Analysis of transposable elements and organellar DNA in male and female genomes of a species with a huge Y chromosome reveals distinct Y centromeres

Aretuza Sousa1,*, Sidonie Bellot2, Jörg Fuchs3, Andreas Houben3 and Susanne S. Renner1
1Department of Biology, University of Munich (LMU), Munich 80638, Germany, 2Plant Biodiversity Research, Technical University of Munich (TUM), Freising, 85354 Germany, and 3Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Stadt Seeland 06466, Germany

Received 9 May 2016; revised 10 June 2016; accepted 23 June 2016; published online 14 October 2016.
*For correspondence (e-mail aretuzasousa@gmail.com).

SUMMARY
Few angiosperms have distinct Y chromosomes. Among those that do are Silene latifolia (Caryophyllaceae), Rumex acetosa (Polygonaceae) and Coccinia grandis (Cucurbitaceae), the latter having a male/female difference of 10% of the total genome (female individuals have a 0.85 pg genome, male individuals 0.94 pg), due to a Y chromosome that arose about 3 million years ago. We compared the sequence composition of male and female C. grandis plants and determined the chromosomal distribution of repetitive and organellar DNA with probes developed from 21 types of repetitive DNA, including 16 mobile elements. The size of the Y chromosome is largely due to the accumulation of certain repeats, such as members of the Ty1/copia and Ty3/gypsy superfamilies, an unclassified element and a satellite, but also plastome- and chondriome-derived sequences. An abundant tandem repeat with a unit size of 144 bp stains the centromeres of the X chromosome and the autosomes, but is absent from the Y centromere. Immunostaining with pericentromere-specific markers for anti-histone H3Ser10ph and H2AThr120ph revealed a Y-specific extension of these histone marks. That the Y centromere has a different make-up from all the remaining centromeres raises questions about its spindle attachment, and suggests that centromeric or pericentromeric chromatin might be involved in the suppression of recombination.

Keywords: centromere, Coccinia grandis, comparative genomics, FISH, histone modification, repetitive DNA, sex chromosomes.

INTRODUCTION
In flowering plants, heteromorphic sex chromosomes are known from some 19 species belonging to the families Cannabaceae, Caryophyllaceae, Cucurbitaceae and Polygonaceae (Ming et al., 2011), with high-quality images of metaphase plates available for 16 (see Table S1). This heteromorphism is triggered by the suppression of recombination between sex-determining regions of the X and Y chromosomes, making natural selection less efficient in those regions and allowing the chromosomes to diverge from each other (reviewed by Bachtrog et al., 2014). Transposable elements (TEs) are expected to accumulate in recombination-suppressed regions, but non-recombining regions can theoretically also contain low-copy types of DNA that became replicated after their integration into a sex chromosome. Such an accumulation of repetitive DNAs on Y chromosomes has so far been shown in Cannabis sativa (Cannabaceae; Sakamoto et al., 2000), Rumex acetosa (Polygonaceae; Shibata et al., 1999; Mariotti et al., 2009) and Silene latifolia (Caryophyllaceae; Hobza et al., 2006; Kubat et al., 2008; Kejnovsky et al., 2009). The proportions of the different kinds of repetitive elements as well as the exact elements that are accumulated on the Y chromosomes of flowering plants appear to vary greatly and do not seem to reflect the repetitive content of the autosomes. However, data are exceedingly sparse. In R. acetosa, which has an X chromosome and two Y chromosomes that are slightly smaller than the X, many retrotransposons that are frequent on the autosomes and the X chromosome are absent on both Ys, which instead have accumulated other TEs, as well as satellites and chloroplast DNA (Steflova et al., 2013, 2014). Also in S. latifolia, which has a Y chromosome that is much larger than the X and autosomes, the retrotransposons Ogre and Retand (Tat) are underrepresented on the Y, which is instead
enriched in other TEs, one tandem repeat (STAR-Y) and chloroplast DNA (Kejnovsky et al., 2006b; Cermak et al., 2008; Filatov et al., 2009; also S. dioica and S. marizii; Filatov et al., 2009). The only species so far studied with a combination of high-throughput sequencing and cytological observations are S. latifolia (Macas et al., 2011) and R. acetosa (Steflova et al., 2013, 2014).

