Cell Reports, Volume *29*

Supplemental Information

Genome-wide Identification and Characterization of a Superfamily of Bacterial Extracellular Contractile Injection Systems

Lihong Chen, Nan Song, Bo Liu, Nan Zhang, Nabil-Fareed Alikhan, Zhemin Zhou, Yanyan Zhou, Siyu Zhou, Dandan Zheng, Mingxing Chen, Alexia Hapeshi, Joseph Healey, Nicholas R. Waterfield, Jian Yang, and Guowei Yang

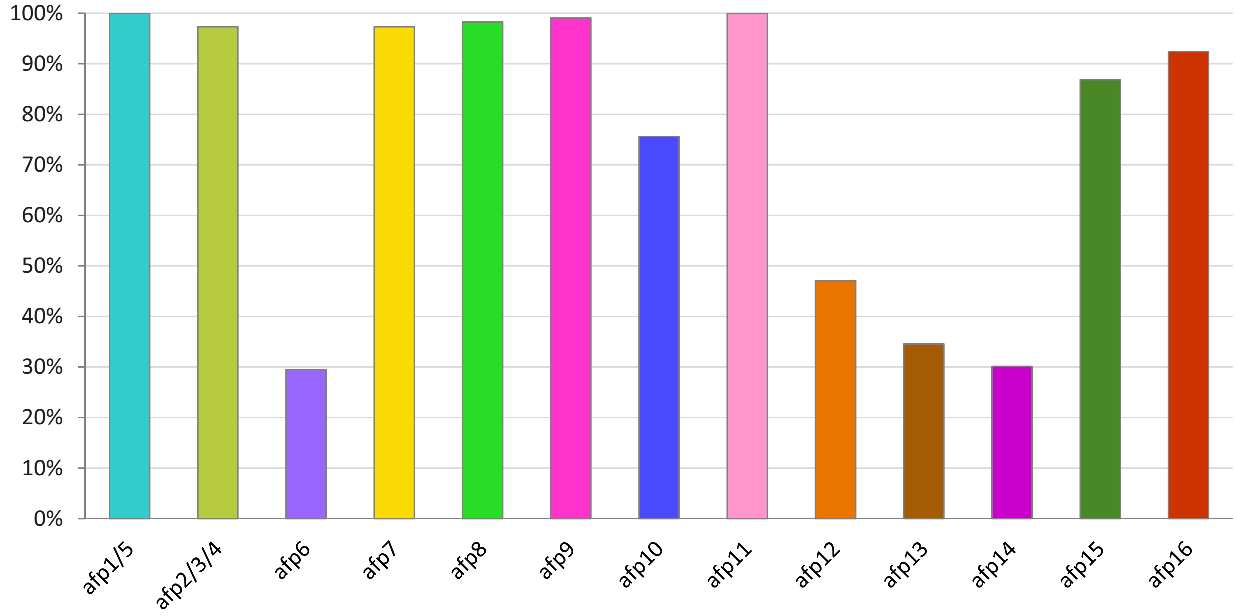


Figure S1. **Proportion of eCIS-like loci encoding each of recognized eCIS subunit components, related to Figure 2.** A column chart of the proportion of loci that encode each of the known component subunits among all 631 loci. Multiple copies within the same loci are counted only once. The color code is consistent with that used in Figure 1.

