Supplementary Information

Supplementary Discussions

Relative abundance estimates in mock community

To test whether our method was able to estimate relative abundances of taxa accurately, we designed our mock community such that each genome was represented by an equal number of 16S-ITS-23S loci (Table S2). The expectation was that all taxa would represented with a similar amount of qtrim ccs reads. However, this was not the case: we could not detect 9 out of 38 taxa (or 50 out of 123 loci), and among detected taxa there was no equal representation (Fig. S9). This is not counting the 4 taxa which were found to contaminate the mock community (see Experimental Procedures). The contamination was only slight: only 5 qtrim ccs reads mapped to 1 locus each of *M. wisconsensis* and *V. parvula*. Primer bias, a phenomenon in which the degree of primer-template matching can lead to preferential amplification of some taxa while possibly ignoring others, does not seem to explain our observation. All 123 loci of both detected and undetected taxa matched the primers perfectly. The exceptions were the loci of *N. aromaticivorans* and *P. saltans* (not detected) and *T. roseus* (detected), which exhibit a single A to G mismatch in the reverse primer. Locus length and locus GC-content do not seem to explain the observation either. We could not find a correlation between either and observed qtrim ccs read counts (Fig. S10).

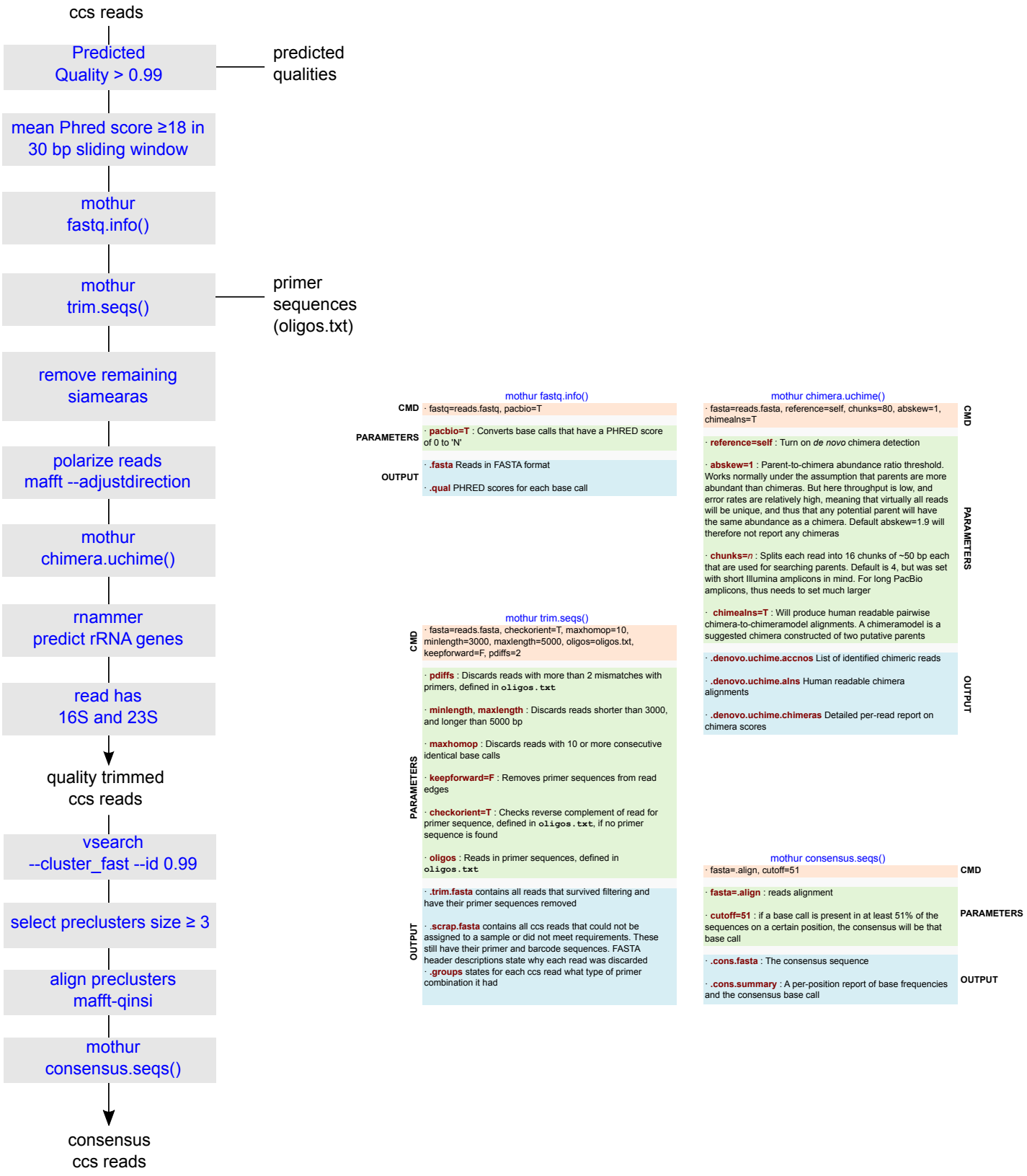


Fig. S1 | Read curation pipeline and mothur function settings. See 'Read curation pipeline' in Methods for a detailed description of the pipeline. See mothur.org/wiki/Sequence_processing for more detailed descriptions and explanations of the mothur functions.

