Table S2. Functional properties of reconstruccted TF regulons in B. subtilis and related Bacillales.

TF name	*Bacillus subtilis sstr. 168*	*Bacillus amyloliquefaciens FZB42*	*Bacillus pumilus SAFR-032*	*Bacillus licheniformis DSM 13*	*Anoxybacillus flavithermus WK1*	*Geobacillus kaustophilus HTA426*	*Bacillus cereus ATCC 14579*	*Bacillus halodurans C-125*	*Bacillus clausii KSM-K16*	*Oceanobacillus iheyensis HTE831*	*Paenibacillus sp. JDR-2*	Functional Class*	Functional role	TF Family	Effector	Effector Evidence (PMID)	Previous Comparative genomics study (PMID)
Part I.	**36 regulators with previosuly unknown regulated genes and TFBS motifs**																
HisR (YerC)	9	9	9	9	9	2	0	9	9	10	8	**A**	Histidine biosynthesis	TrpR	Histidine	-	
BglR (YydK)	3	2	2	0	0	0	0	0	0	0	0	**C**	Beta-glucosides utilization	GntR	Salicin	-	
BglZ (YkvZ)	2	2	1	1	0	0	2	0	0	0	0	**C**	Beta-glucosides utilization	LacI	Beta-glucosides	-	
IolR1 (DegA)	2	0	0	1	0	13	0	6	0	0	7	**C**	Inositol utilization	LacI		-	
FruR	3	3	3	3	3	3	3	3	3	3	0	**C**	Fructose utilization	DeoR	D-fructose-6-phosphate	-	
GamR (YbgA)	3	0	0	7	0	0	0	6	6	6	0	**C**	Glucosamine utilization	GntR	Glucosamine-6-phosphate	-	
MdxR (YvdE)	7	0	0	7	0	0	0	0	0	0	0	**C**	Maltodextrin utilization	LacI		-	
MsmR	5	5	5	5	0	0	0	5	0	0	0	**C**	Alpha-galactosides utilization	LacI		-	
MurR (YbbH)	6	0	6	6	0	0	0	0	0	0	0	**C**	N-acetylmuramate utilization	RpiR	N-acetylmuramate-6-phosphate	-	
RbsR	7	6	6	6	6	6	5	6	6	6	6	**C**	Ribose utilization	LacI	Ribose	-	11750820
RhaR (YulB)	5	0	0	5	0	0	0	4	0	2	4	**C**	Rhamnose utilization	DeoR	Rhamnose	-	
RhgR (YesS)	17	0	0	19	0	0	0	0	0	0	3	**C**	Rhamnogalacturonan utilization	AraC		-	
RmgR (YtdP)	5	0	5	5	0	0	0	11	11	7	3	**C**	Rhamnogalacturonan utilization	AraC		-	
NrdR	4	4	4	4	4	3	3	3	4	3	4	**N**	Deoxyribonucleotide biosynthesis	LacI	Deoxyribonucleotides	-	15949864
YisR	1	1	1	1	1	1	0	0	0	0	0	**R**	Multidrug resistance	AraC		-	
YbzH	2	2	0	2	0	0	0	0	2	0	0	**R**	Hypothetical efflux transporter	ArsR		-	
YczG	3	3	0	3	0	0	2	0	0	0	0	**R**	Aromatic compound detoxification	ArsR		-	
YwrC	3	3	3	3	0	0	0	0	0	0	3	**R**	Metal efflux transporter	AsnC		-	
YhcF	5	5	0	0	0	0	5	3	2	0	4	**R**	Multidrug resistance	GntR		-	
YcxD	2	2	2	2	0	0	0	0	0	0	2	**R**	Multidrug efflux	GntR		-	
YdfD/YisV	4	0	0	0	0	0	4	2	2	0	2	**R**	Hypothetical efflux transporter	GntR		-	
YybR/YdeP	5	3	0	2	0	0	2	2	2	3	2	**R**	Hypothetical substance detoxification pa	HxlR		-	
