

Complete Genomic Sequence of Nitrogen-fixing Symbiotic Bacterium *Bradyrhizobium japonicum* USDA110 (Supplement)

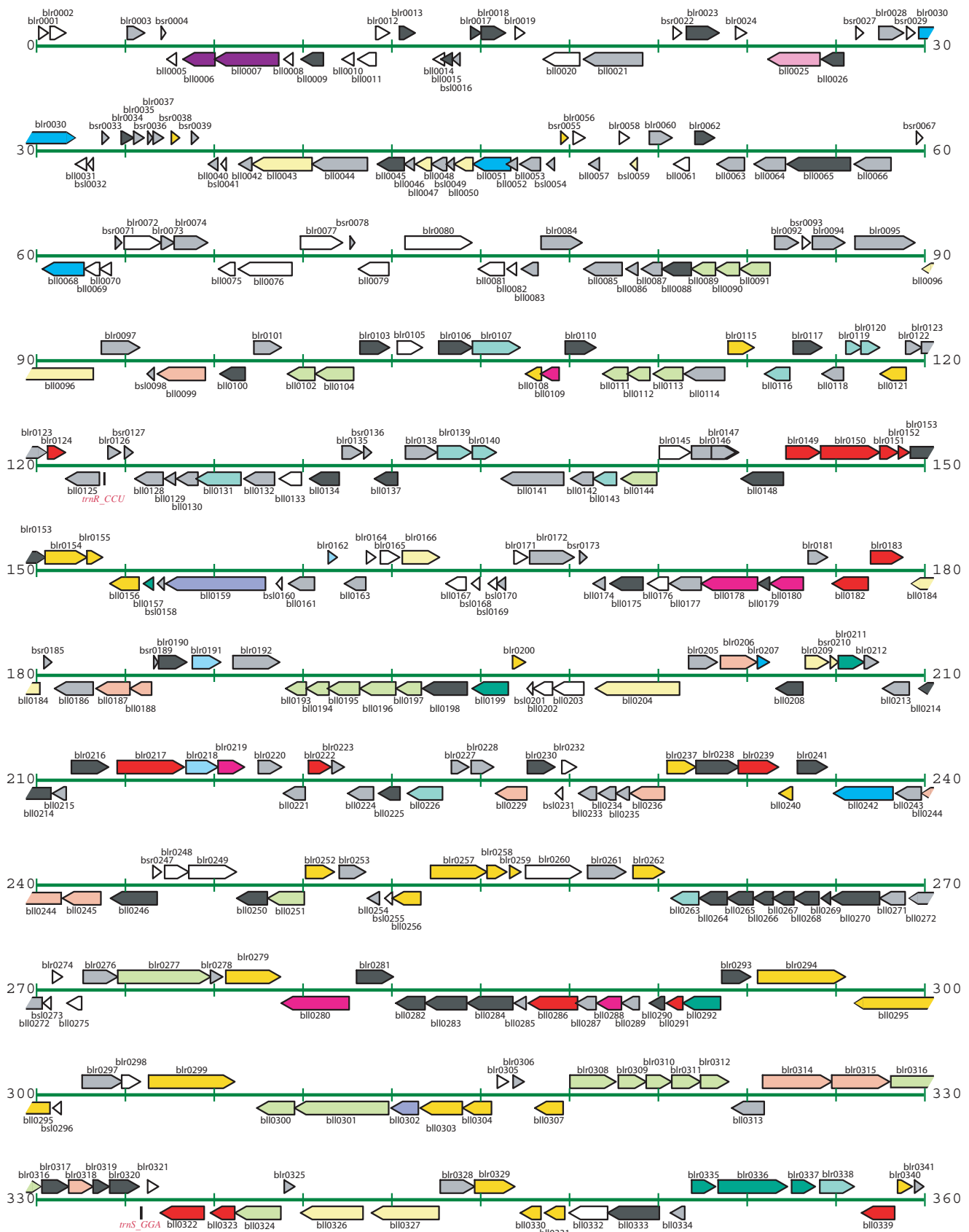
Takakazu KANEKO,¹ Yasukazu NAKAMURA,¹ Shusei SATO,¹ Kiwamu MINAMISAWA,² Toshiki UCHIUMI,³ Shigemi SASAMOTO,¹ Akiko WATANABE,¹ Kumi IDESAWA,¹ Mayumi IRIGUCHI,¹ Kumiko KAWASHIMA,¹ Mitsuyo KOHARA,¹ Midori MATSUMOTO,¹ Sayaka SHIMPO,¹ Hisae TSURUOKA,¹ Tsuyuko WADA,¹ Manabu YAMADA,¹ and Satoshi TABATA^{1,*}

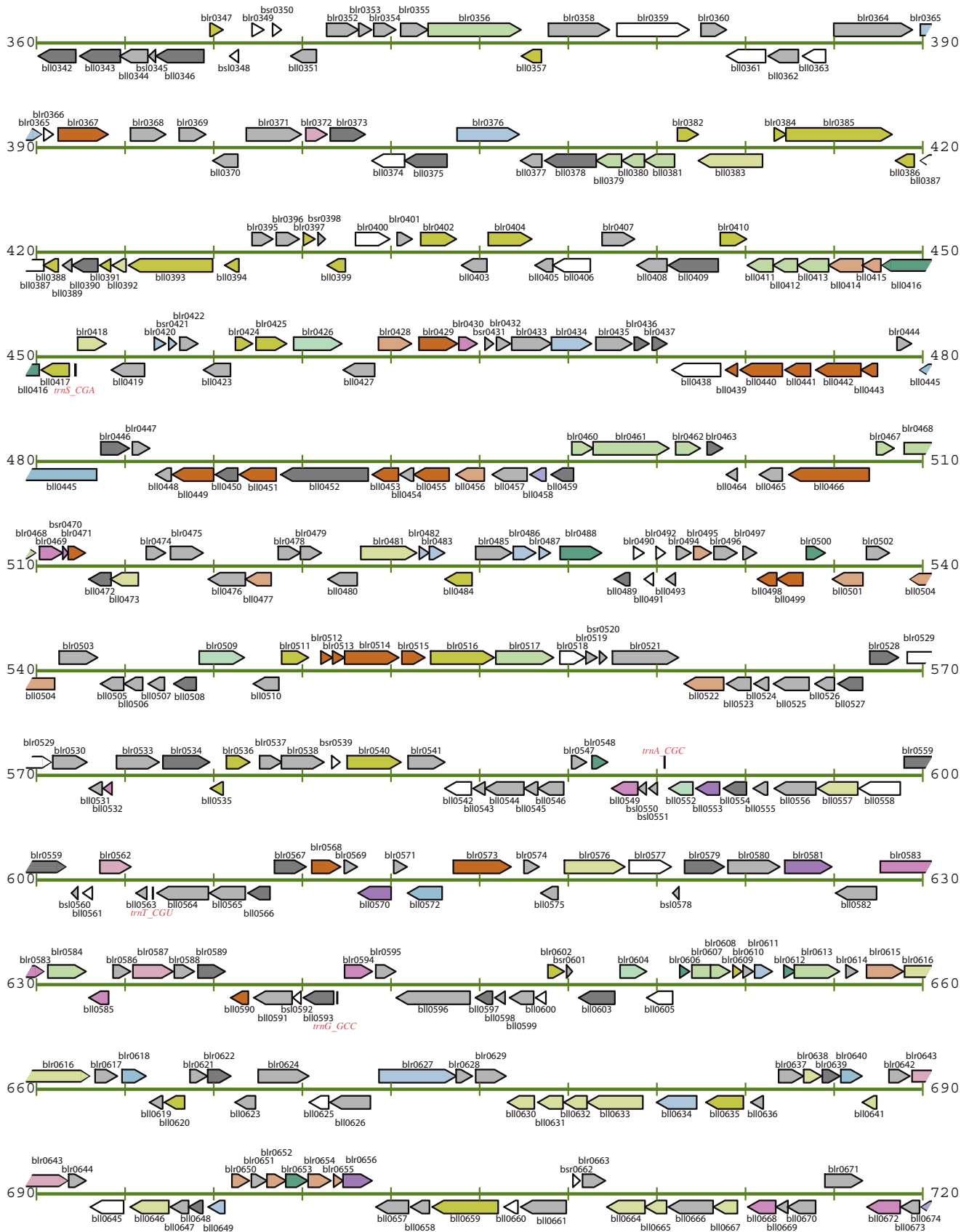
Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan,¹ Graduate School of Life Sciences, Tohoku University, Katahira, Aoba-ku, Sendai 980-8577, Japan,² and Department of Chemistry and BioScience, Faculty of Science, Kagoshima University, Kagoshima 890-0065, Japan³

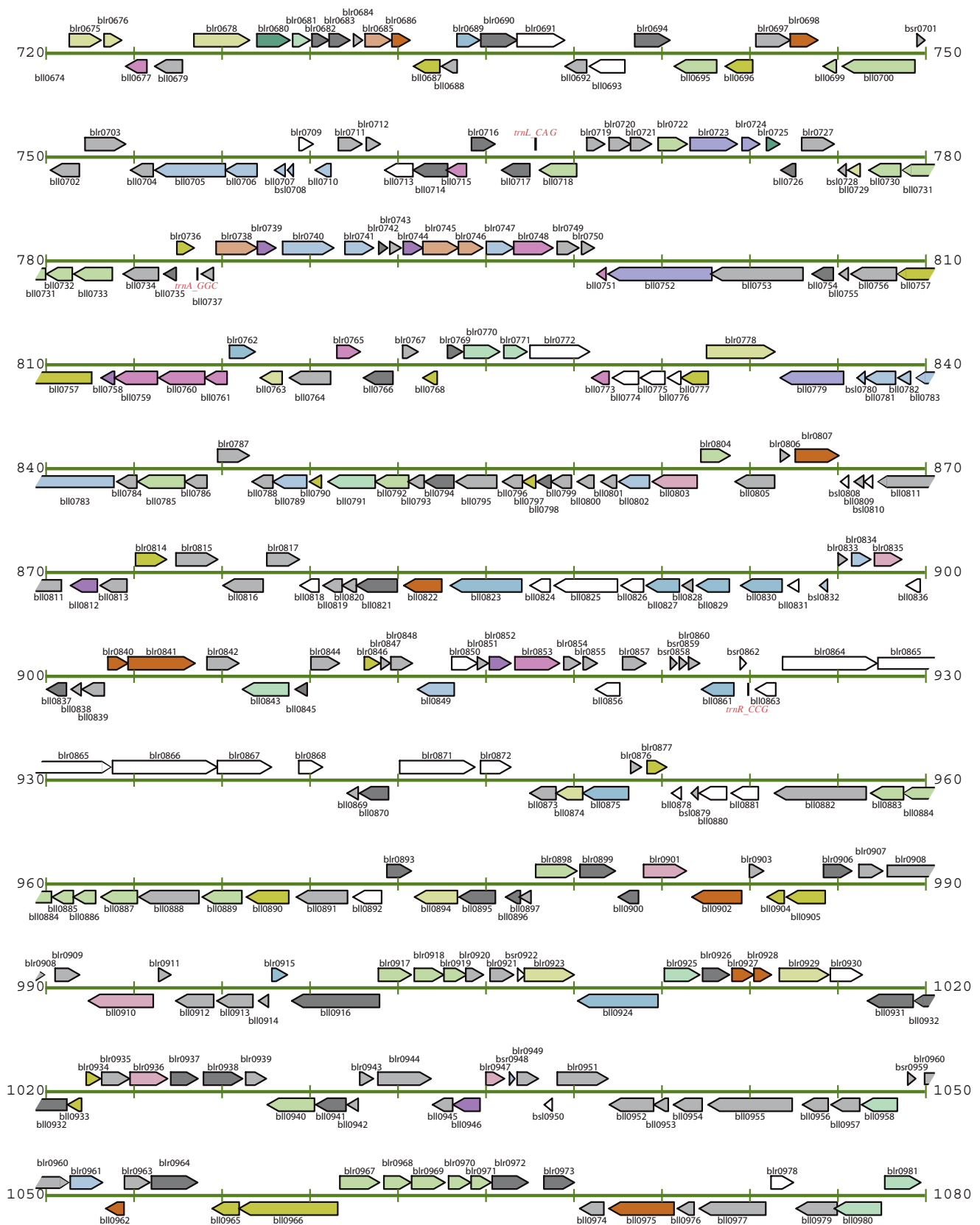
(Received 18 November 2002)

