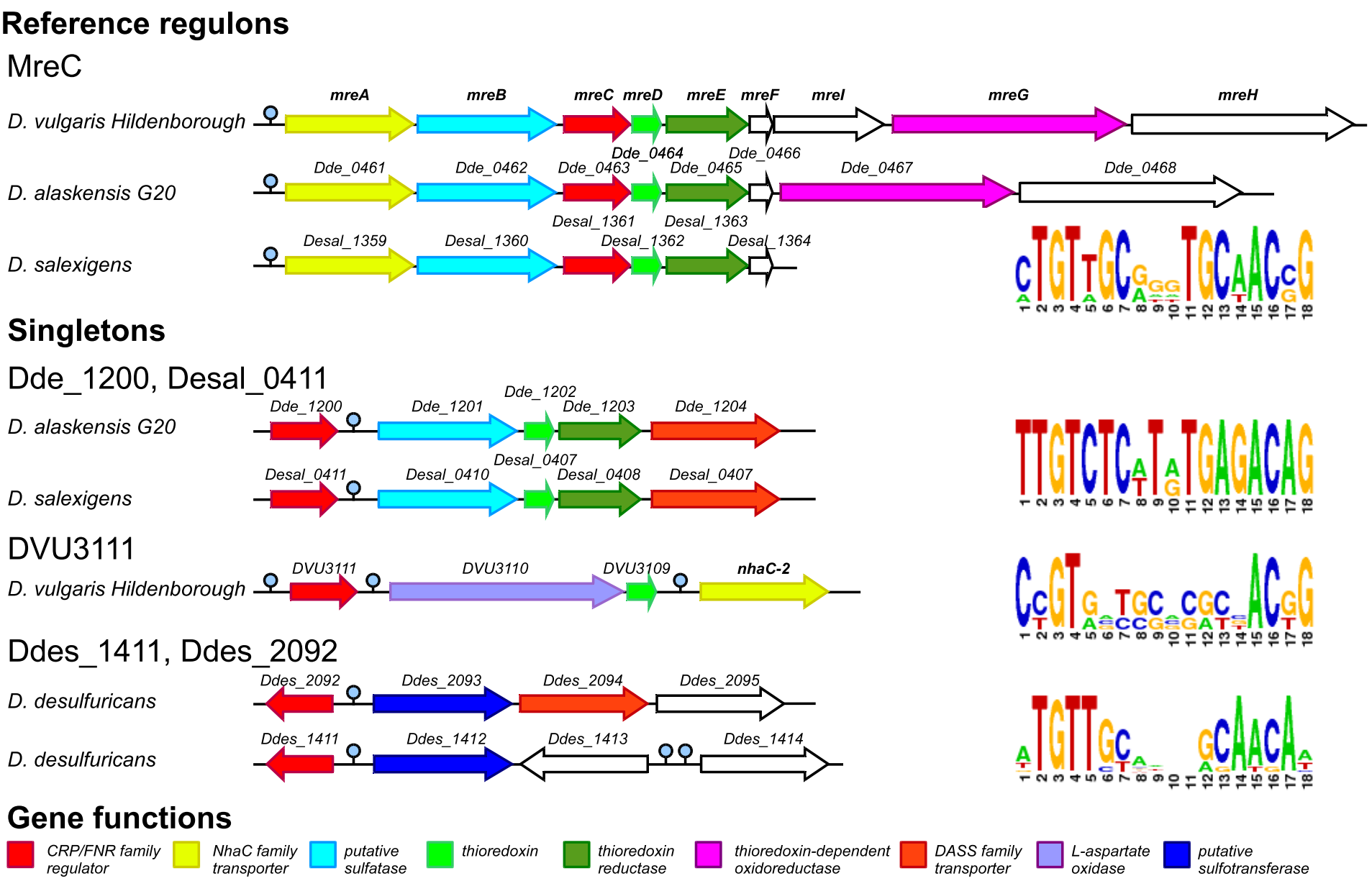
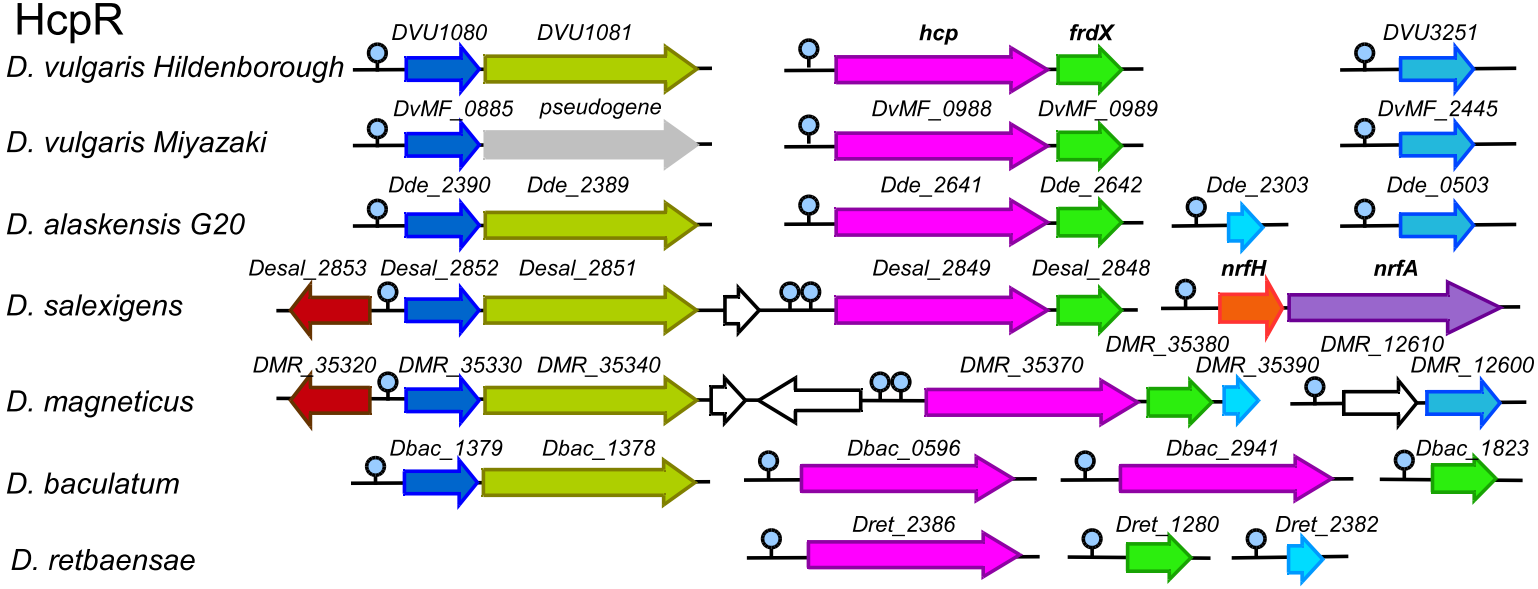
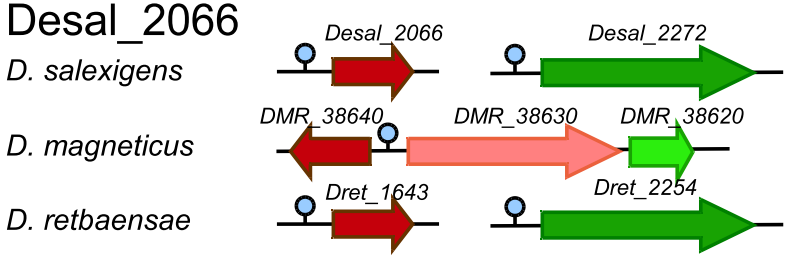


FIG S1 Genomic organization of MreC-like regulons.

Reference regulons

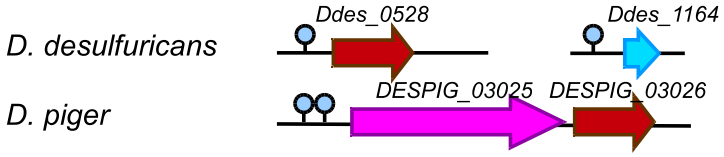


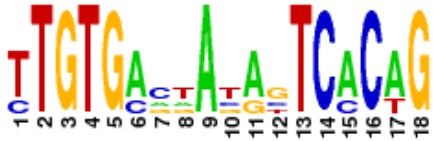
Desal_2066



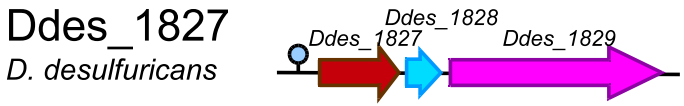
Singletons

Ddes_0528, DESPIG_03026





Ddes_1827



Gene functions

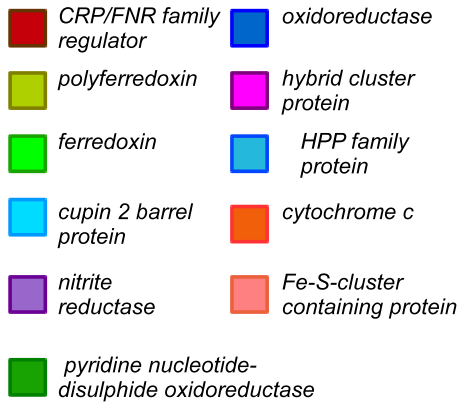


FIG S2 Genomic organization of HcpR-like regulons.

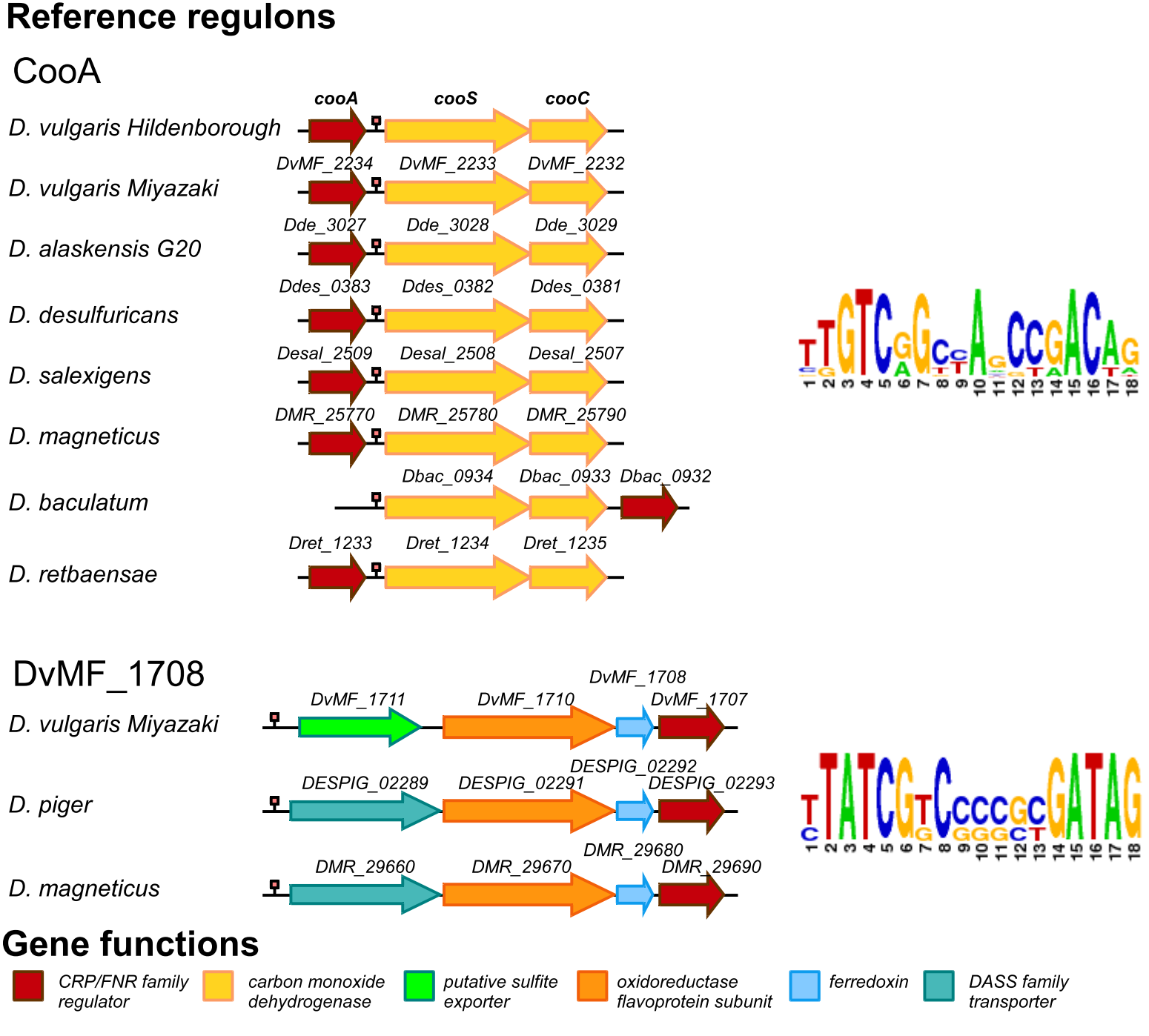


FIG S3 Genomic organization of CooA-like regulons.