YwbI	3	3	3	2	0	0	0	0	0	0	2	**R**	Cell integrity protection	LysR		-	
LytT	2	2	2	2	0	0	2	0	0	0	0	**R**	Cell integrity protection	LytTR	LytS, sensor histidine kinase	-	
YuaC	2	2	2	2	0	0	0	0	2	0	0	**R**	Osmotic stress response	MarR		-	
YvbF/YvaV	10	10	5	5	0	0	0	0	0	0	0	**R**	Osmotic stress response	MarR		-	
YybA	5	4	0	4	0	0	0	3	3	2	2	**R**	Polyamine homeostasis	MarR		-	

YdfL	1	0	0	1	0	0	1	1	0	0	0	**R**	Multidrug resistance	MerR		-	
YrkD	6	7	7	11	19	18	9	0	0	6	0	**R**	Hypothetical substance detoxification pa	RcnR		-	
GlcR	2	2	0	2	0	0	0	2	0	2	0	**U**		DeoR		-	
YdfF	2	0	0	3	0	0	0	2	3	0	0	**U**		ArsR		-	
YhdI/YdeL	4	2	0	2	0	0	4	0	0	0	0	**U**		GntR		-	
YhgD	2	2	2	2	2	3	1	2	0	0	0	**U**		TetR		-	
YizB	3	0	0	0	0	0	0	0	0	2	0	**U**		PadR		-	
YtcD	3	3	3	4	0	0	0	0	0	0	3	**U**		HxlR		-	
YvfU	4	0	0	6	0	0	4	0	0	0	4	**U**		LuxR	YvfT, sensor histidine kinase	-	
Part II.	**27 regulators with previously known regulated genes but unknown TFBS motifs**																
AlsR	3	3	3	3	0	0	0	0	0	0	0	**C**	Acetoin production	LysR		-	
AnsR	3	3	3	3	0	0	3	0	0	0	0	**A**	Asparagine degradation	Xre	Asparagine	-	
ArsR	4	0	0	0	0	0	4	2	2	2	0	**R**	Arsenic resistance	ArsR	Arsenite; Cadmium, ion (Cd2+); Sil	15948947	
AseR	2	2	0	2	3	6	0	3	3	3	0	**R**	Arsenic resistance	ArsR	Arsenite	15948947	
BirA	10	10	5	7	7	7	10	7	3	2	7	**O**	Biotin biosynthesis	BirA	Biotin	8892842	12368242
BsdA	5	5	0	5	0	0	0	0	0	0	0	**R**	Salicylic acid resistance	LysR	Salicilate	17295427	
CcpB	6	2	0	0	0	0	0	0	0	0	0	**C**	Carbon catabolism	LacI		-	
CitR	2	2	0	2	0	0	0	0	2	0	0	**C**	Citrate metabolism	LysR		-	
CssR	4	4	4	4	0	0	0	0	0	3	0	**R**	Protein secretion stress response	OmpR	CssS, sensor histidine kinase	-	
CzrA	3	3	1	2	2	5	3	0	2	2	2	**R**	Zinc resistance; Cadmium resistance; C	ArsR	Silver ion, (Ag+); Cadmium, ion (Cd	15948947	
DctR	1	1	1	0	0	1	0	0	0	0	0	**C**	C4-dicarboxylate transport	DeoR		-	
FatR	2	2	2	2	0	0	1	0	0	0	0	**R**	Toxic fatty acid stress response	TetR	Oleic acid; γ-Linolenic acid; Linolei	11734890	
FrlR	6	7	0	0	0	0	0	0	0	0	0	**A,C**	Fructoselysine utilization	GntR	Fructoselysine 6-phosphate	-	
GanR	6	0	6	6	0	0	0	5	5	0	0	**C**	Galactan utilization	LacI		-	
GltR	2	0	0	0	0	0	0	0	0	0	3	**U**		LysR		-	
GlvR	3	3	0	3	0	0	0	0	0	0	0	**C**	Maltose utilization	RpiR	Maltose-6-phosphate	-	
GmuR	8	7	5	6	0	5	0	5	0	0	0	**C**	Glucomannan utilization	GntR		-	
GudR	6	0	0	6	0	0	0	0	6	7	0	**C**	Glucarate utilization; Galactarate utilizat	GntR		-	
KipR	7	7	6	7	0	5	7	2	0	0	0	**S**	Sporulation	IclR		-	
