

Contrasting Patterns of Transposable Element and Satellite Distribution on Sex Chromosomes (XY₁Y₂) in the Dioecious Plant *Rumex acetosa*

Pavlina Steflava^{1,2}, Viktor Tokan¹, Ivan Vogel^{1,2}, Matej Lexa², Jiri Macas³, Petr Novak³, Roman Hobza^{1,4}, Boris Vyskot¹, and Eduard Kejnovsky^{1,2,*}

¹Department of Plant Developmental Genetics, Institute of Biophysics ASCR, Brno, Czech Republic

²Laboratory of Genome Dynamics, CEITEC—Central European Institute of Technology, Masaryk University, Brno, Czech Republic

³Biology Centre ASCR, Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic

⁴Laboratory of Molecular Cytogenetics and Cytometry, Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Olomouc, Czech Republic

*Corresponding author: E-mail: kejnovsk@ibp.cz.

Accepted: March 25, 2013

Data deposition: This project has been deposited at GenBank under accession nos. SRX118072, SRX118073, KC310863, KC310864, KC310865, KC310866, KC310867, KC310868, KC310869, KC310870, KC310871, and KC310872.

Abstract

Rumex acetosa is a dioecious plant with the XY₁Y₂ sex chromosome system. Both Y chromosomes are heterochromatic and are thought to be degenerated. We performed low-pass 454 sequencing and similarity-based clustering of male and female genomic 454 reads to identify and characterize major groups of *R. acetosa* repetitive DNA. We found that Copia and Gypsy retrotransposons dominated, followed by DNA transposons and nonlong terminal repeat retrotransposons. CRM and Tat/Ogre retrotransposons dominated the Gypsy superfamily, whereas Maximus/Sireviruses were most abundant among Copia retrotransposons. Only one Gypsy subfamily had accumulated on Y₁ and Y₂ chromosomes, whereas many retrotransposons were ubiquitous on autosomes and the X chromosome, but absent on Y₁ and Y₂ chromosomes, and others were depleted from the X chromosome. One group of CRM Gypsy was specifically localized to centromeres. We also found that majority of previously described satellites (RAYS I, RAYS II, RAYS III, and RAE180) are accumulated on the Y chromosomes where we identified Y chromosome-specific variant of RAE180. We discovered two novel satellites—RA160 satellite dominating on the X chromosome and RA690 localized mostly on the Y₁ chromosome. The expression pattern obtained from Illumina RNA sequencing showed that the expression of transposable elements is similar in leaves of both sexes and that satellites are also expressed. Contrasting patterns of transposable elements (TEs) and satellite localization on sex chromosomes in *R. acetosa*, where not only accumulation but also depletion of repetitive DNA was observed, suggest that a plethora of evolutionary processes can shape sex chromosomes.

Key words: sex chromosomes, sorrel (*Rumex acetosa*), transposable elements, satellites.

Introduction

Sex chromosomes are the genomic regions undergoing specific evolutionary processes (Charlesworth 1991). There is also extraordinary variability in the patterns of sex chromosomes: not only the XY system dominating in mammals and the ZW system in lepidoptera and birds but also many variants with multiple X or Y chromosomes found in both animals and plants with the extreme example of five Xs and five Ys in the male platypus (McMillan et al. 2007; Ming et al. 2011). The unifying feature of the Y and W chromosome is partial or

complete loss of recombination with their partner X and Z chromosome, respectively, which leads to genetic degeneration of the Y or W chromosome and accumulation of repetitive DNA combined with expansion (Charlesworth et al. 1994; Gvozdev et al. 2005; Kejnovsky, Hobza, et al. 2009). In plants, sex chromosomes are found in several dioecious species and often represent an early evolutionary stage (Ming et al. 2011). An incipient stage of sex chromosomes is represented by homomorphic sex chromosomes present in some plants (e.g., *Carica papaya* or *Bryonia dioica*). Other

plants have evolutionarily older heteromorphic sex chromosomes with either large Y chromosome (*Silene latifolia*, *Coccinia grandis*, *Rumex acetosa*, and *Cannabis sativa*) or small Y chromosome (*Cycas revoluta*, *Humulus lupulus*, and *Marchantia polymorpha*, for review see Ming et al. 2011).

Sorrel (*R. acetosa*) is a dioecious plant with the XY₁Y₂ system. Dioecy in *Rumex* genus (XY system) arose about 16 Ma and the *acetosa* clade with multiple XY₁Y₂ system originated 12–13 Ma (Navajas-Pérez et al. 2005a). The X chromosome is the largest in male metaphase, but both Y chromosomes together are bigger than the X chromosome. Five satellites have been found in *R. acetosa*—RAYS I, RAY II, RAYS II (specific for Y₁ and Y₂, Shibata et al. 1999; Navajas-Pérez et al. 2005b), RAE180 (Y₁, Y₂ and one autosome, Shibata et al. 2000), and RAE730 (autosomes, Shibata et al. 2000). RAYS I and RAE180 are the main components of the Y heterochromatin (Shibata et al. 1999, 2000). RAYS I is also common in other species with multiple XY₁Y₂ systems (*R. papillaris*, *R. intermedius*, *R. thyrsoideus*, and *R. tuberosus*) but absent in species with an XY system such as *R. acetosella* and *R. suffruticosus* (Navajas-Pérez et al. 2005b; Cunado et al. 2007). RAE180 is expanded on the Y₁ chromosome in *R. acetosa*. It is also amplified on one autosome in *R. suffruticosus* and dispersed in low copy number in *R. acetosella* (Shibata et al. 2000; Cunado et al. 2007; Navajas-Pérez et al. 2009). RAYS I, RAYS II, RAYS III, and RAE730 satellites arose by different ancestral duplications and reshufflings from the same 120-bp unit (Navajas-Pérez et al. 2005b; Mariotti et al. 2009). Intraspecific variability of Y-associated satellites such as RAYS I and RAE180 is much higher than that in the autosomal RAE730 satellite, which indicates a particular mode of evolution of satellites in a nonrecombining genomic context (Navajas-Pérez et al. 2005b, 2005c). To date, no TEs have been described in *Rumex* species. Only four clones originating from degenerate polymerase chain reaction (PCR) on microdissected sex chromosomes exhibited homologies with Gypsy (DOP-47 and 61), Copia (DOP-60), and non-LTR retrotransposons (DOP-8, Mariotti et al. 2005). *Rumex acetosa* has two 45S rDNA loci on two autosomal pairs (Lengerova and Vyskot 2001).

Repetitive DNA forms a significant proportion of eukaryotic genomes. This is particularly evident in plants, which have faster genome dynamics than animals (Kejnovsky, Leitch, et al. 2009). However, the rules governing genome size and repeat composition are not fully understood. Even closely related species often significantly differ in composition of their transposable elements or satellites (Neumann et al. 2006). The chromosomal localization of repetitive DNA was previously thought to be only a result of selection, but recent findings show that other factors such as targeting of TEs into specific chromosomal niches are also important (for review see Heslop-Harrison and Schwarzacher 2011; Kejnovsky et al. 2012). The Y or W sex chromosomes often accumulate various repetitive DNA as has been proven for humans (Skaletsky

et al. 2003), drosophila (Steinemann M and Steinemann S 1992), fish (Cioffi et al. 2011), and reptiles (Pokorna et al. 2011). In plants, tandem repeats (Hobza et al. 2006), microsatellites (Kubat et al. 2008), and transposable elements (Cermak et al. 2008) are accumulated on the Y chromosome of *S. latifolia*, whereas tandem repeats are gathered on both Y chromosomes in *R. acetosa* (Shibata et al. 1999; Mariotti et al. 2009), and transposable elements are accumulated on the Y chromosome in *C. sativa* (Sakamoto et al. 2000). However, repetitive DNA can have also other patterns than simple accumulation on the Y chromosome. For example, Ogre retrotransposon is ubiquitous on all autosomes and the X chromosome but is absent on the Y chromosome in *S. latifolia* (Cermak et al. 2008). Microsatellites are accumulated on the X chromosome rather than Y chromosome in fish *Hoplias malabaricus* (Cioffi et al. 2011), and some microsatellites are absent on the W chromosome in the lizard *Eremias velox* despite their presence on other chromosomes (Pokorna et al. 2011).

In this study, we analyzed the structure, genomic proportion, expression, and chromosomal localization of the main classes of TEs and satellites in the dioecious plant *R. acetosa*. We found that Maximus/Sireviruses (among Copia elements) and Chromoviruses (among Gypsy elements) predominate and their chromosomal localization exhibits various contrasting patterns, not only accumulation on the Y chromosomes.