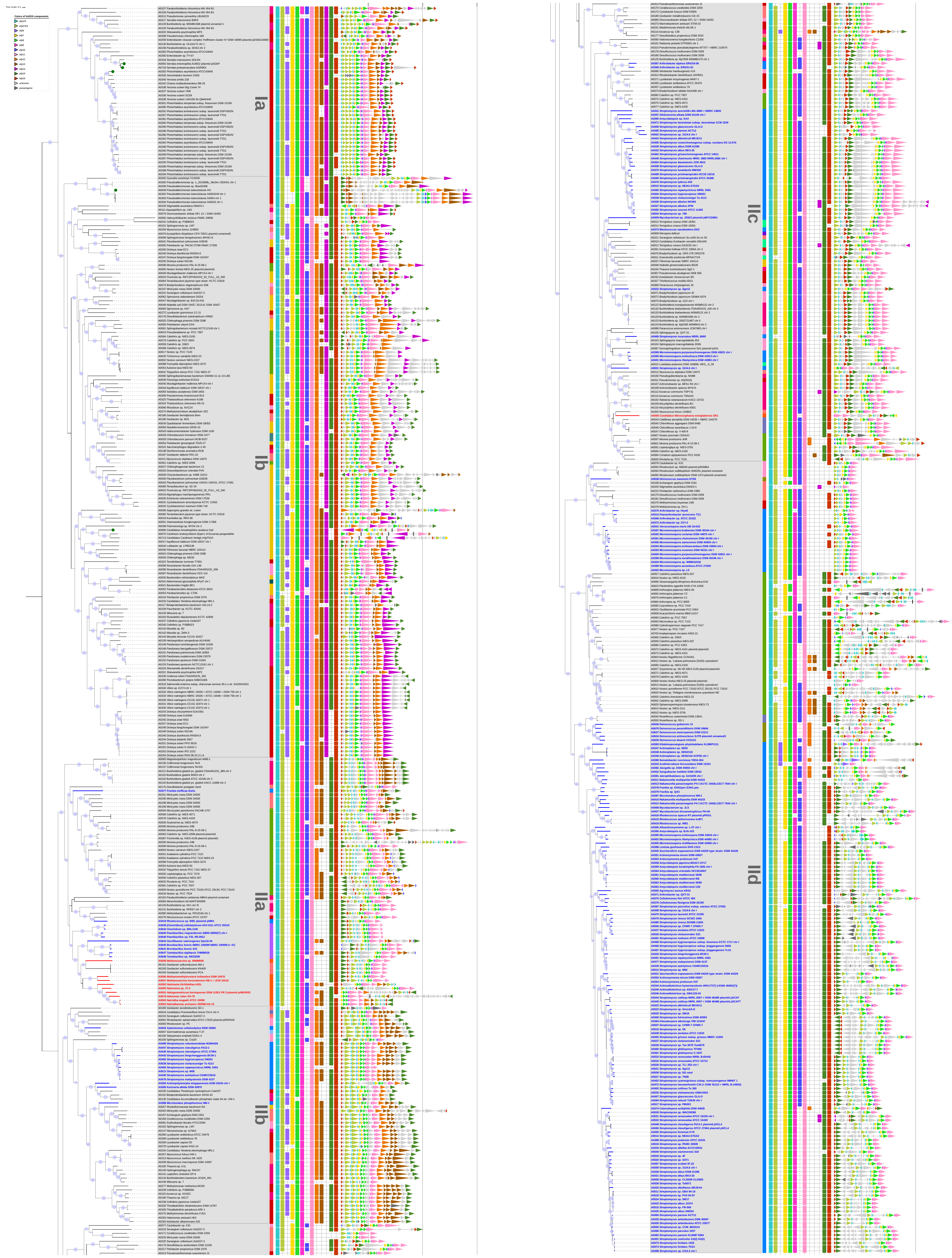


Figure S2. The detailed phylogenetic tree of the 631 eCIS-like loci. Bootstrap values are indicated with proportional light purple bubbles on individual branches. Loci derived from archaeal and Gram-positive bacterial genomes are highlighted with red and blue branches, respectively. Previously reported eCIS loci are indicated by solid green circles at the leaf. The dbeCIS ID of each eCIS-like loci is provided before the genome name. The subtypes are highlighted with grey background shadows. The left-most colored bar denote the taxonomic groups using the same schema as Figure 3. The 13 colored squares after each node indicate the presence and absence of the known eCIS components in each locus with the same color code as Figure 1. Detailed genetic organization of each locus is provided as right-most schematic linear map (to scale).

Figure S3. The detailed sub-tree of subtype IIa, related to Figure 4. Internal conserved genes are denoted as colored solid circles after the known components squares and highlighted as colored pentagram within the right-most linear map.