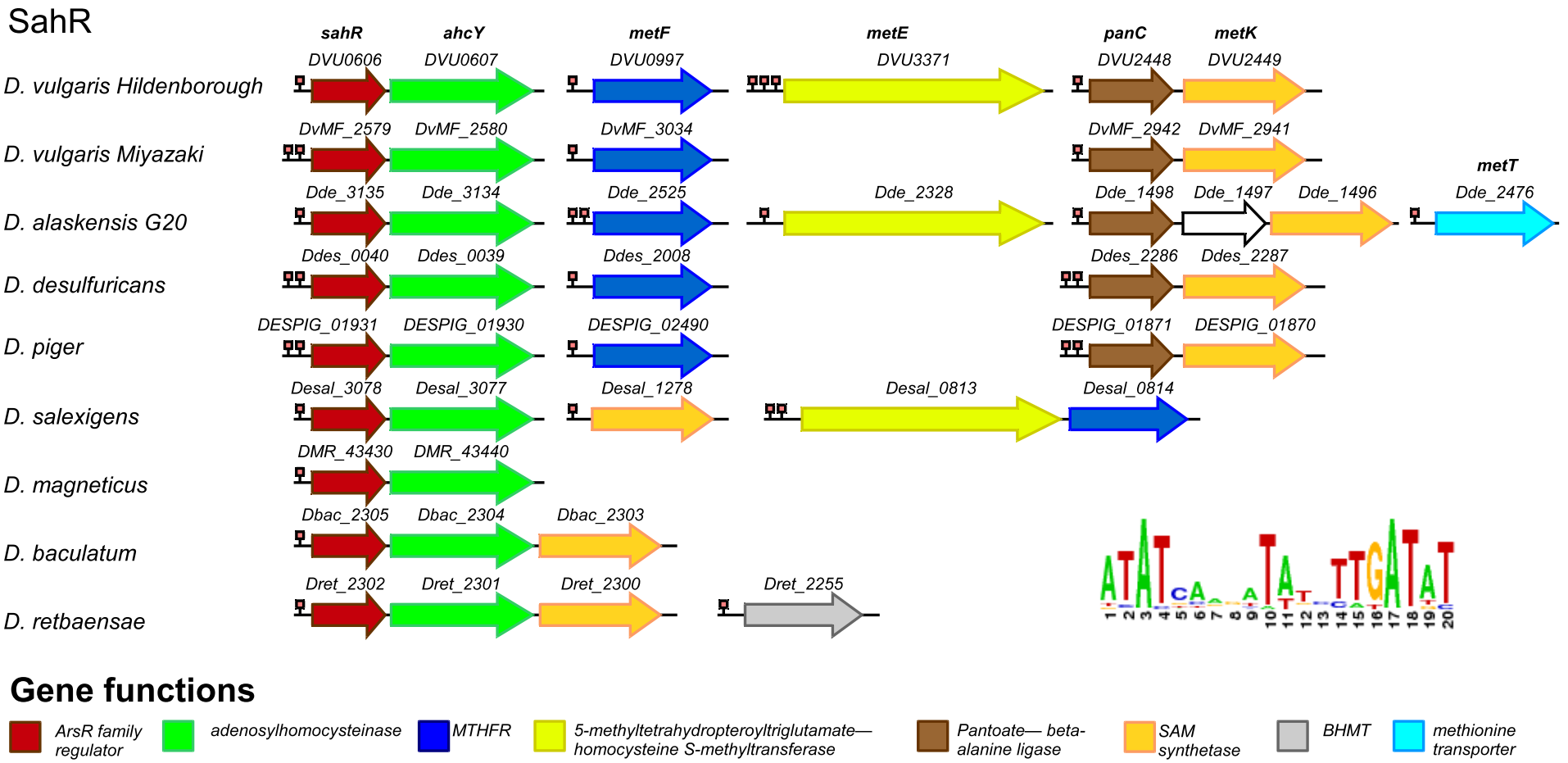


FIG S4 Genomic organization of SahR regulon.

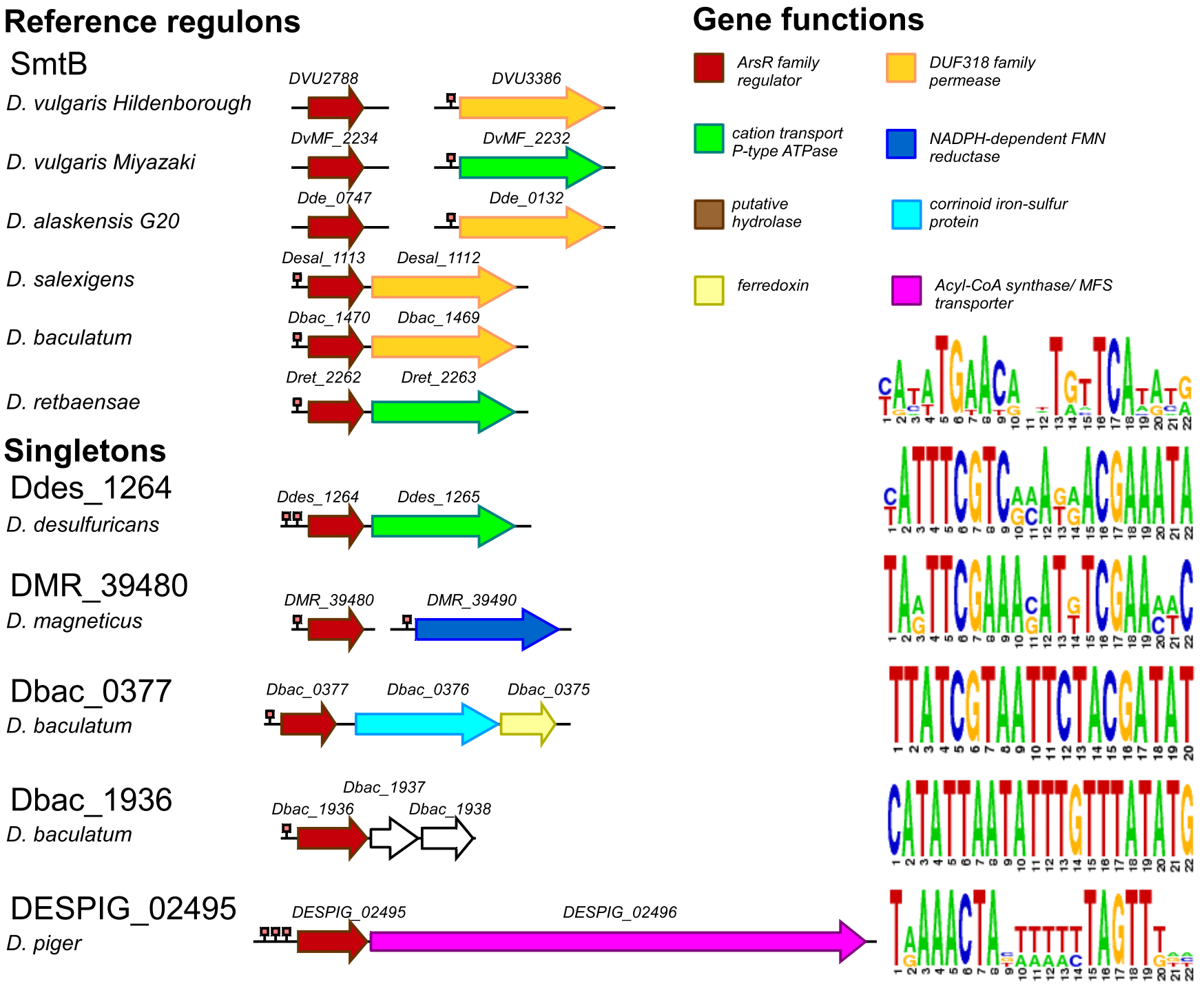


FIG S5 Genomic organization of SmtB regulon and ArsR family singletons.

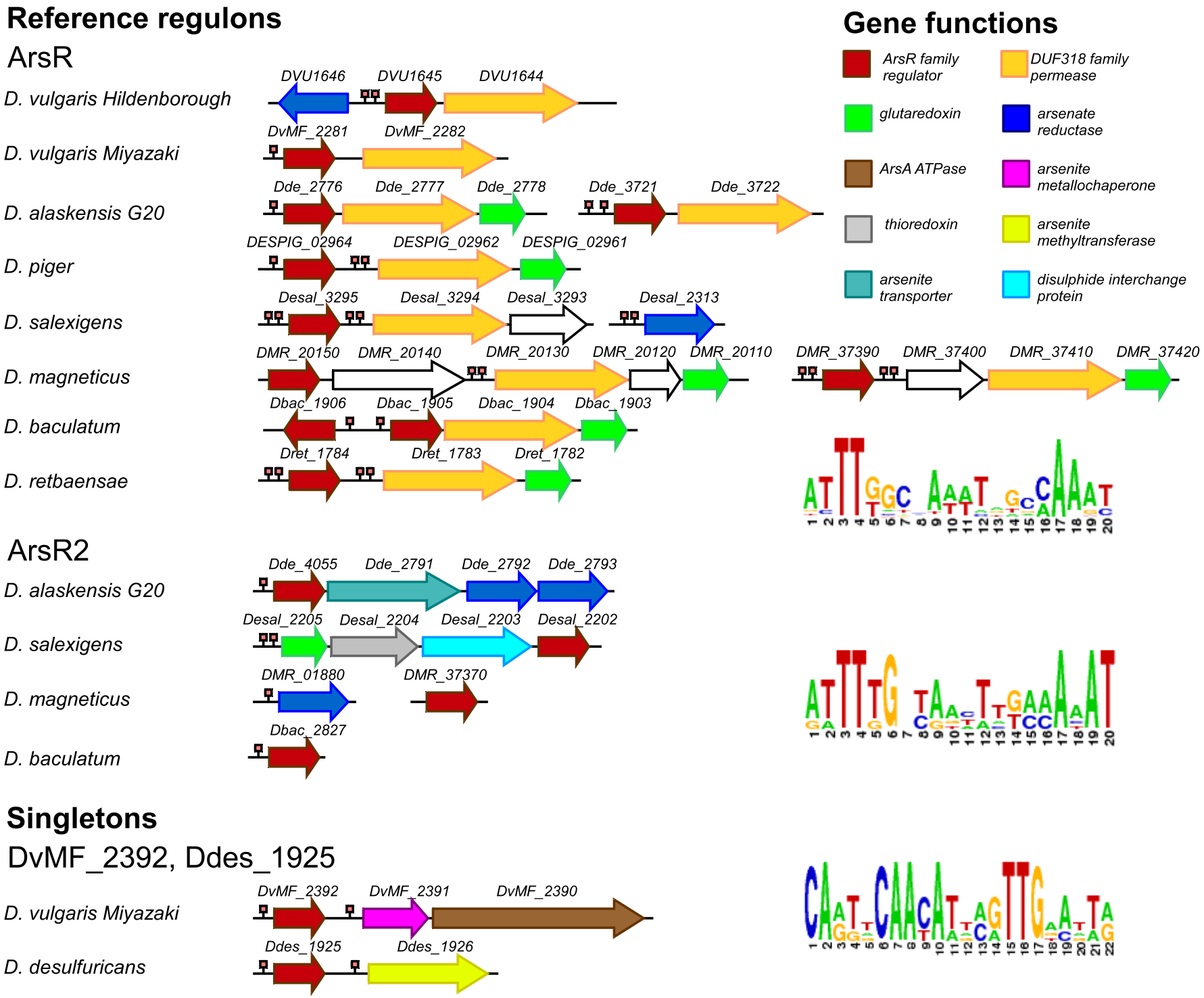


FIG S6 Genomic organization of arsenic resistance regulatory systems.

