JB Accepts, published online ahead of print on 11 September 2009
J. Bacteriol. doi:10.1128/JB.00976-09

Identification of functional LsrB-like autoinducer-2 receptors

Catarina S. Pereira[1,2], Anna K. de Regt[3], Patrícia H. Brito[1], Stephen T. Miller[3 *], Karina B. Xavier[1, 2, *]

[1]Instituto Gulbenkian de Ciência. 2781-901 Oeiras, Portugal.

[2]Instituto de Tecnologia Química e Biológica. 2781-901 Oeiras, Portugal.

[3]Department of Chemistry and Biochemistry, Swarthmore College, Swarthmore, PA, 19081 USA

[*]To whom correspondence should be addressed

Karina B. Xavier, email kxavier@igc.gulbenkian.pt; Tel. (+351) 21 446 4655; Fax (+351) 21 440 7970.

Stephen T. Miller, email smiller1@swarthmore.edu; Tel. (610) 957-6063; Fax (610) 328-7355

Running title: LsrB-like Autoinducer-2 receptors

Key words: AI-2, *luxS,* quorum sensing, signal transduction, lateral gene transfer

ABSTRACT

Although a variety of bacterial species have been reported to use the inter-species communication signal autoinducer-2 (AI-2) to regulate multiple behaviors, the molecular mechanisms of AI-2 recognition and signal transduction remain poorly understood. To date, two types of AI-2 receptors have been identified: LuxP, present in *Vibrio* spp, and LsrB, first identified in *Salmonella typhimurium*. In *S. typhimurium*, LsrB is the ligand binding protein of a transport system that enables internalization of AI-2. Here, using both sequence analysis and structure prediction, we establish a set of criteria for identifying functional AI-2 receptors. We test our predictions experimentally, assaying key species for their ability to import AI-2 *in vivo* and test their LsrB orthologs for AI-2 binding *in vitro*. Using these experimental approaches, we were able to identify AI-2 receptors in organisms belonging to phylogenetically distinct families such as Enterobacteriaceae, Rhizobiaceae, and Bacillaceae. Phylogenetic analysis of LsrB orthologs indicates that this pattern could result from one single origin of the functional LsrB gene in a γ-proteobacterium, suggesting possible posterior independent events of lateral gene transfer to the α-proteobacteria and Firmicutes. Finally, we used mutagenesis to show that two AI-2 interacting residues are essential for AI-2 binding ability. These two residues are conserved in the binding site of all the functional AI-2-binding proteins but not in the non-AI-2-binding orthologs. Together, these results strongly support our ability to identify functional LsrB-type AI-2 receptors, an important step in investigations of this inter-species signal.

INTRODUCTION

Autoinducer-2 (AI-2) is a small molecule produced and secreted by a large number of bacterial species belonging to very widespread branches within the Bacteria kingdom (15, 46, 64). AI-2 or its synthase, LuxS, has been implicated in the regulation of many bacterial behaviors including biofilm formation, virulence, competence, and production of secondary metabolites like antibiotics (17, 60, 64). While in some cases AI-2 is clearly acting through a canonical quorum sensing mechanism (61), in others a role in central metabolism has been proposed (62). One of the obstacles to understanding the function of AI-2 in any given species is a lack of knowledge of the molecular mechanisms of AI-2 recognition, signal transduction, and/or processing.

Undoubtedly, one of the major difficulties in identifying AI-2 receptors is the complexity of the chemistry of this signal molecule. The product of the reaction catalyzed by LuxS is 4,5-dihydroxy-2,3-pentadione (DPD) which, in solution, spontaneously re-arranges into a variety of chemically distinct forms collectively called AI-2 (31, 46). We have shown that these forms are in equilibrium and can thus interconvert and that the availability of the different forms of AI-2 is highly dependent on the chemistry of the environment (31). Additionally, different organisms recognize distinct forms of this molecule (12).

So far, two types of AI-2 receptors have been identified and are classified by their ability to bind chemically distinct DPD derivatives, the LuxP- and LsrB-type of receptors characterized first in *Vibrio harveyi* and *Salmonella typhimurium*, respectively (12, 31). The crystal structure of the *V. harveyi* LuxP-AI-2 complex revealed that the ligand recognized by this receptor is a furanosyl borate diester (12), a cyclic form of DPD bound to borate, while crystal structures of the LsrB-AI-2 complexes from *S.*

typhimurium and *Sinorhizobium meliloti* show that these species recognize a DPD adduct that does not contain boron and has different stereochemistry ((2*R*,4*S*)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran (*R*-THMF)) (31, 37). The structures of the LsrB-type receptors bound to AI-2 further showed that six residues were responsible for hydrogen-bonding with AI-2 and that these residues were completely conserved between the two species (31, 37). These residues are distinct from those in the LuxP AI-2 binding site, contributing to the specificity of each receptor for the form of AI-2 recognized by a given species.

LuxP is a periplasmic binding protein (PBP) that, upon binding to AI-2, modulates the activity of a membrane sensor histidine kinase, LuxQ. Together, LuxPQ regulate a signal transduction cascade which controls the AI-2 quorum sensing regulon in organisms belonging to the Vibrionales like *V. harveyi, Vibrio cholerae* and *Vibrio anguillarum* (5, 13, 32, 33); to date, however, LuxP-type receptors have not been found outside of the Vibrionales.

The LsrB-type receptors also belong to the large family of PBPs but have a low homology to LuxP (the sequence identity between the *V. harveyi* LuxP and the *S. typhimurium* LsrB AI-2 receptors is only approximately 11 %). The function of the LsrB protein has been characterized in the two closely related enteric bacteria *S. typhimurium* (56, 57) and *Escherichia coli* (65), the plant symbiont *S. meliloti* (37), and the oral pathogen *Aggregatibacter* (*Actinobacillus*) *actinomycetemcomitans* (48). In all these organisms it is thought that LsrB acts as the substrate binding protein of an ABC (ATP-binding cassette) transport system responsible for AI-2 internalization. Due to the homology with other ABC transport systems, it is predicted that the Lsr transporter is composed of LsrB, two transmembrane proteins (LsrC and LsrD) which form a channel, and a cytoplasmic protein (LsrA) that contains an ABC-binding motif and is thought to

be responsible for ATP hydrolysis during transport. Once inside the cell, AI-2 is phosphorylated by the kinase LsrK and further processed by the enzymes LsrG and LsrF (56, 66). The genes encoding these proteins (with the exception of LsrK) are all in the same operon, which is regulated by the repressor LsrR. In the absence of phospho-AI-2 (P-AI-2), LsrR represses the transcription of the *lsr* operon; however, when AI-2 is internalized and phosphorylated by LsrK, P-AI-2 binds LsrR causing the de-repression of the operon. Thus, increased expression of the Lsr system leads to increased AI-2 import, resulting in a rapid depletion of AI-2 from the extracellular medium.

It does not appear that AI-2 taken up by this system is used as a carbon source, since cultures of *S. typhimurium* and *S. meliloti* were unable to grow when AI-2 was used as the sole carbon source (37, 57). Rather, AI-2 removal via the Lsr system enables these organisms to terminate their own AI-2 signaling system and to regulate the AI-2-dependent gene expression of other organisms in the vicinity. Thus, in cultures composed of different species, bacteria with a functional Lsr system are capable of interfering with AI-2-mediated group behaviors of the other species (63).

Recently, two studies have undertaken database sequence analysis to identify LsrB orthologs (41, 50). These studies showed that orthologs to the Lsr system are not broadly conserved across the Bacteria Kingdom, while identifying hypothetical LsrB receptors in some organisms belonging to very distinct families such as Enterobacteriaceae, Pasteurellaceae, Rhizobiaceae, Rhodobacteraceae and Bacillaceae.