Materials and Methods

454 Sequencing

One sequencing run of the 454 GS FLX platform (454 Life Sciences, Roche) was performed for each male and female genomic DNA isolated from healthy young leaves, resulting in 280,954 and 295,993 quality-filtered reads, respectively, with average read length 332 nucleotides for the male and 338 nucleotides for the female sample (accession numbers SRX118072 and SRX118073). Male and female read sets were combined for the purpose of complex analysis, providing a total of 193.4 Mb of sequencing data. Given the genome size of *R. acetosa* 7.0 pg in female and 7.5 pg in male (2C) (Blocka-Wandas et al. 2007), this represents 5.7% of the genome. The sequencing reads were clustered on the basis of similarity (as described by Novak et al. 2010; Macas et al. 2011), and clusters containing at least 57 reads (representing around 0.01% of the genome) were used for further analysis.

Illumina Sequencing

Pair-end sequencing was performed for two male and two female genomic DNA. These were isolated from leaves representing parents and their single male and female progenitors (deposited under SRA062840). The leaves from the same individuals were then used for RNA-Seq experiment (deposited under SRA058606) resulting in four pair-end libraries of

transcriptomic data. Both genomic and transcriptomic classes of reads were then analyzed using FastQC (available at: <http://www.bioinformatics.babraham.ac.uk>) quality control tool. The reads were trimmed and filtered on the basis of quality using FASTX-toolkit (available at: http://hannonlab.cshl.edu/fastx_toolkit/), and the redundant reads were removed from all data sets. Both genomic and transcriptomic libraries were then mapped to the identified clusters of genomic 454 data using BLAT (Kent 2002). The BLAT analysis was run with default parameters, except for the stepSize parameter, which was reduced to nine. This was to ensure greater sensitivity of the mapping analysis (the genome is sampled with higher sampling frequency). To eliminate redundancy in the obtained alignments, the following steps were taken: Only alignments with an alignment e value of less than 10⁻²⁰ and 10⁻¹⁵ for genomic and transcriptomic data were considered. The BLAT output was then sorted according to e value, percent identity, and alignment score. Only alignments that fitted these criteria best were chosen for future analysis to ensure that every Illumina read was mapped only once to one of the genomic reads (locations). The numbers of mapped Illumina genomic versus 454 genomic reads (table 1) or transcriptome reads versus 454 genomic reads (fig. 6) were counted for every cluster and subsequently for every identified TE family of *R. acetosa*. The weighted average (considering library sizes) of

the relative expression and genomic proportions of repetitive families was counted. In-house computational pipeline including custom-made Bash and Python scripts was used to sort and filter the alignments.

Phylogeny and Classification

The reconstructed DNA sequences were analyzed for the presence of a reverse transcriptase (RT) domain by sequence similarity. Nucleotide sequences of RT cores were then used to place the clusters into a phylogenetic tree of LTR-retrotransposon RT domains. The identification of RT cores was based on a collection of consensus amino acid sequences of known RT domains available at Gypsy Database (Llorens et al. 2011) and TREP database (<http://wheat.pw.usda.gov/ITMI/Repeats/>). This collection was used to create a BLAST+ (Camacho et al. 2009) database and searched using the blastx command with DNA sequences of the LTR elements in question. Regions having e value < 10⁻³ were cut out making sure they were unique and fell into the 500–1,000-bp range observed for the best matches of well-known elements. The extracted RT cores were subsequently analyzed using the Geneious Pro Alignment tool (Drummond et al. 2011) to generate a multiple nucleotide sequence alignment. Once aligned, the Neighbor-Joining distance model of the Geneious Pro Tree Builder was used to build a phylogenetic tree.

Structural Annotation of LTR Elements

The reconstructed nucleotide sequences were first analyzed for the presence of structural features typical for specific classes of repetitive sequences, namely LTRs, *gag* and *pol* genes and their individual protein domains (GAG, AP, RT, RH, and INT), other ORFs, PBS, and/or PPT. The presence of typical protein domains was detected by sequence similarity, in the same way as the detection of RT cores in the previous paragraph, except for using the appropriate consensus sequences. The recognition of *gag* and *pol* genes relied on the combined evidence of predicted ORFs using the FrameD++ software package (Schiex et al. 2003) used because of its tolerance to reading frame interruptions and the presence of protein domains. However, no exact delimitation of the ORFs/genes was attempted because of the nature of the analyzed sequences (e.g., averaged from multiple loci, presence of nonfunctional but autonomous LTR elements). The PBS and PPT sequences were detected using the LTR finder software (Xu and Wang 2007).

Preparation of Probes for Fluorescence In Situ Hybridization

Specific primers were designed, usually for RT or the transposase domain of individual TEs. In the first step, template DNA was amplified using PCR with a mix containing 1× complete PCR buffer, 0.1 mM dNTPs, 0.1 μM primers, 0.5 U Taq polymerase (Top Bio), and 10–15 ng of template DNA. Reaction

Table 1

Repeat Composition in *Rumex acetosa* Genome Estimated from Illumina Sequencing Data

Repeat Type	Classification		Genome Proportion	
	Super Family	Family	Male (%)	Female (%)
Retroelements	Gypsy	Chromovirus—CRM	5.73	5.38
		Chromovirus—Tekay/Del	0.34	0.36
		Athila	0.48	0.41
		Tat/Ogre	5.36	5.69
	Copia	Maximus/Sire	34.92	35.58
		Bianca	0.19	0.20
		TAR	0.10	0.10
LINE		0.01	0.01	
DNA transposons	Mutator		1.30	1.33
	CACTA		0.26	0.25
Total transposable elements			48.69	49.31
Satellites	RAYS I		0.79	0.0002
	RAYS II		0.061	0
	RAYS III		0.44	0.00013
	RA160		0.61	0.76
	RA690		0.29	0.27
	RAE 180		2.72	1.04
	RAE 730		0.24	0.47
Total satellites			5.15	2.54
rDNA			0.18	0.21

conditions were as follows: 94 °C/4 min 34× (94 °C/50 s + 55 °C/50 s + 72 °C/1 min) + 72 °C/5 min. PCR products were checked by gel electrophoresis, cleaned using the PCR purification kit (Qiagen), cloned into pDrive vector (Qiagen), and transformed to *Escherichia coli*. Clones were sequenced to verify the presence of a specific product. Selected clones were then used for preparation on probes for fluorescence in situ hybridization (FISH) by PCR and labeling using Nick Translation Kit (Roche).

Fluorescence In Situ Hybridization

FISH was performed on mitotic metaphase chromosomes, prepared from root tip cells. The hybridization mix contained 50% formamide, 2× SSC, and 10% dextran sulphate. The labeled DNA (1–5 ng/μl) was denatured, added to slide, and hybridized at 37 °C for 18 h. Slides were then washed 2 × 5' in 2 × SSC at 42 °C, 2 × 5' in 0.1× SSC at 42 °C, 2 × 5' in 2× SSC at 42 °C, 5' in 2× SSC at room temperature, 7' in 4× SSC + 1% Tween, and finally washed in 1× PBS. The chromosomes were counterstained with DAPI, viewed in Olympus AX70 fluorescent microscope, scanned by CCD camera, and analyzed by ISIS software.

Satellite DNA Sequence Analysis

The 454 sequencing reads were analyzed for potential repetitive sequence motifs. Known repeats were identified in clustered reads by sequence similarity to known *R.* satellite sequences RAYSI, RAYSII, RAYSIII, RAE180, and RAE730 (Navajas-Pérez et al. 2005b). The sequences were downloaded from PlantSat (Macas et al. 2002) and National Center for Biotechnology Information GenBank (Benson et al. 2012). Because of the prevalence of plastid DNA, mitochondrial, and retroelement sequences in the 454 data, clusters of reads with top basic local alignment search tool (BLAST) hits mapping to known repetitive sequences were eliminated before further analysis. The remaining 454 reads were subjected to k-mer counting and extension by the algorithm of Macas et al. (2010, 2011). Identified repeat motifs were associated with clusters of origin, and the cluster contigs were visually analyzed for tandemly repeated regions using the polydot program from the EMBOSS package at word size = 9 (Rice et al. 2000). Sequences with tandem subrepeats were broken into their respective monomers at the first point of self-similarity, as determined by running BLAST (Altschul et al. 1997) of the sequence on itself with word size = 7 and a threshold of $e=0.001$. The obtained monomer sequences were used for an exhaustive search of additional matches in the 454 sequence reads. The collected sequences were aligned with CLUSTALX (Thompson et al. 1997) and displayed as a sequence logo using the Weblogo 3.3 program (Crooks et al. 2004). Each multiple alignment was used to generate a consensus monomer sequence at 80% identity threshold.

