

SUPPLEMENTARY DATA

Acentric chromosome ends are prone to fusion with functional chromosome ends through a homology-directed rearrangement

Yuko Ohno¹, Yuki Ogiyama^{1,†}, Yoshino Kubota¹, Takuya Kubo² and Kojiro Ishii^{1,3,*}

¹ Graduate School of Frontier Biosciences, Osaka University, Suita, Osaka 565-0871, Japan

² Graduate School of Environmental Science, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan

³ Institute for Academic Initiatives, Osaka University, Suita, Osaka 565-0871, Japan

* To whom correspondence should be addressed. Tel.: +81 6 6879 4660; Fax: +81 6 6879 4660;
Email: ishii@fbs.osaka-u.ac.jp

† Present address: Institute of Human Genetics, CNRS UPR 1142, 34396 Montpellier, France

Supplementary Methods

Bayesian estimation of strain-dependent effects on survivor emergence

To evaluate the joint probability of telomere-fusion survivor emergence at the $\Delta cen1$ screens in wild-type and mutant strain backgrounds, we developed a Bayesian generalised linear mixed model that allowed us to estimate the posterior distributions of strain-dependent effects on survivor emergence as random effects by fitting to the experimental data. Because the random effect parameter for each screen is estimated based on a subset of all of the data, by definition, hierarchical Bayesian modelling can provide a better inference by choosing a variance parameter that refers to the dispersion of whole strains. Here, $p(X, Y, Z)$ was used as a general notation for the joint probability, where X , Y and Z represent the telomere-fusion, removal of $cen1$ -reintegrated false-positive survivors, and initial drug resistant-survivor emergence, respectively (40). The $p(X, Y, Z)$ notation can be decomposed into three probabilities, i.e., $p(X, Y, Z) = p(X | Y, Z) p(Y | Z) p(Z)$, where the $p(A | B)$ notation refers to the conditional probability that event A occurs under the condition that event B occurs. All of these decomposed probabilities, $p(X | Y, Z)$, $p(Y | Z)$, and $p(Z)$, were estimated using a common framework for Bayesian logistic regression, of which the linear predictor was the sum of the intercepts, replicates, and strain-dependent effects. Hierarchical prior distributions were specified for the effects of replicates and strains as the Gaussian distribution around mean zero, whereas the intercepts were non-hierarchical priors and the variances were non-informative priors. The relative effect of each strain was evaluated as the difference in log odds of $p(X, Y, Z)$ between mutant and wild-type posteriors.

The posterior distributions of all parameters in the Bayesian statistical model were estimated using the Monte Carlo Markov chain (MCMC) method. Sampling from the marginal posterior distributions was performed using the Gibbs sampling software JAGS 3.4.0 (<http://mcmc-jags.sourceforge.net/>). The posterior samples were obtained by three independent chains in which 1000 values were sampled with a 5-step interval after 1000 burn-in MCMC steps. The convergences of the MCMC samples were confirmed such that all R-hat indexes for all parameters were close to unity. All statistical significances were checked by evaluating the 95% BCIs of the posterior distributions. The JAGS and R codes that are common for estimation of the probabilities of telomere-fusion, $cen1$ reintegration, and initial survivor emergence are shown below.

```
# BUGS code
for (i in 1:N.sample) {
    Y[i] ~ dbin(p[i], N[i])
    logit(p[i]) <- logit.p[i]
}
for (i in 1:N.sample) {
    logit.p[i] ~ dnorm(m[i], tau[1])
    m[i] <- alpha + beta[Group[i]]
}
alpha ~ dnorm(0, 1.0E-4)
for (j in 1:N.group) {
    beta[j] ~ dnorm(0, tau[2])
}
for (k in 1:N.tau) {
    tau[k] <- 1.0 / (s[k] * s[k])
```

```

        s[k] ~ dunif(0, 1.0E+4)
    }

# R code
library(rjags)

N.sample <- nrow(d)
N.group <- length(levels(d$group))
list.data <- list(
    N.sample = N.sample,
    N.group = N.group,
    N.tau = 2,
    Y = d$y,
    N = d$N,
    Group = d$group_id
)
inits <- list( # parameter initial values
    alpha = 0,
    beta = rnorm(N.group, 0, 0.1),
    s = c(1, 1)
)
n.burnin <- 1000
n.chain <- 3
n.thin <- 5
n.iter <- n.thin * 1000

model <- jags.model(
    file = "model.bug",
    data = list.data,
    inits = inits,
    n.chain = 3 # !!!
)
update(model, n.burnin) # burn in
post.mcmc.list <- coda.samples(
    model,
    c("alpha", "beta", "s"),
    thin = n.thin, n.iter = n.iter
)

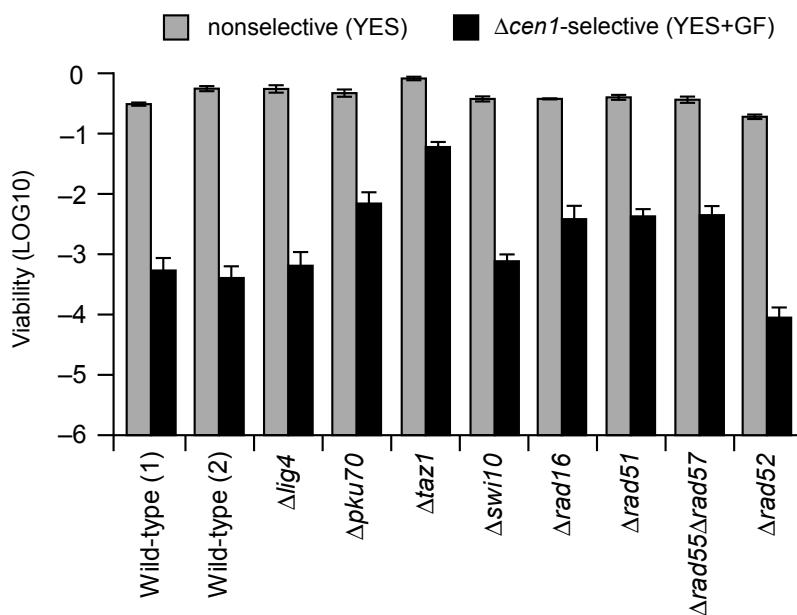
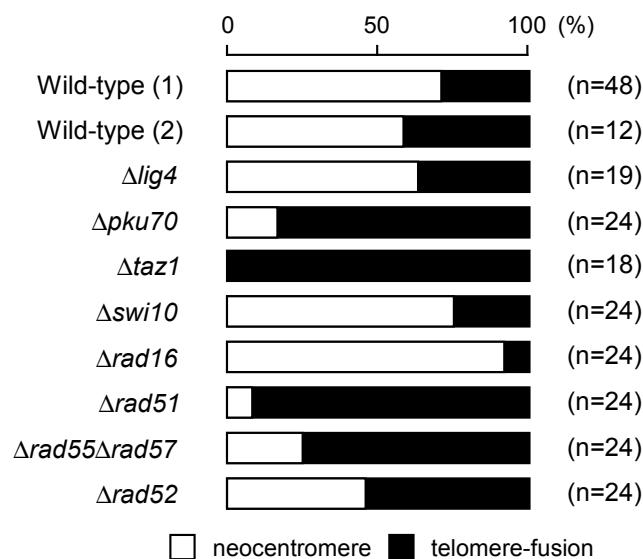
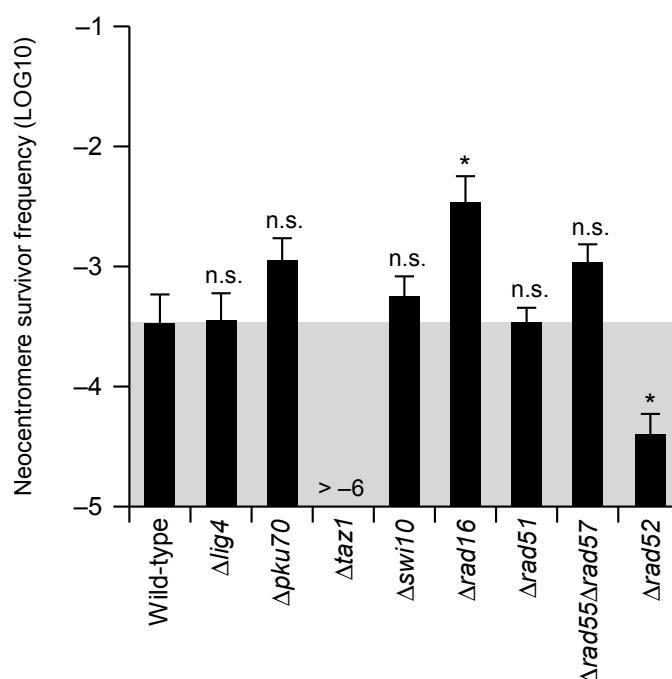
```

Supplementary Table S1. Yeast strain list

Strain name	Genotype
KYP33	<i>h</i> ⁻ <i>leu1 ura4</i>
KYP378	<i>h</i> ⁻ <i>leu1 ura4 cen1L::P^{adh1}-loxP cen1R::ura4⁺-loxP-kanR</i> ^{ORF}
KYP810	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-31 (1R;2L fusion)</i>
KYP812	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-33 (1L;3R fusion)</i>
KYP815	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-36 (1L;2R fusion)</i>
KYP821	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-42 (1L;2R fusion)</i>
KYP826	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-47 (1L;2L fusion)</i>
KYP828	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-49 (1L;3R fusion)</i>
KYP830	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-51 (1R;2R fusion)</i>
KYP832	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-53 (1R;3L fusion)</i>
KYP838	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-59 (1L;2R fusion)</i>
KYP386	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-63 (1R;3L fusion)</i>
KYP387	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-64 (1R;2R fusion)</i>
KYP388	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-65 (1L;2R fusion)</i>
KYP1272	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-73 (1L;2L fusion)</i>
KYP1273	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-77 (1R;2R fusion)</i>
KYP4478	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-96 (1R;2R fusion)</i>
KYP4479	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-97 (1L;2L fusion)</i>
KYP4480	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-98 (1L;2L fusion)</i>
KYP4481	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-99 (1R;2L fusion)</i>
KYP4482	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-100 (1R;2R fusion)</i>
KYP3948	<i>h</i> ⁻ <i>leu1 ura4 cen1L::P^{adh1}-loxP cen1R::ura4⁺-loxP-kanR</i> ^{ORF} <i>rad52-GFP::nat taz1-mCherry::kanR</i>
KYP2481	<i>h</i> ⁻ <i>leu1 ura4 cen1L::P^{adh1}-loxP cen1R::ura4⁺-loxP-kanR</i> ^{ORF} <i>Δpku70::hph</i>
KYP2428	<i>h</i> ⁻ <i>leu1 ura4 cen1L::P^{adh1}-loxP cen1R::ura4⁺-loxP-kanR</i> ^{ORF} <i>Δlig4::hph</i>
KYP3949	<i>h</i> ⁻ <i>leu1 ura4 cen1L::P^{adh1}-loxP cen1R::ura4⁺-loxP-kanR</i> ^{ORF} <i>Δrad16::nat</i>
KYP2339	<i>h</i> ⁻ <i>leu1 ura4 cen1L::P^{adh1}-loxP cen1R::ura4⁺-loxP-kanR</i> ^{ORF} <i>Δswi10::hph</i>
KYP674	<i>h</i> ⁻ <i>leu1 ura4 cen1L::P^{adh1}-loxP cen1R::ura4⁺-loxP-kanR</i> ^{ORF} <i>Δaz1::hph</i>
KYP2428	<i>h</i> ⁻ <i>smt-0 leu1 ura4 cen1L::P^{adh1}-loxP cen1R::ura4⁺-loxP-kanR</i> ^{ORF} <i>Δrad51::hph</i>
KYP2429	<i>h</i> ⁻ <i>smt-0 leu1 ura4 cen1L::P^{adh1}-loxP cen1R::ura4⁺-loxP-kanR</i> ^{ORF} <i>Δrad52::hph</i>
KYP2554	<i>h</i> ⁻ <i>smt-0 leu1 ura4 cen1L::P^{adh1}-loxP cen1R::ura4⁺-loxP-kanR</i> ^{ORF} <i>Δrad55::hph Δrad57::nat</i>
KYP2131	<i>h</i> ⁻ <i>leu1 ura4 cen3L::P^{adh1}-loxP cen3R::ura4⁺-loxP-kanR</i> ^{ORF}
KYP4885	<i>h</i> ⁻ <i>leu1 ura4 cen3L::P^{adh1}-loxP cen3R::ura4⁺-loxP-kanR</i> ^{ORF} <i>SAS(3L)::LEU2</i>
KYP4889	<i>h</i> ⁻ <i>leu1 ura4 cen3L::P^{adh1}-loxP cen3R::ura4⁺-loxP-kanR</i> ^{ORF} <i>SAS(3R)::LEU2</i>
KYP4890	<i>h</i> ⁻ <i>leu1 ura4 cen3L::P^{adh1}-loxP cen3R::ura4⁺-loxP-kanR</i> ^{ORF} <i>SAS(3L)::LEU2 SAS(3R)::LEU2</i>
KYP811	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-32</i>
KYP814	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-35</i>
KYP816	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-37</i>
KYP817	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-38</i>
KYP818	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-39</i>
KYP819	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-40</i>
KYP820	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-41</i>
KYP822	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-43</i>
KYP823	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-44</i>
KYP824	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-45</i>
KYP825	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-46</i>
KYP827	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-48</i>
KYP829	<i>h</i> ⁻ <i>leu1 ura4 Δcen1::P^{adh1}-loxP-kanR cd1-50</i>

Supplementary Table S2. Oligonucleotide list

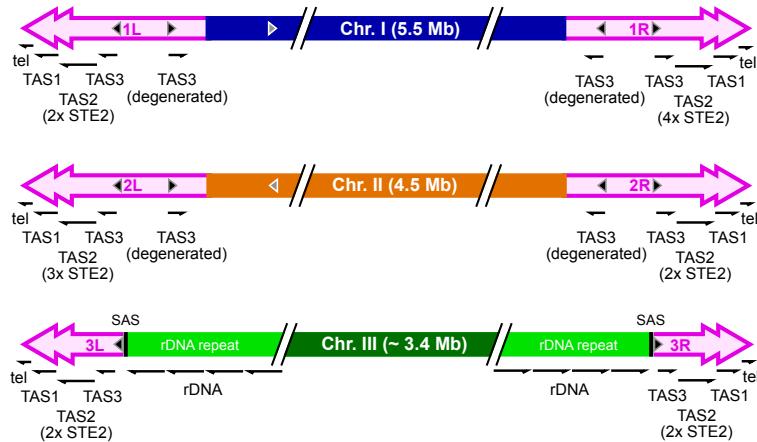
caPCR primers	
7926	5'-GGGTTGCAAAGTATGATTGTGGTAA-3'
14206	5'-GCATAAAGATGGTACTTCAA-3'
31268	5'-TGTTGAATGTCAGAACCAACTGTTGCAT-3'
Asp(1L)	5'-TCCGTACGCAATTATTGCA-3'
Asp(2L)	5'-TAACGGTTGCGTTCTTCC-3'
TAS1_TAIL-II	5'-GGTGAATTCACTAAGTGTAAACAGTAGTAGCAGTG-3'
TAS3_TAS2-F1	5'-GGAGTAAGTAGTGAGTCAG-3'
TAS1_TAS2-F1	5'-TTAACGATTTGGAGAGAG-3'
subtelrD-R4	5'-CCTCTAAGCCAGAACATCCG-3'
60241	5'-TTGAAGTACCATCTTATGC-3'
c-rDNA-t	5'-GAAGTTGTCAATGGAAGGG-3'
SAS-Rv1	5'-CGGGATCCGGCCGCAGGCTGGCTACTGTTTAC-3'
TAS1-Rv1	5'-CGGGATCCGGCCGCACCACGTAACCTTGTAAACC-3'
TAS2 ligation-independent cloning (LIC) primers	
TAS2cen-infusion-KS-R	5'-TTAACCATCCACGATGCATTGATATCGAATTCTGCAGC-3'
TAS2tel-infusion-KS-F	5'-TATTGGAAAGTCAGATGCATGCTTATCGATACCGTCGAC-3'
qPCR primers	
act1-56F	5'- ATCCAACCGTGAGAAGATGACT-3'
act1-56R	5'- AACACAGCTTGAATAGCAACATAAAAG-3'
STE2-1F	5'- GGAGAACAAAGAAGTAAGTAAAGTAAGAAA-3'
STE2-98R	5'- AAAGTACACCGCATGTT CCTATTAT-3'
Telomere probe	
tel	5'-CGTGTAAACCACGTAACCTTGTAAACCCGATC-3'

