

SOFTWARE

Open Access

PePPER: a webserver for prediction of prokaryote promoter elements and regulons

Anne de Jong^{1,2,3}, Hilco Pietersma^{1,2,3}, Martijn Cordes^{1,2,3}, Oscar P Kuipers^{1,2,3} and Jan Kok^{1,2*}

Abstract

Background: Accurate prediction of DNA motifs that are targets of RNA polymerases, sigma factors and transcription factors (TFs) in prokaryotes is a difficult mission mainly due to as yet undiscovered features in DNA sequences or structures in promoter regions. Improved prediction and comparison algorithms are currently available for identifying transcription factor binding sites (TFBSs) and their accompanying TFs and regulon members.

Results: We here extend the current databases of TFs, TFBSs and regulons with our knowledge on *Lactococcus lactis* and developed a webserver for prediction, mining and visualization of prokaryote promoter elements and regulons via a novel concept. This new approach includes an all-in-one method of data mining for TFs, TFBSs, promoters, and regulons for any bacterial genome via a user-friendly webserver. We demonstrate the power of this method by mining WalRK regulons in Lactococci and Streptococci and, vice versa, use *L. lactis* regulon data (CodY) to mine closely related species.

Conclusions: The PePPER webserver offers, besides the all-in-one analysis method, a toolbox for mining for regulons, promoters and TFBSs and accommodates a new *L. lactis* regulon database in addition to already existing regulon data. Identification of putative regulons and full annotation of intergenic regions in any bacterial genome on the basis of existing knowledge on a related organism can now be performed by biologists and it can be done for a wide range of regulons. On the basis of the PePPER output, biologist can design experiments to further verify the existence and extent of the proposed regulons. The PePPER webserver is freely accessible at <http://pepper.molgenrug.nl>.

Background

As early as in 1960 the term operon was coined for a group of genes of which the expression is coordinated by an operator [1]. Experimental methods like Electrophoretic Mobility Shift Assays (EMSA), Surface Plasmon Resonance (SPR), nuclease protection assays (DNase-footprinting) and Chromatin Immuno Precipitation (ChIP) can all be used to demonstrate that an interaction exists between a transcription factor (TF) and DNA [2]. Experimentally proven TFBSs have been described in literature and are available via publicly accessible databases such as DBTBS [3], RegulonDB [4], PRODORIC [5], MicrobesOnline [6], RegTransBase [7] and RegPrecise [8]. Besides experimental proof for the existence of protein-DNA interaction, TFBS discovery algorithms have been developed to uncover conserved

regions that might act as TFBSs (MEME [9], ARCS-Motif [10], GLAM2 [9], W-AlignACE [11], GIMSAN [12], RankMotif++ [13], GAME [14], and Tmod [15]). This so-called motif mining is based on a collection of genes having a certain correlation. Gene-to-gene correlations can be derived e.g., from transcriptome data or from functional relations like belonging to the same metabolic pathway or to certain COG or GO classes. Motif mining consists of a search for conserved DNA patterns in the upstream intergenic regions of the genes or the operons to which the gene(s) belong. A high probability (low *p*-value) that the occurrence of a certain DNA pattern is very specific for a gene set does not necessarily imply that this motif constitutes a TFBS but it is a good lead for biological functional analysis.

Regulons

Genes and operons that are under control of the same TF are members of that TF's regulon. Although methods for the prediction of regulons have been substantially improved [16], they are still far from perfect. Comparative

* Correspondence: jan.kok@rug.nl

¹Department of Molecular Genetics, University of Groningen, Groningen Biomolecular Sciences and Biotechnology Institute, 9747 AG Groningen, The Netherlands

²Top Institute Food and Nutrition, Wageningen, The Netherlands
Full list of author information is available at the end of the article

genomics tools can be used to predict regulons in bacterial genomes but the procedure can lead to incorrect regulon calling. Despite this drawback, several regulon databases are available that are based on comparative genomics methods and lack experimental evidence. Probably the most extended and accurate databases of regulons are DBTBS for *B. subtilis* [3] and RegulonDB for *E. coli* [4]. The latest update of DBTBS brought the total number of *B. subtilis* TFs to 120, promoters to 1475 and regulated operons to 736, of which 463 operons have been experimentally validated [3]. Together, RegulonDB and DBTBS are the major resources for regulon network mining dedicated to prokaryotes. PRODORIC and RegTransBase are the most extended and manually curated databases on gene regulation in prokaryotes in general [5]. Besides regulon information they include TFBSs and bioinformatics tools for prediction, analysis and visualization of gene regulatory networks using ProdoNet [17] and furthermore, PRODORIC offers the tool “virtual footprint”, which can be used to mine for novel regulons. The *in silico* prediction of regulons is usually based on operons that share the same TFBS and the information is supplemented with the results from comparative genomics analysis of known regulons. This method is used in the recently launched webserver RegPrecise [8], which gives access to a database containing a collection of manually curated regulons grouped together by similar properties such as belonging to the same biological process or metabolic pathway. The database is limited to six closely related bacteria (*Shewanella*, *Thermotogales*, *Bacillales* and *Desulfovibrionales*). On the other hand FITBAR [16] is dedicated to TFBS mining and discovery, RegAnalyst [18] and ProdoNet [17] are webservers enabling integration of data on proteomics and metabolic pathways and provide subsequent graphical representation of networks.

In this work, we designed and developed a novel tool, PePPER, to mine for regulons and TFBSs in any sequenced bacterial genome. As a showcase, we extended the existing regulon databases with a database of *L. lactis* regulons that is derived from literature on transcriptional regulation. The latter is accessible via the user-friendly PePPER web interface.

Implementation

Data resources

MolgenRegDB is an integrated *in house* collection of TFs, TFBSs and regulons of *L. lactis* and is available via the PePPER webserver (<http://pepper.molgenrug.nl>). In addition, TF and TFBS data were downloaded from RegulonDB (*E. coli*) and DBTBS (*B. subtilis*) and subsequently reformatted and integrated together with MolgenRegDB in the PePPER database. Data of all publically available bacterial genomes are daily updated from NCBI (<http://www.ncbi.nlm.nih.gov>) and available via the PePPER webserver.