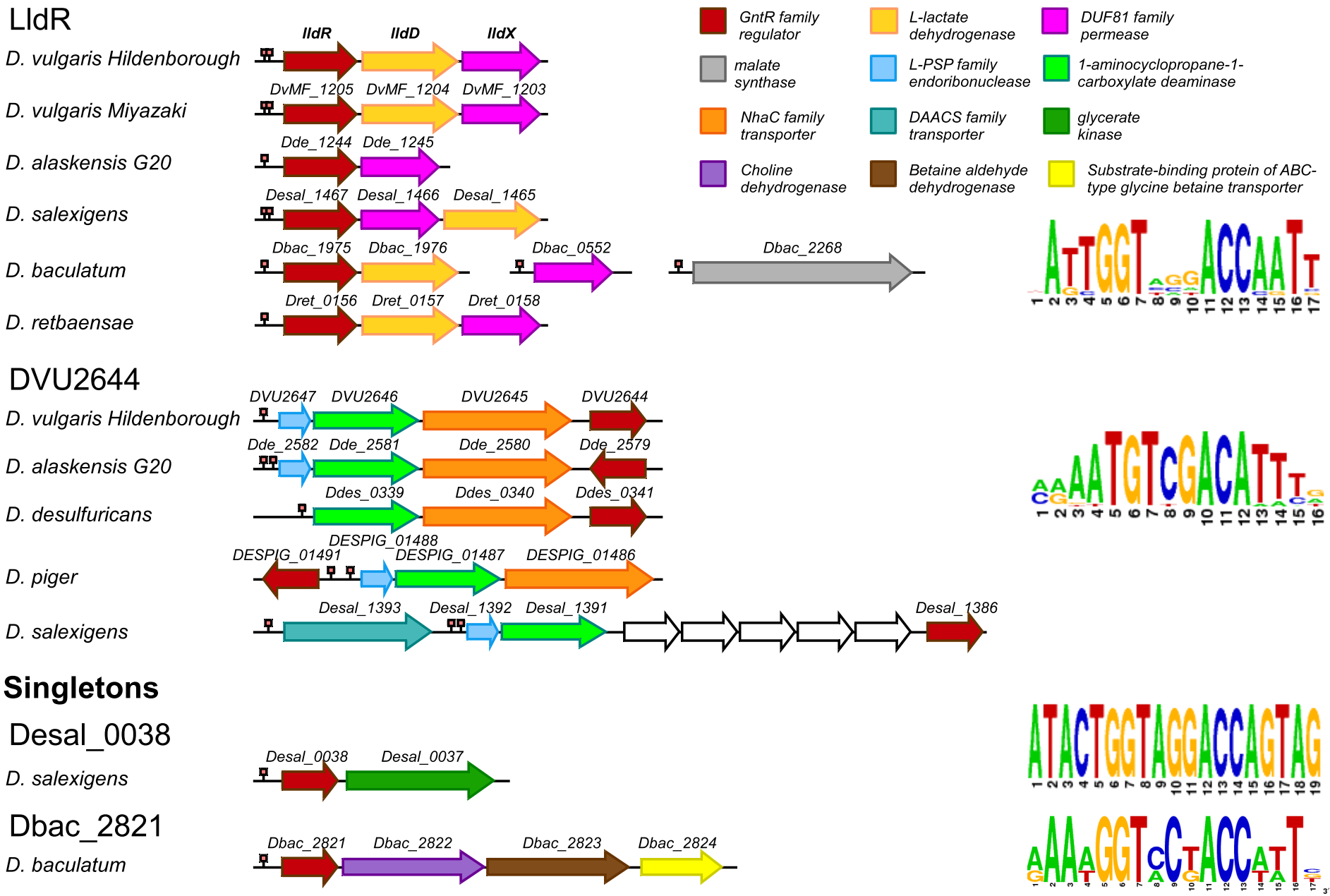


FIG S7 Genomic organization of FadR subfamily regulatory systems.

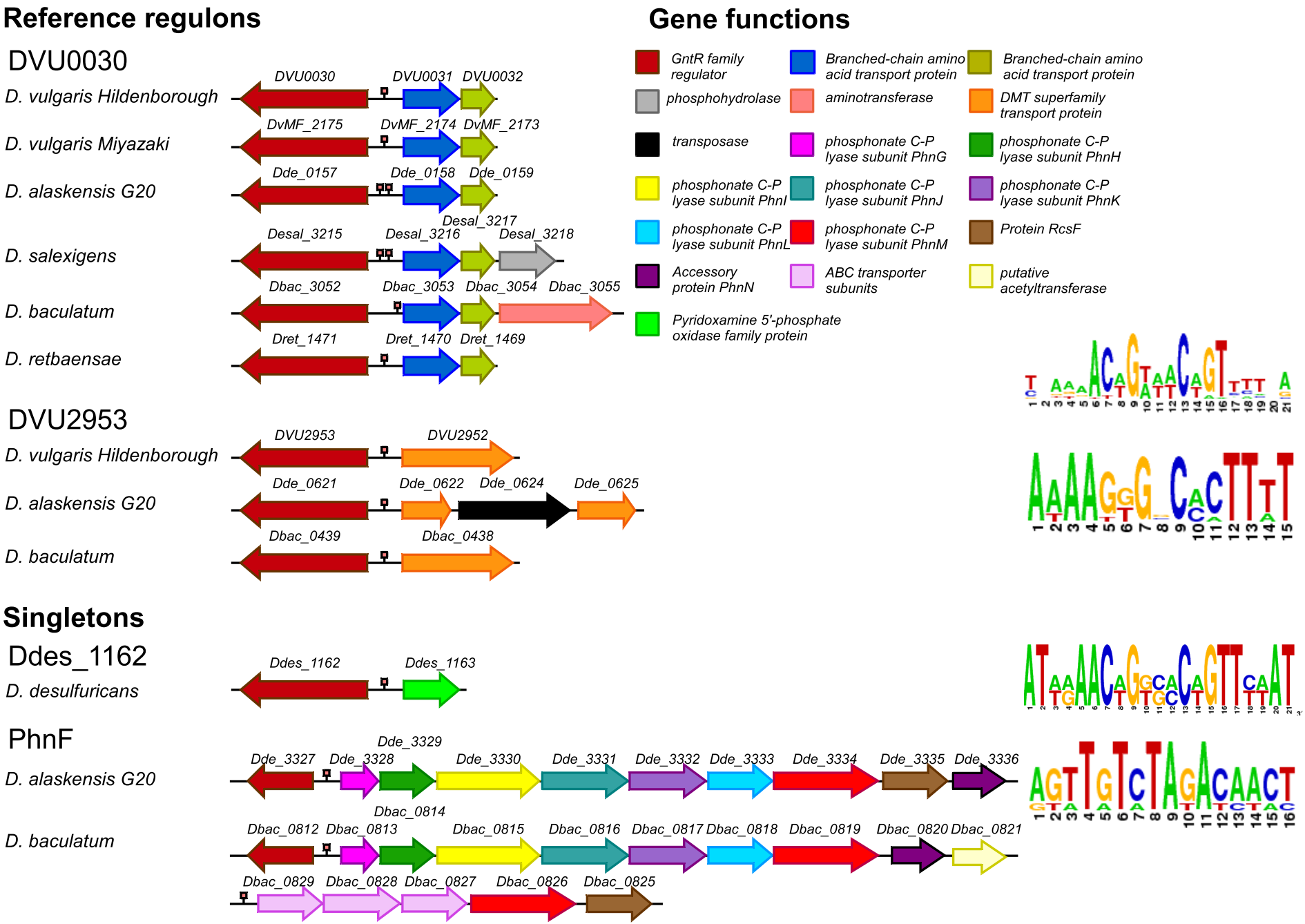


FIG S8 Genomic organization of MocR and HutC subfamilies regulatory systems.

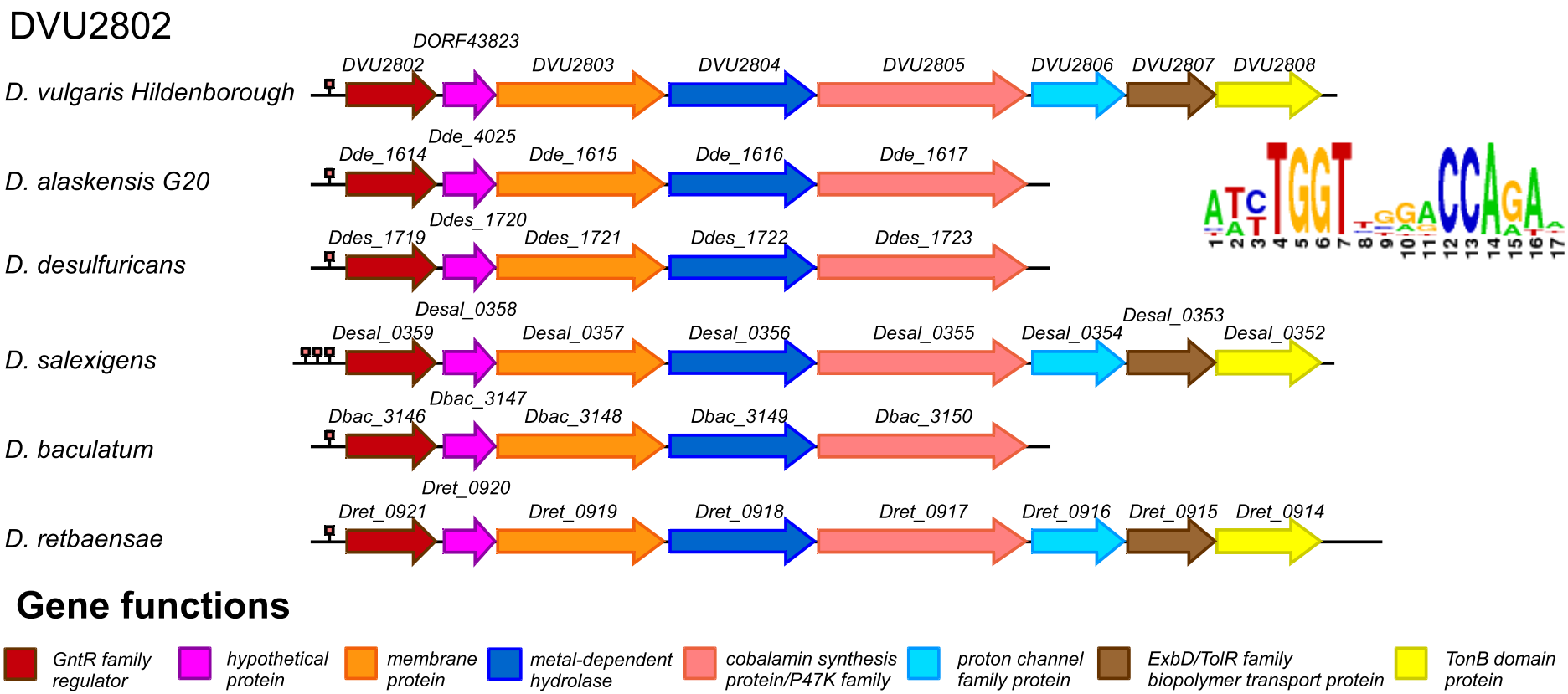


FIG S9 Genomic organization of DVU2802 subfamily regulatory systems.

TABLE S1 Regulatory proteins of Crp/Fnr, ArsR and GntR TF families from *Desulfovibrionales*.

