ABSTRACT

Title of Document: CONTEXTUALIZATION OF THE *E. COLI* LSR SYSTEM: RELATIVE ORTHOLOGY, RELATIVE QS ACTIVITY, AND EMERGENT BEHAVIOR

David Nathan Quan, Doctor of Philosophy, 2014

Directed By: Professor William E. Bentley
Fischell Department of Bioengineering

Within bacterial consortia there exist innumerable combinatorial circumstances, some of which may tip the scale toward pathogenicity, some of which may favor asymptomatic phenotypes. Indeed, the lines and intersections between commensal, pathogenic, and opportunistic bacteria are not always clean. As a foothold to mediate pathogenicity arising from consortia, many have puzzled at communication between bacteria. Primary among such considerations is quorum sensing (QS). Analogous to autocrine signaling in multicellular organisms, QS is a self-signaling process involving small molecules. Generally, QS activation is believed to have pleiotropic effects, and has been associated with numerous pathogenic phenotypes. The research herein focuses on autoinducer-2 (AI-2) based QS signaling transduced through the Lsr system. Produced by over 80 species of bacteria, AI-2 is believed to be an interspecies signaling molecule. Outside of the marine bacteria genera *Vibrio* and *Marinomonas*, the only known AI-2 based QS transduction pathway is the Lsr system. We sought to deepen the characterization of the Lsr system in contexts outside of the batch cultures in which it was originally defined.

First, we interrogated *E. coli* K-12 W3110 Lsr system orthologs relative to the same strain's *lac* system. Both systems are induced by the molecule which they import and catabolize. We searched for homologs by focusing on the gene order along a genome, as gene arrangement can bear signaling consequences for autoregulatory circuits. We found that the Lsr system signal was phylogenetically dispersed if not particularly deep, especially outside of Enterobacteriales and Pasteurellaceaes, indicating that the system has generally been conferred horizontally. This contrasts with the *lac* system, whose signal is strong but limited to a select group of highly related enterobacteria. We then modeled the Lsr system with ODEs, revealing bimodality *in silico,* bolstering preliminary experimental evidence. This bifurcated expression was seen to depend upon nongenetic heterogeneity, which we modeled as a variation of a single compound parameter, *basal,* representing the basal rate of AI-2 flux into the cell through a low flux pathway. Moreover, in our finite difference-agent based models, bimodal expression could not arise from spatial stochasticity alone. This lies in contrast with the canonical LuxIR QS system, which employs an intercellular positive feedback loop to activate the entire population. We examined the consequences of this contrast, by modeling both systems under conditions of colony growth using finite difference-agent based methods. We additionally investigated the confluence of Lsr signaling with chemotactic sensitivity to AI-2, which has been demonstrated in *E. coli*. Finally, the consequences of bimodality in interspecies interactions were assessed by posing two populations containing different Lsr systems against each other. While few natural consortia consist of only two interacting bacteria, these studies indicate that AI-2 based Lsr signaling may mediate a

multitude of transitional intraspecies and interspecies bacterial dynamics, the specifics of which will vary with the context and the homologs involved.

CONTEXTUALIZATION OF THE *E. COLI* LSR SYSTEM: RELATIVE ORTHOLOGY, RELATIVE QS ACTIVITY, AND EMERGENT BEHAVIOR

By

David Nathan Quan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor William E. Bentley, Chair
Associate Professor Michael P. Cummings
Associate Professor Nam Sun Wang
Associate Professor Adam Hsieh
Assistant Professor Ganesh Sriram

Dedication

As a token of love and appreciation, to my parents, to my brother, and my broader family who have sacrificed so much, both small and great—to their unrelenting and unconditional support.

Acknowledgements

I wish to acknowledge all the members of my dissertation committee for their time, advice, and commitment to myself, their own graduate students, and the university. I especially recognize my committee chair and advisor, Dr. William Bentley, under whose guidance I have been allowed to follow my nose and in whose lab I have enjoyed a stimulating research environment. I also wish to acknowledge my fellow lab members both present and past, whose work serves as fundamental bedrock for my own, and whose ideas I have come to recount as readily as my own. In particular, I wish to acknowledge Dr. Chen-Yu Tsao whose development of pCT5 and pCT6 has been utterly important for our lab and has resulted in compelling basic science. I also acknowledge the work of those in the quorum sensing field as whole—it is only by standing upon their shoulders that I can view any forest at all. I also wish to gratefully acknowledge my mentors at the Center for Cell Dynamics, the members of the Alliance for Cellular Signaling, and Dr. Gary Jarvis for giving me a wide berth when trying to do science during my formative years. In particular, I would like to recognize a former mentor at the Center for Cell Dynamics, Dr. Jonathon Alberts, whose Sim2D platform was modified and heavily relied upon here. Finally, I acknowledge the reader: I hope you find what you need or perhaps even something interesting or worthwhile.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

Originally termed autoinduction [1,2], "quorum sensing" (QS) is an admitted misnomer. The first description of a bacterial "quorum" clearly fell within the context of cell concentration [3]. Nonetheless, for the last two decades the popular equivalency of "quorum" to a "sufficiently large population" has been rigorously reevaluated [3]. With the aid of new microfluidic schemes, researchers have shown that QS can be activated for one or only a few isolated bacteria [4–6]. With the benefit of careful consideration, others have indicated that under some circumstances, QS might more appropriately be called "diffusion sensing" or "efficiency sensing" [7–9]. Yet others have argued for paring back the usage of the term "quorum sensing" from an evolutionary point of view [10].

Indeed, QS studies have become home to a broad spectrum of bacterial intraspecies and interspecies autoinduction or autoinduction-like signaling. This ready adoption is probably attributable not only to the fact that QS is an apparently widespread phenomenon, but could also be ascribed to the allure of the underlying paradigm: unicellular organisms using self-secreted molecules to drive multicellular behavior. As an additional point of interest, many QS-driven behaviors, such as toxin production and biofilm formation, are tied to pathogenic phenotypes. The current count of molecules generally considered autoinducers includes autoinducing peptide (AIP) from certain Gram positive bacterial species, autoinducer-1 (AI-1 or acyl homoserine lactones (AHLs)) from myriad beta- and alpha-proteobacteria, autoinducer-2 (AI-2 or (S)-4,5-dihydroxy-2,3-pentandione (DPD) and its interconvertible stereoisomers) from a great

array of bacteria, autoinducer-3 (AI-3, of unidentified composition) from *E. coli*, *Cholera* autoinducer-1 (CA-1) and its analogs found in other *Vibrio* species, and Pseudomonas Quinolone Signal (PQS). Notably, S-adenosyl methionine (SAM) lies directly upstream of both AHLs and AI-2 synthesis, as noted in Figure 1-1. Serving as a methyl donor in the activated methyl cycle, marked cellular abundance, and high reactivity may make it a favored target for such repackaging.

Each QS signal is matched to one or more effector pathways that fall under one of two broad categories, as depicted in Figure 1-2: two component regulatory systems (TCRS) or direct transcription factor mediation. TCRS are typically composed of a transmembrane receptor with histidine kinase functionality that activates a cytosolic regulatory partner [11]. For these systems, autoinducers remain extracellular when generating a response. A separate category of QS activity is affected through the direct mediation of transcription factor activity by autoinducer or modified autoinducer. Occurring in the cytosol, this requires autoinducer internalization through either passive or active means. Such cases include the activity of LuxIR found in numerous beta- and alpha-proteobacteria [12] and Lsr systems found dispersed mainly among gamma-proteobacteria [13,14], respectively.

