



HAL
open science

SalmoNet, an integrated network of ten *Salmonella enterica* strains reveals common and distinct pathways to host adaptation

Aline Métris, Padhmanand Sudhakar, David Fazekas, Amanda Demeter, Eszter Ari, Marton Olbei, Priscilla Branchu, Rob A. Kingsley, Jozsef Baranyi, Tamas Korcsmaros

► To cite this version:

Aline Métris, Padhmanand Sudhakar, David Fazekas, Amanda Demeter, Eszter Ari, et al.. SalmoNet, an integrated network of ten *Salmonella enterica* strains reveals common and distinct pathways to host adaptation. *npj Systems Biology and Applications*, 2017, 3 (1), 10.1038/s41540-017-0034-z . hal-02620203

HAL Id: hal-02620203

<https://hal.inrae.fr/hal-02620203v1>

Submitted on 25 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

TECHNOLOGY FEATURE OPEN

SalmoNet, an integrated network of ten *Salmonella enterica* strains reveals common and distinct pathways to host adaptation

Aline Métris^{1,5}, Padhmanand Sudhakar^{1,2}, David Fazekas^{2,3}, Amanda Demeter^{1,2,3}, Eszter Ari^{3,4}, Marton Olbei^{1,2}, Priscilla Branchu^{1,6}, Rob A. Kingsley¹, Jozsef Baranyi¹ and Tamas Korcsmáros^{1,2}

Salmonella enterica is a prominent bacterial pathogen with implications on human and animal health. *Salmonella* serovars could be classified as gastro-intestinal or extra-intestinal. Genome-wide comparisons revealed that extra-intestinal strains are closer relatives of gastro-intestinal strains than to each other indicating a parallel evolution of this trait. Given the complexity of the differences, a systems-level comparison could reveal key mechanisms enabling extra-intestinal serovars to cause systemic infections. Accordingly, in this work, we introduce a unique resource, SalmoNet, which combines manual curation, high-throughput data and computational predictions to provide an integrated network for *Salmonella* at the metabolic, transcriptional regulatory and protein-protein interaction levels. SalmoNet provides the networks separately for five gastro-intestinal and five extra-intestinal strains. As a multi-layered, multi-strain database containing experimental data, SalmoNet is the first dedicated network resource for *Salmonella*. It comprehensively contains interactions between proteins encoded in *Salmonella* pathogenicity islands, as well as regulatory mechanisms of metabolic processes with the option to zoom-in and analyze the interactions at specific loci in more detail. Application of SalmoNet is not limited to strain comparisons as it also provides a *Salmonella* resource for biochemical network modeling, host-pathogen interaction studies, drug discovery, experimental validation of novel interactions, uncovering new pathological mechanisms from emergent properties and epidemiological studies. SalmoNet is available at <http://salmonet.org>.

npj Systems Biology and Applications (2017)3:31; doi:10.1038/s41540-017-0034-z

INTRODUCTION

The genus *Salmonella* includes pathogens associated with syndromes ranging from gastroenteritis to bacteraemia and enteric fever.¹ Gastroenteritis caused by *Salmonella* is one of the most common foodborne diseases, with nearly 100 million cases per year occurring worldwide.² While enteric fever is rare in developed countries, it is still associated with significant mortality and morbidity in low income countries with over 90,000 deaths worldwide in 2015.³ *Salmonella* pathogenesis depends on a large number of virulence genes including those located on large pathogenicity islands encoding type III secretion systems that translocate effector proteins into the host cell cytoplasm.⁴ *S. enterica* subspecies I includes over 1500 different serovars and accounts for the vast majority of human infections.⁵ Based on the epidemiological record, disease symptoms and observations from experimental infections has resulted in the classification of serovars into two pathovars, namely gastro-intestinal and extra-intestinal. Most serovars of subspecies I are of the former pathovar and most often associated with gastro-intestinal infections. Serovars of gastro-intestinal pathovars often exhibit a broad host range. However, a small number of serovars are host-adapted and

are characterized by an extra-intestinal infection and dissemination beyond the intestinal mucosa followed by colonization of systemic sites of the reticuloendothelial system. As most serovars are of the gastro-intestinal pathovar type, the most parsimonious explanation for the extra intestinal serovars is that they evolved from a gastro-intestinal pathovar ancestor, most likely on multiple occasions. The molecular basis of host adaptation has been studied most extensively in *S. enterica* serovar Typhi (*S. Typhi*), the causative agent of typhoid fever. Adaptation of *S. Typhi* is characterized by the acquisition of a number of virulence associated genes and the loss of coding capacity affecting over 200 genes.^{6,7}

Genes horizontally acquired by *S. Typhi* include a large pathogenicity island (SP-7) encoding biosynthesis genes for the Vi polysaccharide capsule and the TviA regulator protein,⁶ and the typhoid toxin that is encoded outside SPI-7.⁸ Dissemination beyond the intestinal mucosa is in part mediated by evasion of detection by the host innate immune system by expression of the Vi polysaccharide capsule,⁹ and by the down-regulation of flagella expression, a pathogen associated molecular pattern (PAMP), mediated by TviA.¹⁰ The function of the typhoid toxin in pathogenesis is not clear, however many of the symptoms of

¹Quadram Institute Bioscience, Norwich Research Park, Norwich NR4 7UA, UK; ²Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, UK; ³Department of Genetics, Eötvös Loránd University, Pázmány P. s. 1C, H-1117 Budapest, Hungary and ⁴Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary

Correspondence: Tamas Korcsmáros (Tamas.Korcsmaros@earlham.ac.uk)

⁵Present address: Safety and Environmental Assurance Centre, Unilever, Colworth Science Park, Sharnbrook, Bedfordshire, UK

⁶Present address: IRSD, Université de Toulouse, INSERM, INRA, ENVT, UPS, Toulouse, France

Aline Métris, Padhmanand Sudhakar, and David Fazekas contributed equally to this work.

Received: 6 May 2016 Revised: 19 September 2017 Accepted: 22 September 2017

Published online: 18 October 2017

Table 1. Information about the numbers corresponding to the data sources and the reconstructed networks for the reference strain *Salmonella Typhimurium* LT2

Network type	Data source	Number of interactions in <i>S. Typhimurium</i> LT2	
Metabolic	Model validated by flux-balance analysis ⁸²	2312	
	BioModel database ⁷⁵	754	
Regulatory	Experimental evidence in <i>Salmonella</i>	Manual curation of low-throughput experiments	9
		Datasets containing high-throughput experiments	234
	Genome-wide predictions	Based on experimentally verified binding sites in <i>Salmonella</i>	1189
		Based on <i>E. coli</i> binding sites from RegulonDB ¹⁷	1865
PPI	Experimental evidence in <i>Salmonella</i>	Manual curation	27
		IntAct database ⁷²	29
	Proteome-wide predictions	Structure based predictions using the Interactome 3D resource ⁹⁶	290
Orthology based predictions using <i>E. coli</i> PPI data from, ⁹⁸ IntAct ⁷² and BioGRID ⁹⁷		1846	

typhoid fever were induced by injection of the typhoid toxin into mice.¹¹

However, many of the extra-intestinal serovars of *Salmonella* do not encode SPI-7 or the typhoid toxin. Therefore, alternative mechanisms for the systemic dissemination are likely to have evolved in these serovars. This reflects the convergent evolution to an extra-intestinal lifestyle reflected in the phylogenetic relationship of these pathogens. Convergence in genome sequence polymorphisms of extra-intestinal serovars of *S. enterica* has previously been observed in the form of loss of coding capacity (genome degradation) due to deletions and pseudogene formation.⁷ Degradation of coding sequences corresponding to genes associated with the intestinal phase of infection such as anaerobic metabolism, motility and chemotaxis, and enteropathogenesis was over-represented in these serovars. A similar pattern of genome degradation was also observed in a rapidly evolving hypermutator strain of *S. Enteritidis* that was restricted to a systemic site niche in an immunocompromised patient.¹²

Serovars of *S. enterica* subspecies I exhibit divergence in their nucleotide sequences that corresponds to approximately 737,000 SNPs.¹³ In some cases, non-synonymous SNPs alter the function of proteins, and may alter the function of non-coding sequences when the SNPs are present in regulatory sequences or small RNAs. Serovars also encode hundreds of serovar-specific genes, as well as contain varying degrees of genome degradation that result in considerable differences in coding capacities and gene expression regulation. In light of this complexity, there is a need to apply a systems biology approach to compile network information across the metabolic, transcriptional regulatory and protein-protein interaction (PPI) layers in order to address the hypothesis that extra-intestinal serovars exhibit convergence in molecular mechanisms of host adaptation. Integration of interaction information from multiple layers is expected to provide insights into the shared attributes that characterize *Salmonella* pathogenicity and virulence.

