

Complete Genome Structure of the Thermophilic Cyanobacterium *Thermosynechococcus elongatus* BP-1 (Supplement)

Yasukazu NAKAMURA,¹ Takakazu KANEKO,¹ Shusei SATO,¹ Masahiko IKEUCHI,² Hiroshi KATOH,²
Shigemi SASAMOTO,¹ Akiko WATANABE,¹ Mayumi IRIGUCHI,¹ Kumiko KAWASHIMA,¹ Takaharu KIMURA,¹
Yoshie KISHIDA,¹ Chiaki KIYOKAWA,¹ Mitsuyo KOHARA,¹ Midori MATSUMOTO,¹ Ai MATSUNO,¹
Naomi NAKAZAKI,¹ Sayaka SHIMPO,¹ Masako SUGIMOTO,¹ Chie TAKEUCHI,¹ Manabu YAMADA,¹
and Satoshi TABATA^{1,*}

*Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan¹ and
Department of Life Sciences (Biology), The University of Tokyo, Komaba 3-8-1, Meguro, Tokyo 153-8902,
Japan²*

(Received 21 May 2002)

Communicated by Mituru Takanami

* To whom correspondence should be addressed. Tel. +81-438-52-3933, Fax. +81-438-52-3934, E-mail: tabata@kazusa.or.jp

Table 1. Codon usage frequency for the *T. elongatus* genome.

UUU F	25.8 (20037)	UCU S	5.8 (4522)	UAU Y	15.2 (11780)	UGU C	5.9 (4556)
UUC F	9.4 (7290)	UCC S	10.3 (8044)	UAC Y	13.3 (10354)	UGC C	5.5 (4248)
UUA L	9.5 (7351)	UCA S	4.7 (3624)	UAA *	1.0 (809)	UGA *	0.7 (570)
UUG L	24.6 (19150)	UCG S	5.6 (4386)	UAG *	1.4 (1097)	UGG W	16.9 (13172)
CUU L	14.7 (11411)	CCU P	10.4 (8072)	CAU H	10.2 (7901)	CGU R	10.5 (8176)
CUC L	26.9 (20894)	CCC P	27.9 (21718)	CAC H	11.9 (9221)	CGC R	29.3 (22752)
CUA L	11.3 (8786)	CCA P	8.9 (6901)	CAA Q	31.4 (24434)	CGA R	6.1 (4712)
CUG L	32.7 (25425)	CCG P	10.4 (8054)	CAG Q	26.2 (20354)	CGG R	15.7 (12176)
AUU I	37.0 (28779)	ACU T	7.6 (5944)	AAU N	16.1 (12547)	AGU S	13.4 (10407)
AUC I	16.6 (12908)	ACC T	25.3 (19688)	AAC N	12.9 (10035)	AGC S	11.7 (9126)
AUA I	1.7 (1300)	ACA T	9.1 (7045)	AAA K	16.4 (12760)	AGA R	1.8 (1396)
AUG M	18.3 (14254)	ACG T	12.8 (9923)	AAG K	14.6 (11366)	AGG R	2.7 (2129)
GUU V	14.0 (10855)	GCU A	18.0 (13958)	GAU D	30.9 (24042)	GGU G	16.8 (13059)
GUC V	17.8 (13849)	GCC A	43.8 (34037)	GAC D	13.5 (10498)	GGC G	26.6 (20665)
GUA V	6.0 (4676)	GCA A	16.9 (13104)	GAA E	31.2 (24270)	GGA G	8.8 (6849)
GUG V	32.8 (25506)	GCG A	18.6 (14459)	GAG E	27.1 (21084)	GGG G	19.0 (14792)

The frequency of the usage of each codon in the chromosome was shown in permillage of overall counts. Numerals in parentheses indicate the actual number of the codons.

Table 2. Codon-anticodon recognition pattern for the *T. elongatus* genome.

UUU F	<i>trnF-GAA</i>	UCU S	<i>trnS-GGA</i>	UAU Y	<i>trnY-GUA</i>	UGU C	<i>trnC-GCA</i>
UUC F		UCC S		UAC Y		UGC C	
UUA L	<i>trnL-UAA</i>	UCA S	<i>trnS-UGA</i>	UAA -		UGA -	
UUG L	<i>trnL-CAA</i>	UCG S	<i>trnS-CGA</i>	UAG -		UGG W	<i>trnW-CCA</i>
CUU L	<i>trnL-GAG</i>	CCU P	<i>trnP-GGG</i>	CAU H	<i>trnH-GUG</i>	CGU R	
CUC L		CCC P		CAC H		CGC R	<i>trnR-ACG</i>
CUA L	<i>trnL-UAG</i>	CCA P	<i>trnP-UGG</i>	CAA Q	<i>trnQ-UUG</i>	CGA R	
CUG L	<i>trnL-CAG</i>	CCG P	<i>trnP-CGG</i>	CAG Q		CGG R	<i>trnR-CCG</i>
AUU I	<i>trnI-GAU</i>	ACU T	<i>trnT-GGU</i>	AAU N	<i>trnN-GUU</i>	AGU S	<i>trnS-GCU</i>
AUC I		ACC T		AAC N		AGC S	
AUA I	<i>trnI-CAU</i>	ACA T	<i>trnT-UGU</i>	AAA K	<i>trnK-UUU</i>	AGA R	<i>trnR-UCU</i>
AUG M	<i>trnM-CAU</i>	ACG T	<i>trnT-CGU</i>	AAG K		AGG R	<i>trnR-CCU</i>
	<i>trnM-CAU</i>						
GUU V	<i>trnV-GAC</i>	GCU A	<i>trnA-GGC</i>	GAU D	<i>trnD-GUC</i>	GGU G	<i>trnG-GCC</i>
GUC V		GCC A		GAC D		GGC G	
GUA V	<i>trnV-UAC</i>	GCA A	<i>trnA-UGC</i>	GAA E	<i>trnE-UUC</i>	GGA G	<i>trnG-UCC</i>
GUG V	<i>trnV-CAC</i>	GCG A	<i>trnA-CGC</i>	GAG E		GGG G	<i>trnG-CCC</i>

Corresponding amino acids and tRNA species are given by one-letter abbreviations and italic letters.

