REVIEW ARTICLE

A phylogenomic analysis of bacterial helix–turn–helix transcription factors

Catarina L. Santos[1,2], Fernando Tavares[1,2], Jean Thioulouse[3,4] & Philippe Normand[3,5]

[1]IBMC – Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal; [2]Departamento de Botânica, Faculdade de Ciências, Universidade do Porto, Porto, Portugal; [3]Université de Lyon, Lyon, France; [4]Laboratoire de Biométrie et Biologie Evolutive, CNRS, UMR5558, Université Lyon 1, Villeurbanne, France; and [5]Ecologie Microbienne, UMR CNRS 5557, Université Lyon I, Villeurbanne, France

Abstract

Perception by each individual organism of its environment's parameters is a key factor for survival. In a constantly changing environment, the ability to assess nutrient sources and potentially stressful situations constitutes the main basis for ecological adaptability. Transcription regulators are key decision-making proteins that mediate the communication between environmental conditions and DNA transcription through a multifaceted network. The parallel study of these regulators across microbial organisms adapted to contrasting biotopes constitutes an unexplored approach to understand the evolution of genome plasticity and cell function. We present here a reassessment of bacterial helix–turn–helix regulator diversity in different organisms from a multidisciplinary perspective, on the interface that links metabolism, ecology and phylogeny, further sustained by a statistically based approach. The present revision brought to light evidence of patterns among families of regulators, suggesting that multiple selective forces modulate the number and kind of regulators present in a given genome. Besides being an important step towards understanding the adaptive traits that influence the microbial responses to the varying environment on the very first and most prevalent line of reaction, the transcription of DNA, this approach is a promising tool to extract biological trends from genomic databases.

Introduction

With an astonishing capacity to survive and populate virtually every available ecological niche, the diversity of the microbial world is a fascinating living example of high evolutionary capacity. The bacterial genomes, small-sized and nowadays simple to sequence, together with their enormous diversity, are an excellent model to study both phylogenetic adaptation and ecological adaptability. Often confused in the literature, these two expressions hold completely different meanings in evolutionary biology: if adaptation stands for the natural selection of positive phenotypes that, in the long term, results in the modulation of the population genomic pool, adaptability describes the fitness of each organism, i.e. the capacity of that organism to cope with transient changes in the environment.

The amount of genomic data now available allows a deeper and more thorough understanding of the flexibility and functionality of genomes, providing the possibility to realize the full potential of the sequenced genes. The key challenge nowadays is the integration of the available sequences into a multidisciplinary approach, which requires the functional, ecological and phylogenetic positioning of the organisms (Roberts, 2004). In this review, we intend to highlight the clustering of regulatory protein families into distribution patterns among different prokaryotic species, considering their ecological adaptability to different habitats.

Adaptability is based on a set of factors that contribute to the final goal of survival through regulation, and includes modulation of transcription as one of the first lines of response. An exhaustive analysis of 145 prokaryotic genomes has unveiled the dominance of one-component systems as the main transcription factors (TFs) among prokaryotes: not only are they the precursors of the two-component systems, but they also present a larger diversity in terms of domains and a broader distribution among

bacteria and archaea (Ulrich *et al.*, 2005). Among the bacterial one-component TFs, up to 84% of the output domains comprise a DNA-binding helix–turn–helix (HTH) region (Ulrich *et al.*, 2005). So, although not restricted to transcription regulation, the HTH motif assumes a central role in this process (Aravind *et al.*, 2005). Briefly, an HTH motif consists of two α-helices forming an internal angle of around 120° and connected by a short turn of four residues; this turn can, however, be extended to as many as 21 amino acids in the *winged* turns (Rosinski & Atchley, 1999). The whole structure comprises 20 residues and the second helix, known as the 'recognition helix', is involved in the sequence-specific DNA interaction (for a review on HTH structure, see Kohn *et al.*, 1997, and cited references). Although there is a huge functional diversity among TFs carrying an HTH domain, these regulators seem nevertheless to have a mono-phyletic origin, having arisen from duplication followed by divergence and specialization (Rosinski & Atchley, 1999).

Small metabolites and covalent modifications translate the external or internal stimulus(i) into the response(s) by activating or deactivating the TFs, which, in turn, induce or repress the transcription of specific genes or operons. However, this is not a straightforward stimulus–regulator–DNA relationship; the transcriptional regulatory networks are composed of complex interactions that connect factors and effectors in a flexible and dense super network. Whereas the study of the regulators in a single genome is important to understand the intricate networks that drive the organism's flexibility in terms of habitat, nutrient sources and stress conditions, the parallel study of regulators across related genomes is no less valuable. On the contrary, it represents an important tool to understand the evolution of deep life-supporting mechanisms, such as transcription regulation. In fact, evidence has accumulated suggesting that microbial life migrated to become distributed across the earth's various biotopes, and only afterwards was selected on the basis of the specificities of the different environments (von Mering *et al.*, 2007, and references therein). Metaproteomic approaches have revealed differences between the nature of the protein complement extracted from different environments (Wilmes & Bond, 2006, and references therein), which suggests the presence of different regulatory pathways establishing a link between the background genomic content and the adaptation to specific environments. Furthermore, a recent work focused on evolutionary pathways suggests the optimization of protein expression as one important route towards adaptation (Babu & Aravind, 2006), and another recent study presents evidence that TFs evolve much faster than their target genes across phyla (Lozada-Chavez *et al.*, 2006). In fact, according to the cited authors, adaptive mutations affecting the transcription level are more frequent than structural ones, affecting genes involved in biochemical

pathways. All these examples suggest that the distribution of TFs along the different genomes should be the consequence of the selective pressures that have been acting upon the organism. In other words, the adaptability of a given organism, defined in terms of the number and type of regulators present in its genome, should somehow reflect the oscillations that the environment of the organism has undergone over geological time.

The vast amount of unexplored genomic data requires the establishment of artificial limits in phylogenomic approaches, as commonly seen in other studies (Kotelnikova *et al.*, 2005; Ulrich *et al.*, 2005). In this review, we present the functional diversity of 14 one-component bacterial HTH-regulators (hereafter referred to only as TFs or HTH regulators) summarized in Table 1, which cover all the main cellular functionalities and range of frequencies found in bacteria, and analyse their presence among 270 bacterial organisms. These, in turn, include all main phylogenetic groups and ecological niches. We intend to unveil the underlying distribution patterns among the different regulatory families, as well as relate them to the organism's specificities in terms of genome and ecological niche, aiming to integrate ecology, evolution and transcriptional regulation in *Eubacteria*. We believe that this approach merits extension to other protein families and perhaps the development of specific software able to compute these large-scale comparisons more easily.

Transcriptional regulators at the frontline of ecological adaptability

Owing to their high number and importance in the prokaryotic world, one would expect HTH-regulators to be functionally diverse. And so they are: genes that are known to be regulated by these TFs extend to virtually all possible facets of bacterial functioning. As often happens when labelling genes, the name of each family of HTH-regulators reflects the specific function of the first characterized member of that family. For instance, the Fur family of regulators was named after the *ferric uptake regulator* (*fur*) gene in *Escherichia coli*, which codes for a regulator of iron uptake systems (Hantke, 1981). This common naming procedure may, however, bias the reader to think that the Fur family of regulators is limited to iron homeostasis, which is not the case. In fact, besides iron homeostasis, Fur regulators can intervene in a wide range of processes such as metal (other than iron) homeostasis, virulence or acid-shock response (Escolar *et al.*, 1999). Several duplications in the phylogenetic history of the Fur-encoding genes have led to the presence of Fur orthologues specialized in regulating specific functions (Santos *et al.*, 2008). The same happens with other regulator families: although the primary function of their members may be related to the name of the family, some

Table 1. Diversity of one-component HTH regulators

Name	Main regulatory functions*	Total no. of members/ relative frequency†
AraC	Carbon metabolism, pathogenesis, stress response	2717/16%
ArsR	Metal detoxification, efflux and sequestration	943/6%
AsnC	Amino acid biosynthesis, degradation and transport, as well as the production of pili, porins, sugar transporters and nucleotide transhydrogenases	680/4%
Crp	Carbon metabolism	352/2%
DeoR	Carbon metabolism	640/4%
DtxR	Iron acquisition, storage and utilization	100/1%
Fur	Iron acquisition, storage and utilization	411/2%
GntR	Carbon metabolism	2327/14%
IclR	Carbon metabolism and other functions, such as breakdown of cell wall polysaccharides, catabolic pathways of aromatic compounds and sporulation	566/3%
LacI	Carbon metabolism and nucleoside uptake and utilization	1094/6%
LuxR	Biosynthesis, glycerol metabolism and quorum-sensing	2036/12%
MarR	Multidrug resistance (including antibiotics, oxidative stress, organic solvents and household disinfectants)	1402/8%
MerR	Metal detoxification, efflux and sequestration	986/6%
Xre	Wide variety of functions, including plasmid copy and bacteriophage transcription control, methylases and a specific vegetative protein from *Dictyostelium discoideum*	2878/16%

*Sources: SMART, InterPro, cited literature.
†In the present study.

members can also regulate other genes with slightly or markedly different functions. Through a selection of descriptive examples and guided by an ecological perspective, the following sections give an overall picture of several functions in which HTH TFs are involved.

Multidrug resistance

One of the cellular aspects in which HTH-regulators are known to play a fundamental role is the regulation of the resistance to a wide variety of toxic compounds. This resistance may be achieved by acting on the drug (detoxification or extrusion by efflux pumps), on the target (biochemical modification of the targets) or even on the bacterial cell itself (reduction of membrane permeability) (Alekshun & Levy, 1999; Godsey *et al.*, 2002). Multiple antibiotic resistance regulator (MarR) is an HTH family of

regulators that has been classically associated with multi-drug resistance regulation. MarR was initially characterized in *E. coli* as a regulator of antibiotic resistance, hence its name, but it is now known to regulate a vast range of genes involved in the resistance to many other toxic compounds, including organic solvents, household disinfectants and others (reviewed by Alekshun & Levy, 1999). The resistance phenotype is induced upon exposure to several chemicals, including those that contain aromatic ring(s). However, other regulatory families not classically associated with multidrug resistance, such as MarR, may also exert an effect on this phenotype. *Pseudomonas putida*, a model species for studying tolerance to highly toxic chemicals, has one efflux pump – TtgGHI – which has been revealed to be very important for the extrusion of toluene, as well as other aromatic hydrocarbons, aliphatic acids and antibiotics. The regulation of the *ttgGHI* operon is achieved by TtgV, an HTH repressor belonging to the isocitrate lyase regulator (IclR) family of regulators (Guazzaroni *et al.*, 2005). Unlike other IclR proteins, TtgV presents a broad range of coeffectors, such as indole or benzonitrile, a characteristic that seems to be common to multidrug efflux pumps regulators. Upon binding to one of its effectors, TtgV is released from DNA and allows RNA polymerase to transcribe the *ttgGHI* operon (Guazzaroni *et al.*, 2005). BmrR, in turn, is a mercury resistance regulator (MerR) family regulator which responds to what seems to be a broad range of effectors, namely rhodamine 6G and tetraphenylphosphonium, and which regulates the transcription of the multidrug efflux pump Bmr in *Bacillus subtilis* (Ahmed *et al.*, 1994).

