Brown et al; Supporting Information

Table of contents

1. Analysis of the re-arranged karyotype in CBS 2777

1.1 Assembling the individual chromosomes

Chromosome 3 corresponds to chromosome III of the laboratory strain and bears no further discussion.

Chromosome 1 contains

1. The distal left end of chromosome I to about 1.21 Mb: segment I.1
2. Segment I.4 which extends from about 2.26Mb on the left arm to the telomere at the right arm of chromosome I;
3. 150kb of centromere flanking DNA on the right arm of chromosome II (segment II.2) including the *dgdh* repeats that flank the right side of the central core of the centromere of chromosome II.

We assembled these three segments into a single chromosome by linking the right end of II.2 to the left end of I.4 by PCR and the left end of II.2 to the right end of I.1. These links were defined as breakpoints 2 and 5 respectively (Fig. S1a) present at 1,216,710 and 1,388,551 in the CBS 2777 assembly. The chromosome II partner of breakpoint 5 lies within the IMR repeat on chromosome II corresponding to residue 1,627,727 of the laboratory strain, the three other partners of the respective breakpoints were in single copy DNA. Breakpoints 2 and 5 predict that sequences on the right end of segment I.1 (I.1R), II.2 and the left end of segment I.4 (I.4L) would be physically linked and include the tandemly repeated *dgdh* sequences that flank the centromere in the laboratory strain. We confirmed this prediction by pulsed field gel and hybridization analysis; probes cognate for each of thethese four sequences recognized a 950kb SfiI fragment (Fig. S1b). The size of the cognate fragment was consistent with the SfiI sites predicted on the basis of the sequence to be 955,267 bp in length. The cognate AscI and NotI fragments were predicted to be 2.8Mb and 1.6Mb respectively and were not resolved by this gel. We also targeted a cassette containing a *ura4* gene and an AscI site into a CBS 2777 *leu1Δ ura4Δ* derivative strain at either breakpoint 5 or the ectopic chromosome II derived *dgdh* repeats and used these strains to place the respective target using SfiI + AscI double digests, pulsed field gel and hybridization analysis. (Fig. S1b). The sizes of the fragments recognized by probes on either side of these sequences were as predicted at 539kb and 425kb.The sequences of breakpoints 2 and 5 are provided below.

Chromosome 2 contains

1. the region of the right arm of laboratory strain chromosome II (segment II.4) that extends from the right end of II.2 to the telomere
2. a 600kb segment of the left arm of chromosome I termed segment I.3
3. the central core of the centromere of chromosome II (segment II.3) and
4. a short segment of chromosome II (II.1) that is located between residues 569,645 and 689,628 of the laboratory strain chromosome II.

We linked the right end of I.3 and left end of II.4 by PCR thereby defining breakpoint 1 (Fig S2a). We then linked the right end of the central core of chromosome II (segment II.3) to left end of II.1 to define breakpoint 3 and the left end of the central core of chromosome II (segment II.3) to the left end of I.3 to define breakpoint 4 (Fig. S2b) and then joined breakpoints 3 and 4 by amplifying across the central core of chromosome II using primers at the left ends of I.3 and the II.1 to produce a fragment of 8.4kb which included the central core of chromosome II and thus joined I.3, an inversion of II.3 between the IMR sequences and the left end of II.1 (Fig. S2b). The sequence of this product upto and beyond the respective breakpoints is shown below and demonstrates that no further re-arrangements have occurred in this region. The organization of the chromosome determined thus far predicted that the right end of II.1 with respect to the *972h*$^{-}$ assembly would now be telomeric and would thereby define breakpoint 7.

We tested this prediction by restriction enzyme digestion, conventional agarose gel electrophoresis and filter hybridization using a probe predicted to lie adjacent to the right end of segment II.1. Consistent with the prediction the sequences at the right end of segment II.1 are duplicated. One copy is predicted to be non-telomeric and on chromosome 4 and the other is predicted to be telomeric. One of the two cognate Xba I fragments was a 19kb in length and Bal31 sensitive (Fig. S3a). The size of this fragment however was inconsistent with simple healing of telomeric simple sequence DNA suggesting that additional sequences had been healed on this chromosome end. The ends of all fission yeast chromosomes include a single copy of the gene for the Tlh1 helicase and so we tested the possibility that this gene was also present adjacent to breakpoint 7 by PCR. This experiment (Fig. S3b) confirmed that this was the case and thus allowed us to determine the sequence of breakpoint 7 (see below). The PCR product was heterogeneous in size reflecting the fact that the gene at the breakpoint includes at 36bp coding tandem repeat. The breakpoint sequence was determined by primer walking and occurred between residue 689, 628 on chromosome II and 1,135bp of novel A and T rich sequence flanking the 5' of a copy of the *tlh1* gene.

Chromosome 4 contains

1. all of the left arm of chromosome II
2. the central core of the centromere of chromosome II (segment II.3) and
3. a duplicated segment of distal left arm of chromosome I that corresponds to the left end of segment I.3 and extends from 1,216,651 to 1,299,253 and which we term segment I.2.

The sequences that we have identified as present and flanking the centromere on chromosome 2 are all present on chromosome 4 raising the possibility that they have the same sequence organization on the two chromosomes. That this was so was shown by four lines of evidence:

1. Conventional agarose gel electrophoresis and filter hybridization using probes that recognize the central core of the centromere of chromosome II, the right end of II.1 and the left end of I.2 recognize a single sequence (Fig. S2c). The map of this sequence as defined by the sizes of the restriction fragments matches the map of the map of the re-arrangement and position of the relevant restriction sites based upon the assembly of the laboratory strain (Fig. S2d)
2. Pulsed field gel electrophoresis and filter hybridization indicated that one copy of the central core sequence is linked to sequences that are present between 752kb and 753kb on chromosome II. This linkage demonstrates that on chromosome 4 the sequences present on II.1 on chromosome 2 extend into II.5 without further re-arrangement. The second copy of the central core sequence is linked to sequences that are present between 1,397kb and 1,398kb on chromosome I. This linkage demonstrates that on chromosome 2 the sequences that are present on I.2 extend into I.3 without further re-arrangement. (Fig. S4a, b). The sizes of the cognate restriction fragments are consistent with the map of the re-arrangement and positions of the relevant restriction sites based upon the sequence assembly. We also determined the positions of *dps1*, *vac*, *nht1* and *ste4* in the interval duplicated on chromosomes 2 and 4 by sequence targeting a *ura4*AscI cassette into each gene and confirmed that they were placed as predicted and indicated in text figure 1 by hybridization with one or other of the chromosome specific probes (Fig. S4c).
3. These results also imply that II. 5 in chromosome 4 is arranged as a paracentric inversion extending from 569,645bp on the distal left arm of chromosome II to 1,619,574bp in the IMR left (the equivalent of 1,628,841 of IMR right) of centromere II and thus indicate the existence of a cryptic breakpoint on the distal long arm of chromosome 4 which we refer to as breakpoint 6. We confirmed the presence of breakpoint 6 by PCR and identified long range physical linkage between sequences

predicted to lie on either side of the breakpoint (Fig. S5 a, b). This inversion predicts that, as on chromosome 1, the presence of *dgdh* sequences around a non-centromeric breakpoint and again this was confirmed by pulsed field gel and filter hybridization analysis with the single copy and *dh* probes recognizing Asc I and Not I fragments of about 1.3Mb and about 1.2Mb respectively. The observed sizes of the cognate fragments were consistent with those predicted from the sequence assembly; AscI 1.24 Mb, NotI 1.25 Mb and SfiI 0.4 Mb (Fig. S5b) allowing for the presence of a 100kb inversion around the AscI site predicted at 1,239,024 in the assembly.

4. The final feature of the maps illustrated in figure 1B and represented in figure 2C is the existence of a second telomeric breakpoint, breakpoint 8 at the right end of segment I.2. A probe (Fig. S3) flanking the hypothetical breakpoint identified a 10.5 kb EcoRI fragment as predicted from the laboratory strain sequence, a weakly hybridizing fragment at 3.5 kb common to all the strains and an additional restriction enzyme fragment of 15kb present only in CBS2777. None of these fragments were Bal31 exonuclease sensitive. We wondered nevertheless whether the additional CBS 2777 specific fragment was healed with sub-telomeric DNA and confirmed that a copy of *tlh1* lay just beyond the breakpoint sequences just as at the other new telomere breakpoint 7 (Fig. S3). As before primer walking identified the breakpoint sequences and showed that they were within four base pairs of one another on the telomeric side and at 1,299,253 on chromosome I.

Thus we were able to establish the idiogrammatic map of the chromosomal organization shown in figure 1b of the main text.

1.2 CBS 2777 breakpoint sequences

Breakpoints 1 and 2 nearly reciprocal

Breakpoint 1

I: 2259804
cccaaagcatatttttgtccaggcaagggaagttggataattctcttttgatctcagtaagtctttttaattaa
tttagttttgcatttaggttttactgtttgatttttctgccttttaaaagagcatac/ I: 2,259,935 :
II : 1,779,556 /
tgattgattgatttgttgattatagcttcactccattcttcgttgcatccttttaagtttctttttatttcac
aaaa
II: 1,779, 634

Breakpoint 2

II:1779501
Tctttggttttacgagtggtatttgctggttcaaattttaagtgcatataccgattgatt/
II: 1,779,560 : I : 2,259,935 /
Cccaaagt**atg**cagagagacgaggaaagacccattgactttggaatgtctttagaactcccaaaaacaccaaatg
a I: 2,260,010

Breakpoint 8bp 5' of Start codon of meiotic PUF family protein 1

Chromosome II breakpoint lies within four copies of a 4bp tandem repeat which is underlined and in bold

Sequence of chromosome II broken in BP1 indicated by /
ttcaaattttaagtgcatatacc**gat**/**tgattgattgatt**tgttgattatagcttcactcca

Sequence of chromosome II broken in BP2 indicated by /
ttcaaattttaagtgcatatacc**gattgatt**/**gattgatt**tgttgattatagcttcactcca

The region of chromosome II present in the respective breakpoint is highlighted in yellow.

CEN CBS 2777 8386 bp Breakpoint 3 > Central Core Chromosome 2 > Breakpoint 4

II: 569,689
Ctaaagtccatactctcttcttgtgaacggtccattttcaaaacg/ II : 569,645 :
1,628,841: II
/actacgatgtatgcatgtgaataatatttacatttttcattcctatgtctactgtttaaactaagtattgtaaa
tacttataaaatttttattatgatataatgagcttgttctttatttttacaaagcaatatggcttgcatataataca
taggctacaatacaatgtacattcaagtattgaaaagcttttcctgtctttccaattaaaaaacactcaacttca
acgacgtgatataagtataggtatattaaataaagcgtttttttataaccagttccgcaaatgtaggaagttaatc
aattttaaatgttgaaaagttataagaaatagtgatccaattaatcatgccatggaataatttttataaaccggt
aatcgttgcaaagtgcttaccgtttacttttagggcgaaacaacaatacaattaggtagtaccagatcgtttatg
aaactgcttttaggtgggtactttaaaaccacatgaggtttcagtgaacaacgttttgttatttttttaagtaatg
aacttaaactttctttgtttactggttcttatcttactaccataagtattagtaatgtaattttcggtcaaaaga
ggtgtataaaacgacaaaaatgtgttttaaaatttccattctaatttattcatctcatcaattttgtaaagccaa
cgaggtatttatttttgcttgttttttatttttaattagtttacgttaaaatttcaaattatttaattacgatga
atacttgggttaatgtaaaaatagaatgattcgaaacaaaattagttttatcacattcctgttttcgtacttttc