In order to gain further insight into how *Salmonella* host adaptation has evolved there is a need to integrate different levels of knowledge (e.g., metabolism and regulation) as current data resources store the different layers separately, making complex analysis difficult. While a substantial amount of data on regulation, metabolism and protein-protein interactions is available, much of this is curated for model organisms, such as *Escherichia coli*. Integrating different types of information into a complex network remains a challenge for non-model organisms, like *Salmonella*.¹⁴

For example, in the case of transcriptional regulation, widely used resources such as ORegAnno,¹⁵ PAZAR¹⁶, or RegulonDB¹⁷ do not contain information on *Salmonella*. Other resources such as KEGG¹⁸ provides information only on metabolic pathways, and even these reactions are not direct *Salmonella* connections but orthology based inferences using *E. coli*. Furthermore, there are no resources that combine curated and predicted interaction information for *Salmonella*. Thus existing resources are either not comprehensive enough to capture multi-layer information or do not contain *Salmonella* specific interaction data.

We therefore compiled the metabolic, regulatory and PPI networks of 10 representative strains of *Salmonella* comprising 5 gastro-intestinal and 5 extra-intestinal strains. In addition to the interactions corresponding to the manually curated information specific to *Salmonella* pathogenicity islands, the networks also contain regulatory interactions derived from high-throughput experiments and whole-genome motif scans apart from interactions inferred from *E. coli* by orthology.

RESULTS AND DISCUSSION

Networks, data representation, and quality control

In this study, we have established a workflow to collect and infer interaction information from three different network levels (metabolic, regulatory and PPI) based on various sources (Table 1). We used data derived from literature, online databases, as well as genome-wide predictions. To further enrich the dataset, we performed this on the whole genome sequence assemblies of a range of *Salmonella* strains (Table 2) representing two pathovars (gastro-intestinal and extra-intestinal) that exhibit distinct life styles. The resulting networks are available to the scientific community for further analysis and enhancement at <http://salmonet.org>. The networks contain three layers (metabolic, regulatory and PPI) for 5 gastro-intestinal and 5 extra-intestinal strains (Supplementary Table 1).

The SalmoNet database consists of a total of 81,514 interactions involving 30,870 genes across the studied strains (see the strain specific distribution in Table 3). Considering all the annotated genes for the strains (49,472 genes), SalmoNet therefore covers 62% of the coding capacity of all the strains. In terms of the number of individual ortholog groups, SalmoNet contains information on the interacting partners of 132 sets of transcription factors (TFs) in the regulatory network, 1282 sets of proteins in the PPI network, as well as information on 1196 sets of enzymes in the

Table 2. Strains included in the study and their life-style

Serovar	Strain	Lifestyle	N.p ^a	Genome assembly ID ^b
Typhi	CT18	Extra-intestinal, causes typhoid fever in humans	2	000195995.1
Paratyphi	ATCC 9150	Extra-intestinal, second most prevalent cause of typhoid fever	0	000011885.1
Choleraesuis	SC-B67	Extra-intestinal, causes swine paratyphoid	2	000008105.1
Dublin	CT 02021853	Extra-intestinal, bovine-adapted serovar	1	000020925.1
Gallinarum	287/91	Extra-intestinal, causative agent of fowl typhoid in poultry	0	000009525.1
Agona	SL483	Gastro-intestinal	1	000020885.1
Enteritidis PT4	P125109	Gastro-intestinal	1	000009505.1
Heidelberg	SL476	Gastro-intestinal	2	000020705.1
Newport	SL254	Gastro-intestinal	2	000016045.1
Typhimurium	SL1344	Gastro-intestinal	3	000210855.2
Typhimurium	LT2	Gastro-intestinal (reference strain closely related to SL1344)	1	000006945.1

^aN.p. number of plasmids^b GenBank database (<http://www.ncbi.nlm.nih.gov/genbank/>)**Table 3.** Number of genes/proteins and their interactions from the networks for the different *Salmonella* strains

Strain name	Number of proteins	Metabolic network		Regulatory network		PPI network	
		nodes	links	nodes	links	nodes	links
Typhi	4718	1121	2348	1710	2595	1235	1949
Choleraesuis	4607	1137	2390	1650	2913	1223	1953
Dublin	4606	1170	2542	1583	2735	1247	2036
Gallinarum	3943	1140	2432	1484	2628	1200	1924
Paratyphi	4083	1136	2380	1565	2692	1202	1923
Agona	4592	1182	2584	1652	2845	1235	1978
Enteritidis	4192	1206	2653	1680	2921	1266	2062
Heidelberg	4757	1187	2590	1638	2736	1266	2072
Newport	4784	1189	2611	1582	2766	1256	2055
Typhimurium SL1344	4657	1228	2762	1735	3107	1287	2068
Typhimurium LT2	4533	1227	2763	1794	3288	1352	2213
Average	4497	1175	2550	1643	2838	1251	2021

metabolic network. Of the total 6070 unique connections in the regulatory network, 16% were present in all the 10 strains (Supplementary Figure 1) spanning the gastro-intestinal and extra-intestinal pathovars, thus comprising a core subset of the *Salmonella* regulatory networks inferred from our workflow (Fig. 1). The edges in this core are those that represent higher confidence since they follow the principle of cross-strain conservation.¹⁹ This methodology has previously been used as a basis for minimizing false positives. The ratio for PPIs and metabolic interactions present in all the 10 strains were higher: 72.6% and 69.2%, respectively. Variation in the fraction of each network represented by the core in regulatory and PPI/metabolic networks supports the idea that transcriptional regulation evolves at a faster rate than the PPI or metabolic levels,²⁰ although the noise arising from the heterogeneous sources used for the reconstruction of the regulatory network cannot be ignored. The use of the *matrix-quality* tool²¹ to determine customized *P*-values for every TF-strain combination for the transcriptional regulatory (binding site) predictions minimizes the high false positive rates which could otherwise arise from using generic *P*-value thresholds. Due to the low number of true positive sets, Precision-Recall

calculations could not be inferred for most of the transcription factors analyzed in the study. However, to exemplify the reliability of the networks, we determined the target Recall rates (recovery of known targets) for one of the transcription factors SsrB. SsrB regulates the expression of multiple target genes including a number of virulence factors including members of the *Salmonella* pathogenicity islands.^{22,23} 24 instances of the 18 bp SsrB binding motif in *S. Typhimurium* SL1344 have been reported,²³ By performing a random and equal bifurcation of the known binding sites into test and training sets, we were able to achieve recall rates of 75% in the reconstructed regulatory network for *S. Typhimurium* SL1344 (Supplementary Table 2). Furthermore, swapping the test and training sets yielded a recall rate of 64% suggesting that the reconstructed networks are robust in terms of recovery of true positives. With this example, we show that the predicted regulatory interactions in SalmoNet recover previously reported binding sites due to the employed stringencies such as an informed *P*-value. Besides, users can further enhance the networks by choosing only those interactions which occur in multiple strains of each serovar or all the strains in the study depending on their use case.

The scientific community can access the database via the aforementioned dedicated web resource in which the molecular entities can be searched by their gene names, UniProt accession IDs, and locus tags. The original source of the interactions that was used during the data integration workflow is also displayed. To enable the comparison of interactions across strains, the ortholog clustering IDs (generated during this study) are also provided for individual molecular entities. An easy-to-use option is provided for users to download selected interaction sets from particular layers of the network or for particular strains. The core *Salmonella* network (the set of interactions conserved across all the strains) can also be accessed similarly. The files are available for download both in CSV and Cytoscape formats allowing straightforward further filtering and visualization, respectively.

Network dendrograms for comparison among strains

To determine the evolutionary relatedness of the ten *Salmonella* strains in SalmoNet, we constructed a phylogenetic tree based on their nucleotide sequences. All the polymorphic sites found in the common ortholog genes of all the strains were used to build a Bayesian dendrogram (Fig. 2a). Four of the gastro-intestinal strains (*S. Typhimurium* LT2, *S. Typhimurium* SL1344, *S. Heidelberg* SL476 and *S. Newport* SL254) were monophyletic on the polymorphic site based phylogenetic tree but two of them were clustered together with extra-intestinal strains: *S. Enteritidis* P125109 with