A**B****C**

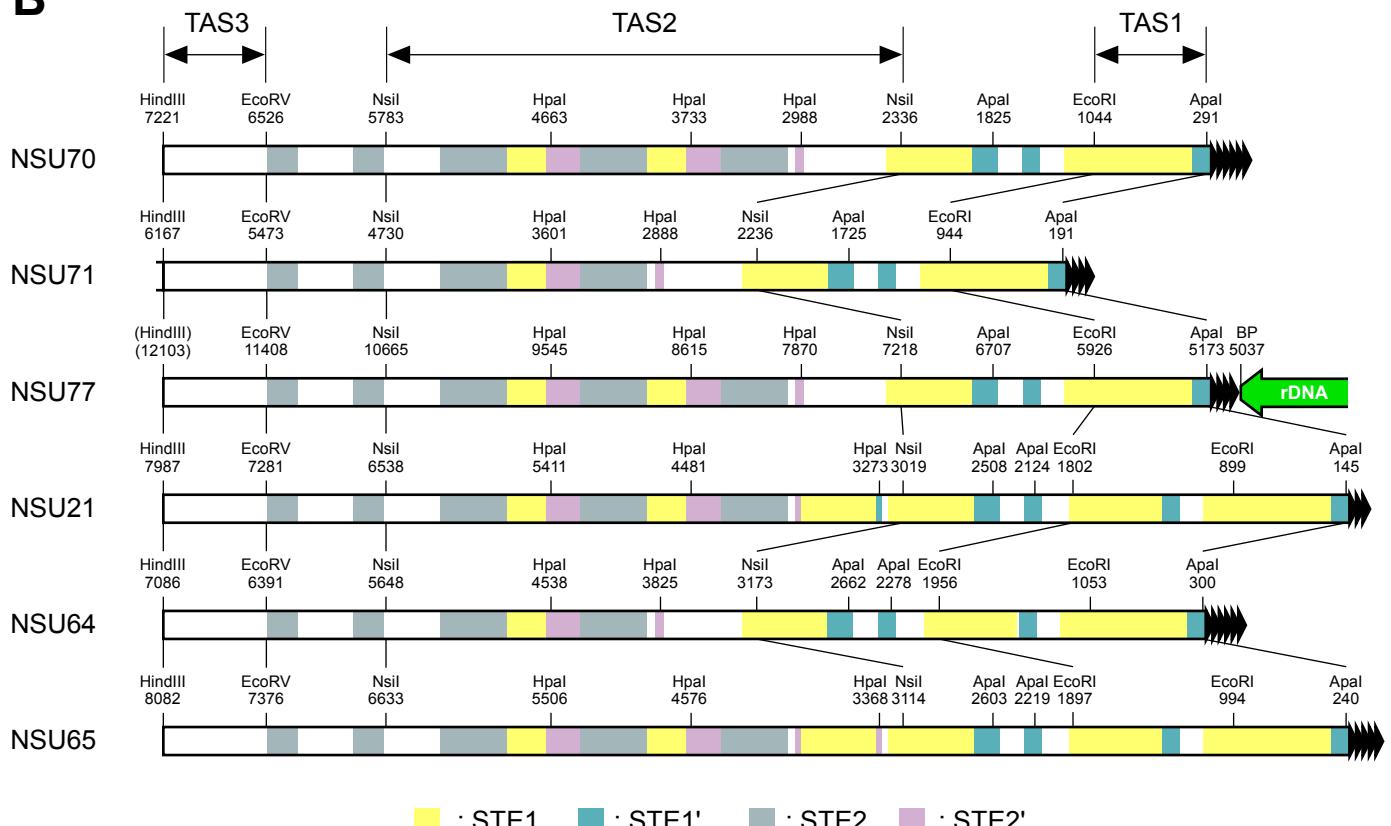
Supplementary Figure S1. Profile of the centromere deletion screen and characterisation of the resulting survivors. (A) The viabilities of total cells (grey) and *cen1*-deleted cells (black), as determined by their colony formation frequency on nonselective YES plates and $\Delta cen1$ -selective G418- and 5-fluoroorotic acid-containing YES (YES+GF) plates. Data are represented as the mean \pm SEMs of $n = 6$ replicates performed simultaneously. (B) The ratios of neocentromere formation and telomere-fusion in the survivors obtained in each screen. The sample size (n) varied depending on the frequency of telomere-fusion events. (C) The frequencies of neocentromere survivor generation upon *cen1* deletion in the indicated mutant backgrounds. Data are represented as the mean \pm SEM of $n = 6$ replicates. *, the 95% BCI does not include zero; n.s., the 95% BCI includes zero (not significant).

Supplementary Figure S2

A



B



C

STE1 (NSU70)

4943	GTAGTGTAGTGTGGTAGTGAAGATGGACAAACAC-----TGAAATGAGTG-GAAGT-GAGTGTGGGATGCAGTAAGTATAATAAGGGGAT	4857
4856	GTAGTGTAGTGAAGATGGACAAAAAG-----TTGAAGTTGATTGAATT-GAGTGTCTGGAGTAGTTAAAGGTGATAGGGACAA	4768
4767	GTAGTGTACTATAATAATTAGGATGGTAAAAATA----TGAAGTTGACTCAGTTTGATTCAAGGTGGGT-----	4701
4017	GTAGTGTAGTGTGGTAGTGAAGATGGACAAACAC----TGAAGATGAGTG-GAAGT-GAGTGTGGGATGCAGTAAGTATAATAAGGGGAT	3931
3930	GTAGTCAATATGGCAATGGAGATGAATGAACAACTTGAAGTTATAGATAAA-GCCTGTGTTGAATGCAGTAAGTTGAATAAGATAC	3838
3837	GCACTGATTATGATAATCAAGATGGATAAAAATA---TGAAGTTGACTCAGTTTGATTCAAGGTGGGT-----	3771
2429	GTAGGAGAA-----GAAGAAGTAATCAA-----GTAAGAGAGTATTAGA-----AAAGTAAAGTAAA-TAAGGATGGATAC	2364
2363	ATAGTGTATATTGGAAAGTCAGATCATAAAAAA----TTGAAGTTGTTGATGTATT-GAGTGTGTTGGAGTAGCGGTAA-----	2290
2289	GTAT-AATAGGGGTAATAAAATGGTAAAAAAA-TTTGAAGTTGTTGAT-GGAAGTTGAGTGTGTTGGAGTACATTAAGTAGATTACAGTTG	2200
2199	GGAGCGTATATGGTAATGAAAATGAAGAAAA---TGAAGTTGGGTTGAATT-GAGCGTGGTAGGATTCAATTATGTTGATAGGAGAT	2111
2110	ATAATGAGATATGGTAATAAAAGTTGAAA-----TGT-GTGGGCTTGA-GT-GCCTTGGGTGAGTAAGTAGAAATAAGGGGC---	2033
2032	GCAGTGTATTATGATAATTTGGATGGAAAAA---TTTGAAGTTCACT-----	1986
1176	-GAGTAAAAGAAAA---GAAGATGAAT-GGATTTAAAGGTGTTGGAGTAGATTAAAGTAGAATACGGAGGAT	1112
1111	GTAGGGTAGTCAATAGTGAAGATGGACAAAAG---TTGAAGTTCATG-GAATTAGACTATGGAAATTCAATTGTAATAAGGTGTT	1024
1023	GTAGTGTATGGGTGAATAAAACGGATGGAAAAA---TTTGAAGTTGATTGAAC-T-GAGTGTGTTAAAGTCAATTGTAATAACGGGTGAT	935
934	GTAGTGTACTATAGTATTAGGATGGTAAAAAA---TTGAATGTTG-GAATT-GAGTGTCTGGAGTAGCTTAAGTATAACGGGTGAG	848
847	GTAGTGTACTATAATAATTAGGATGGTAAAAAA---TTTGAAGTTGTTGAT-GAATT-GAGTGTGTTAGGTTCAATTAGTATAACGGGTGAT	760
759	GTAGTGTACTATAGTATTAGGATGGG-AAAAAA---TTTGAAGTTGTTGAT-GAATT-GAGTGTGTTAGGTTCAATTAGTATAACGGGTGAG	672
671	GTAGTGTACTATAGTATTAGGATGGG-AAAAAA---TTTGAAGTTGTTGAT-GAATT-GAGTGTGTTAGGTTCAATTAGTATAACGGGTGAG	585
584	GTAGTGTACTATAGTATTAGGATGGG-AAAAAA---TTGAATGTTG-GAATT-GAGTATGGTAAATTCAATTAGTGAATAACAGTAGT	499
498	GCAGTGTATTATGATAATTTGGATGGAAAAA---TTTGAAGTTCACT-----	452

Supplementary Figure S2

D

Wang and Baumann (2008) / NSU70

Supplementary Figure S2

NSU71

Supplementary Figure S2

TTTGAAGTTGACTCGTCATGGTGAGCTGGATAACGAAGTAACAATATGGAATAAGGAAATGAAGGGAGAGATAAAGAGTAGGATATAGAAGAGA
GTAGAACATCAAAAAAAATAAGTAAGGAGAG
TAAATTAAATATTAAAACGAATAAATAAAATAAATAAAATAAATAGAATAAAATAAATAAA
GTAGGAGAAT-----GAAGAAGTATCAA-----GTAAGAGAGTATTAGA-----AAGTAAGTAAA-TAAGGATGGATAC
ATAGTGTATATTGGAAAGTCAGATGCATAAAAAA---TTTGAAGTTGGTATGTATT-GAGTGTGTTGGAGTACGGTAA-----
GTAT--AATAGGGGTAATAAAATGGGAAAAAAA-TTTGAAATGTG-GGAAGTTGAGTATGTGAGTACATTAAGTAGATTACAGTTG
GGAGCGTACATGGTAATGAAAATGAATGAAGAAAA---TGAAGTTGGGTTGAATT-GAGCCTGTTAGGATTCTTATGTATGATTAGGAGAT
ATAATGAGATATGGTAATAAAAG-TTGAAA-----TGT-GTGGCTTGA-GT-GCGTTAGGGTCAAGTAGAATAAAGGGGC---
CAGCTGTTATGATAATAAAATGGTAAAGAAAAA---TTTGAAGTTC
CACTAATTAATTGGGTAACGGAGTAACAATATAGAATAAAGGAAATTAGGAAGTGCCTAAGTTGAATAAGAAATAGAATAACGGTATT
CATAAAAAAATAAAATTACTTAAAGTTTTTCAAAATC-AATGCCCAACTA-TTGGCC CACCCGTCAGCCGAGCCGTAAGCGAGTATTGCT
TAAACGATT
GGAGAGAGAATGGATAATGGATGGAGGTAAGAGAGGGTATGAAAGAGTAGAAGATATAGAAGAGAAGAGAAGAAAATGAGGAAGAAGTAGATGAAT
AGAATAGAAGAAACAGAGTAAGGAGAGTAATAAAAGTAAGAGATAAGAAGATAAGTAAGGAGAAAGATAAAAGCAGAGGACTATA
TTGGAGTAGATTAGTACAGTTGTCAGCGT-----AGATGATAATGAAATAAGAAAGATAATTAGCTGCTTATT
TATAAAATTTAAATTACTTAAAG-TTTTTTACACATATAC-AATGCCCAACTACTGGACCCCCACCCGTCAGCCGAGCCGTAAGCGAGTATTGCT
TAAACGTTT
AGAGAGAGAAGGAAATGAAGGAGAGAAGGAAATGAAGGAGAGAAGGAAATGAAGGGAGGAATAGGTAGAGTAGGTAGAGTAGGTAGAGTAGG
TAGAGTAGGTAGGATAGAGTATGGAAGAAAGAGAGGAAGAAGATAATGGAATAAGATAAAATAGAATAAGATAATAAGATAATAAGATAATAAGATA
ATAGAATAAAATAAATAAA
-----GAGTAAAAGAAAA---GAAGATGAAT-GGATTAAGGTTGGAGTAGATTAAGTAGAATAACGGAGAT
GTAGGGTAGTCAATAGTGAAGATGGACAAAAAG----TTGAAGTTCATG-GAATTAGACTATGTTGAATTGGAATTCAACTAATTGATAAAGGTTG
GTAGTGTGTGGGTAAACCGGATGAAAAA---TTGAAGTTGTTGAAC-GAGTGTGTTAAAGTTCATTAAGTATAATACGGTGAT
GTAGTGTACTATAGTATTAGGATGGTAAAAAA---TTGAATGTTG-GAATT-GAGTGTGCTGGAGTACGTTAAAGTATAATACGGTGAG
GTAGTGTACTATAAATTAGGATGGTAAAAAA---TTGAAGTTGTT-GAATT-GAGTGTGTTAGAGTTCTTAAGTATAATACGGTGAT
GTAGTGTACTATAGTATTAGGATGGTAAAAAA---TTGAAGTTGTT-GAATT-GAGTGTGTTAGAGTTCTTAAGTATAATACGGTGAG
GTAGTGTACTATAGTATTAGGATGGG-AAAAAA---TTGAAGTTGTT-GAATT-GAGTGTGTTAGAGTTCTTAAGTATAATACGGTGAG
GTAGTGTACTATAGTATTAGGATGGG-AAAAAA---TTGAATGTTG-GAATT-GAGTATGGTAAATTCAACTAAGTGTAAACAGTAGT
GCAGTGTATTGATAATAAAATGGTAAAAA---TTGAAGTTCACT
CACTAATTAATTGGGTAACGGAGTAACAATATAGAATAAAGGAAATTAGGAAGTGCCTAAGTTGAATAAGAAATAGAATAACGGTATT
CATAAAAAAATAAAATTACTTAAAGTTTTTCAAAATC-AATGCCCAACTA-TTGGCC CACCCGTCAGCCGAGCCGTAAGCGAG
GCTGCGGG
TTACAAGG
TTACGTGG
TTACACGG
TTACAGG
TTACAGGG
TTACAGGGGG
TTACCG
TTACACGG
TTACAGG
TTACAGG
TTACAGGG
TTACAGGGGG
TTACGG
TTACAGG
TTACAGG
TTACAGGGGG
TTACAGGGGG
TTACGG
TTACAGG
TTACAGGGGG

