP1 and UMAP2 represent two dimensions resulting from dimensionality reduction. Separation of different clusters in **b** indicates robustness of cell classification, while overlap of different developmental stages in **c** indicates stable cell-type constitution across three stages. **d**, Flavonoid metabolism in onion epidermal cell. **e**,**f**, Spatial expression of genes (*CHS* and *LTP*) in different onion bulb developmental stages. **g**, Spongy mesophyll cell lineages visualized in spatial coordinates. **h**, Significant biological process GO terms enriched in top 1,000 DEGs with lowest *q* value in spongy mesophyll cell lineages. The *P* values were calculated using the hypergeometric test (one side) and adjusted by the Benjamini–Hochberg method. Scale bars equal to 800 μm for **a** and **e**–**g**.

Based on our spatial transcriptome analysis and previous studies, bulb formation likely involves several cellular processes, including cytoskeleton re-organization[55], cell osmotic pressure regulation[58] and plant hormone interactions[59]. We therefore carried out evolutionary and expression analyses of the genes of these cellular processes.

Under long-day conditions, expansion of leaf sheath cells shifts from longitudinal to lateral during bulb development by cortical microtubule rearrangement[55]. In our study, the adenosine triphosphate motor protein-related genes including tubulin α chain gene (*TUA*), tubulin β chain gene (*TUB*), γ-tubulin complex component 4 gene (*GCP4*), tubulin-folding cofactor C gene (*TFCC*) and fragile fiber/kinesin-like protein KIN-4A gene (*FRA*) involved in microtubule rearrangement were under positive selection in onion[60,61] (Fig. 5a and Supplementary Table 49). *FRA* regulates cortical microtubule stability[62], while *TUA* encodes the tubulin α chain[63]. Positive selection in these two genes in onion may contribute to the phenotypic differences between onion and Welsh onion bulbing; onion has a substantial ball-like bulb with thickened leaf sheaths, whereas Welsh onion has a smooth leaf sheath along its entire

https://doi.org/10.1038/s41588-023-01546-0

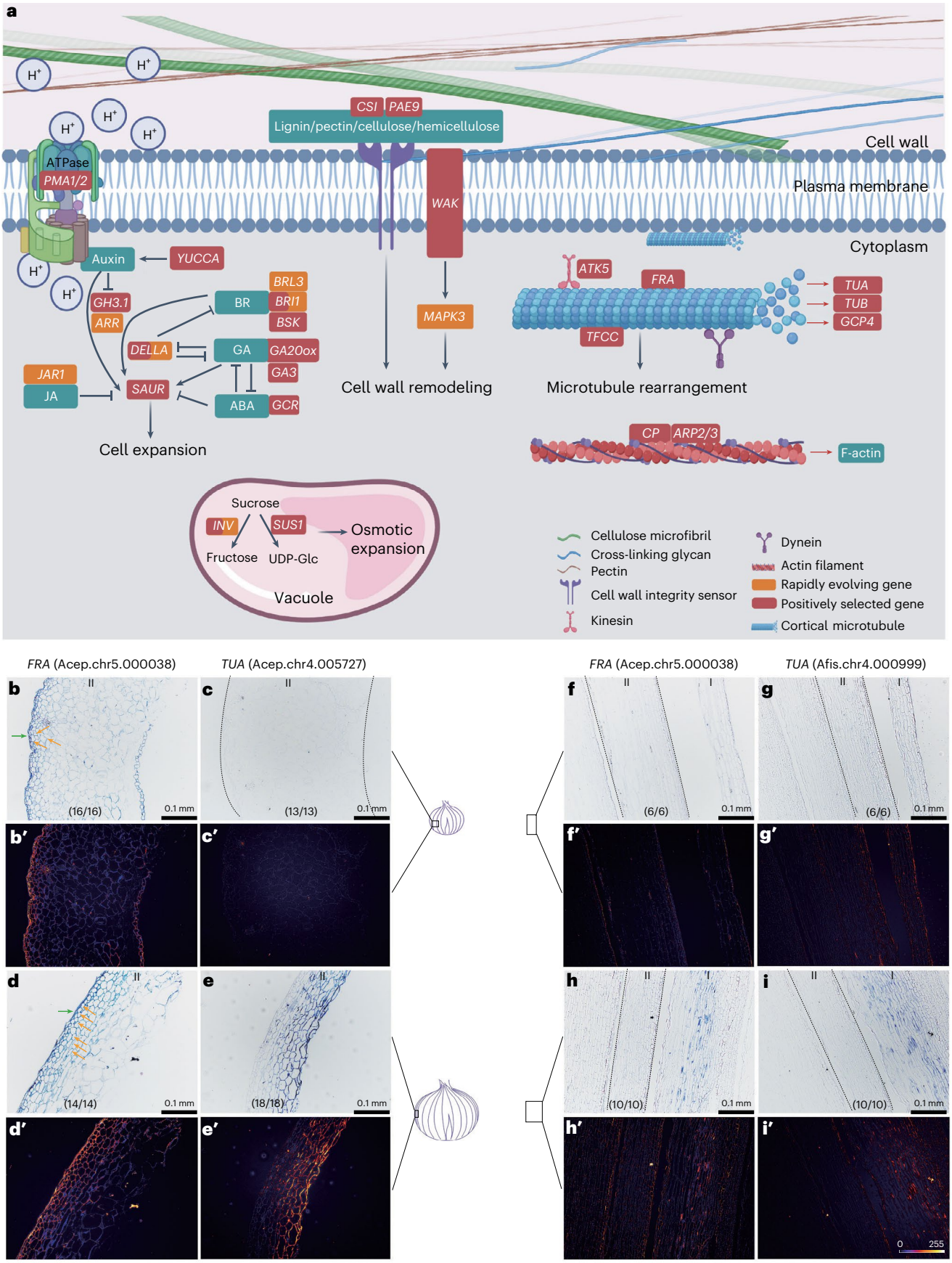


**Fig. 5 | Evolution of genes involved in bulb formation in *Allium* plants and expression pattern divergence. a**, Red genes represent positively selected genes (PSGs). Orange genes represent rapidly evolving genes (REGs). PSGs *FRA*, *TFCC*, *TUA*, *TUB* and *GCP4* are involved in microtubule rearrangement. PSGs *ARP2/3* are involved in enforcing actin filament. PSGs *WAK* and REG *MAPK3* are related to regulators of cell wall remodeling. Genes involved in hormone biosynthesis and signaling pathways form complex interactions to regulate cell expansion, including *YUCCA* (PSG) and *ARR* (REG) in auxin biosynthesis and signal response, *JAR1* (REG) in JA (jasmonic acid) biosynthesis, *BSK* (PSG) and *BRL3* (REG) in BR (brassinolide) signaling, *GA3* (PSG) in GA (gibberellin) signaling response. *INV* (PSG) and *SUS1* (PSG) in sucrose metabolism can induce osmotic expansion. The model was drawn using BioRender (https://biorender.com/). **b**–**i**, Expression patterns of *FRA* and *TUA* during onion bulb formation based on in situ hybridization. **b′**–**i′**, Expression signal intensities of *FRA* and *TUA* in onion and Welsh onion leaf sheaths in **b**–**i** were selected and transformed to pseudo-colors with ImageJ. In onion, *FRA* transcript was observed in epidermal cells (green arrow) and two or three layers of cells (orange arrows) of the second leaf sheath, increasing to five or six layers of cells during late bulb swelling (**b**, **d**, **b′** and **d′**). *TUA* transcript was observed in the inner cells of the second leaf sheath at late bulb formation (**c**, **e**, **c′** and **e′**). In Welsh onion, accumulation of *FRA* and *TUA* transcripts was not observed in second leaf sheaths (**f**–**i**, **f′**–**i′**). Scale bars equal to 0.1 mm for **b**–**i**. Dotted lines are drawn to show the outline of the second leaf sheath. Number in parenthesis represents the number of samples with positive signals over a total number of samples.

length. We further compared the spatiotemporal expression patterns of these two positively selected genes in the leaf base of onion and Welsh onion in the early and late stages of bulbing using RNA in situ hybridization. Interestingly, we found that in the second leaf sheath (from outer to inner) of onion, *FRA* was constantly expressed in the epidermal cells and the outermost two or three layers of cells during early development but extended to the outermost five or six layers of cells in the late bulbing stage (Fig. 5b,d,b′,d′ and Supplementary Fig. 19). Furthermore, *TUA* expression was detected in the inner cells of the second leaf sheath only during late bulb formation (Fig. 5c,e,c′,e′ and Supplementary Fig. 19). Neither gene was observed at either stage in the second leaf sheath of Welsh onion (Fig. 5f–i,f′–i′ and Supplementary Fig. 19), although weak expression was detected in the outermost leaf sheath (Fig. 5h,i,h′,i′ and Supplementary Fig. 19). However, using spatial RNA-seq, we did not observe a clear expression pattern of the *FRA* and *TUA* genes (Extended Data Fig. 10d,e), probably because they are expressed at relatively lower levels. These results suggest that RNA in situ hybridization is an important complement to spatial RNA-seq when studying lowly expressed genes. In addition, the genes related to actin filament movement and actin-related protein 2/3 complex gene (*ARP2/3*)[64] were found to be positively selected in onion (Fig. 5a). Thus, these results indicate that genes associated with the cytoskeleton, including microtubules and actin filaments, may have a role in onion bulb formation.

Furthermore, we found four genes likely involved in cell wall remodeling (wall-associated receptor kinase gene (*WAK*), MAPK gene (*MAPK3*)) and two sucrose metabolism genes likely involved in cellular osmotic pressure regulation (β-fructofuranosidase gene (*INV*) and sucrose synthase 1 gene (*SUS1*)) were positively selected or rapidly evolved in onion and may influence bulb expansion (Fig. 5a).

The above processes are coregulated by many plant hormones, including auxin (IAA), gibberellin (GA), brassinolide (BR), ABA and jasmonic acid (JA). Indeed, we found that several onion genes in the five plant hormone pathways are under positive selection or rapid evolution (Fig. 5a) and the plant hormone signal transduction pathway genes are also enriched in differentially expressed genes (DEGs) during parenchymal cell growth and expansion (Fig. 4h and Supplementary Fig. 18). Taken together, these findings provide evolutionary and experimental evidence of the potential genes responsible for bulb formation.

## Discussion

Our results showed that the three *Allium* species experienced at least four WGD events, more than other closely related lineages such as African lily, *A. officinalis*[33], *Apostasia Shenzhenica*[36] and *Elaeis Guineensis*[65] (Figs. 1c and 2). The $\rho$ WGD event occurred around 70–75 Mya, coinciding with the Cretaceous–Paleogene (K–Pg) boundary, when many angiosperm lineages experienced almost simultaneous polyploidization events[65–69].

Sulfur-containing active compounds constitute a unique defense system in *Allium* species and also underlie the basis of their health benefits in therapeutic usage[3,70]. These compounds are stored as inert precursors (for example, alliin/isoalliin) in the cytosol, while a large number of hydrolyzing enzymes such as alliinase/LFS (accounting for 10% of total protein in some tissues) are stored in the vacuole[2,71]. Upon mechanical injury, this spatial separation is disturbed, and the alliin–alliinase system rapidly releases a large amount of hydrolyzed active compounds, such as allicin/isoallicin and lachrymator[2,3,10]. Although allicin has been reported to only exist in *Allium* species[3,10,12], alliin was found in all four species including African lily (Supplementary Table 45), and homologs of genes involved in the biosynthesis of precursors and genes encoding hydrolyzing enzymes were present in most plant genomes. When *Arabidopsis thaliana* is infected with *Sclerotinia sclerotiorum*, expression of the *FMO* gene is substantially upregulated and involved in the immune response[72]. *PCS1* also has an important role in the immune response of *A. thaliana* by localizing to the site of fungal attack, and mutants of this gene are hypersusceptible to infection[73,74]. Here based on *A. thaliana* transcriptome data[42], we found that the expression levels of *GSH2*, *PCS1*, *GGT3* and *FMO1* were upregulated in plant immune responses (both PTI-induced and intracellular receptor-induced responses) against Pst DC3000, with or without the bacterial intracellular effector avrRpt2 (Extended Data Fig. 7d,e). Thus, we speculate that genes related to alliin biosynthesis and hydrolyzing enzymes constitute a conserved defense system widely distributed across plant lineages. This defense system appears to be strengthened in *Allium*, with dramatic sequence changes (positive selection or rapid evolution) in several biosynthesis genes contributing to more efficient synthesis of precursors and tandem gene duplication leading to many genes encoding hydrolyzing enzymes.

