Mol Genet Genomics (2009) 282:395–406
DOI 10.1007/s00438-009-0472-4

ORIGINAL PAPER

Effect of location, organization, and repeat-copy number in satellite-DNA evolution

R. Navajas-Pérez · M. E. Quesada del Bosque · M. A. Garrido-Ramos

Received: 3 April 2009 / Accepted: 11 July 2009 / Published online: 4 August 2009

Abstract Here, we analyze the evolutionary dynamics of a satellite-DNA family in an attempt to understand the effect of factors such as location, organization, and repeat-copy number in the molecular drive process leading to the concerted-evolution pattern found in this type of repetitive sequences. The presence of RAE180 satellite-DNA in the dioecious species of the plant genus *Rumex* is a noteworthy feature at this respect, as RAE180 satellite repeats have accumulated differentially, showing a distinct distribution pattern in different species. The evolution of dioecious *Rumex* gave rise to two phylogenetic clades: one clade composed of species with an ancestral XX/XY sex chromosome system and a second, derived clade of species with a multiple sex–chromosome system XX/XY_1Y_2. While in the XX/XY dioecious species, the RAE180 satellite-DNA is located only in a small autosomal locus, the RAE180 repeats are present also in a small autosomal locus and additionally have been massively amplified in the Y chromosomes of XX/XY_1Y_2 species. Here, we have found that the RAE180 repeats of the autosomal locus of XX/XY species are characterized by intra-specific sequence homogeneity and inter-specific divergence and that the comparison of individual nucleotide positions between related species shows a general pattern of concerted evolution. On the contrary, both in the autosomal and the Y-linked loci of XX/XY_1Y_2 species, ancestral variability has remained with reduced rates of sequence homogenization and of evolution. Thus, this study demonstrates that molecular mechanisms of non-reciprocal exchange are key factors in the molecular drive process; the satellite DNAs in the non-recombining Y chromosomes show low rates of concerted evolution and intra-specific variability increase with no inter-specific divergence. By contrast, freely recombining loci undergo concerted evolution with genetic differentiation between species as occurred in the autosomal locus of XX/XY species. However, evolutionary periods of rapid sequence change might alternate with evolutionary periods of stasis with variability remaining by the reduced action of molecular mechanisms of non-reciprocal exchange as occurred in XX/XY_1Y_2 species, which could depend on repeat-copy number and the processes involved in their amplification.

Keywords Satellite-DNA · *Rumex* · Sex chromosomes · Concerted evolution · Homogenization

Communicated by S. Hohmann.

R. Navajas-Pérez and M. E. Quesada del Bosque contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-009-0472-4) contains supplementary material, which is available to authorized users.

R. Navajas-Pérez · M. E. Quesada del Bosque · M. A. Garrido-Ramos (✉)
Departamento de Genética, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain
e-mail: mgarrido@ugr.es

Introduction

Satellite-DNA evolution is a key subject in molecular evolutionary genetics. Understanding modes, paths and consequences of satellite-DNA evolution is an open question, which requires a profound analysis using appropriate study systems. In this context, the presence of RAE180 satellite-DNA in the dioecious species of the plant genus *Rumex* makes it a noteworthy trait as RAE180 satellite repeats

have accumulated differentially, showing a distinct distribution pattern in different species. The comparative analysis of the effect of these different patterns of location, organization, and repeat-copy number within different *Rumex* species should enhance the understanding of satellite-DNA evolution, especially in this case, as well as in relation to sex-chromosome evolution.

Satellite-DNA families are comprised tandem non-coding short repeated sequences distributed through the eukaryotic genomes at heterochromatin basically in centromeric and subtelomeric regions as well as chromosome-specific amplified regions (Ugarkovic and Plohl 2002). A satellite-DNA family could arise in a phylogenetically short period by explosive amplification (Bachmann and Sperlich 1993). Then, after its formation, the repeats of the family could follow a gradual mode of sequence evolution during a long evolutionary period (Bachmann and Sperlich 1993). The processes by which satellite-DNA families arise are not well known. A set of molecular-exchange mechanisms has been proposed to be involved in its origin by amplification of a tandem array of multi-copy sequences. Those mechanisms include unequal crossing-over (Smith 1976), transposition (Miller et al. 2000), or extrachromosomal rolling-circle replication and reintegration of tandem arrays into the genome (Feliciello et al. 2005, 2006). A recently originated tandem array is initially homogeneous in sequence because of the multi-copy amplification of the same repeat type. In the course of time, random mutations would accumulate and the repeats would diverge. However, the non-allelic repeats of a satellite-DNA family do not evolve independently, but concertedly.

Mechanisms such as unequal crossing-over or gene conversion (Ohta and Dover 1984) can gradually spread a sequence variant throughout a family within a sexual population in a population-genetic process called molecular drive (Dover 2002). Molecular drive leads to high homogeneity levels in a repetitive DNA family for species-diagnostic mutations, which subsequently gives rise to inter-specific genetic divergence, an evolutionary pattern of repetitive sequences known as concerted evolution (Dover 2002). Empirical observations indicate that the production rate of new sequence variants is a slower process than their rate of spread, while the general paucity of transition stages indicates also that the replacement is relatively fast (Ugarkovic and Plohl 2002). However, the rates of homogenization and fixation of sequence variants, i.e., the rates of sequence change, vary for each satellite-DNA family or even for the same satellite-DNA family within different lineages. Levels of sequence identity between repeats would depend on many factors in each species, such as the rates and biases of transfer between homologous and non-homologous chromosomes, number and distribution of repeat units, physical constraints within the genome, generation time, and effective population size as well as selective constraints (Ohta and Dover 1984). We have analyzed the effects of some of these factors (location, rate of genetic transfer, and repeat number, organization, and distribution) on satellite-DNA evolution in dioecious species of the genus *Rumex*.

Molecular phylogeny classifies *Rumex* in three clades (Navajas-Pérez et al. 2005a). The basal clade is composed by hermaphroditic species. A second clade, derived, is composed of polygamous and gynodioecious species. Finally, a third clade is composed of all Eurasian and American dioecious species. There is an evolutionary trend within the genus *Rumex* from hermaphroditism to dioecy, with polygamy and gynodioecy as intermediate stages (Navajas-Pérez et al. 2005a). We also confirmed that, in the evolution of the genus *Rumex*, there is an evolutionary trend for a chromosome-number reduction from $x = 10$ (the basic ancestral chromosome number for the genus, which persisted in all hermaphroditic as well as in some polygamous *Rumex* species) to $x = 7$ (the basic number found in most dioecious species), with intermediate stages having $x = 9$ (gynodioecious and some polygamous species) and $x = 8$ (the Eurasian dioecious *Rumex suffruticosus*) (Degraeve 1975, 1976, 1980; Navajas-Pérez et al. 2005a). The dioecious clade is subdivided into two subclades (Navajas-Pérez et al. 2005a). One basal subclade is composed of species such as *R. suffruticosus* ($x = 8$) and *Rumex acetosella* ($x = 7$) having a XX/XY sex chromosome system, in which females are XX and males are XY. A second subclade is composed of species such as *Rumex acetosa* and *Rumex papillaris* and represents a derived state with a complex XX/XY_1Y_2 sex-chromosome system, in which females have a karyotype $2n = 12 + XX$ and the males are $2n = 12 + XY_1Y_2$. The origin of the double Y chromosome system might come from a chromosomal translocation between one member of a pair of autosomes and the X chromosome in an ancestral XX/XY species (Smith 1969) or from a process of misdivision of an ancestral Y chromosome (Navajas-Pérez et al. 2006). Taking a mean rate of change in plant nuclear DNA of 0.6% per site per million years (Gaut 1998), the ITS mean-distance corrected estimates between clades supported the origin of dioecy in *Rumex* between 15 and 16 mya, while the divergence time between the two dioecious subclades should be 12–13 mya (Navajas-Pérez et al. 2005a).

