Chromosome-level genome assembly of bunching onion illuminates genome evolution and flavor formation in *Allium* crops

The *Allium* genus is cultivated globally as vegetables, condiments, or medicinal plants and is characterized by large genomes and strong pungency. However, the genome evolution and genomic basis underlying their unique flavor formation remain poorly understood. Herein, we report an 11.27-Gb chromosome-scale genome assembly for bunching onion (*A. fistulosum*). The uneven bursts of long-terminal repeats contribute to diversity in genome constituents, and dispersed duplication events largely account for gene expansion in *Allium* genomes. The extensive duplication and differentiation of alliinase and lachrymatory factor synthase manifest as important evolutionary events during flavor formation in *Allium* crops. Furthermore, differential selective preference for flavor-related genes likely lead to the variations in isoalliin content in bunching onions. Moreover, we reveal that China is the origin and domestication center for bunching onions. Our findings provide insights into *Allium* genome evolution, flavor formation and domestication history and enable future genome-assisted breeding of important traits in these crops.
as lachrymatory factor) with the aid of alliinase (ALL) and lachrymatory factor synthase (LFS), whereas alliin can be spontaneously catalyzed into allicin upon tissue damage. However, the genomic basis of this unique flavor formation in Allium crops remains unclear.

The Allium genus possesses some major cash crops including bunching onions, Welsh onions or storey onions (A. fistulosum), bulb onions (A. cepa), garlic (A. sativum), and shallots (A. ascalonicum). Owing to their morphological resemblance, weak reproductive isolation, and similar chromosome karyotypes, the genetic relationships among Allium species are not well defined. Traditional Allium taxonomy is mostly based on plant morphology. For instance, A. fistulosum is generally classified based on its tillering characteristics and pseudo-stem morphology. Shallots (A. ascalonicum) are often mistaken for bunching onions because of their similar tillering characteristics. Chinese red onions (A. cepa var. pseudoplatanus), because of their similar aerial bulbs. These ambiguities in Allium taxonomy often lead to misunderstanding of the evolution and domestication in this genus, which needs to be clarified at the genomic level.

Natural or artificial interspecific hybridizations are ubiquitous among the Allium genus, which have been empirically and randomly utilized to introduce disease resistance and quality traits. For instance, A. fistulosum is generally classified based on its tillering characteristics and pseudo-stem morphology. Shallots (A. ascalonicum) are often mistaken for bunching onions because of their similar tillering characteristics. Chinese red onions (A. cepa var. pseudoplatanus), because of their similar aerial bulbs. These ambiguities in Allium taxonomy often lead to misunderstanding of the evolution and domestication in this genus, which needs to be clarified at the genomic level.

Multiple genome assessments supported the high quality of A. fistulosum genome assembly. First, Benchmarking Universal Single-Copy Orthologue (BUSCO, 91.0% of the 1614 core eukaryotic genes) and Core Eukaryotic Gene Mapping Approach (CEGMA, 94.35% of conserved genes) analyses indicated the high completeness of the assembled genome (Supplementary Tables 6, 7). Meanwhile, the consensus quality value (QV) and the completeness of the A. fistulosum genome were 37.06% and 93.61%, respectively, indicating the high accuracy of our assembly. By mapping short Illumina reads to the genome assembly, we obtained a mapping rate of 99.65%, a genome coverage of 99.34%, and a SNP rate of 0.0063%, demonstrating the high quality of the assembled genome (Supplementary Table 8). In addition, we observed good collinear relationships between the previous genetic map and the assembled chromosomes (Supplementary Fig. 2), which further supported our high-quality genome assembly. Regardless of the large genome size, the high long-terminal repeat retrotransposon assembly index (LA2, 24.81) suggested the gold continuity of the A. fistulosum genome assembly.

A total of 7885 Mb (69.81% of the total genome) of repetitive sequences was annotated using RepeatMasker, TRF, and RepeatProteinMask (Supplementary Table 9). Most of the repetitive sequences were transposable elements (TEs), whereas long-terminal repeat (LTR) retrotransposons were the most abundant TEs in the A. fistulosum genome (62.18% of the total genome; Supplementary Fig. 3a, Supplementary Table 10). Complementary methods using de novo gene prediction, protein-based homology searches, and transcriptome-based predictions have been used to annotate protein coding sequences. In total, 62,255 genes were predicted, with an average gene length of 5002.2 bp, an average exon length of 208.47 bp, and a gene annotation rate of 98.30% (Supplementary Table 1, and Supplementary Tables 3-4), which represents the best contiguity among the currently released Allium genome assemblies.

Table 1 | The main genome assembly features of three major Allium crops

| Assembly feature | A. fistulosum | A. sativum | A. cepa *
|------------------|--------------|-----------|-----------
| Genome size      | 11,274 Mb    | 16,243 Mb | 14,938 Mb |
| Super scaffold N50 | 1386 Mb  | 1691 Mb   | —         |
| Scaffold N50     | 8.98 Mb     | 725 kb    | 460 Kb    |
| Contig N50       | 7.34 Mb     | 194 kb    | 48.29 kb  |
| Total length of transposable elements | 7410 Mb | 14,786 Mb | —         |
| Gene number      | 62,259      | 57,561    | 86,073    |
| Average gene length | 5002.2 bp | 57,561 bp | 86,073 bp |

*The information of A. cepa genome was updated according to https://www.oniongenome.wur.nl/ (visiting date: 20220804).

Results

Chromosome-scale mega-genome assembly and gene annotations

We assembled a chromosome-level genome of the bunching onion (A. fistulosum) using de novo genome sequencing of the bunching onion landrace. First, the k-mer analysis (k = 17) revealed an estimated genome size of 11.97 Giga base-pairs (Gb), a heterozygosity rate of 0.64%, and the repetitive sequences accounted for 89.89% of the whole genome (Supplementary Fig. 1 and Supplementary Table 2). PacBio sequences (1,649.82 Gb, 138.1 × genome coverage) were generated to assemble the genome into contigs, yielding a draft assembly of 27,972 contigs with a total length of 15.36 Gb and contig N50 of 4.72 mega base-pairs (Mb, Supplementary Table 1). After polishing and quality improvement, an improved assembly (scaffold N50 of 8.98 Mb) was obtained (Table 1 and Supplementary Table 3). Finally, Hi-C interaction datasets (1,136.4 Gb, 95.1 × genome coverage) were performed to construct the genome into super-scaffolds, and 98.71% of the assembled contigs were anchored into eight chromosomes (Fig. 1b and Supplementary Table 1, 3–5). The final genome assembly was 11.27 Gb with a contig N50 of 7.34 Mb and a super scaffold N50 of 1.34 Gb (Fig. 1, Table 1, and Supplementary Tables 3–4), which represents the best contiguity among the currently released Allium genome assemblies. Allium genome evolution and chromosome homology

To explore the genomic evolution of Allium, we selected 13 species from Asparagaleae, Arecales, Poales, Amborellaceae, and Scitamineae for further comparative genomic analysis. Among these plants, we...
discovered that the three Allium crops had the highest number of total and unique gene families (Fig. 2a). In total, 492 single-copy gene families were selected to reconstruct a maximum-likelihood (ML) tree, which illustrated that A. fistulosum and A. cepa diverged about 7.4 million years ago (MYA), and the common ancestor of A. fistulosum and A. cepa diverged from A. sativum about 16.7 MYA (Fig. 2a). Among the selected genomes, the Allium crops exhibited numerous expanded and contracted genes, and 308 gene families were commonly expanded in all three Allium crops. Importantly, we found these gene families predominately enriched in gene ontology (GO) pathways like “Catalytic activity,” “Carbon-sulfur lyase activity,” and “Cysteine-type peptidase activity,” implying that the high sulfur-metabolism activities of Allium crops are responsible for the biosynthesis of sulfur compound related to flavor formation (Fig. 2a, b).