Coccinia grandis (Cucurbitaceae), a species native to tropical Africa and India and a member of a small, entirely dioecious genus (Holstein and Renner, 2011; Holstein, 2015), is characterized by heteromorphic sex chromosomes, resulting in a male genome that is 10% larger than the female (Sousa et al., 2013). Other Coccinia species, e.g. Coccinia sessilifolia and Coccinia hirtella, which diverged from each other about 5 million years ago (Ma), based on a fossil-calibrated phylogeny of the Cucurbitaceae, do not have distinguishable sex chromosomes, and their genome size is unknown (Holstein and Renner, 2011; Holstein, 2015). The sex chromosomes of C. grandis date from about 3 Ma (Holstein and Renner, 2011) and are thus perhaps younger than those of S. latifolia, which date from between 3.5 and 24 Ma (reviewed in Sousa et al., 2013). Among plants with heteromorphic sex chromosomes, C. grandis is unusual in its heterochromatinized Y chromosome (Sousa et al., 2013).

Here we compare the global genome composition of male and female C. grandis plants, using genomic and molecular-cytogenetic approaches, to understand what causes the 0.1 pg difference in genome size between the sexes. This knowledge is the basis for inferring commonalities between plant sex chromosomes, which so far has been elusive because so few have been analyzed with the required high-throughput sequencing methods. Because of this scarcity of data, it has not been possible to assess whether the centromeres of sex chromosomes consistently differ from the remaining centromeres. That this might be the case is suggested by the discovery that in S. latifolia a particular repeat (called STAR-C) is present in the centromeres of the X chromosome and all autosomes, but is absent from the Y centromere (Cermak et al., 2008; Macas et al., 2011); that in Humulus lupulus, the Y chromosome lacks a subtelomeric repeat that is clustered near the centromere on the X chromosome (Divashuk et al., 2011); and that in R. acetosa, a chromovirus element (CL25) that has colonized most centromeres is less abundant on the centromeres of the three sex chromosomes (X, Y1 and Y2; Steflova et al., 2013).

We therefore wanted to know: (i) the C. grandis Y-chromosome repeat composition and compare it with the X chromosome and the autosomes; (ii) the pattern of organellar DNA accumulation in the C. grandis genome; and (iii) if the centromere composition of the Y chromosome in C. grandis differs from the centromeres of the X chromosome or autosomes.

### RESULTS

**Differential distribution of repeats in male and female genomes, and especially in the Y centromere**

Coccinia grandis has a karyotype of 2n = 24, with male individuals (2n = 22 + XY) having a large Y chromosome that is easily seen under a microscope. Female plants, with 2n = 22 + XX, have chromosomes of almost homogeneous size. The X chromosome can only be differentiated in meiosis I when it pairs with the Y chromosome in males.

Table 1 shows the comparative distribution of repeat clusters in the two genomes. The long terminal repeat (LTR) retrotransposon superfamilies Ty1/copia and Ty3/gypsy respectively make up 6.6% and 6.35% of the genome (when averaging male and female values), followed by DNA satellites (2.05%), transposons (0.7%) and long interspersed nuclear elements (LINEs) (0.15%). Plastid and mitochondrial sequences respectively comprise 5.7 and 0.28% of the genome, averaged over male and female cells. Although the male genome is 10% larger than the female genome, both sexes have similar proportions of repetitive elements (Tables 1 and 2, Fig. S1).

Of the in silico data, we selected 21 repetitive elements including mobile elements (belonging to Ty3/gypsy, Ty1/copia, LINEs and DNA transposons) for fluorescence in situ hybridization (FISH) analysis. Four of these elements had a male/female ratio greater than two (more than two-fold enrichment in males; Table 2). They comprise an unclassified repeat cluster (CL260) with four times more reads in the male than in the female genome, a satellite sequence (CL97) and two chromoviruses belonging to the Ty3 gypsy elements (CL153 and CL86). For those four elements, we observed an increased accumulation of FISH

<table>
<thead>
<tr>
<th>Repeat</th>
<th>Genome proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ty1/copia</td>
<td>7.61 5.59 6.6</td>
</tr>
<tr>
<td>Ty3/gypsy</td>
<td>7.5 5.2 6.35</td>
</tr>
<tr>
<td>Satellite</td>
<td>1.86 2.24 2.05</td>
</tr>
<tr>
<td>rDNA</td>
<td>1.73 1.67 1.7</td>
</tr>
<tr>
<td>DNA transposon</td>
<td>0.78 0.62 0.7</td>
</tr>
<tr>
<td>LINE</td>
<td>0.15 0.15 0.15</td>
</tr>
<tr>
<td>Unclassified</td>
<td>9.28 7.72 8.5</td>
</tr>
<tr>
<td>Total</td>
<td>28.91 23.19 26.05</td>
</tr>
</tbody>
</table>

signals along the Y chromosome while the autosomes and the X chromosomes of male and female cells revealed weaker signals at various positions (CL260 and CL86; Figs II, I and S2A, E) or preferentially at pericentromeric (CL97; Fig. 1j, n) or pericentromeric and subterminal positions (CL153; Fig. 1k, o). Among the sequences with a male/female read ratio of less than two, five had prominent signals on the Y chromosome. One of them, CL10, a LTR Ty1/copia/Angela element, was also weakly dispersed on the other chromosomes including the X (Fig. 1c, g), and in interphase nuclei it co-localized with the DAPI-positive heterochromatin (Fig. 2c), which permitted the identification of Y chromosome during interphase. The four others, CL9 and CL37 (two Ty3/gypsy/Athila elements), CL19 (a gypsy Ogre/Tat element) and CL44 (a LINE element), showed clustered signals in the subterminal, or pericentromeric and subterminal, region(s) of all chromosomes (Figs 1b, d, f, h, i, m and S2B, F).