RAE180 satellite-DNA family is present within the genomes of all Eurasian and American dioecious *Rumex* species while being absent of the rest of *Rumex* species. This data support the idea, on the one hand, that the dioecious group is monophyletic and, on the other hand, that RAE180 repeats appeared in the common ancestor of these species, approximately 15–16 mya. However, whereas in the XX/XY dioecious species, RAE180 satellite-DNA is

located only in a small locus of a pair of autosomes (Cuñado et al. 2007), these repeats are present also in a small autosomal locus and additionally have been massively amplified in the Y chromosomes of XX/XY_1Y_2 species (Cuñado et al. 2007; Navajas-Pérez et al. 2009). Comparing satellite-DNA sequences of two species with a XX/XY_1Y_2 system such as *R. acetosa* and *R. papillaris*, we found that the repeats of the RAE180 satellite-DNA families located in the non-recombining Y chromosomes have lower rates of sequence evolution than the repeats of the autosomic RAE730 satellite-DNA family (Navajas-Pérez et al. 2005b). These data imply that patterns of satellite-DNA evolution depend on the chromosomal location of the repeats, as the absence of recombination between the sex chromosomes seriously influenced the rate of satellite-DNA sequence change. These conclusions were derived from a comparison between different satellite-DNA families. The present research opens the possibility of testing the behavior of repeats from the same satellite-DNA family when they are located in different loci. For this task, we isolated and sequenced RAE180 repeats of the autosomic loci of *R. acetosa* and *R. papillaris* (XX/XY_1Y_2 species), and of the autosomic loci of *R. suffruticosus* and *R. acetosella* (XX/XY species) and we have studied their evolutionary patterns in comparison to those previously characterized for RAE180 repeats of the Y chromosomes of *R. acetosa* and *R. papillaris* (Navajas-Pérez et al. 2005b).

Materials and methods

Materials, DNA extraction, and RAE180 repeats isolation

Female and male plants of *R. suffruticosus* and *R. acetosella* (Puerto de Navacerrada, Segovia, Spain), *R. acetosa* (Sierra Nevada, Spain), and *R. papillaris* (Sierra de Baza, Spain) were used in this study. DNA isolation was performed using DNAeasy mini kit (Qiagen) according to the manufacturer's instructions.

For amplification of RAE180 sequences from female and male plants of *R. suffruticosus* and *R. acetosella* and from female plants of *R. acetosa* and *R. papillaris*, we used the pair of primers Rae180a 5′-TCATCGAACTTCATTC AT-3′ and Rae180b 5′-TATAGTAATATCTCGATC-3′ (Navajas-Pérez et al. 2005b). PCR amplifications were performed in 50-µl reactions containing 10 ng of purified DNA, 2 mM of dNTPs, 2 mM of each primer. and 1.25 units of Taq-polymerase in 10 mM Tris–HCl at pH 8.3, 5 mM KCl, 2 mM MgCl reaction buffer. Thermal cycles consisted of 1 min at 94°, 1 min at 55°, and 1 min at 72°. The PCR products were electrophoresed in agarose gels, after which the bands were cut out of the gel, purified using GenElute PCRClean-Up Kit (SIGMA), ligated to the pCR4-TOPO vector, and cloned using the TOPO TA Cloning Kit (Invitrogen), following the manufacturer's instructions. Recombinant clones were sequenced by the dideoxy-sequencing method using the automatic ABI-Prism 377 sequencer (Applied Biosystems). The EMBL accession numbers for all the sequences analyzed in this article are: AM398600 to AM398639, AJ580328 to AJ580351, AJ580453 to AJ580469, and AJ634478 to AJ634526.

Sequence analysis

For sequence analysis, multiple alignments were carried out using Clustal X (Thompson et al. 1997) followed by manual adjustments. For each species, we determined repeat length, nucleotide composition, and intra-specific variability (measured as nucleotide diversity, π, the average number of nucleotide differences per site between two sequences, Nei 1987) using the software satDNA Analyzer (Navajas-Pérez et al. 2007a). Intra-specific variability was estimated by comparison of all RAE180 repeats. The satDNA Analyzer program has been developed for the analysis of the patterns of variation at each nucleotide position among all units of a repetitive family within a species and for comparing a pair of species. Within each species, the program classifies each site accordingly as monomorphic (the same nucleotide in all the repeats of the species) or polymorphic (the site is variable between repeats).

In comparing the repeats of two species, the satDNA Analyzer program (Navajas-Pérez et al. 2007a) discriminates between shared and non-shared polymorphic sites. The program identifies polymorphic sites shared between two species when the same polymorphism is found in both species (see Supplementary material, Sm1, for a graphic representation). When this occurs, we assume that these are ancestral sites that appeared prior to the split between the two species (Navajas-Pérez et al. 2005b). By contrast, non-shared polymorphic sites are autapomorphies, representing different transitional stages in the process of intra-specific sequence homogenization and inter-specific divergence. Under the assumption that concerted evolution is a time-dependent process, the expected stages of transition during the spread of a variant repeat unit toward its fixation can be defined according to the model of Strachan et al. (1985). This is a method of partitioning the variation by analyzing the patterns of variation at each nucleotide site considered independently among all the repeats of a repetitive family when comparing a pair of species (Strachan et al. 1985; Navajas-Pérez et al. 2007a). This method classifies the sites in terms of six stages (Classes I–VI) in the spread of variant repeats through the family and the species (see Supplementary material, Sm1, for a graphic representation). Briefly, the Class I site represents complete homogeneity across all repeat units sampled from a pair of species, whereas

Classes II, III, and IV represent intermediate stages in which one of the species shows a polymorphism. The frequency of the new nucleotide variant at the site considered is low in Class II and intermediate in Class III, while Class IV represents sites in which a mutation has replaced the progenitor base in most members of the repetitive family in the other species. Class V represents diagnostic sites in which a new variant is fully homogenized and fixed in all the members of one of the species while the other species retains the progenitor nucleotide. A Class VI site represents an additional step over the stage of Class V (new variants appear in some of the members of the repetitive family at a site fully divergent between the two species).

Genetic distances between species were calculated according to the Jukes–Cantor method, comparing all the repeats from each species. Rates of RAE180 evolution were determined then according to the divergence times estimated in Navajas-Pérez et al. (2005a). Shared polymorphic sites were excluded from the alignments, and genetic distances and the rates of RAE180 sequence change were estimated from these alignments following Navajas-Pérez et al. (2005b). This procedure avoids the quantification of ancestral variability, prior to the split of two species from their common ancestor, in the determination of genetic distances between two species for a satellite-DNA family.

