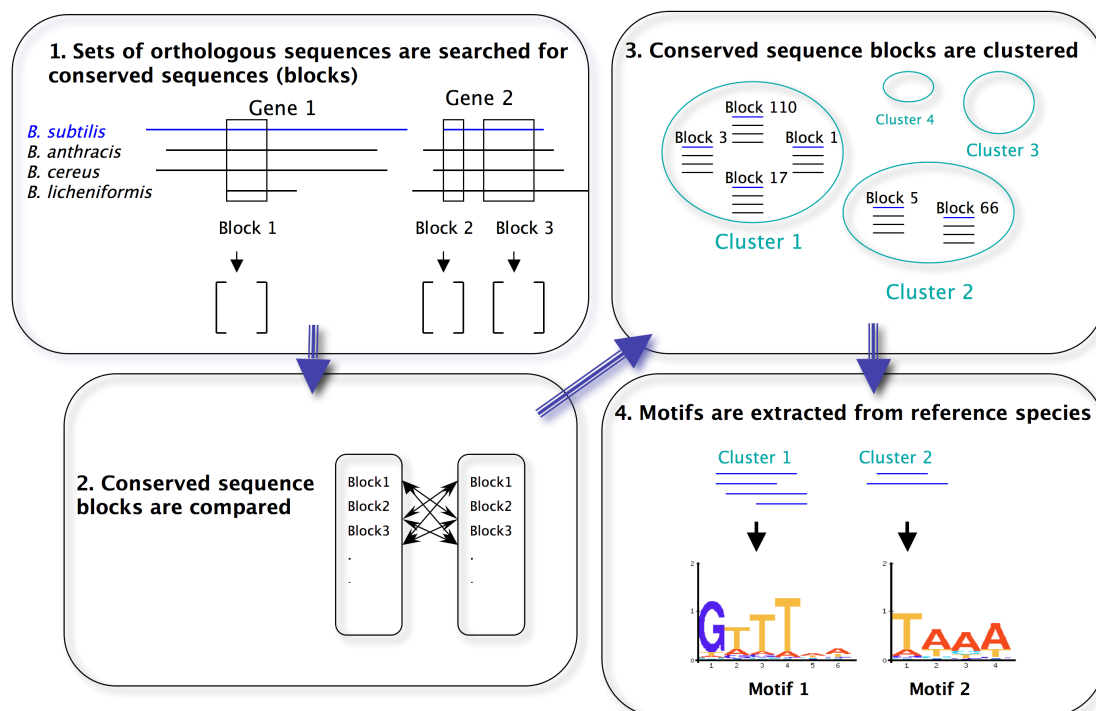# Supporting material

Inferring the transcriptional network of *Bacillus subtilis*

Figure S1. Workflow of the de *novo* motif detection. 1) Search for motifs in the orthologous space: conserved sequences (blocks) are extracted from the orthologous intergenic regions of four Bacillus species by means of BlockSampler, a Gibbs-sampling algorithm. The sequence of *B. subtilis* is indicated in blue. Detected conserved blocks are represented as PWMs; 2) Search for motifs in the co-expression space: All PWMs are aligned pair-wise and similarity scores are obtained 3) Similarity scores are used as input for clustering in order to identify groups of sequences sharing a common motif (or block as represented by a PWM); 4) Each block in a cluster corresponds to a different motif site of one particular regulator, conserved across the four different species. From each block within a cluster the sequence of *B. subtilis* is used to construct a *B. subtilis* specific motif model.

Evaluation of the *de novo* motif detection method
Predicted motif sites were compared with known motif sites reported in DBTBS. There are 605 known motif sites for the genes used in the study (i.e. genes that had orthologs across all four species). When a predicted motif site overlapped with a known site in DBTBS by at least 5 bp, the predicted site was considered a true positive (TP). We defined the following measures

Recall = TP / (TP+FN)

Precision = TP / (TP+FP)

Density of hits = TP / length of concatenated potential motif sites

where FN = false negatives, and FP = false positives. False negatives correspond to sites in DBTBS that were not recovered by our motif detection. Note that in this

calculation all predicted sites that are not overlapping with a known site are considered false positives. The latter assumption underestimates the precision of the method.

The phylogenetic footprinting step (BlockSampler output) resulted in the discovery of 7543 conserved sites in *B. subtilis*. The recall and precision after this step were 89% and 15.7% respectively; the density of hits was 0.68 hits/kb. Clustering similar blocks followed by motif extraction reduced the number of sites to 393, with a recall of 8.4% and a precision of 13.7% and a density of hits of 12.9 hits/kb.

As can be expected, reducing the number of motifs by exploiting the conservation between motifs results in a serious drop of sensitivity. The specificity remains approximately constant but the density of hits increases 19 times. This means that the initial blocks (phylogenetically conserved regions in single genes) contain sequences that are much larger than a motif. By using the co-regulation space these individually conserved regions could successfully be reduced to the cores of the conserved motifs that were shared by multiple genes. Note also that in the calculation of the precision, all novel predictions are considered FP. As a result, the slight drop in precision after reducing the number of motifs means that we lost relatively more known motif models than novel ones, i.e. relatively more novel motifs fulfill the more stringent parameters than known motifs.

