

## Short Communication

# The comparative analysis of the DNA repeat composition among *Cannabis sativa* L., *Humulus lupulus* L. and *Humulus japonicus* Siebold & Zucc. with heteromorphic sex chromosomes

Yulia V. Bocharkina, Oleg S. Alexandrov \*, Olga V. Razumova and Gennady I. Karlov

All-Russia Research Institute of Agricultural Biotechnology, Timiryazevskaya 42, 127550 Moscow, Russia; ju\_football@mail.ru (Y.V.B.); razumovao@gmail.com (O.V.R.); karlov@gmail.com (G.I.K.)

\* Correspondence: alexandrov@iab.ac.ru; Tel.: +7-499-976-6544

**Abstract:** Heteromorphic sex chromosomes are rarely found in plants. They were observed only in 47 species from phylogenetically distant families, suggesting that the evolution of sex chromosomes was independent in these species. It was shown that DNA repeat sequences are one of the major factors driving sex chromosomes evolution, and an accumulation or elimination of the repetitive DNA elements are closely linked with the formation of differences in the sex chromosomes. The goal of this study was to characterize the transposon composition in male and female plants of *Cannabis sativa* L., *Humulus lupulus* L. and *Humulus japonicus* Siebold & Zucc. For the first time, the male and female genomes of *H. japonicus* as well as male genomes of *H. lupulus* and *C. sativa* have been sequenced (there were no open data about them). The analysis of genome-wide sequencing data with using Repeatexplorer2 and author's scripts was carried out. It was shown that accumulation of *Ty3-gypsy* may be associated with speciation in Cannabaceae family which is the opposite of the theory of speciation throw whole-genome duplication. Moreover, the sex-specific DNA repeat clusters in *C. sativa* and *H. japonicus* were found. The analysis also revealed that the concentration of Tekay, Retand and Ikeros repeats in the Y chromosome of *C. sativa* is lower than in the X chromosome and the Angela concentration is higher in the Y chromosome.

**Keywords:** of *Cannabis sativa* L.; *Humulus lupulus* L.; *Humulus japonicus* Siebold & Zucc.; sex chromosomes; DNA repeats; transposons; genome sequencing

## 1. Introduction

Most animals are dioecious species, containing heteromorphic sex chromosomes. Meantime, plant sex dioecy is an extremely rare event. Nearly 6% of land plants have sex and about 90% of plant species are hermaphrodites [1]. Moreover, sex specificity was evolved many times independently, but mechanisms of sex evolution and determination were studied only for few species. The divergence between males and females has been achieved throw the formation of the Sex-Determining-Region (SDR) [2,3]. SDR is located on the sex chromosomes and usually is surrounded by many repetitive sequences. SDR may be a large part of the sex chromosome or be a small part of it. In case it is a small region, chromosomes are homomorphic and contain a small region of absence of recombination. Large SDRs, surrounded by repetitive sequences are the cause of the chromosomal heteromorphism and big region of recombination arrest. Visually homomorphic sex chromosomes do not have strong morphological differences and it was suggested, that these chromosomes are at the early stages of sexual development [4]. Sex chromosomes evolving is achieved by several steps: occurrence of SDR, recombination arrest, repeats accumulation, genes degeneration, deletion of the inactive part of the chromosome. The repetitive sequences are key players of the sex chromosomal evolution [5–7].

Plant genomes are characterized by large amounts of repetitive DNA, referred to as repeatome. This part is known as one of the major sources of changes in genome size [8–

11]. Repeats can be spread through all chromosomes and be more dispersed (the usual situation for transposable elements), but, also, they can be accumulated in one place (like most of the satellites repeats) [12–16]. Repeats are often located in functionally meaningful parts of the genome [17–21]. There are several known main types of repetitive DNA: mobile elements (ME), ribosomal DNA, satellites and organelle repeats. The repeats can differ in copy number and in the length of the motifs. The biggest part of the genome usually consists of Transposable Elements (TEs) [22]. TEs are presented in genomes by two classes. Class I includes retrotransposons (REs), it is remarkable for using reverse transcriptase to be copied from one part of the genome to another part [23,24]. This class is divided into two subclasses: 1. LTR (Long Terminal Repeats), 2. autonomous non-LTR and non-autonomous non-LTR. Class II, known as transposons, moving by the cut-and-paste algorithm [23]. But subclasses of transposons differ in the mechanism of cutting and insertion: exactly cut-and-paste (Terminal Inverted Repeats or TIRs), rolling circle (Helitron), self-synthesizing [8,24,25].

The LTR elements are known as major group in plants. Their fragments range in size from several hundred bp up to 25 thousand bp. LTRs usually consist of 4–6 bp TSD, GAG with its ORFs, structural protein for virus like particles and for POL. POL consists of AP or PROT (aspartic proteinase), RH (RNase H) and INT (DDE integrase). This type of TEs is divided into two superfamilies: *Gypsy* and *Copia*, which have different order of RT and INT domains, GAG-PROT-INT-RT-RH for *Copia* and GAG-PROT-RT-RH-INT for *Gypsy* [26–29].

LTR *Copia* superfamily consists of lineages and families such are: Ale, Alesia, Angela, Bianca, Bryco, Lyco, Gymco, Ikeros (also known as Tork sto-4), Ivana (or Oryco), Osser, SIRE, Tar (also known as Tork), and Tork, while Athila, Clamyvir, Galadriel, Selgy, Tcn1, Reina, Tekay (or Del), CRM (also named Centromeric Retrotransposon), Phygy, and TAT are members of *Gypsy* superfamily [30–32]. The isolation of families was founded through deep analysis of the presence of the domains, size, order, PBS and TSD organization in LTR fragments as well as the amount of ORFs [33,34].