ctttacataataaaaaaaaagaaaaggaattgtaacttgaaatttgggattaatttaagcattagcgccttcaat
aatttatgagtaaatggtaaacaggggtttgatttgattataaagggtatataaacggctagcataacttttgtt
aatccaggcacttttttattgtattctgattttggagctgtgatatgatgcatttgcacattttcgatagaatta
caaatacaaaacattcacttaattgtgatttggttttaattgaccaatatacttcttctcatggaaacattaggg
ttcatattaaattttatcaaaaatcatcaattcgaattcattctacttgtatcacttcaaacaaatgccaaagtt
tttaacttaacaaaacaaaatacatattagtcagttgctatgtagttaatataattcatattctaagaagatctc
gcataaaagtatttctgtgggtaattttacactattcgaattttcttcttttatttttatttttatttttagtta
ttttgattagggaggctattttgctcgctgcgtccttatatgcggcttgatttgtattatgtacattatttccct
gcacctcactgtttctagtaatgagcttagtgctttacaaggttactgctgtttgctttatttcatttttatggg
atcggaagaggacttgaatttttttttttttcataacttaagtatgcattacttctaaaaggttgttaactggttg
cctgactgccttttctatctgtagttatatgatcaaaatttaaagtagtttttctatgttattaattaagtaaa
gcaagcattctacaaataccatttagaatagtcgtcttccttttttaaccaggcttttgcaatttgctgaatgcta
tctcagtaaacaaatttgtaaacaatagtcaatattaggaaaaactagtgctttttatattgacactagttttt
tgcagctttctcaagaagttagtatttgtattcccgatcaaaaatattttacgtaatgatttgagccttcagctt
tttattgactagtagcttatatatttgtaactagataattttgttaaggggtaatttacattatcaatgtccccc
acaaatagttcagcaatttttcttctgctttagtaaattttaaagtacgataaatttgataaaggtaacgtcaac
gagtcggtaaatatttggtattgcatctaaaagccaattcaatcaatttttttaaaaaacgagtaaagaccttcga
atcaaaagaatttatttaaaaaattccaatgatttaggattgccttttttttgggttgtacatgagaataattca
acatcaaccttttttatatggttaaattatatttacgaactttagtaaaacttttaaatgttgagatttgtaaata
gatatacgtttaataattaataatgaaagtttttttctgcatattcgacatcttgagagaatgcgaatccgttttt
aatatttgaagcactactaagattactaattacagactgatgggttacctttttgcgatgataaggctgaattatt
acaatattactaaaggtgacgagtcgctggcagtattaaaattatttttttgctaactggtaactaaacttattag
gccgttaattaaatctggtttataccaaaacgagaagtccatctttttgtttctgatttttaaaaagttattctt
acgtctattaattcagattgactgaaaattgtgatctaggctagttaatttttgggagctgcaaaaaacaaaaaa
aaataaataccttgttaattaaactaccatttaagctgttgtaacagattccataatctaaaagaaaacattgca
gatattaatcatatctttacttttcaaaaacgtaaaaatttttttggtacatgaaactttaccaccatacagttctc
atactaaacattcaatccaaacaatggaccaattactacgatgtgtatgcataacatccaagtatatgtgttcta
tcgtgtatcctagtaaaacatactatacattcagcaacttaaatggtatattaaagatactacatcgctaatcat
tagttttaggagcccttggaaatacctagctttaatgataaaactagatgaatactcaataaagcaaatcaagta
tttcagacagttaaagcagttgacgcaatttgaacgtacaaaatttttcaaaaaattcaatctgagtggtctagtt
atagtcaaaatattgaaaacagtttttcgcacgaacaaactagcagcagctgagacctttggctgttttctttga
aatcatgagttttactgatttccaggaataaatatttttagattttatttttagtagtgaaaacaacagtaagcaa
acactaaagcgaaagcttggactaaattagtcactattaattaaaaggatttcgcattgcattttaaactgttca
tcaaataaatttcttcagagctatctactgggtaaacttttaactgaaaatatagcttcatgataattattgact
ttcattatatgttcacggacacacatgttgttaatacagttgaatatttattgaactgatgaacccgtactaaga
tgtttattcacatagttaatgtacatacgtacaaaaatgtctgtgttttctgctataacccattaaccagaaaat
catttatatagcaaaatattgcttcgaaaactttattctaaactcttaatggttttctcgcgattagtttgtaaa
gtattaacaggtgtttaagattcttttagattatttccgaaaaagcctttcacgatactttaatcaacaaatatc
aagggacaatttttttgtttttttgttgcggtgttttgaaaatatgtgtttcgcatacacccaaagtggttaataat
tagcatttggacgaagcttaagttaattttgataagaaaacaaataaacaagatgaaatacgcattactggcaac
ggagaggcaacacagacccacatagcagcaattggtttgatgcaacatttatttgcgataggtagagctaaacct
tttcacaaatgtgcacgatttttcgaactttctgcaagaaacaaaaaacgtaggggtcctagattattacgtctt
aaatacgctttagttagttttaaatatataggattttcatgtaaacattttaatcgcacataaaataaaacacaa
aaaattgcaaaagtatttttttcgagaattttattttgatttggtcatttcaacgaaataatgaaagaataaattg
ttctggatgatctttagaaactatctttttgaaaatattaataaaaaccatgaaccaaatcacaagcaaaaccaa
acagaaaaaaaaattcactgtttcgcaaatggtctgattacaatatttaacttcgtaatttgcgtaaaatatgag
gctctctgaaattaataaaatttctataataaatttgctttaccaactgatccttcaaactacaacattgtttct
ttaaacgatttgcaacttcttctataccgccatttgagcttttactacggttacagtttggagtttatagctatg
ttaaactgaatgttggtatttatttattaccatcattactaggtttctcagctactagacataaaagtttccctc
tgcgctaattcttgcatgagtttagtaagtgatattttagaaaatgtcaatcacgattgacttactgcaagaatg
gttttgagttttttttttttgtaaccaaatactatcgtcaactaaatatttttttcattctttcctactataatctt
tatttattaagtaaaattaattaagccagcaaattccttgagtaatttcatttgcatgcttggaagctgtgacca
aacacagtgtgccgttgctttgttttatgttttagtaaatttattaaaataattctaactggttaggagtcaaat

tagtaaagttgataatcacaacaattatcataaaactttatgatttgtaaattacaaacaagatatgttaatttt
aaaatttctaatgtagttaatatgcgagttatgtgtatctacataaaaaaaaagaattgatgacatggcgtggaa
agtcatttgtatgaacgcattaaaaaaaggcaattgctaattatctcaatacgccaaagaatggcaattaacaaa
caagtcaatttacccgtattttgatattttctgtgcattcattacttttttttgcccgtattttattaaacgag
ttagcgcttaacaagtataataccagaatttttgctttcaaattaaacattgctaataattattttcttaaggga
aattttataaagtcgagttaaagaaggtattgatttttgcacatattagtttattcaagcacatgaattatcga
aaaaggttcagagaattcgttatttattatttacgttcaattttatttagtgtttcttaaagtcaaaaggtatgt
taaaggaattcgataacttttccagtaatagggtagaaacttaaaggctaatcagtatttagttataaacatata
gtctataaagtaagtttatcttttaaattttgcacacctagcagtttctacagcaatcccttctcctactactct
aacagttggtttgtgttgtaaaatcactaagaagaaatttaaaaaattataataaaaagattattcgaattatgg
tcttaggaatagaaaactgaattatgaggtaatttattgttttgtttctcaaagaagaaggcttttccgtatttg
attaatgttatttaataaaattaatttggattttggaaataatatttttttttcgttggaaatagaaagctgaatt
aagtaaattcaatttaatgggggacattcggtaagaattttaactaatttgagtttttgcattcttcaacgaagg
atggatatgcacgtcgtgtgaatgaataagttgctcagtagcaatctatctaaagtcaaataatttttttacgtga
tgagagagtacaatagatgcaattactaggacaaatagaatatgctaatgcaattaaaattttcaagctttttta
cttgaatgatatgcatatttagtaatagttgttctaaaaatgcattgatataatttacacgtttatcctagtcgg
cattatcatagtaaaaaaaaaaataaaataaaaaatccgggggaagcatctacgttttcgaaaatgaggatagggg
gcattatatgtaattaatagggggagggctatcctgaaacaaatgcttgagaaaatttatgtactccaaatatgct
ggtaaccgtaacttcctagattctgtatgctccttttttaatgaatggcctgattcttgattggaagttagcagta
aacctgaacagaacaaaacttcttgctaacgttaattaatttatattaaatcgttgttttattaactatgttcaa
aataactttgactcttagctcacttatgcgaattaccttaacgctagcactatttttaacatatagcatatttga
tttaatgttgttaattttcataaatgacgggtgttgttaatagattcattttcttataaactacaatatttgcta
tagtcttatttaaattatagctggtttacagactccatttcaagtcaaaatgtgcaataataaggtttcgttagc
aacattatggctttgtatgaattaatacttagctatctagcaagtgtaatagaaatagtaagtttctaaacagga
taatttaatctcttatacataattttgttactcaataaaaccaagttgcctactttggtaggggcatttttgata
tagtaatcacatcaggcaatggtaagttaaatttgcttttaaataaatcatatcttaatatatttacccctcttt
ttttcaaatttgctttatgttttgcgacaagcgtggatttatttaggagcttgctcacttttaccatttcatgt
cattcatcaagtaacaactcggtaaaaattctgaagttttagataaattctattacatatgtgttacttttttaat
tttttaatcgacaggatatttgtgaaaaaagaaaattagtgcatttaacatgaaggataagacactcgccaaaat
cataaaccccccctgaatctatttaccagcaaattcatagatttgagcaataatttcctaaaacatttacaattaa
tgggttatttttttttttctaaaaattgaaggagcttgtttgatctattccgttattagtatgtaaatttaaaaaa
tcgtatccctaactaattcatacttaaacataatatgcagattttaaaataattgtccatatagataacacgcgg
aatacttagaaaagtagaattggccgaagccaaaaaggaaacatagaaatcaaccaggactaaccaatgcttctt
catattaacgagtaacgcggcttacgcatcacaaagcagtacgtgaataaatattattggaaaacatttttcttt
aaccagctaaaccgcttaaagaaaaatgaataagatacaaattttcaggtttttgctttgatgctaaaaactcgt
cgaattctaaaaccaagaattaaacgtacccttaaacttttttttttgattttcgtactaataattatcggagtt
cgcgttttcatttaaaaaaaggtttagtaaatctccaattgcgatacgaacaatgataatgatattttcagata
aacgttacttaatagcatcttatagtcatgggccatctcaatattaccattaaacttttaagcaaaaagatcatt
tctattttaaaaaacctaataaaaattacctttgtaagtatacgtttgctaaacaataaaagaacgtaatacta
aaaaattaaagtcattctagggtttgccaacacaattttctgattatcgttagatgtaaacagtgctcgcttatg
tgtaaataattaagaagaaaattaacccctttatatttgtattagcgcgttcatacgcaagtcttgctcagcaa
aagttgtgtttttaaatgcaataaataaaatagtaactaattaacacgaaatcaccatagaaactaacagttagt
acgtaatttgtcccccagtaattgaaaaacgtaagaacaaaaaaattttcgttttgttttgttaagttaaaaact
ttggcatttgtttgaagtgatacaagtagaatgaattcgaattgatgatttttgataaaatttaatatgaaccct
aatgtttccatgagaaga**agta**/ II : 1,620,686 : I : 1,216,651
Gaagcggaaattacttcgtaatagcaatttaatggatatttttttccacaatttatgcactttgccatgcaacgaa
caaagccaactagtataccaagtcaagaaatctatgagtacagcaaagtatctaagtgttatcctatttagttta
cagtgaccaataaaaattgtatcgtgatattcagattaaaagataattgtcatctattttcagcttcatt
I : 1,216,869

Breakpoint 5

I: 1216531
accgtaaaggcaat**ccagc**aaattcggttntctagaaacatcagtttttcactggtccggatataccgatatcgg
gaaaattaatgtcggtcctaagtttatgaatttnatcgaaacat**agta**/ I:1216650 : II:
1,627,727/
tattggtcaattaaaaccaaatcaaaattaagtgaatgttttgtatttcgtaattctatcgaaaatggttcaaat
ggcatcatatcaaagctccaaaatcagaatacaataaaaaagtgcctgg II: 1,627,851

Breakpoint 6

Underlined residues illustrate two bp region of identity at breakpoint

II : 569,506
aaatcaatttttttatttgagaaaatggtttctaatggtgccngccttttttgttaaaaaattatattgcaacaata
gatagatatgctattccttaaaangtgggtttctnntcattactctttacagctttctgttat**gt**/
II: 569,641 : II: 1,619,578
/agtagtggtcgttcgtagctgataaaattcgcttaaataagtatta
II : 1,619,532

Breakpoint 7 (underlined regions are absent from the laboratory strain assembly)

Tlh 1 2221
gctcttttacagtcctgaactttggcaaacatcccctcatacagagctgtcagctcactcaagtccaatcgcacc
atgnttattgaatgcatctccaatgcacgggtacatagaattacttcgtggctgaaacaatacggtacttgggct
tagtccttgcagtgtatttgcataccttccgttcgaatgatatccctaacgacttttgggtttcgtaaagaac
atcatataaatttctttgaccaagcgtgataaagcaaaacacatgagcataagcaatagaaccagcggtttgaaa
cgatggaatggacaacaagtcgatttgtttttcatcgatcaatcgttgtaactcacatgctgctacaagacgaca
cgtgtatattaggggtgcaaagcaggcagagatttcataggcatatttcgaagtaccatctttatgcaaagacat
gcatgcaagaaactccatgatttccaagtcttgaacatttttagaaaaatcgtattgttgcagaaaagccaattt
taatgcctcgtgcagttccaagtatgccagcttatcatcatcgtcagtattggtgttgttattgttactgttact
gttaatgtcactgccattgtcattgtcattgtcgtcttcagcactatcaaagttgacgacgttgctgatgttact
gctgatgttactgccaccttcttttctctccactgttgctccacacattgcttccaagtgtcttcttatccccga
gtccgctgatggacgacgtaacagaaatattaggtatagtgctgcactttgagcatatctcacaactgtatcctt
ttgctctaagcgacaangtcctcgctcttgngcactatagcttcctccttgagtaaacattcgtcncangcccat
ggtcatncccttagatgacggaatcctcgatagtatacgaatcaactggttacaangnatccaagtacantccat
tgnatantnggatcgaacagctngatcccaatttnctccnagnttttcacaattngccgtccaattgaattgtt
nganttaaaaganttttctctcgtgcattgacaatctcttccttttcctcttcctcttcatcgttatcttcttca
ccatcttcaagaatagtgttcgtgttgttgtctgttgcttggtcaacatcatcttgtcgtttttttattttcacct
ttgtcttgctgctcccgccgctgttgattcttctccataaattttgaaaagcaaaagtccaccattctttccttg
atgttttgtgcatcttctttctcctccccctcctcctcttcttcttcttcctccattcaaaggtacataatcg
actttgaagaactgacatcgttgagcatacttgttcttaaccgtttgcaaagctgttcgacgtatacaattttcc
aattttataacttttccatgagtacgtcgaacatgatggcgaaaggtatccatgattgcatgtacgaatcctgtt
ccgtttttttgtgcacggcacacactcatatccatttaaaagtactggtagtcctctgatgtatgggtatacatgg
gtttgagaagaatgagtttgtaaaacatttgttggacttttttaactttaaagtccttagttcttgaaaccacaca
agatcttcgtttaactctagtttatgcacaatttgcatatgttgcgcagtgtgaatcacgtttaacaagcattca
cattctacgcacatcaaagcatgaatagacaaaatggagagtccataattagccaactngtgattcaagtcagta
taatttgatgatgggctgntgtccaatttggngncatttgnacttggcnaaataagccgngcagtcccaagtcgt
catttgcacttgtagtgagagtagagagagcagaaagagcagagagagcgggtagttgacgctccttggaagaat
tgcaagcttcctcactggtagcctcactgacagcctcaacgatgattggcttatccttttctcgtcctgaacga
aagagctgggttcatagacgacgatatcatcgtctttacaagtgtcgtctctgtatcgacattccttgaaatgtt
gaacgatcctttctacattatcaaattgagttccacattgacaagtaaagcatccggcgatatctctagcagttt
ttgaagcgac/ Tlh 1 28: I
catggcagtagaaatgtgaaaacaacaaaatgaagggataaataatgtatcaacaaataaaaaaaagtttgcatt
tatataacgaatgtaaaacagtagccagcccagtcagccatcgcttatcaaaccctgtttgaaccagcctaaaaa
gtgaaatagaccagccatacgaaatgacatgtcaaagagcatccaaatataaaaaagattgaaacttttttcaac
ttgacaagtttcgacttgaaaaattttttttctcctccctctaccccttacatactacctttgacctagttttc

ctcgattttgtagaaagtggtgctactttgtgaatttcatccatatcggtgcttttagggtagggtttgacattt
gtcgaacgcactacgcagtactacagaaaaaatccacttccaccaaagggttttgcctcacccttcctattactt
gaataccaaaaagacaatccatttctgtctaaaatctttttgtatagcgaatttaataaaattaaaatgatttct
tttctttcaacaaatttttaaaggtttatgaaatcatttcgaaaagtaataacaagttgtgttagaaacaattcc
attttattgcaatgaaaaaaagatgagtaatacaattgtttttttcgaaatttgtaattttatatttatttaaac
aatttcattgaataaactatttttattcattgtaatttatttaatttatttttatttttatttttattttttat
ttattcatttttttctatttttatttttttttttaattcatttttttctatttctatttttatttttatttttat
ttttttttcgtttttccctaccctccttattcataactgagatgggactcatttcataactgagatgtcctttt
ttccttactccacgttaatcgtaagtgagatggcatcgaaacgatgactctcgtccgttctcgttcctcatcgtt
gattctcgtcaaaatacgttctctatccctaagtctccctattcataactgagataggacatgaaaaataactga
gatagactcgttctctctcgttaaataccaattcttacgtcatttccccatctcacaaaccattcaatcctaccc
atctccacct/
/ II: 689628/
ctagtaattggtgtggatgtgttcaaagcagttgaactagtaattggtgtagaactgttcagaatactagaactagta
II : 689551

