Targeting Transposition: At Home in the Genome

Suzanne Sandmeyer

Genome Res. 1998 8: 416-418
Access the most recent version at doi:10.1101/gr.8.5.416

Targeting Transposition: At Home in the Genome

Suzanne Sandmeyer[1]

Department of Biological Chemistry, College of Medicine, University of California, Irvine, California 92717-4700 USA

The completion of the *Saccharomyces* genome sequence in 1996 signaled the beginning of a new chapter in repetitive element sequence analysis (Cherry et al. 1997). The genomic analysis in the article by Kim et al. (1998) in this issue provides new insights into the nature and distribution of repetitive DNA sequences in the *Saccharomyces* genome. It not only assesses the repetitive sequences present in the genome but also clarifies what is not found. *Saccharomyces* has five long terminal repeat (LTR)–retrotransposon families (Ty elements) but no nuclear LINE-like or SINE retroelements and no identifiable DNA-based transposable elements. Remarkably, each Ty family displays insertion specificity. The parallels and differences between the organization of transposable elements in the *Saccharomyces* genome, and what is thus far known of the organization of elements in the genomes of other eukaryotes, suggest that this chapter will have many interesting sequelae.

The genomic analysis by Kim et al. (1998) further defines the relationships of the Ty families of yeast comprised of four copia-like (Ty1, Ty2, Ty4, and Ty5) and one gypsy-like (Ty3) families. The copia-like families have been divided further based on similarity of encoded proteins and LTRs. Ty1 and Ty2 share significant similarity, particularly in the capsid domain, and share a common LTR, called δ. In addition, Kim et al. showed for the first time, that Ty1 and Ty2 δ elements are distinct—differing consistently at a single position. Unlike retrovirus LTRs, yeast element LTRs flanking the internal domain undergo recombination, thereby deleting one copy of the LTR and the internal domain sequence. Interestingly, among the families, there are widely differing copy numbers of complete elements and LTRs: 32 Ty1 elements and 217 Ty1-type δ elements; 13 Ty2 elements and 34 Ty2-type δ elements; two Ty3 elements and 41 σ elements; 3 Ty4 elements and 32 τ elements; and one Ty5 element and seven ω elements. Thus, ratios of LTR sequences to complete elements range from 7 for Ty1 to 21 for Ty3. Because there are many more Ty1 insertions than Ty3 insertions and because the Ty1-type δ elements are more degenerate than the σ elements, Ty1 probably represents an older class of elements within yeast. The fact that the ratio of LTR to complete copies is so much higher for Ty3 suggests that generation of isolated LTRs occurs more frequently for some sequences than for others.

With stunning completeness, the genome sequence reveals the nonrandom nature of the retroelement insertions for each family (see Fig. 1) (Chalker and Sandmeyer 1992; Devine and Boeke 1996; Zou et al. 1996; Bryk et al. 1997; Smith and Boeke 1997). In addition to tRNA genes, 5S, U6, and PI RNA genes were also found associated with insertions. The Tyl, Ty2, Ty3, and Ty4 elements target the upstream region of RNA polymerase 111-transcribed genes: 196 of 217 Ty1; 28 of 34 Ty2; 40 of 41 Ty3; and 30 of 32 Ty4 insertions are within 750 bp of a gene transcribed by RNA polymerase III. This listing is conservative because it does not exclude the possibility that some of the seemingly exceptional insertions are associated with unidentified targets, such as unknown or degenerate polymerase III-transcribed genes. Ty5 does not target tRNA genes. Instead, Ty5 inserts into silenced regions of the genome, the telomeres and mating type loci.

In addition to these important insights, the analysis of Kim et al. (1998) raises provocative questions concerning the relationship between repetitive elements and their habitats: Why is *Saccharomyces* apparently lacking major classes of elements, including quite ancient ones found in bacteria; what is the impact of these elements on genome maintenance; and why do these elements target particular genomic regions? These questions are addressed in the remainder of this article.

Why Does the Nuclear Yeast Genome Appear to Lack DNA Elements and SINE and LINE Classes of RNA elements?

DNA transposons (and DNA viruses), which are found in prokaryotic and eukaryotic organisms, including plants, have not been identified in the yeast genome sequence. One possible explanation is that DNA transposons do not amplify and integrate themselves as efficiently as retrovirus-like RNA elements. Cut-and-paste DNA transposons amplify by hopping into sister chromosomes that have not lost the donor copy. Replicative elements, other than viruses, transpose one new copy without excising the donor element. RNA elements are more similar to viruses in that they generate multiple genomic copies, although the first step is transcription, rather than replication. Thus, some organisms might escape or ultimately erase DNA elements while being colonized by their more prolific cousins, the RNA elements. Integration may provide a second point of discrimination against DNA elements. A fundamental distinction between prokaryotes, where DNA transposons abound, and eukaryotes is, of course, the nuclear membrane. Although the nuclear membrane disintegrates during mitosis in many species, it does not disintigrate in budding yeast. Therefore, DNA transposase would require mechanisms both for nuclear entry and for finding the DNA target. These requirements are also potentially more restrictive for DNA elements than for RNA elements that potentially associate integrase and substrate cotranslationally.