Cannabaceae family contains three species of undoubted agricultural significance and all these three are dioecious, they have a different genome size and sex chromosomal organization. *Cannabis sativa* L. ( $2n=20$ ), also known as hemp, has a haploid genome size equal to 818 Mb for female, 843 Mb for male [35]. The Y chromosome is bigger than the X chromosome [35–37]. *Humulus lupulus* L. ( $2n=20$ ), also known as hop, has a female haploid size equal to 2960 Mb, 2730 for a male [25]. The X chromosome is bigger than the Y chromosome [25]. *Humulus japonicus* Siebold & Zucc (Japanese hop, ♀ $2n=16$ ; ♂ $2n=17$ ) has a haploid genome size for female 1568 Mb, for male it is 1722 Mb [35]. The sex organization system here contains 3 different chromosomes: X which is bigger than Y1, Y1 is bigger than Y2 [38]. Differences of species-specific subtelomeric repeat distribution in the sex chromosomes were revealed in a range of fluorescent *in situ* hybridization (FISH) experiments [36,38–40]. Thus, Cannabaceae family is a good model for investigations of plant sex chromosome evolution.

In this study, the genomes of *C. sativa*, *H. lupulus* and *H. japonicus* were analyzed, and the DNA repeat compositions were compared between species and between male and female samples. This research may become a starting point for further analysis of the sex-evolution model based on repeats accumulation in this family and other families of plants.

## 2. Materials and Methods

### 2.1. Plant material and DNA isolation

In the work, the following plants were used: *C. sativa* var. 'Zenitza', female; *C. sativa* var. 'Zenitza', male; *H. lupulus*, female (common hop, natural sample from Moscow region 55°83'75.32", 37°55'46.50"); *H. lupulus*, male (common hop, natural sample from Moscow region 55°83'77.38", 37°55'51.70"); *H. japonicus* var. 'Samurai', female; *H. japonicus* var. 'Samurai', male. All plants were grown in a greenhouse. Small young leaves from all samples were collected and used for DNA isolation, according STAB-protocol (see Figure S1).

## 2.2. Sequencing and processing of data, search of the DNA repeats and their analysis

Sequencing was carried out with Illumina standard protocol with 1 run for each plant (150 bp fragment length). The coverages were the following: *C. sativa*, female – 1.5; *C. sativa*, male – 1.46; *H. lupulus*, female – 3.98; *H. lupulus*, male – 3.2; *H. japonicus*, female – 1.93; *H. japonicus*, male – 2.19.

For the preprocessing of raw sequence data, the FASTA-files were interlaced and sampled. Amount of the reads used for analysis was 1000000000 per sample. For the generation of the sampled and interlaced input file, the python3 custom script written by I. Kirov was used [41].

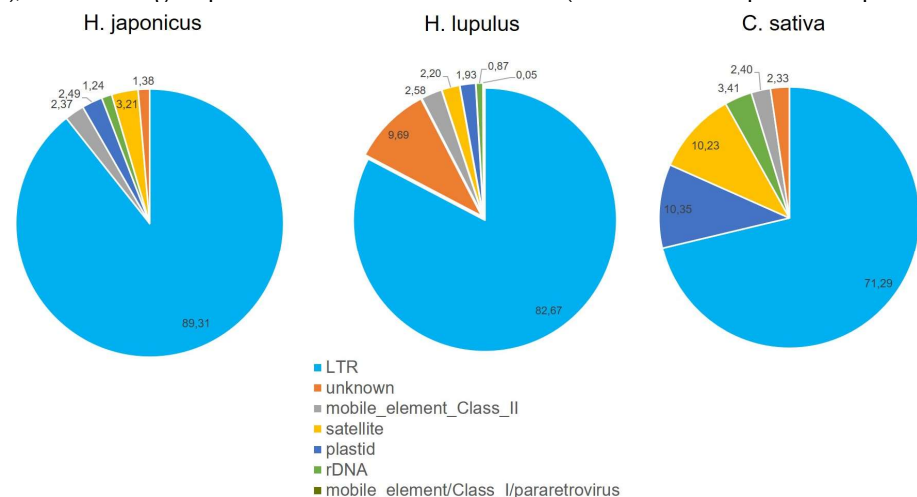
The preprocessed data were used to construct the graph-based clusters by RepeatExplorer2 [42,43] with custom database and strict parameters: 90% of similarity among not less than 54 % of length of the reads. The cluster information was collected based on these virtual graphs. The abundance of the clusters was calculated based on the proportion of reads.

## 3. Results

### 3.1. The comparison of the DNA repeat combinations among *C. sativa*, *H. lupulus* and *H. japonicus*.

At the beginning, the proportion of DNA repeats in the genomes of the studied species was calculated as the ratio of reads containing repeats to the total number of reads. The three studied species differ in this indicator. In *C. sativa*, 80% of reads include repetitive DNA and 60% of reads are in top clusters. In *H. japonicus* and *H. japonicus*, proportion of read with DNA repeats was equal to 95% (69% of reads are in top clusters) and 88% (62% of reads are in the top clusters), respectively.

