The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization

Stefan Heckmann¹, Jiri Macas², Katrin Kumke¹, Jorg Fuchs¹, Veit Schubert¹, Lu Ma¹, Petr Novák², Pavel Neumann², Stefan Taudien³, Matthias Platter³ and Andreas Houben¹*

¹Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, 06466 Gatersleben, Germany, ²Biology Centre of the Academy of Sciences of the Czech Republic, Institute of Plant Molecular Biology, Branísovská 31, České Budějovice 37005, Czech Republic, and ³Leibniz Institute for Age Research – Fritz Lipmann Institute, Beutenbergstraße 11, 07745 Jena, Germany

Received 28 August 2012; revised 11 October 2012; accepted 16 October 2012; published online 12 December 2012.

*For correspondence (e-mail houben@ipk-gatersleben.de).

**SUMMARY**

In higher plants, the large-scale structure of monocentric chromosomes consists of distinguishable eu- and heterochromatic regions, the proportions and organization of which depend on a species’ genome size. To determine whether the same interplay is maintained for holocentric chromosomes, we investigated the distribution of repetitive sequences and epigenetic marks in the woodrush *Luzula elegans* (3.81 Gbp/1C). Sixty-one per cent of the *L. elegans* genome is characterized by highly repetitive DNA, with over 30 distinct sequence families encoding an exceptionally high diversity of satellite repeats. Over 33% of the genome is composed of the Angela clade of Ty1/copia LTR retrotransposons, which are uniformly dispersed along the chromosomes, while the satellite repeats occur as bands whose distribution appears to be biased towards the chromosome termini. No satellite showed an almost chromosome-wide distribution pattern as expected for a holocentric chromosome and no typical centromere-associated LTR retrotransposons were found either. No distinguishable large-scale patterns of eu- and heterochromatin-typical epigenetic marks or early/late DNA replicating domains were found along mitotic chromosomes, although super-high-resolution light microscopy revealed distinguishable interspersed units of various chromatin types. Our data suggest a correlation between the centromere and overall genome organization in species with holocentric chromosomes.

**Keywords:** holocentric genome organization, holokinetic chromosome, centromere, histone marks, repetitive DNA, *Luzula elegans*.

**INTRODUCTION**

Most studied organisms feature one single size-restricted centromere per chromosome (monocentric chromosomes), but in certain independent eukaryotic lineages, holocentric chromosomes occur (Melters *et al.*, 2012). These holocentric chromosomes lack a primary constriction, and, in contrast to monocentric chromosomes, they form holokinetic kinetochores (also called diffuse or non-localized kinetochores) that are distributed along almost the entire poleward surface of the chromatids, to which the spindle fibers attach (Guerra *et al.*, 2010; Heckmann and Houben, 2012).

Centromere functions are highly conserved between mono- and holocentric chromosome species, and similar kinetochore components have been found in the active centromeres of both types (Maddox *et al.*, 2004; Nagaki *et al.*, 2005; d’Alençon *et al.*, 2011). However, structural analysis of mitotic chromosomes in the holocentric plant genus *Luzula* challenged the notion of a ‘diffuse’ centromere organization along holocentric chromosomes (Nagaki *et al.*, 2005; Heckmann *et al.*, 2011). Instead, a longitudinal centromere-like groove that was positive for CENH3 [a mark for active centromeres (Kalitsis and Choo, 2012)] was found along each sister chromatid, discontinued at each sub-terminal end. Consistently, entire mitotic chromosomes of *Luzula* (Gernand *et al.*, 2003; Nagaki *et al.*, 2005) and *Rhyynchospora tenuis* (Guerra *et al.*, 2006) displayed a cell cycle-dependent uniform histone H3S10/S28 phosphorylation mark, illustrating a chromosome-wide ‘pericentromeric-like’ structure (Houben *et al.*, 2007a).

The DNA of centromeres is highly variable, and, except for budding yeast (Clarke and Carbon, 1985), the sequences are neither necessary nor sufficient for centromere formation. However, satellite DNA repeats and
specific families of long terminal repeat (LTR) retrotransposons are usually associated with centromeres in monocentric plant species (Houben and Schubert, 2003; Neumann et al., 2011). In contrast, for holocentric species, centromere-specific DNA sequences have not yet been reported (Gassmann et al., 2012).

In addition to centromeres, heterochromatin-forming repeats are typically found at telomeres, nucleolar organizing regions, sub-terminal and interstitial regions in monocentric species (Schmidt and Heslop-Harrison, 1998). The genome organization is reflected by the distribution of epigenetic marks, such as DNA methylation and post-translational histone modifications. Typically, methylation of lysine residues 9 and 27 of histone H3 corresponds to heterochromatin, while euchromatin is marked by methylation of lysine residues 4 and 36 of histone H3 (Fuchs et al., 2006). In most monocentric species with small genomes (1C < 500 Mbp), e.g. Arabidopsis, strong dimethylation of H3K9 is primarily limited to pericentromeric heterochromatin, while larger genome species show a uniform H3K9me2 distribution. In contrast, dimethylation of H3K4 is exclusively enriched at euchromatic regions along chromosome arms in monocentric species, regardless of their genome size. This observation suggests that genome size in monocentric species is a factor that significantly influences the global distribution of histone methylation marks at transcriptionally less active regions (Houben et al., 2003; Fuchs et al., 2006). However, studies on the chromosomal distribution of typical euchromatin and heterochromatin histone marks are lacking in holocentric plants.