The distribution of Ks values of homologous pairs in the syntenic regions of A. fistulosum and A. sativum was similar to those of A. sativum and Asparagus officinalis (Supplementary Fig. 4), indicating that, similar to garlic. A. fistulosum also underwent three WGD events. In addition, most of the genes of the three major Allium crops corresponded to a single block in the collinearity analysis. Therefore, no unique WGD event occurred among these three genomes after divergence (Supplementary Fig. 5). Previously, a high level of macro-synteny between A. fistulosum and A. cepa has been reported using 103 anchor markers. Here, we observed an almost one-to-one syntenic relationship at the chromosome level, except for Chr4, which underwent a clear inversion between the two Allium crops (Fig. 2c, Supplementary Fig. 6). In addition, four A. fistulosum chromosomes (Chr3, Chr6, Chr7, and Chr8) exhibited nearly one-to-one syntenic relationships with those of A. sativum (Chr2, Chr8, Chr3, and Chr1; Fig. 2c, Supplementary Fig. 6). The remaining four chromosomes exhibited several rearrangements. In details, a clear inversion was observed between Chr1 of A. fistulosum and Chr7 of A. sativum. Chr2 of A. fistulosum was partly syntenic to Chr5 and Chr6 of A. sativum, Chr4 of A. fistulosum showed partially syntenic relationships with Chr4 and Chr5 of A. sativum, and Chr5 of A. fistulosum was partially rearranged from Chr4 and Chr6 of A. sativum (Fig. 2c, Supplementary Fig. 6). Even though there were significant gaps among their genome sizes, high chromosome relatedness was observed among the three Allium genomes, which indicated that they were evolutionarily close to each other.

**Insights into genome expansions and gene duplications**

Allium plants have long been known for their large genomes, which are among the largest in vegetable crops. Therefore, we analyzed the key factors potentially associated with Allium genome expansion. We identified LTRs as the leading genomic constituents in the A. fistulosum (Supplementary Fig. 3a and Supplementary Table 10) and A. sativum genomes. Gypsy was the predominant LTR and burst mostly within the last 2 million years (Fig. 3a and Supplementary Fig. 7). Within Gypsy, the Tekey and Tat clades were the most abundant LTRs in both A. fistulosum and A. sativum genomes. The CRM clade contributed significantly (10.38%) to the LTR components only in A. fistulosum (Fig. 3b). Although Copia LTRs comprised a small proportion of repeat sequence in both crops, their components significantly differed between the A. fistulosum and A. sativum genomes (Fig. 3a, Supplementary Figs. 7 and 8). Therefore, the accumulation of LTRs, especially the Gypsy-type, is responsible for genome enlargement in Allium crops, while uneven expansions of different LTR clades lead to diversity in Allium genome constituents.

The three Allium crops showed the highest number of total and unique gene families among the 13 selected genomes (Fig. 2a). To this end, we investigated the global landscape of gene duplications in 16 species with genome sizes ranging from 0.2 to 28 Gb. The large genomes had a relatively high frequency of dispersed duplication (DSD) events, demonstrating that DSD events play an important role in the gene expansion of large genomes (Fig. 3c, Supplementary Fig. 9). After classifying the duplicated paralogs into ancestral and new DSD genes based on synteny, we identified 13,676, 17,817, and 8424 gene pairs (ancestral/new) in A. fistulosum, A. sativum, and A. cepa, respectively. The ancestral DSD genes exhibited significantly higher expression levels than the new genes (Fig. 3d). The expression bias between ancestral and new DSD genes might result from their different environmental responses or functional redundancies, which also implies the possible pseudogenization, sub-functionalization, or neo-functionalization after gene duplications. We propose that ubiquitous DSD events are responsible for gene expansion and divergence in the Allium genomes.

**Expansions and differentiations of flavor-related genes**

To explore the evolutionary events of the putative genes involved in Allium flavor formation, we identified all ACSO-related genes in the 14 plant genomes (Fig. 4a). Seven γ-glutamylcysteine synthetase (GSH1) orthologs, one glutathione synthetase (GSH2) ortholog, one phytochelatin synthase (PCS) ortholog, three γ-glutamyl transpeptidase (GGT) orthologs, and three flavin-containing monooxygenase (FMO) orthologs were identified in the A. fistulosum genome (Fig. 3a). Among these genes, AfGSH1b, AfGSH1g, AfGSH2, AfPCS1, AfGGT1, AfGGT2, AfFMO1, and AfFMO2 were constitutively expressed in the roots, pseudo-stems, leaves, floral stalks, and flowers, indicating that these
genes might play roles in ACSOs biosynthesis in *Allium fistulosum* (Fig. 4b). The ALL gene encoding a key enzyme for ACSOs hydrolysis was found to be extensively expanded in bunching onion (56 *AfALL*), bulb onion (65 *AcALL*), and garlic (48 *AsALL*) genomes compared with other selected genomes (Fig. 4a, Supplementary Fig. 10). Importantly, 43, 27, and 29 *LFS* orthologs were identified in the genomes of *A. fistulosum*, *A. cepa*, and *A. sativum*, respectively, whereas no *LFS* genes were found in species other than *Allium* among the selected genomes (Fig. 4a, d).