The presence of these bacterial multidrug resistance mechanisms has implications for human health, of which the worldwide trend of increasing antibiotic resistance is the most notorious. The regular exposure to low doses of antibiotics may induce the production of resistant phenotypes in bacteria, often mediated by antibiotic-sensing molecules and transcriptional regulators. For instance, the genome of *Neisseria gonorrhoeae*, the causative agent of gonorrhoea, which according to the Center for Disease Control and Prevention (CDC) is the second most commonly reported notifiable disease in the USA, encodes an efflux pump composed of MtrC–MtrD–MtrE that is able to export diverse hydrophobic antimicrobial agents (HAs). Its corresponding operon is regulated both positively and negatively. Among many other factors, MtrA is an arabinose regulator (AraC) family regulator that has been proved to be essential for the activation of *mtrCDE* expression in response to sublethal doses of Triton X-100 and nonoxynol-9 (Rouquette *et al.*, 1999). So, if on the one hand it is true that TFs may mediate resistance to several drugs, on the other hand there is a possibility of interfering with the transcription regulatory pathways of pathogenic bacteria to develop more efficient therapies. Knowledge of the involved

regulators and coeffectors can thus represent an important step towards coping with the increasing prevalence of antibiotic resistance.

Metal homeostasis

Keeping transition metals in a life-supporting range of concentrations is a paradigmatic dilemma. Although some of these metals are essential nutrients for cellular metabolism, an excess of most of them is highly toxic, if not lethal. A very thin line separates metal starvation from metal toxicity. To survive, bacterial organisms require functional metal uptake and efflux pumps, which have to be tightly regulated by a sensitive and discriminating system. Metalloregulators can be so important for survival that their null mutants become nonviable: this has been suggested for *ideR* in *Mycobacterium tuberculosis*, a diphtheria toxin repressor (DtxR)-family TF (Rodriguez *et al.*, 2002), and for *fur* in *Vibrio anguillarum* (Tolmasky *et al.*, 1994) and *N. gonorrhoeae* (Berish *et al.*, 1993). Furthermore, the need to maintain transition metals within homoeostatic levels is often strongly correlated with the virulence capacity of pathogenic bacteria, implying that the regulators acting on each of these situations are also deeply related. In fact, a common nonspecific host defence response to infection involves the sequestration of iron and other metals, leaving the pathogens deprived of these nutrients and justifying the interaction of metal and virulence regulatory pathways. Particularly in *M. tuberculosis* it has been proved that some genes induced during macrophage infection are also over-expressed *in vitro* under iron-deficient conditions (Gold *et al.*, 2001).

Among metalloregulators, MerR has been widely studied as a typical metal resistance operon regulator (Hobman *et al.*, 2005). This regulatory protein is able to sense small amounts of Hg(II) and to regulate resistance genes responding to a very narrow range of Hg(II) concentrations. However, there is more in the MerR family than just MerR itself. ZntR and PMTR are MerR-family regulators that control zinc homeostasis in *E. coli* and *Proteus mirabilis*, respectively, and CoaR controls cobalt homeostasis in *Synechocystis* PCC 6803 (Outten *et al.*, 2000). Also belonging to the MerR family is SoxR, a major oxidative stress regulator that upon nitrosylation or oxidation of its [2Fe-2S] centre activates expression of *soxS*, which in turn induces the expression of a wide range of genes related to oxidant defence (Pomposiello & Demple, 2001). In the same way, arsenic regulator (ArsR) is mostly known as a family of repressors controlling arsenic intracellular concentration (Chen & Rosen, 1997). However, CzrA from *Staphylococcus aureus* is an ArsR homologue that seems to be involved in zinc homeostasis (Kuroda *et al.*, 1999). Although discriminating within each organism at levels that may vary across

different species, the families of metal-responsive TFs present the plasticity to sense and regulate the concentration of different metals.

Finally, *ideR* is an example of the overlap between virulence and metal homeostasis regulation. This gene encodes a DtxR regulator in *M. tuberculosis* and controls a wide range of genes, not all of them encoding proteins directly related to iron homeostasis (Rodriguez *et al.*, 2002). Being an essential protein, IdeR is a possible target for biomedicine approaches. Furthermore, and taking into consideration the oxidative burst known to occur in animal and plant host cells upon bacterial intrusion, the essential role of IdeR in oxidative stress response (Rodriguez *et al.*, 2002) supports its importance during infection. Another example concerning DtxR can be found in *Corynebacterium diphtheria*, where this regulator was first characterized as the regulator of *tox*, a gene that codes for the diphtheria toxin and now known to be conserved across a considerable range of corynebacterial species (Oram *et al.*, 2004). Interestingly, the signal to which DtxR responds is the concentration of iron. Another typical example of a pathogenesis-involved metalloregulator is Fur. For instance, in *Listeria monocytogenes* the separate disruption of two *fur* homologues has led to significantly less virulent strains upon infection of an animal model. Furthermore, both these mutants presented an increased resistance to hydrogen peroxide when compared with the wild type, and one of them was severely impaired for growth under low-iron conditions. Finally, these mutants were used to infect esculetin-treated mice, in which iron availability is higher than usual, and the results have shown that one of the mutants had its virulence fully restored, whereas that of the second one was closer to the wild-type (Rea *et al.*, 2004). Thus, it seems that the role of these two Fur homologues in virulence is directly influenced by the iron homeostasis. Moreover, evidence has been gathered suggesting that another *M. tuberculosis* Fur, *furA*, regulates *katG* and possibly other virulence factors encoding genes (Pym *et al.*, 2001).

Bioremediation and sensing of anthropogenic compounds

Being the major recyclers of the organic chemicals normally present in nature, bacteria have a crucial role in maintaining the ecological cycles running under dynamic equilibrium. However, and especially since the industrial revolution in the late 18th century, the anthropogenic input of synthetic compounds into the environment has been increasing to near-unsustainable levels. The chemical-degradation potential of microorganisms has thus gained a new role, as the need to metabolize or stabilize human-generated toxic compounds has been raised to one of the top priorities of the industrialized world. The knowledge of the regulatory

networks that control these catabolic pathways is essential to understand the bioremediation potential of each strain and to optimize it. A detailed and organized overview of transcriptional regulators (including representatives of the IclR, AraC, GntR, MarR and Crp families) involved in degradation of aromatic compounds has been published elsewhere (Tropel & van der Meer, 2004), and only a single example of each will follow. In *Pseudomonas aeruginosa* 142, an IclR regulator (OhbR) is suggested to control the transcription of two genes located downstream of it and involved in oxygenolytic ortho-dehalogenation of chlorobenzoates (Tsoi *et al.*, 1999; Tropel & van der Meer, 2004), whereas in *P. putida* WCS358, a *p*-hydroxybenzoate hydroxylase gene is regulated by an AraC regulator (PobC) in response to *p*-hydroxybenzoic acid (Bertani *et al.*, 2001). In *E. coli* W, *paaX* encodes a GntR regulator that responds to phenylacetyl-CoA and regulates the phenylacetic acid catabolic operon (Ferrandez *et al.*, 1998). BadR, in turn, is a MarR-family regulator that is likely induced by benzoyl-CoA and interferes with the regulation of benzoate degradation in *Rhodopseudomonas palustris* (Egland & Harwood, 1999). In *Desulfitobacterium dehalogenans* a Crp regulator (CrpK) is involved in the regulation of a cluster of genes responsible for reductive dehalogenation (Smidt *et al.*, 2000).

As in so many other situations, the regulation of catabolic pathways is not a straightforward relationship between a stimulus, a regulator and a specific gene/operon. On the contrary, multifaceted control of a specific pathway by different regulators responding to different environmental signals has often been reported. One such case is transcriptional control of the polyethylene glycol (PEG) catabolic machinery in *Sphingopyxis macrogoltabida*. In this organism, two different regulators seem to act in a coordinated way to promote the degradation of PEG: a LacI regulator, together with a histone-like protein, negatively regulate at least two genes of the *peg* operon, one of which encodes PegR, an AraC regulator that activates the transcription of the entire *pegBCDAE* operon. The LacI regulator is PEG-induced, derepressing *pegA* and *pegR* genes when in the presence of this common man-made polymer (Charoenpanich *et al.*, 2006). Some other illustrative examples concern the catabolism of chlorinated benzoic acids (CBA) by several bacteria: in particular, *Comamonas testosteroni* BR60 presents a set of enzymes, encoded by the operon *cbaABC*, able to degrade 3-chlorobenzoate. Upstream of this operon a gene was found encoding for a MarR-family regulator, CbaR, which was proved to participate in the regulation of *cbaABC* transcription together with an unidentified regulator (Providenti & Wyndham, 2001). CbaR was also found to be responsive both to benzoate and to similar compounds.

The ability of some bacterial strains to degrade organic compounds, some of which are exclusively man-made,

offers a promising field for biotechnological approaches (Tropel & van der Meer, 2004; Galvao & de Lorenzo, 2006). The optimization of pre-existing degradation pathways is one such approach, which aims at increasing the natural bacterial potential for bioremediation. Another approach is the development of micro-biosensors, which consists in the fusion of the promoters of bacterial catabolic pathways to easily measurable reporter genes, thus allowing the detection and quantification of specific pollutants. Further engineering of these promoters could increase their sensitivity or the range of factors to which they respond. Whatever the case, these applied approaches require a detailed knowledge of the regulatory cascades operating in catabolic bacterial routes, as they interfere precisely at the transcription level. For instance, several mutations of an AraC regulator from *P. putida* mt-2, known to regulate the transcription of a catabolic operon upon binding to benzoate and m-toluate, have been screened to detect altered response to effectors. Some mutants have demonstrated different sensitivities, with increased responsiveness to some compounds and an almost undetectable response to others (Galvao & de Lorenzo, 2006).

Virulence/pathogenicity

Pathogenicity is a complex and challenging phenotype, and thus one would expect the virulence factors to be under a dense and tight network of regulators. In fact, and in order to efficiently colonize a given host, the genomic repertoire of a pathogen has to include the ability to perceive the proximity of a susceptible host; the capacity to colonize it, despite the presence of diverse and more or less complex immunity mechanisms; and the machinery necessary to obtain the nutrients from the host tissue while at the same time keeping it alive as long as possible. The integration of these various signals and functions requires a well coordinated network of TFs, able to regulate the different steps of the pathogenesis process as a whole. A study has been done regarding the regulation of the virulence genes in *Listeria*, which are known to be encoded by a 9-kb fragment known as *Listeria* pathogenicity island 1 (LIPI-1). This island includes *prfA* that codes for PrfA, a cyclic AMP receptor protein (Crp) family protein which has the ability to activate the transcription of all genes in LIPI-1, including itself, and which when mutated leads to complete avirulence (Kreft & Vazquez-Boland, 2001). Moreover, the authors have studied the physico-chemical signals able to affect the expression of virulence genes. Most of them interfere with PrfA and possibly other regulators, and their diversity illustrates how broad the range of factors that intervene in the transcriptional control of virulence is: whereas high temperatures, limitation of growth, high osmolarity and host cell signals act as activation signals, low temperature, the presence of

fermentable sugars and low pH seem to be inhibitory signals (Kreft & Vazquez-Boland, 2001).