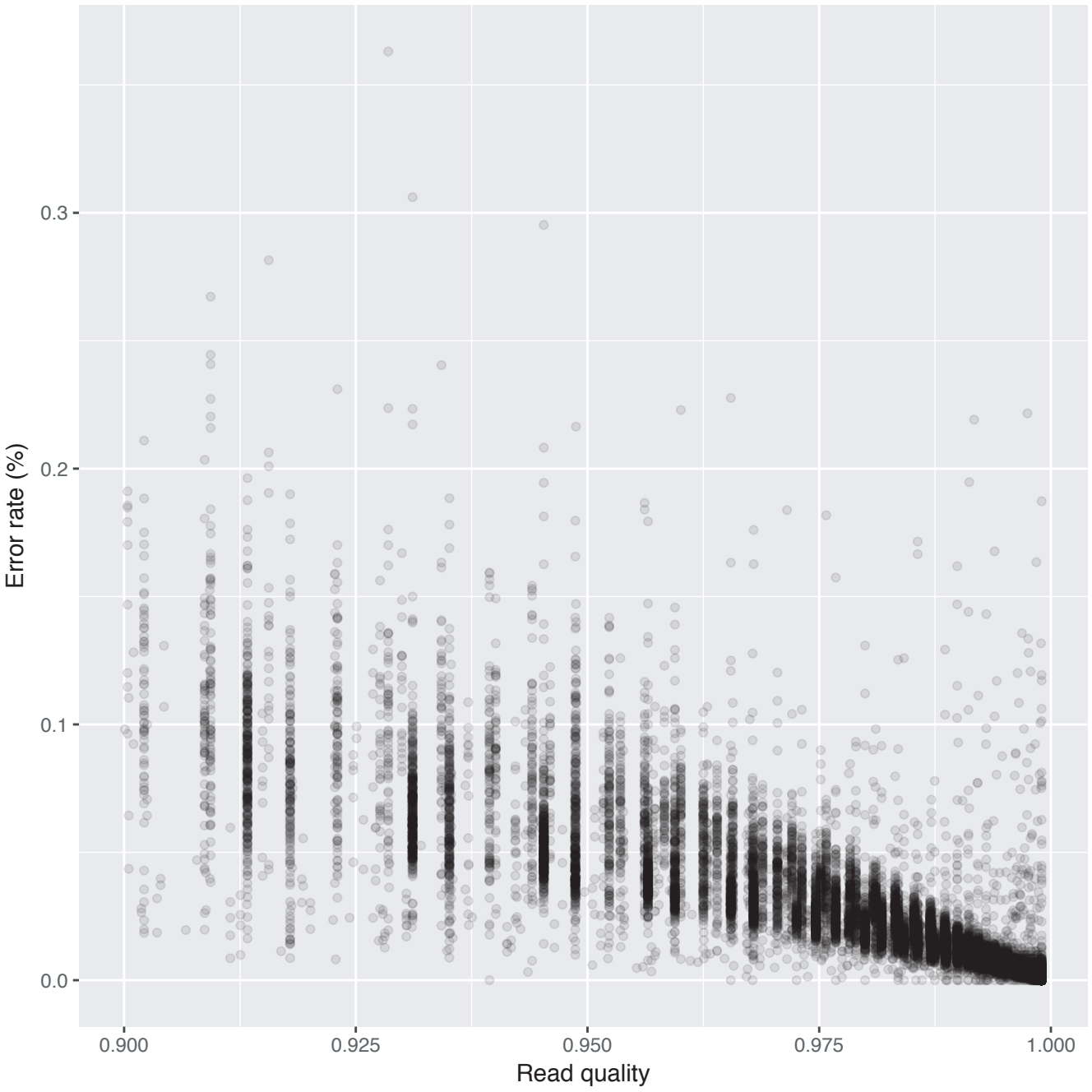


Fig. S2 | Relationship between ccs read quality and observed error rate