All other investigated sequences were not preferentially enriched over the entire length of the Y chromosome, being instead interstitially located on the Y chromosome and scattered in the subterminal regions of many chromosomes, including the X (Figs S2 and S3). Interestingly, the most abundant satellite, CL1, was a 144-bp repeat that hybridized to the centromeres of all chromosomes except the Y (Figs 1a, e and S4). This repeat, which we named the C. grandis centromere repeat or CgCent, is slightly under-represented in the male compared with the female genome (Tables 1 and 2).

### Table 2 Repeat clusters selected for fluorescence in situ hybridization (FISH), with their numbering from the RepeatExplorer pipeline. M and F stand for male and female reads. Clusters are ranked according to their male specificity calculated as the sum of male and female sequence reads divided by total male reads. Datasets I to V reflect five RepeatExplorer pipeline runs (see Experimental Procedures). Check marks in the rightmost column indicate FISH signals accumulated on the Y chromosome

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Dataset</th>
<th>M</th>
<th>F</th>
<th>(M + F)/M</th>
<th>M/F</th>
<th>Annotation</th>
<th>FISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL260</td>
<td>IV</td>
<td>74</td>
<td>17</td>
<td>0.81</td>
<td>4.35</td>
<td>Unclassified</td>
<td></td>
</tr>
<tr>
<td>CL57</td>
<td>III</td>
<td>594</td>
<td>237</td>
<td>0.71</td>
<td>2.51</td>
<td>Satellite</td>
<td></td>
</tr>
<tr>
<td>CL153</td>
<td>I</td>
<td>302</td>
<td>139</td>
<td>0.68</td>
<td>2.17</td>
<td>Ty3/gypsy/chromovirus</td>
<td></td>
</tr>
<tr>
<td>CL86</td>
<td>I</td>
<td>1436</td>
<td>686</td>
<td>0.68</td>
<td>2.09</td>
<td>Ty3/gypsy/chromovirus</td>
<td></td>
</tr>
<tr>
<td>CL173</td>
<td>IV</td>
<td>121</td>
<td>83</td>
<td>0.69</td>
<td>1.45</td>
<td>Mitochondrial sequence</td>
<td></td>
</tr>
<tr>
<td>CL37</td>
<td>I</td>
<td>2651</td>
<td>2039</td>
<td>0.75</td>
<td>1.30</td>
<td>Ty3/gypsy/Athila</td>
<td></td>
</tr>
<tr>
<td>CL140</td>
<td>I</td>
<td>326</td>
<td>286</td>
<td>0.53</td>
<td>1.14</td>
<td>LTR/caulimovirus</td>
<td></td>
</tr>
<tr>
<td>CL121</td>
<td>III</td>
<td>278</td>
<td>262</td>
<td>0.51</td>
<td>1.06</td>
<td>Ty1/copia/Ael/retrofit</td>
<td></td>
</tr>
<tr>
<td>CL247</td>
<td>II</td>
<td>79</td>
<td>78</td>
<td>0.50</td>
<td>1.01</td>
<td>DNA Hat-Ac</td>
<td></td>
</tr>
<tr>
<td>CL10</td>
<td>II</td>
<td>5048</td>
<td>5038</td>
<td>0.50</td>
<td>1.00</td>
<td>Ty1/copia/Angela</td>
<td></td>
</tr>
<tr>
<td>CL9</td>
<td>II</td>
<td>4884</td>
<td>6035</td>
<td>0.45</td>
<td>0.81</td>
<td>Ty3/gypsy/Athila</td>
<td></td>
</tr>
<tr>
<td>CL19</td>
<td>II</td>
<td>3485</td>
<td>4515</td>
<td>0.44</td>
<td>0.77</td>
<td>Ty3/gypsy/Ogre/Tat</td>
<td></td>
</tr>
<tr>
<td>CL76</td>
<td>II</td>
<td>1083</td>
<td>1419</td>
<td>0.43</td>
<td>0.76</td>
<td>Ty1/copia/Tork</td>
<td></td>
</tr>
<tr>
<td>CL1</td>
<td>I</td>
<td>15026</td>
<td>20023</td>
<td>0.43</td>
<td>0.75</td>
<td>Satellite</td>
<td></td>
</tr>
<tr>
<td>CL89</td>
<td>II</td>
<td>1198</td>
<td>1657</td>
<td>0.42</td>
<td>0.72</td>
<td>DNA CMC-EnSpm</td>
<td></td>
</tr>
<tr>
<td>CL44</td>
<td>II</td>
<td>1798</td>
<td>2567</td>
<td>0.41</td>
<td>0.70</td>
<td>LINE</td>
<td></td>
</tr>
<tr>
<td>CL22</td>
<td>II</td>
<td>3008</td>
<td>4478</td>
<td>0.40</td>
<td>0.67</td>
<td>Ty1/copia/Bianca</td>
<td></td>
</tr>
<tr>
<td>CL92</td>
<td>II</td>
<td>834</td>
<td>1292</td>
<td>0.39</td>
<td>0.65</td>
<td>Ty1/copia/Ivana/Oryco</td>
<td></td>
</tr>
<tr>
<td>CL50</td>
<td>II</td>
<td>1479</td>
<td>2464</td>
<td>0.38</td>
<td>0.60</td>
<td>Ty1/copia/Ael</td>
<td></td>
</tr>
<tr>
<td>CL36</td>
<td>II</td>
<td>1890</td>
<td>3343</td>
<td>0.36</td>
<td>0.57</td>
<td>Ty1/copia/Maximus/SIRE</td>
<td></td>
</tr>
<tr>
<td>CL38</td>
<td>V</td>
<td>N/A</td>
<td>927</td>
<td>0.00</td>
<td></td>
<td>Ty1/copia/TAR</td>
<td></td>
</tr>
</tbody>
</table>

