THE PLANT CELL

*The Plant Cell*, 2025, **37**, koaf142
**https://doi.org/10.1093/plcell/koaf142**
Advance access publication 10 June 2025
**Research Article**

# Pancentromere analysis of *Allium* species reveals diverse centromere positions in onion and gigantic centromeres in garlic

Kiyotaka Nagaki,[1,*,†] Koichiro Ushijima,[2,†] Takashi Akagi,[2] Keisuke Tanaka,[3,‡] Hisato Kobayashi[3,§]

[1]Institute of Plant Science and Resources, Okayama University, Kurashiki 710-0046, Japan
[2]Graduate School of Environmental and Life Science, Okayama University, Okayama 710-8530, Japan
[3]NODAI Genome Research Center, Tokyo University of Agriculture, Setagaya 156-8502, Japan
*Author for correspondence: nagaki@rib.okayama-u.ac.jp
†These two authors contributed equally to this work.
‡Present address: Department of Informatics, Tokyo University of Information Sciences, Chiba, Japan.
§Present address: Department of Medical Genome Science, Dokkyo Medical University, Mibu, Japan.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plcell/pages/General-Instructions) is Kiyotaka Nagaki (nagaki@rib.okayama-u.ac.jp).

## Abstract

In eukaryotes, centromeres interact with the kinetochore for distribution of genetic information in cell division, yet their sequence and size are diverse among species. However, their position on chromosomes is considered to be conserved within a species. In this study, we analyzed the centromeres of 3 *Allium* species, namely, Welsh onion (*Allium fistulosum*), onion (*Allium cepa*), and garlic (*Allium sativum*) via pancentromere analysis and repetitive sequence analysis of centromeres and their neighborhoods and revealed their mobility, sequence organization, and size. Among the 3 species, Welsh onion and garlic had stable centromeres, but the onion centromere appeared to be polymorphic and frequently differed in position by up to 28.0 Mb among cultivars and between multiple individuals of the same cultivar. This mobility was stabilized by hybridization with Welsh onions. Furthermore, these 3 species have very different centromere sequence organization, including differences in the existence and maturity of centromeric satellites, and differences in centromere size, with Welsh onion having a centromere of 1.9 Mb, and garlic having a centromere of ~10.6 Mb, the largest of any organism with monocentric chromosomes analyzed to date. Our pancentromere analysis of these *Allium* species reveals the variation in sequence organization, size, and position of this important chromosomal region.

## Introduction

The centromere is a genetic position or the DNA sequence present at the position to distribute the genetic information encoded in the genome evenly to daughter cells during somatic cell division and meiosis (Choo 1997). The centromeric DNA sequence is the DNA sequence that interacts with the nucleosome containing the centromere-specific histone H3 variant (CENH3 which originally found as Centromere Protein A [CENP-A] in humans), which is the cornerstone of kinetochore formation (Richards and Dawe 1998; Jiang et al. 2003; Furuyama and Biggins 2007; Black and Bassett 2008; Samel et al. 2012; Comai et al. 2017; Sundararajan and Straight 2022; Naish and Henderson 2024). At the beginning of centromere analysis, centromere DNA sequences coexisting with centromere-specific histone H3 variants explored in a limited number of model organisms were often tandem repetitive DNA (satellite) sequences (Choo 1997; Jiang et al. 2003; Melters et al. 2013), leading to the belief that such repetitive sequences were necessary to stabilize the newly epigenetically formed centromere (neocentromere) (Wong and Choo 2001; Nagaki et al. 2004).

The development of next-generation sequencing technologies has made it easy and inexpensive to analyze the genome sequences of many species. For example, as of November 2024, the NCBI DNA database (https://www.ncbi.nlm.nih.gov/datasets/genome/) contains more than 18,000 eukaryotic reference genomes, of which more than 4,500 are assembled at the chromosome level and 270 at the telomere-to-telomere level. The data indicate that accurate reference genomes are no longer the prerogative of model organisms and that reference genomes are now available for species across many broad taxonomic groups. Centromere sequence data from such a wide range of species revealed that the stereotypical satellite-bearing centromere was not the only form of centromere (Nasuda et al. 2005; Shang et al. 2010; Gong et al. 2012; Nergadze et al. 2018; Oliveira and Torres 2018). In other words, stable centromeres do not always have satellite repeat sequences, and the relationship between centromere stability and the presence of satellite repeat sequences is uncertain.

More recently, examples of intraspecific shifts in centromeres, which are thought not to have moved within the same species except in special cases such as neocentromeres, have been reported (Rocchi et al. 2012; Schneider et al. 2016; Gent et al. 2017; Chen et al. 2023; Liu et al. 2023; Logsdon et al. 2024). In humans, a comparison of centromere positions in 2 reference genomes revealed

Received April 17, 2025. Accepted April 27, 2025.

discrepancies of up to 2.4 Mb (Logsdon et al. 2024). In soybean (*Glycine max*), pancentromere analysis via chromatin immunoprecipitation (ChIP)-seq using CENH3 revealed that centromere positions were polymorphic among cultivars, suggesting that centromeres were epigenetically shifted by up to 8 Mb (Liu et al. 2023). Furthermore, centromeres were also observed at different positions among individuals after 9 generations of hybridization, indicating that centromere shifts can occur at the interindividual level in the same genotype.

In the *Allium* genus, 1,077 species, including garlic (*Allium sativum*), onion (*Allium cepa*), and Welsh onion (*Allium fistulosum*), were listed in Plants of the World Online (http://www.plantsoftheworldonline.org/) as of November 2024. These species are native to the Northern Hemisphere, and some of them are used worldwide as spices, savory vegetables because of their distinctive aroma, and ingredients for medicine because they produce bactericidal substances (Fenwick et al. 1985; Jones et al. 2004; Swangsri et al. 2024). In addition, because of their large size and small number of chromosomes, these plants have long been used as educational materials for cytogenetics. Centromeric DNA sequences of these species are repetitive sequences present in the centromeric regions of all Welsh onion chromosomes and some onion chromosomes (Nagaki et al. 2012; Kirov et al. 2020). In addition, a reference genome assembled at the chromosome level, which was previously thought to be impossible because of its large genome size, genome sequences of the 3 species were recently released and available from GenBank (Hao et al. 2023). However, in this work, very few analyses have been performed on the centromeric regions of these species.

In this study, detailed analyses of Welsh onion (*Allium fistulosum*), onion (*Allium cepa*), and garlic (*Allium sativum*) were performed via tandem repeat analysis and pancentromere analysis using an anti-CENH3 antibody and 8 Welsh onion, 9 onion, and 2 garlic cultivars. In addition, frequently migrating centromeres in onion were analyzed at the individual level within each cultivar and using hybrids of Welsh onion and onion.

# Results

## Centromere mapping on Welsh onion chromosomes

To map centromeres, ChIP-seq was conducted via an anti-AfiCENH3 antibody, and the antibody recognizes CENH3 in Welsh onion, onion, and garlic and shows distinct centromeric signals on the chromosomes of these *Allium* species by immunostaining (Nagaki et al. 2012). ChIP was performed with this antibody and chromatin from 8 Welsh onion cultivars (Af1-8, Supplementary Table S1), and the immunoprecipitated DNA fragments were sequenced on the NovaSeq or MiSeq platform along with control DNA samples extracted from chromatin preparations prior to ChIP (input control). These reads were subsequently mapped onto a Welsh onion reference genome (ASM3073781v1) (Hao et al. 2023), and those mapped reads were counted every 100 kb bin window. The relative enrichment (RE = ChIP/Input) was calculated based on those numbers, and RE was plotted on the reference genome (Supplementary Fig. S1). In this study, centromeres were defined as regions showing RE more than 5, and 1 RE peak per chromosome was observed on all 8 Welsh onion chromosomes of the 8 cultivars.