In terms of the inter-relationship between centromere organization and chromosome structure, we wished to determine whether the higher-order composition of a holocentric chromosome displays the same characteristics as a monocentric one, and to address this question, we selected the woodrush Luzula elegans Lowe (formerly L. purpurea) as a model species, due to the low number and large size of its holocentric chromosomes (Heckmann et al., 2011). Illumina sequencing, combined with bioinformatic and cytogenetic approaches, revealed a unique genome organization. In addition, an interspersed arrangement of eu- and heterochromatin marks, and of early and late replicating DNA, was found on the large scale. Our findings suggest interplay between the centromere and large-scale genome organization, and will therefore assist in the understanding of holocentric genome organization and its implications for genome evolution and centromere biology.

RESULTS

The repetitive DNA fraction of the L. elegans genome

The DNA content of nuclei was determined to be 7.80 pg per 2C using flow cytometry, corresponding to a 1C genome size of 3.81 Gbp, which is in close agreement with the measurements of Barlow and Nevin (1976). To obtain insight into sequence composition of repetitive sequences in this relatively large genome, high-throughput shotgun sequencing was performed on the Illumina GAIIx platform. The original Illumina sequencing data are available under study accession number ERP001569 (http://www.efd.ac.uk/ena/data/view/ERP001569) at the Sequence Read Archive (http://www.efd.ac.uk/ena/). A randomly sampled proportion (1.5 million) of generated reads was then subjected to bioinformatical analysis, implemented within the clustering-based repeat identification pipeline (Novak et al., 2010). This analysis resulted in thousands of clusters, or groups of reads, with overlapping sequences, each representing a single repeated element or part of it. Following repeat classification within major clusters, the global repeat composition of the genome was determined by taking into account the sizes (number of reads) of individual clusters, which are proportional to the genomic abundance of corresponding repeats. The L. elegans genome was found to be rich in repeated sequences, with highly and moderately repeated elements represented by clusters with genome proportions of at least 0.01% and collectively making up 61% of the genome. The majority of these sequences were classified into established groups of repetitive elements, revealing the Angela clade of Ty1/copia LTR retrotransposons to be a dominant repeat representing over 33% of the genome (Table 1). Except for the satellite repeats, the proportions of all other repeat types, including various groups of LTR retrotransposons, did not exceed a few per cent of the genome. Some repeats were completely absent, including the CRM clade of Ty3/gypsy retrotransposons that are known to be specifically associated with plant centromeres (Neumann et al., 2011). The observed proportion of plastid DNA reads (3.4%) most likely originated from contamination of nuclear DNA preparations by the chloroplast genome, because no signals were detectable on chromosomes by fluorescence in situ hybridization (FISH) using plastid DNA as a probe (data not shown), although we cannot rule out the possibility of low-copy insertions of plastid DNA into the nuclear genome.

Table 1 Repetitive DNA composition of the L. elegans genome

<table>
<thead>
<tr>
<th>Type of repeat</th>
<th>Proportion of the genome (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ty1/copia, Angela clade</td>
<td>33.4</td>
</tr>
<tr>
<td>Ty1/copia, Maximus clade</td>
<td>0.9</td>
</tr>
<tr>
<td>Ty3/gypsy</td>
<td>1.1</td>
</tr>
<tr>
<td>LTR unclassified</td>
<td>2.0</td>
</tr>
<tr>
<td>Long interspersed elements (LINEs)</td>
<td>0.3</td>
</tr>
<tr>
<td>DNA transposons</td>
<td>1.1</td>
</tr>
<tr>
<td>rDNA</td>
<td>0.2</td>
</tr>
<tr>
<td>Satellites</td>
<td>9.9</td>
</tr>
</tbody>
</table>
Although the overall genome proportion of satellite DNA was not exceptional compared to other plant species, there was an extraordinary sequence diversity of tandem repeats found in the *L. elegans* genome. Thirty-seven families of putative satellite repeats differing in their monomer lengths and sequence composition were identified among 291 major repeat clusters (genome proportions of at least 0.01%). The characteristics of the 20 largest satellite DNA clusters are provided in Table 2, and their dot-plot sequence comparison is shown in Figure S1 (assembled contigs representative for each cluster are listed in Data S1). Although some of the identified families had monomer sizes in the range of hundreds of nucleotides, which is typical for the majority of known plant satellites (Macas *et al.*, 2002), 12 of the 20 families had shorter monomer sequences, ranging from 90 bp down to the size of microsatellite repeats (Table 2). The tandem genomic organization of some of the identified satellite repeats was verified by Southern hybridization to partially digested genomic DNA, revealing typical ladder-like hybridization patterns as demonstrated for LeSAT4, LeSAT11 and LeSAT9 + 21 (Figure S2).

