Volume 14 Number 8 1986 Nucleic Acids Research

A 'hot-spot' for Ty transposition on the left arm of yeast chromosome III

John R.Warmington, Rashida Anwar, Carol S.Newlon[1], Richard B.Waring[2], R.Wayne Davies[3], Keith J.Indge and Stephen G.Oliver*

Department of Biochemistry and Applied Molecular Biology, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK and [1]Department of Microbiology UMDNJ - New Jersey Medical School, 100 Bergen Street, Newark, NJ 07103, USA

Received 12 February 1986; Revised and Accepted 27 March 1986

ABSTRACT

The small ring derivative of Saccharomyces cerevisiae chromosome III, which was formed by a cross-over between HML on the left arm and HMR on the right arm, contains three Ty elements. The class II element Ty 1-17 lies immediately centromere-distal to LEU2 on the left arm while two class I elements are tandemly arranged distal to PGK on the right arm. We have sequenced the regions of chromosome III surrounding Ty 1-17 and have defined a region where a number of transposition events have occurred. This region is flanked by the 5' ends of two tRNA genes, $tRNA_3^{Glu}$ on the centromere distal side and $tRNA_3^{Leu}$ immediately in front of LEU2. Close to the $tRNA_3^{Glu}$ gene there is a region containing degenerate delta sequences organised in opposite orientations. Immediately distal to Ty 1-17 there are two complete solo delta elements, one inserted into the other. The sequence indicates that these two delta sequences were inserted into chromosome III by separate transposition events. A model is presented to explain how this structure arose and the role of solo delta elements in transposon propagation and maintenance is discussed.

INTRODUCTION

Chromosome III of the yeast Saccharomyces cerevisiae contains three loci which carry mating-type information: the expression locus MAT, close to the centromere on the right arm, and the silent cassettes HML and HMR which lie close to the left and right ends of the chromosome, respectively (1). Since all three of these loci contain regions of nucleotide sequence homology (2), circular derivatives of chromosome III can be produced by single cross-over events between HML and MAT, or between HML and HMR (3,4). The intrachromosomal recombination event between HML and MAT generates a 63μ covalently-closed circular DNA molecule (3), the "small ring". This molecular species has been isolated (5) and complete gene banks prepared from it (6). These gene banks have been probed for the presence of the repetitive DNA sequences Ty and delta.

Most laboratory strains of S.cerevisiae contain some 30 copies of the Ty transposon (7). In general, these elements consist of a 5.2kb unique

'epsilon' region flanked by identical direct repeats, called 'delta' sequences, of ca.330 bp (7). Homologous recombination events between the two delta elements of a single Ty transposon result in the excision of the 'epsilon' region together with one copy of 'delta', a 'solo delta' element being left behind in the chromosome. Laboratory strains of yeast contain ca.100 such solo delta elements (8).

The small ring form of chromosome III contains three Ty elements (see Fig.1). Two class I elements are tandemly arranged on the right arm, centromere distal to the PGK gene. On the left arm there is a single class II element (Ty 1-17) which is immediately centromere distal to LEU2, from which it is separated by a $tRNA_3^{Leu}$ gene (9). We have published previously the complete nucleotide sequence of this class II element (10). In this paper we examine the sequences flanking this element and demonstrate that the region of chromosome III centromere-distal to LEU2 contains a number of solo delta elements in addition to Ty 1-17. Examination of the sequence provides evidence for successive transposition events and indicates that this region of the chromosome may be a preferred target for the yeast Ty transposon.

MATERIALS AND METHODS

Screening of Chromosome III plasmid library for Ty1 and delta sequences.

