The FASEB Journal article fj.08-123513. Published online January 13, 2009.

The FASEB Journal • Research Communication

Genomic environment influences the dynamics of the *tirant* LTR retrotransposon in *Drosophila*

Marie Fablet,* Emmanuelle Lerat,* Rita Rebollo,* Béatrice Horard,† Nelly Burlet,* Sonia Martinez,* Émilie Brasset,‡ Eric Gilson,† Chantal Vaury,‡ and Cristina Vieira*,[1]

*CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, Villeurbanne, France; †CNRS UMR 5161/ENS Lyon, Faculté de médecine Lyon Sud, Laboratoire de Biologie Moléculaire de la Cellule, Université Lyon 1/HCL, Pierre Bénite, France; and ‡UMR/CNRS 6247, INSERM, U931, Faculté de Médecine, BP38, Clermont Université, Clermont-Ferrand, France

ABSTRACT **Combining genome sequence analysis and functional analysis, we show that some full-length copies of *tirant* are present in heterochromatic regions in *Drosophila simulans* and that when tested *in vitro*, these copies have a functional promoter. However, when inserted in heterochromatic regions, *tirant* copies are inactive *in vivo*, and only transcription of euchromatic copies can be detected. Thus, our data indicate that the localization of the element is a hallmark of its activity *in vivo* and raise the question of genomic invasions by transposable elements and the importance of their genomic integration sites.—Fablet, M., Lerat, E., Rebollo, R., Horard, B., Burlet, N., Martinez, S., Brasset, E., Gilson, E., Vaury, C., Vieira, C. Genomic environment influences the dynamics of the *tirant* LTR retrotransposon in *Drosophila*. *FASEB J.* 23, 000–000 (2009)**

Key Words: chromatin · endogenous retrovirus · natural populations

Transposable elements (TEs), which are DNA sequences that can move and multiply along the chromosomes, are now considered to be full-fledged components of genomes, able to play various and numerous functional roles (1, 2). However, their dynamics within a given genome and natural populations is far from being fully understood (3, 4), since TE amount and genomic distribution vary considerably between different species and populations. In most organisms, a high proportion of the heterochromatic genomic compartment is composed of TEs, which are usually organized in clusters (5, 6) and mainly correspond to deleted elements. However, an increasing amount of data shows that heterochromatin also harbors complete copies of TEs, which may constitute a "reservoir" of new elements that may be reactivated (7–9) and thus are potentially able to invade gene-rich regions like the euchromatin, which may have a considerable evolutionary impact. This idea is supported by data showing that the distribution of TEs between euchromatin and heterochromatin is variable and depends on the populations or strains analyzed (10–13). However, we still do not have a refined analysis of the structure and activity of TEs located in different chromatin conformations within a given genome and between genomes. From this perspective, *tirant*, an endogenous retrovirus from *Drosophila* belonging to the *gypsy*-like long terminal repeat (LTR) retrotransposon subclass, is an interesting model, since its genomic copy number varies considerably in different natural populations of *Drosophila simulans*, ranging from 0 euchromatic insertion in most worldwide populations, to 2 to 5 in East African populations (10, 13). A previous study of the *tirant* regulatory region in natural populations of *Drosophila melanogaster* and *D. simulans* revealed two subfamilies of *tirant* regarding the 5′ LTR-untranslated region (UTR) (14). One subfamily, called C type, corresponds to the euchromatic insertions in African populations, and is found in the heterochromatin of all populations worldwide. This C type has been shown to be expressed when located in euchromatin. The other subfamily, called S type, is found in all populations, at very low copy numbers (14), and despite its high levels of sequence conservation between populations, is not found to be actively transcribed.

In an attempt to understand the dynamics of *tirant* and the influence of genomic localization on its activity, we relate data on the *in vitro* expression of *tirant* and on genome walking and sequence analysis. We show that the chromatin environment is a hallmark for expression and that heterochromatin harbors full-length elements that have invasive features. We tested whether the different 5′ LTR-UTR variants were able to promote expression using the reporter gene technique and correlated this to the genomic location of each *tirant* insertion and local chromatin structure. We then identified the genomic context of the insertions in the *D. simulans* sequenced genome, and described the copies of *tirant*.

MATERIALS AND METHODS

Drosophila natural populations

We worked on fly samples collected from several geographically distinct natural populations of *D. melanogaster* and *D. simulans*. A list of the populations analyzed is provided in Supplemental

[1] Correspondence: Université de Lyon, F-69000, Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France. E-mail: vieira@biomserv.univ-lyon1.fr

doi: 10.1096/fj.08-123513

Table 1. These populations were maintained in the laboratory at 17°C as isofemale lines or small mass cultures with around 50 pairs in each generation.

Reporter gene assays

Constructions

We used 3 variants of the *tirant* 5′ LTR-UTR region previously amplified by polymerase chain reaction (PCR) from natural population samples (14) (GenBank accession numbers: AY756122, AY756118, and AY756121; see **Fig. 1** for a detailed diagram). We refer to these sequences as S/C-*i,j*, where S or C is the type, *i* is the number of repeats of a 19-bp motif in the LTR, and *j* is the number of repeats of a 102-bp motif in the 5′ UTR: S-1,5 (AY756122), C-2,2 (AY756121), and C-2,4 (AY756118) (Fig. 1). The 5′ LTR-UTR regions were cloned separately upstream of a *lacZ* reporter gene into the *Sph*I-*Pst*I (for S-1,5) and *Sph*I-*Xba*I (for C-2,2 and C-2,4) sites of the *pPelican* plasmid (15). Plasmids were amplified in TOP 10 competent cells from Invitrogen (Carlsbad, CA, USA) and purified with Plasmid Mini/Midi kits from Qiagen (Courtaboeuf, France) (final elution with water). The presence of the 5′ LTR-UTR regions in the plasmids were confirmed by PCR and restriction profiles.