According to our ACSOs contents analysis and relevant report, isoalliin was the major sulfur-containing bioactive substance in both *A. fistulosum* and *A. cepa*, whereas alliin was not detected in these two crops. However, alliin was the dominant sulfoxide in *A. sativum* and a small amount of isoalliin was detected (Fig. 4c). LFS is the key enzyme...
Fig. 3 | Extensive genome expansions and gene duplications in Allium crops. 
a Distribution of insertion times of Gypsy and Copia retrotransposons in A. fistulosum. 
B The phylogenetic relationships of Gypsy LTR-retrotransposons identified in A. fistulosum (left) and A. sativum (right) genomes. 
c The distribution of duplicated genes in three Allium genomes. DSD dispersed duplication, PD proximal duplication, TD tandem duplication, TRD transposed duplication, WGD whole-genome duplication. 
d The comparison of the expression levels of ancestral and new DSD genes in three Allium genomes (gene numbers: \(n = 13,676\): A. fistulosum, \(n = 5424\): A. cepa, and \(n = 17,817\): A. sativum). In the box plots, the minima, maxima, center bounds of box showed 25%, 75% quartile values and the average values, while the whiskers showed maxima and minima values of expression levels. \(P\) values were calculated using the two-tailed Wilcoxon test. R roots, PS pseudo-stems, B bulbs, L leaves, FS floral stalks, F flowers, GS garlic sprout. Source data are provided as a Source Data file.
Fig. 4 | Massive duplication and differentiation of sulfur-related genes in *Allium*. a The biosynthesis and hydrolysis pathway of ACSOs and the number of each flavor-related gene in the 14 genomes. The pathway in blue relates to the biosynthesis of isoalliin that dominates in garlic. The pathway in green relates to the biosynthesis of alliin that dominates in garlic. The pathway in blue relates to the biosynthesis of isoalliin that dominates in bulb onion and bunching onion. The gene details are listed in Supplementary Data 1. b Expression patterns of ACSOs biosynthesis genes were investigated in three *Allium* crops by ultra-high-performance liquid chromatography. The pie chart exhibited the percentages of three major ACSOs (alliin, isoalliin, and methiin). c Quantification of ACSOs among three *Allium* crops by ultra-high-performance liquid chromatography. The pie chart exhibited the percentages of three major ACSOs (alliin, isoalliin, and methiin). d The phylogenetic tree and the expression pattern of LFS gene families from *A. fistulosum*, *A. cepa*, and *A. sativum* genomes. The outer rings highlight the genes from *A. fistulosum* (red color), *A. cepa* (blue color), and *A. sativum* (green color). The heatmap exhibited the gene expressions in different tissues: R roots, L leaves, PS pseudo-stems, B bulbs, FS floral stalks, F flowers. Source data are provided as a Source Data file.
that catabolizes hydrolyzed isosalliin into the characteristic propanthial S-oxide (lachrymatory factor) in Allium crops. We classified the LFS gene family into four groups, LFS-a, LFS-b, LFS-c, and LFS-d, according to their phylogenetic structure annotated in three major Allium crops (Fig. 4d). In addition, BAC21275.1 was verified as a functional LFS in bulb onions and was clustered into the LFS-c family (Fig. 4d). A ALFS01, the best hit with BAC21275.1 in bunching onions, exhibited the highest transcriptional levels in most tissues, except roots, among all the ALFS genes examined (Fig. 4d). Importantly, eight ALFS, seven AclFS and three AslFS genes in this group presented similar expression patterns (Fig. 4d). Based on the close phylogenetic relationships with BAC21275.1 and their similar expression patterns, we suppose that the LFS-c group might play synchronized roles in ACSOs hydrolysis in Allium crops. However, as the phylogenetically closest duplications, ALFS15 and ALFS16 showed different expression patterns against ALFS51 (Fig. 4d), suggesting the potential for functional differentiation of these homologous genes. We also observed different expression biases of the LFS groups in these crops. A. fistulosum and A. cepa preferred to extensively express LFS-group members, whereas A. sativum predominately expressed LFS-a group members (Fig. 4d).

Similarly, three proximally duplicated genes (AALL47, AALL48, and AALL33) in Chr4 of bunching onion, four AcALL genes (ACALL55, ACALL68, ACALL98, and ACALL109), four AAD26853.1 and six ASALL genes (ASALL31, ASALL32, ASALL33, and ASALL34) belong to the same cluster as known ALL genes (AAD26833.1, AAB32477.1, and ACN78381.1; Supplementary Fig. 10), all of which exhibited high expression levels in most tissues. These findings suggested that this gene cluster may be functional ALL genes in these three crops.

The LFS-b group consisted of nine ALFS genes that exhibited extremely low gene expression in all tissues (Fig. 4d). In contrast, 12 ASlFS genes clustered in an independent clade in the LFS-d group and were mainly expressed in the roots (Fig. 4d). Furthermore, the microcollinearity analysis indicated that several LFS and ALL genes in bunching onion and garlic exhibited reciprocal one-to-many relationships (Fig. 5a, b). Thus, we inferred that the independent expansion and differentiation of LFS and ALL gene families after the divergence of bunching onion and garlic from their common ancestor might relate to their flavor diversity (Figs. 4d, 5a, b and Supplementary Fig. 10). Notably, the dispersed duplication, tandem duplication, and proximal duplication events of the LFS and ALL gene families were attributed to gene duplications (Fig. 5c, d, Supplementary Data 1). In particular, 33 ALFSs, among 43 duplicates, were clustered in the form of tandem repeats within a 10 Mb region on Chr5 and the intertwined coexistence of LTRs and LFS genes in this region suggested that LTRs likely contribute to LFS gene family proliferation (Fig. 5c, Supplementary Fig. 11). We concluded that the independent expansion and differentiation of sulfur-metabolism-related genes prevail in Allium genomes, which might relate to the characteristic pungent flavor evolution in Allium crops.

Population structure and migration routes of A. fistulosum
We re-sequenced 135 diverse Allium accessions to elucidate the phylogenetic relationships of A. fistulosum crops and their close relatives (Supplementary Figs. 12–13 and Supplementary Data 4). We generated 16,453,99 Gb Illumina short-reads with an average depth of 9.49 × genome coverage. In total, 48,218,339 SNPs and 18,182,723 insertions and deletions (InDels) with an average of 4.08 SNPs and 1.54 InDels per kb were identified. The maximum-likelihood tree of 135 Allium accessions revealed four main clusters centered on Chinese red onions (A. cepa var. proflorum), shallots (A. ascalonicum), A. fistulosum Group I, and A. fistulosum Group 2 (Fig. 6a). All nine Chinese red onions and the A. altaicum accession clustered tightly into one group. Twenty-two shallots clustered together and displayed two distinct subgroups, whereas the remaining two shallots were close to the Chinese red onion group (Fig. 6a). Chinese red onions and shallots clustered apart from A. fistulosum, fitting well with their relatively low genome mapping rate (79–89%) to our reference genome. Therefore, we concluded that Chinese red onions and shallots are close relatives or sibling species of A. fistulosum. Ninety-six A. fistulosum accessions were clustered into two groups; the remaining five A. fistulosum accessions were clustered with shallots and Chinese red onions in the cladogram, which was indicative of ambiguity of mis-assignment in the Allium taxonomy (Fig. 6a and Supplementary Data 4).

To clarify the evolution and migration of A. fistulosum, we performed phylogenetic relationships and population structure analyses with 96 affirmed A. fistulosum accessions, using A. altaicum as an outgroup. The population structure of A. fistulosum was investigated and the cross validation (CV) error was minimized when K = 5 (Supplementary Fig. 14). The 96 A. fistulosum accessions could be categorized into five subgroups and fell into two groups (Fig. 6a, b). Group 1 (G1) could be divided into three subgroups with distinctive geographical distributions based on the population structure. G1 accessions were mainly distributed in western China, G2-1 accessions were distributed in Middle Asia and the Russian Federation, and G3 accessions were distributed in the northern and northeastern areas of China. Group 2 (G2) also displayed two differentiable geographic subgroups, with accessions from southeastern China clustered in subgroup G2-1. Accessions distributed in Japan, America, and Europe showed high genomic similarities and were clustered as G2-2 (Fig. 6b, c).

