



OPEN

Molecular organization of recombinant human-Arabidopsis chromosomes in hybrid cell lines

Yikun Liu^{1,6}, Yeng Mun Liaw^{1,6}, Chee How Teo², Petr Cápál³, Naoki Wada⁴, Kiichi Fukui⁵, Jaroslav Doležel³ & Nobuko Ohmido^{1✉}

Although plants and animals are evolutionarily distant, the structure and function of their chromosomes are largely conserved. This allowed the establishment of a human-Arabidopsis hybrid cell line in which a neo-chromosome was formed by insertion of segments of Arabidopsis chromosomes into human chromosome 15. We used this unique system to investigate how the introgressed part of a plant genome was maintained in human genetic background. The analysis of the neo-chromosome in 60- and 300-day-old cell cultures by next-generation sequencing and molecular cytogenetics suggested its origin by fusion of DNA fragments of different sizes from Arabidopsis chromosomes 2, 3, 4, and 5, which were randomly intermingled rather than joined end-to-end. The neo-chromosome harbored Arabidopsis centromeric repeats and terminal human telomeres. Arabidopsis centromere wasn't found to be functional. Most of the introgressed Arabidopsis DNA was eliminated during the culture, and the Arabidopsis genome in 300-day-old culture showed significant variation in copy number as compared with the copy number variation in the 60-day-old culture. Amplified Arabidopsis centromere DNA and satellite repeats were localized at particular loci and some fragments were inserted into various positions of human chromosome. Neo-chromosome reorganization and behavior in somatic cell hybrids between the plant and animal kingdoms are discussed.

The unique structure of eukaryotic chromosomes played important roles in genome evolution, speciation, and transmission of genetic information to progenies. The study of chromosome functions, including replication, segregation, gene expression, and inheritance, is necessary to reveal the differences between plant and animal chromosome organization and function. Understanding chromosome conservation among different organisms can contribute not only to genetics, but also to studies of the biosynthesis and genome evolution of the organisms.

Cell fusion is a useful technique commonly used in biomedical research for gene mapping^{1,2}. It has also been used to produce hybridoma cells through fusion of murine myeloma cells with human lymphocytes to produce monoclonal antibodies³. However, cell fusion often induces chromosomal instability. Production of somatic hybrid cell lines is often accompanied by chromosome loss whose mechanism is not clear. Human-rodent somatic cell hybrids are useful tools for mapping human chromosomes and understanding gene functions. Human-mouse hybrid cell lines⁴ showed preferential loss of human chromosomes and at least 75% of the human component was lost. The clones had a relatively stable karyotype, and to some extent, the human chromosome content was found to be dependent on the culture conditions. After the hybrid population was formed, Weiss and Green⁴ noted that continued subculture resulted in slow elimination of the human chromosome component. However, no evidence of chromosomal rearrangement was found in a majority of hybrid cell lines.

In a mouse chimeric line, several human chromosome fragments were retained and expressed in chimeric mice⁵. A human chromosome fragment of about 1–2 Mb carrying human Ig genes was maintained as an independent chromosome in the chimeric mice, and chromosomal rearrangement occurred in some of the clones.

¹Graduate School of Human Development and Environment, Kobe University, Kobe, Hyogo 657-8501, Japan. ²Centre for Research in Biotechnology for Agriculture, Universiti Malaya, 50603 Lembah Pantai, Kuala Lumpur, Malaysia. ³Institute of Experimental Botany of the Czech Academy of Sciences, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, 779 00 Olomouc, Czech Republic. ⁴Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Tokushima, Tokushima 770-8503, Japan. ⁵Graduate School of Pharmaceutical Sciences, Osaka University, Suita, Osaka 565-0871, Japan. ⁶These authors contributed equally: Yikun Liu and Yeng Mun Liaw. ✉email: ohmido@kobe-u.ac.jp

Most of the chimeras retained the transferred chromosome fragment in all tissues and one of the derived fragments showed germline transmission up to the F4 generation with no observable phenotype defects.

The stability of different human chromosomal fragments in different cell types has also been studied. Shinohara et al.⁶ constructed a mouse library carrying human chromosome fragments. The estimated sizes were 5–20 Mb for human chromosome fragments hCF-2 and hCF-11 and 90 Mb for hCF-14. Interchromosomal rearrangements were not detected in mouse embryonic stem cells and all the transferred human chromosomes were maintained as independent extra copies. hCF-14 had high stability, but hCF-2 and hCF-11 were relatively unstable in mouse embryonic stem cells. The unstable hCF-2 and more stable hCF-14 were tested in human fibrosarcoma HT1080, HeLa, and chicken DT40 B lymphoma cells under prolonged culture. The retention rate of hCF-14 was relatively high in HT1080, HeLa, and DT40, whereas hCF-2 was lost with different kinetics in HT1080 and HeLa. The stability of hCF-14 in various cell types may indicate the role of specific DNA elements in determining the retention stability.

Genome behaviour in hybrids has been investigated among evolutionarily very distant organisms⁷. Allshire et al.⁸ produced a yeast-mouse hybrid cell line carrying a large yeast chromosomal segment in a mouse background. The spheroplast fusion cell line was originally made by fusing fission yeast *Schizosaccharomyces pombe*, carrying an integrated mammalian selectable marker, with a mouse tumor cell line⁸. This cell line carried an insertion of *S. pombe* DNA at a single location on mouse chromosome 10⁹. McManus et al.⁹ investigated the composition of the yeast transgenomes and the modification of chromatin structure of the yeast DNA in the mouse cells. Chromatin of a large foreign DNA introduced into mammalian cells showed distinct condensation from that of the surrounding host mammalian DNA. The insertions that originated from foreign DNA were coincident with the presence of high levels of the heterochromatin marker histone H3 trimethylated on lysine 9¹⁰. In some cases, such as protoplast fusion of interspecific yeast hybrids, chimeric chromosomes may arise because of chromosome replacement, rearrangement, and recombination in chimeric hybrid yeasts¹¹.

Previous interkingdom cell fusion lines of human HeLa cells with carrot protoplasts¹² and hen erythrocytes with yeast protoplasts¹³ were maintained as heterokaryons that existed as multinucleate cells, although some interphase nuclei were fused in the human-carrot hybrid cells. An interphylum somatic hybrid of human HeLa cells with mosquito *Aedes aegypti* by ultraviolet-inactivated *Sendai* virus were maintained as mononucleate hybrid cells in ten subcultures with the loss of insect chromosomes, but a near diploid number of human chromosomes¹⁴.

In our previous work, we created a human-plant hybrid cell line by fusing *A. thaliana* protoplasts and human HT1080 cells¹⁵. The human-Arabidopsis hybrid cells stably maintained the plant-derived neo-chromosomes (PD chromosomes) and number of Arabidopsis genes were expressed in the human genetic background¹⁵. The results revealed that Arabidopsis-derived sequences fused to human chromosome 15 via a process called terminal translocation. However, the detailed characterization of its genome has not been performed yet.

In this study, we investigated the structure of the Arabidopsis-derived neo-chromosome as well as changes in its structure after 240 days culture. Recently, next generation high-throughput sequencing (NGS) and copy number variation (CNV) analysis have been applied to investigate how alien chromosomes maintain and function in host organism. Structural variation detection methods, including fluorescence in situ hybridization (FISH), microarrays, and WGS, have been used to investigate chromosomal changes¹⁶. We also performed whole-genome sequencing on 60- and 300-day-old cell cultures; that is, 60 and 300 days after the initial cell-protoplast fusion event. The reorganization of integrated Arabidopsis genome regions in the hybrid cell line was investigated by FISH and WGS coupled with CNV and DNA repeat analysis. The results provided new insights into the neo-chromosome formation, and the maintenance and alteration of plant DNA in human genetic background.

Results

Elimination and reorganization of the Arabidopsis genome in a human-plant hybrid cell line. To characterize DNA sequences in the hybrid cell line, and molecular organization of the human-Arabidopsis neo-chromosome in particular, whole-genome Illumina sequencing was done on 60- and 300-day-old hybrid cell lines at 50× and 30× coverage, respectively. Mapping DNA sequence reads of 60-day hybrid cell to the *A. thaliana* genome assembly (TAIR10) revealed the presence of fragments from all Arabidopsis chromosomes, including fragments of the short and long arms (Chr1), complete centromeric repeat regions (Chr3 and Chr5) and telomeric regions (Chr2, Chr3, and Chr5) (pink and red zones in Fig. 1). However, 300-day hybrid cell showed considerable elimination of Arabidopsis sequences from Chr1, Chr3 and Chr4 (pink zones in Fig. 1), while remaining alien fragments of Chr2 and Chr5 were unaffected by the elimination process (red zones in Fig. 1). Majority of long arm of Chr2 present in 60-day cell remained entirely in 300-day-old culture, together with more than 80% of Chr5 including its centromeric region was kept in 300-day-old culture (Table 1). As DNA sequences of the entire long arm of Chr3 was absent in the de novo assembly WGS data of the 300-day hybrid cell line, we concluded that the retained centromeric repeats originated from Chr5. When compared with the microarray data from the original hybrid cell line¹⁵, the WGS of the Arabidopsis genome performed on our 60-day cell line correlated well with the previous microarray results. Their microarray analysis identified more than 400 gene loci (Wada et al.¹⁵, Fig. S5) in five Arabidopsis chromosomes and all of these loci were also detected in our WGS analysis, as represented in Fig. 1.