NSU77

-----TCGTTAAAAAAAGTTAAGGGTAGGATAAAGCCAAGTGAAGGACGTTAGCGATGAATAAGAGGGAGTAATAAGTGTAAAAAGGAATGACAATTAG
AAAGATAGAAAAGAGATAGAAGATGAAAATATTGGAAGTAGATGAATTAGTCATTCTATTACACTAAACAACTAAATTATTGAAAAAACAGTCGTTAACATTACTGGTGATGTGTCAGCTGGAATGAAATGTCAGATAGAGAAAGTACAACATTGTTATTGTTGGAA
AGTAAAAAAATATAAGTAGGAAATAATAGTAACAGATAAGAACAACTATGAGAAAAGATTAAATTGCTTAATTAAACAACTATAAATTAAATAAATA
AATAAATAAATATAATTAAAGATAAATGAACTCAAAGAAAAGTAATTAAAGAAAAGTAAATTAAACAACTATAAATTAAATAAATTAAACAACTATAAATTAAATA
TAAAACTTGATAAAATAACTAAAATAAAACACATAAAATTCACTAAATTATAACCGTTAAATAAACTATAATGGTTACGGTTATTAGGTG
ATGTAAAGTGGAAAGTGAATAAATAGAATCGAATAAAAGTAGTAAAGAACAAAAGGTTGAAACTGGTTGAGCATCTGTCAGGAGTAAAGACAGTAA
TAGATATCACACAC-----
ATTTAAAAATAACTATAATTAAATTAAATTAAATAAAATTATTGGAAAAGCTCGGTCAGTTGATGTCGCCCCAAAAGGGAGATGGTAAATG
TGGAGGAATGGTAAATCGATAGAGTAGAGTAAAGCTAGATAACGAAAATGGATAAGCTCGTTATGAAAAGAGGAAGGAGAATGAAAGGACAAGTGG
AGTAAGTAGTGAGTCAGTCAGTTAATGAGTCATGAGAAGAGATGGTAGAGGCAGCGAGTAACAAAGTAACAAAGAGAGAATAAGAATGAGAATGAA
ATGGGGTAGAGTAGAGAGGAGTAGGGAGTAGAGTAGAGAGTAGAATGTTGAAATCTGAAATGTAATTAGACAGTAATCTGAAATGAGAATGAG
AGAAAAGAAAATAGAAGAAAACAAAATAACAGAAATAATTAAATTAGATTCAATTACAGAAAGAAAACAAAAGTATTGAAATAGTGTGGGCA
AG-----ATAAAAGATAATAAAATAACAACTTGGAAAATAACAGAAAATAATTATTGTCGATGAAACAGCAACAGTAGAGAGTTGAAA
ATTGAAAAGAAAGCATGAGTTTATTAGATTAGTTAGAAAATAATTATTGAGAAAAGCTCGTTAGTTGATGTTGCGGGGAGATGAGAAAATGTAGGGG
AGGTTCAAT**ATGCA**CGTGGAAATGGTTAAGGTGGAAAAGCAGAAGTAAAGTGTGTTGGGATTCATTAAAGTAAATAGAGATATAGTGTGATATGATAAT
TGAGAAGGTGAAAATTAGAGTTGAGTTGACTCATTTGATGTGAGCAGTAACAGGCCATAAGGCAGTAATCGTTAACGGCTTGAAGGAGGAGATAAGG
AAAATAAGGAAGTAGAGTAGAAGAGGAAGTAGTAAAGGAGAGTAGAAGAGGAGTAGAAGAGGAGAGTAGAAGAGGAGTAGAAGGGTGGAGGATAGAAGTAA
ATAAGAGAATAAGGAAATAAGAGCAATAACGAGGAAAAGAAAGTAAAGAGATAAAAGAGATAAAGGGATAAAAGTAA
GGAGAATAAGAAGTAGAAGTAAAGTAAAGAAAAGAAGTTGATTAAGTTAA-AGTATGTTGGAGTGAGCTAGTAATAGGAAACATGCGGTGACTTTG
ATAAAATGAAGTGTGTTGAAATTAGAGCTTATGTTAGATGTTAAATAATAATGTTGCGTAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA
TAAATAAAA---TATTAAACAAAGCAATAATATAAGAAATAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA
ACAACATATTGGAACACAAATAAGACAATAATATAATGATAAGAAAATAATGAAAATAATTATTGTCGATGAAACACAGCAGGGTCAATTG
ATTGAAAATAATTATAAAATGATTAATAAAAGAAAAAAATTATTGAAAAGTTGAGTGTGATATGTCGTCGGGAGAATG
GTAGTGTAGTGTGGTAGTGAAGATGGACAAAACAC-----**TGAATAGTGT-GAAGT-GAGTGTGTTGGAGTGCAAGTAGTATAATAAGGGGAT**
GTAGTGTGCAAGTAGAATAAAACGGATGAAAAA-----TTGAAGTGTGTTGAAATT-GAGTGTGTCGGAGTAGCTTAAAGGTGATAGGGACAA
GTAGTGTGACTATAATAATTAGGATGTTAAAAAAATA-TGAAGTGTGACTAGTGTGTTGAGGGT-----

Supplementary Figure S2

Supplementary Figure S2

TTGCTTGTAGGCCAATACCATGATCTGAATCACCGGTTCTCGTACTAAGTTAACATTGGACGACACTTCATCAGTAGGGTA
AAACTAACCTGTCTACGAGGCTAAACCCAGCTACGTTCTTATTAGTGGGTGAAACAATCCAACGCTTACCGGAATTCTCGTCTCGTATGAGGAA
GAGCCGACATCGAAGAATCAAAAAGAACCTGCCTAGAACCTGGTGCACAGGCTTACCGACAAGCCAGTTACCTCTGTGGTAACTTCTGGACCTCTGTCTCA
AATTCTGAGGAAACAAAGGATCGATAGGGCACACTTCTAGGGTTGATACACTGAAATCAAATAAAGGGGACTTTACCTTTTATTCTACTCG
AGATTCTGTTCTCGATGAGTCCCCCTAGGACACCTCGGTTACTTTAACAGATGTGCCGCCAGGCAAACCTCCCACCTGACAAATGTCATCAACG
CGGATCACTTCGCGAAGAACCTTAAACGCTAGAATATGGGAAATAATCCAACTTCGCTTCAATTGATTAAGGAAAGAACGATAAAGGGTAGTGTGATT
TCACCGGGCTATGGGAAACATACTCCCACCTTACCCCTCTACGGTCTTCAACATGTCAAACTAGAGTCAGCTAACAGGGTCTTCTTCTCC
CGCTGATTTCGCTGGCCCTTCCCAGGGCTGGTTCTAGTAACTGAGATAGGAGACAGTGGGAATCTCGTAACTCATCGCGCTACTAGTT
AGATGACGAGGCAATTGGTACCTTAAGAGAGTCATAGTTACTCCCGCTTACCCGCTTGGTTGAATTCTTCACTTTGACATTAGAGCACTGG
CGACAAATCACATTGCGTAACACCAACTTCTGGCATCGCAATGCTATTTAACAGTCAGATCCCCCTGTGGTACCACTGTTCAAGTTGG
TTGTTAAACGTCAGGCACTAAGGACATACTCCGAGAACAGGAAAGGGAAAGCCGCCAACAGGTTCCCGGCAAGGTCACAGGCTTCAACATCCT
TCTCACCCGAAAGGGAGAACATCGGTCACGCTCAGTCCGCAACCGGTTCAAGGGCAACGTTCAACCTTGTAGGCCAATCTT
ATCCGAAGTTCAGGATCTTTGCGACTCCCTTACTACATTGTTCTACAACTAGAGGCTGTTCACCTGGAGACCTGTCGCGTTATGAGTAC
GACCAGATGTGAAAACAAGGACCGAAGGTCATTCTCGTTGGATTTCAGGGCGTCAGAGCGCACCGGATTAGCATGAGGCGCTGAACCTT
CCAAACACCATCTACCTGTCGGATAAACCGATTCTGGGTTAGTCTAACCTGGTTTAAAGGAAAGAACACTTCCCGGAGCTCCGCCAGCAGTC
TCCAACCTTCTTACGTTGGCTGGTATTGTCACCCCTTCGGTCCGGAAATTAAACCGGATTCTCGTACAGGAGCAGAAAATAGTCGCAACT
CATACGGAGCTCCCTATCTTGGATTCGACTAACCTGTCAGTCTGGTACATGGGAACTTCCCTGGTACAGGAGCAGAAAATAGTCGCAACT
ATATTGCTACTCCACCAAGATCTGCACTAGAGGCTGTTGACCCAGGCTCACGCCAACGGCTTCAACAAACCTCCACGCCCTACTCGTCTGA
GCTTCTAAAGCTAACCCAGACGGTGAGGTATGGGTAGTACGCTTAAGGCCATCATTTCAGGGCTAGTTCTGGCGAGTGGTACACACT
CCTTAGCGGATTCCGACTCTCATGGCCACCGCTCTGGTCTAGATGAACTAACACCTTTCTGGTCTGATGAGGCTGATCTCCGGCACCTTAAACCT
CACGGTCTGGTATCCGCACTGCCAGTCTGGTCTACCAAAATGGCCCAACTGAAACACTCTTCGAGTGGCCACGGTCAATTAGGCGAACAGGGCT
CTTACATTTAAAGTTGAGAATAGGTTGAGGAATTCTTCCCAAGACCTCTAACATTCGTTAACCTCATAAAACGATCTGAGTTCTGCTAT
CTTGAGGGAAACTTCGGCAGGAAACAGCTACTAGATGGTCAATTAGCTTCCGCCCCATACCCAAATTGAGATCAGTTGCACTGCAAGATCTT
ACGGAGCCTCACCGAGATTTCTGGCTTCCGACCATTCAGGCTACCATCTTCCGGTCCAAACAGCTACTGCTTACTCAAACCCCTTCCAT
TGACAAAATTCTGGGTCGGTCTAGGTCAGCCACAGGGCTTCCACCTGCACTTCTGGGTTTCTACCCAAACACTCG
CATAGATGCTAGACTCTTGTCCGTGTTCAAGACGGGCAATTGAAAACATTAGTCAGCATCTTGGCACAGGCCAGTCTCAGTCCCAGTG
GACGTATTACAACAAGGGATAAAACACTCCAGGCGAACCCGAAAGCCACCTCCCTTATTCTTTCCGCCCCGAAAACGATGCTGACACTAC
CACAGGGTCAAGTGCATGCTCGAGAGGACTACTGATCTACCGAGCTAACGTTCAAGGACTGTTCACTGGCTTACGAAATTACGACTATT
AACTCTTTTCAAAAGTTCTGGTACACTTCTGCACTACTGTTCTGCTACTGGCTTACCGGCAATTAGCTTCTGGCTTACCGGACTATT
TAGAGCTGATTCCAAAACACTCGACTCTCGAAAAGGGCTTATAGGCTATACCAATGCAACAAAGGCCGTTCTCACCCCTCTGACGCTCTGGT
CCAAGGAACATTAGACCGATGCTACACTCAAGCAGTTCTGAAATTACAACCTGGACAAACAAAGAAATGTTGCCAGTTCAAAATTGAGGTTT
CCGCTTCACTCGCGGTTACTGGGAACTATGGTTATTTCTGGCTTATGGTATGCTTAAGTTCACTGGCTAGTCTACCTGATTGAGG
TCAAAATAAAAAGGACTCTGGAAAAAAATAACTTTCTGGTAAAGGTTCAAAATAAAAGGAGATTAACCTTTAGTTATTCTTCTCTCAA
TTCTTTCTATCAAAAGGTTGAAACCTTACGTTCAATGAAAGAAAAGGAGGATCAAGAAAATAATTCAACATTCTCCTTGT
TTCAACATCGATTCTAAACTAAATTATTTAAAAAAACAAATTCTGGTCAACACCTCATCAAAATAATTAAAAAAACATTTTGTG
AGAAGATTGTAATGACACTAAACAGGCATGCCCTGGTAGAACCAGGCAAGGCCAATGTCGTTCAAGGATTCTGACGGAATTCTGCAATTCA
CATAGTCTGCACTTCCGCTGGTCTTACGTTCAAGGAGGCAACAGGAGATCGGTGCTGAAAGTTAAATAATTATAGATAATAATT
TCAGACTTCAAAACAAATTCTGGTAAAGGTTTAAATTAAATTCTGGTCTTCTGGTCAACACCTCATCAAAATAATTAAAAAAAGAGAATGAA
AATAAAATAAAAGTAAAGGAAAGTATGATGCTGGCATGCAACAAAACACCAAAACACCAAAACACCAAAACACCAAAACACCAAAAC
TTTCTATCTTCTCTTCTAAACACTTTTTTAAATAAAAAAAATAAGAAAAAGAAAAAAATCAAATTGAAATTT