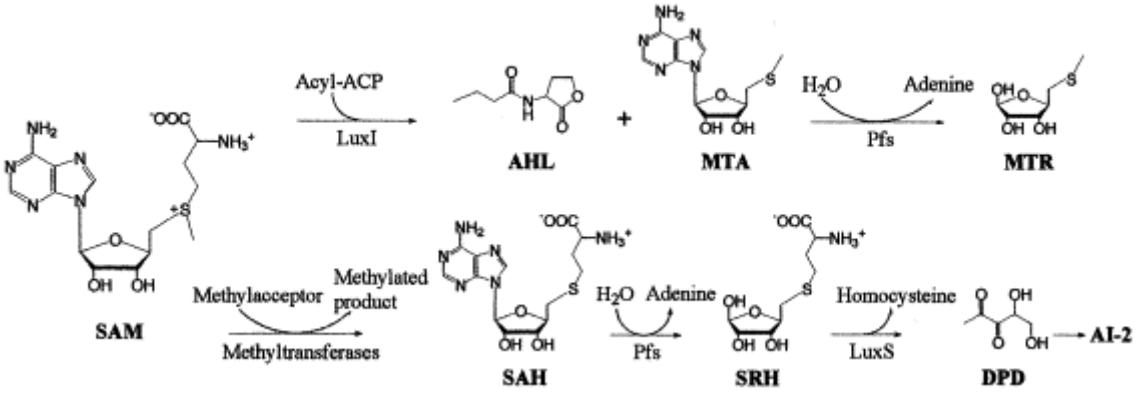


Figure 1-1. Synthesis of AHL and DPD (from [15])**.** AHL and DPD share a common upstream reactant, SAM, from the activated methyl cycle.

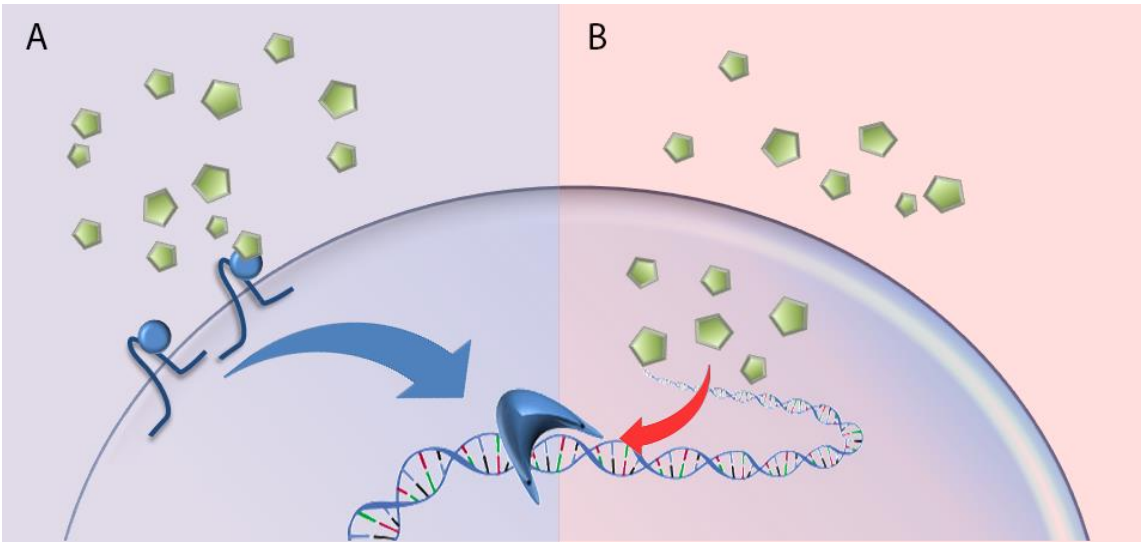


Figure 1-2. QS molecule activation of QS signaling pathway. *A.* Signaling molecules activate the cell through receptor mediation at the cell surface, usually through a two component regulatory system. *B.* Signaling molecules traverse the membrane and interact with transcription factors directly or after slight modification.

Perhaps reflecting the complexity of their environment, many bacteria appear to operate under the influence of multiple QS signal-effector units. Sometimes this appears as topological redundancy or at least functional complementarity. In other cases, QS signals appear to compete for influence, although this too could be a type of complementarity, when discordant phenotypes are separately timed.[16]

1.1.1 QS in *Vibrio harveyi*

For instance, *Vibrio harveyi* QS is known to rely on at least three autoinducers: 3OHC4-HSL (an AHL), DPD, and a *V. harveyi* version of CA-1. In this instance, all three autoinducers interact with separate cognate TCRS pairs, the signals of which are all funneled into and filtered through the same overlapping negative feedback loops, preventing premature activation under low autoinducer conditions and limiting expression at high autoinducer concentrations, ultimately phosphorylating and activating a single regulatory protein, LuxO.[17,18] Despite this canalization, evidence suggests that different autoinducers accumulate at different rates and help propel separate genetic programs at distinct phases of growth [16], possibly triggering activity in combination with distinct sets of transcription factors.

Although signal crosstalk between species can occur when AHLs are structurally similar[19] and is likely where the same molecule has been synthesized by different species [20], whether this necessarily represents QS *qua* QS is open to interpretation [10]. As a simplifying generalization, the activity of AHLs are commonly described as limited to the species from which they are synthesized [21,22]. For example, as far as is known, *V. harveyi*'s AHL, 3OHC4-HSL [23], conforms to this paradigm. Furthermore, orthologs for

V. harveyi's 3OHC4-HSL synthase and receptor, *luxM* and *luxN* respectively, are phylogenetically confined to a small number of *Vibrios* [24–26], and produce distinct AHLs.

The limited scope of AHL signaling contrasts sharply with the more universally produced AI-2—which is to say that DPD synthase (LuxS) homologs are found in myriad species [27]. While DPD exists as a specific chemical species, it spontaneously interconverts between several distinct products, each possibly producing a different level of QS activity [28]. In *V. harveyi*, AI-2 interacts with the TCRS, LuxPQ [29], which according to iterative protein BLAST searches is highly abundant among *Vibrio* and *Marinomonas* species.

A particular derivative formed in marine environments where borate is abundant is borated DPD [29]. DPD is produced by practically all *Vibrio* species, and borated DPD appears to be the cognate AI-2 molecule for those same species [30].

A third *V. harveyi* QS system involves *cqsA* and *cqsS*, a receptor-synthase pair that is conserved widely but exclusively among *Vibrio* species. As with AHLs, the resulting autoinducers (CA-1-like) are not known to have wide signaling efficacies outside of the species by which they are produced [31].

Mediation of QS processes at the membrane is a distinctive characteristic of these QS systems in comparison to other known QS machinery, and makes *V. harveyi* and the *Vibrio* and *Marinomonas* species (all marine bacteria) that share these architectures outliers.

Along the separate continuum of QS processes that feedback onto autoinducer concentration, *V. harveyi* represents an intermediate in the QS landscape. On one end of

this spectrum exists QS systems like the canonical *luxIR* pair, where AHL synthase expression is driven by autoinducer-activated transcription, producing a strong intercellular positive feedback loop. On the other end is QS activation that is not only dissociated from autoinducer synthesis but also drives uptake, producing a strong negative intercellular feedback loop.

1.1.2 LuxIR QS

Widely utilized in synthetic biology, the LuxIR system is composed of an AHL synthase, LuxI, and an AHL sensitive transcription factor, LuxR. Its operations are depicted in Figure 1-3. Freely diffusing across the membrane, AHL extracellular concentration is reflected intracellularly and vice versa. Given sufficient AHL with which to complex, LuxR becomes less prone to degradation, accumulates, and promotes the expression of both itself and LuxI, increasing AHL synthesis, thereby completing an intracellular positive feedback loop [21]. Furthermore, this sharp increase in AHL synthesis leads to a corresponding increase in local extracellular concentration, coordinating the QS activation of neighboring cells which in turn influence their own neighbors, generating positive feedback at a multicellular scale. This system core and variations thereof represent the dominant pathway by which AHLs influence downstream phenotypes.