The 200 kb circular derivative of chromosome III which carries sequences between the HML and MAT loci has been cloned and restriction mapped (6). A series of plasmids each containing one of the eighteen BamHI fragments which comprise the ring chromosome cloned into the BamHI site of YIp5 were used in these studies. Ty1 - containing plasmids were initially identified as plasmids whose yeast inserts hybridized strongly to more than 20 bands in yeast genomic DNA digested with BamHI or EcoRI. Confirmation of Ty1 homology was obtained by digesting each of the ring chromosome plasmids with BamHI and EcoRI, running the digests on 0.7% agarose gels and hybridizing blots of the gels with the 5.6 kb XhoI fragment containing Ty1-17 which had been labelled with ^{32}P by nick translation. Conditions for gel electrophoresis, blotting, nick translation and hybridization have been described in ref.5. The hybridization with the Ty1-17 probe confirmed the presence of three Ty1 elements in the circular chromosome and revealed the presence of a number of additional weekly hybridizing fragments which were presumed to contain solo delta sequences. Confirmation of the presence of delta sequences in these fragments was

obtained by hybridization with the 2.5 kb BamHI-HindIII fragment from the plasmid pPM5 (12) which contains the SUP-RL1 locus as well as a solo delta sequence (S. Sandmeyer, per.comm.).

DNA Sequencing

The DNA sequence of the four contiguous BamHI fragments of the small ring form of chromosome III, A5C,G4B,D8B and C2G (see Fig.1) was determined using a modification (13) of the dideoxynucleotide chain termination method of Sanger et al.,(14). Random sub-clones of these fragments were prepared by the method Deininger (15). Full details of the methods employed in our laboratory for sequence determination and computer analysis are given in refs. 16 and 17. The sequences described in this paper, as well as the longer sequence which has been made available to the editors, is the result of sequencing each nucleotide at least twice. Most nucleotides have been sequenced four or more times and greater than 95% of the sequence has been determined in both orientations.

Chemicals and enzymes

Restriction endonucleases were obtained from BCL Ltd., Pharmacia and BRL Ltd. T4 DNA polymerase was purchased from Pharmacia. T4 DNA ligase and Klenow fragment of DNA polymerase I were from BCL Ltd. [α^{32}p]dATP (sp.activity 400μCi mMol^{-1})was purchased from Amersham International.

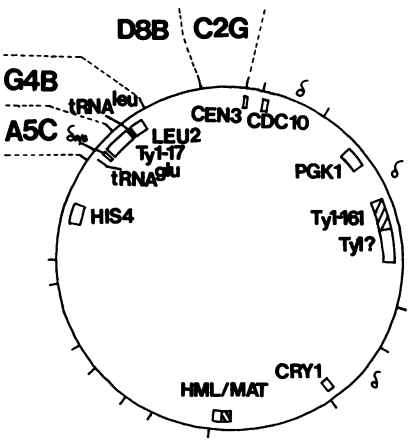

Fig.1 Map of the small ring form of chromosome III. The lines divide the circle into contiguous BamHI fragments and the 4 fragments detailed in this paper (A5C,G4B,D8B and C2G) are named. The approx. positions of the 3 Ty and 4 solo delta elements in the ring are marked.

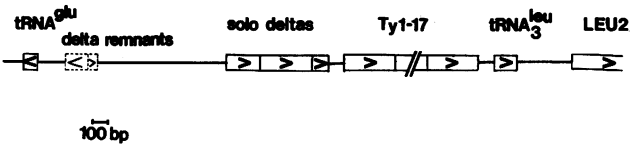

Fig.2 Fine structure map of chromosome III distal to LEU2. The chevrons (>) indicate the orientation of delta elements or of the coding regions of tRNA and structural genes.

RESULTS

Location of Ty and delta sequences in the small ring form of chromosome III

Southern hybridization analysis revealed the presence of three Ty elements, one (Ty1-17) on the left arm centromere-distal to LEU2 and two on the right arm which are tandemly arranged distal to PGK1 (Fig.1). In addition, two regions immediately distal to Ty1-17 on the left arm and three on the right arm were found which showed homology to delta sequences. Their approximate positions are indicated in Fig. 1. DNA sequence analysis has revealed a number of sequences showing homology to delta in the region immediately distal to Ty1-17 on the left arm. A detailed description of this region is given in this paper.

