Supplemental Material

Title: Telomerase-independent survival leads to diverse and complex subtelomere rearrangements in *Chlamydomonas reinhardtii*

Frédéric Chaux[1,3], Nicolas Agier[1,*], Clotilde Garrido[1,*], Gilles Fischer[1], Stephan Eberhard[2] and Zhou Xu[1]

Affiliations:

[1]Sorbonne Université, CNRS, UMR7238, Institut de Biologie Paris-Seine, Laboratory of Computational and Quantitative Biology, 75005 Paris, France

[2]Sorbonne Université, CNRS, UMR7141, Institut de Biologie Physico-Chimique, Laboratory of Chloroplast Biology and Light-Sensing in Microalgae, 75005 Paris, France

[3]present address: Laboratory for Marine and Atmospheric Biogeochemistry, Ruđer Bošković Institute, 10000 Zagreb, Croatia

*These authors made equal contributions to this work.

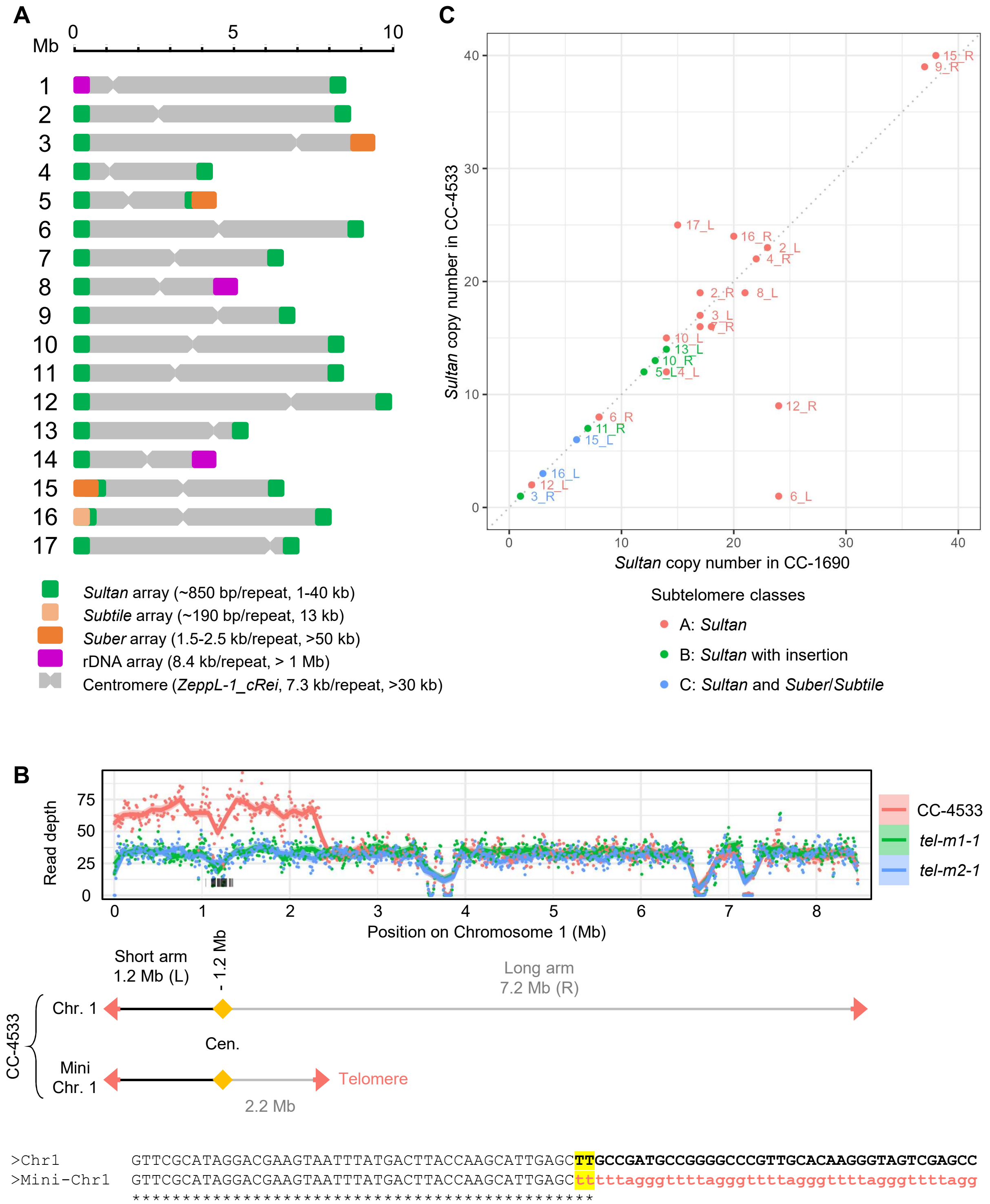
Table of Contents

Strain	Read count	Total Gb	Depth	Mean read length (kb)	N50 (kb)
CC-4533	541 694	5.42	35×	10.0	19.9
tel-m1-0	308 009	3.23	25×	10.5	24.1
tel-m1-1	917 886	4.78	35×	5.2	14.3
tel-m2-1	381 156	4.18	35×	11.0	19.9

Supplemental Table 1. CC-4533 and telomerase mutants Nanopore sequencing statistics. Shown are the “passed” datasets (read quality score >7). “N50” is the median read length weighted by base count.