**Satellite repeats tend to localize at chromosome termini**

Fluorescence *in situ* hybridization probes (Table S1) derived from conserved regions of the 20 selected satellite repeats (Table 2) were used to investigate their distribution on mitotic chromosomes. All of them provided signal patterns consisting of multiple discrete bands, or spots, that are typical of satellite repeats organized in long monomer arrays. Most satellites were detected on all three pairs of chromosomes; although chromosome-specific families were also observed labeling two chromosome pairs (LeSAT23, 25, 27, 28, and 99), or one chromosome pair only (LeSAT43 and 63). There were also families that labeled particular chromosome regions, i.e. LeSAT7, 11 and 109, which labeled the terminal regions, and LeSAT4, which clustered on each chromosome at almost symmetrical distal positions. As all chromosomes are equally sized, and are not distinguishable by their morphology, two probes, LeSAT28 and LeSAT63, were combined to provide a hybridization pattern facilitating their discrimination. LeSAT28-specific signals occurred distally on chromosome 1, opposite the 45S rDNA locus, and (sub)terminally on chromosome 2 (Figure 1a). LeSAT63 localized interstitially on chromosome 2, being slightly more distal towards the LeSAT28-bearing chromosome region (Figure 1a). Although the minor LeSAT28 signal on chromosome 2 was not always detectable, the combination of LeSAT28 and 63 allowed discrimination of all chromosomes. The resulting karyogram based on these probes is shown in Figure 1(b).

None of the tested satellite repeats displayed a distribution pattern expected for the almost chromosome-wide distribution of the centromere in *Luzula* (Figure 1 and Figure S3). Instead, we noticed a tendency for distal satellite clustering, especially at the centromere-free chromosome termini. To confirm this observation, distinct FISH signals of all satellites were classified into centromeric and

<p>| Table 2 Characteristics of the most abundant satellite repeats in the <em>L. elegans</em> genome |
|---------------------------------|---------------------------------|-------------------|-------------------|-------------------|</p>
<table>
<thead>
<tr>
<th>Satellite</th>
<th>Cluster</th>
<th>Cluster size (reads)</th>
<th>Genome proportion (%)</th>
<th>Monomer (bp)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>LeSAT4</td>
<td>4</td>
<td>33.681</td>
<td>2.25</td>
<td>190/220/360</td>
<td>Several closely related sequence variants differing in monomer length</td>
</tr>
<tr>
<td>LeSAT7</td>
<td>7</td>
<td>17.900</td>
<td>1.19</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>LeSAT9 + 21</td>
<td>9 + 21</td>
<td>21.068</td>
<td>1.40</td>
<td>43</td>
<td>LeSAT9 and LeSAT21 contain two highly similar sub-families of the same satellite</td>
</tr>
<tr>
<td>LeSAT11</td>
<td>11</td>
<td>13.038</td>
<td>0.87</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>LeSAT16</td>
<td>16</td>
<td>7495</td>
<td>0.50</td>
<td>178/195/...</td>
<td>Heterogeneous repeat, contains several sub-families differing in monomer size</td>
</tr>
<tr>
<td>LeSAT17</td>
<td>17</td>
<td>7210</td>
<td>0.48</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>LeSAT18</td>
<td>18</td>
<td>6156</td>
<td>0.41</td>
<td>Variable</td>
<td>Non-homogenized tandem repeat, contains tandem sub-repeats of various lengths (tens of bp)</td>
</tr>
<tr>
<td>LeSAT22</td>
<td>22</td>
<td>5262</td>
<td>0.35</td>
<td>51/167</td>
<td></td>
</tr>
<tr>
<td>LeSAT23</td>
<td>23</td>
<td>4721</td>
<td>0.31</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>LeSAT25</td>
<td>25</td>
<td>4084</td>
<td>0.27</td>
<td>6</td>
<td>Simple sequence repeat (CATAAAA)$_n$</td>
</tr>
<tr>
<td>LeSAT27</td>
<td>27</td>
<td>3946</td>
<td>0.26</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>LeSAT28</td>
<td>28</td>
<td>3815</td>
<td>0.25</td>
<td>390/730</td>
<td></td>
</tr>
<tr>
<td>LeSAT36</td>
<td>36</td>
<td>3031</td>
<td>0.20</td>
<td>6</td>
<td>Simple sequence repeat (CATAAC)$_n$</td>
</tr>
<tr>
<td>LeSAT38</td>
<td>38</td>
<td>2846</td>
<td>0.19</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>LeSAT43</td>
<td>43</td>
<td>2494</td>
<td>0.17</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>LeSAT63</td>
<td>63</td>
<td>1433</td>
<td>0.10</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>LeSAT72</td>
<td>72</td>
<td>1192</td>
<td>0.08</td>
<td>4</td>
<td>Simple sequence repeat (CATA)$_n$</td>
</tr>
<tr>
<td>LeSAT89</td>
<td>89</td>
<td>945</td>
<td>0.06</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>LeSAT99</td>
<td>99</td>
<td>816</td>
<td>0.05</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>LeSAT109</td>
<td>109</td>
<td>710</td>
<td>0.05</td>
<td>33</td>
<td>Partial similarity to LeSAT99</td>
</tr>
</tbody>
</table>
non-centromeric clusters (Table 3). To ensure a clear distinction between centromeric and non-centromeric chromosome regions, only clusters within the terminal 5% of chromosome ends were counted as non-centromeric regions, i.e. both terminal regions correspond to approximately 10% of the total chromosome length. Note, the centromere discontinues at each sub-terminal chromosome end and represents on average 75% of metaphase chromosome length as shown by CENH3-immunolabeling and scanning electron microscopy, respectively (Figure S3; Heckmann et al., 2011). Thus, the percentage of non-centromeric clusters may be under-estimated. Of the 122 identified clusters of satellites, 39 (approximately 32%) localized to terminal non-centromeric chromosome regions accounting for approximately 10% of the total chromosome size. Asterisks indicate FISH signals counted as terminal (non-centromeric) satellite clusters (see Table 3). Scale bar = 10 μm.