Phylogenetic analyses were conducted using MEGA version 4.1 (Tamura et al. 2007) and trees constructed by the neighbor-joining method (Saito and Nei 1987). Thousand bootstrap replicates (Felsenstein 1985) were performed to assess internal support for nodes.

Satellite-DNA quantification

To determine the relative percentage of RAE180 repeats within different genomes (males and females of the four species analyzed), a quantitative PCR (qPCR) approach was developed. PCR conditions were optimized for each assay with regard to annealing temperature and concentration of $MgCl_2$. Conventional PCR reactions for agarose gel-based detection were carried out in 25 µl using 90 ng/µl genomic DNA template, 0.2 mM dNTPs, 200 mM each oligonucleotide primer (Rae180a and Rae180b), optimized $MgCl_2$, and 1 unit Taq DNA polymerase. Each sample was subjected to an initial denaturation of 10 min at 95°C, followed by 40 amplification cycles of denaturation at 95°C for 15 s, optimized annealing for 30 s, followed by extension at 72°C for 30 s. Quantitative PCR reactions were carried out using SensiMix/dT Kit, which includes the internal reference (ROX) (Quantace). Each qPCR reaction comprised 12.5 µl SensiMix/dT, forward and reverse primers (Rae180a and Rae180b) at optimized concentrations of 10 pmol (final concentration), 1 µl SYBR Green I solution 50×, 90 ng/µl genomic DNA template, and sterile water up to a final volume of 25 µl. Each sample was subjected to an initial denaturation of 10 min at 95°C as a pre-cycle heat activation of DNA polymerase, followed by 40 cycles of denaturation at 95°C for 15 s, annealing temperature 56.7°C for 30 s, followed by extension 30 s at 72°C. Quantitative PCR experiments were performed using a Real-Time PCR Cromo 4 thermocycler (BioRad).

Real-time kinetic quantitative PCR determines, for each sample well, the C_t, the fractional cycle number at which the well's accumulating fluorescence crosses a set threshold that is several standard deviations above baseline fluorescence. A plot of C_t versus log(amount of input target DNA) is linear, allowing simple relative quantification of unknowns by comparisons to a standard curve derived from amplification, in the same plate, of serial dilutions of a reference DNA sample. The standard curve was generated from a 7-fold $\log_{10}$ dilution series of a solution of 50 ng/µl of recombinant plasmid DNA containing one RAE180 repeat using Rae180a and Rae180b as PCR primers. All samples, including the external standards and non-template control, were run in triplicate. The amplification efficiency for each primer was determined from the linear slope of standard curve; only primers with a standard curve slope of around −3.3 were used for quantification. When analyzed by qPCR, the dilution series produced a set of standard curves, which were used to calculate the slope value using the Opticon3 software (BioRad).

Results

Isolation and analysis of RAE180 sequences from *R. suffruticosus* and *R. acetosella*

The main PCR product obtained for *R. suffruticosus* with Rae180a and Rae180b primers had an approximate length of 330 bp, although some PCR products of 520 bp were also obtained. Each amplified product was comprised one (330-bp product) or two (520-bp product) RAE180 repeat units plus a fragment of each 5′ and 3′ surrounding repeats. In the case of *R. acetosella*, the amplified PCR product had about 700 bp, composed of three repeat units plus a fragment of each 5′ and 3′ surrounding repeats. PCR products were cloned and sequenced. The number of clones and repeats sequenced and the results on base-pair composition, repeat length, intra-specific variability (measured as nucleotide diversity, π, the average number of nucleotide differences per site between two sequences, Nei 1987) and quantification of RAE180 repeats in the genomes of *R. acetosella* and *R. suffruticosus* are summarized in Table 1. Specifically, intra-specific variability was low for both species (5.1 and 5.4%, respectively), being similar for repeats isolated from different sexes. RAE180 sequences

Table 1 Number of clones and repeats sequenced and results on base pair composition, variability, repeat length, and quantification of RAE180 repeats in the genomes of four dioecious *Rumex* species

Species	Number of clones			Number of repeats			%AT	Variability			Repeat length	Percentage	
	Female	Male	Total	Female	Male	Total		Female	Male	Total		Female	Male
R. acetosella	5	8	13	15	24	39	64.4	0.058	0.054	0.054	167–183	0.0007	0.0007
R. suffruticosus	9	7	16	16	12	28	65.4	0.053	0.040	0.051	181–184	0.04	0.04
R. acetosa	18	32[a]	51	18	42[a]	60	68.4	0.199	0.199[a]	0.209	181–186	0.01	5.60
R. papillaris	15	23[a]	38	15	31[a]	46	67.9	0.153	0.211[a]	0.219	182–184	0.01	5.60

[a] Taken from Navajas-Pérez et al. (2007b)

were poorly represented in the genome of *R. acetosella* (7×10^{-4}% of the genome), while this satellite-DNA family represents the 0.04% of the genome of *R. suffruticosus*.

RAE180 sequences of these two species were phylogenetically analyzed (Figs. 1, 2; see also supplementary material, Sm2, for alignments). RAE180 repeats of *R. suffruticosus* were not specifically clustered by sequence type or source gender (Fig. 1). However, in the case of *R. acetosella*, the grouping of sequences resembled the tandem arrangement of each repeat within each sequenced clone in the order A–B–C (Fig. 2). This indicates the organization of RAE180 monomers in a higher-ordered repeated (HOR) supra-organization constituted by three or more repeats of the RAE180 satellite-DNA. It is worth mentioning here that the third repetition of the HOR underwent a 3′ terminal deletion of 16 bp in all the clones analyzed, giving rise to 167-bp monomers in all cases (Fig. 2b). Notably also, in all clones analyzed, the next repeat in the tandem, which was sequenced partially because of the PCR procedure, also had a 25-bp fixed deletion at the 5′ end (Fig. 2b). In this species, sequence differentiation between the three types of sequences was between 6 and 7%, while the variation between sequences of the same type ranged between 2 and 3%.