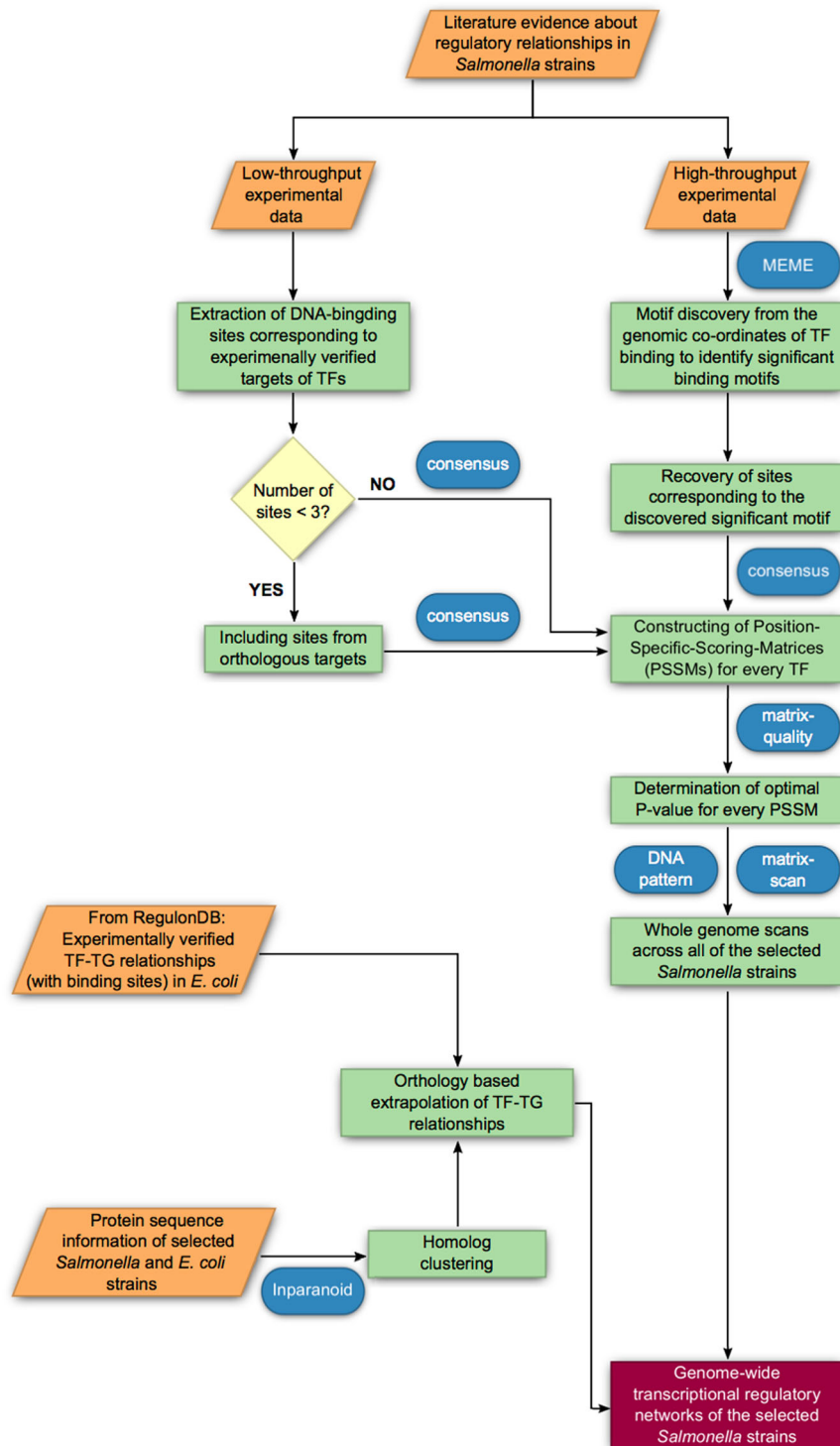


Fig. 1 Workflow depicting the steps followed in the reconstruction of the transcriptional regulatory networks of the *Salmonella* strains

S. Gallinarum 287 91; and *S. Agona* SL483 with *S. Typhi* CT18 and *S. Paratyphi* A ATCC 9150. The tree was constructed assuming an approximately equal rate of mutation in each strain, and based on this assumption, the common ancestor of these strains is central within the genome based tree. Strains from extra-intestinal and gastro-intestinal serovars could not be distinguished based on the topology of the genome based dendrogram as observed in previous studies.²⁴ This is consistent with these pathogens independently emerging as extra-intestinal pathogens and that

their attributes arise multiple times during evolution by a process of convergence in genome degradation in the anaerobic metabolism as also described by Nuccio et al.⁷

Next, we compared the phylogenetic relationship of the extra-intestinal and gastro-intestinal pathogens with their metabolic, regulatory and PPI networks to determine if network adaptations converge or if they reflect the evolutionary history of the strains. We used the matrix representation of the networks to infer Bayesian trees corresponding to hierarchical classifications

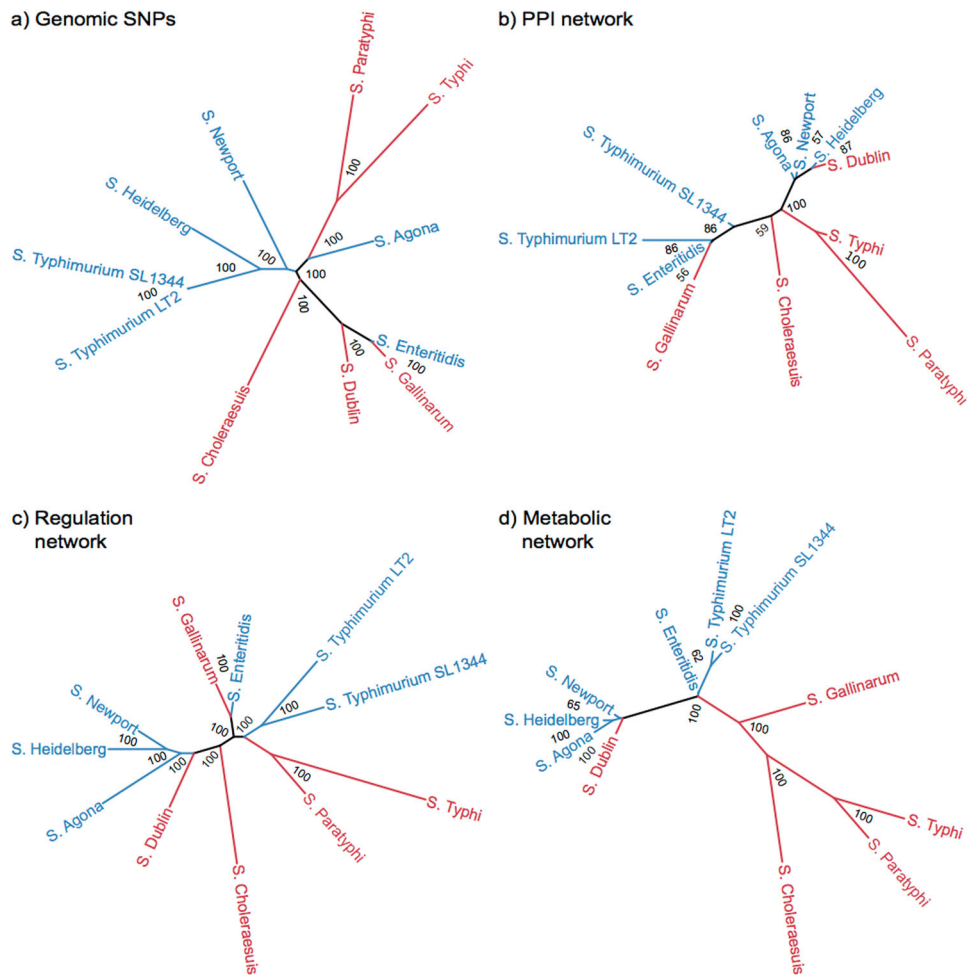


Fig. 2 Genome-based phylogenetic tree and hierarchical classification of networks. To distinguish different serovar types, gastro-intestinal serovars were colored to blue and extra-intestinals to red. Posterior probability values (as percentages) are shown on each node. **a** Bayesian phylogenetic tree based on the polymorphic sites of all common genes. **b-d** Hierarchical classification trees based on the matrix representation of protein-protein interaction networks **b**, regulatory networks **c**, and metabolic networks **d**. We note that four strains (Heidelberg, Agona, Newport and Dublin) form a cluster in all the three network based dendrograms due to technical reasons (see details in the main text)

(Fig. 2b–d). The dendrograms for each network were in all cases well established with nearly all posterior probability percentages at the nodes greater than 85. However, none of the networks resulted in the clustering of the extra-intestinal and gastro-intestinal strains. The hierarchical classification based on the metabolic networks separate the two pathovar types most pronouncedly, with only the *S. Dublin* metabolic network exhibiting greater similarity to gastro-intestinal pathovars than extra-intestinal pathovars. This is consistent with the loss of shared metabolic pathways that can be dispensed with by all extra-intestinal pathovars that have in common the loss of intestinal colonization as a mode of pathogenesis. The loss of metabolic pathways associated with the intestinal phase of infection is relatively strongly indicated. There is no evidence for loss of PPIs and regulatory sub-networks in the extra-intestinal pathovars. This could reflect the absence of changes in these networks in response to the dispensing of functions required specifically for the intestinal phase of infection or changes to these networks associated with adaptation to the extra-intestinal environment may be distinct in each extra-intestinal pathovar with weak or absent convergence of mechanisms.

We note that four strains (Heidelberg, Agona, Newport, and Dublin) form a cluster in all the three network based dendrograms (see details in the main text). This is most likely due to the absence

of particular genes having interactions to some key genes not present in the data sources used in our pipeline. As SalmoNet only contains genes with interaction data, if one of the interacting pair is missing, and the other interactor has no other interactions, SalmoNet does not contain that particular gene. For these four strains, this methodological limitation resulted in leaving out 31 genes, and because of that, these strains were clustered together.