Table 3. Genes relating to photosynthesis.

gene name	copy number		
	<i>T.elongatus</i>	<i>Synechocystis</i>	<i>Anabaena</i>
<i>apcA</i>	1	1	2
<i>cpcC</i>	1	2	1
<i>cpcG</i>	3	2	4
<i>pecA</i>	NI	NI	1
<i>pecB</i>	NI	NI	1
<i>pecC</i>	NI	NI	1
<i>nblA</i>	1	2	1
<i>psaB</i>	1	1	2
<i>psaI</i>	1	1	2
<i>psaK</i>	1	2	2
<i>psaX</i>	1	NI	1
<i>psbA</i>	3	3	5
<i>psbC</i>	1	1	4
<i>psbV</i>	2	1	1
<i>psb28</i>	1	1	2
<i>petC</i>	1	3	4
<i>petE</i>	NI	1	1
<i>petI</i>	1	1	3
<i>petL</i>	NI	1	1
<i>isiB</i>	NI	1	1
<i>ndhD</i>	4	6	5
<i>ndb</i>	1	3	5
<i>ctaC</i>	1	2	3
<i>ctaD</i>	1	2	3
<i>ctaE</i>	1	2	3
CAB/ELIP/HLIP	5	5	10
<i>ccmA</i>	1	1	2
<i>prk</i>	1	1	2
<i>rca</i>	NI	NI	1
<i>apcB, apcC, apcD, apcE, apcF, cpcA, cpcB, cpcD, cpcE, cpcF, psaA, psaC, psaD, psaE, psaF, psal, psalL, psalM, btpA, psbB, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbO, psbP, psb T, psbU, psbV, psbX, psbY, psbZ, isiA, ndhA, ndhB, ndhC, ndhE, ndhG, ndhH, ndhI, ndhJ, ndhK, ndhL, petA, petB, petD, petG, petH, petM, petN, cydA, cydB, cytM, ctaB, atpA, atpB, atpC, atpD, atpE, atpF, atpG, atpH, atpI, ccmL, ccmM, ccmN, rbcL, rbcS</i>		1	
<i>psbD, nblB</i>		2	
<i>ndhF</i>		3	
<i>ccmK</i>		5	

NI; not identified gene

Table 4. List of group II introns.

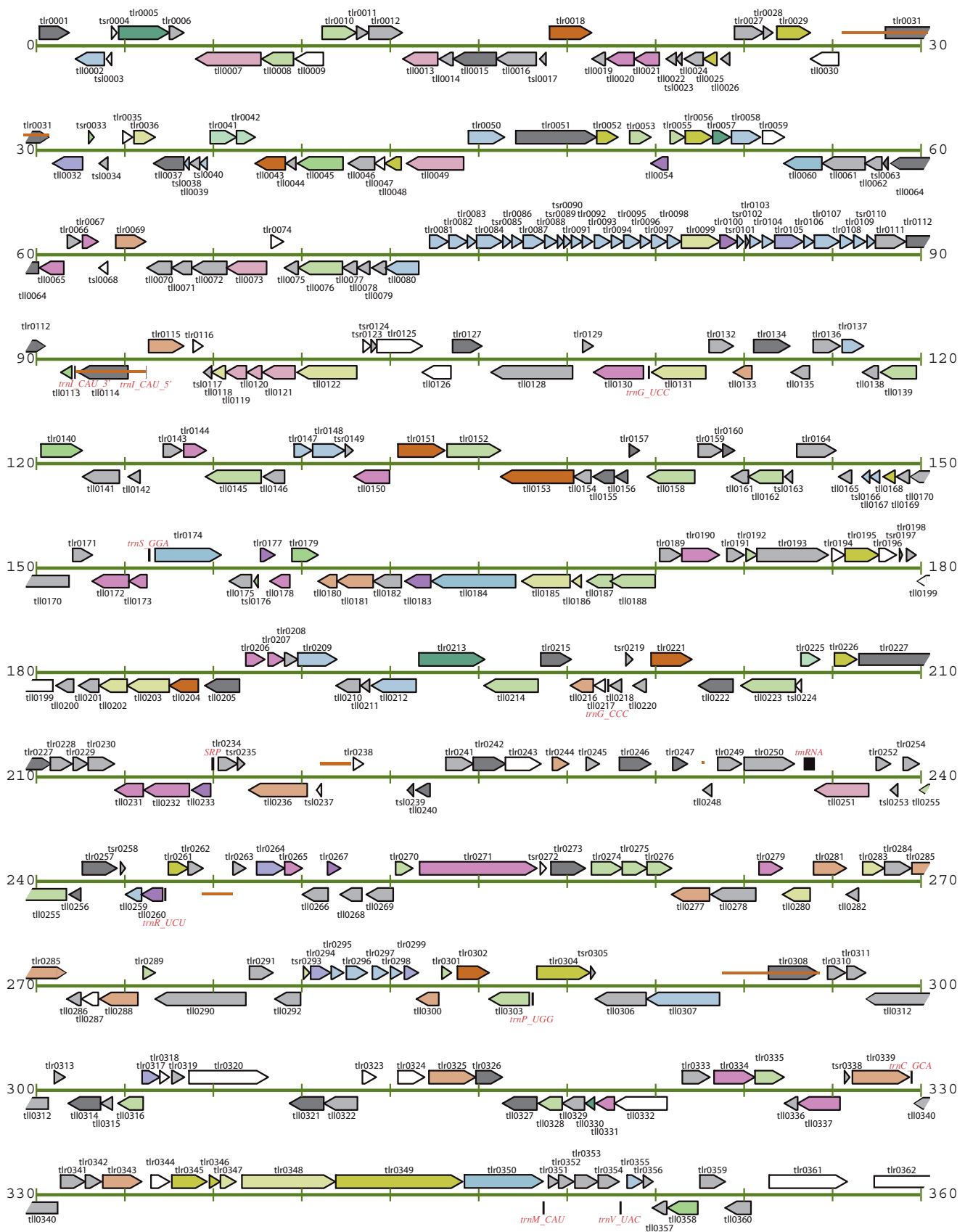
name	length (bp)	direction	position
TelI3a	838	>>	27972 to 28810
TelI3b	838	>>	294034 to 294872
TelI3c	838	>>	520111 to 520949
TelI3d	838	>>	1148767 to 1149605
TelI3e	838	>>	1197582 to 1198420
TelI3f	844	>>	219750 to 220594
TelI3g	74	>>	232653 to 232727
TelI3h	845	>>	444131 to 444771 ...join... 445611 to 445814
TelI3i	838	>>	444772 to 445610
TelI3j	844	>>	1040548 to 1041392
TelI3k	837	>>	1159085 to 1159922
TelI3l	845	>>	1683646 to 1684491
TelI3m	145	>>	1715514 to 1715659
TelI3n	408	>>	1717072 to 1717480
TelI3o	838	>>	2082690 to 2083528
TelI3p	849	>>	2169568 to 2170417
TelI3q	844	<<	245778 to 246622
TelI3r	849	<<	1578164 to 1579013
TelI3s	844	<<	2465187 to 2466031
TelI3t	844	<<	2564637 to 2565481
TelI4a	2384	>>	27344 to 27971 ...join... 28811 to 30566
TelI4b	2384	>>	293406 to 294033 ...join... 294873 to 296628
TelI4c	2384	>>	519483 to 520110 ...join... 520950 to 522705
TelI4d	2386	>>	1148139 to 1148766 ...join... 1149606 to 1151363
TelI4e	2384	>>	1196954 to 1197581 ...join... 1198421 to 1200176
TelI4f	2384	>>	624954 to 627338
TelI4g	2384	>>	1023460 to 1024515 ...join... 1025360 to 1026687
TelI4h	2385	<<	91363 to 93748