The peaks indicate the locations of the centromeres in the reference genome, and the compositions of the repetitive sequences contained within or adjacent to the centromeres were investigated (Fig. 1). The CENH3 peaks were distributed on each chromosome, with sizes ranging from 0.9 Mb (Chr6) to 2.6 Mb (Chr2 and 5) and 1.86 Mb on average (Table 1 and Fig. 1), and the distributed regions overlapped with repetitive regions of tandem repeats. The tandem repeats presented variations in sequence and repeat unit length, and their variants were present in all centromeres while they formed tracts (Fig. 1 and Supplementary Figs. S2 to S4). The sequence similarity within the same tract was >98%, and that among tracts was >80%. The size of the repeat units ranged from 1,074 bp (in tract C on Chr5) to 4,452 bp (in tract A on Chr5), and they were either duplicate or deletion variants of the 2 kb units (located on the centromeres of almost all the chromosomes except Chr6) or combinations of them as higher-order repeats (Supplementary Fig. S4). The 2 kb unit, e.g. the subfamily present in tract A on Chr1 (AfCENchr1A), had a repeat unit of 2,088 bp (Supplementary Figs. S4 and S5) and had sequence similarity with previously reported Welsh onion centromeric repeats, Afi11, 19, and 54 (Nagaki et al. 2012) and AfCEN1 (Kirov et al. 2020). Since they are present in all the centromeres and encompass the Welsh onion centromeric repeats that have been reported as fragments, the Welsh onion centromeric repeats were named AfCEN.

To confirm if the CENH3 chromatins contain repetitive sequences other than AfCEN, the ChIP-seq reads were analyzed via the similarity-based de novo clustering program RepeatExplorer2 (http://www.repeatexplorer.org) (Novák et al. 2020). In this analysis, 9 clusters with REs of 4 or more were formed (Supplementary Fig. S6, A and B). Among the 9 clusters, AfCL33 showed homology to the CAT36 sequence, a previously reported pericentromeric tandem repeat of Welsh onion (Kirov et al. 2020), and the remaining 8 clusters all showed sequence similarity to AfCEN (Supplementary Fig. S5). The RE of CAT36 is considerably smaller than that of the AfCEN clusters but larger than that of the noncentromere control, 45SrDNA. CAT36 was repeated in the vicinity of the AfCEN tracts on Chr5 and 6, so it is possible that a small amount of CENH3 was distributed adjacent to the AfCEN tracts in CAT36 (Fig. 1 and Supplementary Fig. S2A).

## Centromere positions in onion differ among cultivars and individuals

Next, the ChIP-seq reads of 6 onion (Ac1–6) and 3 shallot (an onion variant, *A. cepa* L. var. aggregatum, Ac7–9) cultivars (Supplementary Table S1) were mapped to the onion reference genome (GCA_030765085.1) (Hao et al. 2023) (Fig. 2A and Supplementary Fig. S7). To determine the locations of ancient centromeres in *Allium*, onion chromosomes and their homologous Welsh onion chromosomes were aligned, and CENH3-distributed regions were mapped to these chromosomes (Supplementary Fig. S7). Onion and Welsh onion chromosome sequences were relatively conserved and could be aligned in almost all regions, and onion and Welsh onion centromeres were mapped to homologous regions on the alignments. However, multiple CENH3 peaks per chromosome in onion including shallot are often observed, even within a single cultivar, and their locations vary among cultivars (Fig. 2 and Supplementary Figs. S2 and S7). CENH3-distributed regions were observed in many cultivars at the position of the ancient centromere, but in many cultivars, the CENH3 distribution regions shifted from the ancient centromeres. Even when CENH3 peaks remained in ancient positions, other CENH3 peaks were present within the same individual in many cultivars. Double peaks observed on 1 chromosome may be in an allelic relationship as observed in human chromosomes (Logsdon et al. 2024). For some chromosomes, 3 or more multiple peaks were observed. These cases may have a single

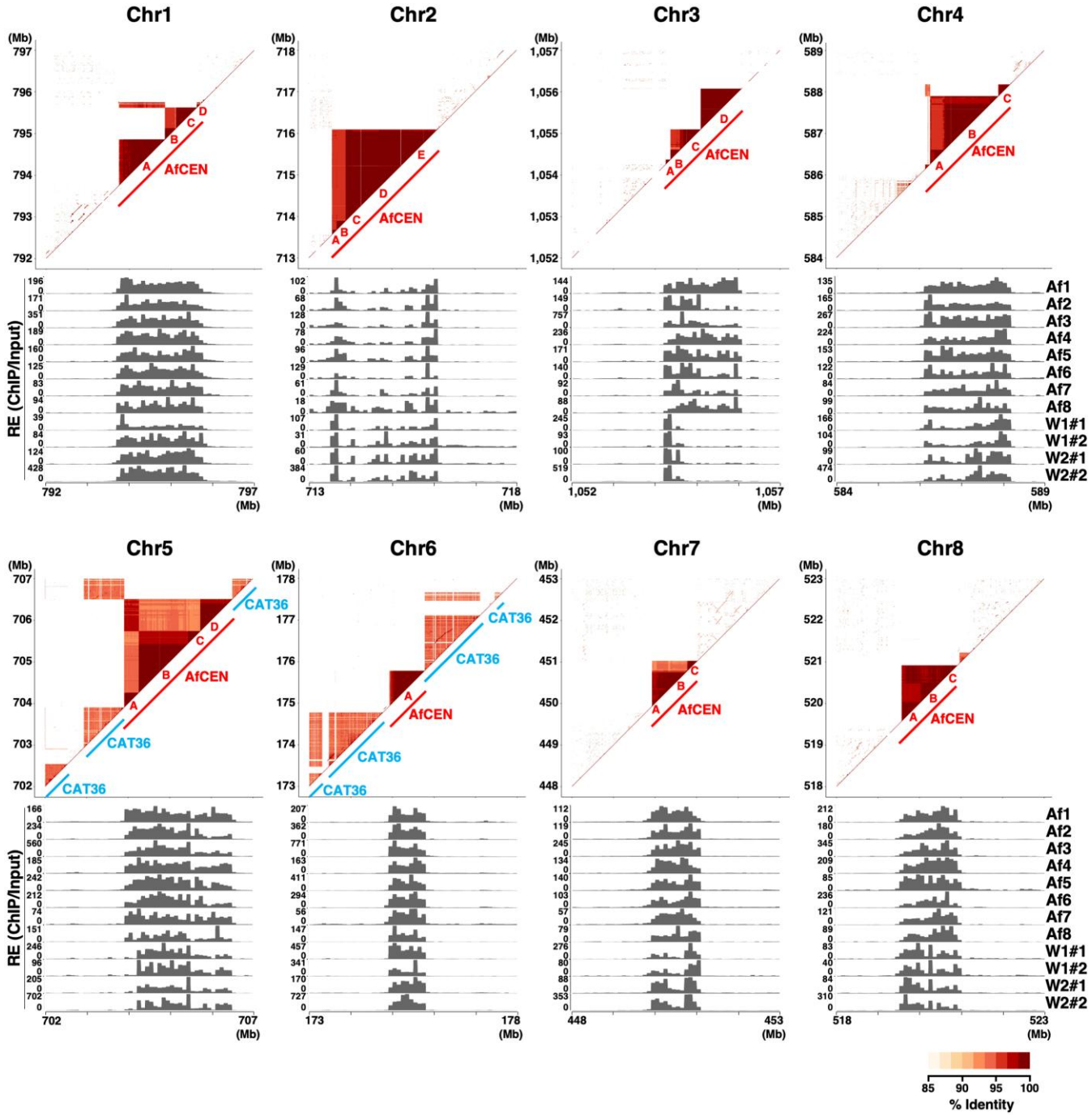


**Figure 1.** Mapping of CENH3 and identification of tandem repeats in Welsh onion centromeres. Centromeres of 8 Welsh onion (Af: *A. fistulosum*) and 2 Wakegi (W: *A.* × *proliferum*) cultivars were identified via ChIP-seq (RE=ChIP/Input) mapping of CENH3 on the chromosomes (Chr). Tandem repeats in the 5 Mb regions around those CENH3 peaks were analyzed using the moddotplot. The detected tandem repeat tracts **(A** to **E)** were labeled onto the moddotplot results.

centromere-like structure composed of multiple CENH3 peaks reported in other plants (Liu et al. 2023; Dias et al. 2024). Migration of the CENH3 peaks to the same position was observed among the different cultivars, indicating that there may be preferred regions where the CENH3 peaks migrate (Gent et al. 2017). When the position of the ancient centromere was used as a reference point, the distance from the point to the migrated centromere reached 28.0 Mb on Chr4 (Fig. 2, A and B). In addition, the distribution of CENH3 was investigated between 2 individuals of 7 cultivars (Ac2, 3, 4, 5, 7, 8, and 9, Fig. 2A and Supplementary Figs. S2 and S7). No differences were detected between the individuals in Ac7, but polymorphisms in CENH3-distributed regions were detected between individuals of the same cultivar in Ac2, 3, 4, 5, 8, and 9. These data suggest that onion CENH3-distributed regions are highly polymorphic and undergo frequent shifts. The CENH3 peaks ranged from 0.8 to 10.4 Mb in size (Av. 5.0 Mb), and the peaks including the ancient positions were larger (Av. 5.8 Mb) than those in the new positions (Av. 4.1 Mb) (Fig. 2C). Although both epigenetic shifts and structural rearrangements are possible causes of these shifts in the CENH3 peak, the shifts frequently observed between individuals