Furthermore, the principal components analysis (PCA) revealed that G2-1 and G2-2 clustered individually, whereas the three China-originated groups (G1-1, G1-3, and G2-2) showed a closer relationship (Fig. 6c). G1-3 possessed the lowest linkage disequilibrium (LD) decay, indicating that north and northeast China might be a diversity center of A. fistulosum (Supplementary Fig. 15). In contrast, G2-1 showed a unique structural pattern (when K = 3, 4, 5, and 6) with relatively low nucleotide diversity (π) and high LD decay (Fig. 6b, d and Supplementary Fig. 15). Furthermore, the fixation indices (FST) of G2-2/G1-1 (0.0783), G2-1/G1-3 (0.0891), G1-2/G2-1 (0.1095), and G2-1/G2-2 (0.1491) were higher than those of other subgroup pairs (Fig. 6d). The long leading branch of G2-1 (Fig. 6b) suggested that there might be an early divergence and independent domestication of G1-2 group, which also partly explains the relatively independent PCA clustering of G1-2 (Fig. 6b, c). Moreover, four accessions from western China (G1-1) clustered with Middle Asian accessions (G1-2, G1-6a), and there was a clear gene flow from the G1-1 group to the G1-2 group (Supplementary Fig. 16), indicating that Middle Asian accessions might have migrated from western China. Besides, the genetic structure of G2 exhibited a gradually changing trend from G2-1 to G2-2. When K = 2 and 3, G2-1 possessed both G1-3/G1-1 and G2-2 genetic structures, which illustrated that the phylogenetic relationship between G2-1 and G1 was closer than that between G2-1 and G2 (Fig. 6b). In addition, FST values between G1 and G2-2 were lower than those between G1 and G2-1 (Fig. 6d). Together with their geographical locations, we speculated that the Japanese, American, and European accessions (G2-2) migrated or were introduced from southeast China (G2-1, Fig. 6e).

Genomic signals of isosalliin variations in A. fistulosum
We quantified the isosalliin contents of 91 A. fistulosum accessions to determine the genomic basis of the isosalliin variations. The isosalliin content of G1-2 accessions were observed to be significantly higher than those of G1-3 accessions (Fig. 7a). By comparing the selective sweeps between these two groups, we discovered 2718 and 2133 genes located in the selective regions of G1-2 and G1-3, respectively. Several sulfur-flavor-formation-related genes, such as AFGSU2, AFM03, AFGGT3, AFA1L48, AFA1L49, AFA1L55, and AJLF51 were under selection in G1-2. APCS1 and AFA1L42 were identified to be under selection in G1-3 (Fig. 7b). Moreover, a positive correlation between the endogenous isosalliin amount and expression levels of several ACSO-biosynthesis or
-hydrolysis genes was observed in the leaves of six accessions with significant difference in isoalliin content, among which AfGSH2 and AfPCS1 exhibited the highest correlation (Fig. 7c). Among the 25 flavor-formation-related genes that were highly expressed in low-isoalliin accessions (A12, C155, and C16), 19 ACSOs hydrolysis-related genes (ALL and LFS) were identified, implying that both weak biosynthesis and strong hydrolysis may contribute to their lower isoalliin accumulation. Among these 19 ACSOs hydrolysis-related genes, differential LFS or ALL gene expression patterns in six accessions were observed (Fig. 7c), implying that the extensive expansion of ACSOs hydrolysis-related genes might offer more evolutionary opportunities for the variation of isoalliin contents in bunching onion.

**Discussion**

In the present study, we generated a chromosome-level mega-genome of *A. fistulosum*, with a genome size of 11.27 Gb and contig NS0 of 7.34 Mb (Table 1). Our reference genome exhibited superior quality based on both major assembly indicators and several assessments (BUSCO, CEGMA, LAI, and Merqury; Supplementary Tables 6 and 7). Undoubtedly, the achievement of this genome assembly will accelerate our understanding of the evolution and relatedness of major *Allium* crops. The genome enlargements of both bunching onion and garlic crops arise from bursts of common *Gypsy*-type LTRs, but they exhibit uneven expansions of the LTR subclades. Meanwhile, DSD events dominated the contributions to gene expansion in major *Allium*
Fig. 6 | Population structure and the plausible migration routes of *A. fistulosum*. a The maximum-likelihood tree of 135 re-sequenced *Allium* accessions. b The maximum-likelihood tree and population structure of 96 re-sequenced *A. fistulosum* accessions. c The principal component analysis of 96 re-sequenced *A. fistulosum* accessions. d Nucleotide diversity (π) and population divergence (*F*ST) across the five subgroups. The numbers in red font were π values and the numbers in blue font were *F*ST values. e The possible migration routes for bunching onion (the map was downloaded from Mapswire, https://mapswire.com/world/physical-maps/). The dots indicated the habitat of the close relatives of *A. fistulosum*. 

*Note: The text is a partial translation and may not perfectly match the original content. The full article can be accessed via the provided DOI.*
crops (Fig. 3). As highly conserved genome synteny has been observed in bunching onion and bulb onion with an SSR-markers based genetic map\textsuperscript{21}, high genome collinearity between bulb onion, bunching onion, and garlic was shown at the chromosome level. The chromosomes of the bunching onion (\textit{A. fistulosum}) and bulb onion (\textit{A. cepa}) showed almost one-to-one collinearity (Fig. 2c, Supplementary Fig. 6), indicating high consanguinity between these two \textit{Allium} crops. Therefore, the estimated sequence of each chromosome of the bunching onion can be used as a template for onions with high commonality in the chromosome structure. Meanwhile, this genomic evidence well explains the weak reproductive isolation and ubiquitous existence of interspecific hybridization among \textit{Allium} crops\textsuperscript{14–17}, which provides valuable information for directing future genome-aided breeding of \textit{Allium} crops.

\textbf{The \textit{Allium} genus is characterized by unique flavors and medicinal properties owing to its specific cysteine-derived sulfur compounds\textsuperscript{3–5}. We observed that sulfur-metabolism-related genes in GO terms of “Carbon-sulfur lyase activity” and “Cysteine-type peptidase activity” were specifically expanded in the three \textit{Allium} crops examined (Fig. 2b). Importantly, extensive expansion of \textit{ALL} and \textit{LFS} gene families is a striking feature of \textit{Allium} crops. In intact tissues, ACSOs are stored in the cytosol of mesophyll cells. Upon tissue damage, the vacuolar \textit{ALL} enzyme contacts and hydrolyzes ACSOs to produce sulfenic acids, which are further converted into various sulfur-containing bioactive compounds via spontaneous reactions\textsuperscript{5}. Thus, we speculated that the extensive expansion of ACSOs hydrolysis-related genes might lead to the over-accumulation of related transcripts and resultant enzymes, which might be helpful for quick responses to external stimuli. In particular, the \textit{LFS} gene family was detected only in \textit{Allium} crops among the 14 selected genomes (Fig. 4a). These results strongly support that the \textit{Allium} crops evolved large and special \textit{ALL} and \textit{LFS} gene families possibly to resist diseases and insects, by utilizing its volatile and pungent end-products\textsuperscript{5–7}.

The species-specific clustering of \textit{ALL} and \textit{LFS} gene families indicated their independent expansions in the three \textit{Allium} crops (Fig. 4, Supplementary Fig. 10), implying that their pungent flavor might have resulted from convergent evolution. It is known that the activity of \textit{LFS} is reduced to 1/100 or less even by replacing only one amino acid\textsuperscript{29}. Therefore, we speculate that the extensive duplication and differentiation of the \textit{LFS} gene family may lead to the difference of protein activities or functional diversities. \textit{LFS} genes with high activity and high expression levels are likely involved in the production of tearing...
properties and characteristic flavors in *Allium* crops. Those members with no or negligible LFS activity, might act in other secondary metabolic pathways instead. It has been reported that different terpene synthase (TPS) subgroups participate in different bioactive terpene biosynthesis. Similarly, *A. fistulosum* and *A. cepa*, preferred to extensively express LFS-c group members, which may function in isolaillin hydrolysis. However, *A. sativum* predominately expresses both LFS-a and LFS-c members (Fig. 4c, d), which might act on alliin and isolaillin decompositions. ACSOs can be decomposed into various sulfur constituents including diallyl and methyl allyl, thus, we conjecture that different clades of the LFS and ALL gene families might play roles in formulating other sulfur-containing bioactive compounds, which need to be further verified in future genetic and biochemical studies.