Implemented mining tools

Overrepresented DNA motifs are identified using MEME [9] and the position-specific probability matrices (PSPMs) obtained were converted to position weight matrices (PWMs) that are compatible with MOODS [19]. BLAST 2.2 [20] is used for protein comparisons. Glimmer3 [21] is used for automated gene detection (open reading frame or ORF calling) and Ribosomal Binding Sites (RBSs) are detected using RBSfinder [22]. In case of *de novo* ORF calling, the translation start is adapted to match the RBS prediction, otherwise the original annotation is used. TransTermHP [23] is implemented for the discovery of putative transcription terminators. Possible secondary RNA structures are predicted and plotted using RNAfold and RNAplot of the Vienna package [24]. A new prokaryote promoter prediction tool was developed and is based on PWMs and Hidden Markov Models (HMMs) of -35 and -10 consensus sequences and various sigma factor binding sites. PWMs and HMMs of *B. subtilis* and *E. coli* promoters are used as reference for Gram-positive and Gram-negative bacteria, respectively. A collection of individual tools used by PePPER are accessible via the webserver.

Building the *L. Lactis* TFBS library

A database of validated *L. lactis* TFBSs of regulons derived from literature data was made after which for each TFBS a PSPM was calculated using MEME and subsequently transposed to a MOODS compatible PWM format. To that end we used the upstream intergenic regions plus the first 20 bases of their genes as input for MEME in order to search for overrepresented DNA motifs. These motifs ranged in length from 6 to 18 bases and a database of all intergenic regions of *L. lactis* MG1363 was used as a background model. Subsequently, the overrepresented DNA motifs were manually compared to the literature data. Only those DNA motifs that resemble the experimentally verified TFBSs were included in the database, including the MOODS cutoff values. An overview of TFBSs of regulons, including WebLogos, is shown in Additional file 1: Table S1; the database containing all the PSPM profiles is available via the PePPER webserver.

The PePPER toolbox

A powerful toolbox has been created in PePPER and is accessible via the PePPER webserver. By selecting a regulon on the basis of its TF and one or more genomes, the program will perform a Blast analysis between the proteins of the known regulon and all the proteins encoded by the genes in the selected genome(s). PePPER provides a clear overview in colors, of the predicted regulon(s), in each genome, which is based on the degree of protein homologies; detailed information is given in

attached tables. More details about scoring and the color scheme are given on the PePPER webserver.

PePPER all-in-one

PePPER all-in-one is a parameter-free pipeline of the individual PePPER tools allowing fully automatic intergenic annotation combined with analysis of regulons. A schematic overview of the PePPER all-in-one process is shown in Figure 1. Two input formats are accepted for analysis: i) plain DNA sequences in FastA format, ii) a fully annotated file in the Genbank file format. DNA sequences lacking ORF information will be automatically annotated using Glimmer3 to discriminate between ORFs and intergenic regions. Input files in Genbank format, either uploaded or selected from the PePPER library of genomes, will produce the most extended results, including hyperlinks to NCBI resources such as protein annotation, protein domains and genomic context of the genes. The output is organized into three tables and one figure: i) Table 1, the “Summary of Results” contains links to detailed information on analysis of regulons, TFBS, promoters, transcription terminators, RNA folding and motif analysis using MEME, ii) Table 2 and Table 3, “Files available for download”, iii) Table 4, Combined results of the TFBS and regulon mining. Figure 1 gives a graphical presentation of the intergenic regions.

Promoter prediction using PePPER

A universal prokaryote transcription initiation DNA motif does not exist [2], but a common DNA pattern (the Pribnow box) 10 base pairs upstream of the transcription start site (TSS) and a conserved sequence 35 base pairs upstream of the TSS are overrepresented in promoter regions. These patterns are searched for separately, after which putative promoters are only taken into account if the spacing between their -35 and -10 motifs is 16 to 18 bases. Although many different sigma factors binding sites are known (especially from *B. subtilis*) these are not used in the promoter prediction routine used here; they are implemented as conserved DNA motifs in the TFBS mining tool. The resulting promoter prediction algorithm is universal for prokaryotes, but we do offer the possibility to discriminate between Gram-positive and Gram-negative bacteria to improve the accuracy of the prediction algorithm. Furthermore, “incomplete” promoters, in which only a -35 or a -10 sequence is predicted are also shown in the results.

Webserver

PePPER (<http://pepper.molgenrug.nl>) can be accessed through a user-friendly web interface for querying and browsing. The server runs on a linux platform (Ubuntu server LTS 10.04) with an Apache webserver (version