Breakpoint 8

tlh1 2221

gctcttttacagtcctgaactttggcaaacatcccctcatacagagctgtcagctcactcaagtccaatcgcacc
atgnttattgaatgcatctccaatgcacgggtacatagaattacttcgtggctgaaacaatacggtacttgggct
tagtccttgcagtgtatttgcatacccttccgttcgaatgatatccctaacgacttttgggtttcgtaaagaac
atcatataaatttctttgaccaagcgtgataaagcaaaacacatgagcataagcaatagaaccagcggtttgaaa
cgatggaatggacaacaagtcgatttgtttttcatcgatcaatcgttgtaactcacatgctgctacaagacgaca
cgtgtatattaggggtgcaaagcaggcagagatttcataggcatatttcgaagtaccatctttatgcaaagacat
gcatgcaagaaactccatgatttccaagtcttgaacatttttagaaaaatcgtattgttgcagaaaagccaattt
taatgcctcgtgcagttccaagtatgccagcttatcatcatcgtcagtattggtgttgttattgttactgttact
gttaatgtcactgccattgtcattgtcattgtcgtcttcagcactatcaaagttgacgacgttgctgatgttact
gctgatgttactgccaccttcttttctctccactgttgctccacacattgcttccaagtgtcttcttatccccga
gtccgctgatggacgacgtaacagaaatattaggtatagtgctgcactttgagcatatctcacaactgtatcctt
ttgctctaagcgacaangtcctcgctcttgngcactatagcttcctccttgagtaaacattcgtcncangcccat
ggtcatncccttagatgacggaatcctcgatagtatacgaatcaactggttacaangnatccaagtacantccat
tgnatantnggatcgaacagctngatcccaatttnctccnagnttttcacaattngccgtccaattgaattgtt
nganttaaaaganttttctctcgtgcattgacaatctcttccttttcctcttcctcttcatcgttatcttcttca
ccatcttcaagaatagtgttcgtgttgttgtctgttgcttggtcaacatcatcttgtcgtttttttattttcacct
ttgtcttgctgctcccgccgctgttgattcttctccataaattttgaaaagcaaaagtccaccattctttccttg
atgttttgtgcatcttctttctcctccccctcctcctcttcttcttcttctcctccattcaaaggtacataatcg
actttgaagaactgacatcgttgagcatacttgttcttaaccgtttgcaaagctgttcgacgtatacaattttcc
aattttataacttttccatgagtacgtcgaacatgatggcgaaaggtatccatgattgcatgtacgaatcctgtt
ccgtttttgtgcacggcacacactcatatccatttaaaagtactggtagtcctctgatgtatgggtatacatgg
gtttgagaagaatgagtttgtaaaacatttgttggacttttaactttaaagtccttagttcttgaaaccacaca
agatcttcgtttaactctagtttatgcacaatttgcatatgttgcgcagtgtgaatcacgtttaacaagcattca
cattctacgcacatcaaagcatgaatagacaaaatggagagtccataattagccaactngtgattcaagtcagta
taatttgatgatgggctgntgtccaatttggngncatttgnacttggcnaaataagccgngcagtcccaagtcgt
catttgcacttgtagtgagagtagagagagcagaaagagcagagagagcgggtagttgacgctccttggaagaat
tgcaagcttcctcactggtagcctcactgacagcctcaacgatgattggcttatccttttctcgtcctgaacga
aagagctgggttcatagacgacgatatcatcgtctttacaagtgtcgtctctgtatcgacattccttgaaatgtt
gaacgatcctttctacattatcaaattgagttccacattgacaagtaaagcatccggcgatatctctagcagttt
ttgaagcgac/
tlh1 28: I
catggcagtagaaatgtgaaaacaacaaaatgaagggataaataatgtatcaacaaataaaaaaaagtttgcatt
tatataacgaatgtaaaacagtagccagcccagtcagccatcgcttatcaaaccctgtttgaaccagcctaaaaa
gtgaaatagaccagccatacgaaatgacatgtcaaagagcatccaaatataaaaaagattgaaacttttttcaac
ttgacaagtttcgacttgaaaaattttttttctcctccctctaccccttacatactaccttgacctagttttc
ctcgattttgtagaaagtggtgctacttgtgaatttcatccatatcggtgcttttagggtagggtttgacatttg
tcgaacgcactacgcagtactacagaaaaaatccacttccaccaaagggttttgcctcacccttcctattacttg
aataccaaaaagacaatccatttctgtctaaaatctttttgtatagcgaatttaataaaattaaaatgatttctt
ttctttcaacaaatttttaaaggtttatgaaatcatttcgaaaagtaataacaagttgtgttagaaacaattcca

ttttattgcaatgaaaaaaagatgagtaatacaattgttttttcgaaatttgtaattttatatttatttaaaca
atttcattgaataaactatttttattcattgtaatttatttaatttatttttatttttatttttattttatt
tattcatttttttctatttttatttttttttttaattcatttttttctatttctatttttatttttattttatt
ttttttcgtttttccctacccctccttattcataactgagatgggactcatttcataactgagatgtccttttt
tccttactccacgttaatcgtaagtgagatggcatcgaaacgatgactctcgtccgttctcgttcctcatcgttg
attctcgtcaaaatacgttctctatccctaagtctccctattcataactgagataggacatgaaaaataactgag
atagactcgttctctctcgttaaataccaattcttacgtcatttccccatctcacaaaccattcaatcctaccca
tctccacctcta/
I: 1299253 /
aaaagcaacgcttctcgggatgtcgtgctatcaaacaaagataacctcgctaaagcacttcttaacaacttaacc
ttggaggaattgatagctgttttaattgcgcaagctatttcataggctgaaatatcagaagttctttccatgaaa
ttaaaggcaatgtctttttgaaatttatacaaacatcaacccggagcccgaaaaacc I: 12299034

2 Experimental Methods

2.1 Micro-array and PCR analysis of the re-arranged karyotype. For micro-array analysis, chromosomal DNA was size-fractionated by pulsed field gel electrophoresis as described (*1*), electro-eluted from the gel into dialysis tubing, concentrated using butan-2-ol and then amplified by the Qiagen REPL1-g mini kit before labelling using the Agilent DNA ULS labelling kit (5190-0419), then purified and hybridized to the Agilent *S. pombe* 4x44k ChIP-on ChIP array (G4810) using un-fractionated reciprocally labelled *972h*$^{-}$ DNA as competitor. All steps were carried out according to the manufacturer's instructions. Arrays were scanned using an Agilent Scanner and the data analysed using the Agilent Genomic Workbench. 5.0.14. Oligonucleotide primers were designed on the basis of the reference strain sequence at each end of the segments indicated in Fig. 1a and then linked to one another using conventional PCR for breakpoints 1-6. For breakpoints 7 and 8 the ends of segments I.2 and II.1 were linked to the *tlh1* gene using the Expand dNTP kit (Roche). Breakpoints 3 and 4 were linked similarly. Details of the results, primer sequences (Table S1) and sequences across the breakpoints are contained within the Information.

2.2 Strain construction and yeast culture. The *ura4* and *leu1* genes in CBS 2777 were sequentially deleted using sequence targeting and a kanMX6 cassette of Bahler and colleagues (*2*) flanked by ϕC31 integrase *attP* and *attB* sites. A codon optimized ϕC31 integrase expression vector (*3*) was used to sequentially delete and allow recycling of the kanMX6 cassette. A cassette containing the ura4 cassette and an Asc I site was then targeted to the *dps1*, *vac8, nht1* and *ste 4* genes on chromosomes 2 and 4 of CBS 2777 *leu 1Δ ura4Δ* strain. The targeting reactions were screened by PCR, checked by restriction enzyme digestion and filter hybridization analysis and rechecked by restriction enzyme digestion and filter hybridization analysis following pulsed field gel electrophoresis (*4*). Silencing was

assayed using FOA as described at http://www-bcf.usc.edu/~forsburg/plasmids.html. *S. pombe* were otherwise grown in YE5S, handled and stored as described (*1*),(*2*).

2.3 DNA sequencing and analysis Sequencing was provided by Genome Enterprise Limited, the trading subsidiary of The Genome Analysis Centre. Libraries were constructed using the Illumina DNA TruSeq protocol, with a true insert size of 359bp (mean size) or 479bp including Illumina adapters (60bp on either end of the fragments). The library was sequenced on a single lane of the Illumina HiSeq 2000 using 100bp paired-end reads which resulted in 84.3 million pairs of reads. The Q30 quality scores for read one was up to base 85 and up to base 70 for read 2. For details of the sequence assembly and bio-informatics see the data.

2.4 Micro-array expression analysis. Total RNA was extracted from either CBS 2777 or CBS 2776 cells with hot phenol, the size and integrity was checked by agarose gel electrophoresis, cRNA probe prepared and labelled with CY3 or CY5 respectively using the Agilent Low Input Quick Amp labelling kit. The probes were purified, fragmented and analysed by competitive hybridization to a custom 4x44K gene expression micro-array (design ID 033946; courtesy of Jurg Bähler and colleagues). The results were scanned using an Agilent micro-array scanner, processed and analysed using the Agilent Genespring software.

2.5 ChIP. As described in main text

2.6 Restriction site mapping and PCR

Restriction site mapping by conventional and pulsed field gel electrophoresis was carried out as previously described (*4, 5*). PCR was carried out using either Taq polymerase (homemade or from Yorkshire Biosciences) or for amplicons in excess of 3kb using the dNTP pack Expand polymerase from Roche. Primer sequences are indicated in data table S1

2.7 DNA extraction, sequencing and analysis : DNA was extracted from the 5ml of yeast cultures using a protocol kindly supplied by Jacob Dalgaard of the Marie Curie Research Institute. 5ml of saturated culture was concentrated by centrifugation, spheroplasted using zymolyase 20T in 100µl 1M sorbitol, 50mM EDTA, concentrated by centrifugation once again, re-suspended in 0.2mL of DNAzol, vortex mixed, and the DNA precipitated with an equal volume of cold ethanol. The crude DNA was digested with ribonuclease and then pronase in 10mMTris.HCl, 1mM EDTA, pH8.0 ; 0.1%SDS, extracted between three and five

times with a 1:1 mixture of phenol and chloroform and then precipitated with ethanol once again prior to use. Pulsed field and conventional agarose gel electrophoresis were carried out as previously described.

Number	Name	Sequence	Comment	Position
Defining breakpoints				
108	76:1.3R	tgtatgcaagggaacgtgaa	BP1	I: 2259719-2259738
119	76:2.4L	gta gat gaa ata cca aac acg a	BP1	II:1779676-1779655
109	76:1.4L	tgt ctt ttg cgt cat cca ag	BP2	I:2260435-2260416
116	76:2.2R	taaccatttacactgctattgc	BP2	II:1779253-1779274
118	76:2.3R	tat aag gac gca gcg agc	BP3	II:1627446-1627463
222	2.1L new BP-76	ctaaagtcaatactctcttcttg	BP3	II: 569667-569689
106	76:1.2L	gcctaaatcaaaatgaagctga	BP4	I:1216880-1216859
117	76:2.3L	cgt tca tac gca agt ctt gt	BP4	II:1620962-1620943
105	76:1.1R	ttt gaa gca agg aaa gcc t	BP5	I:1216338-1216356
247	IMR2 1628928R	caa gc ttg taa gca taa tat gtt cag a	BP5 and 6	II: 1628928R
248	Chr2 569451F	cttgttgtcatacgaatgcactgt	BP6	II: 569451F
250	ChrI 1299001F	agtc gcggccgc cgattatttagctactcaactctgg	BP8	I:1299001F
251	ChrII 697111F	agtc gcggccgc tactagttctagtattctgaacagttc	BP7	II: 697111F
133	ReqQ3841R	aac gaa atg ggt aat agt ttt tcc ac	BP7 and 8	telomeric
Construction of targeting plasmids				
304	Vac8 LF	agct gcggccgc ctcgag tgtgaatgaccttatgagatttgaa		II:582794-5828818
305	Vac8 LR	acgt ggatcc att ttt att ttt att ggc aca tta tat tac		II:583765-583794
306	Vac8 RF	acgtggatcc tctgactttgtttatatgattgtt		II:585909-585932
307	Vac8 RR	ctgc gagctcgag cc aat cca ttt tgg aaa aag ttt c		II:586911-586934
309	Dps1 LF	agct gcggccgc ctcgag tgtaaacagatcagctatgtctga		II:683143-683166
310	Dps1 LR	acgt ggatcc ttt cat ttc tta ttt tca aaa gta ac		II:684118-684142
311	Dps1 RF	acgtGGATCC gacgagtttagtacgctgcatcttt		II:685359-685383
312	Dps1 RR	ctgc gagctcgag tca cta tga tcc aag cga cag aat a		II:687362-687386
314	Nht1 LF	agct gcggccgc ctcgag attccctatgatgttatcaatattg		I: 1218101-1218125
315	Nht1 LR	acgt ggatcc ag agc ttt ctg atg cag gca tca g		I:1219641-1219664
316	Nht1 RF	acgtggatcc atggtatggagggttttccatg		I:1220777-1220798
317	Nht1 RR	ctgc gagctcgag tca gga gtc aat tag tga gct tac		I:1221777-1221797
319	Ste4 LF	agct gcggccgc ctcgag cttttcgcttctactttggcatc		I:1294558-1294580
320	Ste4 LR	acgt ggatcc cta agc ttt tta aaa gcc atc ta		I:1295535-1295557
321	Ste4 RF	acgtggatcc taattttgcttaggatagcaatgta		I: 1296511-1296535
322	Ste4 RR	ctgc gagctcgag cac ttg aac aca tat ata ata ag		I: 1297508-1297527
324	BP5 LF	agct gcggccgc ctcgag tat gat ata atg agc ttg ttc ttt a		II: 1628726-1628750
325	BP5 LR	acgt ggatcc agtatattggtcaattagaaccaa		II: 1627727-1627750
326	BP5 RF	acgtggatcc tac tat gtt tcg ata aat tca ta		I: 1216628-1216650

