Insect Molecular Biology

Insect Molecular Biology (2015) **24**(3), 319–330 doi: 10.1111/imb.12158

Non-random expression of ribosomal DNA units in a grasshopper showing high intragenomic variation for the ITS2 region

M. Ruiz-Estévez[1], F. J. Ruiz-Ruano[1], J. Cabrero, M. Bakkali, F. Perfectti, M. D. López-León and J. P. M. Camacho

Departamento de Genética, Universidad de Granada, Granada, Spain

Abstract

We analyse intragenomic variation of the ITS2 internal transcribed spacer of ribosomal DNA (rDNA) in the grasshopper *Eyprepocnemis plorans*, by means of tagged PCR 454 amplicon sequencing performed on both genomic DNA (gDNA) and RNA-derived complementary DNA (cDNA), using part of the ITS2 flanking coding regions (5.8S and 28S rDNA) as an internal control for sequencing errors. Six different ITS2 haplotypes (i.e. variants for at least one nucleotide in the complete ITS2 sequence) were found in a single population, one of them (Hap4) being specific to a supernumerary (B) chromosome. The analysis of both gDNA and cDNA from the same individuals provided an estimate of the expression efficiency of the different haplotypes. We found random expression (i.e. about similar recovery in gDNA and cDNA) for three haplotypes (Hap1, Hap2 and Hap5), but significant underexpression for three others (Hap3, Hap4 and Hap6). Hap4 was the most extremely underexpressed and, remarkably, it showed the lowest sequence conservation for the flanking 5.8-28S coding regions in the gDNA reads but the highest conservation (100%) in the cDNA ones, suggesting the preferential expression of mutation-free rDNA units carrying this ITS2 haplotype. These results indicate that the ITS2 region of rDNA is far from complete homogenization in this species, and that the different rDNA units are not expressed at random, with some of them being severely downregulated.

Keywords: Amplicon sequencing, B chromosomes, ITS2, rDNA expression.

Introduction

Among the high variety of repetitive DNA sequences making up eukaryote genomes (Britten & Kohne, 1968), ribosomal DNA (rDNA) is one of the most abundant tandem repeats, as a high number of copies allows the synthesis of the massive numbers of ribosomes required during rapid growth periods (Eickbush & Eickbush, 2007). Hundreds or thousands of rDNA units are present in most eukaryote species, but only a fraction of them are active at a given moment (Reeder, 1999). Each 45S cistron of this multigene family is constituted by three rRNA genes (18S, 5.8S and 28S) separated by two internal transcribed spacers (ITS1 and ITS2) and preceded by an external transcribed spacer and an intergenic spacer (Long & David, 1980). For many years, it was believed that every copy in the rDNA arrays showed an identical sequence because natural selection would have constrained the evolution of the coding regions; however, the finding of high sequence similarity among rDNA units also in the non-coding regions, suggested the existence of a mechanism for their active homogenization. Brown *et al.* (1972) first proposed that the rRNA gene copies evolved horizontally, with novel variants arising by mutation and spreading to other units. This homogenization process was later called concerted evolution (Zimmer *et al.*, 1980), and it was suggested to occur through unequal crossover and gene conversion (Dover, 1982; Eickbush & Eickbush, 2007).

Extensive intragenomic variation has been reported for the ITS regions in a growing number of organisms, however, including bacteria (Stewart & Cavanaugh, 2007), fungi (Simon & Weiss, 2008; James *et al.*, 2009; Li

First published online 7 January 2015.

Correspondence: Juan Pedro M. Camacho, Departamento de Genética, Universidad de Granada, 18071 Granada, Spain. Tel.: +34 958248925; fax: +34 958244073; e-mail: jpmcamac@ugr.es

[1]These authors contributed equally to this work.

© 2015 The Royal Entomological Society

et al., 2013), plants (Mayol & Rosselló, 2001; Feliner *et al*., 2004), sponges (Wörheide *et al*., 2004), grasshoppers (Keller *et al*., 2006) and Drosophila (Stage & Eickbush, 2007), which suggests that the efficiency of concerted evolution is variable among organisms.

A suitable approach to analyse intragenomic variation for the ITS regions is whole-genome shotgun sequencing (WGSS), but this is only useful for small genomes. For instance, Ganley & Kobayashi (2007) found that the level of rDNA copy variation was extremely low in five fungi species that had been sequenced by this method, but WGSS is not feasible in species with large genomes that harbour many rDNA units. Alternatively, next-generation sequencing (NGS) of ITS amplicons is increasingly being used for thorough analysis of intragenomic rDNA variation, both in small (Lindner *et al*., 2013) and large (Keller *et al*., 2008) genomes, as NGS provides more information than do traditional methods (e.g. PCR, cloning and Sanger sequencing) by several orders of magnitude, and recent studies suggest that the results of traditional and NGS (454 and Illumina) methods are comparable (Hřibová *et al*., 2011; Matyášek *et al*., 2012). Nevertheless, a number of studies (Gilles *et al*., 2011; Brodin *et al*., 2013; Niklas *et al*., 2013) have highlighted the need to control for sequencing errors in order to avoid false-positives.

Ribosomal RNA genes are among the most highly expressed genes in both prokaryotic and eukaryotic genomes, because their product is highly demanded by cells (Eickbush & Eickbush, 2007); however, it is as yet unknown whether the different rDNA units are randomly expressed or if there is some kind of preference for the expression of some of them. Neither it is clear whether rDNA expression depends on its genomic location or DNA sequence. In many species, rDNA units are highly homogenized and it appears irrelevant whether one or another rDNA unit is expressed, however, a high intragenomic variation makes it conceivable for there to be evolution of some quality control through the preferential expression of some rDNA units over others.

Recently, Teruel *et al*. (2014) have shown extensive intragenomic variation for the ITS regions in the grasshopper *Eyprepocnemis plorans*. This species shows high variation in the number of standard (A) chromosomes carrying rDNA units among populations, from only two chromosome pairs carrying rDNA in Dagestan to all 12 chromosome pairs carrying it in Morocco (López-León *et al*., 2008). The Torrox population, where the present research was performed, is located in southern Spain and shows an intermediate situation, with seven chromosome pairs carrying rDNA (Cabrero *et al*., 2003). The genome of this species is also large (>10 Gb) (Ruiz-Ruano *et al*., 2011) and carries supernumerary (B) chromosomes, i.e. dispensable elements present in ~15% of eukaryotic species (Camacho, 2005). Traditionally, B chromosomes have been considered genetically inactive elements, but recent findings have shown the transcription of some DNA sequences contained in B chromosomes, including rDNA (van Vugt *et al*., 2003; Leach *et al*., 2005; Ruiz-Estévez *et al*., 2012, 2013), pseudogene-like fragments (Banaei-Moghaddam *et al*., 2013) and protein-coding genes (Trifonov *et al*., 2013; Valente *et al*., 2014), suggesting that B chromosomes are not completely inactive.

The finding of high intragenomic variation for the ITS regions in the grasshopper *E. plorans* (Teruel *et al*., 2014) inspired the present research in which we focus on ascertaining whether the different rDNA units within a same genome are expressed at random. For this purpose, we performed tagged PCR and 454 amplicon sequencing of the ITS2 region in 18 *E. plorans* individuals. The analysis of the sequence reads obtained from both genomic DNA (gDNA) and complementary DNA (cDNA) from the same individuals allowed us to get an estimate of the expression efficiency for each sequence variant (i.e. haplotype) as the ratio between its relative frequency in cDNA and gDNA, and this provided an appropriate test for randomness of rDNA unit expression.