There are two tRNA genes centromere-distal to LEU2

Fig.2 shows the fine structure map of the region of chromosome III immediately centromere-distal to LEU2. Andreadis et al.,(9) had previously reported the presence of a $tRNA_3^{Leu}$ gene 350 bp away from the start of the LEU2 structural gene and our own sequence data has confirmed this. In addition, we have discovered a second tRNA gene, for $tRNA_3^{Glu}$, some 8 kb further away from the centromere. The sequence of this new gene has been indicated by heavy underlining in Fig. 3. The coding sequence of this gene is exactly the same as that of 3 other yeast $tRNA_3^{Glu}$ genes described previously (18,19). It is noteworthy that all 4 of these genes were associated with Ty or delta sequences. By analysing two independent yeast gene banks we estimate that between 20% and 50% of $tRNA_3^{Glu}$ genes lie not more than 5 kb away from a Ty or delta sequence (JRW and SGO, unpublished observations). Others (18) have suggested that nearly all such genes are associated with these repeated elements. Apart from all having the RNA polymerase III termination signal of a tract of deoxyadenosine residues (20), there is little in common between the chromosomal sequences in which these 4 $tRNA_3^{Glu}$ coding sequences are

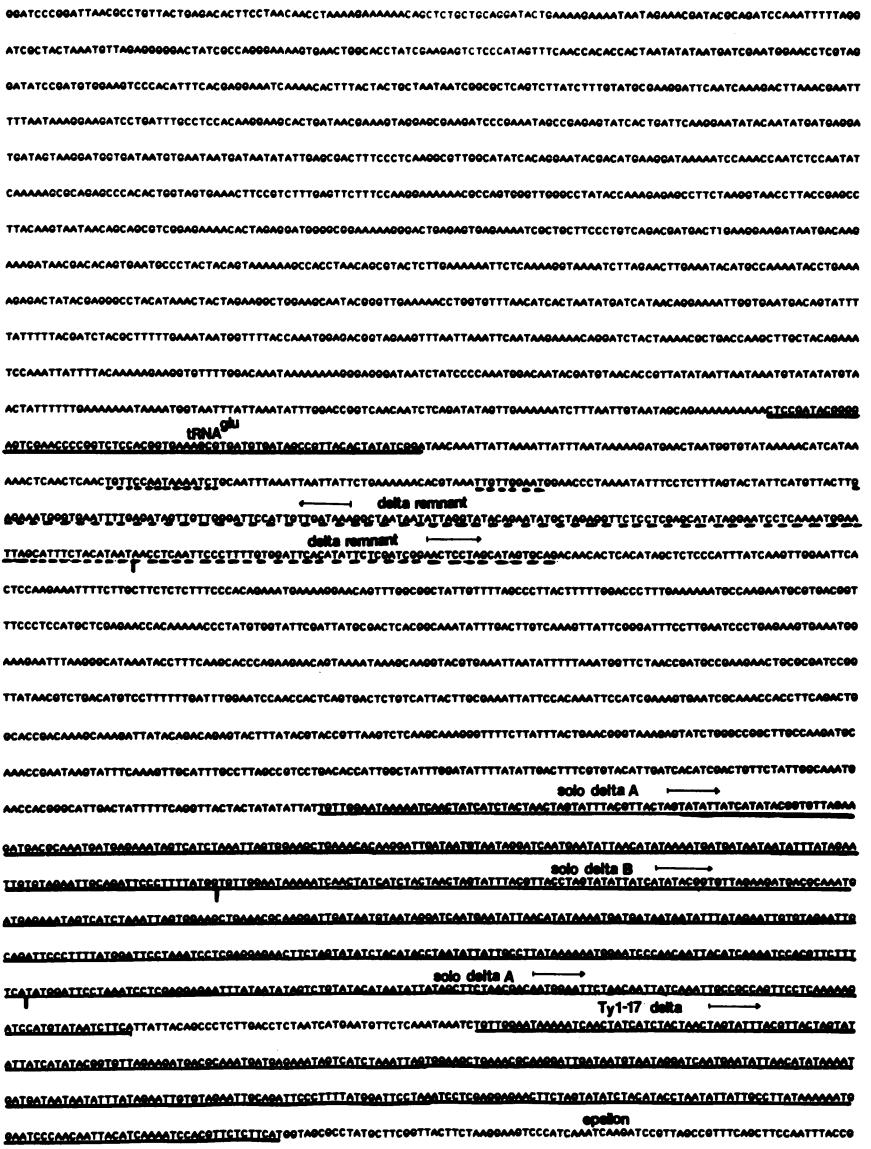

Fig.3 DNA sequence of chromosome III distal to LEU2. This sequence covers a similar region to the map in Fig.2. It shows the position of the $tRNA_3^{Glu}$ gene (underlined) and of the double delta insert (underlined). The degenerate delta sequence close to the tRNA gene are indicated by the dotted underlining. Arrows above the line indicate the orientation of all the delta elements and remnants. A complete sequence of the chromosome between $tRNA_3^{Glu}$ and $tRNA_3^{Leu}$ has been supplied to the editors and will appear in the public data banks, portions of this larger sequence appear in refs. 9 and 10.

