Polymorphisms on the right arm of yeast chromosome III associated with Ty transposition and recombination events

John R.Warmington, Rebecca P.Green*, Carol S.Newlon[1] 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, 185 South Orange Avenue, Newark, NJ 07103, USA

ABSTRACT

The region of Saccharomyces cerevisiae chromosome III centromere-distal to the PGK gene is the site of frequent chromosome polymorphisms. We have sequenced this region from fragments of chromosome III isolated from three different yeast strains, GRF88, CN31C and CF4-16B. The sequence analysis demonstrates that these polymorphisms are associated with the presence of Ty and delta elements and defines a region of the chromosome which is a hot-spot for transposition events (the RAHS). The three strains can be arranged into a logical evolutionary series in which successive transposition and recombination events insert Ty elements and fuse them with consequent deletions of chromosome and of transposon sequences. The influence of such events on yeast genome evolution is discussed.

INTRODUCTION

Most laboratory strains of the yeast Saccharomyces cerevisiae contain ca. 30 copies (1) of the yeast retrotransposon, Ty (for review, see ref. 2). the general structure of these elements consists of a 5.2 kb unique 'epsilon' region flanked by two directly repeated 'delta' sequences of ca. 330 bp each which have identical, or nearly identical, nucleotide sequences. The sequence homology between delta sequences in the same element permits cross-over events to occur between them which results in the excision of the epsilon region together with one copy of delta, a 'solo delta' sequence being left behind in the chromosome. Such delta-delta recombination events appear to be relatively common since laboratory strains of yeast usually contain some 100 solo delta sequences (3). The distribution of Ty and solo delta elements within the yeast genome appears to be random (1). However, the preferential insertion of Ty elements into promoter sequences (4) and adjacent to tRNA genes (5,6) has been recorded.

Most Ty elements whose chromosomal context has been studied in detail are associated with some particular class of mutation for which the investigators selected. In order to examine the dynamics of the interaction

of Ty with the yeast genome and to understand the factors involved in maintaining an equilibrium number of these transposons within it, an analysis of unselected events is required. Such an analysis is provided by the determination of the DNA sequence of yeast chromosome III which is being performed in our laboratories.

In the course of these sequencing studies, we have discovered two regions of chromosome III which appear to act as transposition 'hot-spots' (Fig. 1a). One of these, the left-arm hot-spot (LAHS) lies immediately centromere distal to the LEU2 gene. DNA sequencing (7) showed the LAHS to be flanked by the 5' ends of two tRNA genes, tRNA^{Leu}3 and tRNA^{Glu}. It contains the complete class II Ty element, Ty 1-17 (ref. 8), a number of delta fragments and a structure in which two solo deltas are inserted into one another. We interpreted this structure to be the result of two successive rounds of transposition and delta-delta recombination and suggested that the delta sequence itself may be a preferred target for transposition events. Our own data (8) and that of others (3,9) have shown that Ty and other repeated elements, such as tau (6) have a propensity to insert into the same region of delta DNA between base pairs 212 and 227. If delta is a preferred target for transposition, then the steady-state number of complete Ty elements would be critically dependent on events occuring within delta sequences. An equilibrium would be established between the deletion of Ty elements by delta-delta recombinations and promotion of new transposition events by the solo delta sequences left behind. Such a scheme has been described as the 'Ty cycle' (10) and in this paper we present a further examination of this cycle based on our sequence analysis of the right arm transposition hot-spot (RAHS). The RAHS lies centromere-distal to PGK and we have sequenced this region from three different isolates of chromosome III. The data demonstrate that successive transposition and recombination events involving Ty and delta sequences are responsible for the polymorphisms found in this region of the chromosome.

METHODS

(1) <u>Strains</u>: the region of chromosome III analysed in this communication was recovered from three yeast strains. GRF 88 was derived from S288c and carries the his4-38 mutation (M. Rose, pers. comm.). The genealogy of S288c is described in ref. 11. CF4-16B was obtained from S.C. Falco. The strain is disomic for chromosome III and carries complementing his4C alleles. It was also derived from S288-related strains. CN31C is a segregant of a

Table 1

Strain	Genotype
GRF88	MATα, his4-38, (gal2?)
CF4-16B	MATa his4-290 _____ ura3-52, ade2-1 (gal2?) MATα his4-280
248	MATa, cry1-3, his4, leu2-1, lys2, ade6
YNN27	MATα, trp1-289, ura3-52, gal2
CN31C	MATa, cry1-3, his4, leu2-1, ade6, ura3-52 (gal2?)

cross between strain 248 (ref. 12) and YNN27 (ref. 13). It carries the four chromosome III markers from strain 248 and the ura3-52 allele from YNN27. The genealogy of strain 248 can be traced back through at least five crosses and its background includes strains from the Cold Spring Harbor Yeast Course, T. Manney and C. McLaughlin (J. Strathern, pers. comm.). The cry1-3 (and perhaps much of the right arm of the strain 248 chromosome III) originated in a strain obtained from C. McLaughlin. The full genotypes of these strains are given in Table 1.

(2) Plasmids: CB15A-2D was recovered from a library constructed in yeast shuttle vector YCp50 from GRF88 DNA partially digested with Sau3A by M. Rose and P. Novick (pers. comm.), using a probe from the centromere-distal end of plasmid 62B5-2D (Fig. 1b). Extensive Southern blot analysis (data not shown) demonstrated that this plasmid is homologous to the ring III chromosome from strain XG1-24b which has been cloned and mapped (12,14). Plasmids 31C-GF and 16B-GF were constructed by gap repair, in yeast , of a linearized plasmid containing sequences that flank the region of interest in the chromosome (15). The plasmid used for gap filling, B52D-H9G-GF, was constructed in the shuttle vector YCp50 and contains the centromere proximal 1.8kb BamHI-XhoI fragment of plasmid H9G adjacent to the centromere distal 2.8 kb EcoRI-BamHI fragment of plasmid 62B5-2D in their chromosomal orientation (see Fig. 1b). The plasmid was linearized at the unique BamHI site which separates the chromosome III fragments by digestion with BamHI and used to transform strain CN31C or strain CF4-16B. Ura[+] transformants containing the gap-repaired plasmid of interest were distinguished from those containing only the recircularized starting plasmid by isolating covalently closed circular DNA using the procedure of

Devenish and Newlon (ref. 16) and looking for large plasmids carrying sequences within the region of interest by Southern blotting and hybridizing with ^{32}P labelled single copy from plasmid CB15A-2D. All of the Ura$^+$ transformants screened contained extrachromosomal plasmid and approximately half of the transformants carried the gap-filled plasmid. The chromosome III sequences in the plasmids were shown to be the same as those in the intact chromosome by comparing restriction maps of the plasmid with those deduced for the chromosome by Southern analysis.

Southern Blotting

Plasmid DNAs were purified, digested with various combinations of restriction enzymes and fractionated on 0.7% TBE-agarose gels (17). The fractionated DNAs were transferred to Hybond N by the procedure of Southern (18). The filters were probed with a Tyl-15 specific probe which comprised a 9.6 kb HindIII restriction fragment from the plasmid pMA91-301 (19) and labelled with ^{32}P-α-dCTP (3000 Ci/mol) by nick translation (20). Filter hybridisation conditions used were as recommended by Amersham for Hybond N. After autoradiography, the Ty probe was removed from the filters and they were re-hybridised with a 0.6 kb BglIII-EcoRI fragment from 31C-GF.

Sequencing of the d_1 delta element and its flanking DNA in CB15A-2D and H9GSTy

The 1.4 kb XhoI-PvuII restriction fragments containing the d_1 element of Tyl-161 in CB15A-2D and the single Ty in 16B-GF, were isolated, end repaired with DNA polymerase I (Klenow fragment), and blunt end ligated into the SmaI restriction site of Ml3mp18. The fragments were sequenced in both orientations using the 'Bandaid' technique of Dale et al. (21).

Sequencing of the d_2 element and its flanking DNA from CB15A-2D.

The 2.6 kb SalI restriction fragment spanning the two tandem Ty elements was isolated, and cleaved with BglII and PvuII. The 500 bp BglII-PvuII DNA fragment containing the d_2 sequence was sub-cloned into SmaI, BamHI cleaved Ml3mp18 and mp19 and sequenced in both orientations. The 2.6 kb SalI fragment was also cleaved with XhoI and the SalI-XhoI fragments were subcloned into SalI cut Ml3mp18 and sequenced.

Sequencing of the d_3 element and its flanking sequences in CB15A-2D and 16B-GF.

The 2.7 kb EcoRI restriction fragment containing d_3 delta sequences was isolated, and cleaved with either BglII, ClaI or XhoI. The resulting fragments were subcloned into Ml3mp18 and mp19 cleaved with the appropriate combinations of EcoRI, BamHI, AccI and SalI and sequenced.