Single-cell RNA-seq is difficult to perform in plants[75–77]. However, based on high-resolution Stereo-seq technology, our data provide new insights into onion bulb development and metabolism even at the subcellular nanoscale, with 100-fold higher resolution than Visium Spatial Gene Expression of 10x Genomics[44]. Our Stereo-seq results showed that onion bulb epidermal cells co-expressed many flavonoid synthesis genes, indicating that epidermal cells are the primary source of onion bulb flavonoids. In addition, the onion epidermis showed high expression of the *LTP* gene family, with *LTP*, *LTP2* and *LTP3* involved in the formation of onion epidermal wax. These findings provide fundamental knowledge for further research on onion pest control and the breeding of pest-resistant lines.

Taking the four species together, this is an extensive genomic study, with a chromosome-level assembly totaling 52.96 Gb. These results provide a fundamental update to the genetic resources and molecular mechanisms of trait evolution in *Allium*, which will facilitate crop breeding.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-023-01546-0.

## References

1. Chase, M. W. et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181**, 1–20 (2016).

2. Jones, M. G. et al. Biosynthesis of the flavour precursors of onion and garlic. *J. Exp. Bot.* **55**, 1903–1918 (2004).
3. Yoshimoto, N. & Saito, K. S-Alk (en)ylcysteine sulfoxides in the genus *Allium*: proposed biosynthesis, chemical conversion, and bioactivities. *J. Exp. Bot.* **70**, 4123–4137 (2019).
4. Reis, A. C. et al. rDNA mapping, heterochromatin characterization and AT/GC content of *Agapanthus africanus* (L.) Hoffmanns (Agapanthaceae). *An. Acad. Bras. Cienc.* **88**, 1727–1734 (2016).
5. Sharaibi, O. J. & Afolayan, A. J. Micromorphological characterization of the leaf and rhizome of *Agapanthus praecox* subsp. *praecox* Willd. (Amaryllidaceae). *J. Bot.* **2017**, 1–10 (2017).
6. Fenwick, G. R., Hanley, A. B. & Whitaker, J. R. The genus *Allium*—part 1. *Crit. Rev. Food Sci. Nutr.* **22**, 199–271 (1985).
7. Boulos, L. *Flora of Egypt* (Al Hadara Publishing, 1999).
8. Harris, J., Cottrell, S., Plummer, S. & Lloyd, D. Antimicrobial properties of *Allium sativum* (garlic). *Appl. Microbiol. Biotechnol.* **57**, 282–286 (2001).
9. Capasso, A. Antioxidant action and therapeutic efficacy of *Allium sativum* L. *Molecules* **18**, 690–700 (2013).
10. Borlinghaus, J., Albrecht, F., Gruhlke, M. C., Nwachukwu, I. D. & Slusarenko, A. J. Allicin: chemistry and biological properties. *Molecules* **19**, 12591–12618 (2014).
11. Fu, J. et al. Identification and characterization of abundant repetitive sequences in *Allium cepa*. *Sci. Rep.* **9**, 16756 (2019).
12. Khandagale, K. et al. Omics approaches in *Allium* research: progress and way ahead. *PeerJ* **8**, e9824 (2020).
13. King, J., Bradeen, J., Bark, O., McCallum, J. & Havey, M. J. A low-density genetic map of onion reveals a role for tandem duplication in the evolution of an extremely large diploid genome. *Theor. Appl. Genet.* **96**, 52–62 (1998).
14. Jakše, J. et al. Pilot sequencing of onion genomic DNA reveals fragments of transposable elements, low gene densities, and significant gene enrichment after methyl filtration. *Mol. Genet. Genomics* **280**, 287–292 (2008).
15. Shigyo, M., Khar, A. & Abdelrahman, M. (eds) *The Allium Genomes* pp. 99–112 (Springer, 2018).
16. Kiseleva, A., Kirov, I. & Khrustaleva, L. Chromosomal organization of centromeric Ty3/gypsy retrotransposons in *Allium cepa* L. and *Allium fistulosum* L. *Russ. J. Genet.* **50**, 586–592 (2014).
17. Hertweck, K. L. Assembly and comparative analysis of transposable elements from low coverage genomic sequence data in Asparagales. *Genome* **56**, 487–494 (2013).
18. Peška, V., Mandáková, T., Ihradská, V. & Fajkus, J. Comparative dissection of three giant genomes: *Allium cepa*, *Allium sativum*, and *Allium ursinum*. *Int. J. Mol. Sci.* **20**, 733 (2019).
19. Sun, X. et al. A chromosome-level genome assembly of garlic (*Allium sativum*) provides insights into genome evolution and allicin biosynthesis. *Mol. Plant* **13**, 1328–1339 (2020).
20. Ohri, D., Fritsch, R. M. & Hanelt, P. Evolution of genome size in *Allium* (Alliaceae). *Plant Syst. Evol.* **210**, 57–86 (1998).
21. Duchoslav, M., Šafářová, L. & Jandová, M. Role of adaptive and non-adaptive mechanisms forming complex patterns of genome size variation in six cytotypes of polyploid *Allium oleraceum* (Amaryllidaceae) on a continental scale. *Ann. Bot.* **111**, 419–431 (2013).
22. Khrustaleva, L., Kudryavtseva, N., Romanov, D., Ermolaev, A. & Kirov, I. Comparative tyramide-FISH mapping of the genes controlling flavor and bulb color in *Allium* species revealed an altered gene order. *Sci. Rep.* **9**, 12007 (2019).
23. Shigyo, M., Khar, A. & Abdelrahman, M. (eds) *The Allium Genomes* pp 197–214 (Springer, 2018).
24. Ricroch, A., Yockteng, R., Brown, S. C. & Nadot, S. Evolution of genome size across some cultivated *Allium* species. *Genome* **48**, 511–520 (2005).
25. Liao, N. et al. Chromosome-level genome assembly of bunching onion illuminates genome evolution and flavor formation in *Allium* crops. *Nat. Commun.* **13**, 6690 (2022).
26. Finkers, R. et al. Insights from the first genome assembly of onion (*Allium cepa*). *G3 (Bethesda)* **11**, jkab243 (2021).
27. Nystedt, B. et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579–584 (2013).
28. Neale, D. B. et al. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* **15**, R59 (2014).
29. Guan, R. et al. Draft genome of the living fossil *Ginkgo biloba*. *GigaScience* **5**, 49 (2016).
30. Li, G. et al. A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nat. Genet.* **53**, 574–584 (2021).
31. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).
32. Ming, R. et al. The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* **47**, 1435–1442 (2015).
33. Harkess, A. et al. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat. Commun.* **8**, 1279 (2017).
34. Zhang, Y. et al. Chromosome-scale assembly of the *Dendrobium chrysotoxum* genome enhances the understanding of orchid evolution. *Hortic. Res.* **8**, 183 (2021).
35. Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L. & Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207**, 437–453 (2015).
36. Zhang, G. Q. et al. The *Apostasia* genome and the evolution of orchids. *Nature* **549**, 379–383 (2017).
37. Wang, X. et al. Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. *Mol. Plant* **8**, 885–898 (2015).
38. Sensalari, C., Maere, S. & Lohaus, R. ksrates: positioning whole-genome duplications relative to speciation events in $K_S$ distributions. *Bioinformatics* **38**, 530–532 (2022).
39. Yamaguchi, Y. & Kumagai, H. Characteristics, biosynthesis, decomposition, metabolism and functions of the garlic odour precursor, S-allyl-L-cysteine sulfoxide. *Exp. Ther. Med.* **19**, 1528–1535 (2020).
40. Michelmore, R. W. & Meyers, B. C. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* **8**, 1113–1130 (1998).
41. Guo, L. et al. The opium poppy genome and morphinan production. *Science* **362**, 343–347 (2018).
42. Yuan, M. et al. Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* **592**, 105–109 (2021).
43. Lescot, M. et al. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res.* **30**, 325–327 (2002).
44. Chen, A. et al. Large field of view-spatially resolved transcriptomics at nanoscale resolution. Preprint at *bioRxiv* https://doi.org/10.1101/2021.01.17.427004 (2021).
45. Xia, K. et al. The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in *Arabidopsis* leaves. *Dev. Cell* **57**, 1299–1310 (2022).
46. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).
47. Khandagale, K. & Gawande, S. Genetics of bulb colour variation and flavonoids in onion. *J. Hortic. Sci. Biotech.* **94**, 522–532 (2019).
48. Goławska, S., Sprawka, I., Łukasik, I. & Goławski, A. Are naringenin and quercetin useful chemicals in pest-management strategies? *J. Pest Sci.* **87**, 173–180 (2014).

49. Kurepa, J., Shull, T. E. & Smalle, J. A. Quercetin feeding protects plants against oxidative stress (version 1; peer review: 1 approved, 1 approved with reservations). *F1000Res.* **5**, 2430 (2016).
50. Sossountzov, L. et al. Spatial and temporal expression of a maize lipid transfer protein gene. *Plant Cell* **3**, 923–933 (1991).
51. Suh, M. C. et al. Cuticular lipid composition, surface structure, and gene expression in *Arabidopsis* stem epidermis. *Plant Physiol.* **139**, 1649–1665 (2005).
52. DeBono, A. et al. *Arabidopsis* LTPG is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface. *Plant Cell* **21**, 1230–1238 (2009).
53. Yeats, T. H. & Rose, J. K. The biochemistry and biology of extracellular plant lipid-transfer proteins (LTPs). *Protein Sci.* **17**, 191–198 (2008).
54. Heath, O. Formative effects of environmental factors as exemplified in the development of the onion plant. *Nature* **155**, 623–626 (1945).
55. Mita, T. & Shibaoka, H. Changes in microtubules in onion leaf sheath cells during bulb development. *Plant Cell Physiol.* **24**, 109–117 (1983).
56. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
57. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
58. Zhang, C. et al. Transcriptome analysis of sucrose metabolism during bulb swelling and development in onion (*Allium cepa* L.). *Front. Plant Sci.* **7**, 1425 (2016).
59. Atif, M. J. et al. Mechanism of *Allium* crops bulb enlargement in response to photoperiod: a review. *Int. J. Mol. Sci.* **21**, 1325 (2020).
60. Shibaoka, H. Plant hormone-induced changes in the orientation of cortical microtubules: alterations in the cross-linking between microtubules and the plasma membrane. *Annu. Rev. Plant Biol.* **45**, 527–544 (1994).
61. Zhong, R., Burk, D. H., Morrison, W. H. & Ye, Z. H. A kinesin-like protein is essential for oriented deposition of cellulose microfibrils and cell wall strength. *Plant Cell* **14**, 3101–3117 (2002).
62. Ganguly, A., Zhu, C., Chen, W. & Dixit, R. FRA1 kinesin modulates the lateral stability of cortical microtubules through cellulose synthase–microtubule uncoupling proteins. *Plant Cell* **32**, 2508–2524 (2020).
63. Rao, G., Zeng, Y., He, C. & Zhang, J. Characterization and putative post-translational regulation of α-and β-tubulin gene families in *Salix arbutifolia*. *Sci. Rep.* **6**, 19258 (2016).
64. Goley, E. D. & Welch, M. D. The ARP2/3 complex: an actin nucleator comes of age. *Nat. Rev. Mol. Cell Biol.* **7**, 713–726 (2006).
65. Singh, R. et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* **500**, 335–339 (2013).
66. Paterson, A. H., Bowers, J. E. & Chapman, B. A. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl Acad. Sci. USA* **101**, 9903–9908 (2004).
67. Tuskan, G. A. et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604 (2006).
68. Vanneste, K., Baele, G., Maere, S. & Van de Peer, Y. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Res.* **24**, 1334–1347 (2014).
69. D'hont, A. et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* **488**, 213–217 (2012).
70. Putnik, P. et al. An overview of organosulfur compounds from *Allium* spp.: from processing and preservation to evaluation of their bioavailability, antimicrobial, and anti-inflammatory properties. *Food Chem.* **276**, 680–691 (2019).
71. Ellmore, G. S. & Feldberg, R. S. Alliin lyase localization in bundle sheaths of the garlic clove (*Allium sativum*). *Am. J. Bot.* **81**, 89–94 (1994).
72. Stotz, H. U. et al. Role of camalexin, indole glucosinolates, and side chain modification of glucosinolate-derived isothiocyanates in defense of *Arabidopsis* against *Sclerotinia sclerotiorum*. *Plant J.* **67**, 81–93 (2011).
73. Hématy, K. et al. Moonlighting function of phytochelatin synthase1 in extracellular defense against fungal pathogens. *Plant Physiol.* **182**, 1920–1932 (2020).
74. Matern, A. et al. A substrate of the ABC transporter PEN3 stimulates bacterial flagellin (flg22)-induced callose deposition in *Arabidopsis thaliana*. *J. Biol. Chem.* **294**, 6857–6870 (2019).
75. Birnbaum, K. et al. A gene expression map of the *Arabidopsis* root. *Science* **302**, 1956–1960 (2003).
76. Jean-Baptiste, K. et al. Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* **31**, 993–1011 (2019).
77. Shulse, C. N. et al. High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* **27**, 2241–2247 (2019).
78. Olsen, J. L. et al. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* **530**, 331–335 (2016).
79. Shusei, S. et al. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641 (2012).
80. Project, A. G. et al. The *Amborella* genome and the evolution of flowering plants. *Science* **342**, 1241089 (2013).