Here, we expand upon the previous bioinformatic studies (41, 50) with additional analysis, based not only on sequence but also on structure prediction, that allow us to establish a set of criteria for predicting which orthologs of LsrB are functional AI-2 receptors. We then present experimental evidence that confirms a set of these

predictions and demonstrates the presence of functional AI-2 receptors in the Enterobacteriaceae, Rhizobiaceae and Bacillaceae families.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The strains used are listed in Table 1. Bacteria from the Enterobacteriaceae family *(E. coli* MG1655 and UTI89 UPEC) and the Bacillaceae family (*Bacillus cereus* ATCC 10987 and *Bacillus anthracis* Sterne 34F2 vaccine strain) were grown in Luria-Bertani (LB) medium with shaking at 37°C. The bacteria from the Rhizobiaceae family (*S. meliloti* Rm1021, *Agrobacterium tumefaciens* C58, *Rhizobium etli* CFN42 and *Rhizobium leguminosarum* bv viciae 3841) were cultured with shaking at 30ºC in their optimal cultured medium respectively: LBMC (LB supplemented with 2.5 mM $MgSO_4$ and 2.5 mM $CaCl_2$), LB, YEM (10 g L^{-1} mannitol, 0.5 g L^{-1} yeast extract, 0.2 g L^{-1} $MgSO_4.7H_2O$ and 1 g L^{-1} NaCl) and TYC (5 g L^{-1} tryptone, 3 g L^{-1} yeast extract, 0.5 g L^{-1} $CaCl_2$).

Databases analysis. The KEGG SSDB (Kyoto Encyclopedia of Genes and Genomes - Sequence Similarity DataBase, http://www.genome.jp/kegg/ssdb/) was used to search for protein orthologs of LuxS and the Lsr operon from *S. typhimurium* LT2 in January 2009. This database provides amino acid sequence similarities between all protein-coding genes in the complete genomes in the GENES database and all possible pairwise genome comparisons are performed by the SSEARCH program (36) available at http://www.genome.jp/kegg/ssdb/). In this study, we have selected gene pairs that were best bidirectional hits and had a Smith-Waterman similarity score of at least 100. To be considered a best bidirectional hit, the relationship of gene x in genome A with gene y in genome B must be such that, when x is compared against all genes in genome B, y is

found as the top scoring and the reverse is also true. Pairs that met these criteria were scored as orthologs proteins.

Structure Prediction. All LsrB protein orthologs were submitted to the fold-recognition server PHYRE (22) for structure prediction. In the majority of cases, *S. typhimurium* LsrB was identified as one of the top ten fold templates and, thus, the server returned a structure-based sequence alignment between LsrB and the query sequence. Alignments were examined to determine if residues previously shown to form hydrogen bonds with *R*-THMF in *S. typhimurium* LsrB (K35, D116, D166, Q167, P220, and A222, (31)) were conserved in the predicted structure. For the one-third of group II orthologs where PHYRE did not return an alignment with LsrB, simple sequence alignments were calculated using NCBI-blastp (1, 18) and checked for conservation of the residues listed above. Such cases are noted in Table S1.

AI-2 activity in bacterial cultures. To monitor AI-2 activity in *E. coli* and *Bacillus* cell cultures during growth, overnight cultures were grown to saturation and diluted (1:100) into 25 ml of LB medium in 250 ml Erlenmeyer flasks. In Rhizobiaceae species, cultures in exponential phase were diluted to optical density (OD_{600}) = 1 into the appropriate medium with 80 μM chemically synthesized AI-2 (47, 66). In both cases, aliquots were collected at the indicated times and cell-free culture fluids were prepared by filtration of liquid cultures (51, 52) which were analyzed in duplicate for AI-2 activity using the *V. harveyi* BB170 bioluminescence reporter assay, as described previously (4, 5). AI-2 activity is reported as fold induction of light production compared with the background light obtained with the appropriate growth medium (as previously explained in (37)).

159 **Protein expression and purification.** The genes encoding LsrB orthologs in *R. etli*, *R.*
160 *leguminosarum, A. tumefaciens, E. coli* MG1655 and *E. coli* UTI89 were cloned from
161 genomic DNA into the plasmid pProEX HTb for expression as polyhistidine-tagged
162 proteins. The *B. anthracis* LsrB ortholog was cloned into the plasmid pET151/D-TOPO
163 using The Champion pET Directional TOPO Expression Kit (Invitrogen) for expression
164 as a polyhistidine-tagged fusion protein. N-terminal signal peptides for secretion, as
165 determined by the program SignalP 3.0 (6), were excluded from the constructs.
166 Plasmids were transformed into *E. coli* strains BL21 and FED101 (BL21 *luxS* null
167 mutant) and expression was induced with 0.1 mM Isopropyl beta-D-1-
168 thiogalactopyranoside (IPTG) when the cultures reached an OD_{595} of 0.9. The bacteria
169 were harvested after expressing for 5 hours at 22º. Pellets were resuspended in 50 mM
170 NaH_2PO_4, pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM DTT, 0.36 mg/mL
171 leupeptin, 0.36 mg/mL aprotinin, 0.36 mg/mL DNase and lysed using an M-110Y
172 Microfluidizer (Microfluidics). The lysate was centrifuged and the tagged protein
173 purified using Ni-NTA affinity chromatography (Qiagen). Protein was eluted from the
174 column using 50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 250 mM imidazole and then
175 buffer swapped using Sephadex-G25 agarose into 50 mM NaH_2PO_4, pH 8.0, 300 mM
176 NaCl, 1 mM DTT. Purified protein was concentrated to 10 mg mL^{-1}. The genes
177 encoding the *S. typhimurium* and *B. cereus* LsrB orthologs, were cloned in pGEX-4T1,
178 transformed, expressed and purified as described previously (31, 37). The primers used
179 for cloning the respective genes are listed in Table 2.

180 **AI-2 binding assay.** Proteins tested for AI-2 binding were denatured (70º, 10 min) to
181 release any bound ligand and pelleted (12). The *V. harveyi* strain BB170 was used to
182 test for the presence or absence of AI-2 in the resulting supernatants as previously
183 described (4, 5).

***B. anthracis* mutagenesis.** The mutations D171N and A227T were introduced into two separate *B. anthracis*/pET151 constructs using the QuikChange Lightning Site-Directed Mutagenesis Kit (Stratagene). Primers used for creating the mutations are given in Table 2. The same kit was then used to create the double mutant D171N/A227T. The mutant proteins were expressed and purified as described above for the *B. anthracis* wild-type LsrB ortholog.

Phylogenetic Analyses. The evolutionary history of the *lsrB* gene was studied by analyzing the phylogenetic relationship of the functional orthologs identified in this study and contrasting it with the phylogeny of *rpoB* (RNA polymerase β-subunit). *rpoB* is generally accepted to provide a good representation of the phylogenetic relationships within Bacteria (11), as it provides a comparable phylogenetic resolution to that of 16S rRNA with the advantage of being a single-copy gene. To construct the organismal tree, the *rpoB* gene sequences from all the organisms in Table S1 and representative species of all major phyla of Bacteria were downloaded from the KEGG database and aligned with ClustalW (59) using the translated protein sequences. Alignments were carried out with default parameters and visually inspected in Molecular Evolutionary Genetics Analysis (MEGA), version 4 (58). Hypervariable regions with ambiguous alignment were excluded from analysis. The *lsrB* gene tree was made with all the sequences identified as functional *lsrB* orthologs (group I, Table 3) and the tree was inferred using maximum likelihood (ML) in PAUP* 4.0b10 (53) using heuristic searches, 10 random taxon-additions, and TBR branch-swapping. Mrbayes 3.1.2 (44) was used to infer branch support by running two simultaneous sets of four Markov chains for 1 million generations sampled every 100 generations. The distribution of the log likelihoods was used to evaluate the stationarity of this parameter and to determine burn-in values. Modeltest 3.7 (38) and MrModeltest 2.2 (34) were used to select the

best-fitting evolutionary models for phylogenetic analyses. The *rpoB* phylogeny was estimated with a total dataset of 83 species. This dataset was translated to amino acids and analyzed using Neighbor-Joining (45) with the Poisson correction distances (68) and a gamma distribution rate variation among sites. Nodal support was estimated with non-parametric bootstrap (1000 replicates). The *rpoB* trees were rooted with *Thermotoga maritima* (Thermotogales). These analyses were carried out in MEGA.

RESULTS

LsrB orthologs in completely sequenced bacterial genomes. To search for orthologs of LsrB we carried out a reciprocal best hit analysis against all 809 completely sequenced bacteria genomes present in the KEGG database as of January 2009 using the protein sequence of LsrB from *S. typhimurium* (STM4077). The reciprocal best hit strategy of sequence similarity comparisons has been employed previously for this type of studies because it allows the distinction between orthologs and paralogs (10). The organisms with proteins identified as orthologs are shown in Table 3 (KEGG proteins identities and E-values are provided in supplementary Table S1). Sorting these organisms in order of percentage of identity of the LsrB orthologs with the *S. typhimurium* AI-2 receptor clearly revealed two distinct groups of LsrB orthologs: a first group with high percentage identity (>60%, E-value bellow 1E-103), and a second group with percentage identity below 36% (E-value higher than 1E-44) which we termed group I and group II, respectively.