**LINE**, long interspersed nuclear element. SIRE and TAR are phylogenetic clades of Ty1/Copia plant LTR retrotransposons.

**What makes the Y chromosome different?**

Probes matching the large single copy (LSC) region of the C. grandis plastome (Fig. S5) yielded strong signals on the Y chromosome (Fig. 2a, b) and weaker signals on the pericentric and subterminal regions of the X and the other chromosomes. In female cells, the same probes yielded a strong hybridization signal on most chromosomes, sometimes covering an entire arm (Fig. S6A, B); the same pattern was seen in male cells after artificially (using Photoshop) increasing the brightness of the entire image. Short probes (about 4000 bp) matching the inverted repeat (IR) and the small single-copy (SSC) region were accumulated mostly on the Y chromosome (Fig. 2c–e), while in the other chromosomes and female cells, they localized to the pericentric and subterminal regions (Fig. S6C–E). Long probes (about 17 000–24 000 bp) matching the IR and SSC regions hybridized strongly on both arms of the Y chromosome close to the telomeric ends (Fig. 2f, g). On the other chromosomes of both sexes, these long probes yielded strong IR and weak SSC signals in the pericentric region (Figs 2f, g and S6F, G). The mitochondrial probe (cluster CL173) gave strong signals on the Y chromosome.
Figure 1. Abundant repeats (as detected by RepeatExplorer) detected using fluorescence in situ hybridization. Sequence annotation is given at the top of each column, and the asterisk stands for ‘Ty3/gypsy’, while the symbol § stands for ‘Ty1/copia.’ Roman numerals in square brackets represent different analyses performed in RepeatExplorer. The inset in (c) shows heterochromatin blocks in the nucleus stained intensely with the CL10 probe. Arrows indicate the Y chromosome. The sex of each karyotype is indicated in the upper right-hand corner. Bars correspond to 5 μm.
chromosome, with respectively weak or strong signals on the male and the female autosomes (Figs 2a and S6H).

The Y chromosome shows an atypical distribution of histone H2AThr120ph and H3Ser10ph marks

Immunostaining with antibodies against the cell cycle-regulated histone modifications H2AThr120ph and H3Ser10ph gave signals on all chromosomes in both sexes (Figs 3 and S7). Both histone marks are diagnostic for active centromeres (Houben et al., 1999; Demidov et al., 2014). In more distended chromosomes, H2AThr120ph signals were located between signals of H3Ser10ph, and both antibodies cross-reacted with the pericentromeres of all chromosomes including the Y, on which they also labeled the entire arms (Fig. 3a, b). Unexpectedly, the H2AThr120ph signals (red signals in Fig. S7, right-most column) were dispersed rather than restricted to the primary constriction (Fig. S7, white arrow). Sequential immunostaining–FISH experiments confirmed the centromeric position of CgCent (Fig. S8).