Selecting parameter values for DISTILLER

To determine optimal parameter settings for DISTILLER, we performed several test runs in which the main parameters were varied (see table S1). Predicted interactions resulting from each run where compared with known interactions documented in DBTBS. For each run, recall and precision were estimated as follows:

- Recall = TP / (TP+FN)
- Precision = TP / (TP+FP)

where TP = true positives, FN = false negatives, and FP = false positives. All used parameter settings tested are listed in Table S1, and the results are plotted in Figure S2. Test 14 resulted in the best trade off between precision and recall (measured by precision x recall). Therefore the parameter settings corresponding to test 14 were applied in our analysis: minimum number of genes in a module = 4; minimum number of conditions in a module = 30; number of randomizations = 100,000; box p-value threshold = 0.001; bandwidth = 0.00001.

In general, the effect of varying parameter settings on the ability of DISTILLER to recall known interactions versus predicting new ones was as follows:

1. Increasing the minimum number of genes in a module from 3 to 4 did not have a definitive effect.
2. Increasing the minimum number of arrays in a module from 30 to 50 did not have a definitive effect.
3. Increasing the motif threshold from 0.995 (p-value of 0.005) to 0.999 (p-value of 0.001) had a better recall x precision. This is not surprising, since the known motif sites have a score of 1, and increasing the threshold results in excluding many predicted instances.
4. Increasing the bandwidth threshold from a p-value of 0.0001 to 0.00001 yielded worse recall x precision. Bandwidth is a measure of the "tightness" of the genes' co-expression in the sets of arrays tested, and here it seems that

relaxing the threshold could recall more of the known interactions.

Table S1 Different parameter settings that were used to test the parameter sensitivity. Tests printed in gray could not be completed because of computational complexity.

Test	Motif threshold	No of genes	No of conditions	Bandwidth p-value	Randomizations
test 1	0.995	3	30	0.0001	100,000
test 2	0.995	3	30	0.00001	100,000
test 3	0.995	3	50	0.0001	100,000
test 4	0.995	3	50	0.00001	100,000
test 5	0.995	4	30	0.0001	100,000
test 6	0.995	4	30	0.00001	100,000
test 7	0.995	4	50	0.0001	100,000
test 8	0.995	4	50	0.00001	100,000
test 9	0.999	3	30	0.0001	100,000
test 10	0.999	3	30	0.00001	100,000
test 11	0.999	3	50	0.0001	100,000
test 12	0.999	3	50	0.00001	100,000
test 13	0.999	4	30	0.0001	100,000
test 14	0.999	4	30	0.00001	100,000
test 15	0.999	4	50	0.0001	100,000
test 16	0.999	4	50	0.00001	100,000

Figure S2 Assessing different parameter settings on the precision and recall of DISTILLER. For the different parameter sets in table S1 precision (B), recall (A), and the product of precision and recall (C) were plotted for up to 150 best scoring modules. Parameter setting for test 14 was used in our analysis; it showed the best recall and near optimum precision.

A

Precision for different parameter settings

B

Recall for different parameter settings

C

Precision x recall for different parameter settings

4
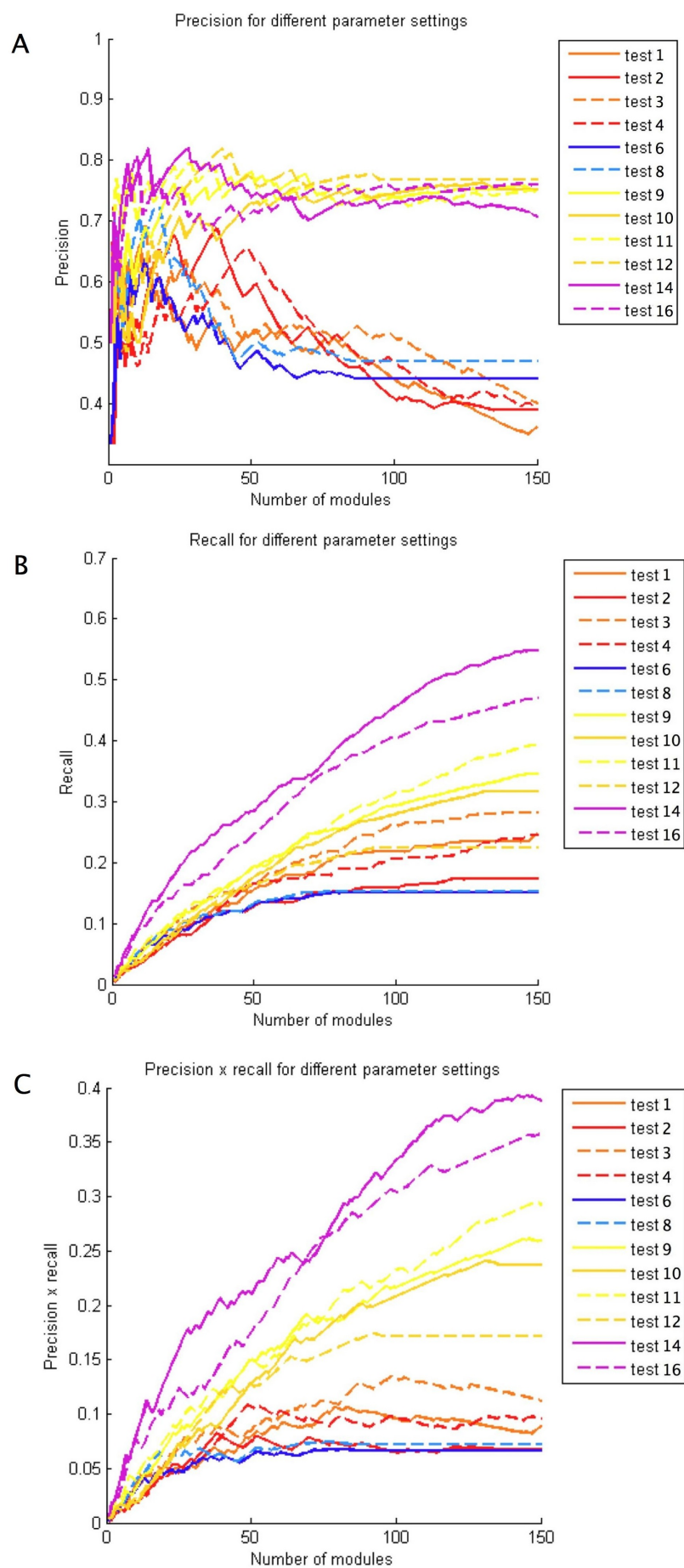
Table S2. Choosing an optimum genomic distance to assign regulators to groups of genes.