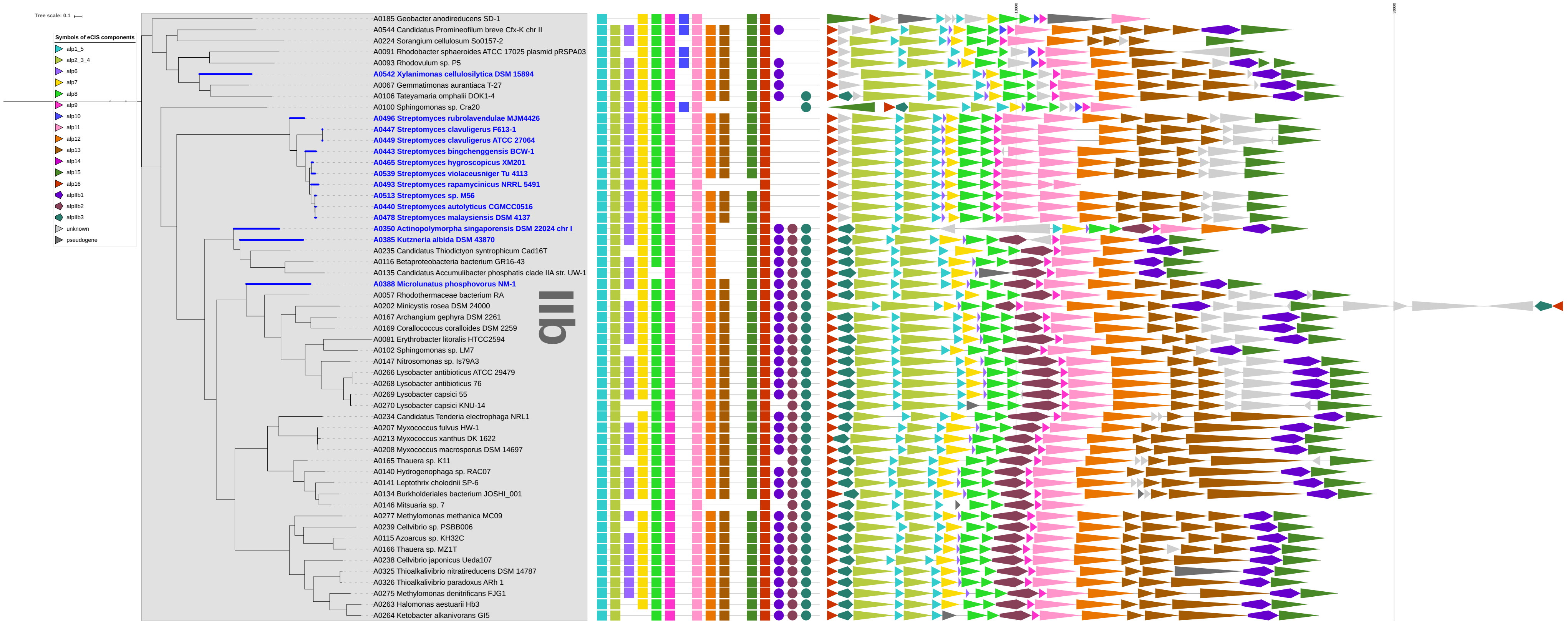


Figure S4. The detailed sub-tree of subtype IIb, related to Figure 4. Internal conserved genes are denoted as colored solid circles after the known components squares and highlighted as colored pentagram within the right-most linear map.