Transposition of DNA elements is initiated by nicking at the ends of the genomic donor DNA, whereas the first genomic incision in retrovirus-like transposition is a concerted strand transfer of donor into target DNA. Nicked genomic

[1]E-MAIL sbandme@uci.edu; FAX (949) 824-2688.

8:416–418 ©1998 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/98 $5.00; www.genome.org

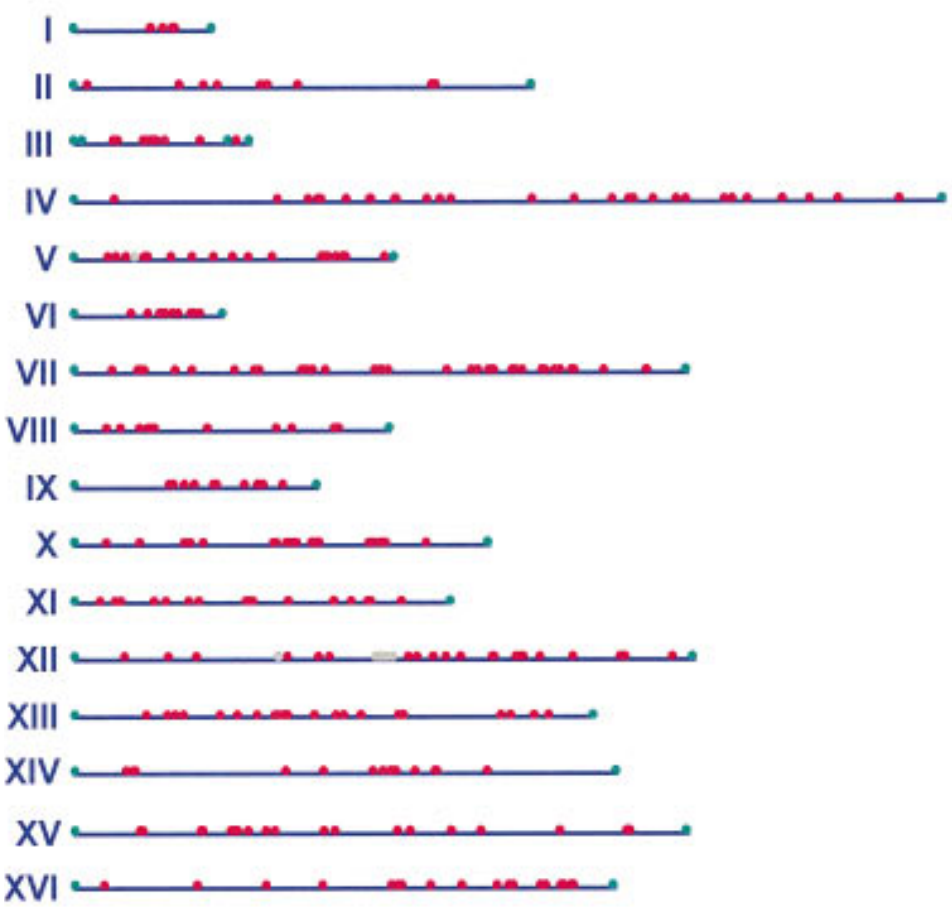

Figure 1 Ty elements transpose into widely dispersed chromosomal targets. The upstream flank of RNA polymerase III-transcribed genes including tRNA (red dots), 5S, U6, and P1 RNA (gray dots) genes are targets for Ty1, Ty2, Ty3, and Ty4 transposition. Ty5 transposes to silenced loci including telomeres and mating-type cassettes on chromosome III (green dots).

DNA is likely unstable and therefore subject to the reverse reaction or DNA repair. Properties of modern DNA transposons reflect the labile nature of the transposition intermediate. For example, the bacterial element, Tn*7*, does not initiate transposition with donor nicking until the target site is identified (Bainton et al. 1993). Furthermore, many transposons in prokaryotes and eukaryotes transpose more efficiently into sites proximal to the donor site. Because the initiation steps for DNA transposition and retroelement transposition differ, they represent another point where yeast might discriminate against the proliferation of DNA elements.