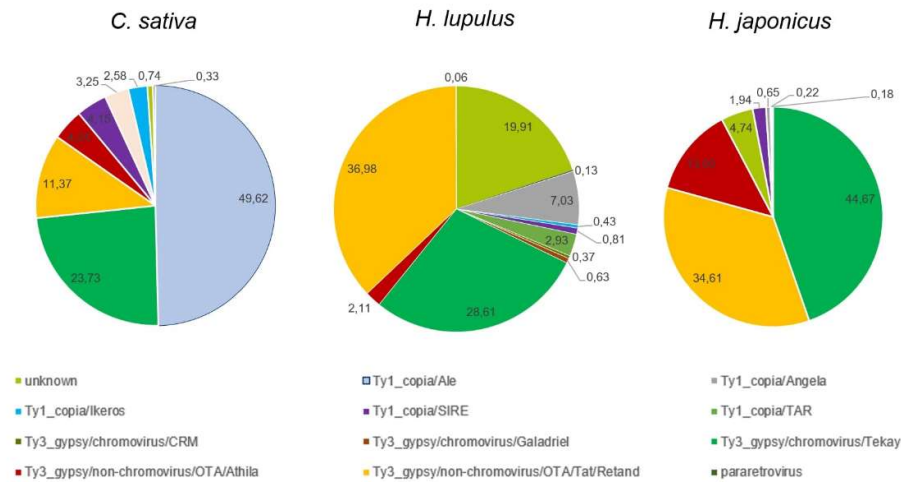
Next, the proportions of the different DNA repeat groups were analyzed for each of the studied species (Figure 1). The analysis revealed that LTRs have the largest proportion in all cases (71,29%–89,31%). In *H. japonicus*, proportions of other DNA repeat groups were significantly low (in the range of 1,24%–3,21%). In *C. sativa*, three groups have the same proportions (mobile\_elements\_Class\_I, mobile\_elements\_Class\_II, and unknown repeats), but two groups showed near 10% result (satellites and plastid repeats).



**Figure 1.** The ratios of different groups of DNA repeats in the *C. sativa*, *H. lupulus* and *H. japonicus* genomes.

Since the majority of the reads were mapped into the LTR group, the next step was to perform a comparative analysis of all LTR families in each species separately. The Figure 2 demonstrate that the species differ greatly in the type of retrotransposons. The majority (almost half) of the reads for *C. sativa* belong to different families of *Ty1-copia* type. In *H. lupulus*, the biggest parts of the LTR compartment are *Ty3-gypsy*/OTA/Tat/Retand,

*Ty3\_gypsy/Tekay*. *H. japonicus*, also has the majority of reads, belonging to *Ty3-gypsy* families. These are *Ty3-gypsy/OTA/Tat/Retand*, *Ty3-gypsy/Tekay* and *Ty3\_gypsy/Athila*. However, proportions of these repeats differed between the two species of *Humulus*, and the differences were significant.



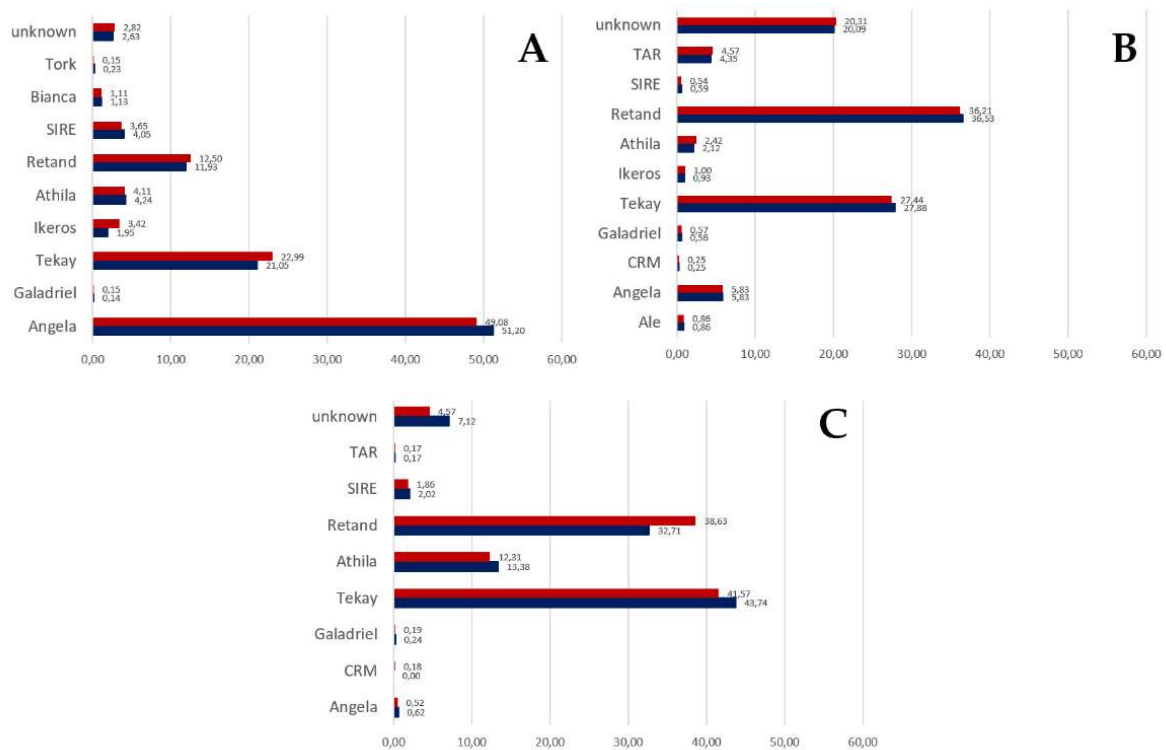
**Figure 2.** The ratios of the different LTR families in the *C. sativa*, *H. lupulus* and *H. japonicus* genomes.