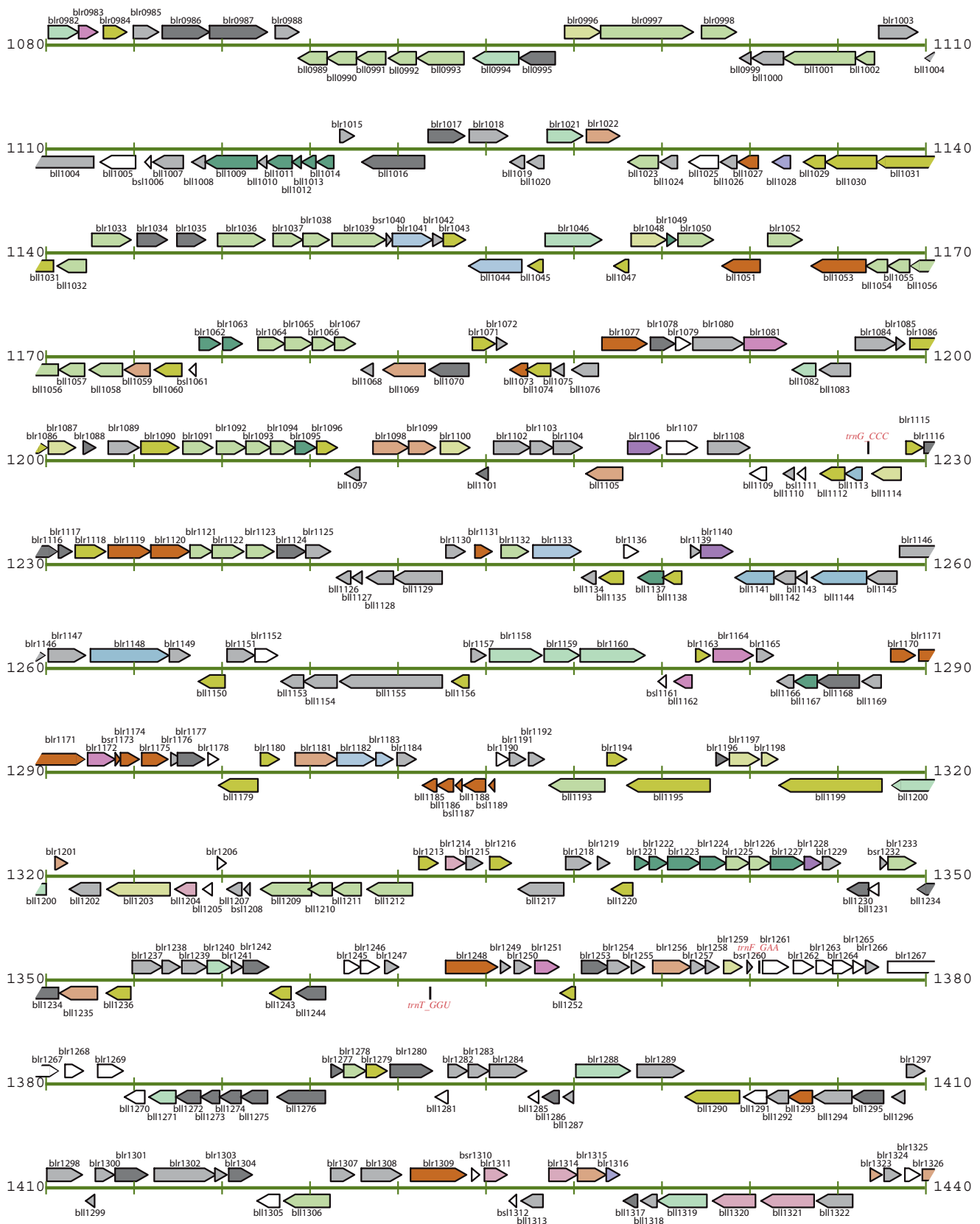
Communicated by Mituru Takanami

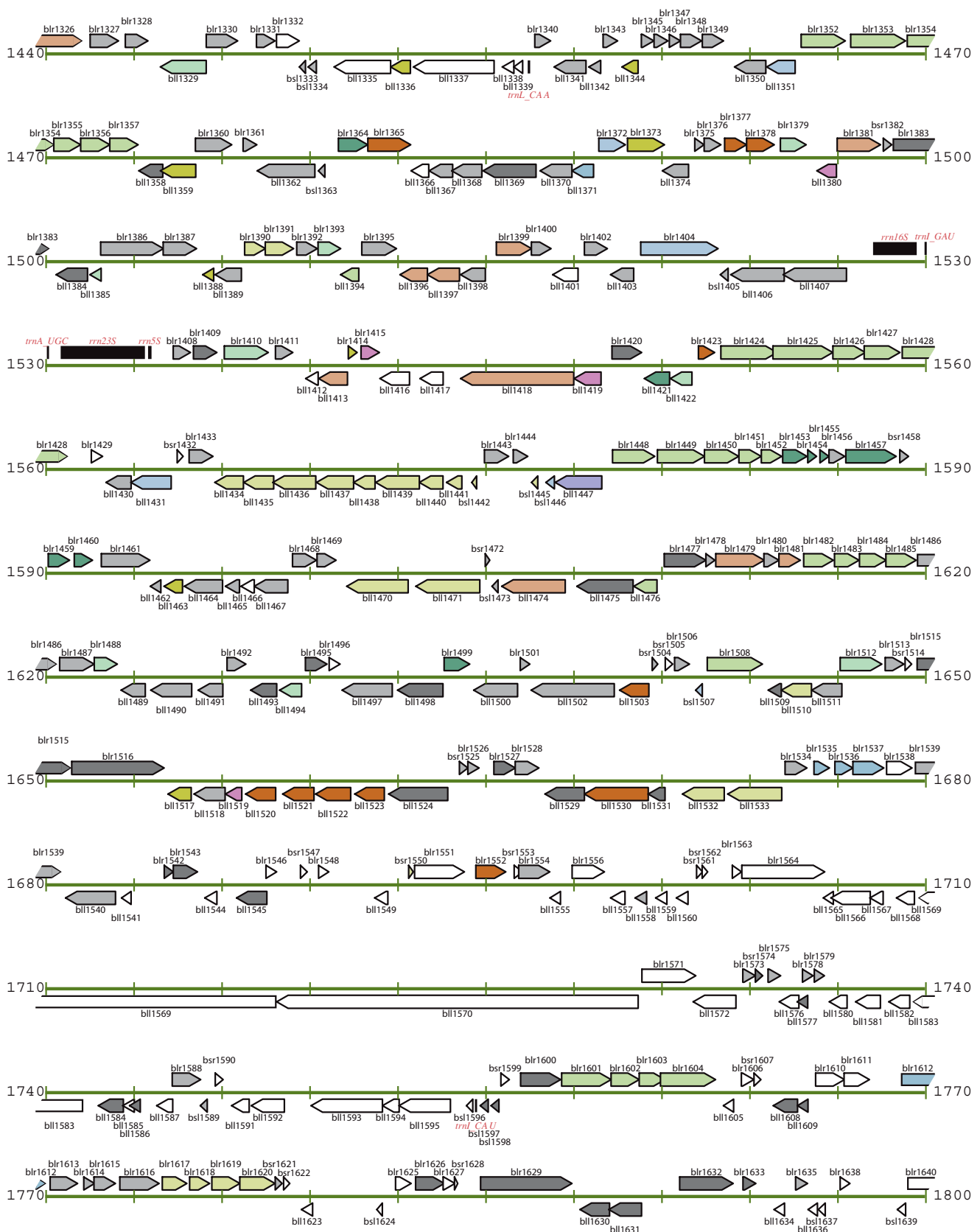
* To whom correspondence should be addressed. Tel. +81-438-52-3933, Fax. +81-438-52-3934, E-mail: tabata@kazusa.or.jp