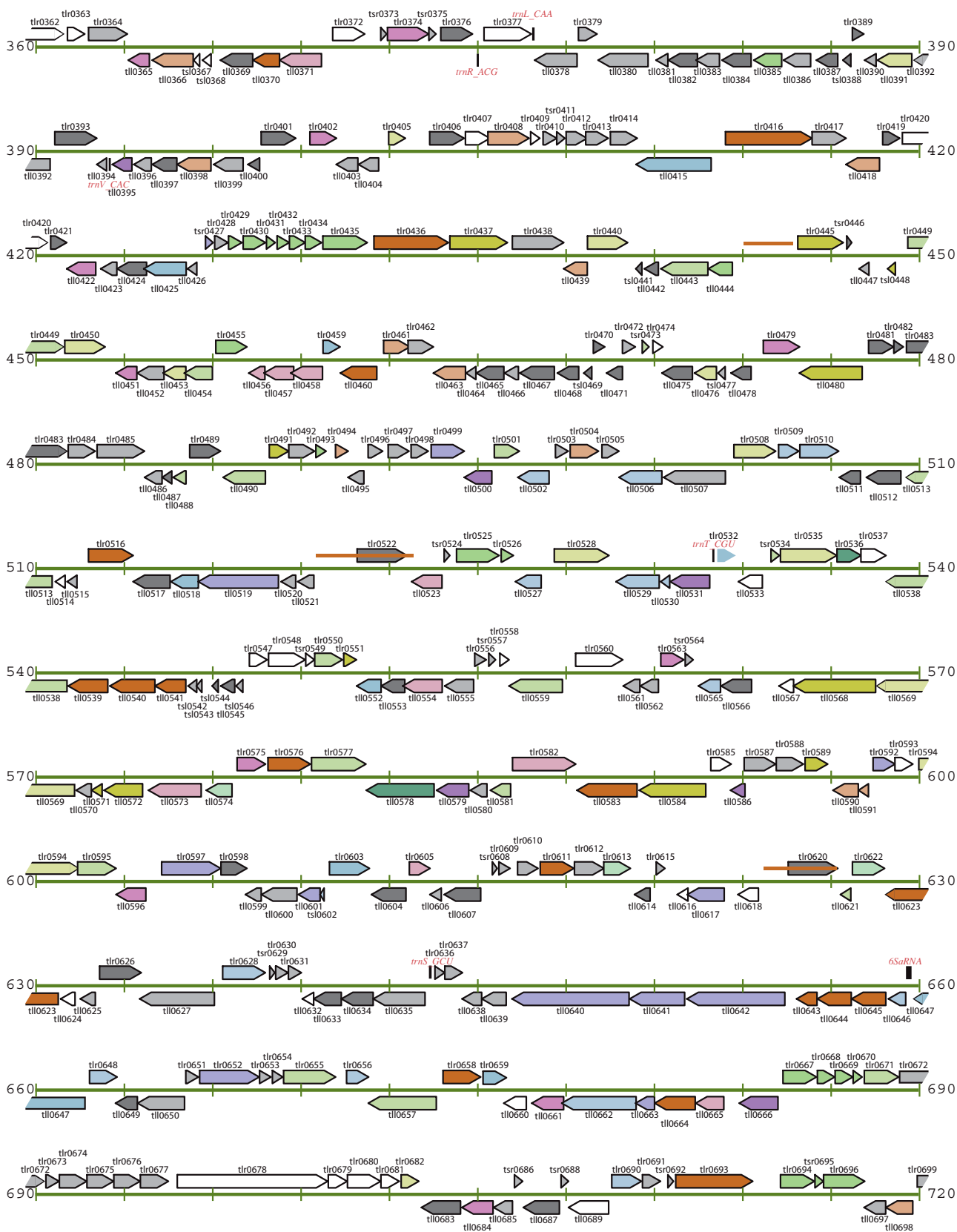
Table 5. Structural features of insertion sequences.