To identify fusion breakpoints in the neo-chromosome, structural variations were called using DELLY¹⁷ and the predicted breakpoints were confirmed by PCR followed by Sanger sequencing. The results showed that a segment of Arabidopsis Chr5 at position 21,627,782 joined a segment of Chr2 at position 15,076,777 of their original TAIR10 coordinates (marked by asterisk in Fig. 1, Supplementary Fig. S1 and Table S1). The results were consistent for both the 60- and 300-day-old cell lines. The interrogated junction comprised terminal 198 bp from Chr5 fragment and 64 bp from the middle of the Chr2 fragment joined in a 3' to 5' orientation (Supplementary Fig. S1), while remaining parts of Chr2 were scattered on different positions of neo-chromosome and other

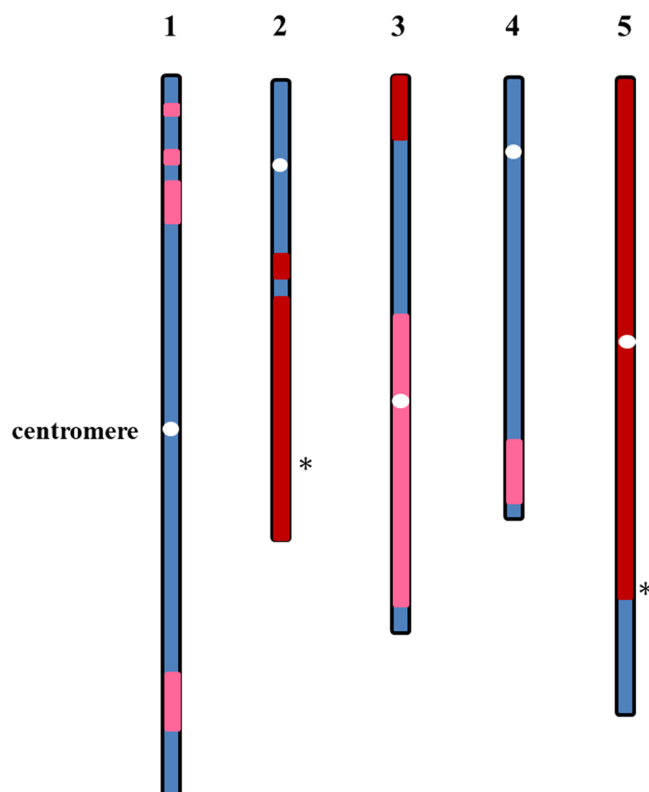


Figure 1. *Arabidopsis thaliana* sequences present in the 60-day and 300-day-old hybrid cell line. The idiogram shows the regions larger than 200 kb of the *Arabidopsis* chromosomes that were 60-day present (pink + red), 300-day-old present (red) and absent (blue) in the hybrid cell line. White ovals indicate centromere positions; asterisks (*) indicate fusion positions that join Chr2 and Chr5 in the hybrid genome.

<i>A. thaliana</i> chromosome	Location in <i>A. thaliana</i> genome (bp)	Sequence length (bp)	60-day	300-day	Markers
1	1,341,549–1,850,658	509,109	+	–	
1	3,339,470–3,876,463	536,993	+	–	
1	4,933,240–6,138,109	1,204,869	+	–	
1	16,509,811–17,148,282	638,471	+	–	
1	25,384,354–27,494,134	2,109,780	+	–	
2	7,507,287–8,318,501	811,214	+	+	
2	9,554,225–19,698,289	10,144,064	+	+	BAC T1J8 ^a
3	96–2,693,711	2,693,615	+	+	BAC T9J14 ^a , <i>Bsd</i> ^b , GFP ^c
3	8,348,362–22,733,522	14,385,160	+	–	
4	9,257,087–9,867,323	610,236	+	–	
5	0–21,629,605	21,629,605	+	+	BAC T22P22 ^a

Table 1. *Arabidopsis thaliana* genome regions present in the 60- and 300-day-old hybrid cell lines. Markers of *Arabidopsis* genomic region: ^aBacterial artificial chromosome (BAC); ^bBlasticidin S resistance (*Bsd*-resistance) gene and ^cGreen fluorescence protein (GFP) gene localized by Wada et al.¹⁵ in hybrid cell. +: present, –: not present.

human chromosome. The structural variations detected using DELLY were supported by four split-reads and seven discordant paired-end reads that aligned to both chromosomes.

Calculated using SAMtools, the WGS data showed a higher proportion of the *Arabidopsis* genome in the 60-day-old culture (55 Mbp, 46%) than in the 300-day-old culture (29 Mbp, 24%), confirming that *Arabidopsis*-derived genome fragments in the hybrid cell line were lost during the 240 days of culture. To further investigate this phenomenon, we analyzed CNV and characterized DNA repeat content in the *Arabidopsis*-derived fragments. To characterize *Arabidopsis* DNA repeat composition, we analyzed de novo assembled genome of the hybrid cell line to identify *Arabidopsis* repeats using RepeatMasker¹⁸. In general, all known *Arabidopsis* repeat families (e.g., retroelements, DNA transposons) were present in the hybrid genome in lower numbers compared

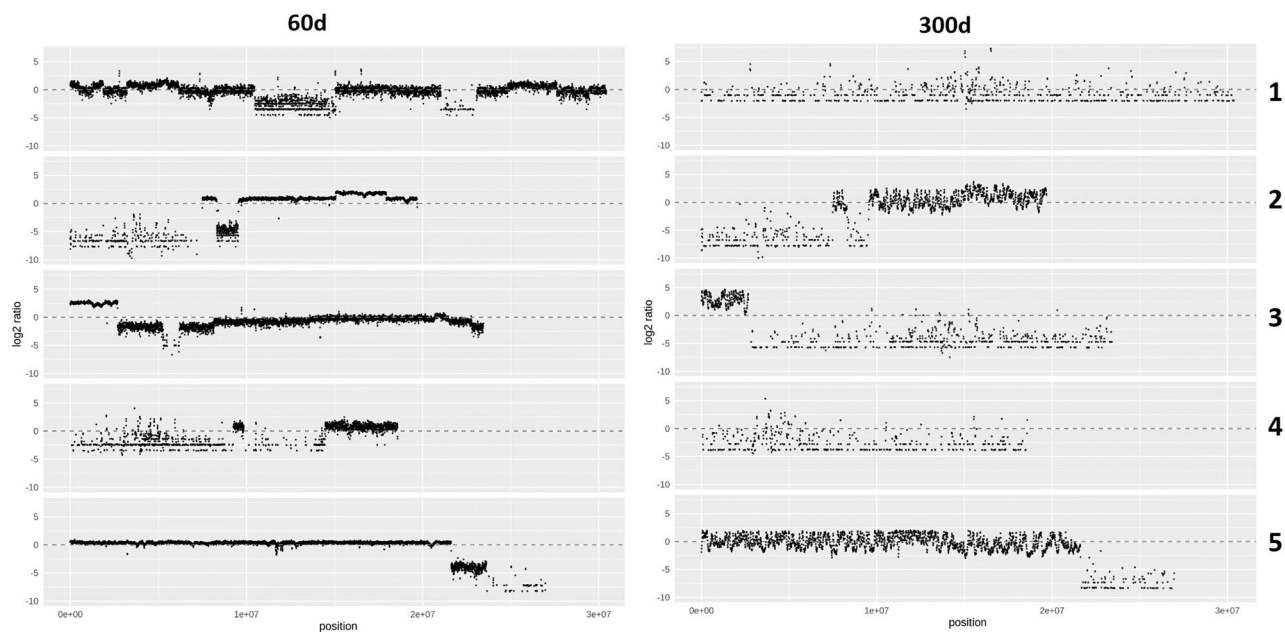


Figure 2. Copy number variation of Arabidopsis genome fragments in the 60- and 300-day-old hybrid cell lines. Log₂ ratio of zero (gray dashed line) indicates diploid level. X-axis: DNA sequence of Arabidopsis pseudomolecules; Y-axis: log₂ ratio relative to the Arabidopsis reference genome TAIR10. Numbers on the right side of the graphs indicate Arabidopsis chromosome numbers. Differences in plot density were caused by the adjustment of window size in CNV-seq according to coverage depth.

to the wild-type Arabidopsis genome (Supplementary Fig. S2 and Supplementary Table S2). The repeat composition was similar in the 60- and 300-day-old hybrid cell lines, suggesting that the elimination of repeat families occurred at an earlier stages of cell culture. Interestingly, the cumulative length of satellite repeats was measurably higher in both the 60- and 300-day-old cell lines (4,879,652 bp and 3,740,017 bp, respectively) compared to the Arabidopsis (976,448 bp) and human (1,485,509 bp) genomes.

Changes in copy number of the introgressed Arabidopsis genome segments were determined by CNV-seq¹⁹ after normalization with the Arabidopsis reference genome TAIR10 (Fig. 2). Plot density was adjusted by CNV-seq according to the coverage depth. The WGS data from the 60-day-old hybrid cell line showed that the genome of hybrid cell line contained DNA fragments from all Arabidopsis chromosomes (Fig. 1, Table 1). CNV-seq data showed a relatively stable copy number state for 60-day-old hybrid cell, with no more than two copy number states detected for Arabidopsis Chr2 and Chr5. The CNV cluster with the log₂ value of -4 around the end of the long arm of chromosome 5 indicates that the copy number difference between this cluster in wild type Arabidopsis and hybrid cells is at least 16 fold different. This suggested that the majority of the hybrid cells in the highly heterogeneous hybrid cell population were lacking in this chromosome region.