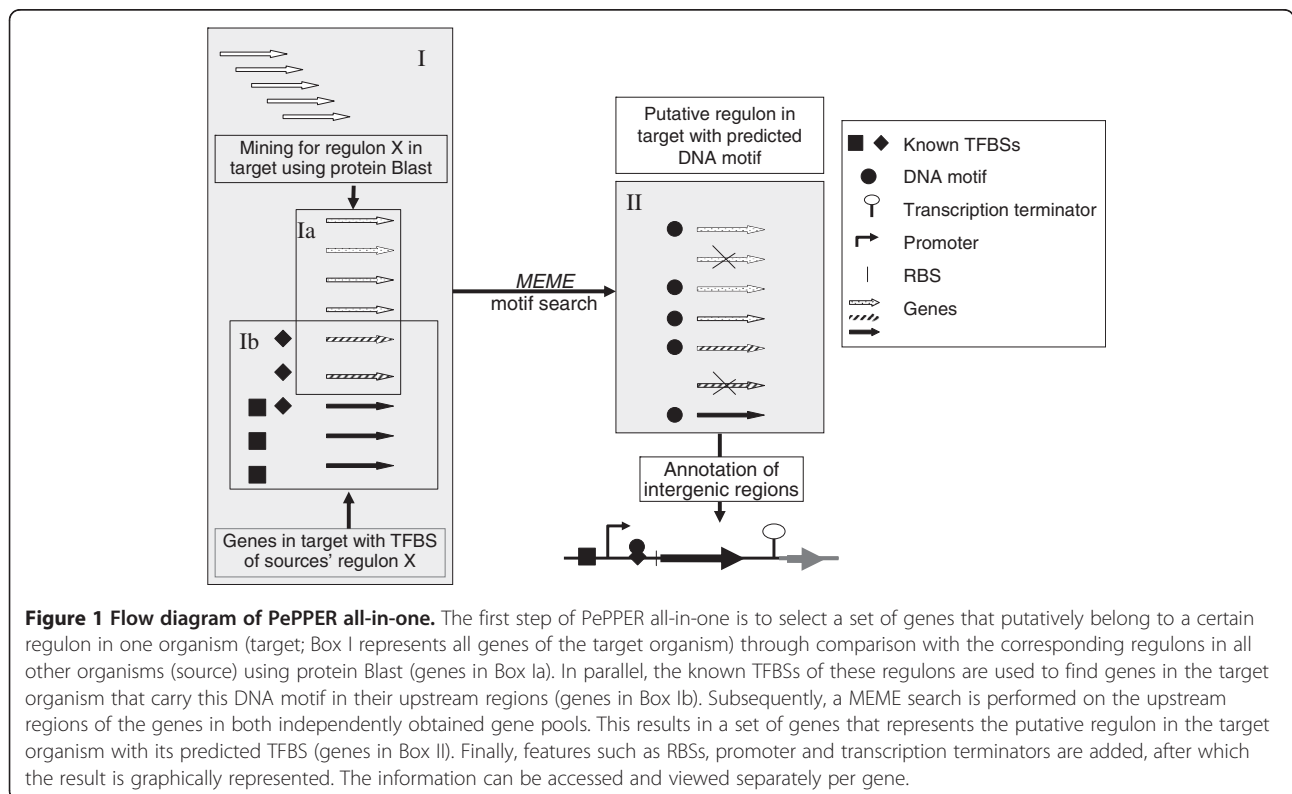


Table 1 Regulators of which the regulons have been studied in *Lactococcus lactis* ssp. *cremoris* MG1363 and *Lactococcus lactis* spp. *lactis* IL1403 and their literature references. -, strain/subspecies not specified

Gene	Literature
AhrC	MG1363 [25,26]
ArgR	MG1363 [25,26]
BusR	[-27,28]
CcpA	MG1363 [29]
CodY	MG1363 [30,31]
ComX	[-32]
CopR	IL1403 [33]
CtsR	MG1363 [34,35]
FhuR	IL1403 [36]
FlpA	MG1363 [37]
FlpB	MG1363 [37]
FruR	[-38]
GadR	[-39,40]
GntR	MG1363 [41]
HdiR	MG1363 [42]
HisZ	[-43,44]
LlrA	MG1363 [45]
LlrB	MG1363 [45]
LlrC	MG1363 [45]
LlrD	[-46]
LlrE	MG1363 [45]
LlrF	MG1363 [45]
LlrG	MG1363 [45]
LmrR	MG1363 [47,48]
MalR	[-49]
PhoU	[-50]
PurR	[-51,52]
PyrR	MG1363 [53]
RcfB	[-54]
SpxA	[-55,56]
XylR	[-57]
ZitR	MG1363 [58]

2.2) and a MySQL server (version 5.1) and Blast 2.2. Programming was done using PHP 5.0, Perl 5.12 and BioPerl 1.8. A combination of Joomla and jQuery 1.4 was used to build the user-friendly web interface.

Results and discussion

Regulons in *Lactococcus lactis*

Each of the 154 known or predicted TFs of *L. lactis* subsp. *cremoris* MG1363 [59] will probably regulate the transcription of one or more genes or operons. The

Table 2 Analysis of regulons

<i>B. subt</i>	<i>L.lactis</i> MG1363				Blast e-value
	locus tag	gene name	Protein ID	GeneID	
<i>walk</i>	llmg_0414	<i>llrC</i>	YP_001031764.1	4797664	2.00E-91
<i>walR</i>	llmg_0413	<i>kinC</i>	YP_001031763.1	4798420	1.00E-95
<i>yycJ</i>	llmg_0412	<i>vicX</i>	YP_001031762.1	4798732	4.00E-78
<i>yycK</i>	llmg_2419	<i>htrA</i>	YP_001033660.1	4797497	3.00E-74
<i>ftsA</i>	llmg_2061	<i>ftsA</i>	YP_001033316.1	4797264	3.00E-74
<i>ftsZ</i>	llmg_2060	<i>ftsZ</i>	YP_001033315.1	4798073	5.00E-106
<i>phoP</i>	llmg_0414	<i>llrC</i>	YP_001031764.1	4797664	2.00E-67
<i>phoR</i>	llmg_0413	<i>kinC</i>	YP_001031763.1	4798420	2.00E-51
<i>tagB</i>	llmg_1603	<i>tagB</i>	YP_001032887.1	4798977	1.00E-32
<i>tagD</i>	llmg_1606	<i>tagD2</i>	YP_001032890.1	4798976	3.00E-33
<i>tagF</i>	llmg_1604	<i>tagF</i>	YP_001032888.1	4798736	2.00E-73
<i>yjeA</i>	llmg_0293	<i>xynD</i>	YP_001031648.1	4797603	3.00E-43
<i>yocH</i>	llmg_2194	<i>llmg_2194</i>	YP_001033444.1	4798120	2.00E-22

A. Comparison of the *B. subtilis* WalRK regulon to the *L. lactis* MG1363 genome.

functionality of 32 TFs of *L. lactis* MG1363 and *L. lactis* subsp. *lactis* IL1403 has been reported in literature, using techniques ranging from DNA microarray analysis to DNA footprinting. Although the two lactococcal subspecies are closely related, not each regulator or regulon of one is present or similar in the other. The majority of the TFs in MG1363 and IL1403 show a high degree of mutual similarity. Of the 154 TFs in *L. lactis* MG1363, 22 are not present in *L. lactis* IL1403 while 20 out of the 143 TFs identified in *L. lactis* IL1403 are not found in MG1363 (Tables 2 and Table 3). Analysis performed by PePPER showed that large regulons (those of CodY, CcpA, CmbR, CesSR, ArgR, and PurR) as well as some small regulons (those of RcfB, ZirR, BusR and LmrR) are well conserved in the two strains. The conservation of regulons between the closely related subspecies is illustrated by the CmbR regulon of cysteine and methionine biosynthesis, which has been studied in detail in both *L.*

Table 3 Comparison of the WalRK TCS of *B. subtilis* to the *L. lactis* orthologs using PePPER's multiple genome regulon mining tool

<i>B. subtilis</i>	<i>L. lactis</i>			
	MG1363	IL1403	SK11	KF147
<i>walR</i>	<i>llrC</i>	<i>llrC</i>	<i>LACR_0444</i>	<i>tcsR</i>
<i>walk</i>	<i>kinC</i>	<i>kinC</i>	<i>LACR_0443</i>	<i>tcsK</i>
<i>yycJ</i>	<i>vicX</i>	<i>yeaA</i>	<i>LACR_0441</i>	<i>yeaA</i>
<i>yycK</i>	<i>htrA</i>	<i>htrA</i>	<i>LACR_2439</i>	<i>htrA</i>

B. Note that a nomenclature mix up took place in the *B. subtilis* 168 NC_000964 file (release Feb. 2011); in this release, YycG (locus BSU40400) is called Walk, but it is described as "two-component sensor histidine kinase YycF" while YycF is labeled WalR "two-component response regulator YycG". The names WalR and Walk have recently been corrected. Due to this temporary swap, the annotation of *yycF* and *yycG* could still be wrong in the annotation of other bacterial genomes.