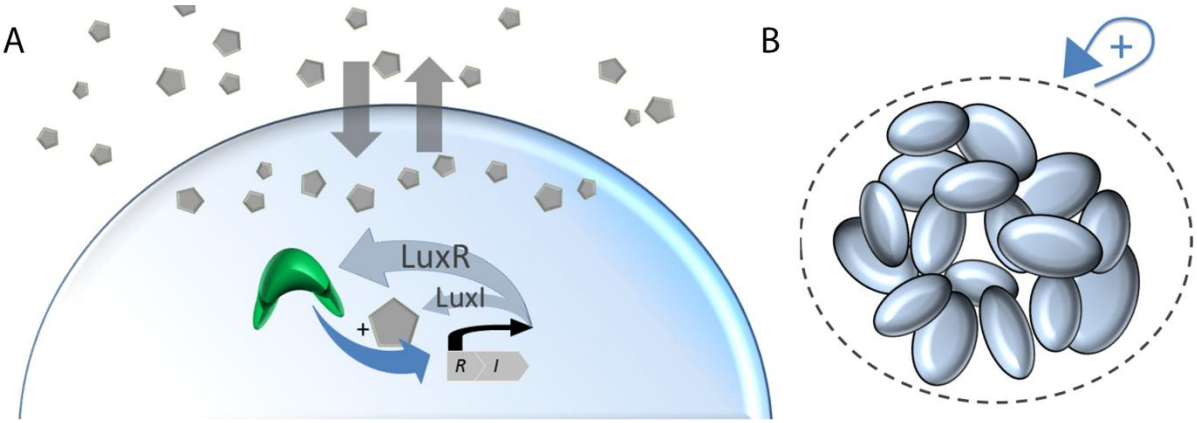


Figure 1-3. LuxIR activity at single cell and multicellular hierarchies. *A.* AHL freely traverses the membrane and stabilizes the transcription factor, LuxR. This activates a positive feedback loop, increasing LuxR expression as well as that of LuxI, the AHL synthase. *B.* This positive feedback operates at the multicellular level as well.

LuxIR homologs are widespread throughout proteobacteria. As of 2008, 26% of sequenced bacterial genomes contained a complete set of homologs [12]. Among them are the pathogens *Yersinia pestis*, *Agrobacteria tumefaciens*, and *Pseudomonas aeruginosa*. Many species contain multiple and distinct *luxIR* pairs [12], which were likely acquired independent of each other even if their signaling converges downstream [32].

Effective at organizing population expression, *luxIR* systems have often been repurposed toward applied ends. Initial efforts usually involved adding exogenous genes under the *luxR* promoter, such as an enzymatic degrader of AHL to create an oscillating signal [33] or chloramphenicol to limit population density despite sufficient nutrient in the environment [34]. In certain cases, *luxIR* QS was used to ensure complete induction within a population [35]. More advanced designs included separating *luxIR* sensing and autoinducer production capacities into multiple bacterial carriers whether for synthetic biology or therapeutic ends [36,37].

In addition to complete *luxIR* pairs, most species contain more *luxR* homologs than *luxI* homologs.[12] Possibly, at least some of these unmatched *luxR* homologs are not redundant, but serve as a means of testing for different species' AHLs. This is the case for *Rhizobium leguminosarum* bv. *viciae* which depends upon exogenous N-(3-hydroxy-7-cistetradecenoyl)-L-homoserine lactone (3-OH-C14:1-HSL) to induce conjugation [38]. The *luxR* homolog, *sdiA*, found in *E. coli*, *Salmonella*, and related enterobacterial species [39], also appears to support such a role. While these species do not synthesize any AHL, they nevertheless respond to several AHLs through SdiA [40–42]. As might be expected of a quorum sensing regulator, *sdiA* is at least in some species tied to biofilm activity [43], although the extent of its reported signaling has varied depending on whether the study

involved a knockout [44], plasmid based overexpression [42], or reporter fusion into the chromosome [45].

1.1.3 AI-2 QS

Perhaps even more broadly distributed than AHL based associated QS is AI-2 driven QS. However, this is not without caveats. For whereas many bacteria have AHL receptors without producing the specific AHL, many bacteria appear to produce AI-2 without an apparent ability to perceive the signal. This inversion has previously led some to suggest that AI-2 may represent a metabolic by-product rather than a true signaling molecule. Indeed, LuxS in combination with Pfs are an integral part of regenerating homocysteine after methyl donation as part of the activated methyl cycle, as seen in Figure 1-4 [46]. This pathway serves as an alternative to S-adenosyl homocysteine (SAH) hydrolase (SahH). While eukaryotes use the SahH pathway exclusively, the proportion of bacteria expressing *sahH* is roughly split with that utilizing the luxs/pfs pathway [46]. With only the rarest exception are *luxS* and *sahH* ever coincident to the same genome [27,47]. Importantly, bacteria containing *sahH* are unable to produce AI-2 except by conversion from ribulose-5-phosphate [48,49]. Whether the flux from this secondary synthesis pathway is sufficient to effect even limited QS operations remains unknown [49]. Still, while AI-2 production may not be universal, *luxS* is nonetheless well represented among bacterial genomes, and like AHL, AI-2 signaling is not restricted to the bacteria from which it is synthesized. Moreover, whether various combinations of exogenous *in vitro* synthesized [27] or purely synthetic [50] AI-2 and exogenous *sahH* expression can rescue *luxS* mutant phenotypes provides an interesting point of study.

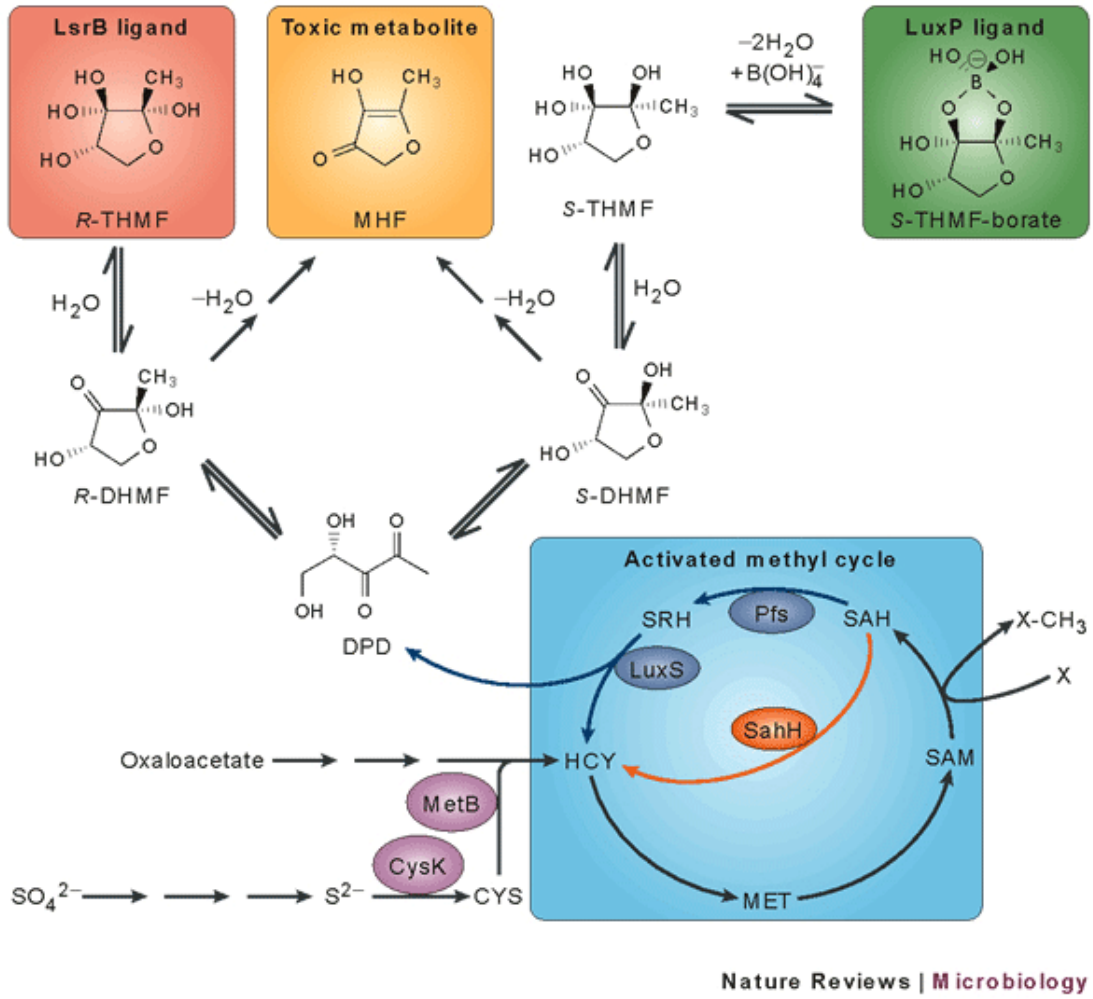


***Figure 1-4. The activated methyl cycle and the derivation of DPD* (from [51]).** AI-2 is an umbrella term for a number of spontaneously cyclizing DPD derivatives. DPD itself is a byproduct of the activated methyl cycle during the regeneration of homocysteine, at least in some bacteria. Most other cells with an activated methyl cycle utilize the alternative, SahH, which bypasses the production of DPD.