Conversely, CNV was higher in the 300-day-old hybrid cell line and large fluctuations in copy number state were detected in DNA fragments from all Arabidopsis chromosomes. Considerably less DNA from Chr1, Chr4 and Chr3 was found in the 300-day-old cell line. Larger variation in DNA copy number, especially for Chr5-derived sequences, indicated DNA loss and amplification. The log₂ ratio of Chr3 in the 60-day-old cell line indicated 2- to eightfold variation in copy number, whereas in the 300-day-old cell line, the copy number varied up to 32-fold. Contrary to the segmented manner that indicated block-by-block fluctuations of copy number observed in 60-day-old culture, in the 300-day-old culture continuous copy number fluctuations with an unclear segment boundary were detected. In entire human chromosome, 300-days old hybrid cell showed a slightly higher fluctuation compared to the 60-days old cell (Supplementary Fig. S3). However, in the 300-days old hybrid cell line, the copy number changes of the human chromosome (log₂ ratio; 1 to -2.5) are smaller in the comparison to the Arabidopsis chromosome 3 (0 to 5) and 5 (2.5 to -2.5), respectively. No copy number gains on the human chromosome 15 and others during the 240-days culture was observed. Therefore, it is deduced that genome instability occurred in the whole hybrid genome in cell culture.

Neo-chromosome structure and cytogenetic landmarks. Neo-chromosomes were classified into three subtypes, PD type-T, type-S and type-A based on the number and orientation of the Arabidopsis genomic signals on the neo-chromosomes¹⁵. We used cell culture followed by FISH analysis with Arabidopsis centromeric probes to estimate cell heterogeneity by identifying the neo-chromosomes, classified as PD type-T, type-A, and/or type-S. This was performed on both 60-day and 300-day cell cultures in this study. 95% type-T and only 5% type-A chromosomes in 21 chromosome spreads (Supplementary Fig. S4) were detected in the 300 days cell cultures. No type-S was detected. To ensure reliability of the neo-chromosomes detection, we performed the same FISH analysis on a 60-day cell line only two days after cells were revived from cold storage. We observed a

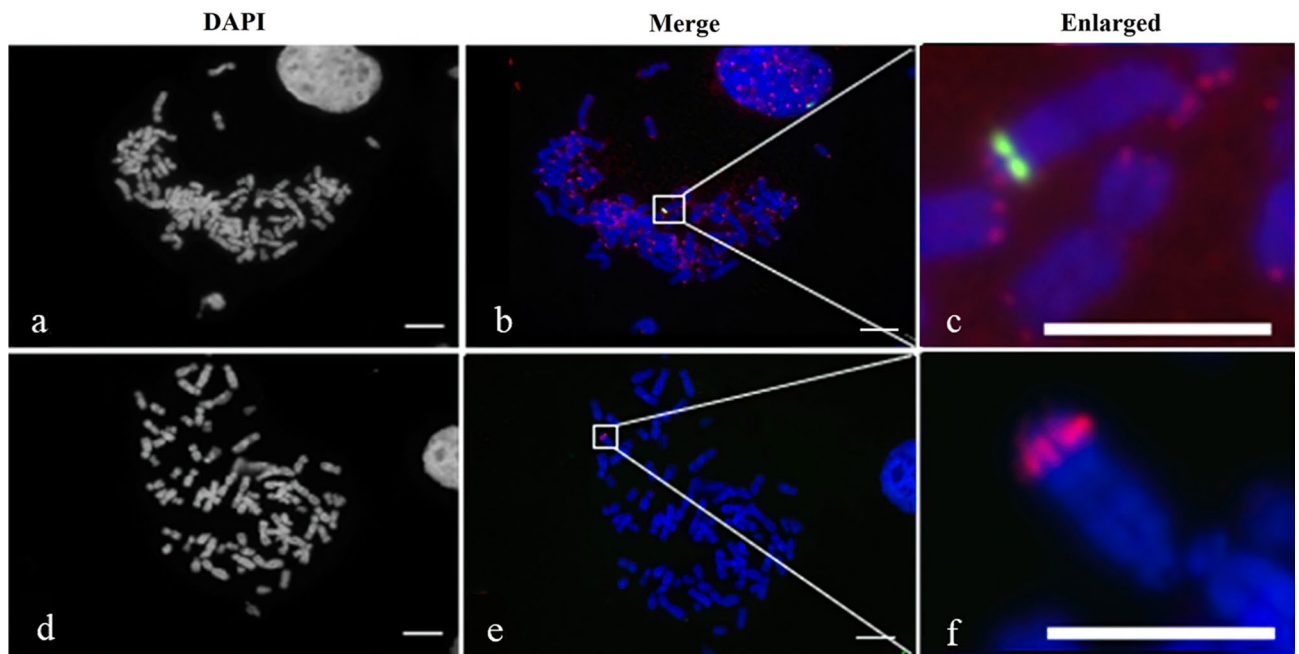


Figure 3. Genomic distribution of DNA repeats in the 300-day hybrid cell line. Fluorescence in situ hybridization (FISH) was done on mitotic metaphase plates using probes for Arabidopsis centromere (Atcen, green) and human telomere (red) (a–c), and the Atcen (red) and Arabidopsis telomere (green) (d–f). Chromosomes were counterstained using DAPI (gray pseudocolor in a,d and blue in b,c,e,f). Scale bars, 10 μ m (a,b,d,e) and 5 μ m (c,f).

similar trend, with 88% of type-T and 12% of type-A chromosomes in a total of 25 chromosome spreads counted (Supplementary Fig. S4). No type-S was observed.

We applied FISH on mitotic metaphase spreads to detect the orientation of the Arabidopsis chromosome fragments, confirm telomere type and centromere origin of the neo-chromosome in the 300-day-old hybrid cell line (Fig. 3). FISH failed to detect Arabidopsis telomeric repeats (TTTAGGG)_n²⁰ in cells of the 300-day-old culture (Fig. 3d–f), which indicated the Arabidopsis telomere sequence units were absent, or only few were present and below a detection limit of FISH. On the other hand, the human telomere probe (TTAGGG)_n was detected at both termini of the neo-chromosome with clear and strong hybridization signals (Fig. 3a–c). Moreover, a weak human telomere signal was detected in the proximal region of the neo-chromosome in the 60 day-old cell line, but this signal was not detected in the 300-day old cell line, indicating a rapid loss of this locus during the culture. A strong hybridization signal of Arabidopsis centromere (Atcen, 180 bp unit) repeats was observed next to the human telomere loci. The human telomere loci on the neo-chromosomes appeared near the Atcen signals probably because of low spatial resolution due to high chromosome condensation. These Atcen repeat signals were much stronger than in the wild-type Arabidopsis mitotic metaphase plates, suggesting amplification of the Atcen sequence in the hybrid cell line (Fig. 3c,f).

To localize particular Arabidopsis DNA sequences on the neo-chromosome, pooled probes Ch2P, Ch3P, and Ch5P were used for FISH. Each pooled probe consisted of five PCR amplicons targeting single-locus DNAs covering approximately 5 kb physical distance in the TAIR10 genome (Supplementary Table S3). Their hybridization signals were located close to each other on the distal part of neo-chromosome. To determine the orientation of Ch2P and Ch3P, they were labelled with a different fluorochrome and detected simultaneously by FISH on the metaphase chromosome of 300-day-old hybrid cell line (Fig. 4a). The Ch2P fluorescence signal overlapped with the Ch3P and Arabidopsis centromere signals, respectively (Fig. 4a, bottom plates). The hybridization signals of each single probe were detected on the same regions (Supplementary Fig. S5). It would be due to the short physical distances between the Arabidopsis genome segments as represented by FISH analysis using the pooled probes.

The FISH results are summarized in Fig. 4b. The neo-chromosome was formed from human chromosome 15 (HSA15)¹⁵ into which Arabidopsis genome segments had integrated. The mitotic neo-chromosome resembled the normal HSA15 structure, with human telomere signals on both termini and a human centromere in the primary constriction. A compact region included the Atcen sequence localized on the long arm of a distal part of HSA15 and Arabidopsis genome sequences from Chr2, Chr3, and Chr5 with human-specific telomeres.