Results

Variation for the ITS2 region

The 454 pyrosequencing experiment yielded 150 685 reads, of which 130 742 matched the ITS2 sequence (66 617 gDNA and 64 125 cDNA reads). This implied almost 4× coverage in the gDNA as the 0B males (i.e. those lacking B chromosomes) of this population carry ~15 000 rDNA units and the B chromosome carries ~3000 units (Montiel *et al*., 2014).

The 123 nt 5.8-28S coding region was identical in 97.72% of the whole collection of gDNA reads, suggesting a 2.28% error rate in the whole experiment, if we assume the conservative criterion that all the variants observed in the coding region were attributable to sequencing errors. In great contrast to the coding regions, the most frequent haplotype in the ITS2 region included only 35.79% of total reads. After selecting those gDNA haplotypes showing a frequency higher than 2.28% per male, the total sample of 18 individuals contained six ITS2 haplotypes (Hap1–Hap6; Table 1). As a whole, the 62 322 reads for these six haplotypes accounted for ~93.55% of all gDNA reads. A total of 63 040 reads corresponding to these six haplotypes were obtained from the cDNA, with 3502 per individual on average (SE = 363;Table 1). As an additional test for ITS2 variation, we carried out whole-genome shotgun 454 sequencing of the entire gDNA of an *E. plorans* 2B male (i.e. a male carrying two B chromosomes) from the same population. In spite of the low estimated coverage of that genome (only 0.01×), we found

Table 1. Proportion of reads for ITS2 amplicons obtained from genomic DNA and complementary DNA in the 18 males analysed

Id	Bs	Hap1	Hap2	Hap3	Hap4	Hap5	Hap6	Total reads
				gDNA				
51	0	0.5046	0.3266	0.1477	0	0.0018	0.0193	**2281**
53	0	0.3305	0.2232	0.1786	0	0.2416	0.0260	**6729**
57	0	0.3708	0.3514	0.2185	0	0.0419	0.0174	**2934**
65	0	0.4106	0.3731	0.1542	0	0.0427	0.0194	**5733**
70	0	0.3419	0.2536	0.1947	0	0.1930	0.0168	**2855**
80	0	0.4883	0.3128	0.1638	0	0	0.0352	**4690**
24	1	0.3206	0.3054	0.1041	0.2535	0.0019	0.0146	**3160**
44	1	0.4211	0.2695	0.1826	0.0940	0	0.0328	**5574**
46	1	0.2865	0.4011	0.2053	0.0993	0	0.0076	**1047**
48	1	0.4279	0.3296	0.0973	0.1262	0	0.0190	**3216**
49	1	0.4022	0.1911	0.3022	0.0378	0.0556	0.0111	**450**
54	1	0.4209	0.3127	0.1258	0.1323	0	0.0083	**1081**
66	1	0.3693	0.3378	0.1126	0.1540	0	0.0263	**2513**
69	1	0.4728	0.2664	0.1243	0.1153	0.0010	0.0202	**3018**
55	2	0.3116	0.1902	0.1762	0.3094	0	0.0126	**1351**
63	2	0.2815	0.1588	0.0697	0.4429	0.0349	0.0122	**9295**
50	3	0.4898	0.1342	0.0997	0.2627	0	0.0136	**1915**
62	3	0.3536	0.1567	0.1507	0.3208	0	0.0183	**4480**
Mean		**0.3891**	**0.2719**	**0.1560**	**0.1305**	**0.0341**	**0.0184**	
SE		0.0167	0.0189	0.0130	0.0319	0.0164	0.0018	
				cDNA				
51	0	0.5977	0.2189	0.1733	0	0.0002	0.0099	**4144**
53	0	0.2950	0.3647	0.0437	0	0.2967	0	**5289**
57	0	0.4746	0.1121	0.1942	0	0.2133	0.0057	**2461**
65	0	0.5111	0.3977	0.0263	0	0.0649	0	**4745**
70	0	0.2918	0.3391	0.0746	0	0.2944	0	**2666**
80	0	0.5882	0.3737	0.0377	0	0.0004	0	**4849**
24	1	0	0.9885	0	0	0.0115	0	**261**
44	1	0.1285	0.7991	0.0699	0	0	0.0024	**3276**
46	1	0.2779	0.6625	0.0558	0	0	0.0038	**2616**
48	1	0.7615	0.2041	0.0284	0	0	0.0061	**2960**
49	1	0.0814	0.5539	0.1078	0.0089	0.2463	0.0017	**2923**
54	1	0.6042	0.2385	0.1292	0.0177	0	0.0104	**5091**
66	1	0.3837	0.3359	0.1276	0	0.1527	0	**1277**
69	1	0.4187	0.2844	0.0927	0	0.2041	0	**3375**
55	2	0.6618	0.2377	0.0961	0	0.0003	0.0041	**3652**
63	2	0.7623	0.1320	0.0477	0.0137	0.0403	0.0041	**6877**
50	3	0.7627	0.1099	0.1001	0.0229	0	0.0044	**2748**
62	3	0.6334	0.1509	0.1757	0.0345	0	0.0055	**3830**
Mean		**0.4575**	**0.3613**	**0.0878**	**0.0054**	**0.0847**	**0.0032**	
SE		0.0561	0.0577	0.0131	0.0024	0.0270	0.0008	

gDNA, genomic DNA.

27 complete ITS2 sequences, among which all but one of the six previously identified haplotypes were represented (Hap1–Hap5).

To test a possible association between haplotype frequency and the number of B chromosomes, we performed a set of regression analyses with haplotype read proportion in the gDNA as a dependent variable. This showed significant correlations only for Hap2 ($r = -0.688$, $SE = 0.181$, $t = -3.79$, $df = 16$, $P = 0.0016$, Bonferroni-corrected $P = 0.008$) and Hap4 ($r = 0.859$, $SE = 0.128$, $t = 6.7$, $df = 16$, $P < 0.001$, Bonferroni-corrected $P <$ 0.001). The negative correlation between Hap2 frequency and the B chromosome number explained 47% of the observed inter-individual variance in genomic Hap2 content, which suggests that the B chromosomes are impoverished in this haplotype, compared with A chromosomes. By contrast, the positive correlation between Hap4 frequency and the B chromosome number explained 73.7% of the observed variance between individuals for the frequency of this haplotype. This, together with the complete absence of sequencing reads for this haplotype in 0B individuals, suggests that Hap4 is exclusive of the B chromosomes.

All six haplotypes showed a conserved secondary structure composed of three arms (helix I, helix II and helix III), in coincidence with previous findings in this species

Figure 1. Secondary structure of Hap1 and the five other ITS2 haplotypes (Hap2–Hap6) found in *Eyprepocnemis plorans* after analysing 18 males from the Torrox population. Note that, in respect to Hap1, Hap6 carries a hemicompensatory change in position 97, which is also present in Hap3. But the latter also carries a hemicompensatory change in position 151 and a noncompensatory change in position 153. These two latter changes are also present in Hap2, but it lacks the change in position 97. Hap4 carries an adenine insertion around positions 169–171, whereas Hap5 carries a CAA repetitive insertion of this triplet in positions 168–170. The variation in Gibbs free energy (ΔG) is indicated in Kcal/mol.