Figure 1. Distribution of satellite DNA on L. elegans mitotic metaphase chromosomes studied by FISH. (a) LeSAT28 and LeSAT63 (red) localized individually in relation to a 45S rDNA probe (green), and relative to each other (green, LeSAT28; blue, LeSAT63). (b) Karyogram based on probes used in (a), allowing discrimination of the three equally sized chromosomes. (c) FISH experiments with various LeSATS (pink) together with LeSAT28 (green) and LeSAT63 (blue), and the resulting schematic ideograms. DAPI-stained DNA is shown in gray. The gray-shaded chromosome ends in schematic ideograms represent defined terminal (non-centromeric) chromosome regions accounting for approximately 10% of the total chromosome size. Asterisks indicate FISH signals counted as terminal (non-centromeric) satellite clusters (see Table 3). Scale bar = 10 μm.
regions, while 83 (approximately 68%) localized to interstitial centromeric regions (Table 3). The amount of non-centromeric clusters varied between chromosomes, i.e. from approximately 27% on chromosome 2 to approximately 38% on chromosome 1 (Table 3). As the terminal non-centromeric regions represent only approximately 10% of the total chromosome length, the abundance of satellite clusters in these regions is on average 4.2-fold higher than in interstitial centromeric regions.

In contrast to the satellite repeats, the probes derived from mobile elements showed uniformly dispersed chromosome patterns (Figure 2). The probes were either prepared from shotgun-cloned *L. elegans* genomic fragments, corresponding to partial sequences of the Ty1/copia element Angela (Table S2), or from PCR products using primers based on the contig CL8c28 (assembled contigs are listed in Data S1) corresponding to LTR and gag–pol regions of elements from the chromovirus clade of Ty3/gypsy retrotransposons (Table S1). A clone containing an abundant microsatellite motif \([\text{TA}]_n\) generated a similar pattern (Table S2).

### Table 3 Satellites are proportionally more frequent in non-centromeric than in centromeric chromosomal regions

<table>
<thead>
<tr>
<th></th>
<th>Chromosome 1</th>
<th>Chromosome 2</th>
<th>Chromosome 3</th>
<th>All chromosomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of clusters</td>
<td>42</td>
<td>37</td>
<td>43</td>
<td>122</td>
</tr>
<tr>
<td>Number of centromeric clusters (%)</td>
<td>26 (61.9)</td>
<td>27 (73)</td>
<td>30 (69.8)</td>
<td>83 (68)</td>
</tr>
<tr>
<td>Number of non-centromeric cluster (%)</td>
<td>16 (38.1)</td>
<td>10 (27)</td>
<td>13 (30.2)</td>
<td>39 (32)</td>
</tr>
<tr>
<td>Ratio of relative cluster abundance in non-centromeric versus centromeric regions*</td>
<td>5.5:1</td>
<td>3.3:1</td>
<td>3.9:1</td>
<td>4.2:1</td>
</tr>
</tbody>
</table>

FISH signals of all studied satellites are classified into centromeric and non-centromeric clusters (see Figure 1). Their abundance per chromosome is given. Defined centromeric and non-centromeric chromosome parts represent approximately 90% and 10% of the total chromosome length, respectively (see Figure 1).

*The relative cluster abundance is calculated based on the absolute values of satellite clusters in non-centromeric and centromeric regions given their different lengths.

Ty3/gypsy retrotransposons (Table S1). A clone containing an abundant microsatellite motif \([\text{TA}]_n\) generated a similar pattern (Table S2).

### Uniformly mixed genome organization of *Luzula* on a large scale, with defined subunits at a lower chromatin organization level

To examine whether the genome of *L. elegans* is organized on a large scale into eu- and heterochromatin-enriched sub-regions, we used antibodies against typical euchromatin- and heterochromatin-specific histone marks. In contrast to previous reports for monocentric species (Houben *et al.*, 2003), immunostaining with a euchromatin-specific antibody such as H3K4me2 revealed uniform labeling of the entire chromosome and interphase nuclei (Figure 3).

![Figure 2](https://example.com/image2.png)

**Figure 2.** Distribution of LTR retrotransposons on *L. elegans* mitotic chromosomes. Probes were derived from sequences of Angela-like elements and chromoviruses, representing the most abundant clades of Ty1/copia and Ty3/gypsy elements in the *L. elegans* genome, respectively. Note some interstitial centromere-atypical clustering of Ty3/gypsy elements indicated by arrows. Two Ty3/gypsy-specific probes representing either the gag-pol (CL8c28_1) or LTR (CL8c28_2) regions showed similar patterns. CL8c28_1 (see Table S1) and clone 5 (see Table S2) are shown as Ty3/gypsy- and Angela-specific probes, respectively. In the merged image, DAPI-stained DNA is shown in blue and LTR retrotransposons appear red. Scale bar = 10 μm.