Transfection into S2 cells

Transfections were done using Celfectin (Invitrogen). Cells were then incubated for 72 h at 22°C. S2 cells were cotransfected with *tirant* constructs and the *pGL3-Control* vector (Promega, Madison, WI, USA), consisting of a luciferase gene downstream of an SV40 promoter. β-Galactosidase activity was normalized *vs.* luciferase activity.

The ratios of the reported β-galactosidase activity to the luciferase activity were compared for the various constructions. After cell lysis, the β-galactosidase and luciferase activities were measured using the β-gal reporter gene assay, chemiluminescence kit (Roche Diagnostics, Mannheim, Germany), and the Luciferase Assay System kit (Promega), respectively.

Each transfection was repeated six times. Three wells were transfected with *pPelican* alone, as a negative control, and three others were transfected with a vector carrying a *lacZ* gene downstream of an SV40 promoter, as a positive control.

Genome-walking analysis

We determined the insertion site sequences for each copy of *tirant* with the Universal Genome Walker kit (Clontech, Mountain View, CA, USA) in three genomes from natural populations of *D. simulans*: the Makindu population from Kenya (5 euchromatic copies), the Chicharo population from Portugal (0 euchromatic copies), and the Zimbabwe population (2 euchromatic copies). Genomic DNA was extracted from 20 female adults from each population. For each population, 4 genomic libraries were obtained using the restriction enzymes supplied by the provider: *Dra*I, *Eco*RV, *Pvu*II, and *Stu*I, with 9, 5, 0, and 2 restriction sites, respectively, along the *tirant* reference sequence. Adaptors were ligated according to the supplier's instructions, and two nested PCRs were subsequently done, with primers specific to *tirant* LTR and to the adaptors. The external *tirant*-specific primer can amplify both C and S types: 5′ GTT TAG AGG CGT GGG GGT TTA GAA TC 3′. The internal *tirant*-specific primers specifically used for the C and S types are 5′ TGT AAG CAT AAT GAA CAT GCC GAC TC 3′ and 5′ TGT AAA CAT AAT TTC CAT GCC ACT TC 3′, respectively. PCRs were done in 2 steps using the Advantage Genomic PCR kit (Clontech).

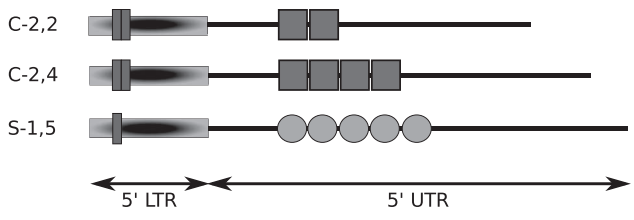

Figure 1. Structure of the variants of the 5′LTR-UTR region tested. Rectangles and ovals stand for the different tandem repeats. Variants are named as follows: S/C-*i,j*, where S or C is the subfamily, *i* is the number of 19-bp tandem repeats in the LTR, and *j* is the number of 102-bp repeats in the 5′UTR.

PCR amplification of *tirant* entire copies

To amplify each copy of *tirant*, we designed specific primers from the flanking sequences of each insertion (Supplemental Table 2). We realized 3-step PCRs with the Expand Long Template PCR System (Roche) with an annealing temperature of 60°C and the System 1 buffer, provided by the supplier.

Populational screening of *tirant* insertions

For some of the insertion sites determined by genome walking in the Makindu population that corresponded to potentially full-length copies of *tirant*, we designed a pair of primers with the forward primer in the *env* gene, and the reverse primer in the 3′ flanking region. We looked for the presence of PCR products in a pool of 3 flies for each natural population, which would indicate that the corresponding insertion of *tirant* was shared by several populations. The Makindu population was systematically used as a positive control for the PCR assays. The screened populations and the list of the primers are presented in Supplemental Table 3. We used the EuroBlueTaq enzyme from Eurobio (Les Ullis, France). The PCR was a 3-step reaction run in 30 cycles with an annealing temperature of 57°C. Amplified products were migrated on a 1% agarose gel.

Local chromatin structure analysis

Chromatin immunoprecipitation (ChIP)

Extraction of chromatin from 16-h *D. simulans* Makindu embryos and immunoprecipitation (IP) were adapted from Sandmann *et al.* (16). Cell lysis buffer was changed to 5 mM PIPES pH 8, 85 mM KCl, 0.5% Nonidet P-40 supplemented with protease inhibitors. Chromatin was shared with a Bioruptor sonicator water bath (Diagenode, Liège, Belgium) for 6 × 30-s on/30-s off cycles at high power in order to have random fragments from 1 kb to 500 bp. Shared chromatin was incubated overnight at 4°C with antibodies recognizing H3K9me2 (Millipore, Billerica, MA, USA; 07441), H3K27me3 (Millipore 07449), H3K4me2 (Millipore 07030), H3 (Abcam, Cambridge, UK; ab1791), and rabbit IgG (Sigma-Aldrich, St. Louis, MO, USA; I5006). The antigen-antibody complexes were washed as described before (16), but a second washing solution was modified as followed: TE 2×, 500 mM NaCl, 1% Triton, 0.1% SDS.

Real-time PCR

To quantify each IP, real-time PCR was performed using QuantiTect SYBR Green PCR kit (Qiagen) on a MXP3000P PCR system (Stratagene, La Jolla, CA, USA). Reactions were done in duplicates, and standard curves were calculated on serial of input chromatin. To evaluate the relative enrichment of C and S copies after IP, we calculated the difference in cycles between the IP-enriched sample and the input DNA for *tirant* copies and for a control (*actin*-CG4027). C 2 and S 5 copies (see **Table 1,** Makindu population) were amplified using primers in the flanking region and in the *tirant* LTR (C 2 specific primers: forward, 5′ GTG TTC CAG TTG CCG TCT TC 3′; reverse, 5′ TTT TCT GGG GTG TTG TAC GC 3′; S 5 specific primers: forward, 5′ TGC TCT CAA CTG CGC GCG AGT TAC 3′; reverse, 5′ GTA AA ATA ATT TCC ATG CCA CTT C 3′).