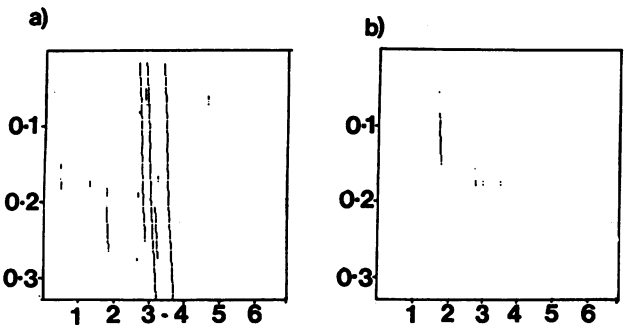

Fig.4 DIAGON dot matrix comparison to search for delta homology.
a) Compares the Ty1-17 delta sequence with that of the A5C BamHI fragment. The orientation of the delta being the same as that shown in Fig.4.
b) Shows the same comparison but using the reverse orientation for the delta sequence. Proportional matching and a span length of 83 were used in both comparisons. The probability of a match occuring by chance is 10^{-6}.

embedded. However, a consensus sequence, 5' T Pu A/T N T T Pu T A/T A/-3', may be derived for the 5' flanking region of these 4 genes.

DNA sequence of chromosome III between the two tRNA genes

The complete nucleotide sequence of the left arm of chromosome III

```
     1         10         20         30         40         50
ATTATTGTTGGAATAAAAATCAACTATCATCTACTAACTAGTATTTACGTTACTAGTATA
    60         70         80         90        100        110
TTATCATATACGGTGTTAGAAGATGACGCAAATGATGAGAAATAGTCATCTAAATTAGTG
   120        130        140        150        160        170
GAAGCTGAAACaCAAGGATTGATAATGTAATAGGATCAATGAATATTAACATATAAAATG
   180        190        200        210        220   226          9
ATGATAATAATATTTATAGAATTGTGTAGAATTGCAGATTCCCTTTTATGGTGTTGGAAT
          20         30         40         50         60
AAAAATCAACTATCATCTACTAACTAGTATTTACGTTACcTAGTATATTATCATATACGG
          80         90        100        110        120
TGTTAGAAGATGACGCAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGC
         140        150        160        170        180
AAGGATTGATAATGTAATAGGATCAATGAATATTAACATATAAAATGATGATAATAATAT
         190        200        210        220        230
TTATAGAATTGTGTAGAATTGCAGATTCCCTTTTATGGATTCCTAAATCCTCGAGGAGAA
         250        260        270        280        290
CTTCTAGTATATCTACATACCTAATATTATTGCCTTATAAAAAATGGAATCCCAACAATT
         310        320      226 230        240        250
ACATCAAAATCCACGTTCTtTTCATATGGATTCCTAAATCCTCGAGGAGAAtTTaTAaTA
260        270        280        290        300        310
TAgTCTgtATACaTAATATTATaGC-TTcTAAcgAcAATGGAATTtCtAACAATTA--TCA
320        330        340        350
AAtTgCCgCcaGTTCcTC(AAAAagaTCCAtGTaTaatC)TTC-ATTAT
```

Fig.5 The double-delta insertion. Delta-A is marked in bold type and delta-B in normal type. Target site duplications are underlined. Lower case letters indicate deviations from the Ty1-17 delta consensus sequence. The absence of bases found in this consensus is indicated by a dash (-). The sequence in parentheses at the end of the delta-A sequence represents a partial internal duplication and accounts for the longer than normal length of this element.