We then performed the reciprocal best hit analysis against all genomes using each LsrB protein sequence from group I as a reference (i.e. instead of the LsrB from *S. typhimurium*). In all cases, the only hits with greater than 57 % identity were the other

protein sequences included in group I from the first analysis. Thus, the group I orthologs are consistent regardless of the LsrB sequence used as reference.

The genomes of the organisms with LsrB orthologs were further analyzed to identify orthologs of the other proteins of the Lsr operon. As shown in Table 3, all the species of group I have orthologs of all the proteins in the Lsr operon (with the exception of LsrF, a putative AI-2 processing protein, in *Rhodobacter sphaeroides*), whereas none of the group II organisms have orthologs of the complete operon, lacking at least two proteins encoded by genes from this operon in all cases. LsrE was not included in this analysis because the protein seems to be exclusive to the *Salmonella* genus and a LsrE knockout mutant in *S. typhimurium* showed no phenotype related to the regulation of the *lsr* operon or AI-2 production (56, 57).

Reasoning that conservation of the residues that formed hydrogen bonds with AI-2 (31) would be crucial to LsrB function, we next used a fold recognition-based server to predict structures for the LsrB orthologs. The sequences of the LsrB orthologs were submitted to the PHYRE web server (22), which returned structure predictions and structure-based alignments based on each of the ten best scoring template PDB structures available in the PHYRE library. For all of the orthologs in group I and two-thirds of the orthologs in group II, the structure of *S. typhimurium* LsrB was returned as one of these top ten templates. The alignments with *S. typhimurium* LsrB were then examined to determine if residues previously shown to form hydrogen bonds with *R*-THMF in *S. typhimurium* LsrB (K35, D116, D166, Q167, P220, and A222, (31)) were predicted to be structurally conserved. Strikingly, as shown in the last column of Table 3, these six residues were completely conserved in all of the orthologs in group I and differed in at least two positions in all cases for group II. Residue D166 (numbering based on *S. typhimurium* LsrB) was not conserved in any of the group II orthologs, most

typically being replaced with an aspargine. The other most common substitution was A222T (a full listing of the non-conserved amino acids is given in Supplementary Material Table S1).

Based on these results, we hypothesize that the species in group I, which have >60% identity, orthologs to the proteins of the Lsr operon, and all six AI-2 binding site residues conserved, have functional LsrB-like AI-2 receptors, whereas group II proteins are likely to have a different function.

Profiles of AI-2 removal from extracellular medium. Previous studies in *S. typhimurium* (57), *E. coli* (65), *S. meliloti* (37), and *A.* actinomycetemcomitans (48)revealed that the *lsr* operon in these organisms encodes proteins involved in an ABC transport system that imports extracellular AI-2. Thus, in the presence of these organisms, AI-2 does not persist in the extracellular medium but is internalized by the cells and further modified. Our analysis, described above, indicated that all the organisms predicted to have functional LsrB receptors (group I, Table 3) also had orthologs to all the proteins in the Lsr operon. Thus, we predicted that the organisms in group I have a functional Lsr system for AI-2 internalization and that these organisms would rapidly remove AI-2 from culture fluids. In contrast, for organisms from group II, which lack orthologs to some of the proteins in the Lsr operon and presumably do not have a functional AI-2 transport system, we predicted that AI-2 would persist in the extracellular media. To test these predictions, we compared the profile of AI-2 removal of a set of organisms from groups I and II.

Our analysis revealed that almost all *E. coli* strains (11 out of 13) analyzed belong to group I. However, two *E. coli* strains (*E. coli* APEC and *E. coli* UT189 / UPEC) have LsrB orthologs with very low sequence identity and lack orthologs to several of the

proteins from the *lsr* operon and therefore are classified as members of group II (Table 3). We tested an *E. coli* strain (MG1655) from group I for AI-2 uptake and found, as had been previously shown (65), that this strain removed AI-2 from culture fluids (Fig. 1A). We then compared the AI-2 removal profile in *E. coli* UT189 / UPEC strain (from group II) with the profile from *E. coli* MG1655 strain and observed that, while *E. coli* MG1655 efficiently cleared AI-2 from culture fluids by 6 hours, the *E. coli* strain UT189 / UPEC strain cleared little, if any, AI-2 by 10 h (Fig. 1A). This supports our prediction that the uropathogenic strain UT189 / UPEC, though belonging to the same species as MG1655, is a member of group II and accordingly does not have a functional Lsr transport system for AI-2 uptake.

Like *E. coli* MG1655, two *Bacillus* strains, *cereus* (ATCC 1087) and *anthracis* (vaccine Sterne 34F2), have orthologs classified as group I. Putative AI-2 receptors have been identified in these species previously (41, 50), but not confirmed experimentally. We tested these strains for AI-2 removal and, as expected, they were able to completely remove AI-2 from culture fluids (Fig. 1B), supporting the premise that organisms in group I have functional AI-2 transporters.

To further test the premise that group II organisms are unable to incorporate AI-2, we compared AI-2 removal in organisms from the Rhizobiaceae family from group II (*R. etli, R. leguminosarum,* and *A. tumefaciens*) with AI-2 in the only Rhizobiaceae from group I (*S. meliloti*). None of the Rhizobiaceae in Table 3 has LuxS orthologs, and thus we expected that none of these species would produce AI-2. This was confirmed by the fact that cell-free culture fluids collected from these bacteria produced only low levels of bioluminescence induction in a *V. harveyi* BB170 bioassay (data not shown). However, as we have previously shown in the case of *S. meliloti*, non-AI-2 producing species can still be capable of taking up AI-2 produced synthetically or by other species

(37). Thus, in order to compare AI-2 removal profiles in these species, we cultured these bacteria to the same cell density (OD_{600} = 1), supplied chemically synthesized AI-2, and measured AI-2 activity in the culture fluids over time (Fig. 2). Over the time of the measurements, *S. meliloti* effectively removed the exogenously provided AI-2 while the other three species did not, supporting the prediction that the bacteria from group II (*R. etli, R. leguminosarum and A. tumefaciens*), and likely all group II species, do not have Lsr systems capable of taking up AI-2.

***In vitro* AI-2 binding to LsrB orthologs**. While the above results support our ability to identify species with functional AI-2 transporters, they do not directly show that the identified LsrB ortholog is responsible for AI-2 binding. In order to directly test for AI-2 binding ability, we cloned the LsrB orthologs from the same organisms tested in the previous section (three belonging to group I and four belonging to group II) and compared their ability to bind AI-2 with that of LsrB from *S. typhimurium*. The candidate proteins were overexpressed in both an *E. coli* strain that produces AI-2 and, as a negative control, in a *luxS* mutant *E. coli* strain that does not make AI-2. These proteins were then purified and tested for the ability to bind AI-2 using a previously developed assay (12) in which the protein is heat denatured to release any bound ligand. The denatured protein is then pelleted and the resulting supernatants are added to a reporter strain of *V. harveyi* that bioluminesces in response to AI-2. As shown in Figure 3, all three orthologs from group I (i.e. that have >60% identity, a complete set of orthologs to the *lsr* genes, and the six amino acids from the binding pocket conserved), *E. coli* MG1655, *B. cereus* and *B. anthracis*, showed a LuxS-dependent AI-2 binding ability similar to that observed for the previously characterized *S. typhimurium* LsrB protein (first four pairs of bars in Fig. 3). Conversely, no AI-2 binding activity was detected in the candidates from the group II (*R. etli, R. leguminosarum, A. tumefaciens,*

E. coli UT189 / UPEC, last four pairs of bars in Fig. 3). Thus, as predicted from sequence analysis and structure prediction (above), LsrB orthologs from group I demonstrate AI-2 binding ability while group II orthologs lack this ability.