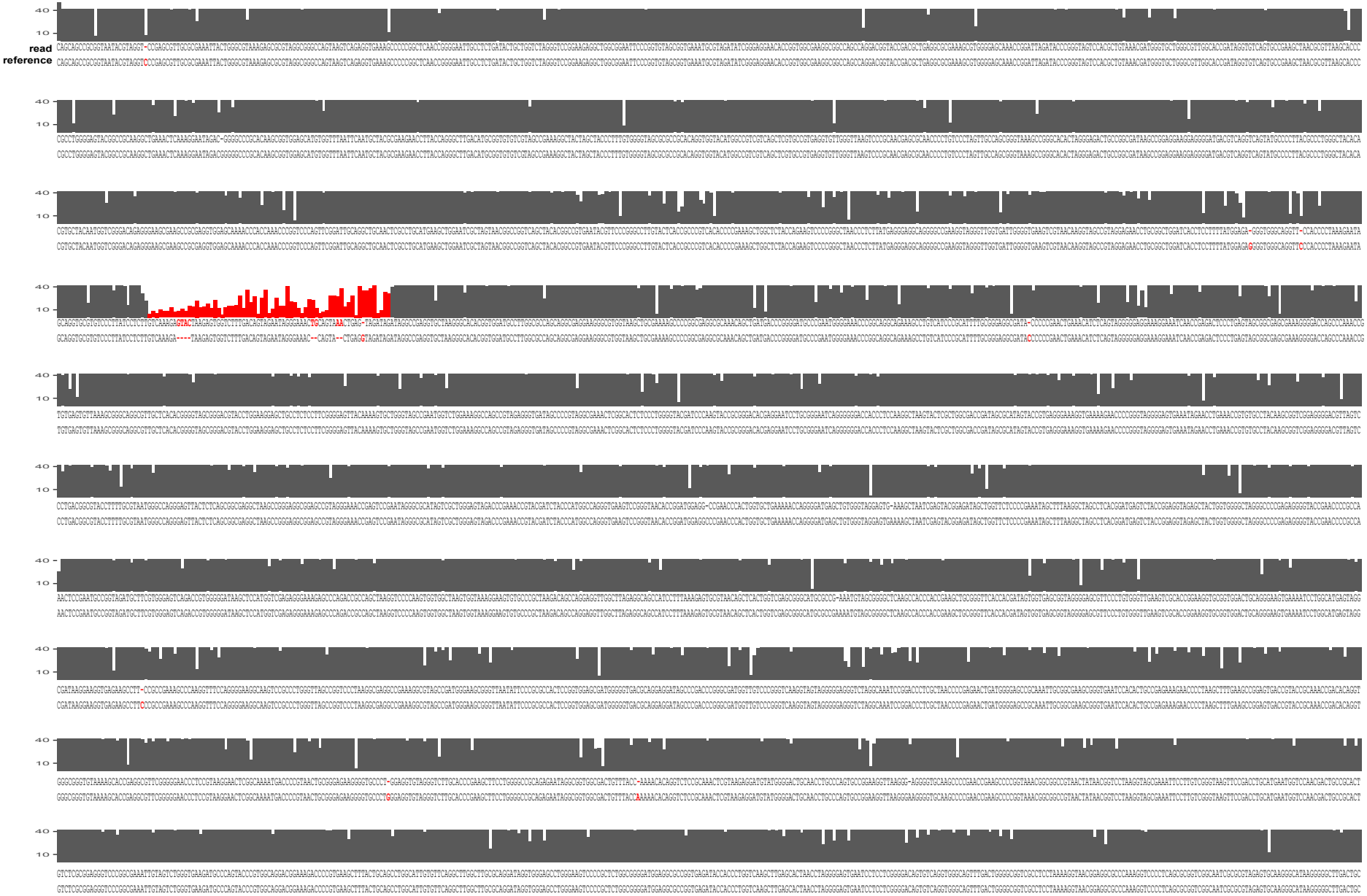
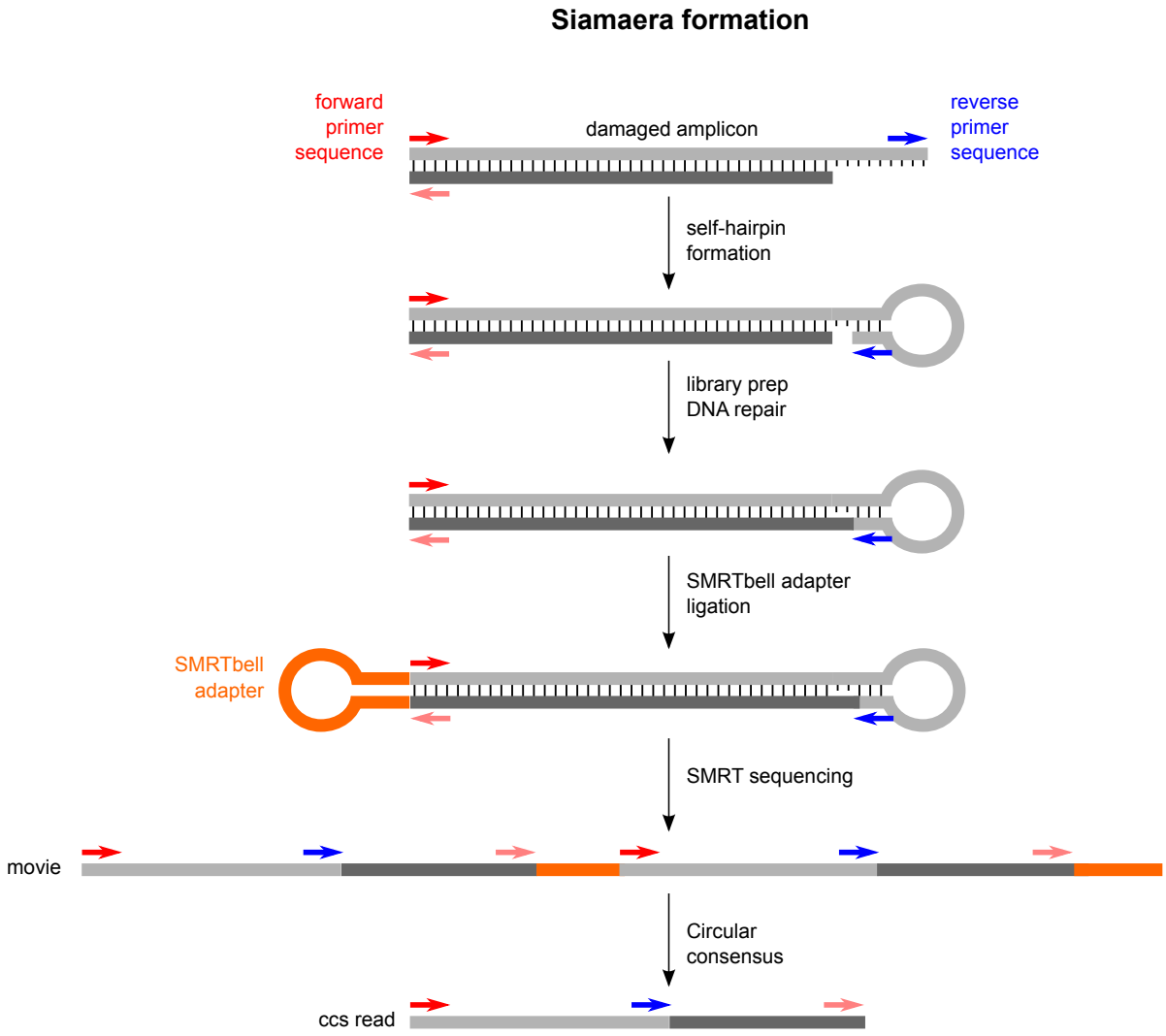
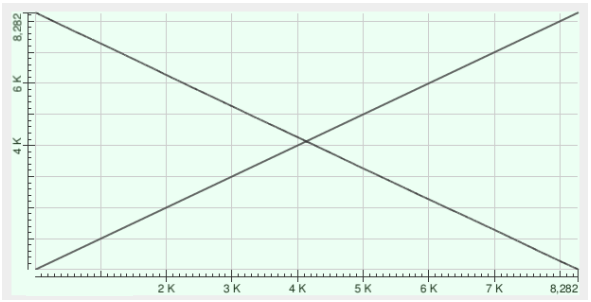