The same sequences were analyzed with CLANS clustering software (CLuster ANalysis of Sequences, Frickey and Lupas 2004) to reveal families and subgroups of all seven satellites. In each analysis, the software was used to cluster sequencing reads, position the clusters for optimal visualization, and show sequencing reads from male and female plants in contrasting colors.

Results

Genomic Proportion and Composition of Repetitive DNA

We performed one 454 GS FLX platform sequencing run for each male and female genomic DNA and similarity-based clustering of the reads. The first 260 clusters (with more than 57 reads) contained 335,924 reads and represented 58.3% of genome. We obtained also 22,555 smaller clusters (with 2–57 reads) that contained 74,605 reads (12.9% of genome). The other 166,079 reads that remained as singlets represented 29% of genome (fig. 1). We found the main groups of transposable elements, satellites, rDNA loci, and chloroplast DNA. The chloroplast genome was represented by eight clusters (fig. 1). The majority of chloroplast DNA reads probably originated in contaminating cpDNA, even though a proportion might have come from nuclear cpDNA insertions. For this reason, we removed chloroplast DNA reads from further analysis of the nuclear genome.

We focused on TEs and satellites for which we identified individual families together with their genome proportions in male and female individuals (table 1). Although reconstruction of elements was done using 454 data, genome proportions were estimated from Illumina data, which provide more representative results (Macas et al. 2011). The most abundant were Maximus/Sire family of Copia retrotransposons (34.9% in male and 35.6% in female genomes) followed by Chromovirus/CRM and Tat/Ogre families of Gypsy retrotransposons (5.7% and 5.4% in male). LINE elements and two superfamilies of DNA transposons—Mutator and CACTA—were found to make up considerably smaller genome proportions (table 1). All transposable elements represented together about 49% of the genome.

All seven types of satellites together comprised 5.15% of male and 2.54% of female genomes. The most abundant were RAE180 satellite representing 2.72% of the male genome (table 1). The proportion of RAYSI, RAYSII, and RAYSIII was much higher in males in agreement with their Y-specific localization. RAE180 was more abundant in males because of accumulation on both Y chromosomes. Other tandemly arranged sequences are rDNAs that are located on two autosomal pairs in *R. acetosa* (Lengerova and Vyskot 2001). The unexpected difference in proportion of rDNA in male (0.18%) and female (0.21%) was probably caused by higher sensitivity of GC-rich sequences (such as rDNA) to quality of sequencing as was demonstrated by Macas et al. (2011).

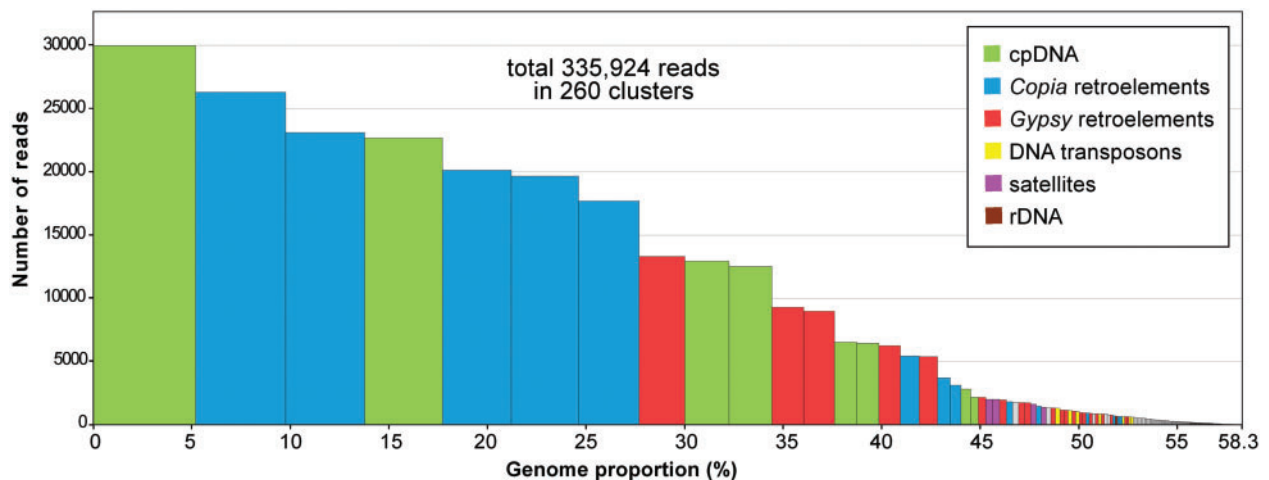


Fig. 1.—Repeat composition of clusters and their genomic proportions. The height of columns represents number of reads in the each clusters, and the width of column indicates genomic proportion of cluster.

To classify Copia and Gypsy elements in more detail, we aligned their RT domains in individual clusters and constructed phylogenetic trees for both superfamilies (fig. 2). Both trees contained subfamilies identified in our clusters (in red) together with representatives of known subfamilies of Copia or Gypsy from other plant species (in black). Among Copia, we identified nine subfamilies of Maximus/Sireviruses, one TAR subfamily, and one Bianca subfamily (fig. 2A). Chromoviruses were dominant among Gypsy elements with seven CRM subfamilies and one Tekay subfamily. We compared *R. acetosa* CRM subfamilies with other CRM elements published by Neumann et al. (2011). The phylogenetic tree based on RT of CRM elements showed that all seven subfamilies found in *R. acetosa* clustered together with group A (supplementary fig. S1, Supplementary Material online), which is known to represent CRM elements having a CR motif and localized in the centromere. We found the CR motif in the seven *R. acetosa* CRM subfamilies, well-conserved in five subfamilies (CL25, 42, 51, 15, and 67), whereas only partially preserved in two subfamilies (CL28 and 48, supplementary fig. S2, Supplementary Material online). In addition to CRM elements, we found one Tekay/Del subfamily (CL37), which also belonged to the chromoviruses, two Tat subfamilies (CL 11 and CL17), and one Athila subfamily (fig. 2B).

We analyzed the structure of selected TE families reconstructed from 454 sequencing data (fig. 3). We were able to discern all main features characteristic for the specific family—gag and pol genes, LTRs, PBS, and PPT regions (fig. 3). In some elements (CL2 and CL17), LTR regions were assembled into one LTR, whereas in other clusters (CL5 and CL25), right and left LTR were distinguished. As examples of very abundant Copia retrotransposons, we present the reconstructed Maximus/Sire subfamilies from CL2 and CL5. We found an extra ORF in the 3'-UTR of Maximus/Sire subfamily corresponding to CL5. As an example of Gypsy retrotransposons,

we used the Tat subfamily (CL17), which has a long 5'-UTR region. We measured the coverage of all these elements with male (blue) and female (red) genomic reads (fig. 3, lines below reconstructed elements). We were unable to reconstruct the whole CRM elements with LTRs from CL42. However, coverage of elements with genomic reads was higher in male than in female, which is consistent with the accumulation of this CRM subfamily on both Y chromosomes (fig. 4J).