NSU21

AAGCTTCGTAAAAAAAGTTAAGGGTAGGATAAGCCAAGTGAAGGCAGTCAGCGATGAATAAGAGGAGTAATAAGTATGAAAAAGGAATGACAATTAG
AAAGATAGAAAAGAGATAAGAGTAAAGGAAATTGGAGAAGTAGTGTCAATTCTTACACTAAACAAATCAACTAAATTATTGAAAAAAA
GTCGTACAAATTACTGGTGTGTCAGCTGGCAATGAATATGCAATTAGGAATTAGAATACGATATAAGGAAAGTACAACATTGGTTATTGTTGAA
AGTAAAAAATTAAGAGAATAAATAGAACAGATAAGAACAATGATGAAACAAATGATGAAACAAATGAGAAAAAGGAAAGTAAATTGGTTATTGTTGAA
AAAATAAAATAAATAAATATATAATGAAATGAATCAAAGAAAAGTAAATTAAAAGAAAAGTAAATTAAAATAAAATAAAACAATATAAATT
AAAAAAATTTAAAGGAAACTAAATAAAATAAATCATAAAATTCACTAAATTAAACAGCTTAAATAAACTAAATATGTTACG
GTTATTAGGTGATGATAAGTGGAGGTGAATAATAGAATCGATAAAAGTGAAGAACAAGGGTGGAAACTGGTGTGAGCATGTCAGAGGT
AAAGAGCTAATAGATAACACAC- - - - - AAAAATATTATTGTCGAAACACATGCAAGCGTAGGCAATTGAAA
ATTTAAAATAACTATAAATTATAAAATTTAAATAAAATTATTGGAAAAGCTCGGTCAAGTGTGATGTCGCGGGGAAAAGAGGAATGGTAATG
TGGAGGAATGGTAAATGCGATAGAGTATAGTAAAGCTAGATAACGAAATGGATAAGCTCGTTATTGAAAAGAGGAAGGAAAGTATGAAGGACAAGTGG
AGTAAGTGTAGCTAGTCAGTTAATGAGTCAGGAAGAGTGGTAGAGGCCAGGGAGTAACAAAGTAACAAAGAGAGAGAATAAGAATGAGAA
ATGGGGATAGAGTAGAGAGAGTAGGGAGAGTAGAGAGAGTAGAGTAAGTGTGAAATACGTAATGAAATAATTAGACAAGTAACATATAAGA
AGAAAAGAATAGAAGAACAAAACAGAAATAATTAAATTAGATTCGAAACAGAAACAAAGTGTGAAATAGTGTAGGGGCAA
AG- - - - - ATAAAGATAATAAATAAAACAATTGAAAATAACAGAAAATTATTGCAATGAACAGCAACAGTAGAGAGGTTGAAA
ATTGAAAAGAAGAGCATGAGTTTATTAGATTAGGAAAATTATTGAGAAAAGCTCGTTCAGTTGTTGTCGGGGGAAAGTGAAGGAAATGTAGGGG
AGGTTCAT**ATGCAT**CGTGGAAATGGTAAAGGAGGAGCAGAAGTAAAGTGTGTTGGGATTCAATTAGATAATAGAGATATAGTGTGATGATGATAAT
TGAGAAGGGTAAAGGATTAGTGTACTATTGATGTCAGGAGGAGTACCGAGGCTATAAGGGAGTAACTGTTAACGCTGGTTAGGAGGAGGAA
AATAAAGGAATGAGAGTAGAGAGGAAAGTAGTAAAGGAGAGTAGAGAGGAGTAGAGAGGAGTAGAGAGGAGTAGAGAGGAGTAGAGGAGTAGAG
TAAGAGAATAAGAGAATAAGGGAAATAAGAGCAATAAACGAGGAAAAGGAAAGTAGAAGAGGAAATAAGAGGATAAGGGGATAAGA
GGGAGAATAAGAGAATAAGTAAAGGAGGAAAGAGAGTAGTAAAGTAA- - - AGTATGTTGGAGTCAGTAAGTATAATAGGAACATGCGGTGACTTTG
ATAATGAAAGTGTGTTGAAATTGAGCTTATGTTAGATGTTAAATAAATATGATGTCGAAAATAAATAAATAAATAAATAAATAAATAAATAA
TAAAATAAATATTAAAACGAAATAATATGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
ACAAATATTGGAACACAAAAGACAAATTAAATAATGATAGAAAATAATGAAAATAATTGCAATGAATGACAAACAGCAGGGCAGTTGAAA
ATTGAAAATAAATATAAATGTTAAATAAAGGAGGAAAGGAGGAAAGTAGTAAAGGAGGAAAGTAGTAAAGGAGGAAAGTAGTAAAGGAGGAA
GTAGTGTAGTGTGAGTGAAGATGGACAAACAC- - - - - TGAATGAGT-GAAGT-GAGTGTGTTGGGATGCGAGTAAGTATAATAGGGGG
GTAGTGTGATGAGTGAATAAACCGGATGAAAAA- - - - TTGAGTTGAGTTGAAATT-GAGTGTGCTGGAGTACGTTAAAGGTGATAGGGGCAA
GTAGTGTACTATAATAATTAGGATGGTTAAAATA- - - TGAAGTTGACTCAGTTGATTAGGTGGGT- - -
AACGAGCAGTAAAGCGGATATCGTTAACAT-GTAAAG**GAAATATAAAAGAGGTAACATAGGAAGTATAAATATTG**
ATAGGTTAGGAAATAAGTAAAGC- - - AGATAGTGTAGA
- - - GTAGAATAGTGTAGTGGAGTAAAGTGA
GTAGAGTATGGAAGAAGAG
GGAGAACAAAGAAGTAAGTAAAGTAAGGAG
ATAAAATGAGTGTGAAATTGGCTTATGTTAGTCAGTTAAACAAATAATGTTGTTGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
TAAAATAA
ACAAATATTGGAACACAAAAGACAAATTAAATAATGATAGAAAATAATGAAAATAATTGCAATGAATGACAAACAGCAGGGCAGTTGAAA
ATTGAAAATAATTATAAAATGTTAAATAAAGGAG
GTAGTGTAGTGTGAGTGAAGATGGACAAACAC- - - - - TGAATGAGT-GAAGT-GAGTGTGTTGGGATGCGAGTAAGTATAATAGGGGG
GTAGTGTGCAATTGGCGAATGGAGGAGTGAATGAAAGTAAACTGGTGTGTTAGAATAA-GGCTGTGTTGATAGTGTGAGTAAGTGAATAAGGAG
GCGAGTGTATTGATGATAATGAGTGTGAGTAAAGATA- - - TGAAGTTGACTCAGTTGATTAGGTGGGT- - -

Supplementary Figure S2

NSU64

AAGCTTCGTTAAAAAAAGTTAAGGGTAGGATAAAGCCAAGTGAAGGCAGTTAGCGATGAATAAGAGGGAGTAATAAGTGTGAAAAGGAATGACAATTAG
AAAGATGAAAAGAGATAGAACAAATTGGAAAGTAGATGAAATTAGTCATTACACTAAACAAATCAACTAAATTATTGAAAAAACAC
GTCGTTCAACATTCTGGTGATGTCAGCTGGCAATGAAATGTCAAATAGGAATTGATACAGTATAGAGGAAGTACAACATTGTTTGTGGA
AGTAAACAAATTAAAGTAGAACATAAAGTACAGATAAATGAAACAAATGATGAAACAAATAGGAAAAGTAAATTCTGGTAAATTAAAATAAAT

Supplementary Figure S2

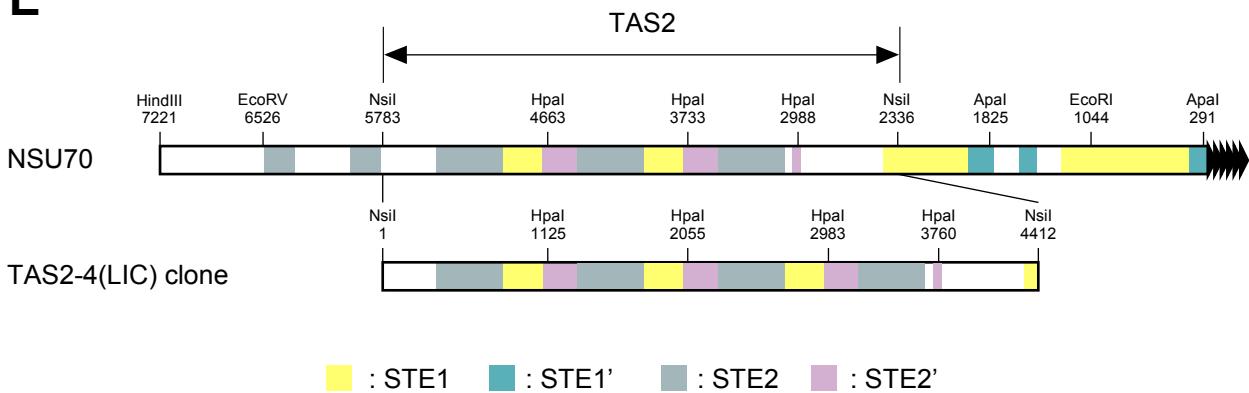
Supplementary Figure S2

TTACAGG
TTACAGG
TTACAGGGGG
TTACGG
TTACAGGG
TTACAGGGG
TTACGG
TTACAGGG
TTACAG
TTACACGG
TTACAG
TTTACGG
TTACAGGG
TTACACGG
TTACAG
TTACAGG
TTACAGGG
TTACAGGGG
TTACGG
TTACAGG
TTACAGGG
TTACAGGGG
TTACAGGGGG
TTACAGGG
TTACAGGG
TTACAGGGG
TTACAGGGGG
TTACAGGG
TTACACGG
TTACAGGG
TTACAGGGGG
TT

NSU65

Supplementary Figure S2

GAAATTTGAGGATGGTCACTTAATAAAGGAAAAAAATTCAAAGTCAGGTGTGAAGCAAGT
AAGCAGCGCTAAGCGAGTAATCGTTAAC
GTAGTTAGAAGAAGTAAAGGAAATGA-AGGAGAAA-----TGAAG-GGGCAGAATAAAGGAGAGAAGG-----AAGGAGGAGAATAG
TAGAACAGA--TAGAGTAGATGGAGAAA-----GAAGAG-----GAAGAAAATAATGAAATAAG-----AAGA-----
GTAGGAGAAT-AAGAAAGTAAATGAATAAAA-----TGTGT-----AGG-AATAAGGTTG
GTAGTGTACATGGATAAAACGGATGAAAAA---TTGAAGTGGTTGAAGT-GAGTGTGTTAGGTTACTAAGGTAATAAGGAGT
ATAGTGTACATGGTGATAAAACGAATGAAAAA---TTGAAGTGGTTGAATT-GAGTGTGTTAGGTTACTAAGGTAATAAGGAGT
GCAGCGTACTATGTAATGAGTACGGTTGAGAAA---TTGAAGTGGTGAAGTCACTGATGTTGAGGGTGGG-----
AAGCAGCTAAGCGAGTATTCGTTAACAT-GTTAAC GAAA
GATAAAGGAGGATAAAAGAAAATAGAATATAGAATATAGAATATAGAATATAGAATATAGAATATAGAATACAGAATA
GAGTAGAGAGTAGGGAGTAGGGAGTAGGGAGTAGGGAGTAAGGAATAAATAAAATAA
ATAGAATAAGATGAAT---AAAGAAGTAA-----TTAAAGTAAGA-----GAGTATTAGAAAGTAAAGTAAT-AAGGATGGATAC
ATAGTGTATATTGGAAAGTCAGATGCAATGAAAGGAAATGGTAAAGGAAATGGTAAAGTGTATT-GAGTGTGTTGGAGTACGGTAA-----
GTATAATAGGG-GTAAATAAAATGGTAAAGGAAATGGTAAAGGAAATGGTAAAGTGT-GGAAGTTGAGTATGTTGGAGTACATTAAGTGAATTACAGTTG
GGAGCGTACTATGTAATGAAAGTGAAGGAAA---TGAAGTGGTTGAATT-GAGCCTGTTAGGATTCAATTGATGAT
-TAGGAGATAATAGAGATAATGGTAAATGAAAGGAAATGGTAAAGTGT-GGGCTT-GAGTATGTTGAAATTCAACTAAGTGTAAACAGTAGT
GCAGCTATTGATAATTAAATGGTAAAGGAAATGGTAAAGGAAATGGTAAAGTGT-GGGCTT-GAGTATGTTGAAATTCAACTAAGTGTAAACAGTAGT
CACTAAGGAAATGGTAAACGAGTAACAAATAGAATAAGGAAATTAGGAAGTGCCTGAAGTGTAAAGGAAATAGAATAAGGAAATACGGTATT
CATAAAAAAATAAATTACTTAAGT TTTTTTCAAAATAC-AATGCCCAACTA-TTGGCC CACCCGTCAGCCGAGCCGTAACGGCAGATTCTG
TAAACGTTT
GGAGAGAGGAAATGGATAATGGATGGAGGTAAGAGAGGATGAAAGAGTAAAGATATAGAAGGAGAGAGAAGAAATGGAGAAGAGTAAGATGAAT
AGAATAAGGAAACAGAGTAAGGAGATAAAATAAAGTAAGAGATAAAAGATAAGTAAGGAGAAGAGATAAAAGCAGAGGACTATA
TTGGAGTAGATTAGTGTACAGTTGTCAGCGTAGTGTATAATGAAG-----ATAAAAGAAAGATAATTAGCTGCGTTATT
TATAAAATTTAAATTACTTAAG-TTTTTTCAATACATAC---ATGCCCAACTA-TTGGCC CACCCGTCAGCCGAGCCGTAACGGCAGATTCTG
TAAACGTTT
AGAGAGAGGAAAGGAAATGAAGGAGAGAAGGAAATGAAGGAGGAGAATAGGTAGAGTAGGTAGGGTATGGAAGAAAGAAGAGGAAAGGAAAGTAATGGAAT
AGAATAATAGAATAAATAGAATAAATAAATAAATAAATAAATAA
-----GAGTAAAGAAGGAAATGGATAATGGAGGAGGAGAATGGATAATGGAGGAGGAGAATGGATAATGGAGGAGGAGAATGGATAATGGAGGAGGAG
GTAGGGTAGTCAATAGTGAAGATGGACAAAAG-----TTGAAGTTCATG-GAATTAGACTATGTTGGAATTCAACTAATTGTAATAACGGTGG
GTAGTGTGATGGTGATAAAACGGATGAAAAA---TTGAAGTGGTTGAACG-GAGTGTGTTAAAGTGTAAAGTGTAAATACGGTGAT
GTAATGTACTATAGTATTAGGATAGATAAAA-----TTGAAGTGTAT-GAATT-GAGTGTGTTGAACGTTAAAGTGTAAATACGGTGAT
GTAGTGTACTATAATAATTAGGATGGTAAAGGAAATGGTAAAGGAAATGGTAAAGTGTAAAGTGTAAATACGGTGAT
GTAGTGTACTATAGTATTAGGATGGTAAAGGAAATGGTAAAGGAAATGGTAAAGTGTAAAGTGTAAATACGGTGAT
GAGCCTGTAATGGTAATGGTAAAGGAAATGGTAAAGGAAATGGTAAAGTGTAAAGTGTAAATACGGTGAT
CACTAAGGAAATGGTAAACGAGTAACAAATGAGAATAAGGAGCAATTAGAACGTCAGGTAAATAGAATAAGGAAAGATAATTAGCTGCGTTATT
TATAAAATTTAAATTACTTAAG-TTTTTTCAATATATAAAATGCCCAACTA-TTGGCC CACCCGTCAGCCGAGCCGTAACGGCAGATTCTG
TAAACGTTT
AGAGAGAGGAAATGAAGGAGAGAAGGAAATGAAGGAGAGAAGGAAATGGATAATGGAGGAGAATGGATAATGGAGGAGAAGGAAAGTAATGGAAT
AGAATAATAGAATAAATAGAATAAATAAATAAATAAATAA
-----GAGTAAAGAAGGAAATGGATAATGGAGGAGGAGAATGGATAATGGAGGAGGAGAATGGATAATGGAGGAGGAGAATGGATAATGGAGGAG
GTAGGGTAGTCAATAGTGAAGATGGACAAAAG-----TTGAAGTTCATG-GAATTAGACTATGTTGGAATTCAACTAATTGTAATAACGGTGG
GTAGTGTGATGGTGATAAAACGGATGAAAAA---TTGAAGTGGTTGAACG-GAGTGTGTTAAAGTGTAAAGTGTAAATACGGTGAT
GTAATGTACTATAGTATTAGGATAGATAAAA-----TTGAAGTGGTTGAACG-GAGTGTGTTAAAGTGTAAAGTGTAAATACGGTGAT
GTAGTGTACTATAATAATTAGGATGGTAAAGGAAATGGTAAAGGAAATGGTAAAGTGTAAAGTGTAAATACGGTGAT
GTAGTGTACTATAGTATTAGGATGGTAAAGGAAATGGTAAAGTGTAAAGTGTAAATACGGTGAT
GTAGTGTACTATAGTATTAGGATGGG-AAAA-----TTGAAGTGGTTGAACG-GAGTGTGTTAAAGTGTAAAGTGTAAATACGGTGAG
GAGCCTGTAATGGTAATGGTAAAGGAAATGGTAAAGGAAATGGTAAAGTGTAAAGTGTAAATACGGTGAT
CACTAAGGAAATGGTAAACGAGTAACAAATGAGAATAAGGAGCAATTAGAACGTCAGGTAAATAGAATAAGGAAAGATAATTAGCTGCGTTATT
CATAAAAAAATAAATTACTTAAGT TTTTTTCAAAATAC-AATGCCCAACTA-TTGGCC CACCCGTCAGCCGAGCCGTAACGGCAG
GCTCGGG
TTACAGG
TTACGTGG
TTACACGG
TTACAGG
TTACAGGGGG
TTACGG
TTACAGGG
TTACAGGGGG
TTACAGG
TTACAGG
TTACAG
TTACGG
TTACAGGGGG
TTACGG
TTACACGG
TTACAGG
TTACAGGGGG
TTACAGG
TTACAGG
TTACAGGGGG
TTACAGGGGG
TTACACGG
TTAC