Among others, the AraC HTH-transcriptional regulatory family has been associated with pathogenicity (Gallegos *et al.*, 1997). For instance, in *M. tuberculosis* a mutant disrupted in an AraC homologue (Rv1931c) has been proved to be impaired in its survival abilities both in macrophages and after intravenous infection of mice, a phenotype that was confirmed by complementation assays indicating that Rv1931c expression is important for virulence (Frota *et al.*, 2004). Interestingly, some AraC-regulators that act as virulence activators seem to interfere with DNA in a particular way. Rns is an AraC-TF that regulates the expression of CS1 and CS2 pili in enterotoxigenic *E. coli*. Surprisingly, this protein does not bind to a small region upstream of the promoter it activates, as other HTH-regulators do, but instead has the capacity to bind to an unprecedented variety of DNA regions, both upstream and downstream of the regulated gene (Munson *et al.*, 2001; Egan, 2002). Substitution analyses between Rns and other regulators have shown that Rns is not an unique case, but instead is the archetype of a subfamily of AraC virulence activators (Munson *et al.*, 2001).

Besides AraC, many other regulators from different families may assume a crucial role in this phenotype. *Brucella melitensis* is the causative agent of brucellosis in domestic animals and, in some cases, may cause a systemic and febrile illness in humans known as Malta fever. Recently, a noteworthy work was undertaken to identify HTH transcription regulators involved in virulence of *Brucella*, by creating 88 integrative disruption-mutants from 10 different families and evaluating their ability to survive in mice (Haine *et al.*, 2005). After intraperitoneal inoculation of BALB/c mice, 10 of the tested mutants were found to be attenuated when compared with the wild type. Interestingly, six of these corresponded to bacteria that had been mutated in gluconate regulator (GntR) homologues (Haine *et al.*, 2005). Another study done on *B. melitensis* has revealed the presence of three putative Crp homologues, *narR*, *nnrA* and *nnrB* (Haine *et al.*, 2006). Whereas NarR regulates the nitrate reductase regulon, an *nnrA* mutant was shown to be strongly attenuated after infection in mice. So not only is NnrA involved in the regulation of the three last reductases in the denitrification pathway, but its presence has also been proved to be essential for both full virulence and resistance to NO (Haine *et al.*, 2006).

As stated above, one of the abilities necessary for pathogens is to sense the proximity of a susceptible host. To achieve this, bacteria have evolved different mechanisms that include the activation of a TF by a host molecule. *Proteus mirabilis* is a causative agent of urinary tract infections that, when untreated, may produce kidney and bladder stones. Such precipitation of calcium and magnesium take place when the pH increases due to the production of ammonia from the hydrolysis of urea. The enzyme that catalyses this reaction – urease – is therefore considered to be an important virulence factor of *P. mirabilis*. Its transcription can be positively activated by UreR, an AraC-regulator. Studies done on infected CBA mice with insertionally inactivated *ureR* mutants have shown that in the presence of the wild-type strain, the mutants were always outcompeted in an infection (Dattelbaum *et al.*, 2003). It seems therefore that UreR is responsible for the transcription of a crucial virulence factor – urease – and is activated by a host molecule – urea. Another example of this host perception by the pathogen comes from a phytopathogen: *Streptomyces scabies*. This causative agent of potato scab produces thaxtomin, a plant toxin and a determinant of virulence. An AraC homologue, TxtR, was found to activate thaxtomin biosynthetic genes and to be essential for pathogenicity: in fact, tobacco plants inoculated with *txtR*-knockout mutants displayed almost no symptoms (Joshi *et al.*, 2007). Moreover, these mutants were also unable to activate thaxtomin by nitration, thus suggesting that TxtR also mediates the transcription of genes responsible for the generation of nitrogen-active species. Interestingly, the cellulose building-block cellobiose was found to bind to TxtR. It is likely that this molecule can be found in regions where cellulose is being produced for the expansion of plant tissue, such as growth zones, which are precisely the sites of thaxtomin synthesis (Joshi *et al.*, 2007). Finally, *Xanthomonas oryzae* pv. *oryzae*, responsible for the bacterial leaf blight disease in rice, is also a good example of this communication across kingdoms. This species possesses a quorum-sensing luminescence regulator (LuxR) homologue (OryR) that, instead of binding a bacterial acyl-L-homoserine lactones (AHL), seems to respond to the presence of a molecule present in rice (Ferluga *et al.*, 2007). *oryR* mutants in different strains have shown reduced virulence in rice when compared with the wild type. Interestingly, OryR was induced in the presence of macerated rice, therefore supporting the hypothesis that it is a molecule from the host itself that alerts the pathogen to the possibility of infection and initiates transcription of the necessary genes.

Metabolism and development

By definition, cellular metabolism is the sum of all biochemical reactions by which the cells process the nutrients into energy and biomass. Excluding situations where life is maintained under a dormant state such as in bacterial spores, a continuous metabolic flux is necessary for survival, as is the transcription control of its elements. Sensing of available nutrients in the environment and coupling that perception to the synthesis of the appropriate enzymatic apparatus, together with the downregulation of unneeded

metabolic routes, allow bacteria to optimize the exploration of its surroundings. The paradigm of carbon metabolism regulation in bacteria is centred on the Crp family of TFs, which has been reviewed in detail elsewhere (Korner *et al.*, 2003). Different members of this family respond to specific signals, such as cAMP, O_2, NO, CO and 2-oxoglutarate, and regulate a wide variety of metabolic-related phenotypes in different organisms, which include nitrogen fixation in symbiosis, stationary phase survival, use of arginine as an energy source, acquisition of polysaccharides and halorespiration (Korner *et al.*, 2003). Inactivation of these regulators, by interfering with basic life-supporting biochemical pathways, can probably cause drastic effects on the growth and survival of the mutated strains. It has been demonstrated that a *crp* knockout mutant in *E. coli* is severely impaired for growth on glucose, even though it did not display any specific trend on the cellular metabolic flux distribution (Perrenoud & Sauer, 2005). On the other hand, carbon metabolism regulation may also interfere with virulence of pathogenic bacteria. *Erwinia chrysanthemi* is a plant pathogen that can cause soft-rot disease in various plants and can grow using pectin as its sole carbon source. An *E. chrysanthemi* Crp participates in the transcriptional control of *paeX*, which encodes a pectin acetylesterase that facilitates pectin degradation by removing its acetyl groups. Besides being important for carbon metabolism, the breakdown of pectin promotes the tissue maceration observed in infected plants (Shevchik & Hugouvieux-Cotte-Pattat, 2003).

Besides Crp, other TFs play important roles in the regulation of bacterial metabolism. The asparagine synthase regulator (AsnC) family, also known as AsnC/Lrp after its first two characterized members, is one such case, and its members are often referred to as feast/famine regulatory proteins (FFRPs). In fact, in *E. coli* Lrp senses the extracellular availability of leucine and affects the regulation of nutrient acquisition via the cell membrane as well as the termination of autotrophic pathways, participating therefore in the adaptation to nutrient-rich or -depleted environments, whereas AsnC senses asparagine and regulates its biosynthesis (Yokoyama *et al.*, 2006). In *Neisseria meningitidis*, however, a protein annotated as AsnC binds L-leucine and L-methionine instead. A knockout mutant in this regulator presents a slower growth under nutrient-poor conditions when compared with the wild-type strain, although its growth on rich medium remains almost unaffected, which indicates that this TF is likely involved in the adaptation to starvation (Ren *et al.*, 2007). Interestingly, it seems that in the AsnC family members the presence of the ligand acts by stabilizing an oligomeric form of the regulator, instead of changing its conformation (Yokoyama *et al.*, 2006; Ren *et al.*, 2007), which is more common among the other HTH regulators.

Other families of regulators that have functions concerning the transcription of genes involved in metabolism are DeoR, LacI and GntR, and a few examples are described below to illustrate each case. In *E. coli*, use of L-ascorbate is made possible by proteins encoded in the regulon *ula*, which is composed of two divergently transcribed operons. UlaR, a regulator that belongs to the deoxyribonucleoside synthesis operon regulator (DeoR) family, seems to repress the transcription of these genes by binding simultaneously to the promoters of both operons and mediating the formation of a DNA-loop (Campos *et al.*, 2004). In *Bifidobacterium lactis*, the presence of raffinose and sucrose was shown to induce the transcription of three genes involved in sucrose use, including *scrR*, a LacI transcription regulator. Studies in a heterologous system indicated that ScrR functions as a positive regulator of the operon (Trindade *et al.*, 2003). Finally, in *Corynebacterium glutamicum* a GntR family regulator (FarR) was shown to be involved in the regulation of amino acid biosynthesis (Hanssler *et al.*, 2007).

As virtually everything that happens inside a cell is dependent on the presence of energy, it is expected that the regulators involved in primary metabolism influence, directly or indirectly, many other cellular reactions. The ability of *Vibrio harveyi* to emit luminescence is a paradigmatic case of a quorum-sensing controlled function, and a LuxR regulator is known to play a crucial role in this. Nevertheless, this phenomenon also requires individual cells to invest energy and resources. In fact, it has been demonstrated that two regulators involved in sensing and controlling the nutritional status of the cell are also important for bioluminescence regulation (Chatterjee *et al.*, 2002). Whereas a Crp functions as an activator, a LysR regulator acts as a repressor. Interestingly, electrophoretic mobility shift assays have demonstrated that these regulators actually bind to the *lux* promoter (Chatterjee *et al.*, 2002), which suggests that the relationship between different HTH-regulators can be quite complex.

If cellular metabolism leads to an increase in biomass and energy, the cell can then use these resources either to grow and multiply, or to develop towards a certain state. Being a direct consequence of metabolism, it follows that the regulators that control development are closely related to those involved in metabolism. Although multicellular development is not common among bacteria, one can actually find some instances of structural differentiation, such as the maturation of flagella, which allows motility in some bacterial species, or the pleiomorphic behaviour of certain bacterial colonies. In *E. coli*, for instance, a *crp* mutant was shown to be nonmotile. This was attributed to the reduced expression of flagellin and flagellar sigma factor, which resulted into a complete lack of flagellin accumulation (Soutourina *et al.*, 1999). Similarly, *Streptomyces coelicolor* is known to develop an aerial mycelium that ends up differentiating into chains of exospores. However, a mutant in a GntR-like regulator, DevA, was shown to produce a

sparse aerial mycelium, with short and rare hyphae and spores with an unusual shape (Hoskisson *et al.*, 2006). Finally, the establishment and maintenance of a symbiotic state is also under regulation at the transcription level. Nitrogen fixation being an oxygen-labile process, oxygen is expected to play a crucial role in the regulation of the nitrogen-fixing apparatus. And in fact, in *Rhizobium etli* CFN42 at least three regulators belonging to the Crp family are involved in the microaerobic transcription of *fixNOQP*d, an essential operon for symbiotic nitrogen fixation (Lopez *et al.*, 2001).

Quorum-sensing

Initially thought to be strictly solo organisms, bacteria were viewed as unicellular beings originating colonies morphologically defined as a dense cluster of individual cells surviving purely on their own. Nowadays, increasing research in the field has shown that in fact these bacterial colonies can function as a kind of a 'super-organism', where an intricate network of cell-to-cell communication signals drives individual regulatory cascades leading to the concerted activation or repression of certain pathways. This phenomenon, known as quorum-sensing, is mainly dependent on cell density and can influence several bacterial functions, including secondary metabolism, development of mutualistic/symbiotic relationships, virulence efficiency and even monitoring of other microbial species' populations (Fuqua *et al.*, 1996; Fuqua & Greenberg, 1998). Members of the LuxR family of TFs function as quorum-sensors by binding the signal molecules and regulating the expression of cell density-dependent genes. These signals, the synthesis of which is mediated by LuxI homologues, are often *N*-AHLs in *Proteobacteria*, whereas in *Firmicutes* and *Actinobacteria*, amino acids and short post-translationally processed peptides are more common. One of the first quorum-sensing systems to be studied was in *Vibrio fischeri*, that allowed the establishment of a basic model of bacterial cell-to-cell communication. Briefly, as the bacterial density increases, so does the amount of AHLs produced and exported to the medium, and consequently the rate of AHL influx to neighbouring bacteria. After reaching a threshold, the AHLs bind to LuxR homologues and activate them, which in turn regulate the expression of several genes (Fuqua *et al.*, 1996; Zheng *et al.*, 2006). Quorum-sensing functions do not fit in any defined field of bacterial activity: instead, they affect several different bacterial pathways in different organisms. Therefore, in this subsection, we will give a brief perspective on the variety of functions quorum-sensing regulation can affect, many of which have been mentioned in previous sections.