(Teruel *et al.*, 2014), with only a few changes located in helix III (see Fig. 1). Thermal stability (ΔG) was also highly conserved, ranging from −93.15 for Hap1 to −87.17 for Hap2, with −92.57 on average (SE = 1.47).

A minimum spanning tree showing the relationships among the six ITS2 haplotypes indicated that Hap1 is the most similar haplotype to that of the *E. p. meridionalis* outgroup (Fig. 2), meaning that it might be the ancestral haplotype in this population, as is also suggested by its central position in the tree. The co-existence of six different haplotypes in a single population, with four differences (including indels and substitutions) between the most divergent ones, suggests low rDNA sequence homogenization efficiency at the ITS2 region in this species, spe-

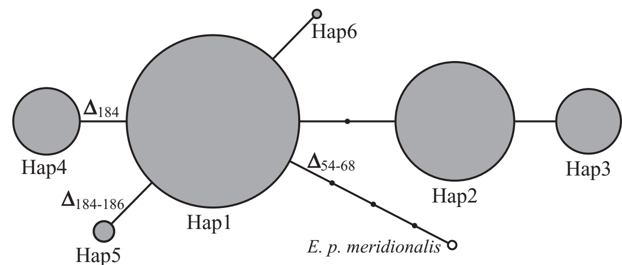

Figure 2. Minimum spanning tree showing the relationships between the six ITS2 haplotypes found in 18 *Eyprepocnemis plorans* males from Torrox, based on their DNA sequences. The nature of these mutational changes is indicated in Figure 1. In brief, they were substitutions, excepting an indel of one nucleotide in Hap4, another of three nucleotides in Hap5, and another of 15 nucleotides in the *E. p. meridionalis* haplotype. Circle diameter is proportional to haplotype abundance.

Table 2. Expression efficiency for the six haplotypes in the 18 individuals analysed

Id	Bs	Hap1	Hap2	Hap3	Hap4	Hap5	Hap6
51	0	1.18	0.67	1.17		0.14	0.51
53	0	0.89	1.63	0.24		1.23	0.00
57	0	1.28	0.32	0.89		5.09	0.33
65	0	1.24	1.07	0.17		1.52	0.00
70	0	0.85	1.34	0.38		1.53	0.00
80	0	1.20	1.19	0.23			0.00
24	1	0.00	3.24	0.00	0.00	6.05	0.00
44	1	0.31	2.97	0.38	0.00		0.07
46	1	0.97	1.65	0.27	0.00		0.50
48	1	1.78	0.62	0.29	0.00		0.32
49	1	0.20	2.90	0.36	0.24	4.43	0.15
54	1	1.44	0.76	1.03	0.13		1.25
66	1	1.04	0.99	1.13	0.00		0.00
69	1	0.89	1.07	0.75	0.00	205.37	0.00
55	2	2.12	1.25	0.55	0.00		0.33
63	2	2.71	0.83	0.68	0.03	1.16	0.33
50	3	1.56	0.82	1.00	0.09		0.32
62	3	1.79	0.96	1.17	0.11		0.30
Mean		**1.19**	**1.35**	**0.59**	**0.05**	**25.17**	**0.25**
SE		0.16	0.20	0.09	0.02	22.54	0.07

The expression efficiency has been calculated as the quotient between the cDNA and gDNA counts. Id = Individual; Bs = Number of B chromosomes.

cifically much lower than the homogenization observed for the partial 5.8-28S coding region included in this work.

Differential expression between haplotypes

An index of expression efficiency for each haplotype was calculated as the ratio between the proportions of reads found in gDNA and cDNA. Table 2 shows the values obtained for each haplotype in each individual. We discarded individual no. 24 for expression analysis because we recovered only 261 reads from its cDNA (i.e. about one order of magnitude lower than the number of reads obtained in any of the remaining individuals). Individual no. 24 also showed a very odd expression pattern. It should be noted that including this individual does not significantly change the results. One sample Student *t*-test, with an expected mean equal to 1, showed random expression of Hap1, Hap2 and Hap5, but significant underexpression of Hap3, Hap4 and Hap6 (Table 3). An additional, indirect, evidence of the nonrandom usage of rDNA units comes from the fact that haplotype diversity for the ITS2 region (Table 4) was significantly lower in the cDNA reads than in the gDNA ones (two-sample *t*-test: $t = 5.05$, $df = 16$; $P < 0.001$), indicating that only a subset of rDNA units is expressed. Remarkably, the B-specific haplotype (Hap4) showed the lowest expression level, even when considering only the five individuals showing Hap4 reads in the cDNA (Fig. 3); therefore, the expression efficiency of this B-specific haplotype was extremely low in B-carrying individuals as a whole.

We then analysed the degree of sequence conservation for the 123 nt of the 5.8S and 28S regions flanking the ITS2 reads, as the percent of reads showing an identical sequence to the canonical rDNA sequence found in *E. plorans* and other grasshopper species (Ruiz-Ruano *et al.*, unpublished data). This showed significantly higher conservation for the reads obtained from the cDNA (99.2%) than for those obtained from the gDNA (98.4%; see values in Table S3; two-sample Student *t*-test, $t = 3.92$, $df = 16$; $P = 0.001$). A comparison of the cDNA reads between the six haplotypes (Table S3) showed that only Hap4 displayed 100% conservation in the flanking coding reads. Remarkably, Hap4 was the haplotype showing the least conserved flanking coding regions in the gDNA (Table S3). This further highlights the non-random expression of the rRNA genes.

In all 12 individuals carrying B chromosomes, we analysed about 20 diplotene cells by means of silver impregnation. A nucleolus was seen attached to the rDNA region of a B chromosome in the five individuals yielding 454 reads with Hap4 in their cDNA (Fig. S1). The remaining individuals showed no Hap4 cDNA reads and six of them showed no nucleoli attached to Bs, whereas one individual showed B-nucleoli in eight out of the 20 diplotene cells analysed. This could be attributable to tissue differences in rDNA expression level among grasshopper body parts, as the nucleoli were visualized in the testis and the cDNA reads were obtained from the rest of the body.

Table 3. One-sample Student *t*-tests on the quotients between the proportions of reads found in the cDNA and gDNA for each haplotype. The null hypothesis is random expression, on which basis we should expect a mean proportion equal to 1

Haplotype	Mean	SE	N	t	df	*P*	Pb
Hap1	1.26	0.15	17	1.75	16	0.0998	0.299
Hap2	1.24	0.18	17	1.35	16	0.1945	0.195
Hap3	0.63	0.09	17	−4.14	16	0.0008	0.003
Hap4	0.05	0.02	11	−40.08	10	<0.0001	<0.001
Hap5*	27.56	25.41	7	1.66	6	0.1487	0.297
Hap6	0.26	0.08	17	−9.72	16	<0.0001	<0.001

*For Hap5, the individual no. 69 was not used because being an extreme outlier. Note that Hap5 showed a very high variation between individuals, for which reason the test failed to reject the null hypothesis, with the available sample. df = degrees of freedom; Pb = Probability after applying the Sequential Bonferroni method.