**Table 1.** Positions and sizes of CENH3-distributed regions and AfCEN arrays in the Welsh onion (*A. fistulosum*) genome

| Chromosome number | Chromosome size (Mb) | Centromere location (Mb) | Centromere size (Mb) | AfCEN arrays (Mb) |
|---|---|---|---|---|
| 1 | 1,690 | 793.7 to 795.8 | 2.1 | 2.0 |
| 2 | 1,625 | 713.5 to 716.1 | 2.6 | 2.4 |
| 3 | 1,607 | 1,054.2 to 1,056.1 | 1.9 | 1.8 |
| 4 | 1,405 | 586.1 to 588.2 | 2.1 | 2.0 |
| 5 | 1,386 | 703.9 to 706.5 | 2.6 | 2.3 |
| 6 | 1,173 | 175.0 to 175.9 | 0.9 | 0.8 |
| 7 | 1,157 | 449.8 to 451.1 | 1.3 | 1.3 |
| 8 | 1,127 | 519.5 to 521.0 | 1.5 | 1.5 |

of the same cultivar suggest that these shifts are more likely to be epigenetic shifts than structural rearrangements in the centromere region where recombination is suppressed.

The repetitive sequences within or around onion centromeres were investigated (Supplementary Fig. S2). Unlike the Welsh onion centromeres, the onion centromeres did not contain highly (>90% similarity) conserved tandem repeats, with the exception of Chr5. Sprouting tandem repeats with low conservation (<90% similarity) were, however, observed in some centromeres. A repeat sequence with a 1.8 kb repetitive unit was repeated on the centromeres of Chr1, 4, 6, and 8, and the sequences were partially similar to those reported for AcCEN1K (an ortholog of AfCEN1K; Kirov et al. 2020; Supplementary Fig. S8). Additionally, a 1.5 kb sequence with reconstructed fragments of this sequence was repeated on Chr2. These repeats with sequences similar to AcCEN1K were collectively referred to as AcCEN. The AcCEN tracts and CENH3-distributed regions overlapped on Chr1, 2, 4, and 6 but not on Chr8 (Fig. 2A). In addition to AcCEN, a repeat sequence not similar to AcCEN with a 1,975 bp repeat unit (named CentAc) was repeated on Chr3, 5, and 7 and colocalized with the CENH3 peaks.

To confirm if the CENH3 chromatins contain repetitive sequences other than AcCEN and CentAc, the ChIP-seq reads were analyzed by RepeatExplrer2 as Welsh onion via Ac4 ChIP-seq reads. In this analysis, 7 clusters with REs of 35 or more were formed (Supplementary Fig. S6, C and D). Among the 7 clusters, AcCL28 showed homology to AcCEN sequences, and the remaining 6 clusters all showed homology to CentAc. Those CentAc clusters presented a greater RE than did AcCEN. BLAST searches using those CentAc sequences as queries revealed chromosome specificity (AcCL16 for Chr3, AcCL62, −72, and −81 for Chr5, AcCL67 for Chr7, and AcCL75 for Chr5), and budding tandem repeat units were found in those BLAST hit regions (CentAcChr3: 3,605 bp, CentAcChr5: 1,975 bp, and CentAcChr7: 1,780 bp). The CENH3 peaks containing and not containing those repeats were examined for size differences, and the peak containing CentAc was larger than the peaks containing AcCEN and not containing those tandem repeats (Fig. 2D).

## Onion and Welsh onion chromosomes in Wakegi (*Allium* × *proliferum*) show a single CENH3 peak on each chromosome

In this study, Welsh onions presented uniform centromere positions, whereas onions presented cultivar differences in centromere positions (Figs. 1 and 2 and Supplementary Figs. S1 and S7). What will be the mobility of centromeres in the hybrids of those species? Wakegi (*Allium* × *proliferum*) are natural diploid hybrids of Welsh onion and shallot (Friesen and Klaas 1998). These species are the results of ancient hybridizations and, because of their sterility, have since been maintained by vegetative reproduction. Two independent lineage Wakegi (W1: Japanese and W2: Taiwanese) (Sugiharto Arifin et al. 2000) were used for ChIP-seq, and the ChIP-seq reads of 2 individuals from those cultivars of Wakegi (W1#1 and 2 and W2#1 and #2) were mapped to the Welsh onion and onion reference genomes (Figs. 1 and 2 and Supplementary Figs. S1 and S7). Since Wakegi are hybrid diploids of onion and Welsh onion, it possesses 1 set of onion and Welsh onion chromosomes each. One CENH3 peak per chromosome was observed on onion chromosomes in the 2 independent cultivars of Wakegi. However, the CENH3 peak of Wakegi on Chr2 was anchored to a different part of its broad distribution on Chr2 in shallot. No polymorphisms in CENH3 distribution were observed among individuals of these cultivars, suggesting that CENH3 movement was suppressed in Wakegi after hybridization with Welsh onion. These results suggest that the Chr2 peaks were fixed during independent hybridization events of the 2 cultivars. For the Welsh onion chromosomes in Wakegi, the distribution was similar to that of all the Welsh onion cultivars for almost all the chromosomes except Chr3, 5, and 7 (Fig. 1 and Supplementary Fig. S1). On Chr3, CENH3 was distributed in all 4 AfCEN tracts in the Welsh onion, but there was only a major peak in tract A and a minor peak in tract B in Wakegi. On Chr5, its peak disappeared from tract A, and on Chr7, its peak was biased toward tracts A and C. Although the hybridization timings of Wakegi are unknown, these data suggest that the position of centromeres within Wakegi has been maintained relatively stable, but not completely, by vegetative reproduction.

## Garlic possesses giant centromeres

Subsequently, the ChIP-seq reads of 2 garlic cultivars (As1 and 2) were mapped to the garlic reference genome (GCA_030737875.1; Hao et al. 2023; Fig. 3 and Supplementary Fig. S9). First, the centromeres were mapped to the alignment of the garlic chromosome and the homologous chromosome of the Welsh onion to determine whether there was migration of the garlic centromere from the positions of the ancient centromeres of *Allium* (Supplementary Fig. S9). Alignment was possible in many regions, although the homology between garlic chromosomes and Welsh onion chromosomes was lower than that between onion chromosomes and Welsh onion chromosomes. However, except for Chr4, the centromeric regions had gaps. The centromere positions of both species intersected on the alignment on Chr4, but on Chr1, 2, 3, and 5, the intersections were in the gaps, and the centromeres were mapped on different ends of the gap on Chr6 to 8. These results indicate that the centromeres of both species essentially remain at their ancient centromere positions but that structural changes, including insertions, that have occurred in the centromeres make their relative positions difficult to determine.