3.2. The comparison of the DNA repeats between male and female plants of *C. sativa*, *H. lupulus* and *H. japonicus*.

In the next step of analysis, the proportions of DNA repeats in the genomes of the studied species were separately calculated for male and female samples. The results are presented in Figures S2. The significant differences in repeat combinations and proportions between male and female samples were not observed in all cases.

The detailed analysis of LTRs was also conducted for male and female samples (Figure S3). The comparison of proportions did not reveal significant differences in *Ty1\_copia*, *Ty3\_gypsy* proportions between samples of different sexes.

To find the significant differences between species as well as male and female plants the analysis of LTR families was carried out. The number of LTR families differs between species. Altogether, twelve families were observed in the Cannabaceae family, among them: (1) 6 families were observed in all three species (*Angela*, *Galadriel*, *Tekay*, *Athilla*, *Retand*, *SIRE*); (2) *CRM* and *TAR* were presented in both species of *Humulus* genus; (3) *Ikeros* was a common family for *C. sativa* and *H. lupulus*; (4) *Bianca* and *Tork* were presented in *C. sativa* only; (5) *Ale* family was specifically presented in *H. lupulus* only. The significant sex-specific differences were observed for *C. sativa*. *Tekay*, *Retand* and *Ikeros* repeats showed larger fractions in females and *Angela* in males (Figure 3A). In the case with *H. japonicus*, the results were obtained from two sequenced plants only (high-quality comparison need more NGS data to analyze the sex-specificity of this species' LTR families). However, it can be preliminarily concluded that *Retand* has larger proportion in female as well as *Athilla*, *Tekay* and *SIRE* in male plant (Figure 3C). Unlike these two species, there were not observed the sex-specific differences among LTR families in *H. lupulus* (Figure 3B).



**Figure 3.** The ratios of the different LTR families in the male and female genomes of (A) *C. sativa*, (B) *H. lupulus* and (C) *H. japonicus*.

#### 4. Discussion

The first significant result of this study was the receipt of updated data on the genome structures for three Cannabaceae species – *C. sativa*, *H. lupulus* and *H. japonicus*. It was previously reported, the *C. sativa* genome contains 64%–65% of repetitive sequences, among them the biggest part was LTR-repeats (about 50%), rDNA takes place in ~3% of repeatome and a big part (43%) of repeatome consists of low-complexity sequences [44]. Our results differ from the previous ones: we found that the genome contains 80% of repetitive sequences. The LTR repetitive elements make up 75% of repeatome, and the rDNA amount is 3%–4%. According to Natsume et al (2015), the *H. lupulus* genome contains ~60% of repetitive sequences, among them there were identified about 93% of LTRs [45]. Padgitt-Cobb et al. (2021) estimated the size of the repeatome as “above 70%”. Among them, there were 76.5% of LTRs [46]. Our results strongly suggest that 88% of the *H. lupulus* genome consists of repeats that differ from the previous reports. But in the same time, we found that the LTR amount make up 81% of the repeatome. This value is approximately the average of the values obtained by Natsume’s and Padgitt-Cobb’s groups. Since the *H. japonicus* genome is firstly sequenced in this study, the data of the genome component ratios were obtained for the first time. The analysis revealed that the *H. japonicus* repeatome takes up more space in the genome (95%) than the *C. sativa* and *H. lupulus* repeatomes in their genomes. The LTR amount was also the highest among these three species (89,31%).

It was proven that the speciation of *Humulus* and *Cannabis* occurred around 27 million years ago [36]. Wherein, the *C. sativa* genome is three times smaller than *H. lupulus* one, and *H. japonicus* genome is two times smaller than the *H. lupulus* one [35,47,48]. It was previously suggested that the species speciation occurred with whole-genome duplication [46]. However, the detailed analysis of LTRs suggests that it was achieved throw the accumulation of *Ty3-gypsy* repeats. *C. sativa* has ~38% repeats of *Ty3-gypsy* type, *H.*



*lupulus* has ~68 % and *H. japonicus* has ~90-92%. This type of accumulation of one particular type of repeats contradicts the hypothesis of whole-genome duplication. Therefore, the proliferation of *Ty3-gypsy* type might cause the speciation as it was proven for *Capsicum* spp [49].

In the analysis of the LTR families among three studied species, five groups were observed. The first group included the families which are presented in all three species (Angela, Galadriel, Tekay, Athila, Retand, SIRE). A similar set of the common LTR families is usually found in closely related species. For example, Borreda et al. (2019) reported about the set of the Retrofit, Oryco, SIRE, Tork, CRM, Reina, Del, Galadriel, Athila, and Tat LTR families, which are common for *Citrus* spp [50]. Such members of this set as SIRE, Galadriel, and Atilla were also determined as common for Cannabaceae spp. that may be linked with their higher activity in comparison with the others. The second group of the Cannabaceae LTR families included CRM and TAR, which were presented in both species of *Humulus* genus. Apparently, their activity started after the speciation of *Humulus* and *Cannabis*. The similar conclusions were made by Wawrzynski et al. (2008), when they compared the activity of transposons after the *Glicine max* (L.) Merr. and *G. tomentella* Hayata speciation [51]. The third identified group of Cannabaceae LTRs consists of one family (Ikeros), which is found in *C. sativa* and *H. lupulus*. An absence of the Ikeros activity in *H. japonicus* genome may be linked with the independent start of the Ikeros activity in *C. sativa* and *H. lupulus* genomes after the *H. lupulus* and *H. japonicus* divergence (3.74–6.78 MYA) [36,52]. This result once again confirms the correctness of the idea of the independent fate of TE during genome differentiation, which was, for example, shown in the analysis of the evolution of the genomes of *Nicotiana* or *Capsicum* spp. [53–55]. Finally, this theory can explain finding of the fourth and fifth groups of the Cannabaceae LTRs, which were *C. sativa*- (Bianca and Tork) and *H. lupulus*-specific (Ale), respectively.