Table 4. Number of haplotypes, haplotype diversity, nucleotide diversity and average number of nucleotide differences for the ITS2 sequences in the 18 individuals analysed

		gDNA				cDNA			
Id	Bs	H	Hd	π	k	H	Hd	π	k
51	0	5	0.617	0.007	1.28	5	0.565	0.007	1.25
53	0	5	0.750	0.007	1.29	4	0.690	0.006	1.05
57	0	5	0.690	0.007	1.34	5	0.679	0.006	1.17
65	0	5	0.666	0.007	1.28	4	0.576	0.006	1.03
70	0	5	0.744	0.007	1.32	4	0.708	0.006	1.11
80	0	4	0.636	0.007	1.32	4	0.513	0.006	1.04
24*	1	6	0.719	0.006	1.18	2	0.023	0.000	0.05
44	1	5	0.707	0.007	1.33	4	0.340	0.003	0.59
46	1	5	0.706	0.007	1.29	4	0.481	0.005	0.92
48	1	5	0.683	0.006	1.19	4	0.378	0.004	0.78
49	1	6	0.707	0.008	1.43	6	0.614	0.006	1.09
54	1	5	0.692	0.007	1.22	5	0.561	0.006	1.17
66	1	5	0.713	0.007	1.23	4	0.701	0.007	1.22
69	1	6	0.677	0.006	1.20	4	0.694	0.006	1.11
55	2	5	0.740	0.007	1.24	6	0.496	0.006	1.07
63	2	6	0.693	0.005	0.86	6	0.398	0.004	0.69
50	3	5	0.663	0.005	0.92	5	0.396	0.005	0.85
62	3	5	0.725	0.006	1.13	5	0.544	0.006	1.18
Mean		**5.17**	**0.696**	**0.007**	**1.22**	**4.50**	**0.520**	**0.005**	**0.96**
SE		0.27	0.001	<0.001	0.02	0.97	0.029	<0.001	0.09

H, haplotypes; Hd, haplotype diversity; π, nucleotide diversity; k, nucleotide differences; cDNA, complementary DNA; gDNA, genomic DNA.
*Note the extreme outlying values for cDNA, complementary DNA in male no. 24 justifying its exclusion from statistical analyses.

Discussion

Extensive intragenomic variation for the ITS2 region

The existence of six different ITS2 haplotypes in *E. plorans* from a single population is a clear indication of a remarkable intragenomic variation in this species. One of the haplotypes (Hap4) was specific to a B chromosome, because it was not found in any of the B-lacking individuals analysed, and read proportions for it were significantly associated with the number of B chromosomes. This haplotype was unique in showing a characteristic adenine insertion in which previous molecular detection of B-specific transcripts was based (Ruiz-Estévez *et al*., 2012).

The secondary structure of ITS2 plays an important role in defining the cleavage sites for release of rRNA during its maturation (Musters *et al*., 1990; van der Sande *et al*., 1992). The secondary structure of the six haplotypes found in *E. plorans* was conserved, as in previous findings by Teruel *et al*. (2014), and ΔG was approximately similar in the six haplotypes, suggesting that they are putatively functional, and all of them were actually found in the cDNA. The coexistence of six different haplotypes suggests that homogenization is poorly efficient in this species.

At first, it might appear that the fact that the *E. plorans* genome needs to homogenize high amounts of rDNA units could justify its poor homogenization for the ITS2 region. The extensive intraspecific variation for ITS regions reported by Keller *et al*. (2006, 2008) in the grasshopper *Podisma pedestris*, which also shows a gigantic genome (16.56 Gb), would support this inference. Another grasshopper species with a huge genome, however, *Stauroderus scalaris* (15.98 Gb; Belda *et al*., 1991), which carries even more rDNA than *E. plorans* (López-León *et al*., 1999), shows a highly homogenized ITS2 rDNA

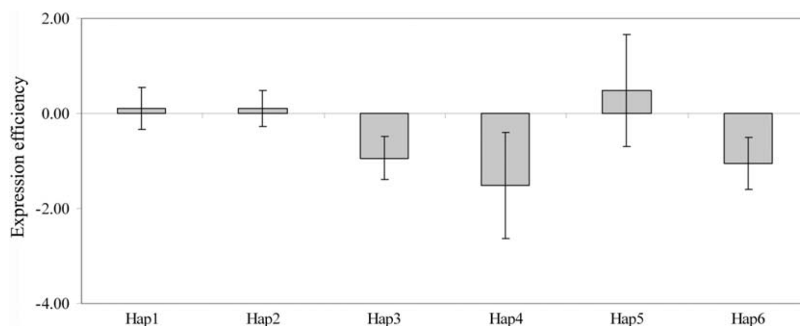

Figure 3. Expression efficiency of the six ITS2 haplotypes (Hap1–Hap6) found in 18 *Eyprepocnemis plorans* males from Torrox, calculated as the quotient between the proportions of reads found in the cDNA and gDNA (transformed to log2 to see overexpressions as positive values and underexpressions as negative values). Error bars indicate 95% CIs.

region (Ruiz-Ruano *et al.*, unpublished data). This suggests that causes other than rDNA quantity or genome size need to be invoked to explain the ITS2 variation found in *E. plorans*. A possible explanation could stem from the recent intragenomic expansion of rDNA between different A chromosomes across the Mediterranean area, from only two in Dagestan (Caucasus; chromosomes 9 and 11) to three in Armenia (with rDNA also in chromosome 10), four in Greece (also in the X chromosome), five in Turkey (also in chromosome 1), up to nine in Spain (seven in Torrox) and all 12 chromosome pairs in Morocco, suggesting the progressive intragenomic spread of rDNA from eastern to western populations (López-León *et al.*, 2008). The analysis of microdissected A and B chromosomes (Teruel *et al.*, 2014) suggests that, if the haplotype structure already existed in rDNA arrays from chromosomes 9 and 11 before the intragenomic spread, then the rDNA in chromosome 9 appears to have been more expansive than that in chromosome 11, since the haplotype found in chromosome 9 (Hap1) was also found in chromosomes 8, X and B, whereas Hap2 was only found in chromosome 11. These results suggest the existence of uneven distribution of ITS2 haplotypes among chromosomes, on which non-random expression could be based (see below).

The positive correlation found between genome size and the amount of rDNA in plants and animals (Prokopowich *et al.*, 2003) might suggest some optimization for both genomic variables, but a massive intragenomic spread of rDNA, such as that which occurred in *E. plorans*, could break this relationship impeding efficient homogenization among non-homologous chromosomes, especially for a genome which, before such spread, only had to homogenize the rDNA units in two non-homologous chromosomes (9 and 11). This situation appears to be ancestral in this species as it is similar in the South African subspecies *E. plorans meridionalis* (López-León *et al.*, 2008). The possibility that homogenization between homologous chromosomes works well in this species is contradicted by the existence of six different haplotypes since, if chromosomes 9 and 11 would have been properly homologously homogenized, prior to intragenomic rDNA spread, we would expect a maximum of two different haplotypes, one coming from chromosome 9 and other from chromosome 11. Alternatively, this variety of haplotypes could be the result of population mixture prior to the recent expansion of this species through the Western Mediterranean area which led to the spread of a same type of B chromosome (B_1) across Tunisia, Sicily, Balearic Islands, the Iberian Peninsula and Morocco (Cabrero *et al.*, 2014).