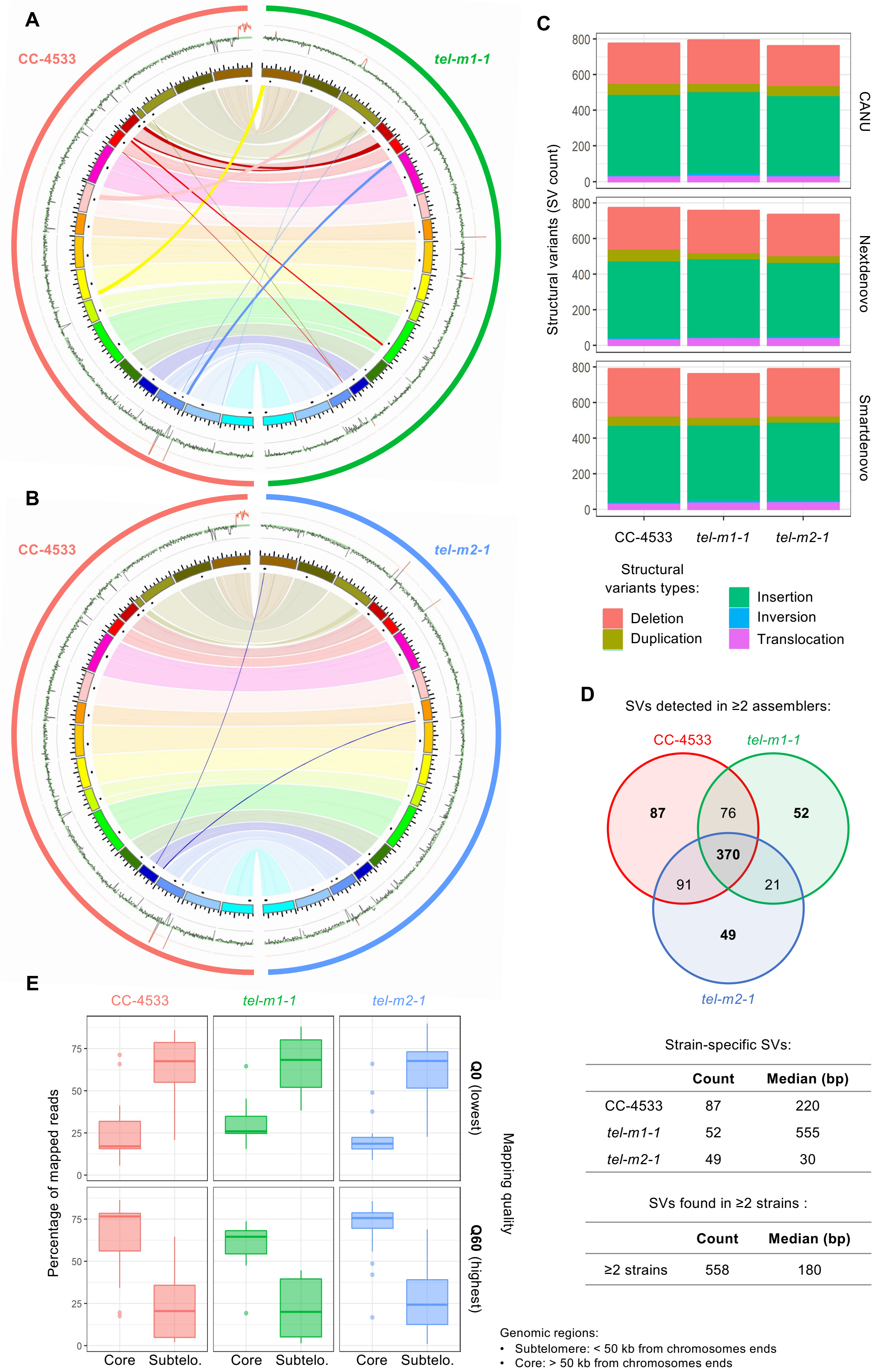
Strain	Assembler	Genome size (Mb)	Count of contigs	N50 (Mb)	Contig length (kb)					
					Minimum	Average	Maximum	Quartiles		
								Q1	Q2	Q3
CC-4533	Canu	117.7	88	6.1	16	1,337	10,184	54	100	1,606
	NextDenovo	114.0	54	6.1	26	2,112	11,858	237	605	3,220
	SMARTdenovo	114.6	109	2.0	13	1,052	5,141	189	698	1,537
tel-m1-1	Canu	121.7	248	2.4	2	491	8,057	26	52	290
	NextDenovo	116.8	71	3.6	15	1,645	12,157	155	807	2,358
	SMARTdenovo	119.9	138	2.7	12	869	7,197	59	194	1,000
tel-m2-1	Canu	118.3	162	3.6	3	730	9,721	40	71	624
	NextDenovo	113.7	59	3.6	14	1,927	9,884	207	1,015	3,109
	SMARTdenovo	114.2	106	3.2	17	1,077	7,623	80	239	1,416

Supplemental Table 2. Assembly-level statistics for CC-4533 and telomerase mutants. The assembly statistics for each strain using 3 assemblers (Canu, NextDenovo and SMARTdenovo) are shown. The displayed N50 corresponds to the contig N50. For each chromosome, the assembly giving the most colinear structure with CC-1690 was selected to build the genome model.



Supplemental Figure S1. The long-term culture of wild-type CC-4533 shows little rearrangements compared to CC-1690. (A) Schematic representation of all 17 chromosome of *C. reinhardtii* and the composition of their subtelomeres. Chromosomes are drawn at scale, but not the subtelomeres. (B) Duplication of a centromere-containing fragment of Chromosome 1 in the long-term culture of CC-4533, termed "mini-Chromosome 1". (*Upper part*) Read depth of the indicated strains over

Chromosome 1, with the doubled coverage for CC-4533 between 0 and 2.2 Mb. (*Middle part*) Schematic representation of mini-Chromosome 1. (*Lower part*) Junction sequence to the new telomere on the right arm of mini-Chromosome 1, with a 2-bp microhomology highlighted in yellow. (C) Comparison between the *Sultan* copy number at each class A, B and C subtelomere between the CC-1690 and the long-term culture of CC-4533.