Multiple localization of the Ch2-2 probe suggests translocation and amplification of Arabidopsis genome segments in the human genome. To check the distribution of each of the Arabidopsis single-copy genome segments, three 400–500 bp long DNA probes (Ch2-2, Ch3-2, and Ch5-2) were selected from 15 single copy probes for FISH analysis (Supplementary Table S3). The Ch3-2 and Ch5-2 probes always localized only on the expected locus on the neo-chromosome in both the 60- and 300-day-old cell lines (Supplementary Fig. S5). In the 60-day-old hybrid cell line, only one expected signal for Ch2-2 was observed and no

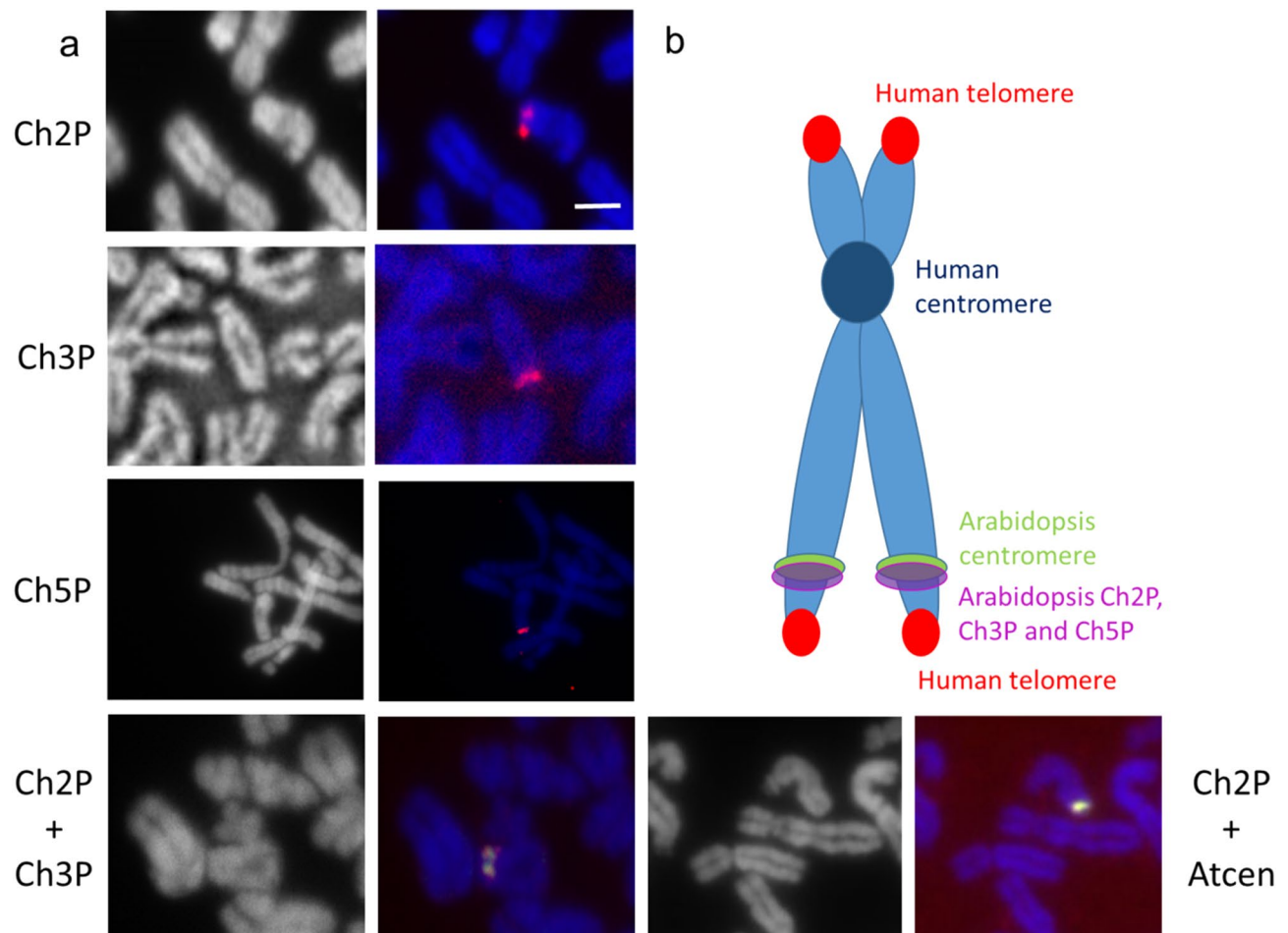


Figure 4. Molecular organization of the neo-chromosome. (a) Localization of Arabidopsis genome segments in mitotic metaphase chromosomes by FISH using chromosome-specific pooled probes singly, Ch2P (red), Ch3P (red), Ch5P (red), the combination Ch2P (red)/Ch3P (green), and Ch2P (red)/AtCen (green). Chromosomes were counterstained using DAPI (gray pseudocolor or blue). (b) Diagrammatic representation of the neo-chromosome. Scale bar, 5 μ m.

probe hybridization signal was found in the original hybrid cell line HT1080 (Supplementary Fig. S6). Surprisingly, in the 300-day-old cell line, the Ch2-2 hybridization signal was observed on three to ten positions within the genome, not only on the neo-chromosome, but also on other human chromosomes (Fig. 5a). Moreover, the Ch2-2 signal positions on the human chromosomes varied; some were close to the centromere while others were near telomeric regions (Fig. 5b). The chromosomes of hybrid cell line were thus classified into three types according to the number of Ch2-2 hybridization signals and the frequency of each type (Fig. 5c,d). Type 1 (75% of the chromosomes) had only one signal next to the centromere, type 2 (15% of the chromosomes) had two signals near telomeres, and type 3 (10% of the chromosomes) had more than two signals. The signals found on the human chromosomes other than those on HSA15 suggested that translocation and duplication occurred during the course of cell culture. The variable count of hybridization signals in different cells indicated chromosome instability in the hybrid cell line.

Simultaneous FISH with Ch2-2 and Atcen probe resulted in multiple hybridization signals of Ch2-2, but none of these signals detected on the neo-chromosome with Atcen signal (Fig. 6). Conversely, when pooled probes for Arabidopsis Ch2P were combined with Atcen probes, the hybridization signals overlapped and Ch2P probes cross hybridized in the same locus with the Atcen probes (Fig. 4, bottom right). Although Ch3-2 and Ch5-2 were supposed to be close to Ch2-2, the Ch3-2 and Ch5-2 probes did not block the Ch2-2 signal (middle and bottom plates in Fig. 6), possibly because the Ch2-2 locus was closer to the Atcen region or even overlapped on metaphase chromosomes.

Discussion

We report new insights into the long-range molecular organization of a unique neo-chromosome that originated in a human-plant hybrid cell line. The neo-chromosome was first described by Wada et al.¹⁵ and comprises human chromosome 15 and regions of Arabidopsis chromosomes 2, 3, and 5 as originally revealed by BAC-FISH and microarray analysis. We confirmed this principal composition of the neo-chromosome and we have identified Arabidopsis genome sequence maintained in the hybrid cells using illumina sequencing (Fig. 1, Table 1). We

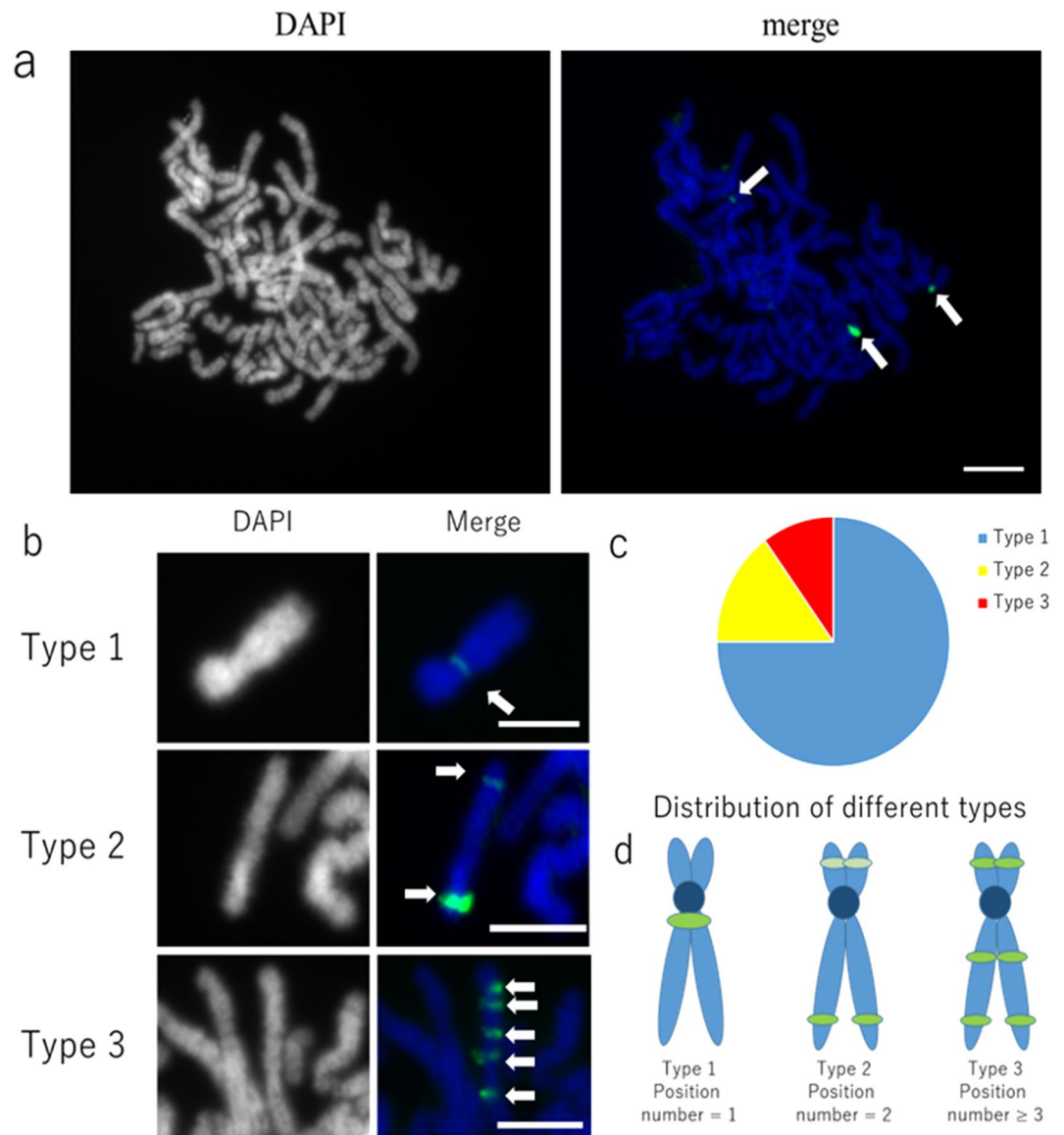


Figure 5. Localization of Arabidopsis genome segment Ch2-2 by FISH to human chromosomes in the hybrid cell line. **(a)** Probe Ch2-2 (green) hybridized to different chromosomes (arrows). Chromosomes were counterstained using DAPI (gray pseudocolor or blue). **(b)** Different types of chromosomal location of the Ch2-2 probe. **(c)** Relative proportion of chromosomes with different Ch2-2 signal distributions. 20 cells were examined by FISH for Ch2-2 localization. **(d)** Diagrammatic representation of the three types of chromosomes. Scale bars, 5 μ m.

have also employed FISH to determine the orientation of the inserted Arabidopsis chromosome segments and to outline the evolution of the neo-chromosome during the culture of the hybrid cell line (Figs. 3, 4, 5).