An illustrative example of the diversity of cell density-regulated genes can be found in *Serratia* spp. This *Enter-*

obacteriaceae family comprises several organisms that can be easily found in soil, water, plant surfaces and raw food materials. Some *Serratia* spp. can become severe opportunistic pathogens and are the causative agents of a series of food-borne diseases (Van Houdt *et al.*, 2007a, and references therein). Quorum-sensing activation of genetic expression in *Serratia* spp., mediated by LuxR homologues, can intervene at several metabolic levels. In *Serratia marcescens* MG1, a LuxR homologue (SwrR) and its cognate LuxI homologue (SwrI) were found to be involved in the regulation of cellular motility. These cells have two different kinds of flagellum-dependent motility: swimming and swarming. A *swrI* knockout mutant had a reduced swarming ability, which could be restored upon the addition of one of its cognate AHLs, C4-HSL (Van Houdt *et al.*, 2007a). Indeed, the *swrI/swlR* system has been implicated in the production of serrawettin W2, which reduces surface tension and thus allows swarming (Van Houdt *et al.*, 2007a, and references therein). On the other hand, this same SwrI/SwrR system in *S. marcescens* MG1 and its corresponding SprI/SprR in *Serratia proteamaculans* B5a were found to regulate transcription of *lipB*, a type I secretion apparatus. This control of extracellular enzyme component may contribute to pathogenesis and food spoilage processes (Van Houdt *et al.*, 2007a, b). Furthermore, in *Serratia plymuthica* RVH1, a LuxI homologue – SplI – was found to synthesize one AHL (3-oxo-C6-HSL) and to influence the production of two others (C4-HSL and C6-HSL). Using mutants that could be rescued by the different AHLs, it was demonstrated that SplI/SplR was involved in the regulation of the production of an extracellular chitinase, a nuclease, a protease and an unidentified antibacterial compound (Van Houdt *et al.*, 2007b). Adding to this, both *S. plymuthica* RHV1 *splI* mutant and *S. marcescens* *swrI* mutant presented a reduced ability to ferment glucose to 2,3-butanediol, as is common for some *Enterobacteriaceae* including *Serratia* spp. As a result, when growing in the presence of fermentable sugars, these mutant strains produced acidic end products, leading to the acidification of the medium and to an early growth arrest (Van Houdt *et al.*, 2007a, b).

Besides the phenotypes described above, quorum-sensing control also interferes in the regulation of *Serratia* spp. secondary metabolism. Two secondary metabolites of this genus have been widely studied, mainly due to their biomedical potential: prodigiosin and carbapenem antibiotics. Prodigiosin is a reddish pigment; its role(s) in bacterial cells is(are) still unclear but it has shown antibacterial, antiprotozoal, antimalarial, antifungal, anticancer and immunosuppressive activities. Carbapenem belongs to the beta-lactam broad-spectrum antibiotic family. In *Serratia* sp. American Type Culture Collection 39006, a mutant for a LuxI homologue, *smaI*, displayed a phenotype with reduced production of both prodigiosin and carbapenem (Thomson

et al., 2000). Additionally, prodigiosin production was also found to be cell-density dependent in *S. marcescens* SS-1, whereas in *S. plymuthica* RVH1 and *S. marcescens* st. 12, production of both antibiotics were found to be controlled by quorum-sensing (Van Houdt *et al.*, 2007a, and references therein).

In *Mesorhizobium tianshanense*, a quorum-sensing LuxI/LuxR system (MrtI/MrtR) was found to be crucial for symbiosis. Despite the fact that *mtrI* and *mtrR* knockout mutants could grow normally in tryptone yeast (TY) medium, their AHL production was severely impaired (Zheng *et al.*, 2006). Moreover, both these mutants had an aberrant root hair adherence phenotype and were unable to nodulate their host plants (Zheng *et al.*, 2006). This example illustrates the importance of quorum-sensing for symbiosis. On the other hand, in *P. aeruginosa*, a quorum-sensing system seems to be important for the adaptation to starvation. Indeed, the stringent response induction activates the expression of quorum-sensing regulated genes in a cell-density independent fashion, including the virulence factor LasB, among others. This early induction of virulence factors allows the organism in a nutrient-poor medium to infect the host, escaping to a much richer environment, or provides new nutrients due to the enzymatic activity of the proteins expressed (van Delden *et al.*, 2001).

Finally, quorum-sensing regulators have a crucial role in some pathogenesis processes. *Vibrio cholerae*, the causative agent of cholera, has been extensively studied in this regard. The known pathogenesis pathway of this organism involves two sensory proteins, ToxR and TcpP, which in the presence of certain environmental signals activate the transcription of ToxT. This in turn induces the production of virulence factors such as the cholera enterotoxin and the toxin-coregulated pilus. The presence of a LuxI/LuxR system (LuxO/HapR) has been characterized in *V. cholerae*, and proved to influence the pathogenicity of this organism. Briefly, it seems that at low bacterial densities, LuxO inhibits the expression of *hapR*, allowing the transcription of *tcpP* and activating the ToxR regulon. On the other hand, at high bacterial densities, LuxO is inhibited by the presence of high signal concentrations and the transcription of *hapR* increases. HapR, in turn, represses the expression of *tcpP* and consequently that of the *toxR* regulon, leading to a decrease in virulence (Zhu *et al.*, 2002). Furthermore, HapR also induces the expression of an HA protease, which may contribute to the detachment of *V. cholera* cells, therefore promoting the search for new infection entry points or even new hosts. Conversely, *hapR* mutants have the ability to form thicker biofilms, which promotes their persistence in the host (Zhu *et al.*, 2002). Another instance of the importance of quorum-sensing regulation in pathogenicity occurs with *Pectobacterium carotovorum* spp. *carotovorum*, a phytopathogen with a wide range of hosts that requires the presence of two LuxR homologues for effective infection. These two LuxR regulators (ExpR1 and ExpR2), in the absence of their corresponding AHL, upregulate the expression of *rsmA* (Mole *et al.*, 2007 and references therein). RsmA is involved in post-transcriptional regulation and destabilizes mRNA transcripts of several genes that code for plant cell-wall degrading enzymes, and therefore its presence inhibits virulence. *Pseudomonas syringae*, in turn, is a saprophytic organism that lives in the soil and can infect a narrow range of plants, causing the bacterial speck disease. For *P. syringae*, a quorum-sensing system involving a LuxR TF (AhlR) seems to be crucial throughout the entire pathogenesis process (Mole *et al.*, 2007 and references therein). In fact, by controlling extra-polysaccharide production and swarming motility, which accounts for the dispersion of epiphytic bacteria, AhlR is essential for host colonization. Moreover, in later phases of infection, *ahlR* mutants were unable to macerate host tissues, which means that the need for quorum-sensing in *P. syringae* virulence is not restricted to part of the infection cycle (Mole *et al.*, 2007 and references therein).

The crucial role of quorum-sensing systems in virulence has started to attract the attention of pharmaceutical companies that aim to develop new therapies based on the inhibition of these processes. The above-mentioned problem of antibiotic-resistant bacteria has led investigators to consider as an alternative the decrease of the virulence of infective bacteria, instead of bactericide or bacteriostatic traditional approaches. This virulence decrease may be achieved by targeting three different points of the quorum-sensing process: the signal generator, the signal itself, or the signal receptor, a LuxR homologue. Experiments in this field have yielded very promising results: not only are quorum-sensing inhibitors able to reduce virulence and promote elimination of pathogenic bacteria in animal models, they can also make bacterial biofilms more susceptible to traditional therapies, therefore opening the possibility for complementary strategies (Rasmussen & Givskov, 2006). Interestingly, both algae and higher plants have been found to produce AHL inhibitors (Koch *et al.*, 2005); these could then be used as prototypes for synthesis of new drugs.

Patterns of HTH-regulator distribution

HTH-regulators, encoded by genes present in considerable numbers in bacterial genomes, are involved in a large variety of functions, as highlighted above. In this context, statistics is a good approach to discern tendencies and to determine the presence and the nature of the factors that may influence their distribution. To assess the frequency of each regulatory family, a query sequence carrying a single protein domain corresponding to each family (except in the case of Crp, where a cyclic nucleotide-monophosphate binding domain

– cNMP – has been characterized, thus making two domains) was chosen and submitted to hidden Markov model searches on the SIMPLE MODULAR RESEARCH ARCHITECTURE TOOL (SMART) website (March 2007) in the Genomic Mode (Letunic *et al.*, 2006); afterwards, all proteins with similar domain composition were displayed and counted individually for each genome, with manual curation. In parallel, all organisms involved were classified according to several features, covering both genomic and ecological aspects.

An overall analysis considering the total number of members of each regulatory family across all the analysed genomes revealed that AraC and Xre (16% and 17%, respectively) are the most abundant families, closely followed by GntR (14%) and LuxR (12%). DtxR is the less abundant family (1%), together with Crp and Fur (2% each), IclR (3%), AsnC and DeoR (4% each). Finally, 8% of the regulators belong to MarR family, and MerR, ArsR and LacI account for 6% of the total regulators each. It should be noted, however, that the weight a TF has on the whole regulatory pathways is far from being directly related to its abundance: not only are a few regulators able to modulate the expression of a large number of genes, but also some TFs are responsible for regulating other TFs, therefore amplifying their range of action by indirectly regulating a large number of genes. For instance, previous studies have shown that despite its low frequency, Crp has a considerable impact on *E. coli* TRNs. In fact, although there are only two regulators belonging to the Crp family in the *E. coli* genome, one of them is able to regulate directly up to 200 genes and is beyond doubt the TF that regulates the highest number of TFs (Martinez-Antonio & Collado-Vides, 2003).

The data on the frequency of each family of regulators was further analysed with correspondence analysis (CA) and, interestingly, a statistically significant ($P = 0.001$) division of the regulators into two groups, separated by the first component, was observed (Fig. 1). This division suggests that regulatory families are not evenly distributed along the genomes, but rather have genome-dependent patterns of distribution. The moon-shaped group is constituted of AraC, AsnC, Crp, DeoR, GntT, IclR, LacI and LuxR, and the round-shaped group includes ArsR, DtxR, Fur, MarR, Mer and Xre (Fig. 1). Biologically, this division is sustained by function. According to the description given in Table 1, the first group of regulators is composed of basic metabolism-related regulators, i.e. TFs involved in the transcription of genes needed for the basic functioning of the cells, including growth, reproduction and communication. The second group of regulators comprises TFs involved either in metal homeostasis of the organism or in the response to stress situations, including the presence of drugs or reactive oxygen species (ROS). So this division of the regulators, which considers the distribution pattern of each one across the different genomes, has indeed a functional basis.
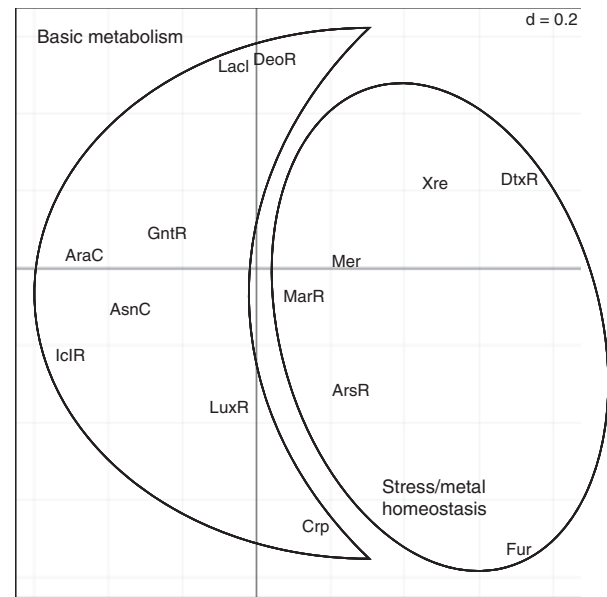
**Fig. 1.** CA analysis of the frequency table of the regulators across genomes.