Gene ID	Genome	TF family	Orthology group	Singleton TF
DVU2097	Desulfovibrio vulgaris str. Hildenborough	Crp/Fnr	CooA	
DVU0379	Desulfovibrio vulgaris str. Hildenborough	Crp/Fnr	MreC	
DVU3111	Desulfovibrio vulgaris str. Hildenborough	Crp/Fnr	DVU3111	S
DVU2547	Desulfovibrio vulgaris str. Hildenborough	Crp/Fnr	HcpR	
DvMF_2234	Desulfovibrio vulgaris str. Miyazaki	Crp/Fnr	CooA	
DvMF_0993	Desulfovibrio vulgaris str. Miyazaki	Crp/Fnr	HcpR	
DvMF_1708	Desulfovibrio vulgaris str. Miyazaki	Crp/Fnr	DvMF_1708	
Dde_3027	Desulfovibrio alaskensis str. G20	Crp/Fnr	CooA	
Dde_0463	Desulfovibrio alaskensis str. G20	Crp/Fnr	MreC	
Dde_2644	Desulfovibrio alaskensis str. G20	Crp/Fnr	HcpR	
Dde_1200	Desulfovibrio alaskensis str. G20	Crp/Fnr	Dde_1200	S
Ddes_0383	Desulfovibrio desulfuricans ATCC 2774	Crp/Fnr	CooA	
Ddes_1411	Desulfovibrio desulfuricans ATCC 2774	Crp/Fnr	Ddes_1411	S
Ddes_2092	Desulfovibrio desulfuricans ATCC 2774	Crp/Fnr	Ddes_1411	S
Ddes_0528	Desulfovibrio desulfuricans ATCC 2774	Crp/Fnr	Ddes_0528	S
Ddes_1827	Desulfovibrio desulfuricans ATCC 2774	Crp/Fnr	Ddes_1827	S
DESPIG_03026	Desulfovibrio piger ATCC 29098	Crp/Fnr	Ddes_0528	S
DESPIG_02293	Desulfovibrio piger ATCC 29098	Crp/Fnr	DvMF_1708	S
Desal_2509	Desulfovibrio salexigens DSM 2638	Crp/Fnr	CooA	
Desal_1361	Desulfovibrio salexigens DSM 2638	Crp/Fnr	MreC	
Desal_2853	Desulfovibrio salexigens DSM 2638	Crp/Fnr	HcpR	
Desal_0411	Desulfovibrio salexigens DSM 2638	Crp/Fnr	Dde_1200	S
Desal_0494	Desulfovibrio salexigens DSM 2638	Crp/Fnr	Desal_0494	S
Desal_3734	Desulfovibrio salexigens DSM 2638	Crp/Fnr	Desal_3734	S
Desal_2066	Desulfovibrio salexigens DSM 2638	Crp/Fnr	Desal_2066	
DMR_25770	Desulfovibrio magneticus RS-1	Crp/Fnr	CooA	
DMR_35320	Desulfovibrio magneticus RS-1	Crp/Fnr	HcpR	
DMR_29690	Desulfovibrio magneticus RS-1	Crp/Fnr	DvMF_1708	
DMR_38640	Desulfovibrio magneticus RS-1	Crp/Fnr	Desal_2066	
Dbac_0932	Desulfomicrobium baculatum DSM 4028	Crp/Fnr	CooA	
Dbac_2118	Desulfomicrobium baculatum DSM 4028	Crp/Fnr	HcpR	
Dret_1233	Desulfohalobium retbaense DSM 5692	Crp/Fnr	CooA	
Dret_1279	Desulfohalobium retbaense DSM 5692	Crp/Fnr	HcpR	
Dret_1643	Desulfohalobium retbaense DSM 5692	Crp/Fnr	Desal_2066	
DVU1645	Desulfovibrio vulgaris str. Hildenborough	ArsR	ArsR	
DVU2788	Desulfovibrio vulgaris str. Hildenborough	ArsR	SmtB	
DVU0606	Desulfovibrio vulgaris str. Hildenborough	ArsR	SahR	
DvMF_2281	Desulfovibrio vulgaris str. Miyazaki	ArsR	ArsR	
DvMF_1207	Desulfovibrio vulgaris str. Miyazaki	ArsR	SmtB	
DvMF_2579	Desulfovibrio vulgaris str. Miyazaki	ArsR	SahR	
DvMF_2392	Desulfovibrio vulgaris str. Miyazaki	ArsR	ArsR3	S

Dde_3721	Desulfovibrio alaskensis str. G20	ArsR	ArsR	
Dde_2776	Desulfovibrio alaskensis str. G20	ArsR	ArsR	
Dde_4055	Desulfovibrio alaskensis str. G20	ArsR	ArsR2	
Dde_0747	Desulfovibrio alaskensis str. G20	ArsR	SmtB	
Dde_3135	Desulfovibrio alaskensis str. G20	ArsR	SahR	
Ddes_0040	Desulfovibrio desulfuricans ATCC 2774	ArsR	SahR	
Ddes_1264	Desulfovibrio desulfuricans ATCC 2774	ArsR	Ddes_1264	S
Ddes_1925	Desulfovibrio desulfuricans ATCC 2774	ArsR	ArsR3	S
DESPIG_02964	Desulfovibrio piger ATCC 29098	ArsR	ArsR	
DESPIG_01931	Desulfovibrio piger ATCC 29098	ArsR	SahR	
DESPIG_02495	Desulfovibrio piger ATCC 29098	ArsR	DESPIG_02495	S
Desal_3295	Desulfovibrio salexigens DSM 2638	ArsR	ArsR	
Desal_2202	Desulfovibrio salexigens DSM 2638	ArsR	ArsR2	
Desal_1113	Desulfovibrio salexigens DSM 2638	ArsR	SmtB	
Desal_3078	Desulfovibrio salexigens DSM 2638	ArsR	SahR	
DMR_20150	Desulfovibrio magneticus RS-1	ArsR	ArsR	
DMR_37390	Desulfovibrio magneticus RS-1	ArsR	ArsR	
DMR_37370	Desulfovibrio magneticus RS-1	ArsR	ArsR2	
DMR_43430	Desulfovibrio magneticus RS-1	ArsR	SahR	
DMR_39480	Desulfovibrio magneticus RS-1	ArsR	DMR_39480	S
DMR_45340	Desulfovibrio magneticus RS-1	ArsR	DMR_45340	S
Dbac_1905	Desulfomicrobium baculatum DSM 4028	ArsR	ArsR	
Dbac_1906	Desulfomicrobium baculatum DSM 4028	ArsR	ArsR	
Dbac_2827	Desulfomicrobium baculatum DSM 4028	ArsR	ArsR2	
Dbac_1470	Desulfomicrobium baculatum DSM 4028	ArsR	SmtB	
Dbac_2305	Desulfomicrobium baculatum DSM 4028	ArsR	SahR	
Dbac_0377	Desulfomicrobium baculatum DSM 4028	ArsR	Dbac_0377	S
Dbac_1936	Desulfomicrobium baculatum DSM 4028	ArsR	Dbac_1936	S
Dret_1784	Desulfohalobium retbaense DSM 5692	ArsR	ArsR	
Dret_2262	Desulfohalobium retbaense DSM 5692	ArsR	SmtB	
Dret_2302	Desulfohalobium retbaense DSM 5692	ArsR	SahR	
DVU0030	Desulfovibrio vulgaris str. Hildenborough	GntR	DVU0030	
DVU2644	Desulfovibrio vulgaris str. Hildenborough	GntR	DVU2644	
DVU2802	Desulfovibrio vulgaris str. Hildenborough	GntR	DVU2802	
DVU2953	Desulfovibrio vulgaris str. Hildenborough	GntR	DVU2953	
DVU2785	Desulfovibrio vulgaris str. Hildenborough	GntR	LldR	
DVU2086	Desulfovibrio vulgaris str. Hildenborough	GntR	DVU2086	S
DvMF_2175	Desulfovibrio vulgaris str. Miyazaki	GntR	DVU0030	
DvMF_1205	Desulfovibrio vulgaris str. Miyazaki	GntR	LldR	
DvMF_0513	Desulfovibrio vulgaris str. Miyazaki	GntR	DvMF_0513	S
DvMF_1479	Desulfovibrio vulgaris str. Miyazaki	GntR	DvMF_1479	S
DvMF_1994	Desulfovibrio vulgaris str. Miyazaki	GntR	DvMF_1994	S
DvMF_2022	Desulfovibrio vulgaris str. Miyazaki	GntR	DvMF_2022	S

DvMF_2930	Desulfovibrio vulgaris str. Miyazaki	GntR	DvMF_2930	S
DvMF_3005	Desulfovibrio vulgaris str. Miyazaki	GntR	DvMF_3005	S
Dde_0157	Desulfovibrio alaskensis str. G20	GntR	DVU0030	
Dde_2579	Desulfovibrio alaskensis str. G20	GntR	DVU2644	
Dde_1614	Desulfovibrio alaskensis str. G20	GntR	DVU2802	
Dde_0621	Desulfovibrio alaskensis str. G20	GntR	DVU2953	
Dde_0749	Desulfovibrio alaskensis str. G20	GntR	LldR	
Dde_1244	Desulfovibrio alaskensis str. G20	GntR	LldR	
Dde_3327	Desulfovibrio alaskensis str. G20	GntR	PhnF	S
Ddes_0341	Desulfovibrio desulfuricans ATCC 2774	GntR	DVU2644	
Ddes_1719	Desulfovibrio desulfuricans ATCC 2774	GntR	DVU2802	
Ddes_1162	Desulfovibrio desulfuricans ATCC 2774	GntR	Ddes_1162	S
Ddes_0854	Desulfovibrio desulfuricans ATCC 2774	GntR	Ddes_0854	S
DESPIG_01491	Desulfovibrio piger ATCC 29098	GntR	DVU2644	
DESPIG_00074	Desulfovibrio piger ATCC 29098	GntR	DESPIG_00074	S
DESPIG_01888	Desulfovibrio piger ATCC 29098	GntR	DESPIG_01888	S
DESPIG_02282	Desulfovibrio piger ATCC 29098	GntR	DESPIG_02282	S
DESPIG_02737	Desulfovibrio piger ATCC 29098	GntR	DESPIG_02737	S
Desal_3215	Desulfovibrio salexigens DSM 2638	GntR	DVU0030	
Desal_1386	Desulfovibrio salexigens DSM 2638	GntR	DVU2644	
Desal_0359	Desulfovibrio salexigens DSM 2638	GntR	DVU2802	
Desal_1467	Desulfovibrio salexigens DSM 2638	GntR	LldR	
Desal_0038	Desulfovibrio salexigens DSM 2638	GntR	Desal_0038	S
Desal_2890	Desulfovibrio salexigens DSM 2638	GntR	Desal_2890	S
Desal_2919	Desulfovibrio salexigens DSM 2638	GntR	Desal_2919	S
Desal_3557	Desulfovibrio salexigens DSM 2638	GntR	Desal_3557	S
Desal_3745	Desulfovibrio salexigens DSM 2638	GntR	Desal_3745	S
DMR_35550	Desulfovibrio magneticus RS-1	GntR	DVU0030	
DMR_17410	Desulfovibrio magneticus RS-1	GntR	DMR_17410	S
LI0606	Lawsonia intracellularis PHE/MN1-00	GntR	LI0606	S
Dbac_3052	Desulfomicrobium baculatum DSM 4028	GntR	DVU0030	
Dbac_3146	Desulfomicrobium baculatum DSM 4028	GntR	DVU2802	
Dbac_0439	Desulfomicrobium baculatum DSM 4028	GntR	DVU2953	
Dbac_1975	Desulfomicrobium baculatum DSM 4028	GntR	LldR	
Dbac_0812	Desulfomicrobium baculatum DSM 4028	GntR	PhnF	S
Dbac_1979	Desulfomicrobium baculatum DSM 4028	GntR	Dbac_1979	S
Dbac_2821	Desulfomicrobium baculatum DSM 4028	GntR	Dbac_2821	S
Dret_1471	Desulfohalobium retbaense DSM 5692	GntR	DVU0030	
Dret_0921	Desulfohalobium retbaense DSM 5692	GntR	DVU2802	
Dret_0156	Desulfohalobium retbaense DSM 5692	GntR	LldR	
Dret_0942	Desulfohalobium retbaense DSM 5692	GntR	Dret_0942	S

TABLE S2 Predicted binding sites of Crp/Fnr family TFs. [a]Positions relative to the start of translation