Functional enrichment analysis of regulons point to pathovar specific transcription factor functions

Host adapted extra-intestinal pathovars are exposed to distinct host environments and conditions compared to the gastro-intestinal counterparts which are confined to the environment of the intestinal lumen and mucosa. Evolution to this alternative lifestyle was likely accompanied by plasticity in the regulation of functions in the extra-intestinal pathovars. Extra-intestinal pathovars mediating systemic infections are associated with increased severity and distinct pathogenicity.^{25–29} Enrichment analysis of the putative regulons revealed regulation of different functional processes in the two pathovar types by the same orthologous transcription factor (Fig. 3a–b). For instance, the virulence modulating regulator (CsgD) is known to control the expression of various pathogenicity related genes which are important for

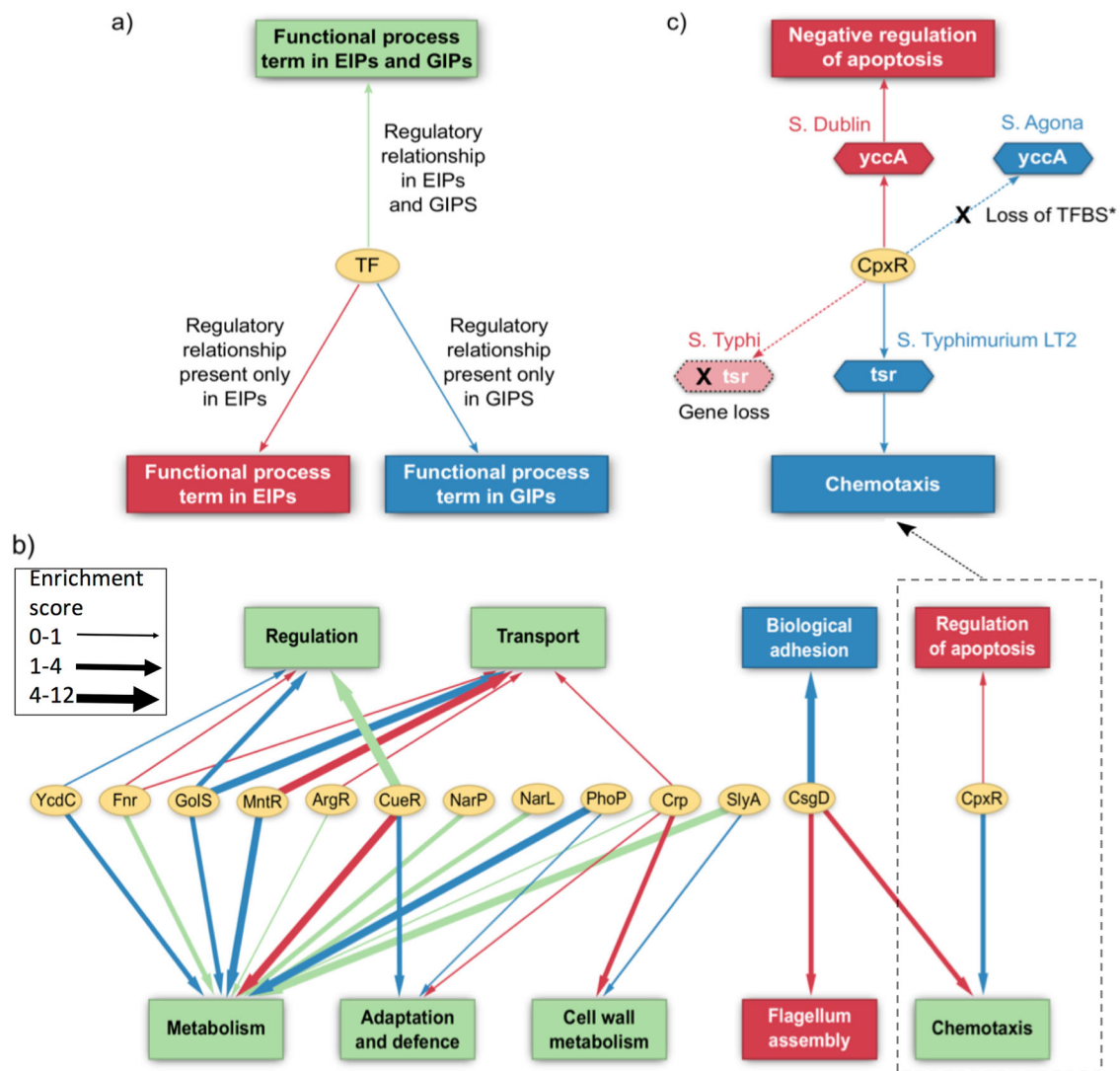


Fig. 3 Network of pathovar specific enriched functions and transcription factors. **a** Network legend for the figure. **b** Graphical representation of the functional processes predicted to be commonly and differentially modulated by orthologous transcription factors (TFs) in extra-intestinal and gastro-intestinal pathovars. **c** A specific example, enlarged from **b** demonstrating the loss in gastro-intestinal and extra-intestinal pathovars of regulatory relationships between cpxR and genes involved in the negative regulation of apoptosis and chemotaxis respectively. *TF* transcription factor, *TFBS* transcription factor binding site

virulence, persistence and biofilm formation.^{30,31} In our analysis, the set of genes putatively regulated by CsgD were found to be enriched with the specific functional process of 'Biological adhesion' in gastro-intestinal pathovars. However, the functional process of 'Flagellum assembly' was over-represented among the putative CsgD targets in extra-intestinal pathovars while 'Chemotaxis' was over-represented in both extra-intestinal and gastro-intestinal pathovars representing specific differences and commonalities in the role of CsgD between the two pathovars. Similarly, comparative analysis of the CpxR regulons revealed pathovar-specific enrichment of functions within the regulons. CpxR encodes a cognate response regulator and forms part of the CpxAR two component system involved in the sensing of and response to various cell envelope stresses and stimuli.^{32,33} In accordance with already known information that CpxR regulates motility and chemotaxis,³⁴ the set of putative targets of the CpxR regulon in gastro-intestinal pathovars was enriched with the functional process of chemotaxis. In the extra-intestinal pathovars, however, the functional process of regulation of apoptosis was found to be over-represented as a result of distinct cis-regulatory differences. For example, the gene encoding the YccA protein in

the extra-intestinal serovar harbored a CpxR binding site in its promoter region while the YccA ortholog in the gastro-intestinal serovar was observed to have a complete loss of the CpxR binding site due to truncation of its promoter (Supplementary Figure 2, Fig. 3c). YccA is homologous to the human anti-apoptotic protein BI-1 which represses the activity of the tumor suppressor protein Bax.³⁵ Due to the high conservation between *E. coli* YccA and the human BI-1, the YccA protein was even able to protect yeast cells against apoptosis induced by ectopically expressed human Bax protein³⁶ thus suggesting that YccA is associated with the modulation of host apoptosis. Other apoptotic related factors regulated by CpxR include genes or their orthologs encoding proteins such as the periplasmic serine endoprotease DegP/HtrA,³⁷ Hemolysin expression-modulating protein Hha,³⁸ and the toxicity modulator TomB with which Hha forms a putative toxin-antitoxin pair.³⁹ CpxR also modulates the expression of members of two distinct classes of proteins namely porins (such as OmpF) and chemotaxis related proteins (such as CheA, CheW) which are known to modulate apoptosis in the host cell upon infection.^{40,41}

Apoptosis of macrophages is a common host response once *Salmonella* has established an infection systemically but this is tightly regulated as the over-induction of apoptosis is detrimental to *Salmonella*.⁴² Hence, given that extra-intestinal pathovars cause systemic infections, it may be beneficial for the extra-intestinal pathovars to down-regulate apoptosis. This is one possible explanation for the over-representation of apoptosis regulation genes within the CpxR regulon in extra-intestinal pathovars and could indicate that CpxR plays a role in modulating apoptosis in extra-intestinal pathovars. The importance of CpxR in extra-intestinal pathovars is also demonstrated by studies which point out the use of CpxR as a safe and effective vaccine candidate against fowl typhoid caused by *Salmonella Gallinarum*, an extra-intestinal serovar. The results from the compositional analysis of the regulons indicate that the regulons of the two pathovars may have evolved to adapt to their respective pathogenic niches. The differences with respect to the functional processes regulated by *Salmonella* transcription factors could essentially be due to the expected adaptive evolution of extra-intestinal pathovars in contrast to the gastro-intestinal pathovars, which are mostly confined to the gut.

Applications of the database

The molecular interactions forming the biological interface between *Salmonella* and its host play a significant role in the colonization and infection process. *Salmonella* pathogenesis depends on its ability to adhere to host epithelial cells and the Type III secretion system mediated injection of effector proteins, which then cause the re-arrangement of the host cell cytoskeleton and internalization.^{43–46} *Salmonella* resides, survives and multiplies in specialized membrane bound vacuoles.^{47,48} Various genes known to be involved in *Salmonella* virulence and pathogenesis have been implicated in the interactions of *Salmonella* with the host cells.^{43–46} From the regulatory networks in SalmoNet, the transcription factors, which potentially regulate the virulence genes whose products are involved in the interactions with the host, can be identified. Moreover, by merging the regulatory networks with an increasing number of datasets such as the PPIs between *Salmonella* and the human host,⁴⁹ it is expected to enhance our understanding of the increasing number of mechanisms employed by *Salmonella* to infect and survive inside host cells.