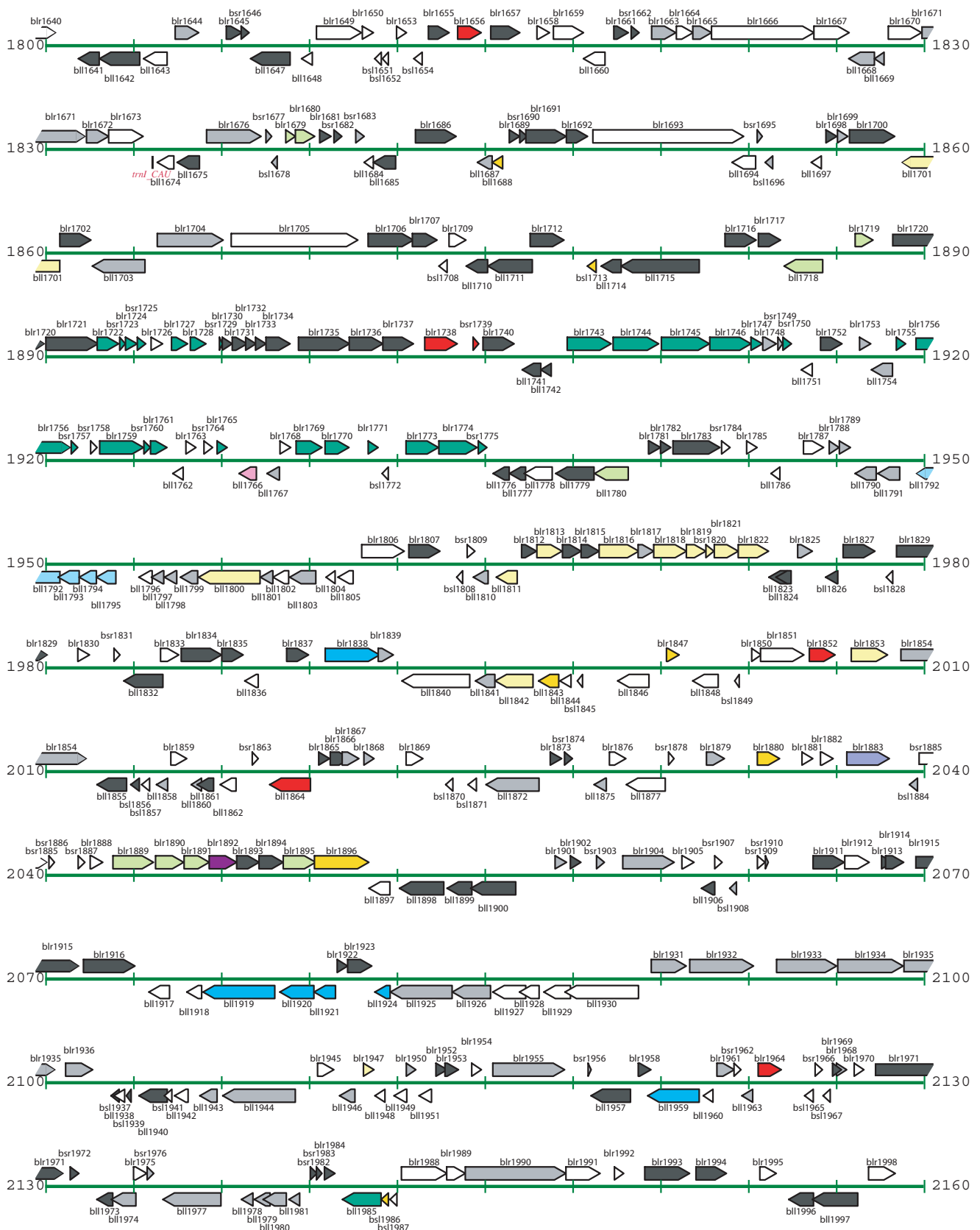


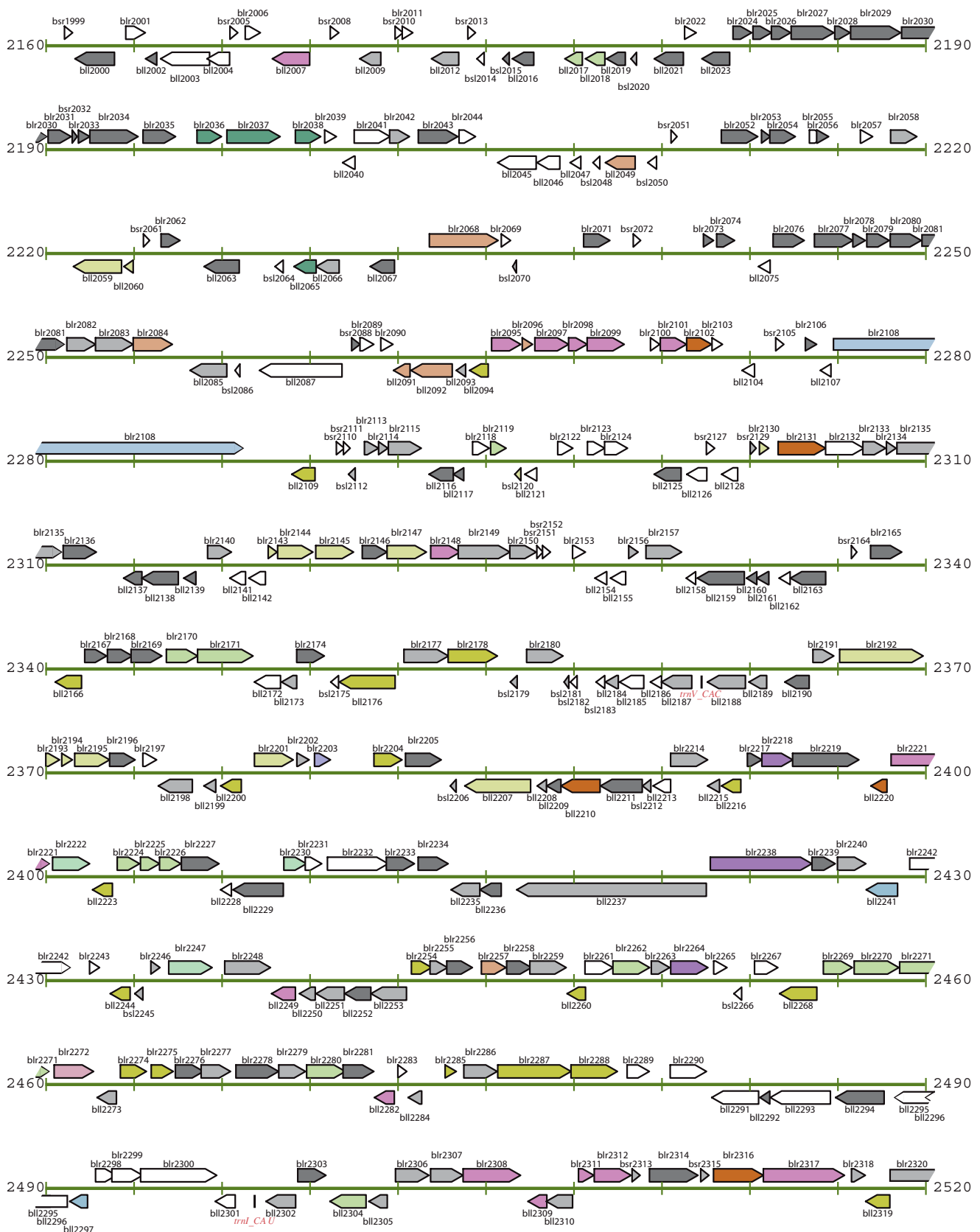


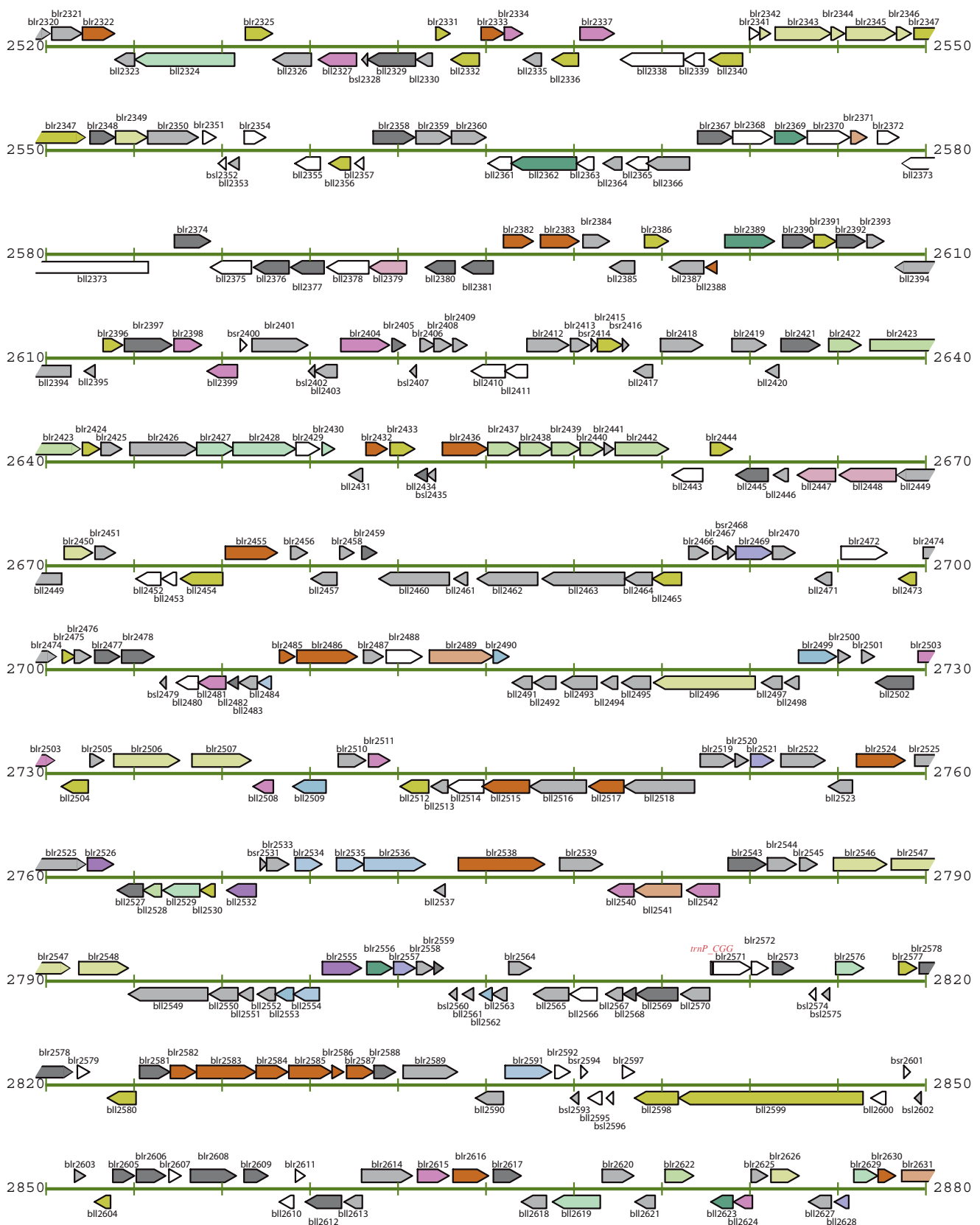


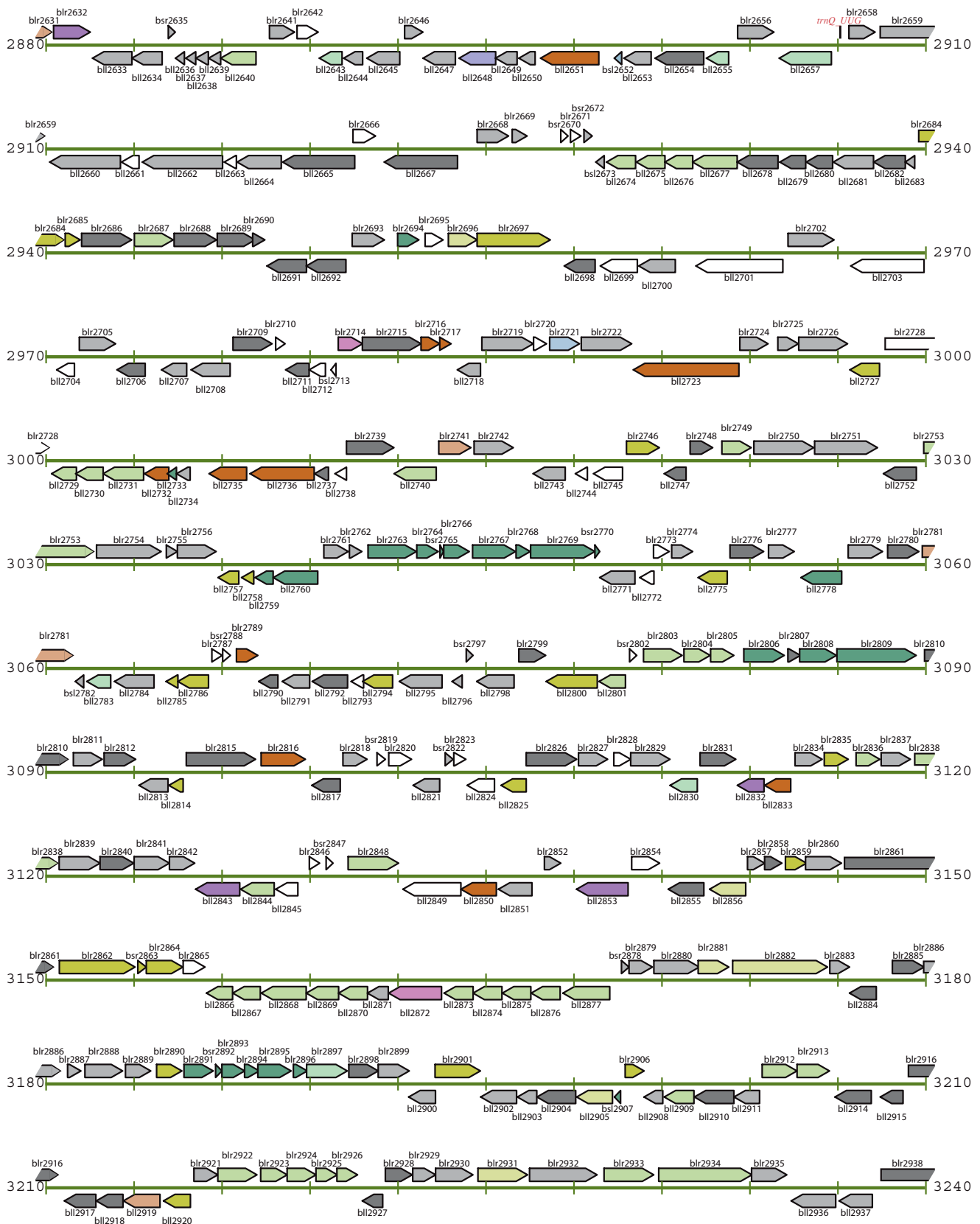


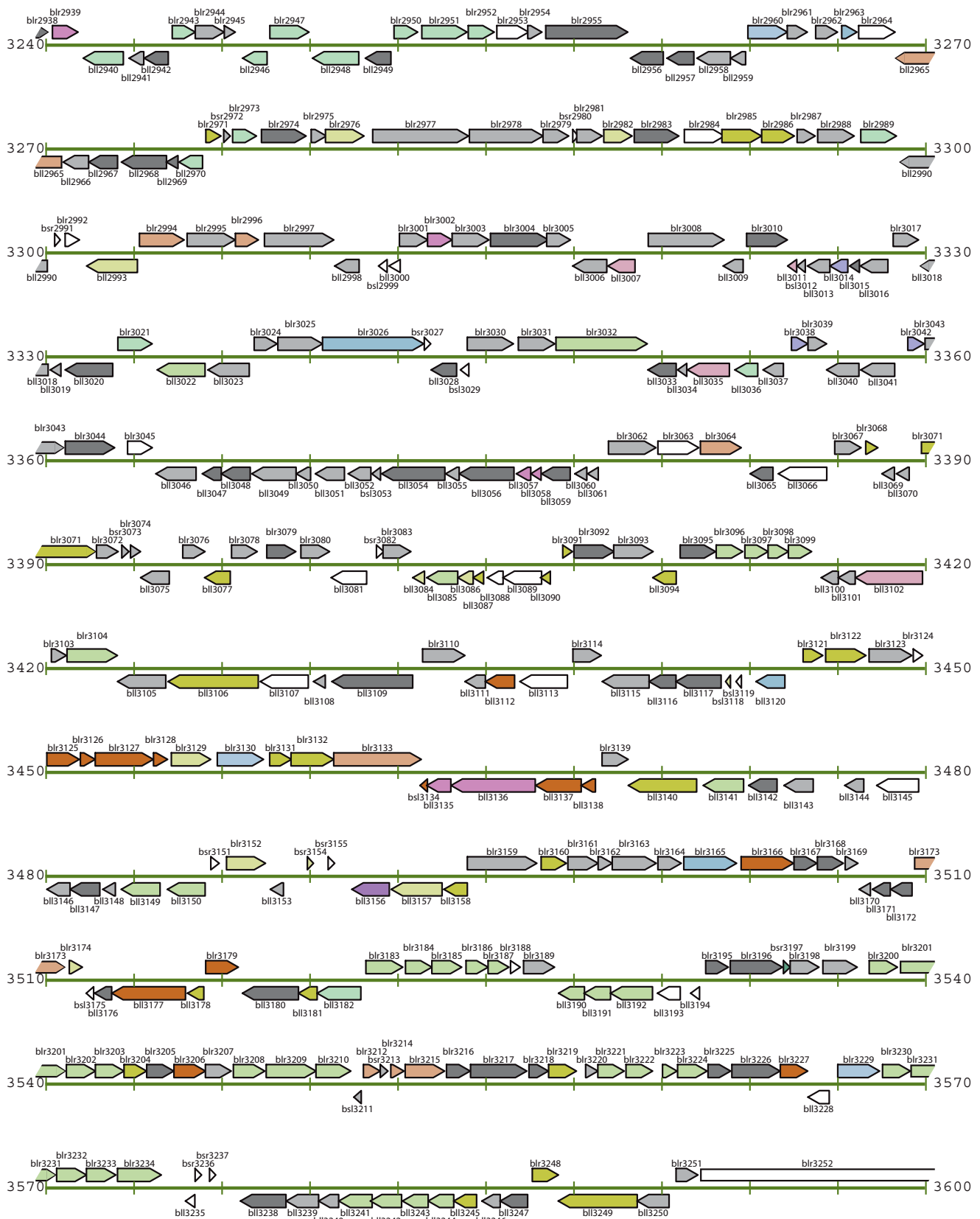


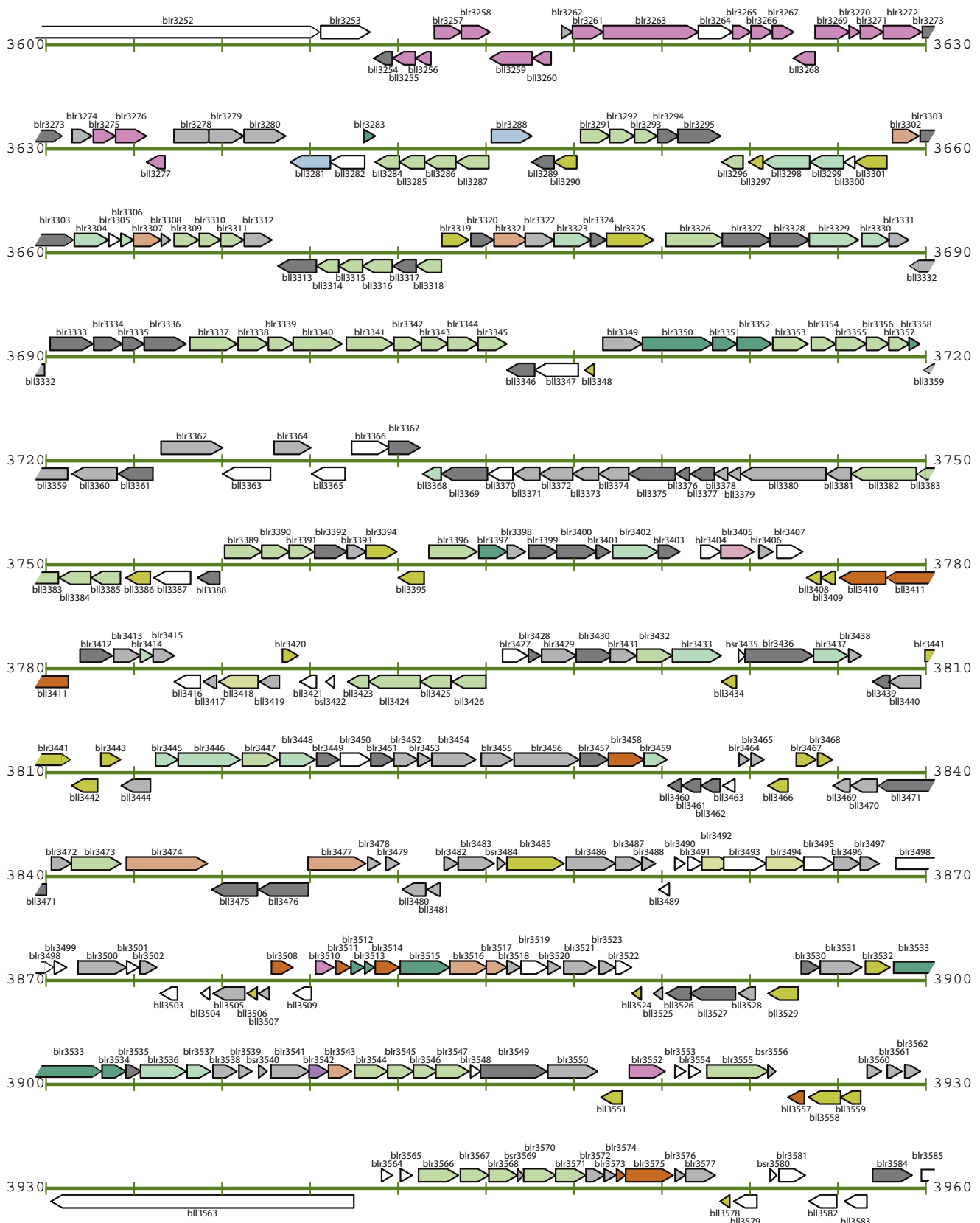


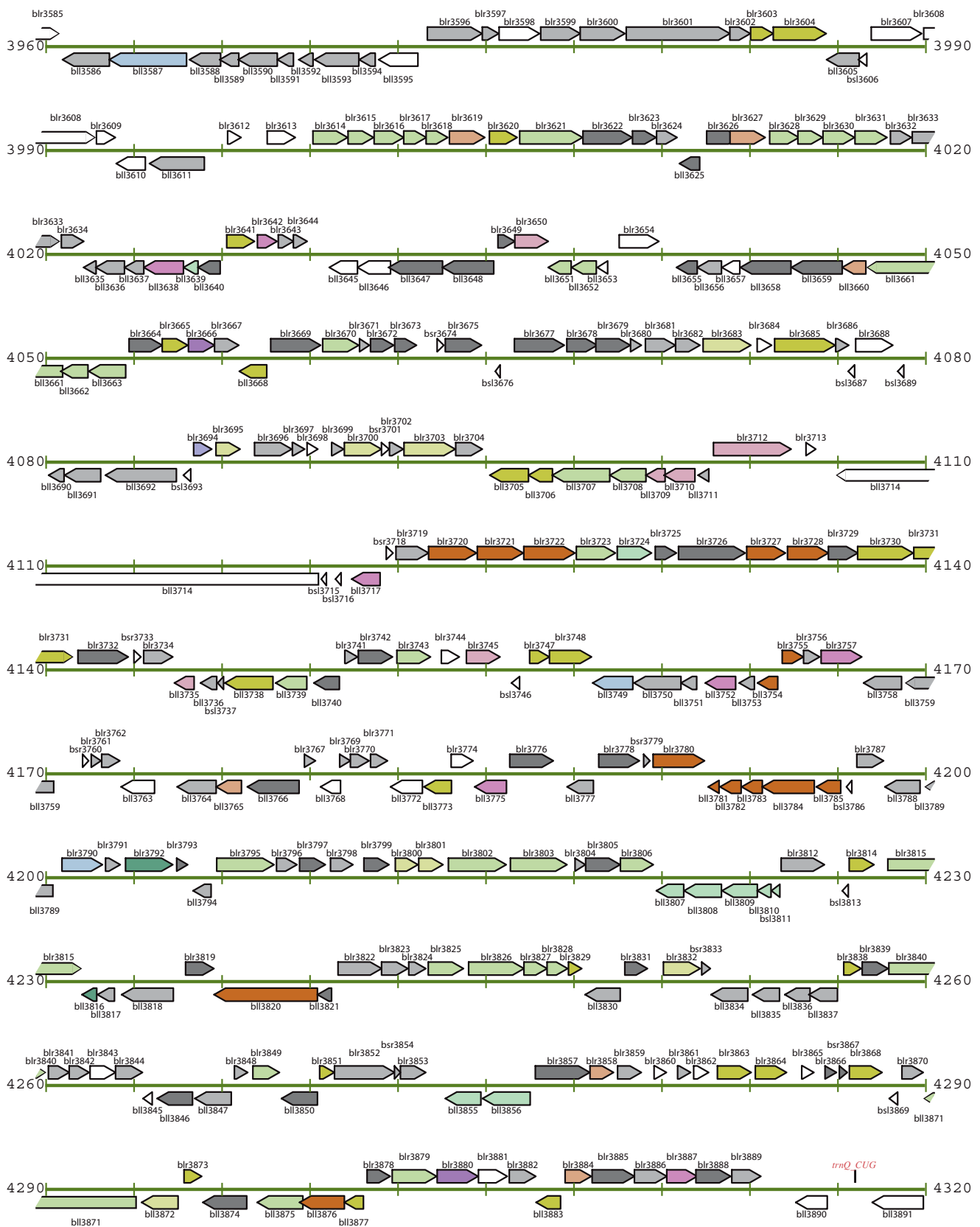


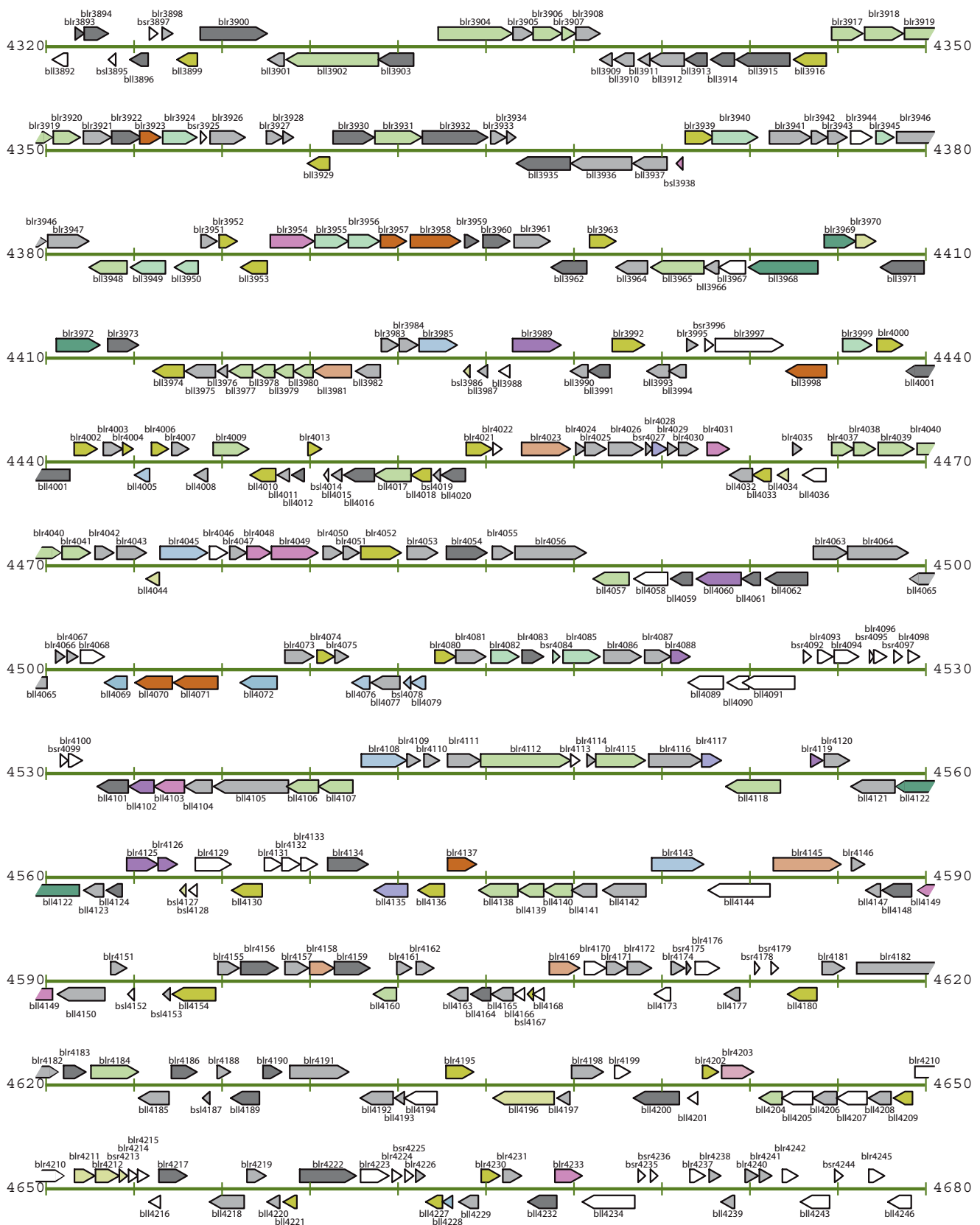


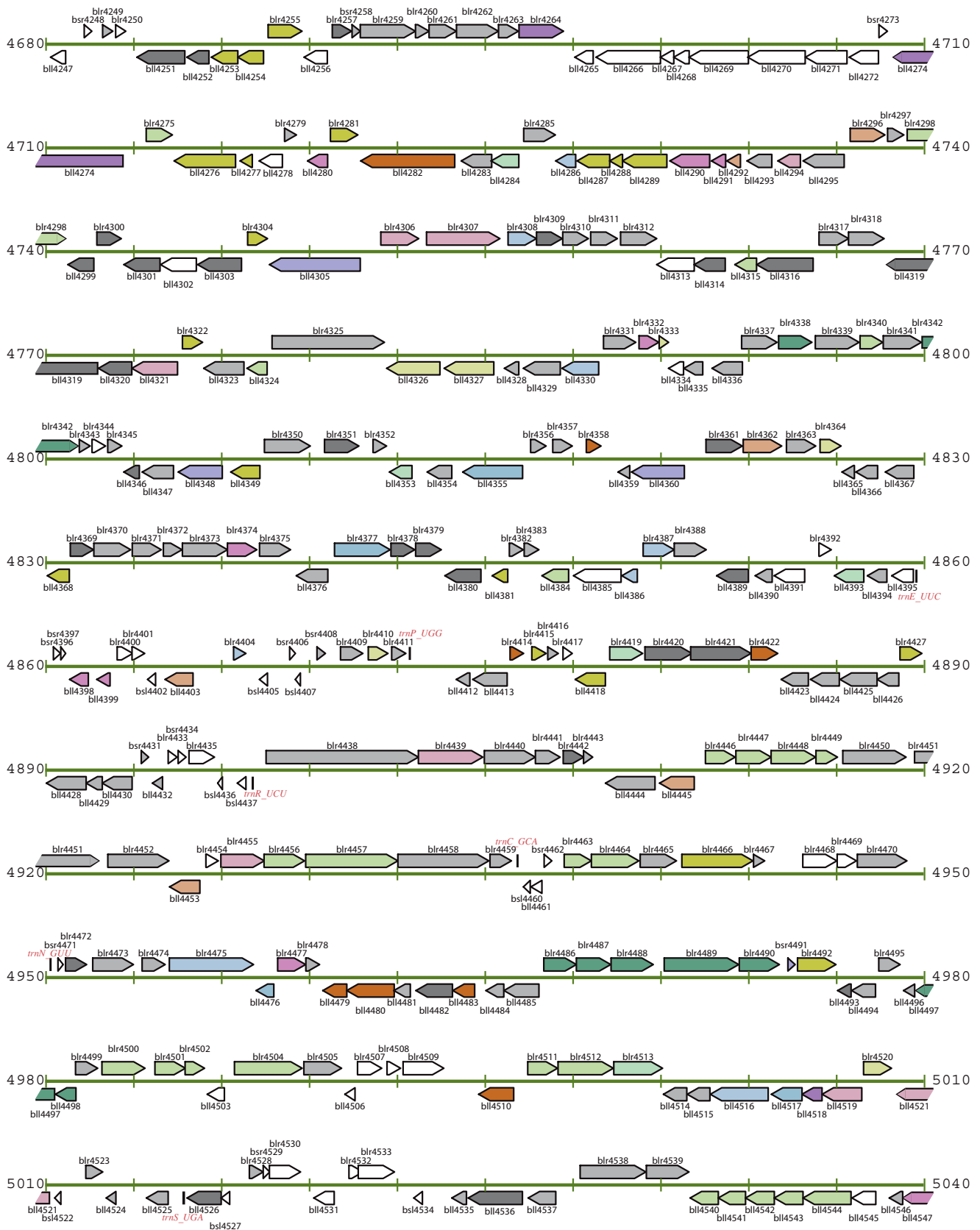


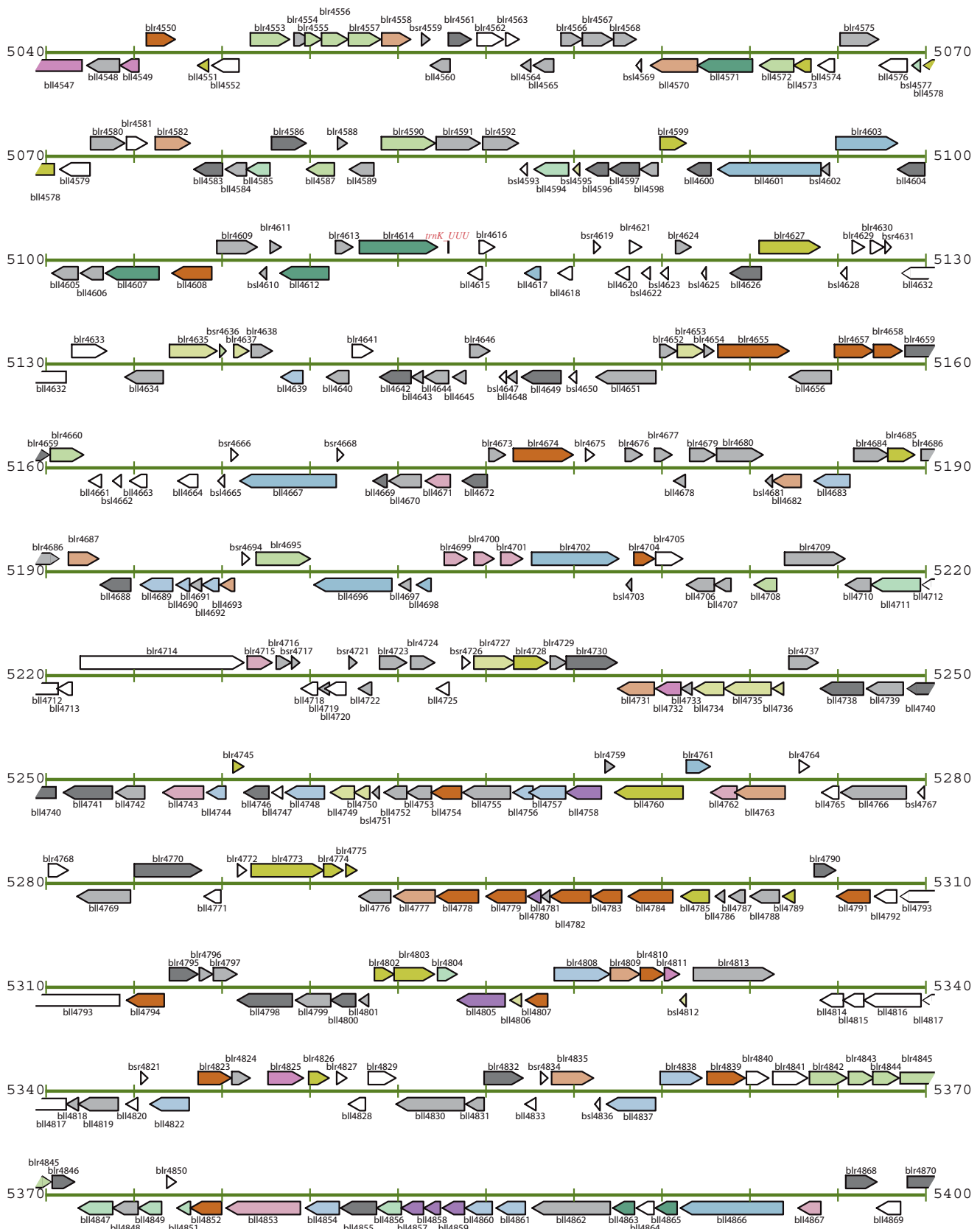


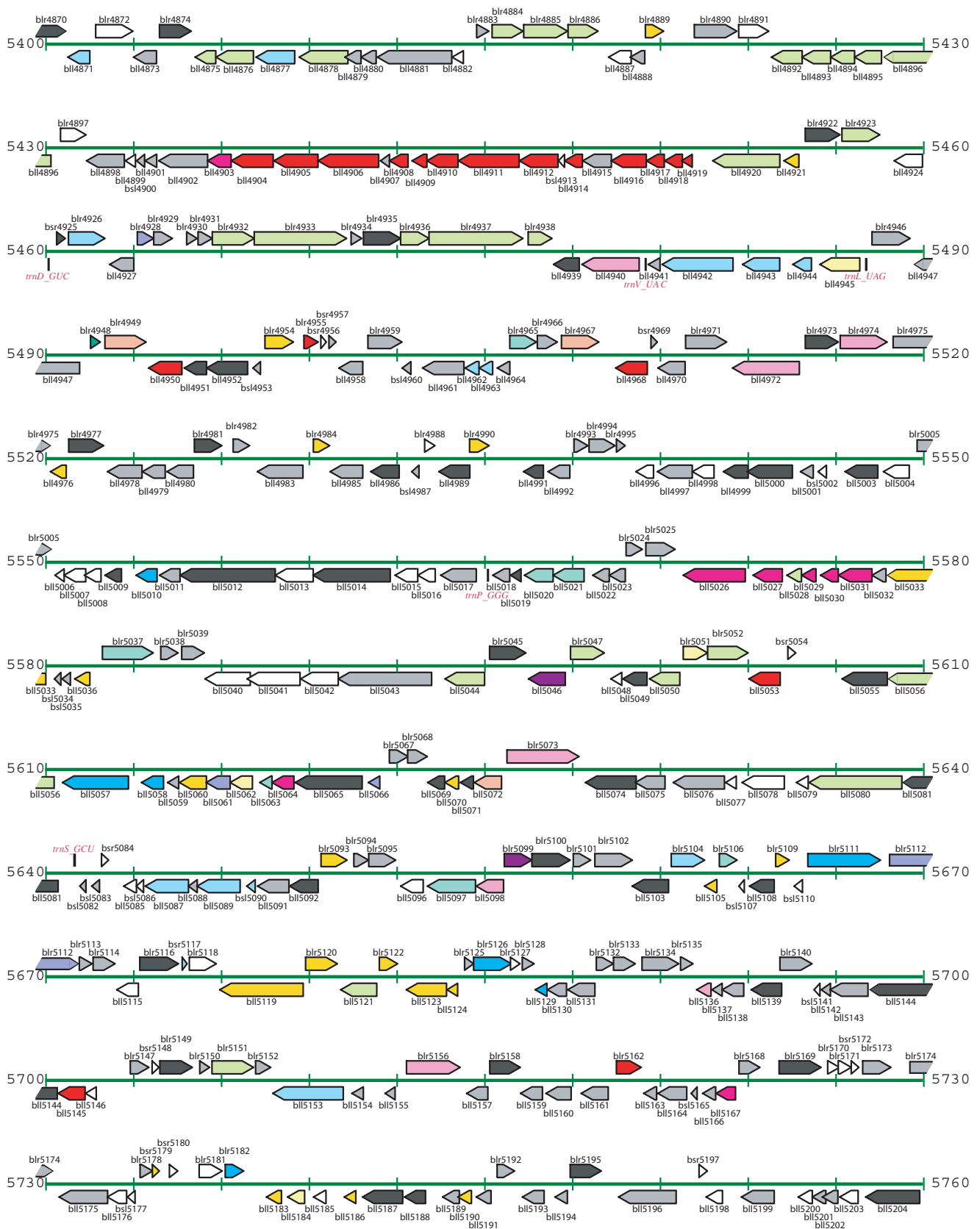


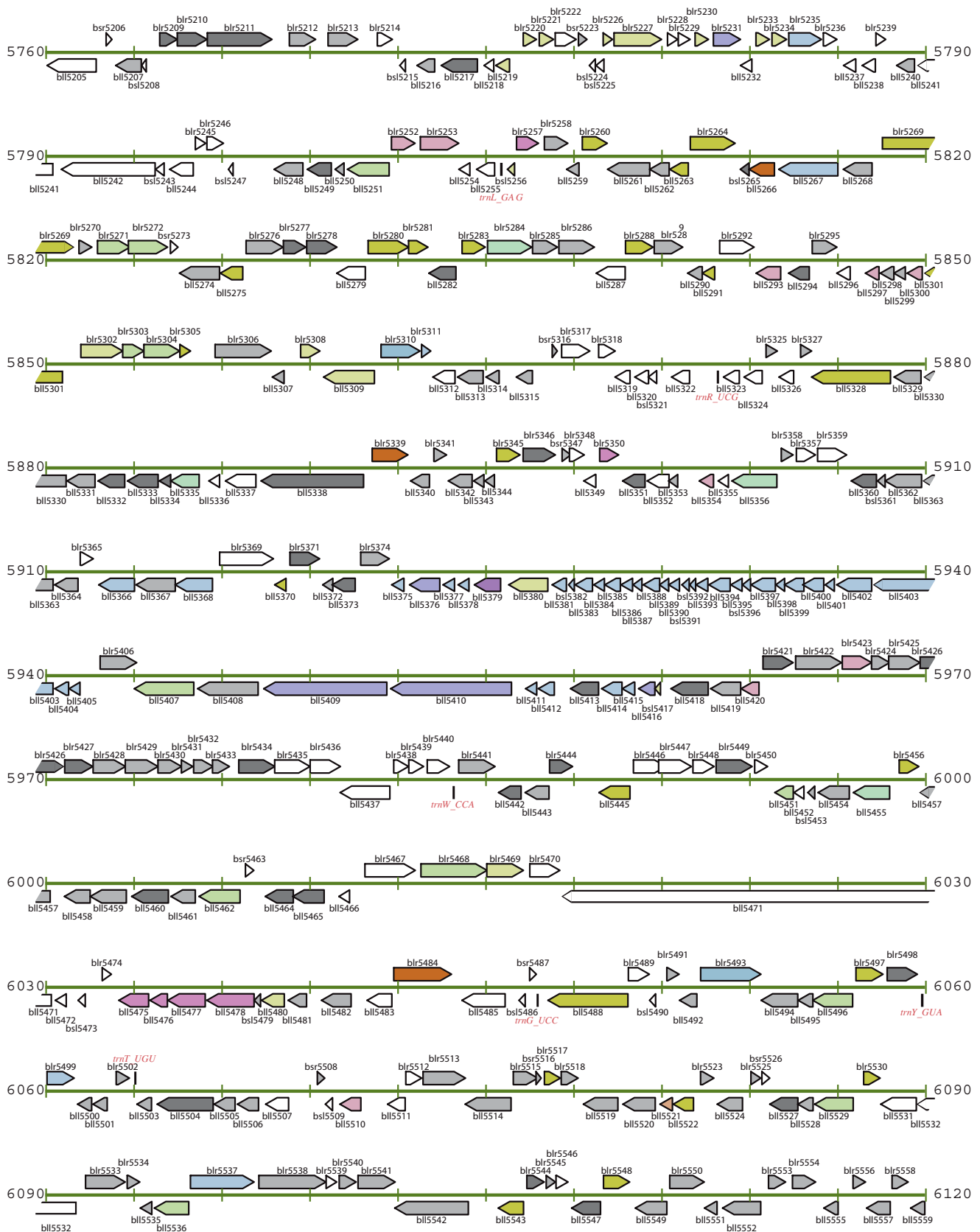