(See figure on previous page.)

Figure 2 Presence of regulon in query organisms by protein Blast. WalRK regulon genes of *B. subtilis* (a) and *S. pneumoniae* (b) were used to perform a protein Blast in order to examine whether a WalRK regulon might be present in Lactococci and Streptococci. Colors from light to dark blue are indicative of increasing protein similarity; for more details see the PePPER webserver.

lactis IL1403 [60] and *L. lactis* MG1363 [36]. Analysis of both CmbR regulons shows that 16 out of 17 proteins in the IL1403 CmbR regulon have high similarity to MG1363 proteins (data not shown). Finally, all known TFs and TFBSs of *L. lactis* were collected in one database, the MolgenRegDB. This is currently the most comprehensive manually curated regulon database of *L. lactis*; it is available via the PePPER webserver (<http://pepper.molgenrug.nl>).

Prediction of the WalR regulon in four *L. Lactis* strains

The *B. subtilis* operon *walRKyycHIJK* is a 6-cistron operon encoding among others the two-component system (TCS) WalRK that controls the expression of 23 genes. These genes represent the WalR regulon [61-64]. This signal transduction pathway is crucial for the regulation of cell wall metabolism and is one of the few TCS known to be a virulence factor in *S. pneumoniae* [61]. The presence of the WalR regulon has never been described in *L. lactis*. We validated PePPER by comparing its results to literature data and subsequently used it to unravel the putative WalR regulons and cognate TFBS in 4 sequenced strains of *L. lactis*. PePPER showed that the products of 4 of the genes of the *walRKyycHIJK* of *B. subtilis* are orthologous to *kinC*, *lrrC* and *vicX*, *htrA* of *L. lactis* MG1363 (see Table 3). Furthermore, PePPER showed that 13 out of the 23 proteins of the WalR regulon of *B. subtilis* show high similarity (Blast e-value $<10^{-20}$) to proteins in *L. lactis* MG1363; they are organized in 6 operons (Table 2). Using PePPER's multiple genome mining tool, it is clear that orthologs of the WalRK TCS and part of the WalR regulon genes of *B. subtilis* are present in all other fully sequenced *L. lactis* strains: IL1403, SK11 and KF147 (Table 3).

Streptococci are phylogenetically closely related to Lactococci and therefore we performed an analysis of the WalRK regulons in both groups using the PePPER toolbox. The result (Figure 2) shows the similarity of the known WalRK regulons (*B. subtilis*, *S. pneumoniae*, *Staphylococcus aureus*) with the predicted regulons of Lactococci and Streptococci. The *B. subtilis* regulon (Figure 2a) showed the high conservation between the two groups where 6 genes were not found and 9 genes out of 23 were found in all species (protein blast e-value cutoff of 10^{-20}). The difference between the two groups is shown in Figure 2b where 11 out of 24 genes of the *S. pneumoniae* R6 WalRK regulon were not found in Lactococci.

TFBS prediction

The WalR binding site (TGTAAn6-TGTAAn) was mapped using DNaseI footprinting and EMSAs in 4 organisms; *B. subtilis*, *Staph. aureus*, *S. pneumoniae* and *S. mutans* [61]. We added the WalR TFBSs derived from these 4 organisms separately, as well as an averaged (combined) WalR motif (WalR_[combined]) to the PePPER database and subsequently screened the genomes of the *L. lactis* strains for the presence of these DNA motifs. The upstream DNA regions of two genes of *L. lactis* MG1363 that are orthologous to WalRK regulon members carry the WalR_[combined] TFBS, namely *xynD* (TGTAT-n6-TGTTA) and *htrA* (TGAAA-n6-TGAAG). In the upstream DNA region of the other 4 WalRK operon orthologs no WalR_[combined] was found. Interestingly, WalR_[combined] (Figure 3) was present upstream of the cell wall hydrolase genes *acmA* and *acmB* [65,66], which could indicate that these genes might be regulated