Although the PPI and metabolic networks were inferred by orthological extrapolation, the original sources of data from which the extrapolation was performed are reliable due to their experimental basis even though the source data for the PPI networks were derived from high-throughput techniques. Hence, given the lack of PPI data for these *Salmonella* strains in this study, the inferred PPI and metabolic networks can be considered as a primary starting point for hypothesizing and uncovering new mechanisms. The PPIs are very interesting in this regard since previous studies have shown that *Salmonella* modulates many post-translational modifications such as ubiquitination of host proteins in order to avoid host responses such as autophagy.⁵⁰ The multi-layered nature of the SalmoNet resource can also be exploited in order to uncover potential biological insights by which *Salmonella* subverts host mechanisms. Recent evidence points to the modulation of metabolism (both of the *Salmonella* and the host) as yet another mechanism employed by *Salmonella* to acquire nutrients, evade host defense and survive under harsh intracellular conditions.^{51–59}

An integrative analysis of the regulatory and metabolic networks has the potential to reveal new insights into the transcriptional regulatory modulation of metabolic enzymes and could identify new metabolic drug targets as an intervention strategy. Integrating the PPI and regulatory networks not only provides a combined view of signal transduction and gene

regulation but also help us to shortlist upstream regulators of genes involved in establishing infection and metabolic functions. The activities of such individual regulators and transcription factors can be taken up for testing and screened for inhibition by small molecules/antibiotics.⁶⁰ Besides, SalmoNet also provides information on binding sites which can be used to design transcription factor decoys (anti-sense nucleotides of the transcription factor binding motif⁶¹ that prevent the binding of transcription factors to their targets. Anti-sense oligonucleotides have been used to modulate gene expression in a wide variety of intracellular bacterial pathogens such as *E. coli*,^{62,63} *Salmonella Typhimurium*,⁶² *Enterococcus faecalis*,⁶⁴ and *Klebsiella pneumoniae*.⁶⁵ Clinical trials have also been performed using anti-sense oligonucleotides for the treatment of human diseases^{66,67} thus suggesting that the potential of using anti-sense oligonucleotides against bacterial infections looks promising. In addition, the proteins and enzymes involved in critical PPIs and metabolic reactions respectively can also be subjected to the classical or systems-based drug-discovery pipelines.⁶⁸ The multi-layered network of SalmoNet allows discovering new molecular targets and strategies for therapeutic or prophylactic interventions based on the emergent properties of the networks. Modern drug and target discovery pipelines^{69,70} advocate a systemic approach, which relies on the integration of various heterogeneous data such as expression profiles and other multiple prior knowledge networks. SalmoNet satisfies this need by providing the prior knowledge networks for various strains of *Salmonella*.

As a source of both experimental and predicted interactions, SalmoNet contains information on the transcriptional regulatory targets of various transcription factors based on genome-wide motif scans. In addition, predicted targets of recently characterized transcription factors, such as RtsB,⁷⁰ which regulates the expression of invasion and flagellar genes, are also provided for future experimental verification and validation. Similar experimental testing and verification can also be performed on the predicted PPIs as well given the importance of PPIs in the survival of *Salmonella* inside the host cells.

From an epidemiological perspective, information on networks of multiple strains and strain-specific interactions further enriches *Salmonella* epidemiology studies. Traditional epidemiological approaches rely on tracking genetic polymorphisms and loss/gain of virulence genes specific to certain contexts and conditions. Hence, interaction networks could help to evaluate the effects of genetic polymorphisms in a systematic way, and thus, help in filling the gap between observed phenotypes of mutated strains and their genotypes. For example, SalmoNet can be used to further investigate the effect of cis-regulatory mutations on interactions, as well as the network level properties which determine the virulence characteristics of different strains of *Salmonella*.

Benchmarking *Salmonella* network data

There is no single resource storing *Salmonella*-specific protein-protein interactions (PPIs) but they are captured in general databases such as STRING⁷¹ and IntAct.⁷² In STRING, PPIs are based on different types of data such as genomic context, co-occurrence, co-expression, data derived text-mining and imported data from IntAct and other PPI resources. STRING contains only 237 experimentally verified interactions in addition to 1014634 predicted interactions based on criteria such as neighborhood, gene-fusion, gene-co-occurrence, co-expression, and text mining among *Salmonella* proteins. In IntAct, which contains manually curated and also imported PPI data from other databases, there are only 31 PPIs for *Salmonella* proteins.

In the case of transcriptional regulatory networks, RegulonDB¹⁷ stands out as one of the most comprehensive repositories for prokaryotic gene regulation. However, RegulonDB is restricted to

E. coli. RegPrecise⁷³ contains information for multiple bacterial species using genome-wide predictions based on manually curated reconstruction of regulons (which are set of genes whose expression is regulated by a transcription factor). Unfortunately, RegPrecise does not provide the original source of data used for the predictions, making further application of the data difficult. While well-known sources such as ORegAnno¹⁵ and PAZAR¹⁶ also capture regulatory interactions for multiple species, they do not contain any interactions for *Salmonella*.

As for the metabolic networks, there are numerous resources such as KEGG,¹⁸ MetaCyc/BioCyc⁷⁴ and the BioModels databases⁷⁵ containing seemingly *Salmonella* specific metabolic reactions. However, these databases are either not curated systematically or are not based on experimental results. KEGG for example contains information on pathway reactions and their entities for a large number of *Salmonella* strains but the *Salmonella* pathway annotations are based on computational predictions and not on experimental data. Similarly, coliBASE⁷⁶ captures comparative genomic data in terms of whole genome alignments and ortholog gene lists for *Salmonella*. Further information on bacterial metabolism can be found in specialized databases such as PATRIC for pathogens⁷⁷ or TRACTOR DB for Gamma proteobacteria.⁷⁸ However, most of the above mentioned resources contain limited interaction information for *Salmonella* and do not enable researchers for comparative network analysis or systems biological modeling of processes other than metabolism (e.g., they do not provide regulators of metabolic processes).

CONCLUSION

We present the first public biological network resource for *Salmonella* research. SalmoNet contains network data (metabolic, regulatory, protein-protein interaction) for 10 representative pathogenic *Salmonella* strains. To elucidate the virulence program of *Salmonella* for either discovering knowledge on biological mechanisms or for therapeutic interventions, it is rewarding to integrate the different network layers that capture emergent properties of the system. SalmoNet represents a resource which contains information on interactions from multiple layers of biological organization that can be analyzed as such, or as a topological backbone to be integrated with new -omic datasets for analyzing the dynamics. SalmoNet opens the possibility for systems-level studies of the pathogen *Salmonella* with unprecedented details in a standardized and well documented format. The resource can be browsed and downloaded as a whole or in user-defined interaction sets at <http://salmonet.org>. The analysis of SalmoNet could go far beyond fundamental biological and systems biology research. SalmoNet can be applied by medical microbiologists and epidemiologists to understand the strain specific differences of *Salmonella* and can serve as a starting point for further experimental investigations and systems medicine based drug discovery.

METHODS

Strains and orthology

Five strains of serovars with a predominantly gastro-intestinal lifestyle of *Salmonella* and five strains of serovars of extra-intestinal lifestyle were selected (Table 2). We included *Salmonella enterica* subsp. *enterica* serovar *Typhimurium* str. SL1344 as a sixth gastro-intestinal strain since it is widely used as a reference strain. We determined orthologous proteins among the selected strains, as well as for the model organism *E. coli* K12 with InParanoid.⁷⁹ We used a reciprocal best hit approach using BLAST to identify homologous protein sequences including those from plasmids corresponding to the selected strains. The protein sequences were downloaded from UniProt⁸⁰ as of January 2015. The results from

the comparison of proteins one by one among any pair of strains were used to derive the ortholog clusters. Sequence similarity was set at $\geq 95\%$ in order to minimize false positives given that the chosen strains belonged to the same species. Clustered groups contained both paralogs and orthologs (Supplementary Table 3). We removed the pseudo-genes listed in ⁷ from the resultant ortholog list.

Reconstruction of networks

We developed metabolic, regulatory and protein-protein interaction networks for *Salmonella*, using complementary approaches followed by merging the three layers into a unified *Salmonella* network. We performed this process for the 10 strains separately that resulted in 10 strain specific molecular networks.