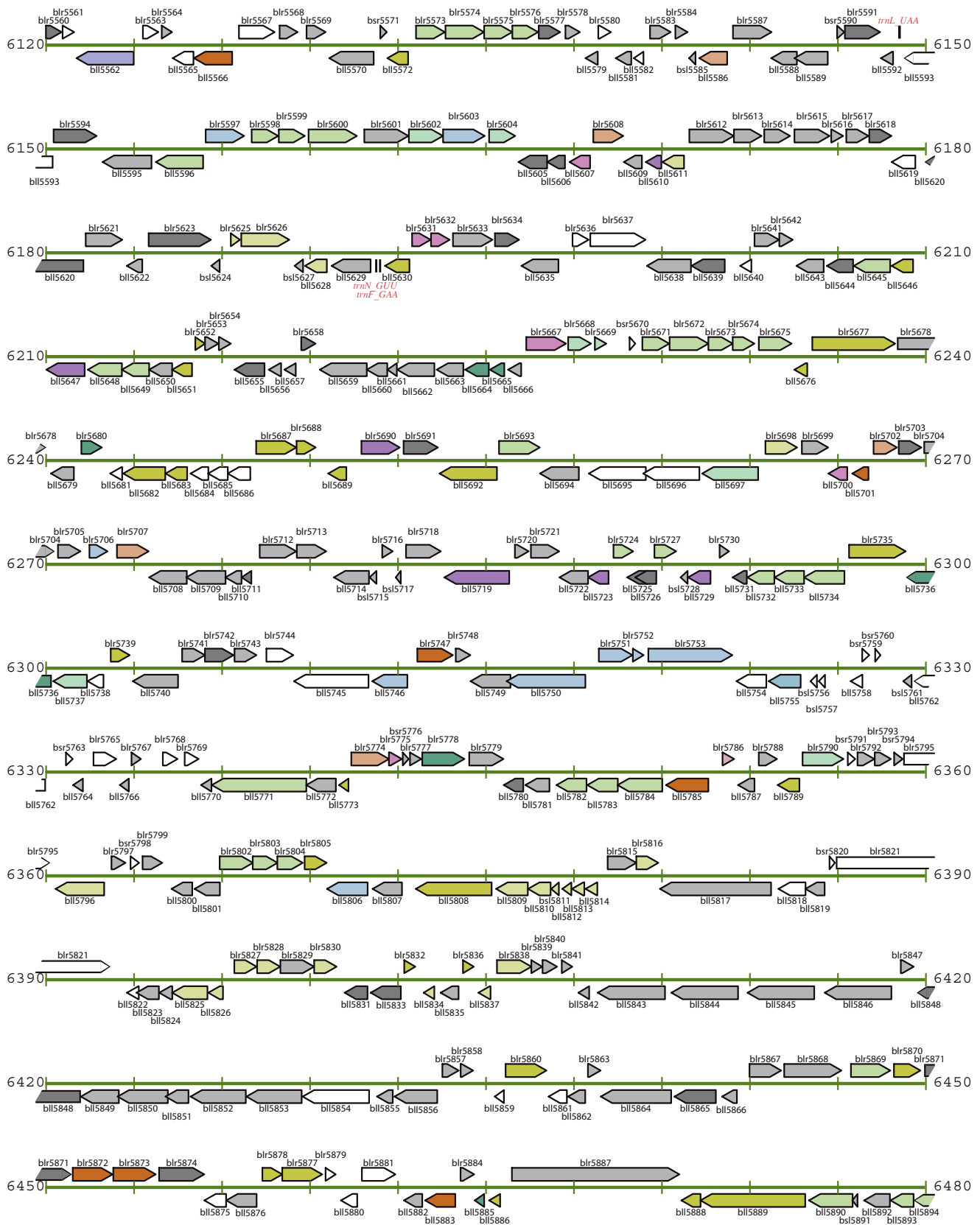


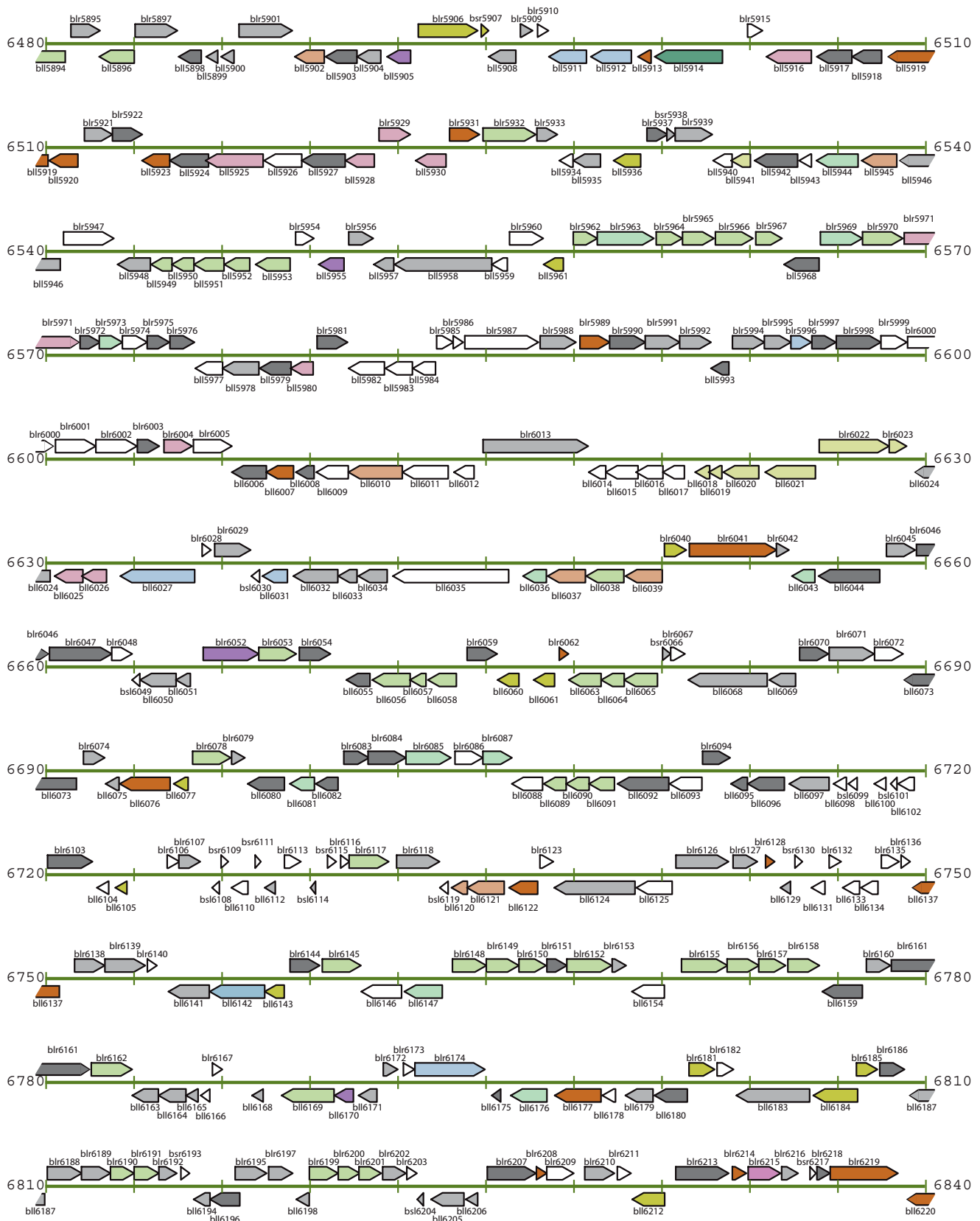


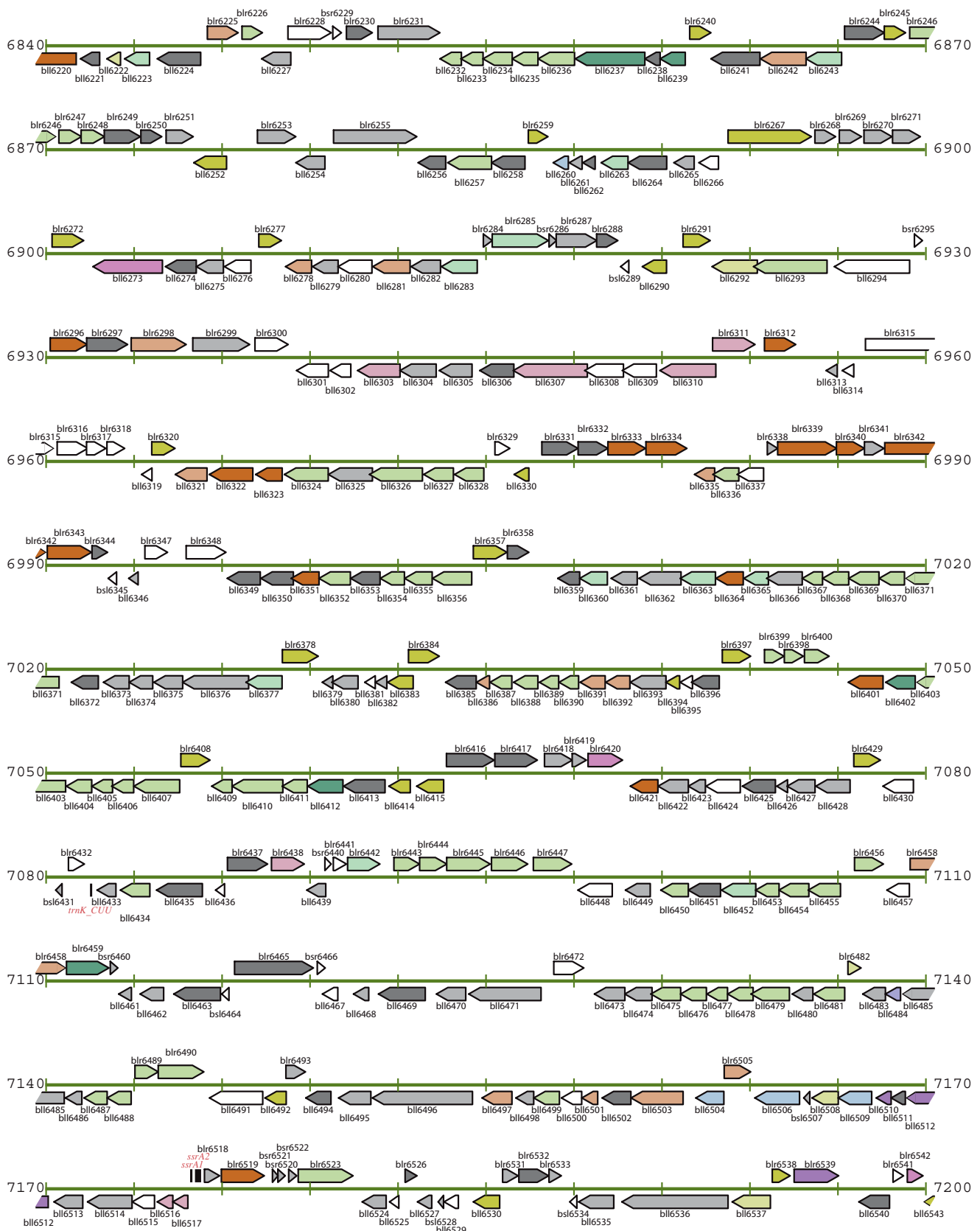


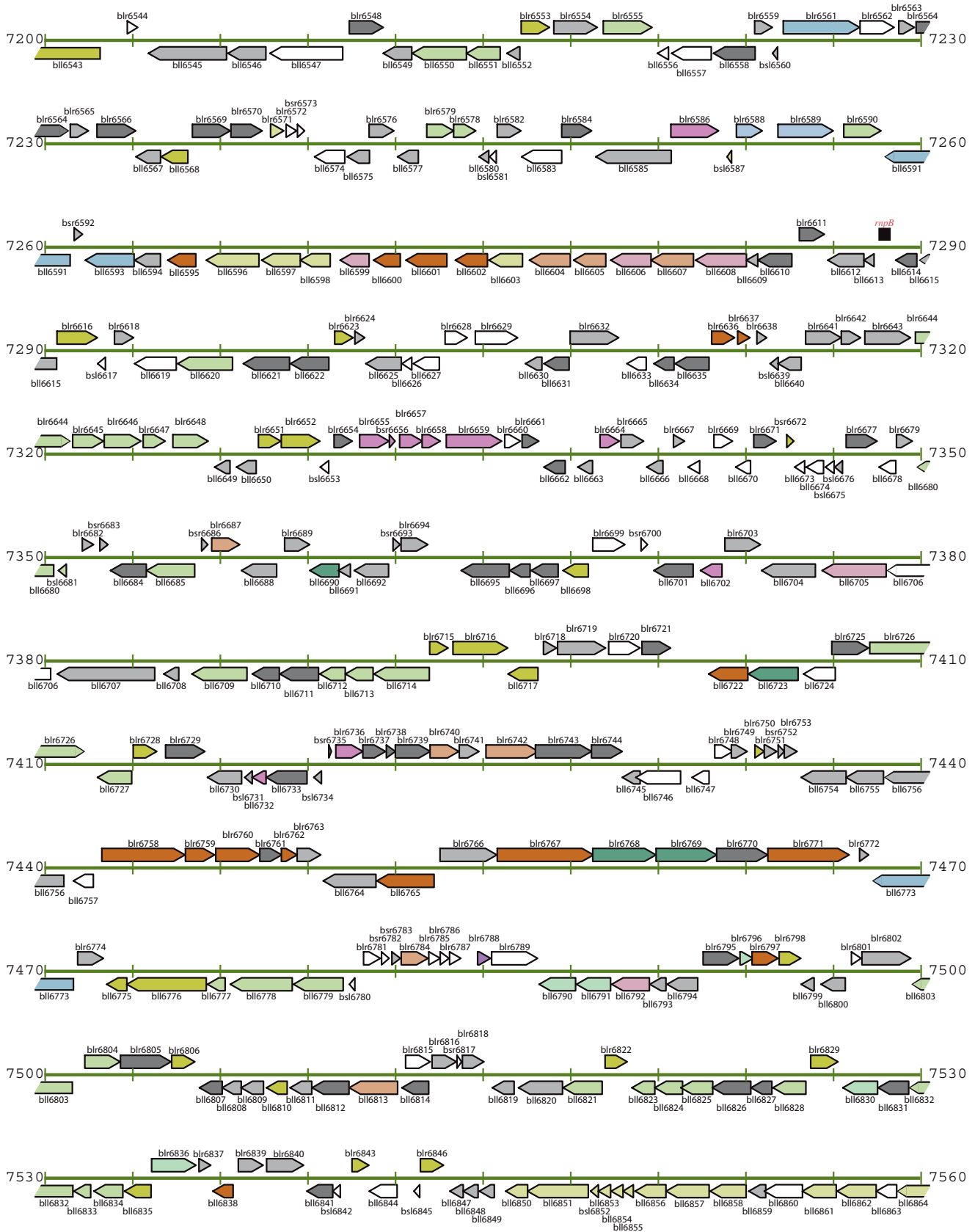


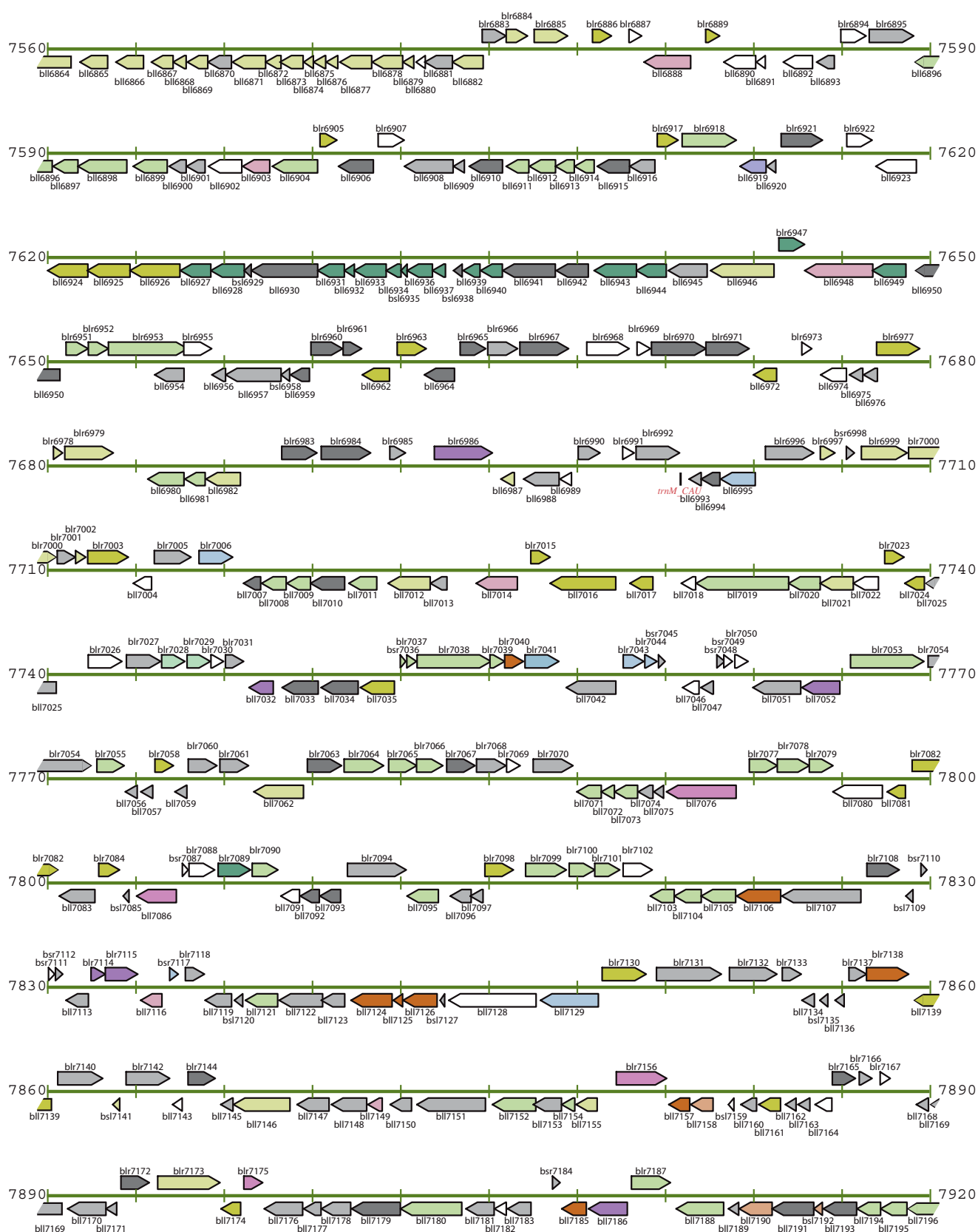


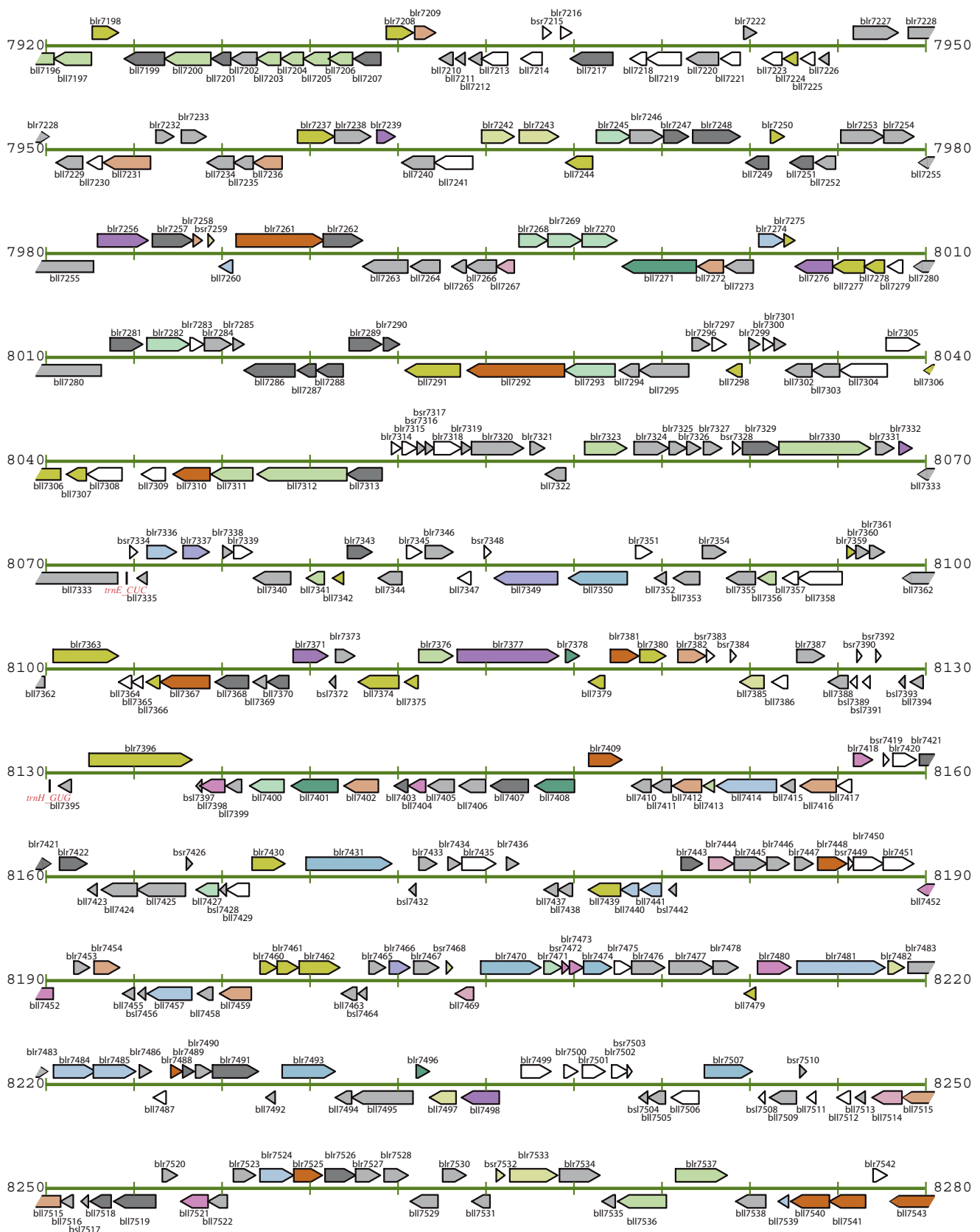


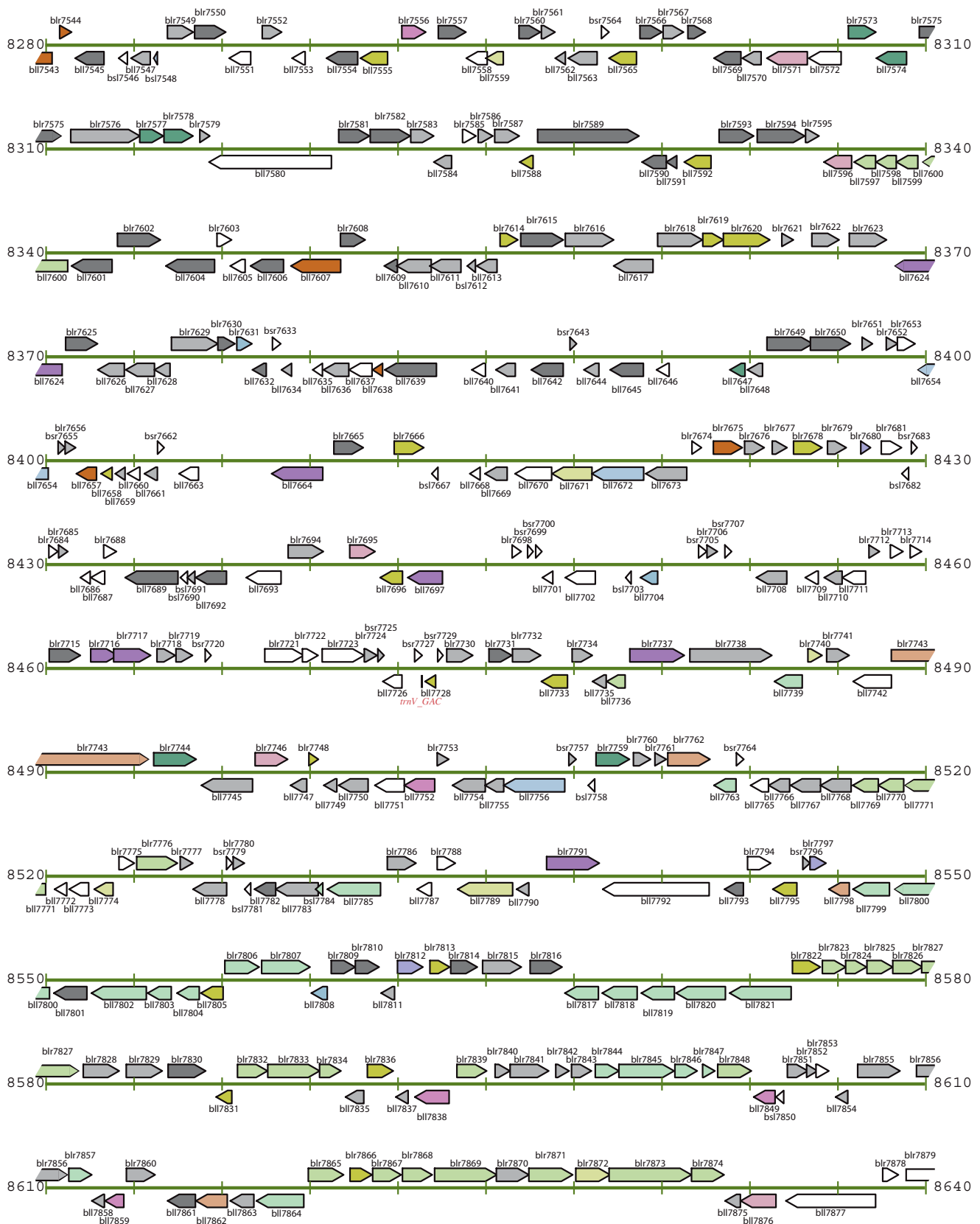


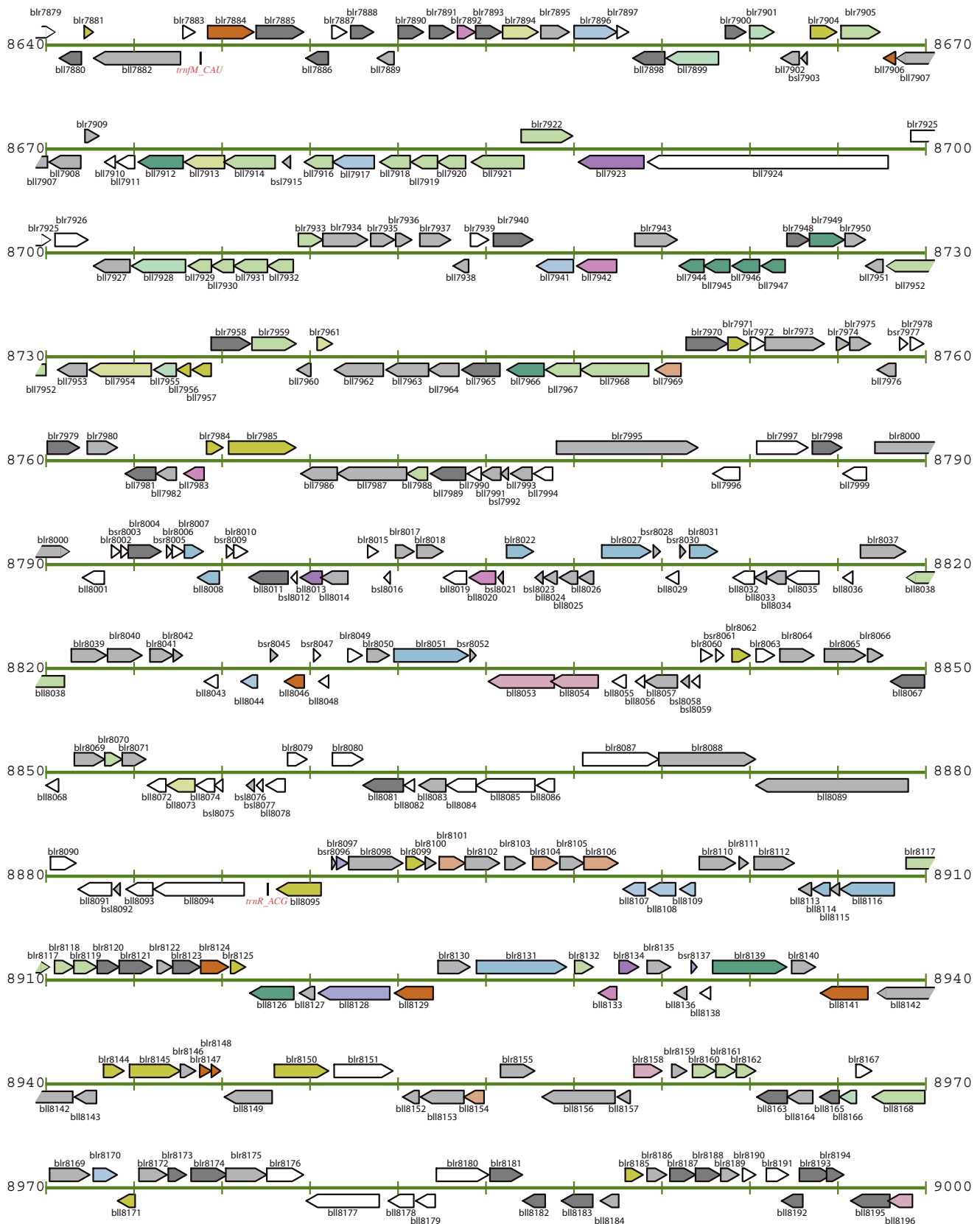












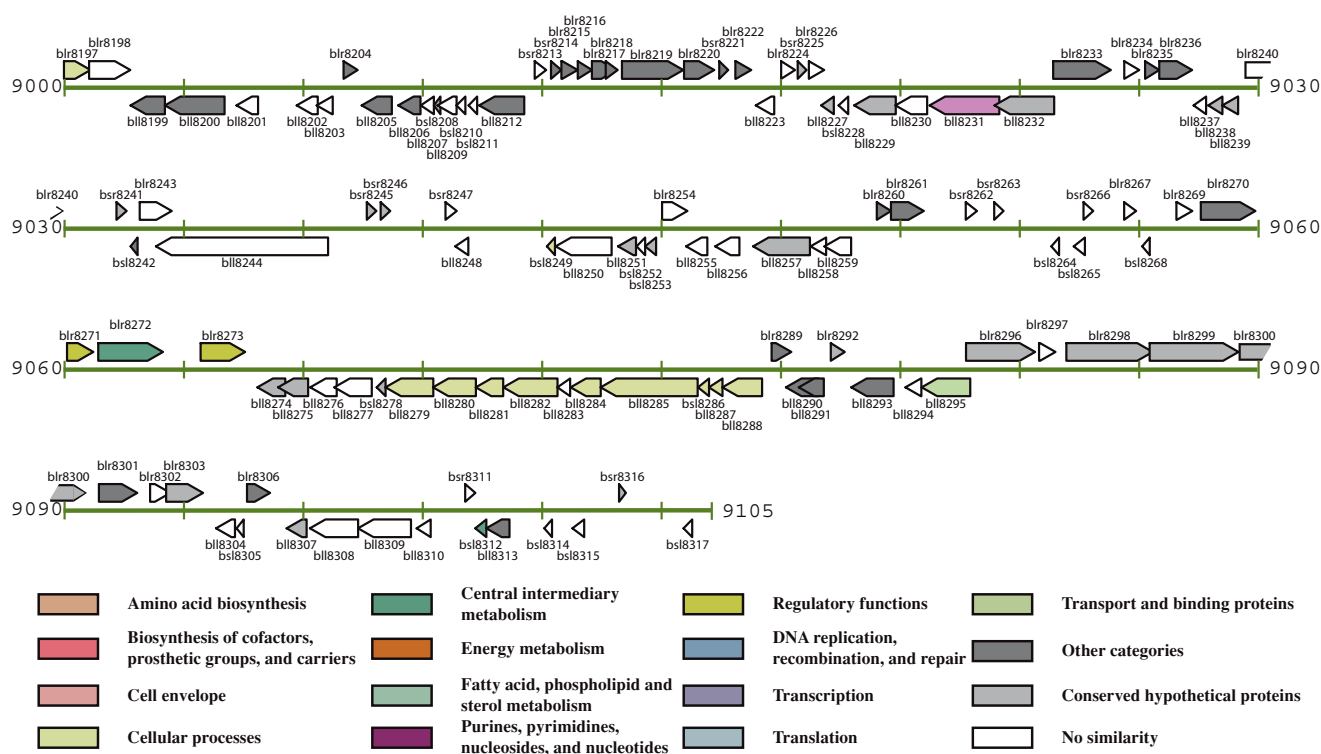


Figure 1. The gene map of the genome of *Bradyrhizobium japonicum* USDA110. The circular chromosome (9,105,828 bp) was opened at the *Pac* I recognition site, and is represented by a linear map starting from this junction (map position 0/100). Green bars show the scale in 3-kb increments with numerals in kb. On both sides of the scale, the potential protein-encoding genes assigned on the basis of computer prediction and the similarity search are shown as boxes with arrowheads indicating the reading direction. The detailed assignment procedures are described in the main article in this issue (pp. 189–pp. 197). The potential genes whose function could be evaluated by similarity search were classified into 14 functional categories, and are represented by different color codes. The RNA-encoding genes are represented by black bars and boxes. The sequences as well as the gene information shown in this paper are available in the Web database, RhizoBase, at <http://www.kazusa.or.jp/rhizobase/>. The sequence data analyzed in this study have been registered in DDBJ/GenBank/EMBL under accession number BA000040.