In the analysis of the LTR combinations in male and female samples, the significant sex-specific differences were found in *C. sativa*. It has been revealed that Tekay, Retand and Ikeros families have larger fractions in females and Angela in males. The similar results were presented in *Spinacia oleracea* L. (with homomorphic sex chromosomes) and *S. tetrandra* Stev. (with heteromorphic sex chromosomes) [56]. Li et al. (2021) showed that the proportion of Ale and Retand was higher in the female genome than in the male genome, while the proportion of Angela, Bianca, CRM, Tekay, LINE, EnSpm\_CACTA, and rDNA was higher in male genome than in the female genome of *S. oleracea*. At the same time, Li's group indicated that the proportion of Angela, Athila, Ogre, CRM, Tekay, and EnSpm\_CACTA was higher in the female genome than in the male genome, and the proportion of rDNA was higher in the male genome than in the female genome of *S. tetrandra*. Thus, *C. sativa* has a higher proportion of Retand in females, like females of *S. oleracea*, and Tekay, like females of *S. tetrandra* and males of *S. oleracea*. And, the male genome of *C. sativa* demonstrates a higher proportion of Angela, like males of *S. oleracea* and females of *S. tetrandra*. The Ikeros repeats were not described in comparison of the male and female genomes of *Spinacia* spp. and their different proportion in males and females may be a feature of *C. sativa* genome.

In general, all the differences noted above in the proportions of repeats between the male and female genomes of *C. sativa*, *S. oleracea*, and *S. tetrandra* can be associated with different concentrations of these repeats in the Y chromosome. Indeed, if the concentration of a repeat is lower in the Y chromosome than in the X chromosome, then the proportion of this repeat in the female genome will be higher than in the male, since the female genome contains two X chromosomes. Conversely, if the concentration of a repeat is higher in the Y chromosome than in the X chromosome, then the proportion of this repeat in the male genome with XY sex chromosomes will be higher.

As a sum, this study is a new step in the investigation of the genome divergence and sex chromosome formation among dioecious species of Cannabaceae family. The obtained results can be a basis for further analysis of the sex-evolution model based on repeats accumulation in this family and other families of plants.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Figure S1: The DNA isolation protocol; Figure S2: The ratios of the DNA repeat types in the male and female genomes of *C. sativa*, *H. lupulus* and *H. japonicus*; Figure S3: The ratios of the LTR superfamilies in the male and female genomes of *C. sativa*, *H. japonicus* and *H. lupulus*.

**Author Contributions:** O.V.R. and G.I.K. conceived and designed the investigation; O.V.R. collected the plant material and isolated DNA; Y.V.B. processed sequencing data, searched for repeats, analyzed the results, wrote the text of the paper; O.S.A. analyzed the results, formulated the discussion and wrote the text of the paper.

**Funding:** This work was financially supported by a grant from the Russian Foundation for Basic research, agreement № 20-316-70018\19, dated 19.11.2019.