Less phylogenetically dispersed than AI-2 synthesis, receptors for the molecule are nonetheless found in numerous bacteria. Predominantly counted among bacteria with receptors are gamma-proteobacteria, for which the *luxS* transgenomic signal appears to be the most monophyletic [46,52]. This is partly attributable to LuxP receptors found among numerous *Vibrio* and *Marinomonas* species, as previously discussed. Among non-*Vibrio* and non-*Marinomonas* species, the only other known AI-2 QS receptor is the periplasmic receptor protein, LsrB, the binding component of an ABC type transporter. Originally identified in *Salmonella* along with the rest of the LuxS regulated (Lsr) system [53], LsrB has been shown through reciprocal BLAST searching to be most prevalent, but not universal in, nor restricted to Pasteurellaceaes and Enterobacteriales [13]. Just as *Vibrios* and *Marinomonas* species exist in a shared environment, it could be suggested that LsrB has been transferred horizontally among bacteria sharing the same environment [14].

In its sum, the Lsr system acts as a sugar-like importer system, composed of divergently arranged operons that share an intergenic region, *lsrRK* and *lsrACDBFGE* in *Salmonella* [53,54]. *lsrE* is absent in *E. coli*, which otherwise has a nearly exact duplicate [55–57]. *LsrR* belongs to the *deoR* family of transcriptional repressors, interacting with the intergenic region through a helix-turn-helix motif, keeping system expression to a minimum when sufficient substrate is unavailable. LsrA is the nucleotide binding component of the ABC type transporter, binding and using ATP to drive AI-2 import. LsrC and LsrD are transmembrane proteins comprising a pore for AI-2 between the cytosol and the periplasm. As described previously, homology indicates that LsrB binds AI-2 in the periplasm, where the complex joins to the transmembrane components. Once inside the cytoplasm, AI-2 is phosphorylated by LsrK. Notably, *lsrK* appears to have its

own constitutive promoter, albeit one that does not drive significant expression [56]. Phosphorylated AI-2 (AI2-P) de-represses system expression by destabilizing LsrR's interaction with DNA. AI2-P is also broken down into multiple products by LsrF and LsrG. The exact function of *lsrE* is unknown, although it is commonly alternatively annotated as ribulose-3-phosphate epimerase. This, and its operon position suggests it may operate on AI2-P [51]. Unlike *lsrF* and *lsrG*, however, it fails to effect apparent AI2-P levels [54].

The Lsr system is also affected by multiple factors outside its apparent regulon. For example, cAMP is required for system expression [55]. Additionally, AI-2 export requires the transmembrane protein YdgG, without which AI-2 remains sequestered within the cell [58]. YdgG appears to belong to a well conserved superfamily of exporters that have no other ascribed function [59]. Finally, it is believed that an additional low flux importer pathway is required for import when the Lsr system is inactive [55,60].

A few additional proteins have also been shown to mediate AI-2 activity at the cell membrane. These include RbsB from the ribose importer system in *Aggregatibacter actinomycetecomitans* [61], TlpB which allows *Helicobacter pylori* to be chemotactically repelled by AI-2 [62], and the PTS system in *E. coli* [63]. While TlpB is a broad array chemoreceptor, both RbsB and PTS are involved in the import of sugars across the membrane. RbsB was shown to compete with LsrB for AI-2 and that it was required for Lsr activation [61]. In a certain sense, however this is unsurprising insofar as many high affinity transporters for one monosaccharide act as low affinity transporters for related monosaccharides. For example, GalP imports glucose and galactose at high rates, and

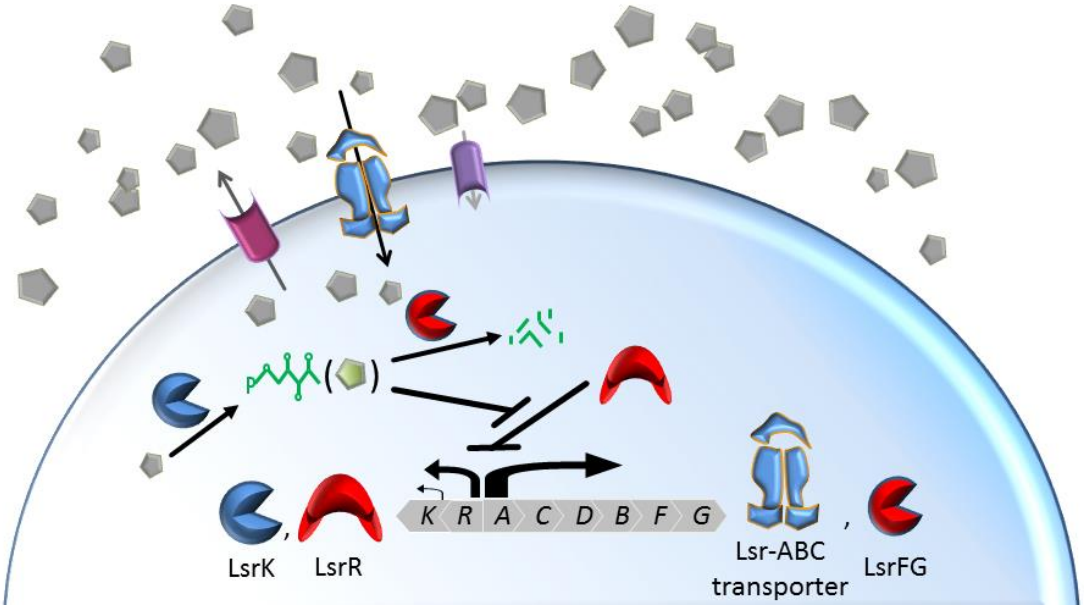


Figure 1-5. The Lsr system found in many E. coli. The ABC-type transporter and LsrK serve as positive feedback elements, working serially to produce AI-2P. LsrR and LsrF/G, on the other hand, serve as negative feedback elements, encouraging system repression or discouraging system de-repression respectively. Outside of these elements exist the transporters YdgG (an exporter) and a low affinity alternative importer pathway (consisting, at least, of the PTS system).

lactose and fructose at lower rates when both of its primary targets are absent [64,65]. This also obtains for ALBP, which primarily transports allose across the membrane but also doubles as a low flux pathway for ribose [66]. Both allose and ribose contain aldehyde functional groups. That both ribose and AI-2 are both 5 carbon molecules lends credence to the possibility of such a secondary role for RbsB. While RbsB is at least indirectly involved in *Aggregatibacter actinomycetemcomitans*'s AI-2 response [67], the regulatory pathway is unknown. Possibly, the *rbs* system comprises the alternative low flux importer affecting Lsr response, but this remains a subject of further research. Interestingly, a system annotated as *rbs* in *Haemophilus influenzae* also appears to regulate AI-2 influx. Additionally, this *rbs*'s expression is also controlled by AI-2. In the reported studies, none of the pentose sugars tested, including ribose, induced competitive inhibition, however, indicating that this may simply be a case of improper annotation [68], serving as a reminder of possible annotation biases associated with precedence. Many of the same ideas regarding AI-2 cross-reactivity with *rbs* apply to the PTS system which serves as a flux pathway for a variety of sugars, and upon which Lsr signaling is dependent in *E. coli* [63].

1.1.4 Lsr regulon

The exact pathway by which signal is transduced by the Lsr system appears to vary from species to species. Two separate studies have assessed the nature and number of potential LsrR binding sites. According to ChIP-CHIP analysis, the LsrR regulon is limited to the Lsr system itself in *Salmonella* Typhimurium [69]. In *E. coli,* the LsrR regulon is ostensibly more expansive. While the footprinting differential between *lsrR* knockouts and wildtype cells is strongest near the *lsr* intergenic region, additional

binding sites include regions near *yegE*, *mppA*, and *yihF* [70]. Both *yegE* and *mppA* assist in the transition from exponential growth toward a sessile lifestyle. A probable diguanylate cyclase, *yegE* appears to help coordinate the concomitant downregulation of flagellar genes and upregulation of *curli* [71]. Separately, *mppA* is involved in recycling membrane associate peptides [72].