# Methods

## Plant materials and DNA preparation

Four species (onion, garlic, Welsh onion and African lily) were collected for genome sequencing and assembly. For onion, we used a doubled haploid line obtained from the Department of Horticulture, University of Wisconsin, which was originally developed and selected based on gynogenesis discussed in ref. 81 from the Department of Plant Breeding and Genetics, Cornell University. For Welsh onion, an inbred line obtained from the Liaoning Academy of Agricultural Sciences was used. For garlic and African lily, cultivars were obtained from a farmers' market. Total DNA was extracted from fresh leaves using the standard cetyltrimethylammonium bromide method[82], and the DNA quality and quantity were measured using a NanoDrop spectrophotometer and Qubit fluorometer, respectively.

## Genome sequencing and library construction

For the four genomes, genomic DNA was sequenced using the Sequel II system (PacBio) in HiFi mode. SMRTbell library construction and sequencing were performed at Berry Genomics. We prepared ~20 kb SMRTbell libraries according to the official protocols of PacBio. After sequencing, raw data were analyzed with SMRT Link v7.1 or v8.0 using the circular consensus sequencing (CCS) protocol with a cutoff minimum of three passes and estimated accuracy of 0.99. High-quality reads were used for genome assembly.

To avoid potential plastid or mitochondrial genome contamination, we aligned HiFi reads to chloroplast and mitochondrial genomes with minimap2 (ref. 83, because neither Welsh onion nor garlic has a published mitochondrial genome, and African lily lacks one as well, we used mitochondrial and chloroplast genomes of onion for alignment). For each read, if the ratio between genome length covered by the aligned region and read length was greater than 98%, the reads were derived from the plastid genome and were filtered out before genome assembly.

For *k*-mer estimation of genome size, DNA libraries with insert sizes of 300–500 bp were sequenced on the MGISEQ-2000 platform (BGI) for onion, garlic, Welsh onion and African lily. In addition, we also carried out quality control with FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) and removed poor-quality reads (reads with *n* bases and QVs lower than 20) and over-represented sequences.

## Hi-C library construction and sequencing

Hi-C libraries of the four species were prepared following previously published methods[84]. Briefly, fresh leaves were fixed and cross-linked in formaldehyde solution (1%). After cell lysis, cross-linked chromatin was digested using the restriction enzyme DpnII. Subsequently, the sticky ends generated by DpnII were filled in with biotinylated nucleotides, and the intramolecular blunt ends were then ligated to allow the formation of chimeric junctions. DNA was physically sheared into ~350 bp fragments, and the biotin-labeled fragments were purified with streptavidin beads and used to construct the sequencing libraries. The libraries were sequenced on the MGISEQ-2000 platform. As above, the BGI sequencing data were also checked for microbial contamination, with no contamination found.

## Genome assembly and chromosome construction

PacBio HiFi reads were used to perform de novo genome assembly for onion, garlic, Welsh onion and African lily with Hifiasm (v0.9-r289 with default parameters)[85]. For onion, garlic and Welsh onion, one set of the two haploid contigs with longer length from the Hifiasm output was used for further assembly. For African lily, due to differences between the two haploid sets of contigs and estimated genome size, a heterozygous genome was used for subsequent assembly. Finally, Hi-C sequencing data were used to anchor the assembled contigs into pseudochromosome molecules. The Hi-C data were first mapped to the polished genome assembly with Juicer (v1.6)[86]. Next, chromosome-level assembly was generated using 3D-DNA (v180922) with default parameters[87]. To check the quality and further improve chromosome-scale assembly, manual review and refinement of the candidate assembly was performed by Juicebox (v1.6) Assembly Tools[86]. For African lily, the longer of the two haploid chromosomes was used as the haploid genome.

## Assembly quality evaluation

We use three approaches to evaluate the assembly quality of the four genomes. First, we performed BUSCO assessment of the genome sequences and then predicted gene models using BUSCO5 (ref. 31) with the embryophyta_odb10 or liliopsida_odb10 data set. Second, we used Merqury[88] to assess genome completeness and consensus QV. Third, to explore telomere sequences on the chromosomes of the four species, we extracted 20 Mb at both ends of each chromosome. We then built a nucleotide database for each extracted 20 Mb sequence using the command 'formatdb -p F -i extracted.fa' and aligned the self-constructed telomere sequences to each database by 'blastall (v2.2.25) -p blastn -a 10 -e 10 -m 8 -F F -d extracted.fa -i telomere.fa -o result.m8,' respectively. The telomere sequences used as the query to search the database were constructed by concatenating 100,000 copies of the current known telomeres (t1, CTCGGTTATGGG; t2, TTAGGG)[89]. The alignment results were filtered to retain only hits with identities ≥90%, alignment length ≥100 bp and score≥100. We also checked potential microbial contamination of the assemblies to be aligned to the NCBI nucleotide database by BLASTN and found no microbial contamination.

## Prediction and functional annotation of protein-coding genes

The protein-coding genes of onion, garlic, Welsh onion and African lily were identified using an annotation strategy that integrated de novo prediction and homology-based and transcript-based evidence. Augustus (v3.3.2)[90], GlimmerHMM (v3.0.4)[91] and SNAP (v2.0)[92] were used to de novo identify gene structures from the genome sequences. For homology-based gene prediction, protein sequences from ten published genomes, that is, *A. thaliana*, *Oryza sativa*, *A. officinalis*, *A. shenzhenica*, *Aegilops tauschii*, *Brachypodium distachyon*, *Ensete ventricosum*, *Phalaenopsis equestris*, *Phoenix dactylifera* and *Xerophyta viscosa*, were used to search against the genomes using TBLASTN[93] with an *E* value cutoff of $1 \times 10^{-5}$ and a minimum query coverage of 25%. GeneWise (v2.4.1)[94] was used to identify protein-coding regions based on sequence alignments. For the transcript-based method, the Trinity package (v2.10.0)[95] was used to conduct de novo transcriptome assembly of RNA-seq data, and the PASA pipeline (v2.4.1)[96] was then used to predict gene models based on transcript alignments. Finally, evidence from the above approaches was integrated to generate the consensus gene models using EvidenceModeler (v1.1.1)[97]. High-confidence gene models were further filtered to remove short gene models (less than 50 amino acids) and retain only gene models supported by at least three homologous genes ($E$ value $\leq 1 \times 10^{-5}$, coverage ≥80%) or transcript evidence. The final high-confidence gene sets were functionally annotated against the SwissProt, GO, KEGG and InterPro databases[98].

## Phylogenetic tree reconstruction

Phylogenetic relationships among the selected species were inferred by two independent methods, that is, concatenation and coalescent approaches. For concatenation, 346 single-copy orthologous genes from ten species (including four Amaryllidaceae species, four monocots outside Amaryllidaceae, one eudicot and one basal angiosperm *Amborella*) were identified by OrthoMCL[99] and extracted to conduct protein multiple sequence alignments (MSAs) of each single-copy gene family using the MUSCLE program[100]. The alignments were converted into corresponding CDS MSAs, which were then concatenated to generate a supergene alignment. A maximum-likelihood species tree was reconstructed using IQ-TREE (v2.0.3, http://www.iqtree.org)

with the concatenated single-copy gene families. For the coalescent method, reciprocal BLAST was applied to identify putative orthologs, and a set of 1,456 gene families from reciprocal best hits (RBH) among the selected ten species were considered as orthologous genes. A maximum-likelihood gene tree was inferred for each of the putative orthologous groups using IQ-TREE (v2.0.3, http://www.iqtree.org). Astral (v5.6.1)[101] was used to infer species trees from the gene trees. Estimation of divergence time was carried out using the MCMCTree module incorporated in the PAML package (v4.9e)[102]. Divergence times in the TimeTree database (http://www.timetree.org) and fossil records were applied as time calibration points. The common ancestral node of *Amborella trichopoda* and *Solanum lycopersicum* was constrained to 173–199 Mya, the common ancestral node of *S. lycopersicum* and *Zostera marina* was constrained to 139–156 Mya[103] and the common ancestral node of *A. officinalis* and African lily was constrained to 55–71 Mya.

### Whole-genome synteny

MCScanX[104] was applied to recognize syntenic blocks using default parameters. Syntenic blocks were visualized with MCScan (https://github.com/tanghaibao/jcvi/wiki/MCscan). Chromosome-scale syntenic block plots and dot plots were constructed using the Python version of MCScan. We also performed intergenomic comparison with three sister monocots (*A. shenzhenica*, *Cymbidium goeringii* and *E. guineensis*) using MCScan with default parameters to further characterize polyploidization events.

### Assessment of WGD

Syntenic blocks were identified based on detected homologous gene pairs, and synonymous substitutions per synonymous site (*Ks*) between collinear genes were estimated using the Nei–Gojobori approach[105] with the WGDI pipeline (v0.4.7)[106]. We used average *Ks* values to represent each syntenic block. Multipeak fitting of the curve was performed using the Gaussian approximation function (cftool).

To detect relics of possible polyploidization[107] in the four genomes, we first identified the syntenic blocks of homologous genes within each genome by performing intragenomic comparisons of the four species. Next, we calculated *Ks* between homologous genes within syntenic blocks and drew the density distribution of *Ks*. Each *Ks* distribution peak represents one polyploidization event and the time of the event was calibrated using *Ks* peaks between species and their corresponding divergence times.

We obtained the collinear genes of each block using *A. trichopoda* as the reference genome and constructed the collinear gene tree using IQ-TREE2 (ref. 108). To estimate species trees based on gene trees, we used a coalescent-based approach with Astral (v5.6.1)[101]. We analyzed discordance in the gene tree of syntenic blocks to test the two hypotheses of whether *Allium* and African lily share the $\rho$ WGD event using DiscoVista[109].

Using the 'ksd' module in wgd[110], we calculated *Ks* for the orthologous gene pairs between species constructed according to the RBH protocol. The ksrates[38] program was used to perform ortholog *Ks* adjustment according to branch length differences from the common ancestral nodes to terminal nodes representing current species. A phylogenetic tree of six species was used (*A. thaliana*, (*A. shenzhenica*, (African lily, (garlic, (onion, Welsh onion)))), with African lily as the focal species for *Ks* adjustment.

### Repeat annotation and LTR retrotransposon analysis

Repetitive sequences and transposable elements (TEs) in the genome were identified using homology-based and de novo-based annotation. The homology-based repeat library was generated from a known repeat library using RepeatMasker (v.4.1.0)[111] and ProteinMask (v4.0.5, a package within RepeatMasker)[111]. Next, we used RepeatModeler (v1.0.5)[112], RepeatScout[113], Piler (v0.1-1)[114] and LTR_FINDER (v1.07)[115] to construct a de novo repetitive-element library. We used RepeatMasker (v4.1.0)[111] to identify repeats using both the de novo repeat database and Repbase. We then combined the de novo- and homology-based predictions of TEs according to the coordinates in the genome. Tandem repeats were predicted in each genome by RepeatMasker (v4.1.0)[111] and Tandem Repeats Finder (v4.07)[116]. Muscle was used to align repeat sequences, and the evolutionary distance between the repeat consensus and copies was calculated using the Kimura two-parameter model with the dnadist program implemented in PHYLIP (v3.695)[117].

Full-length LTR retrotransposons were identified in *O. sativa*, *A. officinalis* and the four genomes using LTR_FINDER (v1.07)[115] and LTRharvest[118] with default parameters. The full-length LTRs were then translated into amino acids in six frames and mapped against the Ty1_Copia (PF07727) and Ty3_Gypsy (PF000078) domains in the Pfam database using HMMER (v3.3.1)[119] with an $E$ value $\leq 1 \times 10^{-9}$. The sequences mapped to the Copia and Gypsy superfamilies were aligned using MAFFT (v7.407)[120] with default parameters. The phylogenetic trees of the two superfamilies were constructed using FastTree (v2.1)[121].

### Spatial RNA-seq

We collected onion bulbs at three bulbing stages: that is, early, midterm and mature bulb. A 1 cm × 1 cm window was cut near the base of each longitudinal onion bulb block containing more than five layers of leaf sheaths connected by the bulb base. After cutting, tissue blocks were embedded in cold Tissue-Tek OCT (Sakura, 4583) on dry ice before cryosectioning. Using a Leika CM1950 cryostat, bulb tissues were further cut at a thickness of ~10 µm. We ligated 25 nt random coordinate identity (CID) barcodes, 10 nt molecular identifiers (MIDs) and 22 nt polyT to DNB arrays for 1 cm × 1 cm chip preparation. These oligos can identify DNB position, count expression and capture mRNA. We then mounted the bulb tissue slice onto the chip for imaging and RNA release. RNA released from the slices was locally captured and reverse-transcribed into cDNA and amplified by PCR on DNBs of the chip. The amplified products were collected and used as templates for the construction of sequencing libraries using an MGI DNBSEQ-Tx sequencer. For further details on chip and library preparation, please see ref. 44.