Regulon	Genome	Operon	Distance	Site
CooA	Desulfohalobium retbaense DSM 5692	Dret_1234-Dret_1235	-93	TTGTCAGCCAGCTGACAG
	Desulfomicrobium baculatum DSM 4028	Dbac_0934-Dbac_0933	-130	CTGTCAGCTAGGCGACAG
	Desulfovibrio alaskensis G20	Dde_3028-Dde_3029	-113	GTGTCAGCCAGCCGACAA
	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_0382-Ddes_0381	-118	TTGTCAGCCAGCCGACAG
	Desulfovibrio magneticus RS-1	DMR_25780-DMR_25790	-99	TTGTCGGCAATCTGACAG
	Desulfovibrio salexigens DSM 2638	Desal_2508-Desal_2507	-108	TTGTCGGGTACCCGACAA
	Desulfovibrio vulgaris Hildenborough	DVU2098-DVU2099	-187	TTGTCGGCTAGCCGACAG
	Desulfovibrio vulgaris str. Miyazaki F	DvMF_2233-DvMF_2232	-182	CTGTCGGCTAGCCGACAT
Dde_1200	Desulfovibrio alaskensis G20	Dde_1201-Dde_1204	-171	TTGTCTCTTGTGAGACAG
Desal_0411	Desulfovibrio salexigens DSM 2638	Desal_0410-Desal_0407	-103	TTGTCTCATATGAGACAG
Ddes_0528	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_0528	-162	TTGTGACTACAGTCACAG
		Ddes_1164	-99	TTGTGCGTATACTCACAG
Ddes_1411	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_1413	-227 -101	TTGTTGTTGGTGCAACAT ATGTTGCATGGGCAACAA
		Ddes_1412	-128	TTGTTGCAATGACAACAT
		Ddes_1414	-125	ATGTTGCACCAACAACAA
		Ddes_1411	-189 -39	ATGTTGTCATTGCAACAA ATGTTGCAACAGCATGAA
Ddes_1827	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_1828-Ddes_1829	-71	CCGTAACAATTGTTACGG
Ddes_2092	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_2093-Ddes_2095	-129	ATGTTGCATACGGAACAC
		Ddes_2092	-97	GTGTTCCGTATGCAACAT
Desal_0494	Desulfovibrio salexigens DSM 2638	Desal_0493	-81	CAGTGACCTGCGTCATTG
Desal_2066	Desulfohalobium retbaense DSM 5692	Dret_2254	-305	CATTGATCTCAATCAATG
		Dret_1643	-85	CATTGATGCAGATCAAGG
	Desulfovibrio magneticus RS-1	DMR_38630-DMR_38620	-82	CGTTGATCCACATCAATG
		DMR_38640	-87	CATTGATGTGGATCAACG
	Desulfovibrio salexigens DSM 2638	Desal_2066	-40	CATTGATCTCGATCAACA
		Desal_2272	-97	CGTTGATCTGGATCAATG
DESPIG_03026	Desulfovibrio piger ATCC 29098	DESPIG_03025- DESPIG_03026	-107 -87	CTGTGAAAAGATTCCCTG TTGTGACGATGGTCACAG
DvMF_1708	Desulfovibrio magneticus RS-1	DMR_29660-DMR_29690	-83	TTATCGTCCCCGCGATAG
	Desulfovibrio piger ATCC 29098	DESPIG_02289	-86	CTATCGGCGGGCTGATAG
	Desulfovibrio vulgaris str. Miyazaki F	DvMF_1711-DvMF_1707	-151	TTATCGTCCCCGCGATAG
MreC	Desulfovibrio alaskensis G20	Dde_0461-Dde_0467	-258	CTGTAGCAAATGCTACGG
	Desulfovibrio salexigens DSM 2638	Desal_1359-Desal_1364	-154	ATGTTGCATTTGCAACGG
	Desulfovibrio vulgaris Hildenborough	DVU0381-DVU0374	-163	CTGTTGCGGGTGCAACCG

DVU3111	Desulfovibrio vulgaris Hildenborough	DVU3110-DVU3109	-74	CTGTAGCCCAGGCTACGG
		DVU3108	-101	CCGTGATGCCCGCCACGG
		DVU3111	-122	CCGTGCTGGGCATGACTG
HcpR	Desulfohalobium retbaense DSM 5692	Dret_2386	-121	TTTTGACTCAGGTCAAAG
		Dret_2382	-105	TTTTGATGCAGGTCAAAG
		Dret_1280	-93	ATTTGATCCGCGTCAAAG
	Desulfomicrobium baculatum DSM 4028	Dbac_1379-Dbac_1378	-52	TTTTGACCTAGGTCAAGG
		Dbac_1823	-74	TTTTGACATGGCTCAAAG
		Dbac_0596	-94	TTTTGATCTGCGTCAAAG
		Dbac_2941	-82	TTTTGATCCAGCTCAAAG
	Desulfovibrio alaskensis G20	Dde_2303	-111	TTTTGACGTGGGTCACGC
		Dde_2390-Dde_2389	-80	ATGTGACCTGCATCACAG
		Dde_0503	-50	ATGTGACGCAGGTCACAG
		Dde_2641-Dde_2642	-153	TTGTGACTCCGGTCACAT
	Desulfovibrio magneticus RS-1	DMR_35370-DMR_35390	-94	TTTTGACTTAGGTCAAAC
		DMR_35330-DMR_35340	-93	GCTTGACGTAAGTCAAGG
		DMR_12610-DMR_12600	-64	ATTTGACTTAGGTCATGT
		DMR_35320	-41	CCTTGACTTACGTCAAGC
	Desulfovibrio salexigens DSM 2638	Desal_0751-Desal_0750	-151	TTTTGACTTAAGTCATAG
		Desal_2853	-51	CCCTGACTTAAGTCATTT
		Desal_2849	-238 -88	CCTTGACCTGAATCAAAT TTTTGATCTGAATCAAAT
		Desal_2852-Desal_2851	-77	AAATGACTTAAGTCAGGG
	Desulfovibrio vulgaris Hildenborough	DVU2543-DVU2544	-184	ATTTGACGCACGTCACAA
		DVU1080-DVU1081	-234	GTGTGACCCGCGTCACAG
		DVU3251	-95	CTGTGACCCACGTCACGG
	Desulfovibrio vulgaris str. Miyazaki F	DvMF_0988-DvMF_0989	-211	TTTTGATGTCCGTCACAT
		DvMF_2445	-122	ATGTGACCCCGATCACGG

TABLE S3 Predicted binding sites of ArsR family TFs. [a]Positions relative to the start of translation

Regulon	Genome	Operon	Distance	Site
ArsR	Desulfohalobium retbaense DSM 5692	Dret_1784	-20 -12	ATTTGGTTACTTTGCCAAGT ACTTTGCCAAGTATGCAAAC
		Dret_1783-Dret_1782	-49 -42	ATTTGGCTTTTTGGGAAAAT TTTTTGGGAAAATAGAAAAT
	Desulfomicrobium baculatum DSM 4028	Dbac_1906	-160 -38	ATTTGCACAAATGACCAAAT GCTTGGCAATATTGCCAAGT
		Dbac_1905-Dbac_1903	-154 -32	ACTTGGCAATATTGCCAAGC ATTTGGTCATTTGTGCAAAT
	Desulfovibrio alaskensis G20	Dde_3721-Dde_3722	-88 -39	ATTTGGCTGAATCGCCAAAT ATTTGGCGAAAAAGCCAAAT
		Dde_2777-Dde_2778	-31	ATTTGGCGAATCAGCCAAAC
		Dde_2776-Dde_2778	-11	GCTTGGCCTAATTGCCAAAT
	Desulfovibrio magneticus RS-1	DMR_37400-DMR_37420	-84 -76	ATTTTGCTATTTGGCAAAAC ATTTGGCAAAACATACAAAT
		DMR_37390	-47 -39	TTTTGGCTATTTTGCCAAAT ATTTTGCCAAATAGAAAACG
		DMR_20130-DMR_20110	-57 -49	ATTTTGCCTTTTGACCAAAT TTTTGACCAAATAGCAAATC
	Desulfovibrio piger ATCC 29098	DESPIG_02964	-31	ATTTTGCCAAATGACCAAAT
		DESPIG_02962	-30 -22	ATTTGGCAATTTTGCCAATA ATTTTGCCAATAAGGAATAT
	Desulfovibrio salexigens DSM 2638	Desal_3294-Desal_3293	-39 -31	AGTTTCGTATTTGGCAAAAT ATTTGGCAAAATATAAAGAC
		Desal_2313	-38 -30	GTTTTCCTCTTTGGCAAAAT CTTTGGCAAAATATCCAAAC
		Desal_3295	-42 -34	ATTTTTATATTTTGCCAAAC ATTTTGCCAAACAGGAAAGT
	Desulfovibrio vulgaris Hildenborough	DVU1645-DVU1644	-44 -36	ATTTGTCTATTTTGCCAAAT ATTTTGCCAAATAATCAAGT
		DVU1646	-142 -134	ACTTGATTATTTGGCAAAAT ATTTGGCAAAATAGACAAAT
	Desulfovibrio vulgaris str. Miyazaki F	DvMF_2281	-12	ATTTGGCGAAACATGCAAAT
ArsR2	Desulfomicrobium baculatum DSM 4028	Dbac_2827	-46	GTTTTGACATCTTGACATAT
	Desulfovibrio alaskensis G20	Dde_2790-Dde_2793	-143	ATTTTGATGACACTCAAAAT
	Desulfovibrio magneticus RS-1	DMR_01880	-226	ATTTTGGTAGTTTGAAAAAT
	Desulfovibrio salexigens DSM 2638	Desal_2205-Desal_2203	-42 -34	ATTTGGTTAATTTGCCAAAT AATTTGCCAAATATAAACAT

ArsR3	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_1925	-29	CAATTCAATATAAGTTGAAATA
		Ddes_1926	-43	CAAGGCAACATTAGTTGAATAG
	Desulfovibrio vulgaris str. Miyazaki F	DvMF_2391-DvMF_2390	-65	CAGTACAACAAGCATTGGCGTG
		DvMF_2392	-29	CAGTTCAACATTCGTTGTATTA
SahR	Desulfohalobium retbaense DSM 5692	Dret_2302-Dret_2300	-172	TTATCAACTTTTTCAGATAT
		Dret_2255	-102	ATATCAAAATATCTTGATAT
	Desulfomicrobium baculatum DSM 4028	Dbac_2305-Dbac_2303	-45	ATATCAGGATATCTTGATAT
	Desulfovibrio alaskensis G20	Dde_1498-Dde_1496	-18	ATATGCGGATACATTGATAT
		Dde_3135-Dde_3134	63	ACATCAACTTTTCTTGATTC
		Dde_2328	-293	ATATTGATGAATCTTGATAT
		Dde_2525	-69 -45	ATATTACAATAGGTCGATAT ATATAAGGCTATCTTGATAT
		Dde_2476	-131	TTATCAACTTAGTTTGATAT
	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_0040-Ddes_0039	-57 -35	ATATGCGGATAACTTGATAC GTATCAGCATAAATTGATTT
		Ddes_2008	-80	ATATCAATTTATCTTGATAT
		Ddes_2286-Ddes_2287	-18 -10	ATATCACAATACCTTGATAT ATACCTTGATATGCAGATAT
	Desulfovibrio magneticus RS-1	DMR_43430-DMR_43440	-18	ATATCAGGATATCTTGATAT
	Desulfovibrio piger ATCC 29098	DESPIG_02490	-73	TTATCAAGTTATATTGATAT
		DESPIG_01931- DESPIG_01930	-83 -61	ATATCCAGATAGGCTGATAT GCATCAAGCAAATTTGATAC
		DESPIG_01871- DESPIG_01870	-18 -10	ATATCACCATACCTTGATAT ATACCTTGATATGCAGATAT
	Desulfovibrio salexigens DSM 2638	Desal_1278	-111	ATATGCAAATTTATTTATAT
		Desal_0813-Desal_0814	-282 -274	ATATAACAATATTTTTATAT ATATTTTTATATATTGAATT
		Desal_3078-Desal_3077	-39	ATATAAAGGTTTTGTTATGT
	Desulfovibrio vulgaris Hildenborough	DVU3371	-341	ATATCAGTATTTCCTGATGT
		DVU0997	-87	ATATCAATCTATCTTGATAT
		DVU2448-DVU2449	-18	ATATGTGGATAGGTTGATAT
		DVU0606-DVU0607	-52	ACATCAAGTTTTCTTGATTT
	Desulfovibrio vulgaris str. Miyazaki F	DvMF_3034	-208	ATATCAAGATAGCTTGATAT
		DvMF_2579-DvMF_2580	-103	AAATCAAGTTTTCTTGATGT
		DvMF_2942-DvMF_2941	-18	ATATTCGGATAACTTGATAT