Moreover, the expansion of ALL and LFS gene families were possibly synchronized with the proliferation of LTRs (Supplementary Fig. 11)\(^{23,24}\). In addition, FMO genes, involved in the biosynthesis of both ACSOs and glucosinolates, underwent extensive expansions only in *Arabidopsis* among the 14 selected genomes (Fig. 4a), which might be related to the accumulation of sulfur-containing glucosinolates in Cruciferous crops\(^{31}\). However, the absence of several glucosinolate-related genes in *Allium* crops suggested that they might select different evolutionary directions toward the biosynthesis of sulfur-containing bioactive compounds, compared with *Cruciferous* plants (Fig. 4a and Supplementary Fig. 17).

Traditional *Allium* categories cause confusion in evolutionary studies and exploitation of agricultural usage, mainly because of the lack of genome information. Previously, *A. altaicum* was commonly considered the ancestor of *A. fistulosum* based on its phenotypic resemblance and noncoding chloroplast DNA similarity\(^{25,34}\) (Supplementary Fig. 12). Here, we found that it did not cluster with *A. fistulosum* in either SNP-based phylogeny or PCA analyses, but was phylogenetically closer to the Chinese red onion (Fig. 6a and Supplementary Data 4). Thus, we postulated that *A. altaicum* is closely related to *A. fistulosum*, rather than its proposed ancestor. In contrast to earlier morphological classifications, three storey onions (*A. fistulosum* var. *viiparum*) accessions and the Chinese red onions (*A. cepa var. proliferum*), propagated by aerial bulbs, were clustered into two different groups (Fig. 6a, Supplementary Fig. S18 and Supplementary Data 4). Their characterized aerial bulbs, which could be used for asexual propagation, may have resulted from convergent evolution.

The geographic positions of GI-1 are close to the major diversity center (ranging from the Mediterranean Basin to Central Asia) for *Allium* genus, and the highest nucleotide diversity was observed in this subgroup (Fig. 6d). These findings are consistent with the historical records that northwestern China was the origin center of *A. fistulosum*\(^{35,36}\). In addition, the literature on the cultivation of *A. fistulosum* in China dates back to the third century BC. The Japanese *A. fistulosum* accessions were considered to be introduced from China at approximately 720 AC. Thereafter, *A. fistulosum* was spread to Western Europe during or at the end of the Middle Ages (1000-1400 AC) from Asia\(^{35,36}\). These records are supportive of our phylogenetic results that the G2-2 (accessions from Japan and America and European countries) might have migrated from southern China.

Furthermore, among *A. fistulosum*, diverse accessions exhibited distinct geographic clustering rather than morphological classification in phylogenic analysis (Fig. 6b). Middle Asian accessions (GI-2), which are distributed in Siberia and its adjacent areas, might migrate from western China (GI-1), whereas Japanese, American, and European accessions (G2-2) migrated or were introduced from southeast China G2-1 (Fig. 6e). Collectively, we validate that Northwestern China (GI-1) is the single origin center of *A. fistulosum*.

In summary, the high-quality chromosome-level genome of *A. fistulosum* greatly advances our understanding of genome diversity and expansion of *Allium* crops. The highest diversification of China-originated accessions and the existence of close relatives indicate that China is the primary origin and domestication center for *A. fistulosum* crops. The expansion and differentiation of the alliinase and lachrymatory factor synthase gene families might relate to *Allium*-specific flavor formation. Our current study will enable further functional genomic studies and genomic selection of important agronomic traits in *Allium* crops.

**Methods**

**Plant materials and sequencing libraries preparations**

The bunching onion accession “SXSJC” (grown in Hangzhou, Zhejiang Province, China) was sequenced for genome assembly. Detailed information on the 135 *Allium* accessions used for re-sequencing is provided in Supplementary Data 4. A total of 101 *A. fistulosum* accessions were distributed worldwide, including in China, Japan, and European, American, and Central Asian countries. Shallots and Chinese red onions were collected from China, and the feral *A. altaicum*, which was considered the closest relative and potential ancestor for cultivated *A. fistulosum*\(^{1,33}\), was collected from the Altai Mountains (59.44E, 47.05N; Xinjiang, China).

For genome assembly, genomic DNA was extracted from the young leaves of bunching onions using a DNAsecure Plant Kit (Tiangen Biotech, Beijing, China), and DNA quality was detected by agarose gel electrophoresis. Single molecular real-time (SMRT) PacBio sequencing libraries were built following the standard protocols of Pacific Biosciences. Briefly, high-quality genomic DNA was sheared to ~20 kb, the damaged ends were repaired and ligated with a blunt-end adaptor, and the resulting libraries were sequenced using the PacBio Sequel platform.

For Bionano sequencing, DNA was extracted using the suggested kits (Bionano Prep Plant DNA Isolation Kit), and the specific sequences across the entire genome were labeled. The labeled DNA was transferred into a cartridge for scanning, and the images were converted into molecules.

Hi-C libraries were constructed according to the suggested procedure. Briefly, leaf samples were fixed with formaldehyde solution before chromatin extraction, and chromatin was digested with 400U of DPNII restriction enzyme at 37°C. DNA ends were labeled with biotin, and DNA ligation was performed using T4 DNA ligase. After ligation, proteinase K was added for reverse crosslinking. The DNA fragments were then purified and dissolved. The purified DNA was fragmented to 300–500 bp, and the DNA ends were repaired. Biotin-labeled DNA fragments were separated using Dynabeads M-280 Streptavidin (Life Technologies). Hi-C libraries were controlled for quality and sequenced on an Illumina HiSeq X Ten sequencer.

For Illumina sequencing, DNA libraries were constructed according to a previously described procedure (Illumina). Briefly, DNA was broken randomly into segments using a Covaris ultrasonic crusher, DNA ends were repaired, and poly (A) adaptors were added. PCR was performed using purified DNA, and the final libraries were sequenced on an Illumina platform after assessment.

**Genome assembly and quality assessment**

The genome size, heterozygosity ratio, and repeat sequence ratio were evaluated by *k-mer* distribution analysis (k = 17) using Illumina short-reads. PacBio reads were used to assemble the contig-level genome using CANU\(^{30}\) (v1.9, parameters: genomeSize = 12 G, corOutCoverge = 40 G). The contigs were polished with PacBio reads using NextPolish (v1.2.4, rerun = 3, https://github.com/Nextonics/NextPolish) and Illumina short-reads using Pilon (v1.2.2)\(^{31}\). Subsequently, redundant sequences were removed using Purge_dups (v1.2.3)\(^{32}\). Bionano data were used for auxiliary assembly using Bionano Solve (v3.5.1, DeNovo Assembly: -i 5 -F 1 -W 1 -c 1, Hybrid Scaffold: -B 2 -N 2) to improve the accuracy of the
assembly\textsuperscript{41}. Hi-C data were used to assemble the chromosome-level genome using ALLHIC (v0.9.8, https://github.com/tangerzhang/ALLHIC) software.