After sequencing, reads with MIDs containing either N bases or >2 bases with quality scores <10 were filtered out. The filtered reads were mapped to the reference genome using Spliced Transcripts Alignment to a Reference (STAR) software (v2.7.9a)[122]. Mapped reads with mapping quality (MAPQ) >10 were retained for expression quantification. Reads with the same CID mapped to the same gene locus were collapsed, allowing one mismatch. The number of MIDs in reads collapsed under the same CID of the same gene was counted to represent the expression level of the gene at the site coded by CID. The CID sequences in the reads were mapped to the chip CID code matrix, allowing one base mismatch to identify the location of each CID on the chip. Average cell diameter of the onion bulbs was ~50 µm, which can be covered by 80 × 80 DNBs on the chip. Thus, the expression profile matrix was divided into nonoverlapping bins covering an area of 80 × 80 DNBs to represent each cell.

Finally, for each developmental stage, we obtained a spatial expression matrix of two biological replicates, named S1, S2 (early bulb), M1, M2 (midterm bulb) and L1, L2 (mature bulb).

### The analysis of spatial transcriptome data

With the spatial transcriptome data, bins with gene numbers below 50 were filtered out as low-quality bins. The gene expression levels of each bin were normalized using the SCTransform function in Seurat v4.0.4 (ref. 123). For the integration of spatial RNA-seq data, we used Harmony[124] to correct principal component analysis (PCA) embeddings. To reduce data dimensions, we performed PCA on highly variable gene matrices using the Seurat RunPCA function. Based on the top 30 PCA dimensions, the bins were clustered into 12 clusters by FindNeighbors and FindClusters (resolution = 0.3). Each cluster was assigned to a

tissue-type group based on their location in the tissue image. To identify DEGs among clusters, we used the FindAllMarkers function with default parameters. The expression matrix was reduced to two dimensions for visualization using RunUMAP (reduction = 'pca,' dims = 1:30).

We selected spongy mesophyll cell bins based on cell clusters and cell morphology from the three different bulbing stages and reran UMAP for dimensionality deduction and development trajectory visualization because of its ability to learn and maintain local and global distances in low-dimensional space. We then reconstructed lineage trajectories with the monocle3 (refs. 56,57), order_cells and learn_graph functions and chose two points located in the inner layer of S2 as the starting points of pseudotime with coordinates (218, 50) and (200, 50). Finally, we obtained the DEGs in pseudotime among different bulbing stages using the graph_test monocle3 function.

### Data statistical analysis

GO, KO and KEGG pathway enrichment analysis showed that the proportion of the number of genes annotated to certain GO terms, KO terms or KEGG pathway in all target genes was substantially greater than that in all background genes[125–127]. Enrichment significance (*P* value) was calculated using the hypergeometric test (one side), which can be calculated as:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

where *N* is the number of genes with GO terms, KO terms or KEGG pathway annotation among all genes (background genes); *n* is the number of target genes in *N*; *M* is the number of genes annotated for a specific GO term, KO term or KEGG pathway among all genes (background genes); *m* is the number of target genes in a particular pathway.

The calculated *P* value will be further corrected by multiple checks and get corrected *P* value (Benjamini–Hochberg method). Corrected *P* value ≤ 0.05 was taken as the threshold, and the GO terms, KO terms or KEGG pathway satisfying the condition was defined as the GO terms, KO terms or KEGG pathway substantially enriched in the target genes.

GraphPad Prism (v 8.0.2) was used for plotting and statistical analysis. Asterisks indicate statistically significant differences from control. Statistical analyses were performed using Student's *t* test. The values are means ± s.e.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The raw sequencing data of onion, garlic, Welsh onion and Africa lily were deposited in the National Center for Biotechnology Information Sequence Read Archive under the accession PRJNA948806 and in the National Genomics Data Center (https://ngdc.cncb.ac.cn/?lang=en) under the accession PRJCA016760. The assemblies of the four genomes reported in this paper have been deposited in the National Center for Biotechnology Information under accessions JASDDO000000000, JASFAV000000000, JASFAW000000000, JASFAX000000000 and in the Genome Warehouse in National Genomics Data Center[128,129], Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under accession GWHCBHY00000000, GWHCBHZ00000000, GWHCBIA00000000 and GWHCBIB00000000 that are publicly accessible at https://ngdc.cncb.ac.cn/gwh.

## Code availability

All software used in the study are publicly available on the Internet as described in the Methods and Reporting Summary.

## References


81. Hyde, P. T., Earle, E. D. & Mutschler, M. A. Doubled haploid onion (*Allium cepa* L.) lines and their impact on hybrid performance. *HortScience* **47**, 1690–1695 (2012).
82. Healey, A., Furtado, A., Cooper, T. & Henry, R. J. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **10**, 21 (2014).
83. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
84. Xie, T. et al. *De novo* plant genome assembly based on chromatin interactions: a case study of *Arabidopsis thaliana*. *Mol. Plant* **8**, 489–492 (2015).
85. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm. *Nat. Methods* **18**, 170–175 (2021).
86. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
87. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
88. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 1–27 (2020).
89. Fajkus, P. et al. *Allium* telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG) n is synthesized by telomerase. *Plant J.* **85**, 337–347 (2016).
90. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
91. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
92. Tichenor, C. A new software metric to complement function points: the software non-functional assessment process (SNAP). https://apps.dtic.mil/sti/pdfs/ADA592012.pdf (2013).
93. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
94. Birney, E. & Durbin, R. Using GeneWise in the *Drosophila* annotation experiment. *Genome Res.* **10**, 547–548 (2000).
95. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
96. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
97. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
98. Hunter, S. et al. InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37**, D211–D215 (2009).
99. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
100. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
101. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 15–30 (2018).
102. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
103. Bell, C. D., Soltis, D. E. & Soltis, P. S. The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* **97**, 1296–1303 (2010).

104. Xu, Y. et al. Corrigendum #2 to 'VGSC: a web-based vector graph toolkit of genome synteny and collinearity'. *BioMed Res. Int.* **2019**, 2150291 (2019).
105. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).
106. Sun, P. et al. WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol. Plant* **15**, 1841–1851 (2022).
107. Soltis, P. S. & Soltis, D. E. Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* **30**, 159–165 (2016).
108. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
109. Sayyari, E., Whitfield, J. B. & Mirarab, S. DiscoVista: interpretable visualizations of gene tree discordance. *Mol. Phylogenet. Evol.* **122**, 110–115 (2018).
110. Zwaenepoel, A. & Van de Peer, Y. wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* **35**, 2153–2155 (2019).
111. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **25**, 1–14 (2009).
112. Sato, K., Tanaka, T., Shigenobu, S., Motoi, Y. & Itoh, T. Improvement of barley genome annotations by deciphering the Haruna Nijo genome. *DNA Res.* **23**, 21–28 (2016).
113. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
114. Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic repeats. *Bioinformatics* **21**, i152–i158 (2005).
115. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
116. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
117. Wang, H. et al. Identification of antibiotic resistance genes in the multidrug-resistant *Acinetobacter baumannii* strain, MDR-SHH02, using whole-genome sequencing. *Int. J. Mol. Med.* **39**, 364–372 (2017).
118. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).
119. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121 (2013).
120. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
121. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
122. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
123. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
124. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
125. Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J. & Pascual-Montano, A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol.* **8**, R3 (2007).
126. Huang, T. et al. Deciphering the effects of gene deletion on yeast longevity using network and machine learning approaches. *Biochimie* **94**, 1017–1025 (2012).
127. Chen, L., Li, B.-Q. & Feng, K.-Y. Predicting biological functions of protein complexes using graphic and functional features. *Curr. Bioinform.* **8**, 545–551 (2013).
128. Chen, M. et al. Genome warehouse: a public repository housing genome-scale data. *Genom. Proteom. Bioinform.* **19**, 584–589 (2021).
129. CNCB-NGDC Members and PartnersDatabase resources of the National Genomics Data Center, China National Center for Bioinformation in 2023. *Nucleic Acids Res.* **51**, D18–D28 (2023).

## Acknowledgements

We thank M. J. Havey from the Department of Horticulture, University of Wisconsin (Madison, WI, USA) for providing the onion seeds of the doubled haploid line and L. Cui from the Liaoning Academy of Agricultural Sciences (Shenyang, Liaoning, China) for providing the inbred lines of Welsh onion. We thank Y. Gu, L. Chen, Q. Lin, L. Chen, B. Mu, H. Sun, X. Wei, J. Li, S. Li, H. Lu and S. Zhang for general technical assistance or discussion. We thank G. Zhang from Zhejiang University and K. Wang from Northwestern Polytechnical University (NWPU) for their comments and suggestions when preparing our manuscript. We thank Y. Zeng from BGI-Shenzhen for sharing his idea on *Allium* genome study. This work was supported by the Thousand Talents Plan (5113190037), the Talents Team Construction Fund of NWPU and the Fundamental Research Funds for the Central Universities (3102019JC007) to J.C.; the Talents Team Construction Fund of NWPU and the Projects of Interdisciplinary of NWPU (0202022GH0306) to W.W.; National Natural Science Foundation of China (21801206), the Joint Research Funds of Department of Science & Technology of Shaanxi Province and NWPU (2020GXLH-Z-015) to Z.R.; Guangdong Provincial Key Laboratory of Genome Read and Write (2017B030301011) to X.X.

## Author contributions

F.H. and J.C. managed the project. J.C., W.W., R.M., H.Y. and X.X. conceived the study. F.H. and J.C. wrote the manuscript with contributions from all other authors. B.Z. and H.Z. were responsible for genome assembly into contigs and Hi-C scaffolding. X.L. handled genome annotation and analysis of phylogeny, transcriptome, gene family evolution, allicin/isoallicin pathway and immune system. B.Z., Z.T., X.L. and Z.L. were responsible for the analysis of whole-genome synteny and whole-genome duplication (WGD). Z.T. and P.Z. handled repeat annotation and conducted analysis of LTR retrotransposons. L.Z., J.H., J.Q., Q.L., Y.Z. and K.W. were responsible for in situ hybridization experiments. L.Z. focused on studying the evolution and formation of tunicated bulbs. H.Z. was responsible for genome evaluation and conducted ultra-high-performance liquid chromatography/tandem mass spectrometry (UHPLC-MS/MS) analysis. H.Z., K.X., X.G., L.L., W.S., B.Z., S.L. and L.P. conducted spatial RNA sequencing and analyzed spatial transcriptome data. Y.P., W.Z., F.L., Z.R. and J.M. were in charge of plant material collection and DNA/RNA preparation. H.Y., L.H. and W.C. supervised genome sequencing and library construction and also collected plant material for bulb development.

## Competing interests

The authors declare no competing interests.

## Additional information

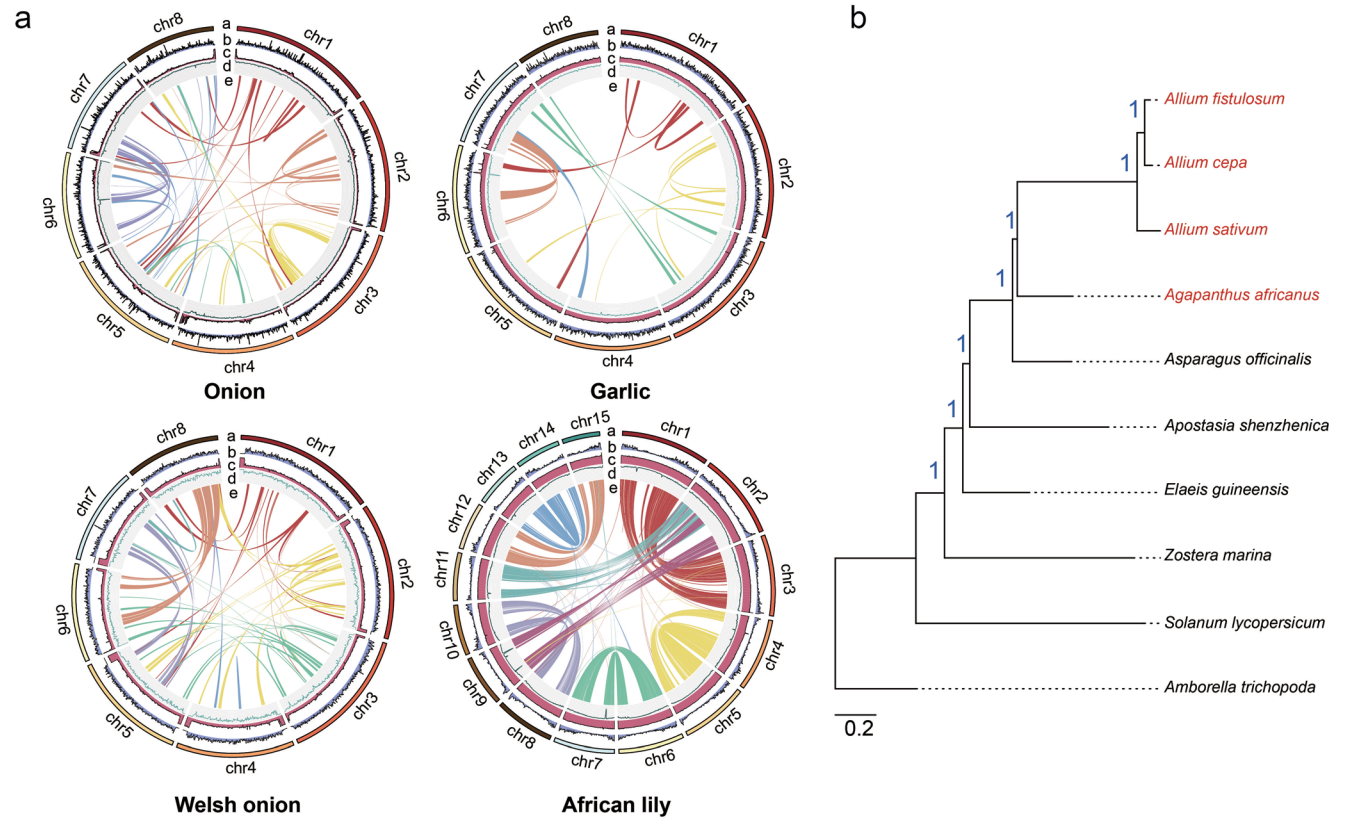
**Extended data** is available for this paper at https://doi.org/10.1038/s41588-023-01546-0.