Additional means by which Lsr may influence downstream phenotypes exist in the form of LsrF's and LsrG's enzymatic outputs. While LsrF is believed to be an aldolase, its specific products remain unknown [73]. LsrG is thought to act as an isomerase under anaerobic conditions to produce 3,3,4-trihydroxy-2-pentanone-5-phosphate [74], whereas it cleaves AI2-P under aerobic conditions to produce 2-phosphoglycolic acid (PG) and an unidentified 3 carbon molecule [75]. PG acted upon by the phosphatase *gph* reenters metabolism as glycolic acid [76]. Whether any of these products affects downstream phenotypes remains unknown.

1.1.5 Known Lsr Associated Phenotypes

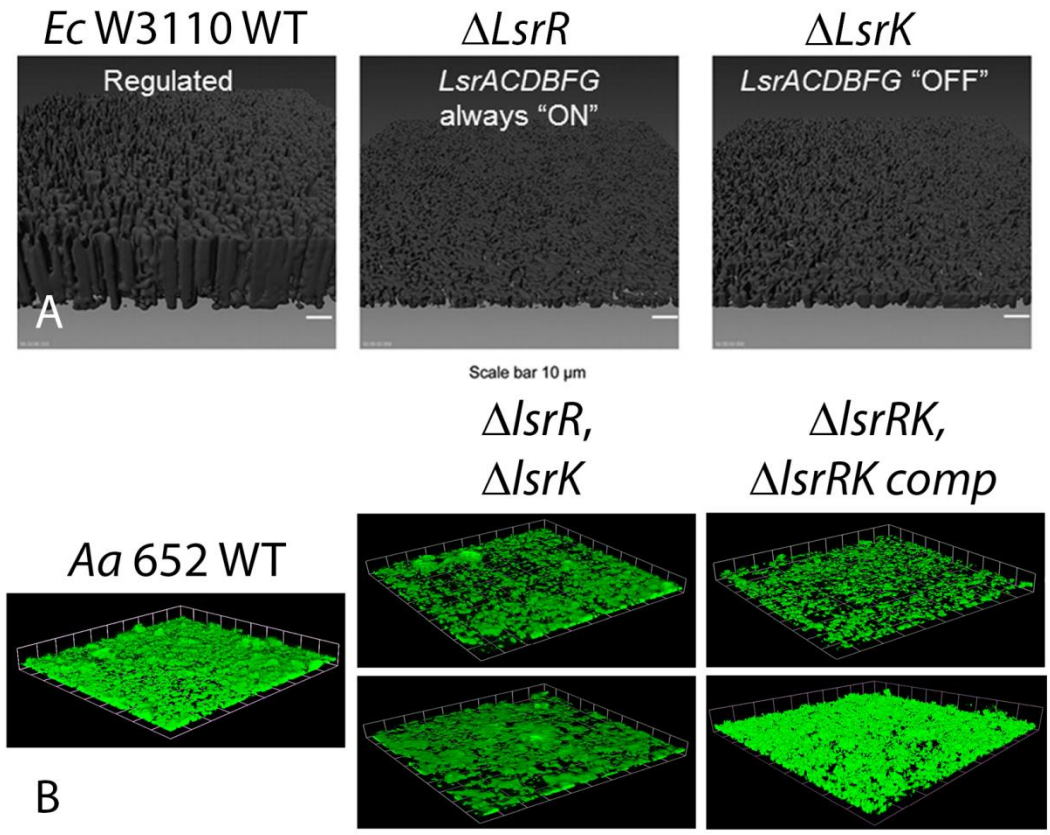
The reported effects of LsrR mutation and other Lsr system modifications mirrors the ChIP-CHIP results. LsrR did not have an apparent impact on the measured phenotypes, *invF* and flagella gene expression, in *Salmonella* Typhimurium except when *luxS* was also deleted or when LsrR was overexpressed [77]. This suggests that LsrR binding to DNA may be easily destabilized by the presence of AI2-P in *Salmonella*, perhaps pointing to a reason why LsrR's direct regulon may be limited in this species. On the other hand, in *E. coli*, deletion of either *lsrR* or *lsrK* led to marked inhibition of biofilm development [60] (Figure 1-6a). Other experiments involving *luxS* knockouts with

homocysteine and SAM controls indicated that exogenous DPD restored biofilm height while not fully complementing biomass accumulation [78]. Interestingly, biofilm phenotype was also inhibited by *lsrR* deletion in *Aggregatibacter actinomycetemcomitans*, a bacteria common to the oral cavity [79]. In this case, although *lsrRK* knockout led to more dramatic changes than *lsrR* deletion alone, statistically significant changes arising from *lsrK* deletion remained unidentified, leading to a suspicion that an additional kinase may be working on cytosolic AI-2 [79] (Figure 1-6b). This small cohort of studies indicates that the Lsr system can influence population scale phenotypes in a manner expected of QS systems.

1.1.6 Broader AI-2 Associated Phenotypes

In general, QS drives behaviors that work most efficiently when orchestrated at multicellular scales [80]. Along with swarming and sporulation, among the more strikingly cooperative of QS guided behaviors is biofilm formation. In addition to the apparent cooperativity involved, biofilms are also known for their protective role, which lends to bacteria's robustness against environmental insults (e.g. oxidation and antibiotics) and also plays a role in refractory infections. This is a common theme in QS signaling, which appears to influence numerous aspects of virulence including motility switching, bacterial adhesion, invasion, and toxin production. More broadly, AI-2 signaling might be expected to control numerous aspects of bacterial interactions in natural consortia.

In addition to the species previously mentioned, AI-2 in particular has been shown to influence biofilm formation and other cooperative behaviors in many bacteria, often in a manner similar to that in *E. coli*, where, as discussed previously, exogenous



322

323 ***Figure 1-6. Lsr system effects biofilm development in multiple species.*** *A.* As adapted
324 from [60], *E. coli* biofilm development appears stalled upon deletion of either *lsrR* or *lsrK*.
325 *B.* The same appears to be true for *A. aggregatibacter*, except the effect from *lsrK*
326 mutation was not statistically significant. Dual deletion resulted in a sparser biofilm, and
327 complementation on a plasmid restored that defect.

328

329

330

331

DPD partially complemented *luxS* deletion. This inexact restoration may be attributable either to the dual signaling and metabolic role played by LuxS or to the transport limitations of exogenous DPD supplementation at both macroscales and microscales (i.e. around biofilm architecture or across the cell membrane).

A couple studies have focused on bacterial species synthesizing AI-2 and containing Lsr homologs, without explicitly studying Lsr based expression. DPD complemented *luxS* mutation in *Photorhabdus luminescens* for some differentially expressed genes to an unspecified degree while others were unaffected. Fully or partially restored expression was found in virulence genes such as hemolysin and pyocin, oxidative response genes such as *msrB* (Peptide methionine sulfoxide reductase B) and *uvrC* (Exonuclease ABC subunit C), and motility related genes such as *flhC* (Flagellum biosynthesis transcription activator). In this report researchers were additionally unable to duplicate DPD induced changes with homocysteine supplementation.[81] In *Bacillus cereus*, exogenously added AI-2 affected biofilm growth in a closely titrated manner, initially increasing biofilm accumulation but also inhibiting biofilm growth when exposed to a higher concentration within the same order of magnitude [82].