Analysis of *tirant* copies from the *D. simulans* sequenced genome

We retrieved the sequences of the chromosome arms 2L, 2R, 3L, 3R, 4, X, and the unassigned part (named U), of the first release of the mosaic assembly of the genome of *D. simulans* available at the ftp site of the Genome Sequencing Center at the Washington University Medical School (http://hgdownload.cse.ucsc.edu/downloads.html#droSim). This mosaic assembly corresponds to different *D. simulans* strains. In the following, we refer to the *tirant* copies using the chromosome name and the start position of the copy (*e.g.*, chr2L_21040234 corresponds to a copy found on chromosome 2L and that starts at position 21040234).

To search for any kind of copy of *tirant* of the two variant types in the *D. simulans* sequenced genome, we used the sequences of the S type (GenBank AY756122) and the C type (GenBank AY756123) cloned by Fablet *et al.* (14). As these sequences corresponded only to the 5′ LTR-UTR and 5′ region of the first open reading frame (ORF) *gag*, we also used the complete sequenced *tirant* from *D. melanogaster* (GenBank AY928610) (17), which is a C type, to identify sequences corresponding to internal parts and to potentially complete copies of *tirant* in *D. simulans*. With all these sequences, we searched for *tirant* copies in the *D. simulans* sequenced genome using Blastn (18). Matches with an e_value $< 10^{-10}$ were retained, and those with distances $<$ 300 bp were merged. The detected copies were compared pairwise with the three query sequences using the program matcher (EMBOSS website; http://emboss.bioinformatics.nl/cgi-bin/emboss/matcher), and the corresponding percentage identity was computed with the dnadist module from the PHYLIP package (19). The detection of the ORFs was performed using the ORF Finder program at the National Center for Biotechnology Information website (www.ncbi.nlm.nih.gov/projects/gorf/).

RESULTS

Reporter gene assays

Three types of *tirant* 5′ LTR-UTR regions were tested for the promotion of *lacZ* expression in S2 cells: S-1,5, which is an S type, and C-2,2 and C-2,4, which are C types with two and four 102-bp motifs in the 5′ UTR, respectively (see Fig. 1). Note that the number of repeats of the 102-bp motif is always 5 for all S types previously identified in natural populations (14). As shown in **Fig. 2**, the activity ratios for S-1,5, C-2,2, and C-2,4 were significantly different from both the positive and negative controls (see Materials and Methods for

TABLE 1. *Tirant insertion sites determined by genome walking*

Insertion	Type	Location	Comments	Length
Makindu population				
1	C	X	Unannotated DNA	Full length
2	C	2L	*tkv* gene intron	Full length
3	C	3R	Unannotated DNA	Full length
4	C	Centromeric DNA	Maupiti islands	Full length
5	S		*R1* non-LTR retrotransposon	Full length
6	C	2L	*hobo* transposon	Short fragment
7	C	2L	*frogger* LTR retrotransposon	Short fragment
8	S	2R	Heterochromatin	Short fragment
9	C		*MAX* LTR retrotransposon	Short fragment
10	C	3R	Unannotated DNA	NA
Zimbabwe population				
1	S		*R1* non-LTR retrotransposon	Full length
2	C	3	Heterochromatin	Long fragment
3	C	2L	3′ from CG13786 gene	Short fragment
4	C	2L	Heterochromatin	Short fragment
5	C	3R	Unannotated DNA	Short fragment
6	C	Centromeric DNA	Maupiti islands	Short fragment
7	C	3	Heterochromatin	NA
Chicharo population				
1	C	Centromeric DNA	Maupiti islands	Full length
2	C		Heterochromatin	Long fragment
3	C		*diver2* LTR retrotransposon	Short fragment
4	S		*Rt1b* non-LTR retrotransposon	NA

It is not possible to determine the position on the chromosomes of some sites, especially those corresponding to insertions into other transposable elements. NA, electrophoresis of the PCR product shows multiple bands, possibly indicating embedded insertions of *tirant*.

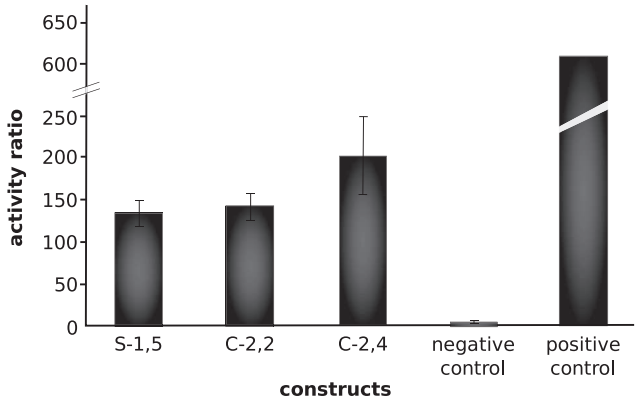

Figure 2. Reporter gene assays, transfection into S2 cells results. Activity ratio: ratio of the β-galactosidase and luciferase activities measured. Negative control corresponds to transfection of an empty *pPelican* vector, and positive control to transfection of a vector carrying the *lacZ* gene downstream of an SV40 promoter. No SD could be calculated for the positive control because 2 of 3 measurements were out of range for the machine used.

details). One-tailed pairwise *t* tests revealed a significant difference between C-2,2 and C-2,4 (P=0.012), and between S-1,5 and C-2,4 (P=0.006), whereas C-2,2 and S-1,5 were not significantly different (P=0.216). This shows that both the C and S types are able to promote the expression of the reporter gene and are therefore theoretically able to promote *tirant* expression in an endogenous context.

