

RESEARCH ARTICLE

Open Access

# Genomic reconstruction of transcriptional regulatory networks in lactic acid bacteria

Dmitry A Ravcheev<sup>1,2</sup>, Aaron A Best<sup>3</sup>, Natalia V Sernova<sup>2</sup>, Marat D Kazanov<sup>2</sup>, Pavel S Novichkov<sup>4</sup> and Dmitry A Rodionov<sup>1,2\*</sup>

## Abstract

**Background:** Genome scale annotation of regulatory interactions and reconstruction of regulatory networks are the crucial problems in bacterial genomics. The *Lactobacillales* order of bacteria collates various microorganisms having a large economic impact, including both human and animal pathogens and strains used in the food industry. Nonetheless, no systematic genome-wide analysis of transcriptional regulation has been previously made for this taxonomic group.

**Results:** A comparative genomics approach was used for reconstruction of transcriptional regulatory networks in 30 selected genomes of lactic acid bacteria. The inferred networks comprise regulons for 102 orthologous transcription factors (TFs), including 47 novel regulons for previously uncharacterized TFs. Numerous differences between regulatory networks of the *Streptococcaceae* and *Lactobacillaceae* groups were described on several levels. The two groups are characterized by substantially different sets of TFs encoded in their genomes. Content of the inferred regulons and structure of their cognate TF binding motifs differ for many orthologous TFs between the two groups. Multiple cases of non-orthologous displacements of TFs that control specific metabolic pathways were reported.

**Conclusions:** The reconstructed regulatory networks substantially expand the existing knowledge of transcriptional regulation in lactic acid bacteria. In each of 30 studied genomes the obtained regulatory network contains on average 36 TFs and 250 target genes that are mostly involved in carbohydrate metabolism, stress response, metal homeostasis and amino acids biosynthesis. The inferred networks can be used for genetic experiments, functional annotations of genes, metabolic reconstruction and evolutionary analysis. All reconstructed regulons are captured within the *Streptococcaceae* and *Lactobacillaceae* collections in the RegPrecise database (<http://regprecise.lbl.gov>).

**Keywords:** Transcriptional regulatory network, Comparative genomics, Carbohydrate metabolism, Lactobacillaceae, Streptococcaceae, Lactic acid bacteria, Regulon, Transcription factor

## Background

Regulation of gene expression in response to external and internal stimuli is a crucial mechanism for adaptation of microorganisms to changes of environmental conditions and intracellular states. In Bacteria, regulation of gene expression at the transcriptional level is usually mediated by transcription factors (TFs) that recognize their cognate TF-binding sites (TFBSs) in the promoter regions of regulated genes. A set of target genes under direct control of a certain

TF is defined as a regulon. All regulons in a single organism establish the transcriptional regulatory network (TRN), a fine-tuned system for complex regulation of gene expression in response to environmental changes and physiological needs of the cell.

Reconstruction of TRNs in bacterial genomes involves identification of regulatory interactions between target operons and TFs that requires genome-wide definition of all respective TFBSs. Various approaches for TRN reconstruction have been developed including traditional bottom-up genetic methods [1-3] and new top-down techniques based on the large-scale expression data [4] and/or automated inference of TFBS motifs [5-8]. On the other hand, the growing number of available complete genomic

\* Correspondence: [rodionov@burnham.org](mailto:rodionov@burnham.org)

<sup>1</sup>Sanford-Burnham Medical Research Institute, La Jolla, CA 92037, USA

<sup>2</sup>A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia

Full list of author information is available at the end of the article

sequences opens opportunities for comparative genomic analysis of transcriptional regulation and subsequent TRN inference (reviewed in [9,10]). This analysis can be efficiently used for both propagation of known regulons to previously uncharacterized organisms and for *ab initio* discovery of novel regulons. Combination of the comparative genomic-based regulon reconstruction with other genome context analysis techniques and multiple available experimental datasets define a knowledge-driven approach for reconstruction of TRNs in a set of related bacterial genomes. This approach has been successfully used for TRN inference in two groups of related genomes including 16 *Shewanella* species [11] and six species from the *Staphylococcaceae* family [12]. The comparative genomics approach not only allowed us to reconstruct TRNs in multiple genomes but also resulted in prediction of functions for previously uncharacterized genes.

In this study, we applied the knowledge-driven comparative genomic approach for reconstruction of TRNs in lactic acid bacteria belonging to the *Lactobacillales* order of the Firmicutes phylum. In spite of the large number of complete genomes and huge economic impact of this group (reviewed in [13-15]), most studies of gene regulation in these bacteria are limited to individual regulons in some model species; large-scale TRN reconstructions were attempted for *Lactobacillus plantarum* [16] and *Lactococcus lactis* [17]. On the other hand, availability of complete genomic sequences and multiple experimental data on gene regulation provide an opportunity for application of comparative genomic-based techniques for reconstruction of TRNs in lactic acid bacteria. To reconstruct genome-wide TRNs in the set of 30 *Lactobacillales* genomes we used a modified approach with three innovations. First, we subdivided the analyzed group of genomes into two phylogenetically distinct groups: the *Streptococcaceae* group including species from the *Streptococcus* and *Lactococcus* genera, and the *Lactobacillaceae* group that also includes two closely-related genomes from the *Leuconostocaceae* family. Second, to compose the initial training sets of TF-regulated genes we used the available experimental data on transcriptional regulation from 14 model species. Third, the coordinated reconstruction of a large number of TF regulons was carried out by a community of annotators using the RegPredict platform [18] with subsequent curation and quality control. Using this combined approach we reconstructed regulons for 102 orthologous TFs, including 47 novel regulons predicted for the first time in this study and awaiting further experimental validation.

## Results and discussion

We selected a set of 30 complete genomes in the *Lactobacillales* order for TRN reconstruction. Based on the

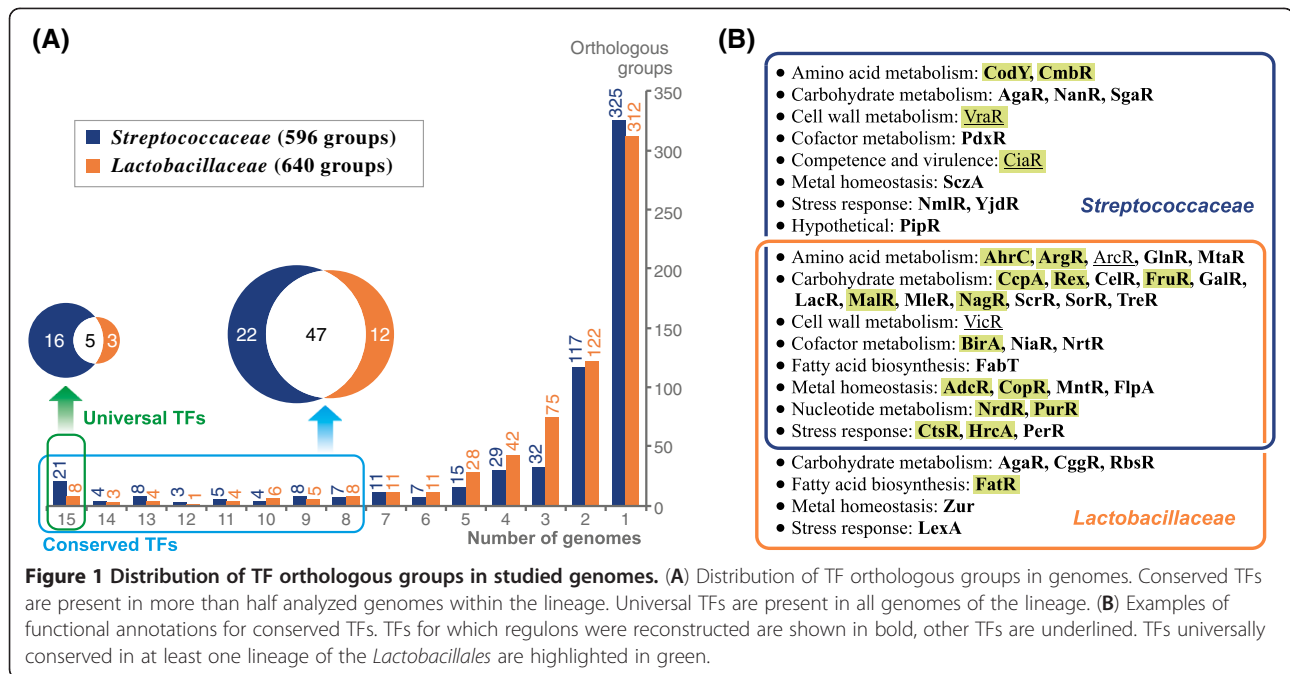
phylogenetic species tree (Additional file 1) all studied genomes were divided into two groups called the *Streptococcaceae* and *Lactobacillaceae*. The *Streptococcaceae* family includes 13 *Streptococcus* spp. and 2 strains of *Lactococcus lactis*. The second group includes 13 genomes from the *Lactobacillaceae* family and 2 genomes from the *Leuconostocaceae* family that are phylogenetically close to each other [19].