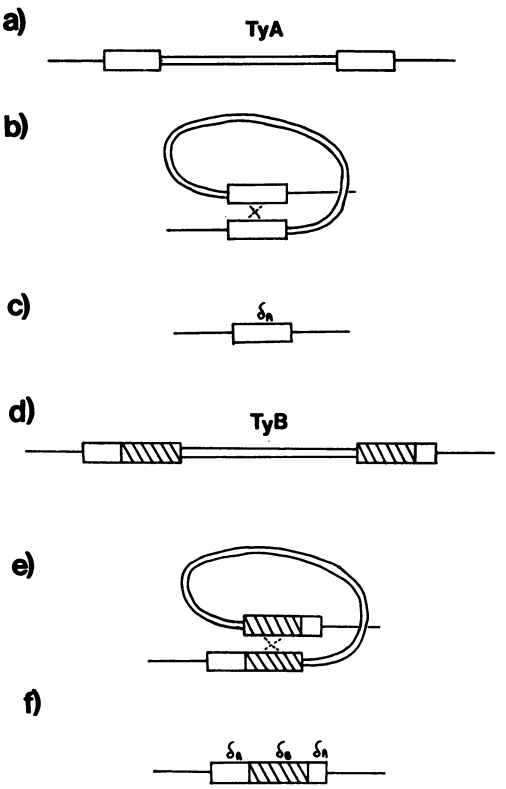

Fig.6 Model for the derivation of the double delta insertion. a) Resident Ty A element which was inserted into chromosome III by a transposition event, duplicating a chromosomal target sequence. b) Loss of Ty A epsilon region by homologous recombination to leave delta-A as a solo delta element (c). d) Insertion of Ty B into delta-A by a transposition event duplicating bases 221 to 226. e) Loss of Ty B epsilon region leaving solo delta-B element (hatched) inserted in delta-A (f).

between $tRNA_3^{Glu}$ and the delta sequence at the left end of Ty1-17 is presented in Fig.3. The entire sequence for the two adjacent BamH1 fragments A5C and G4B (see Fig.1) has been made available to the editors of this journal and will be entered into the established data libraries. The region between the two tRNA genes was searched for sequence homology with delta using the DIAGON dot matrix analysis (21) and the results of that search are shown in Fig.4. There is a region immediately adjacent to $tRNA_3^{Glu}$ which shows a number of blocks of homology to the delta consensus sequence. These blocks of homology occur in both orientations (indicated by arrows above the sequence in Fig.3) and one portion of the 5' end of the

delta sequence appears to be repeated 4 times. The delta sequences in this region are so degenerate, however, that it is impossible to tell whether these remnants originated from a series of transposition events or resulted from the rearrangement of a single delta sequence.

A second region, immediately in front of the 5' delta of Ty1-17, also shows homology with the delta consensus (Fig.4). A detailed inspection (Fig.5) of the sequence demonstrates that this region contains two complete delta elements which we call delta-A (marked in bold type in Fig.5) and delta-B (normal type in Fig.5). The delta-B sequence is wholly contained within that of delta-A. While delta-B shows near-perfect homology to the consensus delta sequence, the delta-A degenerates at its 3' end. (Deviations from the consensus are indicated by lower case letters in Fig.5). We interpret this to mean that delta-A is older than delta-B. Further inspection of the sequence indicates that delta-B was inserted into delta-A as a result of a transposition event. There is a 5 bp duplication of the delta-A target sequence, 5'TATGG3', at either end of delta-A. Such 5 base pair duplications are diagnostic for transposition events (8). Delta-A is itself flanked by a 5 bp duplication of its chromosomal target sequence (5'ATTAT3'), although the 3' end of delta-A is quite degenerate and the exact end of the element is difficult to recognise. (These target sequences are underlined in Fig.5).

The simplest interpretation of the origin of the insertion of delta-B into delta-A is as follows: Delta-A arose from a delta-delta cross-over between the two ends of a Ty element, Ty-A, which had transposed into this region of chromosome III. The solo delta left behind (delta-A) then acted as the target for a second transposition event which inserted a new Ty element, Ty-B. A cross-over between the two delta sequences of Ty-B then left behind a solo delta element, delta-B, inserted into delta-A. This sequence of events is diagrammed in Fig.6.