The amino acids Aspartate 166 and Alanine 222 are required for AI-2 binding. Based on predicted structure-based sequence alignments (above), the amino acids that form hydrogen bonds with AI-2 are completely conserved in all of the LsrB orthologs that demonstrated the ability to bind AI-2. In contrast, all the proteins that were unable to bind AI-2 in our *in vitro* assays lacked at least two of these residues. Specifically, in *R. etli, R. leguminosarum* and *A. tumefaciens,* there are predicted to be two substitutions: D166N and A222T (numbering follows LsrB from *S. typhimurium*). Indeed, the majority of the proteins in group II have these substitutions, though other substitutions are observed (see Table S1 for detailed information). The complete conservation of AI-2 hydrogen binding residues in orthologs of group I but not group II is apparent in a multiple sequence alignment of all of the LsrB orthologs for which we have experimental data (purple in Supplementary Fig. S1). It is worth noting that 29 non-binding site residues are completely conserved across all groups in this alignment (yellow in Fig. S1). However, structural analysis shows that these residues are not clustered. Moreover, these residues are disproportionately Gly and Pro (10 and 4 conserved occurrences respectively) suggesting that, unlike the six residues in the binding site, these residues are conserved for structural rather than functional reasons.

We interpreted this to indicate that residues D166 and A222 are essential for AI-2 binding ability, and to test this idea we introduced the above mutations (D166N and A222T) into the *B. anthracis* LsrB ortholog, both individually and together, and assayed for AI-2 binding ability. As shown in Figure 4, while the wild type protein is capable of binding AI-2, no AI-2 activity was present in the binding pockets of any of the mutants

as measured by the *V. harveyi* bioassay. As a complementary experiment, we tested the ability to create AI-2 binding capacity in the distantly related LsrB ortholog of *R. etli* by mutating the putative binding site residues to mimic the binding site of the proteins from group I. These mutants failed to show AI-2 binding in the *V. harveyi* bioassay (data not shown), indicating that these proteins have already diverged to such a degree that other aspects of the protein structure important for AI-2 binding are missing.

These results show that D166 and A222, conserved in all the LsrB-orthologs we have shown to bind AI-2, are necessary (though not sufficient) for the ability of these proteins to bind AI-2, and thus provide a useful criterion for the identification of other LsrB-like AI-2 receptors. It is possible that more conservative mutations would still allow AI-2 binding, but such mutations are not observed in our list of orthologs. Further, these results support the hypothesis that the proteins in group II are incapable of AI-2 binding and are therefore very unlikely to function as AI-2 receptors *in vivo*.

Evolution of functional LsrB-like AI-2 receptors. Our sequence/structural and functional studies lead us to predict that all the organisms from group I have LsrB orthologs that function as LsrB-AI-2 receptors. This group contains members from the evolutionary distant orders of the Enterobacteriales, Pasteurellales, Rhizobiales, Rhodobacterales, and Bacillales. To infer the evolutionary history of the *lsrB* gene we determined the phylogenetic tree of all the *lsrB* gene orthologs from group I (Fig. 5) and compared it to the *rpoB* housekeeping gene organismal tree constructed with representatives of all major phyla of Bacteria (Fig. 6). Importantly, the organismal tree recovers all major phyla and classes with high bootstrap support. The relationship among phyla has a lower bootstrap support but this does not influence our analysis because the phylogenetic relationship between all species with functional *lsrB* genes (highlighted in grey boxes Fig. 6) is also well supported in this tree.

This analysis indicates that the phylogenies of *lsrB* and *rpoB* largely overlap in their diversification patterns, although with some important exceptions. The majority of the species included in group I of Table 3 clustered within the Enterobacteriales and Pasteurellales (both γ-proteobacteria) and the diversification pattern of the *lsrB* gene mimics the phylogenetic relationships obtained in the *rpoB* organismal tree within this group (compare distributions in Fig. 5 and 6); that is, the *lsrB* gene tree recovers all species groups and the relationship within Enterobacteriales and Pasteurellales is largely congruent between gene trees. Additionally, the widespread occurrence of LsrB within the Enterobacteriales and Pasteurellales strongly suggest a single origin for this AI-2 receptor that occurred in an ancestor of these organisms after the diversification of the Enterobacteriales and Pasteurellales from the Vibrionales. Nonetheless, the presence of *lsrB* genes in the Enterobacteriales and Pasteurellales is not ubiquitous, as shown by by *Erwinia carotovora* and two *E. coli* (UTI89 and APECO1) suggesting independent events of gene loss (Fig. 6).

The major discordance between the *lsrB* and *rpoB* phylogenies relates to the occurrence of functional LsrB in *S. meliloti* (Rhizobiales, α-proteobacteria), *R. sphaeroides* (Rhodobacterales, α-proteobacteria), and three species of *Bacillus* (Bacillales, Firmicutes). Specifically, *lsrB* genes from these species cluster with strong nodal support (Bayesian posterior probability of 1.0; Fig. 5) with specific clades of the Enterobacteriales and Pasteurellales. Thus, these species appear "misplaced" in the *lsrB* gene phylogeny (Fig. 5) in contrast with the organismal phylogeny (*rpoB* tree, Fig. 6). This type of incongruence is consistent with LGT events (9, 54).

In the case of the *Bacillus* species, the phylogenetic pattern of the *lsrB* gene tree reveals that these species cluster with the Pasteurellales. Thus, the occurrence of the *lsrB* gene in the *Bacillus* lineage could be explained by a putative LGT event from bacteria of the

family Pasteurellaceae. The occurrence of this gene within so many *Bacillus* species indicates that, if such a transfer occurred, enough time has passed for the lineage to diversify into at least three different species (Figs. 5 and 6).

The two species from the α-proteobacteria, (*S. meliloti* and *R. sphaeroides*) are nested within the Enterobacteriales clustering with the *Klebsiella* and *Enterobacter*. Given the phylogenetic distance that separates *S. meliloti* and *R. sphaeroides* (Fig. 6) it is surprising that the *lsrB* gene topology clusters these two species together. The most likely explanation for this occurrence requires at least more than one LGT event. Such pattern could be obtained if two sequential LGT events had occurred; for example first from one Enterobacteria (most likely an ancestor of *Klebsiella* and *Enterobacter*) to a *Sinorhizobium* and a second to a *Rhodobacter,* or from one Enterobacter first to *Rhodobacter* and then to *Sinorhizobium.* However, with the data at hand it is difficult to predict the specific order of these events. Furthermore, we predict that the proposed LGT to *S. meliloti* and *R. sphaeroides* must have been quite recent events, given that no further α-proteobacteria species were identified with group I LsrB orthologs. Alternatively, we could postulate one LGT event to the ancestor of these α-proteobacteria with a massive number of gene losses, but we find this possibility very unlikely.

DISCUSSION

A variety of bacterial species have been shown to be capable of responding to AI-2 by regulation of a range of niche-specific functions, but the mechanisms for AI-2 detection have been characterized in only a few cases (17, 64). This constitutes a major obstacle in work towards understanding of the function of AI-2. While sequence analysis of

bacterial genomes reveals the presence of orthologs of LsrB-like AI-2 receptors in Gram-negative as well as Gram-positive bacteria (this study and (41, 50)), establishing which orthologs are, in fact, functional as AI-2 receptors is important for determining if and how these species use AI-2 as a chemical signal. Thus, after analyzing sequences and predicted structures of LsrB orthologs, we identified criteria for predicting which LsrB orthologs are functional AI-2 receptors and assayed the AI-2 binding ability of selected candidates to test our criteria. Our results not only support our predictions, but also provide the first biochemical confirmation of the presence of functional AI-2 receptors in Gram-positive bacteria specifically in *B. anthracis* and *B. cereus*.

Our sequence and structural analyses allowed us to categorize the organisms with LsrB orthologs into two different groups. Members of group I have: 1) LsrB orthologs with greater than 60% sequence identity with *S. typhimurium* LsrB, 2) orthologs to the other key transport proteins of the Lsr operon, and 3) complete conservation of all 6 residues which hydrogen bond with AI-2 in *S. typhimurium* LsrB (based on structure prediction). On the other hand, in organisms belonging to group II the LsrB orthologs have a sequence identity below 36%, are missing orthologs to key proteins of the Lsr operon, and lack at least 2 of the 6 residues in the AI-2 binding pocket. These characteristics led us to hypothesize that the organisms from group I had functional AI-2 binding proteins, whereas the LsrB orthologs in group II were likely to have a different function. In all organisms where the function of either the LsrB protein or its gene has been studied, LsrB has been shown, along with other proteins that form the Lsr transport system, to participate in the uptake of AI-2 (37, 48, 57, 65); thus, we further predicted that organisms with a functional LsrB and orthologs to all the proteins from the Lsr system would take up AI-2. Accordingly, all the organisms from group I tested for binding of AI-2 by LsrB or for *in vivo* AI-2 removal (*S. typhimurium*, *S. meliloti* (37), *E. coli* K-12

(MG1655), *B. cereus* and *B. anthracis*) were capable of both of these functions. None of the proteins from the organisms we tested from group II (*E. coli* UT189 / UPEC, *R. etli, R. leguminosarum, A. tumefaciens*) were capable of binding AI-2, nor were these organisms able to take up AI-2. In addition, our analysis of predicted structures of the LsrB orthologs identified key binding site residues that are not conserved in group II organisms. Mutagenesis of the *B. anthracis* LsrB ortholog (classified as group I and demonstrated to bind AI-2) with the two most common group II substitutions (D166N, and A222T) confirmed that these residues are critical for AI-2 binding. This result strongly supports our use of binding site conservation as a key criterion in identifying class I orthologs.