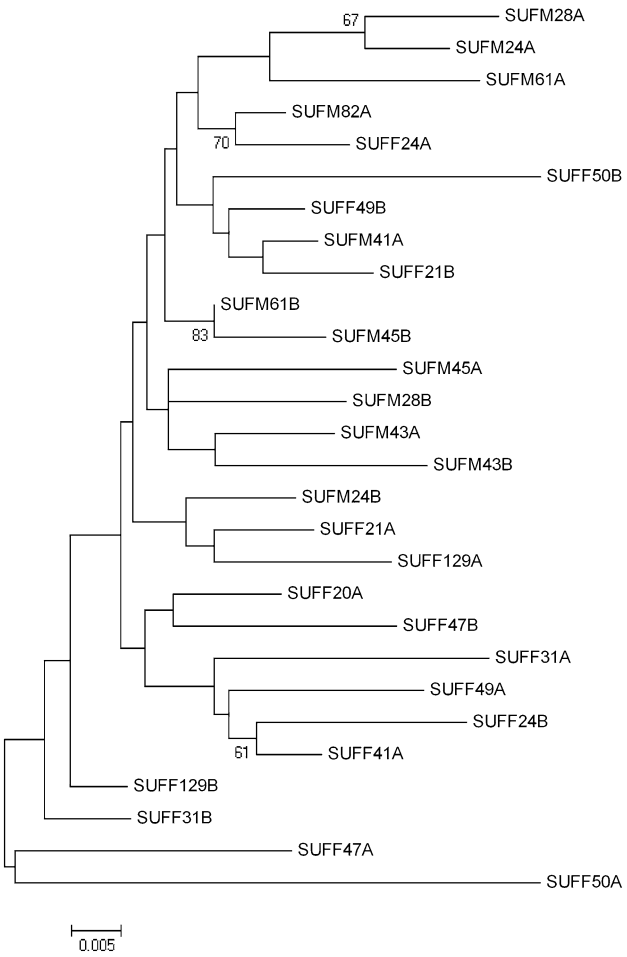
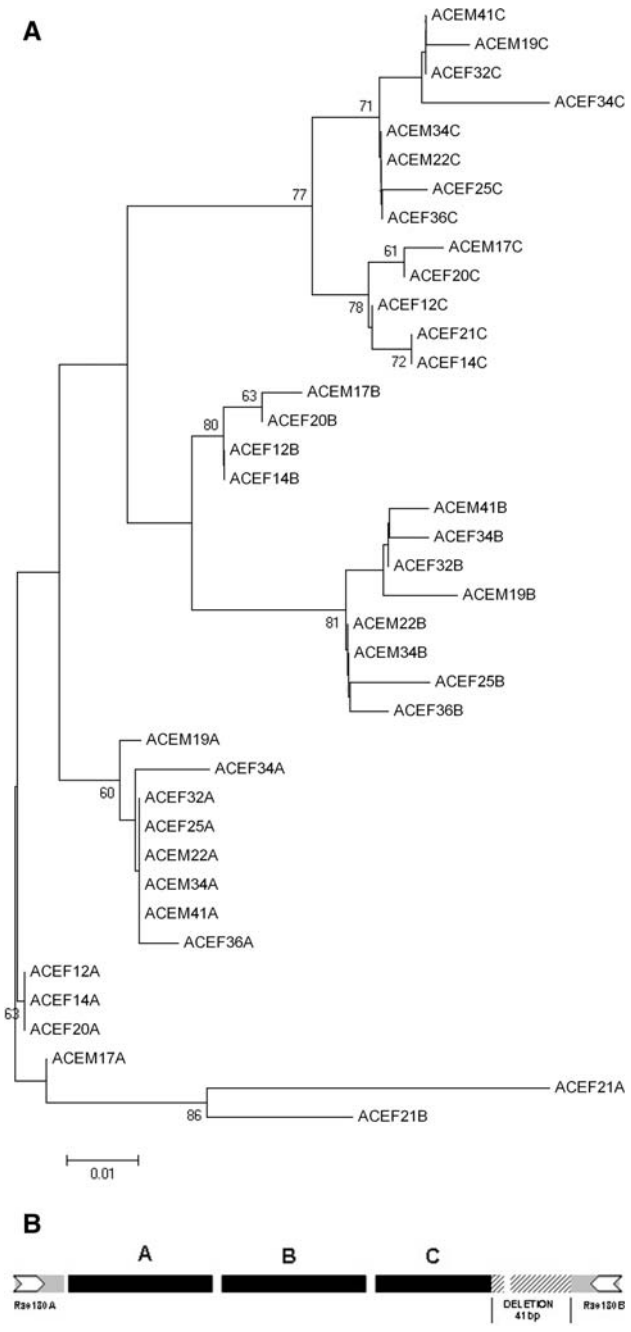


Fig. 1 Neighbor-joining tree of RAE180 repeats isolated from *R. suffruticosus*. RAE180 repeats were not specifically clustered by sequence type or source gender. Sequence code: first three letters (SUF) for species name (*R. suffruticosus*), next letter indicates gender (F, female; M, male) and the number indicates the number of the clone from which the repeat was sequenced. In some cases, within a clone, there were two adjacent RAE180 repeats (see text). In these cases, a final letter in the sequence code indicates the order of the repeat in the clone (A, first and B, second). 1,000 bootstrap replicates were performed to assess internal support for nodes. Only bootstrap values above 60% are indicated

Isolation and analysis of autosomic RAE180 sequences from *R. acetosa* and *R. papillaris*

Females of *R. acetosa* and *R. papillaris* have only the minor autosomic locus in contrast to males, which have two Y chromosomes enriched in RAE180 repeats besides the autosomic locus (Cuñado et al. 2007). To analyze the evolutionary behavior of RAE180 in the autosomic location in such species, we then isolated by PCR amplification the RAE180 repeats from the genomes of females of both species. In both cases, the only PCR product obtained with Rae180a and Rae180b primers had an approximate length of 330 bp, consisting of one RAE180 repeat unit plus a fragment of each 5′ and 3′ surrounding repeats. The products were cloned and sequenced. Results, including quantification of RAE180 repeats in the genome of both females and males, are summarized in Table 1 (data for males were taken from Navajas-Pérez et al. 2007b). In this case, intraspecific variability was 20.9% for *R. acetosa* and 21.9% for



◀ **Fig. 2** **a** Neighbor-joining tree of RAE180 repeats isolated from *R. acetosella*. The grouping of sequences resembles the tandem arrangement of each repeat within each sequenced clone in the order A–B–C, which suggest the organization of RAE180 repeats in a higher-ordered repeated (HOR) supra-organization constituted by three or more repeats of the RAE180 satellite-DNA. Sequence code: first three letters (ACE) for species name (*R. acetosella*), next letter indicates gender (F, female; M, male), the number indicates the number of the clone, and the last letter indicates the order of the repeat in the clone (A, B, or C), as explained in (**b**). Only bootstrap values above 60% are indicated. **b** Schematic representation of the repeat disposition in clones of the *R. acetosella*-RAE180 library. The clone insets comprised three repeat units (A, B, and C, *black boxes*) plus a fragment of each 5′ and 3′ surrounding repeats (*grey boxes*; *arrows* indicate PCR primer localizations). The repeat C underwent a 3′ terminal deletion of 16 bp in all the clones analyzed, giving rise to 167-bp monomers in all cases. Notably also, in all clones analyzed, the next repeat in the tandem, which was sequenced partially because of the PCR procedure, also had a fixed deletion of 25 bp in the 5′ end

R. papillaris, being similar in both females and males. While 0.01% of the genome of females is composed of RAE180 repeats, these sequences represent the 5.6% of the genome of males in both species.

Concerted-evolution analysis of autosomic RAE180 repeats

The mean genetic distance for RAE180 sequences between *R. suffruticosus* and *R. acetosella* proved to be 0.226 (0.247 after excluding shared polymorphic sites). Considering a divergence time of 4 million years between the two species (Navajas-Pérez et al. 2005a), we can estimate the evolutionary rate for this satellite-DNA family at 28×10^{-9} substitutions per site per year, almost three-fold higher than that estimated for RAE180 sequences when comparing males of *R. acetosa* and *R. papillaris* (Navajas-Pérez et al. 2007b). Also, we have studied the sequence divergence by analyzing the patterns of variation at each nucleotide position considered independently among all RAE180 repeats when comparing the two species. Thus, we have classified each site within each species as monomorphic or polymorphic, discriminated between shared and non-shared polymorphic sites between the two species, and classified each non-shared polymorphic site according to the model proposed by Strachan et al. (1985) (see "Materials and methods" section). Table 2 summarizes all these data. A high percentage of Strachan sites of the categories IV, V, and VI were found (14% of the sites per repeat), while 24% of the sites per repeat were Strachan transition stages (II + III). There were eight shared polymorphic sites between the two species. Figure 3 shows the neighbor-joining tree of RAE180 sequences from these two species, which are clustered together according to taxonomic affinity into two main clades supported by 100% of bootstrap replicates.