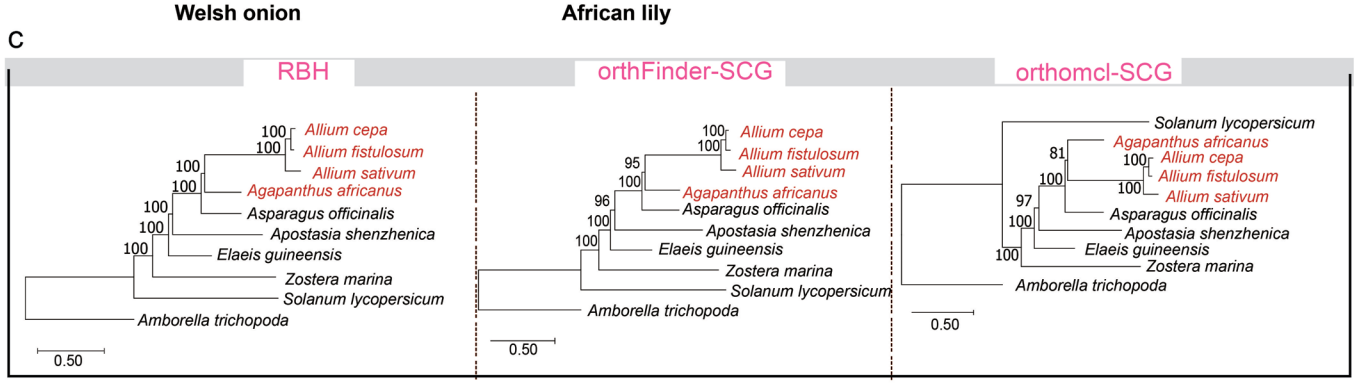
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41588-023-01546-0.

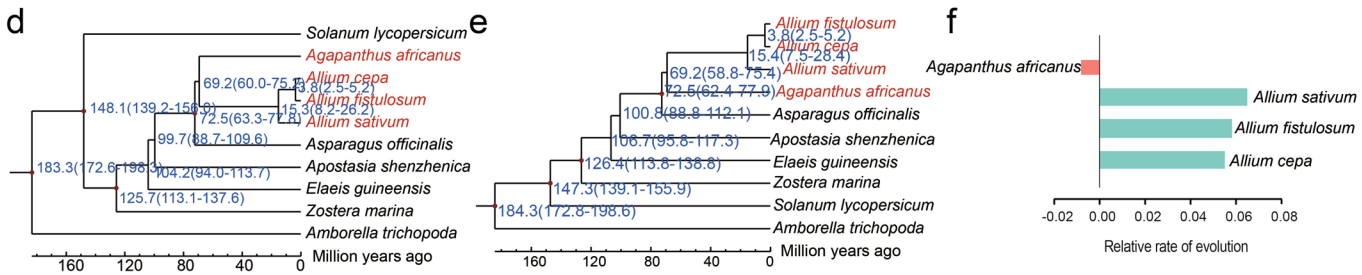
**Correspondence and requests for materials** should be addressed to Xun Xu, Hui Yang, Richard C. Macknight, Wen Wang or Jing Cai.

**Peer review information** *Nature Genetics* thanks Zhangjun Fei and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41588-023-01546-0

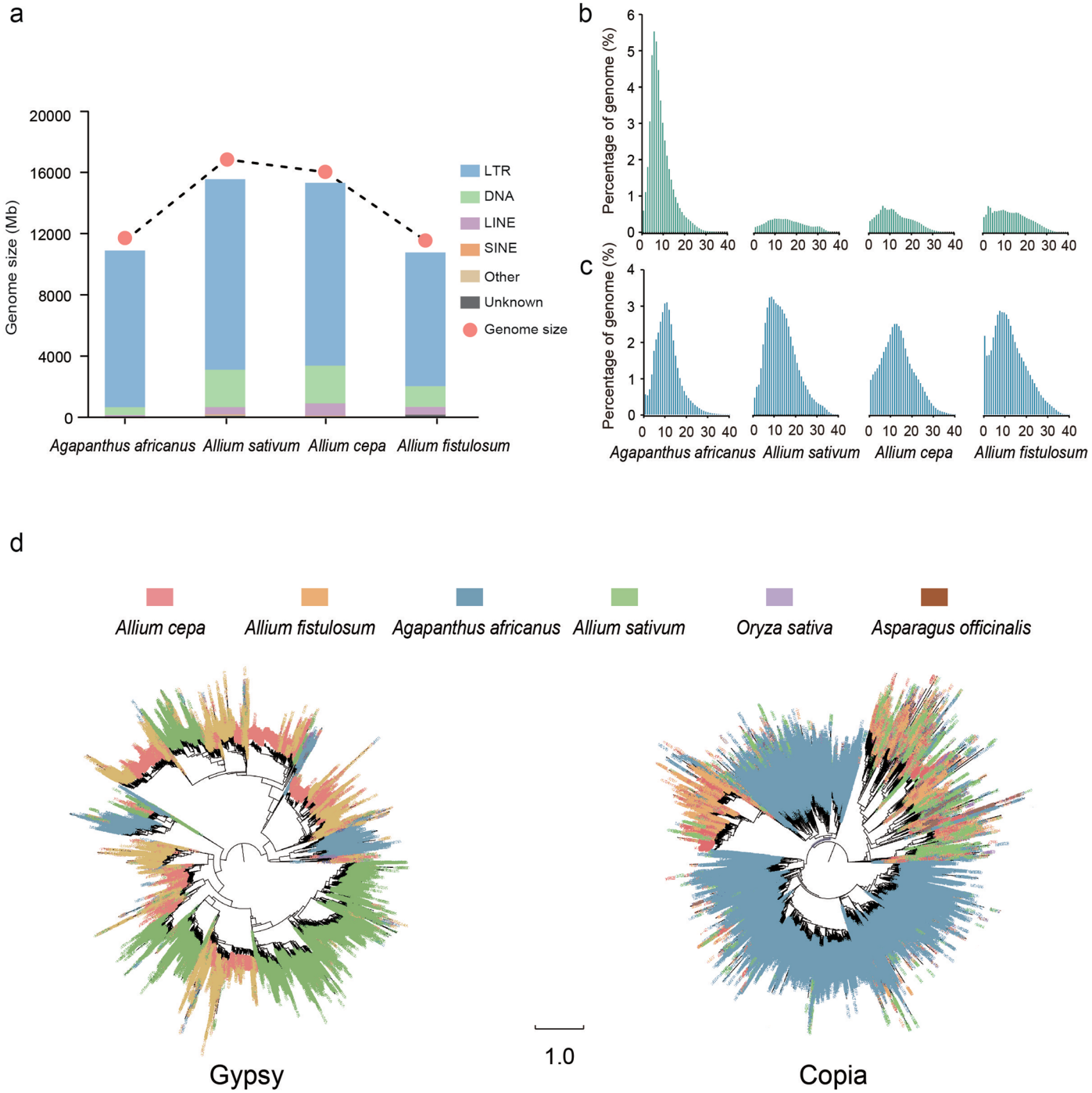


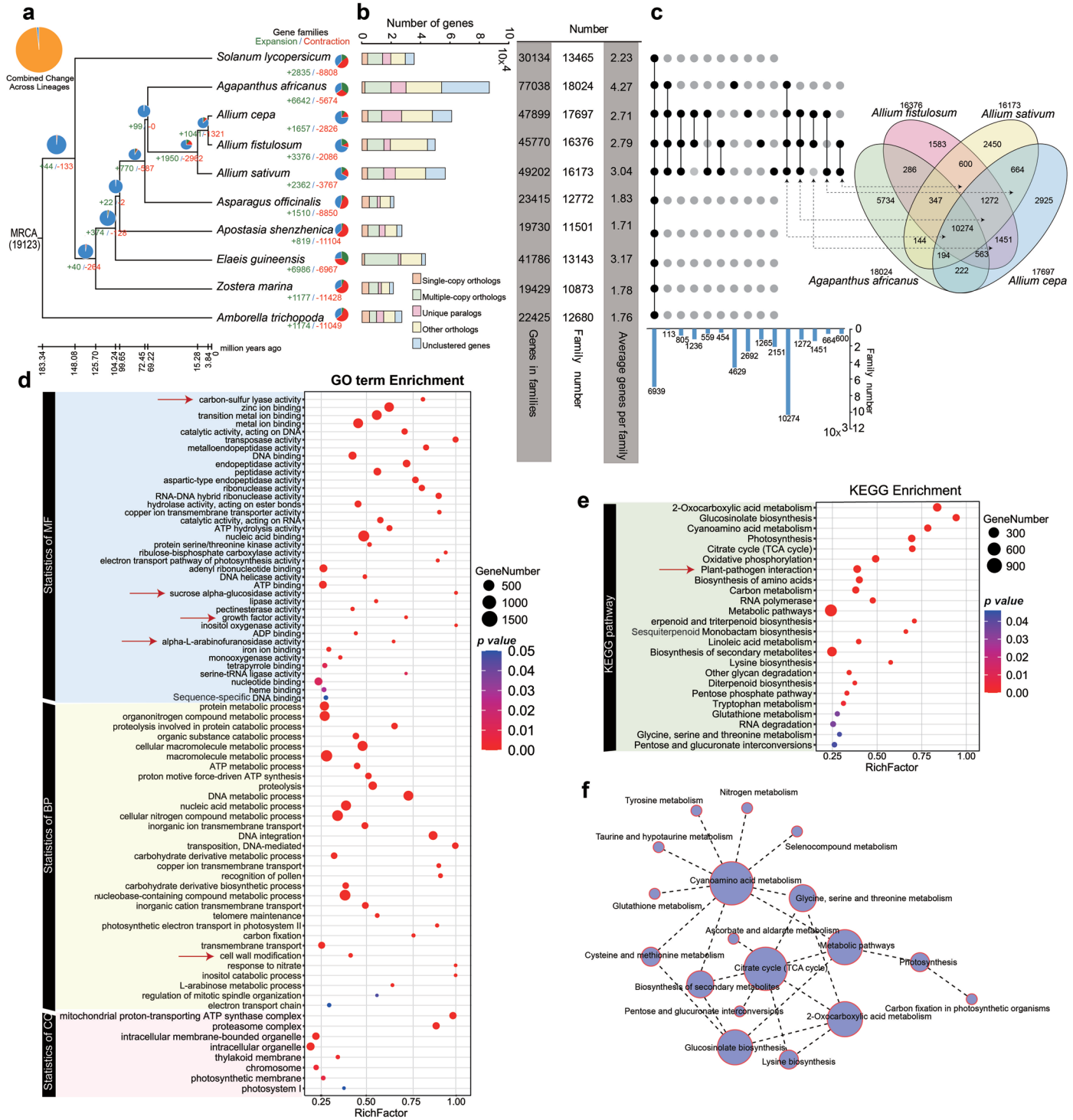

**Extended Data Fig. 1 | Genome overview and phylogenetic analyses of *Allium* species and African lily, related to Fig. 1. a**, Overview of genomes of *Allium* species and African lily: track a corresponds to chromosome length. From b to d, three rings represent density distribution of different genomic features, including coding gene, GC, and repeat sequences, respectively. Track e corresponds to syntenic blocks. **b**, Species tree by summarizing each gene tree with Astral. **c**, Phylogenies from three supergene matrice, each of which was constructed by concatenating ortholog gene sets defined by reciprocal best hit (RBH) by BLAST, single-copy genes (SCG) identified using OrthoFinder and single-copy genes (SCG) identified using OrthoMCL, respectively. **d**, Single-copy genes identified using OrthoMCL. **e**, Single-copy genes identified using OrthoFinder. **f**, Relative evolutionary rate comparison among three *Allium* species and African lily.