These results offer experimental evidence that functional LsrB-AI-2 receptors are present in particular members of the Enterobacteriaceae (*S. typhimurium*, and *E. coli*), Rhizobiaceae (*S. meliloti*), and Bacillaceae (*B. cereus* and *B. anthracis*) and, given the correlation of our experimental results with our classification scheme, we predict that all the other LsrB orthologs from group I are functional AI-2 receptors and that these organisms are competent for AI-2 uptake. Accordingly, we expect that the members of the Pasteurellaceae and Rhodobacteraceae families in group I (Table 3) also have functional AI-2 transporters. On the other hand, we believe it is likely that all group II members have orthologs that are not involved in AI-2 transport, and thus that these organisms do not uptake AI-2 via an LsrB-type mechanism. The criteria described here can be used to predict the presence (or absence) of functional LsrB-like AI-2 receptors in newly sequenced species, and as new species are sequenced we expect the number of organisms in group I to increase.

The large majority of the organisms from group I belong to the Enterobacteriales and the Pasteurellales. This, coupled with the fact that the diversification pattern of the *lsrB*

gene largely mimics the bacterial phylogenetic relationships within this group, is consistent with a single origin for the LsrB-AI-2 receptor that likely occurred in an ancestor of these organisms after the diversification of the Enterobacteriales and the Pasteurellales from the Vibrionales. Thus, the occurrence of LsrB receptors in one species of Rhizobiales (*S. meliloti*), Rhodobacterales (*R. sphaeroides*), and three species of Bacillales was very surprising and immediately raised the possibility of LGT. The hypothesis of LGT between organisms from the Enterobacteriales or the Pasteurellales and these three orders was supported by the comparison of the *lsrB* gene tree and the *rpoB* organismal tree. Specifically, in the *lsrB* gene tree the *Bacillus* are clustered with the Pasteurellales, and the *S. meliloti* and *R. sphaeroides* are nested within the Enterobacteriales. These are nested patterns where species appeared to be "misplaced" in the gene phylogeny and can be interpreted as an indication of events of LGT. Often, genes that have been acquired by LGT have atypical nucleotide distribution (reflected in GC content or codon usage) when compared with the rest of the genome (25). However, in this case analysis of GC usage and codon bias provided no information to argue for or against the hypothesis of LGT (data not shown). Certainly, other occurrences such as convergent evolution by natural selection or ancient origin of *lsrB* at the base of the Bacteria tree with a large number of events of gene loss could also explain the observed patterns, but since we do not have specific data to support a particular explanation over the others, we favor LGT as the most parsimonious explanation as it requires the minimum number of assumptions. LGT events are now well accepted as a major force in the evolution of bacterial genomes (8, 23) leading to an increment in the number of genes (35) and pathways (19) and often enabling bacteria to acquire new functions, such as traits associated with pathogenicity, that allow adaptation to novel environments. In the specific cases of *S. meliloti* and *R. sphaeroids*, it is intriguing that that these

organisms have acquired the AI-2 receptor but not its synthase (LuxS); thus, these organisms have potentially gained the ability to eavesdrop on their neighbors signal as previously suggested (37, 41). It will also be interesting to determine the adaptive value of this new function and explore its impact in the physiology of these organisms. LGT has been proposed for other autoinducer receptors and regulators from the LuxI/LuxR family of species-specific quorum sensing proteins, where it was proposed that the acquisition of this family of proteins has benefited certain bacterial species by allowing them to gain an efficient mechanism for regulating virulence genes (8, 16, 26).

Interestingly, the LsrB ortholog in *R. leguminosarum* bv trifolii, which we identified as belonging to group II, has been shown to be essential for rhamnose (a methyl-pentose sugar) uptake and growth in this sugar, and is thus likely to be a rhamnose binding protein (42, 43). Motivated by this finding, we used the protein sequence of *R. leguminosarum* bv trifolii (KEGG ID pRL110413) to carry out a reciprocal best hit analysis against all the genomes sequences used in the previous analysis. We found that there are 12 orthologs to the *R. leguminosarum* binding protein (along with the proteins from the rhamnose transport operon) present in group II of Table 3. Thus, these 12 binding proteins are orthologs to both LsrB of *S. typhimurium* and the rhamnose binding protein of *R. leguminosarum*. These proteins have more than 65% sequence identity with the *R. leguminosarum* protein but less than 36% identity with *S. typhimurium* LsrB. We interpret this as strong evidence that these 12 proteins in group II are functioning as rhamnose binding proteins, in agreement with our prediction that they are not AI-2 receptors (these proteins are highlighted in the supplementary Table S1). These 12 organisms correspond to species belonging to α-proteobacteria that cluster together in the organismal *rpoB* tree (highlighted by the dashed box in Fig. 6). Interestingly, *S. meliloti* is the only organism that has an LsrB ortholog belonging to group I and also a

different set of proteins which are orthologs to the *R. leguminosarum* proteins from the rhamnose transport operon, further corroborating our hypothesis that the acquisition of LsrB occurred by LGT in *S. meliloti.*

While the presence of a functional LsrB ortholog does not prove that AI-2 import is involved in control of AI-2 mediated behavior, it is suggestive. Accordingly, the function of the Lsr system in AI-2 signaling has already been shown for a member of the Pasteurellaceae, the *A. actinomycetemcomitans*, (an organism not present in Table 3 because, to date, its genome is not present in the KEGG database). Demuth and co-workers have shown that this oral pathogen is capable of internalizing AI-2 via the Lsr system and, importantly, that LsrB is required to mediate the complete AI-2-dependent activation of biofilm formation in this organism (48, 49). In other cases like *Photorhabdus luminescens,* an insect pathogen belonging to the Enterobacteriaceae, transcription of the *lsr* operon was shown to be induced by AI-2, and AI-2 has also been implicated in the regulation of biofilm formation and motility (24). However, it remains to be demonstrated whether or not the Lsr system is involved in mediating these AI-2 regulated behaviors. Likewise, it will be interesting to determine whether the Lsr system is involved in mediating AI-2 signal transduction in *B. cereus* and *B. anthracis*, where AI-2 has been implicated in regulating biofilm formation (2) and growth rate (21). Certainly, the results presented here give support to that possibility.

This study, along with the two previous studies based on sequence analysis (41, 50), also reveals that certain bacteria like *Helicobacter pylori* (39), *Streptococcus mutans* (55), *Staphylococcus epidermidis* (27), *Porphyromonas gingivalis* (20, 67), *Pseudomonas aeruginosa* (14), *Bacillus subtilis* (28) which have been shown to respond to AI-2 do not have either of the known types of AI-2 receptors (neither LuxP nor LsrB), and thus we expect that other receptors for AI-2 remain to be discovered. These

556 receptors may be of entirely new classes or may be promiscuous receptors for other
557 small molecules. Novel receptor classes are likely to be identified by approaches that
558 rely on genetic screens to isolate mutants involved in modulating AI-2-regulated
559 phenotypes, and as shown here integration with approaches that use sequence analysis
560 coupled with biochemical assays may prove very useful. Clearly, elucidation of the
561 proteins involved in AI-2 recognition and signal relay is essential for studying the
562 potential functions of this class of signal molecule in intra- and inter-species cell-to-cell
563 communication and/or intra- and inter-cellular signal transduction. The identification
564 and experimental confirmation of functional LsrB receptors in this study opens the door
565 to the understanding of the molecular basis of AI-2 mediated behavioral regulation in a
566 variety of new species.