Complementary methods were employed to evaluate the quality of the genome assembly. First, genome completeness was assessed based on conserved plant genes in the BUSCO (v3.0.2) and CEGMA (v2.5) databases. Second, Illumina short-reads were mapped to the assembled genome using BWA (v0.7.8)\textsuperscript{41} to assess coverage rate and average depth. The LTR_retriever (v1.0.7)\textsuperscript{52} package was used to evaluate genome quality. Mequery (v1.3) was used to assess the consensus quality (QV) value and completeness of the genome assembly\textsuperscript{50}. The markers of genetic map\textsuperscript{21} were mapped to our genome using Chromomoncer (v1.07, https://github.com/jieluyer/chromonomer_workflow), and collinear relationships were displayed using Python scripts.

**Genome annotations**

RepeatMasker and RepeatProteinMask (v4.07)\textsuperscript{66} were used to identify TEs by alignment to the repeat library (Repbase v15.02), whereas de novo prediction of TEs was performed using RepeatModeler (v1.05, http://www.repeatmasker.org/RepeatModeler/, version 1.0.5), RepeatScout, andLTR-FINDER. Tandem Repeats Finder (TRF, v4.09) analysis was performed to identify the tandem repeats in the genome.

De novo prediction was performed using five ab initio gene prediction programs: Augustus (v3.2.3)\textsuperscript{46}, Geneid (v1.4, https://genome.crg.cat/software/geneid/index.html), Genescan (v1.0, http://genes.mit.edu/GENSCAN.html), GlimmerHMM (v3.0.2, https://ccb.jhu.edu/software/glimmerhmm/), and SNAP (v2013.11.29, https://snap.cuf.edu/software/snap/index.html), and the ends of the LTRs were aligned using MUSCLE (v3.8.31, http://www.drive5.com/muscle/) with the suggested parameters. We classified LTRs into Copia, Gypsy, and other superfamilies using LTRdigest (v1.07)\textsuperscript{52}, while the secondary LTRs were determined according to their phylogenetic properties.

**Insertion time and phylogenetic analysis of LTRs**

The ends of the LTR retrotransposons were aligned using MUSCLE (v3.8.31, http://www.drive5.com/muscle/), and the insertion time (T) of the LTRs was calculated using the formula T = K/2r. K is the genetic distance calculated using the formula K = −0.75ln(1−A/3). The nucleotide divergence rate (λ) between the two LTRs less than 0.75 was retained for further analysis. The nucleotide substitution rate “r” was set to 1.3e-8 substitutions per site per year. Uncorrected pairwise distances were used to construct a neighbor-joining unrooted phylogenetic tree using TreeBest (v1.9.2, http://treesoftware.sourceforge.net/treebest.shtml) with the suggested parameters. We classified LTRs into Copia, Gypsy, and other superfamilies using LTRdigest (v1.07)\textsuperscript{52}, while the secondary LTRs were determined according to their phylogenetic properties.

**Identification of ancestral and new DSD genes**

Gene duplication events were classified as dispersed, proximal, WGD, tandem, and transposed duplications using the DupGen_finder (https://github.com/DXDR/DupGen_finder) pipeline. We investigated the global landscape of gene duplications in 13 selected genomes and several large genomes, including three large genomes of Ginkgo biloba\textsuperscript{50}, rye\textsuperscript{51}, and Chinese pine\textsuperscript{52}. Chromosome collinearity was analyzed to determine whether a gene was located in the collinearity region based on JCVI\textsuperscript{53}. Both A. sativum and A. cepa used A. fistulosum as the reference genome, whereas A. sativum was selected as the reference genome for A. fistulosum. Genes in collinear regions were considered ancestral DSD genes, whereas those outside the collinear regions were considered new DSD genes. We selected dispersed gene pairs in which one was an ancestral gene and the corresponding gene was a new gene to analyze the expression difference.

**Identification and evolution analysis of flavor-related genes**

To identify genes involved in the sulfur-metabolism biosynthesis pathways, the key parameters of BLASTp (v 2.2.26) were E-value <1e-5 and the identity >50%. The protein domains of the candidate genes were identified using HHMmer (v3.1b1) software, and genes with different domains were filtered. The protein sequence of bulb onion LFS gene (GenBank accession no. AB089203) and garlic ALL genes (GenBank accession no. AAB32477.1, ACN78838.1, AAD26853.1, and AAA32639.1) from the NCBI database were used to identify the candidate orthologs. The query sequences of glucosinolate-related genes were obtained using OrthoMCL (v1.4, http://orthomcl.org/orthomcl/). A super-alignment matrix was obtained by aligning single-copy gene families with MUSCLE (v3.8.31, http://www.drive5.com/muscle/), which was subsequently used for ML phylogenetic tree construction using RaxML (ML tree, Model GTR+GAMMA, v8.2.12, http://sco.h-its.org/exelixis/web/software/raxml/index.html). Mcmctree (burn-in = 10,000, sample number = 100,000, sample frequency = 2; http://abacus.gene.ucl.ac.uk/software/paml.htm) was used to calculate the divergence time based on the TimeTree database (http://www.timetree.org/) and a previous report\textsuperscript{56}. Gene family expansion and contraction were detected using CAFE (v4.2)\textsuperscript{60} software. Genomic collinearity between garlic and bunching onions was analyzed using the MCSIcaan pipeline\textsuperscript{61}.

**Spatial expression of the sulfoxide-related genes**

Illumina RNA-seq was performed on the roots, pseudo-stems, leaves, floral stalks, and flowers of bunching onions, and the roots, bulbs, leaves, and pseudo-stems of bulb onions with four biological replicates. Gene expression datasets were obtained using the Tophat (http://ccb.jhu.edu/software/tophat/index.shtml) and

---

**Note:** The content provided is a natural language representation of the document. The URLs and references have not been included as they are part of the supplementary material and are not essential for the understanding of the text provided. The document analysis focuses on the main findings and methodologies described in the text, with references to specific tools and databases used for genome analysis and annotation.
Cufflinks (http://cufflinks.cbcb.umd.edu/) pipelines. The expression patterns of the target genes were visualized using a heatmap library (https://CRAN.R-project.org/package=heatmap) in the R software with log (FPKM + 1) values.

Re-sequencing and population genomics analysis
A total of 350 bp pair-end reads from 135 accessions were generated using the Illumina HiSeq2500 platform. The mapping rates of *A. ascalonicum* and *A. cepa* var. *proliferum* were less than 90, and 4 × genomic coverage was less than 60%. We used 96 *A. fistulosum* accessions to filter SNPs. The SNPs were subjected to a quality control procedure following these steps: (1) removing the Illumina library construction adapters, (2) removing the reads containing more than 10% unknown bases (N bases), and (3) removing the low-quality reads containing more than 50% of low-quality bases (sequencing quality value ≤5). BWA (v0.7.8) was used to align the clean reads to the reference genome, and Samtools (https://github.com/samtools/samtools) was used to call SNPs/InDels with the following parameters (-q 1 -C 50 -m 2 -F 0.002 -d 1000). The raw SNP/InDel sets with a mapping quality <20, or a depth of the variant position <3, or a missing rate >0.05, or an SNP frequency <5% were filtered. After filtering, we selected high-quality SNPs from the 135 Allium accessions for subsequent analyses. ANNOVAR (v20191024, https://annovar.openbioinformatics.org/en/latest/) was used for the functional annotation of variants. A phylogenetic tree was constructed using SNPhylo (v20180901)66. The population structure of the accessions was investigated using Admixture (v1.23). In addition, PCA was performed using VCFools (v0.1.12b, https://vcftools.github.io/index.html). Linkage disequilibrium analysis (LD decay) was calculated using PoplDecay (v0.3.4)67. The XP-CLR scores were calculated using XP-CLR (v1.0) with a sliding window size of 40 kb and a step size of 20 kb. The top 5% region was selected as the candidate selection region.