DISCUSSION

There are only two regions of the small ring form of chromosome III which contain sequences hybridizing to Ty probes (see Fig.1). Although at least two his4 mutants (his4-912 and his4-917;22,23) have been found to be due to Ty insertions, no Ty elements or delta sequences were found in this region of the ring chromosome. Thus Ty insertions adjacent to HIS4 do not appear to be particularly common events. This contrasts with the region immediately in front of the LEU2 gene which has been examined in detail in

this paper. This region contains the complete transposon, Ty1-17 (10,11), which was found to lie between the 5' ends of two tRNA genes, $tRNA_3^{Leu}$ (9) and $tRNA_3^{Glu}$ (Fig.3). The segment of the chromosome between these two genes appears to be a "hot-spot" for Ty transposition. In addition to Ty1-17 itself it contains two other regions which display sequence homology to delta elements. These both lie between the $tRNA_3^{Glu}$ gene and Ty1-17. Both hybridization and sequence analyses demonstrated that there are no more delta elements between the $tRNA_3^{Leu}$ gene and the centromere. Further, hybridization analysis has demonstrated (CSN, unpublished) that there are no other regions of delta homology on the left arm of the chromosome between the $tRNA_3^{Glu}$ gene and HML.

Since Ty elements appear to insert at random positions within the yeast genome (6,7), it is necessary to analyse why transposition events are so frequent in this region of chromosome III. The double delta element immediately centromere distal to Ty1-17 appears to be the result of two successive rounds of transposition and recombination and the region of delta homology further centromere distal to it is very degenerate but may be the product of multiple transposition events.

Ty elements and delta sequences are frequently associated with tRNA genes (18). Indeed, all four copies of $tRNA_3^{Glu}$ which have been sequenced so far are adjacent to delta sequences. Thus a region of chromosome between the 5' ends of two tRNA genes may represent a particularly favourable environment for transposition. A study of the DUR1,2 region of chromosome II by Genbauffe et al.,(19) revealed the presence of a delta sequence, the related element, tau, and part of a sigma element (24,25) between the 5' ends of a $tRNA_3^{Glu}$ gene and a $tRNA^{Cys}$ gene.

It is interesting to note that the tau element on chromosome II had been inserted into a delta sequence, just as delta-B on chromosome III had been inserted into delta-A (Fig.5). Moreover both the insertion events occurred within 5 base pairs of one another in the region of the delta sequence between the TATA box and the transcriptional start site. The tau insertion duplicated base pairs 219 to 223, whereas the delta-B insertion produced a duplication of base pairs 223 to 227. In their original description of target site duplications by Ty transposons, Gafner and Philippsen (8) described an example where a complete Ty element was inserted, in opposite orientation, into a solo delta sequence. Again, the transposition occurred in the same region; nucleotides 212-216 of the solo delta being duplicated. If this target region is found to be conserved for other insertions into

delta sequences, then it may indicate that delta elements themselves are preferred targets for transposition events. The steady-state level of whole Ty elements in the yeast genome would then depend critically on events occuring within delta sequences. A balance would be struck between the deletion of Ty elements by delta-delta recombinations and the promotion of transposition by the solo delta sequences left behind.

An alternative hypothesis is that transpositions accumulate in delta elements because they represent sites where insertions are selectively neutral. This may be the case for the transposons of the slime moulds In the true slime mould, Physarum polycephalum, the majority of the 2,000-5,000 copies of the Tp1 transposon are inserted into other copies of the same element, always within the unique sequence between the two terminal repeats (26). A similar result was obtained with the cellular slime mould, Dictyostelium discoideum, where 5/6 copies of the DIRS-1 transposon isolated were inserts into part of the unique region of the same element (27). The authors considered that DIRS-1 represented a "hot-spot" for transposition events. In an effort to learn more about the role of delta sequences in the spread of Ty elements through the yeast genome, we are currently sequencing the tandem Ty elements on the right arm of chromosome III.

ACKNOWLEDGEMENTS

This work was supported by SERC grant no. GR/C34656 (to SGO and RWD), by NIH grant no. GM21510 (to CSN) and by a NATO Collaborative Award (to SGO and CSN). We thank Nita Patel and Norman Hardman for useful discussions.