In meiosis I metaphase, the CgCent centromere-specific repeat revealed the structure of the bivalents, with the centromeres of the homologous chromosomes pointing to the opposite poles (Fig. 4a). In late meiosis II, before the formation of tetrade cells, the two nuclei containing the Y chromatids could be identified by the absence of CgCent signals (Fig. 4b), making it possible to differentiate male and female pollen nuclei.

DISCUSSION

The large size of the *C. grandis* Y chromosome is mainly due to the accumulation of five TEs, one satellite and an unclassified repeat

Genomic in silico analyses combined with in situ detection in plants with heteromorphic sex chromosomes have now been carried out in three species of flowering plants: *S. latifolia* (Cermak et al., 2008; Macas et al., 2011; Kralova et al., 2014; Steflova et al., 2014), *R. acetosa* (Steflova et al., 2013, 2014) and *C. grandis* (this study). Given that the difference between the male and female genomes of *C. grandis* is about 10% of the genome (Sousa et al., 2013), our chance of detecting Y-specific repeats was similar to that of a study of *S. latifolia*, where the Y chromosome also represents about 10% of the total genome (Macas et al., 2011).
The repetitive fraction of both genomes of *C. grandis* is mainly composed of Ty1/copia (6.6%) and Ty3/gypsy (6.3%) LTR elements, with the Ty1/copia/Angela element (CL10) and three Ty3/gypsy elements (CL9, CL19, CL37 and CL86) being found in much greater abundance on the Y than on the other chromosomes. In *R. acetosa* (2n = 14/15, XY1Y2/XX), Ty1/copia elements are slightly more abundant than Ty3/gypsy elements (Steflova et al., 2013), while in *S. latifolia* (2n = 24, XX/XY), Ty3/gypsy elements are about 2.7-fold more frequent than Ty1/copia elements (Macas et al., 2011), yet only Ty1/copia has accumulated on the Y chromosome (Cermak et al., 2008). The Y chromosome of *C. grandis* resembles the *R. acetosa* Y in being rich in AT, and accumulating mainly Copia elements (Ruiz Rejón et al., 1994; Shibata et al., 1999; Steflova et al., 2013; this study). In *S. latifolia*, as in *R. acetosa*, the Y chromosome is euchromatic (Grant et al., 1994; Matsunaga et al., 1999; Cermak et al., 2008).

Other elements, satellites and rDNA make up only 4.6% of the *C. grandis* genome. In comparison, the sex chromosomes of *S. latifolia* and *R. acetosa* are much richer in satellites, which might be linked to their greater age (see the Introduction). On the Y chromosome of *S. latifolia*, four tandem repeats have been identified (Bužek et al., 1997; Hobza et al., 2006; Cermak et al., 2008), while in *R. acetosa*, seven satellites have been described (Shibata et al., 1999, 2000; Mariotti et al., 2009; Steflova et al., 2013). In both species, most of these tandem repeats are Y-specific. In contrast, in *C. grandis*, we detected only one satellite (CL97), and it is not Y-specific (Fig. 6b).
Accumulation of organellar DNA in the *C. grandis* nuclear genome

Transfer of chloroplast and mitochondrial DNA from the cytoplasm to the nucleus and its integration into the nuclear genome has been observed in many plant species (Timmis and Scott, 1983; Timmis *et al.*, 2004). In *Arabidopsis thaliana*, a large insert of mitochondrial DNA occurs at the pericentromeric region of chromosome 2 and almost the entire plastome is duplicated in the nucleus (Stupar *et al.*, 2001; Noutsos *et al.*, 2005); in rice, the pericentromeric regions of all chromosomes have inserts of 15–131 kb of plastid-derived sequences (Matsuo *et al.*, 2005); in rye, plastid- and mitochondrial-derived sequences have accumulated on a supernumerary B chromosome (Martis *et al.*, 2012), fitting with their expected accumulation in regions with low recombination rates (Michalovova *et al.*, 2013). In *C. grandis*, our FISH studies revealed signals of hybridization to plastid-like regions across the entire genome, and the high abundance and diversity of plastid-like sequences with nuclear-like coverage in our *in silico* data suggest that those signals correspond to various plastid regions rather than to a single plastid region that would have colonized the whole nuclear genome. These results corroborate differential subtraction chain analyses performed in our lab that revealed nuclear DNA regions in *C. grandis* consisting of plastid genes; one matched the *psbA* gene [333-bp open reading frame (ORF) without any stop codon], one the *cemA* gene (201-bp ORF), both from the LSC; another region corresponded to the *ndhH* gene (573-bp ORF), which is in the SSC region, and other regions matched the *ndhB* gene (343-bp ORF) and the *rrn23* gene (271- to 378-bp ORF), both in the IR region. In our current FISH study, the subterminal regions of both arms of the Y chromosome contain parts of the IR and SSC regions. In *R. acetosa*, the Y1 and Y2 chromosomes have also accumulated plastid DNA, while in *S. latifolia*, pieces of the IR region are found along the Y chromosome and pieces of the SSC region in its centromere (Kejnovsky *et al.*, 2006a; Steflova *et al.*, 2014).