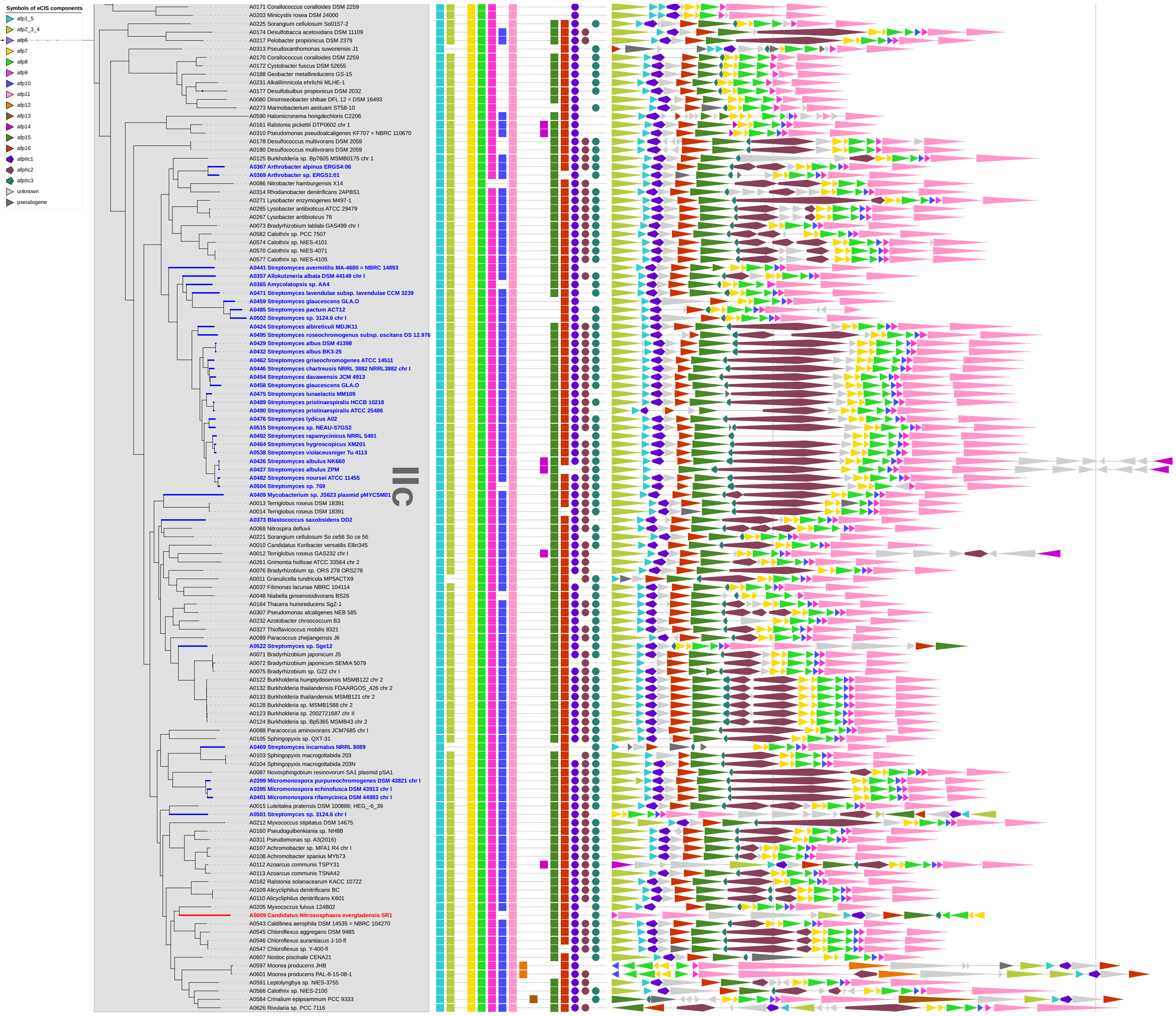


Figure S5. The detailed sub-tree of subtype IIc, related to Figure 4. Internal conserved genes are denoted as colored solid circles after the known components squares and highlighted as colored pentagram within the right-most linear map.