Aminoacyl Stem	D domain stem loop	stem	Anticodon domain stem loop	stem	Variable region	T _ψ C domain stem loop	stem	Aminoacyl stem
trnR-CCU (122357-122281)	GGTCCC	TA GCTC AGCCGGATA	GAGC	G	GCGGT TTCCTAA ACCGT AGGTC	GGATG	TTCGAGT CATCC	CGGGATC G
trnS-GGA (333607-333518)	GGAGAGA	TG GCC GAGTGGCTTAA	GGC	G	CACGC TTGAAA GCGTG TGTGCGGAAACCGTACC	GTGGG	TTCGAAT CCCAC	TCTCTCC G
trnS-CGA (451390-451301)	GGAGAGG	TG GCA GAGTGGTTGAA	TGC	A	CCGCA CTCGAAA TGCGG CATAGGTGCAACCGCTATC	GGGGG	TTCGAAT CCCTC	CCTCTCC G
trnA-CGC (591251-591326)	GGGGCCA	TA GCTC AGCTGGGA	GAGC	G	CGTGC TTCGAAA TGACG AGGTC	GGCGG	TTCGATC CCGCC	TGGCTCC A
trnT-CGU (604005-603930)	GCCGCAA	TA GCTC AGTCGGTA	GAGC	A	CGTCA TTCGTAA TGACG GGGTC	GGGGG	TTCGAAT CCCTC	TTGCGCC A
trnG-GCC (640199-640125)	GCGGGTG	TA GCTC AGTGGA	GAGC	A	CGACC TTGCCAA GGTCG GGGTC	GAGGG	TTCGAGC CCCTT	CGCCCCG T
trnL-CAG (766675-766761)	GCCCAGG	TG GTG GAATTGGTAGA	CGC	G	CTGGC TTCAGGT GCCAG TGGCTTAACGGCCST	GAAGG	TTCGAGT CCTTT	CCTGGCC A
trnA-GGC (785224-785149)	GGGGCCA	TA GCTC AGCTGGGA	GAGC	G	CTTGC ATGGCAT GCAAG AGGTC	GGCGG	TTCGATC CCGCC	TGGCTCC A
trnR-CCG (924007-923931)	GCACCCG	TA GCTC AGCTGGATA	GAGC	G	TTGCC CTCGGAA GGCAA AGGTC	ACACG	TTCGAAT CGTGT	CGGGTCC G
trnG-CCC (1228026-1228099)	GCGGGCG	TA GTTC AATGGTA	GAAC	G	GCAGC TTCCAA GCTCG ATAC	GAGGG	TTCGATT CCCTT	CGCCCCG T
trnT-GGU (1363162-1363088)	GCTGCCG	TA GCTC AGTGGA	GAGC	A	CTCCA TTGGTAA TGGAG AGGTC	GACAG	TTCAATC CTGTC	TGGCAGC A
trnF-GAA (1374308-1374383)	GCCCAGG	TA GCTC AGTTGGTA	GAGC	A	TGCGA CTCGAAA TCGCA GTGTC	GGTGG	TTCGATT CCGCC	CCTGGCC A
trnL-CAA (1456534-1456450)	GCCCCTG	TG GCG GAACCTGGTAGA	CGC	G	CTCGA CTCAAAA TCGAG TTCGCCAAGGAGT	GCTGG	TTCGATT CCGCC	CAGGGCC A
trnI-GAU (1529942-1530018)	GGGCTTG	TA GCTC AGTTGGTA	GAGC	G	CQCGC TTGATAA GCGTG AGGTC	GGAAG	TTCAAGT CTTC	CAGGCCC A
trnA-UGC (1530052-1530127)	GGGGCCA	TA GCTC AGCTGGGA	GAGC	G	CGTGC TTTGCAA GCAAT AGGTC	GTCCG	TTCGATC CCGTC	TGGCTCC A
trnI-CAU (1754724-1754649)	GGGCTTG	TA GCTC AATGGTA	GAGC	C	GGCCG CTCATAA CGGTC TGGTT	GCAGG	TTCGAGT CCTGC	CGGGCCC A
trnI-CAU (1833703-1833628)	GGGCTTG	TA GCTC AATGGTA	GAGC	C	GGCCG CTCATAA CGGTC TGGTT	GCAGG	TTCGAGT CCTGC	CGGGCCC A
trnV-CAC (2362416-2362342)	GGGGCCA	TA GCTC AGCCGGA	GAGC	G	TTCCC TTCACAC GGGAG AGGTC	CAAGG	TTCGATC CCTTG	TGCGCCC A
trnI-CAU (2497176-2497101)	GGGCTTG	TA GCTC AATGGTA	GAGC	C	GGCCG CTCATAA CGGTC TGGTT	GCAGG	TTCGAGT CCTGC	CGGGCCC A
trnF-CGG (2812727-2812803)	CGGAGTG	TG GCTC AGCCCGGTA	GAGC	A	CTGCG TTCGGGA CGCAG GGGTC	GCAGG	TTCGAAT CCTGC	CACTCCG A
trnQ-UUG (2907070-2907144)	TGGGGCG	TA GCC AGCCGGTAA	GGC	A	GCGGA TTTTGTAT TCCCG CATTC	GGAGG	TTCGATC CCTCC	CGCCCCA G
trnQ-CUG (4317579-4317652)	TGGGAAA	TC GTCT AACGGTA	GCAC	A	ACAGA CTCGTAC TCTGT TTGT	CTTGG	TTCGAAT CAAGG	TTCCCCA G
trnE-UUC (4859741-4859667)	GCTCCCT	TC GTCT ATCGGTA	GGAC	G	CCACC CTTTCAC GGTGG AGAG	AGCCG	TTCGATT CCGCT	AGGGAGC G
trnP-UGG (4872412-4872488)	CGGGGTA	TA GCGC AGCCTGGTA	CGCG	G	GCAGT TTTGGGT ACTCG AGGTC	GTGGG	TTCGAAT CCAGC	TGCCCCG A
trnR-UCU (4897125-4897049)	GGTCCCA	TA GCTC AACTGGATA	GAGT	A	CGGGA TTTCTAC TCCCG GGGTT	GCAGG	TTCGAGT CTGC	TGGGATC G
trnC-GCA (4936093-4936166)	GGCCACG	TG GCG GAGTGGTTA	CGC	GC	CGGT CTCGAAA ACCG TTTAC	TCGGG	TTCGAGT CCCGA	CGTGGCC T
trnN-GUU (4950144-4950218)	TCCTCGG	TA GCTC AGCCGTA	GAGC	A	TCCGA CTGTTAA TCGGA TGGTC	GCTGG	TTCGAAT CCAGC	CGGGGGA G
trnS-UGA (5014774-5014685)	GGAAGGG	TG GCC GAGTGGTTAA	GGC	A	CCGGT CTTGAAA ACCGG CGTGCCTGCAAGGGTACC	GTGGG	TTCGAAT CCCAC	CCCTTCC G
trnK-UUU (5113708-5113783)	GAGCGCG	TA GCTC AGCCGTA	GAGC	A	CGTGA CTTTTAA TCACG GGGTC	CTGGG	TTCGAGC CCCAG	CGCGCTC A
trnD-GUC (5460155-5460079)	GCGGGAG	TA GCTC AGTTGGTA	GAGC	G	CCGGC CTGTAC GCGCG AGGTC	GCGGG	TTCGAGC CCCGT	CTCTCCG G
trnV-UAC (5480560-5480484)	GGGCGGT	TA GCTC AGCTGGTA	GAGC	A	TCTCG TTTACAC CGAGA GGGTC	CGCGG	TTCGAAT CCGTG	ACCGCCC A
trnL-UAG (5488101-5488017)	GCGGGCG	TG GCG GAACCTGGTAGA	CGC	G	CTGGA TTTAGGT TCCAG TGACGAAAGTTGT	GGGGG	TTCGAGT CCCTC	CGCCCCG A
trnP-GGG (5565170-5565093)	CGGAGCG	TG GCGC AGCCCGGTA	GCGC	A	CTAGT CTGGGAG ACTAG GGGTC	GGAGG	TTCAAAT CCTCT	CGCTCCG A
trnS-GCU (5640958-5641050)	GGAGAGG	TG GCC GAGTGGCTGAA	GGC	A	ACGCT TTGCTAA AGCGT CATACGGCTCAAGCTGTATC	GAGGG	TTCGAAT CCCTC	CCTCTCC G
trnL-CAG (5805593-5805509)	GCGCTCG	TG GCG GAACCTGGTAGA	CGC	G	CTGCC TTGAGGT GGCAG TGAGTAAATCGT	GGGGG	TTCGAGT CCCTC	CGAGCCG A
trnR-UCG (5872973-5872897)	GCGCTCG	TA GCTC AGCTGGATA	GAGC	A	TGCGA TTTCCGAT TCCGA GGGTC	GGAGG	TTCGAAT CCTTC	CGAGCCG G
trnW-CCA (5983963-5983888)	AGGAGTG	TA GCTC AATTGGTA	GAGC	A	CCGGT CTCGAAA ACCGG GGGTC	GCAGG	TTCGAGC CTGC	CACTCTC G
trnG-UCC (6046825-6046752)	GCGGGTG	TA GCTC AATGGTA	GAGC	AG	CAGC CTTCCAA GCTG AATAC	GAGGG	TTCGATT CCCTT	CACCCGC T
trnY-GUA (6059945-6059860)	GGGGGAG	TG TCCC GAGTGGCAAA	GGG	A	GCTGA CTGTAAA TCACG CGCCTCATGGCTTC	GCAGG	TTCGAGT CCTGC	CTCCCCC A
trnT-UGU (6063033-6063107)	GCCGGCT	TA GCTC AGCCGTA	GAGC	AG	CGGT TTTGTAA ACCG AAGGTC	GGGGG	TTCAATC CCCTC	AGCCCGC A
trnL-UAA (6149083-6149171)	GCGGGCA	TG GCG GAATTGGTAGA	CGC	A	AGGGA CTTAAAA TCCCT CGGTGCGCAAGCGCTGT	GCCGG	TTCGACC CCGCC	TGCCCGC A
trnN-GUU (6191320-6191247)	CAAGGGA	TA GCTC AGTTGGTA	GAGC	A	GATGA CTGTTAA TCAAT TTGTC	GCGGG	TTCGATT CCGCC	TCCATCG C
trnF-GAA (6191390-6191318)	GCAGGAC	TA GCTC ATGGA	GAGC	G	TCGGG CTGAACG CCCGA AGGC	ATCCG	TTCGATT CCGGT	CTCTTCC A
trnK-CUU (7081599-7081524)	GGGGCCA	TA GCTC AGTTGGTA	GAGC	A	GCTGA CTCTTAA TCACG GGGTC	CCAGG	TTCGAGC CTGCG	TGCGCCC A
trnM-CAU (7701576-7701500)	GCGGGGG	TA GCTC AGCTGGTA	GAGC	A	CGGGA ATCATAA TCCTG GGGTC	GGGGG	TTCGAGT CCCTC	TCCCCTC A
trnE-CUC (8072813-8072739)	GCTCCCT	TC GTCT AGCCGGTA	GGAC	G	CGGCC CTCGCAA GGCTG AAAC	AGGGG	TTCGATT CCCTT	AGGGAGC G
trnH-GUG (8130188-8130112)	GCAGTTG	TA GCTC AGTTGGTA	GAGC	G	CCTGT CTGTGGA ACAGG AGGTC	GGTGG	TTCGAGC CCACC	CAACTGT A
trnV-GAC (8472878-8472804)	GGGCACG	TA GCTC AGCCGGA	GAGC	A	CTACC TTGACAT GGTAG GGGTC	ACAGG	TTCGATC CTGT	CCTGCCC A
trnEM-CAU (8645335-8645259)	CGCGGGG	TG GAGC AGCCCGGTA	GCTC	G	TCAGG CTCATAA CCTGA AGGTC	ATAGG	TTCAAAT CCTAT	CGCCGCA A
trnR-ACG (8887625-8887549)	GCGCCCG	TA GCTC AGCTGGATA	GAGC	A	TCAGA CTCGAAA TCTGA GGGTC	GGACG	TTCGAAT CGTTC	CGGGCCG G

Figure 2. Structure of the tRNA genes in the genome of *B. japonicum*. The nucleotide sequences, nucleotide positions and the structural domains for the 50 tRNA genes in the genome are tabulated.