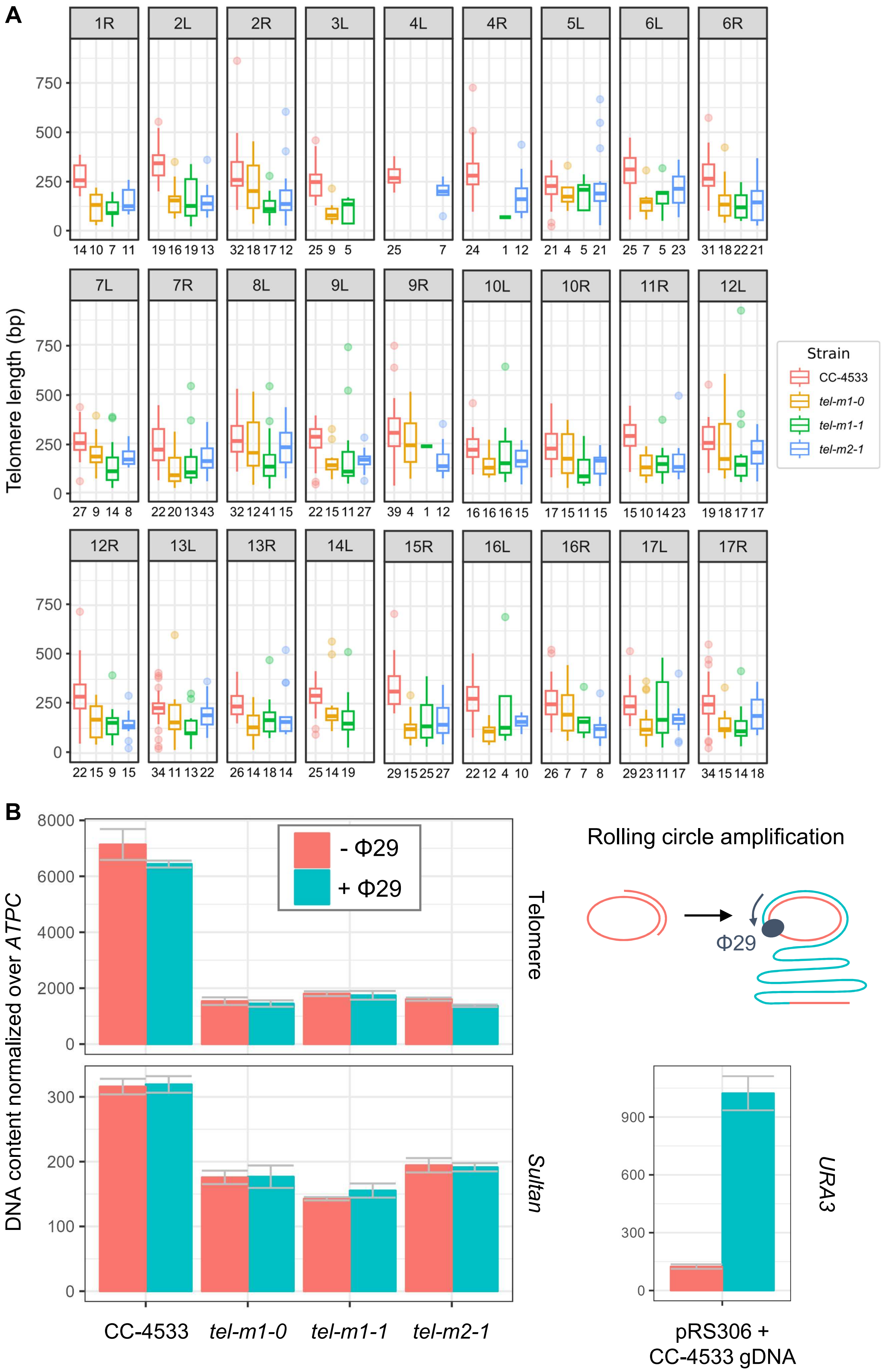
Strain-specific SVs:

	Count	Median (bp)
CC-4533	87	220
tel-m1-1	52	555
tel-m2-1	49	30

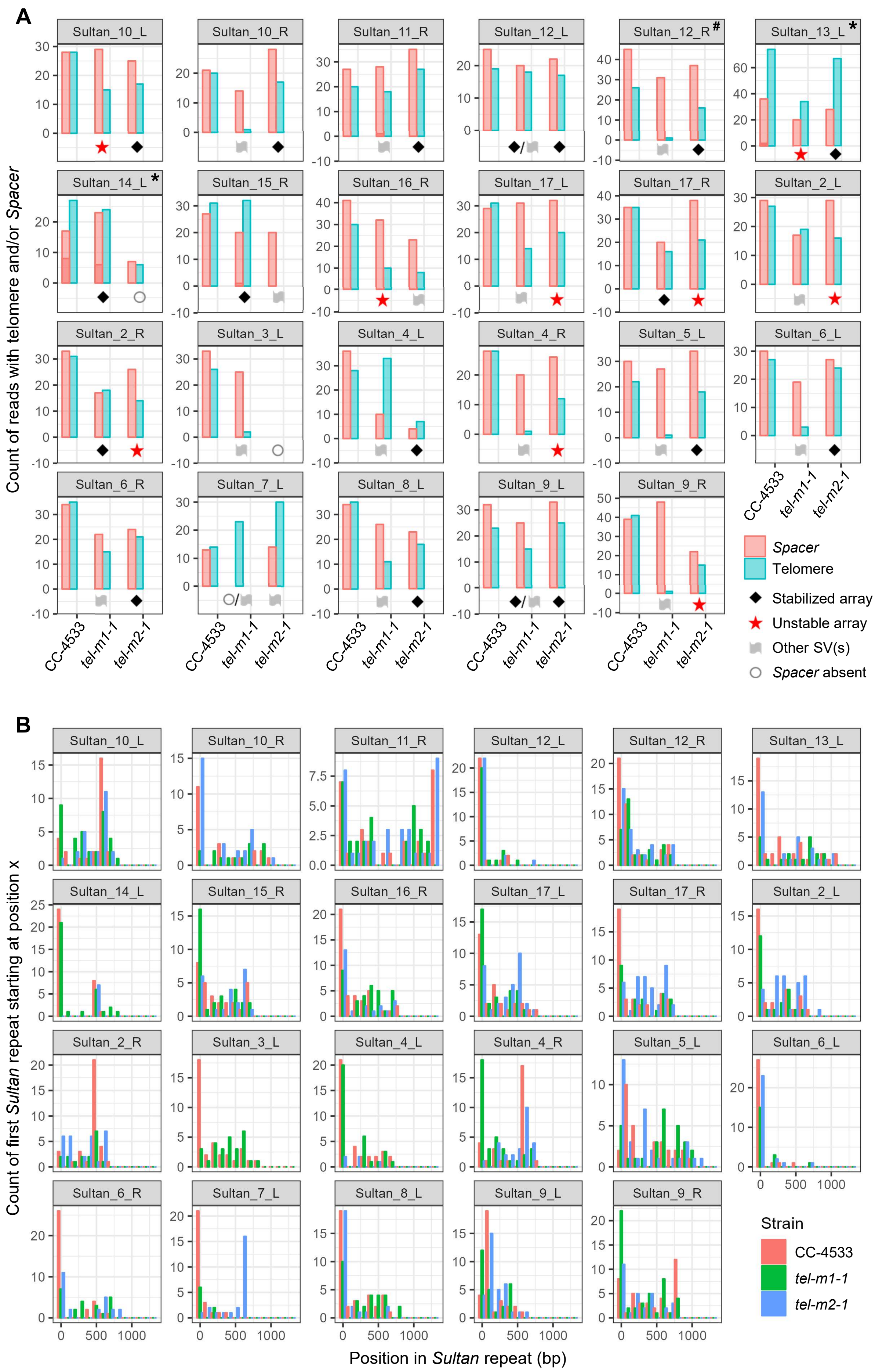
SVs found in ≥2 strains :

	Count	Median (bp)
≥2 strains	558	180

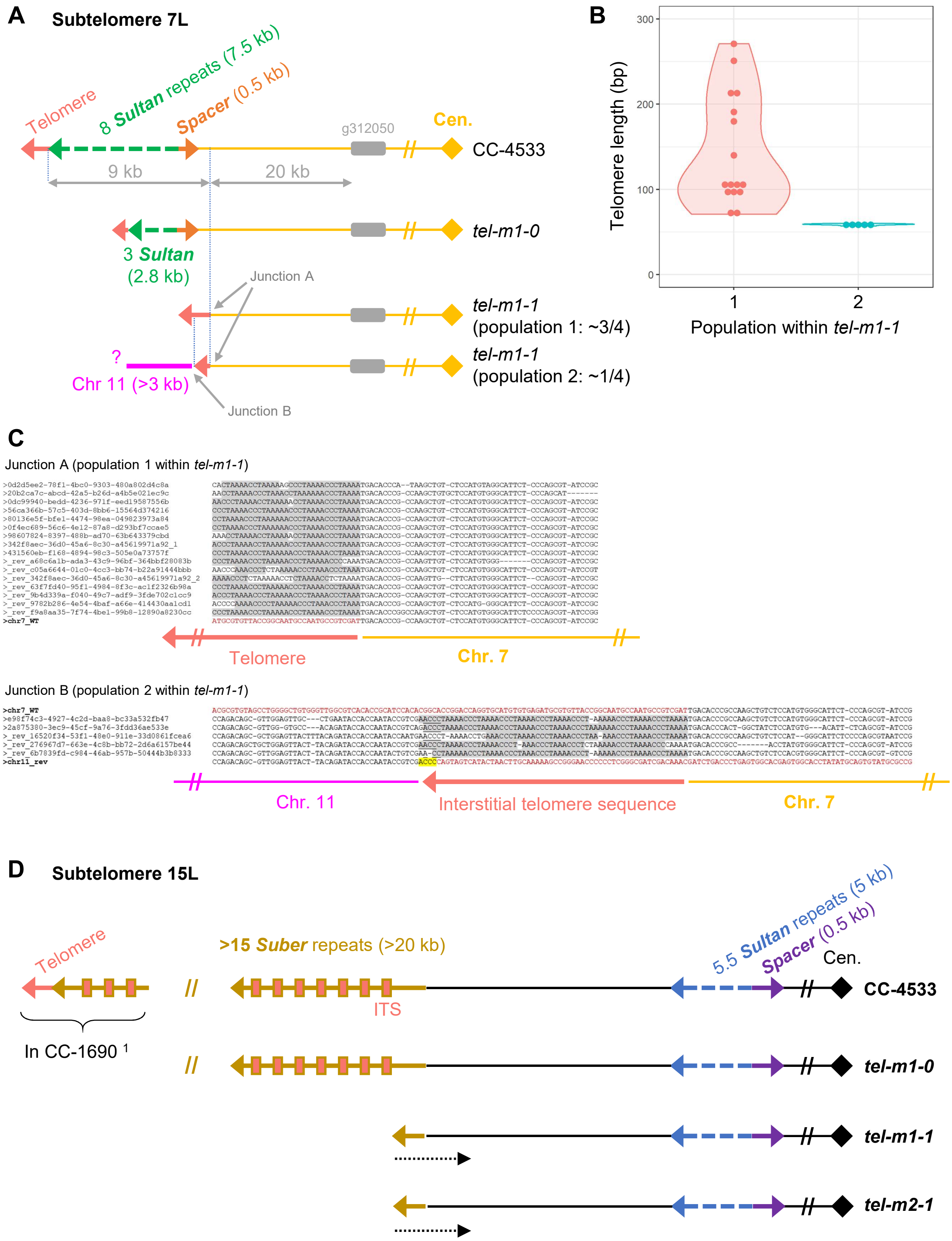
Supplemental Figure S2. Analysis of SVs in the genome assemblies. (A) Two-way Circos plot representation of chromosome-scale SVs between CC-4533 and *tel-m1-1*, as already displayed in Fig. 1C. (B) Same as (A) for CC-4533 and *tel-m2-1*. (C) Count of SVs of the indicated types detected by MUM&Co in the assemblies of CC-4533, *tel-m1-1* and *tel-m2-1*, obtained from 3 assemblers (Canu, NextDenovo and SMARTdenovo), and compared to CC-1690. (D) Venn diagram representation of unique and common SVs detected in the 3 strains, for SVs found in the genomes built by at least 2 assemblers. The tables give the median size of the SVs. (E) Boxplots of the percentage of reads of the lowest mapping quality (Q0) and of the highest (Q60), for reads mapping to subtelomeres and the rest of the genome.