DNA sequence analysis of the unique 31C-GF region.

This region along with some flanking chromosome III DNA was isolated on
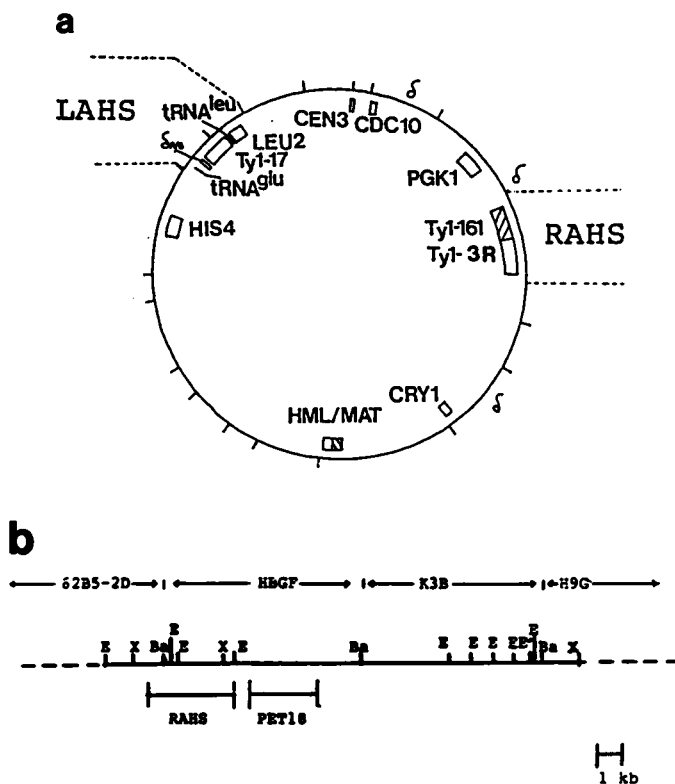
Figure 1 Transposition hot spots on chromosome III.
a) A map of the 200 kb circular derivative of chromosome III from strain GRF88 which has been cloned and mapped (ref. 14) LAHS and RAHS indicate the positions of the left arm and right arm Ty element insertion hot spots.
b) An expanded map of the region from strain CN31C containing the RAHS. Designations of the cloned BamHI fragments are given above the map. Fragments K3B and H9G are the same in the three strains examined in this report. The polymorphic RAHS reported in this paper extends from the EcoRI site shown in fragment 62B5-2D through the XhoI site shown in fragment HBGF. The pet18 mutation is a deletion of the region shown (ref. 38).

a 4kb XhoI fragment, and an overlapping 2 kb EcoRI fragment (see Fig. 2c). The XhoI fragment was subcloned into the SalI site of M13mp19 and the EcoRI fragment was subloned into the EcoRI site of M13 mp18. The two DNA fragments were sequenced in both orientations using the unidirectional deletion strategy of Henikoff et al. (22) except that Mung Bean Nuclease was used in place of S_1 nuclease. The M13 mp18 subclones with BamHI and PstI, prior to ExoIII digestion.

Northern blotting and hybridisation

Total RNA was prepared from yeast cultures grown on 1×10^7 cells/ml in YEPD medium as described by Elder et al. (23). Total yeast RNA (10 μg) was subjected to electrophoresis on a 1.2% formaldehyde agarose gel and transferred onto Hybond N (Amersham). The filter was hybridised with the 1.4 kb XhoI-BamHI restriction fragment from 31C-GF (fig. 2c) labelled by nick translation. The electrophoresis, blotting and hybridisation procedures were as described by the Amersham 'Hybond N' booklet.

MATERIALS

Restriction endonucleases and DNA modifying enzymes were purchased from BCL, Bethesda Research Labs and Pharmacia. Reaction conditions were as recommended by the manufacturer. Hybond N and radionucleotides were from Amersham International. Other chemicals were purchased from BDH.

RESULTS

Isolation and characterisation of the RAHS from three different chromosomes III

The region centromere distal to the PGK locus on the right arm of chromosome III (Fig. 1) shows considerable restriction enzyme site polymorphism (12). Most of this can be attributed to the presence or absence of one or more Ty elements. It appears that this region of the chromosome is a 'hot spot' for transposition events and, therefore, it was designated the RAHS (for Right Arm Hot Spot). We have cloned the RAHS region from the chromosomes III of three different strains of Saccharomyces cerevisiae: GRF88, which is homologous in this region to the donor of the ring chromosome in the heterozygous diploid XG1-24B (ref 12), CF4-16B (a relative of S288C) and CN31C (which carries chromosome III from strain 248, ref. 12). Our analysis demonstrated that these forms of chromosome III contained, respectively, two, one and no Ty elements in the RAHS region. The evidence for this comes from restriction mapping and Southern hybridisation experiments and is described briefly below.

Plasmid CB15A-2D contains the RAHS region of the GRF88 chromosome III, cloned as an 18 kb Sau3A fragment which was derived from a partial digest of genomic DNA and inserted into the BamHI site of YCp50. The RAHS from the other two strains were also isolated in YCp50 using gap repair (see Methods). Plasmid 31C-GF contains a 21 kb RAHS insert from the chromosome III of strain CN31C, while 16B-GF contains a 26 kb RAHS fragment from strain CF4-16B. Each

plasmid was analysed using a range of restriction enzymes, both singly and in combination, and the fragments containing Ty sequences determined by Southern hybridisation (18), using a nick- translated (20) HindIII fragment that contains the class I Ty element, Ty 1-15 (ref. 19).

Plasmid CB15A-2D was found to contain two class I Ty elements. The centromere-proximal (left-hand) member of the pair, we designated Ty 1-161 since it occupies the same position as the element of that name originally described by Kingsman et al. (24). Like the original Ty 1-161, the XhoI fragment carrying this element is longer than XhoI fragments carrying other class I transposons and by restriction analysis the element appears to have 1.2 kb insertion at its 5' (left) end. The centromere-distal Ty in the pair we named Ty 1-3R (see fig. 2). No sequences homologous to the Ty 1-15 probe were found in plasmid 31C-GF although prolonged exposure of the Southern blots did reveal some faint bands which subsequent sequence analysis (see below)
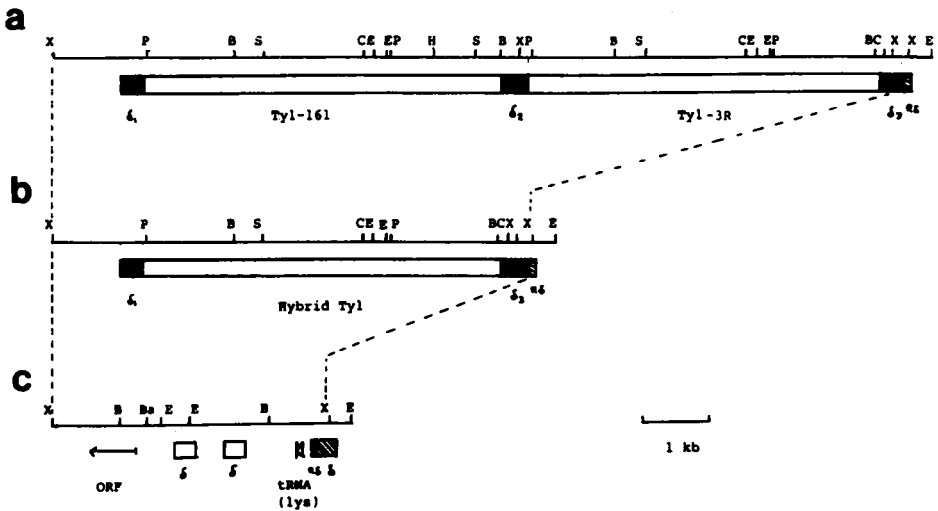
Figure 2 Detailed restriction maps of the RAHS from 3 yeast strains, showing Ty element related polymorphisms.
a) The RAHS of strain GRF88 from the plasmid CB15A-2D containing two tandemly linked Ty1 elements.
b) The RAHS of strain CF4-16B from the plasmid 16B-GF, which contains a single hybrid Ty1 element.
c) The RAHS of strain CN31C from the plasmid 31C-GF. This strain's RAHS contains 3 complete solo delta elements, a tRNA^{lys} gene and an ORF of unknown function.
B = BglII, Ba = BamHI, C = ClaI, E = EcoRI, H = HindIII, P = PvuII, S = SalI, X = XhoI.