https://doi.org/10.1038/s41588-023-01546-0



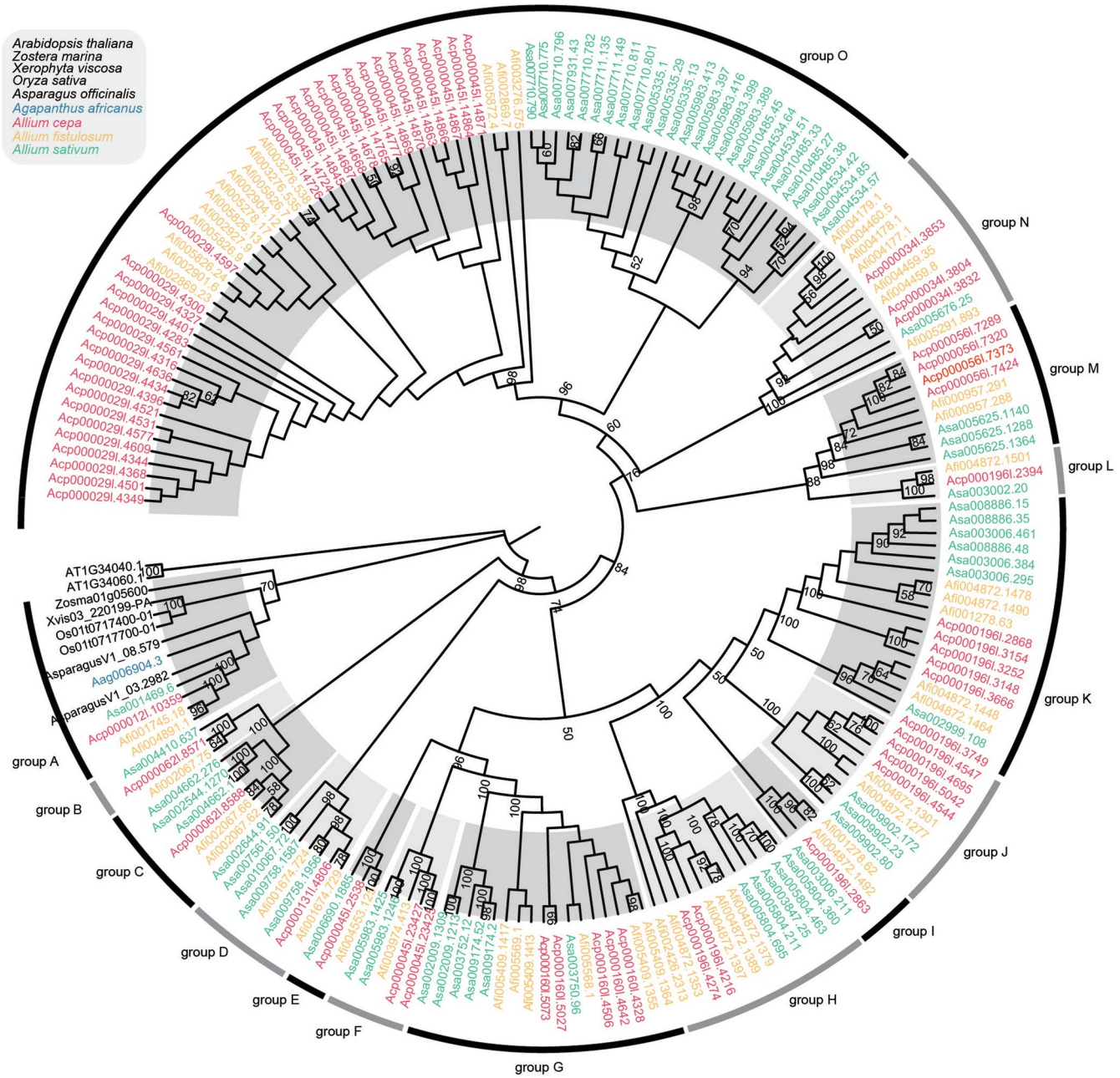

**Extended Data Fig. 2 | Analysis of repetitive sequences in three *Allium* species and African lily. a**, Overall composition of repetitive elements in different genomes. DNA, DNA element; LINE, long interspersed nuclear element; LTR, long terminal repeat transposable element; SINE, short interspersed nuclear element. Divergence distribution of Copia (**b**) and Gypsy (**c**) retrotransposons in four assembled genomes. X-axis represents divergence measured in percentage of sequence differences with consensus in TE library. **d**, Phylogenetic relationships of Gypsy and Copia retrotransposon domains across genomes of asparagus, rice, and four species sequenced in this study (Supplementary Data 3 for alignments of sequences and a version with support values on branches).

https://doi.org/10.1038/s41588-023-01546-0



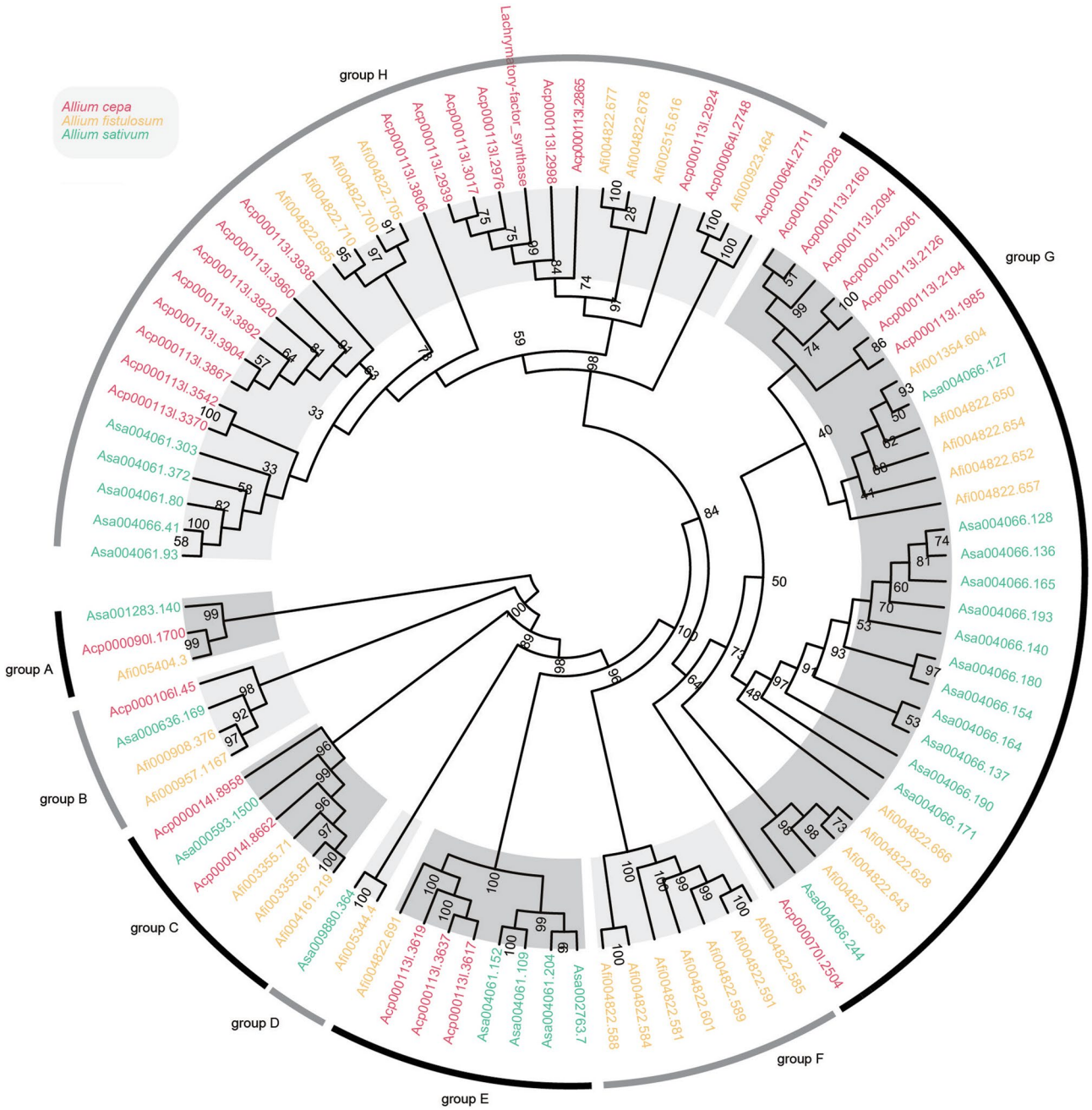

**Extended Data Fig. 3 | Evolution of gene families, related to Fig. 3. a**, Pie diagram on each branch of tree represents the proportion of gene families undergoing gain (green) or loss (red) events. Numbers below pie diagram denote total number of expanded and contracted gene families. **b**, Distribution of single-copy, multiple-copy, unique, and other orthologs in 10 plant species, and three columns on right show number of genes in families, family number, and average genes per family in 10 plant species. **c**, Point-line and Venn diagrams represent shared and unique gene families among 10 species or four closely related Amaryllidaceae species (Welsh onion, garlic, onion, and African lily). Each number represents number of gene families. **d**, Significant GO terms of molecular function, biological process, and cellular component enriched in expanded gene families in ancestral branch of three *Allium* plants. **e**, Significant KEGG pathways enriched in expanded gene families in ancestral branch of three *Allium* plants. **f**, Network diagram of significant KEGG pathways enriched in expanded gene families in ancestral branch of three *Allium* plants. GO terms or KEGG pathways probably related to alliinase, plant immune system, and bulb development discussed in this study are labeled with red arrows. The *p* values (unadjusted, one side) presented in **d** and **e** were calculated using the hypergeometric test.

https://doi.org/10.1038/s41588-023-01546-0



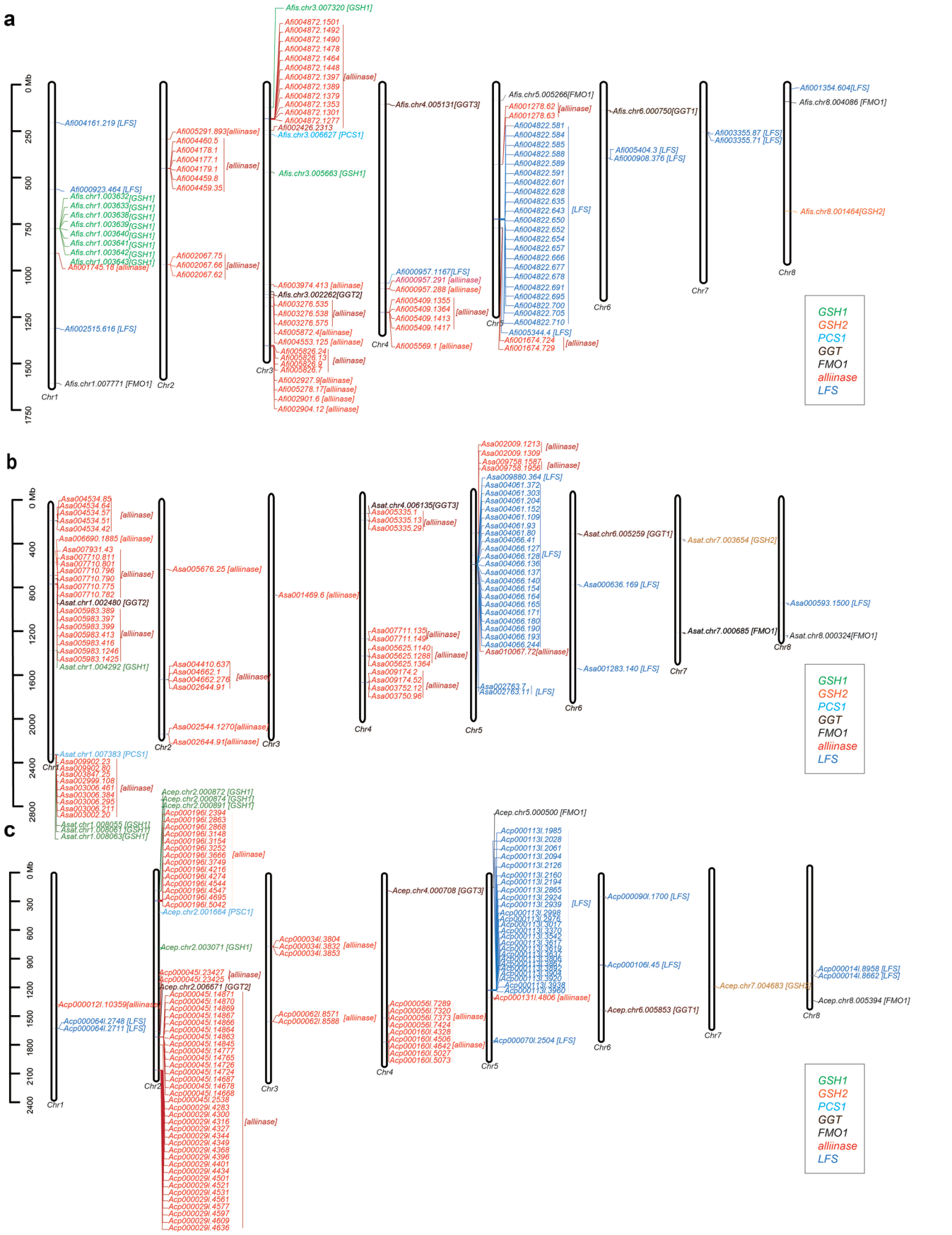

**Extended Data Fig. 4 | Clustering of alliinase gene family in *Allium* and outgroup species, related to Fig. 3.** Phylogenetic analysis of alliinase gene family, with alliinase genes forming 15 distinct groups (A–O). Bootstrap values are shown on each branch. For the alliinase gene family, we constructed a gene tree using standard maximum-likelihood phylogenetic analysis implemented in IQ-TREE with default parameters and 1,000 bootstraps.