Although the structure and mode of action seem to vary within the two groups, the functional kind of regulatory pathways in which each of the regulators is engaged is maintained.

The data on the qualitative classification of the studied organisms was analysed with multiple correspondence analysis (MCA). From the resulting plot (Fig. 2a) the genome size emerges as the main trait influencing the distribution of the organisms, defining the first component. Large genomes are positioned on the right side and a progressive decrease in size is observed towards the left side of the plot.

Size does matter: the influence of the genome dimension

A cluster analysis using the Ward method (Ward, 1963) with a threshold cut of the MCA scores revealed the presence of six groups of genomes automatically defined, which were plotted on the regulators CA (Fig. 3). These groups were distributed along the first component, and taking into consideration that this component corresponds globally to a variation in genome size, the formed groups can be seen as six size classes (with all the other associated characteristics). The same can be seen when plotting the groups of organisms defined by their genome size on the regulators CA plot (Fig. 4a).

Increase in the number of genes in a genome has been postulated to be driven by the need for phenotypic flexibility (Bertin *et al.*, 2008), which implies also a greater need of TFs to regulate all the different functions. Such a correlation has been shown by Aravind *et al.* (2005). According to this
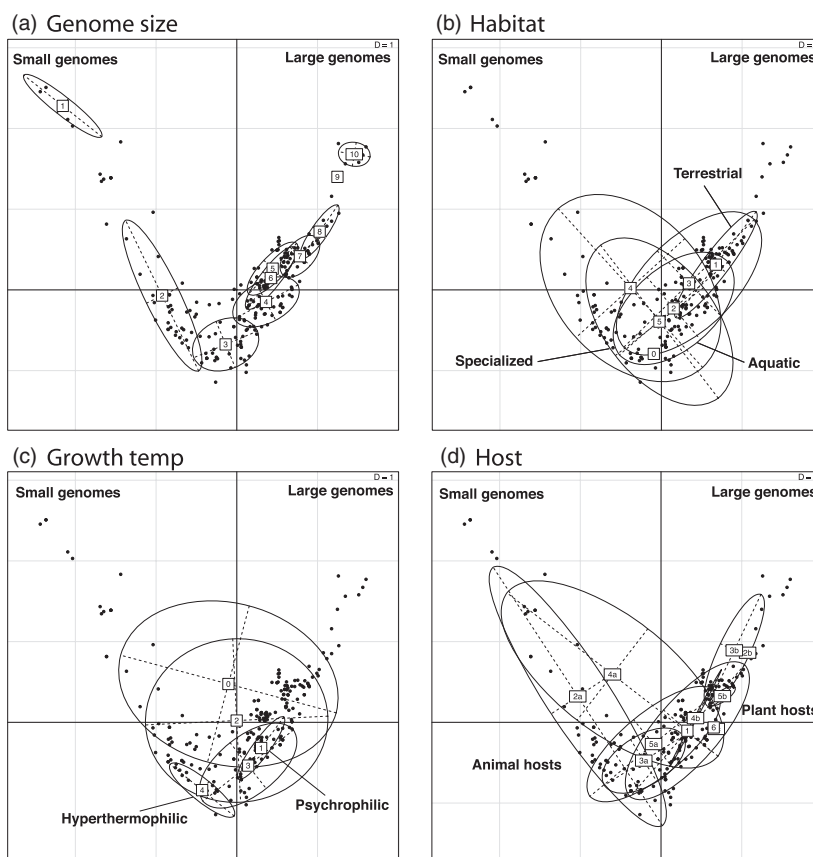
Fig. 2. MCA analysis of the qualitative classification of the organisms involved in the statistical approach employed. (a) Distribution of the organisms according to their genome size, based on the data in the Genome Project for each genome present in the National Centre for Biotechnology Information (NCBI): 1, genome $\leq$ 1 Mbp; 2, genome = 1.1–2 Mbp; 3, genome = 2.1–3 Mbp; 4, genome = 3.1–4 Mbp; 5, genome = 4.1–5 Mbp; 6, genome 5.1–6 Mbp; 7, genome = 6.1–7 Mbp; 8, genome = 7.1–8 Mbp; 9, genome = 8.1–9 Mbp; 10, genome = 9.1–10 Mbp; 11, genomes $\geq$ 10.1 Mbp. (b) Distribution of the organisms according to their habitat, based in the classification parameters present in the Genome Project for each genome available in NCBI: 1, terrestrial; 2, aquatic; 3, multiple; 4, host-associated; 5, specialized. (c) Distribution of the organisms according to their optimal growth temperature, based on the classification parameters of the Genome Project for each genome available in NCBI: 0, unknown/undefined; 1, psychrophilic; 2, mesophilic; 3, thermophilic; 4, hyperthermophilic. (d) Distribution of the organisms according to their relationship with the host, based on the information in the Genome Project for each genome available in NCBI, as well as in the papers describing the sequencing of each specific genome: 1, no association; 2a, strict symbiosis/commensalism with animals; 2b, strict symbiosis/commensalism with plants; 3a, facultative symbiosis/commensalism with animals; 3b, facultative symbiosis/commensalism with plants; 4a, strict pathogen in animals; 4b, strict pathogen in plants; 5a, facultative pathogen in animals; 5b, facultative pathogen in plants; 5c, facultative pathogen in bacteria; 6, bacterial commensalism. As a criterion, bacteria that could grow either with the hosts or under free-living conditions were considered facultative symbionts, whereas opportunistic pathogens, which can also establish a commensal relationship with the host, were considered to be facultative pathogens.

study, the number of HTH motifs encoded by a certain genome increases with its size independently of their specific nature. However, and interestingly, the nature of HTH proteins considered interferes in the relationship with genome size: whereas in HTH involved in two-component systems there is a proportional increase in the number of motifs with the increase in genome size, HTH involved in one-component systems and sigma factors present a nonlinear increase with gene number (Aravind *et al.*, 2005).

More than showing just a nonlinear overall increase in the number of one-component HTH-regulators, the statistical approach used in this review reveals that this increase is not indiscriminate, but is rather biased towards a specific group of regulators, the metabolic ones. Indeed, by observing the positioning of the different groups in the CA plot, one can distinguish a correlation between small genomes and metal homeostasis/stress response regulators and between large genomes and basic metabolism regulators. This situation finds support in the fitness potential of the different genome sizes. Small genomes have less coding capacity and, consequently, fewer regulators. As such, the organisms carrying these genomes have fewer genomic resources to explore the
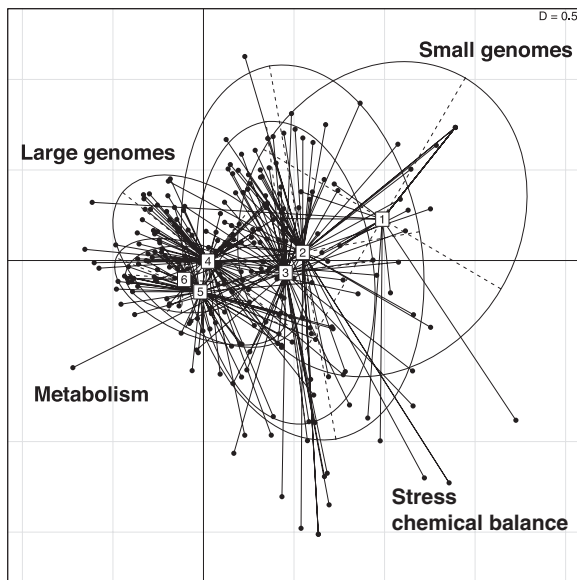
Fig. 3. Six groups formed by cluster analysis applying the Ward algorithm to the overall MCA scores and plotted into the regulators' CA.

surrounding environment and their functionalities are limited to what is actually essential for any living organism: survival. For that reason, a correlation is expected between small genomes and regulators engaged in situations that may interfere with the chances of survival, such as the chemical balance of metals and the ability to avoid stress situations. On the other hand, large genomes have a greater potential for exploration of different compartments of their environment, which results in higher metabolic rates and complexity, and therefore their correlation with metabolic regulators.

## Living underneath: perceptible trends in soil bacteria

From a biological point of view, survival depends on adaptation to a particular ecological niche. The habitat imposes a set of primary selective forces which define and shape its biodiversity. The statistical analyses carried out revealed that bacteria inhabiting both terrestrial (soil) and specialized environments show a correlation with Fur and Crp, and secondarily with LuxR (Fig. 4b).

In addition to this, and by overlapping the patterns generated by MCA analyses (Fig. 2), a correlation between large genomes and soil habitat can be observed.

Concerning genome size, it should be kept in mind that small genomes are more competitive in terms of replication efficiency. On the other hand, a complex environment implies different trophic resources, which in turn results in larger genomes. Soil is, indeed, a complex habitat, involving

particular ecological niches with unique characteristics, such as the presence of various plants roots, and a large flow of organic matter and inorganic nutrients. Under these conditions, the presence of large genomes is made possible by their functional plasticity and diversity. On the other hand, decrease in genome size is mainly driven by selective pressures related to metabolic streamlining, where phosphate can act as the limiting nutrient in DNA replication, reduction of energy consumption and space needed and replication speed (Cavalier-Smith, 2005; Ranea, 2006). Whereas these pressures are particularly conspicuous in pathogens and symbionts, they are negligible in the soil, where the constant recycling of nutrients accounts for energy and phosphate needs and no external selective forces towards fast replication are known. Finally, the increase in genome size has two main underlying causes: gene duplication and horizontal gene transfer (HGT). Whereas duplication should not depend on the habitat characteristics, HGT is more likely related to the number and diversity of species occupying the same ecological niche. Soil is known to have both high diversity and density of microbial species, both of which might favour HGT. However, in other environments, either because the microbial cells are protected by an external barrier (e.g. intracellular parasites), or because the medium is loose and does not favour proximity between different organisms, or even because it has extreme characteristics so that few species can withstand them, rates of HGT can be significantly lower. In conclusion, two main factors seem to account for the large size of the soil bacteria genomes: the absence of selective pressures towards genome reduction, and the possibility of high rates of HGT.