Novel Satellite Sequences

K-mer frequency analysis (Macas et al. 2010) of the 454 sequencing reads from clusters not mapped to known sequences helped us to identify two candidates for novel tandem repeats. The first candidate originated from cluster CL45, whereas the second candidate belonged to cluster CL65 (later detected also in CL38 and CL68). These candidate sequences were mapped to assembled contigs in their respective clusters and further adjusted to match existing reads in their size and composition as described in Materials and Methods section, limiting the sequences to a single monomer of the repeat (consensus monomer sequence, supplementary fig. S1, Supplementary Material online). This procedure had led to the discovery of two novel DNA satellites in *R. acetosa*. We named the new satellites RA160 (CL45) and RA690 (CL38, CL65, and CL68), based on their species of origin and their approximate monomer length. Monomer consensus sequences of two novel satellites obtained from multiple alignments of individual 454 reads and the sequences of previously described satellites RAYSI-III, RAE180, and RAE730 assembled by our approach were used to design PCR primers. Using PCR we obtained representative genomic *Rumex* sequences for each family as described in Materials and Methods section (see Preparation of Probes for Fluorescence In Situ Hybridization). These sequences are available under

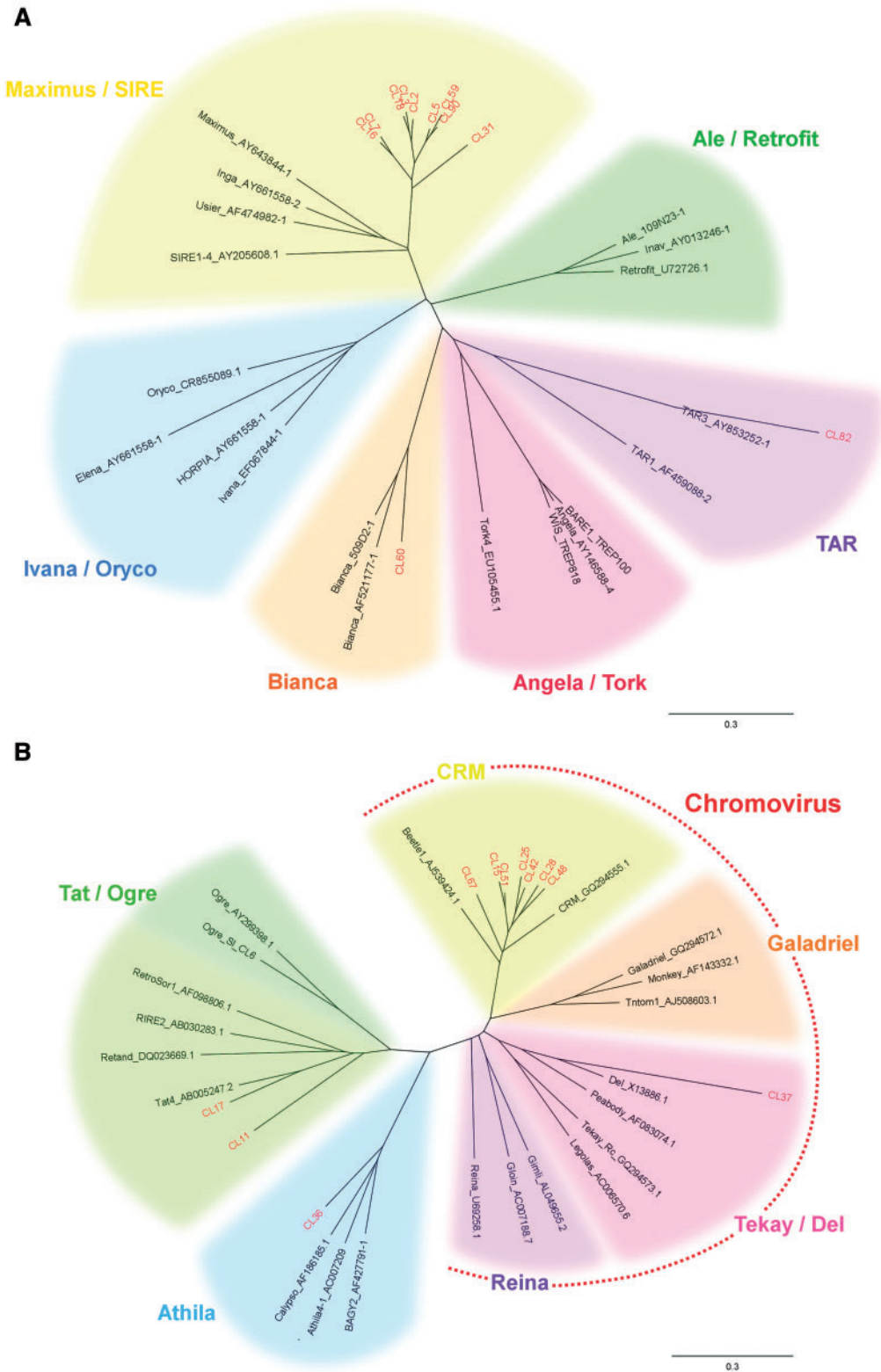


Fig. 2.—Phylogenetic trees of *Rumex acetosa* Copia (A) and Gypsy (B) retrotransposons based on RT sequences. Retrotransposons reconstructed from 454 reads in this study are in red, and representative Copia and Gypsy retrotransposons from other plant species (from GenBank) are in black. Individual families are highlighted by different colors.

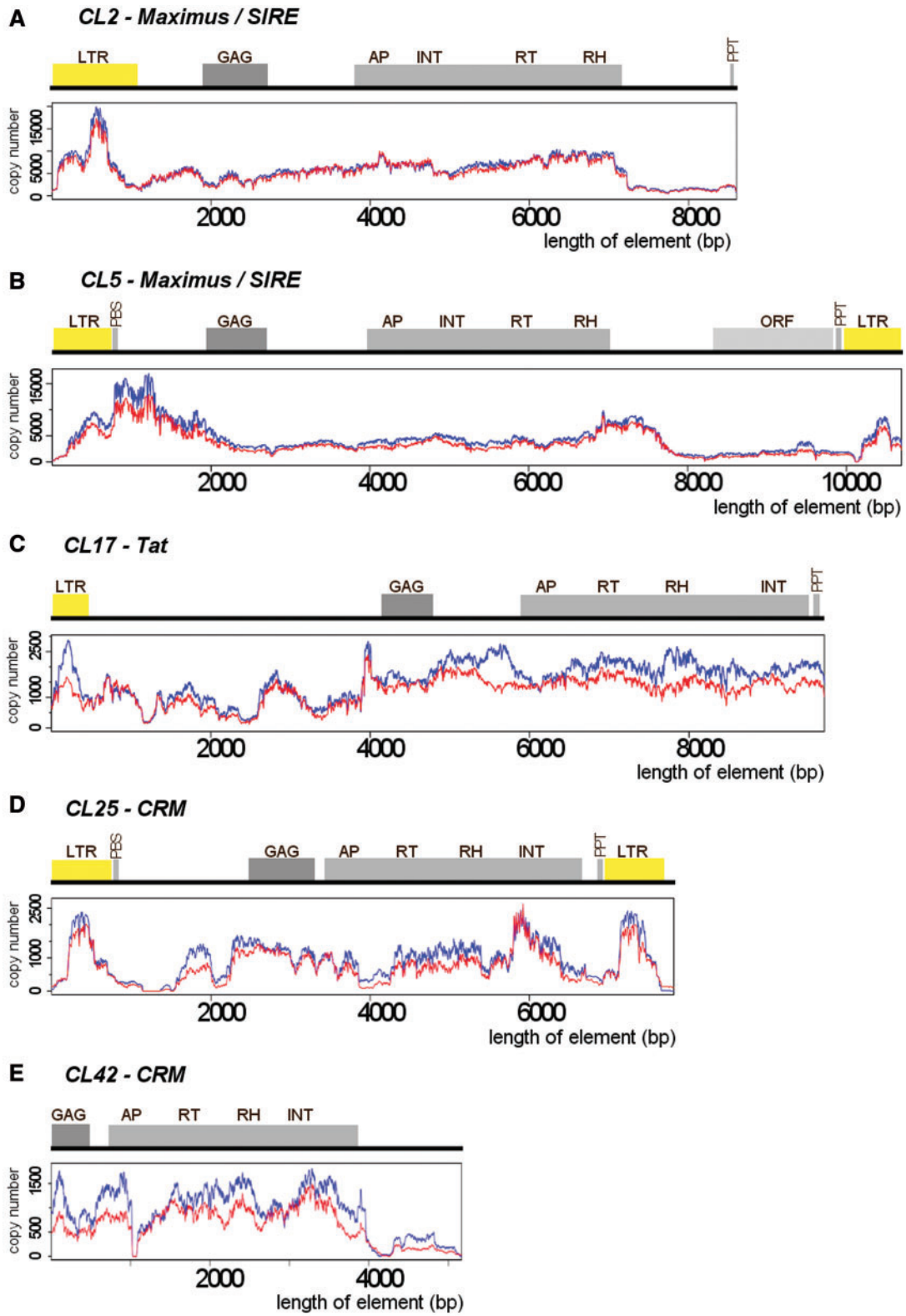


Fig. 3.—Comparison of structure of selected retrotransposon families (A–E). Graphs of coverage by male (in blue) and female (in red) genomic reads are shown under the structure of each element.