Distance	Recall	Precision	Recall x Precision
0	0.4	0.338028169	0.135211268
1047	0.491666667	0.286407767	0.140817152
2000	0.625	0.262237762	0.163898601
2100	0.641666667	0.264604811	0.169788087
2300	0.666666667	0.261437908	0.174291939
2400	0.675	0.255520505	0.172476341
2500	0.691666667	0.25382263	0.175560653
2600	0.691666667	0.250755287	0.173439074
2700	0.691666667	0.246290801	0.170351137
2800	0.691666667	0.238505747	0.164966475
2900	0.691666667	0.233802817	0.161713615
3000	0.691666667	0.229916898	0.159025855
3500	0.708333333	0.222513089	0.157613438
4000	0.716666667	0.209756098	0.150325204
4500	0.725	0.201388889	0.146006944
6000	0.75	0.18	0.135
1763378	0.991666667	0.00895208	0.008877479
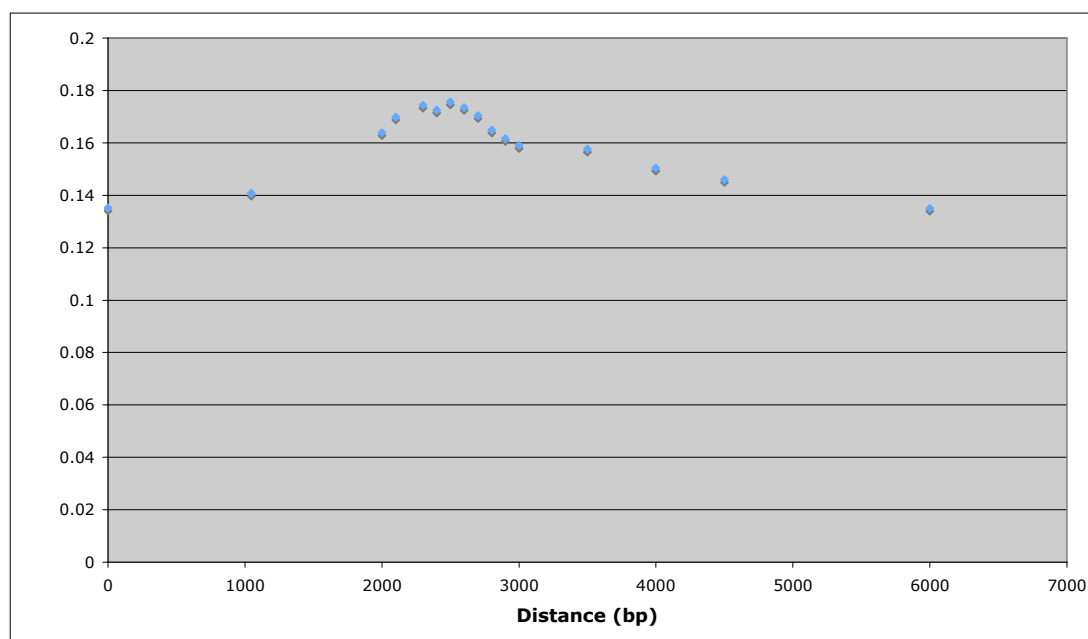
Figure S3. Choosing an optimum genomic distance to assign regulators to groups of genes.

Caption for Table S2 and Figure S3.

The distance between a regulator and its closest target gene was calculated for all documented regulators in DBTBS. The distance measure was taken from the middle of the coding sequences of the genes, taking circularity of the genome into account. The maximum, median, and minimum of those distances were 1763378, 1047, and 0 (auto-regulators), respectively (underlined in table S1). The recall and precision of the method of assigning a regulator to a group of target genes based on the genomic distance was calculated according to the formulas:

Recall = TP/(TP+FN)

Precision = TP/(TP+FP)
where TP = true positives, FN = false negatives, and FP = false positives.
True positives were calculated as follows: for all the known regulons of regulators documented in DBTBS, the number of the assigned regulators based on a certain genomic distance that match the true ones are true positives. False negatives are all those true regulators that were missed by our method, and the false positives are all the regulators wrongfully assigned to the regulons. To assess which of the distances measured best fit our method, the value of (recall x precision) was plotted for each distance, and the distance of 2500 bp had the highest value (highlighted in table S1).

Table S3. List of predicted novel interactions and their mode of regulation. Each predicted regulator-gene interaction is listed in the corresponding columns. A positive or negative mode of regulation is listed in the column 'Regulation', as well as the module in which this interaction was predicted.