XhoI

```
CTCGAGAAAT CGGACATTAA AACGGTGCAA ATACTGTCGA GAAATTATAA CTTACCAAGT CAAAAAGCGA ATGAGAATCT ATTGTTGCGA TTTTTTTCGT
     10         20         30         40         50         60         70         80         90        100

TACTACTTCT TACCCGAACT AACTTTACAT AATAAAAAAG CATGAAGCGC ATATTCAGCG ACTCAAAAAT CCCAAACTGG CACTTCTCTA TATACCGATG
    110        120        130        140        150        160        170        180        190        200

AGCATATTTT GATTGTAAAA AACAACATTA AAAAGAACTG CAAACCTGAA CCAAGAATAC TTAATTTCAA AACCGAGAAA TGAGGTACAA CTTGTAGAAG
    210        220        230        240        250        260        270        280        290        300

TTCACTTAAA CTTATTGAGC ATACAATCTC AACAAAACTA ACTGCAAAGT TTCTTGCTAT AGTCACATAG TAATACTAGC AGGAAGCCAT TAATCAAAAA
    310        320        330        340        350        360        370        380        390        400

GAAACTCCAT ATAAAACGTA ATAAATATTA TGAATAATAG TACAATATAA GAACAAAGAA GAAGAAAATA CATTATAACT GAAGTGCCAA TATTAGAAAA
    410        420        430        440        450        460        470        480        490        500
                                                               ┌─→ end ORF
AAAAAGGTGG AATATCCTAG TGGAAAAATA TCATCTATGC GGTGAAAGAA|TTAAATAGAT A|CTATGGATT GTTTCGAATC TATCCTTTTA TCTGCCCCTT
    510        520        530        540        550        560        570        580        590        600

GCCTTTTTTT TTCATTCTCC TTCTTCAGTC TTTTTCCATA GGCGAGTCCA CAGGGACTAC AAAGTTCCCT GTTTTGGTCG GGTCCACTAC GCCATTGAAT
    610        620        630        640        650        660        670        680        690        700

TGTCCAAGTG TCTTTACATT TACAGCCATTC CTTTATCTCA CCATTTGGGA GCATTGTTTG TCTCTTTTCT CTTTTAGGAC GACTAGTAGT CTTTTTGCTG
    710        720        730        740        750        760        770        780        790        800

TATTTCTTCG TAATACTGGC AGACCGACTA GTTCCTTGTG CACAACTATG GTAGCCTTCA AGATTGTTTG CTTCTTGGCT AGTCGTACGT CTCTCGGAAT
    810        820        830        840        850        860        870        880        890        900
                                        ↓Tyl-161 insertion site                                BglII
TGCATGCACT AATAACACGA TTCTTTGACG GTGTAATTTC GTTCAAAGAG CTGGTTTTGT TCTTAACAAT ATTTCTTGTA AGATCTCTTA GTTGTTCGGC
    910        920        930        940        950        960        970        980        990       1000

ATTCTCAATT TTTTTTTTCAA CATCAACAGC CCTTATTTGA GAAAAGGATA CTTCAATAAA TGAGCAACTT GCATAAACTC TTCCGCGTTC CTTATATGCC
   1010        1020       1030       1040       1050       1060       1070       1080       1090       1100

TTCTTTTTAT CGTTTGACTT CCAGTCATCA ATTTCTCTAT TTAATAATAA CGTTTCTCC ATATCTGACT TTTTAACTAC TCTAGTTAAG AAGGAACTTC
   1110       1120        1130       1140       1150       1160       1170       1180       1190       1200
       start ORF  ←────────────┐
CAGAGTTGTC ATAATTATTA TATCGCATAC TAGCAATAGT AAATGTCTGG TTATTTACCT AGCAACCTCT TAACAAAGCG ATAACAATGA AACAGCAATA
   1210       1220       1230        1240       1250       1260       1270       1280       1290       1300

TAAGTCTATT TAATTGACAA CATCATTGAT AAAAGTTTGA ACAAGCCGTCT AAAAGTCGAA GAAGTTCATT TCAAAGTTTT GTGTATCGAG CTTGATGATA
   1310       1320       1330       1340       1350       1360       1370       1380       1390       1400
         BamHI
CTTATTTAGG GATCCGCTCT AGTTTACTGT GATACAACTG AAAAAACATA CAAAGCTATT GCGCCTGCAT TCTCATATTA GTGGCGTATG TTTGTGTCAC
   1410       1420       1430       1440       1450       1460       1470       1480       1490       1500
                                                                                                 EcoRI
ATTCGATACT TCTACATAAT ACAATGACTG AACCCTTACC GCTGGCACAG CAATACGCGC CGCGAAAGCA GGGTCTAGGA GAAACAATAG AATTCTTGGC
   1510       1520       1530       1540       1550       1560       1570       1580       1590       1600

GATGAGAGAG GGGAAGGCTT TAATAAAAAA AAATAGCCGCT TATCCGATTC TGTTGGAATA AGTGGCTATC AATATTTACA TGTTGATATG CTAGTAAGGA
   1610       1620       1630       1640       1650       1660       1670       1680       1690       1700

TATAAATGCT TATATTGAAT ATTTGGAATA AGCACTAACT ATCACTTATT CAAAAGGACT CGAACAAAAA AATATGAAAT GTCGAAAGCC ACATCACTGA
  5 1710       1720       1730       1740       1750       1760       1770       1780       1790       1800
                                                        solo delta 31C-GF1
GAAGT|TGTTG GATTGAAGCT CTAAAATGTC ATCTTTTAGT AGTATTCGTG TTACTAGTAT ATTATCACAT GCGGTGTAAA AAGATGACGT AAAGATTGAG
   1810       1820       1830       1840       1850       1860       1870       1880       1890       1900

AAACAGTCAT CAACTGTTAT GGAAGCTGAA ATGCAAATAT TGATAATGTA ATAGGATAAT GAAACGTATA AAATGAAAAA GAATAATAAA ATATTATTAT
   1910       1920       1930       1940       1950       1960       1970       1980       1990       2000
                 EcoRI
GTAGAAATAC CGATTCCATT TTGAGAATTC CTATATCCTT GAGAAGAACT TCTAGTATAT TCTGTATACA TAATATTACT GCCTTTATCA GACATGGAAT
   2010       2020       2030       2040       2050       2060       2070       2080       2090       2100
                          3 ↓
GCCAAAAATT ATCAAAATAT CACCCAATAT TCA|TTTTAT TGTTGTGAAG ACATCAAGCC CTAGATATAT ATCATAGTAG GATAAAATAG TAGAATAAAA
   2110       2120       2130       2140       2150       2160       2170       2180       2190       2200

CTATTCATAT ATCTACGTGA AAGCATCAGA GATTGTAATG TTGAGCGCCG TGCCGATGGA AGCGCGCAGG GCTCATAACC CTGATGTCCT ATCGAAACCG
   2210       2220       2230       2240       2250       2260       2270       2280       2290       2300

AGCGGCGCTA TTAAATTTTT TGAACTGTCC ATTAATAATA ATGGAACTGA CCTCGTCAAT TACCGGTAGC TTTTTTGCCC ACATTTTAA TCATCTTTTA
   2310       2320       2330       2340       2350       2360       2370       2380       2390       2400

AGTATTTGAG TATTTGTTGT TCACTAGAGT TTTACTAATT AAGGTTCTGA GAAAATAAGA TATGCATTTA GAAAAGTCGG TTCTACACCG CAAATTTCTT
   2410       2420       2430       2440       2450       2460       2470       2480       2490       2500
                                       5 ↓2440                 2470
                                                       solo delta 31C-GF2
TATTACAGAA ACATTTATAC AAGAGGATAC ATAGT|ATTG GAAAGAAAAT CAATTATCTT TTTGTTGGGT AGTATGAATA TTACTTGTAC ATTATCATAT
   2510       2520       2530       2540       2550       2560       2570       2580       2590       2600

ACAATATAAA AAGTGTCATA GAGTTGTGAA ATATTTATTA TATATTAATA AAAGCTGAAG CGCAATGATT GAAAACGTTT TAGGGTGATT AATGATAAAG
   2610       2620       2630       2640       2650       2660       2670       2680       2690       2700

TACAAAATGA CGAGAAGAAA TAATAATAAC ATAAAGATGG AAATACCCGT TATCTTTTAA GGATATTATA ACCTGGAGGA AAACTTCGTG GGTTCTCTGT
   2710       2720       2730       2740       2750       2760       2770       2780       2790       2800
                                                            3 ↓
ATATTCAACA CTTTTCCGTT ATCAAAAATA GAATCCCAAC AGTATCCCCA TTTTCGTCAA TATCTCCATA ACGACAAGGT CTTCTAGATA CCATATAATG
   2810       2820       2830       2840       2850       2860       2870       2880       2890       2900

ACAATAACCA AGCCACCATT TAGGTGGATT TTTTGTGTGG TTTTGAAATT AAAACATTTT TAGAGGAAAA TTGTGATCAC AGAATTTCAC ATTCCCAGTT
   2910       2920       2930       2940       2950       2960       2970       2980       2990       3000

ACTTTAATAA TATTCATAGT TAGAGATGAA GAGGTACTAT TAATATGAAG CTTAAGTATG AATCTTCTCA TGCTTCTATA AGCGAGGGAT ATAAAACATC
   3010       3020       3030       3040       3050       3060       3070       3080       3090       3100

GCCAACTGTT TCATCTTATT TATTATGTCAAT ATTATGTGTC TTTGGAGCTT ACTATACAGT TAATGTGATA ATTAATATAA CTCTTGTATA ACTACCATAT
   3110       3120       3130       3140       3150       3160       3170       3180       3190       3200
```
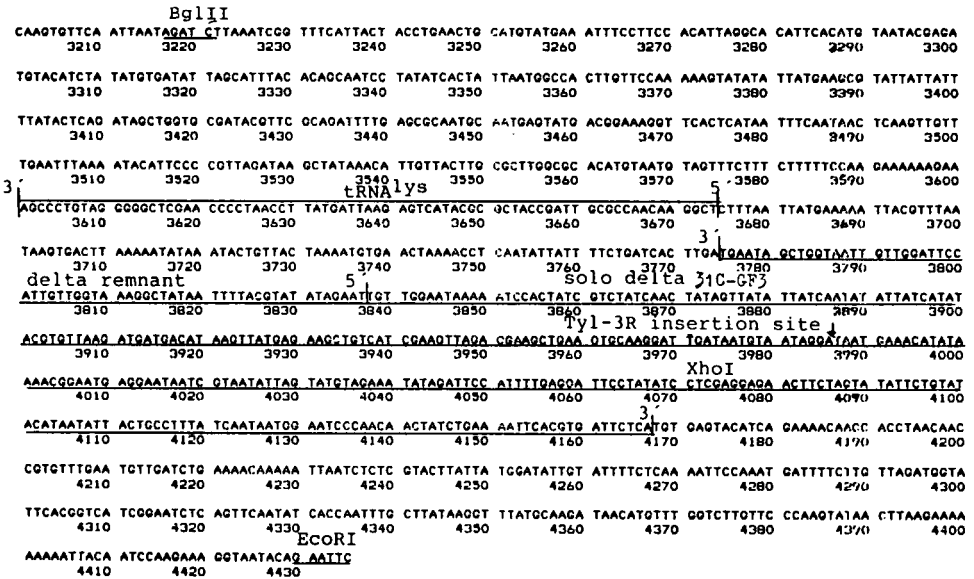
BglII
CAAGTGTTCA ATTAATAGAT CTTAAATCGG TTTCATTACT ACCTGAACTG CATGTATGAA ATTTCCTTCC ACATTAGGCA CATTCACATG TAATACGAGA
       3210   3220   3230   3240   3250   3260   3270   3280   3290   3300