![Figure 3](https://example.com/image3.png)

**Figure 3.** Distribution of histone H3K4me2, H3K9me2 and H3K27me3 in *L. elegans*. Double immunolabeling with H3K4me2 and H3K27me3, as well as with H3K9me2 and H3K27me3, for interphase nuclei and mitotic chromosomes. High-resolution imaging by structured illumination microscopy (SIM, centre and right) compared to normal fluorescence wide-field microscopy (WFM, left). Enlargements of the regions delimited by the white boxes are shown. DNA appears blue and histone marks appear red and green. Scale bar = 1 μm.
A comparable distribution was found for the heterochromatin marks H3K9me2 and H3K27me3 (Figure 3), as well as for methylated DNA (Figure S4).

Next, immunolabeled cells were analyzed by structured illumination microscopy (SIM) to obtain a higher optical resolution of eu- and heterochromatin-specific immunosignals at approximately 100 nm. Various sub-domains were detected: hetero- or euchromatin-containing domains and intermingled eu- and heterochromatin domains (Figure 3). The two heterochromatin marks H3K9me2 and H3K27me3 also revealed a subunit organization into chromatin units that shared both marks, or were free of one or the other or both marks (Figure 3). Interphase nuclei showed either uniform chromatin or DAPI-enriched ‘chromocenters’, most likely based on the cell cycle or cell type. However, at the chromosomal level, chromatin is mostly uniformly shuffled on a global view. Thus, *L. elegans* appears to represent an interspersed arrangement of euchromatic and heterochromatic domains on a large scale, while a distinct chromatin sub-organization into chromatin subunits is detectable using super-high-resolution light microscopy.

The DNA replication behavior was studied to test whether early- and late-replicating chromosome regions occupy distinct chromosomal regions, as in monocentric species (Costas et al., 2011). *Luzula* seedlings were therefore treated with 5-ethynyl-2’–deoxyuridine (EdU) for 5–135 min, and then seedlings were incubated in the absence of EdU so that the cells in S phase progressed to mitosis. The approximate time from S phase until entry into mitosis was estimated to be between 7.5 and 9 h, similar to other studies (Bernardini and Lima-de-Faria, 1967). Independently of the length of the EdU pulse, almost uniformly labeled chromosomes and nuclei were found (Figure 4 and Figure S5). Previous reports in *L. elegans* also described more or less randomly distributed DNA (late-) replicating sites along chromosomes (Bernardini and Lima-de-Faria, 1967; Ray and Venketeswaran, 1979). However, in the case of nuclei, a sub-fraction of 192 EdU-positive nuclei after a 5 min EdU pulse were found to be labeled with spots. A variable spotted pattern occurred in 29.8% of counted nuclei, while the remaining ones showed uniform labeling (Figure 4). Thus, it seems likely that replication clusters may occur, depending on the stage of the cell cycle or the cell type, as also indicated by the ‘chromocenters’ observed by SIM (Figure 3). However, a mostly dispersed and equally distributed replication pattern was observed for all chromosome pairs, and, on the large scale, the chromosomes of *L. elegans* are not compartmentalized into clearly distinguishable early- and late-replicating sub-regions. It is likely that early- and late-replicating sequences are equally distributed along the chromosomes. This result probably reflects the uniformly mixed genome organization, as also indicated by eu- and heterochromatin-specific marks.

DISCUSSION

**Do terminal satellites play a role in holocentric chromosome structure and function?**

We performed a comprehensive characterization of the repetitive fraction of a holocentric plant genome. The content of highly repetitive DNA in *L. elegans* (61% of the 3.81 Gbp/1C-sized haploid genome) is in the same range as in monocentric plants with a comparable genome size that were analyzed with the same method (i.e. *Silene latifolia* males (2.93 Gbp/1C), 61.4%; females (2.87 Gbp/1C), 63.3% (Macas et al., 2011), and *Pisum sativum* (4.3 Gbp/1C), 35-48% (Macas et al., 2007)). Holocentric species with smaller genomes contain less repetitive DNA, such as nematodes (i.e. *Caenorhabditis briggsae* (104 Mbp), 22.4%; *C. elegans* (100.3 Mbp), 16.5% (C. elegans Sequencing Consortium, 1998; Stein et al., 2003)) and *Bombbyx mori* (432 Mbp, 43.6% (Xia et al., 2008)). Thus, in holocentric species, the content of highly repetitive DNA is also greater in species with larger genomes.

Transposable elements largely account for genome size differences in plants (Hawkins et al., 2006; Tenaillon et al., 2011). In *L. elegans*, Ty1/copia elements are much more abundant than Ty3/gypsy elements, with the Angela clade elements the most abundant Ty1/copia element and constituting by far the most abundant repetitive fraction in the genome. *L. elegans* contains a rather large and heterogeneous fraction of satellites compared to plants with a
comparable genome size. In *S. latifolia*, only four families of satellite DNA were identified using the same in silico method (Macas et al., 2011). It is likely that *L. elegans*-specific amplification of Ty1/copia elements from the Angela clade and of satellite repeats is the driving force behind the exceptionally large genome size of the genus *Luzula*. Other members of the genus are characterized by smaller genomes, ranging from 0.26 to 1.99 Gbp/1C (Bennett and Leitch, 2010).