Regulator	Gene	Regulation	Module
AbrB	*appB*	Negative	12
AbrB	*appC*	Negative	12
AbrB	*appD*	Negative	12
AbrB	*pspA*	Negative	12
AbrB	*ydjG*	Negative	12
AbrB	*ydjI*	Negative	12
AbrB	*ywoF*	Negative	12
AbrB	*pksC*	Negative	37
AbrB	*pksD*	Negative	37
AbrB	*qcrA*	Negative	37
AbrB	*qcrB*	Negative	37
AbrB	*qcrC*	Negative	37
AbrB	*rapG*	Negative	37
AbrB	*rapH*	Negative	37
AbrB	*rapK*	Negative	37
AbrB	*sacT*	Negative	37
AbrB	*ybaJ*	Negative	37
AbrB	*ycbJ*	Negative	37
AbrB	*yheJ*	Negative	37
AbrB	*ykuU*	Negative	37
AbrB	*yqeB*	Negative	37
AbrB	*yqfA*	Negative	37
AbrB	*yurI*	Negative	37
AbrB	*yvzA*	Negative	37
AbrB	*ywfB*	Negative	37
AbrB	*ybdG*	Negative	60
AbrB	*yczC*	Negative	60
AbrB	*yocD*	Negative	61
AbrB	*yqjZ*	Negative	61
AbrB	*yxiF*	Negative	61
AbrB	*yxxF*	Negative	61
AbrB	*appA*	Negative	62
AbrB	*appF*	Negative	62
AbrB	*yjcH*	Negative	62
AbrB	*yjiB*	Negative	62
AbrB	*yocS*	Negative	62

AbrB	yueE	Negative	62
AbrB	ywaA	Negative	62
AbrB	acpA	Negative	104
AbrB	cspB	Negative	104
AbrB	tnrA	Negative	104
AbrB	veg	Negative	104
AbrB	yceC	Negative	104
AbrB	yceD	Negative	104
AbrB	yceK	Negative	104
AbrB	ykoF	Negative	104
AbrB	yktD	Negative	104
AbrB	ykuR	Negative	104
AbrB	ykzG	Negative	104
AbrB	ylmC	Negative	104
AbrB	ymfC	Negative	104
AbrB	yobQ	Negative	104
AbrB	yraM	Negative	104
AbrB	yceE	Negative	130
AbrB	yceF	Negative	130
AbrB	yokG	Negative	130
AbrB	yuiE	Negative	130
AbrB	yrdC	Negative	131
AbrB	hom	Negative	132
AbrB	mdr	Negative	132
AbrB	oppA	Negative	132
AbrB	ybgG	Negative	132
AbrB	ycgM	Negative	132
AbrB	ydeI	Negative	132
AbrB	yfhP	Negative	132
AbrB	ygaF	Negative	132
AbrB	yoaE	Negative	132
AbrB	yqgB	Negative	132
AbrB	yrpB	Negative	132
CcpA	glpQ	Negative	51
CcpA	glpT	Negative	51
CcpA	infC	Negative	51
CcpA	jag	Negative	51
CcpA	kbl	Negative	51
CcpA	rocB	Negative	51
CcpA	rplT	Negative	51
CcpA	rpmB	Negative	51
CcpA	rpmI	Negative	51
CcpA	spoIIIJ	Negative	51
CcpA	ybxF	Negative	51
CcpA	ycbF	Negative	51
CcpA	ycsA	Negative	51
CcpA	yesL	Negative	51
CcpA	yfiG	Negative	51
CcpA	yulD	Negative	51
CcpA	yvdF	Negative	51
CcpA	yvdH	Negative	51
CcpA	yvdJ	Negative	51
CcpA	msmR	Negative	52

CcpA	*mtlA*	Negative	52
CcpA	*pmi*	Negative	52
CcpA	*sacA*	Negative	52
CcpA	*sacP*	Negative	52
CcpA	*ysbA*	Negative	52
CcpA	*epr*	Negative	83
CcpA	*lplA*	Negative	83
CcpA	*manP*	Negative	83
CcpA	*manR*	Negative	83
CcpA	*punA*	Negative	83
CcpA	*ydhM*	Negative	83
CcpA	*ydhT*	Negative	83
CcpA	*yncC*	Negative	83
CcpA	*yugN*	Negative	83
CcpA	*xsa*	Negative	84
CcpA	*ydeR*	Negative	84
CcpA	*ywcE*	Negative	84
CcpA	*ywcJ*	Negative	84
CcpA	*ywhE*	Negative	84
CcpA	*dhaS*	Negative	86
CcpA	*gutR*	Negative	86
CcpA	*ykoM*	Negative	86
CcpA	*yslB*	Negative	86
CcpA	*yurN*	Negative	86
CcpA	*citH*	Negative	114
CcpA	*galK*	Negative	114
CcpA	*pta*	Negative	114
CcpA	*yrzF*	Negative	114
CcpA	*yshB*	Negative	114
CcpA	*gcaD*	Negative	115
CcpA	*prs*	Negative	115
CcpA	*ybcF*	Negative	115
CcpA	*ylbN*	Negative	115
CcpA	*ytkP*	Negative	115
CcpA	*ytxD*	Negative	115
CcpA	*cccA*	Negative	116
CcpA	*cstA*	Negative	116
CcpA	*rocA*	Negative	116
CcpA	*rocC*	Negative	116
CcpA	*rocD*	Negative	116
CcpA	*rocE*	Negative	116
CcpA	*rocF*	Negative	116
CcpA	*yckB*	Negative	116
CcpA	*ydzA*	Negative	116
CcpA	*yodJ*	Negative	116
CcpA	*yojA*	Negative	116
CcpA	*ypmQ*	Negative	116
CcpA	*ytcP*	Negative	116
CcpA	*dctB*	Negative	136
CcpA	*sigI*	Negative	136
CcpA	*yydK*	Negative	136
CodY	*appC*	Negative	55
CodY	*appD*	Negative	55