Non-random expression of rDNA units

The simultaneous analysis of haplotype frequency in gDNA and cDNA of the same individuals allowed the analysis of expression efficiency for each haplotype, demonstrating that the different rDNA units in a single genome are not randomly expressed. To our knowledge, this is the first study performing this kind of combined analysis. Our null hypothesis was that if the different rDNA units were randomly expressed, we would observe ITS2 haplotypes in cDNA at similar proportions to those observed in gDNA. This occurred for Hap1, Hap2 and Hap5, although the latter showed high variation between individuals. The three remaining haplotypes (Hap3, Hap4 and Hap6), however, were significantly underexpressed. The most apparent case of nonrandom expression was Hap4, which is exclusive to the B chromosome and is extremely underexpressed, with only five of the 12 B-carrying males showing cDNA reads, and in much lower proportions than in gDNA. Remarkably, these five males showed the presence of nucleoli attached to B chromosomes in silver-stained diplotene cells, indicating that the rDNA contained in the B chromosomes is able to yield its phenotype, i.e. the nucleolus. Our results also indicate that B chromosome rDNA is completely silenced in some males and is still highly downregulated in those males carrying B chromosomes expressing a nucleolus. In fact, a survey in 11 natural populations showed B-NOR expression in 18 males from seven populations, representing 11.66% of all 156 males analysed (Ruiz-Estévez *et al.*, 2013). This frequency is much lower than the 48% observed in the Torrox population (Ruiz-Estévez *et al.*, 2012).

It would be interesting to know about the possible functionality of the rRNA transcribed from the B chromosome since cells tightly regulate the total amount of nucleolar area (Teruel *et al.*, 2007, 2009), so that the production of rRNA copies by the B chromosome would be paralleled by a decrease in the number of copies produced by the A chromosomes. As mentioned above, our analysis of secondary structure appears to indicate that the ITS2 region produced by the B chromosome (Hap4) is as functional as that transcribed from other chromosomes, although this is important only in terms of ITS2 elimination from the transcript; however, the higher degree of sequence conservation for the flanking 5.8-28S coding regions in the cDNA, compared with that in gDNA, suggests some preference for the expression of those rDNA units in the B chromosome showing the most conserved coding regions. This kind of nonrandom expression of rDNA units is consistent with the observation by Flavell *et al.* (1988) that unmethylated cytosines are not distributed at random in the rDNA but is related to the activity of the NOR in which they reside. Recently, Zhou *et al.* (2013) have proposed a model by which a genome containing copies of the rDNA-specialized R2 retrotransposon selects transcription domains in the region containing the fewest R2 insertions, which is also consistent with nonrandom expression of rDNA units. The fact that Hap4 (the B-specific haplotype)

was the only one associated with 100% of completely conserved flanking coding regions in the cDNA, suggests that this selection is more stringent when the rDNA is transcribed from a B chromosome. In species showing high intragenomic variation for rDNA, the preferential expression of some rDNA units can be crucial for fitness because only a part of the rDNA copies are usually active (Reeder, 1999) and random expression would lead to the production of some defective rRNA molecules.

Interestingly, in the *E. plorans* genome, R2 retroelements are preferentially located in B chromosomes, but Bs actually constitute a sink for R2 since the rDNA of the B chromosomes is rarely active (Montiel *et al.*, 2014) and, as shown here, the B-specific ITS2 haplotype (Hap4) is highly repressed. The fact that the few Hap4 copies escaping silencing (thus observed in the cDNA) show flanking coding regions completely conserved suggests that the genome is extremely careful about which rDNA copies are active, especially in a genome with so much rDNA which only needs to express a small proportion of it. The fact that the five individuals showing Hap4 reads in the cDNA also showed nucleoli attached to the B chromosomes suggests that the RNA molecules produced by transcription of the rDNA in the B chromosome are fully functional. The present results demonstrate, however, that the proportion of B-rRNA molecules is extremely low even in these individuals, in consistency with recent findings by quantitative PCR (Ruiz-Estévez *et al.*, 2014a).

How is haplotype differential expression possible bearing in mind that, at least in *Arabidopsis thaliana*, the units of rRNA regulation are NORs rather than individual rRNA genes (Lewis *et al.*, 2004)? In previous analysis by means of silver impregnation, we showed the interdependence for expression between several chromosome NORs in *E. plorans* (Teruel *et al.*, 2009). If the unit of regulation is the NOR, the differential expression among haplotypes is only possible if haplotypes are not distributed at random in the different chromosomes carrying rDNA. The analysis of the 45S (ITS1-5.8S-ITS2) sequences obtained by microdissection and PCR-cloning from individual chromosomes (Teruel *et al.*, 2014) shows that 42 out of the 72 sequences reported matched five of the six haplotypes analysed here, with Hap1 being the most ubiquitous haplotype found in chromosomes 8, 9, X and B, whereas Hap2 was found only in chromosome 11, and Hap4 was exclusive to the B chromosome. Although more ITS2 sequences need to be individually analysed from each chromosome, this result suggests some structure in ITS2 haplotype distribution among chromosomes. This is presumably a consequence of poor non-homologous homogenization of the ITS2 region in *E. plorans*, although this fact does not apply to the 5.8-28S coding regions which are highly homogenized, presumably because of higher surveillance by natural selection (Eickbush & Eickbush, 2007; Ganley & Kobayashi, 2007). In any case, bearing huge amounts of not completely homogenized rDNA is not a problem for a genome as long as it is able to select conserved coding regions for expression, irrespective of whether they are accompanied by ITS2 regions of one haplotype or another, or even if they are located in a B chromosome.

Recent bioinformatic analysis of hundreds of full sequenced human genomes and transcriptomes has shown extensive variation in rDNA dosage which is positively associated with gene expression for chromatin components targeting the nucleolus, and negatively with mitochondrial DNA abundance (Gibbons *et al.*, 2014). We have recently quantified the number of rDNA copies present in B chromosomes from the Torrox population and found high variation between individuals (Ruiz-Estévez *et al.*, 2014b). Variation in the proportion of Hap4 reads found between B-carrying males (Table 1) also suggests differences in B-rDNA dosage between individuals from this population. This raises the possibility that rDNA dosage in the B chromosome could have something to do with the observed variation between B-carrying individuals for the expression or silencing of the rDNA contained in the B chromosome and, therefore, in the capability of the latter to yield a nucleolus. This is an interesting prospect for future research.

Experimental procedures

Biological samples and karyotypic characterization

We collected 19 adult males of the grasshopper *E. plorans* in the Torrox population (Málaga, Spain). They were anaesthetized prior to dissection to take out the testes which were fixed in freshly prepared 3:1 ethanol : acetic acid and stored at 4 °C for cytological analysis. The bodies of 18 *E. plorans* males were divided into two somatic hemibodies for separate extraction of DNA and RNA, and further tagged PCR 454 amplicon sequencing. The remaining male was similarly processed but the full body was used for extraction of DNA destined to WGSS. All bodies were frozen in liquid nitrogen and stored at −80 °C until DNA and RNA extraction. We determined the number of B chromosomes by squashing two testis follicles in 2% lacto-propionic orcein and visualizing primary spermatocytes at first meiotic prophase or metaphase, under a BX41 Olympus microscope coupled to a DP70 digital camera.

gDNA and RNA extractions, cDNA synthesis and rDNA activity analysis

Genomic DNA and total RNA extractions from frozen hemibodies were performed using ‘GenElute Mammalian Genomic DNA Miniprep Kit’ (Sigma) and ‘Real Total RNA Spin Plus kit’ (Durviz), following manufacturer’s recommendations. We submitted total RNA to a second 20U DNase treatment (Real Star kit, Durviz) after extraction to eliminate any traces of DNA contamination. A PCR analysis on RNA confirmed the absence of DNA in the samples. Quantity and quality (absorbance 260:280 nm = 1.9-2) of gDNA and RNA were measured using Tecan’s Infinite 200

NanoQuant and in a denaturing agarose gel to ensure the absence of RNA degradation. Complementary DNA (cDNA) was synthesized with random hexamers using a SuperScript III First-Strand Synthesis SuperMix Kit (Invitrogen, Carlsbad, CA, USA).

rDNA activity in the B chromosome was analysed cytologically by silver impregnation of testis follicles following the protocol reported by Rufas *et al*. (1982). In grasshoppers, this technique reveals nucleoli attached to active nucleolus organizer regions (NORs, i.e. the chromosome locations of rDNA) during first meiotic prophase. This allows ascertaining which chromosome NORs are active in each cell. We also detected the activity of the rDNA in the B chromosome by the PCR amplification method devised by Ruiz-Estévez *et al*. (2012).