A separate group of bacteria producing AI-2 but lacking a known AI-2 receptor also appear to have AI-2 sensitive phenotypes. Exogenously added DPD increased adhesion in a dose dependent fashion in both *luxS* mutant and wildtype *Actinobacillus pleuropneumoniae* populations [83]. In *Streptococcus pneumonia*, researchers were able to demonstrate partial reestablishment of biofilm biomass to wildtype levels in a *luxS* mutant treated with exogenous DPD [84]. In *Streptococcus epidermidis*, exogenous AI-2 complementation restored PSM (pro-inflammatory phenol soluble modulin peptides)

expression in addition to that of many metabolic genes that were differentially expressed in a *luxS* knockout compared to a WT strain [85]. In *Streptococcus intermedius*, AI2 exogenously added to *luxS* mutants again produced a distinct biofilm phenotype compared to WT and *luxS* mutants in both the presence and absence of antibiotics. Moreover, AI-2 conferred some ability to form a basal level of biofilm in the face of several classes of antibiotic.[86] Among these reports, one involving *Borrelia burgdorferi* serves as a particularly illuminating example. For this species and in many of its relatives, the activated methyl cycle is incomplete. In particular, *B. burgdorferi* cannot generate methionine from homocysteine. Nonetheless, AI-2 is produced through the LuxS/Pfs pathway, and moreover, DPD can complement *luxS* deletion, restoring expression of *vlsE* and *erp*, both of which translate into proteins that mediate the host-pathogen interaction.[87] In addition to its ability to chemorepel from AI-2, *H. pylori* also appears to regulate multiple flagellar genes with AI-2, as AI-2 complements *luxS* deletion and also enhances these flagellar gene expression levels when added to wildtype cells. [88]

A number of other bacteria that do not synthesize AI-2 also respond to exogenously supplied DPD. Ranking among these is *Sinorhizobium meliloti*, which contains an Lsr homolog [14] and can deplete AI-2 from the extracellular environment without apparent consequence to colonization or growth phenotypes [89]. *Pseudomonas aeruginosa* expression of virulence genes is influenced by exogenous AI-2 [90]. Finally, DPD was shown to influence biofilm production in a dose responsive fashion via the induction of an oxidative stress response pathway in *Mycobacterium avium* [91]. Unlike *S. meliloti* neither *P. aeruginosa* nor *M. avium* has an identified AI-2 response pathway.

1.1.7 QS in an Ecological Context

From an evolutionary point of view, QS and AI-2 based-QS in particular represent a peculiarity due to perceived, or at least potential, cooperation between nonrelated bacterial species. By way of illustration, pure culture QS signaling that encourages biofilm formation is an entirely suitable method by which to organize population-wide expression of public goods in an isolated setting. In such instances, QS based cooperation can be attributed to kinship.[92] When considering QS related mutations and the likely mixture of genetic backgrounds and species within a bacterial consortium, however, the exact manner in which evolutionary pressures inform and maintain QS signaling becomes less straightforward. In even moderately more complex circumstances, questions of which bacteria produce signal, which bacteria respond to signal, and which bacteria benefit from any resulting product can quickly overwhelm the boundaries of current understanding.

1.2 Research Motivation

Lsr based QS in *E. coli* represents a practical model for studies herein. Among the better characterized of the Lsr systems [49,55–58,60,63] and effecting virulence phenotypes [60,78], the study of such a system is of natural interest for a diverse generalist bacteria like *E. coli*, certain strains of which are a common source of food supply contamination. Arising from such concerns, the identification of strains containing Lsr system homologs and how Lsr based QS might play a role in infection processes among these strains is of interest. Yet even more broadly, it is important to identify other species in which the Lsr system may play a signaling role. While much research has been conducted to identify AI-2 responsive bacteria, sometimes through the Lsr system, questions of generalizability, while somewhat epistemological in nature, are also useful when answers can be had. By the identification of useful homology, and determining how homologs are constituted, one may begin a line of inquiry into questions of generalizability.

The phylogenetic dispersion of AI-2 receptors has been assessed as a means of determining which bacteria might be responsive to AI-2 as an external signaling stimulus [13]. The effectiveness of such searching is largely dependent upon the increasingly rapid pace of whole genome sequencing, and while many such receptors have been identified there remain numerous AI-2 responding bacteria with unknown receptors. While it is possible additional sugar transporters moonlight as high affinity AI-2 transporters, sugar transporters acting as a lone pathway for AI-2 import has yet to be demonstrated in any species. In general, the perception of AI-2 is of natural interest, as some cells may recognize it at lower concentrations than others, etc; and here, perception is meant to not only indicate reception, but also encompasses subsequent gene expression changes. Here

421 we wish to only consider possible Lsr gene expression changes, while remaining aware
422 that the Lsr system may effect a broader regulon. Toward this end the identification of
423 extant Lsr genes, operon organization [93], and the phylogenetic distribution of variants
424 among Lsr homologs is of interest.

425 A second cohort of studies described herein were initially motivated by
426 experiments indicating that stable bimodal Lsr system expression may arise in pure
427 cultures [94] (Figure 1-7a). While it has been suggested that a transient bimodality may
428 arise as a consequence of intracellular signaling topology [95], a more permanent
429 population bifurcation may also develop from hyperlocal competition for AI-2 between
430 cells, as depicted in Figure 1-7b. Bimodal expression within a pure culture would
431 represent a built-in population of QS cheats, and depending on the expected benefit could
432 speak to issues about free riders and other evolutionary conundrums.

433 The picture is further complicated for bacteria that are chemoattracted to AI-2,
434 such as *E. coli* [96,97]. Although QS generally informs population scale behaviors,
435 activation of Lsr and the resulting net recompartmentalization occur on time and spatial
436 scales that intersect with chemotaxis. As a means of isolating this confluence, a mixed
437 finite difference-agent based approach can be used to model emergent behaviors [98].
438 Using such a platform, contextualized QS expression between two competing populations
439 with different genotypically mixed signal/reception competency can be interrogated.
440 Specifically, as a possible consequence of its universality, the literature suggests that AI-
441 2 signaling within a consortia is potentially quite complex. How QS operates in an
442 environment with multiple systems operating concurrently remains an open question.
443 Thus, in addition to investigating QS by pure cultures, this framework can also be used to

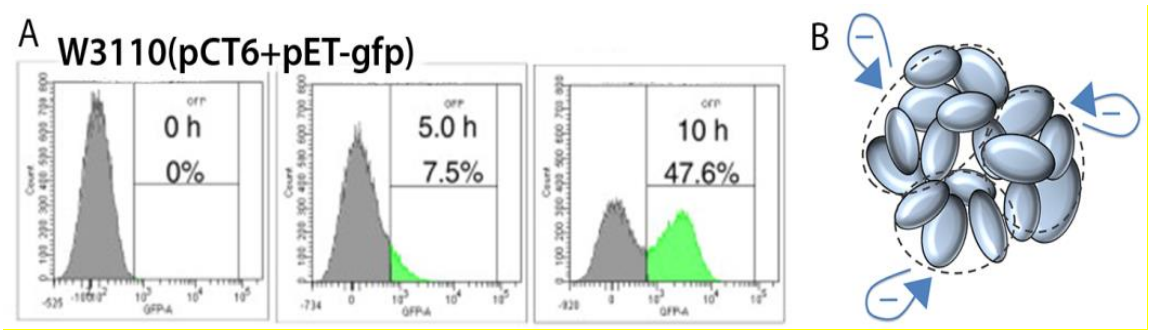


Figure 1-7. Possible bimodality arising from Lsr activity. *A.* FACS results adapted from [94] suggest that a bimodality develops over time. *B.* A diagram of a possible mechanism whereby bimodality could develop, as extracellular negative feedback operates hyperlocally.

investigate basic QS systems acting concurrently in both time and space. Specifically, how different Lsr system variants might compete with one another can be investigated.

Essentially, in addition to determining which species and strains might be responsive to AI-2, there exists additional opportunity to ask how they might be responsive to AI-2. Research presented herein takes a closer look at the phylogenetic distribution of Lsr system homologs and the composing members of its homologs. We then consider more closely how the Lsr system possibly produces bimodal expression through the mathematical modeling of two competing populations. We further consider what the ramifications of system behavior might be in the context of motility, and how multiple bacteria with different Lsr system variants might compete for AI-2 with one another.

1.3 Global Objective, Global Hypothesis, and Specific Aims

The global objective of this dissertation is to further investigate the nature of the Lsr system, using *E. coli* K-12 W3110's Lsr system as a starting point, both interrogating how W3110's Lsr homologs compare to each other and how they compare to other QS systems.

Global Hypothesis: *E. coli*'s Lsr system is one among a spectrum of Lsr systems, the diversity of which bears signaling consequences both in pure cultures and mixed consortia.

Specific Aim 1: Find *E. coli* Lsr system homologs.