LutR	5	5	3	5	5	4	3	4	4	5	0	**C**	Lactate utilization	GntR	L-lactate	-	
ManR	3	3	0	3	0	0	0	0	0	0	0	**C**	Mannose utilization	BglG	ManP, mannose-specific enzyme II	-	
MtlR	3	3	3	3	5	1	0	0	0	4	0	**C**	Mannitol utilization	BglG	MtlA, mannitol-specific enzyme IIC	20444094	
NsrR	4	4	4	5	0	2	0	1	2	2	0	**R**	Nitrosative stress response	Rrf2	Nitric oxide	19006327	16261196
NtdR	4	0	4	4	0	0	0	0	0	0	0	**C**	Neotrehalosadiamine biosynthesis	LacI	3,3'-neotrehalosadiamine	14612444	
YcnK	3	3	0	3	0	0	0	0	0	0	0	**M**	Copper homeostasis	DeoR	Copper ion, (Cu+)	-	
YfmP	2	2	2	2	0	0	2	0	0	0	3	**R,M**	Metal efflux	MerR		14663075	
YvcP	2	0	0	0	0	0	0	2	2	0	2	**R**	Bacitracin resistance	OmpR	YvcQ, sensor histidine kinase	-	
Part III.	**66 regulators with previously known binding motifs**																
AcoR	4	4	4	1	0	0	4	4	0	0	0	**C**	Acetoin utilization	Fis	Acetoin	-	
AdaA	3	3	0	2	0	0	4	0	3	0	4	**R**	Adaptative response to DNA alkylation	AraC		-	
AraR	13	13	5	13	6	7	0	11	0	5	4	**C**	Arabinose utilization	GntR	Arabinose	10417639	11750820

ArgR	18	15	14	16	10	10	11	12	9	11	7	**A**	Arginine biosynthesis; Arginine degrada	ArgR	Arginine	1312212	11305941
BceR	2	2	2	2	0	2	0	2	0	2	0	**R**	Bacitracin resistance	OmpR	BceS, sensor histidine kinase (bacit	-	
BkdR	8	8	8	8	8	8	8	5	5	7	0	**A**	Branched-chain amino acid degradation	Fis		-	
BltR	2	2	2	0	0	0	0	1	0	0	0	**R**	Multidrug resistance	MerR		-	
BmrR	2	0	0	1	0	0	0	0	0	0	0	**R**	Multidrug resistance	MerR		-	
Btr	4	4	2	4	0	0	0	0	4	0	0	**M**	Iron homeostasis	AraC	Bacillibactin; Fe-Bacillibactin	17725565	
CcpA	217	171	144	195	50	44	44	119	119	81	25	**C**	Carbon catabolism	LacI	HPr, phosphocarries protein	19202299	
CcpC	5	4	2	2	0	0	1	0	0	2	0	**C**	Citrate metabolism	LysR	Citrate	14636591	
CcpN	2	2	2	2	1	2	2	2	2	2	0	**C**	Gluconeogenesis	CcpN		-	
CggR	6	6	6	6	6	6	2	6	6	6	6	**C**	Glycolysis	SorC	Fructose-1,6-bisphosphate	12622823	
CitT	1	1	1	1	0	0	1	1	1	3	0	**C**	Citrate utilization	CitB	Citrate	10972810	
CodY	156	143	87	97	62	27	51	40	12	49	34	**A**	Amino acid metabolism	CodY	GTP; Branched-chain amino acids	15228537	
ComA	13	8	7	5	0	0	0	0	0	0	0	**S**	Competence	LuxR	ComP, sensor histidine kinase (Cor	-	
CsoR	3	2	2	2	3	3	2	2	3	4	7	**R,M**	Copper resistance	RcnR	Copper ion, (Cu+)	19249860	
CtsR	8	9	9	10	7	8	10	9	8	7	7	**R**	Heat shock response	CtsR	Heat shock	20852588	
CymR	32	31	14	11	16	4	16	14	11	13	29	**A**	Cysteine metabolism	Rrf2	O-acetyl-L-serine; CysK, cysteine s	16513748	