Determination of ACSOs in *Allium* crops
To prevent ACSO degradation, O-(carboxymethyl) hydroxylamine hemihydrochloride (OCMHA) was used to inhibit the activity of alliinase68. The extraction procedure was as follows: 0.15–0.20 g flesh weight samples were transferred to 2 mL centrifuge tube containing a 5 mm steel ball, and 990 µL OCMHA solution (1.1 g/L methanol/deionized water 4:1, v/v) and 10 µL L-norleucin solution (1 mg/mL water) were added. After rubbing with a tissue grinding machine, the samples were homogenized for 1 min. The resulting slurry was centrifuged (10 min, room temperature, 12,000 rpm) and the supernatant was then diluted 10,000 × g with a mixture of acetonitrile and deionized water (4:1, v/v). The final extracts were stored in vials at −20 °C until analysis.

ACSOs were identified by comparison with the authentic standards (alliin, Beijing Solarbio Life Science Company, Beijing, China; isoalliin, from Chengdu Caoyuankang Biotechnology Company, Chengdu, China; and methiin, from Shanghai Yuanye Biotechnology Company, Shanghai, China). Mixed calibration curves were used at 15.60°C with a ow rate of 0.6 mL/min. The mobile phase consisted of methanol/deionized water 4:1, v/v) and 10 µL L-norleucin solution (1 mg/mL water) maintained at 60 °C with a flow rate of 0.6 mL/min. The mobile phase consisted of (A) water with 0.5% formic acid and (B) acetonitrile with 0.5% formic acid. A gradient elution was used: 0–4 min, 10–15% A; 4–8 min, 15–60% A; 8–15 min, 10% A; injection volume 1 µL. Ionization was conducted in positive ion electrospray mode, and the operating parameters were as follows: capillary voltage, 3000 V; drying gas temperature, 325 °C; drying gas flow, 51/L/min; nebulizer pressure, 45 psi; sheath gas temperature, 350 °C; and sheath gas flow, 11 L/min. In addition, three biological replicates were used for all the accessions.

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

Data availability
All raw sequencing data generated in this study have been deposited to CNSA (https://db.cngb.org/cansa/) under accession CNP0002276. The expression data of *A. sativum* were downloaded from NCBI GEO database under accession GSE145455. Source data are provided with this paper.

References
1. Friesen, N., Fritsch, R. M. & Blattner, F. R. Phylogeny and new intragenic characterization of *Allium* (Alliaceae) based on nuclear ribosomal DNA its sequences. *Aliso* **22**, 372–395 (2006).
2. Li, Q. Q. et al. Phylogeny and biogeography of *Allium* (Amaryllidaceae: Alliinae) based on nuclear ribosomal internal transcribed spacer and chloroplast rps6 sequences, focusing on the inclusion of species endemic to China. *Ann. Bot.* **106**, 709–733 (2010).
3. Eady, C. C. et al. Silencing onion lachrymatory factor synthase causes a significant change in the sulfur secondary metabolite profile. *Plant Physiol.* **147**, 2096–2106 (2008).
4. Fritsch, R. & Keusgen, M. Occurrence and taxonomic significance of cysteine sulfoxides in the genus *Allium* (Alliaceae). *Phytochemistry* **67**, 1127–1135 (2006).
5. Yoshimoto, N. & Saito, K. S-Alk(en)y cysteine sulfoxides in the genus *Allium*: proposed biosynthesis, chemical conversion, and bioactivities. *J. Exp. Bot.* **70**, 4123–4137 (2019).
6. Fu, X., Li, C., Zhou, X., Liu, S. & Wu, F. Physiological response and sulfur metabolism of the *V. dahlieae*-infected tomato plants in tomato/potato onion companion cropping. *Sci. Rep.* **6**, 36444 (2016).
7. Yan, Z. et al. Repellency of forty-one aromatic plant species to the Asian citrus psyllid, vector of the bacterium associated with huanglongbing. *Ecol. Evol.* **10**, 12940–12948 (2020).
8. Kang, M. et al. A chromosome-level *Camptotheca* acuminate genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nat. Commun.* **12**, 3531 (2021).
9. Xia, E. et al. The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol. Plant* **13**, 1013–1026 (2020).
10. Guo, S. et al. Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat. Genet.* **51**, 1616–1623 (2019).
11. Sun, X. et al. A Chromosome-Level Genome Assembly of Garlic (*Allium sativum*) Provides Insights into Genome Evolution and Alliin Biosynthesis. *Mol. Plant* **13**, 1328–1339 (2020).
12. Yusupov, Z. et al. Phylogenomics of Allium section Cepa (Amarillidaceae) provides new insights on domestication of onion. *Plant Divers* **43**, 102–110 (2021).
13. Peffley, E. B. in *High-Tech and Micropropagation III* (ed. Bajaj, Y.P.S.) 244–261 (Springer Berlin Heidelberg, 1992).
14. Budylin, M. V., Kan, L. Y., Romanov, V. S. & Krustaleva, L. I. GISH study of advanced generation of the interspecific hybrids between *Allium cepa* L. and *Allium fistulosum* L. with relative resistance to downy mildew. *Russ. J. Plant Physiol.* **50**, 387–394 (2014).
15. Benke, A. P., Mahajan, V., Manjunathagowda, D. C. & Mokat, D. N. Interspecific hybridization in *Allium* crops: status and prospectus. *Genet. Resour. Crop Evol.* **69**, 1–9 (2022).
16. Umehara, M. et al. Interspecific hybrids between *Allium fistulosum* and *Allium schoenoprasum* reveal carotene-rich phenotype. *Euphytica* **148**, 295 (2006).
17. Stice, S. P. et al. Thiosulfinate tolerance is a virulence strategy of an atypical bacterial pathogen of onion. *Curr. Biol.* **30**, 3130–3140.e6 (2020).

18. Han, T. et al. Polyploidy promotes species diversification of *Allium* through ecological shifts. *N. Phytol.* **225**, 571–583 (2020).

19. Richter, A., Yockteng, R., Brown, S. C. & Nadot, S. Evolution of genome size across some cultivated Allium species. *Genome* **48**, 511–520 (2005).

20. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).

21. Tsukazaki, H. et al. Construction of SSR-based chromosome map in bunching onion (*Allium fistulosum*). *Theor. Appl. Genet.* **117**, 1213–1223 (2008).

22. Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126 (2018).

23. Galindo-González, L., Mhiri, C., Deyholos, M. K. & Grandbastien, M. A. LTR-retrotransposons in plants: Engines of evolution. *Gene* **626**, 14–25 (2017).

24. Xiao, H., Jiang, N., Schaffner, E., Stockinger, E. J. & van der Knaap, E. A Retrotransposon-Mediated Gene Duplication Underlies Morphological Variation of Tomato Fruit. *Science* **319**, 1527–1530 (2008).

25. Finkers, R. et al. Insights from the first genome assembly of Onion (*Allium cepa*). *G3 (Bethesda)* **11**, jkab243 (2021).