Our work revealed complex chromosomal rearrangement which occurred during the formation and maintenance of the neo-chromosome in the human-Arabidopsis hybrid cell line. The Arabidopsis-derived neo-chromosome was assembled by the fusion of Arabidopsis chromosome fragments to human chromosome 15, and contained human and Arabidopsis centromeres, two human telomeres, and highly rearranged chromosomal fragments of Arabidopsis chromosomes 2, 3, and 5 after 300 days of culture. A comparison of the 60- and 300-day-old cell lines revealed that fragments of the Arabidopsis genome were lost and their number decreased from 46% (55 Mb) in the 60-day-old cell line to 24% (29 Mb) in the 300-day-old cell line and that the remaining fragments underwent large fluctuations in copy numbers (Fig. 2). WGS, CNV and FISH revealed that the alien Arabidopsis genome was unstable as culture time progressed and some of the plant DNA fragments were lost. However, some Arabidopsis chromosome fragments, like the T-type, survived long term culture but the repeat sequences present in the Arabidopsis genome became larger than usual.

The WGS results revealed a fragmented manner in which Arabidopsis genome was retained (Supplementary Fig. S1) as well as non-conventional fusion breakpoint structure joining the terminal part of Arabidopsis

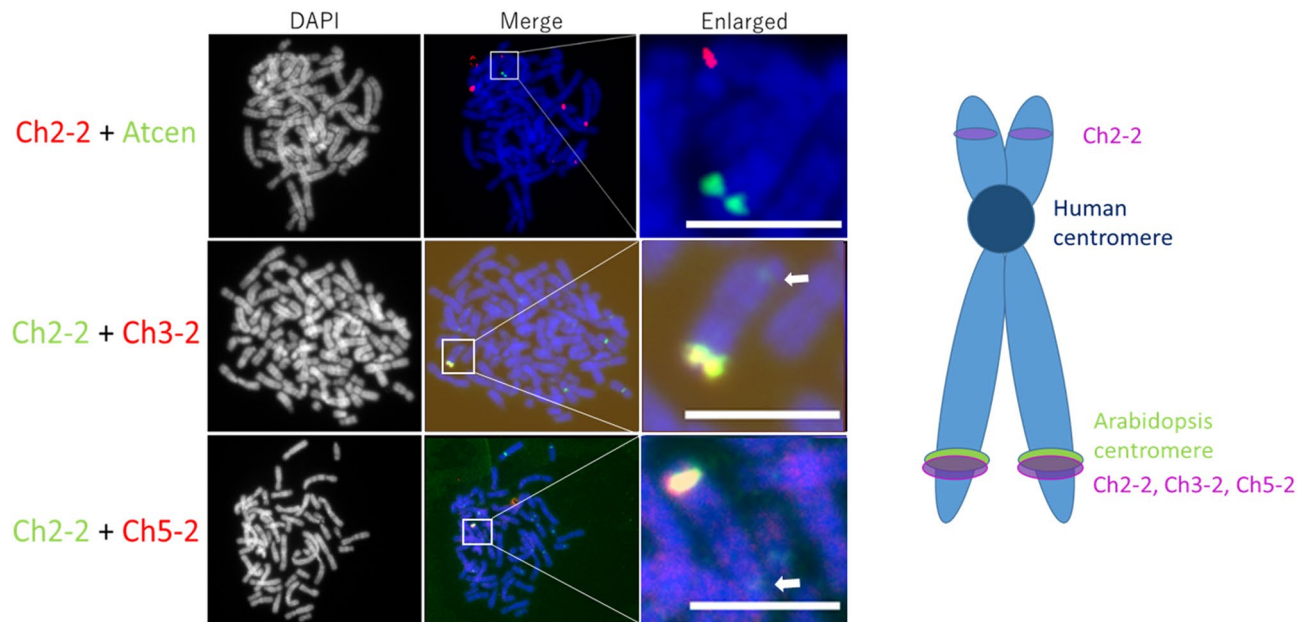


Figure 6. Simultaneous localization of Arabidopsis probes Ch2-2, Ch3-2, Ch5-2 and Arabidopsis centromere (Atcen) probe by FISH on metaphase chromosomes in the hybrid cell line. Arrows indicate weak Ch2-2 signals on the short chromosome arm. Scale bars, 5 μ m.

chromosome 5 and interstitial part of chromosome 2 (Supplementary Fig. S1). The observed overlap of the introgressed Arabidopsis sequences (Figs. 4, 6) indicated that complex processes were involved in the insertion of the Arabidopsis genome fragments. It is also possible that duplication of genome fragments occurred in the hybrid cell line. However, the mechanism responsible for the rearrangement of the Arabidopsis regions in the hybrid cells could not be determined by sequencing because of the low coverage (average 1.7 \times) of the relatively small Arabidopsis regions in the human genome background. For a mouse-yeast hybrid, Fitz-James et al.¹⁰ confirmed that the inserted yeast DNA was composed of rearranged DNA with 10–100 kb fragments (median length 20 kb) of contiguous yeast genome in the mouse fusion cells.

A comparison of the 60- and 300-day-old cell lines revealed that at least 50% of the Arabidopsis genome regions were eliminated during culture (Fig. 2) and a majority of these sequences were from Chr1, Chr3, and Chr4. Moreover, the copy number variation of the remaining Chr2, Chr3, and Chr5 sequences in the 300-day-old cell line was larger than in 60-day-old (Fig. 2). In particular, the region of Chr3 that contained the selective marker genes, blasticidin S (*Bsd*) resistance and green fluorescence protein (GFP) was amplified dramatically up to 32-fold and the Chr5 fragment that contained the centromere repeats was also amplified. These results indicate that the function of these two regions may be selectively controlled. Because cell survival in the media including blasticidin S requires the presence of the *Bsd*-resistance gene that is located on Chr3 in this hybrid cell line¹⁵, all cells presumably contained the Chr3 fragment that contained the resistance gene. However, the reason for preferable retention of other Arabidopsis chromosomal regions is not clear. Previous studies of a human-rodent somatic cell hybrid described preferential loss of complete mouse or human chromosomes^{4,21}. Somatic cell hybrids often exhibit random chromosome loss in cultures, but the mechanism is yet to be elucidated²². Fission yeast *S. pombe* chromosomes were integrated into mouse cells by somatic fusion⁹, and the yeast DNA was found to be integrated into a single site within a mouse chromosome and maintained there. The alien yeast DNA replicated as an independent unit using their own replication system in late S phase, and the replication timing was not synchronized with that of the flanking mouse DNA.

Wada et al.¹⁵ conducted a cytological analysis of the human-Arabidopsis 60-day-old hybrid cell line and concluded that the neo-chromosome originated via a process of end-to-end fusion of HSA15 with Arabidopsis chromosomes, performing three types of plant-derived chromosome (type T, S and A) during the culture process. In the 300-day old culture, the majority of the neo-chromosomes were type-T, most probably due to its preferential retention, as the plant chromosomal region was attached to HSA 15, allowing it is to pass on during cell replication. The previous report by Wada et al.¹⁵ relied on a continuous culture of the hybrid cell line immediately after the fusion event. The cells used in our experiments derived from the original cells but after 4.5 years after storage at -80°C . This long-term cold storage may explain the discrepancies observed in the 60-day cell line between previous Wada et al.¹⁵ and this study. We think that the freeze-thaw process might have caused the retention of the more stable type-T neo-chromosome, rather than type-S and type-A. Previous studies also demonstrated that cryopreservation of sperm could induce cell injury with increased DNA fragmentation^{23,24}. Based on the morphology of the translocated chromosome, we concluded that the acrocentric chromosome observed could only have been human chromosome 13, 14 or 15. Considering the previous results of Wada et al.¹⁵, we therefore estimated that the neo-chromosome was human chromosome 15 in both 60- and 300-day cell line.

Clusters of human telomeric sequence (TTAGGG)_n were observed at both termini of the neo-chromosome and also in its interstitial region. The human telomeric sequence on the distal end of the HSA15 short arm could

be the original human telomere. However, in the 300-day-old hybrid cell line, the Arabidopsis telomeric repeat sequence (TTTAGGG)_n and human telomeric repeat sequence (TTAGGG)_n were not detected in the proximal parts of the neo-chromosome (Fig. 4). Following tandem chromosome fusions, interstitial telomeric sequences have often been found in several plants and animal species^{25–28} and were not functional²⁵. The observed shortening or loss of the tandem repeats may be caused by polymerase replication slippage or unequal crossing over^{29,30}.