CysL	3	3	3	3	11	3	0	5	1	11	2	**I**	Sulfite reduction	LysR	Sulfate; Sulfite; Thiosulfate	-	
DeoR	3	3	3	3	0	0	4	0	0	0	0	**N**	Deoxynucleoside utilization	SorC	Deoxyribose-5-phosphate	10714997	
DesR	3	3	3	3	0	0	0	0	3	1	0	**R**	Cold shock response	LuxR		-	
ExuR	10	5	4	5	0	0	0	0	0	0	0	**C**	Glucuronate/Galacturonate utilization	LacI	Glucuronate; Galacturonate	-	
FadR	11	10	14	14	12	14	10	13	12	10	0	**F**	Fatty acid degradation	TetR	Palmitoyl-CoA; Oleoyl-CoA	17189250	
FapR	10	10	10	13	8	8	15	9	8	11	10	**F**	Fatty acid biosynthesis	FapR	Malonyl-CoA	16932747	
Fnr	10	8	2	13	0	1	1	6	0	0	1	**P**	Anaerobic metabolism	Fnr/Crp	Oxigen	8846791	
Fur	49	36	33	50	7	4	38	28	22	21	15	**M**	Iron metabolism	Fur	Iron ion, (Fe2+)	10400588	
GabR	3	3	4	4	0	0	0	0	0	0	0	**A**	Gamma-aminobutyrate utilization	GntR	Pyridoxal-5-phosphate; Gamma-am	15223311	
GlnL	2	2	4	1	0	0	4	2	0	0	0	**A**	Glutamine/Aspartate utilization	GlnL	GlnK, sensor histidine kinase (gluta	-	
GlnR	6	5	3	3	3	16	4	0	0	2	7	**A**	Nitrogen assimilation	MerR	Feedback-inhibited GlnA, glutamine	18195355	17086994
GltC	3	3	3	3	3	3	0	0	0	0	0	**A**	Glutamate metabolism	LysR	2-oxoglutarate; Glutamate; Glutama	17134717; 17608797	
GntR	4	0	0	4	0	0	0	0	0	3	0	**C**	Gluconate utilization	GntR	D-gluconate; D-glucono-delta-lacto	2492998	
GutR	3	3	0	0	0	0	0	0	0	0	0	**C**	Sorbitol utilization	GutR	Sorbitol	11118449	
HrcA	5	5	5	5	5	5	5	6	6	6	6	**R**	Heat shock response	HrcA	Heat shock	12486078	
HxlR	2	2	2	2	0	0	0	0	0	2	2	**C**	Ribulose monophosphate pathway	HxlR		-	
IolR	12	12	0	12	0	0	0	0	10	0	0	**C**	Inositol utilization	DeoR	2-deoxy-5-keto-D-gluconate 6-phosphate	18310071	
KdgR	6	0	5	6	0	0	0	0	0	0	0	**C**	Pectin utilization	LacI	2-keto-3-deoxy-gluconate	-	
LevR	5	0	0	5	0	0	0	0	0	0	0	**C**	Levan utilization	LevR		-	
LexA	49	42	34	32	29	18	15	21	20	24	15	**R**	SOS response	LexA	DNA damage	9045831	12137951
LiaR	6	6	6	6	5	5	5	6	5	5	3	**R**	Cell wall-active antibiotics stress respon	LuxR	LiaS, sensor histidine kinase (bacitr	15273097	16816187
LicR	4	4	4	0	0	0	0	0	5	5	0	**C**	Beta-glucosides utilization	BglG	LicB, lichenan-specific enzyme IIB	10438772	
LmrA	5	2	0	0	0	0	0	0	0	0	0	**R**	Multidrug resistance	TetR	Quercetin; Fisetin; Ccatechin	17483215	
MalR	4	3	3	3	0	0	0	0	0	0	0	**C**	Malate utilization	CitB	YufL, sensor histidine kinase (malat	-	

MhqR	7	7	5	6	3	1	6	0	3	3	4	**R**	2-Methylhydroquinone and catechol resi	MarR		-	
MntR	5	1	4	1	4	1	4	3	4	4	4	**M**	Manganase transport	DtxR	Manganase ion, (Mn2+)	10760146	