TGTACATCTA TATGTGATAT TAGCATTTAC ACAGCAATCC TATATCACTA TTAATGGCCA CTTGTTCCAA AAAGTATATA TTATGAAGCG TATTATTATT
       3310   3320   3330   3340   3350   3360   3370   3380   3390   3400

TTATACTCAG ATAGCTGGTG CGATACGTTC GCAGATTTTG AGCGCAATGC AATGAGTATG ACGGAAAGGT TCACTCATAA TTTCAATAAC TCAAGTTGTT
       3410   3420   3430   3440   3450   3460   3470   3480   3490   3500

TGAATTTAAA ATACATTCCC CGTTAGATAA GCTATAAACA TTGTTACTTG CGCTTGGCGC ACATGTAATG TAGTTTCTTT CTTTTTCCAA GAAAAAAGAA
3′   3510   3520   3530   3540   3550   3560   3570   5′ 3580   3590   3600
                                            tRNAlys
AGCCCTGTAG GGGGCTCGAA CCCCTAACCT TATGATTAAG AGTCATACGC GCTACCGATT GCGCCAACAA GGCTCTTTAA TTATGAAAAA TTACGTTTAA
     3610   3620   3630   3640   3650   3660   3670   3′ 3680   3690   3700
                                                                              3′
TAAGTGACTT AAAAATATAA ATACTGTTAC TAAAATGTGA ACTAAAACCT CAATATTATT TTCTGATCAC TTGATGAATA GCTGGTAATT GTTGGATTCC
     3710   3720   3730   3740   3750   3760   3770   3780   3790   3800
delta remnant   5′   solo delta 31C-GF3
ATTGTTGGTA AAGGCTATAA TTTTACGTAT ATAGAATTGT TGGAATAAAA ATCCACTATC GTCTATCAAC TATAGTTATA TTATCAATAT ATTATCATAT
     3810   3820   3830   3840   3850   3860   3870   3880   3890   3900
                                                                   Tyl-3R insertion site
ACGTGTTAAG ATGATGACAT AAGTTATGAG AAGCTGTCAT CGAAGTTAGA CGAAGCTGAA GTGCAAGGAT TGATAATGTA ATAGGATAAT GAAACATATA
     3910   3920   3930   3940   3950   3960   3970   3980   3990   4000
                                                                        XhoI
AAACGGAATG AGGAATAATC GTAATATTAG TATGTAGAAA TATAGATTCC ATTTTGAGGA TTCCTATATC CTCGAGGAGA ACTTCTACTA TATTCTGTAT
     4010   4020   4030   4040   4050   4060   4070   4080   4090   4100

ACATAATATT ACTGCCTTTA TCAATAATGG AATCCCAACA ACTATCTGAA AATTCACGTG ATTCTCATGT GAGTACATCA GAAAACAACC ACCTAACAAC
     4110   4120   4130   4140   4150   4160   4170   4180   4190   4200
                                                                   3′
CGTGTTTGAA TGTTGATCTG AAAAACAAAAA TTAATCTCTC GTACTTATTA TGGATATTGT ATTTTCTCAA AATTCCAAAT GATTTTCTTG TTAGATGGTA
     4210   4220   4230   4240   4250   4260   4270   4280   4290   4300

TTCACGGTCA TCGGAATCTC AGTTCAATAT CACCAATTTG CTTATAAGGT TTATGCAAGA TAACATGTTT GGTCTTGTTC CCAAGTATAA CTTAAGAAAA
     4310   4320   4330   4340   4350   4360   4370   4380   4390   4400
                              EcoRI
AAAAATTACA ATCCAAGAAA GGTAATACAG AATTC
     4410   4420   4430

Figure 3. Nucleotide sequence of the RAHS from strain CN31C. The region of plasmid 31C-GF sequenced is that shown in Fig. 2c. The solo delta sequences are underscored and the tRNA^{lys} sequence is overscored.

showed to be due to the presence of solo delta sequences. Prolonged exposure of Southern blots from CB15A-2D and 16B-GF did not reveal any equivalent bands. BglII-EcoRI double digests of 31C-GF produced a 0.6 kb fragment which was not seen in similar digests of the other two plasmids. Further analysis, using this fragment as a probe, demonstrated that 31C-GF contains a 3 kb region of chromosomal DNA which is absent in both CB15A-2D and 16B-GF. Detailed maps for the RAHS region of all three inserts are presented in Fig. 2.

The RAHS from strain CN31C contains three solo delta elements and a tRNA gene

The sequence presented in Fig. 3 represents 4431 bp from the RAHS region of the chromosome III of strain CN31C obtained by sequencing part of the 21 kb insert in plasmid 31C-GF. While this region lacks a complete Ty element, use of the DIAGON dot matrix comparison program of Staden (25) revealed the presence of three delta sequences. These are underscored in Fig. 3 and are also marked on the restriction map shown in Fig.2. None of these three solo delta elements is flanked by a 5bp chromosomal 'target site' duplication (3). The first delta element, 31C-GF1 (bp 1806-2133) is flanked by the sequences 5'GAAGT 3' and 5'ATTTT 3' the second, 31C-GF2 (bp 2536-2859) by 5'ATAGT 3' and
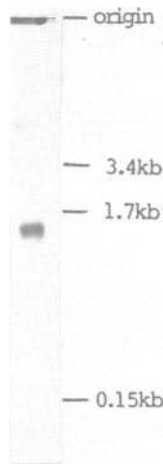
Figure 4. Autoradiograph of a Northern blot of total yeast RNA prepared from strain CN31C, probed with the 1.4 kb XhoI-BamHI fragment from plasmid 31C-GF (Fig. 2c and 4).