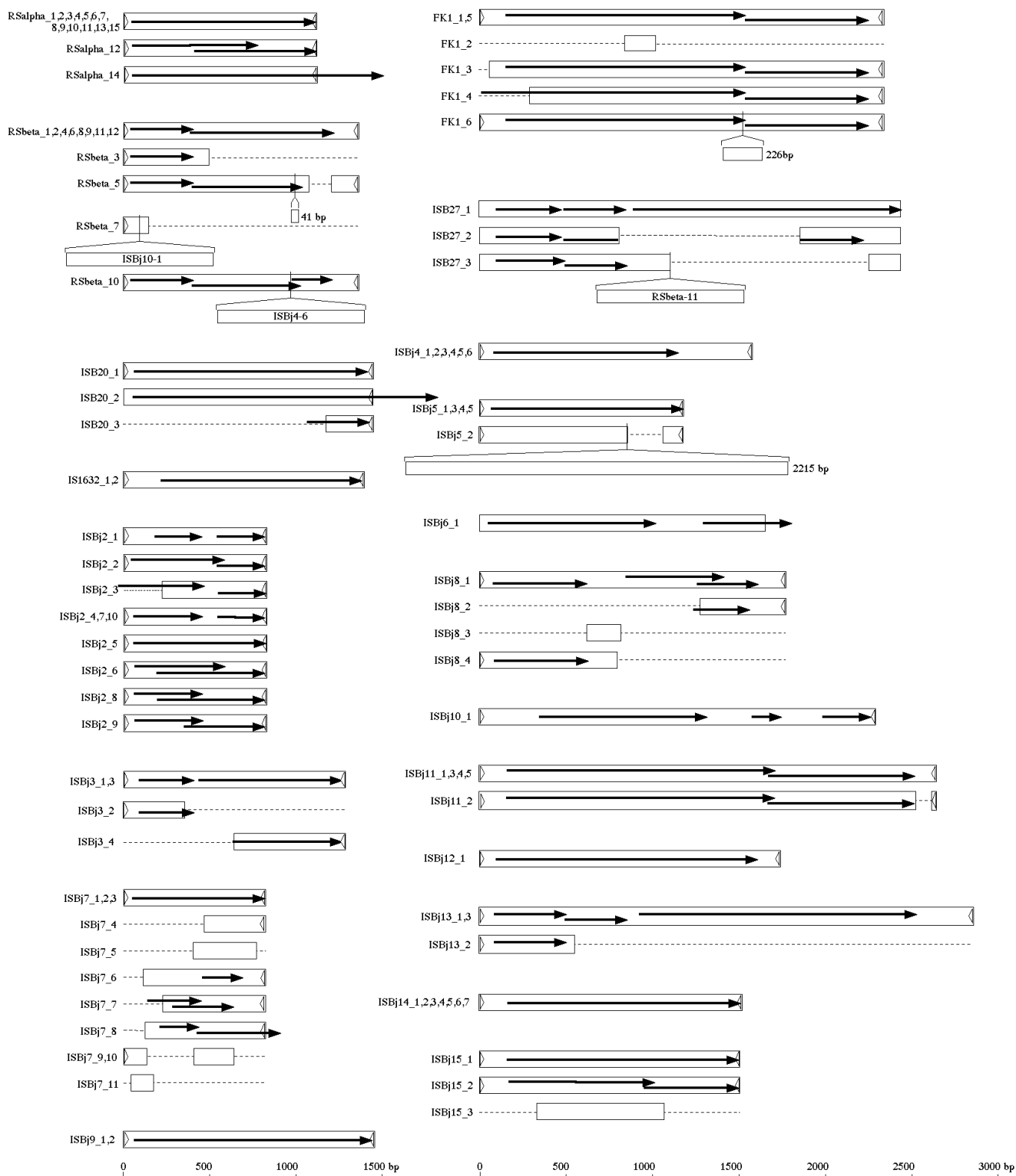


Figure 3. Structure of insertion sequences. The structures of insertion sequences found in the *B. japonicum* genome are depicted as open boxes. Horizontal broken lines show the deleted regions. Open triangles at the ends of the open boxes represent the terminal inverted repeats. Horizontal arrows indicate the positions of genes for putative transposases. The positions of insertions are indicated as vertical bars with open boxes.

Table 1. Codon usage frequency for the *B. japonicum* genome. The frequency of the usage of each codon in the chromosome is shown as permillage of overall counts. Numerals in parentheses indicate the actual number of the codons.

UUU F	6.5 (17263)	UCU S	2.0 (5317)	UAU Y	10.2 (27071)	UGU C	1.3 (3537)
UUC F	30.7 (81106)	UCC S	12.8 (33935)	UAC Y	11.8 (31127)	UGC C	7.8 (20715)
UUA L	0.7 (1875)	UCA S	2.8 (7526)	UAA *	0.5 (1340)	UGA *	1.8 (4730)
UUG L	8.5 (22439)	UCG S	21.4 (56669)	UAG *	0.9 (2247)	UGG W	13.1 (34723)
CUU L	7.8 (20633)	CCU P	3.7 (9855)	CAU H	9.2 (24305)	CGU R	6.6 (17570)
CUC L	34.5 (91273)	CCC P	14.1 (37353)	CAC H	11.4 (30119)	CGC R	39.8 (105051)
CUA L	1.6 (4240)	CCA P	3.4 (8930)	CAA Q	5.7 (15151)	CGA R	4.2 (11080)
CUG L	45.2 (119368)	CCG P	31.2 (82485)	CAG Q	26.1 (68871)	CGG R	15.2 (40248)
AUU I	6.3 (16538)	ACU T	2.5 (6558)	AAU N	8.6 (22608)	AGU S	2.2 (5824)
AUC I	44.8 (118461)	ACC T	28.0 (73936)	AAC N	18.9 (50070)	AGC S	16.6 (43933)
AUA I	1.4 (3660)	ACA T	3.2 (8533)	AAA K	5.9 (15570)	AGA R	1.7 (4427)
AUG M	23.8 (62797)	ACG T	20.4 (53809)	AAG K	30.2 (79816)	AGG R	4.3 (11479)
GUU V	6.1 (16052)	GCU A	7.6 (20128)	GAU D	18.3 (48252)	GGU G	8.3 (22013)
GUC V	37.3 (98480)	GCC A	52.3 (138210)	GAC D	36.2 (95729)	GGC G	58.2 (153914)
GUA V	2.0 (5388)	GCA A	10.8 (28623)	GAA E	15.6 (41303)	GGA G	7.1 (18733)
GUG V	29.3 (77305)	GCG A	53.0 (139975)	GAG E	36.8 (97131)	GGG G	9.6 (25256)

Table 2. Codon-anticodon recognition pattern for the *B. japonicum* genome. Corresponding amino acids and tRNA species are given as a one-letter abbreviation and italic letters. Numerals in parentheses indicate the number of the genes on the chromosome.

UUU F	<i>trnF-GAA</i> ⁽²⁾	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	<i>trnR-ACG</i>
CUC L		CCC P		CAC H		CGC R	
CUA L	<i>trnL-UAG</i>	CCA P	<i>trnP-UGG</i>	CAA Q	<i>trnQ-UUG</i>	CGA R	<i>trnR-UCG</i>
CUG L	<i>trnL-CAG</i>	CCG P	<i>trnP-CGG</i>	CAG Q	<i>trnQ-CUG</i>	CGG R	<i>trnR-CCG</i>
AUU I	<i>trnI-GAU</i>	ACU T	<i>trnT-GGU</i>	AAU N	<i>trnN-GUU</i> ⁽²⁾	AGU S	<i>trnS-GCU</i>
AUC I		ACC T		AAC N		AGC S	
AUA I	<i>trnI-CAU</i> ⁽³⁾	ACA T	<i>trnT-UGU</i>	AAA K	<i>trnK-UUU</i>	AGA R	<i>trnR-UCU</i>
AUG M	<i>trnM-CAU</i> <i>trnM-CAU</i>	ACG T	<i>trnT-CGU</i>	AAG K	<i>trnK-CUU</i>	AGG R	<i>trnR-CCU</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	<i>trnE-CUC</i>	GGG G	<i>trnG-CCC</i>

Table 3. List of insertion sequences.

name	length (bp)	direction	position	name	length (bp)	direction	position
RSα1	1116	>>	1860456 to 1861571	ISBj3_1	1284	>>	1683969 to 1685252
RSα2	1116	>>	1883172 to 1884287	ISBj3_2	356	<<	1837736 to 1838091
RSα3	1116	>>	1904907 to 1906022	ISBj3_3	1284	>>	4320931 to 4322214
RSα4	1116	>>	1962375 to 1963490	ISBj3_4	645	<<	4322817 to 4323461
RSα5	1116	>>	1979002 to 1980117				
RSα6	1116	>>	2066180 to 2067295	ISBj4_1	1585	<<	5122910 to 5124494
RSα7	1116	>>	2244775 to 2245890	ISBj4_2	1584	>>	5747864 to 5749447
RSα8	1116	>>	2338099 to 2339214	ISBj4_3	1587	>>	6668584 to 6670170
RSα9	1116	<<	4666443 to 4667558	ISBj4_4	1585	<<	8289146 to 8290730
RSα10	1116	<<	5141398 to 5142513	ISBj4_5	1585	>>	8460055 to 8461639
RSα11	1116	>>	5397253 to 5398368	ISBj4_6	1586	<<	8987162 to 8988747
RSα12	1117	<<	6289836 to 6290952				
RSα13	1116	>>	8285049 to 8286164	ISBj5_1	1183	>>	21909 to 23091
RSα14	1114	>>	9024824 to 9025937	ISBj5_2	978	>>	1975539 to 1976396
RSα15	1116	<<	9079784 to 9080899				...join... 1978610 to 1978729
				ISBj5_3	1183	>>	2702540 to 2703722
RSβ1	1364	>>	14651 to 16014	ISBj5_4	1183	>>	8293173 to 8294355
RSβ2	1364	<<	1764678 to 1766041	ISBj5_5	1183	>>	8985099 to 8986281
RSβ3	495	>>	1845813 to 1846307				
RSβ4	1364	>>	1897147 to 1898510	ISBj6_1	1656	>>	9090881 to 9092536
RSβ5	1232	<<	1906041 to 1906198				
			...join... 1906200 to 1906281	ISBj7_1	813	>>	1812994 to 1813807
			...join... 1906323 to 1907314	ISBj7_2	823	<<	1834493 to 1835316
RSβ6	1364	>>	2079936 to 2081299	ISBj7_3	815	<<	1841201 to 1842016
RSβ7	148	>>	2102124 to 2102213	ISBj7_4	365	>>	1863302 to 1863667
			...join... 2104519 to 2104576	ISBj7_5	507	>>	2057308 to 2057815
RSβ8	1364	<<	2292940 to 2294303	ISBj7_6	707	<<	2063237 to 2063944
RSβ9	1364	<<	8330204 to 8331567	ISBj7_7	592	>>	2126971 to 2127563
RSβ10	1363	>>	9012887 to 9013858	ISBj7_8	707	<<	9009467 to 9010174
			...join... 9016459 to 9016849	ISBj7_9	371	>>	1854166 to 1854304
RSβ11	1364	>>	9027142 to 9028505				...join... 1854304 to 1854537
RSβ12	1364	>>	9050401 to 9051764	ISBj7_10	371	<<	9031638 to 9031871
							...join... 9031871 to 9032009
FK1_1	2343	<<	1801035 to 1803377	ISBj7_11	130	<<	2061993 to 2062123
FK1_2	175	>>	1814969 to 1815143				
FK1_3	2284	>>	1846314 to 1848597	ISBj8_1	1776	<<	2014819 to 2016594
FK1_4	2051	<<	1874302 to 1876352	ISBj8_2	496	>>	2329929 to 2330424
FK1_5	2342	>>	1984494 to 1986835	ISBj8_3	195	>>	9006817 to 9007011
FK1_6	2366	>>	9013859 to 9015407	ISBj8_4	806	<<	9004162 to 9004967
			...join... 9015634 to 9016450	ISBj9_1	1454	>>	9058516 to 9059969
IS1632_1	1395	<<	1937403 to 1938797	ISBj9_2	1454	<<	5740825 to 5742278
IS1632_2	1395	>>	1876354 to 1877748	ISBj10_1	2298	<<	2102214 to 2104511
ISB20_1	1447	<<	1806977 to 1808423	ISBj11_1	2649	>>	1870866 to 1873514
ISB20_2	1437	>>	2071254 to 2072690	ISBj11_2	2555	<<	2051930 to 2051956
ISB20_3	273	>>	9007146 to 9007418				...join... 2053700 to 2056228
ISB27_1	2440	>>	1856563 to 1859002	ISBj11_3	2649	<<	2155260 to 2157908
ISB27_2	1384	>>	2019241 to 2020624	ISBj11_4	2649	<<	3891046 to 3893694
ISB27_3	1286	<<	9026963 to 9027141	ISBj11_5	2649	<<	5543048 to 5545696
			...join... 9028506 to 9029612	ISBj12_1	1743	<<	2051957 to 2053699
ISBj2_1	832	<<	1754822 to 1755653	ISBj13_1	2865	>>	1940519 to 1943383
ISBj2_2	830	>>	1806143 to 1806972	ISBj13_2	555	>>	2314664 to 2315218
ISBj2_3	617	>>	1819675 to 1820291	ISBj13_3	2865	<<	2331891 to 2334755
ISBj2_4	832	>>	1839309 to 1840140				
ISBj2_5	830	>>	1884292 to 1885121	ISBj14_1	1524	<<	2118632 to 2120155
ISBj2_6	831	<<	1974704 to 1975534	ISBj14_2	1524	<<	2161005 to 2162528
ISBj2_7	832	>>	2027184 to 2028015	ISBj14_3	1524	>>	2202550 to 2204073
ISBj2_8	828	>>	2068499 to 2069326	ISBj14_4	1524	>>	3958041 to 3959564
ISBj2_9	831	>>	2113280 to 2114110	ISBj14_5	1524	<<	5146231 to 5147754
ISBj2_10	832	>>	4286526 to 4287357	ISBj14_6	1524	<<	7370865 to 7372388
				ISBj14_7	1524	>>	8715111 to 8716634
				ISBj15_1	1509	<<	1982676 to 1984184
				ISBj15_2	1507	<<	2131756 to 2133262
				ISBj15_3	738	<<	8995098 to 8995835

Table 4. Structural features of insertion sequences.

name	size (bp)	inverted repeat (bp)	direct repeat (bp)	family	number of full-length copies	number of partial copies
RS α	1116	5	4 ^{a)}	IS630 family	15	0
RS β	1364	13	3	IS3 family	9	3
FK1	2342	48	6 or 8	IS21 family	3	3
IS1632	1395	23	4 or 9	IS256 family	2	0
ISB20	1447	15	4 ^{a)}	IS4 family	1	2
ISB27	2440	0	8	IS66 family	1	2
ISBj2	832	4	4 ^{a)}	IS3 family	9	1
ISBj3	1284	26	3	IS3 family	2	2
ISBj4	1585	25	2 ^{b)}	IS110 family	6	0
ISBj5	1183	3	4 ^{a)}	IS630 family	4	1
ISBj6	1656	0	2	unknown	1	0
ISBj7	823	16	0	IS6 family	3	8
ISBj8	1776	14	4 ^{a)}	IS1380 family	1	3
ISBj9	1454	15	4 ^{a)}	IS4 family	2	0
ISBj10	2298	13	7	IS21 family	1	0
ISBj11	2649	25	6 or 7	IS21 family	4	1
ISBj12	1743	10	0	unknown	1	0
ISBj13	2865	28	8	IS66 family	2	1
ISBj14	1524	16	6	IS5 family	7	0
ISBj15	1509	16	6	IS5 family	2	1

^{a)}These recognized “CTAG” as the target sequence, except for RS α 9, RS α 12, ISBj9-1.

^{b)}All of couple of bases was “TC.”

Table 5. List of DNA elements inserted in tRNA genes.

name	length (bp)	GC contents	position	duplication (bp)	tRNA gene
trnR1 element	30529	56.5%	924008 to 954536	40 ^{a)}	<i>trnR</i> -CCG ^{c)}
trnF1 element	9063	54.7%	1374384 to 1383446	17	<i>trnF</i> -GAA
trnI1 element	51727	59.5%	1702922 to 1754648	49	<i>trnI</i> -CAU copy3 ^{c)}
trnI2 element	23912	59.4%	2473189 to 2497100	49	<i>trnI</i> -CAU copy1 ^{c)}
trnQ1 element	6144	59.5%	4317653 to 4323796	17	<i>trnQ</i> -CUG
trnS1 element	8796	60.3%	5014775 to 5023570	32	<i>trnS</i> -UGA
trnK1 element	48532	60.2%	5113784 to 5162315	45	<i>trnK</i> -UUU
trnP1 element	17798	57.1%	5547295 to 5565092	45	<i>trnP</i> -GGG
trnR2 element	4214	59.5%	5868683 to 5872896	48	<i>trnR</i> -UCG
trnK2 element	34462	61.1%	7047062 to 7081523	45 ^{b)}	<i>trnK</i> -CUU
trnK3 element	52391	62.5%	6994626 to 7047016	45	<i>trnK</i> -CUU
trnK4 element	31318	62.1%	6963263 to 6994580	45 ^{b)}	<i>trnK</i> -CUU
trnM1 element	81959	62.4%	7619541 to 7701499	16	<i>trnM</i> -CAU
trnR3 element	96530	62.6%	8791019 to 8887548	51	<i>trnR</i> -ACG

^{a)}The duplication stretched more 43 bp upstream of gene. ^{b)}The patterns contained 1 base of mismatch either duplication. ^{c)}The elements occurred to duplicate at 5'-end of tRNA genes.