The CENH3-distributed regions in garlic were essentially 1 peak on each chromosome, but 3 peaks of 1.0, 3.4, and 7.8 Mb were found on Chr7 (Table 2 and Fig. 3 and Supplementary Fig. S9). The distances between these 3 peaks were 1.4 and 1.0 Mb, making it difficult to determine whether the centromere should be considered 14.6 Mb, including the distance between the peaks, 12.2 Mb simply by adding the 3 peaks, or 7.8 Mb, the largest peak. Considering that Chr7 was excluded, the size of the garlic CENH3-distributed regions ranged from 8.8 (Chr6) to 13.2 (Chr2) Mb, with an average size of 10.5 Mb (Table 2 and Fig. 3). Their sizes

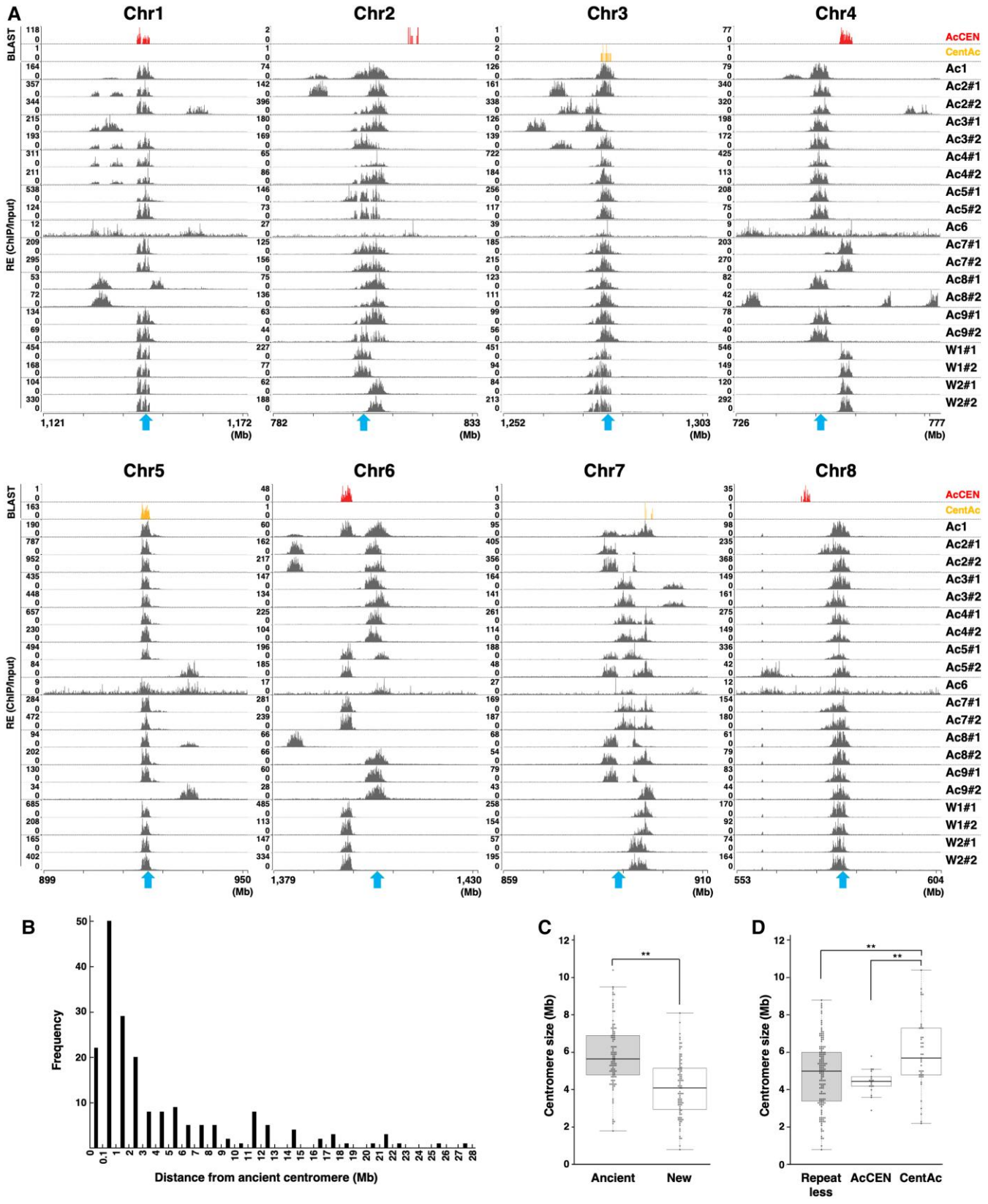


**Figure 2.** Distribution of CENH3 and centromere-specific repetitive sequences in onion centromeres. **A)** Centromeres of 9 onion and 2 Wakegi cultivars were identified via ChIP-seq (RE=ChIP/Input) mapping of CENH3 on the chromosomes (Chr). Two centromere-specific repetitive sequences detected by the moddotplot analysis (see Supplementary Fig. S2) were mapped via BLAST to the same region. The upward arrows indicate ancient centromeres identified via Harplot analysis (see Supplementary Fig. S7). **B)** Distance of the detected CENH3 peaks from the ancient centromere on the same chromosome ($n = 196$). Sizes of centromeres at ancient or new locations **C)** that were colocalized with or without repetitive sequences **D)**. Center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; data points are plotted as dots. $n = 104$ (Ancient in **C)**, 92 (New in **C)**, 137 (Repeat less in **D)**, 18 (AcCEN in **D)**, and 41 (CentAc in **D)** sample points. Significance according to the *t*-test (2-sided) is indicated in the plots (**$P < 0.01$).

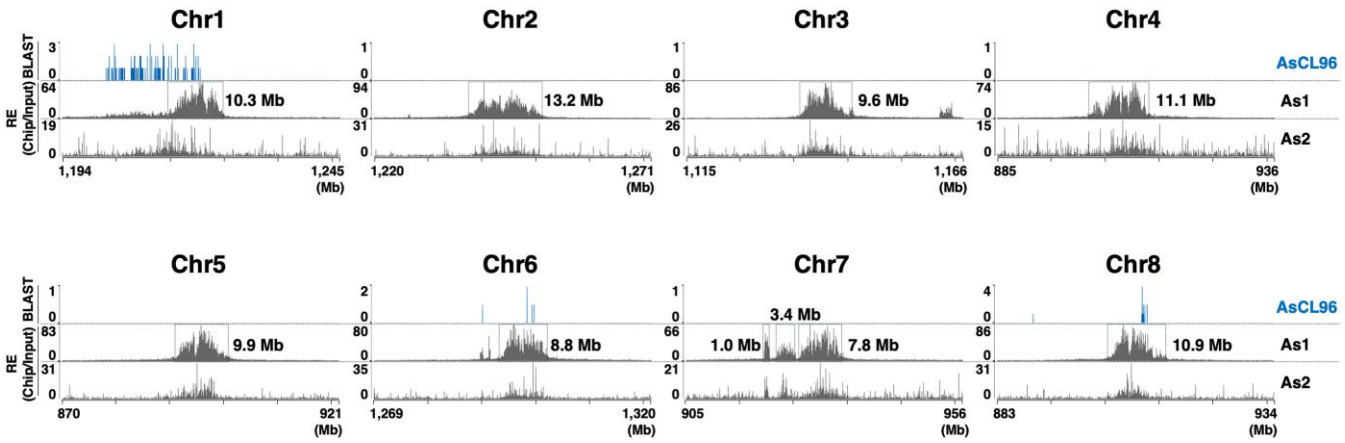


**Figure 3.** Distribution of CENH3 and a repetitive sequence in the garlic centromeres. Centromeres of 2 garlic cultivars were identified via ChIP-seq (RE=ChIP/Input) mapping of CENH3 on the chromosomes (Chr). The dotted squares and numbers indicate centromere locations and sizes, respectively. Repetitive sequence (AsCL96) that was preferentially present in the centromere identified via RepeatExplorer2 analysis (see Supplementary Fig. S6E and F) was mapped via BLAST to the same region.

**Table 2.** Positions and sizes of CENH3-distributed regions in the garlic (*A. sativum*) genome

| Chromosome number | Chromosome size (Mb) | Centromere location (Mb) | Centromere size (Mb) |
|---|---|---|---|
| 1 | 2,424 | 1,213.5 to 1,223.8 | 10.3 |
| 2 | 2,273 | 1,237.5 to 1,250.7 | 13.2 |
| 3 | 2,194 | 1,136.1 to 1,145.7 | 9.6 |
| 4 | 1,944 | 902.0 to 913.1 | 11.1 |
| 5 | 1,913 | 891.0 to 900.9 | 9.9 |
| 6 | 1,813 | 1,292.3 to 1,301.1 | 8.8 |
| 7 | 1,583 | 919.4 to 920.4 | 1.0 |
| | | 921.8 to 925.2 | 3.4 |
| | | 926.2 to 934.0 | 7.8 |
| 8 | 1,379 | 903.4 to 914.3 | 10.9 |

are more than 5.5 times larger than those of Welsh onion and twice as large as the sizes of the centromeres of bread wheat (1.9–8.1 Mb, Av. 5.0 Mb), the largest centromeres on monocentric chromosomes measured by ChIP-Seq to date (Su et al. 2019). Onion CENH3 peak sizes (0.8–10.4 Mb and Av. 5.0 Mb) range between Welsh onion and garlic, but it is notable that there is an approximately 5.5-fold difference in centromere size between closely related species, Welsh onion and garlic. To test the acceptability of garlic giant centromeres, the compatibility of these centromere sizes with a proposed relationship concerning centromere size was confirmed. Bennett et al. (1981) found a proportional relationship between centromere volume and genome size as measured by electron microscopy in plant species. Later, this relationship was revalidated using immunostaining with anti-CENH3 antibodies and ChIP-seq data, and was shown to be applicable to eukaryotes other than plants (Zhang and Dawe 2012; Plačková et al. 2021). Based on the relationship between centromere size and genome size, regression analysis was performed on recently published solid reference genome sequences and CENH3 distribution data for *Arabidopsis* (Naish et al. 2021), rice (Shang et al. 2023), maize (Chen et al. 2023), soybean (Liu et al. 2023), and bread wheat (Su et al. 2019), confirming the compatibility of garlic and Welsh onion data (Supplementary Fig. S10). The analysis revealed that the centromere size of garlic revealed in this study conforms to the relationship; rather, the centromere size of Welsh onion is an outlier.