Fig. S3 | Example of a ccs read with a local stretch of low quality base calls that are enriched in sequencing errors. Bars represent Phred scores, red base calls correspond to errors. Red bars indicate local stretch of low quality base calls



Example of a full length siamaeric read



Example of a partial siameric read

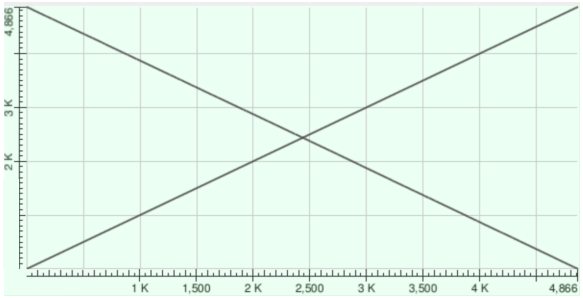


Fig. S4 | Hypothetical mechanism leading to a siamaeric read. In the bottom two examples of siamaeric reads, recognized by BLASTN dot plots

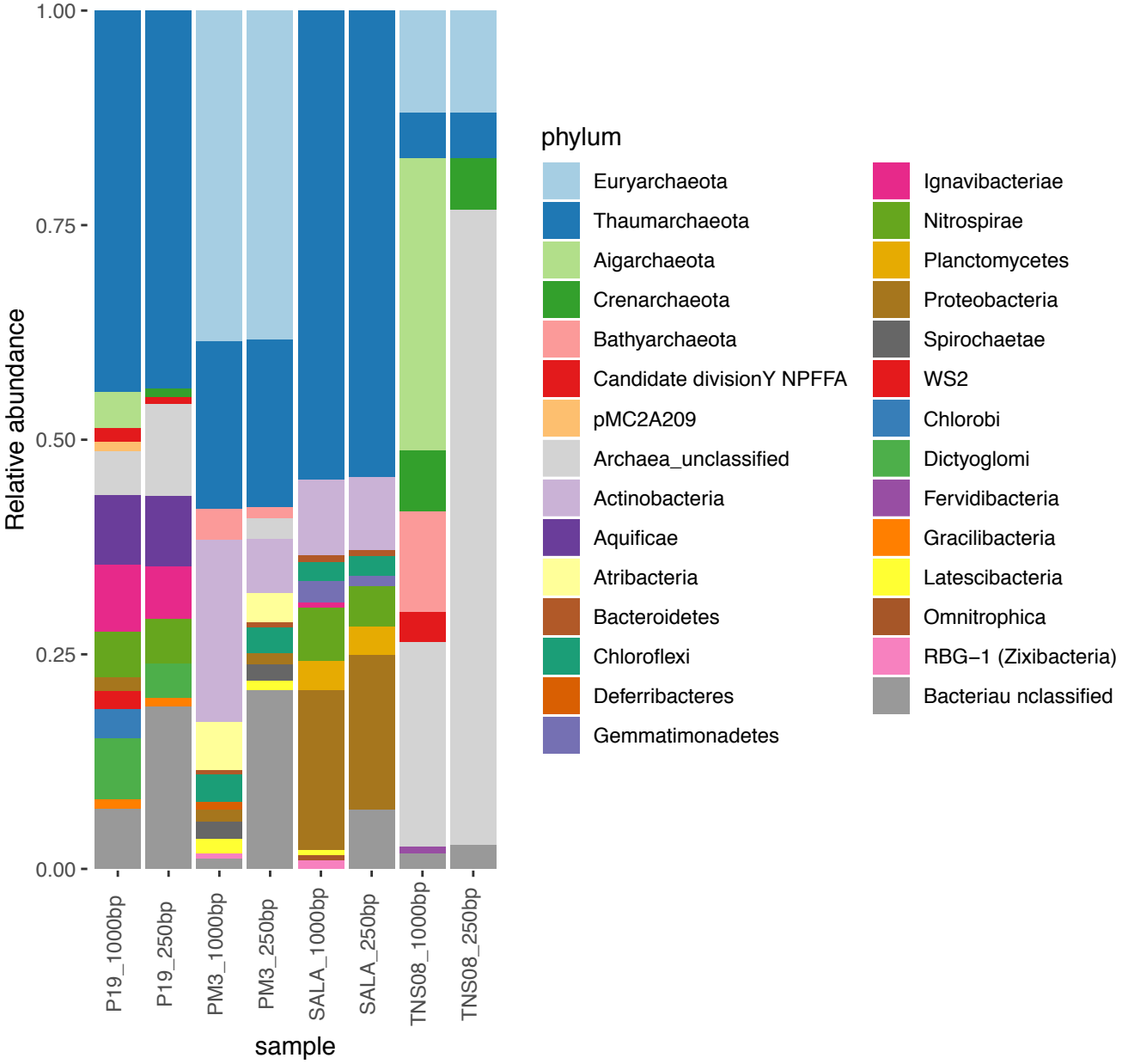


Fig. S5 | Bar charts reflecting estimated relative abundances of bacterial and archaeal phyla with ≥0.5% abundance. *_1000bp: relative abundances estimated from ~1000 bp 16S rRNA gene sequence. *_250bp: relative abundances estimated from 250 bp 16S rRNA gene sequence fragments spanning the V4 region.