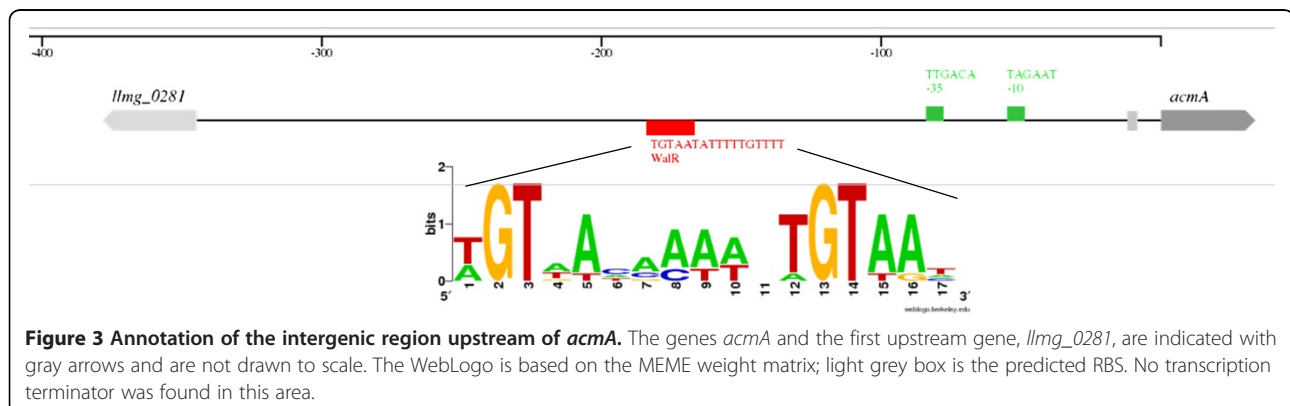
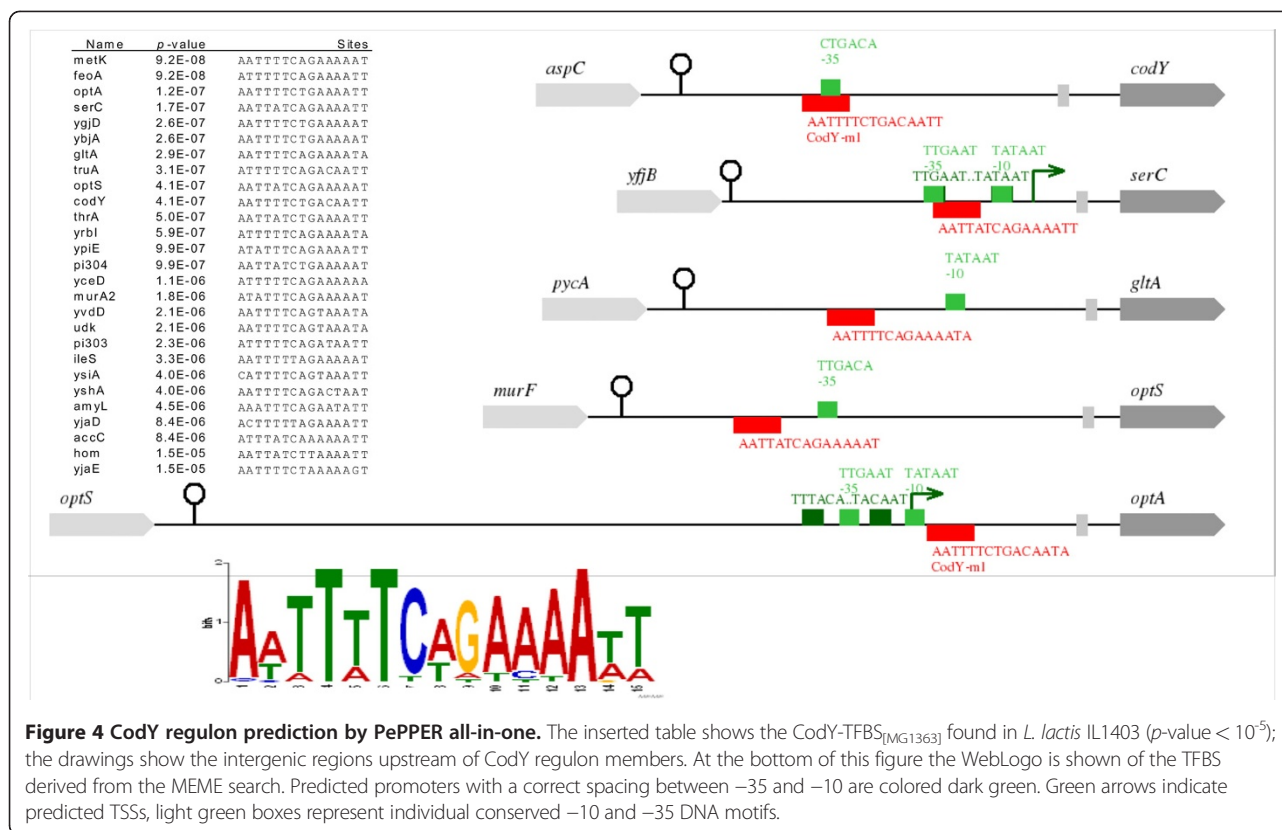


Figure 3 Annotation of the intergenic region upstream of *acmA*. The genes *acmA* and the first upstream gene, *lmg_0281*, are indicated with gray arrows and are not drawn to scale. The WebLogo is based on the MEME weight matrix; light grey box is the predicted RBS. No transcription terminator was found in this area.



by WalRK and that the WalRK stress response system of *L. lactis* also influences their expression.

PePPER all-in-one case study

The well-studied global transcriptional regulator CodY of *L. lactis* MG1363 [67,68] binds to the CodY-box (AATTTTCWGAAAATT) and influences the expression of genes involved in (branched-chain) amino acid uptake and biosynthesis as well as several other genes involved in N-metabolism (proteolysis and peptide uptake). The CodY regulon of *L. lactis* MG1363 was used in the PePPER all-in-one system to mine the *L. lactis* IL1403 genome for the presence of homologs of the CodY_[MG1363] regulon and the CodY-TFBS_[MG1363]. Subsequently, a MEME search was performed. The results showed that a CodY-TFBS is present in the upstream intergenic regions of 5 genes/operons in *L. lactis* IL1403, namely *codY*, *serCAB*, *gltA-citB-icd*, *dppA* and *dppPBCDF*. In *L. lactis* IL1403, Dpp, erroneously annotated as Opt [69], functions as a di/tripeptide transporter, with DppA as the substrate binding protein, and as an oligopeptide transporter employing DppP. The *dppP* gene in the *dppA-dppPBCDF* gene cluster of *L. lactis* MG1363 is mutated; in this strain CodY binds upstream of *dppA* but not upstream of *dppP*. Oligopeptide uptake in *L. lactis* MG1363 is encoded by the

oppDFBCA operon, which is under CodY control [68]. The Opp system in *L. lactis* IL1403 is present but non-functional [69] and no CodY-TFBS_[MG1363] was found in the DNA region upstream of *opp*. Despite the differences in the activities of these transport systems, our analysis indicates that in both lactococcal strains CodY regulates di/tri- and oligopeptide transport. The graphical overview of intergenic regions (Figure 4) shows that CodY represses gene expression by binding in or closely downstream of the promoter regions. The DNA binding motif that was identified in *L. lactis* IL1403 resembles CodY-TFBS_[MG1363] [68] and the CodY-TFBS_[IL1403] reported by Guedon et al. [67].

We used the CodY_[MG1363] regulon to screen for the presence of a similar regulon in a less closely related Gram-positive bacterium, the pathogen *S. pneumoniae* D39. The analysis revealed that seven genes/operons (*ilvD*, *ilvE*, *asd*, *hom-thrB*, *amiACDEF*, *SPD_1878-thrC*, *livJHMGF*) involved in amino acid transport or biosynthesis carry a sequence closely related to CodY-TFBS_[MG1363] in their upstream DNA regions.

Conclusions

PePPER uses a novel approach, in which all available information on prokaryotic regulons and TFBSs is used to

identify regulons in any query bacterium. In addition it offers a user-friendly web interface making the data provided by PePPER easily accessible for non-bioinformaticians. PePPER offers, next to all fully sequenced bacterial genomes, the possibility to upload un-annotated data, which is then processed automatically. Furthermore, prediction of intergenic region elements such as promoters, transcription terminators, sigma factor binding sites, RBSs, as well as that of possible secondary DNA structures therein, will lead to more detailed knowledge of the DNA regions under study. By adding our knowledge on *L. lactis* regulons as well as DBTBS and RegulonDB regulon data to the PePPER database, we provide an extended database of bacterial regulons and TFBSs. PePPER can be used to pinpoint a wide range of putative regulons and their cognate TFBSs in any bacterial genome on the basis of existing knowledge. This regulon information can subsequently be used by biologists to help them design experiments to authenticate the proposed regulons.