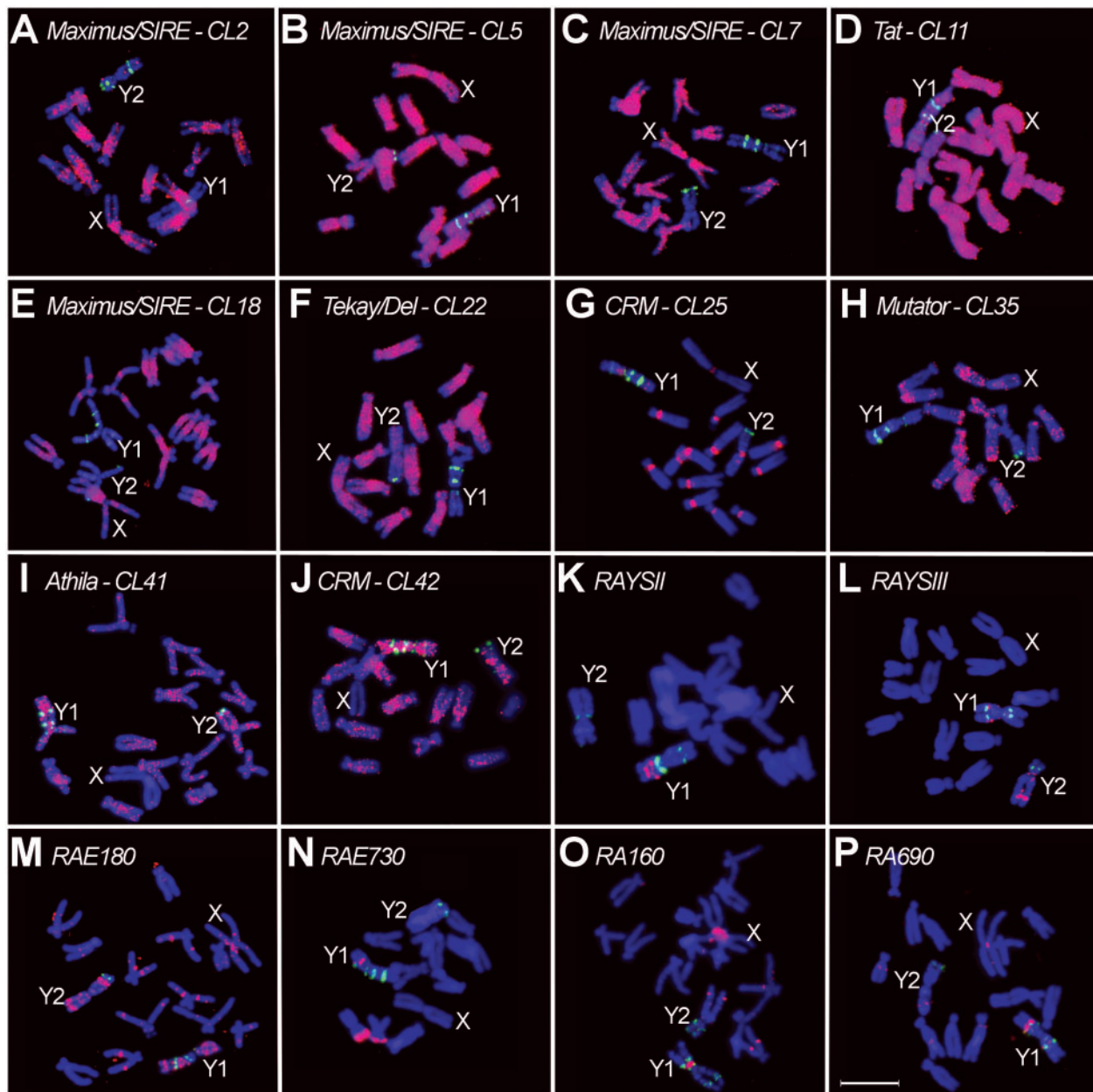


Fig. 4.—Localization of transposable elements and satellites on metaphase chromosomes of *Rumex acetosa* using FISH. The name of transposable element family together with the number of corresponding cluster is inside each figure. Bar indicates 10 μ m.

accession numbers KC310873–KC310879. Genomic proportions of RA160 and RA690 in males are 0.61% and 0.29%, respectively (table 1).

Chromosomal Localization of Transposable Elements and Satellites

To find the chromosomal localization of all main types of transposable elements, we prepared probes representing various parts of individual TE families (supplementary table S1, Supplementary Material online) from the first 60 clusters and

used them for FISH on metaphase chromosomes of the male *R. acetosa*. We obtained several contrasting patterns of chromosomal distribution. The most typical patterns are shown in figure 4. The most abundant subfamilies of Maximus/Sireviruses were distributed on all chromosomes but were absent (CL2, CL7, and CL18) or depleted (CL5) on Y₁ and Y₂ chromosomes (fig. 4A–C and E). Respective subfamilies differed in signal intensity and the extent of subtelomeres labeling—For example, the subfamily corresponding to CL18 was present only a short distance from centromeres (fig. 4E), whereas the two subfamilies (CL5 and CL7) covered the whole

chromosome but the tip. The absence of Maximus/Sire on the Y chromosomes was consistent with its slightly lower genome proportion in male than female individuals (table 1).

The patterns of hybridization of Gypsy elements were more variable. Tat elements (CL11) and Tekay/Del (CL22 and CL37), like Copia elements, were absent from both Y chromosomes (fig. 4D and F). Surprisingly, Athila (CL41) was even absent on the X chromosome (fig. 4I). The CRM elements (CL42) showed accumulation on both Y chromosomes compared with a slight signal on all the other chromosomes (fig. 4J). Accumulation of elements on the Y chromosome or their absence on the X chromosome caused a higher genomic proportion in males than females as calculated from Illumina sequencing data (table 1). Another CRM subfamily (CL25), coming from the same clade in the phylogenetic tree as the Y-accumulated CRM subfamily (CL42), gave specific centromeric signals on all chromosomes—Signals on all autosomes were discrete and much stronger than on either Y chromosome with the weakest signal in the centromere of the X chromosome. There were additional signals to centromeric ones on both Y chromosomes (fig. 4G). The most abundant DNA transposon—the Mutator superfamily—was preferentially located in the subtelomeres of the majority of chromosomes (fig. 4H).

RAYSI satellite was used in most samples as the Y chromosome marker (fig. 4A–P). RAYSI (CL30) was localized in four loci on each arm of the Y₁ chromosome and in two large loci at the p arm and two minor loci on the q arm of the Y₂ chromosome (fig. 4K). RAYSII (CL221) was present as two signals in the middle of the p arm of the Y₁ chromosome but was absent on the Y₂ chromosome (fig. 4K). RAYSIII (CL109, CL126, and CL158) was found in four strong loci on the Y₂ chromosome and three minor loci on the Y₁ chromosome (fig. 4L). RAE180 (CL32 and CL73) was found in many loci on both Y chromosomes and on almost all autosomes and the X chromosome (fig. 4M). RAE730 (CL24) was present as a strong signal on one autosomal pair and as a minor signal on both arms of the Y₁ chromosome (fig. 4N). RA160 gave two strong and one weak signals on the p arm and two minor signals on the q arm of the X, weaker signal on both arms of the Y₁ chromosomes, three minor signals on the q arm of the Y₂ chromosome, and minor signals on two autosomal pairs (fig. 4O). RA690 was localized in two bands on the q arm of the Y₁ chromosome, and one minor signal was present on the q arm of the Y₂ chromosome, in the centromere of the X chromosome, and on two autosomal pairs (fig. 4P). Localization of all studied satellites on the Y₁, Y₂, and X chromosomes is summarized in a schematic map (fig. 5).

Sequence Homogeneity of Satellites and Their Putative Y-Linked Variants

To assess the homogeneity/variability of satellites and in an attempt to reveal potential male-specific (Y linked) satellite

variant(s), we clustered sequence reads corresponding to all seven satellites present in *R. acetosa* genome using CLANS software (Frickey and Lupas 2004). Reads originating from male plants are shown in blue, whereas reads from female plants are shown in red (fig. 6). We found that out of all analyzed satellites, RAYSI, RAYSII, and RAYSIII were the most related ones. Because these satellites are localized mostly on the Y chromosomes, blue symbols prevail in the RAYSI, RAYSII, and RAYSIII clusters. The homogeneity was highest in RAYSI (specifically localized on both Y chromosomes) and in RAE730 (localized only on one autosomal pair, figs. 4 and 6A). On the other hand, RAE180 showed the highest variability (present on all chromosomes, figs. 4 and 6A). A more detailed analysis of the RAE180 cluster revealed a male-specific variant that corresponded to RAE180 satellites present on the Y chromosomes (fig. 6B). The larger size of the subcluster formed by the blue dots indicates that putative Y-linked RAE180 satellites are more diverged compared with their X-linked and autosomal counterparts (fig. 6B). Similarly, we found two subclusters inside RAYSI and RAYSIII clusters (fig. 6C and D). It remains to be determined whether these two clusters correspond to satellite variants localized on the Y₁ and Y₂ chromosomes or represent two subgroups localized on both Y chromosomes. Sequence logos show sequence differences of putative Y-linked variants of RAE180 satellite compared with variant localized on autosomes and the X chromosome (fig. 6E).