Tagged PCR and amplicon next-generation sequencing

The ITS2 region was amplified from gDNAs and cDNAs of 18 males carrying different numbers of B chromosomes. We then sequenced the amplicons in a Roche 454 run (1/8 of a plate). 6-mer-length tags were added to the 5′ end of the forward ITS3 (5′GTCGATGAAGAACGCAGC3′) and reverse ITS4 (5′ATATGC TTAAATTCAGCGGG 3′) primers (anchored in the 5.8S and the 28S genes, respectively) and, for each different sample, we used a specific combination of both forward- and reverse-tagged primers to separate the sequencing results *in silico*. We designed the six 6mer tags using EDITTAG (Faircloth & Glenn, 2012), differing in four or more nucleotides (Table S1) and we amplified 36 samples (18 gDNAs and 18 cDNAs).

PCR reactions contained 20 ng gDNA or 30 ng cDNA, 0.4 μM of each forward- and reverse-tagged primer (Table S2), 0.2 mM dNTPs, 1X Phusion HF Buffer and 0.4U Phusion® High-Fidelity DNA Polymerase (Thermo Scientific, Waltham, MA, USA) in a final volume of 25 μl. PCR amplifications were carried out in an Eppendorf Mastercycler ep Gradient S (Eppendorf) under the following conditions: initial denaturation for 30 s at 98 °C, 30 cycles of 15 s at 98 °C, 30 s at 60 °C and 10 s at 72 °C, followed by a final extension of 7 min at 72 °C. The amplicons were visualized by electrophoresis in a 1.5% agarose gel, and the bands of about 350 bp were excised and purified using the GenElute Gel Extraction Kit (Sigma-Aldrich, St Louis, MO, USA). We performed two separated reactions with the same combination of tags for each sample to reduce PCR bias, and mixed the product in equimolar amounts for sequencing in a 454 GS FLX Titanium equipment (Roche Diagnostics, Indianapolis, IN, USA), by Macrogen Inc. We submitted the raw reads to the National Center for Biotechnology Information's SRA database with accession number SRR1202012.

Amplicon sequencing data analysis

We wrote a series of custom Python scripts and used them for counting the number of reads corresponding to each sequence (haplotype) in each sample (https://github.com/fjruizruano/amplicon-pyroseq). This was carried out in five consecutive steps. (1) To select the ITS2 amplicons, local alignments of the initial half of the sequencing reads against the forward and reverse primers were carried out using the Smith-Waterman algorithm, as implemented in the EMBOSS suite (Rice *et al*., 2000). We only considered reads with high identity with one of the primers. Additionally, this alignment served to determine the orientation of each read in respect of the aligned primer, because we had reads in both orientations because the 454 adaptors were not included in the primers. We substituted the reads matched to the reverse primer with their reverse-complementary in order to have all the amplicons in the same orientation. (2) We assigned each read to a given sample according to its tag combination (Table S2) using local alignments of the complete reads against the tagged primers. To avoid missassignation, we sorted the alignments from higher to lower similarity and assigned the sample to a given read only if the best alignment showed less than four sequence differences (i.e. the minimum number of differences between tags) in both tagged primers. Selected reads were stored in different files according to their combination of tags. The alignments also allowed us to trim tags and primers from the sequences. (3) We used the Acacia software (Bragg *et al*., 2012) to correct typical 454 pyrosequencing errors, i.e. substitutions and indels in homopolymeric regions. We then generated a file with all found haplotypes per sample and their frequency, and we searched for chimeric sequences with UCHIME (Edgar *et al*., 2011) using the default options for the *de novo* algorithm. (4) To annotate the 5.8S, ITS2 and 28S regions in all the haplotype files, for each sample, we aligned the sequences using MAFFT v7 (Katoh & Standley, 2013) with LINSI options including an additional *E. plorans* ITS2 sequence (accession number JN811835.1) as a reference, annotating its ITS2 region with the ITS2 Database III annotation tool (Koetschan *et al*., 2010). (5) Since the 454 reads included partial regions of the 5.8S and 28S rRNA genes, summing up 123 nt, in addition to the ITS2 sequence, we used these partial coding regions as internal control for sequencing errors. This provided a way to avoid false-positives in identifying genuine ITS2 haplotypes (i.e. variants for the ITS2 region sequence) in the gDNA. For stringency, we considered as sequencing errors all the variation found in these 123 nt, compared with the sequence found by Teruel *et al*. (2014) which is conserved in all *E. plorans* rDNA sequences (accession number JN811835.1). The proportion of reads carrying any variation in respect to the conserved coding sequence was thus our estimate of the maximum error rate of the experiment. To select the genuine ITS2 haplotypes in the different samples of the experiment, we calculated the error rate in the whole experiment and then applied it to every male to avoid discarding genuine haplotypes that belong to only one or few males. The reads were then classified according to the ITS2 haplotype to which they belong.

Whole-genome shotgun sequencing-next-generation sequencing

To test for PCR induced bias, we analysed the genomic ITS2 diversity after direct 454 sequencing of the whole gDNA extracted from another male from the same population. That male carried two B chromosomes and its genome was sequenced in an 1/8 of the 454 GS FLX Plus plate (accession number SRR1200829). We mapped the reads to the *E. plorans* ITS2 (accession number JN811835.1) using Roche's GS Mapper software, and selected those showing at least 90% identity to the complete ITS2 and those that appeared more than twice.

Secondary structure and genetic diversity analyses

We predicted the ITS2 secondary structure of each haplotype and its stability measured by the change in Gibbs free energy (ΔG)

using MFOLD v2.3 (Zuker, 2003) with a folding temperature of 30 °C. The folds were drawn using VARNA (Darty *et al*., 2009). We performed sequence diversity analysis, considering indels, with DnaSP v5.05 (Librado & Rozas, 2009). A minimum spanning tree was built with the different haplotypes based on pairwise differences, using ARLEQUIN v3.5.1.3. (Excoffier & Lischer, 2010) and was visualized with HAPSTAR v0.7 (Teacher & Griffiths, 2011). An additional haplotype from *E. p. meridionalis* (accession number: JX445147), a B-lacking subspecies located in South Africa, was included for minimum spanning tree anchoring.

Statistical analyses

After haplotype selection, we calculated the proportion of reads obtained for each haplotype in the gDNA and cDNA of each individual. We then tested for possible associations between the different haplotypes and the number of B chromosomes by means of regression analysis.