CodY	*bioI*	Negative	55
CodY	*comGD*	Negative	55
CodY	*yjcG*	Negative	55
CodY	*ykvW*	Negative	55
CodY	*yqzE*	Negative	55
CodY	*yrkF*	Negative	55
CodY	*yuiA*	Negative	55
CodY	*yuxL*	Negative	55
CodY	*ywcE*	Negative	55
CodY	*yxiA*	Negative	55
CodY	*yxjG*	Negative	55
ComK	*pksI*	Positive	22
ComK	*ywfM*	Positive	22
ComK	*yydF*	Positive	22
ComK	*lexA*	Positive	81
ComK	*parC*	Positive	81
ComK	*parE*	Positive	81
ComK	*ruvA*	Positive	81
ComK	*ruvB*	Positive	81
ComK	*uvrA*	Positive	81
ComK	*uvrB*	Positive	81
ComK	*yhaM*	Positive	81
ComK	*yhaO*	Positive	81
ComK	*yqjY*	Positive	81
ComK	*yckE*	Positive	138
ComK	*yjbF*	Positive	138
ComK	*yqzG*	Positive	138
ComK	*bioA*	Positive	139
ComK	*bioB*	Positive	139
ComK	*bioD*	Positive	139
ComK	*bioF*	Positive	139
ComK	*bioI*	Positive	139
ComK	*bioW*	Positive	139
ComK	*clpP*	Positive	139
ComK	*dnaK*	Positive	139
ComK	*grpE*	Positive	139
ComK	*hrcA*	Positive	139
ComK	*metS*	Positive	139
ComK	*pabA*	Positive	139
ComK	*pabB*	Positive	139
ComK	*pabC*	Positive	139
ComK	*rph*	Positive	139
ComK	*slr*	Positive	139
ComK	*spoVB*	Positive	139
ComK	*yacF*	Positive	139
ComK	*ycgM*	Positive	139
ComK	*yisT*	Positive	139
ComK	*yokB*	Positive	139
ComK	*ytbQ*	Positive	139
ComK	*ywfB*	Positive	139
ComK	*ywrO*	Positive	139
ComK	*yxiF*	Positive	139
ComK	*pksC*	Positive	141

ComK	rapK	Positive	141
ComK	ybdN	Positive	141
ComK	ybfO	Positive	141
ComK	ycbJ	Positive	141
ComK	yfmK	Positive	141
ComK	yvaW	Positive	141
ComK	yvaX	Positive	141
ComK	yvaY	Positive	141
ComK	yvaZ	Positive	141
ComK	yvbA	Positive	141
ComK	yvzA	Positive	141
ComK	yweA	Positive	141
ComK	ywqC	Positive	141
ComK	yybJ	Positive	141
ComK	yybK	Positive	141
ComK	yybL	Positive	141
ComK	yybM	Positive	141
ComK	yybN	Positive	141
ComK	yydG	Positive	141
ComK	yydH	Positive	141
ComK	yydI	Positive	141
ComK	yydJ	Positive	141
ComK	ycgL	Unknown	80
ComK	ydfL	Unknown	80
ComK	ydhH	Unknown	80
ComK	ywbO	Unknown	80
ComK	ywoB	Unknown	80
ComK	atpI	Unknown	82
ComK	hemL	Unknown	82
ComK	kinB	Unknown	82
ComK	rpsL	Unknown	82
ComK	rsbT	Unknown	82
ComK	sipV	Unknown	82
ComK	tnrA	Unknown	82
ComK	yabD	Unknown	82
ComK	yceK	Unknown	82
ComK	ydiQ	Unknown	82
ComK	ynfE	Unknown	82
ComK	yobQ	Unknown	82
ComK	yodF	Unknown	82
ComK	yogA	Unknown	82
ComK	yqeL	Unknown	82
ComK	yrvI	Unknown	82
ComK	yvzD	Unknown	82
ComK	ywqB	Unknown	82
ComK	yybB	Unknown	82
ComK	degS	Unknown	109
ComK	ilvB	Unknown	109
ComK	ilvC	Unknown	109
ComK	leuA	Unknown	109
ComK	leuB	Unknown	109
ComK	leuC	Unknown	109
ComK	leuD	Unknown	109