5'ATATC 3'; and the third, 31C-GF3 (bp 3838-4167) by 5'AGAAT 3' and 5'TGTGA 3'. This last solo delta is flanked on its centromere-proximal (left) side by the last 64 bp of another delta element arranged in the opposite orientation. This supports the idea that there has been considerable delta-delta interaction in this region via either transposition or recombination. Such delta-delta rearrangements probably explain the absence of chromosomal target site duplications at the ends of these elements. This absence might be the result of sequence degeneration, but this seems unlikely in the case of 31C-GF1 and 31C-GF2 both of whose sequences are highly conserved.

The LAHS region is flanked by the 5'ends of two tRNA genes (7) and others (5,6) have pointed out an association between Ty elements and such genes. A search for tRNA genes in the 31C-GF sequence revealed a tRNA^{Lys} at bp 3601 to 3674 (overscored in Fig. 3). This gene is embedded in a very A + T rich domain which is abutted by the 64 bp delta remnant 5' to the tRNA gene (see Fig. 3). This arrangement is consistent with the previously described association between yeast transposons and transfer RNA genes.

The RAHS from 248 contains a transcribed gene of unknown function

A search of the 31C-GF sequence for potential protein-encoding genes revealed a single open reading-frame (ORF) of 222 amino acids between bp 551 and 1228 and oriented towards the centromere (see Fig. 3). The sequence context of this ORF encourages the belief that it is expressed. Putative TATA

boxes occur at positions 1300 and 1330. Several CT tracts, which are associated with a number of yeast promoters (26), are found downstream of these TATA boxes. The 3' flanking region of the ORF contains two sequences which have been suggested to be yeast transcriptional terminators: first, TAGT at position 377 followed by a T-rich stretch and a TTT at position 309 (ref. 27); second, the sequence TTTTATA is found at position 410 (ref. 28). A Northern blot of total yeast RNA using a XhoI-BamHI fragment (bp 1 to 1410) from 31C-GF probe confirmed the transcription of this putative gene, revealing a transcript of ca. 1.0 kb in length (Fig. 4). The codon bias of the ORF is typical of a poorly expressed yeast gene (29). A search of the Doolittle (30) and PIR (31) protein sequence data banks failed to identify any known proteins with significant homology to the 31C-GF ORF.

Sequence analysis of the RAHS from GRF88 which contains two Ty elements

The restriction enzyme and Southern blot analysis described earlier demonstrated that GRF88 contains a deletion of chromosomal sequences in the RAHS region of chromosome III. In order to determine the nature of the deletion event it was necessary to determine the sequences flanking the tandem Ty insert in CB15A-2D and to compare them with the 31C-GF sequence.

The 1.4 kb XhoI-PvuII fragment containing the left (centromere- proximal) delta elements of Ty1-161 was isolated, end-repaired using Klenow polymerase and the resulting blunt ended fragment cloned into the SmaI site of M13mp18. The 'Bandaid' technique (21) was then used to sequence this DNA in both orientations. The sequence obtained is presented in Fig. 5a. It demonstrates that Ty 1-161 is not a variant class I element but is of normal length. The left delta of Ty 1-161 (d_1 in Fig. 2) has lost its characteristic XhoI site and the XhoI site that was previously inferred to lie within the left delta is in fact contained within unique chromosome III DNA some 940 bp upstream of the 5' end of the left delta of Ty1-161. This delta is in the middle of the 31C-GF reading-frame at position 939 as recorded in Fig. 3.

The junction between the tandem Ty elements was investigated by isolating it on a 2.6 kb SalI fragment (see Fig. 2), and then sub-cloning it as a 0.5 kb BglII - PvuII fragment into M13mp18 and mp19 cleaved with SmaI and BamHI. This sequence was confirmed using two further sub-clones of the SalI fragment which were obtained by XhoI digest. Fig. 5b gives 790 bp of sequence from the junction region and shows that Ty1-161 and Ty1-3R are joined by a single delta sequence (d_2 in Fig. 2a) of 322 bp with no intervening chromosome III DNA.

The centromere-distal junction between the tandem pair of Ty elements and

a

XhoI
```
CTCGAGAAAT CGGACATTAA AACGGGTGCAA ATACTCTGGA GAAATTATAA CTTACCAAGT CAAAAAGCCGA ATGAGAATCT ATTGTTGCGA TTTTTTTTCGT
         10         20         30         40         50         60         70         80         90        100

TACTACTTCT TACCCGAACT AACTTTACAT AATAAAAAAG CATGAAGCGC ATATTCAGGG ACTCAAAAAT CCCAAACTCG CACTTCTGTA TATACCGATG
        110        120        130        140        150        160        170        180        190        200

AGCACTACTTT GATTGTAAAA AACAACATTA AAAAGAACTG CAAACCTGAA CCAAGAATAC TTAATTTCAA AACCGAGAAA TGAGGTACAA CTTGTAGAAG
        210        220        230        240        250        260        270        280        290        300

TTCACTTAAA CTTATTGAGC ATACAATCTC AACAAAACTA ACTGCAAAGT TTGTTGCTAT AGTCATATAG TAATACTAGC AGGAAGCCAT TAATCAAAAA
        310        320        330        340        350        360        370        380        390        400

GAAACTCCAT ATAAAACGTA ATAAATATTA TGAATAATAG TACAATATAA CAACAAACAC GAAGAAAATA CATTATAACT GAAGTGCCAA TATTAGAAAA
        410        420        430        440        450        460        470        480        490        500
```
end truncated ORF →
```
AAAAAGGTCG AATATCCTAG TGGAAAAATA TCATCTATGC GGTGAAAGAA TTAAATAGAT ACTATGGATT GTTTCGATCT ATCCTTTTAT CTGCTGCTTG
        510        520        530        540        550        560        570        580        590        600

CCTTTTTTTT TCATTCTCCT TCTTCAGTGT TTTTCCATAG GCCAGTCCAC AGCGACTACA AAGTTCCCTG TTTTGGTCGG GTCCACTACG CCATTGAATT
        610        620        630        640        650        660        670        680        690        700

GTCCAAGTGT CTTTACATTT AGAGCATTCC TTTATCTCAC CATTTGGGAG GATTGTTTGT CTCTTTTCTC TTTTAGGACG ACTAGTAGTC TTTTTGCTGT
        710        720        730        740        750        760        770        780        790        800

ATTTCTTCGT AATACTGGCA GACCGACTAG TTCCTTGTGC ACAACTATCG TAGCCTTCAA GATTGTTTGC TTCTTGGCTA GTCGTACGTC TCTCGGAATT
        810        820        830        840        850        860        870        880        890        900
```
SphI   5' start ORF ←   Tyl-161 left delta
```
GCATGCACTA GTTACACGAT TCTTTGAGGG TGTAATTTTG TTGGAATAAA AATCAACTAT CATCTACTAA CTAGTATTTA CGTTACTAGT ATATTATCAT
        910        920        930        940        950        960        970        980        990       1000

ATACGGTGTT AGAAGATGAC GCAAATGATG AGAAATAGTC ATCTAAATTA GTGGAAGCTG AAACGCAAGG ATTCGATAATG TAATAGGATC AATGAATATT
       1010       1020       1030       1040       1050       1060       1070       1080       1090       1100

AACATATAAA ACGATGGATAA TAATATTTAT AGAATTGTGT AGAATTGCAG ATTCCCTTTT ATGGATTCCT AAATCCTTGA GGAGAACTTC TAGTATATTC
       1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
```
3'   epsilon region
```
TGTATACCTA ATATTATAGC CTTTATCAAC AATGGAATCC CAACAATTAT CTCAACATTC ACCCATTTCT CATGGTACCG CCTGTGCTTC GGTTACTTCT
       1210       1220       1230       1240       1250       1260       1270       1280       1290       1300

AAGGAAGTCC ACACAAATCA AGATCCGTTA GACGTTTCAG CTTCCAAAAC AGAAGAATGT GAGAAGGCTT CCACTAAGGC TAACTCTCAA CAGACAACAA
       1310       1320       1330       1340       1350       1360       1370       1380       1390       1400
```
PvuII
```
CACCTGCTTC ATCAG
       1410
```