327	BP5 RR	ctgc gagctcgag aatcaatgtccctgtacgcaaag		I: 1215628-1215650
618	CNP1 promoterR	GAA AAG TTC TTC TCC TTT ACT GTT AAT TAA tg cca tat taa gtt gtt cct atc aat t		Includes homology for SLiCE
619	CNP1 promoterF	ccgggtgacccggcggggacgaggcaagctaa ac tcg acc gtt tat gtt taa aac act tg		Includes homology for SLiCE
620	CNP1 LF	tcactatagggagaccggcagatccgcggc GGATCC tcg acc gtt tat gtt taa aac act tg		Includes homology for SLiCE
621	CNP1 LR	CGA TAC TAA CGC CGC CAT CCA GTT TAA ACGAttaa GTT GTT CCT ATC AAT TTC TTT TG		Includes homology for SLiCE
622	CNP1 RF	Gattacacatggcatggatgaactatacaaa ATG GCA AAG AAA TCT TTA ATG GCT		Includes homology for SLiCE
623	CNP1 RR	ATA CAC ATA CGA TTT AGG TGA CAC TAT AGA GGATCC GTT TAA TAA TTC TTT GGT AGA TAG		Includes homology for SLiCE
626	RAD21-PK9-L-R 1345014R	GGGTTAGGAATACCTCTAGCAGCAGAACCGGATAG TGA TGA AAG TAG CAT TCC AC		Includes homology for SLiCE
627	RAD21-PK9-L-F 1344301	GTTGTAAAACGACGGCCAGTGAATTCGAG**CTCGAG** Tgcttgaatacatcttccatc		Includes homology for SLiCE
628	RAD21-PK9-R-F 1345018	CCG GCG GGG ACG AGG CAA GCT AAA CAG ATC Gaggtcggttaatatttttttcaaaatc		Includes homology for SLiCE
629	RAD21-PK9-R-R 1345563R	CTC GAG GCC AGA AGA CTA AGA GGT GAA AGA ctcga GAA CTT TTC AAA TTC AAT ATC CC		Includes homology for SLiCE
640	CENP-CPK9-R-F	CCG GCG GGG ACG AGG CAA GCT AAA CAG ATC CAA TAC TAA TAG TGT GTT ATG GAT TTC		Includes homology for SLiCE
641	CENP-CPK9-R-R	CTC GAG GCC AGA AGA CTA AGA GGT GAA AGA ctcgag Tactagtttcgtttgtatctc		Includes homology for SLiCE
643	CENP-CPK9-L-R	GGGTTAGGAATACCTCTAGCAGCAGAACCGGA TCG TTC GTT TGG AAA ATC CCC TAT TCC		Includes homology for SLiCE
644	CENP-CPK9-L-F	GTTGTAAAACGACGGCCAGTGAATTCGAG**CTCGAG** CCA GCA CTA CCG GAA GTA AAG CAG		Includes homology for SLiCE

Table S1. Primers used in this work

3 Sequence assembly, analysis and bioinformatics

3.1 DNA Sequencing A single genomic library with an average insert size of 200 base pairs was prepared from CBS2777 and sequenced to a mean depth of 1000X using the Illumina GAIIX sequencing platform. This generated 168,752,150 72 bp paired end reads. Initial filtering of reads based on quality scores was performed using custom perl scripts resulting in approximately 163 million paired end reads for subsequent assembly. To assemble these sequences we used both *de novo* and reference based assembly approaches.

3.2 *De novo* assembly. All reads were assembled using CLC assembly cell (CLC Bio, Aarhus, Denmark). Initial *de novo* assemblies of these reads generated 5,850 contigs with an N50 of 56,116 base pairs (see Table S2 below). The longest single contig obtained was 278,267 base pairs in length, which was shown by blast to match to chromosome I of S. pombe with 99% query coverage and 99% maximum identity. Systematic investigation of these contigs revealed the presence of minimal *Escherichia coli* contamination in the reads. Although these reads made up less than 0.02 % of the total reads, they had a significant effect on the quality of the resulting assembly as judged by the length and number of contigs obtained (see Table S2). To remove these contaminating reads, all reads were assembled to a reference *E coli* genome sequence (K-12 substr. MDS42 DNA, AP012306) using CLC assembly cell. This revealed approximately 7-fold coverage of the *E. coli* genome. We therefore removed these reads from the read collection. In a final round of assembly optimisation, we investigated the consequences of sequence depth on assembly. Here we found that reducing the depth of coverage in conjunction with removing contaminating sequences further enhanced our assemblies. Given this, we artificially reduced the coverage of reads by 10 fold, which also allowed us to directly compare the data with that generated for *S. kambucha* and NCYC132 (*6*). Each collection of sequences generated a slightly different assembly and so a typical example is shown in Table S2 below. For comparison, Table S3 shows assembly statistics for *S. kambucha* and NCYC132 non-paired end sequences obtained from the Broad Institute.

No. of Contigs	N50	No. of Contigs > N50	Min	Median	Mean	Max	Total Length	Comments
5,850	56,116	73	116	573	2,905	278,267	17,069,490	Initial assembly with quality trimmed reads
1,404	79,573	49	133	492	9,331	390,466	13,112,921	Assembly after cleaning E. coli contamination
457	84,157	46	172	5,880	26,965	278,266	12,324,361	Illustrative assembly with approximately 100 fold coverage

Table S2 − Basic *de novo* sequence assembly statistics for CBS2777. The total length of the three chromosomes of the laboratory reference strain is 12,571,820 bp.

No. of Contigs	N50	No. of Contigs > N50	Min	Median	Mean	Max	Total Length	Comments
3,551	6,684	563	200	1,960	3,431	31,550	12,184,235	NCYC132
5,558	3,984	929	200	1,313	2,200	24,570	12,231,981	S. kambucha

Table S3 – Basic *de novo* sequence assembly statistics for NCYC132 and *S. kambucha* sequences. The raw reads were obtained directly from the *Schizosaccharomyces* group Database at the Broad Institute (*6*) The total length of the three chromosomes of the laboratory reference strain is 12,571,820 bp.

Given the depth of sequence coverage available to us, we utilised IMAGE (Iterative Mapping and Assembly for Gap Elimination) to refine our contigs (*7*). This tool attempts to close gaps in assemblies by comparison with a reference sequence. Using IMAGE resulted in the generation of a set of 275 contigs with significantly improved coverage statistics (see Table S4).

No. of Contigs	N50	No. of Contigs > N50	Min	Median	Mean	Max	Total Length	Comments
275	128,002	28	203	15,040	44,433	480,962	12,220,211	Post IMAGE

Table S4 – *De novo* sequence assembly statistics after processing contigs and reads with IMAGE.

The reliance of IMAGE on a reference sequence causes problems when dealing with rearranged strains as IMAGE should fail at known rearrangements as it will be unable to traverse the rearrangement. Therefore the contigs generated both pre and post IMAGE processing were screened for the presence of each of the known rearrangements within CBS2777. Every rearrangement could be identified in the pre IMAGE processed contigs and in all but one case were also present in the IMAGE processed contigs. The loss of one breakpoint in the IMAGE processed contigs reflects the proximity of the breakpoint to repeated sequences.

Every contig was checked by blast analysis against the reference laboratory strain using custom perl scripts. All contigs mapped appropriately to the reference sequence unless they contained a known rearrangement. In addition, the endpoint of every contig was checked. All the endpoints fell within known repeated sequences, low complexity sequence or tRNA sequences. The *de novo* assemblies failed to reveal any further evidence of additional rearrangements within the genome of CBS2777.

3.3 Testing for rearrangements. A variety of packages are available for using paired-end sequence data to investigate structural rearrangements. We utilised GASV (Geometric Analysis of

Structural Variants) (*8*) and SVDetect (*9*) to investigate if we could detect any additional rearrangements in these sequences not previously identified in the microarray analysis. Both packages were able to detect the previously identified rearrangements. Additional candidate rearrangements were identified by both packages. In every case, these corresponded to known distributed repeated regions within the genome and were not supported by evidence from the microarray study. Taken together with the *de novo* assemblies outlined above we therefore concluded no further rearrangements between single copy sequences exist in CBS2777, although we cannot exclude the possibility of additional inversions between repeated sequences.

3.4 Reference based assembly. To determine the nature of CBS2777 when compared with other S. pombe strains, reference based assembly was used to generate a sequence unique to CBS2777. For simplicity of analysis, two reference sequences representing CBS2777 were generated, one corresponding to the 3 chromosomes of the type strain, the second corresponding to the rearranged 4 chromosomes of CBS2777. The generation of a 3 chromosome unrearranged CBS2777 reference sequence was important as it allowed us to directly compare sequences between the laboratory reference, *S. kambucha* and NCYC132. The four chromosome reference sequence template was generated by manually editing the type strain sequence to correspond with the known rearrangements determined by microarray and confirmed by *de novo* sequence analysis. The same cleaned reads utilised for de novo assembly were also used for both reference-based assemblies (Tables S5 and S6).

Strain	**Chr1**	**Chr2**	**Chr3**	**Total Length**	**Comment**
Lab Ref Strain	5,579,133	4,539,804	2,452,883	12,571,820	
CBS2777	5,579,438	4,540,069	2,453,042	12,572,549	Assembled with respect to 3 chromosome reference
NCYC132	5,579,036	4,539,819	2,452,813	12,571,668	
S. kambucha	5,579,114	4,539,750	2,452,889	12,571,753	

TableS 5

Strain	**Chr1**	**Chr2**	**Chr3**	**Chr4**	**Total Length**	**Comment**
CBS2777	4,655,817	3,947,365	2,447,592	1,713,746	12,764,520	Assembled with respect to a 4 chromosome reference

Table S6 – Chromosome lengths for the reference based assembly of CBS2777. Here CBS2777 is assembled to the 4 chromosome reference sequence.

3.5 SNP Calling The CLC assembly cell basic SNP caller (find variations) was used for SNP calling across all three strains. An assembly file (.CAS format) was generated using CLC assembly cell which identified only uniquely mapping reads, thus removing repeat regions from our analysis. The assembly cell simple SNP caller has few options available. Standard parameter settings were used with the exception that a coverage threshold of 50% of the average coverage over the genome was used to limit calling of SNPs in poorly sequenced regions. To verify the quality of SNP calling, coding SNPs in NCYC132 and *S. kambucha* were compared with those identified at the time of sequencing and found to be consistent (*6*). Custom perl scripts were used to compare the SNPs between strains by systematic comparison with the laboratory reference sequence (see Table S7). Due to the inherent difficulties in correctly placing in/dels, we removed these from our analysis. To investigate the distribution of differences along the chromosomes the numbers of coding SNPs were determined in 10kb windows sliding 2 kb along each chromosome. The differences between two test strains can be inferred by systematic comparison with the laboratory reference. In this way, we were able to generate identity plots across each chromosome between each pairwise combination of strains representing the identity of coding sequence in these regions (see Table S7). These data visually demonstrate the similarity of CBS2777 to the other *S. pombe* strains tested. Furthermore, the distribution of identity suggests that these strains originated from a limited population of progenitors.

Chr 1	**Lab Ref Strain**	**CBS2777**	**NCYC132**	***S. kambucha***
Lab Ref Strain	-			
CBS2777	99.65	-		
NCYC132	99.58	99.63	-	
S. kambucha	99.73	99.78	99.66	-

Chr 2	**Lab Ref Strain**	**CBS2777**	**NCYC132**	***S. kambucha***
Lab Ref Strain	-			
CBS2777	99.5	-		
NCYC132	99.55	99.6	-	
S. kambucha	99.68	99.72	99.66	-

Chr 3	**Lab Ref Strain**	**CBS2777**	**NCYC132**	***S. kambucha***
Lab Ref Strain	-			
CBS2777	99.42	-		
NCYC132	99.54	99.6	-	
S. kambucha	99.59	99.72	99.72	-

Table S7 – The average percent identity between coding sequences for each pairwise combination of strains. Note that in/dels are excluded from this comparison.

3.6 Sequencing summary Taken together, the data presented here confirm that the 4 chromosome rearranged CBS2777 is a bona fide S. pombe strain. We can detect no additional rearrangements, although we cannot exclude inversions between repeated sequences, and we do not observe any significant differences between strains in terms of number or distribution of SNPs. Considering the effect of non-synonymous SNPs on coding sequences, we note that NCYC132 has the greatest number of coding sequences with at least one non-synonymous change. *S. kambucha* has the least number of non-synonymous changes. CBS2777 is closer to *S. kambucha* than to NCYC 132.