### Repertoire of TFs in *Lactobacillales* genomes

To estimate the scale and diversity of the TF mediated regulatory networks in the *Lactobacillales* genomes, we performed a genetic census of their putative DNA-binding TFs using similarity search and the existing prokaryotic TF compilations (Additional files 2, 3). The total number of putative TFs varies broadly within the *Lactobacillales* genomes, from ~60 in *S. thermophilus* and *L. helveticus* to ~240 in *L. plantarum* and ~150 in *S. gallolyticus* (Additional file 4).

The putative TFs identified in the *Lactobacillales* are distributed between 49 protein families and about 90% of these TFs belong to 24 major families with at least two representatives per genome. The largest number of TF representatives was observed for the Xre family (298 TFs total, ~19 TFs per genome). Among other large families with more than 4 TFs per genome are the TetR, GntR, MarR, OmpR, LacI, LysR, MerR and AraC families. Comparison of TF repertoires between the two *Lactobacillales* groups reveals 42 TF families that have representatives in both groups. Among the lineage-specific TF families, three families (CodY, PF04394, YobV) are present only in the *Streptococcaceae*, whereas four families (LexA, SdaR, SfsA, ComK) are unique for the *Lactobacillaceae*. Interestingly, the LexA and CodY regulators are both present in the *Staphylococcaceae*, *Bacillaceae*, and *Enterococcaceae* families of Firmicutes, suggesting the family-specific loss of these TFs and their regulons in the *Streptococcaceae* and *Lactobacillaceae* families, respectively.

The entire set of 3445 TFs identified in 30 studied genomes was broken into 596 orthologous groups in the *Streptococcaceae* and 640 orthologous groups in the *Lactobacillaceae* (Additional files 2, 3). The numbers of universal TFs present in all analyzed *Streptococcaceae* and *Lactobacillaceae* genomes are 21 and 8, respectively (Figure 1). At that, only 5 regulators are shared between these two groups of universal TFs including the global regulator for catabolite repression CcpA [20,21], the copper uptake regulator CopR [22], the purine biosynthesis regulator PurR [23], the redox control global regulator Rex [12], and the aminosugar utilization regulator NagR [24]. In each taxonomic group we defined all TFs that are present in more than half of the analyzed genomes as conserved TFs. Among the conserved TFs, 47



are shared by both studied lineages, whereas the *Streptococcaceae* and *Lactobacillaceae* groups have respectively 22 and 12 conserved TFs that are lineage-specific. The regulators of carbohydrate and amino acid metabolism form the most populated functional group of conserved TFs (Figure 1).

The studied two taxonomic groups of lactic acid bacteria demonstrate different distributions of orthologous TFs. The fractions of universal and conserved TFs are significantly higher in the *Streptococcaceae* than in the *Lactobacillaceae*. On the other hand, the *Lactobacillaceae* genomes are equipped by a confidently higher fraction of sporadically distributed TFs that are present in 2 to 6 genomes. Thus, the *Lactobacillaceae* are characterized by higher variability of TFs orthologous groups than the *Streptococcaceae* that is in correspondence with the larger phylogenetic distances between species in the *Lactobacillaceae* family (Additional file 1).

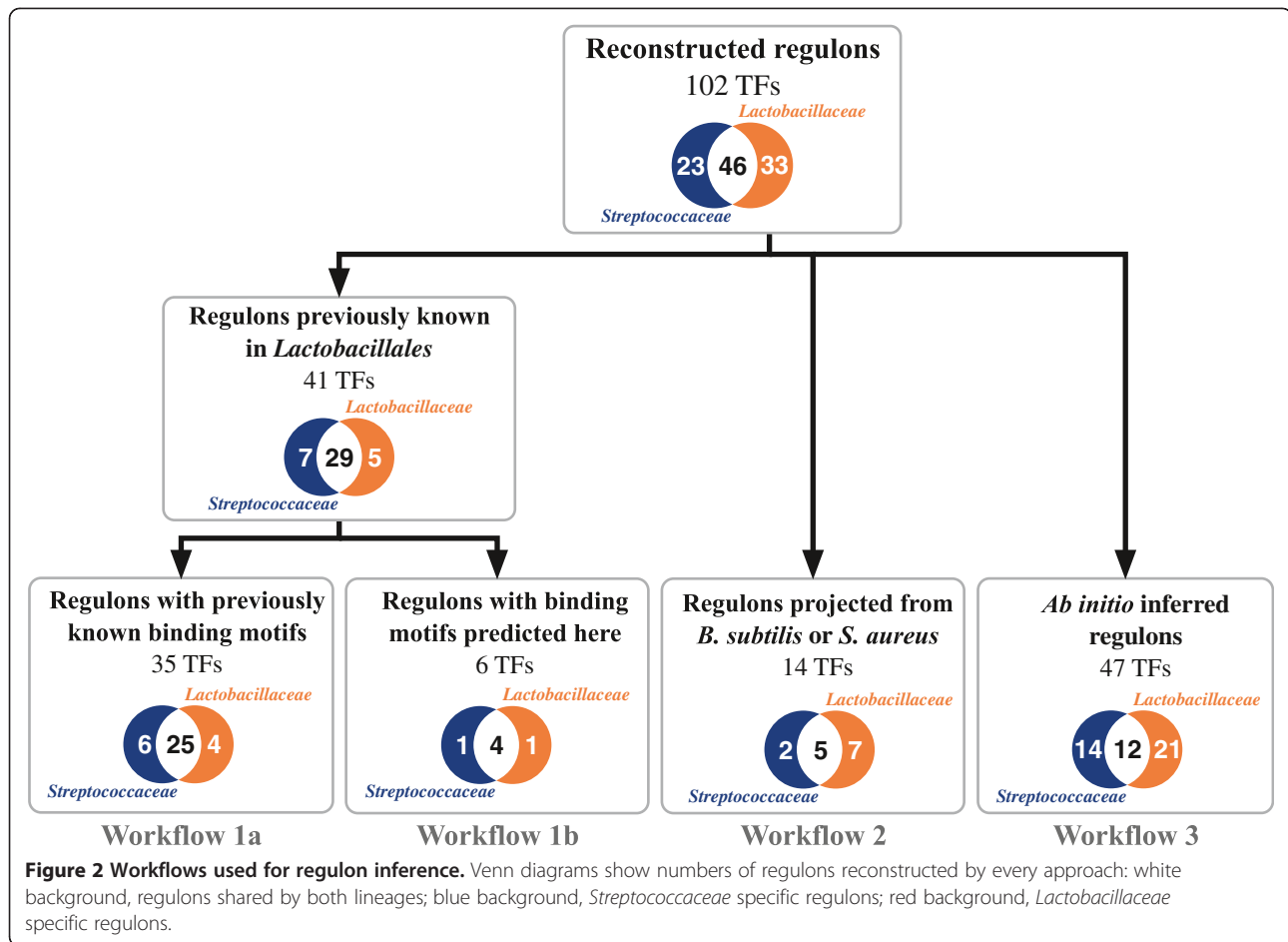
#### Reconstruction of regulons in two lineages of Lactobacillales

A comparative genomic approach implemented in the RegPredict Web server [18] was applied for regulon inference in the *Streptococcaceae* and *Lactobacillaceae* groups of genomes. Totally, 102 orthologous TF regulons were reconstructed in the studied genomes (Additional file 5). Initially we collected the published experimental data on transcriptional regulation in model *Lactobacillales* species (Additional file 5). Depending on the availability of experimental data, we applied three different workflows for regulon inference: (1) expansion and projection of TF regulons previously characterized

in *Lactobacillales*, (2) the reconstruction of regulons for TFs that have orthologs previously characterized in *B. subtilis* or *S. aureus*, and (3) *ab initio* inference of regulons for previously uncharacterized TFs (Figure 2).

A significant number of the studied *Lactobacillales* genomes, including eight *Streptococcus*, two *Lactococcus* and four *Lactobacillus* species, were considered as model species with previously characterized TFs and regulons (Additional file 5). By using workflow 1, we propagated the previously established regulatory interactions for 41 TFs in the model *Lactobacillales* species, and predicted new regulon members by the comparative genomics approach. Using workflow 2, we inferred regulons for twelve TFs that have orthologs previously experimentally investigated in *B. subtilis* and two TFs that were previously studied in *S. aureus*. Finally, using workflow 3, we predicted and described 47 novel TF regulons (Figure 2). Thus, all TF regulons studied in this work are either entirely predicted by computational analysis regulons or partially predicted regulons with some regulatory interactions supported by experimental data in model species.

The resulting set of reconstructed regulons varies drastically between the individual genomes (Table 1). The average number of reconstructed TF regulons per genome is 35.8, whereas the minimal and maximal numbers are 18 and 46 that were inferred in *L. delbrueckii* and *L. plantarum*, respectively (Figure 3). An average regulatory network of the reconstructed TF regulons includes ~250 genes per genome. The minimal number of genes in a reconstructed TRN was observed in *L. delbrueckii* (69 genes), whereas the maximal number of regulated genes in a network was in *S.*



*suis* (366 genes). In summary, the number of regulated genes in the reconstructed TF regulons is significantly higher in the *Streptococcaceae* genomes. However, *L. plantarum* from the *Lactobacillaceae* group has numbers of TFs and target genes in the reconstructed TRN comparable with *Streptococcaceae*. The observed variability in reconstructed TRNs of studied microorganisms can be explained by multiple factors including the diversity of their respective ecological niches and nutrient availability, and different metabolic capacities of individual species.