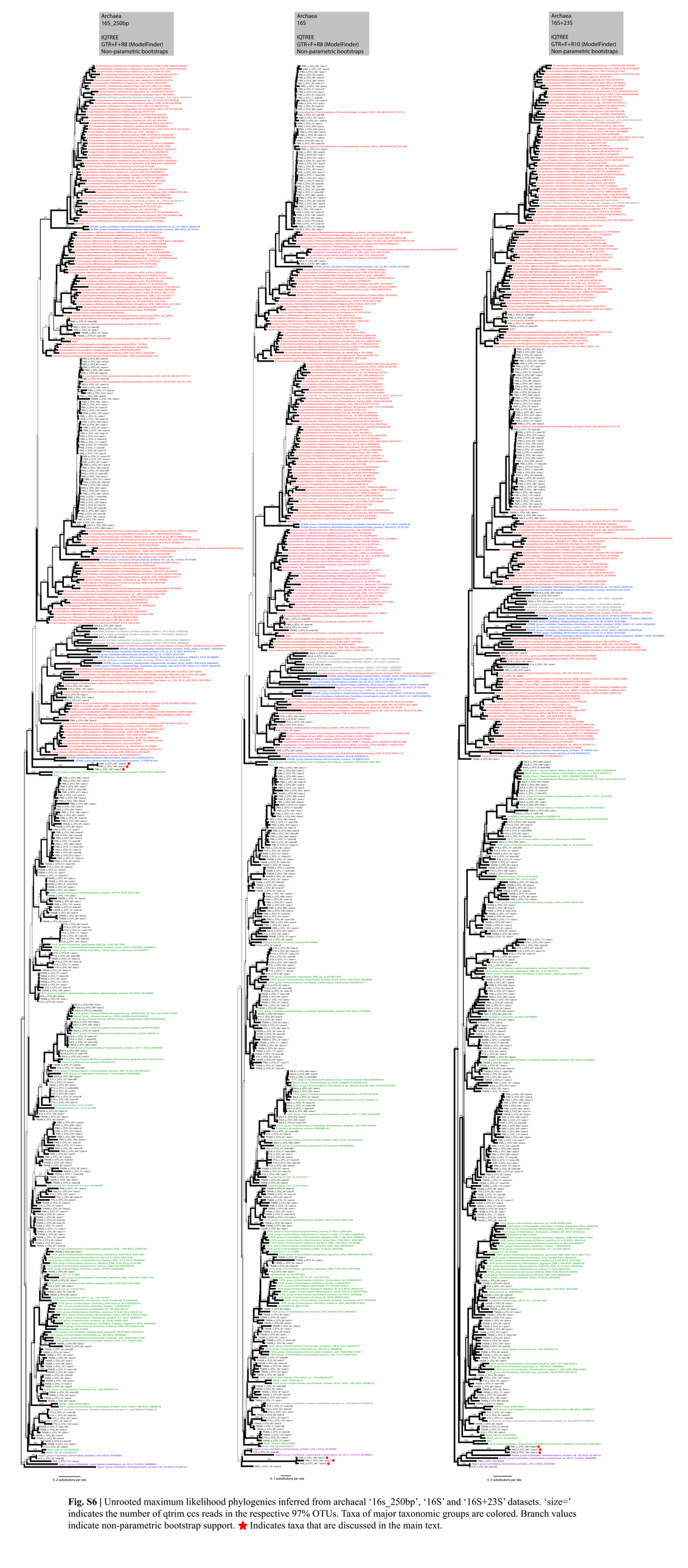


Fig. S6 | Unrooted maximum likelihood phylogenies inferred from archaeal '16s_250bp', '16S' and '16S+23S' datasets. 'size=' indicates the number of qtrim ccs reads in the respective 97% OTUs. Taxa of major taxonomic groups are colored. Branch values indicate non-parametric bootstrap support. ★ Indicates taxa that are discussed in the main text.

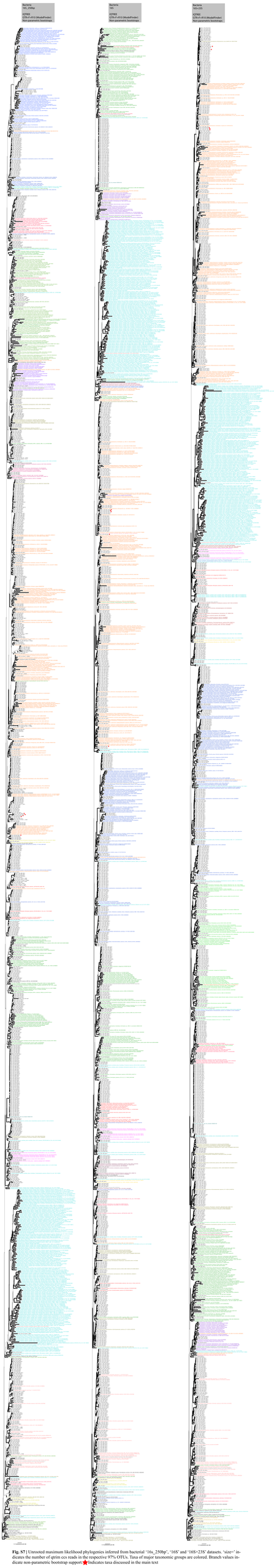


Fig. S7 | Unrooted maximum likelihood phylogenies inferred from bacterial '16s_250bp', '16S' and '16S+23S' datasets. 'size=' indicates the number of qtrim ccs reads in the respective 97% OTUs. Taxa of major taxonomic groups are colored. Branch values indicate non-parametric bootstrap support ★indicates taxa discussed in the main text