3.7 Small RNA Mapping and analysis. cDNA Libraries of were prepared from siRNA of strains CBS2776 and CBS 2777 using the NEBNext Small RNA Sample Prep kit using two rounds of size selection and sequenced using the SOLiD4 . SOLiD4 50bp colour space reads for each strain were processed with the small RNA pipeline of LifeScope 2.5.1. Briefly, the small RNA adaptor sequence was trimmed from reads before filtering against S. pombe sequences for tRNA, rRNA, snoRNA, snRNA1-5, telomerase RNA and the SRP RNA gene. Filtered reads where then aligned to the genome reference sequence. The LifeScope 'seed-extend' mapping algorithm allows for all seed length subsequences within each read to act as an alignment anchor to which neighbouring bases within the read are then added. Seed parameters were modified to better accommodate filtering and mapping of smRNAs of approx. 18bp in length. The default alignment seed length and miss-match (mm) allowance for read filtering was changed from 25bp & 3mm to 20bp & 1mm. The genome alignment seed was changed from 20bp & 1mm to 18bp & 1mm. Up to 30 mapping positions each read were subsequently recorded in BAM format for downstream analysis.

siRNA coverage at each base position for each strain was determined by processing the BAM files with bedtools (*10*) . All mapping positions for a particular siRNA, not just primary mapping position, were utilized as in Djupedal et al (*11*). To enable a comparison between CBS2777 and CBS2776, reads from each were mapped to the three chromosome reference genome. A custom perl script generated SVG files showing the coverage over the three centromeres, normalising for differences in depth of sequencing between the two libraries (Fig. S9). For simplicity of visualisation, coverage is averaged in five base pair windows. Visual inspection revealed no significant differences between the distribution of siRNAs between the rearranged CBS2777 and the un-rearranged CBS2776 and the distribution of reads over the three centromeric regions is equivalent in both strains. To quantify this similarity, Pearson's r was calculated for each chromosome at single base pair resolution, excluding those sites with no coverage in both strains, giving chromosome I: 0.97, chromosome II:0.92 and chromosome III:0.92 the visual inspection. We also confirmed that the relative proportion of reads mapping to each chromosome was unchanged between CBS2777 and CBS2776 (see Table S8).

The distribution of reads over each of the three centromeric regions corresponds with that previously reported by Djupedal et al (see Fig. S8). We do observe notable coverage peaks at three loci, two on chromosome I and one on chromosome II. Each of these peaks shows coverage of 4 times or more that of the nearest neighbours. The sequences at these regions were identified and in one case is within a known ncRNA. In the second case, the sequence is in close proximity to known ncRNAs. The remaining sequence is not currently annotated as a known siRNA or ncRNA (see Table S9). These data reveal no differences in the production of siRNA from either strain.

	CBS2777	**CBS2776**
Chr1	15%	15%
Chr2	25%	24%
Chr3	60%	61%

Table S8 - The relative proportions of siRNA reads mapping to each chromosome is the same in CBS2777 and CBS2776.

Position	**Sequence**	**Comment**
1:3763480-3763580	GACCGCACTAAAAGCATGGTACCAAAGCTCGAACATAGAAAGAAATCCAGAAAGGGATTTAAAGATTGACTTTTTCGACAAACTTCATGTTACAAGTCTTA	non-coding RNA, centromeric (predicted) [Source:PomBase Gene ID;Acc:SPNCRNA.232]
2:1605260-1605330	ATGAATATGGTTTTACTTATTTTACTACCCATGATGTCGTTGGTTAAAGACATGATGTTAAGGGTGAACCG	Close to two known ncRNAa - SPNCRNA.359 and SPNCRNA.360
1:3779770-3779820	GAAACCATTTTTCTTGTTTCTGCATGCATTCTTAGAAGAGATTTAAGCTTT	

Table S9– Sequences corresponding to the unusual peaks of siRNA coverage identified in both CBS2777 and CBS2776.

3.8 ChIP Seq. Analysis Reads were mapped to either the three chromosome laboratory reference strain or the four chromosome CBS2777 strain described here. Stampy and BWA were used to map reads with default settings (Stampy Ref pmid 20980556, BWA ref pmid: 19451168). This results in the reporting of primary mapped reads with non uniquely mapped reads being assigned quality scores of less than 25 (ref pubmed id: 22709551). Where appropriate, non-uniquely mapping reads were filtered out on the basis of quality score to generate uniquely mapped reads. Comparisons between reads mapped to the 3 chromosome and 4 chromosome strains were complicated by the presence of duplicated DNA in the 4 chromosome strain. Thus read counts were normalised first to total number of reads sequenced for each library and then by the number of reads mapping to a unique portion of the genome in both strains, namely centromere 3. Coverage plots were created in R from WIG files generated by BEDTools (BEDTools reference pmid: 20110278). Coverage plots were smoothed using

the default running medians scatter plot smoothing in R. For coverage plots across entire chromosomes data points were plotted every 100 basepairs. To calculate the appropriate normalisation parameters we relied on the data shown in table S10. In brief, unmapped reads were removed from the bamfiles and the total number of reads mapping to the laboratory three chromosome reference strain using samtools. Reads mapping to each centromere were then determined. This was straightforward for centromeres 1 and 3 (I: 3752589- 3823195 and III: 1069903- 1151934 respectively) which are each present in single copy in both CBS2776 and CBS2777. However, centromere 2 presents complexities due to its rearrangement and duplication within CBS2777. We therefore counted reads mapping to three regions within the genome and assigned them to centromere 2 on the basis that in at least one of the two strains these regions contained reads mapping around a centromere. Centromere 2 in CBS2776 maps within II:1518797-1741522. In the rearranged CBS2777 additional centromeric reads are found at I: 1216651-1266651 and II: 569545-619645. It is important to note that these read counts represent individual reads, not the captured sequenced fragments. Greater than 99% of the reads have a mapped pair leaving less than 1% of the data as singleton reads. Thus a reasonable approximation of the number of sequences captured is half the number of reads shown in table S10. For the purposes of normalisation where we are interested in the percentage of total reads mapping to each centromeric region, this is of no consequence. To investigate the impact of reads which map to more than one location, we generated this data for reads with a mapping quality of 40 or greater (uniquely mapping high quality reads) or allowed all quality scores (therefore including the primary mapping position for multi-mapped reads.

Normalized read coverage following CNP1 and CNP3 ChIP-seq of strains CBS 2777 and CBS 2776

CenP1	Centromere 1	I:3752589-3764483	I:3764484-3769077	I:3769078-3773195	I:3773196-3777789	I:3777790-3790520	I:3715078-3752589	I:3790520-3823195						Chip Efficiency Correction	
		flanking repeats left	imrL	central core	IMR R	flanking repeats right	single copy L	single copy R			Totals	Sum Totals		2.580645129	
	CBS 2777	1510.67	7016.96	14884.64	7050.95	1249.05	1443.53	1202.86			34358.66	98475.40			
		2699.17	12820.71	28538.85	12938.12	2290.00	2631.45	2198.44			64116.74				
	CBS 2776	4628.86	27068.33	60745.75	27185.72	4191.61	3071.00	2544.20			129435.46	264165.24			
		4425.22	28460.08	64411.27	28470.08	4061.33	2656.56	2245.24			134729.78				
	Centromere 2	II:1598512-1616568	II:1616671-1620805	II:1620806-1627610	II:1627611-1631744	II:1631745-1647353	II:1570806-1598512	II:1647353-1677610	II:569545-619645	I:1216651-1266651					
		flanking repeats left	imrL	central core	IMR R	flanking repeats right	single copy L	single copy R	rearranged left	rearranged right	Totals	Sum Totals		Corrected Counts	Ratio
	CBS 2777	1708.78	3603.80	44956.17	4159.49	1113.17	570.57	470.66	12304.69	9386.27	78273.60	225230.61		581240.28	1.84
		3026.48	6885.54	86295.38	7847.78	2084.08	1171.99	984.74	21920.47	16740.54	146957.01				
	CBS 2776	6472.93	14826.96	105265.47	14491.07	5082.85	2337.31	2273.69	1988.45	2026.52	154765.24	316598.03		316598.03	
		6239.81	15665.30	111993.45	15825.29	4901.25	1954.17	2002.15	1637.72	1613.63	161832.79				
	Centromere 3	III:1069903-1092667	III:1092668-1097059	III:1097060-1101934	III:1101935-1106326	III:1106327-1140810	III:1047060-1069902	III:1140811-1151934							
		flanking repeats left	imrL	central core	IMR R	flanking repeats right	single copy L	single copy R			Totals	Sum Totals			
	CBS 2777	3884.65	7097.65	14710.87	7176.61	3706.27	725.56	356.11			37657.72	108237.23			
		6615.02	13404.32	28201.77	13540.87	6842.65	1329.78	645.09			70579.51				
	CBS 2776	8276.72	26354.78	59173.42	26375.11	10052.29	2016.41	914.18			133162.91	269286.71			
		7510.45	27025.56	62281.59	27109.34	9481.36	1890.34	825.16			136123.80				
CenP3	Centromere 1	I:3752589-3764483	I:3764484-3769077	I:3769078-3773195	I:3773196-3777789	I:3777790-3790520	I:3715078-3752589	I:3790520-3823195						Chip Efficiency Correction	
		flanking repeats left	imrL	central core	IMR R	flanking repeats right	single copy L	single copy R			Totals	Sum Totals		2.21	
	CBS 2777	2992.52	14640.47	21012.43	14705.29	2515.70	3257.26	2182.05			61305.71	119390.48			
		2718.71	14064.30	19817.56	14122.77	2308.15	3011.77	2041.51			58084.77				
	CBS 2776	3978.90	38321.58	66932.59	38530.80	3545.07	3264.31	2841.88			157415.12	282464.74			
		3785.23	30268.81	51789.47	30296.91	3431.24	2964.39	2513.57			125049.62				
	Centromere 2	II:1598512-1616568	II:1616671-1620805	II:1620806-1627610	II:1627611-1631744	II:1631745-1647353	II:1570806-1598512	II:1647353-1677610	II:569545-619645	I:1216651-1266651					
		flanking repeats left	imrL	central core	IMR R	flanking repeats right	single copy L	single copy R	rearranged left	rearranged right	Totals	Sum Totals		Corrected Counts	
	CBS 2777	3427.62	4880.64	53481.06	5699.60	2446.93	968.15	782.83	16743.23	10872.34	99302.40	192329.80		425874.90	1.41
		3119.58	4632.77	50748.15	5399.46	2247.03	1183.08	989.27	14941.92	9766.15	93027.41				
	CBS 2776	6322.82	19714.95	106965.45	20471.63	5102.45	2491.82	2487.94	1558.62	1568.37	166684.05	301456.48		301456.48	
		5958.12	16609.40	82917.80	17056.85	4765.00	2310.85	2208.42	1487.64	1458.35	134772.43				
	Centromere 3	III:1069903-1092667	III:1092668-1097059	III:1097060-1101934	III:1101935-1106326	III:1106327-1140810	III:1047060-1069902	III:1140811-1151934							
		flanking repeats left	imrL	central core	IMR R	flanking repeats right	single copy L	single copy R			Totals	Sum Totals			
	CBS 2777	7617.87	16008.55	21511.54	16100.04	6530.79	1191.41	608.73			69568.93	136059.62			
		6538.55	15691.91	20733.23	15713.33	5894.62	1340.48	578.58			66490.69				
	CBS 2776	7379.53	34276.08	65400.48	34354.09	9239.65	2712.56	1021.74			154384.13	283177.11			
		6872.03	29863.76	50265.95	29900.77	8626.37	2372.01	892.09			128792.98				

Table S10 The figures refer to the normalized read coverage for the respective regions in the indicated experiments. The raw read numbers in the CBS 2777 experiments were corrected for differences in the efficiencies of the ChIP between by taking the respective ratios of the numbers of the total of the CEN1 and CEN3 reads (ChIP Efficiency Correction = (2776 (CEN1 +CEN3) / 2777 (CEN1 +CEN3)). This calculation assumes all reads outside the specified intervals represent non-specific background and is necessary to allow comparison between the CBS 2777 and CBS 2776 experiments .

Fig. S1. Characterization of the breakpoints 2 and 5 in strain CBS 2777 by PCR and restriction site mapping.

A. Breakpoints 2 and 5 were initially identified as specific to CBS 2777 by PCR using the indicated primers on the basis of the microarray data in figure 2. Primers are specified using the notation established in figure 4 immediately above.

B. Physical linkage between sequences predicted to flank breakpoints 2 and 5 was established by pulsed field gel electrophoresis, filter transfer and hybridization.

C. Mapping the Bp5 and ectopic *dgdh* repeats on CBS chromosome 1 using sequence targeting and restriction enzyme digestion. The BP5 and the ectopic *dgdh* repeats in CBS 2777 chromosome 1 were targeted with a *ura4* gene flanked by an AscI site and DNA from the resulting targeted clones together with that from CBS 2777 was digested with SfiI alone or SfiI and AscI and analysed by filter hybridization following pulsed field gel electrophoresis. The sizes of the cognate fragments are as predicted from the sequence assembly.

Fig. S2. Characterization of the breakpoints 1, 3 and 4 in strain CBS 2777 by PCR and restriction site mapping.

A. Breakpoint 1was established by PCR using the indicated primers.

B. Breakpoints 3 and 4 were initially identified individually by PCR and subsequently linked by long range PCR which identified a specific 8.4kb product linking segments II.1 and I.2 and the absence of the native arrangement of the central core of centromere II.

C. DNA extracted from CBS 2777 was digested with the indicated restriction enzymes size fractionated by agarose gel electrophoresis, filer transferred and hybridized with the indicated probes which are predicted on the basis of the sequence of the laboratory strain to lie adjacent to the breakpoints 3 and 4. The sizes of the restriction fragments match those predicted on the basis of the assembly of the laboratory strain (d)

D. indicates the arrangement of the probes and restriction sites predicted to occur in the sequences adjacent to the breakpoints 3 and 4. The restriction fragments observed in a precisely match those predicted in b. As elsewhere the arrows represent the orientation of the corresponding DNA with respect to the laboratory strain sequence. The figures above the diagram refer to the breakpoints on the sequence of the laboratory strain with " ' " indicating that corresponding sequence was in an inverted or complementary arrangement. Genes are indicated as red boxes, disposed as to the coding strand.

Fig. S3. Characterization of the breakpoints 7 and 8 in strain CBS 2777 by PCR and restriction site mapping.

A Breakpoints 7 and 8 were initially identified by restriction enzyme digestion, agarose gel electrophoresis and filter hybridization using probes predicted on the basis of the CGH analysis and other data to lie in a sub-telomeric position. The fragment specific to CBS 2777 at breakpoint 7 was Bal31 nuclease sensitive but that at breakpoint 8 was not.

B Breakpoints 7 and 8 were flanked by the *tlh1* gene. PCR using primers predicted on the basis of the micro-array data to be immediately centromere proximal of the respective breakpoints were used together with *tlh1* gene primers to demonstrate linkage of the respective sequences.

Fig. S4. Characterization of the centromeric DNA on chromosomes 2 and 4 in strain CBS 2777 by pulsed field gel electrophoresis, restriction analysis and sequence targeting.