567

568 **ACKNOWLEDGMENTS**

569 The work performed in the laboratory of K.B.X. was supported by Fundação para a
570 Ciência e Tecnologia, Portugal (FCT) Grant PTDC/BIA-BCM/73676/2006. C.S.P. and
571 P.H.B. were supported by FCT awards SFRH/BD/28543/2006 and
572 SFRH/BPD/26852/2006 respectively. The laboratory of S.T.M. gratefully
573 acknowledges grants from the NIH (Grant AI074041), the Camille and Henry Dreyfus
574 Foundation and Swarthmore College. We also would like to thank program FLAD/NSF
575 Proj. 600-10/2006 for sponsoring the stay of S.T.M. in the laboratory of K.B.X. during
576 his sabbatical.

577 We thank Chengetai Mahomva and Bella Liu for technical assistance. We are very
578 grateful to Adriano O. Henriques, Gladys Alexandre, James P. Shapleigh, Jeffrey I.
579 Gordon and Marcus B. Jones for providing the strains used in this study.

We thank Bonnie Bassler, Michiko Taga, Brian Hammer and Isabel Gordo for critically reading the manuscript. We also thank Bonnie Bassler for receiving C.S.P. in her laboratory.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403-410.
2. **Auger, S., E. Krin, S. Aymerich, and M. Gohar.** 2006. Autoinducer 2 affects biofilm formation by *Bacillus cereus*. Appl. Environ. Microbiol. **72:**937-941.
3. **Baek, S. H., A. Hartsock, and J. P. Shapleigh.** 2008. *Agrobacterium tumefaciens* C58 uses ActR and FnrN to control *nirK* and *nor* expression. J. Bacteriol. **190:**78-86.
4. **Bassler, B. L., M. Wright, R. E. Showalter, and M. R. Silverman.** 1993. Intercellular signalling in *Vibrio harveyi*: sequence and function of genes regulating expression of luminescence. Mol. Microbiol. **9:**773-786.
5. **Bassler, B. L., M. Wright, and M. R. Silverman.** 1994. Multiple signalling systems controlling expression of luminescence in *Vibrio harveyi*: sequence and function of genes encoding a second sensory pathway. Mol. Microbiol. **13:**273-286.
6. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340:**783-795.
7. **Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453-1474.
8. **Boucher, Y., C. J. Douady, R. T. Papke, D. A. Walsh, M. E. Boudreau, C. L. Nesbo, R. J. Case, and W. F. Doolittle.** 2003. Lateral gene transfer and the origins of prokaryotic groups. Annu. Rev. Genet. **37:**283-328.

9. **Brown, J. R.** 2003. Ancient horizontal gene transfer. Nat Rev Genet **4:**121-132.

10. **Cannon, S. B., and N. D. Young.** 2003. OrthoParaMap: distinguishing orthologs from paralogs by integrating comparative genome data and gene phylogenies. BMC Bioinformatics **4:**35.

11. **Case, R. J., Y. Boucher, I. Dahllof, C. Holmstrom, W. F. Doolittle, and S. Kjelleberg.** 2007. Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. Appl. Environ. Microbiol. **73:**278-288.

12. **Chen, X., S. Schauder, N. Potier, A. Van Dorsselaer, I. Pelczer, B. L. Bassler, and F. M. Hughson.** 2002. Structural identification of a bacterial quorum-sensing signal containing boron. Nature **415:**545-549.

13. **Croxatto, A., J. Pride, A. Hardman, P. Williams, M. Camara, and D. L. Milton.** 2004. A distinctive dual-channel quorum-sensing system operates in *Vibrio anguillarum*. Mol. Microbiol. **52:**1677-1689.

14. **Duan, K., C. Dammel, J. Stein, H. Rabin, and M. G. Surette.** 2003. Modulation of *Pseudomonas aeruginosa* gene expression by host microflora through interspecies communication. Mol. Microbiol. **50:**1477-1491.

15. **Federle, M. J., and B. L. Bassler.** 2003. Interspecies communication in bacteria. J. Clin. Invest. **112:**1291-1299.

16. **Gray, K. M., and J. R. Garey.** 2001. The evolution of bacterial LuxI and LuxR quorum sensing regulators. Microbiology **147:**2379-2387.

17. **Hardie, K. R., and K. Heurlier.** 2008. Establishing bacterial communities by 'word of mouth': LuxS and autoinducer 2 in biofilm development. Nat. Rev. Microbiol.

18. **Henikoff, S., and J. G. Henikoff.** 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA **89:**10915-10919.

19. **Iwasaki, W., and T. Takagi.** 2009. Rapid pathway evolution facilitated by horizontal gene transfers across prokaryotic lineages. PLoS Genet. **5:**e1000402.

20. **James, C. E., Y. Hasegawa, Y. Park, V. Yeung, G. D. Tribble, M. Kuboniwa, D. R. Demuth, and R. J. Lamont.** 2006. LuxS involvement in the regulation of genes coding

for hemin and iron acquisition systems in *Porphyromonas gingivalis*. Infect. Immun. **74:**3834-3844.

21. **Jones, M. B., and M. J. Blaser.** 2003. Detection of a *luxS*-signaling molecule in *Bacillus anthracis*. Infect. Immun. **71:**3914-3919.

22. **Kelley, L. A., and M. J. Sternberg.** 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. **4:**363-371.

23. **Koonin, E. V., K. S. Makarova, and L. Aravind.** 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu. Rev. Microbiol. **55:**709-742.

24. **Krin, E., N. Chakroun, E. Turlin, A. Givaudan, F. Gaboriau, I. Bonne, J. C. Rousselle, L. Frangeul, C. Lacroix, M. F. Hullo, L. Marisa, A. Danchin, and S. Derzelle.** 2006. Pleiotropic role of quorum-sensing autoinducer 2 in *Photorhabdus luminescens*. Appl. Environ. Microbiol. **72:**6439-6451.

25. **Lawrence, J. G., and H. Ochman.** 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **95:**9413-9417.

26. **Lerat, E., and N. A. Moran.** 2004. The evolutionary history of quorum-sensing systems in bacteria. Mol. Biol. Evol. **21:**903-913.

27. **Li, M., A. E. Villaruz, V. Vadyvaloo, D. E. Sturdevant, and M. Otto.** 2008. AI-2-dependent gene regulation in *Staphylococcus epidermidis*. BMC Microbiol. **8:**4.

28. **Lombardia, E., A. J. Rovetto, A. L. Arabolaza, and R. R. Grau.** 2006. A LuxS-dependent cell-to-cell language regulates social behavior and development in *Bacillus subtilis*. J. Bacteriol. **188:**4442-4452.

29. **Meade, H. M., S. R. Long, G. B. Ruvkun, S. E. Brown, and F. M. Ausubel.** 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. J. Bacteriol. **149:**114-122.

30. **Miller, L. D., C. K. Yost, M. F. Hynes, and G. Alexandre.** 2007. The major chemotaxis gene cluster of *Rhizobium leguminosarum* bv. viciae is essential for competitive nodulation. Mol. Microbiol. **63:**348-362.

31. **Miller, S. T., K. B. Xavier, S. R. Campagna, M. E. Taga, M. F. Semmelhack, B. L. Bassler, and F. M. Hughson.** 2004. *Salmonella typhimurium* recognizes a chemically distinct form of the bacterial quorum-sensing signal AI-2. Mol. Cell. **15:**677-687.

32. **Neiditch, M. B., M. J. Federle, S. T. Miller, B. L. Bassler, and F. M. Hughson.** 2005. Regulation of LuxPQ receptor activity by the quorum-sensing signal autoinducer-2. Mol. Cell. **18:**507-518.

33. **Neiditch, M. B., M. J. Federle, A. J. Pompeani, R. C. Kelly, D. L. Swem, P. D. Jeffrey, B. L. Bassler, and F. M. Hughson.** 2006. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. Cell **126:**1095-1108.

34. **Nylander, J. A. A.** 2004. MrModeltest version 2, p. Program distribuited by the author Evolutionary Biology Centre, Uppsala University.

35. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405:**299-304.

36. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444-2448.

37. **Pereira, C. S., J. R. McAuley, M. E. Taga, K. B. Xavier, and S. T. Miller.** 2008. *Sinorhizobium meliloti*, a bacterium lacking the autoinducer-2 (AI-2) synthase, responds to AI-2 supplied by other bacteria. Mol. Microbiol. **70:**1223-1235.

38. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817-818.

39. **Rader, B. A., S. R. Campagna, M. F. Semmelhack, B. L. Bassler, and K. Guillemin.** 2007. The quorum-sensing molecule autoinducer 2 regulates motility and flagellar morphogenesis in *Helicobacter pylori*. J. Bacteriol. **189:**6109-6117.