Analysis of the genomic context of each copy of *tirant*

The reporter gene assay results cannot be directly extrapolated to the endogenous copies of *tirant* without taking into account the genomic localization of each copy (*i.e.*, in euchromatin or heterochromatin), which could influence transcription. We therefore determined the genomic localization of each insertion of *tirant* by the genome-walking technique.

The genome-walking protocol allowed us to identify 10, 7, and 4 *tirant* insertion sites in the genomes of the Makindu, Zimbabwe, and Chicharo populations, respectively. The results are presented in Table 1. The number of insertions was slightly higher than that obtained by *in situ* hybridization (10). This is due to the following facts: *1*) only euchromatic copies on the chromosome arms were detected by *in situ* hybridization, *2*) full-length copies as well as solo LTRs could be detected by genome walking, whereas the detection size threshold for *in situ* hybridization excluded that for solo LTRs, and *3*) the genome walking was done on DNA extracted from a pool of 20 individuals, thus revealing the sum of the insertions resulting from individual variability, whereas insertion sites were estimated for single individuals by the *in situ* hybridization technique.

Since the sequenced genome of *D. simulans* was very poorly annotated at the time of the study, we used the annotations of the corresponding regions in the *D. melanogaster* sequenced genome. The only shared site among the genomes of the 3 natural populations analyzed in the present study corresponded to an insertion into the Maupiti Islands centromeric DNA, which is an A + T-rich region displaying many insertions of TEs, mainly non-LTR retrotransposons, which showed a high degree of similarity with the *F* and *doc* elements (9). Of the other 18 insertions of *tirant* identified in the 3 genomes, 7 were found within other TEs, such as *hobo* (20), *frogger* (21), *MAX* (21), *diver2* (22), *R1* (23), and *Rt1b* (24). Five of the 18 insertions were heterochromatic, and the remaining 6 were found in unannotated DNA which probably corresponds to noncoding DNA. With the data of the exact insertion site for each copy of *tirant*, we were able to design primers specific to the flanking region of each individual copy of *tirant*, in order to amplify it by long PCR in the different populations. As expected from previous Southern blot results (14), we obtained full-length amplicons, as well as shorter ones, for each of the 3 populations tested. The detailed results and the correspondence with insertion sites are shown in Table 1. Five of the insertions in the Makindu population were of the expected size for a full-length *tirant*, *i.e.*, 8.5 kb. In the other cases, the amplicons corresponding to short fragments were assumed to be solo LTRs (the amplicons were sequenced for insertions Makindu 7 and Zimbabwe 4). We used a PCR test to check whether the potentially full-length insertions were present in other populations than Makindu, and we found that these insertions were specific to the Makindu population and were not found in the other populations.

To determine whether local chromatin environment could vary between C and S copies, we identified histone modification marks on two different *tirant* full-length copies: C 2 and S 5 (see Table 1). ChIP followed by copy-specific real-time PCR analysis was performed using antisera against permissive (H3K4me2) and repressive (H3K9me2 and H3K27me3) histone marks (25). We observed that the S-type copy of *tirant* is strongly associated with the repressive histone marks H3K9me2 and H3K27me3. In contrast, the C-type copy of *tirant* is characterized by the coenrichment of the repressive mark H3K27me3 and the permissive H3K4me2 (**Fig. 3**). These observations suggest that the only full-length S-type copy present in the Makindu genome is embedded in a repressive chromatin environment, while the C-type copy can be in a bivalent chromatin domain.

D. simulans sequenced genome analysis

Among the 20 copies of *tirant* identified in the *D. simulans* sequenced genome, 5 are of the S type and 15 of the C type (**Table 2** and **Fig. 4*A*, *B***). For each type, we found one almost-complete sequence. The full-length S-type copy U_8880559 is located in the U part of the genome and corresponds to a sequence of 8508 bp. It is flanked by two LTRs of 427 bp sharing 99.30% identity, which indicates that this is a recent insertion. Two ORFs correspond to potentially complete *gag* (positions 2028 to 2978) and *env* (positions 6847 to 8118) genes. The *pol* ORF (positions 3218 to 6455) displays some frameshifts that result in inframe stop

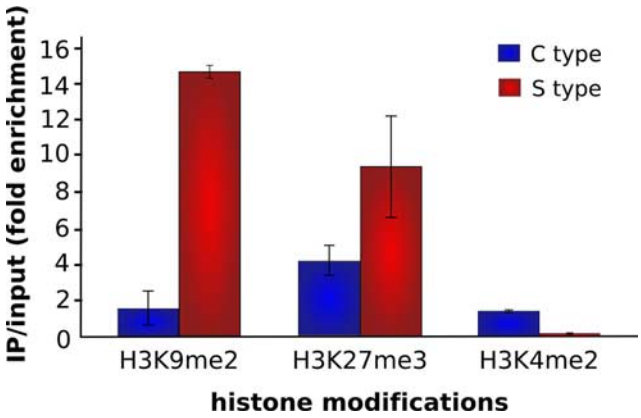

Figure 3. Histone marks association with full-length C- and S-type copies of *tirant*. ChIP analysis of C- and S-type copies of *tirant* in *D. simulans* embryos from the population of Makindu. See Materials and Methods for fold enrichment computation. H3K9me2, H3K27me3, and H3K4me2 are epigenetic marks characteristic of constitutive heterochromatin, nonconstitutive heterochromatin, and euchromatin, respectively.

codons. The full-length C-type copy U_10384153 is also located in the U part of the genome, and corresponds to a sequence of 6803 bp, surrounded by undetermined bases, which made it impossible for us to determine the presence of LTRs at the extremities. Two ORFs correspond to potentially complete *gag* (positions 911 to 2044) and *env* (positions 5693 to 6801) genes; the *pol* ORF (positions 2284 to 5515) displays a large internal deletion that creates a frameshift.