We assessed the degree of expression for each haplotype by comparing the proportions of reads found in both gDNA and cDNA. For this purpose, we calculated an index of expression efficiency for each haplotype as the ratio between the read proportions in the cDNA and gDNA of each male (Table 2). Under the hypothesis of random expression, this ratio should be equal to 1 for all haplotypes. Kolmogorov–Smirnov tests showed that the distributions for each haplotype cDNA/gDNA ratio fitted a normal distribution ($P > 0.05$ in all six cases), for which reason we tested the random expression hypothesis by using Student's one-sample test, with 1 as the expected mean. Finally, the sequential Bonferroni method was applied to minimize type I errors. These analyses were carried out using the STATISTICA software v6 (StatSoft, Inc., 2007). To improve the graphical display of the results, we log2 transformed the obtained values per haplotype and individual, so that values close to zero indicated random expression, whereas those significantly higher or lower than zero indicated overexpression or underexpression, respectively.

Integrity of research

All experiments complied with the current Spanish laws. Mercedes Ruiz-Estévez, Francisco Ruiz-Ruano, Josefa Cabrero, Mohammed Bakkali, Francisco Perfectti, Mª Dolores López-León, and Juan Pedro M. Camacho declare that they have no conflict of interest. All institutional and national guidelines for the care and use of laboratory animals were followed.

Acknowledgements

We thank Tom Eickbush for useful comments on the manuscript, Tatiana López for technical assistance and Karl Meunier for language revision. This work was supported by grants from the Spanish Ministerio de Ciencia y Tecnología (CGL2009-11917) and Plan Andaluz de Investigación (CVI-6649), and was partially performed by FEDER funds. M Ruiz-Estévez was supported by a FPU fellowship from the Spanish Ministerio de Ciencia e Innovación.

References

Banaei-Moghaddam, A.M., Meier, K., Karimi-Ashtiyani, R. and Houben, A. (2013) Formation and expression of pseudogenes on the B chromosome of rye. *Plant Cell* **25**: 2536–2544.

Belda, J.E., Cabrero, J., Camacho, J.P.M. and Rufas, J.S. (1991) Role of C-heterochromatin in variation of nuclear DNA amount in the genus *Chorthippus* (Orthoptera, Acrididae). *Cytobios* **67**: 13–21.

Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P. and Tyson, G.W. (2012) Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat Methods* **9**: 425–426.

Britten, R.J. and Kohne, D.E. (1968) Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161**: 529–540.

Brodin, J., Mild, M., Hedskog, C., Sherwood, E., Leitner, T., Andersson, B. *et al*. (2013) PCR-Induced transitions are the major source of error in cleaned Ultra-Deep Pyrosequencing Data. *PLoS ONE* **8**: e70388. doi: 10.1371/journal.pone.0070388.

Brown, D.D., Wensink, P.C. and Jordan, E. (1972) A comparison of the ribosomal DNAs of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J Mol Biol* **63**: 57–73.

Cabrero, J., Perfectti, F., Gómez, R., Camacho, J.P.M. and López-León, M.D. (2003) Population variation in the A chromosome distribution of satellite DNA and ribosomal DNA in the grasshopper *Eyprepocnemis plorans*. *Chromosome Res* **11**: 375–381.

Cabrero, J., López-León, M.D., Ruiz-Estévez, M., Gómez, R., Petitpierre, E., Rufas, J.S. *et al*. (2014) B_1 was the ancestor B chromosome variant in the western Mediterranean area in the grasshopper *Eyprepocnemis plorans*. *Cytogenet Genome Res* **142**: 54–58.

Camacho, J.P.M. (2005) B chromosomes. In *The Evolution of the Genome* (Gregory, T.R., ed.), pp. 223–226. Academic Press, New York.

Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**: 1974–1975.

Dover, G.A. (1982) Molecular drive: a cohesive mode of species evolution. *Nature* **299**: 111–117.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.

Eickbush, T.H. and Eickbush, D.G. (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* **175**: 477–485.

Excoffier, L. and Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resources* **10**: 564–567.

Faircloth, B.C. and Glenn, T.C. (2012) Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS One* **7**: e42543. doi: 10.1371/journal.pone.0042543.

Feliner, G.N., Larena, B.G. and Aguilar, J.F. (2004) Fine-scale geographical structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in Armefia (Plumbaginaceae). *Ann Bot* **93**: 189–200.

Flavell, R.B., O'Dell, M. and Thompson, W.F. (1988) Regulation of cytosine methylation in ribosomal DNA and nucleolus organizer expression in wheat. *J Mol Biol* **204**: 523–534.

Ganley, A.R.D. and Kobayashi, T. (2007) Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res* **17**: 184–191.

Gibbons, J.G., Branco, A.T., Yu, S. and Lemos, B. (2014) Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. *Nat Commun* **5**: 4850. doi: 10.1038/ncomms5850.

Gilles, A., Meglécz, E., Pech, N., Ferreira, S., Malausa, T. and Martin, J.F. (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* **12**: 245.

Hřibová, E., Čížková, J., Christelová, P., Taudien, S., de Langhe, E. and Doležel, J. (2011) The ITS1-5.8S-ITS2 sequence region in the Musaceae: structure, diversity and use in molecular phylogeny. *PLoS One* **6**: e17863. doi: 10.1371/journal.pone.0017863.

James, S.A., O'Kelly, M.J.T., Carter, D.M., Davey, R.P., Van Oudenaarden, A. and Roberts, I.N. (2009) Repetitive sequence variation and dynamics in the ribosomal DNA array of *Saccharomyces cerevisiae* as revealed by whole genome resequencing. *Genome Res* **19**: 626–635.

Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.

Keller, I., Chintauan-Marquier, I.C., Veltsos, P. and Nichols, R.A. (2006) Ribosomal DNA in the grasshopper *Podisma pedestris*: escape from concerted evolution. *Genetics* **174**: 863–874.

Keller, I., Veltsos, P. and Nichols, R.A. (2008) The frequency of rDNA variants within individuals provides evidence of population history and gene flow across a grasshopper hybrid zone. *Evolution* **62**: 833–844.

Koetschan, C., Förster, F., Keller, A., Schleicher, T., Ruderisch, B. and Schwarz, R. (2010) The ITS2 Database III – Sequences and structures for phylogeny. *Nucleic Acids Res* **38** (Database issue): D275–D279.

Leach, C.R., Houben, A., Bruce, F., Pistrick, K., Demidov, D. and Timmis, J.N. (2005) Molecular evidence for transcription of genes on a B chromosome in *Crepis capillaris*. *Genetics* **171**: 269–278.

Lewis, M.S., Cheverud, J.M. and Pikaard, C.S. (2004) Evidence for nucleolus organizer regions as the units of regulation in nucleolar dominance in *Arabidopsis thaliana* interecotype hybrids. *Genetics* **167**: 931–939.

Li, Y., Jiao, L. and Yao, Y.J. (2013) Non-concerted ITS evolution in fungi, as revealed from the important medicinal fungus *Ophiocordyceps sinensis*. *Mol Phylogenet Evol* **68**: 373–379.

Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Lindner, D.L., Carlsen, T., Nilsson, R.H., Davey, M., Schumacher, T. and Kauserud, H. (2013) Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecol Evol* **3**: 1751–1764.

Long, E.O. and David, I.B. (1980) Repeated genes in eukaryotes. *Annu Rev Biochem* **49**: 727–764.

López-León, M.D., Cabrero, J. and Camacho, J.P.M. (1999) Unusually high amount of inactive ribosomal DNA in the grasshopper *Stauroderus scalaris*. *Chromosome Res* **7**: 83–88.

López-León, M.D., Cabrero, J., Dzyubenko, V.V., Bugrov, A.G., Karamysheva, T.V., Rubtsov, N.B. *et al.* (2008) Differences in ribosomal DNA distribution on A and B chromosomes between eastern and western populations of the grasshopper *Eyprepocnemis plorans plorans*. *Cytogenet Genome Res* **121**: 260–265.