Specific Aim 2: Determine if the known components of the Lsr system, along with auxillary proteins can lead to stable, bimodal Lsr activation.

Specific Aim 3: Assess how Lsr's recompartmentalization dynamics may differentiate it from other QS systems within the context of population activation, emergent phenomenon, and competition between bacteria containing different Lsr systems.

1.4 Dissertation Outline

Chapter 2 describes a novel algorithm to find homologs of modular networks and an analysis of the results when using the *E. coli* K-12 Lsr system. This is contrasted to the results for the *lac* system from the same strain.

Chapter 3 describes the development of a set of ODE's that describes the development of bimodal system expression given parameter variation.

Chapter 4 describes the development and results of a finite difference-agent based model contrasting LuxIR activation to Lsr activation within the context of emergent behaviors arising from the confluence of QS and varying motility modes.

Chapter 5 summarizes the previous chapter's work and indicates challenges to future work, denoting the larger implications and significance of the work.

Chapter 2: Comparison of Homolog Identification for the sugar importing Lac System and the QS Lsr System from *E. coli* K-12 W3110

2.1 Abstract

Bacterial cell-cell communication is mediated by small signaling molecules known as autoinducers. Importantly, autoinducer-2 (AI-2) is synthesized via enzyme LuxS in over 80 species, some of which mediate their pathogenicity by recognizing and transducing this signal in a cell density dependent manner. AI-2 mediated phenotypes are not well understood however, as the means for signal transduction appears varied among species, while the AI-2 synthesis process appears conserved. Approaches to reveal the recognition pathways of AI-2 will shed light on pathogenicity as we believe recognition of the signal is likely as important, if not more, than the signal synthesis. LMNAST (Local Modular Network Alignment Similarity Tool) uses a local similarity search heuristic to study gene order, generating homology hits for the genomic arrangement of a query gene sequence. We develop and apply this tool for the *E. coli lac* and LuxS regulated (*lsr*) systems. Lsr is of great interest as it mediates AI-2 uptake and processing. Both test searches generated results that were subsequently analyzed through a number of different lenses, each with its own level of granularity, from a binary phylogenetic representation down to trackback plots that preserve genomic organizational information. Through a survey of these results, we demonstrate the identification of orthologs, paralogs, hitchhiking genes, gene loss, gene rearrangement within an operon context, and also horizontal gene transfer (HGT). We also found a variety of operon structures that are consistent with our

hypothesis that the signal can be perceived and transduced by homologous protein complexes, while their regulation may be key to defining subsequent phenotypic behavior.

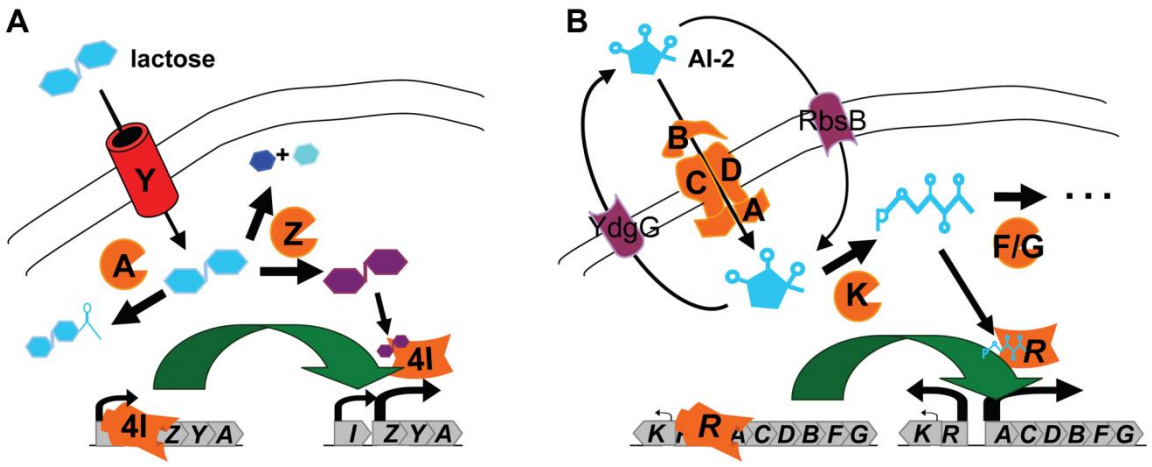
2.2 Introduction

Comparing prokaryotic whole genome sequences to identify operons is a mature area of research [99–102]. Orthologous operon identification can imply a secondary degree of relation between components, reaffirming Clusters of Orthologous Groups (COG) and other assignments of function as well as suggesting essentiality [103]. This conservation of components also speaks to the conservation of signaling capacity in orthologous modular signaling operon-based units. That is, we are interested in ascertaining the genetic modularity of signal transduction processing, in particular those that operate within known, putative regulons. Drawing partly on previous work investigating microsynteny and gene neighborhoods [101,104,105], we developed a general similarity search approach, we call a Local Modular Network Alignment Similarity Tool (LMNAST). LMNAST applies a BLAST-like heuristic to gene order and arrangement. Resultant search hits help capture the conservation and phylogenetic dispersion of a given query modular network.

Using, as queries, contiguously abutting genes of prokaryotic modular signaling networks, LMNAST identifies and scores hits based on the minimum number of frank mutations in gene organization needed to arrive at a given putative system homolog starting from the query. Here, homology refers to similarity in relative gene order and relative transcriptional direction, after nucleotide level threshold filtering of gene elements based on BLAST [106] E-value.

For the purpose of evaluation, two small modular systems were used as test inputs: one was the *E. coli lac* system and the other was the LuxS regulated (Lsr) system. In some ways, the two systems are quite similar (Figure 2-1). Both import and

catabolize the small molecules that induce system expression. For the *lac* system, this small molecule is, of course, lactose. For the Lsr system, the small molecule is autoinducer-2 (AI-2). AI-2 is a signaling molecule common among at least eighty bacterial species [27]. As mediated either through the Lsr system or LuxPQ, bacteria are believed to use AI-2 to guide population based phenotypes, a phenomenon termed quorum sensing [54]. LuxPQ is a histidine kinase two component system, the regulon of which is distinct from Lsr and is not considered further. Lsr is an interesting query because its distribution should help elucidate its putative, modular quorum sensing function [27] and because the known homologs differ in gene organization.[54,55,82]