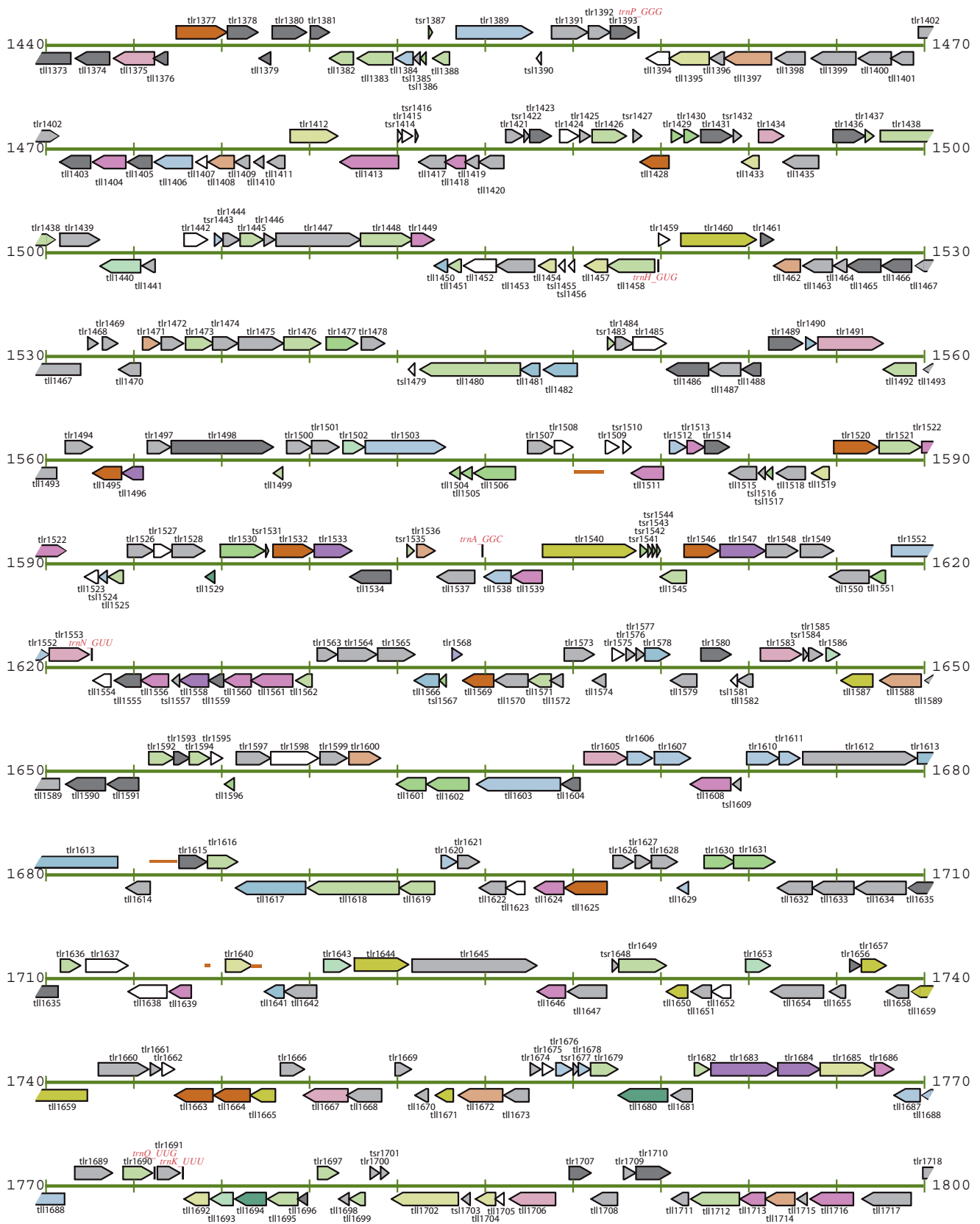
name	size (bp)	inverted repeat (bp)	direct repeat (bp)	family	number of full-length copies	number of partial copies
ISEL1	1327	0	0	IS200	19	3
ISEL2	841	19	3	IS5	8	0
ISEL3	1679	0	0	IS200 & IS605-TnpB	10	6
ISEL4	1101	27	6	IS481	14	4
ISEL5	1083	13	0	IS4	1	5

Table 6. List of insertion sequences.