A
Telomere length (bp)
750
500
250
0
1R
2L
2R
3L
4L
4R
5L
6L
6R
14 10 7 11
19 16 19 13
32 18 17 12
25 9 5
25 7
24 1 12
21 4 5 21
25 7 5 23
31 18 22 21
7L
7R
8L
9L
9R
10L
10R
11R
12L
27 9 14 8
22 20 13 43
32 12 41 15
22 15 11 27
39 4 1 12
16 16 16 15
17 15 11 15
15 10 14 23
19 18 17 17
12R
13L
13R
14L
15R
16L
16R
17L
17R
22 15 9 15
34 11 13 22
26 14 18 14
25 14 19
29 15 25 27
22 12 4 10
26 7 7 8
29 23 11 17
34 15 14 18
Strain
CC-4533
tel-m1-0
tel-m1-1
tel-m2-1
B
DNA content normalized over ATPC
8000
6000
4000
2000
0
- Φ29
+ Φ29
Telomere
300
200
100
0
Sultan
CC-4533
tel-m1-0
tel-m1-1
tel-m2-1
Rolling circle amplification
Φ29
900
600
300
0
URA3
pRS306 +
CC-4533 gDNA

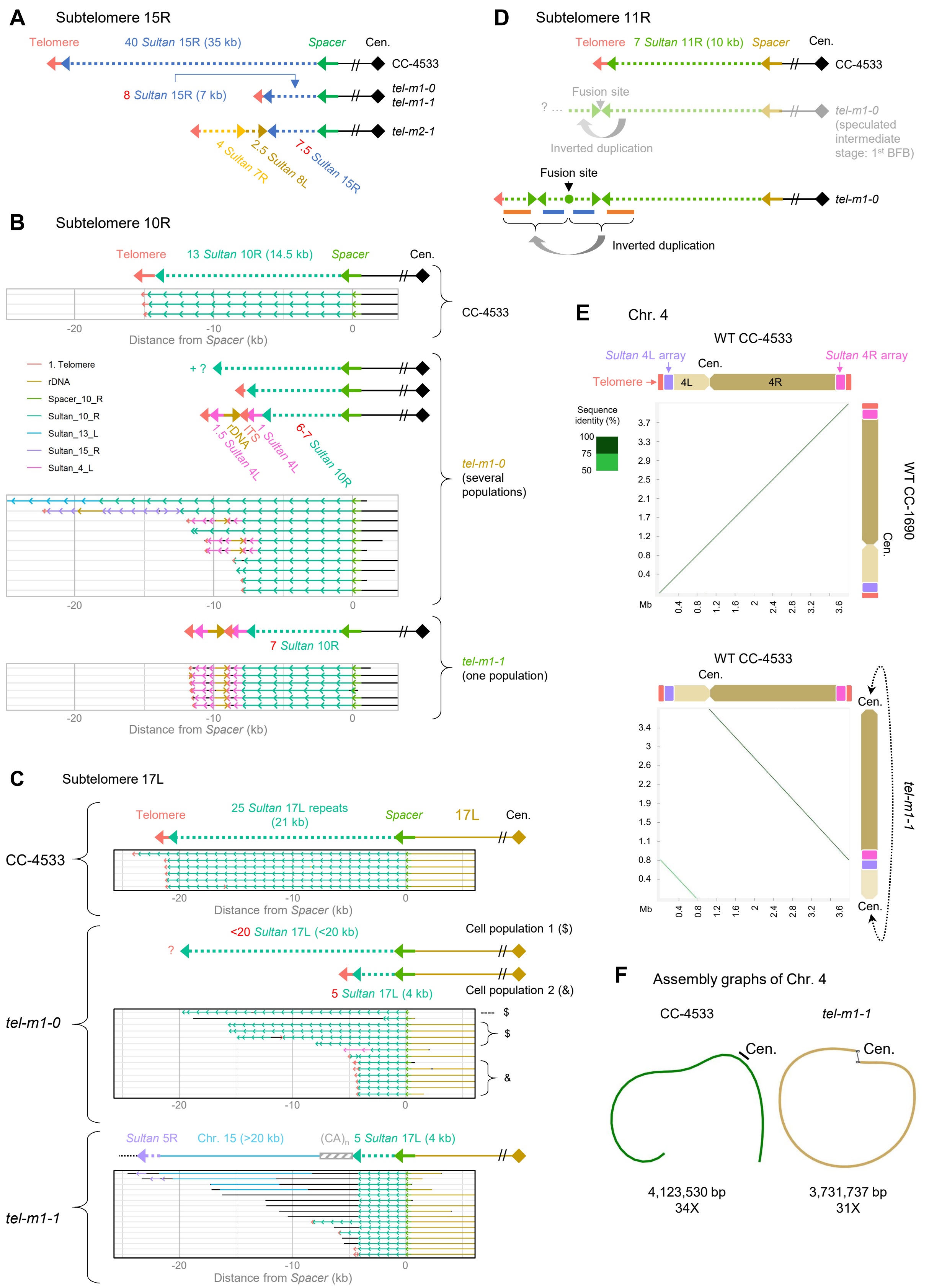
Supplemental Figure S3. Analysis of telomere length at the read level. (A) Telomere length distribution for each *Sultan* subtelomere, *i.e.* each uniquely identifiable class A and class B subtelomere is represented as a boxplot for each indicated strain. The number of corresponding reads is indicated under the boxes. (B) Rolling circle amplification assay performed on CC-4533 and the telomerase mutant samples, based on ϕ29-dependent amplification of partially single-stranded or nicked DNA circles as illustrated in the top right. Telomeric and *Sultan* contents were measured by qPCR, adjusted for genomic DNA quantity by using the single-copy gene *ATPC*. Each bar represents the mean normalized value derived from qPCR measurements (performed in triplicates) of n = 3 independent ϕ29 or no-ϕ29 reactions and the error bar indicates the standard deviation. As a positive control, the plasmid pRS306 containing the *URA3* gene was spiked into CC-4533 genomic DNA and its quantity was measured by qPCR targeting *URA3*.