Matyášek, R., Renny-Byfield, S., Fulneček, J., Macas, J., Grandbastien, M.A., Nichols, R. *et al.* (2012) Next generation sequencing analysis reveals a relationship between rDNA unit diversity and locus number in *Nicotiana* diploids. *BMC Genomics* **13**: 722.

Mayol, M. and Rosselló, J.A. (2001) Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Mol Phylogenet Evol* **19**: 167–176.

Montiel, E.E., Cabrero, J., Ruiz-Estévez, M., Burke, W.D., Eickbush, T.H., Camacho, J.P.M. *et al.* (2014) Preferential occupancy of R2 retroelements on the B chromosomes of the grasshopper *Eyprepocnemis plorans*. *PLoS ONE* **9**: e91820. doi: 10.1371/journal.pone.0091820.

Musters, W., Boon, K., Van der Sande, C.A., Van Heerikhuizen, H. and Planta, R.J. (1990) Functional analysis of transcribed spacers of yeast ribosomal DNA. *EMBO J* **9**: 3989–3996.

Niklas, N., Pröll, J., Danzer, M., Stabentheiner, S., Hofer, K. and Gabriel, C. (2013) Routine performance and errors of 454 HLA exon sequencing in diagnostics. *BMC Bioinformatics* **14**: 176.

Prokopowich, C.D., Gregory, T.R. and Crease, T.J. (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46**: 48–50.

Reeder, R.H. (1999) Regulation of RNA polymerase I transcription in yeast and vertebrates. *Prog Nucleic Acid Res Mol Biol* **62**: 193–327.

Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: the European molecular biology open software suite. *Trends Genet* **16**: 276–277.

Rufas, J.S., Iturra, P., de Souza, W. and Esponda, P. (1982) Simple silver staining procedure for the localization of nucleolus and nucleolar organizer under light and electron microscopy. *Arch Biol* **93**: 267–274.

Ruiz-Estévez, M., López-León, M.D., Cabrero, J. and Camacho, J.P.M. (2012) B-Chromosome Ribosomal DNA is functional in the grasshopper *Eyprepocnemis plorans*. *PLoS ONE* **7**: e36600. doi: 10.1371/journal.pone.0036600.

Ruiz-Estévez, M., López-León, M.D., Cabrero, J. and Camacho, J.P.M. (2013) Ribosomal DNA is active in different B chromosome variants of the grasshopper *Eyprepocnemis plorans*. *Genetica* **141**: 337–345.

Ruiz-Estévez, M., Badisco, L., Vanden Broeck, J., Perfectti, F., López-León, M.D., Cabrero, J. *et al.* (2014a) B chromosomes showing active ribosomal RNA genes contribute insignificant amounts of rRNA in the grasshopper *Eyprepocnemis plorans*. *Mol Genet Genomics* **289:** 1209–1216.

Ruiz-Estévez, M., Cabrero, J., Camacho, J.P.M. and López-León, M.D. (2014b) B chromosomes in the grasshopper Eyprepocnemis plorans are present in all body parts analyzed and show extensive variation for rDNA copy number. *Cytogenet Genome Res* **143**: 268–274.

Ruiz-Ruano, F.J., Ruiz-Estévez, M., Rodríguez-Pérez, J., López-Pino, J.L., Cabrero, J. and Camacho, J.P.M. (2011) DNA amount of X and B chromosomes in the grasshoppers *Eyprepocnemis plorans* and *Locusta migratoria*. *Cytogenet Genome Res* **134**: 120–126.

Simon, U.K. and Weiss, M. (2008) Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Mol Biol Evol* **25**: 2251–2254.

Stage, D.E. and Eickbush, T.H. (2007) Sequence variation within the rRNA gene loci of 12 *Drosophila* species. *Genome Res* **17**: 1888–1897.

StatSoft, Inc. (2007) STATISTICA (data analysis software system), version 8.0. http://www.statsoft.com/support/blog/entryid/6/statsoft-announces-version-8-of-statistica. Accessed on 31 March 2007.

Stewart, F.J. and Cavanaugh, C.M. (2007) Intragenomic variation and evolution of the internal transcribed spacer of the rRNA operon in bacteria. *J Mol Evol* **65**: 44–67.

Teacher, A.G.F. and Griffiths, D.J. (2011) HapStar: automated haplotype network layout and visualisation. *Mol Ecol Res* **11**: 151–153.

Teruel, M., Cabrero, J., Perfectti, F. and Camacho, J.P.M. (2007) Nucleolus size variation during meiosis and NOR activity of a B chromosome in the grasshopper *Eyprepocnemis plorans*. *Chromosome Res* **15**: 755–765.

Teruel, M., Cabrero, J., Perfectti, F. and Camacho, J.P.M. (2009) Quantitative analysis of NOR expression in a B chromosome of the grasshopper *Eyprepocnemis plorans*. *Chromosoma* **118**: 291–301.

Teruel, M., Ruiz-Ruano, F.J., Marchal, J.A., Sánchez, A., Cabrero, J., Camacho, J.P.M. *et al*. (2014) Disparate molecular evolution of two types of repetitive DNAs in the genome of the grasshopper *Eyprepocnemis plorans*. *Heredity* **112**: 531–542.

Trifonov, V.A., Dementyeva, P.V., Larkin, D.M., O'Brien, P.C.M., Perelman, P.L., Yang, F. *et al*. (2013) Transcription of a protein-coding gene on B chromosomes of the Siberian roe deer (*Capreolus pygargus*). *BMC Biol* **11**: 90.

Valente, G.T., Conte, M.A., Fantinatti, B.E.A., Cabral-de-Mello, D.C., Carvalho, R.F., Vicari, M.R. *et al*. (2014) Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Mol Biol Evol* **31**: 2061–2072.

Van der Sande, C.A., Kwa, M., van Nues, R.W., van Heerikhuizen, H., Raué, H.A. and Planta, R.J. (1992) Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. *J Mol Biol* **223**: 899–910.

van Vugt, J.F.A., Salverda, M., de Jong, H. and Stouthamer, R. (2003) The paternal sex ratio chromosome in the parasitic wasp *Trichogramma kaykai* condenses the paternal chromosomes into a dense chromatin mass. *Genome* **46**: 580–587.

Wörheide, G., Nichols, S.A. and Goldberg, J. (2004) Intragenomic variation of the rDNA internal transcribed spacers in sponges (Phylum Porifera): implications for phylogenetic studies. *Mol Phylogenet Evol* **33**: 816–830.

Zhou, J., Eickbush, M.T. and Eickbush, T.H. (2013) A population genetic model for the maintenance of R2 retrotransposons in rRNA gene loci. *PLoS Genet* **9**: e1003179. doi: 10.1371/journal.pgen.1003179.

Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W. and Wilson, A.C. (1980) Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. *Proc Natl Acad Sci USA* **77**: 2158–2162.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Primary spermatocytes of the grasshopper *Eyprepocnemis plorans*, at diplotene, submitted to silver impregnation and showing the absence (a) and presence (b) of a nucleolus attached to the B chromosome. Arrows point to nucleoli. Bar = 5 μ.

Table S1. Sequence of the six tags used for the ITS2 primers. Note that they differ in four or more nucleotides.

Table S2. Tag combinations for the primers used to amplify each sample (F: forward; R: reverse). Id= Individual, Bs= Number of B chromosomes.