In agreement with that, experiments on digital organisms have shown that learning to know the environment is a crucial factor towards selectivity, as it increases the fitness of an organism. The more complex an habitat is, the more information it contains. Perceiving multifaceted environments requires the storage of larger amounts of data and thus results in an increase in genome size (Ofria *et al.*, 2003). Perceptibly, the increase in coding sequences and ability to respond to different situations can only be optimized with the development of specific and finely tuned regulatory systems. On the other hand, a recent study focused on the number of regulators has highlighted that even though *Trichodesmium erythraeum* and *Sinorhizobium meliloti* have genomes with approximately similar sizes (7.7 and 6.7 Mbp, respectively), they have contrasting numbers of one-component regulatory systems (69 and 390, respectively). Their ecological niches, which are the upper levels of tropical oceans for *T. erythraeum*, considered to be mild, and the soil for *S. meliloti*, considered to be physically and chemically challenging, was presented as one of the factors justifying the greater need for regulators in *S. meliloti* (Ulrich *et al.*, 2005).
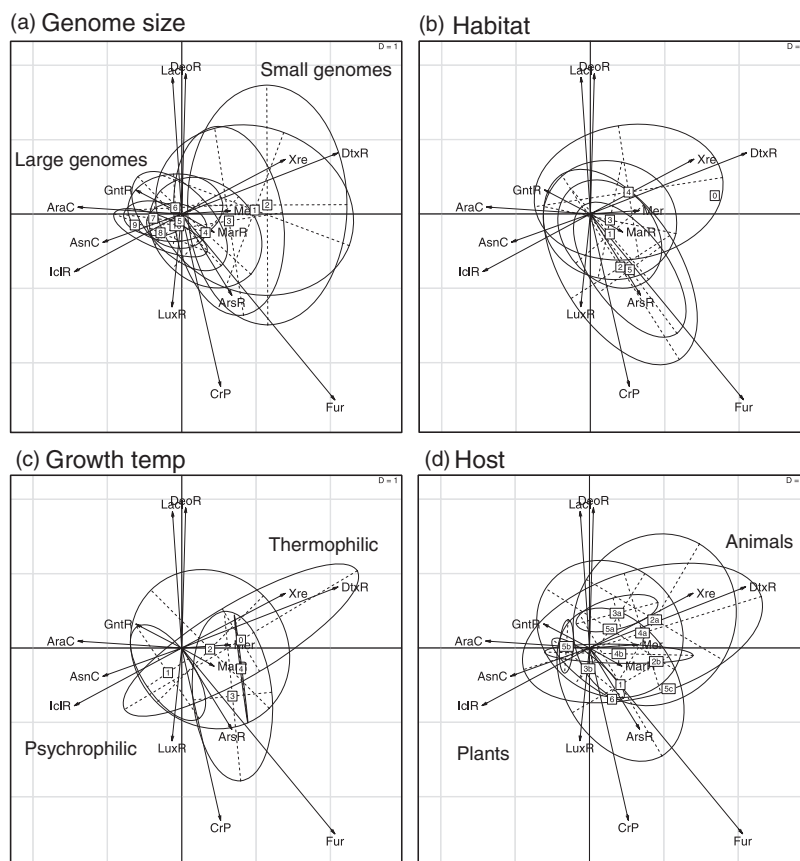
Fig. 4. Qualitative groups of organisms plotted on the regulators' CA. (a) Organisms defined by genome size. (b) Organisms defined by habitat. (c) Organisms defined by optimal growth temperature. (d) Organisms defined by their relationship with the host. For the description of the numbers, please see the legend for Fig. 2.

Concerning the distribution of regulators, soils are biotopes known for constant nutrient recycling, where bacterial communities are responsible for the decomposition of organic matter, providing nitrogen (N), phosphorus (P) and carbon (C). This important flux of nutrients through bacterial cells is the most likely factor driving their need of a strict control on their own metabolism as well. Crp is an important carbon-metabolism regulator (Table 1), and Fur has been shown to interact with Crp at least in *E. coli*, integrating iron homeostasis and carbon metabolism signals and transcriptional responses (Zhang *et al.*, 2005), therefore illustrating their importance in the soil adaptation scenario. LuxR, in turn, presents quorum-sensing functions and is therefore most likely correlated with the communication within the highly populated microbial communities present in the terrestrial environment.

The reasons why the same regulators are also correlated with specialized environments, are not as clear. This class of habitat encompasses diverse extreme conditions, from hot springs to gamma-irradiated biotopes, so the explanation for the importance of a certain family of regulators most

likely requires the consideration of each case individually. Nevertheless, it seems that the existence of various stress sources is the most conspicuous characteristic that specialized environments share with each other and the soil, and thus may be the common link resulting in the same differential usage of regulators.

Hot spots in regulatory networks: implications of the optimal growth temperature

The optimal growth temperature of an organism is another factor shown to influence the frequency of the distinct families of transcriptional regulators and to be somewhat correlated with genome size. Both psychrophiles and thermophiles are correlated with medium to large genomes, whereas hyperthermophiles show a tendency towards medium to small genomes (Fig. 2c). On the other hand, whereas thermophiles and hyperthermophiles are correlated with Fur, Xre and to a lesser extent DtxR, ArsR and Crp, evidence concerning psychrophiles suggests positive selection of AsnC, IclR and LuxR (Fig. 4c). Despite the specificity of the

underlying causes that justify these differences, all find support in the rationale that growth temperature has a marked effect on the metabolism of any organism and that extreme temperatures have divergent outcomes.

The main constraints of growing under low temperatures are related to two essential factors – low thermal energy and high viscosity (D'Amico *et al.*, 2006). Both these characteristics strongly impact normal metabolic rates of any non-adapted organism: whereas low thermal energy implies reduced enzyme activity, high viscosity decreases membrane fluidity, which interferes with the uptake of nutrients and disposal of waste and toxic products, to mention just a few effects (Deming, 2002; D'Amico *et al.*, 2006). Living in the cold implies a series of adaptations that enable psychrophiles to maintain life-supporting metabolic rates, including the adjustment of the rate of unsaturated/saturated fatty acids in the membranes and the lowering of the activation energy of the enzymes (Deming, 2002; Chattopadhyay, 2006; D'Amico *et al.*, 2006). Nevertheless, and taking into consideration that low temperatures constitute a strong hindrance to biochemical transformations, it is rational to argue that basic metabolism should be under strict control in psychrophiles, which in turn justifies the statistical correlation found between these organisms and AsnC, IclR and LuxR, three families of regulators involved in basic metabolic pathways (Table 1).

On the other hand, the association between thermophiles and Fur, DtxR, Xre, ArsR and Crp can be related to the important presence of dissimilatory metal reduction in thermophilic microbial communities. In fact, metal reduction was found to be a characteristic common to all thermophilic bacterial communities characterized so far, with emphasis on iron reduction (Slobodkin, 2005). The selective pressures towards the use of inorganic electron acceptors and the integration of metal and carbon response may be related to the positive selection of respectively Fur, DtxR, ArsR and Crp in these bacteria (Table 1).

Animals vs. plants: when the host dictates the rules

Analysing the distribution of the host-association classes in the MCA and CA plots, two trends of a different nature are discernible. The first concerns the correlation between the host type, animals vs. plants, and the genome size. Comparing the MCA plots of the genome size and the relationship with the host (Fig. 2d), one can clearly distinguish a correlation between small genomes/animal-associated bacteria, and big genomes/plant-associated bacteria. On the other hand, drawing the ellipses corresponding to the categories of host-association into the regulators' CA plot, a differential association of the bacteria with the different families of regulators is revealed: whereas animal-associated

bacteria are mainly related to LacI, DeoR and Xre, a group of regulators that can be globally associated with carbon metabolism (Table 1), plant-associated bacteria show a tendency towards Fur, Crp and LuxR (Fig. 4d).

The first important point that can be drawn from these data is that what actually interferes in genome size and in the nature of HTH-regulators is not the kind of the association, i.e. commensalism vs. parasitism, but the host to which the bacteria are associated. In fact, all animal-associated bacteria present the same patterns of distribution, as opposed to those of plant-associated bacteria, regardless of whether the associations are pathogenic or symbiotic. This suggests that pathogenicity has evolved many times from symbiosis: virulence factors are positively selected but the hosts are maintained. This results in the preservation of certain functional and biochemical features, such as the genome size and the need for certain transcription regulators, among the organisms that share the same range of hosts, despite the nature of their association. Actually, this similarity in the genotype of pathogenic and symbiotic bacteria has been noted before (Day & Maurelli, 2006; Dethlefsen *et al.*, 2007), supporting our approach to this matter. The genetic machinery and the regulatory mechanisms that drive bacterial interactions with eukaryotes seem to be similar and, apparently, in a first approach the symbiotic host actually suffers some damage by its microsymbiont partner, reinforcing the idea that despite the development of the interaction, the initial infection mechanisms permitting the colonization of a specific ecological niche are similar (Hentschel *et al.*, 2000).

The division between animal and plant-associated bacteria regulators likely relies on two different aspects that differentiate their hosts: the differences in the immune system and the molecular composition of the tissues. Regarding their immune capacity, animals and plants present striking differences related to their different lifestyles. As sessile organisms, plants cannot move to avoid environmental challenges. They also lack circulating immune cells, as well as phagocytic processes and immunoglobins. Animals, on the other hand, are equipped with such features and are mobile. Their metabolic rate is higher, and so is their need to nourish, which in turn exposes them to a large variety and number of microorganisms and to a higher risk of DNA oxidative damage. Moreover, co-evolution with specialized parasites and repeated genomic duplications, which result in higher genomic instability, also contributed to the development of the acquired immune system in vertebrates (Rolff, 2007). In fact, whereas both animals and plants possess a nonspecific first line of defence against microbial invasion – the innate immunity – only animals have an adaptive immune system, a specific, fast and highly efficient response to infection by pathogens (Nurnberger *et al.*, 2004; Zipfel & Felix, 2005; Iriti & Faoro, 2007).

Therefore, it is not surprising that there are striking functional differences in the bacteria associated with animals and plants, as they have to use different strategies to overcome their hosts' different immune systems. In particular, the fact that specialized cells from the animal immune system can be transported throughout the body in a circulating fluid, together with the increased specificity of this system when compared with the plant systems, may be responsible for imposing a high pressure on the invading microorganisms, affecting particularly their replication time, and therefore favouring the microorganisms carrying a small genome. In plants this pressure is likely relieved, allowing the maintenance of large genomes.

On the other hand, it is known that upon microbial infection, sequestration of iron and possibly other metals works as a nonspecific host defence strategy. However, in plants, besides those mentioned above, iron seems to assume a fundamental role in the host defence strategies by establishing a connection between the formation of cell wall appositions, the production of defensive proteins and the oxidative burst (Liu *et al.*, 2006). This increased importance of iron in plant invasion is likely responsible for the presence of tight selective pressures related to iron homeostasis in plant-associated organisms, justifying their correlation with regulators responsible for mediating transcription of genes related to metal equilibrium. Besides Fur, a known regulator controlling iron uptake, storage and metabolism (Table 1), plant-associated bacteria are also correlated with Crp, which has been shown to act cooperatively or antagonistically with Fur (Zhang *et al.*, 2005), as stated above.

The second factor pointed above as able to discriminate the animal- from the plant-associated bacteria was the molecular composition of the hosts. Once the immune system is overcome, one can think of the host tissues as a nutritive pool with which the associated bacteria can meet their nutritional needs. It is known that the stability of the surrounding environment and the possibility to use the host genomic machinery promotes the loss of nonessential genes from the bacteria, in a long-term process known as evolution by reduction (Coenye *et al.*, 2005, and references therein). Therefore, genes involved in the biosynthesis of products present in the tissues of a specific host would be selectively lost, while others would be maintained. In fact, such situations have been already described in the literature. For instance, the genome of the aphid-associated *Buchnera aphidicola* reveals the loss of biosynthetic pathways for some amino acids that the bacteria could get from the host, but the maintenance of others that could not be found in the phloem, the main nutrient source of the aphids (Lawrence, 2006, and references therein). Therefore, the different composition of plant and animal tissues leads towards a differential loss of biosynthetic pathways and consequently of their underlying regulators in their associated bacteria. If the

lost regulators are different in plant- and animal-associated bacteria, so are the positively selected ones, therefore contributing to the division of the organisms in relation to regulators' usage.