NagR	4	4	4	5	0	3	3	4	2	1	0	**C**	N-acetylglucosamine utilization	GntR	N-acetylglucosamine-phosphate	20047956	
NiaR	6	6	3	6	5	6	6	6	6	0	0	**O**	NAD biosynthesis	NiaR	Niacine	18276644	18276644
OhrR	1	1	1	1	0	1	2	0	0	2	3	**R**	Peroxide stress response	MarR	tert-Butyl hydroperoxide; Diamide; (	11983871	
PadR	3	2	2	1	0	0	0	0	0	0	0	**R**	Phenolic acid stress response	PadR	p-coumaric acid; Ferulic acid	21685295	
PerR	15	18	11	17	12	10	17	11	9	15	13	**R**	Oxidative stress response; Peroxide stre	Fur	Hydrogen peroxide; Manganese ior	19508286	
PurR	24	24	26	27	19	22	21	25	26	22	7	**N**	Purine metabolism	PurR	5-phosphoribosyl 1-pyrophosphate	7638212	
PutR	4	4	3	4	0	0	0	0	0	0	0	**A**	Proline utilization	SrmR	Proline	21840319	
QdoR	5	2	1	2	0	0	0	0	0	0	0	**R**	Multidrug resistance	TetR	Quercetin; Fisetin; Tamarixetin; Gal	17483215	
Rex	10	8	9	12	9	3	8	2	3	8	7	**P**	Anaerobic metabolism	Rex	NADH	17015645	22210771
RocR	7	7	3	4	0	5	1	5	2	0	0	**A**	Arginine degradation; Ornithine degrada	Fis	Ornithine; Citrulline	-	
SdpR	2	0	0	2	2	2	2	2	2	0	2	**S**	SdpC antitoxin system involved in sporu	ArsR	SdpI, signal transduction protein	16469701	
SinR	19	19	17	19	0	0	0	3	0	0	0	**S**	Biofilm formation	Xre	SinI, antirepressor protein	20652907	
TnrA	106	85	49	63	27	32	0	50	61	27	70	**A**	Nitrogen assimilation	MerR	Feedback-inhibited GlnA, glutamine	21435182	17086994
TreR	3	3	3	3	3	3	2	3	2	0	0	**C**	Trehalose utilization	GntR	Trehalose-6-phosphate	8755887	
XylR	5	5	5	10	0	8	0	12	6	7	3	**C**	Xylose utilization	ROK	Xylose	7966270	11750820
YdfI	1	1	0	1	0	0	0	0	0	0	0	**R**	Metabolite transport	LuxR		-	
YetL	2	2	0	0	0	0	0	0	0	0	0	**R**	Flavonoids response	MarR	Flavonoids	-	
YodB/CatR	6	5	6	8	2	0	7	5	0	3	5	**R**	Oxidative stress response	HxlR	Quinone; Diamide	20652907	
YtlI	13	10	0	0	0	0	0	0	0	0	0	**A**	L-cystine transporter	LysR		-	
YtrA	7	7	5	5	3	3	4	0	0	0	0	**R**	Ramoplanin resistance	GntR	Acetoin	21856850	
Zur	10	8	8	10	5	3	5	9	10	7	8	**M**	Zinc homeostasis	Fur	Zink ion, (Zn2+)	9811636	12904577
	953	776	613	801	356	348	393	494	445	402	346		Number of regulated genes				
	131	108	91	115	45	55	61	65	60	59	52		Number of regulons				

***Functional classes: P - respiration; F - fatty acid metabolism; A - amino acid metabolism; R - stress response and drug resistance; C - carbon metabolism;**
S - sporulation, social behavior and competence; M - metal homeostasis; N - nucleic bases metabolism; O - cofactor metabolism; I - miscellaneous; U - unknown