#### Classification of reconstructed regulons by taxonomic distribution

Distribution and properties of 102 TF regulons in 30 genomes of the *Lactobacillales* are summarized in Additional file 5. In each of the two taxonomic groups, the *Streptococcaceae* and *Lactobacillaceae*, we defined TF regulons that are universal (*i.e.* present in all 15 genomes) and the remaining regulons with a mosaic distribution in the analyzed genomes (Table 2). Only five regulons including the global regulons CcpA and Rex were found to be universally conserved in both lineages. Ten regulons that are universal in the *Streptococcaceae*

have orthologs with mosaic distribution in the *Lactobacillaceae*. In contrast, no TF regulons appeared in the group of regulons that are universal in the *Lactobacillaceae* but mosaic in the *Streptococcaceae*. A large set of 31 TF regulons that are mosaic in both lineages contains regulons controlling different sugar utilization pathways.

The remaining TF regulons reconstructed in this work are present only in a single taxonomic group. Among 23 TF regulons present solely in the *Streptococcaceae*, the CmbR, CodY and PipR regulons are universal, whereas the others have a mosaic distribution in 15 analyzed genomes. The SOS response regulon LexA is only lineage-specific regulon that was identified as universal in the *Lactobacillaceae*, whereas the remaining 20 lineage-specific TF regulons have a mosaic distribution in this lineage.

In conclusion, 23 TF regulons are specific for the *Streptococcaceae*, 33 regulons are present only in the *Lactobacillaceae* and 46 regulons have orthologs in both lineages. Thus, among the reconstructed regulons, the *Streptococcaceae* group has significantly larger number of universal TF regulons in comparison with the *Lactobacillaceae* group.

**Table 1 Statistics for reconstructed regulons in studied genomes**

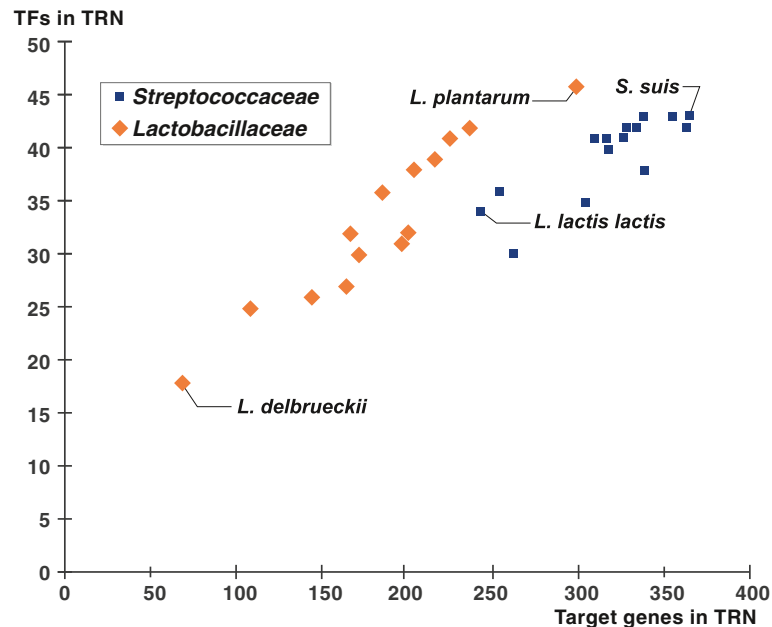
	Genome	TFs	Target genes	Target operons	Regulatory interactions	
<b>Streptococcaceae</b>	<i>L. lactis cremoris</i> SK11	36	255	125	130	
	<i>L. lactis lactis</i> II1403	34	244	128	138	
	<i>S. thermophilus</i> CNRZ1066	30	263	125	141	
	<i>S. agalactiae</i> 2603 V/R	38	340	159	186	
	<i>S. uberis</i> 0140 J	42	330	156	183	
	<i>S. equi</i> MGCS10565	42	334	143	167	
	<i>S. dysgalactiae</i> GGS_124	43	356	160	189	
	<i>S. pyogenes</i> M1 GAS	40	319	150	180	
	<i>S. gallolyticus</i> UCN34	41	328	167	199	
	<i>S. mutans</i> UA159	41	317	147	173	
	<i>S. suis</i> 05ZYH33	43	366	145	173	
	<i>S. mitis</i> B6	35	305	148	174	
	<i>S. pneumoniae</i> TIGR4	42	365	167	206	
	<i>S. gordonii</i> CH1	41	312	167	194	
	<i>S. sanguinis</i> SK36	43	339	163	191	
	<b>Total</b>		<b>591</b>	<b>4773</b>	<b>2250</b>	<b>2624</b>
		<b>Non-overlapping <sup>1</sup></b>	<b>69</b>	<b>779</b>	<b>397</b>	<b>470</b>
<b>Lactobacillaceae</b>	<i>L. sakei</i> 23 K	36	186	92	106	
	<i>L. casei</i> ATCC 334	41	226	110	120	
	<i>L. rhamnosus</i> GG	42	237	106	116	
	<i>L. delbrueckii</i> ATCC BAA-365	18	69	36	37	
	<i>L. acidophilus</i> NCFM	27	165	80	90	
	<i>L. helveticus</i> DPC 4571	21	91	53	55	
	<i>L. johnsonii</i> NCC 533	26	145	78	87	
	<i>P. pentosaceus</i> ATCC 25745	38	205	97	111	
	<i>L. brevis</i> ATCC 367	39	217	112	128	
	<i>L. plantarum</i> WCFS1	46	299	147	170	
	<i>L. fermentum</i> IFO 3956	30	172	85	101	
	<i>L. reuteri</i> JCM 1112	32	167	83	96	
	<i>O. oeni</i> PSU-1	25	109	59	70	
	<i>L. mesenteroides</i> ATCC 8293	32	202	89	103	
	<i>L. salivarius</i> UCC118	31	198	86	96	
	<b>Total</b>		<b>484</b>	<b>2688</b>	<b>1313</b>	<b>1486</b>
		<b>Non-overlapping <sup>1</sup></b>	<b>79</b>	<b>539</b>	<b>289</b>	<b>328</b>

<sup>1</sup>Numbers of orthologous groups of TFs, genes and operons in lineage.

### Classification of reconstructed regulons by TF protein families

The reconstructed regulons are controlled by TFs from 31 protein families. The mostly represented TF family in the obtained set of regulons is the LacI family (26 regulons). As expected, all LacI family regulons reconstructed in this work controls different carbohydrate utilization pathways. Another broadly represented TF family is GntR family (13 regulons) that control carbon and amino acid metabolisms and resistance to toxic compounds. TFs from the

TetR and MarR families (7 and 6 reconstructed regulons, respectively) regulate genes involved in environmental adaptation, multidrug and heavy metal resistance and fatty acids metabolism. The reconstructed regulons from the RpiR family control various carbohydrate utilization pathways (5 regulons), whereas the predicted AguR regulon controls the agmatine utilization pathway. The BglR and DeoR families of TFs include respectively 6 and 5 regulators that control carbohydrate catabolism. Three regulons operated by TFs from the LysR family regulate the cysteine/



**Figure 3** Scatter plot for numbers of TFs and target genes in studied genomes. Genomes with minimal and maximal numbers of target genes in each lineage are signed.

methionine metabolism, whereas one regulon (MleR) controls the malonate utilization pathway.

**Classification of reconstructed regulons by function**

The functional gene content was assessed for each reconstructed TF regulon to tentatively predict its possible biological function and molecular effector. Metabolic reconstruction of the respective biochemical pathways and prediction of functions of co-regulated genes were performed using the subsystem-based approach implemented in the SEED genomic platform [25,26]. The inferred functional annotations of genes constituting regulons were captured in the RegPrecise database [27] within the taxonomic

collections of regulons for the *Streptococcaceae* and *Lactobacillaceae* groups ([http://regprecise.lbl.gov/RegPrecise/collections\\_tax.jsp](http://regprecise.lbl.gov/RegPrecise/collections_tax.jsp)).