Metabolic networks: We defined the metabolic network as follows: if a metabolite is a product of a reaction and substrate in another, the two proteins catalysing the different reactions were linked, as described in ref. 81 We did not consider the links for metabolites appearing in more than 10 reactions as outlined in ref. 81

We retrieved the metabolic reactions from two different sources with different levels of curation: the manually curated metabolic model of *S. Typhimurium* LT2 (referred to as STM),⁸² and predictions from the BioModels database⁷⁵ containing Enzyme Commission (EC) numbers. In the latter, EC number assignments are automatic and include predictions for enzymes present in *Salmonella* spp and not necessarily present in *E. coli*. The STM model was derived from an *E. coli* metabolic model and confirmed by flux balance analysis. For the extrapolation of metabolic reactions from the above mentioned models, we assumed that the same reactions occur in the *Salmonella* strains when orthologous protein(s) of the enzyme(s) involved in the reactions were found to be present in the corresponding strains.

Regulatory networks: Regulatory interactions represent the binding of transcription factors to gene promoters. They consist of both predicted and experimentally verified interactions in our study. We collected experimentally verified DNA-binding sites of *Salmonella* transcription factors (Supplementary Table 4) from the literature with manual curation, as well as information from publicly available datasets. For the high-throughput datasets retrieved from,⁸³ peaks were identified as described elsewhere.^{84,85} For all the other high-throughput datasets, the processed data (transcription factors, their targets and corresponding binding motifs when available) was retrieved from the cited sources (Supplementary Table 4). We then constructed Position Specific Scoring Matrices (PSSMs) from the manually inferred binding sites and sites corresponding to the significant consensus motifs from the low- and high-throughput datasets respectively using the *consensus* tool⁸⁶ with default parameters. PSSMs constructed from too few binding sites have low predictive values. Hence, in instances where the number of binding sites (according to published data) corresponding to a transcription factor were less than three, we used corresponding sites from orthologous targets present in one or more of the other *Salmonella* strains under study for the PSSM construction. Since the predictive capacity and information content varies among PSSMs, we determined specific optimal P-value thresholds for every PSSM-strain combination using the *matrix-quality* tool²¹ (Supplementary Table 5). We used the TRANSFAC-formatted PSSMs via the *matrix-scan* tool⁸⁷ to scan the promoter regions of all the genes from the genomes of the selected *Salmonella* strains. We confined the binding site search to 5000 bp upstream of the start codon of every protein coding gene to capture functionally active transcription factor binding sites in genomic regions including intergenic sequences between convergent genes.⁸⁸ However, sequences which overlap with upstream coding sequences were excluded. The promoters were retrieved

using the “retrieve sequence” function of the RSAT tool suite. For the background model, we used a Markov order of 1, and the model was estimated individually for every strain. Both the strands of the genome were scanned for the presence of putative binding sites and the optimal P-value determined for every TF-strain combination as described previously was used during the corresponding scans. Putative hits with a P-value lower than the corresponding optimal cut-off values were considered to be significant. Based on the principle of “regulogs”,^{89,90} we also inferred transcription factor-target gene relationships in *Salmonella* strains. Regulogs are regulatory interactions first detected in one species (in this case in *E. coli*) and then predicted to be potentially present in another species (in this case in *Salmonella*) based on sequence homology of the transcription factor, the target gene and the transcription factor binding site. Accordingly, we used the *E. coli* transcription factor - target gene binding site information retrieved from RegulonDB¹⁷ in conjunction with the homolog clustering results to extrapolate regulatory interactions from *E. coli* to the *Salmonella* strains. Operon information was retrieved from DOOR.⁹¹ The workflow is presented in Fig. 1.

PPI networks: We performed manual curation to retrieve existing PPI information in *Salmonella* spp from the literature using a curation protocol we developed for the Signalink eukaryotic signaling network resource as previously described.^{92,93} Briefly, we collected signaling interactions involving *Salmonella* proteins from primary research articles identified by using iHOP⁹⁴ and ChilliBot⁹⁵ tools in addition to those articles directly found in PubMed searches. The main text and the abstracts of these articles were examined to retrieve the interactions between *Salmonella* proteins. Experimentally verified *Salmonella* PPIs were retrieved from IntAct.⁷² Proteome-wide predictions to predict PPIs were carried out using 3D protein-based structure predictions of Interactome 3D⁹⁶ and using *E. coli* PPIs for interolog predictions. The interologs were inferred based on *E. coli* PPIs retrieved from IntAct,⁷² BioGrid⁹⁷ and a high-throughput, yeast-2-hybrid screen of the *E. coli* interactome.⁹⁸

Phylogenetic tree construction

Gene sequences corresponding to the *Salmonella* strains considered in this study were downloaded using the *retrieve-sequence* tool from the Regulatory Sequence Analysis Tools.⁸⁷ Out of the 2912 common ortholog gene sets from the strains in this study, 457 ortholog sets were discarded due to discrepancies such as misconverted locus tags/IDs. Ortholog genes that had different lengths across strains were aligned by using ClustalOmega⁹⁹ implemented in the *msa* Bioconductor R package.¹⁰⁰ We identified 85 ortholog gene sets where one or more strains had more than one sequences (due to gene duplication or misannotation). We discarded these extra sequences after manual curation and the sequences that were more similar to the sequences of other strains were retained.

MrBayes v3.2.4¹⁰¹—which is a program for Bayesian inference based selection of evolutionary models—was used to analyze the phylogenetic relationships of the strains using the polymorphic sites from genes in the orthologous gene sets. The parameters of the evolutionary model between the sequence sites were unlinked. Gaps were not considered as polymorphisms since the applied phylogenetic software treated them as missing data. Thus, gaps did not contribute to the phylogenetic information. Ortholog groups whose ratio of polymorphic sites to gene lengths was more than 0.1 (100 genes) were discarded and consequently polymorphic sites from potentially *false* orthologs that had low sequence similarity were excluded. After applying the above filter, 64,531 polymorphic sites from 2360 orthologous genes were used to infer a genome based phylogenetic tree. Metropolis coupling Markov Chain Monte Carlo (MC³) analysis was performed for 10

million generations and 25% of the samples from the beginning of the chain were discarded when applying MrBayes.

Network based dendrograms

The networks (Supplementary Table 6) were represented by interaction matrices containing binary data, where “1”-s represented the presence of an interaction between the same pair of nodes inferred by orthology across the strains and “0”-s stood for missing interactions (Supplementary Table 6). In order to represent the hierarchical classification of strains from network data, we constructed dendrograms based on the metabolic, regulatory and PPI interaction matrices using MrBayes v3.2.4. To calculate network based dendrograms, the same MrBayes MC³ analyses were performed as for the genome-based tree except that the datatype was set to “restriction” and no substitution model was applied.

Functional analysis of the transcriptional regulatory network

In order to understand the biological context within the regulons of the two serovars, functional enrichment analysis was performed to infer the over-represented Gene Ontology (GO) Biological Process Terms within the predicted regulons. Here, we considered only those interactions that were predicted to occur in at least two of the ten studied strains. This was performed to minimize the chances of including possible false positives in our analyses. In addition, we considered only the predicted regulons and GO terms containing at least 10 entities across all the strains within each pathovar. The background set comprised the entire collection of genes with annotated GO terms in the genomes. To determine the enriched GO Biological Process terms, the hypergeometric test with the Bonferroni correction was applied. The significance score for each enrichment event was calculated as the $-\log_{10}$ function of the corrected P-value. Enriched terms with a significance score greater than zero were considered as significant. We retrieved TF-GO relationships, which were exclusive to either of the two serovars. We restricted the analysis to TFs that were predicted to contain different enriched GO Biological Process terms within their putative regulons in extra-intestinal and gastro-intestinal pathovars. We also performed a manual assignment of functional processes derived from the Gene Ontology database for every GO term identified in the previous step. Subsequently, we replaced GO terms with their corresponding functional process(es) in order to simplify the graph without losing information.

Data availability statement

The datasets generated in the study are freely available at <http://salmonet.org/>. The source data as well as the generated datasets are provided as Supplementary tables which are freely available via *NPJ Systems Biology and Applications website*. The tools and resources such as the RSAT suite, Mr. Bayes, InParanoid, iHOP, ChilliBot, ClustaOmega, RegulonDB, Interactome 3D, IntAct, BioGrid, and UniProt which are used in this study are publicly available. Custom codes used in the study are available upon request.

ACKNOWLEDGEMENTS

The authors are grateful for the helpful discussions to the members and visitors of the Baranyi, Korcsmaros, and Kingsley groups, as well as for the gap-filling ChIP-Seq and RNA-seq datasets provided by Joseph Wade (Wadsworth Center, USA). This work was supported by a fellowship to T.K. in computational biology at the Earlham Institute (Norwich, UK) in partnership with the Quadram Institute (Norwich, UK), and strategically supported by the Biotechnological and Biosciences Research Council, UK grants (BB/J004529/1, BB/P016774/1 and BB/CSP17270/1).

AUTHOR CONTRIBUTIONS

A.M. contributed to the design of the work and drafting the manuscript. P.S. carried out the network reconstructions and drafted and revised the manuscript. D.F. contributed to the network reconstructions and set up the web-resource. A.D. performed the curation and testing of the website. E.A. was involved in inferring the classification trees and dendrograms. M.O. contributed to internal testing and quality control. P.B. and R.K. contributed to framing the biological basis of the work and the discussions in the manuscript. J.B. and T.K. conceived and supervised the entire study. All the authors read and approved the final version of the manuscript.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Systems Biology and Applications* website (<https://doi.org/10.1038/s41540-017-0034-z>).