The distal chromosome regions of *L. elegans* tend to be enriched in satellite DNA. Although some central blocks are found, the preferential localization of repetitive DNA and of heterochromatin in (sub)terminal regions seems to be a common feature of holocentric plants (Sheikh and Kondo, 1995; Vanzela and Guerra, 2000; Guerra and Garcia, 2004), and also of holocentric autosomes in animals (*C. elegans* Sequencing Consortium, 1998; Spence et al., 1998; Tartarotti and de Azeredo-Oliveira, 1999; Mola and Papeschi, 2006; Hill et al., 2009). In contrast, in many monocentric organisms, high-copy repeats and heterochromatin typically cluster at centromeres and various other sites (Schmidt and Heslop-Harrison, 1998; Mola and Papeschi, 2006; Lamb et al., 2007). Surprisingly, in *L. elegans*, no correlation between enriched terminal satellites and epigenetic chromatin modifications was found, similar to *C. elegans*, in which heterochromatin is cytologically absent (Albertson et al., 1997). It seems likely that, in these situations, enriched repressive domains are interspersed with active chromatin, and therefore no enrichment of heterochromatin-typical marks was detectable.

It is an obvious question whether a functional inter-relation-ship exists between terminally enriched satellite DNA and holocentricity. First, mutual exclusion of heterochromatin and centromere function may account for repetitive DNA accumulation in non-centromeric chromosome ends, as speculated for the nucleolar organizing region (Heckmann et al., 2011). Secondly, in holocentric species, terminal heterochromatin may be involved in the physical end-to-end association of homologous chromosomes (rod bivalents) during meiotic divisions (Nordenskiold, 1962; Dernburg, 2001; Bongiorni et al., 2004; Guerra et al., 2010; Heckmann and Houben, 2012).

Are there centromere-specific sequences in holocentric species?

The low abundance of Ty3/gypsy repeats and the absence of typical centromere-associated Ty3/gypsy retrotransposons of the CRM clade in *L. elegans* were surprising. Even an in-depth screen for low-copy chromdomain sequences of all available 21.4 million sequence reads revealed no significant hits for CRM elements, which typically colonize centromeric regions in monocentric plants (Neumann et al., 2011). Moreover, none of the identified satellite repeats gave rise to a distribution pattern that was expected for the almost chromosome-wide centromere distribution.

In *Luzula nivea*, a 178 bp tandem repeat sequence (LCS1; also present in nine other *Luzula* species) (Haizel et al., 2005) sharing similarity with the centromeric tandem repeat RCS2 of rice (Dong et al., 1998; Nonomura and Kurata, 2001) has been described. LCS1 clusters into tandem arrays of at least 50 kb at heterochromatic regions along each of the *L. nivea* chromosomes (Haizel et al., 2005). Whether LCS1 plays a centromeric role is uncertain. Interestingly, no LCS1-related sequences were found in a BLAST similarity search of all available 21.4 million *L. elegans* sequence reads in this study.

Given the sequence-independent incorporation of CENH3 in *C. elegans* (Gassmann et al., 2012), the sequence-independent formation of centromeres in nematodes (Howe et al., 2001), and the absence of centromeric sequences even in the genome-sequenced holocentric animals *C. elegans* (Gassmann et al., 2012) and *B. mori* (Xia et al., 2008; d’Alencon et al., 2010), it is also likely that no typical centromeric sequences exist in *L. elegans*. More likely, a centromere-specific chromatin status exists that preferentially associates with CENH3 in *L. elegans*. Even in monocentric plants and animals, the occurrence of neocentromeres demonstrates that the centromeric DNA sequence itself is neither necessary nor sufficient to determine centromere function, and centromeres are determined epigenetically (Guerra et al., 2010; Kalitsis and Choo, 2012).

Holocentric chromosomes: multiple sequence-independent centromeric subunits along chromosomes

A centromeric subunit organization as originally proposed by Zinkowski et al. (1991) for monocentric eukaryotes probably also applies for holocentric species, as *C. elegans* shows a centromeric subunit organization along holocentric chromosomes (Gassmann et al., 2012) as well as *L. elegans*, as indicated by the mixed genome organization at the large scale, and the chromatin subunit organization at a higher resolution. A dispersed CENH3 distribution (dot-like foci) during interphase and prophase is found in *Luzula* (Nagaki et al., 2005; Heckmann et al., 2011) and *C. elegans* (Buchwitz et al., 1999), indicating multiple centromeric subunits that fuse during metaphase to one functional kinetochore unit. Thus, holocentric chromosomes appear to be composed of multiple centromere units interspersed by non-centromeric chromatin all along their length (Figure 5).