ETAS2-4(LIC)

ATGCAT CGTGAATGGTTAAGGTGAAAGACGAAGTAAGTGTGGGATTCTTAAGTATAATAGAGATATAGTGTGATGATAATTGAGAAGG
 ATGAAAAATTGAAG-CTGACTCATGATGTGAGCGAGTAACGAGCATAAGGCAGTAATCGTTAAGCGTTAAGGAGGGATAAGGAAATAAG
 GAATGAGAGTAGAAGAGGAAGTAGTAAAGGAGAGTAGAAGAGGTAGAGTAGAAGAGGAAGTAGAAGAGGTAGAGTAGAAGTAGATAAAGA
 ATAAGGGAATAAGAGCAATAAACGAGGAAAGAAAGTGAAAAGAATAAAAGAGATAAAGGAGATAAAGGATAAAAG
 GGAGAATAAGAGTAAGTAAAGAAGAAAAGAAGTGTAAAGTAA-AGTATGTTGGAGTCAGTAAGTATAATAGGAACATGCGGTGACTTTG
 ATAAATGAAGTGTGAAATTGAGCTTATGTTAGATGTTAAATAATATGTTGCGTAAATAATAAAAATGAAATGAAATAAAA
 TAAAATAAAATAAAATTTAAACGAATAATATAAGAAATAAAATAATGAAATAAAATTAAACAAATAAAATGAAATAAAAATAAAT
 ACAAAATTGGAACACAATAAAGACAATTTAAATAATGATAAGAAAATATTGCAATGAATGACAAACAGCAGGAGTTGAAA
 ATTGAAATAAAATTATAAAATGTTAATAAGGAAAGAAAATTTGAGTGTGATATGTCGTCGGAGAAATG
 GTAGTGTAGTGTGAGATGAAGATGGACAAACAC - - - TGAAGATGAGTG-GAAGT-GAGGTGTTGGGATGCGTAAGTATAAAGGGGAT
 GTAGTGTGAGTGAATAACCGGATGAAAAA---TTTGAAGTTGATTGAAATT-GAGGTGTTGGAGTCAGTTAAAGGTGATAGGGACAA
 GTAGTGTACTATAATAATTAGGATGTTAAAAATA --- TGAAGTTGACTCAGTTGATTGAGGTGGGTTGAGTGTGATAGGGACAA
 AACGAGCAGTAAGCGAGTAATCGTTAACAT-**GTTAAC** GAAATATAAAAGAGGTAAACATAGGAAGTATAATATTGG
 ATAGGTAGAAATAAGTAAAGC-AGAATAGGTAGA
 --- GTAGAATAGGTATAGGTAGAATAGTAA
 GTAGAGTATGAGAAAGAGGAAGAAAGTAATGGAATAAGAAATAATAAAAATG
 GGAGAACAAAGAAGTAAGTAAAGAAGAAAAGAAGATACAATGAAGGGACTATGTTGGAGTCAGTAAGTATAATAGGAACATGCGGTGACTTTG
 ATAAATGAAGTGTGAAATTGAGCTTATGTTAGTGTAAACAAATAATGTTGCGTAAATAATAAAAATGAAATGAAATAAAA
 TAAAATAAAATAAAATTTAAACGAATAATATAAGAAATAAAATAATGAAATAAAATTAAACAAATAAAATGAAATAAAAATAAAT
 ACAAAATTGGAACACAATAAAGACAATTTAAATAATGATAAGAAAATAATGAAATAATTATTGCAATGAATGACAAACAGCAGGAGTTGAAA
 ATTGAAATAAAATTATAAAATGTTAATAAGGAAAGAAAATTATTGAGTGTGATATGTCGTCGGAGAAATG
 GTAGTGTAGTGTGAGTGAAGATGGACAAACAC - - - TGAAGATGAGTG-GAAGT-GAGGTGTTGGGATGCGTAAGTATAAAGGGGAT
 GTAGTGTCAATATGGCAATGGGAGGATGAGTGAATGAAACCTTGAAGTTTATAGAATAA-GGCTGTGTTGAAATGCGTAAGTGAATAAGATAC
 GCAGTGTATTATGATAATCAAGATGGTGAAGGATA --- TGAAGTTGACTCAGTTGATTGAGGTGGGTTGAGTGTGATAGGGACAA
 AACGAGCAGTAAGCGAGTAATCGTTAACAT-**GTTAAC** GAAATATAAAAGAGGTAAACATAGGAAGTATAATATTGG
 ATAGGTAGAAATAAGTAAAGC-AGAATAGGTAGA
 ATAGGTAGAAATAAGTAAAGC-AGAATAGGTAGA
 ATAGGTAGAAATAAGTAAAGC-AGAATAGGTAGA
 GTAGAATATGAGAAAGAGGAAGAAAGTAATGGAATAAGAAATAATAAAAATG
 GGAGAACAAAGAAGTAAGTAAAGAAGAAAAGAAGATACAATGAAGGGACTATGTTGGAGTCAGTAAGTATAATAGGAACATGCGGTGACTTTG
 ATAAATGAAGTGTGAAATTGAGCTTATGTTAGTGTAAACAAATAATGTTGCGTAAATAATAAAAATGAAATGAAATAAAA
 TAAAATAAAATAAAATTTAAACGAATAATACAGAAATAAAATAATAAA --- CAAATAAAATGAAATAAAAATAAAT
 ACAAAATTGGAACACAATAAAGACAATTTAAATAATGATAAGAAAATAATTATTGCAATGAACGACAAACAGCAGGAGTTGAAA
 ATTGAAATAAAATCATAAAATGTTGATAAGGAAAGAAAATTATTGAGGAAACGTTGAGTGTGATATGTCGTCGGGGAGGATGG
 GAAATTGAGGATGGTCATTTAATAAGGAAAGAAAATTATTGAGGAAACGTTGAGTGTGATATGTCGTCGGGGAGGATGG
 AACGAGCAGTAAGCGAGTAATCGTTAACAT-**GTTAAC** GAA
 GACAAGAATGTAACAGATAGAATTAAAGAGTAAATAGAGAGAAGAGAGTAGGTAGAAGGGAGCAGGATATATAAGTAGAATATAGAAGAAATAGA
 GTAAGGAGAG
 TAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA
 GGTAGGAGAATAAGAAGCAATTAAAGTAAGGAGAAAGAGATAAAAGCAGGGACTATATTGGAGTCAGTTAAAGTAGAATAGGGATGAGTAATGAACT
 ATAGTAATTATGATAGACAAGTGTGAAATAATGTTGGGATGCGAGAATATAATAGAGAATTGAGTGTGACTTGGTAATTGAAAGGGGATGAAAA
 ATTTGAAGTTGACTCAGTCAGTGTGAGCTGTTGAGAATCGAGTAACAAATATGGAATAAGGAAATGAGGAGAGATAAAGAGTAGGATATAGAAGAGA
 GTAGAATCAAAATAAAAGTAAGGAGAG
 TAAA-TTAAATTTAAACGAATAATAAAATAAAATAAAATAAAATAAAATAAAATAAA
 GTAGGAGAAT-GAAGAAGTAATCAAATA --- GTAGGAGAGTATTAGA --- AAGTAAAGTAA-TAAGGATGGATAC
 ATAGTGTATATTGAAAGTCAG**ATGCAT**

Supplementary Figure S2. The fission yeast subtelomere DNA structure. (A) Chromosomal view of the repeats located proximal to the telomeres. The nomenclature of the repeats is the same as that used in Figure 1A. The copy number of STE2 in each TAS2 is also indicated. (B) Graphical summary of the DNA structures found in the fission yeast telomeric clones in the Sanger Institute database (pNSU70, pNSU71, pNSU77, pNSU21, pNSU64, and pNSU65; available from the Sanger Institute ftp site: <ftp://ftp.sanger.ac.uk/pub/yeast/sequences/pombe/telomeres/>). The numbers associated with the restriction sites refer to the original numbering of the clones. The repetitive units STE1, STE1', STE2, and STE2' are indicated in yellow, cyan, grey, and lavender, respectively. STE1 is a tandem repeat of approximately 88 bp, whereas the other repeats are essentially composed of unique sequences, although STE2' harbours a direct repeat (see (D) for details). The tel repeat region is represented by a stack of filled triangles. The relative positions of TAS1 (Apal/EcoRI), TAS2 (Nsil/Nsil), and TAS3 (EcoRV/HindIII) are also indicated for each clone. For pNSU77, the fused rDNA sequence and the break point (BP) beyond the tel repeats are indicated. (C) Sequence alignment of the STE1 repeats found in pNSU70. The numbers refer to the original numbering of the clone. (D) Sequence analyses of pNSU70, pNSU71, pNSU77, pNSU21, pNSU64, and pNSU65. The shaded sequences indicate the repetitive units of STE1, STE1', STE2, and STE2', with the same colour coding used in (B). The restriction sites shown in (B) are highlighted in magenta. The primers used for amplification of STE2 by qPCR are underlined in every STE2 repeat (to confirm their capacity to bind to each repeat with one nucleotide degeneration, as indicated by a dashed line). The reference sequence of telomeric DNA (43) takes into account the sequence of pNSU70. (E) The genomic sequence of the longest TAS2 clone. The wild-type genome was digested with Nsil, and the fragments encompassing TAS2 were cloned directly into a vector via ligation-independent cloning (LIC) technology. The structural properties found in the longest TAS2 clone (TAS2-4(LIC)) were colour-coded as described in (B).

A

cd1-31 (1R;2L)

(H1) cen ==> AAAGAGATAGAAGATGAAAATTGGAAAGTA--GATGAAATTAGTCATTCTATTACACTAAAACA
 (H1') tel <== ... GAT C

AATCAACTAAATTATTGAAAAAACAGTCGTTACAATTACTGGTGATGTGTCAGCTGGCAATGAATATGTCAAATAGGAATTG
 A A .

AATACGATA TAGAAAAGTACAACCTTTGTTATTGTTG . T
 GAGAAAAGTACAACCTTTGTTATTGTTATTGAAAGTAAAAAAATAAAGTAGAGAATAAAGTAGTAACAGATAA
 G . => tel (H1)
 TGAAACAATATGAAACAAATAGAGAAAAAGATTAATTCGTTAATTA <== cen (H1')

cd1-65 (1L;2R)

(H3) cen ==> GTAAAATCTCGCTATTGTTATTGTAATGATGAAGAGTCATGGGAGATGAATGTTGAAACG
 (H3') tel <== ... G A .

ATGGCATAGAATTGGTAACGAAAAGTGAACATATTGGGATCAACTATTCAGTATTGTTAAAGAAAATGTTGAACTCGC
 GTTGGGATCAACTATTCAGTATTGTTAAAGAAAATGTTGAACTCGA
 - TC T
 CAAGTAATGAGAGGTGGTCTTCGTTAATAATGAGTGGTGGTACGGTTACAGGATATGATCTGTATGGT GAGAAT
 GTGATGTATGTATTGAGTATAGACAAT => tel (H3)
 <== cen (H3')

cd1-47 (1L;2L)

(H4) cen ==> GCAAATGGAAAGACACGAATCATGATCGC-TACCAAGGCATT CGGACTCGGTATCAACTATATGGGAG
 (H4') tel <== ... A T G G T

TGCGTTAGTAGTACACTATGATTACAGCTTCATCTATGGATTATGTACAGGAGACAGGTGAGCTGGAAAGAGATGGCAA
 AGATTACAGCTTCATCTATGGATTATGTACAGGAGACAGGTGAGCTGGAAAGAGATGGCAA

GTATGCGATTGCAGCATTGTTACAGGAAATATGATTCTACATGGTCGAGCTACGT
 GTATGCGATTGCAGCATTGTTACAGGAAATATGATTCTACATGGTCGAGCTACGTGGAGATTGATGAAACAACTTCTT
 AATGATAATACGATGTGTTGATCGTTCTCGCAAGTGAATGGATGGCGAATGTGTATGTTGATCGTTGCTA
 GTGTTACTGCTCAAGATGCTCAGATTGTTACTGGTGAAGAATCAACTGTGTACGATGTATGGAGTGAACCGACATT
 GCCAGAAACACCGAAACCAGCCATTGCAACACATTGCGTTATAATGCATGTTTCGTTCCACAGGCCAGGG
 A => tel (H4)
 ATAGCAGTGGTATGAGTGCTATGAACACTAACACTAGTACTACGCCAGTGTCT <== cen (H4')

cd1-51 (1R;2R)

(H5) cen ==> TAATGATAATGACAATGATGTGTTCTCAAATTACATTGGCTAAATCTGCTATTAAAAGTATGAGA
 (H5') tel <== ... C C .

CAAAGGCATCTATCTTCAATGAGTTATTGTTTGCTAGTGTAC-ATATCTGCTGGACAACCCAGCCAGAGCAAGAGATG
 A T AT TG .

GTGTATTGGACTTGCGGAATGGCAAGTATAAGACTCGCAATTGTATTGATGTTGAGGCTGATGATTACAGCAGAT
 A T A .

ACGATAAGACTCGTAATATGAAGTTGCTGAAAAGCCAATCCCCAGGTTCTTGAGCCGCTTCCATTAGCACTTCG
 A T .

GTACTATGTTGGTTCGACCATTGGAAGCATTGATGAAGTATGTGACAACCGCTGATAGGTGAAAGTAGCTGTAT- A
 A C GCAC .