Comparing the autosomic RAE180 repeats of *R. acetosa* and *R. papillaris* females, we found the genetic distance between the two species to be 0.197. However, almost half of the polymorphic sites within each species were shared between the females of the two species (62 shared polymorphic sites). A similar situation was found for RAE180 sequences of *R. acetosa* and *R. papillaris* males (Navajas-Pérez et al. 2005b). Excluding the shared polymorphic sites, the mean genetic distance between females of both species was 0.045 and thus the evolutionary rate in this case should be 9×10^{-9} substitutions per site per year. There are no Strachan sites of the classes IV–VI, while 34% of the

Table 2 Interspecific comparative analysis of RAE180 repeats

Species comparison	SP (%)	Strachan sites II–III (%)	Strachan sites IV–VI (%)	Genetic distance	Evolutionary rate
R. acetosa versus *R. papillaris males*[a]	108 (59%)	33 (17%)	1 (0.5%)	0.033	8.25×10^{-9}
R. acetosa versus *R. papillaris females*	62 (34%)	62 (34%)	0 (0%)	0.045	9×10^{-9}
R. acetosella versus *R. suffruticosus*	8 (4%)	45 (24%)	27 (14%)	0.226	28×10^{-9}

The table shows the number and percentage of shared polymorphic sites (SP) between two genomes, the genetic distances and the evolutionary rates as well as the results on the classification of each non-shared polymorphic site according to the model proposed by Strachan et al. (1985) in six different stages of transition during the spread of a variant repeat unit toward its fixation. Genetics distances estimates were after excluding shared polymorphic sites

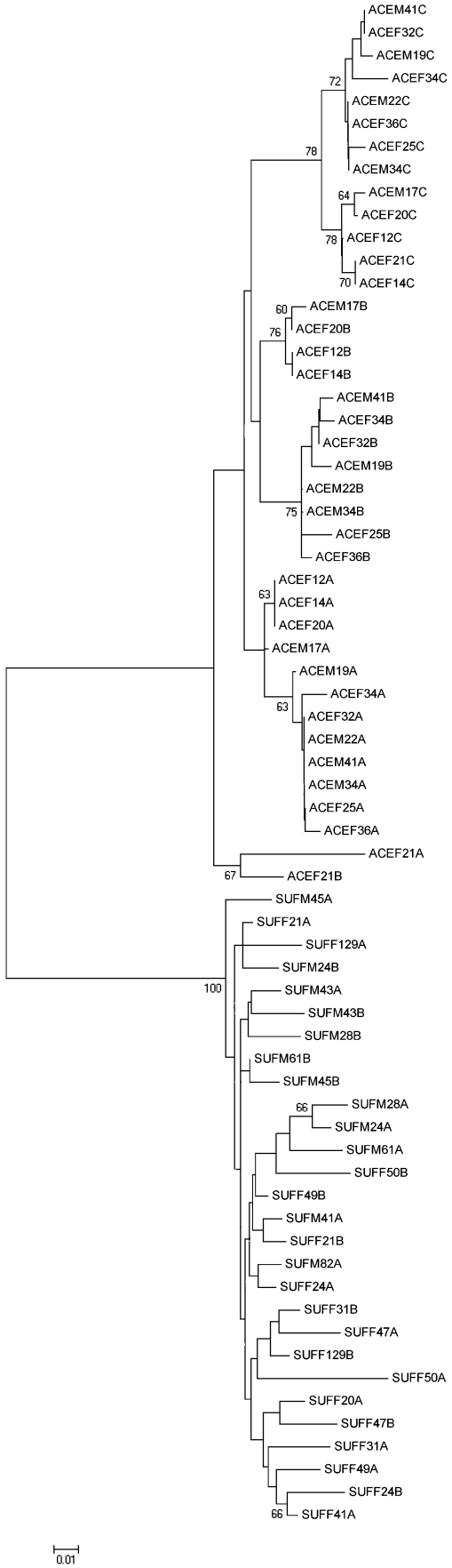
[a] Taken from Navajas-Pérez et al. (2007b)

sites represented Strachan stages II–III (Table 2). Similar to the repeats from males (Navajas-Pérez et al. 2005b), RAE180 repeats from females of both species appear intermixed in the corresponding neighbor-joining tree (Fig. 4; see also Sm3 complementary material for sequence alignments).

Discussion

Satellite-DNA amplification has accompanied the karyotype rearrangements found during species evolution in *Rumex*. Amplification of RAYSI and RAE730 satellite-DNA families occurred in the lineage leading to XX/XY_1Y_2 species (Navajas-Pérez et al. 2005b, 2006). Both satellite-DNA families share a common origin but one is autosomic (RAE730 family) and the other is located in the Y chromosomes (RAYSI) (Navajas-Pérez et al. 2005b, 2006). In turn, the RAE180 family analyzed in this article originated before the split between XX/XY and XX/XY_1Y_2 species and has no relation with the other satellite-DNAs (Navajas-Pérez et al. 2005b). In XX/XY_1Y_2 species, RAE180 repeats are located in the Y chromosomes with an additional small RAE180 autosomic site (Cuñado et al. 2007; Navajas-Pérez et al. 2009). Thus, in the relatively short evolutionary time span from the origin of the XX/XY_1Y_2 clade (12–13 mya; Navajas-Pérez et al. 2005a), the Y chromosomes have undergone an accelerated process of genetic degeneration mediated by the accumulation of RAE180 and RAYSI repetitive sequences (Fig. 5). In fact, we demonstrated here that the 5.6% of the genome of *R. acetosa* and *R. papillaris* males is made up of the RAE180 repeats while the percentage in females is 0.01%. With the assumption of a genome size of 3,234 Mb (Bennett and Leitch 2004) for *R. acetosa*, this would indicate that there are about 1 million copies of RAE180 sequences in the males (repeats located in the Y chromosomes and in the minor locus of a pair of autosomes) and about 2,000 copies of RAE180 repeats in females (only the autosomic locus) of XX/XY_1Y_2 species. *R. acetosa* and *R. papillaris* are two species belonging to a homogeneous group of species characterized by similar morphological and karyological features, including their complex sex-chromosome system plus a sex-determination mechanism based on the X/A balance (Wilby and Parker 1988; Ainsworth et al. 1999). Nevertheless, although both species appear to have karyotype similarities and similar amounts of RAE180 sequences in their Y chromosomes, heterochromatin of the Ys is the subject of rapid evolutionary rearrangement mechanisms that led to differences in satellite-DNA distribution in both species (Navajas-Pérez et al. 2009).