Another mechanism leading to genomic rearrangements is the formation of a dicentric chromosome containing two functional centromeres³¹, and our neo-chromosome consisted of two centromeres, one from human and one from Arabidopsis. Maciejowski et al.³² showed that dicentric chromosomes that caused the formation of chromatin bridges in mitosis can lead to chromothripsis (*i.e.*, chromosome shattering) through cycles of breakage and rejoining called breakage-fusion-bridge cycles. Gisselsson et al.³³ proved that breakage-fusion-bridge events can lead to genetic heterogeneity in tumors. This possibility was also noted by Wada et al.¹⁵ because human histone H3 variant centromere protein (CENP)-A was detected on two subtypes of the neo-chromosome (type S and A) by immunostaining. The expression of Arabidopsis centromere-specific histone H3 (CENH3) was not detected in the microarray data. In 300-day-old hybrid cell, CENH3 gene on Chr1 was completely eliminated. The amplified Arabidopsis centromere derived from Chr5 wasn't functional as Arabidopsis centromere in the 300-day-old hybrid cell line. Dicentric chromosomes are usually formed as a product of telomere fusion. When the two centromeres of a dicentric chromosome migrate to opposite mitotic spindle poles during, breakage-fusion-bridge cycle is initiated, resulting in complex chromosome rearrangements³⁴. However, because the neo-chromosome contained only one active human centromere, the complex chromosomal rearrangement was most probably not a consequence of dicentric chromosome formation.

Interestingly, several inter-chromosomal translocations were found by FISH using Arabidopsis chromosome-specific DNA probes (Table S3) and karyotype heterogeneity was observed in the hybrid cell population (Fig. 5). One possible reason behind these translocations is the involvement of transposable elements, which can be activated when a host genome fails to suppress their activity³⁵. As the sequence of Ch2-2 fragment is single-locus, it did not coincide with any transposable element in *A. thaliana* Repeat Maps³⁶. Interspecific translocation also was reported in mouse-human cell lines, where the long arm of human chromosome 17, or a portion of human X chromosome was translocated to a mouse chromosome^{1,2}. Friend et al.³⁷ also described several mouse-human hybrids that showed interspecific translocation. They observed the presence of the same hybrid chromosome in majority of the cells within a clone, although they didn't investigate the maintenance of the hybrid chromosome under long culture times^{1,37}.

Considering the arguments mentioned above and the complex molecular composition of the neo-centromere, a highly probable mechanism of its origin seems to be chromothripsis, where up to hundreds of structural rearrangements are acquired in a short time³⁸. This process is characterized by extensive genomic rearrangements and oscillating patterns of DNA copy number levels³⁹. Our WGS data revealed that Arabidopsis DNA was retained in fragments (Fig. 2) and FISH data showed highly overlapped regions of Arabidopsis Chr2, Chr3, and Chr5 (Figs. 4, 6). Thus, we speculate that chromothripsis occurred during cell fusion when the Arabidopsis genome was shattered, stitched, and randomly inserted into the human genome.

In this study, oscillations of copy number states occurred in all the Arabidopsis chromosomes in the hybrid cell lines, and occurred randomly across all chromosomes. A large difference in copy number variation with rapid elimination of large genome regions from the 60- to 300-day-old cultures indicated genomic instability in the hybrid line. The criterion for the occurrence of chromothripsis is clustered chromosomal rearrangements confined in localized genomic regions⁴⁰. However, Gu et al.⁴¹ reported that chromothripsis-like rearrangement also occurred beyond confined regions involving two or three chromosomes. The cause for the occurrence of chromothripsis is still unclear, with possible reasons including dicentric centromere, centromere inactivation or a genomic stress⁴². It is probable that a genomic shock due to the fusion of divergent genomes may lead to large-scale genome restructuring. Fragmented retention of Arabidopsis genome sequences, highly overlapped Arabidopsis chromosomal regions on the neo-chromosome, and multiple insertion of Arabidopsis chromosomal fragments into human genome all may contribute to stress.

To conclude, we demonstrated that fragments of plant genomes can be maintained in human cells. However, a majority of plant DNA sequences are eliminated after long culture because of genomic stress generated from the cross-kingdom hybridization. We suggest that the Arabidopsis genome was fragmented by chromothripsis⁴³ and its fragments were integrated into human chromosomes. Determination of gene expression and epigenetic modifications of the Arabidopsis genome fragments will be important to test how cross-kingdom conservation of mechanism regulate genome function.

Materials and methods

Plant material and DNA extraction. Seeds of *Arabidopsis thaliana* cv. Columbia ($2n=10$) from Kobe University were sterilized in 1 mL 10% (v/v) kitchen bleaching solution (KAO, Tokyo, Japan) for 30 min, rinsed several times, vernalized at 4 °C for 2–3 days and germinated on 0.5×Murashige and Skoog (MS) medium (Wako, Tokyo, Japan) supplemented with 0.5% agar for 10–15 days at 25 °C. DNA was extracted from seedlings in DNA Suisui buffer (RIZO Inc., Tsukuba, Japan) and ethanol precipitated.

Cell culture. The human-Arabidopsis hybrid cell line that contained a neo-chromosome (PD chromosome) was obtained by Wada et al.¹⁵ from 60- and 149-day-old cultures. 60-days and 149-days cell cultures originated from cell stocks kept in storage frozen at –80 °C for 5.5 and 4.5 years, respectively. The 60-day cell culture was used for experiments almost immediately, after only a few days of stabilization. The 300 day culture originated from cells previously grown for 149-day and stored frozen. These cells were revived and propagated every 2–3 days in culture for an additional 150 days before performing WGS and FISH analysis. The hybrid cell line

was cultured in Dulbecco's Modified Eagle Medium (DMEM) (Gibco, Carlsbad, California, USA) supplemented with 10% fetal bovine serum (FBS) (Gibco) and 6 µg/mL blasticidin S (Bsd) (KNF, Tokyo, Japan) at 37 °C in a 5% CO₂ incubator.

DNA extraction from hybrid cells. Genomic DNA was extracted from 60- and 300-day-old hybrid cells according to Miller et al.⁴⁴. Briefly, the cells were lysed, protein digested, RNase treated, and ethanol precipitated. Approximately 8 million cells at 90% confluency in 10-cm plates were lysed with 300 µL nuclei lysis buffer (10 mM Tris-HCl, 400 mM NaCl, 2 mM EDTA, pH 8.2) and digested at 37 °C overnight with 20 µL 10% SDS and 50 µL proteinase K solution (1 mg protease K in 1% SDS and 2 mM EDTA). Then, 100 µL 5 M NaCl was added and shaken vigorously for 15 s to precipitate the protein. The sample was pelleted at 13,000 × g for 15 min and the supernatant was transferred to a new tube. Next, 10 µL RNase A (10 mg/mL) was added to 200 µL of the solution and incubated at 37 °C for 1 h. Two volumes of absolute ethanol were added and the tube was inverted several times and kept overnight at –80 °C. Then, the sample was spun at 13,000 × g for 30 min at 4 °C. The supernatant was discarded and the pellet was washed with 70% cold ethanol and centrifuged again at 13,000 × g for 5 min at 4 °C. The supernatant was discarded and the pellet was left to air dry before being resuspended in distilled water. DNA concentration was measured with Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and its integrity was checked by gel electrophoresis.

Whole-genome sequencing. Genomic DNA isolated from the 60-day-old hybrid cell line was sequenced on Illumina NovaSeq 6000 instrument with 150-bp paired-end reads at the Institute of Experimental Botany⁴⁵. The sequencing library was prepared using NEBNext Ultra II DNA Library Prep Kit for Illumina (E7645) (New England BioLabs Inc., Ipswich, USA) with dual index primers and four rounds of PCR amplification. DNA fragments of 400–1000 bp were selected using Pippin Prep (Sage Science Inc., Beverly, MA, USA). Genomic DNA isolated from the 300-day-old hybrid cell line was sequenced by Macrogen Corp. (Kyoto, Japan) on Illumina NovaSeq 6000 instrument with 150-bp paired-end reads. The sequencing library was prepared using TruSeq DNA PCR-Free kit (Illumina, Inc., Tokyo, Japan).

NGS data processing. Quality control was performed using FastQC^{46,47} and adapter trimming and quality filtering were performed with Fastp and/or Trimmomatic^{46,48}. To avoid unspecific mapping, the filtered reads were mapped to combined genomes of human (GRCh38) and *A. thaliana* (TAIR10) with BWA-MEM v0.7.17¹⁸. Duplicated reads were removed with Picard (<http://broadinstitute.github.io/picard>) after mapping. Reads that aligned to the *A. thaliana* genome were extracted for downstream processing. The proportion of Arabidopsis genome in the hybrid cells was calculated using SAMtools⁴⁹. Long contiguous Arabidopsis region present in hybrid cell was determined using BEDTools⁵⁰ coupled with visual inspection of sequencing reads with reference to CNV data. Figure 1 and Table 1 were generated using BEDtools merge by first merging book-ended/flanking regions and overlapping features. For each of the merged regions, neighbouring regions within 500 bp were merged again, and only regions longer than 10 kb were retained. From these regions, neighbouring regions within 500 kb were again merged and regions exceeding 200 kb were retained. Finally, only regions larger than 500 kb were shown in Fig. 1 and Table 1, respectively.