Expression of Transposable Elements and Satellites

We performed Illumina platform sequencing of RNA isolated from leaves of male and female *R. acetosa* plants. Reads were mapped onto the clusters corresponding to transposable elements, and the relative expression of individual TE families for each sex was measured (fig. 7). The majority of expression reads corresponded to Maximus/Sire followed by CRM elements and Tat/Ogre elements. All these are most abundant in the genome. However, when the relative expression of each TE family was compared with its genomic proportion, it was evident that CRM, TAR/Tork, Bianca, LINE, and CACTA elements were relatively more transcribed at the expense of Maximus/Sire and Tekay/Del elements. The transcription of other elements (Tat/Ogre, Athila, and Mutator) corresponded more or less to their genomic proportions. We found that some satellites were also expressed—Expression of RAE180 corresponded to its genomic proportion, RA690 was overexpressed, and RAYSI, RAYSII, RAYSIII, RAE160, and RAE730 were underexpressed (fig. 7).

Discussion

This study is the first comprehensive characterization of the repetitive fraction of the nuclear genome of *R. acetosa*, a model dioecious plant with a multiple sex chromosomal system (XY₁Y₂). We found that abundant repetitive DNA

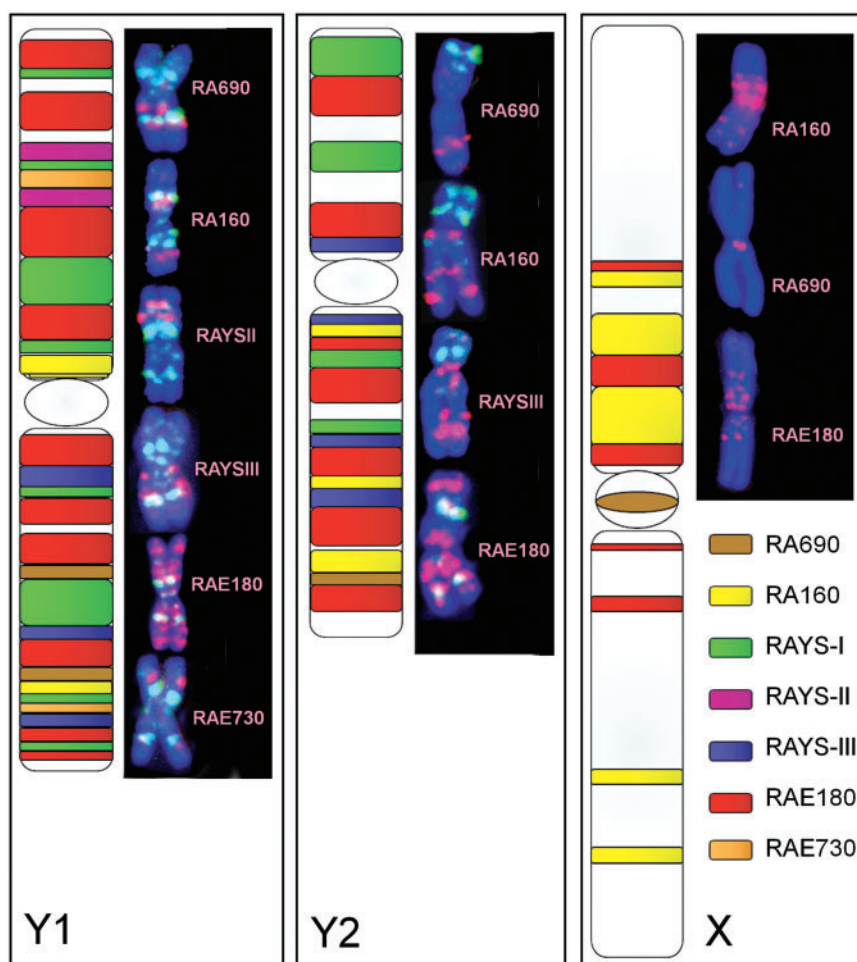


Fig. 5.—Schematic map of satellites localization on the Y1, Y2, and X chromosomes in *Rumex acetosa*. Each of sex chromosomes after FISH with specific satellite probe (red) is shown next to its scheme. Green probe represents RAYS-I in all FISH experiments.

represents at least 49% of the genome. This estimation represents highly and middle abundant repeats found in first 260 clusters, and the proportion of repetitive fraction would be higher if also other clusters with low repetitive fraction are taken into account. We showed that *R. acetosa* genome is composed of Copia LTR retrotransposons, and only smaller proportion is made up of Gypsy retrotransposons, DNA transposons, and satellite DNA. However, it is difficult to conclude why specific (sub)families are more abundant than others because the mechanism(s) governing the colonization of genomes by different groups of TEs are not fully understood. For example, LTR retrotransposons dominate in maize and poplar, non-LTR retrotransposons make up a significant proportion in *Brassica oleracea* and *Gossypium raimondii*, and DNA transposons are most abundant in *Lotus japonicus* and *Fragaria vesca* (for review, see Kejnovsky, Hawkins, et al. 2012).

In our work, we showed that tandem repeats in *R. acetosa* are strongly gathered on Y₁, Y₂, or both Y chromosomes in contrast to the variable chromosomal patterns of TEs. Among

TEs, we showed that although some TEs are accumulated on both Y chromosomes (CRM, CL42), the majority of TEs are missing or under-represented there (Maximus/Sire or Tat/Ogre). It is surprising that despite the fact that all CRM subfamilies contained the CR motif, only the CL25 subfamily was localized in centromeres, indicating that the presence of the CR motif is not a sufficient condition for centromeric localization and that other factors are also important. Our results show that the generally accepted picture of Y chromosomes, as those where all repeats are only accumulated, should be modified. We can explain repeat distribution patterns on sex chromosomes in *R. acetosa* by high rate of colonization of Y₁ and Y₂ chromosomes by satellites that prevented transposable elements from significantly expanding there. Nonetheless, some TEs (CRM, CL42) were able to compete with satellites for Y-linked niches either by higher insertion rate or lower rate of removal.

Our data are relevant to questions on the structure, evolution, and age of sex chromosomes. Known sex chromosomes