b
```
AAATTTAGAA ACAGATTTTT TGGCACAAAG GCAATGAGAC TTAGAGATGA ACTATCAGCT AATAATTTAT ACGTATACTA CATCGAGACC AAGAAGAACA
         10         20         30         40         50         60         70         80         90        100
```
                                                                    Tyl-161 epsilon region
                                                                    BglII
```
TTGCTGATGT GATGACAAAA CCTCTTCCGA TAAAAACATT TAAACTATTA ACAAACAAAT GGATTCATTA GATCTATTAC ATTATGGGTG CTATGTTCGA
        110        120        130        140        150        160        170        180        190        200
```
                                     middle delta   5'
```
ATAAAAATCA ACTATCATCT ACTAAACTAGT ATTTACGTTA CTAGTATATT ATCATATACG GTGTTAGAAG ATGACGCAAA TGATGAGAAA TAGTCATCTA
        210        220        230        240        250        260        270        280        290        300

AATTAGTGGA AGCTGAAACG CAAGGATTGA TAATGTAATA GGATCAATGA ATATTAACAT ATAAAACGAT GATAATAATA TTTATAGAAT TGTGTAGAAT
        310        320        330        340        350        360        370        380        390        400
```
XhoI
```
TGCAGATTCC CTTTTATGGA TTCCTAAATC CTCGAGGAGA ACTTCTAGTA TATCTACATA CCTAATATTA TAGCCTTAAT CACAATCCAA TCCCAACAAT
        410        420        430        440        450        460        470        480        490        500
```
3'   Tyl-3R epsilon region
```
TACATCAAAA TCCCACATTCT CTACATGGTA GCGCCTCGTGC TTCCGTTACT TCTAAGGAAG TCCACACAAA TCAAGATCCG TTAGACGTTT CAGCTTCCAA
        510        520        530        540        550        560        570        580        590        600
```
                                                                      PvuII
```
AATTCAAGAA TATGATAAGG CTTCCACTAA GGCTAACTCT CAACAGACAA CAACACCTGC TTCATCAGCT GTTCCACAGA ACCCCCATCA TGCCTCTCCT
        610        620        630        640        650        660        670        680        690        700

CAACCTGCTT CAGTACCACC TCCACACAAT GGGCCGTACC CACACACTG CATGATGACC CAAAACCAAG CCAATCCATC TGGTTGGTCA TTT
        710        720        730        740        750        760        770        780        790
```

c
```
ACAAGAAACC AATTACTAAA GGATTATTAA CCGACAGTAA ATCTACAATC AGTATAATTA TATCCAATAA TGAAGAGAAA TTTAGGAACA GATTTTTTGG
         10         20         30         40         50         60         70         80         90        100
```
                                        Tyl-3R epsilon region
```
TACTAAAGCA ATGAGATTGA GAGATGAAGT ATCAGGGAAAT CATCTGCACG TATGCTATAT CGAAACCAAA AAGAATATTG CAGACGTAAT GACCAAACCT
        110        120        130        140        150        160        170        180        190        200
```
                                                BglII   5'
```
CTTCCGATAA AAACATTCAA ACTATTAACA AACAAATGGA TTCATTAGAT CTATTACATT ATGGGTGCTA TGTTGGAATA AAAATCCACT ATCGTCTATC
        210        220        230        240        250        260        270        280        290        300
```
Tyl-3R right delta   ClaI
```
AACTAATAGT TATATTATCA ATATATTTCA ATATACGTGT TAAGATGATG ACATAAGTTG TGAGAAGCTG TCATCGGATGT TAGAGGAAGC TGAAACGCAA
        310        320        330        340        350        360        370        380        390        400

GGATTGATAA TGTAATAGGA TCAATGAATA TAAACATATA AAACGGAATG AGGAATAATC GTAATATTAG TATGTAGAAA TATAGATCC ATTTTGAGGA
        410        420        430        440        450        460        470        480        490        500
```
XhoI
```
TTCCTATATC CTCGAGGAGA ACTTCTAGTA TATTCTGTAT ACCTAATATT ATAGCCTTTA TCAACAATGG AATCCCAACA ATTATCTAAT TACCCACCTA
        510        520        530        540        550        560        570        580        590        600
```
3'   delta 31C-GF3 remnant
```
TATCTCATAA TGAAACATAT AAAACGGAAT GAGGAATAAT CGTAATATTA GTATGTAGAA ATATAGATTC CATTTTGAGG ATTCCTATAT CCTCGAGGAG
        610        620        630        640        650        660        670        680        690        700
```
                                                                                                 XhoI
```
AACTTCTAGT ATATTCTGTA TACATAATAT TACTGCCTTT ATCATAATG GAATCCCAAC AACTATCTGA AAATTCACGT GATTCTCATG TGAGTACATC
        710        720        730        740        750        760        770        780        790        800
```
                                                                                        3'
```
AGAAAACAAC CACCTAACAA CCGTGTTTCA ATGTTGATCT GAAAACAAAA ATTAATCTCT CGTACTTATT ATGGATATTG TATTTTCTCA AAATTCCAAA
        810        820        830        840        850        860        870        880        890        900

TGATTTCTT GTTAGATGGT ATTCACGGTC ATCGGAATCT CAGTTCAATA TCACCAATTT GCTTATAAGG TTTATGCAAG ATAACATGTT TGGTCTTGTT
        910        920        930        940        950        960        970        980        990       1000
```
                                                                EcoRI
```
CCCAAGTATA ACTTAAGCAAA AAAAAAAATTA CAATCCAACGA AAGGTAATAC AGAATTC
       1010       1020       1030       1040       1050
```

chromosomal DNA was isolated on a 2.7 kb EcoRI fragment whose sequence is shown in Fig. 5c. It can be seen that the epsilon sequence of Ty 1-3R terminates in a complete delta sequence of 327 bp (d_3 in Fig. 2a). This delta sequence is itself flanked by the last 180 bp of another delta sequence (Rd in Fig. 2a), suggesting that Ty1-3R may originally have been incorporated into chromosome III by transposition into, or recombination with, a pre-existing delta sequence. The 180 bp delta fragment is joined to unique chromosome III DNA at its right end and the sequence is identical to that beginning at position 3986 of 31C-GF. Thus strain GRF88 contains a deletion that extends from nucleotide 939 through 3986 of the 31C-GF sequence shown in Fig. 3 (see Discussion).

The RAHS from the CF4-16B chromosome III which contains a single Ty element

The 0.5 kb SphI - PvuII fragment from 16B-GF that contains the left junction of Ty1-16B was cloned into SphI, SmaI digested M13 mp18 and mp19 and sequenced. It was found that this fragment had exactly the same nucleotide sequence as the equivalent fragment from CB15A-2D and contained 3C bp of chromosome III DNA from the SphI site to the 5' end of the left delta sequence of Ty1-161, the Ty1-161 left delta itself and the first 140 bp of the Ty1-161 epsilon up to the PvuII site. The 0.8 kb BglII - EcoRI fragment that contains the right junction of Ty1-16B was cloned into BamHI, EcoRI digested M13mp18 and mp19 and sequenced. Again, the nucleotide sequence of this fragment was found to be identical to that of the equivalent fragment from CB15A-2D. From the BglII site it contains 20 bp of the 3' epsilon region of Ty1-3R, the Ty1-3R right delta sequence (d_3 in Fig. 2a and b), the delta fragment and then unique chromosome III DNA exactly as shown in the CB15A-2D sequence given in Fig. 5c.

The restriction mapping and sequence data indicated that the left delta and epsilon region of Ty1-16B, up until at least the PvuII site, are derived from Ty1-161 (the left Ty element of CB15A-2D). The right end of Ty1-16B, from at least the HindIII site, appears to be derived from Ty1-3R (the right Ty element of the tandem pair in CB15A-2D).

Figure 5. Nucleotide sequence from the RAHS of GRF88 contained in plasmid CB15A-2D.
a) The left delta (d_1) of the Ty1-161 element and its flanking sequence. Left most XhoI-PvuII fragment in Fig. 2a.
b) The middle delta (d_2) and the flanking Ty epsilon sequences (Fig. 2a).
c) The right delta element of Ty 1-3R (d_3) and its flanking sequences, up to the right most EcoRI site in Fig. 2a.
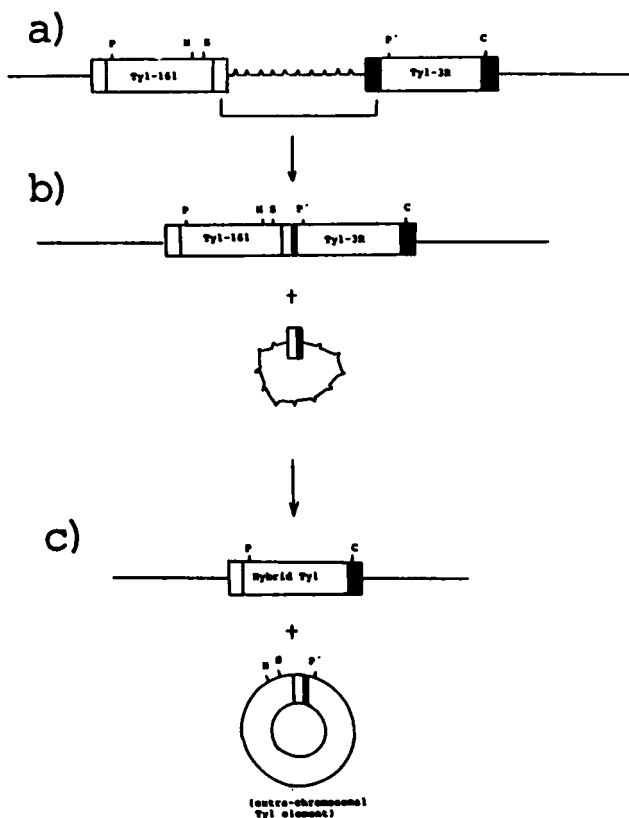
Figure 6. A mechanism to explain the generation of the three Ty element associated polymorphisms characterised.
a) The earliest structure is assumed to be that of the RAHS contained in plasmid 16B-GF from strain CN31C. Two independent transposition events inserted Ty 1-161 and Ty 1-3R into the CN31C RAHS. The two Ty elements are separated by ca. 3 kb of DNA unique to plasmid 16B-GF.
b) A recombination event between the two proximal delta elements brings Ty 1-161 and Ty 1-3R together. This forms the tandem Ty element structure characterised in plasmid CB15A-2D, containing DNA from the RAHS of strain GRF88. The ca. 3 kb of intervening DNA is lost from the RAHS of GRF88.
c) Recombination between the epsilon regions of Ty 1-161 and Ty 1-3R generates the hybrid Ty1 element (Ty 1-16B) found in the RAHS of strain CF4-16B. This event could occur from either stage a) when the Ty1 elements are separated or at b) where they are in tandem array.