Furthermore, the repetitive sequence organizations within the garlic centromeres and their surrounding regions were investigated (Supplementary Fig. S2). In a repeat sequence survey, garlic sequences presented higher overall sequence similarity than Welsh onion and onion sequences did, but no clear tandem repeats were found in the garlic centromeres or their surrounding regions. Spots of >95% sequence similarity were observed on Chr1, which partially overlapped with the CENH3-distributed region, although not in a tandem repetitive fashion. RepeatExplorer analysis was then performed to look for repeat sequences involved in CENH3 chromatin (Supplementary Fig. S6, E and F). Compared with the noncentromeric control, 45SrDNA (AsCL53, RE = 0.6) was present in 1 cluster (AsCL96), which presented an RE = 24.9, and it accumulated in the ChIP. RepeatExplorer classified this cluster as Tekay, and BLAST with this sequence revealed a preference for the centromere and matched the spots on Chr1 (Fig. 3 and Supplementary Fig. S9). These results may indicate the initial stages of AsCL96 maturation as a centromeric sequence.

# Discussion

## Centromere stability

Because centromere positions have long been used as landmarks for karyotypic analysis, these positions have been considered immovable within the species, except in special cases, such as neocentromeres (Rocchi et al. 2012). However, recent advances in NGS technology have made it possible to obtain reference genomes and perform pancentromere analysis, and examples of shifted centromeres have been reported in several species (Schneider et al. 2016; Gent et al. 2017; Liu et al. 2023; Logsdon et al. 2024). In humans, centromere shifts of up to 2.4 Mb were detected in some chromosomes between the 2 reference genomes (Logsdon et al. 2024). In soybean, up to 8 Mb centromere shifts have been observed between cultivars and during hybrid hindgrowth, and these data suggest frequent centromere shifts (Liu et al. 2023). The present study revealed differences in centromere mobility among 3 closely related species of *Allium* (Figs. 1 to 3 and Supplementary Figs. S1, S2, S7, and S9). Welsh onion and garlic centromeres remained stable, with 1 CENH3 peak/chromosome (Figs. 1 and 3 and Supplementary Figs. S1, S2, and S9). In contrast, multiple CENH3 peaks in onion were found on the same chromosome, and their positions differed among cultivars and among

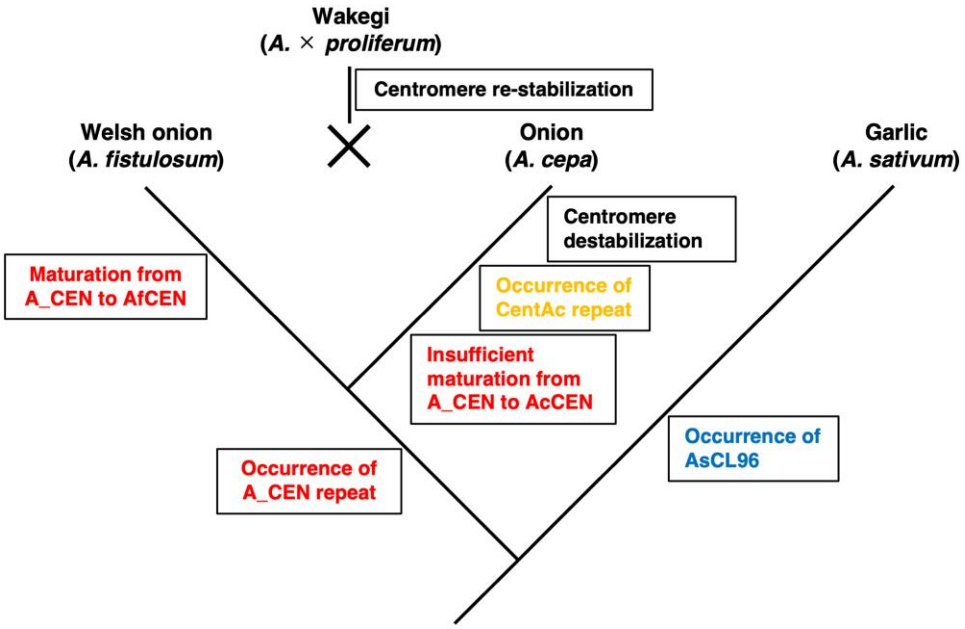


**Figure 4.** Summary diagram of the evolution of centromeric repetitive sequences and changes in the mobility of CENH3 in *Allium* species based on the results of this study. Events based on the centromere sequence composition and mobility revealed in this study are presented on the phylogenetic tree proposed in a previous paper (Hao et al. 2023). The 3 events that occurred on the onion branches were in no particular order.

multiple individuals of the same cultivar up to 28.0 Mb (Fig. 2 and Supplementary Figs. S2 and S7). Additionally, in Wakegi, hybrids of Welsh onion and onion, these multiple peaks disappeared, resulting in 1 peak/chromosome. The change in centromere mobility among closely related species and their hybrids is reminiscent of the loss of centromere stabilizing factors in onions and their resupply in the hybrid (Fig. 4). What factors are involved in these shifts in centromeres? In chickens, centromere drift occurs due to the loss of CENP-U and CENP-S, but this range remains below 500 kb (Hori et al. 2017). However, this shift is small compared with the shifts observed in *Allium* species and soybean (Liu et al. 2023) and only fluctuates within the same peak. Furthermore, since the kinetochore complexes in plants do not involve CENP-U or CENP-S (Naish and Henderson 2024), other factors may be associated with large centromere shifts in those plants. One possible candidate is the loading factor of CENH3. Recently, a CENH3 loading factor was identified in *Arabidopsis*, suggesting that it may be related to de novo deposition (Takeuchi et al. 2024). Another factor associated with centromere migration may be reproductive mode. The onion (Ac7) and Wakegi (W1 and W2) for which no polymorphism was observed between 2 individuals of the same cultivar (Fig. 2 and Supplementary Figs. S2 and S7) are all varieties maintained by vegetative propagation (Supplementary Table S1), and centromere migration may occur only during sexual reproduction. Further research, including genetic analysis, is needed to clarify these issues.

## Centromeric DNA sequences

In the early days of centromere analysis, tandem repetitive DNA sequences and retrotransposons that localize specifically to centromeres were reported in mammals and plants (Maluszynsak and Heslop-Harrison 1991; Choo 1997; Nagaki et al. 1998, 2003; Cheng et al. 2002; Jiang et al. 2003; Iwata et al. 2013; Melters et al. 2013; Comai et al. 2017; Zhang et al. 2017; Wlodzimierz et al. 2023; Naish and Henderson 2024). Often, the repeat units of those tandem repetitive DNA sequences are a few hundred base pairs, and their length is similar to the length of the DNA orbiting the nucleosome, which often favors such repeat units in the periodic arrangement of centromeric nucleosomes (Henikoff et al. 2001; Hasson et al. 2013; Melters et al. 2013; Naish et al. 2021; Liu et al. 2023; Tek et al. 2024). In this study, Welsh onions were shown to have AfCEN tandem repeats that roughly match the distribution of CENH3 (Fig. 1 and Supplementary Fig. S1). However, its basic unit size was ~2 kb, and furthermore, its size varied within chromosomes and between inter-repeating tracts (Supplementary Figs. S3 and S4).