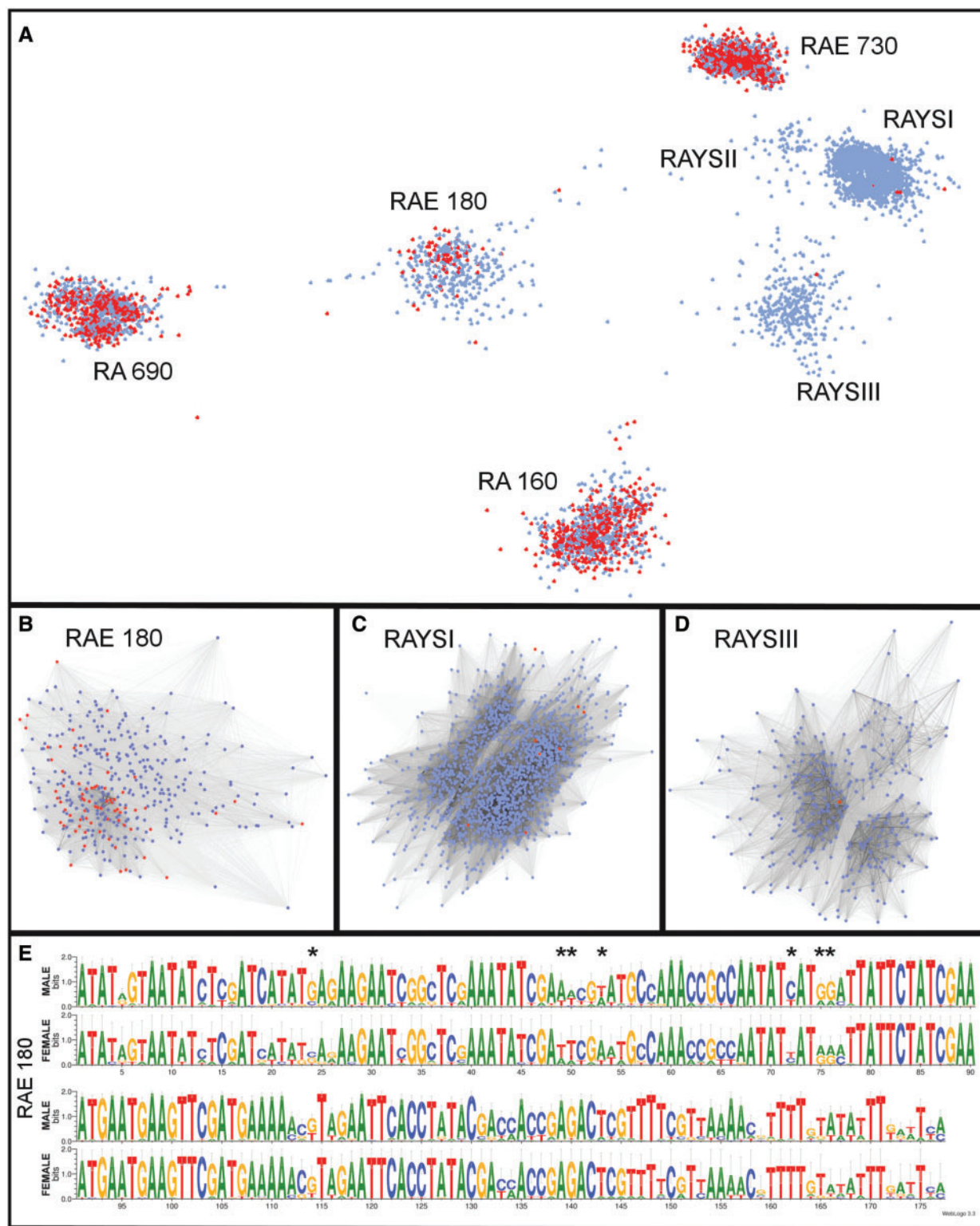


Fig. 6.—Sequence homogeneity of satellites. Clustering of sequence reads originating from male (blue) and female (red) plants using the CLANS software (Frickey and Lupas 2004). Each dot corresponds to a single sequencing read. Reads were mapped by CLANS onto a spherical surface to best represent pairwise sequence similarity and positioned by authors for clear visualization. All satellites together (A), detailed visualization of RAE180 (B), RAYS I (C), and RAYS III (D), and sequence logos of RAE180, where differences between male and female consensus sequences are marked by asterisks (E). Individual clusters were rotated manually into positions showing as much internal structure as possible.

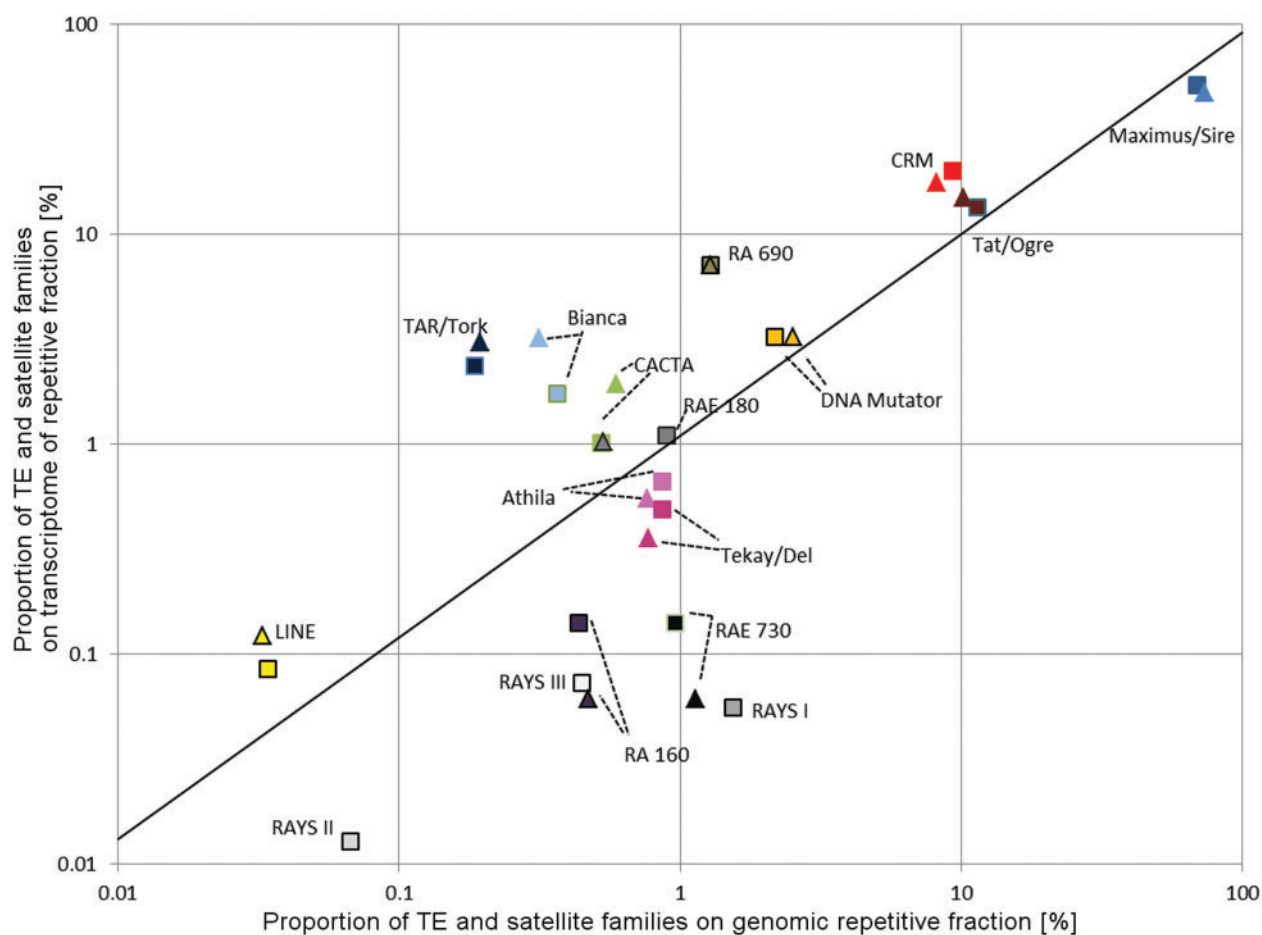


Fig. 7.—Proportion of various TE and satellite families on transcriptome of repetitive fraction plotted against their proportion on repetitive genomic fraction; 100% was represented by all transcripts corresponding only to repetitive DNA (y axis) or to all genomic repetitive fraction (x axis). Expression was measured in *Rumex acetosa* male (squares) and female (triangles) leaves by Illumina platform RNA sequencing and proportion on repetitive DNA fraction was measured by Illumina platform DNA sequencing (see Materials and Methods). Each repeat type is shown by different color.

in plants are mostly in the early stages of their evolution compared with much older mammalian sex chromosomes (Vyskot and Hobza 2004). The young evolutionary age of plant sex chromosomes probably results in some satellites and retrotransposons being weakly accumulated and only slightly enriched on the Y chromosome in the most studied dioecious plant with sex chromosomes—*S. latifolia* (Hobza et al. 2006; Cermak et al. 2008). Our findings show that the situation in *R. acetosa* is different: Some satellites show strong accumulation or even Y chromosome-specific localization and both Y chromosomes that represent together 39% of the genome (Blocka-Wandas et al. 2007). Therefore, the *R. acetosa* Y chromosomes have different sequence composition than the X chromosome and autosomes. They are probably more degenerated and older than sex chromosomes in *S. latifolia*. This view is supported by the finding that both Y chromosomes in *R. acetosa* are heterochromatic, whereas the Y chromosome in *S. latifolia* is euchromatic.

Distribution of various satellites along the whole length of the Y1 and Y2 chromosomes could indicate that there are no evolutionary strata on sex chromosomes in *R. acetosa*, similar to the strata found in human X chromosome (Lahn and Page 1999). If they were present, satellites should accumulate more intensively in a region of the Y chromosomes corresponding to a part of the X chromosome that stopped recombination earlier. However, an exact determination of an existence of the evolutionary strata would need an analysis of dozens of genes located on the X and Y chromosomes. No genes have been identified in *R. acetosa* yet. Accumulation of several satellites at centromere of the X chromosome of *R. acetosa* (fig. 5) could indicate lowered recombination in that region.