A Physical linkage between sequences predicted to flank the re-arranged central core of chromosome II on the basis of the map shown in figure 1B of the main text was established by pulsed field gel electrophoresis, filter transfer and hybridization with the indicated probes.

B Restriction enzyme maps of the regions flanking the centromeres on chromosomes 2 and 4 of CBS 2777 were consistent with the results shown in A.

C Mapping the centromere flanking FNA on chromosomes 2 and 4 using sequence targeting and restriction enzyme digestion. The *dps1, vac8, nht1* and *ste4* genes on chromosomes 2 and 4 (main text figure 1c) were targeted with a *ura4* gene flanked by an AscI site and DNA from the resulting targeted clones together with that from untargeted CBS 2777 was digested with AscI and analysed by filter hybridization with either of the two chromosome specific probes following pulsed field gel electrophoresis. The sizes of the cognate fragments are as predicted from the sequence assembly.

Fig. S5. characterization of the breakpoint 6 in strain CBS 2777 by PCR and restriction site mapping.

A. Breakpoint 6 was initially identified by PCR using the indicated primers

B. Physical linkage of the sequences flanking breakpoint 6 to one another and the presence of the *dgdh* repeats adjacent to the breakpoint was confirmed by pulsed field gel electrophoresis and filter hybridization using the indicated probes.

Fig.S6. The average percentage identity between coding sequences for pairwise combinations of strains shown across the genome. Note that in/dels are excluded from this comparison.

Fig.S7. Mapping of Cnp1 and Cnp3 to the un-rearranged centromeres on chromosomes 1 (A) and 3 (B) in CBS 2777. Mate paired reads were mapped to the respective chromosome assemblies using the methodology used in Fig. 3A of the main text. Ectopic reads on chromosome 1 correspond to the fragment of the *imr2* sequence at breakpoint 5. The presence of reads at the ends of chromosome 3 correspond to rDNA sequence which we ascribe to contamination of ChIP material with this repeated DNA.

Fig. S8. siRNA coverage depth for CBS 2777 (red) against CBS 2776 (blue). The three identified peaks of coverage are indicated with an asterisk. The x-axis scale corresponds to the position in basepairs on each chromosome. The coverage depth is normalised according to the total number of reads sequenced from each library.

Fig. S9. Mapping of Swi6, Rad21 and small RNAs to the un-rearranged centromere on chromosome 2 in CBS 2776. Reads were mapped to the chromosome assemblies using the methodology used in Fig.4 of the main text.

Fig. S10. Correlated binding of Rad 21 and Swi6 across the euchromatic portion of the pericentromeric DNA of CBS 2777 chromosomes 2 and 4 and across the same sequences in CB S2776. Traces illustrate the binding of the respective proteins across the indicated intervals with the annotated genes illustrated below.

Fig. S11. Correlated binding of Rad 21 and Swi6 across the euchromatic portion of the pericentromeric DNA of CBS 2777 chromosomes 2 and 4 and across the same sequences in CBS 2776. Traces illustrate the binding of the respective proteins across the indicated intervals with the annotated genes illustrated below.

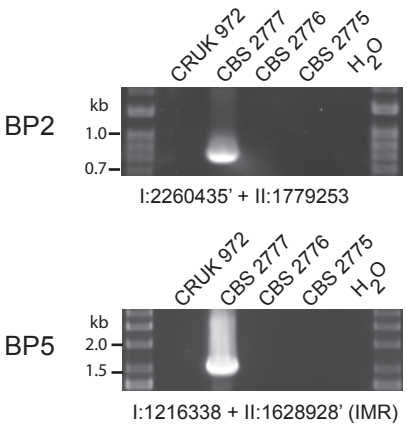
Fig. S12. Correlated binding of Rad 21 and Swi6 across CBS 2777 chromosome 4 showing widespread correlated binding. Duplicates are colour coded; green/yellow or blue/red. The binding to breakpoint 6 arises as a result of the presence of *dgdh* repeats adjacent to this breakpoint

References

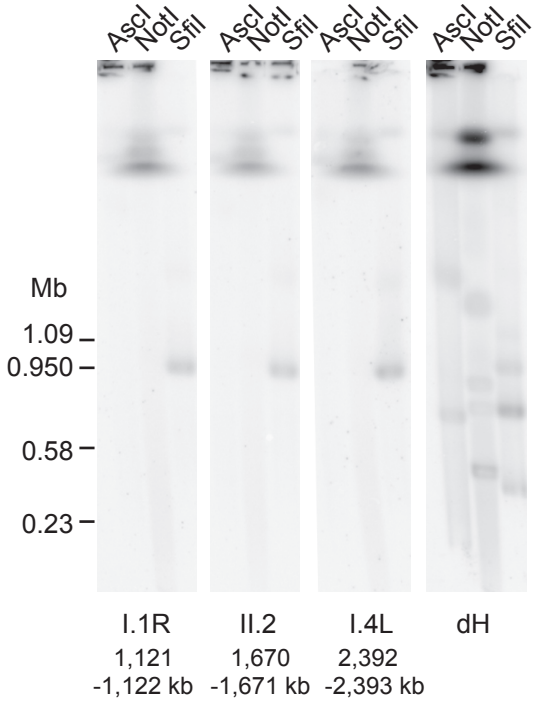
1. **W. R. Brown *et al.*, A Geographically Diverse Collection of Schizosaccharomyces pombe Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity. *G3 (Bethesda)* 1, 615 (Dec, 2011).**
2. **J. Bahler *et al.*, Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast* 14, 943 (Jul, 1998).**
3. **Z. Xu *et al.*, Site-specific recombination in Schizosaccharomyces pombe and systematic assembly of a 400kb transgene array in mammalian cells using the integrase of Streptomyces phage phiBT1. *Nucleic Acids Res* 36, e9 (Jan, 2008).**
4. **W. R. Brown, A physical map of the human pseudoautosomal region. *Embo J* 7, 2377 (Aug, 1988).**
5. **W. R. Brown *et al.*, Structure and polymorphism of human telomere-associated DNA. *Cell* 63, 119 (Oct 5, 1990).**
6. **N. Rhind *et al.*, Comparative functional genomics of the fission yeasts. *Science* 332, 930 (May 20, 2011).**
7. **I. J. Tsai, T. D. Otto, M. Berriman, Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol* 11, R41 (2010).**
8. **S. Sindi, E. Helman, A. Bashir, B. J. Raphael, A geometric approach for classification and comparison of structural variants. *Bioinformatics* 25, i222 (Jun 15, 2009).**
9. **B. Zeitouni *et al.*, SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics* 26, 1895 (Aug 1, 2010).**
10. **A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841 (Mar 15, 2010).**
11. **I. Djupedal *et al.*, Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. *EMBO J* 28, 3832 (Dec 16, 2009).**

Brown et al; Fig. S1

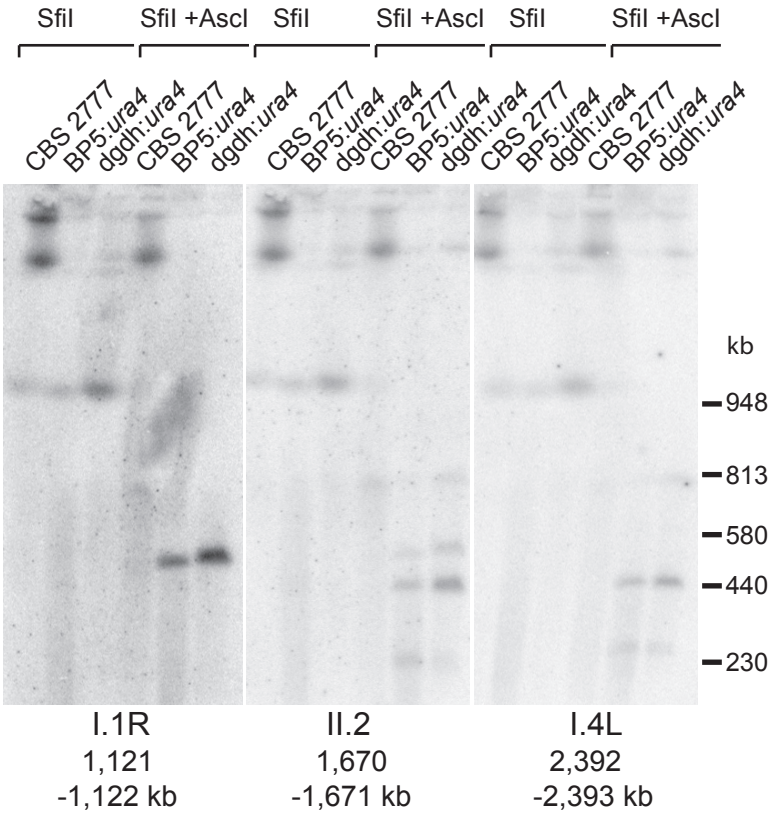
a: breakpoints 2 and 5



b: linkage of sequences around breakpints 2 and 5

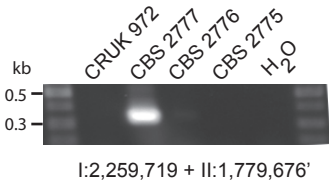


b: placing breakpoint 5 and ectopic dgdh repeats on CBS 2777 chromosome 1

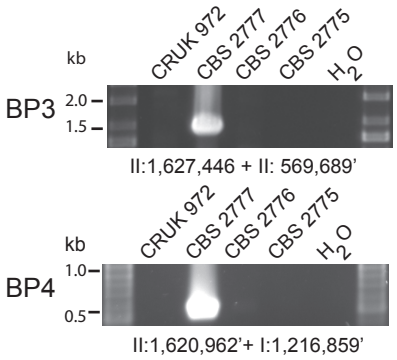


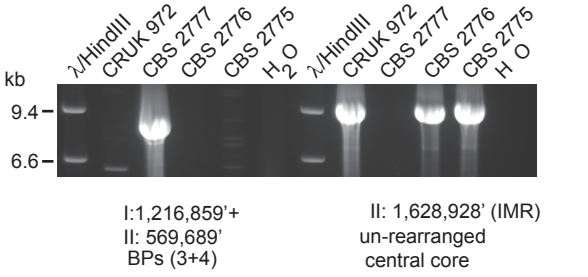
Brown et al; Fig. S2

a: breakpoint 1



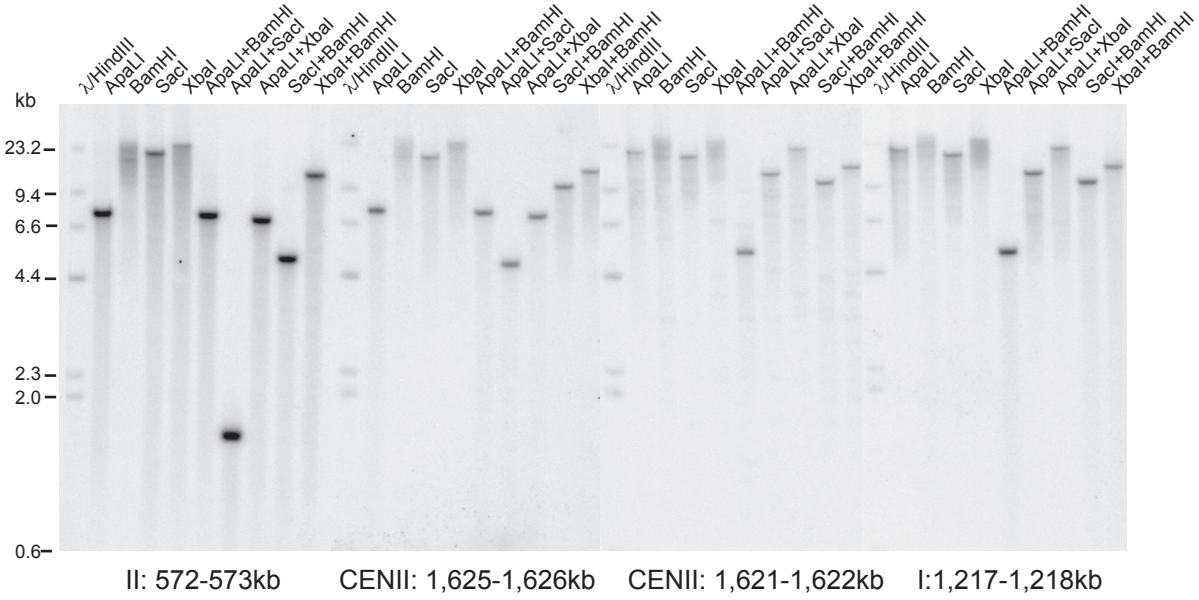
b: rearranged central core of chromosome II, breakpoints 3 and 4