DISCUSSION

The exact relationship between the three strains of S. cerevisiae whose chromosome III sequences we have analysed in the RAHS region is unknown;

```
                10      20      30      40      50      60      70      80      90     100
1 ..................A......A....CT....Q..T...CQ...CT.Q....................-..A......GC..A.Q......ATA
2 ..................A......A....CT....Q..T...CQ...CT.Q....................-..A......GC..A.Q......ATA
3 ..................C......Q....TC....A..Q...TA...TC.A....................-........T......AT..Q.T......GCT
consensus TQTTCGAATAAAAATC-ACTATC-TCTA--AACTA-TA-TTA--TTA--A-TATATTATCATATACQGTQTTAAGA-GATGAC--AA-T-ATGAGAA---

               110     120     130     140     150     160     170     180     190     200
......T.AA.....T................................................T..............--.....-T......AT.T...GA.
......T.AA.....T................................................T..............--.....-T......AT.T...GA.
......Q.TG.....A...............................................A..............Q......CQ.A..T...TT.
GTCATC-A--TTAQ-GGAAQCTGAAACGCAAGGATTGATAATGTAATAGG ATCAATGAATAT-AACATATAAAACGGAATGAG-AATAAT--T-ATA--A

               210     220     230     240     250     260     270     280     290     300
T.G.......T.GC.......C....AT.........A.....T.......................GT.................T.........
T.G.......T.GC.......C....AT.........A.....C...................-..AC.................A..-...
G.A........A.AT.......A....GA.........T.....C......................GT.................T.........
-T-TGTAGAA-T--AGATTCC-TTTT---GGATTCCTA-ATCCT-GAGGAG AACTTCTAGTATATTCT--ATACCTAATATTATAGCCTT-ATCAACAATG
                                       ‾‾‾‾‾‾‾‾‾‾‾‾ cross-over

               310     320     330
..............TC.C.AC.TT...CCATTTCT..
..............CA.C.AA.TC...ATTCTCTA..
..............TC.A.TT.CC...CTATATCT..
GAATCCCAACAATTA--T-A--A--CAC-------CA
```

Figure 7. Alignment of the delta element sequences from the three Ty element deltas from the RAHS of GRF88 (Fig. 2a).
1 = d_1 (31C-GF1); 2 = d_2 (31C-GF2) and 3 = d_3 (31C-GF3) Nucleotides differing between one or more of the delta elements are shown. The consensus delta sequence is marked below. The probable position of the recombination event between the proximal delta sequences of Ty 1-161 and Ty 1-3R is underscored.

however, the sequence data obtained suggests the following evolutionary relationship.

Strain CN31C contains no complete Ty elements in the RAHS region of chromosome III. Nevertheless the presence of three solo delta sequences in the CN31C RAHS, none of which is bounded by a target site duplication and one of which is inserted into a delta fragment, suggests that even in this strain the RAHS has been the site of a number of successive transposition events followed by delta-delta recombinations. The CN31C RAHS contains two functional genes, the 31C-GF ORF and tRNA[lys]. We propose the gene name CTR1 for the ORF, standing for chromosome three reading-frame.

The RAHS from strain GRF88 can be considered as being derived from strain 248 via two independent transposition events. One transposition inserted Ty1-161 into chromosome III at the position numbered 939 in the 31C-GF sequence (Fig. 3), meaning that the element was actually inserted into the coding sequence of CTR1. This may be considered an unusual event given the reported preference for Ty elements to insert into promoter sequences (4). However, we would reiterate that we have examined unselected events whereas previous workers have investigated Ty insertions associated with particular mutations and so may have produced a biased picture. The second transposition into the CN31C RAHS inserted Ty1-3R into a pre-existing solo delta sequence at position 3986 in the H9G sequence (Fig. 3). Our sequence analysis of the LAHS region of chromosome III (ref. 7) had suggested that delta sequences may themselves be preferred targets for transposition events and Ty1-3R provides

further support for that hypothesis. Our data (this paper and ref. 7) and that of others (3,3) delimits the delta transposition target as lying between base pairs 150 and 230. This region contains the delta element promoter and provides a further example of Ty elements inserting into yeast promoter sequences.

After the two transposition events, the RAHS would have contained two class I Ty elements (Ty1-161 and Ty1-3R) separated by ca. 3 kb of unique chromosome III DNA containing part of CTR1, two solo delta sequences and the tRNA^{lys} gene. A recombination event bewteen the right delta sequence of Ty1-161 and the left delta sequence of Ty1-3R could then have occurred fusing the two Ty elements via a single, central, delta sequence and producing the tandem Ty pair found in the RAHS of strain GRF88. Fig. 6 diagrams this sequence of events. A comparison of the sequences of the three deltas contained within that tandem Ty pair (Fig. 7) suggests that the recombination event between the proximal deltas of Ty1-161 and Ty1-3R occurred between base pairs 238 and 244, close to or at the transcription start site. In addition to fusing the two Ty elements, this recombination event also resulted in the deletion of ca. 3 kb of chromosome III DNA. The consequences of this deletion will be considered after the derivation of the RAHS from strain CF4-16B has been explained.

CF4-16B contains a single Ty element, Ty1-16B in its RAHS region. This Ty appears to be the product of a recombination event between the epsilon regions of the tandem Ty pair contained within the chromosome III of GRF88. Strain CF4-16B represents, therefore, the end-point of our proposed evolutionary series. The restriction mapping and sequence analysis of Ty 1-16B presented in the RESULTS section strongly suggests that a recombination event occurred between Ty1-161 and Ty1-3R in the region between the first PvuII site in the epsilon sequence of Ty1-161 and the following HindIII site. We believe this is the first example of an epsilon-epsilon recombination event to be reported. Such recombinations are difficult to detect especially if, as in this case, they occur between two members of the same class of Ty elements. However, a number of factors suggest that epsilon-epsilon recombinations may indeed be rare events. For instance, no element with a restriction pattern which indicates that it is a hybrid between a class I and class II Ty has been reported. Epsilon- epsilon recombinations most frequently result in deletion of chromosomal DNA or produce a translocation. Such events are likely to be selected against and yeast may even have some positive means of repressing them. The epsilon-epsilon recombinations we reported here occurred between

two Ty elements fused in tandem and so produced no deleterious genetic effects. Delta-delta recombinations may occur at a higher frequency than those between epsilon sequences since they usually involve the two delta sequences of a single Ty element and, therefore, are genetically innocuous. However, as we have demonstrated in this paper, recombination between delta sequences on different Ty elements may occur and result in a deletion of chromosomal sequences. Rothstein (33) has presented evidence that the frequency of even delta-delta recombinations is repressed by the action of a chromosomal gene, EDR.