Overall, the reconstructed regulons were classified into 8 functional groups (Additional file 5). The largest group counts 56 regulons for carbohydrate and central carbon metabolism. Three other functional groups of regulons are involved of stress response (11 regulons), metal homeostasis (9 regulons) and amino acid metabolism (8 regulons). Small numbers of regulons were reconstructed for cofactor metabolism (4 regulons), nucleotide metabolism (3 regulons) and fatty acid metabolism (2 regulons). Additionally, 9 reconstructed TF regulons contain genes with unknown

**Table 2** Distribution of TFs with reconstructed regulons for orthologous TFs in *Streptococcaceae* and *Lactobacillaceae* genomes

TF type <sup>1</sup>		TF number	Examples
Streptococcaceae	Lactobacillaceae		
Universal	Universal	5	CcpA, CopR, NagR, PurR, Rex
Universal	Mosaic	10	AdcR, ArgR, BirA, CtsR, FabT, FruR, GlnR, HrcA, MalR, NrdR
Mosaic	Universal	0	n/a
Mosaic	Mosaic	31	CcpB, CelR, FucR, GalR, GutR, LacR, MdxR, MleR, MntR, MtaR, MtlR, MurR, NiaR, NrtR, PadR, PerR, PflR, ScrR, TagR, TreR, UxuR
Universal	—	3	CmbR, CodY, PipR
Mosaic	—	20	AgaR, AlsR, CelQ, HomR, NanR, NmlR, PdxR, RegR, Rgg, RgrA, RliC, SczA, SgaR
—	Universal	1	LexA
—	Mosaic	32	AguR, AraR, CggR, DeoR, ExuR, FatR, HxlR, lolR, NihR, RbsR, RpiR, SdaR, XylR, Zur

Numbers of regulons for orthologous TFs in *Streptococcaceae* and *Lactobacillaceae* are shown. <sup>1</sup> 'Universal' regulons are present in all 15 studied genomes of the group; 'Mosaic' regulons are present in group but in less than 15 genomes; '—' means the absence of regulons in this group.

or hypothetical functions, and thus their specific functional roles and effectors remain unknown. Functional content for the selected subset of reconstructed regulons is briefly described below.

### **Carbohydrate metabolism**

Carbohydrates comprise a key source of carbon and energy for a variety of microorganisms. Genomes from the *Lactobacillales* order encode a large number of sugar catabolic pathways and most of these pathways have a mosaic distribution in individual species [28]. This diversity of sugar catabolic pathways is matched by a large number of regulatory systems that allow sugar-specific induction of expression of these pathways.

Here we reconstructed 56 TF regulons that control the sugar and central carbon metabolism in the *Lactobacillales*. In addition to 19 sugar metabolism regulons that were previously characterized experimentally in model *Lactobacillales* species, 6 regulons were reconstructed by projection from *B. subtilis* or *S. aureus* and 31 novel regulons were predicted for the first time in this work (Additional file 5). Analysis of available experimental data revealed 114 previously known regulatory interactions involving the sugar metabolism regulators in model *Lactobacillales* organisms. Using comparative genomics, we were able to significantly expand the sugar metabolism regulatory subnetwork. As an example, the previously known regulator of hyaluronidase RegR in *S. pneumoniae* [29] was predicted to have an expanded regulon of 12 additional genes involved in the hyaluronate utilization that is conserved in six *Streptococcaceae* genomes (Additional file 6).

About 10 new regulatory interactions per genome were predicted for carbohydrate metabolism genes using the projection of known regulons from *B. subtilis* and *S. aureus*. For example, CcpN in *B. subtilis* was previously described as a regulator of the gluconeogenesis genes *gapB* and *pckA* [30]. The comparative genomics reconstruction of the CcpN regulon revealed the pyruvate phosphate dikinase *ppdK*, the *ccpN* gene and the fructose biphosphatase *fbp* as novel members of the CcpN regulon in the *Lactobacillales* (Additional files 6, 7). Thus, the novel CcpN regulon in the *Lactobacillales* have a set of target genes that is completely different from the known CcpN regulon in *B. subtilis*, although in both lineages it controls the gluconeogenesis pathway.

The reconstructed regulatory network includes 52 TFs that control 38 peripheral sugar utilization pathways, and 10 of these pathways are controlled by more than one TF. For instance, we reconstructed 5 different TF regulons for maltose and maltodextrin utilization pathway and 3 TF regulons for the sucrose catabolism. Two different TF regulons per one sugar metabolic pathway were described for pathways involved in utilization of ascorbate, cellobiose,

gluconate, lactose, N-acetylgalactosamine, ribose and trehalose. The observed redundancy in sugar-specific TFs is explained by (i) non-orthologous replacements of TFs for the same pathway in different genomes, and (ii) existence of alternative pathway variants and multiple paralogs regulated by different TFs in the same genome. For example, two maltose/maltodextrin ABC transporters are controlled by two non-orthologous TFs from the LacI family in the *Streptococcaceae* (Additional file 8). The *malEFG* operon is always regulated by MalR protein. The *malXCDA* operon is regulated by MalR in four *Streptococcus* spp. and under control of MalR2 in three other *Streptococcus* spp.

Three other sugar utilization pathways in the *Streptococcus* spp. are equipped by a redundant set of catabolic genes controlled by multiple non-orthologous regulators. SgaR and SgaR2 from the BglG family control the ascorbate utilization. The GntR family regulators AgaR and AgaR2 were predicted to control the N-acetylgalactosamine utilization gene loci. The cellobiose utilization genes are regulated by CelR from the BglG family [31] and CelQ from the ROK family [32]. These functionally redundant sets of TF regulons indicate a complex evolutionary history of the sugar utilization subsystems in Firmicutes.

### **Stress response**

We reconstructed totally 11 TF regulons involved in various stress responses and drug resistance. These include the CtsR, HrcA, NmlR, PadR and PerR regulons that were previously experimentally described in at least one model *Lactobacillales* genome. The SOS response regulon LexA and multidrug resistance regulon YtrA were projected using the previous knowledge of orthologous regulons in *B. subtilis* [33,34]. Overall, we predicted 13 target operons in the *Lactobacillaceae* LexA regulon including some novel functions (e.g. *parEC*, *addBA*, *nrnA*).

The Fur family regulator PerR was previously studied in *S. pyogenes* and *S. suis*, where it co-regulates genes involved in peroxide stress response and manganese transport [35,36]. Here we predicted some new members of the PerR regulon involved in iron transport (*fhuADBG*, *feoAB*, *fatDCAB*) and iron-sulfur cluster biosynthesis (*sufCDSEB*). The latter iron-sulfur cluster biosynthesis operon was previously identified by us as a novel member of the PerR regulon in the *Staphylococcaceae* family [12]. Interestingly, we also predicted that in some *Streptococcus* spp., PerR controls expression of another TF for manganese homeostasis, MntR, and these two regulators form a potential cascade.

### **Metal homeostasis**

The group of reconstructed regulons for metal homeostasis includes 9 TF regulons. Starting from 24 regulatory interactions previously described in 5 known regulons for model *Lactobacillales* organisms, we expanded and projected the

AdcR, CopR, FlpA, MntR and SczA regulons resulting in prediction of several new regulatory interactions in the genomes. For instance, we predicted regulation of genes encoding manganese (*mntH*) and nickel (*nikABCDE*) transporters, heavy metal-transporting ATPase (*pmtA*), and *mntR* gene by the manganese homeostasis regulon MntR in the *Streptococcaceae*.

In contrast to the AdcR regulon for control of zinc homeostasis in the *Streptococcaceae*, the *Lactobacillaceae* species utilize the typical zinc-responsive regulator Zur from the Fur family. The Zur regulon in the *Lactobacillaceae*, which was reconstructed by projection from the *B. subtilis* Zur regulon, includes the zinc transporter *znuABC*, and the *zur* and *rpsN2* genes, the latter encoding a paralog of the ribosomal protein S14.

Two novel TFs from the ArsR family analyzed in the *Lactobacillaceae* are distantly related to the CzcR regulator of cobalt-zinc-cadmium resistance from *Pseudomonas aeruginosa* [37]. The reconstructed CzcR1 and CzcR2 regulons were predicted to control different sets of genes involved in heavy metal resistance. Another novel regulon, termed NihR (nickel homeostasis regulator), was predicted by its conserved co-localization with the *nikMQO* genes encoding a nickel transport system. The predicted NihR regulon also includes the *nihR* and *nixA* genes, the latter encoding a high-affinity nickel permease (Additional file 7).

#### **Amino acids metabolism**

Among 8 reconstructed TF regulons for control of amino acid metabolism, seven regulons have been previously described in at least one model *Lactobacillales* species. For these TF regulons, we projected 56 regulatory interactions to additional genomes and predicted ~25 novel regulatory interactions per genome. First, we predicted the MtaR- and CmbR-dependent regulation of the *mmuMP* operon involved in S-methylmethionine utilization in the *Streptococcaceae*. Second, we identified multiple new members of the nitrogen metabolism regulon GlnR in both lineages including the glutamine transporter *gluQHMP*, the arginine catabolism genes *arcABC* and the aspartate-ammonia ligase *asnA*. A novel amino acid metabolism regulon that was inferred for the first time in this work is operated by a RpiR-like regulator, termed AguR (agmatine utalization regulator), which was found to be co-localized with agmatine utilization genes in the *Lactobacillaceae* genomes.