Competing interests: The authors declare no competing financial interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

- Sánchez-Vargas, F. M., Abu-El-Haija, M. A. & Gómez-Duarte, O. G. *Salmonella* infections: an update on epidemiology, management, and prevention. *Travel Med. Infect. Dis.* **9**, 263–277 (2011).
- Majowicz, S. E. et al. The global burden of nontyphoidal *Salmonella gastroenteritis*. *Clin. Infect. Dis.* **50**, 882–889 (2010).
- GBD. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1459–1544 (2016).
- Mills, D. M., Bajaj, V. & Lee, C. A. A 40 kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. *Mol. Microbiol.* **15**, 749–759 (1995).
- Aleksic, S., Heinzerling, F. & Bockemuhl, J. Human infection caused by *Salmonella* of subspecies II to VI in Germany, 1977–1992. *Zent. Bakteriol.* **283**, 391–398 (1996).
- Parkhill, J. et al. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848–852 (2001).
- Nuccio, S.-P. & Bäuml, A. J. Comparative analysis of *Salmonella* genomes identifies a metabolic network for escalating growth in the inflamed gut. *MBio* **5**, e00929–14 (2014).
- Haghjoo, E. & Galán, J. E. *Salmonella typhi* encodes a functional cytolethal distending toxin that is delivered into host cells by a bacterial-internalization pathway. *Proc. Natl. Acad. Sci. USA* **101**, 4614–4619 (2004).
- Wilson, J. W. et al. Space flight alters bacterial gene expression and virulence and reveals a role for global regulator Hfq. *Proc. Natl. Acad. Sci. USA* **104**, 16299–16304 (2007).
- Winter, S. E., Raffatellu, M., Wilson, R. P., Rüssmann, H. & Bäuml, A. J. The *Salmonella enterica* serotype Typhi regulator TviA reduces interleukin-8 production in intestinal epithelial cells by repressing flagellin secretion. *Cell. Microbiol.* **10**, 247–261 (2008).
- Song, J., Gao, X. & Galán, J. E. Structure and function of the *Salmonella* Typhi chimaeric A(2)B(5) typhoid toxin. *Nature* **499**, 350–354 (2013).
- Klemm, E. J. et al. Emergence of host-adapted *Salmonella* Enteritidis through rapid evolution in an immunocompromised host. *Nat. Microbiol.* **1**, 15023 (2016).
- Desai, P. T. et al. Evolutionary genomics of *Salmonella enterica* subspecies. *MBio* **4**, e00579–12 (2013).
- Gonçalves, E. et al. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Mol. Biosyst.* **9**, 1576–1583 (2013).
- Griffith, O. L. et al. ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res.* **36**, D107–D113 (2008).
- Portales-Casamar, E. et al. The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. *Nucleic Acids Res.* **37**, D54–D60 (2009).
- Gama-Castro, S. et al. RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res.* **44**, D133–D143 (2016).
- Kanehisa, M. et al. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42**, D199–D205 (2014).
- Karimpour-Fard, A., Detweiler, C. S., Erickson, K. D., Hunter, L. & Gill, R. T. Cross-species cluster co-conservation: a new method for generating protein interaction networks. *Genome Biol.* **8**, R185 (2007).
- Shou, C. et al. Measuring the evolutionary rewiring of biological networks. *PLoS Comput. Biol.* **7**, e1001050 (2011).
- Medina-Rivera, A. et al. Theoretical and empirical quality assessment of transcription factor-binding motifs. *Nucleic Acids Res.* **39**, 808–824 (2011).
- Walthers, D. et al. The response regulator SsrB activates expression of diverse *Salmonella* pathogenicity island 2 promoters and counters silencing by the nucleoid-associated protein H-NS. *Mol. Microbiol.* **65**, 477–493 (2007).
- Tomljenovic-Berube, A. M., Mulder, D. T., Whiteside, M. D., Brinkman, F. S. L. & Coombes, B. K. Identification of the regulatory logic controlling *Salmonella* pathoadaptation by the SsrA-SsrB two-component system. *PLoS Genet.* **6**, e1000875 (2010).
- Timme, R. E. et al. Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. *enterica* inferred from genome-wide reference-free SNP characters. *Genome Biol. Evol.* **5**, 2109–2123 (2013).
- Chiu, C.-H. et al. The genome sequence of *Salmonella enterica* serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. *Nucleic Acids Res.* **33**, 1690–1698 (2005).
- Abbott, S. L., Ni, F. C. Y. & Janda, J. M. Increase in extraintestinal infections caused by *Salmonella enterica* subspecies II-IV. *Emerg. Infect. Dis.* **18**, 637–639 (2012).
- Wilkins, E. G. & Roberts, C. Extraintestinal salmonellosis. *Epidemiol. Infect.* **100**, 361–368 (1988).
- Chen, P. L. et al. Extraintestinal focal infections in adults with nontyphoid *Salmonella* bacteraemia: predisposing factors and clinical outcome. *J. Intern. Med.* **261**, 91–100 (2007).
- Huang, D. B. & DuPont, H. L. Problem pathogens: extra-intestinal complications of *Salmonella enterica* serotype Typhi infection. *Lancet Infect. Dis.* **5**, 341–348 (2005).
- MacKenzie, K. D. et al. Bistable expression of CsgD in *Salmonella enterica* serovar Typhimurium connects virulence to persistence. *Infect. Immun.* **83**, 2312–2326 (2015).
- Zakikhany, K., Harrington, C. R., Nimtz, M., Hinton, J. C. D. & Römling, U. Unphosphorylated CsgD controls biofilm formation in *Salmonella enterica* serovar Typhimurium. *Mol. Microbiol.* **77**, 771–786 (2010).
- Raivio, T. L. & Silhavy, T. J. Transduction of envelope stress in *Escherichia coli* by the Cpx two-component system. *J. Bacteriol.* **179**, 7724–7733 (1997).
- Pogliano, J., Lynch, A. S., Belin, D., Lin, E. C. & Beckwith, J. Regulation of *Escherichia coli* cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. *Genes Dev.* **11**, 1169–1182 (1997).
- Wolfe, A. J., Parikh, N., Lima, B. P. & Zemaitaitis, B. Signal integration by the two-component signal transduction response regulator CpxR. *J. Bacteriol.* **190**, 2314–2322 (2008).
- Xu, Q. & Reed, J. C. Bax inhibitor-1, a mammalian apoptosis suppressor identified by functional screening in yeast. *Mol. Cell.* **1**, 337–346 (1998).
- Chae, H.-J. et al. Evolutionarily conserved cytoprotection provided by Bax Inhibitor-1 homologs from animals, plants, and yeast. *Gene* **323**, 101–113 (2003).
- Hegde, R. et al. Identification of Omi/HtrA2 as a mitochondrial apoptotic serine protease that disrupts inhibitor of apoptosis protein-caspase interaction. *J. Biol. Chem.* **277**, 432–438 (2002).
- Hong, S. H., Lee, J. & Wood, T. K. Engineering global regulator Hha of *Escherichia coli* to control biofilm dispersal. *Microb. Biotechnol.* **3**, 717–728 (2010).
- García-Contreras, R., Zhang, X.-S., Kim, Y. & Wood, T. K. Protein translation and cell death: the role of rare tRNAs in biofilm formation and in activating dormant phage killer genes. *PLoS One* **3**, e2394 (2008).
- Gorga, F., Galdiero, M., Buommino, E. & Galdiero, E. Porins and lipopolysaccharide induce apoptosis in human spermatozoa. *Clin. Diagn. Lab. Immunol.* **8**, 206–208 (2001).
- Rolig, A. S., Carter, J. E. & Ottemann, K. M. Bacterial chemotaxis modulates host cell apoptosis to establish a T-helper cell, type 17 (Th17)-dominant immune response in *Helicobacter pylori* infection. *Proc. Natl. Acad. Sci. USA* **108**, 19749–19754 (2011).
- Takaya, A. et al. Derepression of *Salmonella* pathogenicity island 1 genes within macrophages leads to rapid apoptosis via caspase-1- and caspase-3-dependent pathways. *Cell. Microbiol.* **7**, 79–90 (2005).
- Lara-Tejero, M. & Galán, J. E. *Salmonella enterica* serovar Typhimurium pathogenicity island 1-encoded type III secretion system translocases mediate intimate attachment to nonphagocytic cells. *Infect. Immun.* **77**, 2635–2642 (2009).
- Galán, J. E. *Salmonella* interactions with host cells: type III secretion at work. *Annu. Rev. Cell. Dev. Biol.* **17**, 53–86 (2001).
- Kaur, J. & Jain, S. K. Role of antigens and virulence factors of *Salmonella enterica* serovar Typhi in its pathogenesis. *Microbiol. Res.* **167**, 199–210 (2012).
- Zhang, S. et al. Molecular pathogenesis of *Salmonella enterica* serotype typhimurium-induced diarrhea. *Infect. Immun.* **71**, 1–12 (2003).
- Bakowski, M. A., Braun, V. & Brumell, J. H. *Salmonella*-containing vacuoles: directing traffic and nesting to grow. *Traffic* **9**, 2022–2031 (2008).