Comparison of the cloned flanking regions with those of the copies identified *in silico* showed that the 2L-random-887610 copy (Table 1; Fig. *4B*) is present in both the Makindu and Zimbabwe populations (insertions Makindu 7 and Zimbabwe 4). The sequencing of PCR products from populations shows that this insertion is a solo LTR. However, it is not possible to conclude about the corresponding insertion found in the sequenced genome, since it is flanked downstream by undetermined bases. The other insertions detected in the 3 populations correspond to unique insertions that are not present in the sequenced genome of *D. simulans*.

DISCUSSION

The S and C types of *tirant*, previously characterized in *D. melanogaster* and *D. simulans* (14), appear to behave quite differently in the genome of *D. simulans*. It had previously been shown that the C type was polymorphic, presenting varying numbers of a 102-bp motif in its UTR, whereas the S type was monomorphic, with always 5 repeats of this motif in its UTR. In addition, the S-type copy number was low and homogeneous among *D. simulans* natural populations, while the C-type copy number was high in East African populations, but quite low in the surrounding populations (14).

Tirant activity

Reporter gene assays in S2 cells suggest that both the S and C types are able to promote the expression of the reporter gene, with significant differences in the expression strength between the C-type sequences tested. However, as revealed by previous RT-PCR experiments on gonads, only the C-type sequences are expressed when located in the euchromatin (14). The S type, which was localized in the heterochromatin, was not

TABLE 2. *Positions of the* tirant *copies in the D. simulans sequenced genome*

						Percentage identity			
Chromosome	Strand	Start	Stop	Type	Length (bp)	*tirant* Dm AY928610	C type AY756123	S type AY756122	chrU_8880559
U	−	13736532	13737472	S	942	79.10	72.97	88.03	94.58
U	−	5979828	5980937	S	1110	85.57	NA	NA	99.28
U	−	8880559	8889066	S	8508	80.06	73.35	99.47	/
U	+	3039715	3042707	S	2993	78.43	NA	NA	98.47
U	−	2603202	2604990	S	1791	76.16	NA	NA	99.44
U	+	1610739	1611425	C	687	89.70	89.10	58.48	60.00
3L	−	22549046	22549538	C	493	85.39	NA	NA	82.62
U	−	1396504	1396898	C	422	96.68	99.26	89.01	88.99
U	+	28708	29409	C	702	94.14	NA	NA	86.55
U	+	9615110	9615888	C	780	93.93	NA	NA	85.19
U	+	10384153	10390953	C	6803	98.63	99.26	78.16	80.81
U	−	15708778	15709918	C	1142	95.97	NA	NA	74.82
U	+	3501254	3503485	C	2232	97.17	NA	NA	75.21
U	+	6605489	6606440	C	952	92.30	94.11	79.87	79.38
U	+	7436025	7437064	C	1040	95.76	NA	NA	82.19
2L-random	+	887610	888048	C	440	90.38	90.76	86.11	87.78
2L	+	21040234	21040766	C	533	83.86	NA	NA	74.56
2L	+	21042156	21045168	C	2845	82.96	67.84	65.90	80.52
2L	+	21052705	21053256	C	552	86.79	NA	NA	84.04
2L	+	21205201	21205440	C	240	91.25	NA	NA	86.67

NA, value cannot be computed.