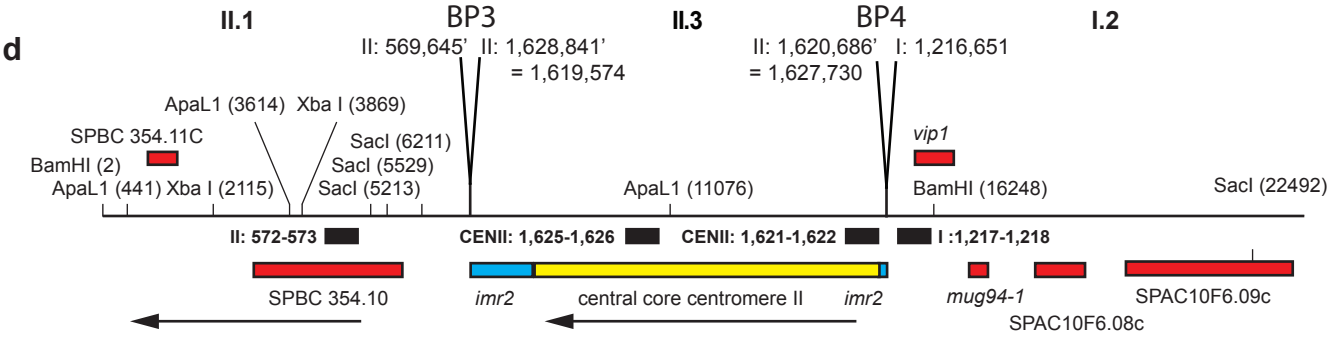


linking BPs 3 and 4 by PCR across the central core

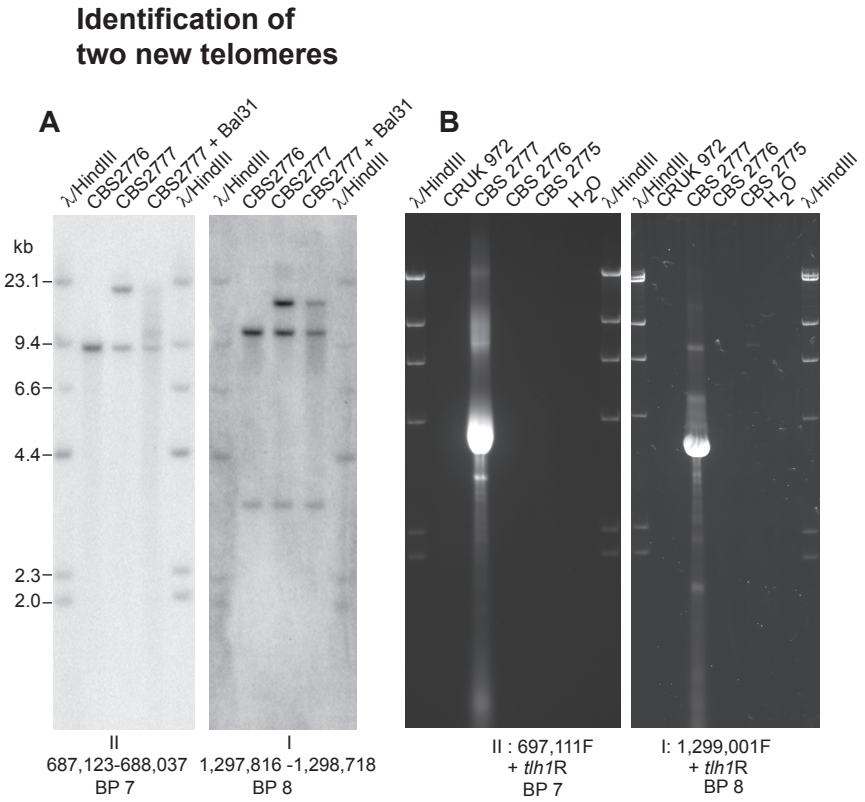
c



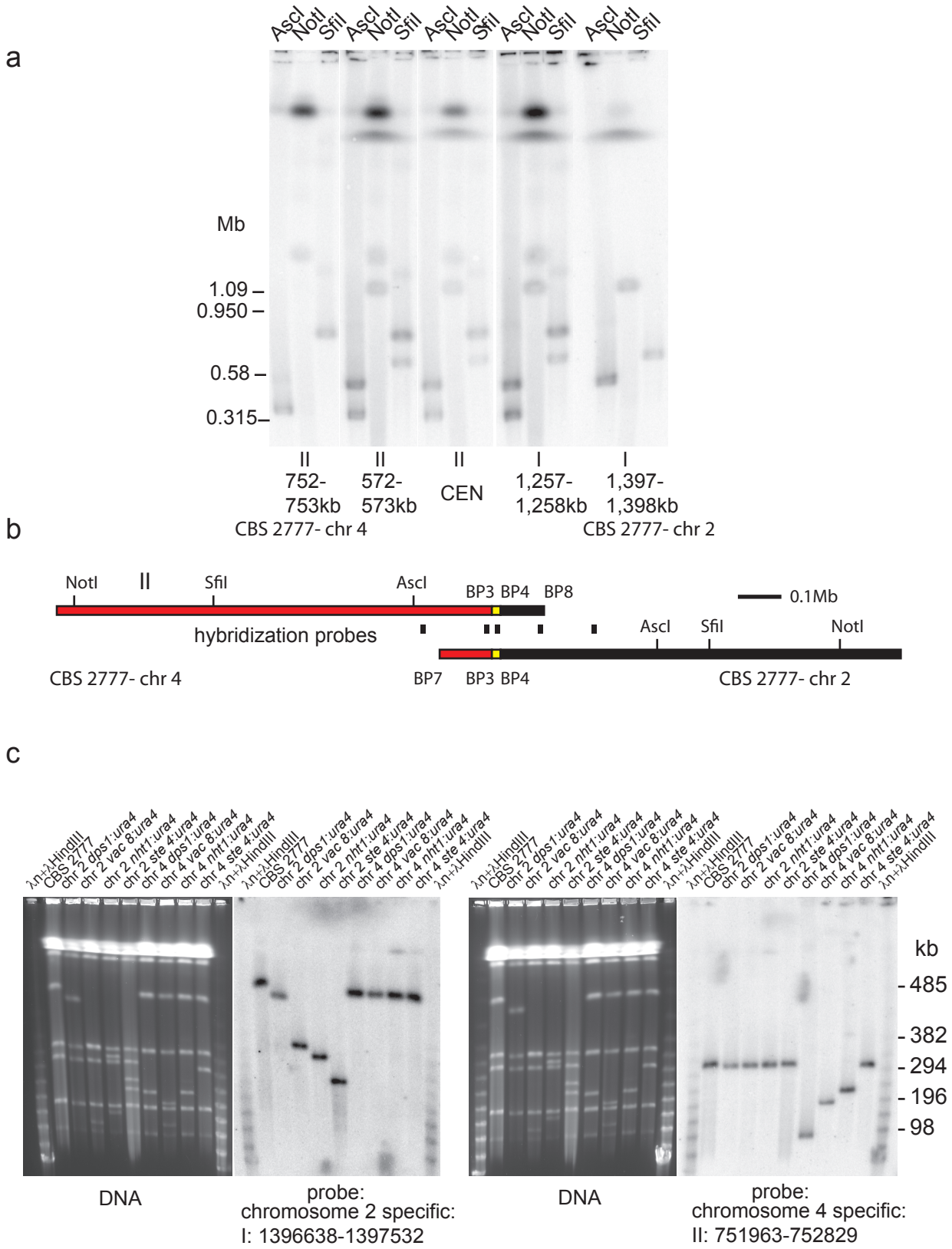
d



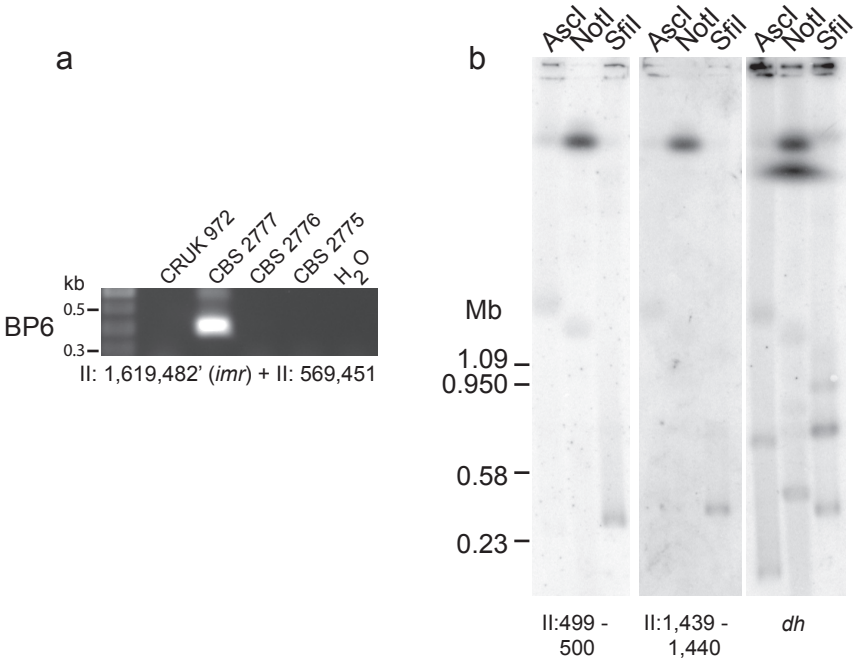
Brown et al; Fig. S3



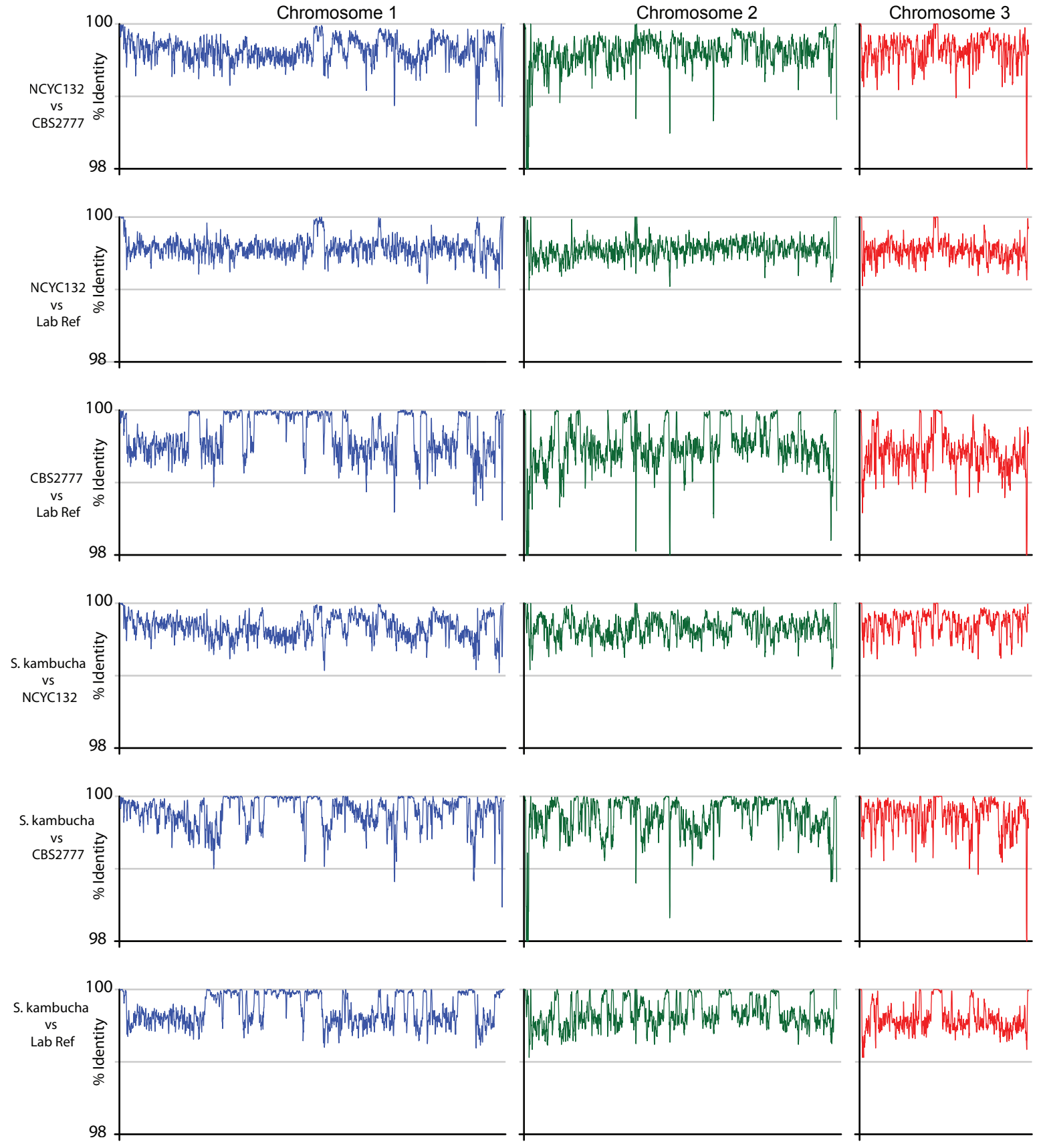
Brown et al; Fig. S4



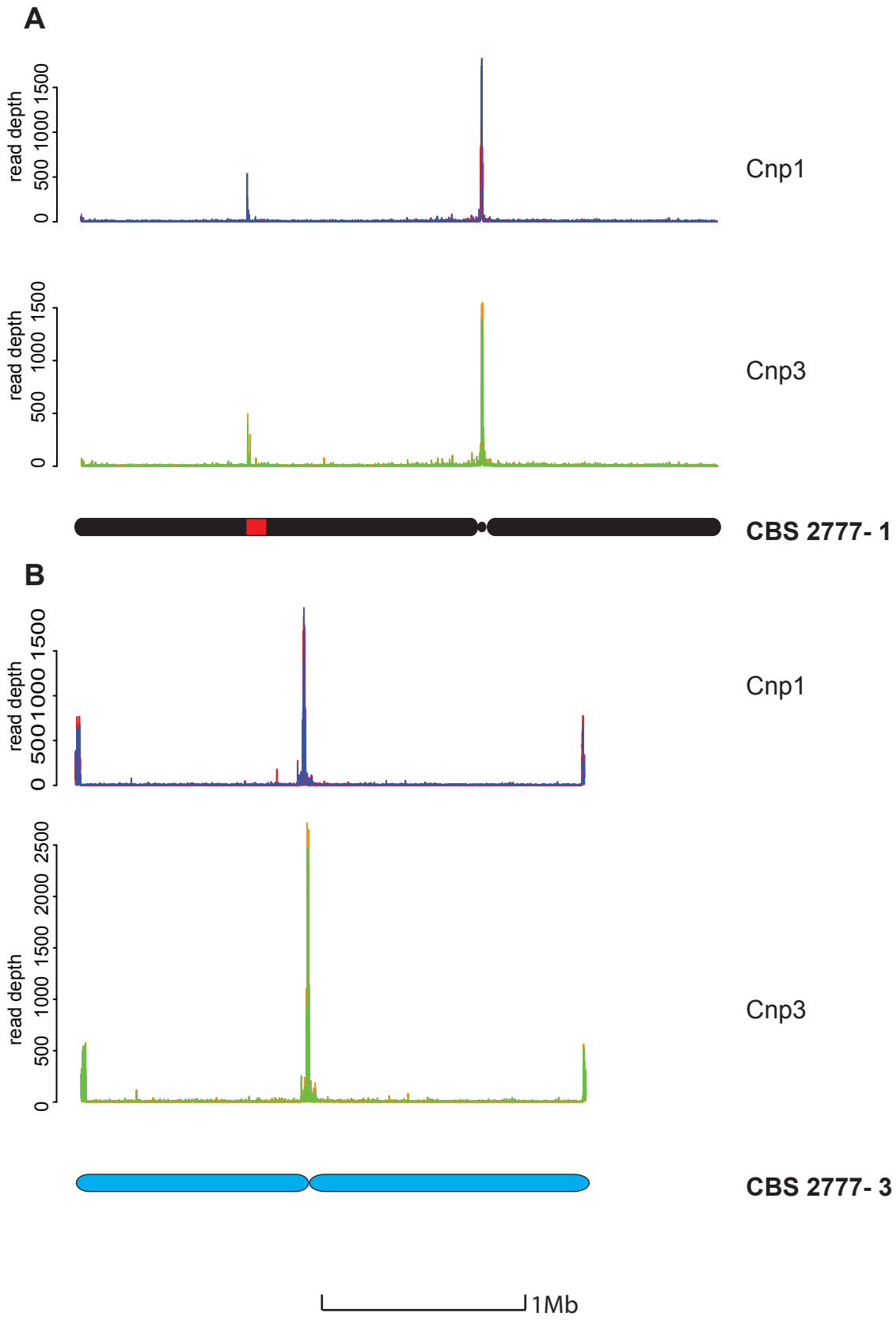