48. Steele-Mortimer, O. The Salmonella-containing vacuole: moving with the times. *Curr. Opin. Microbiol.* **11**, 38–45 (2008).
49. Schlekler, S. et al. The current Salmonella-host interactome. *Proteom. Clin. Appl.* **6**, 117–133 (2012).
50. Rytönen, A. & Holden, D. W. Bacterial interference of ubiquitination and deubiquitination. *Cell. Host. Microbe* **1**, 13–22 (2007).
51. Kim, J. S. et al. Molecular characterization of the InvE regulator in the secretion of type III secretion translocases in *Salmonella enterica* serovar Typhimurium. *Microbiology* **159**, 446–461 (2013).
52. Wynosky-Dolfi, M. A. et al. Oxidative metabolism enables Salmonella evasion of the NLRP3 inflammasome. *J. Exp. Med.* **211**, 653–668 (2014).
53. Antunes, L. C. M. et al. Impact of salmonella infection on host hormone metabolism revealed by metabolomics. *Infect. Immun.* **79**, 1759–1769 (2011).
54. Hernandez, L. D., Hueffer, K., Wenk, M. R. & Galán, J. E. Salmonella modulates vesicular traffic by altering phosphoinositide metabolism. *Science* **304**, 1805–1807 (2004).
55. Dandekar, T. et al. Salmonella-how a metabolic generalist adopts an intracellular lifestyle during infection. *Front. Cell. Infect. Microbiol.* **4**, 191 (2014).
56. DeRubertis, F. R. & Woebber, K. A. Accelerated cellular uptake and metabolism of L-thyroxine during acute *Salmonella typhimurium* sepsis. *J. Clin. Invest.* **52**, 78–87 (1973).
57. Arsenault, R. J., Napper, S. & Kogut, M. H. *Salmonella enterica* Typhimurium infection causes metabolic changes in chicken muscle involving AMPK, fatty acid and insulin/mTOR signaling. *Vet. Res.* **44**, 35 (2013).
58. Mazé, A., Glatzer, T. & Bumann, D. The central metabolism regulator EIAGlc switches Salmonella from growth arrest to acute virulence through activation of virulence factor secretion. *Cell Rep.* **7**, 1426–1433 (2014).
59. Herzberg, M., Jawad, M. J. & Pratt, D. Succinate metabolism and virulence in *Salmonella typhimurium*. *Nature* **204**, 1285–1286 (1964).
60. Berg, T. Inhibition of transcription factors with small organic molecules. *Curr. Opin. Chem. Biol.* **12**, 464–471 (2008).
61. Mann, M. J. Transcription factor decoys: a new model for disease intervention. *Ann. NY Acad. Sci.* **1058**, 128–139 (2005).
62. McKinney, J. S., Zhang, H., Kubori, T., Galán, J. E. & Altman, S. Disruption of type III secretion in *Salmonella enterica* serovar Typhimurium by external guide sequences. *Nucleic Acids Res.* **32**, 848–854 (2004).
63. Tilley, L. D. et al. Gene-specific effects of antisense phosphorodiamidate morpholino oligomer-peptide conjugates on *Escherichia coli* and *Salmonella enterica* serovar typhimurium in pure culture and in tissue culture. *Antimicrob. Agents Chemother.* **50**, 2789–2796 (2006).
64. Shen, N. et al. Inactivation of expression of several genes in a variety of bacterial species by EGS technology. *Proc. Natl Acad. Sci. USA* **106**, 8163–8168 (2009).
65. Kurupati, P., Tan, K. S. W., Kumarasinghe, G. & Poh, C. L. Inhibition of gene expression and growth by antisense peptide nucleic acids in a multiresistant beta-lactamase-producing *Klebsiella pneumoniae* strain. *Antimicrob. Agents Chemother.* **51**, 805–811 (2007).
66. Sharma, V. K., Sharma, R. K. & Singh, S. K. Antisense oligonucleotides: modifications and clinical trials. *Med. Chem. Commun.* **5**, 1454–1471 (2014).
67. Koo, T. & Wood, M. J. Clinical trials using antisense oligonucleotides in duchenne muscular dystrophy. *Hum. Gene Ther.* **24**, 479–488 (2013).
68. Wang, R.-S., Maron, B. A. & Loscalzo, J. Systems medicine: evolution of systems biology from bench to bedside. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **7**, 141–161 (2015).
69. Butcher, E. C., Berg, E. L. & Kunkel, E. J. Systems biology in drug discovery. *Nat. Biotechnol.* **22**, 1253–1259 (2004).
70. Ellermeier, C. D. & Slauch, J. M. RtsA and RtsB coordinately regulate expression of the invasion and flagellar genes in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* **185**, 5096–5108 (2003).
71. Szklarczyk, D. et al. STRINGv10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
72. Orchard, S. et al. The MintAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358–D363 (2014).
73. Novichkov, P. S. et al. RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC. Genom.* **14**, 745 (2013).
74. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **42**, D459–D471 (2014).
75. Juty, N. et al. BioModels: Content, Features, Functionality, and Use. *CPT Pharmacomet. Syst. Pharmacol.* **4**, e3 (2015).
76. Chaudhuri, R. R., Khan, A. M. & Pallen, M. J. colIBASE: an online database for *Escherichia coli*, *Shigella* and *Salmonella* comparative genomics. *Nucleic Acids Res.* **32**, D296–D299 (2004).
77. Wattam, A. R. et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42**, D581–D591 (2014).
78. González, A. D., Espinosa, V., Vasconcelos, A. T., Pérez-Rueda, E. & Collado-Vides, J. TRACTOR_DB: a database of regulatory networks in gamma-proteobacterial genomes. *Nucleic Acids Res.* **33**, D98–D102 (2005).
79. Sonnhammer, E. L. L. & Östlund, G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* **43**, D234–D239 (2015).
80. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
81. Kreimer, A., Borenstein, E., Gophna, U. & Ruppín, E. The evolution of modularity in bacterial metabolic networks. *Proc. Natl. Acad. Sci. USA* **105**, 6976–6981 (2008).
82. Thiele, I. et al. A community effort towards a knowledge-base and mathematical model of the human pathogen *Salmonella typhimurium* LT2. *BMC. Syst. Biol.* **5**, 8 (2011).
83. Smith, C., Stringer, A. M., Mao, C., Palumbo, M. J. & Wade, J. T. Mapping the regulatory network for *Salmonella enterica* serovar Typhimurium invasion. *MBio* **7**, e01024–e01026 (2016).
84. Fitzgerald, D. M., Bonocora, R. P. & Wade, J. T. Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. *PLoS Genet.* **10**, e1004649 (2014).
85. Singh, S. S. et al. Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev.* **28**, 214–219 (2014).
86. Thomas-Chollier, M. et al. RSAT 2011: regulatory sequence analysis tools. *Nucleic Acids Res.* **39**, W86–W91 (2011).
87. Medina-Rivera, A. et al. RSAT 2015: regulatory sequence analysis tools. *Nucleic Acids Res.* **43**, W50–W56 (2015).
88. Haycocks, J. R. J. & Grainger, D. C. Unusually Situated Binding Sites for Bacterial Transcription Factors Can Have Hidden Functionality. *PLoS One* **11**, e0157016 (2016).
89. Rodionov, D. A. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem. Rev.* **107**, 3467–3497 (2007).
90. Yu, H. et al. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res.* **14**, 1107–1118 (2004).
91. Mao, F., Dam, P., Chou, J., Olman, V. & Xu, Y. DOOR: a database for prokaryotic operons. *Nucleic Acids Res.* **37**, D459–D463 (2009).
92. Korcsmáros, T. et al. Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery. *Bioinformatics* **26**, 2042–2050 (2010).
93. Fazekas, D. et al. SignaLink 2 - a signaling pathway resource with multi-layered regulatory networks. *BMC. Syst. Biol.* **7**, 7 (2013).
94. Hoffmann, R. Using the iHOP information resource to mine the biomedical literature on genes, proteins, and chemical compounds. *Curr. Protoc. Bioinform.* Chapter 1, Unit1.16 (2007). <https://doi.org/10.1002/0471250953.bi0116s20>
95. Chen, H. & Sharp, B. M. Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinforma.* **5**, 147 (2004).
96. Mosca, R., Céol, A. & Aloy, P. Interactome3D: adding structural details to protein networks. *Nat. Methods* **10**, 47–53 (2013).
97. Chatr-Aryamontri, A. et al. The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* **43**, D470–D478 (2015).
98. Rajagopala, S. V. et al. The binary protein-protein interaction landscape of *Escherichia coli*. *Nat. Biotechnol.* **32**, 285–290 (2014).
99. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
100. Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C. & Hochreiter, S. msa: an R package for multiple sequence alignment. *Bioinformatics* **31**, 3997–3999 (2015).
101. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017