*To whom correspondence should be addressed
Present addresses: [2]Department of Pharmacology, UMDNJ, Rutgers Medical School, Piscataway, NJ 08854, USA and [3]Allelix Inc., 6850 Goreway Drive, Mississauga, Ontario L4V 1P1, Canada

REFERENCES

1. Herskowitz, I. and Oshima, Y. (1982) in Molecular Biology of the Yeast Saccharomyces, Strathern, J.N., Jones, E.W. and Broach, J.R. Eds., pp 181-289, Cold Spring Harbor Laboratory, New York.
2. Nasmyth, K.A., Tatchell, K., Hall, B.D., Astell, C. and Smith, M.(1981) Nature 289, 244-250
3. Strathern, J.N., Newlon, C.S., Herskowitz, I. and Hicks, J.B. (1979) Cell 18, 309-319
4. Klar, A.J.S., Strathern, J.N., Hicks, J.B. and Prudente, D. (1983) Molec. and Cell. Biol. 3, 803-810
5. Devenish, R.J. and Newlon, C.S. (1982) Gene 18, 277-288
6. Newlon, C.S., Green, R.P., Hardeman, K.J., Kim, K.E., Lipchitz, L.R.,

Palzkill, P.G., Synn, S. and Woody, S.T. (1985) UCLA Symp. on Molec. and Cell Biol. 33 (In press)
7. Cameron, J.R., Loh, E.Y. and Davis, R.W. (1979) Cell 16, 739-751
8. Gafner, J. and Philippsen, P. (1980) Nature 286, 414-418
9. Andreadis, A., Hsu, Y.-P., Kohlhaw, G.B. and Schimmel, P. (1982) Cell 31, 319-325
10. Warmington, J.R., Waring, R.B., Newlon, C.S., Indge, K.J. and Oliver, S.G. (1985) Nuc. Acids Res. 13, 6679-6693
11. Kingsman, A.J., Gimlich, R.L., Clarke, L., Chinault, R.C. and Carbon, J. (1981) J.Molec.Biol. 145, 619-632
12. Olson, M.V., Page, G.S., Sentenac, A., Piper, P.W., Worthington, M., Weiss, R.B. and Hall, B.D. (1981) Nature 291, 464-469
13. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Nat. Acad. Sci. U.S.A. 80, 5463-5467
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467
15. Deininger, P.L. (1983) Analyt. Biochem. 129, 216-223
16. Fulton, A.M., Mellor, J., Dobson, M.J., Chester, J., Warmington, J.R., Indge, K.J., Oliver, S.G., de la Paz, P., Wilson, W., Kingsman, A.J. and Kingsman, S.M. (1985) Nuc. Acids Res. 13, 4097-4112
17. Martin, W.J., Warmington, J.R., Galinski, B.R., Gallagher, M., Davies, R.W., Beck, M.S. and Oliver, S.G. (1985) Bio/Technology 3, 911-915
18. Eigerl, A. and Feldmann, H. (1981) EMBO J. 1, 1245-1250
19. Genbauffe, F.S., Chisholm, G.E. and Cooper, T.G. (1984) J. Biol. Chem. 259, 10518-10525
20. Koski, R.A., Clarkson, S.G., Kurjan, J. and Hall, B.D. and Smith, M. (1980) Cell 22, 415-425
21. Staden, R. and McLachlan, L. (1982) Nuc. Acids Res. 10, 141-156
22. Chaleff, D.T. and Fink, G.R. (1980) Cell 21, 227-237
23. Farabaugh, P.J. and Fink, G.R. (1980) Nature (London) 286, 352-356
24. del Rey, F.J., Donahue, T.F. and Fink, G.R. (1982) Proc. Nat. Acad. Sci. U.S.A. 79, 4238-4142
25. Sandmeyer, S.B. and Olson, M.V. (1982) Proc. Nat. Acad. Sci. 79, 7674-7678
26. Pearston, D.H., Gordon, M. and Hardman, N. (1985) EMBO J. 4, 3557-3562
27. Cappello, J., Cohen, S.M. and Lodish, H.F. (1984) Molec. Cell. Biol. 4, 2207-2213