Additional file

Additional file 1: Table S1. *Lactococcus lactis* TFBS WebLogos. Column 1 presents all known (studied) regulons of *L. lactis*. Alternative names for TFBS are given in column 2. The consensus sequence given in columns 3 were taken from the literature references from Table 1. In column 4, the TFBS identities are given that are used by PePPER. The upstream sequences of the genes of the regulons indicated in column 1 were aligned using MEME [9]. The obtained DNA motifs (WebLogos) are presented in column 5.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AJ and JK devised the PePPER concept and web design and wrote the manuscript. OPK participated in the design of the study and helped to draft the manuscript. HP and MC contributed to the writing of the webserver scripts and to building of the PePPER database. All authors read and approved the final manuscript.

Acknowledgements

We thank Tom Eckhardt and Jan Willem Veening for fruitful discussions. This project was partly supported by grants from the Top Institute Food and Nutrition, Wageningen, the Netherlands and The Netherlands Organisation for Scientific Research (NWO), the Netherlands.

Author details

¹Department of Molecular Genetics, University of Groningen, Groningen Biomolecular Sciences and Biotechnology Institute, 9747 AG Groningen, The Netherlands. ²Top Institute Food and Nutrition, Wageningen, The Netherlands. ³The Netherlands Kluyver Centre for Genomics of Industrial Fermentations, Delft, The Netherlands/Netherlands Consortium of Systems Biology, Amsterdam, The Netherlands.

Received: 11 July 2011 Accepted: 13 April 2012

Published: 2 July 2012

References

- Jacob F, Perrin D, Sanchez C: **MONOD J: Operon: a group of genes with the expression coordinated by an operator.** *C R Hebd Seances Acad Sci* 1960, **250**:1727–1729.

- van Hijum SAFT, Medema MH, Kuipers OP: **Mechanisms and evolution of control logic in prokaryotic transcriptional regulation.** *Microbiol Mol Biol Rev* 2009, **73**(3):481.
- Sierro N, Makita Y, de Hoon M, Nakai K: **DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information.** *Nucleic Acids Res* 2008, **36**(Database issue):D93–D96.
- Gama-Castro S, Salgado H, Peralta-Gil M, Santos-Zavaleta A, Muniz-Rascado L, Solano-Lira H, Jimenez-Jacinto V, Weiss V, Garcia-Sotelo JS, Lopez-Fuentes A, Porron-Sotelo L, Alquicira-Hernandez S, Medina-Rivera A, Martinez-Flores I, Alquicira-Hernandez K, Martinez-Adame R, Bonavides-Martinez C, Miranda-Rios J, Huerta AM, Mendoza-Vargas A, Collado-Torres L, Taboada B, Vega-Alvarado L, Olvera M, Olvera L, Grande R, Morett E, Collado-Vides J: **RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units).** *Nucleic Acids Res* 2010, **1–8**. doi:10.1093/nar/gkq1110.
- Grote A, Klein J, Retter I, Haddad I, Behling S, Bunk B, Biegler I, Yarmolinetz S, Jahn D, Munch R: **PRODORIC (release 2009): a database and tool platform for the analysis of gene regulation in prokaryotes.** *Nucleic Acids Res* 2009, **37**(Database issue):D61–D65.
- Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP: **MicrobesOnline: an integrated portal for comparative and functional genomics.** *Nucleic Acids Res* 2010, **38**(Database issue):D396–D400.
- Kazakov AE, Cipriano MJ, Novichkov PS, Minovitsky S, Vinogradov DV, Arkin A, Mironov AA, Gelfand MS, Dubchak I: **RegTransBase—a database of regulatory sequences and interactions in a wide range of prokaryotic genomes.** *Nucleic Acids Res* 2007, **35**(Database issue):D407–D412.
- Novichkov PS, Laikova ON, Novichkova ES, Gelfand MS, Arkin AP, Dubchak I, Rodionov DA: **RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes.** *Nucleic Acids Res* 2010, **38**(Database issue):D111–D118.
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: **MEME SUITE: tools for motif discovery and searching.** *Nucleic Acids Res* 2009, **37**(Web Server issue):W202.
- Zhang S, Su W, Yang J: **ARCS-Motif: discovering correlated motifs from unaligned biological sequences.** *Bioinformatics* 2009, **25**(2):183–189.
- Chen X, Guo L, Fan Z, Jiang T: **W-AlignACE: an improved Gibbs sampling algorithm based on more accurate position weight matrices learned from sequence and gene expression/ChIP-chip data.** *Bioinformatics* 2008, **24**(9):1121–1128.
- Ng P, Keich U: **GIMSAN: a Gibbs motif finder with significance analysis.** *Bioinformatics* 2008, **24**(19):2256–2257.
- Chen X, Hughes TR, Morris Q: **RankMotif++: a motif-search algorithm that accounts for relative ranks of K-mers in binding transcription factors.** *Bioinformatics* 2007, **23**(13):i72–i79.
- Wei Z, Jensen ST: **GAME: detecting cis-regulatory elements using a genetic algorithm.** *Bioinformatics* 2006, **22**(13):1577–1584.
- Sun H, Yuan Y, Wu Y, Liu H, Liu JS, Xie H: **Tmod: toolbox of motif discovery.** *Bioinformatics* 2010, **26**(3):405–407.
- Oberto J: **FITBAR: a web tool for the robust prediction of prokaryotic regulons.** *BMC Bioinforma* 2010, **11**(1):554.
- Klein J, Leupold S, Munch R, Pommerenke C, Johl T, Karst U, Jansch L, Jahn D, Retter I: **ProdoNet: identification and visualization of prokaryotic gene regulatory and metabolic networks.** *Nucleic Acids Res* 2008, **36**(Web Server issue):W460–W464.
- Sharma D, Mohanty D, Suroliya A: **RegAnalyst: a web interface for the analysis of regulatory motifs, networks and pathways.** *Nucleic Acids Res* 2009, **37**(Web Server issue):W193–W201.
- Korhonen J, Martinmaki P, Pizzi C, Rastas P, Ukkonen E: **MOODS: fast search for position weight matrix matches in DNA sequences.** *Bioinformatics* 2009, **25**(23):3181–3182.
- Johnson M, Zaretskaya I, Raytselis Y, Merezukh Y, McGinnis S, Madden TL: **NCBI BLAST: a better web interface.** *Nucleic Acids Res* 2008, **36**(Web Server issue):W5–W9.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL: **Identifying bacterial genes and endosymbiont DNA with Glimmer.** *Bioinformatics* 2007, **23**(6):673–679.