On the other hand, the case for XX/XY species is different. Contrasting with the advanced state of genetic differentiation between the X and the Y chromosomes in *R. acetosa* and *R. papillaris*, we have found that *R. acetosella* and *R. suffruticosus* represent an earlier stage of genetic differentiation between sex chromosomes and also that the Y chromosome in these species is euchromatic and apparently lacks any type of satellite-DNA family (Cuñado et al. 2007). In fact, RAE180 was the only satellite-DNA family detected in the genomes of XX/XY species (Navajas-Pérez et al. 2006; Cuñado et al. 2007). However, RAE180 repeats are poorly represented within the genome of these species. RAE180 was found to be slightly amplified in *R. suffruticosus* in which the repeats are accumulated in a minor locus in a pair of autosomes (Cuñado et al. 2007). In the case of *R. acetosella*, even when PCR and Southern-blot hybridization techniques revealed the presence of RAE180 sequences within its genome, a FISH assay was unable to detect these repeats in their chromosomes, a situation that might be explained by the fact that RAE180 sequences in *R. acetosella* are underrepresented or non-tandemly organized at a level below the resolution of the FISH technique (Cuñado et al. 2007). These data corroborate the quantification developed in this research. In the case of *R. suffruticosus*, RAE180 repeats amount to 0.04% of the genome (for a rough estimate, if we assume a similar genome size to *R. acetosa* for this species, we could estimate some 7,000 copies of RAE180 sequences in *R. suffruticosus*). In contrast, we found that RAE180 satellite DNA represents



◀ **Fig. 3** Neighbor-joining tree of RAE180 repeats from *R. suffruticosus* and *R. acetosella*. The repeats were clustered according to taxonomic affinity into two main clades supported by 100% of bootstrap replicates, one composed of *R. suffruticosus* repeats and the other composed by the *R. acetosella* repeats. Only bootstrap values above 60% are indicated

7.33×10^{-4}% of the *R. acetosella* genome. Genome size estimated for this tetraploid species is 6,566 Mb (Bennett and Leitch 2004), which could indicate an absolute number of about 200 RAE180 repeats in its genome. The differences estimated here for the percentage of RAE180 satellite DNA between the genomes of these two species, *R. suffruticosus* and *R. acetosella*, could explain the difference in FISH results (Cuñado et al. 2007).

The RAE180 satellite-DNA family is present within the genomes of the Eurasian and American dioecious *Rumex* species, while it is absent from the rest of *Rumex* species. This data support the contention that the RAE180 satellite-DNA family was present in the genome of the common ancestor of all dioecious *Rumex* species thus dating the origin of RAE180 sequences at least in 15–16 mya (Navajas-Pérez et al. 2005a). The differences in the amount of RAE180 repeats through the different species analyzed here may indicate differential events of amplification, depending on the evolutionary time and the physical place of accumulation. For RAE180 satellite DNA, there are three different observable cases (Fig. 5): (i) maintenance of RAE180 repeats in low copy in *R. acetosella*, (ii) moderate amplification in one autosomal locus in *R. suffruticosus*, and (iii) massive amplification in Y chromosomes of XX/XY_1Y_2 species and a moderate amplification in one autosomal locus. Despite the common origin of all RAE180 sequences and the short evolutionary time lapse from the split of the two groups of species—XX/XY versus XX/XY_1Y_2—(12–13 mya—Navajas-Pérez et al. 2005a), there are two situations.

1. RAE180 sequences have not amplified within the genomes of XX/XY species except for a single autosomal locus detectable by FISH only in *R. suffruticosus*. The situation in *R. acetosella* in which there are a few copies of RAE180 repeats not detectable by the FISH technique might be considered an ancestral situation maintained in the XX/XY lineage of dioecious *Rumex*. A derived state within this lineage is the amplification process detected in *R. suffruticosus*. Thus, this amplification process might have occurred after the split of these two species and hence might have occurred in the last 4 million years, according to data gathered by Navajas-Pérez et al. (2005a) for ITS sequence divergence between *R. acetosella* and *R. suffruticosus*. Quantitative changes induced by the amplification of RAE180 satellite repeats could occur in the course of

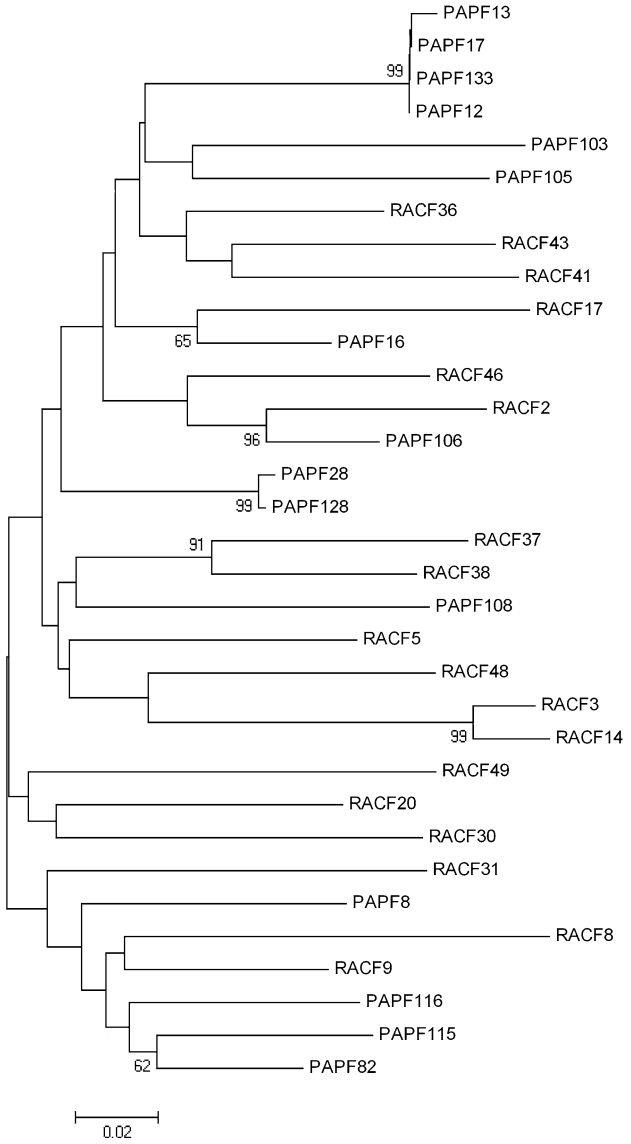


Fig. 4 Neighbor-joining tree of RAE180 repeats isolated from females of *R. acetosa* and *R. papillaris*. RAE180 repeats from females of both species appear intermixed in the tree indicating no differentiation between the two species for these types of sequences