SmtB	Desulfohalobium retbaense DSM 5692	Dret_2262-Dret_2261	-47	TAATTGAACAATTGATCAAGCG
	Desulfomicrobium baculatum DSM 4028	Dbac_1470-Dbac_1469	-64	TACTTGAACAACTATTCAAGCG
	Desulfovibrio alaskensis G20	Dde_0132	-61 -45	CATATGAACGATTGCTCATATG CATATGTATGTTTGTTCATAAA
	Desulfovibrio salexigens DSM 2638	Desal_1113-Desal_1112	-42	TATATGAACAGTTGTTCATATA
	Desulfovibrio vulgaris Hildenborough	DVU3386	-19	CGCATGAACAGATGTTCATATG
	Desulfovibrio vulgaris str. Miyazaki F	DvMF_2066	-89	CATATGAACACGTGTTCATATG
Dbac_0377	Desulfomicrobium baculatum DSM 4028	Dbac_0377-Dbac_0375	-45	TTATCGTAATTCTACGATAT
Dbac_1936	Desulfomicrobium baculatum DSM 4028	Dbac_1936-Dbac_1938	-49	CATATTAATATTTGTTTATATG
Ddes_1264	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_1264-Ddes_1265	-177 -128	CATTTCGTCGCATGACGAAATA TATTTCGTCAAAGAACGAAATA
DESPIG_02495	Desulfovibrio piger ATCC 29098	DESPIG_02495- DESPIG_02496	-89 -81 -55	TAAAACTATAAAACTAGTTTTT TAAAACTAGTTTTTTAGTTGAC TGAAACTACTTTTTTAGTTTGA
DMR_39480	Desulfovibrio magneticus RS-1	DMR_39490	-42	TAATTCGAAAGATTTCGAAATC
		DMR_39480	-11	TAGTTCGAAACATGTCGAACAC

TABLE S4 Predicted binding sites of GntR family TFs. [a]Positions relative to the start of translation

Regulon	Genome	Operon	Distance	Site
Dde_1162	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Dde_1162	-119	ATTGAACAGGGACAGTTTTAT
		Dde_1163	-74	ATAAAACTGTCCCTGTTCAAT
Desal_0038	Desulfovibrio salexigens DSM 2638	Desal_0038-Desal_0037	-88	ATACTGGTAGGACCAGTAG
DVU0030	Desulfohalobium retbaense DSM 5692	Dret_1470-Dret_1469	-372	TCGAAATTGTATCTGTCTCAG
		Dret_1471	-40	CTGAGACAGATACAATTTCGA
	Desulfomicrobium baculatum DSM 4028	Dbac_3053-Dbac_3055	-64	CGCAAACTGTATCTGTACCGC
		Dbac_3052	-536	GCGGTACAGATACAGTTTGCG
	Desulfovibrio alaskensis G20	Dde_0158-Dde_0159	-127 -109	TAAAAACTGTATCTGTTCTGA TGAAAATTGTATCTGTGTTTG
		Dde_0157	-62 -44	CAAACACAGATACAATTTTCA TCAGAACAGATACAGTTTTTA
	Desulfovibrio magneticus RS-1	DMR_35550	-71	CCATCACAGACACAGTTTTCA
	Desulfovibrio salexigens DSM 2638	Desal_3216-Desal_3218	-118 -93	CTAAAACTGTTACGGTTCATA TAAAATCTGTATCTGTACTGA
		Desal_3215	-64 -39	TCAGTACAGATACAGATTTTA TATGAACCGTAACAGTTTTAG
	Desulfovibrio vulgaris Hildenborough	DVU0031-DVU0032	-431	TAAAAACTGTATCTGTCATCA
		DVU0030	-40	TGATGACAGATACAGTTTTTA
	Desulfovibrio vulgaris str. Miyazaki F	DvMF_2174-DvMF_2173	-124	GTTTGACTGTATCTGTTCAGG
		DvMF_2175	-196	CCTGAACAGATACAGTCAAAC
DVU2644	Desulfovibrio alaskensis G20	Dde_2582-Dde_2580	-176 -64	CGAATGTCGACATTTT AAATTGTCGACAATCG
	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_0339-Ddes_0340	-137	CAAATGTTGACATTTA
	Desulfovibrio piger ATCC 29098	DESPIG_01491	-282 -204	CGAATGTCGACATTTT CGAATGTCGACATACT
		DESPIG_01488-DESPIG_01486	-127 -49	AGTATGTCGACATTCG AAAATGTCGACATTCG
	Desulfovibrio salexigens DSM 2638	Desal_1393	-255	ATAATGTCGACATTTG
		Desal_1392-Desal_1391	-101 -67	GAGATGTCGACATATA AAAATGTCGACATTAG
	Desulfovibrio vulgaris Hildenborough	DVU2647-DVU2644	-42	AAAATGTCGACATTTA
DVU2802	Desulfohalobium retbaense DSM 5692	Dret_0921-Dret_0914	-108	AATTGGTCTATCCAAAA
	Desulfomicrobium baculatum DSM 4028	Dbac_3146-Dbac_3150	-140	TTTTGGTTCTACCAAAA
	Desulfovibrio alaskensis G20	Dde_1614-Dde_1617	-17	AATTGGTGGGACCAGAG
	Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774	Ddes_1719-Ddes_1723	-142	CTTTGGTATGACCAGAA
	Desulfovibrio salexigens DSM 2638	Desal_0359-Desal_0352	-286 -122 -17	ATCTGGTCCAACCAGAC ATCTGGTTGGACCAGTT AACTGGTTGGACCAGAT
	Desulfovibrio vulgaris Hildenborough	DVU2802-DVU2808	-16	ATCTGGTTGGGCCAGAA

DVU2953	Desulfomicrobium baculatum DSM 4028	Dbac_0439	-53	ATAAGGGCCAATTTT
		Dbac_0438	-72	AAAATTGGCCCTTAT
	Desulfovibrio alaskensis G20	Dde_0621	57	ATAAGGGACACTTTT
		Dde_0622-Dde_0625	-75	AAAAGTGTCCCTTAT
	Desulfovibrio vulgaris Hildenborough	DVU2953	-269	AAAAGGGGCACTTTT
		DVU2952	-71	AAAAGTGCCCCTTTT
LldR	Desulfohalobium retbaense DSM 5692	Dret_0156-Dret_0158	-35	CATTGGTAGGACCAATT
	Desulfomicrobium baculatum DSM 4028	Dbac_1975-Dbac_1976	-32	TATTGGTAGGACCAATT
		Dbac_0552	-70	GATTGGTCCTACCAATT
		Dbac_2068	-42	TATTGGTAGGACCAATT
	Desulfovibrio alaskensis G20	Dde_1244-Dde_1245	-31	AAGTGGTCAGACCAATT
	Desulfovibrio salexigens DSM 2638	Desal_1467-Desal_1465	-47 -17	TATCGGTCAGACCAATT AATTGGTAGGACCCATT
	Desulfovibrio vulgaris Hildenborough	DVU2785-DVU2783	-65 -49	GATTGGTACAACCAATT TATTGGTTGGACCAATC
	Desulfovibrio vulgaris str. Miyazaki F	DvMF_1205-DvMF_1203	-77 -61	AATTGGTTCAACCAGTG GAGTGGTTGTACCAATT
PhnF	Desulfomicrobium baculatum DSM 4028	Dbac_0812	-79	ATTTGTATAGACATCT
		Dbac_0813-Dbac_0821	-39	AGATGTCTATACAAAT
		Dbac_0829-Dbac_0826	-93	AGTTGTCTAGACCACT
	Desulfovibrio alaskensis G20	Dde_3327	-112	GGTTGTCTAGATAACT
		Dde_3328-Dde_3336	-40	AGTTATCTAGACAACC

Supplementary text. Detailed description of the reconstructed reference and singleton TF regulons

Crp/Fnr family regulons

Members of the Crp/Fnr family of TFs are involved in control of several functions such as respiration, nitrogen fixation, degradation of aromatic compounds, photosynthesis, and virulence. We constructed the neighbor-joining phylogenetic tree for 34 proteins from the Crp/Fnr family of TFs in *Desulfovibrionales* (Fig. 2). As a result, 24 regulatory proteins were clustered into five reference groups of three or more proteins, whereas the remaining 10 proteins formed singletons (Table 1). We reconstructed reference regulons for the Crp/Fnr family by applying a comparative genomics approach to each of the identified five groups of TFs. Then we used TF binding motifs and functional roles of the reference regulons to infer regulons for singleton TFs. Overall, well-conserved and singleton TFs of the Crp/Fnr family in *Desulfovibrionales* can be classified into three large groups, MreC-like, HcpR-like, and CooA-like. The group of MreC-like proteins includes three MreC orthologs and five singleton TFs. The transcription regulator MreC was previously experimentally described in *D. alaskensis* G20 as a putative activator of the *mre* operon that is involved in the metabolic process of uranium (VI) and chromate reduction (9). MreC orthologs are present in the three genomes, *D. alaskensis* G20, *D. salexigens*, and *D. vulgaris* str. Hildenborough. In all three genomes, the *mreC* regulatory gene is a third gene in the *mre* operon that is a likely target of MreC-dependent regulation (Fig. S1). Analysis of the *mreA* upstream region identified a candidate 18-bp MreC binding motif with the consensus sequence CnGTwGCAnnTGCwACnG. Subsequent genome scanning with this sequence motif did not reveal any additional conserved TFBSs, thus we concluded that MreC is a local regulator of the *mre* operon. The MreC regulon has a potential functional role in metal reduction since the *mreG* gene encodes a NADPH-dependent U(VI)- and Cr(VI)-reducing oxidoreductase, which is involved in the cytoplasmic heavy metal reduction in *D. alaskensis* (9). Other MreC-regulated genes from the *mre* operon are hypothetical NhaC-family transporter *mreA*, putative arylsulfatase *mreB*, thioredoxin *mreD*, and thioredoxin reductase *mreE*.

Five MreC-like proteins were classified as singleton TFs (Fig. S1). Analysis of the genomic context of two singleton TFs, Dde_1200 and Desal_0411, revealed that they are located within two operons containing gene sets that partially overlap with the reconstructed MreC reference regulon. These operons include homologs of *mreB*, *mreD*, and *mreE* genes from MreC regulon, thus suggesting a potential duplication of the *mre* operon in the common ancestor of *D. alaskensis* and *D. salexigens*. However, these operons lack homolog of NhaC-family transporter MreA which is substituted by a putative transporter from the sodium:sulphate transporter family. In upstream regions of both operons, we found an 18-bp pseudopalindromic motif, which is partially similar to a motif of the MreC reference regulon (Fig. S1).

The third MreC-like singleton TF DVU3111 is encoded in a unique genomic locus that also includes a paralog of the *mreA* transporter, a ferredoxin, and a hypothetical oxidoreductase. As the reference MreC regulon also includes a similar *mreA* gene, we built the training set for identification of a putative DVU3111-binding motif by combining the upstream gene regions of *DVU3111*, *DVU3110* and *DVU3108/mreA* genes. A common 18-bp DNA motif identified in this training set demonstrated significant similarity with the consensus motifs of other reconstructed regulons from the MreC subfamily.

The phylogenetic and genome context analysis of the last two MreC-like singleton TFs, Ddes_2092 and Ddes_1411, suggests that these regulators are close paralogs that both are adjacent to genes encoding hypothetical arylsulfotransferases. These two paralogous loci contain different transporter genes: *Ddes_2092* is co-localized with a transporter from the sodium/sulfate family, whereas *Ddes_1411* is co-localized with an MFS-type transporter. The identified 18-bp binding motif for the Ddes_2092 and Ddes_1411 singleton TFs has the consensus sequence wTGTTGCnnnnGCAACAw which is similar to the MreC binding motif.

In summary, the detailed analysis of MreC-like regulators demonstrates general applicability of the novel approach for the inference of local singleton regulons. The results suggest that the set of functional roles of genes controlled by the reference TF regulons can be iteratively enriched by func-

tional roles of genes from the analyzed singleton TF regulons and that this expanding set can be efficiently used for reconstruction of other singleton TF regulons.

A group of HcpR-like proteins from the Crp/Fnr family in the *Desulfovibrionales* includes 13 regulators, ten of which form two orthologous groups, whereas the remaining three regulators were classified as singleton TFs.