Although elements with reverse transcriptases related to LINE reverse transcriptase exist in other simple eukaryotic organisms and sequences encoding reverse transcriptases related to LINE reverse transcriptase are found in yeast mitochondria, SINE- and LINE-type retrotransposons appear to be absent from the yeast nuclear genome. Perhaps, similar to DNA elements, primitive versions of these elements did not compete effectively with DNA repair. According to the Eickbush model for transposition of R2Bm, a site-specific LINE-like element in *Bombyx mori* (Luan et al. 1993), transposition is initiated by nicking of the target DNA by the element-encoded endonuclease followed by cDNA synthesis in situ. In contrast, retrovirus-like elements copy RNA into DNA in discrete particles and integrate that DNA in a concerted strand transfer. It is generally thought that LINE element reverse transcriptases provide the enzymatic means through which SINE elements are mobilized. Thus, the general absence from the yeast genome of SINEs and pseudogenes is consistent with the notion that their derivation is linked to LINEs.

How Does the Presence of Multiple Elements Impact Maintenance of the Genome and Expansion of Element Populations?

Related sequences in nonhomologous contexts undergo ectopic recombination in yeast, but meiotic recombination of Ty1 elements is less than predicted based on studies of other sequences (Kupiec and Petes 1988). Consistent with this, Kim et al. (1998) found no evidence of chimeric Ty1 and Ty2 elements in the yeast genome. Nevertheless, particular Ty1 elements were determined to be recombinant by inference from sequences at their termini. Thus, although it may not be an efficient process, retroelements do medate gross rearrangements of the yeast genome.

This study also confirmed that most Ty1 and Ty2 elements are likely competent for transposition. This is important and interesting because of a long-standing paradox in the Ty1 field (Curcio et al. 1988). A significant percentage of cellular poly(A)$^+$ RNA is Ty1 RNA, yet transposition is normally infrequent. Furthermore, induction of a single active element increases the frequency of transposition by an amount disproportionate to the increase in RNA. These observations could be accounted for if there were a *cis* bias by retroelement proteins for genomic RNA or if most elements were inactive, leading to synthesis of interfering proteins. Previous investigations have shown that transposition of a marked element can be readily complemented in *trans*. This study, although not conclusive, makes the dominant-negative model less appealing, as only 7 of 45 Ty1 and Ty2 elements were found to contain inactivating mutations. Moreover, Kim et al. (1998) identified a discrete subclass of three Ty1 elements, the members of which differ from the main class by a majority of silent mutations, indicating that they are under selection. How did this come about in the presence of abundant Ty1 RNA and in the absence of a *cis*-acting bias? It could be explained if these elements have undergone mutations resulting in proteins and RNA with specificity for one another (i.e., *trans*-acting bias). This could be tested. If it is not observed, perhaps it will be necessary to finally resolve the fundamental mechanism by which retrovirus-like elements and retroviruses

partition genomic from translated RNA copies.

Is the Propagation of Retroelements in the Yeast Genome a Case of Effective Parasitism or Mutualism?

The case for the latter is attractive, although not proven. Tests of the potential effects of increasing Ty1 element copy number in yeast have shown disadvantages of particular insertions but have not shown that increases in copy number—even large ones—are particularly burdensome (Boeke et al. 1991; Wilke and Adams 1992). In eukaryotic cells, transposable element clustering is widespread (Moyzis et al. 1989; Pimpinelli et al. 1995; Andersson et al. 1998) and helps to account for host tolerance of transposable elements. However, the yeast elements put a new twist on this phenomenon by clustering, relative to specific genomic features, RNA polymerase III-transcribed genes and silenced regions. This pattern of facilitated clustering could be particularly important given the compact nature of the yeast genome. Targeting of Ty3 and Ty5, in particular, is likely to be mediated by transcription factors and silencing proteins, respectively. Insertion just upstream of tRNA genes does not disrupt gene expression, and telomeric and mating-type regions targeted by Ty5 are transcriptionally silent. Thus, a single mechanism may guarantee that insertion is both efficient and transparent to host function.

Eukaryotic genomes may have extracted benefits from retroelement targeting as well. Regions dense in transposable element insertions are distributed throughout the chromosomes of eukaryotic organisms. By providing a common sequence at semiregular intervals, these elements could be coopted for a structural role in chromosome organization or movement. In chromosomes of human and *Drosophila* origin, clusters of transposable elements are within heterochromatic regions. Whether these regions contribute to chromosome function is not yet known. In the case of *Drosophila*, HeT and TART elements have been demonstrated to provide telomeric structure (Biessman and Mason 1997). Although clustering and targeting is observed for endogenous transposable elements, external agents may not be constrained. Thus far, although retrovirus insertion hot spots are observed in the genome, strong regional preferences, such as those described for the yeast elements, are not (Withers-Ward et al. 1994; Carteau et al. 1998).