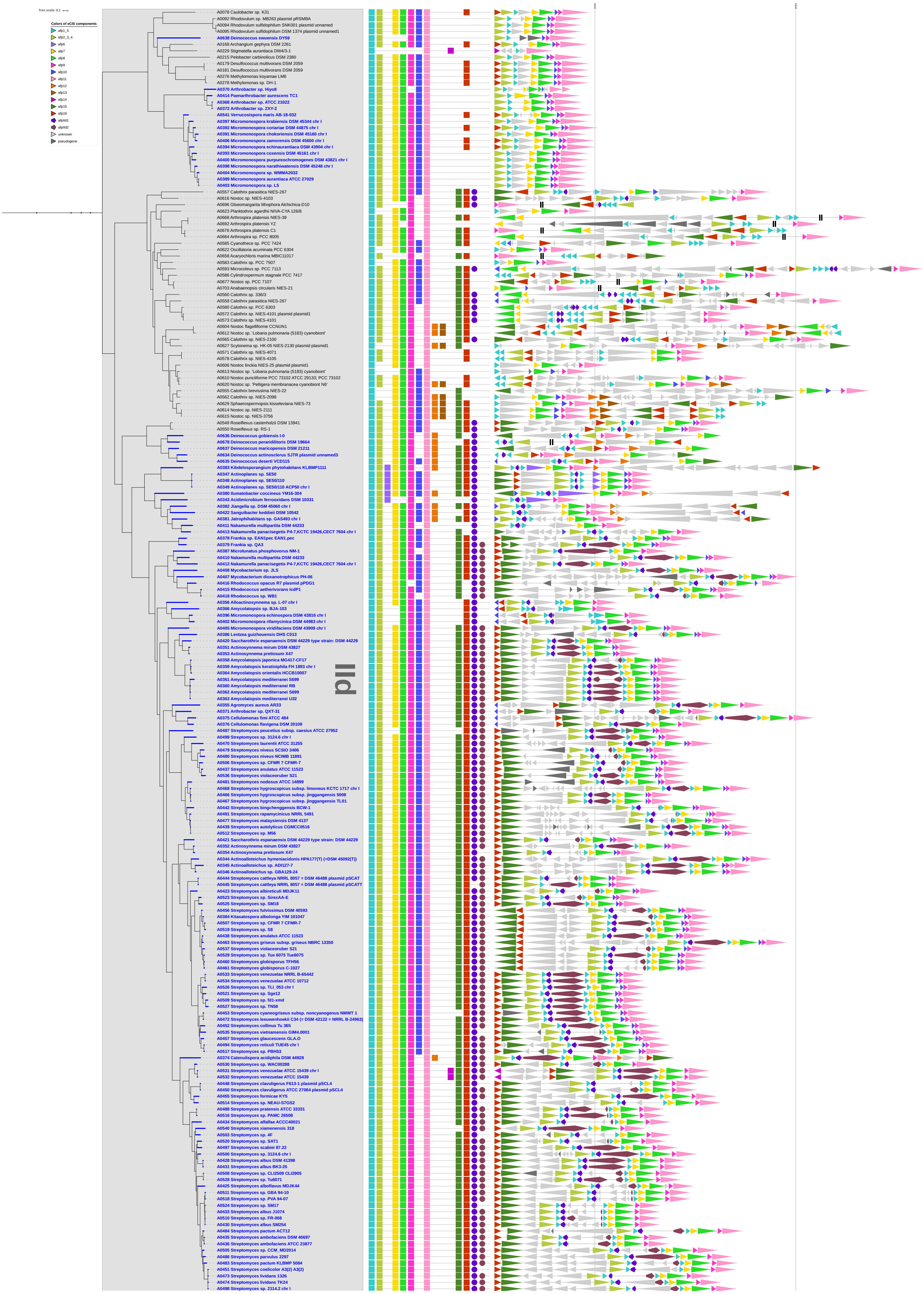
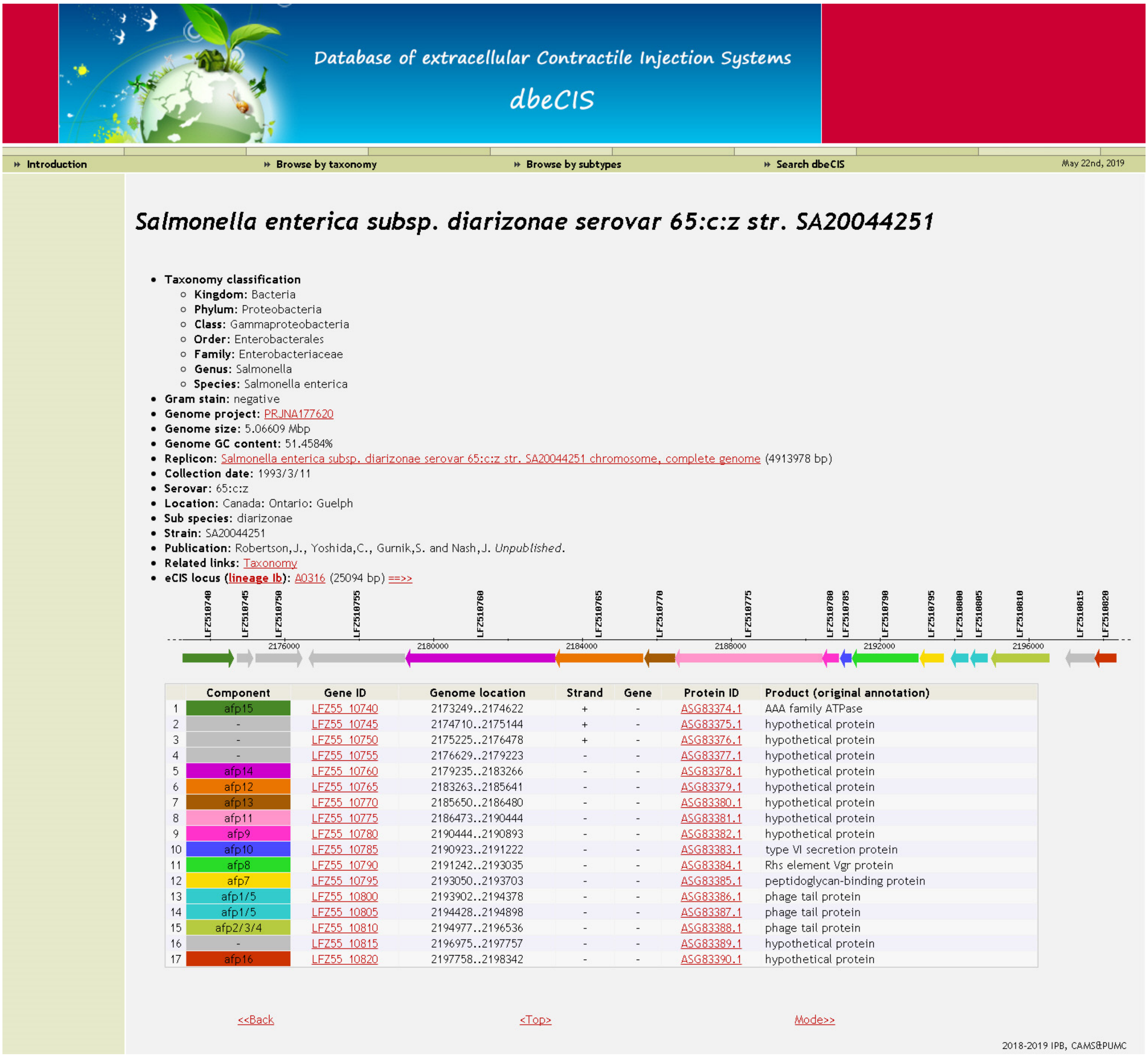


Figure S6. The detailed sub-tree of subtype IId, related to Figure 4. Internal conserved genes are denoted as colored solid circles after the known components squares and highlighted as colored pentagram within the right-most linear map.