Beside plastid DNA, the Y chromosome of *C. grandis* accumulates mitochondrial-like sequences, especially at its interstitial region, and other chromosomes show either strong or weak signals in their pericentromeric regions. Mitochondrial-like probes have not been tested in other plant species with sex chromosomes.

The distinct chromatin composition of the Y (peri)centromeres

Our most surprising result is that the *C. grandis* Y centromere lacks hybridization signals for the CgCent repeat and in this respect is different from all other centromeres. Centromere satellite arrays are among the most difficult regions to identify *in silico* due to their high variability within and among species (Miga *et al.*, 2014). The CgCent monomer has 144 bp, but it can vary slightly in size (Fig. S4). The lack of signals of this monomer from the Y centromere could be caused by its low abundance or by a rearrangement of the monomer arrays, preventing our probe from binding. In the human Y chromosome, the monomers of a centromere satellite that has been analyzed consist of two groups with different array sizes and sequence compositions due to deletions and insertions of TEs (Miga *et al.*, 2014). Similar events might prevent our probe from binding to the Y centromere. Strikingly, in
S. latifolia too, a particular repeat (called STAR-C) is present in the centromeres of the X chromosome and all autosomes, but absent from the Y centromere (Macas et al., 2011), and in R. acetosa, a chromovirus element (CL25) that has colonized most centromeres is less abundant on the centromeres of the three sex chromosomes (X, Y1, and Y2; Steflova et al., 2013).

The distribution of two pericentromeric histone markers, H3Ser10ph and H2AThr120ph, further sets the Y centromere of C. grandis apart from the species’ other centromeres. Both markers only stain active centromeres, as shown for H3Ser10 in plant species with two centromeres and for H2AThr120ph by its co-localization with CENH3 (Houben et al., 1999). Three types of centromeres are known: monocentric, holocentric and meta-polycentric. Meta-polycentric centromeres have an extended primary constriction containing multiple CENH3 domains that span up to a third of the length of the chromosomes (Neumann et al., 2016). In such centromeres, signals of H2AThr120ph occur at chromatin surrounding the CENH3 domains without overlapping them. In C. grandis, the H2AThr120ph signal resembles that in species with such centromeres, except that in C. grandis this histone labels the Y chromosome along its entire length, raising questions about the attachment of the spindle fibers in the Y.

CONCLUSIONS

The large size of the Y chromosome in C. grandis is due to a massive accumulation of repeats, such as members of the Ty1/copia and Ty3/gypsy superfamilies, the unclassified element CL260 and the satellite CL 97, but also plastid and mitochondrial-like sequences. Selection on suppression of recombination to avoid separation of sex-determining genes was presumably the start of sex chromosome dimorphism in C. grandis, and the presence of a single Y chromosome in artificial polyploids of C. grandis results in male plants (Roy and Roy, 1971), implying that at least one key sex gene is located on its Y chromosome. In Cucumis, which is closely related to Coccinia (Schaefer and Renner, 2011), the sex-determining genes G (WIP), A (ACS11) and M (ACS7) are located on three chromosomes (Boualem et al., 2015). Sex-determining mechanisms can be evolutionarily modified very quickly, and it is possible that two of these genes were moved to autosomes after differentiation of the Y chromosome. Perhaps the extension of histone modification (centromeric or pericentromeric chromatin) in the Y chromosome is involved in recombination suppression.

EXPERIMENTAL PROCEDURES

Plant material

Coccinia grandis (Cucurbitaceae) seeds for this study were collected in spring 2011 on the campus of Kakatiya University in Vidyaranyapura, Warangal, India. In Munich, plants were cultivated in the greenhouses of the Botanical Garden, and vouchers have been deposited in the Munich herbarium (official acronym M) as Sousa and Silber 1 and Sousa and Silber 2. Female and male plants were identified by chromosome preparations.