Detection of fusion junction and structural variation. To detect fusion junctions between the introgressed Arabidopsis chromosome segments, structural variations were called using DELLY^{17,46} and filtered with supporting split reads or paired-end reads of at least three pairs. Translocated chromosomal regions were extracted to identify junction sequences between Arabidopsis chromosome fragments. Primers flanking the junction region were designed using Primer3.

De novo assembly. Sequence reads were assembled using Minia³⁵¹ with optimal k-mer determined using KmerGenie (1.7051)⁵². The reads from the 60- and 300-day-old hybrid cell lines were assembled with k83-mer and k73-mer, respectively. The assembled genomes were quality checked using QUAST (5.0.2). The assembled genomes from the 60- and 300-day-old hybrid cell lines consisted of 986,822 contigs with N50 of 3,922 bp and 935,237 contigs with N50 of 4,281 bp, respectively.

Characterization of DNA repeats. De-novo assembled genomes were masked with RepeatMasker (v1.332)¹⁸ to identify Arabidopsis DNA repeat families (Supplementary Table S2). Simple repeats and low complexity repeats were not considered because of considerable overlap with the human genome. The Arabidopsis and human reference genomes were masked with the same Arabidopsis repeat library for comparison.

Copy number variation. CNV of Arabidopsis genome regions in the hybrid cell line was determined by CNV-seq¹⁹ using default parameters with a minimum sliding window of 4, p-value 0.001, and log₂ threshold 0.6. The bin size is pre-determined by CNV-seq based on the read depth. The window size was 9054 bp and 19,748 bp for the 60- and 300-day-old cultures, respectively. The log₂ ratio was obtained by normalizing with the Arabidopsis reference genome TAIR10 using data accessed from run DRR188057 of Bioproject PRJDB8649 in the Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra/>). Publicly available WGS data of HT1080 (SRR9931532) was obtained from the SRA database and used as the reference for both 60-day and 300-day hybrid cells to compare copy number variation changes in the hybrid genome. The analysis output from CNV-seq was then plotted in a programming language 'R' (Fig. 2).

Mitotic cell arrest. Hybrid cells (5×10^7 per 10-cm dish) were seeded and subjected to 2.5 mM thymidine treatment after 1 day of culture. After 24 h, the medium was exchanged with DMEM supplemented with 10% FBS and 6 $\mu\text{g}/\text{mL}$ Bsd. The cells were cultured for 6 h and then recultured for 24 h in the presence of 2.5 mM thymidine. Following this, the medium was exchanged with DMEM supplemented with 10% FBS and 6 $\mu\text{g}/\text{mL}$ Bsd, and synchronized cells were cultured for 7 h, after which the synchronized cells were arrested with 0.05 $\mu\text{g}/\text{mL}$ colcemid (NACALAI TESQUE, Kyoto, Japan) for 4 h⁵³.

Chromosome preparation for FISH. Synchronized cells were spun down for 10 min at 1000 g, re-suspended in 1 mL Hypotonic Buffer N (10 mM Hepes pH 7.5, 2 mM MgCl_2 , 25 mM KCl) containing 1 mM DTT and 1 mM PMSF (added just before use), and incubated on ice for 1 h. The suspension was then centrifuged for 10 min at 500 g. Pelleted cells were fixed in 1 mL fresh Carnoy's solution (3:1 ethanol:glacial acetic acid). Then, 10 μL of chromosome suspension in Carnoy's solution was dropped onto a clean glass slide that was 25 cm below the pipet tip. Before drying out, a drop (10–15 μL) of fresh Carnoy's solution was applied to the slide to spread the chromosomes.

Preparation of probes for FISH. Fifteen probes for Arabidopsis single-copy genome regions, Arabidopsis centromere repeats (Atcen, 180 bp), and telomeric short repeats of human (TTAGGG) and Arabidopsis (TTT AGGG) were amplified by PCR to generate FISH probes^{54,55}. The primer sequences used for the PCRs are listed in Supplementary Table S4. Purified PCR products were labelled with Cy3-dUTP (GE Healthcare, Chicago, Illinois, USA) and Fluorescein-12dUTP (Roche Diagnostics GmbH, Mannheim, Germany) using a Nick Translation kit (Roche Diagnostics). The average fragment size of the probes was 100–300 bp. The Arabidopsis-specific probes are listed in Supplementary Table S4. Three compound probes also were prepared: Ch2P (contains Ch2-1–Ch2-5), Ch3P (Ch3-1–Ch3-5) and Ch5P (Ch5-1–Ch5-5) (Supplementary Table S3).

Fluorescence in situ hybridization (FISH). Slides were pre-treated in 1% formaldehyde/PBS at room temperature for 5 min followed by washing 2 times in PBS for 5 min each and rinsing briefly in water. Then, the slides were denatured in 70% formamide in $2 \times \text{SSC}$ at 70 °C for 4 min followed by dehydration in 70% then 100% cooled ethanol for 5 min each and air-dried. Hybridization mixtures containing labelled probes and HB buffer (50% formamide/ $2 \times \text{SSC}$ /10% dextran sulfate MW 500,000) were denatured at 80 °C for 10 min and immediately cooled on ice. Next, 10 μL of the hybridization mix was applied onto each slide, covered with a glass coverslip (18 \times 18 mm), and sealed. The slides were incubated at 37 °C for 1–2 days in a humid dark box. After hybridization, the coverslips were removed, and the slides were washed 3 times in 20% formamide in $2 \times \text{SSC}$ at 42 °C for 5 min each, followed by washing three times in $0.1 \times \text{SSC}$ at 42 °C for 5 min each. Then, the slides were washed three times in $2 \times \text{SSC}$ at room temperature for 5 min each, briefly rinsed in distilled water, and air-dried in the dark. After drying, the slides were sealed in 10–15 μL Vectashield (Vector Lab, Inc., CA, USA) containing 1.5 $\mu\text{g}/\text{mL}$ DAPI and covered with 24 \times 24 mm glass coverslips. FISH images were photographed directly using a SPOT RT3 CCD camera (SPOT imaging solutions, Inc., MI, USA) with a BX60 microscope (Olympus, Tokyo, Japan) using standard Olympus optical filter sets. Image analyses were done using ImageJ 1.52a (NIH).

Data availability

All data generated or analyzed during this study are included in this article and supplemental data.

Received: 15 October 2020; Accepted: 8 March 2021

Published online: 30 March 2021

References

- Elsevier, S. M. *et al.* Assignment of the gene for galactokinase to human chromosome 17 and its regional localisation to band q21–22. *Nature* **251**, 633–636. <https://doi.org/10.1038/251633a0> (1974).
- Boyd, Y. Characterization and use of somatic cell hybrids with interspecific translocations involving the human X chromosome. *Ann. Hum. Genet.* **51**, 13–26. <https://doi.org/10.1111/j.1469-1809.1987.tb00863.x> (1987).
- Samoilovich, S. R., Dugan, C. B. & Macario, A. J. Hybridoma technology: new developments of practical interest. *J. Immunol. Methods* **101**, 153–170. [https://doi.org/10.1016/0022-1759\(87\)90147-5](https://doi.org/10.1016/0022-1759(87)90147-5) (1987).
- Weiss, M. C. & Green, H. Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc. Natl. Acad. Sci. U. S. A.* **58**, 1104–1111. <https://doi.org/10.1073/pnas.58.3.1104> (1967).
- Tomizuka, K. *et al.* Functional expression and germline transmission of a human chromosome fragment in chimaeric mice. *Nat. Genet.* **16**, 133–143. <https://doi.org/10.1038/ng0697-133> (1997).
- Shinohara, T. *et al.* Stability of transferred human chromosome fragments in cultured cells and in mice. *Chromosome Res.* **8**, 713–725. <https://doi.org/10.1023/a:1026741321193> (2000).
- Wang, D. Y., Kumar, S. & Hedges, S. B. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc. Biol. Sci.* **266**, 163–171. <https://doi.org/10.1098/rspb.1999.0617> (1999).
- Allshire, R. C. *et al.* A fission yeast chromosome can replicate autonomously in mouse cells. *Cell* **50**, 391–403. [https://doi.org/10.1016/0092-8674\(87\)90493-4](https://doi.org/10.1016/0092-8674(87)90493-4) (1987).
- McManus, J. *et al.* Unusual chromosome structure of fission yeast DNA in mouse cells. *J. Cell Sci.* **107**(Pt 3), 469–486 (1994).
- Maximilian, H. F. J. *et al.* Large domains of heterochromatin direct the formation of short mitotic chromosome loops. *eLife* **9**, e57212. <https://doi.org/10.7554/eLife.57212> (2020).
- Kishida, M., Muguruma, T., Sakanaka, K., Katsuragi, T. & Sakai, T. Chromosomal deletion or rearrangement in chimeric hybrids of *Saccharomycopsis fibuligera* and *Saccharomyces diastaticus* obtained by cell fusion. *J. Ferment. Bioeng.* **81**, 281–285. [https://doi.org/10.1016/0922-338X\(96\)80577-0](https://doi.org/10.1016/0922-338X(96)80577-0) (1996).
- Dudits, D., Rasko, I., Hadlaczký, G. & Lima-de-Faria, A. Fusion of human cells with carrot protoplasts induced by polyethylene glycol. *Hereditas* **82**, 121–123. <https://doi.org/10.1111/j.1601-5223.1976.tb01545.x> (1976).