Salmonella enterica subsp. diarizonae serovar 65:c:z str. SA20044251

- **Taxonomy classification**
    - **Kingdom:** Bacteria
    - **Phylum:** Proteobacteria
    - **Class:** Gammaproteobacteria
    - **Order:** Enterobacterales
    - **Family:** Enterobacteriaceae
    - **Genus:** Salmonella
    - **Species:** Salmonella enterica
- **Gram stain:** negative
- **Genome project:** PRJNA177620
- **Genome size:** 5.06609 Mbp
- **Genome GC content:** 51.4584%
- **Replicon:** Salmonella enterica subsp. diarizonae serovar 65:c:z str. SA20044251 chromosome, complete genome (4913978 bp)
- **Collection date:** 1993/3/11
- **Serovar:** 65:c:z
- **Location:** Canada: Ontario: Guelph
- **Sub species:** diarizonae
- **Strain:** SA20044251
- **Publication:** Robertson,J., Yoshida,C., Gurnik,S. and Nash,J. *Unpublished*.
- **Related links:** Taxonomy
- **eCIS locus (lineage Ib):** A0316 (25094 bp) ==>>

	Component	Gene ID	Genome location	Strand	Gene	Protein ID	Product (original annotation)
1	afp15	LFZ55_10740	2173249..2174622	+	-	ASG83374.1	AAA family ATPase
2	-	LFZ55_10745	2174710..2175144	+	-	ASG83375.1	hypothetical protein
3	-	LFZ55_10750	2175225..2176478	+	-	ASG83376.1	hypothetical protein
4	-	LFZ55_10755	2176629..2179223	-	-	ASG83377.1	hypothetical protein
5	afp14	LFZ55_10760	2179235..2183266	-	-	ASG83378.1	hypothetical protein
6	afp12	LFZ55_10765	2183263..2185641	-	-	ASG83379.1	hypothetical protein
7	afp13	LFZ55_10770	2185650..2186480	-	-	ASG83380.1	hypothetical protein
8	afp11	LFZ55_10775	2186473..2190444	-	-	ASG83381.1	hypothetical protein
9	afp9	LFZ55_10780	2190444..2190893	-	-	ASG83382.1	hypothetical protein
10	afp10	LFZ55_10785	2190923..2191222	-	-	ASG83383.1	type VI secretion protein
11	afp8	LFZ55_10790	2191242..2193035	-	-	ASG83384.1	Rhs element Vgr protein
12	afp7	LFZ55_10795	2193050..2193703	-	-	ASG83385.1	peptidoglycan-binding protein
13	afp1/5	LFZ55_10800	2193902..2194378	-	-	ASG83386.1	phage tail protein
14	afp1/5	LFZ55_10805	2194428..2194898	-	-	ASG83387.1	phage tail protein
15	afp2/3/4	LFZ55_10810	2194977..2196536	-	-	ASG83388.1	phage tail protein
16	-	LFZ55_10815	2196975..2197757	-	-	ASG83389.1	hypothetical protein
17	afp16	LFZ55_10820	2197758..2198342	-	-	ASG83390.1	hypothetical protein

Figure S7. Snapshot of dbeCIS information for an example locus, related to STAR Methods. An example of dbeCIS webpage to present the taxonomy, bacterial characteristics, genomic features, related publications and the clickable linear map of an eCIS-like locus.

Table S2. Microorganisms, including Gram-negative/positive bacteria and Archaea, encode “split” eCIS-like loci within their genomes, related to Figure 2.

eCIS ID*	**Genome**	**Kingdom**	**Phylum**	**Family**	**Gram**	**eCIS Subtype**
A0679	*Halovivax ruber* XH-70	Archaea	Euryarchaeota	Natrialbaceae		IIa
A0660	*Amoebophilus asiaticus* 5a2$	Bacteria	Bacteroidetes	Amoebophilaceae	-	Ib
A0675	*Cardinium hertigii* cEper1	Bacteria	Bacteroidetes	Amoebophilaceae	-	Ib
A0713	*Cardinium hertigii* cHgTN10	Bacteria	Bacteroidetes	Amoebophilaceae	-	Ib
A0658	*Acaryochloris marina* MBIC11017	Bacteria	Cyanobacteria	Acaryochloridaceae	-	IId
A0703	*Anabaenopsis circularis* NIES-21	Bacteria	Cyanobacteria	Aphanizomenonaceae	-	IId
A0696	*Gloeomargarita lithophora* Alchichica-D10	Bacteria	Cyanobacteria	Gloeomargaritaceae	-	IId
A0668	*Arthrospira platensis* NIES-39	Bacteria	Cyanobacteria	Microcoleaceae	-	IId
A0676	*Arthrospira platensis* C1	Bacteria	Cyanobacteria	Microcoleaceae	-	IId
A0684	*Arthrospira sp.* PCC 8005	Bacteria	Cyanobacteria	Microcoleaceae	-	IId
A0692	*Arthrospira platensis* YZ	Bacteria	Cyanobacteria	Microcoleaceae	-	IId
A0677	*Nostoc sp.* PCC 7107	Bacteria	Cyanobacteria	Nostocaceae	-	IId
A0678	*Deinococcus peraridilitoris* DSM 19664	Bacteria	Deinococcus-Thermus	Deinococcaceae	+	IId

Note: *, eCIS IDs available from the dbeCIS database; $, experimentally verified (Bock et al. 2017).