Next-generation sequencing of genomic DNA and repeat quantification

Paired-end Illumina Nextera sequencing of male and female total genomic DNA was performed at the IPK, Gatersleben. A total of 10 259 242 and 16 531 172 paired-end sequence reads (average read length 225) were obtained from male and female samples, yielding a genome coverage of 2.7× for the male and 5× for the female genome. We used the RepeatExplorer pipeline (Novák et al., 2013), implemented within the Galaxy environment (http://repeatexplorer.umbr.cas.cz/), to perform the following steps. After checking sequence quality with the FastQC:Read QC tool, adapters were removed and the nucleotides from positions 1–25 and 225–250 were trimmed to obtain high-quality read fragments. Reads were filtered by quality, using a cut-off of 10, and then by length (200 bp). Interlaced paired reads were sampled at random to cover 8.5–31.6% of the genomes of C. grandis and then concatenated to a single dataset, followed by clustering sequence reads with a minimum overlap of 55% and a similarity of 90%. We ran this step five times using different combinations of the male and female datasets to obtain as many distinct repetitive clusters as possible (Table S2). For three of these analyses, mitochondrial and plastid-like sequences were excluded from the datasets before the concatenation and clustering steps. We developed a python script for listing in a single table the number and proportion of reads and their annotation (File S1). Protein domains were identified using the tool Find RT Domains in RepeatExplorer (Novák et al., 2013).

Detection of plastid and mitochondrial sequences

The trimmed reads from the male genome were de novo assembled using the CLC Genomics Workbench 8 (http://www.clcbio.com) with the default values, resulting in 214 603 contigs. Those contigs were BLAST-searched (BLASTn command of BLAST+) version 29, Camacho et al., 2009) against the whole plastome of Cucumis sativus (NC_007144) allowing the identification of four contigs with very high bit scores. The four contigs were then mapped against the plastome of Cucumis melo (NC_015983) using Geneious R8 (Biomatters, http://www.geneious.com/), and concatenated in the order in which they mapped, which was supported by overlapping extremities of the contigs. The junction between the contigs (also corresponding to the junctions between the SSC and IR regions) were double-checked by remapping all the reads against the plastome of C. grandis (obtained for this study from our data), using CLC and a threshold of minimum 90% similarity over at least 70% of the read length. Those steps were repeated for the reads of the female plant, and the two final plastomes were annotated by comparison to C. melo using Geneious R8 and manual checking, and deposited in GenBank (accessions KX147311 and KX147312). Graphical maps were generated with the OrganellarGenomeDRAW software (Lohse et al., 2007). We used the Coccinia plastomes to design primers allowing the PCR amplification of five plastid regions already used as probes in a previous FISH study (Steflova et al., 2014) of S. latifolia (probes Slcp7, Slcp10, Slcp14, and Slcp20) and R. acetosa (probe Racp10), as well as of two larger regions spanning the entire IR and SSC regions. To detect mitochondrial DNA in situ we designed primers on a 4339-bp long contig of the RepeatExplorer (CL173) with more than 86% similarity to mitochondrial sequences in RepeatMasker.
and 97.6% to the Citrullus lanatus mitochondrial genome (NC_014043.1). Coccinia-specific primers were designed using Primer3 (Untergasser et al., 2012) and their sequences and expected product sizes are presented in Table S2. Probes were prepared as described in the next section (Development of FISH probes).

Development of FISH probes

To obtain FISH probes, template DNA was amplified using custom primers (Table S2) by PCR with 1.25 units of Taq DNA polymerase (New England Biolabs, https://www.neb.com/) or Taq Polymerase Q5 Hot Start High-Fidelity (New England Biolabs). The PCR products were checked by gel electrophoresis, and sequenced with an ABI 3130-4 capillary sequencer. Sequences were assembled and edited with Geneious R8 (Biomatters, http://www.geneious.com/). Manual checking and BLAST analysis were performed against the original contigs from the RepeatExplorer in Geneious R8 to verify the identity of the sequenced products.

Coccinia grandis sequences have been deposited in GenBank (see Table S2). Purified PCR products of 200–700 bp were labeled with biotin-16-dUTP or digoxigenin-11-dUTP (Roche, http://www.roche.com/) by PCR. Fragments longer than 800 bp, were labeled with digoxigenin-11-dUTP (Roche) or biotin-14-dATP (Invitrogen, http://www.invitrogen.com/) by nick translation.

Preparation of chromosome spreads and FISH

Root tips were collected from potted plants, pre-treated in 70% ethanol, cycloheximide (Roth, https://www.carroths.com/) in 2 mM 8-hydroxyquinoline for 5 h at 18°C to achieve higher numbers of metaphase cells (Tlaskal, 1980), fixed in freshly prepared 3:1 (v/v) ethanol/glacial acetic acid at room temperature (22°C) for 2 h, transferred to 70% ethanol and kept at room temperature overnight, and afterwards stored at −20°C until use. For chromosome preparations, we used the dropping method of Aliyeva-Schnorr et al. (2015) with the following modification. Fixed root tips were digested with 1% cellulase (w/v; Onozuka RS, Serva, http://www.serva.de/) and 1% cytohelicase (Sigma) in 1× PBS for 25 min, digested with 1% pectolyase (w/v; Sigma, http://www.sigmaaldrich.com/), 0.4% pectolyase (w/v; Sigma, http://www.sigmaaldrich.com/) and 0.4% cytohelicase (w/v; Sigma) in citric buffer, pH 4.8 for 2 h 50 min at 37°C in a humid chamber, and the freshly prepared fixative consisted of 1:1 (v/v) methanol/glacial acetic acid.