The nonresponsiveness of TFs: factors that do not seem to affect their distribution

Some of the factors studied do not seem to affect the distribution of the considered regulators in bacteria, at least with the statistical approach used, with the qualitative classes selected to qualify the organisms and/or with the available information concerning that factor. Oxygen requirement, despite the influence that it must have on the metabolic rates and energetic outcome, does not seem to have a clearcut effect on the distribution of regulators, as all classes have a broad distribution when plotted into regulators CA, except for a tendency of anaerobic organisms towards stress/metal homeostasis regulators (see Supporting Information, Fig. S1). Although some trends can be seen when the phylogenetic groups are plotted into the regulators' CA, the results merely reflect the main characteristics of each class (see Fig. S2). For instance, both *Bacilli* and *Mollicutes* present a correlation with LacI, DeoR and Xre, which overlap with the regulators correlated with animal-associated bacteria. Regarding *Bacilli*, although some organisms inhabit soil and others have multiple environments, most of them are animal-associated, including symbionts and pathogens. *Mollicutes*, in turn, constitute a class composed of strict animal pathogens. Therefore, the overlap between animal-associated bacterial regulators and *Bacilli* and *Mollicutes* regulators is explained by the lifestyle of the organisms belonging to these two classes. On the other hand, *Deltaproteobacteria* for instance have a tendency towards Fur, Crp and LuxR, which reflects the fact that the habitat of these organisms includes the soil, even though at least some of them do have multiple habitats.

In contrast, classes such as *Actinobacteria*, which do not have a dominant trait among the ones shown to influence the distribution of regulators (e.g. *Actinobacteria* include free-living bacteria living in the soil and/or in symbiosis with plants, or facultative animal pathogens, and none of these classes is dominant over the others) present a scattered distribution.

Finally, some cases can be seen that do not reflect the patterns defined above. *Epsilonbacteria* and *Spirochaetes* are mainly constituted of animal-associated bacteria, but nevertheless they present a correlation with Fur, Crp and LuxR. This can result from the small number of organisms representing these classes. Although in the overall sample of animal-associated bacteria they become diluted into the majority, and do not influence the general main tendency,

they reveal unique traits when analysed individually. This also suggests that other factors, besides the ones that were considered in this statistical approach, must be influencing the distribution of the regulators; further analyses are needed to fully understand the observed distribution.

Concluding remarks

Ecological adaptability, i.e. the capacity to perceive and to fit to transient oscillations in the surrounding environment, is crucial for survival and growth. Although the observed responses are short-term, the tools that support this resilience are part of the genetic pool of each individual. Among such tools are the one-component TFs, able to activate or repress the expression of certain genes responding to external or internal stimuli. It has been known since Darwin that evolution selects the most adapted structural features. The main question we intended to address while reviewing TFs was whether evolutionary forces were able to reshape the deep genomic features in such a way that could influence adaptability or, in other words, whether the range of conditions a certain organism can withstand is determined by the selective forces it has been exposed to through geological time. We have addressed the diversity of HTH bacterial-regulators using a statistical approach, and the outcome is consistent with the idea of multiple selective forces modulating the number and kind of regulators present in a given genome. So, recalling Haeckel, it seems that not only ontogeny, but also adaptability, recapitulates phylogeny.

If studies focused on one organism illustrate the individual dynamics of living species, comparative studies across several organisms or their genomes contribute to the knowledge on conserved mechanisms, and on the evolution of certain functionalities or evolutionary pressures that contributed to shape-contrasting characteristics. The constant increase in the number of sequenced genomes provides researchers nowadays with large amounts of valuable data, but also requires that they possess the skills and the tools to explore it in an integrated way. Our statistical approach, combining frequency data on regulator families and ecological/individual characteristics through component and multiple component analysis, represents a powerful and appropriate tool to discern trends in the massive amount of data accumulating in the databases. Furthermore, our phylogenomic overview is focused on the delicate border that separates ecology from genomics: we have given examples of how this interface can be explored analysing several different organisms simultaneously. From our results, we believe that the statistical approach employed is worthy of extension to other protein families as a way to unveil broad range phylogenomic patterns.

References

Ahmed M, Borsch CM, Taylor SS, Vazquez-Laslop N & Neyfakh AA (1994) A protein that activates expression of a multidrug efflux transporter upon binding the transporter substrates. *J Biol Chem* **269**: 28506–28513.

Alekshun MN & Levy SB (1999) The mar regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol* **7**: 410–413.

Aravind L, Anantharaman V, Balaji S, Babu MM & Iyer LM (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**: 231–262.

Babu MM & Aravind L (2006) Adaptive evolution by optimizing expression levels in different environments. *Trends Microbiol* **14**: 11–14.

Berish SA, Subbarao S, Chen CY, Trees DL & Morse SA (1993) Identification and cloning of a fur homolog from *Neisseria gonorrhoeae*. *Infect Immun* **61**: 4599–4606.

Bertani I, Kojic M & Venturi V (2001) Regulation of the *p*-hydroxybenzoic acid hydroxylase gene (pobA) in plant-growth-promoting *Pseudomonas putida* WCS358. *Microbiology* **147**: 1611–1620.

Bertin PN, Medigue C & Normand P (2008) Advances in environmental genomics: towards an integrated view of micro-organisms and ecosystems. *Microbiology* **154**: 347–359.

Campos E, Baldoma L, Aguilar J & Badia J (2004) Regulation of expression of the divergent ulaG and ulaABCDEF operons involved in LaAscorbate dissimilation in *Escherichia coli*. *J Bacteriol* **186**: 1720–1728.

Cavalier-Smith T (2005) Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Ann Bot (London)* **95**: 147–175.

Charoenpanich J, Tani A, Moriwaki N, Kimbara K & Kawai F (2006) Dual regulation of a polyethylene glycol degradative operon by AraC-type and GalR-type regulators in *Sphingopyxis macrogoltabida* strain 103. *Microbiology* **152**: 3025–3034.

Chatterjee J, Miyamoto CM, Zouzoulas A, Lang BF, Skouris N & Meighen EA (2002) MetR and CRP bind to the *Vibrio harveyi* lux promoters and regulate luminescence. *Mol Microbiol* **46**: 101–111.

Chattopadhyay MK (2006) Mechanism of bacterial adaptation to low temperature. *J Biosci* **31**: 157–165.

Chen Y & Rosen BP (1997) Metalloregulatory properties of the ArsD repressor. *J Biol Chem* **272**: 14257–14262.

Coenye T, Gevers D, Van de Peer Y, Vandamme P & Swings J (2005) Towards a prokaryotic genomic taxonomy. *FEMS Microbiol Rev* **29**: 147–167.

D'Amico S, Collins T, Marx JC, Feller G & Gerday C (2006) Psychrophilic microorganisms: challenges for life. *EMBO Rep* **7**: 385–389.

Dattelbaum JD, Lockatell CV, Johnson DE & Mobley HL (2003) UreR, the transcriptional activator of the *Proteus mirabilis* urease gene cluster is required for urease activity and virulence in experimental urinary tract infections. *Infect Immun* **71**: 1026–1030.

Day WA & Maurelli AT (2006) Black holes and antivirulence genes: selection for gene loss as part of the evolution of bacterial pathogens. *Evolution of Microbial Pathogens* (Seifert HS & DiRita VJ, eds), pp. 109–122. ASM Press, Washington, DC.

Deming JW (2002) Psychrophiles and polar regions. *Curr Opin Microbiol* **5**: 301–309.

Dethlefsen L, McFall-Ngai M & Relman DA (2007) An ecological and evolutionary perspective on human–microbe mutualism and disease. *Nature* **449**: 811–818.

Egan SM (2002) Growing repertoire of AraC/XylS activators. *J Bacteriol* **184**: 5529–5532.

Egland PG & Harwood CS (1999) BadR, a new MarR family member, regulates anaerobic benzoate degradation by *Rhodopseudomonas palustris* in concert with AadR, an Fnr family member. *J Bacteriol* **181**: 2102–2109.

Escolar L, Perez-Martin J & de Lorenzo V (1999) Opening the iron box: transcriptional metalloregulation by the Fur protein. *J Bacteriol* **181**: 6223–6229.

Ferluga S, Bigirimana J, Höfte M & Venturi V (2007) A LuxR homologue of *Xanthomonas oryzae* pv. *oryzae* is required for optimal rice virulence. *Mol Plant Pathol* **8**: 529–538.

Ferrandez A, Minambres B, Garcia B, Olivera ER, Luengo JM, Garcia JL & Diaz E (1998) Catabolism of phenylacetic acid in *Escherichia coli*. Characterization of a new aerobic hybrid pathway. *J Biol Chem* **273**: 25974–25986.

Frota CC, Papavinasasundaram KG, Davis EO & Colston MJ (2004) The AraC family transcriptional regulator Rv1931c plays a role in the virulence of *Mycobacterium tuberculosis*. *Infect Immun* **72**: 5483–5486.

Fuqua C & Greenberg EP (1998) Self perception in bacteria: quorum sensing with acylated homoserine lactones. *Curr Opin Microbiol* **1**: 183–189.

Fuqua C, Winans SC & Greenberg EP (1996) Census and consensus in bacterial ecosystems: the LuxR–LuxI family of quorum-sensing transcriptional regulators. *Annu Rev Microbiol* **50**: 727–751.

Gallegos MT, Schleif R, Bairoch A, Hofmann K & Ramos JL (1997) Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol R* **61**: 393–410.

Galvao TC & de Lorenzo V (2006) Transcriptional regulators a la carte: engineering new effector specificities in bacterial regulatory proteins. *Curr Opin Biotech* **17**: 34–42.

Godsey MH, Zheleznova Heldwein EE & Brennan RG (2002) Structural biology of bacterial multidrug resistance gene regulators. *J Biol Chem* **277**: 40169–40172.

Gold B, Rodriguez GM, Marras SA, Pentecost M & Smith I (2001) The *Mycobacterium tuberculosis* IdeR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages. *Mol Microbiol* **42**: 851–865.

Guazzaroni ME, Krell T, Felipe A, Ruiz R, Meng C, Zhang X, Gallegos MT & Ramos JL (2005) The multidrug efflux regulator TtgV recognizes a wide range of structurally different effectors in solution and complexed with target DNA: evidence from isothermal titration calorimetry. *J Biol Chem* **280**: 20887–20893.

Haine V, Sinon A, Van Steen F, Rousseau S, Dozot M, Lestrate P, Lambert C, Letesson JJ & De Bolle X (2005) Systematic targeted mutagenesis of *Brucella melitensis* 16M reveals a major role for GntR regulators in the control of virulence. *Infect Immun* **73**: 5578–5586.

Haine V, Dozot M, Dornand J, Letesson JJ & De Bolle X (2006) NnrA is required for full virulence and regulates several *Brucella melitensis* denitrification genes. *J Bacteriol* **188**: 1615–1619.

Hanssler E, Muller T, Jessberger N, Volzke A, Plassmeier J, Kalinowski J, Kramer R & Burkovski A (2007) FarR, a putative regulator of amino acid metabolism in *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol* **76**: 625–632.