#### **CRISPR-Cas genes under the control of global regulators**

CRISPR-Cas is a recently discovered prokaryotic RNA-based system for adaptive immunity for defense against phages, plasmids and other mobile genetic elements [38,39]. Previously, the expression of CRISPR-Cas genes has been shown to be regulated by different global TFs, such as H-NS, Lrp, LeuO [40,41] and Crp [42] in *E. coli*,

and Rex in *Thermotoga maritima* [43]. In this work we predicted some new cases of regulation of CRISPR-Cas genes in the *Lactobacillales* genomes. The predicted operon *cas9-cas1-cas2-csn2* is preceded by a candidate CcpA-binding site in *S. agalactiae* and *S. mutans*, whereas in *S. agalactiae* and *S. equi* this operon has a candidate CodY-binding site in its promoter region. Another putative CRISPR-Cas operon, *cas5-cas8c-cas7-cas1-cas2*, is predicted to be regulated by CcpA in *S. equi* and *S. pyogenes*. These predicted regulatory interactions for CRISPR-Cas genes extend our understanding of regulatory mechanisms for bacterial immune systems.

#### **Evolution of regulons in Lactobacillales**

##### **Evolution of orthologous TF regulons**

Among 102 TF regulons reconstructed in this work, 46 regulons have orthologous TFs in both studied lineages of the *Lactobacillales* (Additional file 5). We compared the deduced TFBS motifs for the orthologous TF regulons in the *Streptococcaceae* and *Lactobacillaceae* lineages and classified them into three categories (Additional file 9). Category I includes 23 TF regulons with binding motifs that are well conserved or slightly variable in two lineages. Category II contains 17 TF regulons with moderately different motifs (2 to 4 mismatches in the conserved motif positions). Category III has 6 remaining TF regulons with binding motifs that are substantially different between the *Streptococcaceae* and *Lactobacillaceae*. Remarkably, the category I is enriched by universal and highly conserved regulators. Thus, it includes all 5 regulons that are universal in both studied lineages and 10 regulons for TFs that are universal in the *Streptococcaceae* and have a mosaic distribution in the *Lactobacillaceae*. The category I contains the highest number of global regulons (CcpA, Rex) and mid-size regulons that control 3 or more target operons per genome (AdcR, ArgR, CtsR, GlnR, HrcA, MalR, MtaR, PerR and PurR). The remaining 13 regulons in this category are local, i.e. containing less than 3 target operons per genome. In contrast, the categories II and III contain respectively 94% and 100% of local regulons. These observations suggest that the conservation of TFBS motifs have a positive correlation with the regulon size. Similar correlation was previously reported for the *Staphylococcaceae* spp. [12].

Analysis of conservation of gene contents for orthologous TF regulons between the two studied lineages classified all regulons into three different categories. The first group contains 27 strictly conserved regulons that do not demonstrate any difference in their gene content or have only slight changes between the *Streptococcaceae* and *Lactobacillaceae* (e.g. an insertion of additional genes into target operons). The second group includes 16 regulons that have a common core of conserved genes supplemented by unique sets of peripheral genes that are substantially



different between the *Streptococcaceae* and *Lactobacillaceae*. This group of partially conserved regulons includes the global regulons CcpA and Rex and several mid-size regulons (see below). The third group contains 3 orthologous regulons, MdxR, NiaR and NrtR, having completely different sets of target genes between the *Streptococcaceae* and *Lactobacillaceae* groups, although their corresponding biological roles are conserved between the studied groups. High diversity of the latter three regulons can be explained by redundancy of the respective metabolic systems and regulatory mechanisms. Thus, MdxR controls the maltose and maltodextrin utilization and the similar biological role was assigned to several other TF regulons in the *Lactobacillales* (see above). The phylogenetic distributions of two NAD metabolism regulons NiaR and NrtR have an overlapping pattern in many studied genomes (Additional file 5). In conclusion, we did not observe any correlation between the conservation of orthologous TF regulon structure and their cognate binding motifs in the analyzed taxonomic groups.

The 16 TF regulons with a common core and flexible set of peripheral genes between the *Streptococcaceae* and *Lactobacillaceae* groups can be classified into three subgroups: (i) regulons expanded in the *Streptococcaceae* (AdcR, BirA, CopR, GlnR, LacR and MntR); (ii) regulons expanded in the *Lactobacillaceae* (HrcA and ScrR), and (iii) regulons with different peripheral sets of genes in the *Streptococcaceae* and *Lactobacillaceae* (ArgR, CcpA, CtsR, MalR, NrdR, PerR, PurR and Rex). We propose that different sets of peripheral genes in the above 16 regulons have appeared due to independent expansion of regulons in each lineage. A common core of the ArgR regulon is formed by genes for the arginine transport and biosynthesis. The periphery of this regulon in the *Streptococcaceae* consists of the *argR* gene and *arcABC* operon for arginine catabolism, whereas the extended part of the ArgR regulon in the *Lactobacillaceae* includes only carbamoyl-phosphate synthase (*carAB*). The PurR regulon in the *Streptococcaceae* is expanded by genes for metabolism of folate-associated one-carbon compounds, whereas in the *Lactobacillaceae* this regulon has additional genes for the adenine and guanine metabolism (*purB* and *guaB*) and ribose-phosphate pyrophosphokinase (*prsA*). Expansion of the PerR regulon in the *Streptococcaceae* affects genes for the iron and manganese homeostasis, whereas in the *Lactobacillaceae* PerR regulates several additional genes such as a NADH peroxidase. The CcpA regulon demonstrated the largest peripheral sets of genes in the studied lineages, 117 operons in the *Streptococcaceae* and 42 operons in the *Lactobacillaceae*. Noteworthy, the CcpA regulon in the *Streptococcaceae* is expanded by the CRISPR-Cas cassette genes (see above) and some virulence genes (the exfoliative toxin A gene *shetA* and the streptolysin S

biosynthesis operon *sagABCEFGHI*), suggesting that this global regulon supplies the link between the carbohydrate utilization, virulence and anti-phage immunity.

#### **Non-orthologous displacements of TF regulons**

Functional analysis of reconstructed TF regulons in the *Lactobacillales* revealed that many biological processes are regulated by two or more non-orthologous TFs. These metabolic subsystems with redundant TF regulation include 8 distinct sugar utilization pathways (see above) and the zinc homeostasis that are controlled by 20 non-orthologous TFs (Additional file 5). In most cases, patterns of phylogenetic distribution of non-orthologous TFs controlling the same biological subsystem complement each other in the analyzed 30 genomes. Interestingly, there are only four cases in which the non-orthologous TFs belong to different protein families (CelR/CelQ, GntR1/GntR2, LacR/LacR2 and AdcR/Zur), whereas in all remaining cases the identified pairs of non-orthologous TFs belong to the same TF family and thus can be classified as cases of xenologous gene replacement [44]. The largest number of xenologous replacements of TFs was identified within the LacI family (Additional file 8).

#### **Interconnectivity in reconstructed TRNs**

The cross talk between TF regulons can be identified by prediction of TFBSs for two or more TFs within the regulatory region of the same operon. Numerous target operons in the reconstructed TRNs are subject to regulation by multiple TFs. For instance, *S. pneumoniae* has 38 target operons that share TFBSs for two or more TFs. Regulation of several target operons by multiple TFs is evolutionary conserved across a number of related genomes (Table 3).

Three global regulons, CcpA, CodY and Rex, often interconnect with each other and also with multiple local TF regulons. For example, the CcpA regulon overlaps with 31 TF regulons that control the carbohydrate utilization such as MalR, ScrR, TreR, GalR, FruR, and RegR. The CodY regulon overlaps with the GlnR and CmbR regulons that control the nitrogen metabolism and sulfur amino acid biosynthesis, respectively. A similar situation was previously observed in *S. aureus*, where the CcpA regulon overlaps with numerous local regulons for sugar utilization and the CodY regulon overlaps with amino acid metabolism regulons [12,45,46]. Similarly, co-regulation of two heat shock response regulons, HrcA and CtsR, was found both in the *Streptococcaceae* and *Lactobacillaceae* groups, and was also previously observed in *S. aureus* [12,47].

Autoregulation of TFs is a regular feature of the reconstructed regulons. An average portion of the *Lactobacillales* TFs with predicted control of their own expression is 72%. This index slightly varies between the analyzed genomes and is very close to the percentage of