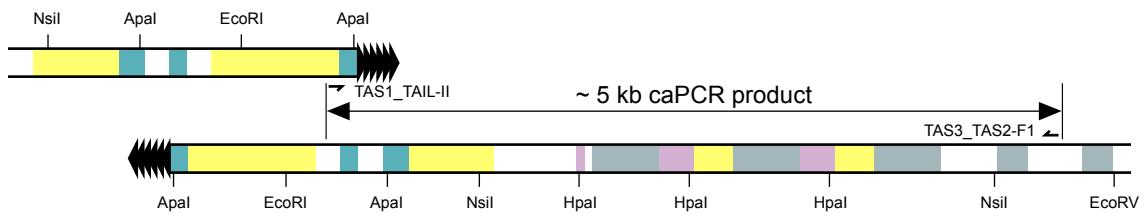
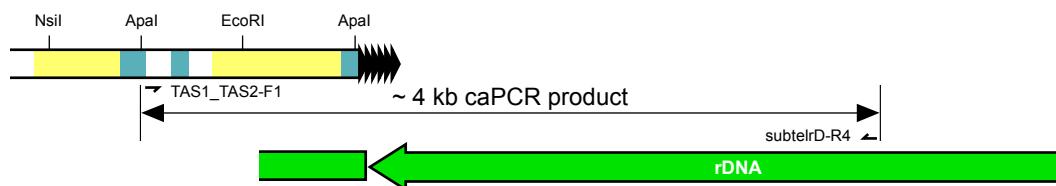
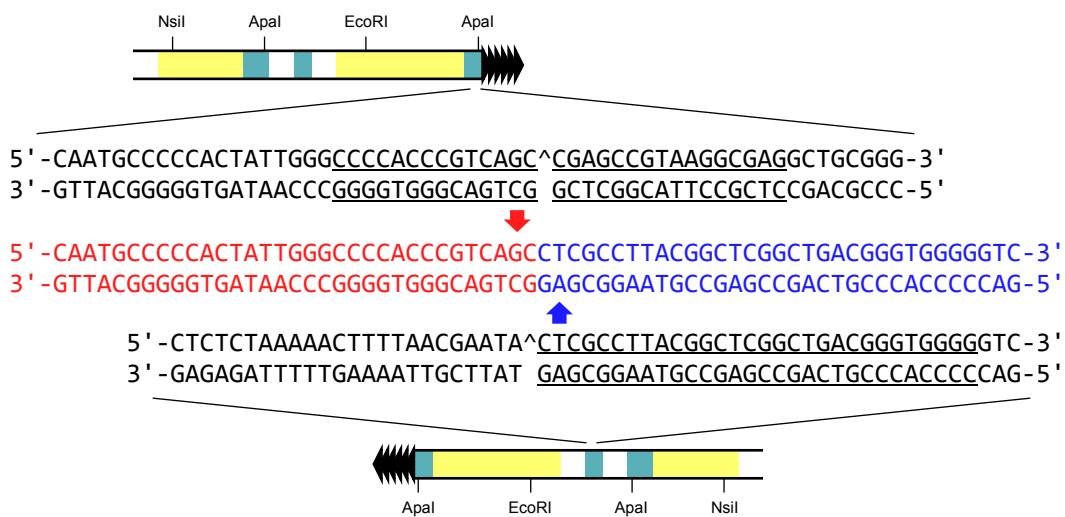
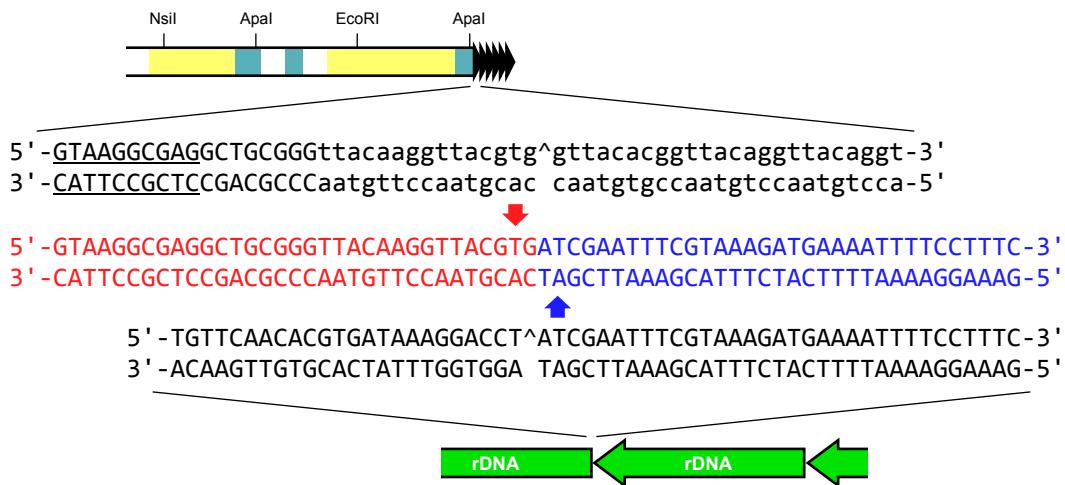
CTTGGATTTCATGTTGATTGCTGGCGAACGATTGCAAAGAGATTACCGTATCGAATTTCCTCAAAGGCCACCTACCAA
 T T .

TGCATTCAAGAAACCGTTGGATTTCAGGAACTACAGACACATTGCTCACTACTTTAAAGAAAAAGAACATCGAGAAAGACATG
 G A .

ACGAGAGAATCTTACAGGCTGACATACAGAACACAGCGCTCACATCTATGGACGCACTATGGACA
 G C .

TGCATTATCTGCCATCGGATTATTCGTCACACTTTCTGCAAGCTATAAGTGGCAGGAACATTGAGATTGAGACAA
 G C .

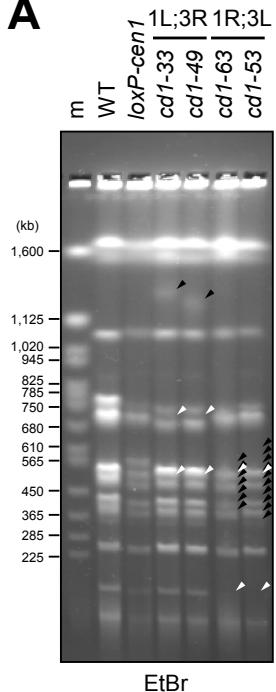
CCCACCCATGGATGTTGGTGAACAAAGCACCCATTCAAGCGAGTTGATCAATTGGAG => tel (H5)
 C G .
 CCCGACCCATGGATGTTGGTGAACAAAGCACCCATTCAAGCGAGTTGATCAATTGGAG <== cen (H5')

B *cd1-98 (1L:2L)**cd1-53, cd1-63 (1R:3L)***C** *cd1-98 (1L:2L)**cd1-53, cd1-63 (1R:3L)*

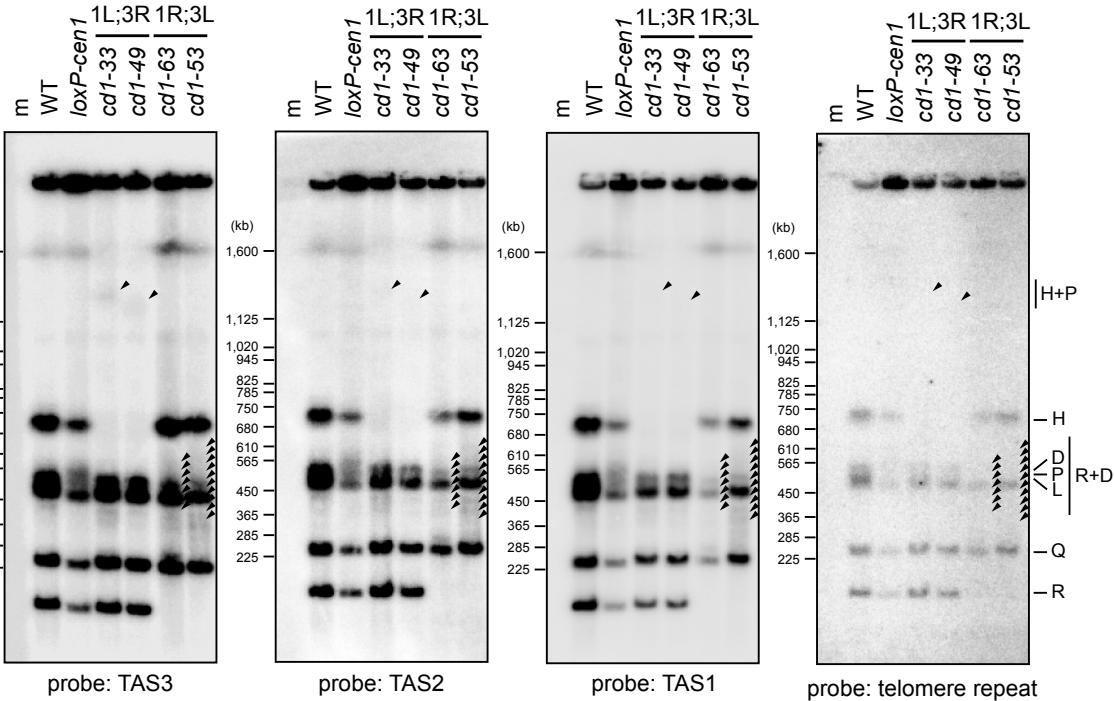
Supplementary Figure S3. Sequence analysis of the chromosome fusion junction. (A) Alignment of the inversely homologous subtelomere sequences and the actual fusion sequences of *cd1-31*, *cd1-65*, *cd1-47*, and *cd1-51* (grey-shaded text). Nucleotide degenerations in the original subtelomere sequences are highlighted in red and blue. The actual fusion sequences were determined by sequencing the gPCR products. The site of crossover in the fusogenic sequence can be confined to a limited segment according to the choice of degenerated nucleotides, which is indicated by the overlapping grey-shaded text. The nomenclature of the homology segments (H1–H5 and H1'–H5') follows that described by Wang and Baumann (43). (B) Schematic diagram of the NHEJ-type fusion (*cd1-98*, *cd1-53*, and *cd1-63*). Colour coding of the DNA structures is the same as that used in Supplementary Figure S2B. The primers sets used for the gPCR analyses of the *Δcen1-f* rearrangements are also shown. (C) Experimentally determined junction sequences. Subcloning and subsequent sequencing of the gPCR products revealed that all of the fusion events were attributable to the canonical NHEJ pathway.

Supplementary Figure S4

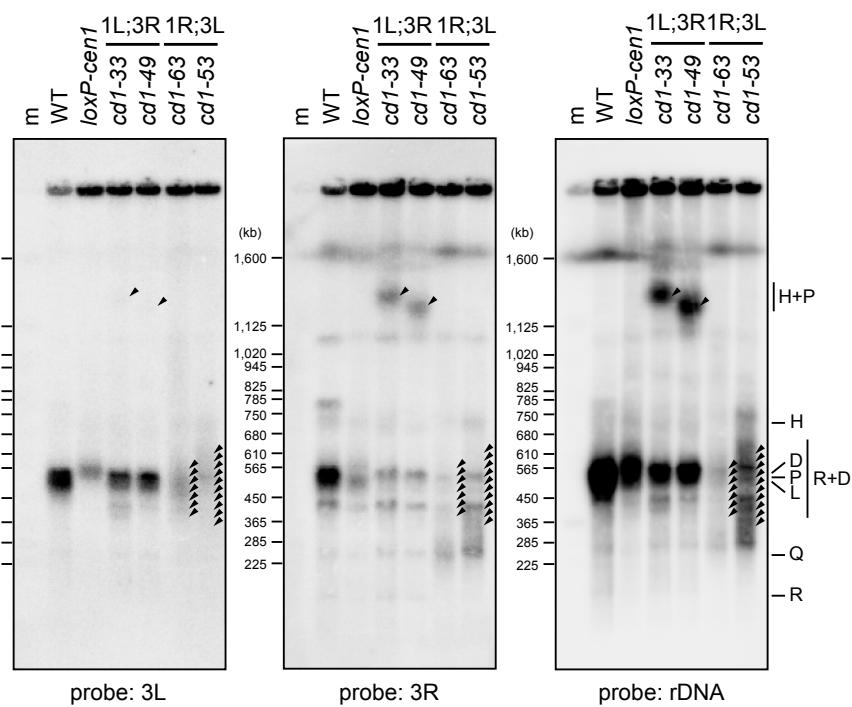
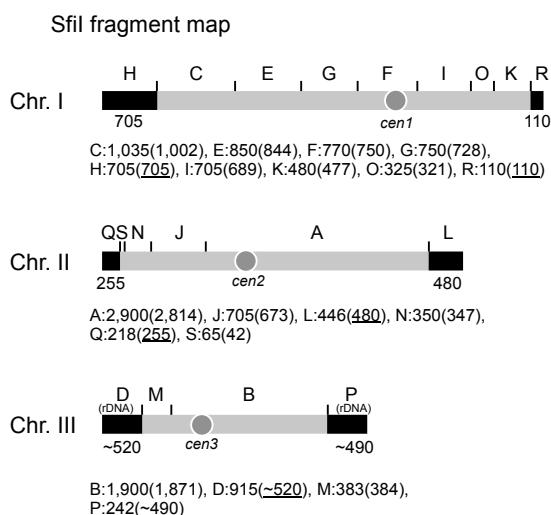
A



B



C

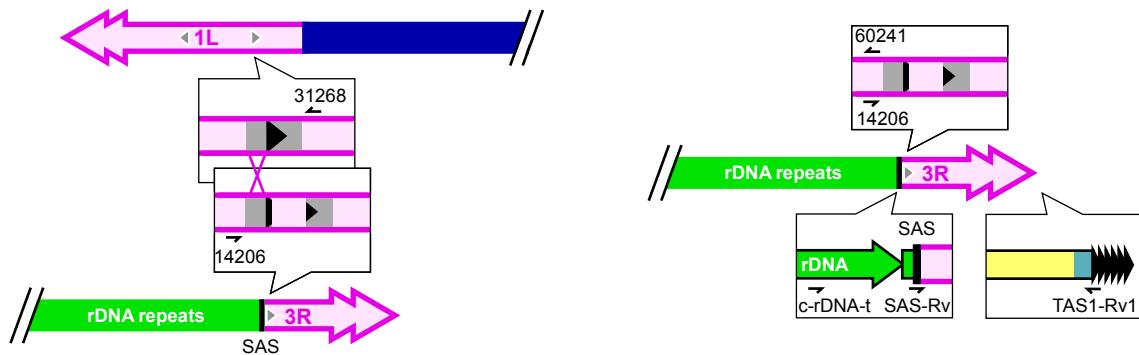


Supplementary Figure S4. Sfil digestion of chromosome III-related fusion survivors. (A, B) PFGE analyses of Sfil-digested chromosomes in the $\Delta cen1-f(1;3)$ survivors. The gel was subjected to EtBr staining (A) and Southern blotting with the indicated probes (B). The white arrowheads indicate the Sfil bands that disappeared in the survivors due to fusion, and the black arrowheads indicate the newly-generated fusion bands. The identities of the bands with altered expression levels are shown at the right-hand side of the gels. The nomenclature of the Sfil fragments follows that described by Fan et al. (64) (see (C)). Because the rDNA repeat length was variable, particularly in the survivors, the band intensity of the Sfil fragment containing rDNA was low on some occasions. m, molecular size marker. (C) Schematic diagram of Sfil fragments localised along the fission yeast chromosomes, with emphasis on the terminal fragments. The Sfil-digested fragment sizes reported by Fan et al. (64) are indicated below each chromosome. The values in parentheses correspond to the fragment sizes calculated from fission yeast whole-genome sequences. The underlined values in parentheses correspond to terminal fragments that were not precisely defined in the database; thus the experimentally determined values from our wild-type laboratory yeast strain and the values described by Fan et al. (64) are shown.