name	length (bp)	direction	position	name	length (bp)	direction	position
ISEL1a	1327	>>	114293 to 115619	ISEL3a	1628	<<	138887 to 140514
ISEL1b	1326	<<	185703 to 187028	ISEL3b	1673	>>	241108 to 242780
ISEL1c	1327	<<	202441 to 203767	ISEL3c	1665	<<	386509 to 388173
ISEL1d	1327	>>	257362 to 258688	ISEL3d	1679	>>	397188 to 398866
ISEL1e	1327	>>	403280 to 404606	ISEL3e	1665	<<	467712 to 469376
ISEL1f	1327	<<	508183 to 509509	ISEL3f	1462	>>	1143402 to 1144863
ISEL1g	1327	<<	611376 to 612702	ISEL3g	1628	>>	1172493 to 1174120
ISEL1h	1327	<<	706533 to 707859	ISEL3h	1673	>>	1211291 to 1212963
ISEL1i	1327	<<	987226 to 988552	ISEL3i	1679	>>	1447268 to 1448946
ISEL1j	1327	>>	1158966 to 1159084	ISEL3j	1679	>>	2113577 to 2115255
			...join...	ISEL3k	1679	>>	2320398 to 2322076
			1159923 to 1161130	ISEL3l	1679	>>	2411784 to 2413462
ISEL1k	1327	<<	1183821 to 1185147	ISEL3m	1679	>>	2511996 to 2513674
ISEL1l	1327	<<	1200403 to 1201729	ISEL3n	193	>>	720543 to 720735
ISEL1m	1327	<<	1440988 to 1442314	ISEL3o	194	>>	1477375 to 1477568
ISEL1n	1325	>>	1554594 to 1555918	ISEL3p	193	<<	2430698 to 2430890
ISEL1o	1326	<<	2091772 to 2093097	ISEL4a	1101	>>	197063 to 198163
ISEL1p	1327	>>	2263103 to 2264429	ISEL4b	83	<<	311724 to 311806
ISEL1q	1327	>>	2266166 to 2267492	ISEL4c	1101	<<	381500 to 382600
ISEL1r	1327	>>	2282980 to 2284306	ISEL4d	1101	<<	471284 to 472384
ISEL1s	1327	>>	2486125 to 2487451	ISEL4e	747	<<	473630 to 474376
ISEL1t	1283	<<	2545413 to 2546695	ISEL4f	1101	<<	563293 to 564393
ISEL1u	281	>>	2169449 to 2169567	ISEL4g	1101	<<	742525 to 743625
			...join...	ISEL4h	1101	<<	951831 to 952931
			2170418 to 2170579	ISEL4i	1101	>>	1043030 to 1044130
ISEL1v	281	<<	1578002 to 1578163	ISEL4j	1101	>>	1145293 to 1146393
			...join...	ISEL4k	656	>>	1405394 to 1406049
			1579014 to 1579132	ISEL4l	1101	>>	1411979 to 1413079
ISEL2a	841	<<	222600 to 223440	ISEL4m	1101	<<	1551623 to 1552723
ISEL2b	841	<<	440384 to 441224	ISEL4n	1101	>>	1642323 to 1643423
ISEL2c	841	<<	545998 to 546838	ISEL4o	1101	<<	1709435 to 1710535
ISEL2d	841	>>	778446 to 779286	ISEL4p	1101	<<	2020021 to 2021121
ISEL2e	841	>>	803944 to 804784	ISEL4q	747	<<	2190540 to 2191286
ISEL2f	841	>>	1024516 to 1025356	ISEL4r	1101	<<	2220512 to 2221612
ISEL2g	841	<<	1290353 to 1291193	ISEL5a	958	>>	231185 to 232142
ISEL2h	841	>>	2022099 to 2022939	ISEL5b	1083	<<	2494995 to 2496077
				ISEL5c	281	>>	229502 to 229782
				ISEL5d	110	<<	1021491 to 1021600
				ISEL5e	125	<<	2565518 to 2565642
				ISEL5f	83	>>	124124 to 124206