In yeast, each Ty family displays de facto clustering as a consequence of targeting. The telomeres and genes transcribed by polymerase III are distributed throughout the chromosomes, as are the clusters of repeated elements in eukaryotic cells, in a pattern that would be compatible with structural functions. Furthermore, tRNA genes, rDNA (containing 5S genes), and telomeres, the targets of Ty insertions, are all known to recombine ectopically and also to repress RNA polymerase II transcription. A zone of Ty elements could provide a barrier protecting the integrity and expression of ORFs adjoining these regions. In yeast there is a telomerase. Nevertheless, in addition to providing a buffer for recombination events, yeast transposable elements could provide redundancy in telomerase function or in repair of double-stranded breaks in the DNA (Kupiec and Petes 1988; Moore and Haber 1996).

Conclusions

Analyses such as the one reported here by Kim et al. (1998) are invaluable. Without careful assessment of the numbers, types, and patterns of repetitive elements, we would surely miss much of the richness of the information made available from the genome project. It has been clear for some time that repetitive elements are found in heterochromatic clusters in many species and that this might be the case in all organisms. However, neither the basis of the clustering nor the basis of the heterochromatic configuration of this DNA nor the implication for genomic structure is understood. These questions are not merely academic. Expression of newly inserted sequences in eukaryotic cells, whether introduced by transformation into plants or by retrovirus vectors into eukaryotic, cells is frequently associated with poor expression. As more complete genome sequence information becomes available, revelation of transposable element patterns should prompt renewed consideration of their interplay with their host genome. This will surely lead to further insights into the workings of chromosomes. In addition, if modification of genomes is sought, we must look carefully at the mechanisms and consequences of nature's ongoing modifications. Perhaps it should not be such a rude surprise that many of our would-be modifications, similar to those in nature, fall silent with time.

REFERENCES

Andersson, G., A.-C. Svensson, N. Setterblad, and L. Rask. 1998. *Trends Genet.* 14: 109–114.

Bainton, R.J., K.M. Kubo, J.-N. Feng, and N.L. Craig. 1993. *Cell* 72: 931–943.

Biessmann, H. and J.M. Mason. 1997. *Chromosoma* 106: 63–69.

Boeke, J.D., D.J. Eichinger, and G. Natsoulis. 1991. *Genetics* 129: 1043–1052.

Bryk, M., M. Banerjee, M. Murphy, K.E. Knudsen, D.J. Garfinkel, and M.J. Curcio. 1997. *Genes & Dev.* 11: 255–269.

Carteau, S., C. Hoffmann, and F. Bushman. 1998. *J. Virol.* (in press).

Chalker, D.L. and S.B. Sandmeyer. 1992. *Genes & Dev.* 6: 117–128.

Cherry, J.M., C. Ball, S. Weng, G. Juvik, R. Schmidt, C. Adler, B. Dunn, S. Dwight, L. Riles, R.K. Mortimer, and D. Botstein. 1997. *Nature* 387: 67–73.

Curcio, M.J., N.J. Sanders, and D.J. Garfinkel. 1988. *Mol. Cell. Biol.* 8: 3571–3581.

Devine, S.E. and J.D. Boeke. 1996. *Genes & Dev.* 10: 620–633.

Kim, J.M., S. Vanguri, J.D. Boeke, A. Gabriel, and D.F. Voytas. 1998. *Genome Res.* (this issue).

Kupiec, M. and T.D. Petes. 1988. *Mol. Cell. Biol.* 8: 2942–2954.

Luan, D.D., M.H. Korman, J.L. Jakubczak, and T.H. Eickbush. 1993. *Cell* 72: 595–605.

Moore, J.D. and J.E. Haber. 1996. *Nature* 383: 644–646.

Moyzis, R.K., D.C. Tomey, J. Meyne, J.M. Buckingham, J.-R. Wu, C. Burks, K.M. Sirotkin, and W.B. Goad. 1989. *Genomics* 4: 273–289.

Pimpinelli, S., M. Berloco, L. Fanti, P. Dimitri, S. Bonaccorsi, E. Marchetti, R. Caizzi, C. Caggese, and M. Gatti. 1995. *Proc. Natl. Acad. Sci.* 92: 3804–3808.

Smith, J.S. and J.D. Boeke. 1997. *Genes & Dev.* 11: 241–254.

Teng, S.-C., B. Kim, and A. Gabriel. 1996. *Nature* 383: 641–644.

Wilke, C.M. and J. Adams. 1992. *Genetics* 131: 31–42.

Withers-Ward, E.S., Y. Kitamura, J.P. Bames, and J.M. Coffin. 1994. *Genes & Dev.* 8: 1473–1487.

Zou, S., N. Ke, J.M. Kim, and D.F. Voytas. 1996. *Genes & Dev.* 10: 634–645.