**Table 3 Examples of target operons regulation by multiple TFs**

Combinations of TFs	Examples	
	Target operon	Genome(s)
<b>Quadruple regulation</b>		
AdcR, CcpA, CodY, Rex	<i>adhB1</i>	<i>S. agalactiae</i> , <i>S. dysgalactiae</i>
<b>Triple regulation</b>		
AdcR, CodY, Rex	<i>adhB1</i>	<i>S. equi</i> , <i>S. mitis</i> , <i>S. pyogenes</i>
MtaR, CmbR, HomR	<i>metEF</i>	<i>S. gallolyticus</i> , <i>S. gordonii</i> , <i>S. mutans</i> , <i>S. pneumoniae</i>
CcpA, MalR, MdxR	<i>malXCDA</i>	<i>L. plantarum</i>
CcpA, MalR3, MdxR	<i>npIT</i>	<i>L. plantarum</i>
CcpA, MalR, MdxR	<i>malT-mapA-pgmB</i>	<i>L. mesenteroides</i>
<b>Double regulation</b>		
CcpA, Rex	<i>adhE</i>	<i>S. mitis</i> , <i>S. sanguinis</i> , <i>L. acidophilus</i> , <i>L. brevis</i> , <i>L. johnsonii</i>
	<i>forT</i>	<i>S. dysgalactiae</i> , <i>S. equi</i> , <i>S. pyogenes</i> , <i>S. sanguinis</i> , <i>S. uberis</i>
CcpA, MalR	<i>ptsG-rgfB</i>	<i>S. agalactiae</i> , <i>S. equi</i> , <i>S. gordonii</i> , <i>S. mutans</i> , <i>S. uberis</i>
	<i>pulA</i>	<i>S. agalactiae</i> , <i>S. equi</i> , <i>S. gordonii</i> , <i>S. mutans</i> , <i>S. pyogenes</i>
HrcA, CtsR	<i>groSL</i>	all <i>Streptococcaceae</i> , <i>L. fermentum</i> , <i>L. reuteri</i> , <i>P. pentosaceus</i>
	<i>hrcA-grpE-dnaKJ</i>	<i>L. fermentum</i> , <i>L. plantarum</i> , <i>L. reuteri</i> , <i>L. sakei</i> , <i>P. pentosaceus</i>
CcpA, ScrR	<i>scrBR</i>	<i>S. gallolyticus</i> , <i>S. gordonii</i> , <i>L. acidophilus</i> , <i>L. johnsonii</i> ,
CcpA, TreR	<i>trePA</i>	<i>S. dysgalactiae</i> , <i>S. pneumoniae</i> , <i>L. mesenteroides</i> , <i>O. oeni</i>
CcpA, GalR	<i>galKETRM</i>	<i>L. casei</i> , <i>L. rhamnosus</i>
	<i>lacLM</i>	<i>L. fermentum</i> , <i>L. plantarum</i> , <i>L. salivarius</i> , <i>P. pentosaceus</i>
CcpA, CodY	<i>livKHMGF</i>	<i>S. gordonii</i> , <i>S. mitis</i> , <i>S. sanguinis</i> , <i>S. suis</i> , <i>S. thermophilus</i>
CodY, GlnR	<i>gluQHMP</i>	<i>S. agalactiae</i> , <i>S. gallolyticus</i> , <i>S. mutans</i>
CcpA, FruR	<i>fruRBA</i>	<i>L. lactis</i> , <i>S. equi</i> , <i>S. suis</i> , <i>L. helveticus</i> , <i>L. sakei</i> , <i>P. pentosaceus</i>
PerR, MntR	<i>mntABC</i>	<i>S. equi</i> , <i>S. pyogenes</i> , <i>S. mutans</i> , <i>S. sanguinis</i> , <i>S. uberis</i>
CcpA, RegR	<i>hylD-ugl-hylEFG-ohl-regR</i>	<i>S. agalactiae</i> , <i>S. pyogenes</i>

autoregulated TFs that was previously reported for *S. aureus* [12].

Multiple regulatory cascades between various TFs were detected in the studied genomes. For instance, CcpA regulates the *codY* gene and multiple genes encoding sugar utilization regulators, such as FruR, ScrR, TreR, and CelR. Among other identified cascades, CodY controls *glnR*, CmbR controls *homR* and PerR controls *mntR* in the *Streptococcaceae*, whereas CtsR controls *hrcA* in the *Lactobacillaceae*. Some cascades, such as the regulation of *fruR*, *galR*, *gutR*, *mtlR*, *scrR* and *treR* genes by CcpA, are conserved between the *Streptococcaceae* and *Lactobacillaceae* lineages, whereas the remaining cases of regulatory TF cascades are lineage specific. Comparison with the previously reconstructed regulatory network of *S. aureus* [12] showed that some of the identified *Lactobacillales* regulatory cascades are also conserved in the *Staphylococcaceae*. For instance, the CcpA-dependent regulation of *fruR*, *rbsR2*, *scrR* and *treR*, as well as the regulation of *glnR* by CodY and *hrcA* by CtsR, are conserved between the two groups. The conservation of cascades between distantly related genomes

points to the importance of these regulatory interactions in the *Lactobacillales* regulatory networks.

## Conclusions

The knowledge-based bottom-up approach and comparative genomics techniques have been previously successively applied for reconstruction of bacterial TRNs in different groups of genomes [11,12]. Here we tentatively defined the reference collection of TF regulons in 30 *Lactobacillales* genomes comprised of 102 orthologous groups of TFs and ~4100 regulatory interactions (~140 per genome). The resulting regulatory network contains ~7500 regulated genes (~250 per genome) that are involved in sugar utilization, stress response, metal homeostasis and metabolisms of amino acids, fatty acids, nucleotides, and cofactors. We used a modified workflow for TRN reconstruction that is characterized by three main innovations: (1) analysis of two taxonomically related groups of genomes (the *Streptococcaceae* and *Lactobacillaceae*), (2) involvement of numerous experimental data from the literature about TF regulation in lactic acid bacteria, and (3)

coordinated reconstruction of multiple TF regulons by a community of annotators using multi-user web interface of the RegPredict tool for regulon analysis [18]. By utilizing the semi-automatic workflow for regulon inference combined with manual curation and regulon annotation we described the largest reference collection of TF regulons in lactic acid bacteria do date. The collection also includes the previously uncharacterized regulons for 47 TFs that comprise ~1000 target genes (12 – 63 genes per genome). Thus, a significant number of predicted regulatory interactions and novel TF regulons await future experimental validation. During preparation of this manuscript, Bitoun *et al.* experimentally analyzed Rex regulon in the *S. mutans* UA159 [48] and confirmed Rex dependent regulation of 4 targets predicted in this work (*adhE*, *rex-guaA*, *ldh*, and *fdCT*).

Comparison of the inferred TRNs in the *Lactobacillales* genomes reveals interesting trends in the evolution of TRNs and individual TF regulons. First, we found a positive correlation between the TFBS motif conservation between the two lactic acid bacteria lineages and the distribution and conservation of the respective TF regulons. Second, we report that in the analyzed collection of *Lactobacillales* regulons non-orthologous displacements of TFs occur more often between structurally related TFs in comparison with regulators that belong to different protein families.

## Methods

Thirty complete genomes of *Lactobacillales* (Additional file 1) were downloaded from MicrobesOnline database [49]. Primary TF sets for each studied genome were extracted from P2TF (<http://www.p2tf.org>) database. Sigma factors and RNA binding proteins were excluded from the collections. Groups of orthologs were pre-counted as following. Initially, groups of orthologous proteins were constructed for every pair of TFs sets. All pairwise comparisons were done using BLASTP, and bidirectional best hits (BBHs) were identified if the protein sequences identity was more 50% and the aligned region was longer than 2/3 of the length of the shorter protein. If two paralogous genes from one genome were more similar to each other than to a BBH partner from another genome, both paralogs were added to the same orthology cluster. Finally, all orthologous clusters containing common genes were joined together. The clusters were formed using ad-hoc software written using Oracle RDBMS Express Edition (PL/SQL codes are available by request). TF families were assigned by analysis of protein domain structure using the following databases: CDD [50], Pfam [51], SMART [52], and MicrobesOnline Domain and Families [49].

For regulon reconstruction we used the previously established comparative genomics approach (reviewed in [10]) implemented in the RegPredict Web server

(<http://regpredict.lbl.gov>) [18]. The approach includes inference of TFBSs, construction of nucleotide positional weight matrices (PWMs) for TFBSs motifs, and reconstruction of regulons in complete genomes on the basis of prediction of putative TFBSs in promoter gene regions. To take into account possible lineage specific changes in TFBSs motifs, we constructed individual PWMs for the *Streptococcaceae* and *Lactobacillaceae* taxonomic groups.

The three major workflows used for TF regulon reconstruction are (1) projection and expansion of previously known regulons from model *Lactobacillales* organisms, (2) projection of known regulons from model organisms belonging to another taxa, and (3) *ab initio* prediction of novel regulons (Figure 2).

In workflow 1, the projection and expansion of previously know TF regulons includes two slightly different workflows. In workflow 1a, both a set of regulated genes and TFBSs motif are known, whereas in workflow 1b, only a set of co-regulated genes is known from the collected experimental data. For previously known TFBSs motifs, a PWM was built and used for identification of additional sites in the analyzed genomes using the Run Profile tool in the RegPredict Web server. All novel true positive TFBSs were added to the training set and the updated PWM was constructed and further used for final regulon reconstruction. For regulons with originally unknown TFBS motifs, we collected a set of upstream regions of known TF-regulated genes and their orthologs and used this set for TFBS identification by the Discover Profile tool in the RegPredict. The TFBS motif discovery tool uses the expectation-maximization algorithm for clustering of all potential motifs with a specified symmetry (palindrome, direct or inverted repeat) and finally optimizes the inferred PWM. In ambiguous cases, putative regulatory elements were validated by phylogenetic footprinting [53] using multiple alignments for upstream non-coding regions of orthologous genes.

In workflow 2, the previously experimentally studied regulons in other model organisms from the Firmicutes phylum (*B. subtilis* or *S. aureus*) were projected to the *Lactobacillaceae* genomes. For TFBS identification, we used training sets of upstream regions of genes that are considered as orthologs to the TF-regulated genes from other model species outside of the *Lactobacillales* lineage.