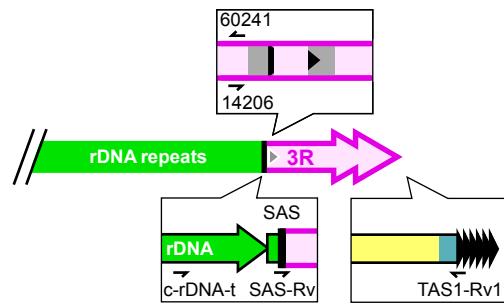
Supplementary Figure S5

A

$\Delta cen1-f(1L;3R)$ (*cd1-33, cd1-49*)



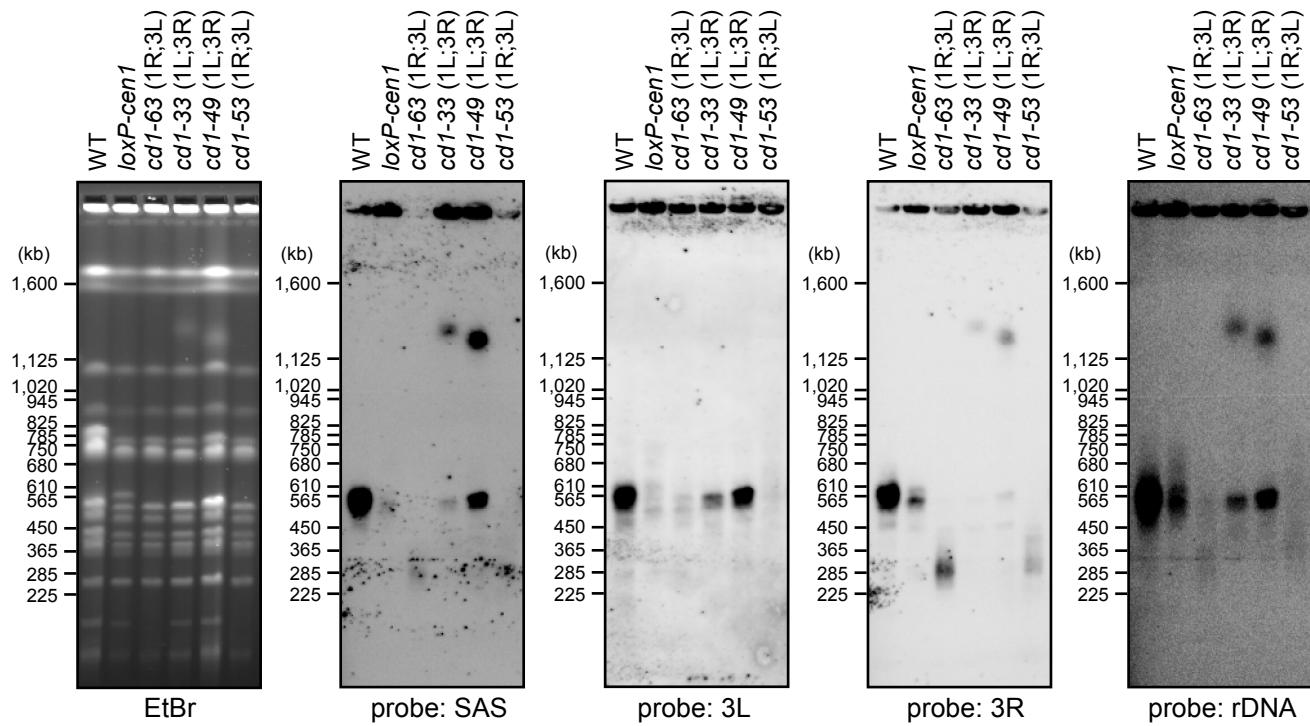
Wild-type



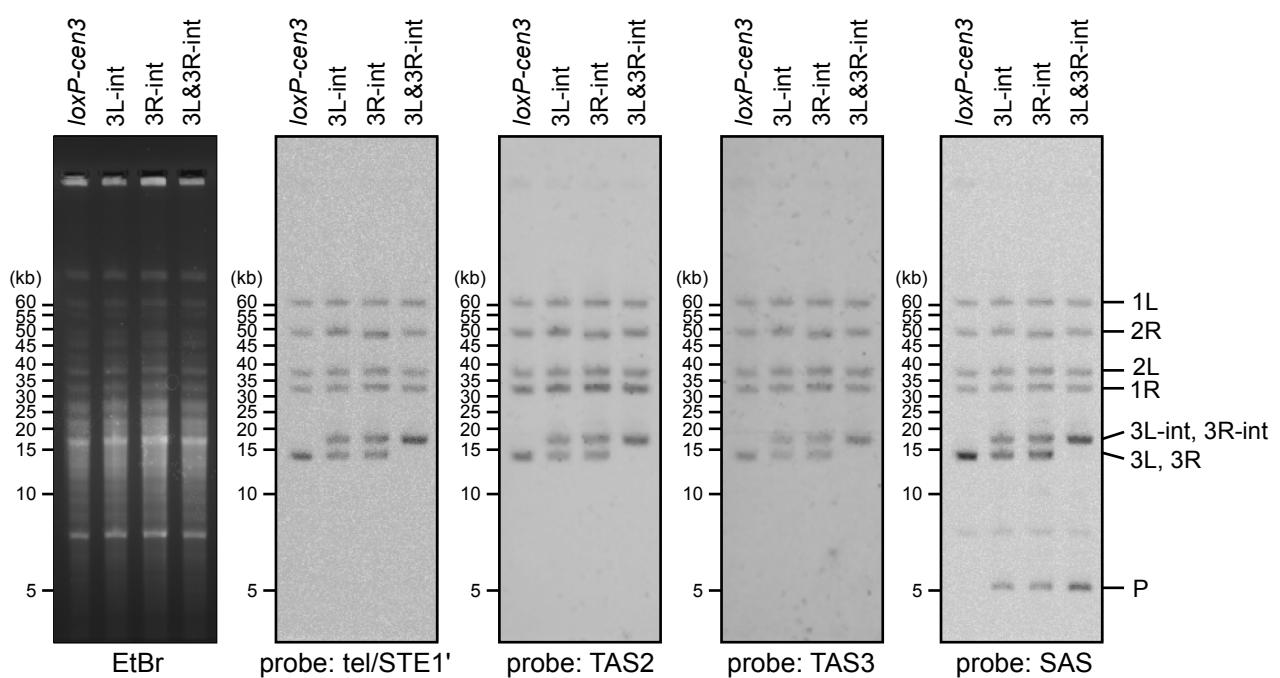
B

(rDNA) ==> TAGAGGTTGGAGATGGTAGGATTGGATGGTTGTGAGATGGGAAATGACGTAAGAATTGAACTTACGAGAGAGAACGAGTCTATCGGGTTTTGG
 CATGTCCTATCTCAGTTGAATAGGGAGAGTTAGGGATAGAGAACGATTTCGATGAGAATCAACGATGAGTAACGAGAGGGAAATGAACTGGAGAACAGACAGACT
 CATCGTTCAATGCAATGCCATGCCATGCAATGCCATCTCACTTAGAATAACGTAGAGTAAGGAAAAAAGGACATCTCAGTTGAAATGAGTCGATCTCACTTATGAA
 TAAGGAGGAGTAGGGAAAAGGAAAAAATAAAAAATAAAAAATAAATAAAATTAAATGAATAAAAATAAAAATAAAAATAAAAATAAAAATAAAAATAAAAATAAAT
 TAAAATGAATAAAAATTGTTATTCAATGAAATTGTTAAATAAATAAAATTACAATTGCAAAAAAAACATTGATTACTTTCTTTTTTTTATTGCAATA
 AAATGGAATTGTTGGTACACAATTATTATTCATGACACCTTAAAAAATTGTTGAATAAAAAGAAGAAAAGAAAAGAAAAGAAATCATTAAATTATCAAATTG
 CTATACATAAAAGATTAGAGGAAATGGATTATCTTGGTATTCAAGAATAGGAAGGGTGAGGCAACCCCTTGGTGGGAAGTGGATTCTGTAGTACTGCGTAT
 TCGCGTCGACAATGTCACCCCTACCCCTAGAACGACCGATATGGGAAATTCAAAAGTAGCACCACCTTCTACAAAATCGGAAAATAGGTCAAAGGTAGTATGAA
 AGGGTAGAGGGAGGAGAATAAAAAGTTCAAGTCGAAACTTGTCAAGGTTAAAGTTCAATCTTTTACTTGGATGCTTTGACATGGCTGATTCCTGT
 CATTTCGATGGCTGGTCTATTCACCTTTAGGCTGGTCAACAGGGTTGATAAGCGATGGCTGACTGGGCTGGTACTGTTTACATTGTTATATAATGCAAACTT
 TTTTATTGTTGATACATTATCCCTCATTTGTTGTTCACATTCTACTGCC ==> (subtelomere)

C

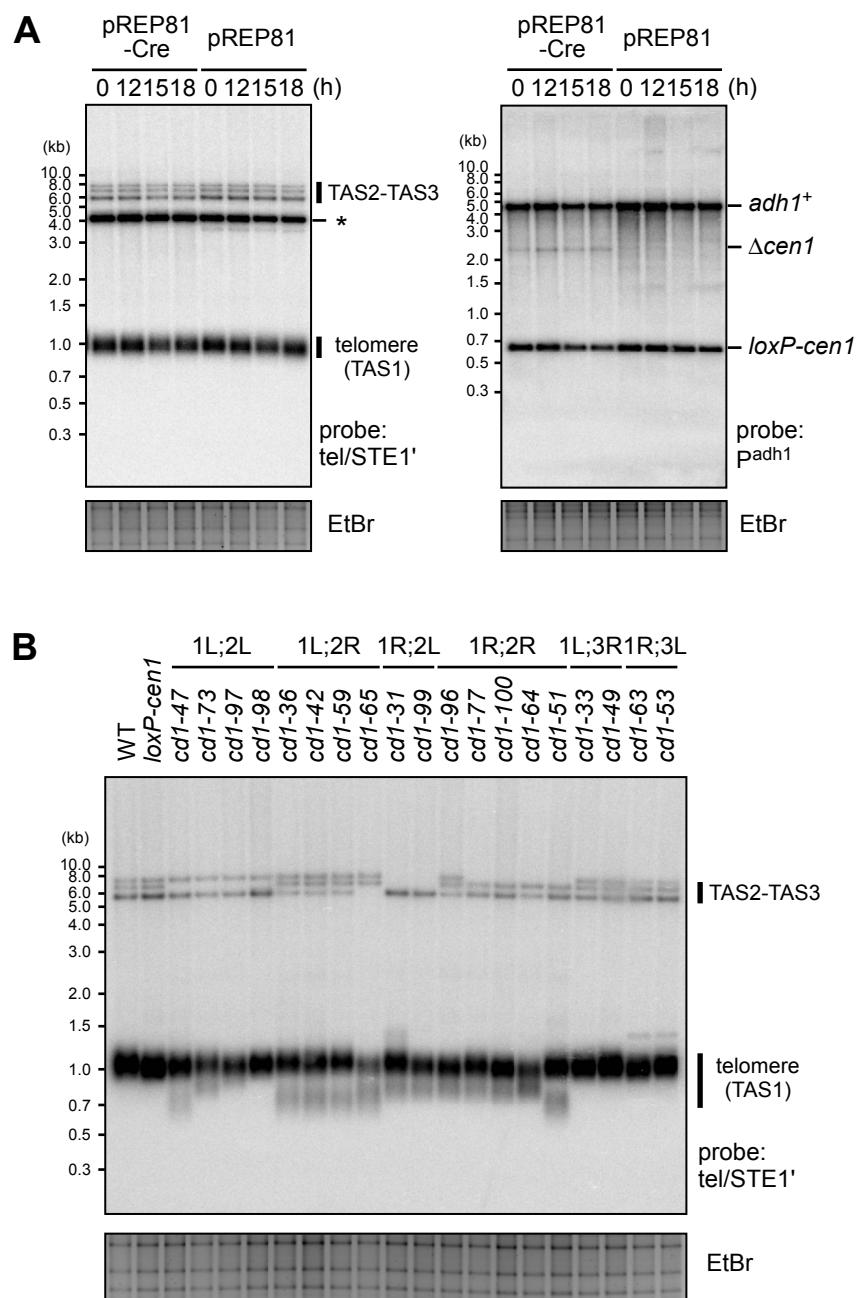


D

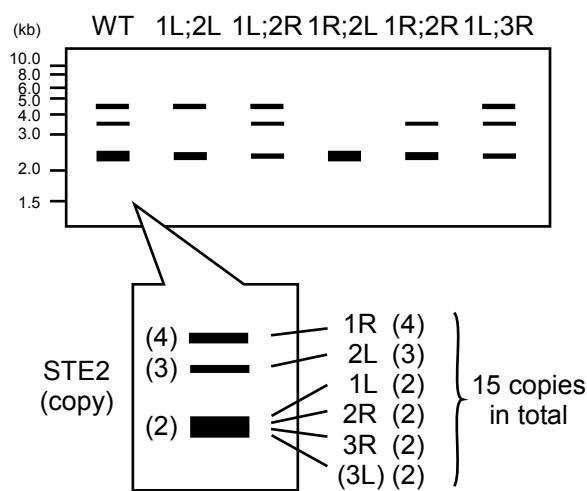
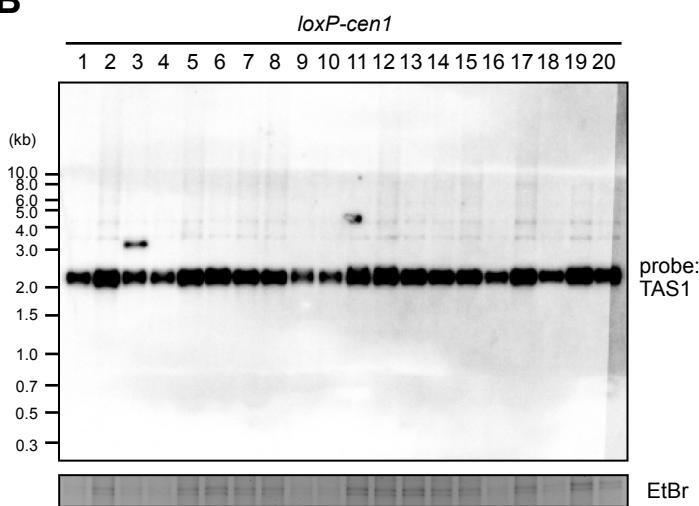
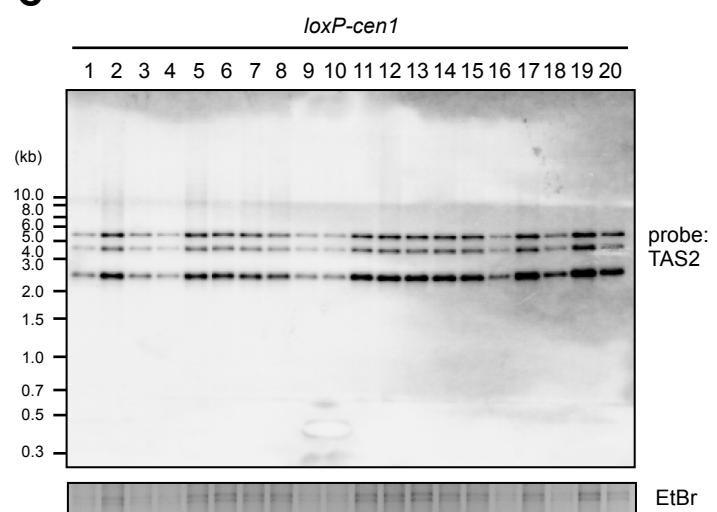
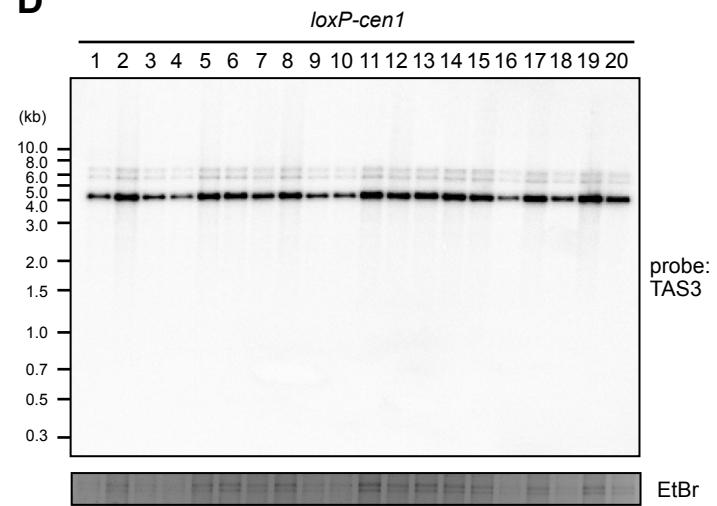