Brown et al; Fig. S5



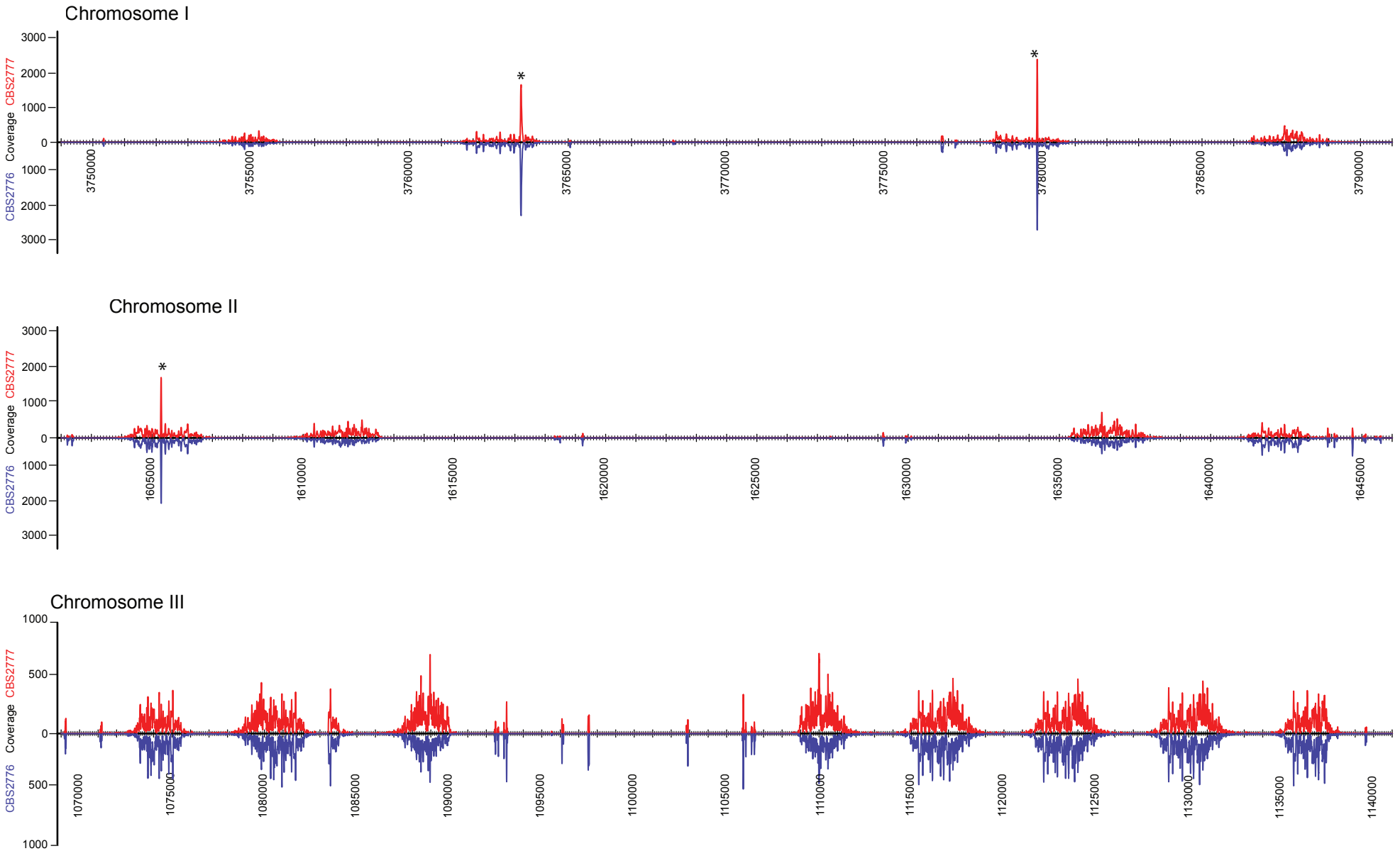
Brown et al; Fig. S6



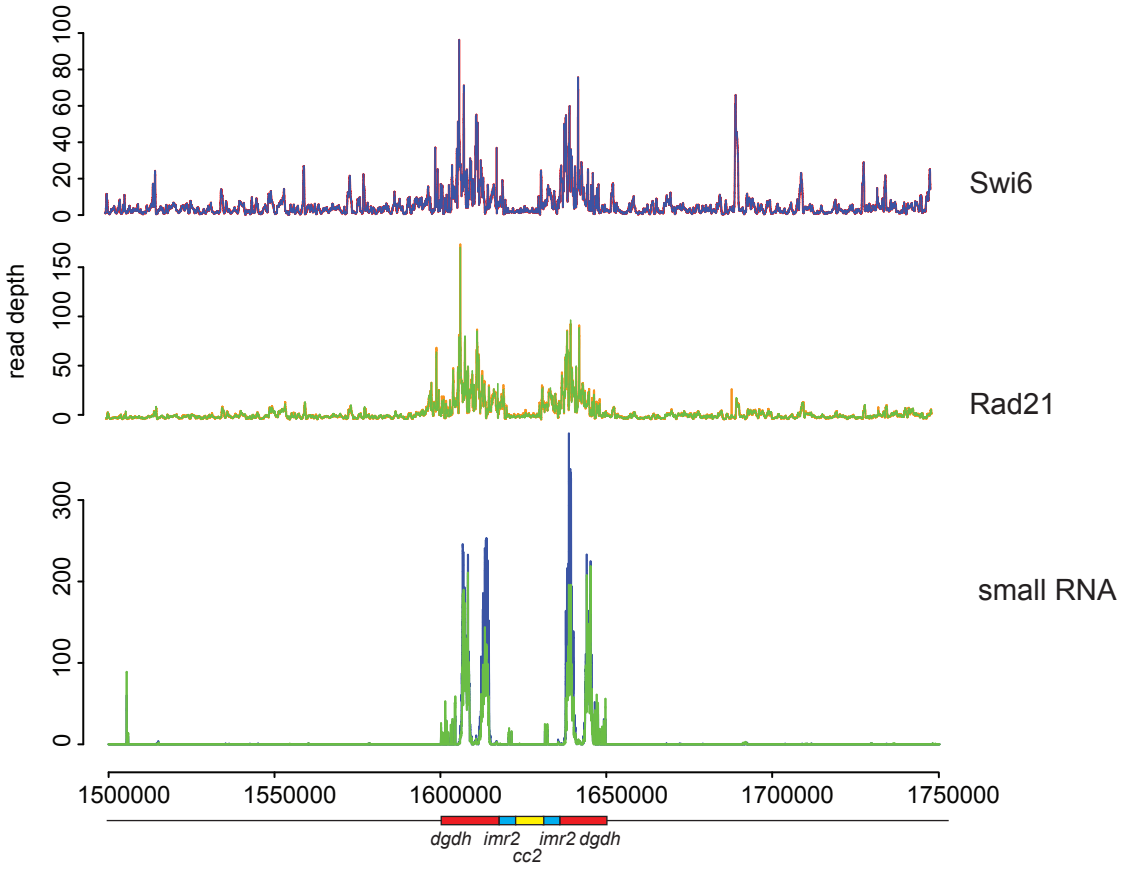
Brown et al; Fig S7



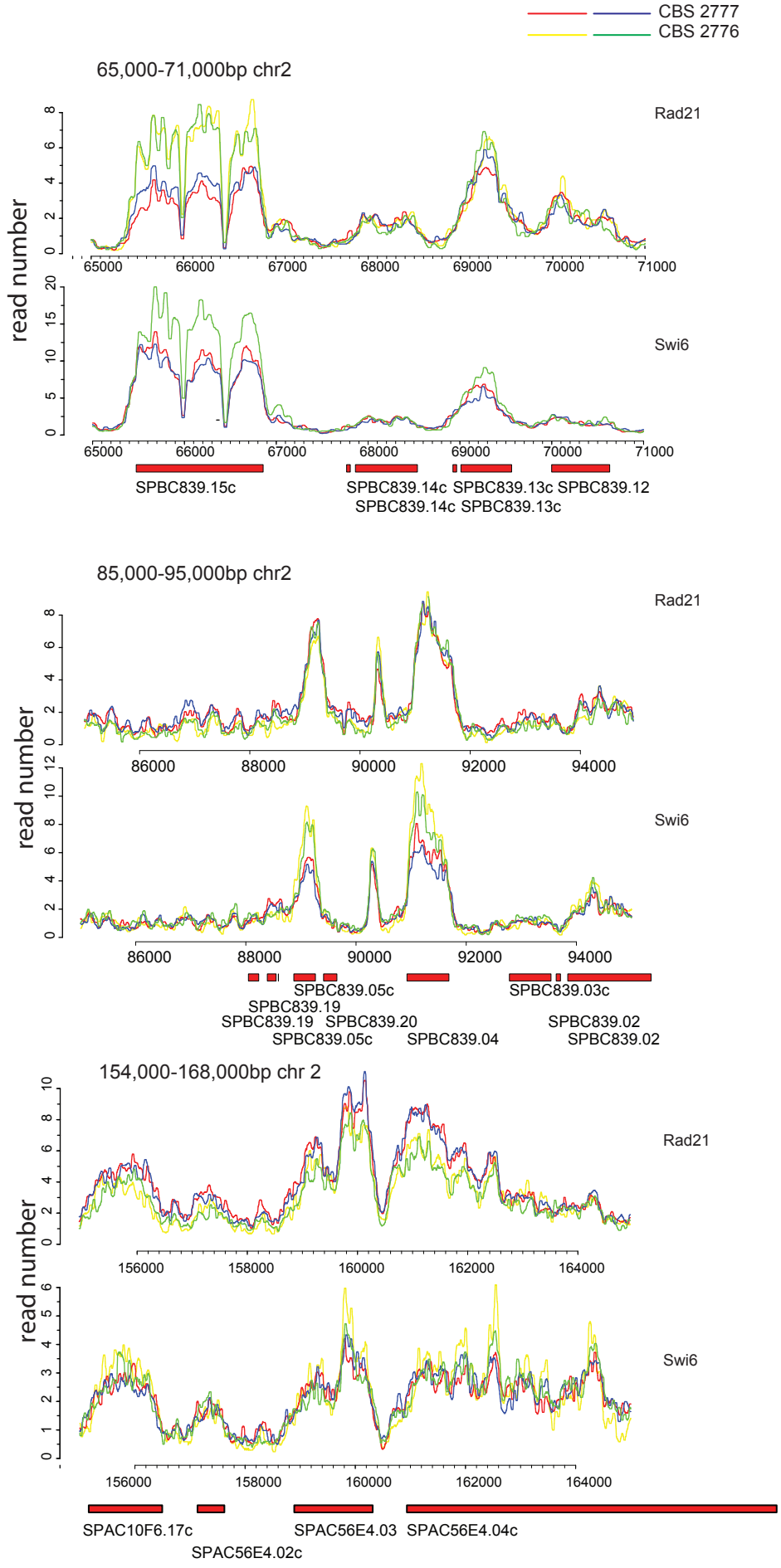
Brown et al; Fig S8



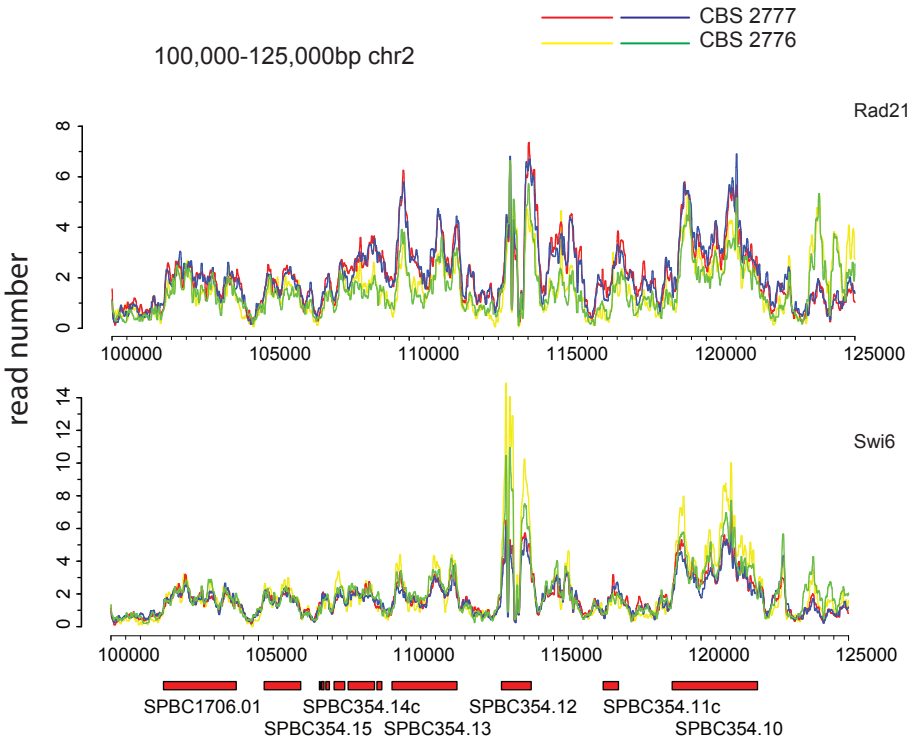
Brown et al; Fig. S9



Brown et al; Fig. S10



Brown et al; Fig. S11



Brown et al; Fig. S12