Workflow 3 was used for *ab initio* prediction of novel TF regulons. Initially, the presumably co-regulated genes were predicted by the analysis of conserved gene neighborhoods around a putative TF gene. Upstream regions of presumably co-regulated genes extracted from multiple *Lactobacillales* genomes were used for identification of TFBSs and PWM construction as described above.

The obtained PWMs for known or predicted TFBS motifs were used for comparative genomics reconstruction

of regulons in two groups of genomes, the *Lactobacillaceae* and *Streptococcaceae*, using the RegPredict Web server [18]. Each studied genome was scanned with the constructed PWMs using Run Profile tool in RegPredict. The threshold for site search was defined as a lowest score observed in the training set. The consistency check approach [10,54] and/or functional relatedness of candidate target operons were used to eliminate false positive TFBS predictions.

A community of annotators consisting of 18 undergraduate students from Hope College (Holland, MI) performed initial reconstruction of multiple TF regulons in the analyzed groups of genomes. This represented a coordinated annotation effort with expert curators in the context of a microbiology course. The resulting draft regulons underwent strict quality control to ensure accuracy of the reconstructions.

Functional gene annotations were uploaded from SEED [25], UniProt [55] and MicrobesOnline [49]. Multiple alignments of protein and DNA sequences were built by MUSCLE [56]. Phylogenetic trees were constructed using maximum likelihood algorithm implemented in PHYLIP package (v 3.69) [57] and visualized via Dendroscope tool [58]. Complete description of the reconstructed regulons including TFs, their target genes and operons, and associated TFBS were uploaded to the RegPrecise database (<http://regprecise.lbl.gov>) [27].

## Additional files

**Additional file 1: Phylogenetic tree of 30 studied *Lactobacillaceae* genomes.** - Description of data: Economic impact for each studied genome is shown in square brackets. The tree is based on approximately 78 universal prokaryotic proteins in the MicrobesOnline database: <http://www.microbesonline.org/cgi-bin/speciesTree.cgi>.

**Additional file 2: Repertoire of DNA-binding transcriptional factors identified in 15 *Streptococcaceae* genomes.** Description of data: Orthologous groups are sorted by TF family and then by conservancy of each TF group. TFs with regulons reconstructed in this work are highlighted by light blue.

**Additional file 3: Repertoire of DNA-binding transcriptional factors identified in 15 *Lactobacillaceae* genomes.** Description of data: Orthologous groups are sorted by TF family and then by conservancy of each TF group. TFs with regulons reconstructed in this work are highlighted by light blue.

**Additional file 4: Distribution of predicted DNA binding transcription factors in studied *Lactobacillales* genomes.**

**Additional file 5: Collection of TF regulons reconstructed in *Lactobacillales*.** Description of data: <sup>a</sup> Novel TF names introduced in this work are marked by asterisks. <sup>b</sup> Presence (+) or absence (-) of TFs orthologs. <sup>c</sup> Workflow 1, expansion and projection of a regulon previously characterized in model *Lactobacillales* organisms: (1a) TFBSs motif was known, (1b) TFBSs motif was predicted in present work; Workflow 2, projection of an orthologous regulon from *B. subtilis* or *S. aureus*; Workflow 3, *ab initio* regulon inference. Regulons previously studied in model organisms highlighted by green. Names of functional groups are highlighted in blue.

**Additional file 6: Functional content, experimental evidences and conservation for reconstructed regulons in *Streptococcaceae*.**

Description of data: Regulons are sorted by regulator names. Novel TF names introduced in this work are marked by asterisks. 'Conservation of regulatory interaction' column shows number of genomes with regulated gene/ operon (number of genomes having orthologs of operon).

**Additional file 7: Functional content, experimental evidences and conservation for reconstructed regulons in *Lactobacillaceae*.**

Description of data: Regulons are sorted by regulator names. Novel TF names introduced in this work are marked by asterisks. 'Conservation of regulatory interaction' column shows number of genomes with regulated operon (number of genomes having orthologs of operon).

**Additional file 8: Analysis of Lacl family TFs in the studied *Lactobacillales* genomes.**

**Additional file 9: Comparison of predicted binding site motifs in *Streptococcaceae* and *Lactobacillaceae* genomes.** Description of data: <sup>1</sup> Sequences Logos were constructed using WebLogo package (<http://weblogo.berkeley.edu/logo.cgi>). <sup>2</sup> NS, number of binding site sequences used to construct Logo. <sup>3</sup> Category reflects a conservancy between TFBS motifs in *Streptococcaceae* and *Lactobacillaceae*: I, highly conserved motifs; II, moderately different motifs; III, substantially different motifs. <sup>4</sup> Number of genomes that contain this regulon. <sup>5</sup> Average number of target operon in regulon per genome.

## Abbreviations

TF: Transcription factor; TFBS: Transcription factor binding site; TRN: Transcriptional regulatory network.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

DARo and PSN conceived and designed the research project. DARa and DARo wrote the manuscript. DARa, DARo, and NVS performed comparative genomic analysis for reconstruction of regulons. DARo also provided the quality control as a curator. MDK computed orthologous clusters for TFs. AAB contributed to the development of the project, writing of the manuscript, and implemented the community annotation effort at Hope College. All authors read and approved the final manuscript.

## Acknowledgements

This work was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under contract DE-SC0004999 with SBMRI and LBNL. Additional funding was provided by the Russian Foundation for Basic Research (12-04-33003 and 12-04-31939), State contract #8135 (application 2012-1.2.2-12-000-1013-079) to MDK, and by the National Science Foundation under award DBI-0850546 to AAB and DARo. We would like to thank the students of the Fall 2011 Microbiology course at Hope College for their work to build initial reconstructions of regulons.

## Author details

<sup>1</sup>Sanford-Burnham Medical Research Institute, La Jolla, CA 92037, USA. <sup>2</sup>A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia. <sup>3</sup>Department of Biology, Hope College, Holland, MI 49423, USA. <sup>4</sup>Lawrence Berkeley National Laboratory, Berkeley, CA 94710, USA.

Received: 16 November 2012 Accepted: 8 February 2013

Published: 12 February 2013

## References

- Goelzer A, Bekkal Brikci F, Martin-Verstraete I, Noirot P, Bessieres P, Aymerich S, Fromion V: **Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*.** *BMC Syst Biol* 2008, **2**:20.
- 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, et al: **RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Sensor Units).** *Nucleic Acids Res* 2011, **39**:D98-D105.