**Acknowledgments:** The authors are grateful to Andrey Pochtovyy for his technical consultations and Karina Bone-Prolesova for her help with the organization of the experiments as well as Tatyana Alexandrova for critically reading the English version of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Renner, S.S. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *Am. J. Bot.* **2014**, *101*(10), 1588–1596, doi: 10.3732/ajb.1400196.
- Renner, S.S.; Müller, N.A. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. *Nat. Plants* **2021**, *7*(4), 392–402, doi: 10.1038/s41477-021-00884-3.
- Sanderson, B.J.; Feng, G.; Hu, N.; Carlson, C.H.; Smart, L.B.; Keefover-Ring, K.; Yin, T.; Ma, T.; Liu, J.; DiFazio, S.P.; Olson, M.S. Sex determination through X-Y heterogamety in *Salix nigra*. *Heredity* **2021**, *126*(4), 630–639, doi: 10.1038/s41437-020-00397-3.
- D. Charlesworth, D. Plant contributions to our understanding of sex chromosome evolution. *New Phytol.* **2015**, *208*(1), 52–65, doi: 10.1111/nph.13497.
- Li, S.F.; Zhang, G.J.; Yuan, J.H.; Deng, C.L.; Gao, W.J. Repetitive sequences and epigenetic modification: inseparable partners play important roles in the evolution of plant sex chromosomes. *Planta* **2016**, *243*(5), 1083–1095, doi: 10.1007/s00425-016-2485-7.
- Zhou, Q.; Ellison, C.E.; Kaiser, V.B.; Alekseyenko, A.A.; Gorchakov, A.A.; Bachtrog, D. The epigenome of evolving *Drosophila* neo-sex chromosomes: dosage compensation and heterochromatin formation. *PLoS Biol.* **2013**, *11*(11), e1001711, doi: 10.1371/journal.pbio.1001711.
- Howell, E.C.; Armstrong, S.J.; Filatov, D.A. Evolution of neo-sex chromosomes in *Silene dioica*. *Genetics* **2009**, *182*(4), 1109–1115, doi: 10.1534/genetics.109.103580.
- Shao, F.; Han, M.; Peng, Z. Evolution and diversity of transposable elements in fish genomes. *Sci. Rep.* **2019**, *9*(1), 15399, doi: 10.1038/s41598-019-51888-1.
- Almojil, D.; Bourgeois, Y.; Falis, M.; Hariyani, I.; Wilcox, J.; Boissinot, S. The structural, functional and evolutionary impact of transposable elements in eukaryotes. *Genes* **2021**, *12*(6), 918, doi: 10.3390/genes12060918.
- Vicent, C.M.; Casacuberta, J.M. Impact of transposable elements on polyploid plant genomes. *Ann. Bot.* **2017**, *120*(2), 195–207, doi: 10.1093/aob/mcx078.
- Kidwell, M.G. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **2002**, *115*(1), 49–63, doi: 10.1023/A:1016072014259.
- Bouzinba-Segard, H.; Guais, A.; Francastel, C. Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. *Proc. Natl. Acad. Sci. USA* **2006**, *103*(23), 8709–8714, doi: 10.1073/pnas.0508006103.
- Garrido-Ramos, M.A. Satellite DNA in plants: more than just rubbish. *Cytogenet. Genome Res.* **2015**, *146*(2), 153–170, doi: 10.1159/000437008.
- Hobza, R.; Lengerova, M.; Svoboda, J.; Kubekova, H.; Kejnovsky, E.; Vyskot, B. An accumulation of tandem DNA repeats on the Y chromosome in *Silene latifolia* during early stages of sex chromosome evolution. *Chromosoma* **2006**, *115*(5), 376–382, doi: 10.1007/s00412-006-0065-5.
- Mariotti, B.; Manzano, S.; Kejnovský, E.; Vyskot, B.; Jamilena, M. Accumulation of Y-specific satellite DNAs during the evolution of *Rumex acetosa* sex chromosomes. *Mol. Genet. Genomics* **2009**, *281*(3), 249–259, doi: 10.1007/s00438-008-0405-7.
- Todorovska, E. Retrotransposons and their role in plant—genome evolution, *Biotechnol. Biotechnol. Equip.* **2007**, *21*(3), 294–305, doi: 10.1080/13102818.2007.10817464.
- Gent, J.I.; Wang, N.; Dawe, R.K. Stable centromere positioning in diverse sequence contexts of complex and satellite centromeres of maize and wild relatives. *Genome Biol.* **2017**, *18*(1), 121, doi: 10.1186/s13059-017-1249-4.
- Roberti, A.; Bensi, M.; Mazzagatti, A.; Piras, F.M.; Nergadze, S.G.; Giulotto, E.; Raimondi, E. Satellite DNA at the centromere is dispensable for segregation fidelity. *Genes* **2019**, *10*(6), 469, doi: 10.3390/genes10060469.
- Hummel, B.; Hansen, E.C.; Yoveva, A.; Aprile-Garcia, F.; Hussong, R.; Sawarkar, R. The evolutionary capacitor HSP90 buffers the regulatory effects of mammalian endogenous retroviruses. *Nat. Struct. Mol. Biol.* **2017**, *24*(3), 234–242, doi: 10.1038/nsmb.3368.

20. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **2016**, *351*(6277), 1083–1087, doi: 10.1126/science.aad5497.
21. Makarevitch, I.; Waters, A.J.; West, P.T.; Stitzer, M.; Hirsch, C.N.; Ross-Ibarra, J.; Springer, N.M. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* **2015**, *11*(1), e1004915, doi: 10.1371/journal.pgen.1004915.
22. Biémont, C.; Vieira, C. Genetics: junk DNA as an evolutionary force. *Nature* **2006**, *443*(7111), 521–524, doi: 10.1038/443521a.
23. Levin, H.L.; Moran, J.V. Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* **2011**, *12*(9), 615–627, doi: 10.1038/nrg3030.
24. Sotero-Caio, C.G.; Platt, R.N.; Suh, A.; Ray, D.A. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* **2017**, *9*(1), 161–177, doi: 10.1093/gbe/evw264.
25. Piégu, B.; Bire, S.; Arensburg, P.; Bigot, Y. A survey of transposable element classification systems – A call for a fundamental update to meet the challenge of their diversity and complexity. *Mol. Phylogenet. Evol.* **2015**, *86*, 90–109, doi: 10.1016/j.ympev.2015.03.009.
26. Piednoël, M.; Carrete-Vega, G.; Renner, S.S. Characterization of the LTR retrotransposon repertoire of a plant clade of six diploid and one tetraploid species. *Plant J.* **2013**, *75*(4), 699–709, doi: 10.1111/tpj.12233.
27. Ustyantsev, K.; Blinov, A.; Smyshlyaev, G. Convergence of retrotransposons in oomycetes and plants. *Mob. DNA* **2017**, *8*, 4, doi: 10.1186/s13100-017-0087-y.
28. Usai, G.; Mascagni, F.; Natali, L.; Giordani, T.; Cavallini, A. Comparative genome-wide analysis of repetitive DNA in the genus *Populus* L. *Tree Genet. Genomes* **2017**, *13*, 96., doi: 10.1007/s11295-017-1181-5.
29. Paz, R.C.; Kozaczek, M.E.; Rosli, H.G.; Andino, N.P.; Sanchez-Puerta, M.V. Diversity, distribution and dynamics of full-length *Copia* and *Gypsy* LTR retroelements in *Solanum lycopersicum*. *Genetica* **2017**, *145*(4-5), 417–430, doi: 10.1007/s10709-017-9977-7.
30. Llorens, C.; Futami, R.; Covelli, L.; Domínguez-Escribá, L.; Viu, J.M.; Tamarit, D.; Aguilar-Rodríguez, J.; Vicente-Ripolles, M.; Fuster, G.; Bernet, G.P.; Maumus, F.; Muñoz-Pomer, A.; Sempere, J.M.; Latorre, A.; Moya, A. The *Gypsy* Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* **2011**, *39*(Database issue), D70–4, doi: 10.1093/nar/gkq1061.
31. Lloréns, C.; Futami, R.; Bezemer, D.; Moya, A. The *Gypsy* Database (GyDB) of mobile genetic elements. *Nucleic Acids Res.* **2008**, *36*(Database issue), D38–46, doi: 10.1093/nar/gkm697.
32. Llorens, C.; Muñoz-Pomer, A.; Bernad, L.; Botella, H.; Moya, A. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biol. Direct.* **2009**, *4*, 41, doi: 10.1186/1745-6150-4-41.
33. Orozco-Arias, S.; Isaza, G.; Guyot, R. Retrotransposons in plant genomes: structure, identification, and classification through bioinformatics and machine learning. *Int. J. Mol. Sci.* **2019**, *20*(15), 3837, doi: 10.3390/ijms20153837.
34. de Castro Nunes, R.; Orozco-Arias, S.; Crouzillat, D.; Mueller, L.A.; Strickler, S.R.; Descombes, P.; Fournier, C.; Moine, D.; de Kochko, A.; Yuyama, P.M.; Vanzela, A.L.L.; Guyot, R. Structure and distribution of centromeric retrotransposons at diploid and allotetraploid coffee centromeric and pericentromeric regions. *Front Plant Sci.* **2018**, *9*, 175, doi: 10.3389/fpls.2018.00175.
35. Sakamoto, K.; Akiyama, Y.; Fukui, K.; Kamada, H.; Satoh, S. Characterization; genome sizes and morphology of sex chromosomes in hemp (*Cannabis sativa* L.). *Cytologia*, **1998**, *63*(4), 459–464, doi: 10.1508/cytologia.63.459.
36. Divashuk, M.G.; Alexandrov, O.S.; Razumova, O.V.; Kirov, I.V.; Karlov, G.I. Molecular cytogenetic characterization of the dioecious *Cannabis sativa* with an XY chromosome sex determination system. *PLoS One* **2014**, *9*(1), e85118, doi: 10.1371/journal.pone.0085118.
37. Ming, R.; Bendahmane, A.; Renner, S.S. Sex chromosomes in land plants. *Annu. Rev. Plant Biol.* **2011**, *62*, 485–514, doi: 10.1146/annurev-arplant-042110-103914.
38. Alexandrov, O.S.; Divashuk, M.G.; Yakovin, N.A.; Karlov, G.I. Sex chromosome differentiation in *Humulus japonicus* Siebold & Zuccarini, 1846 (Cannabaceae) revealed by fluorescence *in situ* hybridization of subtelomeric repeat. *Comp. Cytogenet.* **2012**, *6*(3), 239–247, doi: 10.3897/CompCytogen.v6i3.3261.
39. Razumova, O.V.; Alexandrov, O.S.; Divashuk, M.G.; Sukhorada, T.I.; Karlov, G.I. Molecular cytogenetic analysis of monoecious hemp (*Cannabis sativa* L.) cultivars reveals its karyotype variations and sex chromosomes constitution. *Protoplasma* **2016**, *253*, 895–901, doi: 10.1007/s00709-015-0851-0.
40. Cermak, T.; Kubat, Z.; Hobza, R.; Kobližkova, A.; Widmer, A.; Macas, J.; Vyskot, B.; Kejnovsky, E. Survey of repetitive sequences in *Silene latifolia* with respect to their distribution on sex chromosomes. *Chromosome Res.* **2008**, *16*(7), 961–976, doi: 10.1007/s10577-008-1254-2.
41. Kirov, I.V., prepareReadsRev2.py. Available online: URL: [https://github.com/Kirovez/RepeatExplorer\\_scripts/blob/master/prepareReadsRev2.py](https://github.com/Kirovez/RepeatExplorer_scripts/blob/master/prepareReadsRev2.py). (accessed on 27 November 2021).
42. Novák, P.; Neumann, P.; Pech, J.; Steinhaisl, J.; Macas, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **2013**, *29*(6), 792–793, doi: 10.1093/bioinformatics/btt054.
43. Novák, P.; Neumann, P.; Macas, J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* **2020**, *15*(11), 3745–3776. doi: 10.1038/s41596-020-0400-y.
44. Pisupati, R.; Vergara, D.; Kane, N.C. Diversity and evolution of the repetitive genomic content in *Cannabis sativa*. *BMC Genomics* **2018**, *19*(1), 156, doi: 10.1186/s12864-018-4494-3.
45. Natsume, S.; Takagi, H.; Shiraishi, A.; Murata, J.; Toyonaga, H.; Patzak, J.; Takagi, M.; Yaegashi, H.; Uemura, A.; Mitsuoka, C.; Yoshida, K.; Krofta, K.; Satake, H.; Terauchi, R.; Ono, E. The draft genome of hop (*Humulus lupulus*), an essence for brewing. *Plant Cell Physiol.* **2015**, *56*(3), 428–541, doi: 10.1093/pcp/pcu169.