Supplementary Figure S5. Identification of a novel sequence (SAS) located between the rDNA repeats and the subtelomere region of chromosome III. (A) Overview of the gPCR analyses. The left panel shows a gPCR analysis using primers 31268 and 14206 to identify the fusion point in the $\Delta cen1-f(1;3)$ survivors (*cd1-33* and *cd1-49*). The success of the gPCR analysis suggested the existence of a 31268- or 14206-corresponding sequence at the terminal region of 3R. The right panel shows a gPCR analysis of the wild-type genome using primer 60241, which was complementary to 14206, together with a series of rDNA primers that hybridised in the outward orientation. One of these primers, c-rDNA-t, yielded a gPCR product containing a novel 1169 bp sequence (SAS) located between the rDNA repeats and the subtelomere sequence. The end-adjacency of SAS was confirmed further by a gPCR analysis with an SAS-specific primer (SAS-Rv) and a terminal primer (TAS1-Rv1). (B) Sequence analysis of SAS. The SAS sequence showed 89% homology with a newly identified sequence at a chromosome BP of the *S. pombe* isolate strain CBS2777 (54). It also exhibited weak homology to STE2 (81% identity over 54 bp), possibly causing cross-hybridisation with TAS2 in some Southern blot analyses (see (D)). (C) The existence of the SAS at both ends of chromosome III, as confirmed by Southern blotting of SfiI-digested chromosomes. A pair of chromosome III terminal fragments in our wild-type laboratory yeast strain migrated at almost the same position in PFGE experiments, making it difficult to assess the existence of the SAS. However, fusogenic rearrangement in the $\Delta cen1-f(1;3)$ survivors (*cd1-33* and *cd1-49*) altered the migration position of one of the fragments. The observation of two SAS hybridisation signals in these strains effectively localised the SAS at both ends of chromosome III. (D) The presence of subtelomere repeats at the SAS-distal ends of 3R and 3L. A plasmid containing a BamHI site was integrated into one or both SASs at 3R and 3L in the *loxP-cen3* strain (40). To confirm the presence of TAS3, TAS2, TAS1 (STE1'), and tel repeats in the SAS-containing fragments, BamHI-digested chromosomes from the resulting integrants were subjected to PFGE followed by Southern blotting. The identities of the other subtelomeric bands were determined by PFGE and Southern blot analyses of the integrants in which TAS3 of any of the chromosomal ends could be distinguished by a newly created BamHI site (data not shown). P, plasmid-sized genomic fragment generated due to the sequential integration of multiple plasmids into the target locus of the host genome.

Supplementary Figure S6



Supplementary Figure S6. The telomere length remains constant during the induction of centromere deletion. (A) Southern blot analyses of genomic DNAs recovered from *loxP-cen1* cells harbouring a Cre-inducible plasmid (pREP81-Cre) or empty vector (pREP81) (42). The DNAs were digested with EcoRI (left), or BsiWI plus BssHII (right), at the indicated time (h) after induction. Southern blotting was performed using probes encompassing both the tel repeats and the STE1' repeats (tel/STE1', left), or the *adh1⁺* promoter region (P^{adh1} , right). EtBr-stained images of the gels are shown as loading controls. The TAS2-TAS3 bands in the left-hand gel indicate signals that were most likely derived from the STE1' repeats located between TAS1 and TAS2 (see Supplementary Figure S1). The asterisk indicates a non-specific hybridisation signal that was not reproducible. The *loxP-cen1* signals in the right-hand gel were derived from the intact *cen1* genome, whereas the $\Delta cen1$ signals correspond to *cen1*-excised DNA. The *adh1⁺* signals were derived from the endogenous *adh1⁺* gene, which remained at constant levels. By contrast, decreased *loxP-cen1* and increased $\Delta cen1$ signals were observed upon Cre induction. (B) Southern blot analyses of genomic DNAs recovered from the indicated telomere-fusion survivors. The experiment was performed as described in (A). Variations in the TAS2-TAS3 signal intensity are representative of subtelomere instability in the survivors. Alterations in the telomere signal were stable and specific to a given chromosome; hence, they most likely reflect instability of the TAS1 repeat rather than the tel repeat.

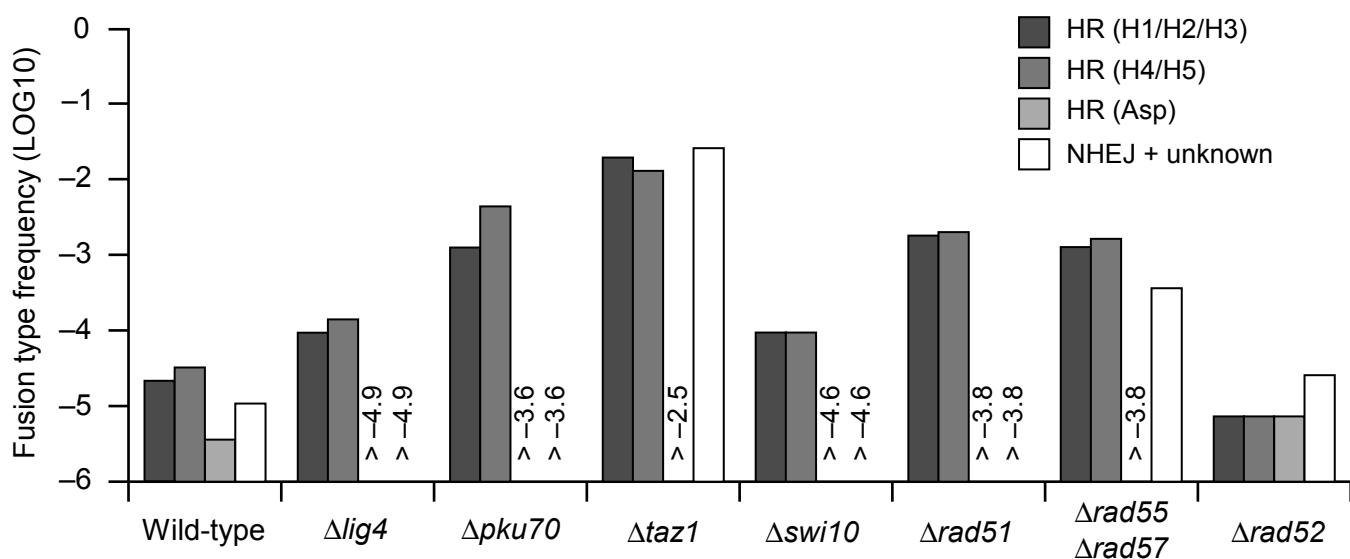
A**B****C****D**

Supplementary Figure S7. The wild-type subtelomere DNA structure. (A) Schematic diagram of representative TAS2 hybridisation patterns in wild-type cells and the indicated fusion strains. The original data refer to (C) and Figure 3B. The identity of each band was deduced based on the disappearance of the common band(s) in assorted fusion survivors, as shown in the balloon. The *STE2* copy numbers are also indicated in parentheses. For rearrangements involving 3L, only NHEJ-type events were observed; therefore, the *STE2* copy number could not be deduced accurately from the TAS2 Southern blot. The identity of the 3L band was determined based on the elimination of the other identified bands. (B–D) Southern blot analyses of genomic DNAs recovered from 20 independent *loxP-cen1* strain isolates of various ages. The DNAs were digested with *Nsi*I and then subjected to Southern blotting with the subtelomeric TAS1 (B), TAS2 (C), and TAS3 (D) probes. EtBr-stained images of the gels are shown as loading controls. Almost no structural variation was observed between the *loxP-cen1* isolates, with the exception of an additional TAS1 band in isolate #3.

Supplementary Figure S8

A

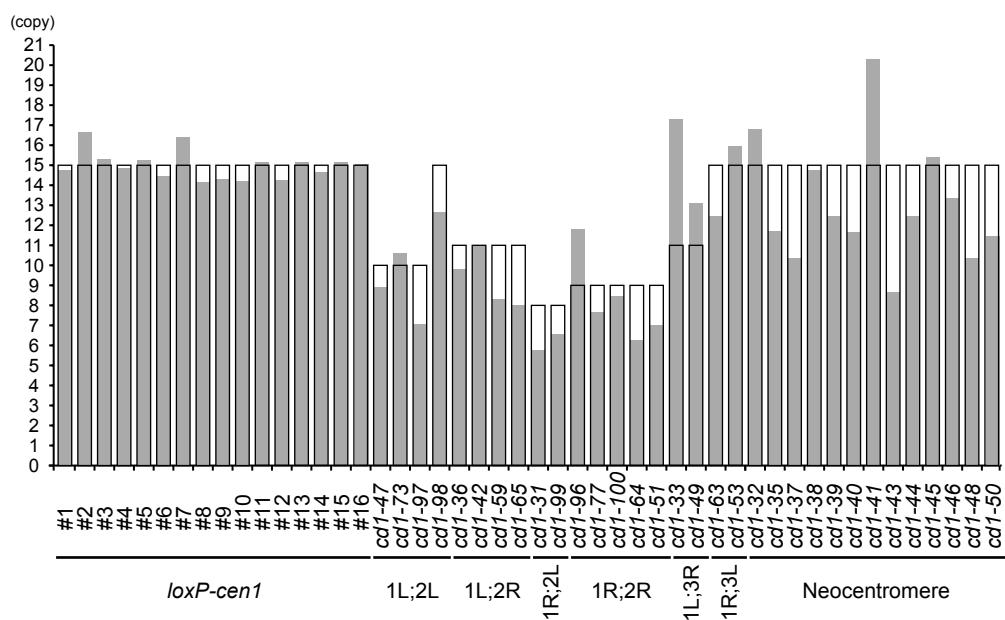
	class	I	II	III	IV	V
	PFGE-Southern blot	tel(–) TAS1(–) TAS2(–) TAS3(+)	tel(–) TAS1(–) TAS2(–) TAS3(+)	tel(–) TAS1(–) TAS2(–) TAS3(–)	tel(–) TAS1(–) TAS2(–) TAS3(–)	tel(+/–) TAS1(+) TAS2(+) TAS3(+)
	gPCR	H1/H2/H3	H4/H5	Asp	–	–
Wild-type	19	6 (31.6)	9 (47.3)	1 (5.3)	0 (>5.0)	3 (15.8)
$\Delta lig4$	5	2 (40.0)	3 (60.0)	0 (>16.6)	0 (>16.6)	0 (>16.6)
$\Delta pku70$	18	4 (22.2)	14 (77.8)	0 (>5.2)	0 (>5.2)	0 (>5.2)
$\Delta taz1$	18	6 (33.3)	4 (22.2)	0 (>5.2)	0 (>5.2)	8 (44.4)
$\Delta swi10$	6	3 (50.0)	3 (50.0)	0 (>14.2)	0 (>14.2)	0 (>14.2)
$\Delta rad51$	19	9 (47.4)	10 (52.6)	0 (>5.0)	0 (>5.0)	0 (>5.0)
$\Delta rad55\Delta rad57$	18	7 (38.9)	9 (50.0)	0 (>5.2)	2 (11.1)	0 (>5.2)
$\Delta rad52$	13	2 (15.4)	2 (15.4)	2 (15.4)	7 (53.8)	0 (>7.1)

B

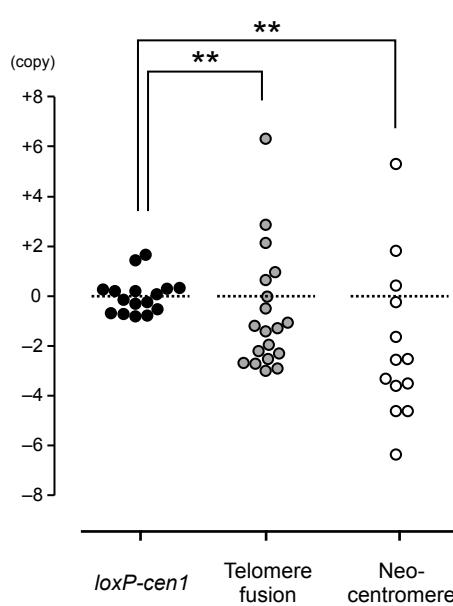
Supplementary Figure S8. Fusion spectrum analysis. (A) Classification of the $\Delta cen1$ -f survivors in the indicated strain backgrounds. The five different types of classification (I–V) were based on the results of PFGE-Southern blot and gPCR analyses. The numbers in parentheses represent the percentage of each class of survivors in respective strains. Classes I–III are the HR-type and class V is the NHEJ-type; with the exception of those in the wild-type, $\Delta rad51$, and $\Delta rad52$ backgrounds, the precise fusion points have not been determined. The identity of class IV remains unclear. (B) The calculated frequency of each fusion type in the indicated strain backgrounds. The calculations were based on the $\Delta cen1$ -f survivor frequencies and fusion spectrum. The frequencies of classes IV and V were combined and are indicated as “NHEJ + unknown”.

Supplementary Figure S9

A



B



Supplementary Figure S9. TAS2 instability occurs also in neocentromere survivors. (A) The predicted and experimentally determined STE2 copy numbers in the *loxP-cen1* clones, the telomere-fusion survivors, and the neocentromere survivors (shown as described in Figure 3D). Like the telomere-fusion survivors, TAS2 was also destabilised in the neocentromere survivors. (B) Scatter plot showing the differences between the predicted and experimentally determined STE2 copy numbers in the strains shown in (A). ** $P < 0.01$ by a Welch's two-tailed t-test.