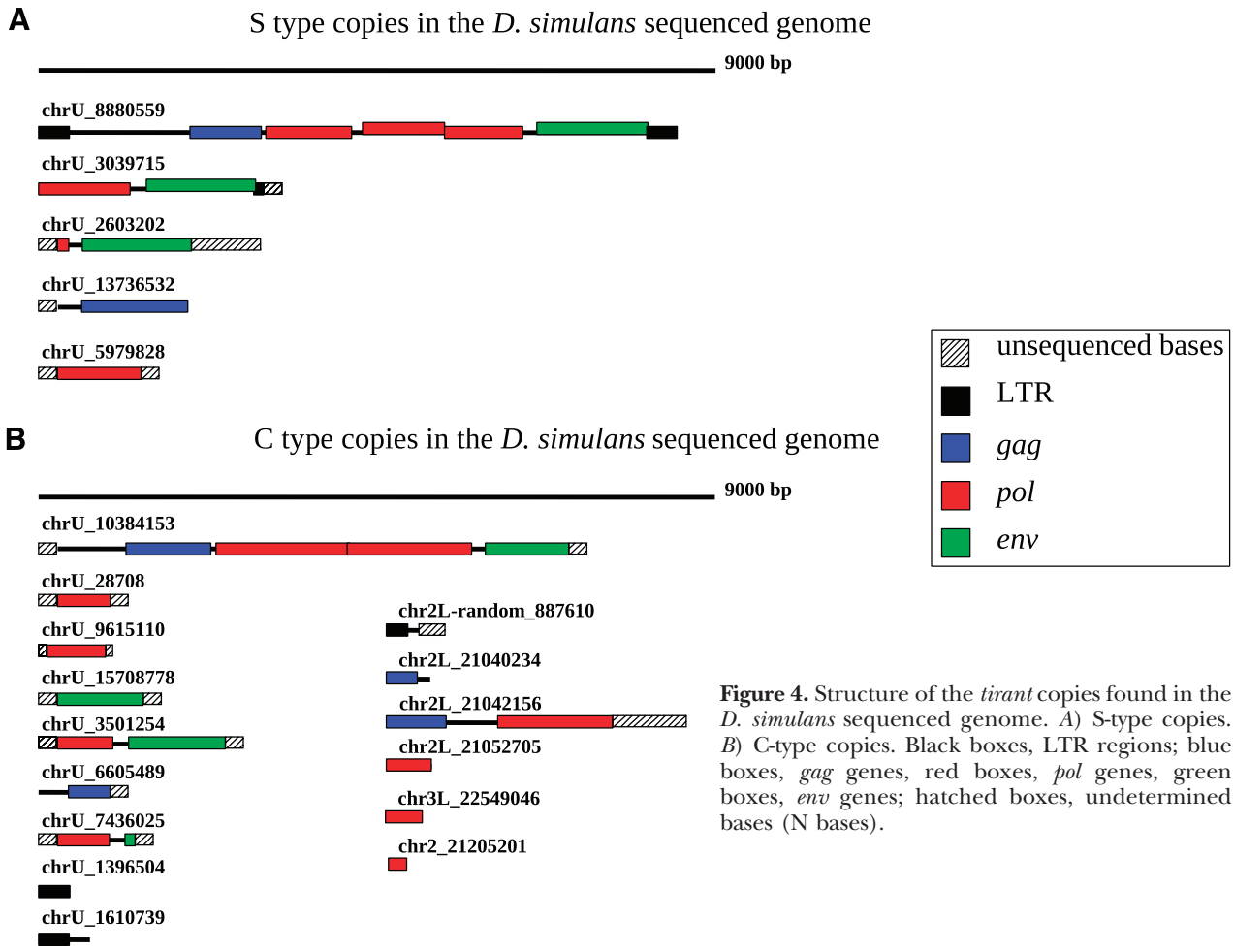

Figure 4. Structure of the *tirant* copies found in the *D. simulans* sequenced genome. *A*) S-type copies. *B*) C-type copies. Black boxes, LTR regions; blue boxes, *gag* genes, red boxes, *pol* genes, green boxes, *env* genes; hatched boxes, undetermined bases (N bases).

expressed in any of the natural populations analyzed (14), but we show in this study, by the *in vitro* experiments, that it is potentially able to drive expression. Two hypotheses can be proposed to explain this discrepancy: either the S type is inhibited in its endogenous natural context (*e.g.*, by a repressive genomic environment, such as heterochromatin), or it is able to promote expression in a *D. melanogaster* genetic context (*e.g.*, S2 cells), but is repressed in a *D. simulans* genetic context. An alternative hypothesis would be that *tirant* expression is allowed in S2 cells, which are derived from embryonic cells, whereas it could be specifically repressed in some tissues, such as ovaries.

The observed situation for *tirant* in *D. simulans* is comparable to cases reported in *D. melanogaster*, in which one line displays genetic instability—also known as the permissive line—with an elevated copy number and a high rate of activity for a particular TE, such as the endogenous retroviruses *gypsy* (26), *ZAM*, and *Idefix* (27). Stable lines—also known as restrictive lines—are not totally devoid of these elements but have few or no copies in the chromosome arms (28, 29). The Makindu population, which has full-length copies of *tirant* in the euchromatin, behaves like a permissive line in which the control of *tirant* is partly or totally impaired, which would explain its activity and relative high copy number in this population. In contrast, the other populations analyzed in this study only have a few inactive heterochromatic copies of *tirant*.

Tirant copies genomic contexts

S type

The genome-walking protocol showed the presence of one or two S-type insertions per genome (*i.e.*, per population), which is consistent with what was previously found by Southern blot analysis (14). One of these insertions is heterochromatic (Makindu 8), and the other three are located in copies of either the *R1* non-LTR retrotransposon (Makindu 5 and Zimbabwe 1) or the *Rt1b* non-LTR retrotransposon (Chicharo 4). *Rt1b*, also known as *Waldo-A*, belongs to the *R1* clade (24). We cannot precisely locate the insertions of these elements, but we know that a high proportion of *R1* copies is located in centromeric heterochromatin (23), or in ribosomal genes that are often inactivated by local heterochromatin formation (30). In this regard, we found that the unique full-length S-type copy of *tirant* in the Makindu genome is associated with H3K9me2 and H3K27me3, two specific repressive histone marks (25). This confirms the assumption that S-type copies of *tirant* are inserted into inactivated genomic regions, mainly heterochromatin, which could explain the ab-

sence of S-type transcripts in endogenous conditions revealed by RT-PCR, even though the S-type 5′ LTR-UTR region was shown to drive expression in the S2 cells.

In the sequenced genome of *D. simulans*, 5 S-type copies were detected, only one of which corresponded to a full-length copy. The shorter ones displayed high-percentage identities with the full-length S-type copy, except the U_13736243 copy (94.58% identity, Table 2), which is the shortest sequence for this type. Moreover, the shorter copies were surrounded by undetermined bases, indicating that they could be longer, and that we are probably underestimating the copy number of the S type. We were unable to identify the exact location of the S-type insertions, since the *D. simulans* sequenced genome was not annotated yet.

C type

Unlike the S type, which appears to be exclusively heterochromatic, insertions of the C type are both euchromatic and heterochromatic. In the Chicharo genome, in which no *tirant* copy can be detected using *in situ* hybridization, 3 C-type copies were found when analyzing the genomic sequence, one in the heterochromatin, another in centromeric DNA (Maupiti Islands), which is also heterochromatic, and the last one in the LTR retrotransposon *diver2*. *Diver2* insertions in the *D. melanogaster* genome are either centromeric or telomeric (22) and so correspond to heterochromatin. Therefore, all *tirant* insertions found by genome walking in the Chicharo genome can be assumed to be heterochromatic.