The HcpR regulon has been previously studied by comparative genomics in *D. vulgaris* and *D. alaskensis* and was predicted to include multiple genes involved in the nitrogen oxide and sulfate reduction (15). In this work, the detailed reconstruction of HcpR regulon in seven *Desulfovibrionales* genomes allowed us to revise both the HcpR-binding motif and the regulon content (Fig. S2). The revised motif associated with the reference HcpR regulog in *Desulfovibrionales* is an 18-bp pseudopalindrome with the consensus sequence nttTGACnnnnGCAaag. We compared the predicted contents of HcpR regulons reconstructed in this study and in our previous work (15). First, we confirmed that in all analyzed genomes the HcpR regulon includes the nitrosative stress response genes *hcp-frdX* encoding a hydroxylamine reductase Hcp and a ferredoxin-like protein FrdX, as well as the *DVU1080-DVU1081* operon encoding a hypothetical oxidoreductase and a membrane-bound polyferredoxin. Second, we found that the sulfate reduction genes *apsAB* and *sat* are not members of the HcpR regulon but they do belong to a global regulon controlled by the NADH-responsive transcription regulator Rex. Putative HcpR-binding sites previously identified in upstream regions of these genes are Rex-binding sites conserved in the nine *Desulfovibrionales* genomes including two bacteria, *D. piger* and *D. desulfuricans*, that lack HcpR protein (12). Third, the revised HcpR regulon was expanded to include a hypothetical transmembrane protein (DVU3251 in *D. vulgaris* str. Hildenborough and its orthologs in three genomes), a hypothetical cupin domain-containing protein (Dde_2303 in *D. alaskensis* and its orthologs in two genomes), and the nitrite reductase NrfHA in *D. salexigens* that has an important role in nitrosative stress response (6).

The second orthologous group of regulators within the HcpR subfamily is represented by the three proteins, Desal_2066, Dret_1643, and DMR_38640. Genomic context analysis of these *hcpR*-like genes did not identified conservatively co-localized genes but we identified a common

pseudopalindromic motif upstream of all three genes, which was similar to the HcpR binding motif in six positions (Fig. S2). A whole-genome search with this motif identified candidate binding sites located upstream of genes encoding a hypothetical FAD-dependent pyridine nucleotide-disulphide oxidoreductase in *D. salexigens* and *D. retbaense* but not in *D. magneticus*. We hypothesized that the DMR_38640 regulator in *D. magneticus* regulates the divergent *DMR_38640* and *DMR_38630-DMR_38620* operons by binding to a candidate site located in their common upstream region. Thus, the novel HcpR-like regulon has two target operons per genome including autoregulated *hcpR*-like genes and putative oxidoreductases that could be involved in the nitrosative stress.

Phylogenetic tree of Crp/Fnr proteins contains three HcpR-like singleton TFs, Ddes_0528, DESPIG_03026, and Ddes_1827. The latter two regulatory genes are co-transcribed with *hcp* genes encoding hydroxylamine reductases (Fig. S2). Analysis of upstream regions of all three these operons identified an 18-bp palindromic TFBSs that have some similarity to HcpR binding motif. An additional member Ddes_0528 singleton regulon in *D. desulfuricans* (*Ddes_1164*) encodes a hypothetical cupin domain-containing protein, which is orthologous to the previously identified member of the reference HcpR regulon. We concluded that the reference HcpR regulon and three HcpR-like singleton regulons in *Desulfovibrionales* have a similar functional role in the control of nitrosative stress response genes.

CooA-like regulators are represented by 11 TFs that are divided into two orthologous groups. The first group of regulators includes orthologs of the carbon monoxide-responsive regulator CooA from *Rhodospirillum rubrum* (17). In our previous analysis of two *Desulfovibrio* genomes (14), the predicted CooA regulon has included the carbon monoxide dehydrogenase (CODH) operon *cooSC* in both species and the CODH-associated hydrogenase operon *cooMKLXUHXF* in *D. vulgaris* (14). In the present study, CooA orthologs were identified in eight *Desulfovibrionales* genomes (all except *D. piger* and *L. intracellularis*). In all these genomes, we found the *cooA* gene to be adjacent to the *cooSC* operon that is preceded by a conserved CooA-binding site. We used these candidate sites from all seven *Desulfovibrionales* genomes to build a common CooA binding site motif with the nTGTCGGnnnnCCGACAn consensus, and then scanned the analyzed genomes for its presence. However, no additional CooA binding

sites have been identified in *Desulfovibrionales*, suggesting that CooA is a local regulator for the *cooSC* operon. Our comparative analysis did not confirm a putative CooA site that has been predicted upstream of the *cooMKLXUHXF* operon in *D. vulgaris*. First, this putative site has three substitutions in conserved positions of the revised CooA binding motif. Second, it is not conserved in other three *Desulfovibrionales* genomes that have orthologs of the *cooMKLXUHXF* operon.

The second orthologous group of CooA-like proteins includes three TFs, DvMF_1708 from *D. vulgaris* str. *Miyazaki F,* DMR_29690 from *D. magneticus*, and DESPIG_02293 from *D. piger*. A genomic context analysis of these CooA-like regulatory genes revealed that they are co-transcribed with genes encoding a hypothetical oxidoreductase and a ferredoxin that are similar to the subunits of adenylsulfate reductase (Fig. S3). These operons in *D. piger* and *D. magneticus* also include an upstream gene encoding a hypothetical CitT-like transporter, whereas the operon in *D. vulgaris* str. Miyazaki F includes transporter from the COG0730 family. A novel 18-bp pseudopalindromic motif was identified upstream of all three operons (Fig. S3). This motif is partially similar to the CooA binding motif, suggesting that it is a candidate binding motif of the CooA-like regulator. No other candidate binding sites were found with this motif in the three genomes, suggesting that DvMF_1708 and its orthologs are the local regulators of adjacent genes that may be linked to utilization of a yet undefined sulfate derivative.

Two singleton TFs in *D. salexigens*, Desal_0494 and Desal_3734, were not assigned to any of the above three Crp/Fnr family sub-groups. A candidate TFBS for the Desal_0494 singleton TF was identified in the upstream region of the downstream *Desal_0493* gene encoding a GlnB-like protein.

In conclusion, the five the Crp/Fnr family reference TF regulons in the *Desulfovibrionales* reconstructed in this work share several common features. First, these regulons use local regulators that were predicted to control from one to three target operons per genome. Second, the candidate binding motifs of the analyzed Crp/Fnr family regulators have the same 18-bp length and a common pair of nucleotides (thymine and adenosine in positions 4 and 15, respectively) that are also conserved in binding motifs of previously characterized regulators from the Crp/Fnr family, such as

Crp, Fnr, Dnr, ArcR, and NtcA (10, Table S2**).** Third, each of these regulons includes at least one gene encoding an oxidoreductase, suggesting that all Crp/Fnr-family TFs in *Desulfovibrionales* can potentially respond to redox-active substances, such as nitric oxide for HcpR (15) and carbon monoxide for CooA (17).

ArsR family regulons

A prototypical regulator from the ArsR family is the plasmid-encoded regulator ArsR that controls arsenical resistance operon from *E. coli* (16). Many ArsR family proteins are metal-sensing transcriptional repressors (reviewed in (11)) but some regulators of this family do not possess metal-binding domains and regulate other cellular functions. The only characterized member of this TF family in *Desulfovibrionales* is the ArsR protein from *D. alaskensis* G20, a regulator of an arsenic resistance *arsRBCC* operon (8).

Phylogenetic analysis of 38 ArsR family proteins from nine *Desulfovibrionales* genomes revealed four reference groups including 30 TFs and eight singleton TFs (Fig. 3). A comparative genomics approach was applied to each of reference groups, and four reference regulons were reconstructed as a result. Then, the obtained TF-binding motifs of reference regulons were used to characterize singletons.

SahR, a new regulator of S-adenosylhomocysteine metabolism. One of the ArsR family TFs, DVU0606, and its orthologs in other *Desulfovibrionales* genomes were found in a conserved operon with the adenosylhomocysteinase gene *ahcY*. We identified a conserved 20-nt motif upstream of the *DVU0606-ahcY* operons and proposed that this motif is a binding site of the DVU0606 regulator that was named the S-AdenosylHomocysteine Regulator SahR. Whole genome scanning with this motif identified additional conserved SahR binding sites upstream of genes involved in S-adenosylhomocysteine and methionine metabolism: *metE* (5-methyltetrahydropteroyltriglutamate homocysteine S-methyltransferase), *metF* (5,10-methylenetetrahydrofolate reductase), and *metK* (S-adenosylmethionine synthetase) (Fig. S4). Additional non-conserved SahR binding sites were also found upstream of the methionine transporter gene *metT* in *D. alaskensis* G20 and the homocysteine S-methyltransferase *bhmT* in *D. retbaensae*. Experimental validation of the predicted novel regulon for methionine metabolism in Desulfovibrionales is currently under way.

Metal detoxification SmtB regulon. The second reference group of ArsR family regulators present in six *Desulfovibrionales* genomes is similar to metal-sensing SmtB protein regulating metallothionein or zinc exporter in cyanobacteria (5, 18). Three of six SmtB-like regulatory genes in *Desulfovibrionales* constitute putative operons with genes encoding permeases, while other three genes are transcribed monocistronically (Fig. S5). A common palindromic motif identified upstream of all three dicistronic operons coincides with the known SmtB binding motif (2). In other three genomes, candidate TFBSs were identified by whole-genome scanning with three SmtB binding sites as a training set.

In *D. vulgaris* str. Miyazaki and *D. retbaensae*, the SmtB TFs regulate cation transport ATPases similar to a zinc exporter protein controlled by SmtB in *Synechocystis* PCC 6803 (18). In other four genomes, SmtB TFs regulate genes encoding predicted permeases from DUF318 family probably involved in an efflux of heavy metal cations.

Arsenic resistance ArsR regulon. Two reference TF groups of ArsR family regulators are chromosomally coupled to arsenic resistance genes and are orthologous similar to the previously described ArsR regulator from *D. alaskensis* G20 (8). For identification purposes, we designated the bigger group of TFs as ArsR, and the smaller group as ArsR2. A phylogenetic analysis clearly distinguished these two clades on the phylogenetic tree (Fig. 3). The reconstructed binding motifs of ArsR and ArsR2 regulators are quite similar, thus we cannot exclude a cross-binding between these regulators to heterologous sites in organisms containing both regulators. The reconstructed ArsR regulons contain between two and six genes per genome, and in most cases include the autoregulated *arsR* genes (Fig. S6). ArsR regulons often include genes encoding predicted permeases of DUF318 family and putative glutaredoxins. The ArsR2 regulons are more variable and may include arsenite efflux permeases, arsenate reductases and thioredoxin and thiol-disulphide interchange-like proteins. We conclude that the *Desulfovibrionales* species use different mechanisms for the arsenic resistance which are regulated by similar ArsR family TFs.

ArsR family singletons. Eight ArsR family proteins from the *Desulfovibrionales* genomes were classified as singleton TFs by the phylogenetic analysis. All ArsR family reference TF binding motifs have common motif structure (20 or 22 bp long with a palindromic repeat), so similar binding motifs were anticipated for ArsR family singleton TFs.

Two orthologous regulatory genes of ArsR family were found near operons encoding arsenic resistance genes, namely the *DvMF_2391-2390* operon encoding arsenite metallochaperone ArsD and arsenite-transporting ATPase ArsA and the *Ddes_1926* gene encoding SAM-dependent arsenic methyltransferase. A common binding motif was identified upstream of these two operons and regulatory genes (Fig. S6). There is no membrane protein associated with the ArsA ATPase in the DvMF_2392 regulon, but ArsR regulon in the same genome include *arsP* gene, and in other bacterial species ArsP protein was found to be associated with ArsA-ArsD transporting system (3). Another singleton regulator, Ddes_1264, is cotranscribed with a gene encoding a P-type ATPase, similar to transporters from SmtB regulon. Two similar strong 22-bp palindormes was identified upstream of the operon.

For reconstruction of other singletons of ArsR family, we looked for strong 22-bp palindromic repeats upstream of singleton regulatory genes that are similar to known TF binding motifs. Single candidate TFBSs sites were found upstream of other ArsR family singleton TFs, DMR_39480, Dbac_0377 and Dbac_1936. Three nearly identical 22-bp palindromes were found upstream of bicistronic operon encoding singleton TF DESPIG_02495 and a fusion protein of acyl-CoA synthase and MFS transporter.

Three ArsR family singleton TFs were predicted to control arsenic or heavy metal resistance, but the roles of other five singletons are unclear. DESPIG_02495 regulates a long fusion protein containing an acyl-CoA transferase domain that may be involved in secondary metabolism. Dbac_0377 is predicted to regulate a thioredoxin and an acetyl-CoA synthase-like enzyme that may participate in methyltransfer reactions. Other singleton regulons contain either enzymes of unknown specificity (hydrolase and FMN reductase) or hypothetical proteins.