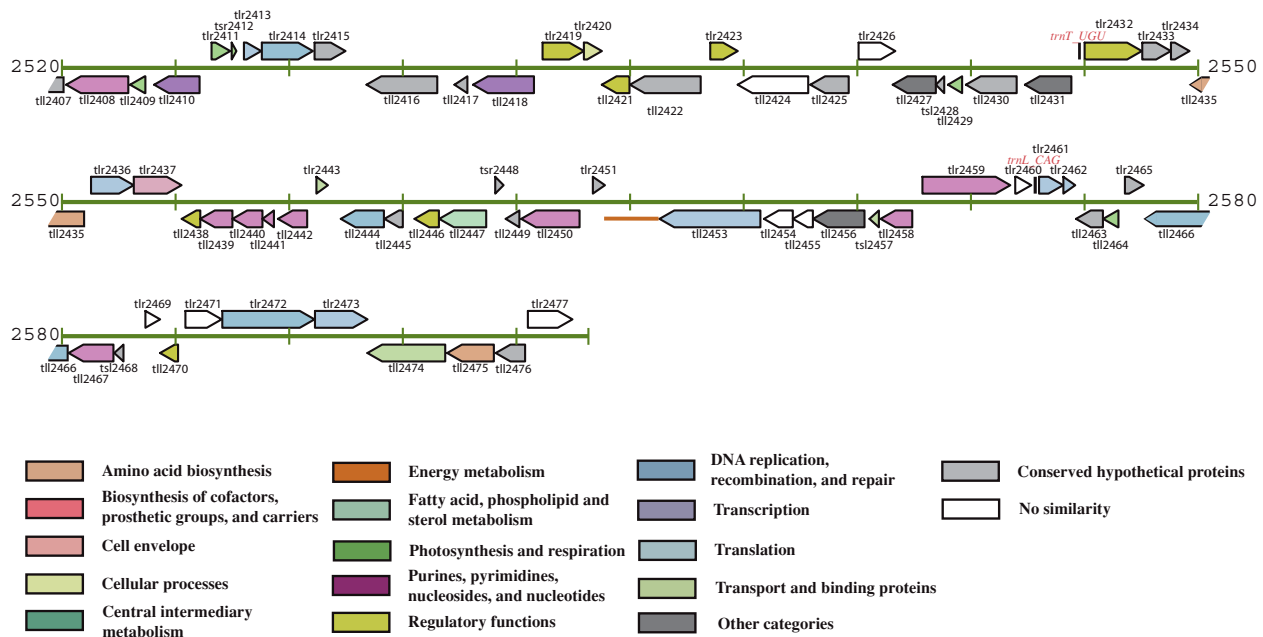


Figure 1. Gene map of the genome of *Thermosynechococcus elongatus* BP-1. The circular chromosome (2,593,857 bp) was opened at the recognition sites of *Swa* I, and is represented by a linear map starting from this junction (map position 0/100). Green bars show the scale in 3 kb with numerals in kb. On both sides of the scale, the potential protein-encoding genes assigned on the basis of computer prediction and the similarity search were shown by boxes with arrowheads indicating the reading direction. The detailed assignment procedures are described in the main article in this issue (pp. 123–pp. 130). The potential genes of which the function could be evaluated by similarity search were classified into 14 functional categories, and indicated by different color codes. The RNA-encoding genes are represented by black bars and boxes. The regions predicted as introns are shown by horizontal red bars. The sequences as well as the gene information shown in this paper are available in the Web database, CyanoBase, at <http://www.kazusa.or.jp/cyanobase/>. The sequence data analyzed in this study have been registered in DDBJ/GenBank/EMBL under accession number BA000039.