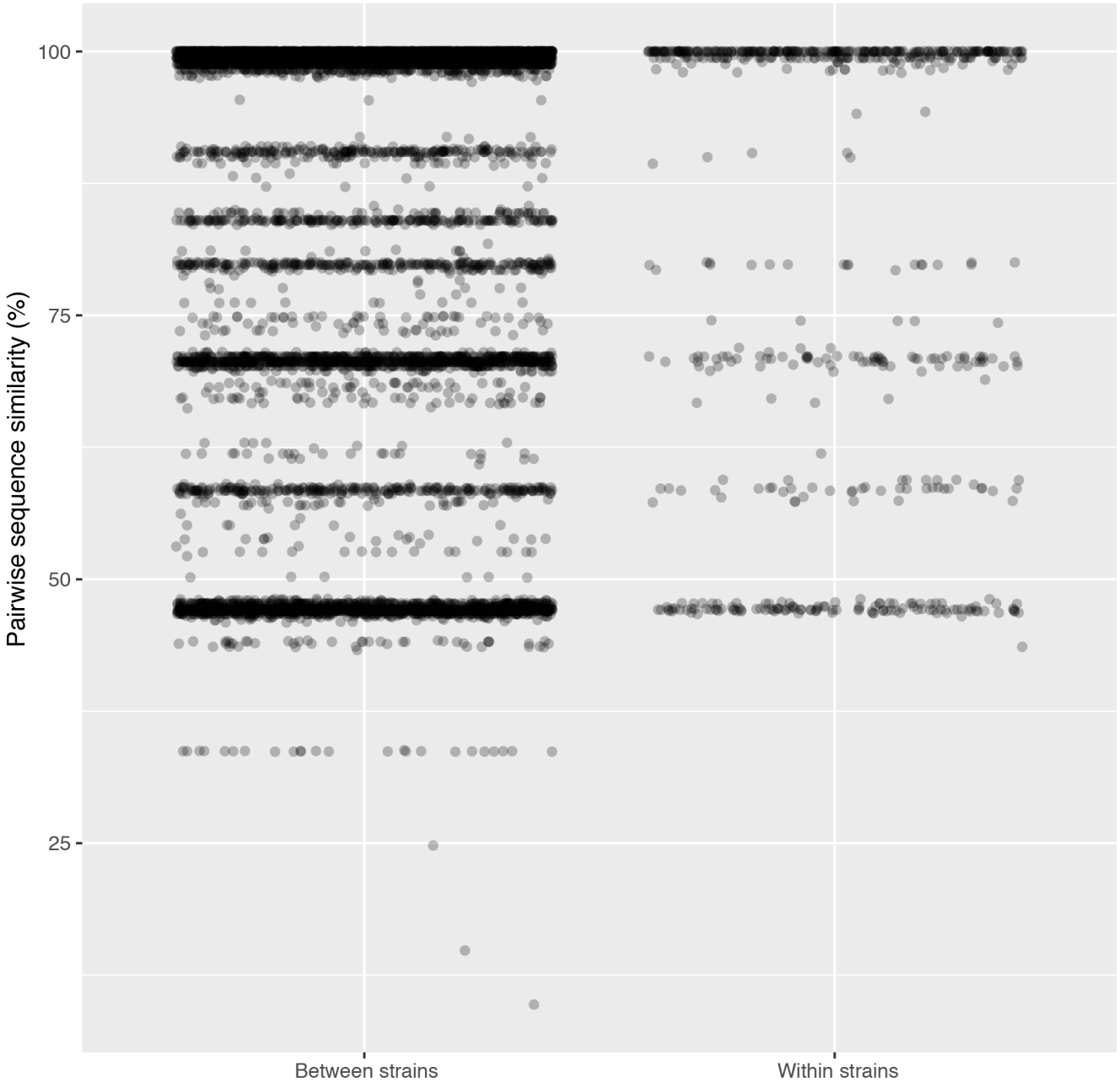


Fig. S8 | Sequence similarities of internal transcribed spacer (ITS) copies either between different strains of the same species (left) or between different ITS copies within the same strain (right). Bacterial and archaeal species used are those that are present in the mock community and their close relatives with complete genomes available.