Table S5. Organisms encoding both classical T6SS and eCIS-like loci, related to Figure 6.

Organism	Accession	T6SS ID*	Completeness of T6SS components	eCIS ID#	eCIS Lineage
Alkalilimnicola ehrlichii MLHE-1	NC_008340	145	no TssH	A0231	IIc
Azoarcus sp. KH32C	NC_020516	736	no TssI	A0115	IIb
Burkholderia gladioli BSR3	NC_015381	433	no TssI	A0119	IIa
	NC_015376	430	no TssJ		
	NC_015383	**434**			
Burkholderia thailandensis MSMB121	NC_021173	758	no TssI	A0133	IIc
	NC_021174	**761**			
		762			
		760			
		759			
Chitinophaga pinensis DSM 2588	NC_013132	912	no TssI, TssM	A0024	Ib
		913	no TssM	A0025	
Collimonas fungivorans Ter331	NC_015856	**455**		A0137	IIa
Corallococcus coralloides DSM 2259	NC_017030	558	no TssH, TssD, TssI, TssA, TssG, TssJ	A0169	IIb
				A0170	IIc
				A0171	
Dechloromonas aromatica RCB	NC_007298	**87**		A0138	Ib
Dickeya chrysanthemi Ech1591	NC_012912	351	no TssI,TssD	A0243	Ib
Dickeya dadantii 3937	NC_014500	409	no TssI,TssD	A0244	Ib
Dickeya zeae Ech586	NC_013592	373	no TssI,TssD	A0258	Ib
Dyadobacter fermentans DSM 18053	NC_013037	911	no TssM, TssO, TssP	A0034	Ib
Flavobacterium johnsoniae UW101	NC_009441	910	no TssM	A0042	Ib
Gemmatimonas aurantiaca T-27	NC_012489	332	no TssJ	A0067	IIb
Geobacter bemidjiensis Bem	NC_011146	296	no TssH	A0186	Ib
Geobacter metallireducens GS-15	NC_007517	107	no TssH, TssE, TssF, TssG, TssJ, TssK	A0188	IIc
		108	no TssD, TssL, TssI, TssA, TssB, TssC, TssM		
Geobacter sp. M21	NC_012918	354	no TssH	A0190	Ib
Geobacter sulfurreducens KN400	NC_017454	602	no TssD, TssL, TssI, TssA, TssB, TssC, TssM	A0193	IIa
		603	no TssH, TssE, TssF, TssG, TssJ, TssK		
Geobacter sulfurreducens PCA	NC_002939	11	no TssD, TssL, TssI, TssA, TssB, TssC, TssM	A0194	IIa
		12	no TssH, TssE, TssF, TssG, TssJ, TssK		
Granulicella tundricola MP5ACTX9	NC_015057	506	no TssJ	A0011	IIc
Leptothrix cholodnii SP-6	NC_010524	**265**		A0141	IIb
Mesorhizobium japonicum MAFF 303099	NC_002678	**7**		A0084	IIa
Methylomicrobium alcaliphilum 20Z	NC_016112	512	no TssD, TssJ	A0274	Ib
Methylomonas methanica MC09	NC_015572	443	no TssL, TssI, TssJ, TssK, TssM	A0277	IIb
		444	no TssH, TssD, TssI, TssA, TssB, TssC, TssE, TssF, TssG		

Myxococcus fulvus HW-1	NC_015711	**450**		A0207	IIb
Myxococcus stipitatus DSM 14675	NC_020126	725	no TssI	A0211	Ib
		797	no TssD, TssI, TssA, TssJ	A0212	IIc
Myxococcus xanthus DK 1622	NC_008095	**127**		A0213	IIb
Niastella koreensis GR20-10	NC_016609	923	no TssM	A0050	Ib
Nitrospira defluvii	NC_014355	**407**		A0068	IIc
Pelobacter carbinolicus DSM 2380	NC_007498	**100**		A0215	IId
Photorhabdus asymbiotica	NC_012962	**355**		A0281	Ia
		356		A0282	
		357	no TssH, TssA, TssE, TssJ	A0283	
		358		A0284	
				A0285	
Photorhabdus luminescens subsp. laumondii TTO1	NC_005126	**50**		A0292	Ia
		51		A0293	
		52	no TssH, TssL, TssA	A0294	
		53	no TssB	A0295	
				A0296	
				A0297	
Pseudogulbenkiania sp. NH8B	NC_016002	**459**		A0160	IIc
Saccharophagus degradans 2-40	NC_007912	**119**		A0315	Ib
Sorangium cellulosum So ce56	NC_010162	248	no TssI	A0221	IIc
		249	no TssH, TssJ		
Sorangium cellulosum So0157-2	NC_021658	815	no TssI	A0222	Ib
				A0223	IIc
				A0224	IIb
				A0225	IIc
Stigmatella aurantiaca DW4/3-1	NC_014623	500	no TssH, TssD, TssI, TssA, TssG, TssJ	A0226	Ib
				A0229	IId
Teredinibacter turnerae T7901	NC_012997	**359**		A0322	Ib