ComK	*lonA*	Unknown	109
ComK	*nprE*	Unknown	109
ComK	*pycA*	Unknown	109
ComK	*sigH*	Unknown	109
ComK	*yfhB*	Unknown	109
ComK	*yjcH*	Unknown	109
ComK	*yjcI*	Unknown	109
ComK	*yjcJ*	Unknown	109
ComK	*yjcS*	Unknown	109
ComK	*yjdA*	Unknown	109
ComK	*yoaD*	Unknown	109
ComK	*yoaE*	Unknown	109
ComK	*yobK*	Unknown	109
ComK	*yokD*	Unknown	109
ComK	*ypdA*	Unknown	109
ComK	*ypfD*	Unknown	109
ComK	*yqgP*	Unknown	109
ComK	*yqjI*	Unknown	109
ComK	*yrpB*	Unknown	109
ComK	*yurK*	Unknown	109
ComK	*ywfH*	Unknown	109
ComK	*ywnF*	Unknown	109
ComK	*yxjG*	Unknown	109
ComK	*hemB*	Unknown	140
ComK	*ytaG*	Unknown	140
Fur	*fhuB*	Negative	1
Fur	*yetG*	Negative	1
Fur	*yxeC*	Negative	8
Fur	*yclM*	Negative	46
Fur	*ykuJ*	Negative	46
Fur	*fhuC*	Negative	105
Fur	*ykvW*	Negative	122
Fur	*ypbB*	Negative	123
Fur	*clpP*	Unknown	58
Fur	*dps*	Unknown	58
Fur	*gspA*	Unknown	58
Fur	*ydaN*	Unknown	58
Fur	*yflA*	Unknown	58
Fur	*ykzA*	Unknown	58
Fur	*yocB*	Unknown	58
Fur	*yqgZ*	Unknown	58
Fur	*yumB*	Unknown	58
Fur	*yvaA*	Unknown	58
Fur	*ywiE*	Unknown	58
Fur	*yxzF*	Unknown	58
GerE	*ylbD*	Unknown	2
GerE	*yisZ*	Unknown	6
GerE	*yitA*	Unknown	6
GerE	*yitB*	Unknown	6
LexA	*ymaC*	Negative	9
LexA	*yprA*	Negative	9
LexA	*yvsG*	Negative	9
LexA	*pcrB*	Negative	128

PerR	dhbA	Unknown	8
PerR	dhbC	Unknown	8
PerR	feuA	Unknown	8
PerR	feuB	Unknown	8
PerR	feuC	Unknown	8
PerR	fhuB	Unknown	8
PerR	fhuD	Unknown	8
PerR	ybbA	Unknown	8
PerR	ydbN	Unknown	8
PerR	yfhA	Unknown	8
PerR	yfiY	Unknown	8
PerR	yfiZ	Unknown	8
PerR	yfmC	Unknown	8
PerR	yhfQ	Unknown	8
PerR	ykuN	Unknown	8
PerR	ykuO	Unknown	8
PerR	yoaJ	Unknown	8
PerR	yusV	Unknown	8
PerR	ywbL	Unknown	8
PerR	ywbM	Unknown	8
PerR	ywbN	Unknown	8
PerR	yxeB	Unknown	8
PerR	yxeC	Unknown	8
PerR	citZ	Unknown	29
PerR	tagB	Unknown	29
PerR	yurY	Unknown	29
PerR	dhbB	Unknown	137
PerR	dhbE	Unknown	137
PerR	dhbF	Unknown	137
PerR	ybbB	Unknown	137
PerR	yetG	Unknown	137
PerR	ykuP	Unknown	137
PerR	yuiI	Unknown	137
ResD	bglP	Positive	45
ResD	hemA	Positive	45
ResD	hemB	Positive	45
ResD	narG	Positive	45
ResD	narH	Positive	45
ResD	narI	Positive	45
ResD	narJ	Positive	45
ResD	ydbL	Positive	45
ResD	yhfO	Positive	45
ResD	yqxI	Positive	45
ResD	yueJ	Positive	45
ResD	yxiE	Positive	45
SigB	rsbR	Positive	7
SigE	yisZ	Positive	33
SigE	yitA	Positive	33
SigE	yitB	Positive	33
SigE	comER	Positive	72
SigE	sspB	Positive	72
SigE	ybaK	Positive	72
SigE	yhcN	Positive	72

SigE	*yogA*	Positive	72
SigE	*yvlA*	Positive	72
SigE	*ylbI*	Positive	74
SigE	*sigG*	Positive	75
SigE	*xsa*	Positive	75
SigE	*cgeC*	Positive	101
SigE	*yfiA*	Positive	101
SigE	*yfnF*	Positive	101
SigE	*yfnG*	Positive	101
SigE	*yfnH*	Positive	101
SigE	*yqfQ*	Positive	101
SigE	*gapB*	Positive	125
SigE	*ykuD*	Positive	125
SigE	*oppB*	Positive	127
SigF	*csbX*	Positive	50
SigF	*yfhE*	Positive	50
SigF	*yfhF*	Positive	50
SigH	*ywfG*	Positive	43
SigH	*rapE*	Positive	87
SigH	*rapG*	Positive	87
SigH	*rapH*	Positive	87
SigH	*yqjZ*	Positive	87
SigH	*citG*	Positive	120
SigW	*appB*	Positive	12
SigW	*appC*	Positive	12
SigW	*appD*	Positive	12
SigW	*ywoF*	Positive	12
SigW	*dnaK*	Positive	28
SigW	*grpE*	Positive	28
SigW	*hrcA*	Positive	28
SigW	*ydgC*	Positive	28
SigW	*yjbI*	Positive	28
SigW	*ypuD*	Positive	28
SigW	*yrzG*	Positive	28
SigW	*yugM*	Positive	28
SigW	*yutM*	Positive	28
SigW	*ywdE*	Positive	28
SigW	*ywsA*	Positive	28
SigW	*yczC*	Positive	56
SigW	*ykuA*	Positive	56
SigW	*yugI*	Positive	57
SigW	*yuzE*	Positive	98
SigW	*yqgU*	Positive	99
SigW	*lytR*	Positive	129
SigW	*sunT*	Positive	129
SigW	*yncM*	Positive	129
SigW	*yxbC*	Positive	129
SigW	*yxbD*	Positive	129
Spo0A	*ywkC*	Positive	31
Spo0A	*yhjM*	Unksown	32
Spo0A	*yodL*	Unknown	32
SpoIIID	*dacB*	Negative	10
SpoIIID	*spoIIP*	Negative	11