22. Suzek BE, Ermolaeva MD, Schreiber M, Salzberg SL: **A probabilistic method for identifying start codons in bacterial genomes.** *Bioinformatics* 2001, **17**(12):1123–1130.
23. Kingsford CL, Ayanbule K, Salzberg SL: **Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake.** *Genome Biol* 2007, **8**(2):R22.
24. Hofacker IL: **Vienna RNA secondary structure server.** *Nucleic Acids Res* 2003, **31**(13):3429–3431.
25. Larsen R, Buist G, Kuipers OP, Kok J: **ArgR and AhrC are both required for regulation of arginine metabolism in *Lactococcus lactis*.** *J Bacteriol* 2004, **186**(4):1147–1157.
26. Larsen R, Kok J, Kuipers OP: **Interaction between ArgR and AhrC controls regulation of arginine metabolism in *Lactococcus lactis*.** *J Biol Chem* 2005, **280**(19):19319–19330.
27. Romeo Y, Bouvier J, Gutierrez C: **Osmotic regulation of transcription in *Lactococcus lactis*: ionic strength-dependent binding of the BusR repressor to the busA promoter.** *FEBS Lett* 2007, **581**(18):3387–3390.
28. Romeo Y, Obis D, Bouvier J, Guillot A, Fourcans A, Bouvier I, Gutierrez C, Mistou MY: **Osmoregulation in *Lactococcus lactis*: BusR, a transcriptional repressor of the glycine betaine uptake system BusA.** *Mol Microbiol* 2003, **47**(4):1135–1147.
29. Zomer AL, Buist G, Larsen R, Kok J, Kuipers OP: **Time-resolved determination of the CcpA regulon of *Lactococcus lactis* subsp. cremoris MG1363.** *J Bacteriol* 2007, **189**(4):1366–1381.
30. den Hengst CD, Curley P, Larsen R, Buist G, Nauta A, van Sinderen D, Kuipers OP, Kok J: **Probing direct interactions between CodY and the oppD promoter of *Lactococcus lactis*.** *J Bacteriol* 2005, **187**(2):512–521.
31. den Hengst CD, Groeneveld M, Kuipers OP, Kok J: **Identification and functional characterization of the *Lactococcus lactis* CodY-regulated branched-chain amino acid permease BcaP (Ctra).** *J Bacteriol* 2006, **188**(9):3280–3289.
32. Wydau S, Dervyn R, Anba J, Dusko Ehrlich S, Maguin E: **Conservation of key elements of natural competence in *Lactococcus lactis* ssp FEMS.** *Microbiol Lett* 2006, **257**(1):32–42.
33. Magnani D, Barre O, Gerber SD, Solioz M: **Characterization of the CopR regulon of *Lactococcus lactis* IL1403.** *J Bacteriol* 2008, **190**(2):536–545.
34. Varmanen P, Ingmer H, Vogensen FK: **ctsR of *Lactococcus lactis* encodes a negative regulator of clp gene expression.** *Microbiology* 2000, **146**(Pt 6):1447–1455.
35. Varmanen P, Vogensen FK, Hammer K, Palva A, Ingmer H: **ClpE from *Lactococcus lactis* promotes repression of CtsR-dependent gene expression.** *J Bacteriol* 2003, **185**(17):5117–5124.
36. Fernandez M, Kleerebezem M, Kuipers OP, Siezen RJ, van Kranenburg R: **Regulation of the metC-cysK operon, involved in sulfur metabolism in *Lactococcus lactis*.** *J Bacteriol* 2002, **184**(1):82–90.
37. Akyol I, Shearman CA: **Regulation of flpA, flpB and rcfA promoters in *Lactococcus lactis*.** *Curr Microbiol* 2008, **57**(3):200–205.
38. Barriere C, Veiga-da-Cunha M, Pons N, Guedon E, van Hijum SA, Kok J, Kuipers OP, Ehrlich DS, Renault P: **Fructose utilization in *Lactococcus lactis* as a model for low-GC gram-positive bacteria: its regulator, signal, and DNA-binding site.** *J Bacteriol* 2005, **187**(11):3752–3761.
39. Sanders JW, Leenhouts K, Burghoorn J, Brands JR, Venema G, Kok J: **A chloride-inducible acid resistance mechanism in *Lactococcus lactis* and its regulation.** *Mol Microbiol* 1998, **27**(2):299–310.
40. Sanders JW, Venema G, Kok J: **A chloride-inducible gene expression cassette and its use in induced lysis of *Lactococcus lactis*.** *Appl Environ Microbiol* 1997, **63**(12):4877–4882.
41. Larsen R, Kloosterman TG, Kok J, Kuipers OP: **GlnR-mediated regulation of nitrogen metabolism in *Lactococcus lactis*.** *J Bacteriol* 2006, **188**(13):4978–4982.
42. Savijoki K, Ingmer H, Frees D, Vogensen FK, Palva A, Varmanen P: **Heat and DNA damage induction of the LexA-like regulator HdiR from *Lactococcus lactis* is mediated by RecA and ClpP.** *Mol Microbiol* 2003, **50**(2):609–621.
43. Bovee ML, Champagne KS, Demeler B, Francklyn CS: **The quaternary structure of the HisZ-HisG N-1-(5'-phosphoribosyl)-ATP transferase from *Lactococcus lactis*.** *Biochemistry* 2002, **41**(39):11838–11846.
44. Champagne KS, Piscitelli E, Francklyn CS: **Substrate recognition by the hetero-octameric ATP phosphoribosyltransferase from *Lactococcus lactis*.** *Biochemistry* 2006, **45**(50):14933–14943.
45. O'Connell-Motherway M, van Sinderen D, Morel-Deville F, Fitzgerald GF, Ehrlich SD, Morel P: **Six putative two-component regulatory systems isolated from *Lactococcus lactis* subsp. cremoris MG1363.** *Microbiology* 2000, **146**(Pt 4):935–947.
46. Martinez B, Zomer AL, Rodriguez A, Kok J, Kuipers OP: **Cell envelope stress induced by the bacteriocin Lcn972 is sensed by the Lactococcal two-component system CesSR.** *Mol Microbiol* 2007, **64**(2):473–486.
47. Agustindari H, Lubelski J, van den Berg van Saparoea HB, Kuipers OP, Driessen AJ: **LmrR is a transcriptional repressor of expression of the multidrug ABC transporter LmrCD in *Lactococcus lactis*.** *J Bacteriol* 2008, **190**(2):759–763.
48. Madoori PK, Agustindari H, Driessen AJ, Thunnissen AM: **Structure of the transcriptional regulator LmrR and its mechanism of multidrug recognition.** *EMBO J* 2009, **28**(2):156–166.
49. Andersson U, Radstrom P: **Physiological function of the maltose operon regulator, MalR, in *Lactococcus lactis*.** *BMC Microbiol* 2002, **2**:28.
50. Cesselin B, Ali D, Gratadoux JJ, Gaudu P, Duwat P, Gruss A, El Karoui M: **Inactivation of the *Lactococcus lactis* high-affinity phosphate transporter confers oxygen and thiol resistance and alters metal homeostasis.** *Microbiology* 2009, **155**(Pt 7):2274–2281.
51. Kilstrup M, Martinussen J: **A transcriptional activator, homologous to the *Bacillus subtilis* PurR repressor, is required for expression of purine biosynthetic genes in *Lactococcus lactis*.** *J Bacteriol* 1998, **180**(15):3907–3916.
52. Jendresen CB: **Characterization of the molecular and genetic regulation of the nucleotide metabolism in *Lactococcus lactis*.** Thesis, Technical University of Denmark; 2010.
53. Martinussen J, Schallert J, Andersen B, Hammer K: **The pyrimidine operon pyrRPB-carA from *Lactococcus lactis*.** *J Bacteriol* 2001, **183**(9):2785–2794.
54. Madsen SM, Hindre T, Le Pennec JP, Israelsen H, Dufour A: **Two acid-inducible promoters from *Lactococcus lactis* require the cis-acting ACiD-box and the transcription regulator RcfB.** *Mol Microbiol* 2005, **56**(3):735–746.
55. Frees D, Varmanen P, Ingmer H: **Inactivation of a gene that is highly conserved in Gram-positive bacteria stimulates degradation of non-native proteins and concomitantly increases stress tolerance in *Lactococcus lactis*.** *Mol Microbiol* 2001, **41**(1):93–103.
56. Veiga P, Bulbarela-Sampieri C, Furlan S, Maisons A, Chapot-Chartier MP, Erkelenz M, Mervelet P, Noiro P, Frees D, Kuipers OP, Kok J, Gruss A, Buist G, Kulakauskas S: **SpxB regulates O-acetylation-dependent resistance of *Lactococcus lactis* peptidoglycan to hydrolysis.** *J Biol Chem* 2007, **282**(27):19342–19354.
57. Erlanson KA, Park JH, Wissam EI K, Kao HH, Basaran P, Brydges S, Batt CA: **Dissolution of xylose metabolism in *Lactococcus lactis*.** *Appl Environ Microbiol* 2000, **66**(9):3974–3980.
58. Morello E, Bermudez-Humaran LG, Llull D, Sole V, Miraglio N, Langella P, Poquet I: ***Lactococcus lactis*, an efficient cell factory for recombinant protein production and secretion.** *J Mol Microbiol Biotechnol* 2008, **14**(1–3):48–58.
59. Wegmann U, O'Connell-Motherway M, Zomer A, Buist G, Shearman C, Canchaya C, Ventura M, Goesmann A, Gasson MJ, Kuipers OP, van Sinderen D, Kok J: **Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. cremoris MG1363.** *J Bacteriol* 2007, **189**(8):3256–3270.
60. Sperandio B, Polard P, Ehrlich DS, Renault P, Guedon E: **Sulfur amino acid metabolism and its control in *Lactococcus lactis* IL1403.** *J Bacteriol* 2005, **187**(11):3762.
61. Dubrac S, Biscicchia P, Devine KM, Msadek T: **A matter of life and death: cell wall homeostasis and the WalkR (YycGF) essential signal transduction pathway.** *Mol Microbiol* 2008, **70**(6):1307–1322.
62. Winkler ME, Hoch JA: **Essentiality, bypass, and targeting of the YycFG (VicRK) two-component regulatory system in gram-positive bacteria.** *J Bacteriol* 2008, **190**(8):2645–2648.
63. Wayne KJ, Sham LT, Tsui HC, Gutu AD, Barendt SM, Keen SK, Winkler ME: **Localization and cellular amounts of the WalRKJ (VicRKX) two-component regulatory system proteins in serotype 2 *Streptococcus pneumoniae*.** *J Bacteriol* 2010, **192**(17):4388–4394.
64. Biller SJ, Wayne KJ, Winkler ME, Burkholder WF: **The putative hydrolase YycJ (WalJ) affects the coordination of cell division with DNA replication in *Bacillus subtilis* and may play a conserved role in cell wall metabolism.** *J Bacteriol* 2010, **193**(4):896–908.