A
B
Count of reads with telomere and/or Spacer
Sultan_10_L
Sultan_10_R
Sultan_11_R
Sultan_12_L
Sultan_12_R #
Sultan_13_L *
Sultan_14_L *
Sultan_15_R
Sultan_16_R
Sultan_17_L
Sultan_17_R
Sultan_2_L
Sultan_2_R
Sultan_3_L
Sultan_4_L
Sultan_4_R
Sultan_5_L
Sultan_6_L
Sultan_6_R
Sultan_7_L
Sultan_8_L
Sultan_9_L
Sultan_9_R
CC-4533
tel-m1-1
tel-m2-1
Spacer
Telomere
Stabilized array
Unstable array
Other SV(s)
Spacer absent
Count of first Sultan repeat starting at position x
Position in Sultan repeat (bp)
Strain
CC-4533
tel-m1-1
tel-m2-1

Supplemental Figure S4. Stabilized and unstable *Sultan* arrays. (A) For the indicated class A and B *Sultan* arrays, the number of reads containing both a telomere sequence and *Sultan* elements is indicated as a blue bar. The number of reads containing both the *Spacer* element and *Sultan* elements is shown as a red bar. In CC-4533, the near equality between these two numbers for most subtelomeres strongly suggests that each of these subtelomeres, starts with a *Spacer* and ends with a telomere sequence at the individual DNA molecule level. It is mostly the case also for stabilized *Sultan* arrays in the mutants, marked by a black diamond. In contrast, unstable *Sultan* arrays (red star) nearly always show more *Spacer*-*Sultan* reads than telomere-*Sultan* reads, suggesting that the chromosome extremity ends with a *Sultan* element. Subtelomeres with additional SVs (gray flag) cannot be subjected to this analysis. Circles indicate the partial loss or complete absence of *Spacer* sequence. *For 13L and 14L, the *Sultan* element was found associated with 2 distinct *Spacer* elements (light and dark red parts of the histogram); besides, both subtelomeres are from class B, thus their *Sultan* elements contain insertions that might map elsewhere in the genome close to ITS, hence the high number of telomeres associated with the *Sultan*. **Subtelomere 12R contains two identical *Spacer* sequences, thus increasing the count of *Sultan* reads associated with a Spacer (see also Supplemental Dataset 1). (B) Distribution of the position of telomere-*Sultan* transition in the first *Sultan* for stabilized *Sultan* arrays or of the first *Sultan* element in unstable arrays, in CC-4533, *tel-m1-1* and *tel-m2-1*.