In addition, 2 more immature tandem-like repeats were detected at and near the centromere in onion (Fig. 2 and Supplementary Fig. S8). Some repetitive regions of these sequences were distributed in a head-to-tail orientation, but in many cases, there were other sequences between adjacent units, or the sequence similarity to adjacent units was lower than that in the AfCEN case. One of those 2 tandem-like sequences, AcCEN, showed sequence similarity to AfCEN. With respect to colocalization with CENH3, the position of CENH3 varied among cultivars, so these tandem-like sequences and CENH3 peaks sometimes coincided and sometimes did not (Fig. 2). Furthermore, no tandem repeat sequences localized to centromeres were found in garlic (Fig. 3). These results suggest that, in *Allium* species, the ancestral repetitive sequence (A_CEN) of AfCEN and AcCEN appeared in the centromeres after the divergence of garlic from the other 2 species and may still be involved in the process of forming repetitive sequences in onion, whereas in Welsh onion, they become complete tandem repetitive sequences (Fig. 4). Considering this situation of tandem repeat sequence maturation together with the difference in centromere stability between the 2 species, 2 possibilities exist: one is the conventional idea that mature tandem repetitive sequences suppress centromere migration, and the other is that frequent centromere migration suppresses tandem repetitive sequence maturation. Furthermore, the fact that garlic has no satellite-type repeat sequences at all (Fig. 3) but has stable

centromeres (Fig. 3 and Supplementary Figs. S2 and S9) calls into question the need for tandem-type repeat sequences for centromere stability.

Centromere-specific retrotransposons have been reported in many plant species (Jiang et al. 2003; Nagaki and Murata 2005; Nagaki et al. 2011, 2015; Fawcett et al. 2023; Naish and Henderson 2024). Among them, gypsy-type LTR retrotransposons, called CRs, are present specifically in the centromeric regions of several plant species (Jiang et al. 2003; Nagaki and Murata 2005; Fawcett et al. 2023; Naish and Henderson 2024; Tsukahara et al. 2025). Among *Allium* species, CR orthologs have also been reported in onion and Welsh onion (Kiseleva et al. 2014). In this study, de novo clustering analysis revealed that the sequences colocalized with CENH3 in Welsh onion and onion were mainly tandem repeat sequences and did not show large-scale colocalization of retrotransposons (Supplementary Fig. S6). Additionally, clusters containing retrotransposon-like sequences that coexist with CENH3 in garlic did not show homology with the CR. The signals in the in situ hybridization images in the original paper also appeared to be localized to pericentromeric regions rather than centromeric locations (Kiseleva et al. 2014). Based on these results, it is likely that the CRs reported in these *Allium* species are not true CRs and may need to be revisited, including checking their distribution within these reference genomes.

### Centromere size

The centromere size was considered to be in the range of several hundred kilo base pairs to 1 Mb because model organisms with small genome sizes and artificial chromosomes were used as research materials (Choo 1997; Nagaki et al. 2003, 2004). In contrast, Bennett et al. (1981) proposed the relationship that the sum of centromere sizes is proportional to genome size. In this study, the relationship was reexamined using recently obtained fine mapping data of CENH3 onto accurate reference genomes and Welsh onion and garlic data (Supplementary Fig. S10). The results showed that the garlic centromere, which appeared to be gigantic, fit the relationship; rather, the Welsh onion centromere was found to be smaller than predicted. The garlic centromere size (Av. 10.6 Mb) mapped in this study is the largest reported to date. If the centromere size is based on the relationship, it is possible that species with larger genomes and fewer chromosomes may have larger centromeres than garlic. The relationship may be based on the following factors: (i) larger centromeres are needed to move larger chromosomes, and (ii) within a species, centromere drives proposed by Henikoff (Henikoff et al. 2000; Wang et al. 2014) or defects in mitosis, such as those observed in Indian muntjac (*Muntiacus muntjak*) (Drpic et al. 2018), occur if centromere sizes are not equalized.

## Materials and methods

### Plant materials

The cultivars of *Allium* species, Welsh onion (*A. fistulosum*), onion (*A. cepa*), shallot (*A. cepa L. var. aggregatum*), Wakegi (*A. ×proliferum*), and garlic (*A. sativum*) used in this study are listed in Supplementary Table S1. For the Welsh onion, plants that were germinated from seeds and grown in a greenhouse or purchased at markets were used in the study. In the case of onion and shallot, plants were germinated from seeds and grown in a greenhouse, seedlings were purchased from seed companies, and leaves grown from bulbs were used in the study. In the case of garlic, leaves grown from the bulb were used in this study. All the greenhouse growth was conducted under natural light.

### Chromatin immunoprecipitation-Seq

ChIP using the anti-AfiCENH3 antibody (Nagaki et al. 2012) was performed as previously described (Nagaki et al. 2015) with minor modifications. Nuclei were isolated individually from 1–3 g of plant leaves and then digested with micrococcal nuclease (Sigma–Aldrich, St. Louis, MO, USA) to produce monomeric chromatin or were sonicated by a Nanoruptor (Diagenode, Denville, NJ, USA) to produce 2–4 mer chromatin. Following incubation of the chromatin with the antibody, the antibody was captured using Dynabeads Protein A (Invitrogen, Carlsbad, CA, USA). DNA was purified from the chromatin with the captured antibodies via phenol/chloroform extraction followed by ethanol precipitation. ChIP-Seq was conducted using precipitated DNA from the input and ChIP fractions. Libraries were constructed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA), and the libraries were read via MiSeq (Illumina), NovaSeq 6000 (Illumina), or NovaSeq X plus (Illumina) with the paired-end 2 × 150 bp protocol. Chip-seq reads were mapped to the reference genome sequence of the *Allium* genus (Hao et al. 2023) using Chromap (Zhang et al. 2021). The coverage and ratio to the input were estimated every 100 kb using Deeptools (Ramírez et al. 2016). The map was plotted with pyGenometracks (Lopez-Delisle et al. 2021).

### Detection of tandem repeats

The detection of tandem repeats around the centromere was carried out using the moddotplot (Sweeten et al. 2024). The sequences around the peaks of the ChIP-seq data were obtained using seqkit (Shen et al. 2024) and subjected to moddoplot with the default settings (85% sequence identity) in the static mode in the case of Welsh onion. Since onion and garlic centromeres have no obvious tandem repeats, the -id option was used to set the lower limit of 80% sequence identity. The static mode sets a minimum window size based on the sequence length as $n/1000$. The matrix with 1,000 × 1,000 sequence pairs was plotted.

### RepeatExplorer analysis

The sequence data were analyzed via the similarity-based de novo clustering program RepeatExplorer2 (http://www.repeatexplorer.org) (Novák et al. 2020). Clusters of repetitive sequences were generated via comparative analysis, an advanced option in RepeatExplorer2 clustering.

### Statistical analysis

Regression analysis on centromere size and genome size was performed by Microsoft Excel using recently published accurate reference genome sequences and CENH3 distribution data for *Arabidopsis* (Naish et al. 2021), rice (Shang et al. 2023), maize (Chen et al. 2023), soybean (Liu et al. 2023), and bread wheat (Su et al. 2019).

## Author contributions

K.N. conceived the study, conducted the ChIP-seq experiment, produced the figures, and wrote a draft of the manuscript. K.T. and H.K. read the sequence. K.U. and T.A. mapped the ChIP-seq data to the reference genome, detected tandem repeat sequences in the reference genome, and drew figures on them. Based on the draft, K.N. and K.U. wrote the manuscript.

## Supplementary data

The following materials are available in the online version of this article.

**Supplementary Table S1.** Information on the plants used in this study.

**Supplementary Figure S1.** Genome-wide mapping of CENH3 and a centromere-specific repetitive sequence in the Welsh onion.

**Supplementary Figure S2.** Analysis of centromeric repetitive sequences of the 3 *Allium* species.

**Supplementary Figure S3.** Intercomparison of repetitive sequences on each centromere in the Welsh onion.

**Supplementary Figure S4.** Comparison of the similarity of AfCEN sequences among each tract.

**Supplementary Figure S5.** Relationships of full-length AfCEN sequences with sequences in previous reports and clusters generated via RepeatExplorer2 analysis.

**Supplementary Figure S6.** RepeatExplorer2 analysis of ChIP-seq data from *Allium* species.

**Supplementary Figure S7.** Comparative analysis of onion chromosomes and Welsh onion chromosomes with sequence similarity and identification of centromere positions.

**Supplementary Figure S8.** Relationship of AcCEN sequences to sequences in a previous report and a cluster generated via RepeatExplorer2 analysis.

**Supplementary Figure S9.** Comparative analysis of garlic chromosomes and Welsh onion chromosomes with sequence similarity and identification of centromere positions.