46. Padgitt-Cobb, L.K.; Kingan, S.B.; Wells, J.; Elser, J.; Kronmiller, B.; Moore, D.; Concepcion, G.; Peluso, P.; Rank, D.; Jaiswal, P.; Henning, J.; Hendrix, D.A. A draft phased assembly of the diploid Cascade hop (*Humulus lupulus*) genome. *Plant Genome* **2021**, *14*(1), e20072, doi: 10.1002/tpg2.20072.
47. Zonneveld, B.J.; Leitch, I.J.; Bennett, M.D. First nuclear DNA amounts in more than 300 angiosperms. *Ann. Bot.* **2005**, *96*(2), 229–244, doi: 10.1093/aob/mci170.
48. Grabowska-Joachim, A.; Sliwinski, E.; Pigula, M.; Skomra, U.; Joachim, A.J. Genome size in *Humulus lupulus* L. and *H. japonicus* Siebold & Zucc. (Cannabaceae) *Acta Soc. Bot. Pol. Pol.* **2006**, *75*(3), 207–214, doi: 10.5586/asbp.2006.024.
49. Kim, S.; Park, M.; Yeom, S.I.; Kim, Y.M.; Lee, J.M.; Lee, H.A.; Seo, E.; Choi, J.; Cheong, K.; Kim, K.T.; Jung, K.; Lee, G.W.; Oh, S.K.; Bae, C.; Kim, S.B.; Lee, H.Y.; Kim, S.Y.; Kim, M.S.; Kang, B.C.; Jo, Y.D.; Yang, H.B.; Jeong, H.J.; Kang, W.H.; Kwon, J.K.; Shin, C.; Lim, J.Y.; Park, J.H.; Huh, J.H.; Kim, J.S.; Kim, B.D.; Cohen, O.; Paran, I.; Suh, M.C.; Lee, S.B.; Kim, Y.K.; Shin, Y.; Noh, S.J.; Park, J.; Seo, Y.S.; Kwon, S.Y.; Kim, H.A.; Park, J.M.; Kim, H.J.; Choi, S.B.; Bosland, P.W.; Reeves, G.; Jo, S.H.; Lee, B.W.; Cho, H.T.; Choi, H.S.; Lee, M.S.; Yu, Y.; Do Choi, Y.; Park, B.S.; van Deynze, A.; Ashrafi, H.; Hill, T.; Kim, W.T.; Pai, H.S.; Ahn, H.K.; Yeom, I.; Giovannoni, J.J.; Rose, J.K.; Sørensen, I.; Lee, S.J.; Kim, R.W.; Choi, I.Y.; Choi, B.S.; Lim, J.S.; Lee, Y.H.; Choi, D. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* **2014**, *46*(3), 270–278, doi: 10.1038/ng.2877.
50. Borredá C, Pérez-Román E, Ibanez V, Terol J, Talon M. Reprogramming of retrotransposon activity during speciation of the genus *Citrus*. *Genome Biol. Evol.* **2019**, *11*(12), 3478–3495, doi: 10.1093/gbe/evz246.
51. Wawrzynski, A.; Ashfield, T.; Chen, N.W.; Mammadov, J.; Nguyen, A.; Podicheti, R.; Cannon, S.B.; Thareau, V.; Ameline-Torregrosa, C.; Cannon, E.; Chacko, B.; Couloux, A.; Dalwani, A.; Denny, R.; Deshpande, S.; Egan, A.N.; Glover, N.; Howell, S.; Ilut, D.; Lai, H.; Del Campo, S.M.; Metcalf, M.; O'Bleness, M.; Pfeil, B.E.; Ratnaparkhe, M.B.; Samain, S.; Sanders, I.; Séguens, B.; Sévignac, M.; Sherman-Broyles, S.; Tucker, D.M.; Yi, J.; Doyle, J.J.; Geffroy, V.; Roe, B.A.; Maroof, M.A.; Young, N.D.; Innes, R.W. Replication of nonautonomous retroelements in soybean appears to be both recent and common. *Plant Physiol.* **2008**, *148*(4), 1760–1771, doi: 10.1104/pp.108.127910.
52. Murakami, A.; Darby, P.; Javornik, B.; Pais, M.S.; Seigner, E.; Lutz, A.; Svoboda, P. Molecular phylogeny of wild hops, *Humulus lupulus* L. *Heredity* **2006**, *97*(1), 66–74, doi: 10.1038/sj.hdy.6800839.
53. Vernhettes, S.; Grandbastien, M.A.; Casacuberta, J.M. The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. *Mol. Biol. Evol.* **1998**, *15*(7), 827–836, doi: 10.1093/oxfordjournals.molbev.a025988.
54. Casacuberta, E.; González, J. The impact of transposable elements in environmental adaptation. *Mol. Ecol.* **2013**, *22*(6), 1503–1517, doi: 10.1111/mec.12170.
55. de Assis, R.; Baba, V.Y.; Cintra, L.A.; Gonçalves, L.S.A.; Rodrigues, R.; Vanzela, A.L.L. Genome relationships and LTR-retrotransposon diversity in three cultivated *Capsicum* L. (Solanaceae) species. *BMC Genomics* **2020**, *21*(1), 237, doi: 10.1186/s12864-020-6618-9.
56. Li, N.; Li, X.; Zhou, J.; Yu, L.; Li, S.; Zhang, Y.; Qin, R.; Gao, W.; Deng, C. Genome-wide analysis of transposable elements and satellite DNAs in *Spinacia* species to shed light on their roles in sex chromosome evolution. *Front. Plant Sci.* **2021**, *11*, 575462, doi: 10.3389/fpls.2020.575462.