SpoIIID	*yloB*	Negative	17
SpoIIID	*yoaN*	Positive	69
SpoIIID	*cotV*	Unknown	13
SpoIIID	*cotW*	Unknown	13
TnrA	*tuaA*	Positive	66
TnrA	*tuaB*	Positive	66
TnrA	*tuaC*	Positive	66
TnrA	*tuaD*	Positive	66
TnrA	*tuaF*	Positive	66
TnrA	*yrbD*	Positive	66

Table S4. Predicted intra-operonic motif sites. All genes predicted to have a motif site upstream are listed, together with their operons, and the predicted motif. If our prediction was listed by de Hoon *et al.* a 'one' is indicated.

Gene	Operon	Predicted motif	Predicted by de Hoon *et al.*
acoB	*acoABCL*	motif_14	0
appC	*appDFABC*	CodY	0
araP	*araABDLMNPQ-abfA*	CcpA	1
bioI	*bioWAFDBI-ytbQ*	CodY	0
cgeB	*cgeAB*	motif_75	0
comEB	*comEABC*	ComK	0
degU	*degSU*	motif_187	1
dppC	*dppABCDE*	CodY	0
flhO	*usd-spoIIID-mbl-flhOP*	motif_106	1
folD	*nusB-folD*	PurR	0
glpK	*glpFK*	CcpA	0
glpK	*glpFK*	motif_106	0
hemC	*hemAXCDBL*	PerR	0
hemL	*hemAXCDBL*	ComK	0
hemL	*hemAXCDBL*	motif_187	0
iolD	*mmsA-iolBCDEF-idh-iolHI-fbaB*	CcpA	1
mbl	*usd-spoIIID-mbl-flhOP*	motif_14	1
mmgD	*mmgABCDE-yqiQ*	SigE	0
nupC	*deoR-dra-nupC-pdp*	CcpA	0
nupC	*deoR-dra-nupC-pdp*	motif_14	0
oppB	*oppABCDF*	SigE	0
punA	*drm-punA*	CcpA	0
purM	*purEKBCSQLFMNHD*	PurR	0
qcrC	*qcrABC*	AbrB	0
rbsC	*rbsRKDACB*	CcpA	0
rocB	*rocABC*	CcpA	0
rplT	*infC-rpmI-rplT-ysdA*	CcpA	0
rpsF	*yyaF-rpsF-ssb-rpsR*	motif_2	1
rpsR	*yyaF-rpsF-ssb-rpsR*	motif_2	0
rsbT	*rsbRSTUVW-sigB-rsbX*	ComK	0
sipV	*yhjE-sipV*	ComK	0
tagB	*tagAB*	PerR	0
trpF	*trpEDCFBA-hisC-tyrA-aroE*	motif_187	0
yacF	*pabBAC-sul-folBK-yazB-yacF-lysS*	ComK	0
yacK	*ctsR-mcsAB-clpC-radA-yacK*	motif_53	0
yceD	*yceCDEFGH*	AbrB	0

ydjI	*pspA-ydjGHI*	SigW AbrB	0
yfhE	*yfhFED*	motif_14	1
yfiA	*malA-yfiA-malP*	SigE	0
yfnD	*yfnHGFED*	motif_187	0
yhaM	*yhaONM*	ComK	1
ylxF	*fla-che*	motif_106	0
ylxR	*ylxS-nusA-ylxRQ-infB-ylxP-rbfA*	motif_2	0
yoxB	*yoxCB-yoaA*	motif_53	0
yqcK	*arsR-yqcK-arsBC*	motif_53	1
yqfA	*yqeZ-yqfA-yqfB*	AbrB	0
yqzE	*comGABCDEFG-yqzE*	CodY	1
yrvI	*relA-yrvI*	ComK	1
ytxH	*ytxGHJ*	motif_260	0
yurG	*yurHG*	motif_3	1
yvfM	*yvfKLM-lacA-yvfO*	motif_14	0
ywcH	*nfrA-ywcH*	motif_53	1
yxiE	*bglPH-yxiE*	ResD	0

Table S5. A list of all arrays and their assigned conditional categories. The column 'Array' refers to the experiment according to the source database (SMD, the first 3 arrays; GEO, the rest), while 'Category' refers to one of the 6 conditional categories, or to a control experiment.