40. **Reigstad, C. S., S. J. Hultgren, and J. I. Gordon.** 2007. Functional genomic studies of uropathogenic *Escherichia coli* and host urothelial cells when intracellular bacterial communities are assembled. J. Biol. Chem. **282:**21259-21267.

41. **Rezzonico, F., and B. Duffy.** 2008. Lack of genomic evidence of AI-2 receptors suggests a non-quorum sensing role for *luxS* in most bacteria. BMC Microbiol. **8:**154.

42. **Richardson, J. S., M. F. Hynes, and I. J. Oresnik.** 2004. A genetic locus necessary for rhamnose uptake and catabolism in *Rhizobium leguminosarum* bv. trifolii. J. Bacteriol. **186:**8433-8442.

43. **Richardson, J. S., and I. J. Oresnik.** 2007. L-Rhamnose transport is sugar kinase (RhaK) dependent in *Rhizobium leguminosarum* bv. trifolii. J. Bacteriol. **189:**8437-8446.

44. **Ronquist, F., and J. P. Huelsenbeck.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:**1572-1574.

45. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406-425.

46. **Schauder, S., K. Shokat, M. G. Surette, and B. L. Bassler.** 2001. The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. Mol. Microbiol. **41:**463-476.

47. **Semmelhack, M. F., S. R. Campagna, M. J. Federle, and B. L. Bassler.** 2005. An expeditious synthesis of DPD and boron binding studies. Org. Lett. **7:**569-572.

48. **Shao, H., D. James, R. J. Lamont, and D. R. Demuth.** 2007. Differential interaction of *Aggregatibacter (Actinobacillus) actinomycetemcomitans* LsrB and RbsB proteins with autoinducer 2. J. Bacteriol. **189:**5559-5565.

49. **Shao, H., R. J. Lamont, and D. R. Demuth.** 2007. Autoinducer 2 is required for biofilm growth of *Aggregatibacter (Actinobacillus) actinomycetemcomitans*. Infect. Immun. **75:**4211-4218.

50. **Sun, J., R. Daniel, I. Wagner-Dobler, and A. P. Zeng.** 2004. Is autoinducer-2 a universal signal for interspecies communication: a comparative genomic and phylogenetic analysis of the synthesis and signal transduction pathways. BMC Evol. Biol. **4:**36.

51. **Surette, M. G., and B. L. Bassler.** 1998. Quorum sensing in *Escherichia coli* and *Salmonella typhimurium*. Proc. Natl. Acad. Sci. USA **95:**7046-7050.

52. **Surette, M. G., and B. L. Bassler.** 1999. Regulation of autoinducer production in *Salmonella typhimurium*. Mol. Microbiol. **31:**585-595.

53. **Swofford, D. L.** 2001. PAUP*, Phylogenetic Analysis Using Parsimony. Sinauer Associates, Sunderland, Massachusetts.

54. **Syvanen, M.** 1994. Horizontal gene transfer: evidence and possible consequences. Annu. Rev. Genet. **28:**237-261.

55. **Sztajer, H., A. Lemme, R. Vilchez, S. Schulz, R. Geffers, C. Y. Yip, C. M. Levesque, D. G. Cvitkovitch, and I. Wagner-Dobler.** 2008. Autoinducer-2-regulated genes in *Streptococcus mutans* UA159 and global metabolic effect of the *luxS* mutation. J. Bacteriol. **190:**401-415.

56. **Taga, M. E., S. T. Miller, and B. L. Bassler.** 2003. Lsr-mediated transport and processing of AI-2 in *Salmonella typhimurium*. Mol. Microbiol. **50:**1411-1427.

57. **Taga, M. E., J. L. Semmelhack, and B. L. Bassler.** 2001. The LuxS-dependent autoinducer AI-2 controls the expression of an ABC transporter that functions in AI-2 uptake in *Salmonella typhimurium*. Mol. Microbiol. **42:**777-793.

58. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596-1599.

59. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673-4680.

60. **Vendeville, A., K. Winzer, K. Heurlier, C. M. Tang, and K. R. Hardie.** 2005. Making 'sense' of metabolism: autoinducer-2, LuxS and pathogenic bacteria. Nat. Rev. Microbiol. **3:**383-396.

61. **Waters, C. M., and B. L. Bassler.** 2005. Quorum Sensing: Cell-to-Cell Communication in Bacteria. Annu. Rev. Cell. Dev. Biol.

743 62. **Winzer, K., K. R. Hardie, and P. Williams.** 2003. LuxS and autoinducer-2: their
744 contribution to quorum sensing and metabolism in bacteria. Adv. Appl. Microbiol.
745 **53:**291-396.
746 63. **Xavier, K. B., and B. L. Bassler.** 2005. Interference with AI-2-mediated bacterial cell-
747 cell communication. Nature **437:**750-753.
748 64. **Xavier, K. B., and B. L. Bassler.** 2003. LuxS quorum sensing: more than just a
749 numbers game. Curr. Opin. Microbiol. **6:**191-197.
750 65. **Xavier, K. B., and B. L. Bassler.** 2005. Regulation of uptake and processing of the
751 quorum-sensing autoinducer AI-2 in *Escherichia coli*. J. Bacteriol. **187:**238-248.
752 66. **Xavier, K. B., S. T. Miller, W. Lu, J. H. Kim, J. Rabinowitz, I. Pelczer, M. F.**
753 **Semmelhack, and B. L. Bassler.** 2007. Phosphorylation and processing of the quorum-
754 sensing molecule autoinducer-2 in enteric bacteria. ACS Chem. Biol. **2:**128-136.
755 67. **Yuan, L., J. D. Hillman, and A. Progulske-Fox.** 2005. Microarray analysis of
756 quorum-sensing-regulated genes in *Porphyromonas gingivalis*. Infect. Immun. **73:**4146-
757 4154.
758 68. **Zuckerkandl, E., and L. Pauling.** 1965. Evolutionary divergence and convergence in
759 proteins, p. pp. 97-166. *In* H. J. V. V. Bryson (ed.), Evolving Genes and Proteins.
760 Academic Press, New York.

761

762

763 **FIGURE LEGENDS**

764 **Figure 1. AI-2 removal profile in bacteria producing AI-2.** Extracellular AI-2
765 activity in cell-free culture fluids from LuxS$^+$ strains (A) *E.coli* MG1655 (triangles) and
766 *E.coli* UPEC (circles) and (B) *B. cereus* (diamonds) and *B. anthracis* (squares) cultures.
767 Aliquots were taken at the specified times. AI-2 activity is reported as fold induction of
768 light produced by *V. harveyi* BB170.

769 **Figure 2. Removal of exogenously supplied AI-2.** *S. meliloti* (triangles), *R.*
770 *leguminosarum* (circles), *R. etli* (squares), and *A. tumefaciens* (crosses) were cultured to OD_{600} =
771 1 in their optimal culture media (LBMC, TYC, YEM and LB respectively). Chemically
772 synthesized AI-2 was then added to all the cultures and aliquots were taken at the
773 specified times. AI-2 activity in cell-free culture fluids is reported as fold induction of
774 light produced by *V. harveyi* BB170.

775 **Figure 3. Binding of AI-2 to potential LsrB-like orthologs.** Proteins were expressed
776 in either $LuxS^+$ (black bars) or $LuxS^-$ (white bars) *E. coli* strains (BL21 and FED101,
777 respectively), purified, and denatured to release the ligand. The released ligand was
778 added to a *V. harveyi* AI-2 reporter strain (BB170) to determine AI-2 activity. AI-2
779 activity is reported as fold induction of light production by *V. harveyi* BB170
780 supplemented with protein supernatant to that of the appropriate buffer. Error bars
781 represent the standard deviations for three independent cultures.

782 **Figure 4. Binding of AI-2 by *B. anthracis* wild type (WT) and mutants D166N and**
783 **A222T LsrB-like proteins.** *B. anthracis* wild type (WT), and mutant D166N and
784 A222T proteins were expressed in either $LuxS^+$ (black bars) or $LuxS^-$ (white bars) as
785 explained in Figure 3. AI-2 activity is reported as fold induction of light production by
786 *V. harveyi* BB170 supplemented with protein supernatant to that supplemented with
787 appropriate buffer. Error bars represent the standard deviations for three independent
788 cultures.