Note: * (T6SS ID) are available from T6SS database - SecReT6 (Li et al. 2015). Description of T6SS components is available from SecReT6 (http://db-mml.sjtu.edu.cn/SecReT6/). Classical T6SS loci containing all essential genes are labeled with underlined T6SS ID number. Most of T6SS loci listed here are lack of several essential T6SS genes. # (eCIS ID) is available from dbeCIS.

Table S6. A list of plasmid-encoded eCIS-like loci in different bacteria species, related to Figure 3.

Bacteria	Phylum	eCIS ID*	Replicon	GenBank accession	eCIS sub-type
Mycobacterium sp. JS623	Actinobacteria	A0409	plasmid pMYCSM01	CP003079	IIc
Rhodococcus opacus R7	Actinobacteria	A0416	plasmid pPDG1	CP008948	IId
Rhodococcus sp. WB1	Actinobacteria	A0418	chromosome	CP015529	IId
		A0419	plasmid pWB1	CP015530	IIa
Streptomyces cattleya DSM 46488	Actinobacteria	A0445	plasmid pSCATT	CP003229	IId
Streptomyces cattleya DSM 46488	Actinobacteria	A0444	plasmid pSCAT	FQ859184	IId
Streptomyces clavuligerus ATCC 27064	Actinobacteria	A0449	chromosome	CM000913	IIb
		A0450	plasmid pSCL4	CM000914	IId
Streptomyces clavuligerus F613-1	Actinobacteria	A0447	chromosome	CP016559	IIb
		A0448	plasmid pSCL4	CP016560	IId
Fischerella sp. NIES-4106	Cyanobacteria	A0587	plasmid plasmid1	AP018299	IIa
Nostoc linckia NIES-25	Cyanobacteria	A0606	plasmid plasmid1	AP018223	IId
		A0605	plasmid plasmid1		Ib
Calothrix sp. NIES-2098	Cyanobacteria	A0561	chromosome	AP018172	Ib
		A0562			IId
		A0563	plasmid plasmid1		IIa
Calothrix sp. NIES-4101	Cyanobacteria	A0573	chromosome	AP018280	IId
		A0574			IIc
		A0572	plasmid plasmid1	AP018274	IId
Scytonema sp. HK-05 NIES-2130	Cyanobacteria	A0627	plasmid plasmid1	AP018195	IId
Deinococcus actinosclerus SJTR	Deinococcus-Thermus	A0634	plasmid unnamed3	CP029777	IId
Halogeometricum borinquense DSM 11551 PR 3	Euryarchaeota	A0001	plasmid pHBOR03	CP001693	IIa
Burkholderia sp. MSMB1588	Proteobacteria	A0128	chromosome 2	CM003770	IIc
		A0129	plasmid unnamed 1	LNJS01000004	Ia
Paraburkholderia caribensis MBA4	Proteobacteria	A0154	plasmid unnamed	CP012748	IIa
Enterobacter roggenkampii DSM 16690	Proteobacteria	A0259	plasmid pDSMZ16690	CP017185	Ia
Rhodobacter sphaeroides ATCC 17025	Proteobacteria	A0091	plasmid pRSPA03	CP000664	IIb
Rhodovulum sp. MB263	Proteobacteria	A0092	plasmid pRSMBA	CP020385	IId
Rhodovulum sulfidophilum DSM 1374	Proteobacteria	A0095	plasmid unnamed1	CP015419	IId
Rhodovulum sulfidophilum SNK001	Proteobacteria	A0094	plasmid unnamed	CP015422	IId
Azospirillum thiophilum CFH 70021	Proteobacteria	A0070	plasmid unnamed3	CP029358	Ib
Novosphingobium resinovorum SA1	Proteobacteria	A0087	plasmid pSA1	CP017076	IIc
Serratia entomophila A1MO2	Proteobacteria	A0650	plasmid pADAP	AF135182	Ia

Note: * (eCIS ID) is available from dbeCIS. Several bacteria species encodes multi-copy eCIS-like loci in both chromosome and plasmid.