13. Ahkong, Q. F. *et al.* Fusion of hen erythrocytes with yeast protoplasts induced by polyethylene glycol. *Nature* **255**, 66–67. <https://doi.org/10.1038/255066a0> (1975).
14. Zepp, H. D., Conover, J. H., Hirschhorn, K. & Hodes, H. L. Human-mosquito somatic cell hybrids induced by ultraviolet-inactivated Sendai virus. *Nat. New Biol.* **229**, 119–121. <https://doi.org/10.1038/newbio229119a0> (1971).
15. Wada, N. *et al.* Maintenance and function of a plant chromosome in human cells. *ACS Synth. Biol.* **6**, 301–310. <https://doi.org/10.1021/acssynbio.6b00180> (2017).
16. Weckselblatt, B. & Rudd, M. K. Human structural variation: mechanisms of chromosome rearrangements. *Trends Genet.* **31**, 587–599. <https://doi.org/10.1016/j.tig.2015.05.010> (2015).
17. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339. <https://doi.org/10.1093/bioinformatics/bts378> (2012).
18. Smit, A. F. A., R., H. & P., G. *RepeatMasker Open-4.0.*, <http://www.repeatmasker.org/> (2013–2015).
19. Xie, C. & Tammi, M. T. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinform.* **10**, 80. <https://doi.org/10.1186/1471-2105-10-80> (2009).
20. Richards, E. J. & Ausubel, F. M. Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* **53**, 127–136. [https://doi.org/10.1016/0092-8674\(88\)90494-1](https://doi.org/10.1016/0092-8674(88)90494-1) (1988).
21. Croce, C. M. Loss of mouse chromosomes in somatic cell hybrids between HT-1080 human fibrosarcoma cells and mouse peritoneal macrophages. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 3248–3252. <https://doi.org/10.1073/pnas.73.9.3248> (1976).
22. Kucherlapati, R. S. & Ruddle, F. H. Mammalian somatic hybrids and human gene mapping. *Ann. Intern. Med.* **83**, 553–560. <https://doi.org/10.7326/0003-4819-83-4-553> (1975).
23. Johnston, S. D., Zee, Y. P., López-Fernández, C. & Gosálvez, J. The effect of chilled storage and cryopreservation on the sperm DNA fragmentation dynamics of a captive population of koalas. *J. Androl.* **33**, 1007–1015. <https://doi.org/10.2164/jandrol.111.015248> (2012).
24. Kopeika, J., Thornhill, A. & Khalaf, Y. The effect of cryopreservation on the genome of gametes and embryos: principles of cryobiology and critical appraisal of the evidence. *Hum. Reprod. Update* **21**, 209–227. <https://doi.org/10.1093/humupd/dmu063> (2015).
25. Meyne, J. *et al.* Distribution of non-telomeric sites of the (TTAGGG)_n telomeric sequence in vertebrate chromosomes. *Chromosoma* **99**, 3–10. <https://doi.org/10.1007/bf01737283> (1990).
26. Reimann, N. *et al.* Evidence that metacentric and submetacentric chromosomes in canine tumors can result from telomeric fusions. *Cytogenet. Cell Genet.* **67**, 81–85. <https://doi.org/10.1159/000133804> (1994).
27. Lee, C., Sasi, R. & Lin, C. C. Interstitial localization of telomeric DNA sequences in the Indian muntjac chromosomes: further evidence for tandem chromosome fusions in the karyotypic evolution of the Asian muntjacs. *Cytogenet. Cell Genet.* **63**, 156–159. <https://doi.org/10.1159/000133525> (1993).
28. Schubert, I., Schriever-Schwemmer, G., Werner, T. & Adler, I. D. Telomeric signals in robertsonian fusion and fission chromosomes: implications for the origin of pseudoaneuploidy. *Cytogenet. Cell Genet.* **59**, 6–9 (1992).
29. Schlötterer, C. & Tautz, D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**, 211–215. <https://doi.org/10.1093/nar/20.2.211> (1992).
30. Messier, W., Li, S. H. & Stewart, C. B. The birth of microsatellites. *Nature* **381**, 483. <https://doi.org/10.1038/381483a0> (1996).
31. Gascoigne, K. E. & Cheeseman, I. M. Induced dicentric chromosome formation promotes genomic rearrangements and tumorigenesis. *Chromosome Res.* **21**, 407–418. <https://doi.org/10.1007/s10577-013-9368-6> (2013).
32. Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J. & de Lange, T. Chromothripsis and kataegis induced by telomere crisis. *Cell* **163**, 1641–1654. <https://doi.org/10.1016/j.cell.2015.11.054> (2015).
33. Gisselsson, D. *et al.* Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5357–5362. <https://doi.org/10.1073/pnas.090013497> (2000).
34. MacKinnon, R. N., Duivenvoorden, H. M. & Campbell, L. J. Unbalanced translocations of 20q in AML and MDS often involve interstitial rather than terminal deletions of 20q. *Cancer Genet.* **204**, 153–161. <https://doi.org/10.1016/j.cancergen.2010.12.001> (2011).
35. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761–764. <https://doi.org/10.1126/science.1146484> (2007).
36. Kapitonov, V. V. & Jurka, J. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica* **107**, 27–37 (1999).
37. Friend, K. K., Dorman, B. P., Kucherlapati, R. S. & Ruddle, F. H. Detection of interspecific translocations in mouse-human hybrids by alkaline Giemsa staining. *Exp. Cell Res.* **99**, 31–36. [https://doi.org/10.1016/0014-4827\(76\)90676-5](https://doi.org/10.1016/0014-4827(76)90676-5) (1976).
38. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40. <https://doi.org/10.1016/j.cell.2010.11.055> (2011).
39. Zhang, C. Z. *et al.* Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184. <https://doi.org/10.1038/nature14493> (2015).
40. Korb, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226–1236. <https://doi.org/10.1016/j.cell.2013.02.023> (2013).
41. Gu, S. *et al.* Mechanisms for complex chromosomal insertions. *PLoS Genet.* **12**, e1006446. <https://doi.org/10.1371/journal.pgen.1006446> (2016).
42. Ly, P. & Cleveland, D. W. Rebuilding chromosomes after catastrophe: emerging mechanisms of chromothripsis. *Trends Cell Biol.* **27**, 917–930. <https://doi.org/10.1016/j.tcb.2017.08.005> (2017).
43. Jovtchev, G., Stergios, M. & Schubert, I. A comparison of N-methyl-N-nitrosourea-induced chromatid aberrations and micronuclei in barley meristems using FISH techniques. *Mutat. Res.* **517**, 47–51 (2002).
44. Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215. <https://doi.org/10.1093/nar/16.3.1215> (1988).
45. Sánchez-Martín, J. *et al.* Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biol.* **17**, 221. <https://doi.org/10.1186/s13059-016-1082-1> (2016).
46. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034. <https://doi.org/10.1093/bioinformatics/btv098> (2015).
47. Simon, A., Pierre, L., Brian, H., & Phil, E. Babraham Bioinformatics—FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2011).
48. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> (2014).
49. Li, H. *et al.* The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> (2009).
50. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842. <https://doi.org/10.1093/bioinformatics/btq033> (2010).
51. Chikhi, R. & Rizk, G. Space-efficient and exact de Bruijn graph representation based on a Bloom filter. *Algorithms Mol. Biol.* **8**, 22. <https://doi.org/10.1186/1748-7188-8-22> (2013).
52. Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**, 31–37. <https://doi.org/10.1093/bioinformatics/btt310> (2014).

53. Sramkoski, R. M. *et al.* A new human prostate carcinoma cell line, 22Rv1. *Vitro Cell. Dev. Biol. Anim.* **35**, 403–409. <https://doi.org/10.1007/s11626-999-0115-4> (1999).
54. Ijdo, J. W., Wells, R. A., Baldini, A. & Reeders, S. T. Improved telomere detection using a telomere repeat probe (TTAGGG)_n generated by PCR. *PCR*. **19**, 4780 (1991).
55. Ohmido, N. & Fukui, K. Visual verification of close disposition between a rice A genome-specific DNA sequence (TrsA) and the telomere sequence. *Plant Mol. Biol.* **35**, 963–968 (1997).

Acknowledgements

This work was supported by the “Kobe University Visiting Researcher project” from Kobe University (to CHT and NO); the ERDF project “Plants as a tool for sustainable global development” (grant number CZ.02.1.01/0.0/0.0/16_019/0000827, to PC and JD); the Mobility Plus Project No. JSPS-18-07 and JPJSBP120203507, to NO; and “Innovative Asia” by JICA (to YML). We thank Margaret Biswas, PhD, from Edanz Group (<https://en-author-services.edanzgroup.com/ac>) for editing a draft of this manuscript.

Author contributions

Y.L.; Collection and assembly of cell materials and FISH data and drafting of the article, Y.M.L.; Collection and assembly of WGS and CNV data and drafting of the article, P.C., and C.H.T.; WGS and CNV analysis and interpretation of data, J.D.; Project PI, and critical revision of the article for important discussion, N.W. and K. F.; Cell materials development, Conception and design of the study, N.O.; Project PI and critical revision of the article for important intellectual content. All authors reviewed 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/s41598-021-86130-4>.

Correspondence and requests for materials should be addressed to N.O.

Reprints and permissions information is available at 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021