Hantke K (1981) Regulation of ferric iron transport in *Escherichia coli* K12: isolation of a constitutive mutant. *Mol Gen Genet* **182**: 288–292.

Hentschel U, Steinert M & Hacker J (2000) Common molecular mechanisms of symbiosis and pathogenesis. *Trends Microbiol* **8**: 226–231.

Hobman JL, Wilkie J & Brown NL (2005) A design for life: prokaryotic metal-binding MerR family regulators. *Biometals* **18**: 429–436.

Hoskisson PA, Rigali S, Fowler K, Findlay KC & Buttner MJ (2006) DevA, a GntR-like transcriptional regulator required for development in *Streptomyces coelicolor*. *J Bacteriol* **188**: 5014–5023.

Iriti M & Faoro F (2007) Review of innate and specific immunity in plants and animals. *Mycopathologia* **164**: 57–64.

Joshi MV, Bignell DR, Johnson EG, Sparks JP, Gibson DM & Loria R (2007) The AraC/XylS regulator TxtR modulates thaxtomin biosynthesis and virulence in *Streptomyces scabies*. *Mol Microbiol* **66**: 633–642.

Koch B, Liljefors T, Persson T, Nielsen J, Kjelleberg S & Givskov M (2005) The LuxR receptor: the sites of interaction with quorum-sensing signals and inhibitors. *Microbiology* **151**: 3589–3602.

Kohn WD, Mant CT & Hodges RS (1997) Alpha-helical protein assembly motifs. *J Biol Chem* **272**: 2583–2586.

Korner H, Sofia HJ & Zumft WG (2003) Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol Rev* **27**: 559–592.

Kotelnikova EA, Makeev VJ & Gelfand MS (2005) Evolution of transcription factor DNA binding sites. *Gene* **347**: 255–263.

Kreft J & Vazquez-Boland JA (2001) Regulation of virulence genes in *Listeria*. *Int J Med Microbiol* **291**: 145–157.

Kuroda M, Hayashi H & Ohta T (1999) Chromosome-determined zinc-responsible operon czr in *Staphylococcus aureus* strain 912. *Microbiol Immunol* **43**: 115–125.

Lawrence JG (2006) Studying evolution using genome sequence data. *Evolution of Microbial Pathogens* (Seifert HS & DiRita VJ, eds), pp. 11–33. ASM Press, Washington, DC.

Letunic I, Copley RR, Pils B, Pinkert S, Schultz J & Bork P (2006) SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* **34**: D257–D260.

Liu G, Greenshields DL, Sammynaiken R, Hirji RN, Selvaraj G & Wei Y (2006) Targeted alterations in iron homeostasis underlie plant defense responses. *J Cell Sci* **120**: 596–605.

Lopez O, Morera C, Miranda-Rios J, Girard L, Romero D & Soberon M (2001) Regulation of gene expression in response to oxygen in *Rhizobium etli*: role of FnrN in fixNOQP expression and in symbiotic nitrogen fixation. *J Bacteriol* **183**: 6999–7006.

Lozada-Chavez I, Janga SC & Collado-Vides J (2006) Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res* **34**: 3434–3445.

Martinez-Antonio A & Collado-Vides J (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol* **6**: 482–489.

Mole BM, Baltrus DA, Dangl JL & Grant SR (2007) Global virulence regulation networks in phytopathogenic bacteria. *Trends in Microbiol* **15**: 363–371.

Munson GP, Holcomb LG & Scott JR (2001) Novel group of virulence activators within the AraC family that are not restricted to upstream binding sites. *Infect Immun* **69**: 186–193.

Nurnberger T, Brunner F, Kemmerling B & Piater L (2004) Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol Rev* **198**: 249–266.

Ofria C, Adami C & Collier TC (2003) Selective pressures on genomes in molecular evolution. *J Theor Biol* **222**: 477–483.

Oram DM, Avdalovic A & Holmes RK (2004) Analysis of genes that encode DtxR-like transcriptional regulators in pathogenic and saprophytic corynebacterial species. *Infect Immun* **72**: 1885–1895.

Outten FW, Outten CE & O'Halloran TV (2000) Metalloregulatory systems at the interface between bacterial metal homeostasis and resistance. *Bacterial Stress Responses* (Storz G & Hengge-Aronis R, eds), pp. 145–157. ASM Press, Washington, DC.

Perrenoud A & Sauer U (2005) Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*. *J Bacteriol* **187**: 3171–3179.

Pomposiello PJ & Demple B (2001) Redox-operated genetic switches: the SoxR and OxyR transcription factors. *Trends Biotechnol* **19**: 109–114.

Providenti MA & Wyndham RC (2001) Identification and functional characterization of CbaR, a MarR-like modulator of the cbaABC-encoded chlorobenzoate catabolism pathway. *Appl Environ Microbiol* **67**: 3530–3541.

Pym AS, Domenech P, Honore N, Song J, Deretic V & Cole ST (2001) Regulation of catalase-peroxidase (KatG) expression, isoniazid sensitivity and virulence by furA of *Mycobacterium tuberculosis*. *Mol Microbiol* **40**: 879–889.

Ranea JA (2006) Genome evolution: micro(be)-economics. *Heredity* **96**: 337–338.

Rasmussen TB & Givskov M (2006) Quorum-sensing inhibitors as anti-pathogenic drugs. *Int J Med Microbiol* **296**: 149–161.

Rea RB, Gahan CG & Hill C (2004) Disruption of putative regulatory loci in *Listeria monocytogenes* demonstrates a significant role for Fur and PerR in virulence. *Infect Immun* **72**: 717–727.

Ren J, Sainsbury S, Combs SE, Capper RG, Jordan PW, Berrow NS, Stammers DK, Saunders NJ & Owens RJ (2007) The structure and transcriptional analysis of a global regulator from *Neisseria meningitidis*. *J Biol Chem* **282**: 14655–14664.

Roberts RJ (2004) Identifying protein function – a call for community action. *PLOS Biol* **2**: 293.

Rodriguez GM, Voskuil MI, Gold B, Schoolnik GK & Smith I (2002) ideR, An essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun* **70**: 3371–3381.

Rolff J (2007) Why did the acquired immune system of vertebrates evolve? *Dev Comp Immunol* **31**: 476–482.

Rosinski JA & Atchley WR (1999) Molecular evolution of helix-turn-helix proteins. *J Mol Evol* **49**: 301–309.

Rouquette C, Harmon JB & Shafer WM (1999) Induction of the mtrCDE-encoded efflux pump system of *Neisseria gonorrhoeae* requires MtrA, an AraC-like protein. *Mol Microbiol* **33**: 651–658.

Santos C, Vieira J, Tavares F, Benson D, Tisa L, Berry A, Moradas-Ferreira P & Normand P (2008) On the nature of fur evolution: a phylogenetic approach in *Actinobacteria*. *BMC Evol Biol* **8**: 185.

Shevchik VE & Hugouvieux-Cotte-Pattat N (2003) PaeX, a second pectin acetylesterase of *Erwinia chrysanthemi* 3937. *J Bacteriol* **185**: 3091–3100.

Slobodkin AI (2005) Thermophilic microbial metal reduction. *Mikrobiologiia* **74**: 581–595.

Smidt H, Van Leest M, van der Oost J & de Vos WM (2000) Transcriptional regulation of the cpr gene cluster in ortho-chlorophenol-respiring *Desulfitobacterium dehalogenans*. *J Bacteriol* **182**: 5683–5691.

Soutourina O, Kolb A, Krin E, Laurent-Winter C, Rimsky S, Danchin A & Bertin P (1999) Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the flhDC master operon. *J Bacteriol* **181**: 7500–7508.

Thomson NR, Crow MA, McGowan SJ, Cox A & Salmond GP (2000) Biosynthesis of carbapenem antibiotic and prodigiosin

pigment in *Serratia* is under quorum sensing control. *Mol Microbiol* **36**: 539–556.

Tolmasky ME, Wertheimer AM, Actis LA & Crosa JH (1994) Characterization of the *Vibrio anguillarum* fur gene: role in regulation of expression of the FatA outer membrane protein and catechols. *J Bacteriol* **176**: 213–220.

Trindade MI, Abratt VR & Reid SJ (2003) Induction of sucrose utilization genes from *Bifidobacterium lactis* by sucrose and raffinose. *Appl Environ Microbiol* **69**: 24–32.

Tropel D & van der Meer JR (2004) Bacterial transcriptional regulators for degradation pathways of aromatic compounds. *Microbiol Mol Biol Rev* **68**: 474–500.

Tsoi TV, Plotnikova EG, Cole JR, Guerin WF, Bagdasarian M & Tiedje JM (1999) Cloning, expression, and nucleotide sequence of the *Pseudomonas aeruginosa* 142 ohb genes coding for oxygenolytic ortho dehalogenation of halobenzoates. *Appl Environ Microbiol* **65**: 2151–2162.

Ulrich LE, Koonin EV & Zhulin IB (2005) One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol* **13**: 52–56.

van Delden C, Comte R & Bally AM (2001) Stringent response activates quorum sensing and modulates cell density-dependent gene expression in *Pseudomonas aeruginosa*. *J Bacteriol* **183**: 5376–5384.

Van Houdt R, Givskov M & Michiels CW (2007a) Quorum sensing in *Serratia*. *FEMS Microbiol Rev* **31**: 407–424.

Van Houdt R, Moons P, Aertsen A, Jansen A, Vanoirbeek K, Daykin M, Williams P & Michiels CW (2007b) Characterization of a luxI/luxR-type quorum sensing system and *N*-acyl-homoserine lactone-dependent regulation of exoenzyme and antibacterial component production in *Serratia plymuthica* RVH1. *Res Microbiol* **158**: 150–158.

von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ, Ward N & Bork P (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**: 1126–1130.

Ward JH (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* **58**: 236–244.

Wilmes P & Bond PL (2006) Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol* **14**: 92–97.

Yokoyama K, Ishijima SA, Clowney L, Koike H, Aramaki H, Tanaka C, Makino K & Suzuki M (2006) Feast/famine regulatory proteins (FFRPs): *Escherichia coli* Lrp, AsnC and related archaeal transcription factors. *FEMS Microbiol Rev* **30**: 89–108.

Zhang Z, Gosset G, Barabote R, Gonzalez CS, Cuevas WA & Saier MH Jr (2005) Functional interactions between the carbon and iron utilization regulators, Crp and Fur, in *Escherichia coli*. *J Bacteriol* **187**: 980–990.

Zheng H, Zhong Z, Lai X, Chen WX, Li S & Zhu J (2006) A LuxR/LuxI-type quorum-sensing system in a plant bacterium, *Mesorhizobium tianshanense*, controls symbiotic nodulation. *J Bacteriol* **188**: 1943–1949.

Zhu J, Miller MB, Vance RE, Dziejman M, Bassler BL & Mekalanos JJ (2002) Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *P Natl Acad Sci USA* **99**: 3129–3134.

Zipfel C & Felix G (2005) Plants and animals: a different taste for microbes? *Curr Opin Plant Biol* **8**: 353–360.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Frequency table of the regulators across the different genomes and the qualitative classification of the organisms involved.

Fig. S1. Distribution of the organisms according to their oxygen requirement, based on the classification parameters present in the Genome Project for each genome available in NCBI.

Fig. S2. Distribution of the organisms according to their phylogenetic class, based on the classification parameters in the Genome Project for each genome available in NCBI.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.