Neumann et al. (2012) showed that the tandem organization of multiple satellites, but not their primary sequence, determines the presence of remarkably large functional centromeres in pea that are ‘intermediate between monocentric and polycentric’. Unlike most other monocentric species (Hall et al., 2004), the centromeric DNA sequence...
composition is highly variable in pea. Thirteen of 19 distinct satellite families, and one family of Ty3/gypsy retrotransposons of the CRM clade, are centromere-associated (Neumann et al., 2012). Thus, pea is the most similar studied organism so far with respect to the number and diversity of satellites. It is tempting to speculate that the larger a centromere gets, the more important the centromere-specific chromatin status or organization becomes, and the less important the primary DNA sequence is. Even in holocentric organisms, there are apparently multiple sequence-independent centromeric subunits along the chromosomes (Gassmann et al., 2012). However, dicentric chromosomes break during anaphase when they reach a critical distance between both active centromeres, thus not forming one functional kinetochoore during metaphase (Higgins et al., 2005; Zhang et al., 2010). The distance between individual kinetochore subunits in holocentrics must therefore be restricted. However, it is not known what the minimal distance between individual kinetochore subunits is for formation of a composite linear-like kinetochore.

Is there interplay between centromere and large-scale genome organization?

Early reports in *L. elegans* described more or less evenly distributed heterochromatin, by C-banding, with a slight tendency for interstitial and terminal heterochromatin clusters and multiple randomly distributed DNA (late-)replicating sites (Bernardini and Lima-de-Faria, 1967; Ray and Venketeswaran, 1978, 1979). We found an overlap of eu- and heterochromatin-specific histone marks, homogenously dispersed DNA methylation and mobile elements along chromosomes on a global view, as well as equally distributed early- and late-replicating sequences all along *L. elegans* chromosomes. Other *Luzula* species are also characterized by dispersed heterochromatin and repetitive DNA along chromosomes (Collet and Westerman, 1984, 1987). Studies in *Luzula flaccida* (Collet and Westerman, 1984) and holocentric animals (Albertson et al., 1997; d’Alencon et al., 2010; Liu et al., 2011) also indicate an interspersed arrangement of eu- and heterochromatic domains along chromosomes.

The almost chromosome-wide distribution of the centromere into multiple centromeric subunits is the most likely reason why holocentric species are characterized by an inter-mixed arrangement and homogenization of chromatin states throughout their genomes. Thus, it is plausible to speculate that the large-scale genome organization differs between monocentric and holocentric species (Figure 5), although further studies in holocentric species are required to verify this hypothesis.

**EXPERIMENTAL PROCEDURES**

**Plant material and culture conditions**

*Luzula elegans* (*2n* = 6) plants (Vouchers at the Herbarium Gatersleben; GAT 7852-7856) were germinated in Petri dishes on wet filter papers under long-day conditions (16 h light/8 h dark, 20°C/18°C), transferred to soil, cultivated for 6-8 weeks under short-day conditions (8 h light/16 h dark, 20°C/18°C), transferred to vernalizing conditions (10 h light/14 h dark, 4°C), and finally returned to long-day conditions (13 h light/11 h dark, 20°C/16°C).

**Genomic DNA isolation, Southern and dot-blot hybridization**

Genomic DNA was isolated using a DNeasy plant max kit (Qiagen, http://www.qiagen.com), according to the manufacturer’s instructions, and, if required, digested using BamHI, EcoRI, HpaII or MspI, electrophoresed in a 0.8% w/v agarose gel and transferred by alkaline transfer to Hybond N+ membranes. Southern and dot-blot hybridization were performed as described by Houben et al. (1996). The membranes were probed with 32P-labeled DNA probes.

**Illumina sequencing**

The genomic DNA was fragmented by nebulization, and a paired-end library with fragment lengths of approximately 500 bp was prepared according to the manufacturer’s instructions (Illumina, http://www.illumina.com). Sequencing was performed on one lane
of an Illumina Genome Analyzer IIx system using Illumina’s paired-end cluster generation and cycle sequencing kits.

**Repeat identification and characterization**

Identification of repetitive sequences was achieved using similarity-based clustering analysis of sequence reads as described by Novak et al. (2010). The analysis was performed using a set of 1.5 million reads randomly selected from a total of 21.4 million high-quality Illumina forward sequence reads. A similarity cut-off of 90% over at least 80% of the read length was used for the clustering, and the reads within individual clusters were assembled and further investigated using a set of custom-written BioPerl and R scripts to determine which type and family of repeats they represent (Macas et al., 2007; Novak et al., 2010). Clusters containing satellite repeats were identified based on the presence of tandem sub-repeats within their read or assembled contig sequences. These satellite repeats were characterized using oligonucleotide frequency analysis of the reads within their clusters as described previously (Macas et al., 2010).

**Plasmid library construction and dot blotting**

Genomic DNA was fragmented by sonication. Jagged ends were removed using BAL-31 exonuclease (New England BioLabs, http://www.neb.com). DNA was purified and electrophoresed in a 1% w/v agarose gel. A 150–600 bp fraction was excised, and the ends were repaired by T4 DNA polymerase and Klenow fragment activity. Repaired fragments were A-tailed by Taq polymerase treatment. Repaired fragments were repaired by T4 DNA polymerase and Klenow fragment activity. repaired fragments were repaired by T4 DNA polymerase and Klenow fragment activity. repaired fragments were repaired by T4 DNA polymerase and Klenow fragment activity. repaired fragments were repaired by T4 DNA polymerase and Klenow fragment activity.