Aminoacyl Stem	D domain stem	loop	stem	Anticodon domain stem	loop	stem	Variable region	TψC domain stem	loop	stem	Aminoacyl stem			
trnI-CAU (93765-93749, 91362-91307)	CCAGGGT	TG GCC	GAGCGGATGA	GGC	A	GCGAA	CTCATAA	TTCGC	CATAG	GCTGG	TTCGACT	CCAGC	ACCCTGG	A
trnG-UCC (110834-110764)	GCGGGCA	TG GTTT	AATGGTA	AAAC	C	CTAGC	CTTCCAA	GCTAG	AGAC	GCGGG	TTCGATT	CCCAG	TGCCCGC	T
trnS-GGA (153813-153897)	GGAGAGA	TG GCC	GAGTGGTCGAA	GGC	G	CAGCA	CTGAAA	TGCTG	TGTGGGGCAACTCACC	GAGGG	TTCGAAT	CCCTC	TCTCTCC	G
trnG-CCC (199454-199383)	GCGGACG	TA ATTC	AGTGGTA	GAAT	G	TCAGC	TTCCCAA	GCTGA	ACGTC	GTGGG	TTCGAGT	CCCAT	CGTCCGC	T
trnR-UCU (244435-244362)	GCCGCCG	TA GCTC	AGCGAATA	GAGC	AC	TCGC	CTTCTAA	GCGA	TTGGTC	GTTGG	TTCGAGT	CCAAC	CGGGCGC	G
trnP-UGG (286896-286820)	CGGGATG	TA GCGC	AGCTTGGTA	GCGC	A	TCTGC	TTTGGGA	GCAGA	GGGTC	GCAGG	TTCAAAT	CCTGT	CATCCCG	A
trnC-GCA (329671-329742)	GGCGGCA	TA GCC	AAGTGGTAA	GGC	A	GAGGT	CTGAAA	ACCTT	TATTC	CCCCG	TTCGAAT	CCGGG	TGCCGCC	T
trnM-CAU (347262-347190)	GGCTCAG	TA GCTC	AGTGGTA	GAGC	AG	GGG	ACTCATAAG	CCC	AAGGTC	GCAGG	TTCGAAT	CCCAG	CTGAGCC	A
trnV-UAC (349873-349801)	GGGCGGT	TA ACTC	AGTGGTA	GAGT	A	TCTCG	TTTACAC	CGAGG	CTGTC	GGGGG	TTCGAGT	CCCTC	ACCGCCC	A
trnR-ACC (375063-374990)	GGGCTTG	TA GCTT	AGTGGTAA	GAGC	G	GCTGA	CTACGGA	TCAGC	AGGTC	GGGGG	TTCGAGT	CCCTC	CAAGCCC	G
trnL-CAA (376874-376955)	GGGCGAG	TG GCG	AAATGGTAGA	CGC	A	CTACA	CTCAAAA	TGTAG	CGGGTTTCCCAT	GTCAG	TTCGAGT	CTGAC	CTCGCCC	A
trnV-CAC (392565-392494)	GGGCGAT	TA GCAC	AGTGGTA	GCGC	A	CTTCC	TTACAC	GGAAG	GGGTC	ACTGG	TTCGAAT	CCAGT	ATCGCCC	A
trnT-CGU (533046-533117)	GCCGATG	TA GCTC	AGTGGTA	GAGC	A	CTCGA	TTCTGTA	TCGAG	CGGTC	GTGAG	TTCGAAT	CTCAT	CATCGGC	T
trnS-GCU (643469-643560)	GGAGAGG	TG GCT	GAGTGGTCGAA	AGC	G	GCAGA	TTGCTAA	TCTGT	TGATAGGCAGGTAACCTCTATCC	GAGGG	TTCGAAT	CCCTC	CCTCTCC	G
trnL-GAG (772101-772021)	GCGGATG	TG GCG	GAATTGGTATA	CGC	G	CACGT	TTGAGGG	GCGTG	TGGCTTTGCCCTT	GCGAG	TTCGAGT	CTCGC	CATCCGC	A
trnP-CGG (1091330-1091257)	CGGGATG	TA GCGC	AGCTTGGTA	GCGC	A	CTTCG	TTCCGGA	CGAAG	GGGTC	GCAGG	TTCAAAT	CCTGT	CATCCCG	A
trnL-UAG (1122824-1122905)	GCGGACG	TG GCG	GAATTGGCAGA	CGC	G	CTAGA	TTTAGGT	TCTAG	TACCCGAAGGTGT	GTGGG	TTCAAGT	CCCTC	CGTCCGC	A
trnV-GAC (1127199-1127126)	GGACGCA	TA GCTC	AGTGGTAA	GAGC	A	CCACG	TTGACAT	CGTGG	GGGTC	ACTGG	TTCGAGT	CCAGT	TGTGTCC	A
trnS-UGA (1240471-1240385)	GGAGAGA	TG GCA	GAGTGGTCGAA	TGC	G	CTCGA	CTTGA	TCGAG	TGTGGCAGTATGCCACC	GAGGG	TTCGAAT	CCCTC	TCTCTCC	G
trnR-CCU (1340921-1340994)	GGGGCTG	TA GCTC	AGTGGTAA	GAGC	G	AGCGC	CTCCTAA	GCGCT	AGGCC	GTGCG	TTCAAAT	CGCAC	CAGTCCC	G
trnW-CCA (1354396-1354468)	GCGTCCT	TA GTTC	AGTGGTA	GAAC	G	TCGGT	CTCCAAA	ACCGG	ATGTC	GGGGG	TTCGAGT	CCTCC	AGGGCGC	G
trnP-GGG (1460223-1460296)	CGGGGCG	TA GCGC	AGCTTGGTA	GCGC	A	CCACT	TTGGGGT	AGTGG	GGGTC	GTGGG	TTCAAAT	CCCAG	CGCTCCG	A
trnH-GUG (1520900-1520828)	GCCGAGC	TA GCC	AAGTGGTAA	GGC	A	GTGGA	TTGTGGT	TCCAC	CATTC	GTGGG	TTCGAGT	CCCAT	CGTCCGC	C
trnA-GGC (1604896-1604968)	GGGGCTA	TA GCTC	AGTGGTA	GAGC	A	CTTCA	ATGGCAT	TGAAG	GGGTC	AGCGG	TTCGAAT	CCGCT	TAGTCC	A
trnN-GUU (1621556-1621627)	TCCTCAG	TA GCTC	AGCGGTA	GAGC	G	GTCGG	CTGTAA	CCGAT	TGGTC	GCTGG	TTCGAAT	CCAGC	CTGGGGA	G
trnQ-UUG (1773740-1773811)	TGGGGTG	TC GCC	AAGCGGTA	GGC	A	GCGGG	TTTTGGT	CCCCG	CATTC	GGGGG	TTCGAAT	CCTCC	CGCCCCA	G
trnK-UUU (1774666-1774737)	GGGTCCG	TA GCTC	AGCGGTA	GAGC	AC	TCGG	CTTTTAA	CCGA	TTGGTC	TTGGG	TTCGAAT	CCCAG	GCGACCC	A
trnG-GCC (1837204-1837133)	GCGGGTA	TA GCTC	AGTGGTA	GAGC	G	CAACC	TTGCCAA	GGTTG	ATGTC	GCGCG	TTCGAAT	CGCGT	TACCCGC	T
trnD-GUC (1921204-1921128)	GGGACTG	TA GTTC	AATCGGTA	GAGC	A	CCGCC	CTGTAC	GGCGG	AAGTT	GCGGG	TTCGAGC	CCCCT	CAGTCCC	G
trnT-GGU (2011688-2011759)	GCCCGTG	TA GCTC	AGTGGTA	GAGC	A	CACCC	TTGGTAA	GGGTG	AGGTC	ACGAG	TTCAATC	CTCGT	CACGGGC	T
trnF-GAA (2057341-2057269)	GCCGGGA	TA GCTC	AGTGGTA	GAGC	A	GAGGA	CTGAAA	TCCTC	GTGTC	GCCAG	TTCAATT	CTGGC	TCCTGGC	A
trnR-CCG (2094075-2094148)	GGACGTG	TG GCTC	AGTGGTAA	GAGC	A	TCGGA	TTCCGGT	TCCGA	GGGTC	GGGGG	TTCGAAT	CCCTC	CACGTCC	G
trnA-CGC (2156389-2156461)	GGGGAGT	TA GCTC	AGTGGTA	GAGC	G	TCGCG	ATCGCAC	CGCGA	AGGTC	AGGGA	TTCGAGT	TCCCT	ACTCTCC	A
trnFM-CAU (2183813-2183886)	CGCGGGG	TA GAGC	AGTGGTAA	GCTC	G	TCGGG	CTCATAA	CCCCG	AGGTC	CATGG	TTCAAAT	CCATG	CCCCGCC	A
trnA-UGC (2335041-2334969)	GGGGGTA	TA GCTC	AGTGGTA	GAGC	A	CCTGC	TTTGCAC	GCAGG	GGGTC	AGCGG	TTCGAGT	CCGCT	TACCTCC	A
trnI-GAU (2335127-2335054)	GGGCTAT	TA GCTC	AGTGGTAA	GAGC	G	CACCC	CTGATAA	GGGTG	AGGTC	CCTGG	TTCGAGT	CCAGG	ATGGCCC	A
trnL-UAA (2356120-2356085, 2355845-2355793)	GGGGGCG	TG GCG	GAATGGTAGA	CGC	T	GCGGA	CTTAAGAAA	TCCGT	TGACCTTAACAAGTCTGT	GAGGG	TTCAAGT	CCCTC	CGCCCCC	A
trnY-GUA (2372804-2372888)	GGGTCCG	TG CCCG	AGTGGTAA	TGGG	G	GCGGA	CTGTAAA	TCCGT	TGGCTACGCTAC	GCTGG	TTCGAAT	CCAGC	CCGGCCC	A
trnS-CGA (2402369-2402455)	GGAGAGG	TG GCA	GAGTGGTAA	TGC	G	CTTGA	CTCGAAA	TCAAAG	TTTAGGGTAACACCTAAC	GGGGG	TTCGAAT	CCCCC	CCTCTCC	G
trnT-UGU (2546857-2546929)	GCTGGTG	TA GCTC	AGTGGTA	GAGC	A	ACTGA	CTTGTAA	TCAGT	AGGTC	GTCGG	TTCGAGT	CCGAC	CATCAGC	T
trnL-CAG (2575683-2575766)	GCGGAAC	TG GCG	GAATGGTAGA	CGC	G	CTAGA	TTTAGGT	TCTAG	TGTCCCAAGGACTT	CGGGG	TTCAAGT	CCCCG	GTCCCGC	A
trnE-UUC (2459178-2459106)	GCCCCCA	TC GTCT	AGTGGCTAA	GGAC	A	CCTCC	CTTTCAC	GGAGG	CGAC	GGGGA	TTCGAAT	TCCCC	TGGGGGT	A

Figure 2. Structure of the tRNA genes in the genome of *T. elongatus*. The nucleotide sequences, nucleotide positions and the structural domains for the 42 tRNA genes in the genome are tabulated.