the speciation process, forming a species-specific profile. Thus, sequence divergence is evident between *R. suffruticosus* and *R. acetosella*. In fact, 14% of the nucleotide positions per repeat are differentiated sites (species-diagnostic mutations or Strachan stages IV–VI) between the two species (Table 2). Only 4% of the nucleotide sites per repeat are ancestral polymorphic sites shared between both species. These data indicate effectiveness of the molecular drive process after species split in the spreading of new sequence variants leading to intra-specific homogeneity (5.1 and 5.4% of sequence variation, respectively; Table 1) and inter-specific divergence (22.6% of sequence divergence between the two species; Table 2), a pattern known as concerted evolution. Transitional stages of differentiation (Strachan stages II and III) represent 24% of the nucleotide positions (Table 2). This data support a gradual process of sequence differentiation between *R. suffruticosus* and *R. acetosella* after RAE180 amplification and confirm the double source of the observed pattern: the effect in the molecular drive process of the relatively recent RAE180 amplification in *R. suffruticosus* and the effect of the molecular-exchange homogenizing mechanisms once amplified the locus in this species. Contrasting with that occurring for RAE180 sequences when they are located in the Y chromosomes (Navajas-Pérez et al. 2005b), the autosomal tandem arrays of RAE180 sequences show high degrees of sequence divergence when comparing these two related species. In fact, we detected a rate of sequence substitution of 28×10^{-9} substitutions per site per year, a rate three-fold higher than that estimated for Y-linked RAE180 sequences (Navajas-Pérez et al. 2007b) but similar to the evolutionary rate found for satellite DNA in other species such as in the *Drosophila obscura* complex (Bachmann and Sperlich 1993). In *Rumex*, other autosomic satellite-DNA family, the RAE730 family had a sequence substitution rate of 22×10^{-9}. The comparison between *R. acetosa* and *R. papillaris* that splits 2.5 mya (Navajas-Pérez et al. 2005a) demonstrated 4% of Strachan stages II–III and 30% of Strachan stages IV–VI (Navajas-Pérez et al. 2007b). Intra-specific variability was 4.6% for *R. acetosa* and 2.8% for *R. papillaris* (Navajas-Pérez et al. 2007b). While we have found similar levels of intra-specific variation in *R. suffruticosus* and *R. acetosella*, the distribution pattern of that variation within tandem arrays of *R. acetosella* is specific with the appearance of HORs, which could indicate certain limitations for molecular drive in this species (Pons and Gillespie 2003), perhaps as a consequence of the presence of RAE180 sequences only in basal amounts (Cuñado et al. 2007; this paper).

2. RAE180 repeats were dramatically amplified within the XX/XY_1Y_2 species clade. This amplification process was an independent event to that occurring in the lineage of the XX/XY species. The amplification of repeat-copy number of RAE180 in the Y chromosomes of the XX/XY_1Y_2 lineage might have occurred recurrently and independently several times in each new lineage, leading to each current XX/XY_1Y_2 species. If this was the case, we would expect differential amplification patterns with genetic divergence between species. Assuming the non-recombining nature of the Y chromosomes, we would also expect a high number of polymorphic sites with a species-specific profile. However, we have found similar amplification levels

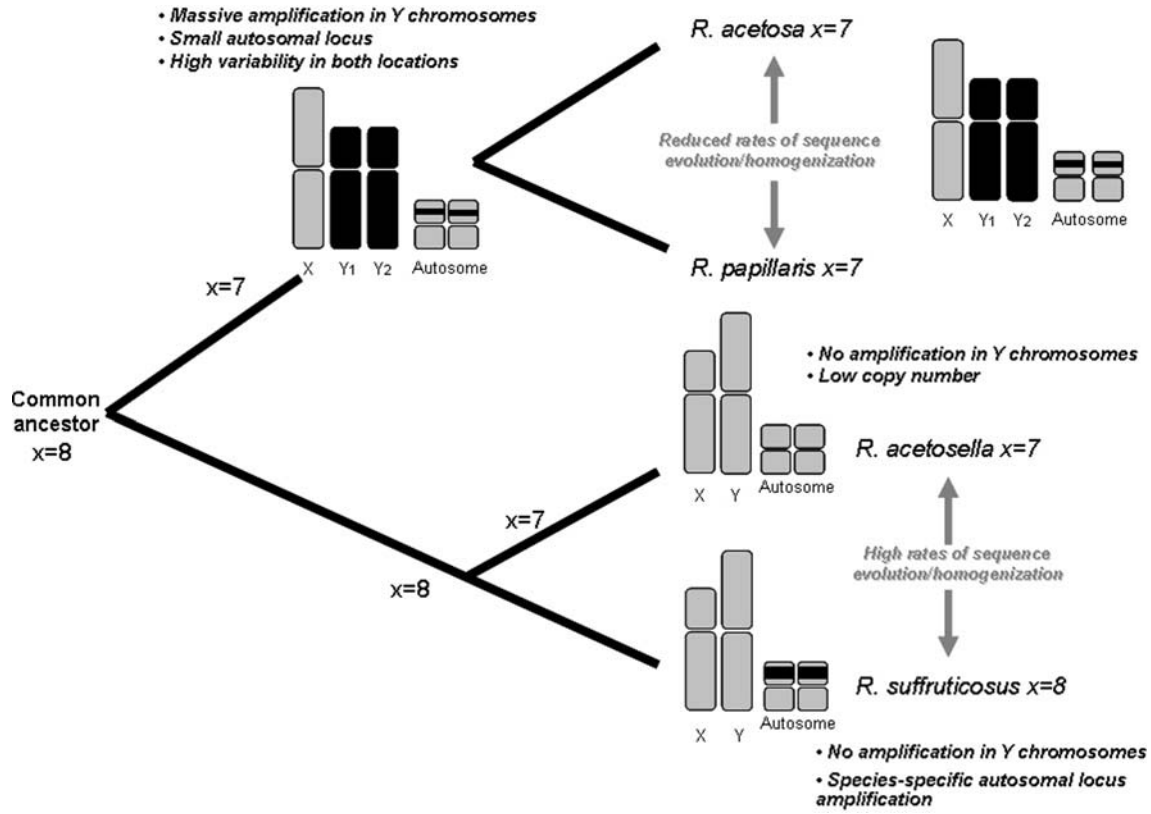


Fig. 5 Schematic representation of the evolutionary dynamics of RAE180 sequences in dioecious *Rumex* species. *Black color* in chromosomes indicates presence of RAE180 sequences

(Table 1) and a high number of shared polymorphic sites (59% of the nucleotide positions per repeat, Table 2) between *R. acetosa* and *R. papillaris*. These two data support the idea: (a) that the main amplification event of the RAE180 satellite-DNA family precedes the split between these two species, despite subsequent new events of amplification and reorganizations (Navajas-Pérez et al. 2009); (b) that most of the intra-specific variability found in each species is ancestral, originated prior to the split of these two species; and (c) that after the split of these two species, processes leading to concerted evolution has failed. In fact, there are high levels of variability in both species (19.9% in *R. acetosa* and 21.1% in *R. papillaris*; Table 1) and only one species-specific diagnostic site (Table 2). These results were explained on the basis of the absence of recombination in the Y chromosomes, a key factor in the molecular drive process (Navajas-Pérez et al. 2005b). The pattern of the autosomal locus of *R. acetosa* and *R. papillaris* resembles Y-linked RAE180 repeat behavior, an unexpected conclusion as a priori we assumed a similar evolutionary scenario for the autosomal loci in both XX/XY and XX/XY$_1$Y$_2$ species. In contrast, the autosomal loci of *R. acetosa* and *R. papillaris* showed high levels of intra-specific variability (23.2 and 15.3%, respectively; Table 1) with no fixed differences between the *R. acetosa* and *R. papillaris* (Table 2). Similar amplification levels (Table 1) and the high number of shared polymorphic sites (34% of the nucleotide positions per repeat, Table 2) between *R. acetosa* and *R. papillaris* again supports the contention that intra-specific variability found in each species at the autosomal locus is ancestral, originated prior to the split of these two species, and, after the split of these two species, processes leading to concerted evolution has failed. This situation might be explained in this case also by reduced rates of recombination for the autosomal locus of *R. acetosa* and *R. papillaris*. It appears that the small autosomal locus found in these two species was singled out of expansion–contraction homogenizing processes that should have lead to concerted evolution. Variability can remain for long evolutionary periods by reduced action of molecular mechanisms of non-reciprocal exchange and sequence variants persist as a library (Mravinac et al. 2002; Meštrović et al. 2006) from which the amplification of a particular divergent variant could give rise later to a new satellite-DNA subfamily (Navajas-Pérez et al. 2006) or to a new enlarged array of the sequence variant (as in *R. suffruticosus*). An alternative explanation

would imply that part of the genetic variability found in autosomes comes from the Y chromosomes by some sort of genetic-transfer mechanism.