**Flow cytometric genome size measurement**

Genome size was estimated as described previously (Fuchs et al., 2008) using *Pisum sativum* cv. Viktoria, Kifejto Borsó (Genebank Accession number PIS 630; 2C = 9.09 pg) (Dolezel et al., 1998), as an internal reference standard. In total, 21 individuals were measured, divided into three independent experiments performed on different days.

**Indirect immunolabeling**

Apical meristems of young plants were fixed for 45 min in ice-cold 4% w/v paraformaldehyde in PBS. After washing 3 times for 10 min in ice-cold PBS, chromosome spreads were prepared by squashing. Immunolabeling was performed as previously described (Houben et al., 2007b). The following dilutions of primary antibodies were used: 1:200 for a rabbit anti-H3K4me2 antibody (Millipore, http://www.millipore.com), 1:200 for a rabbit anti-H3K9me2 antibody (Active Motif, http://www.activemotif.com), 1:50 for a mouse anti-H3K27me3 antibody (Abcam, http://www.abcam.com) and 1:100 for a rabbit anti-LnCENH3 antibody (Nagaki et al., 2005). Immunodetection using 5-methylcytosine (1:300 dilution) (Eurogentec, http://www.eurogentec.com) was performed as described by Marques et al. (2011). A Cy3-conjugated anti-rabbit IgG (Dianova, http://www.dianova.com) and a fluorescein isothiocyanate-conjugated anti-mouse Alexa 488 antibody (Molecular Probes, http://www.invitrogen.com), each at 1:400 dilution, were used as secondary antibodies. Fluorescence images were recorded using an Olympus BX61 microscope (Olympus, http://www.olympus.com) equipped with an ORCA-ER CCD camera (Hamamatsu, http://www.hamamatsu.com). 3D deconvolution microscopy was used to reduce out-of-focus information for globular structures. Image stacks of 10–11 slices per specimen were acquired, and the maximum-intensity projections were processed using the program AnalySIS (Soft Imaging System, http://www.soft-imaging.net). Grey-scale images were pseudocolored using Adobe Photoshop (http://www.adobe.com). To achieve an optical resolution of approximately 100 nm, structured illumination microscopy (SIM) was applied using a C-Apo 63×/1.2 W Korr objective with an Elyra microscope system (Zeiss, http://www.zeiss.com).

**Probe preparation and fluorescence in situ hybridization**

FISH probes were obtained as 5′-Cy3, 5′-Cy5- or 5′-Alexa 488-labeled oligonucleotides (Eurofins MWG Operon, http://www.eurofinsdna.com), or were PCR-amplified. The sequences of all oligonucleotides are listed in Table S1. Nuclear ribosomal 45S rDNA was probed using the clone pTa71 (Gerlach and Bedbrook, 1979), and plastid DNA was probed using the clone HVVMRXALLhC0205G01 (Schulte et al., 2011), respectively. All DNA probes, except oligonucleotides, were labeled with Texas Red-, Cy5- or Alexa 488-dUTP by nick translation as described by Kato et al. (2008).

Chromosome spreads were prepared from apical meristems fixed using 3:1 v/v ethanol/acetic acid. Specimens were dehydrated in an ethanol series, air-dried, and cross-linked using an UV-light illuminator (0.12 J cm⁻²) (Bioteram, http://www.bioteram.com). Probe(s) were then mixed with the hybridization mixture (50% formamide and 20% dextran sulfate in 2× SSC), dropped onto slides, covered with a cover slip, and sealed using ‘fixogum’ (Marabu, http://marabu-kreativ.de). After denaturation on a heating plate at 80°C for 3 min, slides were hybridized at 37°C overnight. Post-hybridization washing was performed in 2× SSC for 20 min at 58°C. After dehydration in an ethanol series, 4′,6-diamidino-2-phenylindole (DAPI) in Vectashield (Vector Laboratories, http://www.vectorlabs.com) was applied. Fluorescence images were obtained as described above.

**DNA replication analysis**

Eighteen-day-old seedlings were incubated in 20 μM 5-ethyl-2′-deoxyuridine (EdU) in H₂O for varying times, followed by multiple washing steps in H₂O allowing EdU-incorporated cells to enter mitosis, and then fixed in 3:1 v/v ethanol/acetic acid overnight. Slides were prepared as described above for apical meristems. EdU incorporation was detected by the click reaction, using the Click-iT EdU Imaging Kit (Invitrogen). After washing in 2× SSC, DAPI in Vectashield (Vector Laboratories) was applied. Fluorescence images were obtained as described above.

**ACKNOWLEDGEMENTS**

We are grateful to Oda Weiß, Marlene Zimmer and Susann Schier (IPK, Department Cytogenetics and Genome Analysis, Gatersleben) for their excellent technical assistance. We also thank Maria Cuacos (Departamento de Genética, Facultad de Biología, Universidad Complutense, Madrid) for help with the Southern experiment, Karin Lipfert (IPK, Department Cytogenetics and Genome Analysis, Gatersleben) for help with figure preparation, and Doreen Stengel (IPK, Department Cytogenetics and Genome Analysis, Gatersleben) for help with uploading genome survey sequences, and to Ingo Schubert (IPK, Department Cytogenetics and Genome Analysis, Gatersleben) for fruitful discussions. This work was supported by the Institut für Pflanzen- genetik und Kulturpflanzenforschung (Gatersleben), the Deutsche