26. Shukla, S. et al. The Onion Genomic Resource: a genomics and bioinformatics driven resource for onion breeding. *Plant Gene* **8**, 9–15 (2016).

27. Lichman, B. R., Godden, G. T. & Buell, C. R. Gene and genome duplications in the evolution of chemodiversity: perspectives from studies of Lamiaceae. *Curr. Opin. Plant Biol.* **34**, 74–83 (2020).

28. Imai, S. et al. An onion enzyme that makes the eyes water. *Nature* **419**, 685–685 (2002).

29. Arakawa, T. et al. Dissecting the stereocorrelated conversion of short-lived sulfuric acid by lachrymatory factor synthase. *ACS Catal.* **10**, 9–19 (2019).

30. Shen, Y. et al. Chromosome-level and haplotype-resolved genome provides insight into the tetraploid hybrid origin of patchouli. *Nat. Commun.* **13**, 3511 (2022).

31. Rose, P., Whiteman, M., Moore, P. K. & Zhu, Y. Z. Bioactive S-alk(en)yl cysteine sulfoxide metabolites in the genus Allium; the chemistry of potential therapeutic agents. *Nat. Prod. Rep.* **22**, 351–368 (2005).

32. Ishida, M., Hara, M., Fukino, N., Kakizaki, T. & Morimitsu, Y. Glucosinolate metabolism, functionality and breeding for the improvement of Brassicaceae vegetables. *Breed. Sci.* **64**, 48–59 (2014).

33. Sanderby, I. E., Geu-Flores, F. & Halkier, B. A. Biosynthesis of glucosinolates gene discovery and beyond. *Trends Plant Sci.* **15**, 283–290 (2010).

34. Friesen, N., Pollner, S., Bachmann, K. & Blattner, F. R. RAPD and noncoding chloroplast DNA reveal a single origin of the cultivated *Allium fistulosum* from *A. altaicum* (Alliaceae). *Am. J. Bot.* **86**, 554–562 (1999).

35. Singh, B. & Ramakrishna, Y. Welsh onion (*Allium fistulosum L.*): a promising spicing-culinary herb of Mizoram. *Indian J. Hill. Farming* **30**, 201–208 (2017).

36. Ford-Lloyd, B. V. & Armstrong, S. J. in *Genetic Improvement of Vegetable Crops* (eds. Kalloo, G. & Bergh, B.O.) 51–58 (Pergamon, Amsterdam, 1993).

37. Kayat, F., Mohammed, A. & Ibrahim, A. M. in *Advances in Plant Breeding Strategies: Vegetable Crops: Volume 10: Leaves, Flowerheads, Green Pods, Mushrooms and Truffles* (eds. Al-Khayri, J.M., Jain, S.M. & Johnson, D.V.) 135–182 (Springer International Publishing, Cham, 2021).

38. Ford-Lloyd, B. V. & Armstrong, S. J. in *Genetic Improvement of Vegetable Crops* (eds. Kalloo, G. & Bergh, B.O.) 51–58 (Pergamon, Amsterdam, 1993).

39. Marcias, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).

40. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

41. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).

42. Guan, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).

43. Lam, E. T. et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* **30**, 771–776 (2012).

44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

45. Tarallo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **25**, 4.10.1-4.10.14 (2009).

46. Stanke, M. et al. AUGUSTUS as ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).

47. Li, S. et al. Chromosome-level genome assembly, annotation and evolutionary analysis of the ornamental plant Asparagus setaceus. *Hort. Res.* **7**, 48 (2020).

48. Harkess, A. et al. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat. Commun.* **8**, 1279 (2017).

49. Yan, L. et al. The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb. *Mol. Plant* **8**, 922–934 (2015).

50. Zhang, G. Q. et al. *The Apostasia* genome and the evolution of orchids. *Nature* **549**, 379–383 (2017).

51. Cai, J. et al. The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* **47**, 65–72 (2015).

52. Al-Mssaleem, I. S. et al. Genome sequence of the date palm *Phoenix dactylifera* L. *Nat. Commun.* **4**, 2274 (2013).

53. Gao, Z.-Y. et al. Dissecting yield-associated loci in super hybrid rice *Oryza sativa* L. *Science* **351**, 922–927 (2015).

54. Ming, R. et al. The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* **47**, 1435–1442 (2015).

55. Albert, V. A. et al. *The Amborella* genome and the evolution of flowering plants. *Science* **342**, 1467 (2013).

56. Belser, C. et al. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* **4**, 879–887 (2018).

57. Haas, B. J. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).

58. Lowe, T. M. & Eddy, S. R. *RNAscan-SE*: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).

59. Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. * Infernal 1.0*: inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009).

60. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. *CAFE*: a computational tool for the study of gene family evolution. *Trends Genet.* **22**, 1269–1271 (2006).

61. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
62. Steinbiss, S., Willhoft, U., Gremme, G. & Kurtz, S. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res. 37, 7002–7013 (2009).
63. Liu, H. et al. The nearly complete genome of Ginkgo biloba illuminates gymnosperm evolution. Nat. Plants 7, 748–756 (2021).
64. Li, G. et al. A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. Nat. Genet. 53, 574–584 (2021).
65. Niu, S. et al. The Chinese pine genome and methylome unveil key features of conifer evolution. Cell 185, 204–217.e14 (2022).
66. Lee, T., Guo, H., Wang, X., Kim, C. & Paterson, A. H. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics 15, 162 (2014).
67. Zhang, C., Dong, S., Xu, J., He, W. & Yang, T. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 35, 1786–1788 (2018).
68. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. Genome Res. 20, 393–402 (2010).
69. Rektorisova, M., Hrbek, V., Jiru, M., Ovesna, J. & Hajslova, J. Variability in S-Alk(en)yl-L-cysteine sulfoxides in garlic within a seven-month period determined by a liquid chromatography - tandem mass spectrometry method. Plant Foods Hum. Nutr. 75, 376–382 (2020).

Acknowledgements
We thank the U.S. National Plant Germplasm System for providing the A. fistulosum accessions. We thank Mrs. Xiaodan Wu and Mei Fang from Analysis Center of Agrobiology and Environmental Science, Zhejiang University for their technical assistance in ACSOs quantifications, as well as Mrs. Quanfang Zhang from Agricultural Technology Extension Center of Fuyang District for providing the A. fistulosum var. viviparum accession "HZC". This work was supported by the China Agriculture Research System of MOF and MARA (CARS-24-B-01, Y-H.Z.) and the Fundamental Research Funds for the Central Universities (+226-2022-00100, M-F.Z.).

Author contributions
J-Q.Y., M-F.Z., and Y-H.Z. conceived of and designed the experiments. N.L., Z.H., X.H., Y.Z., L.K., and Y.W. performed genome assembly and assessment, comparative genome analysis, and other bioinformatic analyses; J-H.Y., X.L., Y-M.Z., G.D., Y-M.M., K.Z., Y-Y.M., M.A., and Y.X. conducted transcriptome assays and re-sequencing; J.M. and H.F. provided most of the A. fistulosum accessions; J.M., H.Y., and M.Z. performed field cultivation and phenotype investigation; N.L. and Z.H. wrote the draft manuscript; J-Q.Y., M-F.Z., and Y-H.Z. revised the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-34491-3.

Correspondence and requests for materials should be addressed to Yan-Hong Zhou, Ming-Fang Zhang or Jing-Quan Yu.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/. © The Author(s) 2022