**Supplementary Figure S10.** Correlation analysis of the genome size and the sum of centromere sizes based on accurate genome and centromere analysis.

## Funding

This research was supported by the Cooperative Research Grant of the Genome Research for BioResources, NODAI Genome Research Center, Tokyo University of Agriculture (number H25-12).

*Conflict of interest statement*. None declared.

## Data availability

The anti-CENH3 ChIP-seq and input data are submitted to the DNA Data Bank of Japan (DDBJ) under accession numbers DRR618576 to DRR618623 and DRR656881 to DRR656895.

## References

Bennett MD, Smith JB, Ward J, Jenkins G. The relationship between nuclear DNA content and centromere volume in higher plants. *J Cell Sci*. 1981:47(1):91–115. https://doi.org/10.1242/jcs.47.1.91

Black BE, Bassett EA. The histone variant CENP-A and centromere specification. *Curr Opin Cell Biol*. 2008:20(1):91–100. https://doi.org/10.1016/j.ceb.2007.11.007

Chen J, Wang Z, Tan K, Huang W, Shi J, Li T, Hu J, Wang K, Wang C, Xin B, et al. A complete telomere-to-telomere assembly of the maize genome. *Nat Genet*. 2023:55(7):1221–1231. https://doi.org/10.1038/s41588-023-01419-6

Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell*. 2002:14(8):1691–1704. https://doi.org/10.1105/tpc.003079

Choo KHA. *The centromere*. New York: Oxford University Press; 1997.

Comai L, Maheshwari S, Marimuthu MPA. Plant centromeres. *Curr Opin Plant Biol*. 2017:36:158–167. https://doi.org/10.1016/j.pbi.2017.03.003

Dias Y, Mata-Sucre Y, Thangavel G, Costa L, Báez M, Houben A, Marques A, Pedrosa-Harand A. How diverse a monocentric chromosome can be? Repeatome and centromeric organization of Juncus effusus (Juncaceae). *Plant J*. 2024:118(6):1832–1847. https://doi.org/10.1111/tpj.16712

Drpic D, Almeida AC, Aguiar P, Renda F, Damas J, Lewin HA, Larkin DM, Khodjakov A, Maiato H. Chromosome segregation is biased by kinetochore size. *Curr Biol*. 2018:28(9):1344–1356.e5. https://doi.org/10.1016/j.cub.2018.03.023

Fawcett JA, Takeshima R, Kikuchi S, Yazaki E, Katsube-Tanaka T, Dong Y, Li M, Hunt HV, Jones MK, Lister DL, et al. Genome sequencing reveals the genetic architecture of heterostyly and domestication history of common buckwheat. *Nat Plants*. 2023:9(8):1236–1251. https://doi.org/10.1038/s41477-023-01474-1

Fenwick GR, Hanley AB, Whitaker JR. The genus allium—part 1. *Crit Rev Food Sci Nutr*. 1985:22(3):199–271. https://doi.org/10.1080/10408398509527415

Friesen N, Klaas M. Origin of some minor vegetatively propagated Allium crops studied with RAPD and GISH. *Genet Resour Crop Evol*. 1998:45(6):511–523. https://doi.org/10.1023/A:1008647700251

Furuyama S, Biggins S. Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proc Natl Acad Sci U S A*. 2007:104(37):14706–14711. https://doi.org/10.1073/pnas.0706985104

Gent JI, Wang N, Dawe RK. Stable centromere positioning in diverse sequence contexts of complex and satellite centromeres of maize and wild relatives. *Genome Biol*. 2017:18(1):121. https://doi.org/10.1186/s13059-017-1249-4

Gong Z, Wu Y, Koblízková A, Torres GA, Wang K, Iovene M, Neumann P, Zhang W, Novák P, Buell CR, et al. Repeatless and repeat-based centromeres in potato: implications for centromere evolution. *Plant Cell*. 2012:24(9):3559–3574. https://doi.org/10.1105/tpc.112.100511

Hao F, Liu X, Zhou B, Tian Z, Zhou L, Zong H, Qi J, He J, Zhang Y, Zeng P, et al. Chromosome-level genomes of three key Allium crops and their trait evolution. *Nat Genet*. 2023:55(11):1976–1986. https://doi.org/10.1038/s41588-023-01546-0

Hasson D, Panchenko T, Salimian KJ, Salman MU, Sekulic N, Alonso A, Warburton PE, Black BE. The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat Struct Mol Biol*. 2013:20(6):687–695. https://doi.org/10.1038/nsmb.2562

Henikoff S, Ahmad K, Malik HS. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*. 2001:293(5532):1098–1102. https://doi.org/10.1126/science.1062939

Henikoff S, Ahmad K, Platero JS, van Steensel B. Heterochromatic deposition of centromeric histone H3-like proteins. *Proc Natl Acad Sci U S A*. 2000:97(2):716–721. https://doi.org/10.1073/pnas.97.2.716

Hori T, Kagawa N, Toyoda A, Fujiyama A, Misu S, Monma N, Makino F, Ikeo K, Fukagawa T. Constitutive centromere-associated network controls centromere drift in vertebrate cells. *J Cell Biol*. 2017:216(1):101–113. https://doi.org/10.1083/jcb.201605001

Iwata A, Tek AL, Richard MM, Abernathy B, Fonsêca A, Schmutz J, Chen NW, Thareau V, Magdelenat G, Li Y, et al. Identification and characterization of functional centromeres of the common bean. *Plant J*. 2013:76(1):47–60. https://doi.org/10.1111/tpj.12269

Jiang J, Birchler JA, Parrott WA, Dawe RK. A molecular view of plant centromeres. *Trends Plant Sci*. 2003:8(12):570–575. https://doi.org/10.1016/j.tplants.2003.10.011

Jones MG, Hughes J, Tregova A, Milne J, Tomsett AB, Collin HA. Biosynthesis of the flavour precursors of onion and garlic. *J Exp Bot*. 2004:55(404):1903–1918. https://doi.org/10.1093/jxb/erh138

Kirov I, Odintsov S, Omarov M, Gvaramiya S, Merkulov P, Dudnikov M, Ermolaev A, Van Laere K, Soloviev A, Khrustaleva L. Functional *Allium fistulosum* centromeres comprise arrays of a long satellite repeat, insertions of retrotransposons and chloroplast DNA. *Front Plant Sci*. 2020:11:562001. https://doi.org/10.3389/fpls.2020.562001

Kiseleva AV, Kirov IV, Khrustaleva LI. [Chromosomal organization of centromeric Ty3/gypsy retrotransposons in Allium cepa L. and Allium fistulosum L]. *Genetika*. 2014:50:670–676. https://doi.org/10.1134/S102279541404005X

Liu Y, Yi C, Fan C, Liu Q, Liu S, Shen L, Zhang K, Huang Y, Liu C, Wang Y, et al. Pan-centromere reveals widespread centromere repositioning of soybean genomes. *Proc Natl Acad Sci U S A*. 2023:120(42):e2310177120. https://doi.org/10.1073/pnas.2310177120

Logsdon GA, Rozanski AN, Ryabov F, Potapova T, Shepelev VA, Catacchio CR, Porubsky D, Mao Y, Yoo D, Rautiainen M, et al. The variation and evolution of complete human centromeres. *Nature*. 2024:629(8010):136–145. https://doi.org/10.1038/s41586-024-07278-3

Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, Ramírez F, Manke T. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics*. 2021:37(3):422–423. https://doi.org/10.1093/bioinformatics/btaa692

Maluszynsak J, Heslop-Harrison JS. Localization of tandemly repeated DNA sequences in Arabidopsis thaliana. *Plant J*. 1991:1(2):159–166. https://doi.org/10.1111/j.1365-313X.1991.00159.x

Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol*. 2013:14(1):R10. https://doi.org/10.1186/gb-2013-14-1-r10

Nagaki K, Cheng Z, Ouyang S, Talbert P, Kim M, Jones K, Henikoff S, Buell C, Jiang J. Sequencing of a rice centromere uncovers active genes. *Nat Genet*. 2004:36(2):138–145. https://doi.org/10.1038/ng1289

Nagaki K, Murata M. Characterization of CENH3 and centromere-associated DNA sequences in sugarcane. *Chromosome Res*. 2005:13(2):195–203. https://doi.org/10.1007/s10577-005-0847-2