65. Buist G, Kok J, Leenhouts KJ, Dabrowska M, Venema G, Haandrikman AJ: **Molecular cloning and nucleotide sequence of the gene encoding the major peptidoglycan hydrolase of *Lactococcus lactis*, a muramidase needed for cell separation.** *J Bacteriol* 1995, **177**(6):1554–1563.
66. Huard C, Miranda G, Wessner F, Bolotin A, Hansen J, Foster SJ, Chapot-Chartier MP: **Characterization of AcmB, an N-acetylglucosaminidase autolysin from *Lactococcus lactis*.** *Microbiology* 2003, **149**(Pt 3):695–705.
67. Guedon E, Sperandio B, Pons N, Ehrlich SD, Renault P: **Overall control of nitrogen metabolism in *Lactococcus lactis* by CodY, and possible models for CodY regulation in Firmicutes.** *Microbiology* 2005, **151**(Pt 12):3895–3909.
68. den Hengst CD, van Hijum SA, Geurts JM, Nauta A, Kok J, Kuipers OP: **The *Lactococcus lactis* CodY regulon: identification of a conserved cis-regulatory element.** *J Biol Chem* 2005, **280**(40):34332–34342.
69. Doeven MK, Kok J, Poolman B: **Specificity and selectivity determinants of peptide transport in *Lactococcus lactis* and other microorganisms.** *Mol Microbiol* 2005, **57**(3):640–649.

doi:10.1186/1471-2164-13-299

Cite this article as: de Jong *et al.*: PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics* 2012 **13**:299.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