The *tirant* C-type insertion 7 from Zimbabwe is of particular interest since its flanking region displays blast matches with the heterochromatic gene *parp* (31). Tulin *et al.* (31) showed that the genomic region where *parp* is located in *D. melanogaster* is rich in TEs, especially *gypsy* elements that have lost their LTRs and insulators. These authors also showed that a mutation in the regulatory region of the gene *parp* deregulates the LTR-retrotransposon *copia*, the transcription level of which increases 50-fold. The idea that the insertion of *tirant* near *parp* in the Zimbabwe genome could regulate other TEs, or promote changes in chromatin structure warrants further investigation. While it was previously assumed from Southern blot data that the Zimbabwe genome had no full-length copy of *tirant*, the long PCR experiments suggest that insertion 1 is potentially full length. This discrepancy could be explained by the presence of restriction polymorphism in this sequence.

In the Makindu genome, the size of PCR products obtained for half of the detected insertions indicates that the copies are (potentially) full length. Insertion sites are varied for these copies: intron of a gene, Maupiti Islands, *R1* transposable element, and unannotated DNA. Interestingly, we observed that one of these full-length C-type *tirant* copies, insertion C 2, harbors simultaneously repressive H3K27me3 and permissive H3K4me2 histone marks. In the past few years, several studies in mouse and human ES cells reported the occurrence of such "bivalent domains" predominantly on key developmental regulators (reviewed in Pietersen and van Lohuizen; ref. 32). It was proposed that this unusual combination of marks keeps genes repressed or expressed at very low level but poised for later activation or complete repression (32). In this regard, we can propose that the Makindu genome contains C-type copies of *tirant* susceptible to be expressed in response to appropriate developmental or cellular cues. Full-length insertions are thus not associated specifically with euchromatin but are also present in heterochromatic regions.

Internal dynamics of *tirant* in the *D. simulans* genome: the cohabitation of two subfamilies

Subfamilies with differing regulatory regions have been reported for various TEs, such as *copia* (33), *blood* (34), and *412* (35) in *Drosophila*, and *Tnt1* in plants (36). In some cases, such as *copia*, subfamilies are associated with different levels of expression (4, 32). The situation observed for *blood* in the *D. melanogaster* subgroup of species (34), is very similar to what we found for *tirant*. One of the *blood* subfamilies, long (L), is mainly heterochromatic, and its insertion sites are shared by most populations. The other subfamily, short (S), is potentially active and euchromatic. One particular population of *D. simulans* has more *blood* insertions on the chromosome arms than the other populations of the species (13), which is interpreted as an invasion of euchromatin by the S subfamily (34). The authors, therefore, assume the existence of competition between the subfamilies, leading to the elimination from euchromatin of the L subfamily, which is replaced by the S subfamily. This situation is quite similar to what we observe for *tirant*, with the S type restricted to the heterochromatin and untranscribed, and the C type that has invaded euchromatin in the East African populations of *D. simulans*.

The dynamics of the *tirant* types is not incompatible with an overall loss of the element from the genome of most *D. simulans* populations worldwide, as was previously proposed (14) but indicates that one population (Makindu), which can be considered as permissive for *tirant*, has been subjected to some deregulation, leading to transposistions, and an increase in copy number. Our results also suggest that the S and C types of *tirant* are subjected to several different mechanisms of regulation, which include chromatin conformation and therefore the epigenetic regulation machinery.

For a long time, heterochromatin has been considered to be the graveyard of TEs, since the absence of recombination would lead to the accumulation of these sequences (37). This model is consistent with data showing that TEs are organized in clusters in the heterochromatin and that most of them are highly rearranged and deleted (6, 38). However, analyses of the *Drosophila* heterochromatic sequences (8, 39) and several experimental studies have shown that full-length elements persist in this part of the genome, and in some cases could be active (7, 9). The identification of full-length S-type copies of *tirant* in the heterochromatin of the *D. simulans* natural populations and sequenced genome, associated with the data obtained for the *in vitro* expression of these type of elements, suggest that the heterochromatin harbors potentially invasive elements, which may be reactivated under particular

conditions that have not yet been determined and that warrant further studies. FJ

We thank Christian Biémont for useful comments, and Monika Ghosh for reviewing the English text. The work in C. Vieira's laboratory was funded by the Centre National de la Recherche Scientifique (UMR 5558 and GDR 2157 on TEs). The work in E.G.'s laboratory was funded by ARC (ARECA program on epigenetic profiling).