https://doi.org/10.1038/s41588-023-01546-0



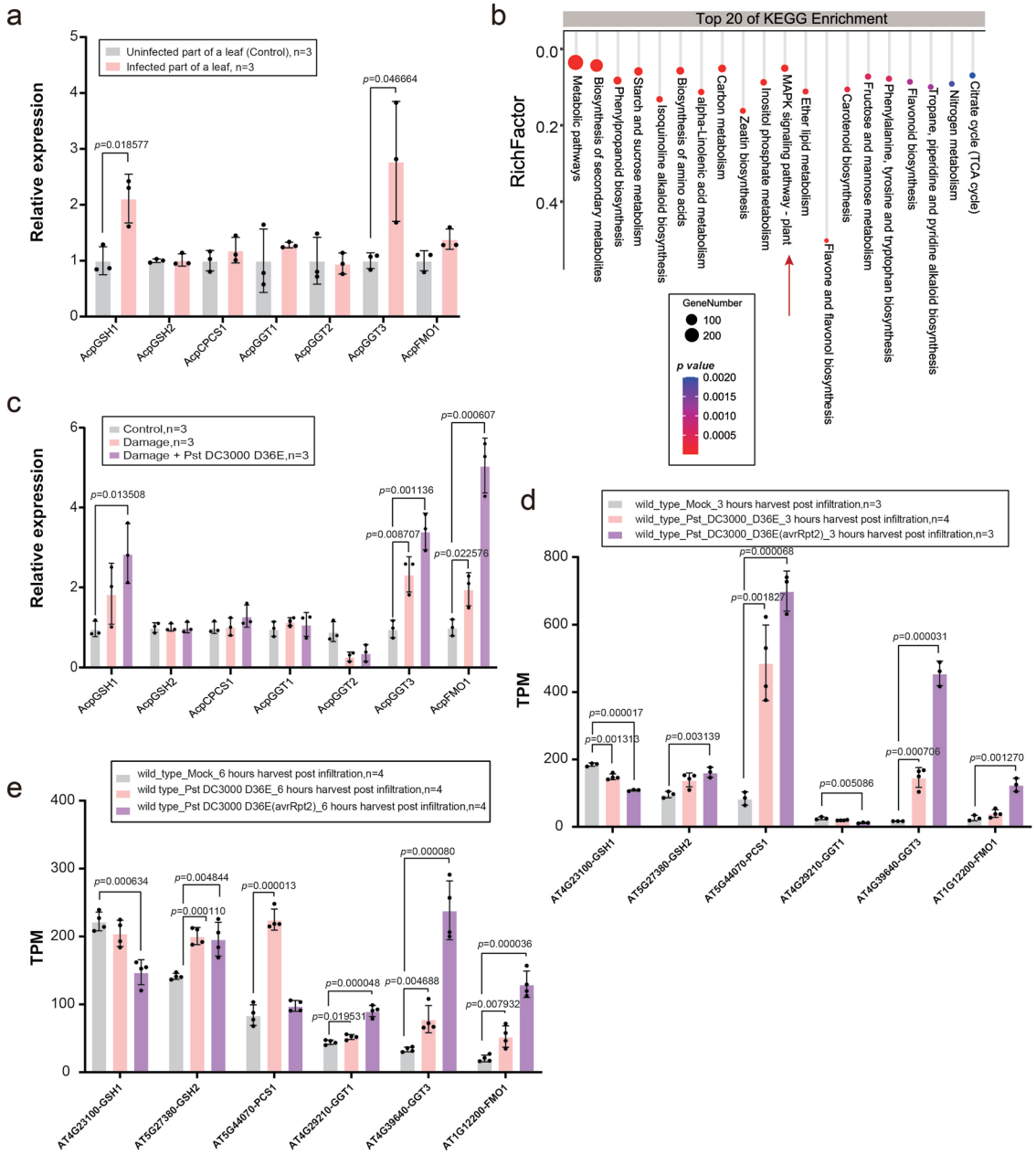

**Extended Data Fig. 5 | Clustering of lachrymatory factor synthase gene family (*LFS*) in *Allium*, related to Fig. 3.** Phylogenetic analysis of *LFS* gene family, with *LFS* genes forming eight distinct groups (A–H). Bootstrap values are shown on each branch. For *LFS* gene family, we constructed a gene tree using standard maximum-likelihood phylogenetic analysis implemented in IQ-TREE with default parameters and 1,000 bootstraps.

https://doi.org/10.1038/s41588-023-01546-0



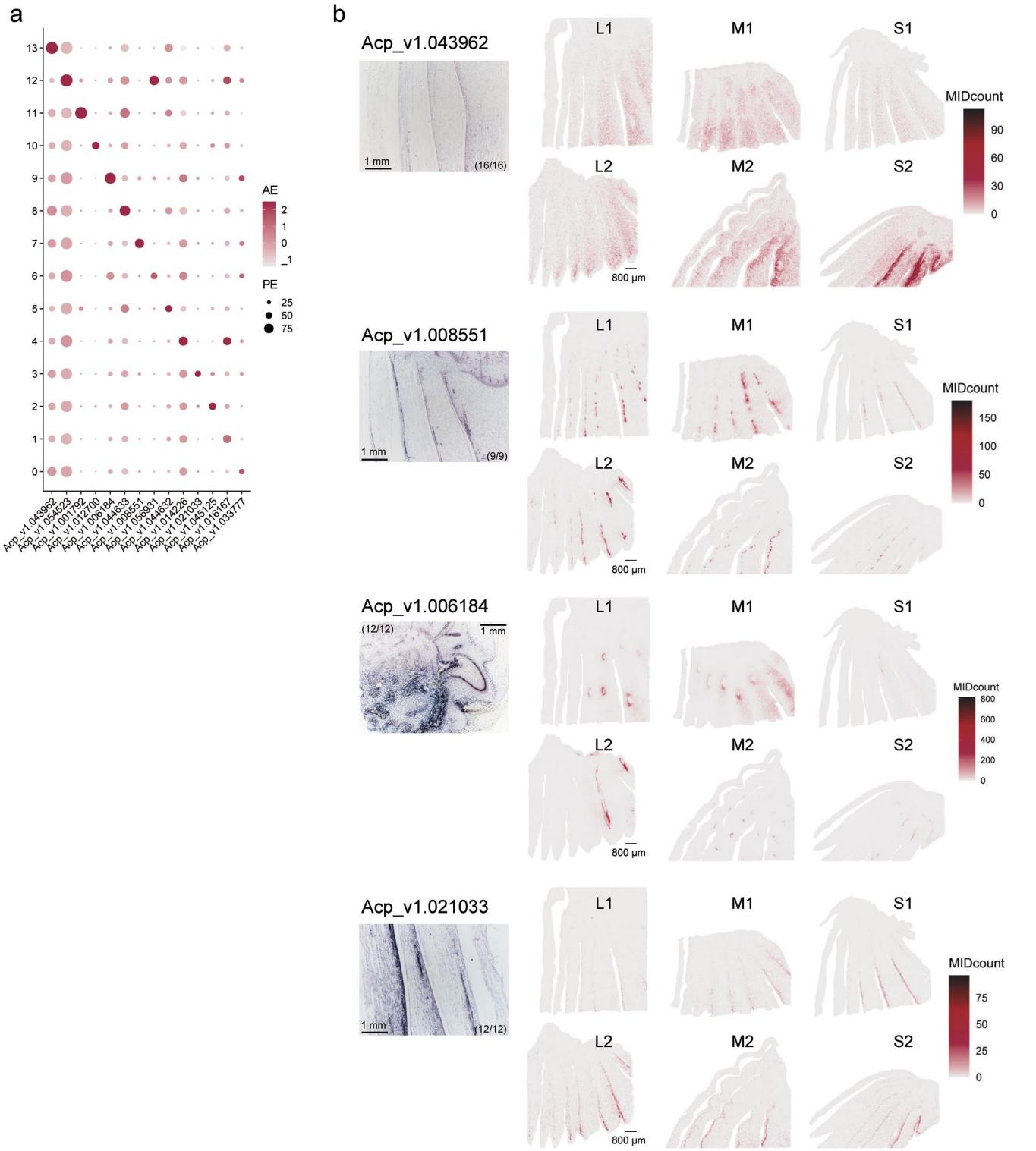

**Extended Data Fig. 6 | Location of genes associated with allicin/isoallicin biosynthesis on chromosomes.** Localization of genes related to allicin/isoallicin biosynthetic pathways on chromosomes of **a**, Welsh onion, **b**, garlic, and **c**, onion.