Another question concerns the origin of two Y chromosomes in *R. acetosa*. Two alternative explanations have been proposed: the splitting of one original Y chromosome and translocation of an autosome onto the X chromosome (Vyskot and Hobza 2004). The same distribution of repetitive

DNA on both Y chromosomes would indicate that their age was the same and supports the splitting hypothesis. In this study, we found that some tandem repeats had different localizations on the two Y chromosomes. However, our recent data show that CA and CAA microsatellites are strongly and evenly accumulated on both Y chromosomes (Kejnovsky et al. 2013). Thus, we cannot conclude whether both Y chromosomes are of the same or a different age, and hence, we cannot support either of the two hypotheses. Identification of the genes localized on the Y₁, Y₂, and X chromosomes of *R. acetosa* together with detailed characterization of the genomic landscape of these sex chromosomes (sequencing of BAC clones) are necessary to shed light on their age, mechanism of origin, and evolutionary trajectories.

Supplementary Material

Supplementary table S1 and figures S1 and S2 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org>).

Acknowledgments

This work was supported by the Grant Agency of the Czech Republic (grants P305/10/0930 to E.K., P501/10/0102 to B.V., and P501/12/2220 to R.H.), by grant AV0Z50040702 and RVO:60077344 from the Academy of Sciences of the Czech Republic, by the project “CEITEC - Central European Institute of Technology” (CZ.1.05/1.1.00/02.0068) from European Regional Development Fund, and by the project OPVK (CZ.1.07/2.3.00/20.0045) and grant no. ED0007/01/01 Centre of the Region Haná for Biotechnological and Agricultural Research.

Literature Cited

- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.
- Benson DA, et al. 2012. GenBank. *Nucleic Acids Res.* 40(Database issue): D32–D37.
- Blocka-Wandas M, Slivinska E, Grabowska-Joachimiak A, Musial K, Joachimiak AJ. 2007. Male gametophyte development and two different DNA classes of pollen grains in *Rumex acetosa* L., a plant with an XXXY₁Y₂ sex chromosome system and female-biased sex ratio. *Sexual Plant Reprod.* 20:171–180.
- Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Cermak T, et al. 2008. Survey of repetitive sequences in *Silene latifolia* with respect to their distribution on sex chromosomes. *Chromosome Res.* 16:961–976.
- Cioffi MB, Kejnovsky E, Bertollo LAC. 2011. The chromosomal distribution of microsatellite repeats in the genome of the wolf fish *Hoplias malabaricus*, focusing on the sex chromosomes. *Cytogenet Genome Res.* 132:289–296.
- Charlesworth B. 1991. The evolution of sex chromosomes. *Science* 251: 1030–1033.
- Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371:215–220.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14:1188–1190.
- Cunado N, et al. 2007. The evolution of sex chromosomes in the genus *Rumex* (Polygonaceae): identification of a new species with heteromorphic sex chromosomes. *Chromosome Res.* 15:825–832.
- Drummond AJ, et al. 2011. Geneious v5.5. Available from: <http://www.geneious.com>.
- Frickey T, Lupas A. 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20:3702–3704.
- Gvozdev VA, Kogan GL, Usakin LA. 2005. The Y chromosome as a target for acquired and amplified genetic material in evolution. *Bioessays* 27: 1256–1262.
- Heslop-Harrison JS, Schwarzacher T. 2011. Organization of the plant genome in chromosomes. *Plant J.* 66:18–33.
- Hobza R, et al. 2006. An accumulation of a tandem DNA repeats on the Y chromosome in an early stages of sex chromosome evolution. *Chromosoma* 115:376–382.
- Kejnovsky E, et al. 2013. Expansion of microsatellites on evolutionary young Y chromosome. *PLoS One* 8:e45519.
- Kejnovsky E, Hawkins JS, Feschotte C. 2012. Plant transposable elements: biology and evolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ, editors. *Plant genome diversity. Volume 1: Plant genomes, their residents and their evolutionary dynamics.* Wien: Springer. p. 17–34.
- Kejnovsky E, Hobza R, Kubat Z, Cermak T, Vyskot B. 2009. The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity* 102:533–541.
- Kejnovsky E, Leitch I, Leitch A. 2009. Contrasting evolutionary dynamics between angiosperm and mammalian genomes. *Trends Ecol Evol.* 24: 572–582.
- Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12: 656–664.
- Kubat Z, Hobza R, Vyskot B, Kejnovsky E. 2008. Microsatellite accumulation on the Y chromosome in *Silene latifolia*. *Genome* 51:350–356.
- Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. *Nature* 286:964–967.
- Lengerova M, Vyskot B. 2001. Sex chromatin and nucleolar analysis in *Rumex acetosa* L. *Protoplasma* 217:147–153.
- Llorens C, et al. 2011. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* 39(1 Suppl):D70–D74.
- Macas J, et al. 2011. Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. *PLoS One* 6: e27335.
- Macas J, Neumann P, Novak P, Jiang J. 2010. Global sequence characterization of rice centromeric satellite based on oligomer frequency analysis in large-scale sequencing data. *Bioinformatics* 26: 2101–2108.
- Macas K, Meszaros T, Nouzova M. 2002. PlantSat: a specialized database for plant satellite repeats. *Bioinformatics* 18:28–35.
- Mariotti B, et al. 2005. Cloning and characterization of dispersed repetitive DNA derived from microdissected sex chromosomes of *Rumex acetosa*. *Genome* 49:114–121.
- Mariotti B, Manzano S, Kejnovsky E, Vyskot B, Jamilena M. 2009. Accumulation of Y-specific satellite DNAs during the evolution of *Rumex acetosa* sex chromosomes. *Mol Genet Genomics.* 281: 249–259.
- McMillan D, et al. 2007. Characterizing the chromosomes of the platypus (*Ornithorhynchus anatinus*). *Chromosome Res.* 15:961–974.
- Ming R, Bendahmane A, Renner SS. 2011. Sex chromosomes in land plants. *Annu Rev Plant Biol.* 62:485–514.
- Navajas-Pérez R, et al. 2005a. The evolution of reproductive systems and sex-determining mechanisms within *Rumex* (Polygonaceae) inferred from nuclear and chloroplastial sequence data. *Mol Biol Evol.* 22: 1929–1939.

- Navajas-Pérez R, et al. 2005b. Reduced rate of sequence evolution of Y-linked satellite DNA in *Rumex* (Polygonaceae). *J Mol Evol.* 60:391–399.
- Navajas-Pérez R, Quesada del Bosque ME, Garrido-Ramos MA. 2009. Effect of localization, organization, and repeat-copy number in satellite-DNA evolution. *Mol Genet Genomics.* 282:395–406.
- Navajas-Pérez R, et al. 2005c. The origin and evolution of the variability in a Y-specific satellite-DNA of *Rumex acetosa* and its relatives. *Gene* 368: 61–71.
- Neumann P, et al. 2011. Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mob DNA.* 2:4.
- Neumann P, Koblížková A, Navrátilová A, Macas J. 2006. Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. *Genetics* 173:1047–1056.
- Novak P, Neumann P, Macas J. 2010. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11(1):378–389.
- Pokorna M, Kratochvil L, Kejnovsky E. 2011. Microsatellite distribution on sex chromosomes at different stages of heteromorphism and heterochromatinization in two lizard species (Squamata: Eublepharidae: *Coleonyx elegans* and Laceridae: *Eremias velox*). *BMC Genet.* 12:90.
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16:276–277.
- Sakamoto K, Ohmido N, Fukui K, Kamada H, Satoh S. 2000. Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. *Plant Mol Biol.* 44: 723–732.
- Shibata F, Hizume M, Kuroki Y. 1999. Chromosome painting of Y chromosomes and isolation of a Y chromosome-specific repetitive sequences in the dioecious plant *Rumex acetosa*. *Chromosoma* 108:266–270.
- Shibata F, Hizume M, Kuroki Y. 2000. Differentiation and the polymorphic nature of the Y chromosomes revealed by repetitive sequences in the dioecious plant, *Rumex acetosa*. *Chromosome Res.* 8:229–236.
- Schiex T, Gouzy J, Moisan A, de Oliveira Y. 2003. FramedD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. *Nucleic Acids Res.* 31: 3738–3741.
- Skaletsky H, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423: 825–837.
- Steinemann M, Steinemann S. 1992. Degenerating Y chromosome of *Drosophila miranda*: a trap for retrotransposons. *Proc Natl Acad Sci U S A.* 89:7591–7595.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876–4882.
- Vyskot B, Hobza R. 2004. Gender in plants: sex chromosomes are emerging from the fog. *Trends Genet.* 20:432–438.
- Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35(Web Server issue):W265–W268.

Associate editor: Bill Martin