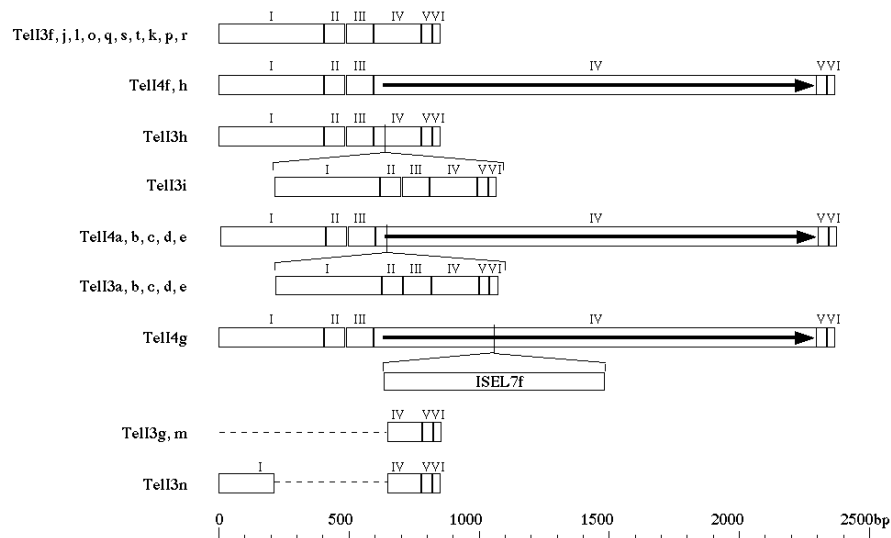


Figure 3. Structure of group II introns. The structures of group II introns found in the *T. elongatus* genome are depicted as open boxes with the positions of the six conserved domains (I to VI). Horizontal broken lines show the deleted regions. Horizontal arrows in domain IV represent presumptive genes for maturase/reverse transcriptase and the direction of transcription. Positions of insertions of other group II introns or IS are indicated as vertical bars with open boxes.

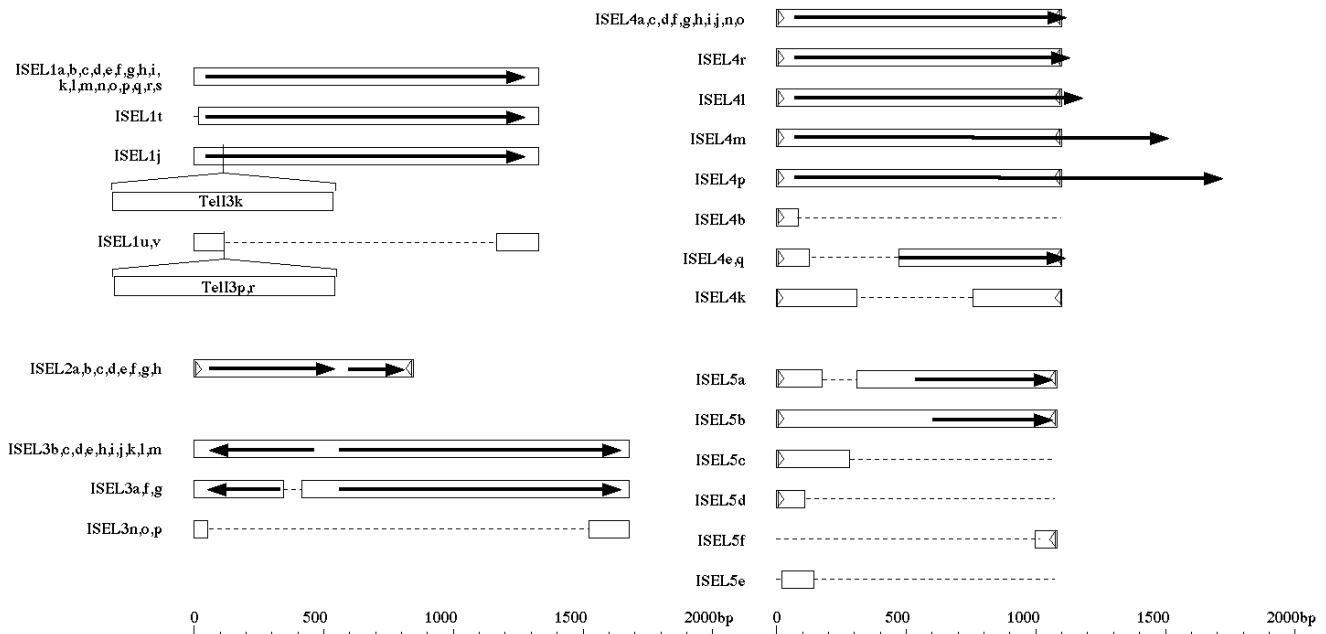


Figure 4. Structure of insertion sequences. The structures of group II introns found in the *T. elongatus* genome are depicted as open boxes. Horizontal broken lines show the deleted regions. Open triangles at the ends of the open boxes represent the terminal inverted repeats. Horizontal arrows indicate the positions of genes for putative transposases. Positions of insertions of group II introns are indicated as vertical bars with open boxes.