789 **Figure 5. Evolutionary history of genes encoding functional LsrB orthologs**
790 **inferred with maximum likelihood.** *lsrB* gene tree constructed with the sequences
791 from all organisms in group I. This is an unrooted phylogram oriented to show

maximum congruence with the organismal tree. Numbers on the nodes indicate posterior probability as estimated with MrBayes.

Figure 6. Molecular phylogeny of Bacteria estimated with *rpoB* gene. *rpoB* gene tree constructed with all the organisms in Table 3 (group I and II) and representative species of all major phyla of Bacteria. This represents our best inference of the organismal tree. Grey boxes indicate species with functional *lsrB* genes (group I, Table 3), and dashed box locates the species with protein sequences in Table 3 likely to function as a rhamnose binding protein. The numbers after species names indicate the number of strains analyzed for the respective species. Taxonomic classifications (Phyla) are shown on the right. This tree was inferred with Neighbor-Joining and the branch lengths are scaled to the number of amino acid substitutions per site. Thickened branches indicate high bootstrap support (higher than 75 %). This is a measurement of phylogenetic strength between nodes and this value reflects a high confidence in the infered relationships between species.

806

807 **Table 1** – Bacterial strains used in this study.

Strain	Source and Reference
Salmonella typhimurium	ATCC 14028
Escherichia coli K-12 MG1655	(7)
Escherichia coli UTI89 (UPEC)	Jeffrey I. Gordon (40)
Bacillus anthracis Sterne 34F2 (vaccine strain)	Martin J. Blaser (21)
Bacillus cereus (ATCC 10987)	Adriano O. Henriques
Sinorhizobium meliloti Rm1021	(29)
Agrobacterium tumefaciens C58	James P. Shapleigh (3)
Rhizobium leguminosarum bv. viciae 3841	Gladys Alexandre (30)
Rhizobium etli CFN42	ATCC 51251

808

809 **Table 2** – Primers used in this study.

Construct - purpose	5'/sense	3'/antisense
R. leguminosarum / pPro – PCR	CGCGGATCCGCCGACATCAAGATCGGT	CCGCTCGAGCGTCAGAAGACCTTGGAGAACTG
A. tumefaciens / pPro – PCR	CGCGGATCCGCAGACGTCAAGATCGC	CCGCTCGAGCAATCTTCGAGAACTGATCGAT
R. etli / pPro – PCR	CGCGGATCCAAGGACATCAAGATCGGC	CCGCTCGAGTCAGAAGACCTTGGAGAACTG
E. coli UPEC / pPro - PCR	CGGGATCCGCGGAAAAAGTCG	CCGCTCGAGTTAATAAAGTGAGTCGATATTGTC
E. coli MG1655 / pPro - PCR	CGCGGATCCGCAGAGCGTATTGCATTT	CCGCTCGAGTCAGAAATCGTATTTGCCGAT
B. anthracis / pET151 - D171N	CTCTAGTCCAACAGTAACGAATCAAAACCAATGGGTAAC	GTTACCCATTGGTTTTGATTCGTTACTGTTGGACTAGAG
B. anthracis / pET151 - A227T	TATTAATGCAGTCATTTGTCCGGATACGACGGCACTTCCAG	CTGGAAGTGCCGTCGTATCCGGACAAATGACTGCATTAATA
S. typhimurium / *pGEX* - PCR	(31)	
B. cereus / *pGEX* - PCR	CGGGATCCAAGAAAAAAGCTGATGATGT	GGAATTCCTAATCAATATTATCCTTCGTAAATACGAC
B. anthracis / pET151 – PCR	CACCGATAAGAAAAAAGCGGA	CTAAAAATTATATTTATCAATAT

810

811 **Table 3 -** Orthologs of the LuxS and Lsr proteins from *S. typhimurium* present in the complete
812 genomes of the KEGG database (January, 2009).

	Species[a]	Orthologs[b]									LsrB identity[c]	Binding site residues[d]
		LuxS	LsrB	LsrA	LsrC	LsrD	LsrK	LsrR	LsrG	LsrF		
Group I	***Salmonella typhimurium* LT2**	+	+	+	+	+	+	+	+	+	100%	6
	***Salmonella enterica* (13 strains)**	+	+	+	+	+	+	+	+	+	100%	6
	***Escherichia coli* (11 strains)**	+	+	+[e]	+	+	+	+	+	+	85%	6
	Escherichia fergusonii	+	+	+	+	+	+	+	+	+	85%	6
	***Yersinia pestis* (7 strains)**	+	+	+	+	+	+	+	+	+	84%	6
	***Yersinia pseudotuberculosis* (4 strains)**	+	+	+	+	+	+	+	+	+	84%	6
	Yersinia enterocolitica	+	+	+	+	+	+	+	+	+	83%	6
	***Klebsiella pneumoniae* (2 strains)**	+	+	+	+	+	+	+	+	+	82%	6
	Photorhabdus luminescens	+	+	+	+	+	+	+	+	+	82%	6
	***Enterobacter* sp. 638**	+	+	+	+	+	+	+	+	+	82%	6
	Pasteurella multocida	+	+	+	+	+	+	+	+	+	80%	6
	Haemophilus influenzae PittEE	+	+	+	+	+	+	+	+	+	80%	6
	***Haemophilus somnus* (2 strains)**	+	+	+	+	+	+	+	+	+	76%	6
	Sinorhizobium meliloti		+	+	+	+	+	+	+	+	72%	6
	***Rhodobacter sphaeroides* (2 strains)**		+	+	+	+	+	+	+		72%	6
	***Bacillus anthracis* (4 strains)**	+	+	+	+	+	+	+	+	+	63%	6
	***Bacillus cereus* (7 strains)**	+	+	+	+	+	+	+	+	+	63%	6
	***Bacillus thuringiensis* (2 strains)**	+	+	+	+	+	+	+	+	+	63%	6
Group II	*Rubrobacter xylanophilus*		+			+			+	+	36%	3
	Ochrobactrum anthropi		+		+	+		+		+	35%	4
	Sinorhizobium medicae		+		+	+		+	+	+	35%	4
	Roseobacter denitrificans		+			+		+	+		34%	4
	Mesorhizobium loti		+			+		+		+	34%	4
	Agrobacterium tumefaciens C58 (2 strains)		+			+		+	+	+	33%	4
	Leptothrix cholodnii		+			+		+		+	33%	4
	Dinoroseobacter shibae		+		+	+	+	+	+		33%	4
	Verminephrobacter eiseniae		+			+		+			33%	4
	Burkholderia phytofirmans		+		+	+		+			33%	4
	Gluconacetobacter diazotrophicus PAl 5 (JGI)		+			+					33%	2
	Rhizobium leguminosarum		+		+	+		+	+	+	33%	4
	Rhizobium leguminosarum bv. trifolii WSM2304		+		+	+		+		+	33%	4
	Gluconacetobacter diazotrophicus PAl 5 (Brazil)		+			+					33%	1
	Rhodococcus sp. RHA1		+		+	+		+	+		33%	4
	Streptomyces coelicolor	+	+		+	+		+	+		33%	4
	Burkholderia xenovorans		+		+	+		+		+	32%	3
	Rhizobium etli		+			+		+	+	+	32%	4
	Dictyoglomus thermophilum		+			+					32%	4
	Rhizobium etli CIAT 652		+			+		+	+		32%	4
	Jannaschia sp. CCS1		+		+	+					32%	4
	Dictyoglomus turgidum		+			+					32%	4
	Acidiphilium cryptum JF-5		+					+			31%	2
	Streptomyces avermitilis		+		+	+		+			31%	4
	Burkholderia phymatum		+		+	+		+		+	31%	4
	Deinococcus geothermalis	+	+		+	+		+			31%	4
	Burkholderia ambifaria MC40-6		+		+	+		+	+	+	31%	4
	Syntrophomonas wolfei		+		+	+		+			30%	1
	Chloroflexus aggregans		+		+	+					27%	4
	Escherichia coli APEC O1	+	+			+	+	+			27%	0
	Escherichia coli UTI89 (UPEC)	+	+			+	+	+			27%	1

813
814 [a] Organisms classified as group I are highlighted in black and group II in grey.

815 [b] Orthologs of both group I and group II are defined as a complete match in the bidirectional best hits and are denoted with +.

816 [c] Percentage of identity using *S. typhimurium* LsrB as reference.

[d] Number of conserved residues in the binding site based on structure prediction using *S. typhimurium* LsrB as reference.

[e] LsrA from *Escherichia coli* E24377A is truncated.