The FISH was performed as in Sousa et al. (2013) with the inclusion of an ethanol series of 70–90–100%, 2 min each, after the final washes. Images were captured with a Leica DMR microscope with a DP70 camera and the KAPPA-CCD camera and the KAPPA software and optimized for best contrast and brightness using Adobe Photoshop CS3 version 10.0.

Immunofluorescence

For immunostaining experiments, root tips were pre-treated as above, fixed in cold 4% paraformaldehyde in phosphate-buffered saline (PBS) for 25 min, digested with 1% pectolyase (Sigma), 0.7% cellulase Onozuka R-10 (Duchefa, https://www.duchefa-biochemie.com/) and 1% cytohelicase (Sigma) in 1× PBS for 8 min at 37°C in a humid chamber, washed three times for 5 min in 1× PBS on ice and squashed in PBS (following Manzanero et al., 2000). After removing the cover slide by freezing, the preparations were deposited in a Coplin jar with 1× PBS before incubation for 1 h at 37°C in blocking solution (5% BSA in 1× PBS + Triton X100). We used rabbit anti-phosphorylated histone H2AThr120 (diluted 1:150) and monoclonal mouse anti-H3Ser10ph (diluted 1: 500, ab14955, abcam, http://www.abcam.com/) as primary antibodies, and Cy3-conjugated anti-rabbit IgG (Dianova, http://www.dianova.com/de/) and fluorescein isothiocyanate-conjugated anti-mouse IgG (Dianova) as secondary antibodies (Houben et al., 1999; Demidov et al., 2014). After final washes in PBS, the preparations were counterstained with DAPI mounted in Vectashield (Vector, https://vectorlabs.com/).

For the sequential staining using FISH, slides were postfixed with freshly prepared 3:1 (v/v) ethanol/glacial acetic acid at room temperature for 30 min, and incubated in 100% ethanol at 4°C overnight. The FISH experiment started from the denaturation step. Images were taken as described above.

ACKNOWLEDGEMENTS

We thank: Mathieu Piednoël (post-doctoral fellow with SSR in Munich), Magdalena Vaio (University of Montevideo) and Jiri Macas (Biology Centre, Academy of Sciences, Budweis, Czech Republic) for instructions in the use of the RepeatExplorer pipeline; Martina Silber (Munich) for help in the lab; Veit Schubert (Gatersleben) for help with super-resolution microscopy; Axel Himmelbach (Gatersleben) for help with next-generation sequencing; D. Filatov and B. Janousek for comments on the manuscript; and the German Science Foundation for funding (DFG RE-603/19-1).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Cluster representation of all analyses from RepeatExplorer.

Figure S2. Mobile element clusters from RepeatExplorer detected using fluorescence in situ hybridization.

Figure S3. Mobile element clusters from RepeatExplorer detected using fluorescence in situ hybridization.

Figure S4. The sequence of the Coccinia grandis centromere-specific CgCent probe (CL1Ctg7).

Figure S5. Map of the Coccinia grandis male chloroplast genome.

Figure S6. Fluorescence in situ hybridization detection of plastid and mitochondrial sequences in mitotic metaphase chromosomes of Coccinia grandis female.

Figure S7. Immunostaining experiment on mitotic metaphase chromosomes of Coccinia grandis, using antibodies against H3Ser10ph and H2AThr120ph on mitotic metaphase chromosomes of C. grandis observed with a super-resolution microscope.

Figure S8. Sequential immunostaining–fluorescence in situ hybridization experiment on mitotic metaphase chromosomes of Coccinia grandis.

Table S1. Land plants with heteromorphic sex chromosomes (Ming et al., 2011) focusing on the type of evidence available for each.

Table S2. Sequences of primers used to generate fluorescence in situ hybridization probes.

File S1. Python script for linear visualization of information contained in the cluster folder, output of the clustering from the RepeatExplorer. Handling instructions are inside the file.

REFERENCES


Manzanero, S., Arana, P., Puertas, M.J. and Houben, A.
Divashuk, M.G., Alexandrov, O.S., Yu Kroupin, P. and Karlov, G.I.
(2007) Organellar Genome-DRAW.
Martis, M.M., Klemme, S., Banaei-Moghaddam, A.M. and Higashiyama, T., Nakao, S., Matsunaga, S., Kawano, S., Michimoto, T.