Array	Category
C_26966	peroxide
C_26967	peroxide
C_26968	peroxide
GSM104507	control
GSM104511	control
GSM104514	control
GSM104517	DNA
GSM104518	DNA
GSM104519	DNA
GSM104520	DNA
GSM104521	DNA
GSM104522	DNA
GSM104523	DNA
GSM104524	DNA
GSM104525	DNA
GSM104526	control
GSM104527	control
GSM104528	control
GSM104727	DNA
GSM104728	DNA
GSM104729	DNA
GSM104730	DNA
GSM104731	DNA
GSM104732	DNA
GSM104733	DNA

GSM104734	DNA
GSM104735	DNA
GSM105249	sporulation
GSM105250	sporulation
GSM105251	sporulation
GSM105252	DNA
GSM105253	DNA
GSM105254	DNA
GSM105255	DNA
GSM105256	DNA
GSM105257	DNA
GSM105258	DNA
GSM105259	DNA
GSM105260	DNA
GSM105261	control
GSM105262	control
GSM105263	control
GSM105264	DNA
GSM105265	DNA
GSM105266	DNA
GSM105267	DNA
GSM105268	DNA
GSM105269	DNA
GSM105270	DNA
GSM105271	quorum response
GSM105272	DNA
GSM105273	DNA
GSM105280	quorum response
GSM105281	quorum response
GSM105282	quorum response
GSM105283	quorum response
GSM105284	quorum response
GSM105285	quorum response
GSM105286	quorum response
GSM105287	quorum response
GSM105288	quorum response
GSM105289	quorum response
GSM105290	quorum response
GSM105326	control
GSM105327	control
GSM105328	control
GSM105329	quorum response
GSM105330	DNA
GSM105331	DNA
GSM105332	DNA
GSM105345	DNA
GSM105346	quorum response
GSM105347	DNA
GSM105348	DNA
GSM105349	quorum response
GSM105350	DNA
GSM105351	DNA
GSM105352	quorum response

GSM105353	DNA
GSM105354	quorum response
GSM105355	quorum response
GSM105356	quorum response
GSM105357	quorum response
GSM105358	quorum response
GSM105359	control
GSM105360	control
GSM105361	control
GSM105416	DNA
GSM105417	DNA
GSM105418	DNA
GSM105419	control
GSM105420	control
GSM105421	control
GSM105422	DNA
GSM105423	DNA
GSM105424	DNA
GSM105425	DNA
GSM105426	DNA
GSM105427	DNA
GSM27217	sporulation
GSM27218	sporulation
GSM27219	sporulation
GSM27278	sporulation
GSM27279	sporulation
GSM27280	sporulation
GSM27281	sporulation
GSM27282	sporulation
GSM27283	sporulation
GSM27285	sporulation
GSM27286	sporulation
GSM27288	sporulation
GSM27514	sporulation
GSM27515	sporulation
GSM27516	sporulation
GSM49830	control
GSM49975	control
GSM49998	phosphate
GSM50000	phosphate
GSM50002	phosphate
GSM50003	phosphate
GSM50008	control
GSM50378	phosphate
GSM50472	phosphate
GSM51432	phosphate
GSM51433	phosphate
GSM51434	phosphate
GSM51435	control
GSM51436	phosphate
GSM51437	phosphate
GSM51438	phosphate
GSM51439	phosphate

GSM51440	phosphate
GSM51441	control
GSM51442	phosphate
GSM51443	phosphate
GSM51444	phosphate
GSM51445	phosphate
GSM51446	control
GSM51447	control
GSM51448	phosphate
GSM51449	phosphate
GSM51450	phosphate
GSM51451	control
GSM51452	control
GSM51453	phosphate
GSM51454	phosphate
GSM51455	phosphate
GSM51456	phosphate
GSM51457	control
GSM51458	control
GSM51459	phosphate
GSM51460	phosphate
GSM51461	phosphate
GSM70016	peroxide
GSM70017	peroxide
GSM70018	peroxide
GSM70019	peroxide
GSM70020	peroxide
GSM70021	peroxide
GSM70022	peroxide
GSM70023	peroxide
GSM70024	heat
GSM70025	heat
GSM70026	heat
GSM70027	heat
GSM70028	peroxide
GSM70029	heat
GSM70030	peroxide
GSM70031	heat
GSM70032	heat
GSM70033	peroxide
GSM70034	heat
GSM70035	peroxide
GSM70036	heat
GSM70037	heat
GSM70038	peroxide
GSM70039	peroxide
GSM70040	peroxide
GSM70041	peroxide
GSM70042	peroxide
GSM70043	peroxide
GSM70044	peroxide
GSM70045	peroxide
GSM70046	peroxide

GSM70047	peroxide
GSM70048	peroxide
GSM70049	peroxide
GSM70050	peroxide
GSM70051	peroxide
GSM70052	heat
GSM70053	peroxide
GSM70054	heat
GSM70055	peroxide
GSM70056	heat
GSM70057	peroxide
GSM70058	peroxide
GSM70059	heat
GSM70060	peroxide
GSM70061	peroxide
GSM70062	peroxide
GSM70063	peroxide
GSM70064	peroxide
GSM70065	peroxide
GSM70066	peroxide
GSM70067	peroxide
GSM70068	peroxide
GSM70069	peroxide
GSM70070	peroxide
GSM70071	peroxide
GSM70072	peroxide
GSM70073	peroxide
GSM70074	peroxide
GSM70075	peroxide
GSM70076	peroxide
GSM70077	peroxide
GSM70078	peroxide
GSM70079	peroxide
GSM70080	peroxide
GSM70081	peroxide
GSM70082	heat
GSM70083	peroxide
GSM70084	heat
GSM70085	peroxide
GSM70086	peroxide
GSM70087	peroxide
GSM70088	peroxide
GSM70089	peroxide
GSM70090	peroxide
GSM70091	peroxide
GSM70092	peroxide
GSM70093	peroxide