https://doi.org/10.1038/s41588-023-01546-0



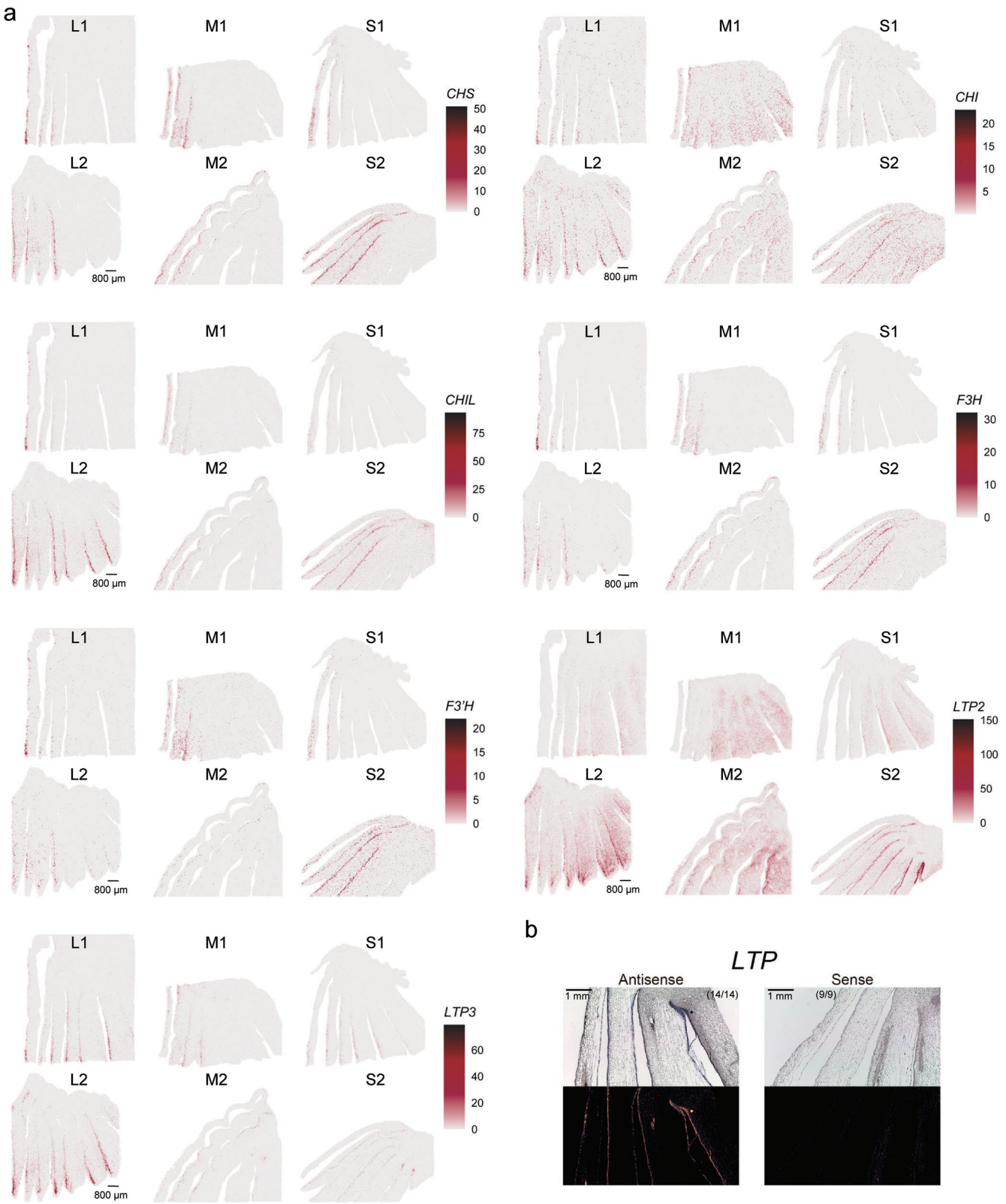

**Extended Data Fig. 7 | Gene expression changes upon pathogen infection in onion. a**, Relative expression levels of alliin biosynthesis genes in transcriptome of fungal-infected blades. **b**, Significant KEGG pathways enriched in up-regulated genes in response to fungal infection. Comparing gene expression profiles between leaf blades of fungal-infected and healthy plants, we identified 1,035 up-regulated genes during infection. **c**, Relative expression levels of alliin biosynthesis genes after artificial injury and exposure to bacterial culture by qRT-PCR. **d** and **e**, Relative expression levels of homologous genes of alliin biosynthesis genes in *Arabidopsis thaliana*. TPM: transcripts per kilobase of exon model per million mapped reads. Transcriptome data were downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142747. Four-week-old *Arabidopsis thaliana* plant leaves were infiltrated with sterile water (Mock) or different Pst strains, then harvested at 3 or 6 h after infiltration. Two bacterial strains, *P. s. pv.* tomato (Pst) DC3000 (avrRpt2), which activates RPS2 (resistance to *P. syringae* 2)-dependent effector-triggered immunity (ETI) and pattern-triggered immunity (PTI), and Pst DC3000 without 'avirulent' effector to activate PTI only in wild-type plants. Data presented in **a**, **c**–**e** are mean ± standard error of three or four independent experiments, and bars with *p* values were significantly different based on t test (two-sided). The *p* values (unadjusted, one side) presented in **b** were calculated using the hypergeometric test.

https://doi.org/10.1038/s41588-023-01546-0



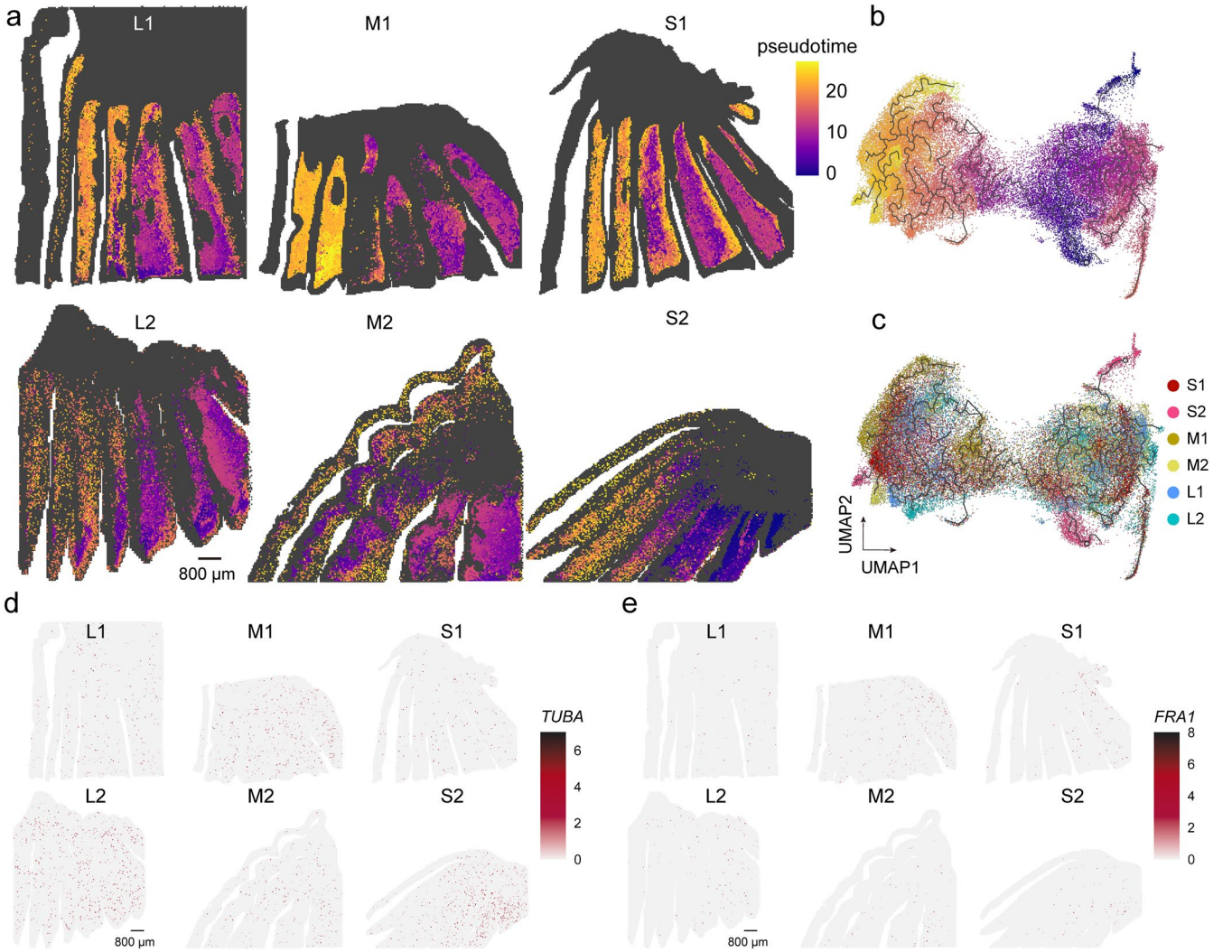

**Extended Data Fig. 8 | Cell marker genes in different clusters, related to Fig. 4. a**, Dot plot showing expression profiles of marker genes in all 14 cell clusters. ID of cell marker genes in onion are below x-axis. Expression level for each bin was calculated by scaled number of molecular identifiers (MIDs) for each marker gene. Average expression (AE) level of all bins in each cell cluster is denoted by dot color. Percentage of bins expressing marker gene (PE) in each cell cluster is denoted by dot size. **b**, Spatial expression patterns of cell marker genes from *in situ* hybridization (left) and spatial RNA-seq (right), showing that *in situ* hybridization and stereo-seq data were highly consistent and supported cell categorization. Scale bars equal to 1 mm for *in situ* hybridization, 800 um for spatial RNA-seq. Number in parenthesis represents number of samples with same signal pattern over total number of samples.

https://doi.org/10.1038/s41588-023-01546-0




**Extended Data Fig. 9 | Spatial expression patterns of genes related to synthesis of flavonoid compounds and cuticular wax, related to Fig. 4e,f.** **a**, Spatial expression patterns of genes (*CHI*, *F3H*, *F3'H*, *FLS*, *LTP2*, and *LTP3*) in different onion bulb development stages based on stereo-seq data. Co-expression of genes in epidermal cells was observed in all samples, showing that flavonoid and cuticular wax biosynthesis is mainly active in epidermal cells. **b**, Expression pattern of *LTP* in leaf base of onion by RNA *in situ* hybridization. Lower panels are images with reverse colors produced by ImageJ based on images in upper panels. Expression patterns based on *in situ* hybridization confirmed *LTP* results in Fig. 4f from large-scale Stereo-seq analysis. Scale bars equal to 800 μm for **a**, 1 mm for **b**. Number in parenthesis represents number of samples with same signal pattern over total number of samples.

https://doi.org/10.1038/s41588-023-01546-0




**Extended Data Fig. 10 | Spongy mesophyll cell lineages and onion bulb formation, related to Figs. 4g,5. a**, Visualization of spongy mesophyll cell lineages, including six sections at three development stages. Sections from different stages and individuals show similar patterns of spongy mesophyll cell development: spongy mesophyll cells from inner to outer layer, and for each layer from base to top and from outer to inner, represent early to late points along the expansion process. Progression of numbers from small to large indicates pseudotime sequence, reflecting progression of spongy mesophyll cell expansion. Same color pattern is also used in panel **b**. **b** and **c**, UMAP dimensionality reduction projection of spongy mesophyll cells grouped by pseudotime scores and sections, respectively. General pattern of pseudotime increasing from left to right can be seen in panel **b**, but no clear pattern can be seen in **c**, suggesting that expansion of spongy mesophyll cell development is a fundamentally short process, starting asynchronously at different development stages as indicated in panel **a**. **d** and **e**, Spatial visualization of expression of indicated genes related to onion bulb formation. Scale bars equal to 800 μm for **a**, **d** and **e**.

nature portfolio

Corresponding author(s): Jing Cai

Last updated by author(s): Sep 18, 2023

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No software had been used for data collection. |
| Data analysis | RepeatMasker v4.1.0, ProteinMask v4.0.5, RepeatModeler v1.0.5, RepeatScout, Piler v0.1-1, LTR_FINDER v1.07, TRF v4.07, Phylip v3.695, LTRharvest, MAFFT v7.407, FastTree v2.1, HMMER2 v3.3.1, Wgdi v0.4.7, MCScanX, ggplot2 v3.3.325, hisat2 v2.0.4, FeatureCounts v2.0.0, DESeq2 v1.23.0, ClusterProfiler v3.16.126, Augustus v3.3.2, GlimmerHMM v3.0.4, SNAP v2.0, PASApipeline v2.4.1, EVidenceModeler v1.1.1, Genewise v2.4.1, Isoseq3 v3.3.0, Trinity v2.10.0, purge_dups v1.0.1, Juicer v1.6, 3D-DNA v180922 , bowtie2 v2.3.4.127, Blast v2.10.0+28, OrthoFinder v2.3.1, PAML v4.9e, RAxML v8.2.12, hifiasm v0.9-r289, CAFE v3.1, GraphPad Prism v 8.0.2, r8s v1.8.1, Astral v5.6.1, R, Python. Perl. Specific parameters used during run-time are provided in the methods. All softwares or scripts are available from official websites or GitHub as indicated in the methods. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The raw sequencing data of onion, garlic, Welsh onion, and Africa lily were deposited in the National Center for Biotechnology Information Sequence Read Archive

under the accession number PRJNA948806 and in the National Genomics Data Center (https://ngdc.cncb.ac.cn/?lang=en) under the accession number PRJCA016760. The assemblies of the four genomes reported in this paper have been deposited in the National Center for Biotechnology Information under accession numbers JASDDO000000000, JASFAV000000000, JASFAW000000000, JASFAX000000000 and in the Genome Warehouse in National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under accession numbers GWHCBHY00000000, GWHCBHZ00000000, GWHCBIA00000000, GWHCBIB00000000 that are publicly accessible at https://ngdc.cncb.ac.cn/gwh.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | No sample size is calculation. The sample sizes of the four species in our study are sufficient for de novo genome sequencing. |
| Data exclusions | Low quality data are excluded for genome assembly. |
| Replication | 6 onions were used for Spatial RNA-seq. For de novo genome sequencing, the sequencing of one sample is enough for genome assembly. |
| Randomization | The samples of onion, garlic, Welsh onion, African lily were randomly collected according to the genus of the species we need, and the detailed information of the sample collection was offered in the methods section. |
| Blinding | Blinding was not applicable to this study because all of our donors were healthy plant donors. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

# Antibodies

| | |
|---|---|
| Antibodies used | Anti-Digoxigenin-AP, Fab fragments antibody (Roche, Cat No #11093274910). Anti-Digoxigenin-AP, Fab fragments are primary antibodies, which contains Fab fragments from polyclonal anti-digoxigenin antibodies, conjugated to alkaline phosphatase. Anti-Digoxigenin-AP, Fab fragments are useful for the detection of digoxigenin-labeled compounds using in situ hybridizations. The antibodies were purchased from MERCK. |
| Validation | The Anti-Digoxigenin-AP, Fab fragments antibody against digoxigenin was validated to be available according to a previous report (Su, S., Zhou, X. & Higashiyama, T. "Whole-mount RNA in situ hybridization technique in Torenia ovules." Plant Reprod 36 (2023): 139–146), in which the Anti-Digoxigenin-AP, Fab fragments antibody is same as we used here. |