REFERENCES

1. Biémont, C., and Vieira, C. (2005) What transposable elements tell us about genome organization and evolution: the case of *Drosophila*. *Cytogenet. Genome Res.* **110,** 25–34
2. Biémont, C., and Vieira, C. (2006) Genetics: junk DNA as an evolutionary force. *Nature* **443,** 521–524
3. Biémont, C., Vieira, C., Hoogland, C., Cizeron, G., Loevenbruck, C., Arnault, C., and Carante, J. P. (1997) Maintenance of transposable element copy number in natural populations of *Drosophila melanogaster* and *D. simulans*. *Genetica* **100,** 161–166
4. Vieira, C., and Biémont, C. (2004) Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *Drosophila simulans*. *Genetica* **120,** 115–123
5. Bergman, C. M., Quesneville, H., Anxolabehere, D., and Ashburner, M. (2006) Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. *Genome Biol.* **7,** R112
6. Vaury, C., Bucheton, A., and Pelisson, A. (1989) The beta heterochromatic sequences flanking the I elements are themselves defective transposable elements. *Chromosoma* **98,** 215–224
7. Kogan, G. L., Tulin, A. V., Aravin, A. A., Abramov, Y. A., Kalmykova, A. I., Maisonhaute, C., and Gvozdev, V. A. (2003) The GATE retrotransposon in *Drosophila melanogaster*: mobility in heterochromatin and aspects of its expression in germline tissues. *Mol. Genet. Genomics.* **269,** 234–242
8. Mugnier, N., Gueguen, L., Vieira, C., and Biémont, C. (2008) The heterochromatic copies of the LTR retrotransposons as a record of the genomic events that have shaped the *Drosophila melanogaster* genome. *Gene* **411,** 87–93
9. Sun, X., Le, H. D., Wahlstrom, J. M., and Karpen, G. H. (2003) Sequence analysis of a functional *Drosophila* centromere. *Genome Res.* **13,** 182–194
10. Biémont, C., Nardon, C., Deceliere, G., Lepetit, D., Loevenbruck, C., and Vieira, C. (2003) Worldwide distribution of transposable element copy number in natural populations of *Drosophila simulans*. *Evolution Int. J. Org. Evolution* **57,** 159–167
11. Charlesworth, B., Jarne, P., and Assimacopoulos, S. (1994) The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. III. Element abundances in heterochromatin. *Genet. Res.* **64,** 183–197
12. Junakovic, N., Terrinoni, A., Di Franco, C., Vieira, C., and Loevenbruck, C. (1998) Accumulation of transposable elements in the heterochromatin and on the Y chromosome of *Drosophila simulans* and *Drosophila melanogaster*. *J. Mol. Evol.* **46,** 661–668
13. Vieira, C., Lepetit, D., Dumont, S., and Biémont, C. (1999) Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol. Biol. Evol.* **16,** 1251–1255
14. Fablet, M., McDonald, J. F., Biémont, C., and Vieira, C. (2006) Ongoing loss of the tirant transposable element in natural populations of *Drosophila simulans*. *Gene* **375,** 54–62
15. Barolo, S., Carver, L. A., and Posakony, J. W. (2000) GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in *Drosophila*. *BioTechniques* **29,** 726, 728, 730, 732
16. Sandmann, T., Jakobsen, J. S., and Furlong, E. E. (2006) ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in *Drosophila melanogaster* embryos. *Nat. Protocols* **1,** 2839–2855
17. Lerat, E., Rizzon, C., and Biémont, C. (2003) Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. *Genome Res.* **13,** 1889–1896
18. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402
19. Felsenstein, J. (1989) PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics* **5,** 3
20. Bonnivard, E., Higuet, D., and Bazin, C. (1997) Characterization of natural populations of *Drosophila melanogaster* with regard to the hobo system: a new hypothesis on the invasion. *Genet. Res.* **69,** 197–208
21. Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D. A., Lewis, S. E., Rubin, G. M., Ashburner, M., and Celniker, S. E. (2002) The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* **3,** RESEARCH0084
22. Kapitonov, V. V., and Jurka, J. (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6569–6574
23. Eickbush, D. G., and Eickbush, T. H. (1995) Vertical transmission of the retrotransposable elements R1 and R2 during the evolution of the *Drosophila melanogaster* species subgroup. *Genetics* **139,** 671–684
24. Busseau, I., Berezikov, E., and Bucheton, A. (2001) Identification of Waldo-A and Waldo-B, two closely related non-LTR retrotransposons in *Drosophila*. *Mol. Biol. Evol.* **18,** 196–205
25. Ebert, A., Lein, S., Schotta, G., and Reuter, G. (2006) Histone modification and the control of heterochromatic gene silencing in *Drosophila*. *Chromosome Res.* **14,** 377–392
26. Kim, A., Terzian, C., Santamaria, P., Pelisson, A., Purd'homme, N., and Bucheton, A. (1994) Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 1285–1289
27. Desset, S., Conte, C., Dimitri, P., Calco, V., Dastugue, B., and Vaury, C. (1999) Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of *Drosophila melanogaster*. *Mol. Biol. Evol.* **16,** 54–66
28. Desset, S., Meignin, C., Dastugue, B., and Vaury, C. (2003) COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics* **164,** 501–509
29. Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995) Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics* **139,** 697–711
30. Strohner, R., Nemeth, A., Nightingale, K. P., Grummt, I., Becker, P. B., and Langst, G. (2004) Recruitment of the nucleolar remodeling complex NoRC establishes ribosomal DNA silencing in chromatin. *Mol. Cell. Biol.* **24,** 1791–1798
31. Tulin, A., Stewart, D., and Spradling, A. C. (2002) The *Drosophila* heterochromatic gene encoding poly(ADP-ribose) polymerase (PARP) is required to modulate chromatin structure during development. *Genes Dev.* **16,** 2108–2119
32. Pietersen, A. M., and van Lohuizen, M. (2008) Stem cell regulation by polycomb repressors: postponing commitment. *Curr. Opin. Cell. Biol.* **20,** 201–207
33. Matyunina, L. V., Jordan, I. K., and McDonald, J. F. (1996) Naturally occurring variation in copia expression is due to both element (cis) and host (trans) regulatory variation. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 7097–7102
34. Costas, J., Valade, E., and Naveira, H. (2001) Amplification and phylogenetic relationships of a subfamily of blood, a retrotransposable element of Drosophila. *J. Mol. Evol.* **52,** 342–350
35. Mugnier, N., Biémont, C., and Vieira, C. (2005) New regulatory regions of *Drosophila* 412 retrotransposable element generated by recombination. *Mol. Biol. Evol.* **22,** 747–757
36. Vernhettes, S., Grandbastien, M. A., and Casacuberta, J. M. (1998) The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. *Mol. Biol. Evol.* **15,** 827–836
37. Charlesworth, B., and Langley, C. H. (1989) The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* **23,** 251–287
38. Daniels, S. B., and Strausbaugh, L. D. (1986) The distribution of P-element sequences in *Drosophila*: the willistoni and saltans species groups. *J. Mol. Evol.* **23,** 138–148
39. Hoskins, R. A., Carlson, J. W., Kennedy, C., Acevedo, D., Evans-Holm, M., Frise, E., Wan, K. H., Park, S., Mendez-Lago, M., Rossi, F., Villasante, A., Dimitri, P., Karpen, G. H., and Celniker, S. E. (2007) Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* **316,** 1625–1628

Received for publication October 7, 2008.
Accepted for publication December 18, 2008.