In conclusion, this study demonstrates the effect of differential location, organization, and repeat-copy number in the evolution of a satellite-DNA family. Molecular drive leads to the concerted-evolution pattern of satellite-DNA repeats. Molecular mechanisms of non-reciprocal exchange are key factors in the molecular drive process leading to the homogenization of tandem arrays within the genome of a species. Satellite-DNAs in the non-recombining Y chromosomes show low rates of concerted evolution and intra-specific variability increase with no inter-specific divergence. By contrast, freely recombining loci undergo concerted evolution with genetic differentiation between species, as occurs between the autosomal loci of *R. acetosella* and *R. suffruticosus*. However, it appears that satellite-DNA evolution might also have evolutionary periods of stasis with variability remaining by reduced action of molecular mechanisms of non-reciprocal exchange, as occurred between the autosomal loci of *R. acetosa* and *R. papillaris*, something that could depend on repeat-copy number and the processes involved in its amplification.

Acknowledgments Our thanks to Dr. Ginés López González (Real Jardín Botánico de Madrid, Spain), and to the Parque Nacional de Sierra Nevada and to the Parque Natural de la Sierra de Baza, for kindly providing the material analyzed in this article. We also are deeply indebted to Dr. Miguel Burgos of the University of Granada for advice and help with the satellite-DNA quantification procedure. This research was supported by grant CGL2006-00444/BOS of the Ministerio of Ciencia e Innovación (Spain) and Fondo Europeo de Desarrollo Regional (FEDER, EU).

References

Ainsworth CC, Lu J, Winfield M, Parker JS (1999) Sex determination by X:autosome dosage: *Rumex acetosa* (sorrel). In: Ainsworth CC (ed) Sex determination in plants. BIOS Scientific Publishers Limited, Oxford, pp 124–136

Bachmann L, Sperlich D (1993) Gradual evolution of a specific satellite DNA family in *Drosophila ambigua, D. tristis*, and *D. obscura*. Mol Biol Evol 10:647–659

Bennett MD, Leitch IJ (2004) Plant DNA C-values database. Release 3.0

Cuñado N, Navajas-Pérez R, de la Herrán R, Ruiz Rejón C, Ruiz Rejón M, Santos JL, Garrido-Ramos MA (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

Degraeve N (1975) Contribution à l′étude cytotaxonomique des Rumex. I Le genre Rumex. Caryologia 28:187–251

Degraeve N (1976) Contribution à l′étude cytotaxonomique des Rumex. IV Le genre Acetosa Mill. Cellule 71:231–240

Degraeve N (1980) Contribution a l'etude cytotaxonomique des Rumex. III Le genre Acetosella Fourr. Genetica 54:29–34

Dover G (2002) Molecular drive. Trends Genet 18:587–589

Feliciello I, Picariello O, Chinali G (2005) The first characterisation of the overall variability of repetitive units in a species reveals unexpected features of satellite DNA. Gene 349:153–164

Feliciello I, Picariello O, Chinali G (2006) Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from *Rana dalmatina*: molecular evidence of a new mechanism of DNA repair acting on satellite DNA. Gene 383:81–92

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Gaut BS (1998) Molecular clocks and nucleotide substitution rates in higher plants. Evol Biol 30:93–120

Meštrović N, Castagnone-Sereno P, Plohl M (2006) Interplay of selective pressure and stochastic events directs evolution of the MEL172 satellite DNA library in root-knot nematodes. Mol Biol Evol 23:2316–2325

Miller WJ, Nagel A, Bachmann J, Bachmann L (2000) Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. Mol Biol Evol 17:1597–1609

Mravinac B, Plohl M, Meštrović N, Ugarković Đ (2002) Sequence of PRAT satellite DNA "frozen" in some Coleopteran species. J Mol Evol 54:774–783

Navajas-Pérez R, de la Herrán R, López González G, Jamilena M, Lozano R, Ruiz Rejón C, Ruiz Rejón M, Garrido-Ramos MA (2005a) The evolution of reproductive systems and sex-determining mechanisms within *Rumex* (Polygonaceae) inferred from nuclear and chloroplastidial sequence data. Mol Biol Evol 22:1929–1939

Navajas-Pérez R, de la Herrán R, Jamilena M, Lozano R, Ruiz Rejón C, Ruiz Rejón M, Garrido-Ramos MA (2005b) Reduced rates of sequence evolution of Y-linked satellite DNA in *Rumex* (Polygonaceae). J Mol Evol 60:391–399

Navajas-Pérez R, Schwarzacher T, de la Herrán R, Ruiz Rejón C, Ruiz Rejón M, Garrido-Ramos MA (2006) The origin and evolution of the variability in a Y-specific satellite-DNA of *Rumex acetosa* and its relatives. Gene 368:61–71

Navajas-Pérez R, Ruiz Rejón M, Garrido-Ramos MA, Aznarte JL, Rubio-Escudero C (2007a) SatDNA Analyzer: a computing tool for satellite-DNA evolutionary analysis. Bioinformatics 23:767–768

Navajas-Pérez R, Rubio-Escudero C, Aznarte JL, Ruiz Rejón M, Garrido-Ramos MA (2007b) SatDNA Analyzer 1.2 as a valuable computing tool for evolutionary analysis of satellite-dna families: revisiting Y-linked satellite-DNA sequences of *Rumex* (Polygonaceae). In: Hochreiter S, Wagner R (eds) Lecture notes in computer sciences, bioinformatics research and development, vol 4414. Springer-Verlag, Berlin, pp 131–139

Navajas-Pérez R, Schwarzacher T, Ruiz Rejón M, Garrido-Ramos MA (2009) Molecular cytogenetic characterization of *Rumex papillaris*, a dioecious plant with an XX/XY1Y2 sex chromosome system. Genetica 135:87–93

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Ohta T, Dover G (1984) The cohesive population genetics of molecular drive. Genetics 108:501–521

Pons J, Gillespie RG (2003) Common origin of the satellite DNAs of the Hawaiian spiders of the genus *Tetragnatha*: evolutionary constraints on the length and nucleotide composition of the repeats. Gene 313:169–177

Pons J, Petitpierre E, Juan C (2002) Evolutionary dynamics of satellite DNA family PIM357 in species of the genus *Pimelia* (Tenebrionidae, Coleoptera). Mol Biol Evol 19:1329–1340

Saito N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Smith BW (1969) Evolution of sex-determining mechanisms in *Rumex*. Chromosomes Today 2:172–182

Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. Science 191:528–535

Strachan T, Webb D, Dover G (1985) Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. EMBO J 4:1701–1708

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Ugarkovic D, Plohl M (2002) Variation in satellite DNA profiles—causes and effects. EMBO J 21:5955–5959

Wilby AS, Parker JS (1988) Recurrent patterns of chromosome variation in a species group. Heredity 61:55–62