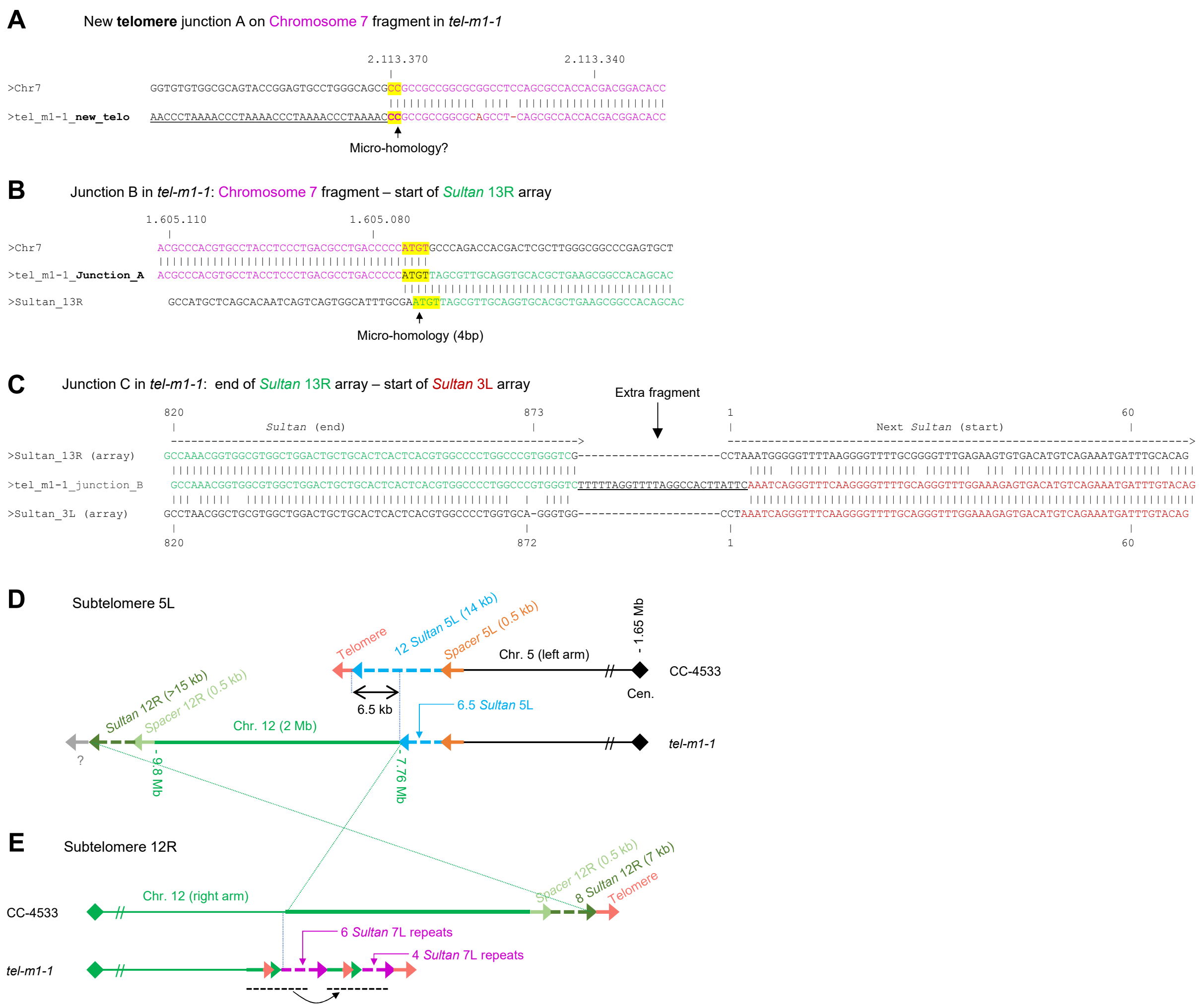


Supplemental Figure S5. Some *Sultan* and *Suber* arrays shorten progressively until complete degradation. (A) Schematic representation of subtelomere 7L in CC-4533, *tel-m1-0* and *tel-m1-1*, showing a decrease in the number of *Sultans* in *tel-m1-0* compared to CC-4533 and a total loss of all *Sultans* and the *Spacer* sequence in *tel-m1-1*, with the formation of a new telomere and even the additional translocation of a sequence from Chromosome 11 in a subpopulation. (B) Telomere length distribution of the terminal telomere in subpopulation 1 and of the ITS in subpopulation 2, as defined in (A). (C) Chromosome 7 to telomere junction in subpopulation 1 of reads and Chromosome 7 to ITS to Chromosome 11 junctions in subpopulation 2 of reads, as defined in (A). The last junction involves a microhomology of 4 bp (highlighted in yellow). (D) Schematic representation of class C subtelomere 15L displaying a structure containing *Suber* elements in *tel-m1-0* similar to CC-4533, which presumably ends with telomeres as shown for CC-1690 (Chaux-Jukic et al. 2021), but with only one partial *Suber* remaining in *tel-m1-1* and *tel-m2-1*, suggesting a progressive loss of *Suber* elements.