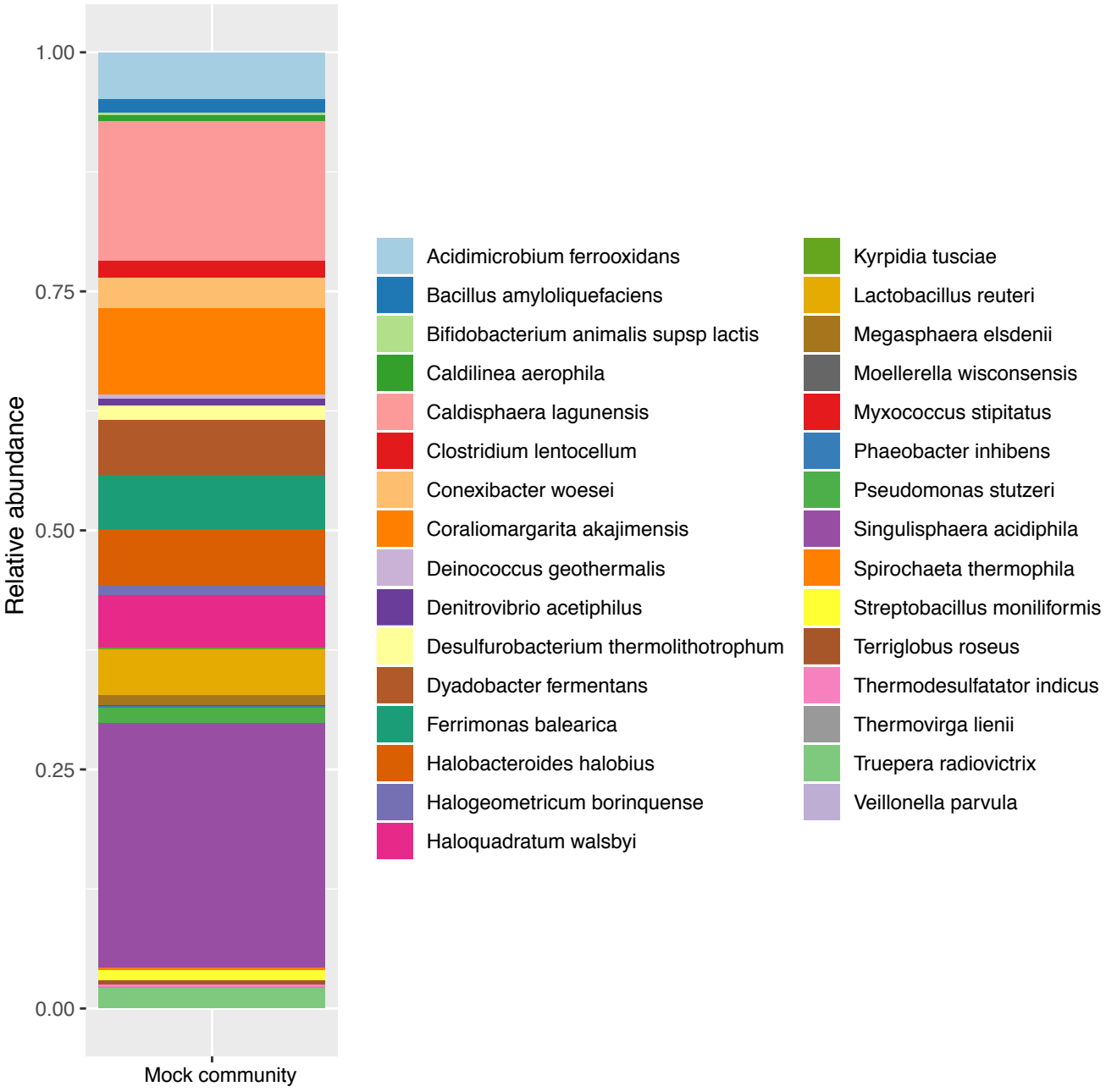


Fig. S9 | Bar chart reflecting the estimated relative abundances of genomes that are part of the mock community.

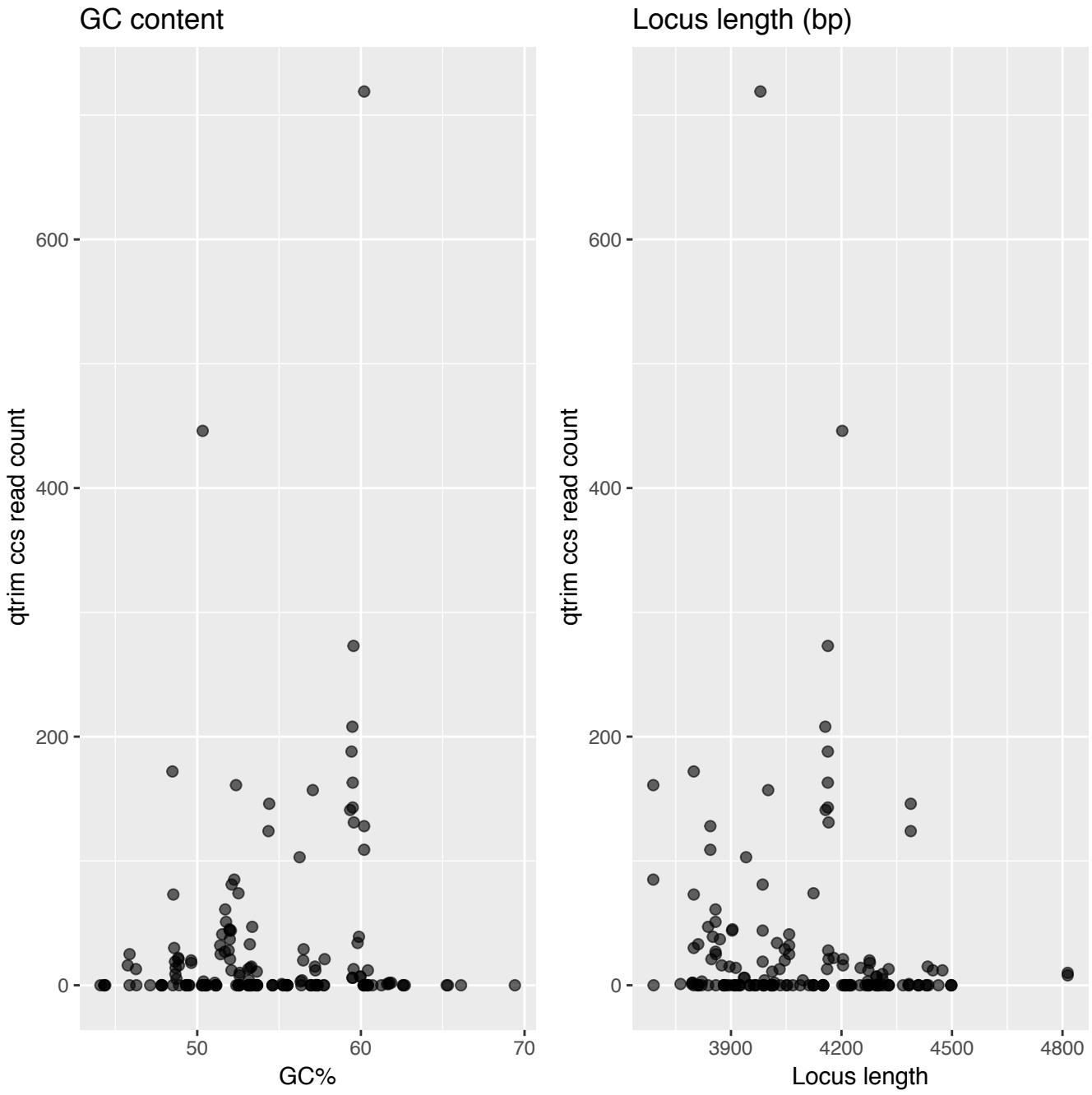


Fig. S10 | Relationships between the %GC (left) or length (right) of the 16S-ITS-23S loci of the mock community genomes and their quality-trimmed ccs read counts.