Nagaki K, Shibata F, Suzuki G, Kanatani A, Ozaki S, Hironaka A, Kashihara K, Murata M. Coexistence of NtCENH3 and two retrotransposons in tobacco centromeres. *Chromosome Res*. 2011:19(5):591–605. https://doi.org/10.1007/s10577-011-9219-2

Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang JM. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres. *Genetics*. 2003:163(3):1221–1225. https://doi.org/10.1093/genetics/163.3.1221

Nagaki K, Tanaka K, Yamaji N, Kobayashi H, Murata M. Sunflower centromeres consist of a centromere-specific LINE and a chromosome-specific tandem repeat. *Front Plant Sci*. 2015:6:912. https://doi.org/10.3389/fpls.2015.00912

Nagaki K, Tsujimoto H, Sasakuma T. A novel repetitive sequence of sugarcane, SCEN family, locating on centromeric regions. *Chromosome Res*. 1998:6(4):295–302. https://doi.org/10.1023/A:1009270824142

Nagaki K, Yamamoto M, Yamaji N, Mukai Y, Murata M. Chromosome dynamics visualized with an anti-centromeric histone H3 antibody in Allium. *PLoS One*. 2012:7(12):e51315. https://doi.org/10.1371/journal.pone.0051315

Naish M, Alonge M, Wlodzimierz P, Tock AJ, Abramson BW, Schmücker A, Mandáková T, Jamge B, Lambing C, Kuo P, et al. The genetic and epigenetic landscape of the Arabidopsis centromeres. *Science*. 2021:374(6569):eabi7489. https://doi.org/10.1126/science.abi7489

Naish M, Henderson IR. The structure, function, and evolution of plant centromeres. *Genome Res*. 2024:34(2):161–178. https://doi.org/10.1101/gr.278409.123

Nasuda S, Hudakova S, Schubert I, Houben A, Endo TR. Stable barley chromosomes without centromeric repeats. *Proc Natl Acad Sci U S A*. 2005:102(28):9842–9847. https://doi.org/10.1073/pnas.0504235102

Nergadze SG, Piras FM, Gamba R, Corbo M, Cerutti F, McCarter JGW, Cappelletti E, Gozzo F, Harman RM, Antczak DF, et al. Birth, evolution, and transmission of satellite-free mammalian centromeric domains. *Genome Res*. 2018:28(6):789–799. https://doi.org/10.1101/gr.231159.117

Novák P, Neumann P, Macas J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat Protoc*. 2020:15(11):3745–3776. https://doi.org/10.1038/s41596-020-0400-y

Oliveira LC, Torres GA. Plant centromeres: genetics, epigenetics and evolution. *Mol Biol Rep*. 2018:45(5):1491–1497. https://doi.org/10.1007/s11033-018-4284-7

Plačková K, Bureš P, Zedek F. Centromere size scales with genome size across eukaryotes. *Sci Rep*. 2021:11(1):19811. https://doi.org/10.1038/s41598-021-99386-7

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016:44(W1):W160–W165. https://doi.org/10.1093/nar/gkw257

Richards EJ, Dawe RK. Plant centromeres: structure and control. *Curr Opin Plant Biol*. 1998:1(2):130–135. https://doi.org/10.1016/S1369-5266(98)80014-9

Rocchi M, Archidiacono N, Schempp W, Capozzi O, Stanyon R. Centromere repositioning in mammals. *Heredity (Edinb)*. 2012:108(1):59–67. https://doi.org/10.1038/hdy.2011.101

Samel A, Cuomo A, Bonaldi T, Ehrenhofer-Murray AE. Methylation of CenH3 arginine 37 regulates kinetochore integrity and chromosome segregation. *Proc Natl Acad Sci U S A*. 2012:109(23):9029–9034. https://doi.org/10.1073/pnas.1120968109

Schneider KL, Xie Z, Wolfgruber TK, Presting GG. Inbreeding drives maize centromere evolution. *Proc Natl Acad Sci U S A*. 2016:113(8):E987–E996. https://doi.org/10.1073/pnas.1522008113

Shang L, He W, Wang T, Yang Y, Xu Q, Zhao X, Yang L, Zhang H, Li X, Lv Y, et al. A complete assembly of the rice Nipponbare reference genome. *Mol Plant*. 2023:16(8):1232–1236. https://doi.org/10.1016/j.molp.2023.08.003

Shang WH, Hori T, Toyoda A, Kato J, Popendorf K, Sakakibara Y, Fujiyama A, Fukagawa T. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res*. 2010:20(9):1219–1228. https://doi.org/10.1101/gr.106245.110

Shen W, Sipos B, Zhao L. SeqKit2: a Swiss army knife for sequence and alignment processing. *Imeta*. 2024:3:e191. https://doi.org/10.1002/imt2.191

Su H, Liu Y, Liu C, Shi Q, Huang Y, Han F. Centromere satellite repeats have undergone rapid changes in polyploid wheat subgenomes. *Plant Cell*. 2019:31(9):2035–2051. https://doi.org/10.1105/tpc.19.00133

Sugiharto Arifin N, Ozaki Y, Okubo H. Genetic diversity in Indonesian shallot (*Allium cepa* var. *ascalonicum*) and *Allium* × *wakegi* revealed by RAPD markers and origin of A. × *wakegi* identified by RFLP analyses of amplified chloroplast genes. *Euphytica*. 2000:111(1):23–31. https://doi.org/10.1023/A:1003732403402

Sundararajan K, Straight AF. Centromere identity and the regulation of chromosome segregation. *Front Cell Dev Biol*. 2022:10:914249. https://doi.org/10.3389/fcell.2022.914249

Swangsri T, Reamtong O, Saralamba S, Rakthong P, Thaenkham U, Saralamba N. Exploring the antimicrobial potential of crude peptide extracts from. *Pharm Biol*. 2024:62(1):666–675. https://doi.org/10.1080/13880209.2024.2395517

Sweeten AP, Schatz MC, Phillippy AM. ModDotPlot-rapid and interactive visualization of tandem repeats. *Bioinformatics*. 2024:40(8): btae493. https://doi.org/10.1093/bioinformatics/btae493

Takeuchi H, Nagahara S, Higashiyama T, Berger F. The chaperone NASP contributes to de novo deposition of the centromeric histone variant CENH3 in Arabidopsis early embryogenesis. *Plant Cell Physiol*. 2024:65(7):1135–1148. https://doi.org/10.1093/pcp/pcae030

Tek AL, Nagaki K, Yıldız Akkamış H, Tanaka K, Kobayashi H. Chromosome-specific barcode system with centromeric repeat in cultivated soybean and wild progenitor. *Life Sci Alliance*. 2024:7(12):e202402802. https://doi.org/10.26508/lsa.202402802

Tsukahara S, Bousios A, Perez-Roman E, Yamaguchi S, Leduque B, Nakano A, Naish M, Osakabe A, Toyoda A, Ito H, et al. Centrophilic retrotransposon integration via CENH3 chromatin in Arabidopsis. *Nature*. 2025:637(8046):744–748. https://doi.org/10.1038/s41586-024-08319-7

Wang K, Wu Y, Zhang W, Dawe RK, Jiang J. Maize centromeres expand and adopt a uniform size in the genetic background of oat. *Genome Res*. 2014:24(1):107–116. https://doi.org/10.1101/gr.160887.113

Wlodzimierz P, Rabanal FA, Burns R, Naish M, Primetis E, Scott A, Mandáková T, Gorringe N, Tock AJ, Holland D, et al. Cycles of satellite and transposon evolution in Arabidopsis centromeres. *Nature*. 2023:618(7965):557–565. https://doi.org/10.1038/s41586-023-06062-z

Wong LH, Choo KH. Centromere on the move. *Genome Res*. 2001:11(4): 513–516. https://doi.org/10.1101/gr.183901

Zhang H, Dawe RK. Total centromere size and genome size are strongly correlated in ten grass species. *Chromosome Res*. 2012:20(4):403–412. https://doi.org/10.1007/s10577-012-9284-1

Zhang H, Song L, Wang X, Cheng H, Wang C, Meyer CA, Liu T, Tang M, Aluru S, Yue F, et al. Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat Commun*. 2021:12(1):6566. https://doi.org/10.1038/s41467-021-26865-w

Zhang W, Zuo S, Li Z, Meng Z, Han J, Song J, Pan YB, Wang K. Isolation and characterization of centromeric repetitive DNA sequences in Saccharum spontaneum. *Sci Rep*. 2017:7(1):41659. https://doi.org/10.1038/srep41659