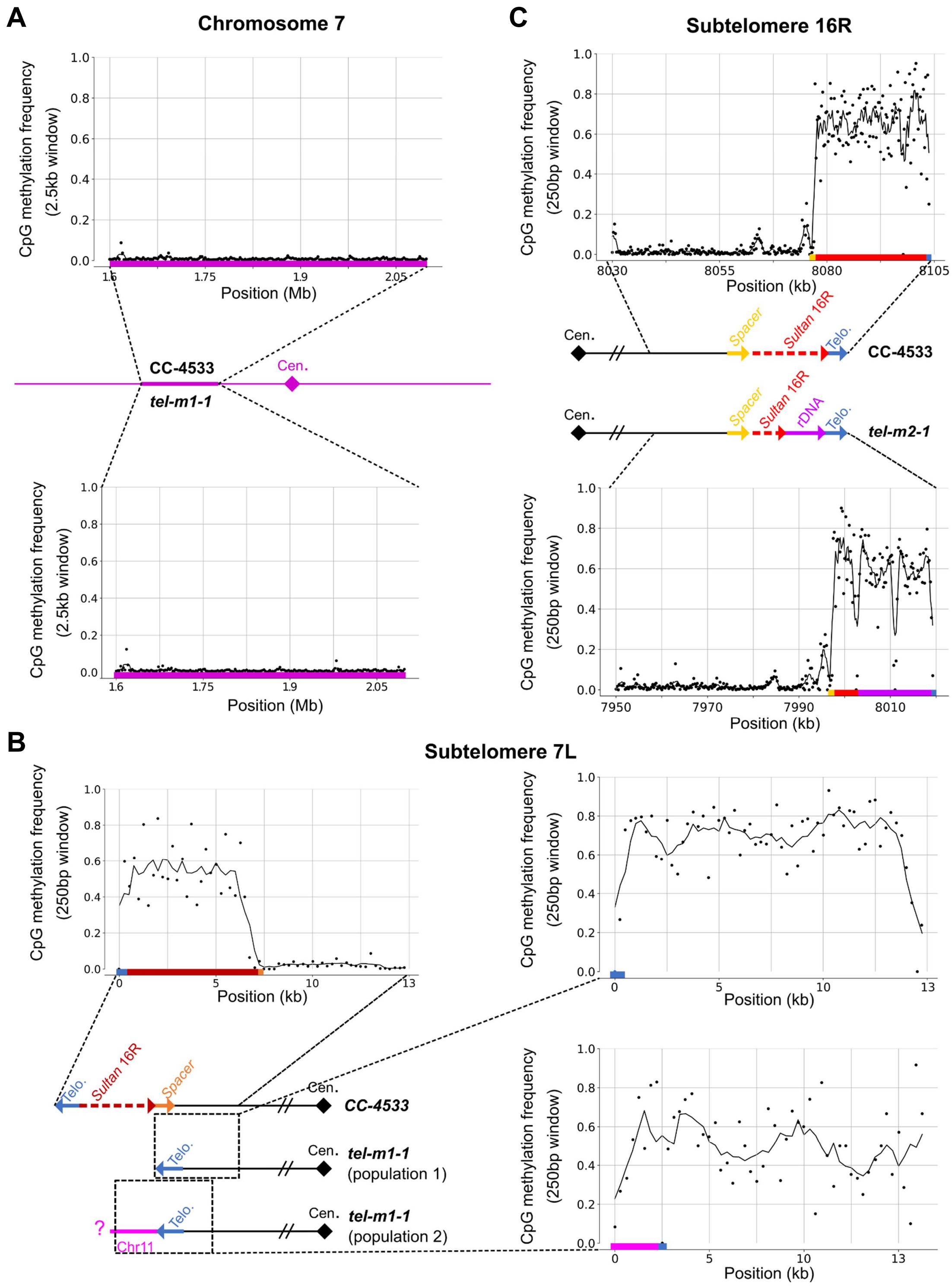


Supplemental Figure S6. Complex multistep rearrangements at chromosome extremities. (A) Schematic representation of subtelomere 15R in CC-4533, *tel-m1-0*, *tel-m1-1* and *tel-m2-1*. (B)

Schematic representation of subtelomere 10R and reads supporting the depicted structures. Several subpopulations are present in *tel-m1-0* and after purifying selection, *tel-m1-1* only comprises one subpopulation. (C) Schematic representation of the dynamics of rearrangements at subtelomere 17L between CC-4533, *tel-m1-0* and *tel-m1-1*. The structure of the subtelomere in *tel-m1-1* appeared to have evolved from the structure in cell population 2 of *tel-m1-0*. (D) Schematic representation of subtelomere 11R in *tel-m1-0* compared to CC-4533. Signature of BFB with multiple inverted repeats suggests at least 2 cycles of BFB. The speculated intermediate stage is shown in shaded colors. (E) Dot plot representation of the alignment of Chromosome 4 between CC-1690 and CC-4533 and CC-1690 and *tel-m1-1*. The extremities of Chromosome 4 of CC-4533 are no longer present in *tel-m1-1*. (F) Graph representation of Chromosome 4 of CC-4533 (linear) and *tel-m1-1* (circular topology).



Supplemental Figure S7. New telomere-capped extremities and junction sequences. (A) Sequence of junction A as marked in Fig. 5B, showing a 2-bp microhomology highlighted in yellow. (B) Sequence of junction B as marked in Fig. 5B, with a 4-bp microhomology. (C) Sequence of junction C as marked in Fig. 5B, with the insertion of a short extra fragment between *Sultan* 13R array and *Sultan* 3L array. (D-E) Schematic representation of subtelomere 5L and 12R in CC-4533 and *tel-m1-1*. Subtelomere 12R was translocated to subtelomere 5L on a short array of *Sultan* elements. The new 12R extremity was formed by a tandem duplication of 9 telomeric repeats, a fragment of Chromosome 12 and *Sultan* elements from subtelomere 7, followed by a terminal telomere sequence.



Supplemental Figure S8. Methylation frequency of other chromosome regions and extremities. (A-C) Analysis of 5mC frequency at CpG sites using reads that unambiguously spanned the indicated regions, in different illustrative cases. (A) At the original region of Chromosome 7 that was duplicated in *tel-m1-1*, shown here for CC-4533 and *tel-m1-1*. (B) At the new 7L extremity of *tel-m1-1* where methylation spreads further towards the interior of the genome in both subpopulations compared to CC-4533 where the hypermethylated domain stops at the *Spacer* sequence. (C) At subtelomere 16R where in *tel-m2-1*, the structure shows a shorter hypermethylated *Sultan* array and a hypermethylated rDNA sequence capped by a telomere, compared to CC-4533 where only the hypermethylated *Sultan* array is present.

Supplemental Dataset 1. *Spacer*-based analysis of subtelomeres. All *Spacer*-containing reads anchored at the *Spacer* element are depicted for each subtelomere for CC-4533, *tel-m1-0*, *tel-m1-1* and *tel-m2-1*. The queried elements are displayed in color. The identifier for each read ("Read_id") is also shown, to facilitate retrieval from the sequencing data.

Supplemental Code. Code and associated files for TeloReader. TeloReader is also available at Github: https://github.com/Telomere-Genome-Stability/Telomere_2023/tree/main/TELOREADER