3. Schroder J, Tauch A: **Transcriptional regulation of gene expression in *Corynebacterium glutamicum*: the role of global, master and local regulators in the modular and hierarchical gene regulatory network.** *FEMS Microbiol Rev* 2010, **34**(5):685–737.
4. Barrett CL, Palsson BO: **Iterative reconstruction of transcriptional regulatory networks: an algorithmic approach.** *PLoS Comput Biol* 2006, **2**:e52.
5. Alkema WB, Lenhard B, Wasserman WW: **Regulog analysis: detection of conserved regulatory networks across bacteria: application to *Staphylococcus aureus*.** *Genome Res* 2004, **14**:1362–1373.
6. Tan K, Moreno-Hagelsieb G, Collado-Vides J, Stormo GD: **A comparative genomics approach to prediction of new members of regulons.** *Genome Res* 2001, **11**:566–584.
7. Conlan S, Lawrence C, McCue LA: **Rhodopseudomonas palustris regulons detected by cross-species analysis of alphaproteobacterial genomes.** *Appl Environ Microbiol* 2005, **71**:7442–7452.
8. Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC, et al: **A predictive model for transcriptional control of physiology in a free living cell.** *Cell* 2007, **131**:1354–1365.
9. Gelfand MS: **Evolution of transcriptional regulatory networks in microbial genomes.** *Curr Opin Struct Biol* 2006, **16**:420–429.
10. Rodionov DA: **Comparative genomic reconstruction of transcriptional regulatory networks in bacteria.** *Chem Rev* 2007, **107**:3467–3497.
11. Rodionov DA, Novichkov PS, Stavrovskaya ED, Rodionova IA, Li X, Kazanov MD, Ravcheev DA, Gerasimova AV, Kazakov AE, Kovaleva GY, et al: **Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus.** *BMC Genomics* 2011, **12**(1):S3.
12. Ravcheev DA, Best AA, Tintle N, Dejongh M, Osterman AL, Novichkov PS, Rodionov DA: **Inference of the transcriptional regulatory network in *Staphylococcus aureus* by integration of experimental and genomics-based evidence.** *J Bacteriol* 2011, **193**:3228–3240.
13. Carr FJ, Chill D, Maida N: **The lactic acid bacteria: a literature survey.** *Crit Rev Microbiol* 2002, **28**:281–370.
14. Kleerebezem M, de Vos WM: **Lactic acid bacteria: life after genomics.** *Microb Biotechnol* 2011, **4**:318–322.
15. Burne RA, Bessen DE, Broadbent JR, Claverys JP: **The seventh international conference on the genetics of streptococci, lactococci, and enterococci.** *J Bacteriol* 2007, **189**:1209–1218.
16. Wels M, Overmars L, Francke C, Kleerebezem M, Siezen RJ: **Reconstruction of the regulatory network of *Lactobacillus plantarum* WCFS1 on basis of correlated gene expression and conserved regulatory motifs.** *Microb Biotechnol* 2010, **4**:333–344.
17. de Jong A, Pietersma H, Cordes M, Kuipers OP, Kok J: **PePPER: a webserver for prediction of prokaryote promoter elements and regulons.** *BMC Genomics* 2012, **13**:299.
18. Novichkov PS, Rodionov DA, Stavrovskaya ED, Novichkova ES, Kazakov AE, Gelfand MS, Arkin AP, Mironov AA, Dubchak I: **RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach.** *Nucleic Acids Res* 2010, **38**:W299–W307.
19. Claesson MJ, van Sinderen D, O'Toole PW: **Lactobacillus phylogenomics—towards a reclassification of the genus.** *Int J Syst Evol Microbiol* 2008, **58**:2945–2954.
20. Zheng L, Chen Z, Itzek A, Ashby M, Kreth J: **Catabolite control protein A controls hydrogen peroxide production and cell death in *Streptococcus sanguinis*.** *J Bacteriol* 2011, **193**:516–526.
21. Zotta T, Ricciardi A, Guidone A, Sacco M, Muscariello L, Mazzeo MF, Cacace G, Parente E: **Inactivation of *ccpA* and aeration affect growth, metabolite production and stress tolerance in *Lactobacillus plantarum* WCFS1.** *Int J Food Microbiol* 2012, **155**(1–2):51–59.
22. Magnani D, Barre O, Gerber SD, Solioz M: **Characterization of the CopR regulon of *Lactococcus lactis* IL1403.** *J Bacteriol* 2008, **190**:536–545.
23. Jendresen CB, Martinussen J, Kilstrup M: **The PurR regulon in *Lactococcus lactis* - transcriptional regulation of the purine nucleotide metabolism and translational machinery.** *Microbiology* 2012, **158**(Pt 8):2026–2038.
24. Bertram R, Rigali S, Wood N, Lulko AT, Kuipers OP, Titgemeyer F: **Regulon of the N-acetylglucosamine utilization regulator NagR in *Bacillus subtilis*.** *J Bacteriol* 2011, **193**:3525–3536.
25. Disz T, Akhter S, Cuevas D, Olson R, Overbeek R, Vonstein V, Stevens R, Edwards RA: **Accessing the SEED genome databases via Web services API: tools for programmers.** *BMC Bioinforma* 2010, **11**:319.
26. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, et al: **The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.** *Nucleic Acids Res* 2005, **33**:5691–5702.
27. 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**:D111–D118.
28. Klaenhammer TR, Barrangou R, Buck BL, Azcarate-Peril MA, Altermann E: **Genomic features of lactic acid bacteria effecting bioprocessing and health.** *FEMS Microbiol Rev* 2005, **29**:393–409.
29. Chapuy-Regaud S, Ogunniyi AD, Diallo N, Huet Y, Desnottes JF, Paton JC, Escaich S, Trombe MC: **RegR, a global LacI/GaIR family regulator, modulates virulence and competence in *Streptococcus pneumoniae*.** *Infect Immun* 2003, **71**:2615–2625.
30. Servant P, Le Coq D, Aymerich S: **CcpN (YqzB), a novel regulator for CcpA-independent catabolite repression of *Bacillus subtilis* gluconeogenic genes.** *Mol Microbiol* 2005, **55**:1435–1451.
31. Shafeeq S, Kloosterman TG, Kuipers OP: **CellR-mediated activation of the cellobiose-utilization gene cluster in *Streptococcus pneumoniae*.** *Microbiology* 2011, **157**:2854–2861.
32. Shafeeq S, Kloosterman TG, Rajendran V, Kuipers OP: **Characterization of the ROK-family transcriptional regulator Roka of *Streptococcus pneumoniae* D39.** *Microbiology* 2012, **158**(Pt 12):2917–2926.
33. Au N, Kuester-Schoeck E, Mandava V, Bothwell LE, Canny SP, Chachu K, Colavito SA, Fuller SN, Groban ES, Hensley LA, et al: **Genetic composition of the *Bacillus subtilis* SOS system.** *J Bacteriol* 2005, **187**:7655–7666.
34. Groban ES, Johnson MB, Banky P, Burnett PG, Calderon GL, Dwyer EC, Fuller SN, Gebre B, King LM, Sheren IN, et al: **Binding of the *Bacillus subtilis* LexA protein to the SOS operator.** *Nucleic Acids Res* 2005, **33**:6287–6295.
35. Grifantini R, Toukoki C, Colaprico A, Gryllos I: **Peroxide stimulon and role of PerR in group A *Streptococcus*.** *J Bacteriol* 2011, **193**:6539–6551.
36. Zhang T, Ding Y, Li T, Wan Y, Li W, Chen H, Zhou R: **A fur-like protein PerR regulates two oxidative stress response related operons *dpr* and *metQIN* in *Streptococcus suis*.** *BMC Microbiol* 2012, **12**:85.
37. Perron K, Caille O, Rossier C, Van Delden C, Dumas JL, Kohler T: **CzcR-CzcS, a two-component system involved in heavy metal and carbapenem resistance in *Pseudomonas aeruginosa*.** *J Biol Chem* 2004, **279**:8761–8768.
38. Koonin EV, Makarova KS: **CRISPR-Cas: an adaptive immunity system in prokaryotes.** *F1000 Biol Rep* 2009, **1**:95.
39. Terns MP, Terns RM: **CRISPR-based adaptive immune systems.** *Curr Opin Microbiol* 2011, **14**:321–327.
40. Medina-Aparicio L, Rebollar-Flores JE, Gallego-Hernandez AL, Vazquez A, Olvera L, Gutierrez-Rios RM, Calva E, Hernandez-Lucas I: **The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in *Salmonella enterica* serovar Typhi.** *J Bacteriol* 2011, **193**:2396–2407.
41. Westra ER, Pul U, Heidrich N, Jore MM, Lundgren M, Stratmann T, Wurm R, Raine A, Mescher M, Van Heereveld L, et al: **H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO.** *Mol Microbiol* 2010, **77**:1380–1393.
42. Shinkai A, Kira S, Nakagawa N, Kashihara A, Kuramitsu S, Yokoyama S: **Transcription activation mediated by a cyclic AMP receptor protein from *Thermus thermophilus* HB8.** *J Bacteriol* 2007, **189**:3891–3901.
43. Ravcheev DA, Li X, Latif H, Zengler K, Leyn SA, Korostelev YD, Kazakov AE, Novichkov PS, Osterman AL, Rodionov DA: **Transcriptional regulation of central carbon and energy metabolism in bacteria by redox responsive repressor Rex.** *J Bacteriol* 2012, **194**:1145–1157.
44. Koonin EV, Makarova KS, Aravind L: **Horizontal gene transfer in prokaryotes: quantification and classification.** *Annu Rev Microbiol* 2001, **55**:709–742.
45. Somerville GA, Proctor RA: **At the crossroads of bacterial metabolism and virulence factor synthesis in *Staphylococci*.** *Microbiol Mol Biol Rev* 2009, **73**:233–248.
46. Sonenshein AL: **Control of key metabolic intersections in *Bacillus subtilis*.** *Nat Rev Microbiol* 2007, **5**:917–927.

47. Chastanet A, Fert J, Msadek T: Comparative genomics reveal novel heat shock regulatory mechanisms in *Staphylococcus aureus* and other Gram-positive bacteria. *Mol Microbiol* 2003, **47**:1061–1073.
48. Bitoun JP, Liao S, Yao X, Xie GG, Wen ZT: The redox-sensing regulator Rex modulates central carbon metabolism, stress tolerance response and biofilm formation by streptococcus mutans. *PLoS One* 2012, **7**:e44766.
49. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, *et al*: MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res* 2010, **38**:D396–D400.
50. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, *et al*: CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 2011, **39**:D225–D229.
51. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R: Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* 1998, **26**:320–322.
52. Schultz J, Milpetz F, Bork P, Ponting CP: SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* 1998, **95**:5857–5864.
53. Shelton DA, Stegman L, Hardison R, Miller W, Bock JH, Slightom JL, Goodman M, Gumucio DL: Phylogenetic footprinting of hypersensitive site 3 of the beta-globin locus control region. *Blood* 1997, **89**:3457–3469.
54. Ravcheev DA, Gerasimova AV, Mironov AA, Gelfand MS: Comparative genomic analysis of regulation of anaerobic respiration in ten genomes from three families of gamma-proteobacteria (Enterobacteriaceae, Pasteurellaceae, Vibrionaceae). *BMC Genomics* 2007, **8**:54.
55. Magrane M, Consortium U: UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011, **2011**:bar009.
56. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004, **32**:1792–1797.
57. Felsenstein J: Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Meth Enzymol* 1996, **266**:418–427.
58. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R: Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinforma* 2007, **8**:460.

doi:10.1186/1471-2164-14-94

**Cite this article as:** Ravcheev *et al.*: Genomic reconstruction of transcriptional regulatory networks in lactic acid bacteria. *BMC Genomics* 2013 **14**:94.

**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](http://www.biomedcentral.com/submit)