The assignment of epsilon–epsilon recombination as the event which converted the adjacent Ty pair (Fig. 6a) or the tandem Ty pair in the RAHS of GRF88 (Fig. 6b) to the single Ty element of CF4-16B (Fig. 6c) relies on evidence involving sequence differences between the two class I Ty elements, Ty1-161 and Ty1-3R. It is instructive to compare the sequence of these two elements from chromosome III with those of the two class I Ty's which have been sequenced previously, Ty 912 (ref. 34) and Ty109 (ref. 35). The centromere-proximal member of the tandem Ty pair, Ty1-161, differs from Ty 912 at only 7 nucleotides within the sequenced region (341 bp) shown in Figs. 5a and b, whereas it differs from Ty109 at 81 nucleotides. In contrast, the nucleotide sequence of Ty1-3R is more similar to that of Ty109 than that of Ty912. In the 538 bp region shown in Figs. 5b and c, it differs from Ty 109 at 14 nucleotides but from Ty 912 at 93 sites. Ty1-161 and Ty912 are easily distinguished from Ty1-3R and Ty109 since the former pair contain a HindIII and a SalI site which the latter two lack. These differences permit the division of class I yeast transposons into two sub-classes, Ty1a (comprising Ty 1-161 and Ty 912) and Ty1b (comprising Ty1-3R and Ty 109).

The delta-delta recombination event which fused Ty1-161 and Ty1-3R resulted in the deletion of 3 kb of unique chromosome III DNA which contained a transcribed open reading-frame (the 31C-GF ORF) and a tRNA^{lys} gene. Since this deletion did not lead to the death of the cell, neither of these genes is essential. This is consistent with the report of Goebl and Petes (36) that 60% of insertions into the yeast genome they studied had no phenotypic effect.

In other regions of chromosome III we have found a number of open reading frames whose deletion or interruption is without lethal effect (ref. 37 and manuscript in preparation). Toh-e el al (38) have reported that the pet18 mutation is associated with another deletion (see Fig. 1b) just centromere distal to Ty1-3R. The pet18 mutation is pleiotropic in its effect and confers the following phenotypes on yeast cells which harbour it: (i) respiratory

deficiency, (ii) inability to maintain killer dsRNA, and (iii) temperature sensitive growth. Sequence comparisons between pet18 and PET18 strains showed that the deletion removed three open reading frames. All of these data indicate a remarkable level of redundancy in a genome which is less than four times the size of that of E. coli. Nevertheless, most or all of these non-essential genes may confer some selective advantage on yeast and investigators must use more sensitive tests (e.g. 39) in order to determine the magnitude of that advantage.

Fig. 6 shows that the delta-delta recombination must have excised the 3 kb fragment of chromosome III DNA on a circular molecule which contains either a solo delta sequence or an intact Ty element. Such a molecule represents a new class of translocatable element since it may be re-inserted into the yeast genome, at any one of at least 160 homologous sites, by a subsequent delta-delta recombination event. If the excised fragment contained an ARS element the probability of its translocation would be increased since it could replicate as an episome before reinsertion into a chromosome. That would not arise in this particular case since there is no functional ARS in the RAHS region (14,40).

The possibility of such translocations offers a new view of the association of tRNA genes with Ty sequences. Translocation of the portion of chromosome III which has been deleted in strain GRF88 would move the tRNA^{lys} gene to a new chromosomal location. A subsequent mating with a partner without the deletion would result in the duplication of the tRNA^{lys} gene. It is possible, therefore, that tRNA genes are often located near Ty elements since this association has served to amplify and spread those genes through the yeast genome and that such amplifications have been selectively advantageous. This is but one example of the likely importance of Ty elements in yeast genome evolution (41). This analysis of a transposition hot-spot from the right arm of chromosome III has demonstrated that Ty insertions and recombinations are a major cause of chromosome polymorphism in yeast.

*Present address: Department of Biochemistry, Washington University School of Medicine, St Louis, MO 63110, USA

REFERENCES
1. Cameron, J.R., Loh, E.Y. and Davis, R.W. (1979) Cell 16, 739-751.
2. Mellor, J. Kingsman, A.J. and Kingsman, S.M. (1986) Yeast 2, 145-152.
3. Gafner, J. and Philippsen, P. (1980) Nature 286, 414-418.
4. Eibel, H., and Philippsen, P., (1984) Nature 307, 386-389.
5. Eigel, A. and Feldmann, H. (1982) EMBO, J. 1, 1245-1250.
6. Genbauffe, F.S., Chisholm, G.E. and Cooper, T.G. (1984) J. Biol. Chem.
    259, 10518-19525.
7. Warmington, J.R., Anwar, R., Newlon, C.S., Waring, R.B., Davies, R.W.,
    Indge, K.J. and Oliver, S.G. (1986). Nuc.Acid.Res. 14, 3475-3485.
8. Warmington, J.R., Waring, R.B., Newlon, C.S., Indge, K.J. and Oliver
    S.G. (1985) Nuc. Acids Res. 13, 6679-6693.
9. Breilmann, D., Gafner, J. and Ciriacy, M. (1985) Curr.Genet. 9, 53-560.
10. Fink, G.R., Boeke, J.D. and Garfinkel, D.J. (1986) Trends in Genet
    118-123.
11. Mortimer, R.K. and Johnston, J.R. (1986) Genetics, 113:35-43.
12. Strathern, J.N., Newlon, C.S., Herskowitz, I. and Hicks, J.B. (1979)
    Cell 18, 309-319.
13. Stinchcomb, D.T., Thomas, M., Kelly, J., Selker, E. and Davis, R.W.
    (1980) Proc. Natl. Acad. Sci. USA 77 4559-4563.
14. Newlon, C.S., Green, R.P., Hardeman, K.J., Kim, K.E., Lipchitz, L.R.,
    Palzkill, T.G., Synn, S. and Woody, S.T. (1986) Structure and
    organisation of yeast chromosome III. UCLA symp. Molec. Cell Biol. New
    Series 33: 211-223.
15. Orr-Weaver, T.L. and Szostak, J.W. (1983) Yeast recombination. Proc.
    Nat. Acad. Sci. USA 80: 4417-4421.
16. Devenish, R.J. and Newlon, C.S. (1982). Gene 18: 277-288.
17. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular cloning:
    A laboratory Manual, Cold Spring Harbor Laboratory, New York.
18. Southern, E.M. (1975). J. Molec. Biol. 98 503-517.
19. 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.
20. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Molec.
    Biol. 113, 5463-5467.
21. Dale, R.M.K., McClure, B.A. and Houchins, J.P. (1985), Plasmid 13
    31-40.
22. Henikoff, S. (1984) Gene 28, 351-359.
23. Elder, R.T., Loh, E.Y. and Davis, R.W. (1983) Proc. Natl. Acad. Sci.
    USA, 80 2432-2336.
24. Kingsman, A.J., Gimlich, R.L., Clarke, L., Chinault, R.C. and Carbon,
    J. (1981) J. Molec. Biol. 145, 619-632.
25. Staden, R. (1982) Nucl. Acids. Res. 10, 2951-1961.
26. Dobson, M.J., Tuite, M.F., Roberts, N., Kingsman, A.J. Kingsman, S.M.,
    Perkins, R.E., Conray, S.C., Dunbar, B. and Fothergill, L.A. Nucl.
    Acids. Res. 10, 2625-2637.
27. Zaret, K. and Sherman, F. (1982) Cell 28, 563-573.
28. Henikoff, S., Kelly, J.D. and Cohen, E.H. (1983) Cell 33, 607-614.
29. Sharp, P.M., Tuohy, T.M.F. and Moaueaki, K.R. (1986) Nucl. Acids. Res.
    14, 5125-5143.
30. Doolittle, R.F. (1981). Science 214, 149-159.
31. Protein Identification Resource National Biomedical Research
    Foundation, Georgetown Univesity Medical Centre, 3900 Reservoir Road,
    N.W., Washington, D.C. 20007, USA.
32. Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. (1985). Cell
    40, 491-500.

33. Rothstein, R.J. (1985) Cold Spring Harbor Symp. Quant. Bie. <u>49</u>: 629-637.
34. Clare, J.J. and Farabaugh, P.H. (1985) Proc. Nat. Acad. Sci. USA <u>82</u>, 2829-2833.
35. Hauber, J., Nelback-Hochstetter, P. and Feldmann, H. (1985) Nuc. Acids. Res. <u>13</u>, 2745-2758.
36. Goebl, M.E., and Petes. T.D. (1986) Cell <u>46</u>, 983-992.
37. Anwar, R. (1987) MSc Thesis, University of Manchester.
38. Toh-E, A. and Sahashi, Y. (1986) Yeast <u>1</u>, 159-172.
39. Mead, D.J., Gardner, D.C.J. and Oliver, S.G. (1986) Molec. Gen. Genet. <u>205</u>, 417-421.
40. Palzkill, T.G., Oliver, S.G. and Newlon, C.S. (1986) Nuc. Acids. Rs. <u>14</u>, 65247-6264.
41. Oliver, S. G. (1987) In <u>Evolutionary Biology of the Fungi</u> (British Mycological Society Symposium, Vol. 12, Eds. A.D.M. Rainer, C.M. Brasier and D. Moore) pp 33-51. Cambridge: University Press.