In summary, the study of ArsR family regulons and singletons demonstrated how novel reconstructed motifs can be used in subsequent rounds of reconstruction, implicating operon predictions and functional annotation information where possible. Prevalence of autoregulated compact local regulons and preservation of binding motif features throughout the family helped us to reconstruct ArsR family regulatory systems in *Desulfovubrionales* genomes (Table S4).

GntR family regulons

We identified fifty three members of the GntR TF family in ten *Desulfovibrionales* genomes. Phylogenetic tree of GntR TFs from the *Desulfovibrionales* containing also the known representatives of all GntR subfamilies (13) was constructed, and distinct branches corresponding to different subfamilies were identified with high bootstrap values (data not shown). Twenty six, six and fifteen GntR family proteins were assigned to the FadR, HutC and MocR subfamilies, respectively (Fig. 4). Six proteins forming a distinct clade on the tree were not assigned to any known subfamily and may represent a new GntR subfamily.

FadR subfamily regulons. Almost a half of GntR family regulators in the *Desulfovibrionales* belong to the FadR subfamily. A phylogenetic analysis of this subfamily identified two reference regulons with five and six regulators, and fifteen singleton TFs. Reference regulons from FadR subfamily were predicted to regulate lactate metabolism and amino acid metabolism.

FadR subfamily members from one of the most conserved groups (DVU2785) are homologs of proteobacterial LldR regulator of lactate metabolism (1), so this reference regulon was designated LldR. In *Desulfovibrionales*, this local regulon typically consists of three genes, *lldR* (regulator), *lldD* (lactate dehydrogenase) and *lldX* (putative lactate transport protein), that are co-transcribed in most genomes. LldR binding sites were identified upstream of these operons in five of six genomes. In *D. baculatum*, an additional LldR binding site was found upstream of the *glcB* gene encoding malate synthase. In *D. alaskensis* G20 genome, a *lldR* paralog, *Dde_1244*, was identified that is co-transcribed with a gene encoding a lactate transporter paralog, Dde_1245. Surprisingly, in this genome LldR binding site was identified upstream of a *Dde_1244-Dde_1245* operon but not up-

stream of a *lldRDX* operon. LldX is a membrane protein homologous to TSUP family transporters (4-Toluene Sulfonate Uptake Permease). This protein is not related to LldP lactate transporters that are regulated by LldR in gamma- and beta-Proteobacteria. Cotranscription and coregulation of the *lldX* gene with *lldD* gene encoding lactate dehydrogenase suggests that main LldX function is a transport of lactate.

Another FadR subfamily gene, DVU2644, is co-transcribed with three other genes that are probably involved in amino acid metabolism. The first gene, DVU2647, annotated as endoribonuclease L-PSP, encodes a homolog of the YjgF protein which was recently shown to inhibit phosphoribosylamine synthesis in *Salmonella* (7). The second gene, DVU2646, encodes putative 1-aminocyclopropane-1-carboxylate deaminase, and we propose that the first gene may be an inhibitor of this enzyme. The third gene encodes transport protein homologous to Na+/H+ antiporters, and some transporters from this family are capable to transport amino acid. These three genes are regulated by DVU2644 and its orthologs in all five *Desulfovibrionales* genomes. DVU2644 TF appears to be functionally linked with amino acid metabolism, but the exact function of this regulon is unknown.

Reconstructed binding motifs of LldR and DVU2644 proteins are not similar to each other, with different length and sequence. To identify potential singleton binding sites, we used these two binding motifs for a site search with a relaxed threshold in upstream regions of all singleton regulatory genes from the FadR subfamily. Potential binding sites slightly similar to the LldR binding motif were found upstream of two regulatory genes, *Desal_0038* and *Dbac_2821*, but not upstream of any other FadR subfamily gene. Desal_0038 protein from FadR subfamily probably regulates an autoregulated operon encoding transcriptional regulator and glycerate kinase. We cannot also exclude the possibility of regulation of a divergently transcribed gene encoding methyl-accepting chemotaxis sensory transducer by Desal_0038. The singleton TF Dbac_2821 regulates an operon containing this regulator and a choline dehydrogenase, a betaine aldehyde dehydrogenase and a substrate-binding subunit of a glycine betaine transporter.

MocR subfamily regulons. Characteristic feature of the MocR subfamily is a presence of a long C-terminal domain similar to a pyridoxal phosphate-dependent transferase that probably has no enzymatic activity but required for ligand sensing. Two reference TF groups, DVU0030 and DVU2953, (found in seven and three genomes, respectively) and five singleton TFs were identified in this subfamily by phylogenetic analysis. TF binding motifs for reference regulators were identified by search for conserved motifs in upstream regions of regulatory genes. Both reference regulons control transport proteins possibly involved in amino acid metabolism. DVU0030 regulon contains a two-gene operon located divergently from the regulatory gene in each genome (Fig. S8). This operon encodes two putative amino acid transport proteins, but in *D. baculatum* an additional gene in the operon, *Dbac_3055*, encodes a putative aminotransferase. An autoregulation of a gene encoding the regulatory protein cannot be ruled out. So, the most probable function of this regulon is transport of unidentified amino acid or similar substance. The second MocR subfamily reference regulon, DVU2953, probably consists of the regulatory gene and a divergent gene encoding DMT family permease (Fig. S8). In *D. alaskensis* G20, the permease gene is disrupted by insertion of transposon. The function of this regulon may be linked with amino acid transport because DMT transporters have been shown to transport different amino acids in *E. coli*. Another possible function of this transport protein is drug resistance.
A candidate TFBS for the MocR subfamily singleton TF Ddes_1162 was identified upstream of the divergent *Ddes_1162* and *Ddes_1163* genes, the latter encoding a pyridoxamine 5'-phosphate oxidase-like protein.

HutC subfamily singletons. Six HutC-subfamily regulatory proteins were identified in the *Desulfovibrionales* genomes, and all of them were classified as singleton TFs by phylogenetic analysis. Two genes encoding similar HutC subfamily regulators, Dde_3327 and Dbac_0812, are adjacent to the two phosphonate metabolism operons. The PhnF regulator from the HutC subfamily has been shown to regulate the phosphonate metabolism operon in *Mycobacterium smegmatis* (4). Since no reference regulon was found in *Desulfovibrionales*, a PhnF binding motif from *M. smegmatis* was used for a singleton reconstruction. Binding sites weakly similar to PhnF were predicted upstream of both operons and upstream of a phosphonate transport operon in *D. baculatum* (Fig. S8). Search for similar sites upstream of other regulatory genes from HutC subfamily was unsuccessful.

DVU2802 regulator may represent a new subfamily within GntR family. Comparative genomics reconstruction of the DVU2802 regulon allowed us to find a candidate TFBSs upstream of only one operon in each genome (Fig. S9). In all six *Desulfovibrionales* genomes that contain DVU2802 orthologs, the first gene of a regulated operon encodes ortholog of DVU2802 TF. Other gene of this regulon, DVU2804, encodes secreted hydrolase from metallo-beta-lactamase family. Another members of the regulon probably participate in maturation (DVU2805) and transport (DVU2806-DVU2807) of the hydrolase protein outside of the cell. Gene content of the regulon varies between genomes, since three of six regulons lack genes encoding the transport system. A function of the regulon is probably determined by the specificity of secreted hydrolase DVU2804 which is unknown yet.

It should be noted that the GntR family analysis revealed a significant diversity of TF binding motifs throughout the family. In contrast to non-conserved regulators from the Crp/Fnr and ArsR families, only nine of thirty one GntR family singleton regulons were successfully reconstructed (Table S3). No common features such as a motif length, parity or conservation of distinct position were preserved at a family level, hampering prediction of binding motifs for non-conserved regulators. All reconstructed GntR family regulons are local and compact, with typically one or two binding sites per genome. Amino acid metabolism and transport are the most common functions for many members of GntR family regulons, but other metabolic functions may be regulated by GntR family membera as well.

References

1. **Aguilera L, Campos E, Giménez R, Badía J, Aguilar J, Baldoma L.** 2008. Dual role of LldR in regulation of the *lldPRD* operon, involved in L-lactate metabolism in *Escherichia coli*. J. Bacteriol. **190**:2997-3005.

2. **Bose M, Slick D, Sarto MJ, Murphy P, Roberts D, Roberts J, Barber RD.** 2006. Identification of SmtB/ArsR cis elements and proteins in archaea using the Prokaryotic InterGenic Exploration Database (PIGED). Archaea. **2**:39-49.

3. **Castillo R, Saier MH Jr.** 2010. Functional Promiscuity of Homologues of the Bacterial ArsA ATPases. Int. J. Microbiol. **2010**:187373.

4. **Gebhard S, Cook GM.** 2008. Differential regulation of high-affinity phosphate transport systems of Mycobacterium smegmatis: identification of PhnF, a repressor of the phnDCE operon. J. Bacteriol. **190**:1335-1343.

5. **Huckle JW, Morby AP, Turner JS, Robinson NJ.** 1993. Isolation of a prokaryotic metallothionein locus and analysis of transcriptional control by trace metal ions. Mol. Microbiol. **7**:177-187.

6. **Kern M, Volz J, Simon J.** 2011. The oxidative and nitrosative stress defence network of *Wolinella succinogenes*: cytochrome c nitrite reductase mediates the stress response to nitrite, nitric oxide, hydroxylamine and hydrogen peroxide. Environ. Microbiol. **13**:2478-2494.

7. **Lambrecht JA, Browne BA, Downs DM.** 2010. Members of the YjgF/YER057c/UK114 family of proteins inhibit phosphoribosylamine synthesis in vitro. J. Biol. Chem. **285**:34401-34407.

8. **Li X, Krumholz LR.** 2007. Regulation of arsenate resistance in *Desulfovibrio desulfuricans* G20 by an arsRBCC operon and an arsC gene. J. Bacteriol. **189**:3705-3711.

9. **Li X, Krumholz LR.** 2009. Thioredoxin is involved in U(VI) and Cr(VI) reduction in *Desulfovibrio desulfuricans* G20. J. Bacteriol. **191**:4924-4933.

10. **Novichkov PS, Laikova ON, Novichkova ES, Gelfand MS, Arkin AP, Dubchak I, Rodionov DA.** 2010. RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. Nucleic Acids Res. **38**(suppl 1):D111-D118.

11. **Osman D, Cavet JS.** 2010. Bacterial metal-sensing proteins exemplified by ArsR-SmtB family repressors. Nat. Prod. Rep. **27**:668-680.

12. **Ravcheev DA, Li X, Latif H, Zengler K, Leyn SA, Korostelev YD, Kazakov AE, Novichkov PS, Osterman AL, Rodionov DA.** 2012. Transcriptional regulation of central carbon and energy metabolism in bacteria by redox-responsive repressor Rex. J. Bacteriol. **194**:1145-1157.

13. **Rigali S, Derouaux A, Giannotta F, Dusart J.** 2002. Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J. Biol. Chem. **277**:12507-12515.

14. **Rodionov DA, Dubchak I, Arkin A, Alm E, Gelfand MS.** 2004. Reconstruction of regulatory and metabolic pathways in metal-reducing delta -proteobacteria. Genome Biol. **5**:R90.

15. **Rodionov DA, Dubchak IL, Arkin AP, Alm EJ, Gelfand MS.** 2005. Dissimilatory metabolism of nitrogen oxides in bacteria: comparative reconstruction of transcriptional networks. PLoS Comput. Biol. **1**:e55.

16. **San Francisco MJ, Hope CL, Owolabi JB, Tisa LS, Rosen BP.** 1990. Identification of the metalloregulatory element of the plasmid-encoded arsenical resistance operon. Nucleic Acids Res. **18**:619-624.

17. **Shelver D, Kerby RL, He Y, Roberts GP.** 1995. Carbon monoxide-induced activation of gene expression in *Rhodospirillum rubrum* requires the product of cooA, a member of the cyclic AMP receptor protein family of transcriptional regulators. J. Bacteriol. **177**:2157-2163.

18. **Thelwell C, Robinson NJ, Turner-Cavet JS.** 1998. An SmtB-like repressor from *Synechocystis* PCC 6803 regulates a zinc exporter. Proc. Natl. Acad. Sci. U. S. A. **95**:10728-10733.