Table S1. 16S-ITS-23S loci in Archaea and Bacteria

# 16S-ITS-23S loci	Archaea	Bacteria
0	96	4 106
1	176	4 692
2	50	1 135
3	36	719
4	5	546
5	1	321
>5	0	1 076
Total	364	12 595
Fraction ≥1	0,74	0,67

Table S2. Composition of the mock community

Taxon	DSM no.	16S-ITS-23S loci	genome length (bp)	molecular weight genome (Da or ng/nmol)	measured concentration (ng/µl)	measured concentration (amol/µl)	"operonmol"/µl	nl/"operonmol"	Volume %	µl to pipette
Methanocaldococcus jannaschii	2661	2	1664970	1082230500	23,4	21,6	43,2	0,023	1,1	1,6
Archaeoglobus fulgidus	4304	1	2178400	1415960000	24,6	17,4	17,4	0,058	2,7	4,0
Pyrococcus furiosus	3638	1	1908256	1240366400	8,1	6,5	6,5	0,154	7,1	10,6
Haloquadratum walsbyi	16790	2	3132494	2036121100	22,1	10,9	21,7	0,046	2,1	3,2
Hyperthermus butylicus	5456	1	1667163	1083655950	28,0	25,8	25,8	0,039	1,8	2,7
Halogeometricum borinquense	11551	2	2820544	1833353600	24,2	13,2	26,4	0,038	1,7	2,6
Caldisphaera lagunensis	15908	1	1546846	1005449900	14,2	14,1	14,1	0,071	3,3	4,9
Novosphingobium aromaticivorans	12444	3	3561584	2315029600	19,4	8,4	25,1	0,040	1,8	2,8
Deinococcus geothermalis	11300	1	574127	373182550	5,5	14,6	14,6	0,068	3,2	4,7
Lactobacillus reuteri	20016	6	1999618	1299751700	26,3	20,2	121,2	0,008	0,4	0,6
Dictyoglomus turgidum	6724	2	1855560	1206114000	18,6	15,5	30,9	0,032	1,5	2,2
Bifidobacterium animalis subsp. lactis	10140	4	1938483	1260013950	10,0	7,9	31,6	0,032	1,5	2,2
Dyadobacter fermentans	18053	4	6967790	4529063500	25,6	5,6	22,6	0,044	2,0	3,1
Acidimicrobium ferrooxidans	10331	2	2158157	1402802050	21,9	15,6	31,2	0,032	1,5	2,2
Streptobacillus moniliformis	12112	5	1662578	1080675700	21,1	19,6	97,8	0,010	0,5	0,7
Conexibacter woesei	14684	1	6359369	4133589850	23,0	5,6	5,6	0,180	8,3	12,4
Denitrovibrio acetiphilus	12809	2	3222077	2094350050	21,1	10,1	20,1	0,050	2,3	3,4
Coraliomargarita akajimensis	45221	2	3750771	2438001150	21,6	8,9	17,7	0,056	2,6	3,9
Kyrpidia tusciae	2912	5	3384766	2200097900	20,1	9,2	45,8	0,022	1,0	1,5
Truepera radiovictrix	17093	1	3260398	2119258700	21,2	10,0	10,0	0,100	4,6	6,9
Ferrimonas balearica	9799	7	4279159	2781453350	25,6	9,2	64,3	0,016	0,7	1,1
Bacillus amyloliquefaciens	7	10	3980199	2587129350	11,5	4,4	44,4	0,023	1,0	1,6
Pedobacter saltans	12145	4	4635236	3012903400	6,7	2,2	8,9	0,113	5,2	7,8
Desulfurobacterium thermolithotrophum	11699	2	1541968	1002279200	20,6	20,5	41,1	0,024	1,1	1,7
Clostridium lentocellum	5427	8	4714237	3064254050	21,1	6,9	55,0	0,018	0,8	1,3
Thermodesulfatator indicus	15286	2	2322224	1509445600	21,7	14,4	28,7	0,035	1,6	2,4
Thermotoga thermarum	5069	1	2039943	1325962950	9,4	7,1	7,1	0,141	6,5	9,8
Megasphaera elsdenii	20460	7	2474718	1608566700	24,4	15,2	106,1	0,009	0,4	0,7
Thermovirga lienii	17291	3	1967774	1279053100	22,0	17,2	51,7	0,019	0,9	1,3
Caldilinea aerophila	14535	2	5144873	3344167450	22,1	6,6	13,2	0,076	3,5	5,2
Pseudomonas stutzeri	4166	4	4689946	3048464900	13,5	4,4	17,7	0,056	2,6	3,9
Spirochaeta thermophila	6578	2	2560222	1664144300	12,3	7,4	14,8	0,068	3,1	4,7
Terriglobus roseus	18391	2	5227858	3398107700	23,4	6,9	13,8	0,073	3,3	5,0
Phaeobacter inhibens	17395	4	3821831	2484190150	18,9	7,6	30,5	0,033	1,5	2,3
Singulisphaera acidiphila	18658	8	9629675	6259288750	22,8	3,6	29,1	0,034	1,6	2,4
Halobacteroides halobius	5150	5	2649255	1722015750	22,9	13,3	66,5	0,015	0,7	1,0
Myxococcus stipitatus	14675	3	10350586	6727880900	18,4	2,7	8,2	0,122	5,6	8,4
Desulfovibrio gigas	1382	1	3693999	2401099350	12,5	5,2	5,2	0,192	8,8	13,3
								Sum	**Sum**	**Sum**
								2,2	100,0	150,0