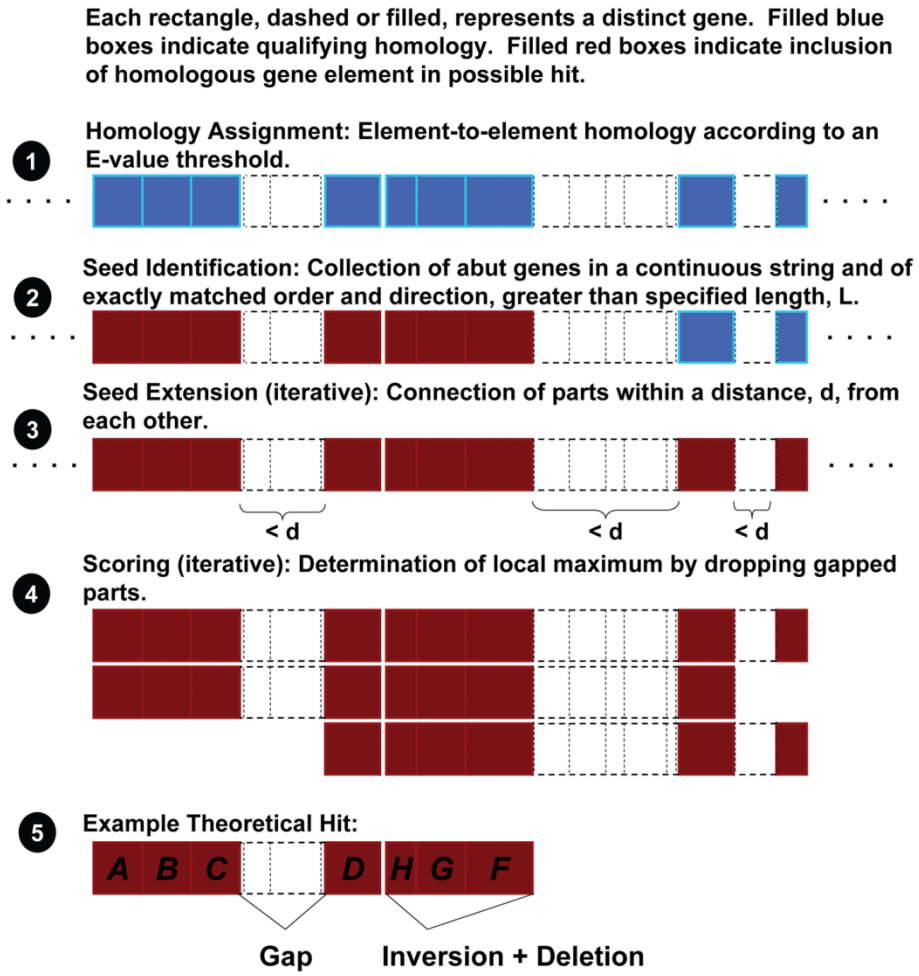
592

593 ***Figure 2-1. Test queries: lac operon and Lsr system.*** **A** The *lac* operon is composed of
594 beta-galactosidase (LacZ), the lactose importer (LacY), and a beta-galactoside
595 transcetylase (LacA). Upstream of the operon, the operon repressor (LacI) is expressed
596 in a co-directional orientation. The primary function of the *lac* system is as a regulated
597 importer/processing unit. Lactose brought in through LacY is converted into allolactose
598 or hydrolyzed into glucose and b-galactose. Both reactions are catalyzed by LacZ.
599 Allolactose then acts to release the repression of the system by LacI. **B** The Lsr system is
600 composed of two divergent operons. One operon consists of an AI-2 kinase, and a system
601 repressor. The other operon consists of an AI-2 transporter and phospho-AI2 (AI2-P)
602 processing genes. Contextual system behavior is partly governed by separately regulated
603 parts including an alternative importer [61], an exporter [58], and the AI-2 synthase gene.
604 Relative to the canonical *lac* system, the Lsr system is complicated by the fact that the
605 cell synthesizes, exports, and imports AI-2, and by the negative regulation associated
606 with the divergently arranged structure. AI-2 exported by a mechanism involving YdgG
607 traverses the outer membrane through a porin and enters the periplasmic space. Through
608 the ABC-type importer, LsrACDB, AI-2 is then transported back into the cytosol. Once

there, AI-2 is phosphorylated by LsrK. This phosphorylated form (AI2-P) de-represses the lsr system and is catabolized by LsrF and LsrG into separate downstream products.

2.3 Methods and Algorithm

As previously indicated, LMNAST evaluates nucleotide records for similarity to a query network using a BLAST-like heuristic. Specifically, it captures the gene order of query networks as a string of characters. Queries are therefore not necessarily restricted to defined networks insofar as any gene ordering may be a query. A standard heuristic of penalties for various rearrangements of orthologous system is employed. For searches described herein, a loose threshold was used to generate an exhaustive set of hits. The overall scheme is depicted in Figure 2-2. The program itself is available at http://www.bentley.umd.edu.



637

638 ***Figure 2-2. LMNAST heuristic.*** LMNAST operates in a BLAST-like manner, using the

639 results of BLAST searches themselves as a curated database.

640

641 1. For each member of the query, in any nucleotide record, a homolog's membership to a

642 character type is assigned by scoring below a specified BLAST E-value threshold. Genes

643 assigned to characters are highlighted blue. Genes without sufficient homology to any

644 character are represented by dashed boxes.

645 2. Sufficiently long stretches of adjacent characters are identified as seeds (red).

3. Sufficiently proximal characters are connected to seeds or seeds are connected to each other when at a base pair distance < d.

4. Rearrangements, losses, and deletions are scored according to a standard similarity heuristic. Noncontinuous elements are dropped iteratively until a maximum score is achieved, arriving at…

5. An LMNAST hit.

2.3.1 Input

Input consisted of an ordered list of gene elements (for example, *lacIZYA*). For each gene element a BLAST result file was generated using tblastn to search the nr/nt database for hits with E-values less than 0.1, narrowing the search space. Each BLAST hit was assigned a character corresponding to the gene element queried. BioPerl [107] was used to query Genbank databases and process data from retrieved files. Nucleotide records with sufficiently proximal characters were investigated further.

2.3.2 Scoring Heuristic

The degree of similarity between a putative hit and a query was evaluated according to the number of deletions, insertions, and rearrangements required to generate the putative hit using the query as a starting point. Intra-hit gene duplications were disallowed as a simplification. Consequently, deletion could be noted by character type inclusion. Insertions of uncharactered elements between gene homologs were scored according to an affine gap rule whereby a portion of the deduction was scaled to the insertion length. Rearrangements refer to altered relative order and relative gene direction. Changed relative direction was only considered when relative order was maintained. When this criterion was satisfied, relative order was evaluated in terms of adjacent homolog distance, disregarding insertions and deletions. For each such structural dissimilarity there was a standard deduction in score. Noncontiguous elements were dropped iteratively until a maximum score was reached for each putative hit. When putative hit versions elicited equal scores in the same round, the version of the hit with

the most characters was retained. Putative hits with scores greater than zero were retained.

2.3.3 Weak and Stringent Criteria

For evaluation purposes and to find a suitable balance between false positives and coverage completeness, each test query was run under both weak and stringent conditions. Stringent criteria searches assumed accurate annotation. Contrarily, weak criteria did not require genes to lie within the annotated coding sequence. Moreover, characters annotated as "pseudo" or bounded outside "gene" annotation were accepted as homologous characters. Weak criteria searches also allowed multiple genes to co-exist within the same annotation. Additionally, as a concession to the possibility of longer range interactions between genes, reduced gap penalties were used in weak criteria searches. Results described herein were derived using a gap penalty of 1 and 2 with an extension penalty of 0.3 and 1, for weak and stringent criteria searches respectively.

2.3.4 Ancillary LMNAST Search Tools

Mean element homology (meH) is a normalized, ancillary measure of string similarity as evaluated by BLAST. Useful for contrasting BLAST results to LMNAST hits, meH was calculated by normalizing each gene homolog's bit score to the maximum bit score for the entire corresponding BLAST result with a background subtraction of the minimum bit score. These normalized bit scores were then averaged for all gene elements within an LMNAST hit. A score of one indicates exact likeness whereas zero indicates the least degree of similarity.

Also, widening the query beyond the system of interest to include a nominal number of flanking genes, here termed “extended window searching,” afforded additional contextualization of LMNAST hit results.

Finally, in evaluating certain low homology hits, nonscoring synonyms were used. Nonscoring synonyms are elements with equivalent gene annotation but insufficient homology according to the initial E-value filter. This is somewhat analogous to replacement in blastp.

2.4 Results

2.4.1 *E. coli* K-12 W3110 *lac* Operon Query

We began evaluation of LMNAST by searching for the well characterized *E. coli lac* operon. Specifically, the *E. coli lac* genes *lacI* (BAE76127), *lacZ* (BAE76126), *lacY* (BAE76125), and *lacA* (BAE76124) (spanning bp 360473 to 366734 of the Genbank nucleotide record AP009048) were used as a query. The stringent criteria search yielded fewer hits than the corresponding weak criteria search (189 vs. 236). Of the hits derived from the stringent criteria search, complete and perfectly arranged *lac* systems were found in 26 unique *E. coli* strains and *S. enterica* arizonae serovar 62:z4,z23 (meH 0.8), the only *Salmonella enterica* serovar represented among all *lac* system hits, in keeping with its significant divergence from other serovars [74]. A representation of *E. coli* hits in a phylogenetic context is available in Figure 2-3a. The average meH ($0 < meH < 1$) for these complete systems was 0.98. An extended window query with five additional genes on either side of the original search frame, revealed eight complete systems with a hitchhiking, proximal cytosine deaminase after losing all other proximal genes. Only one system with all four characters was entirely removed from the original query's proximal gene set, suggestive of negligible stability for the canonical system outside of a highly limited phylogenetic domain.

An additional 28 hits were bereft one *lac* system character (average meH 0.74). In all but three of these cases that missing gene was *lacA*. Of the hits without *lacA*, ten had an additional frank structural change to a divergent expression pattern originating

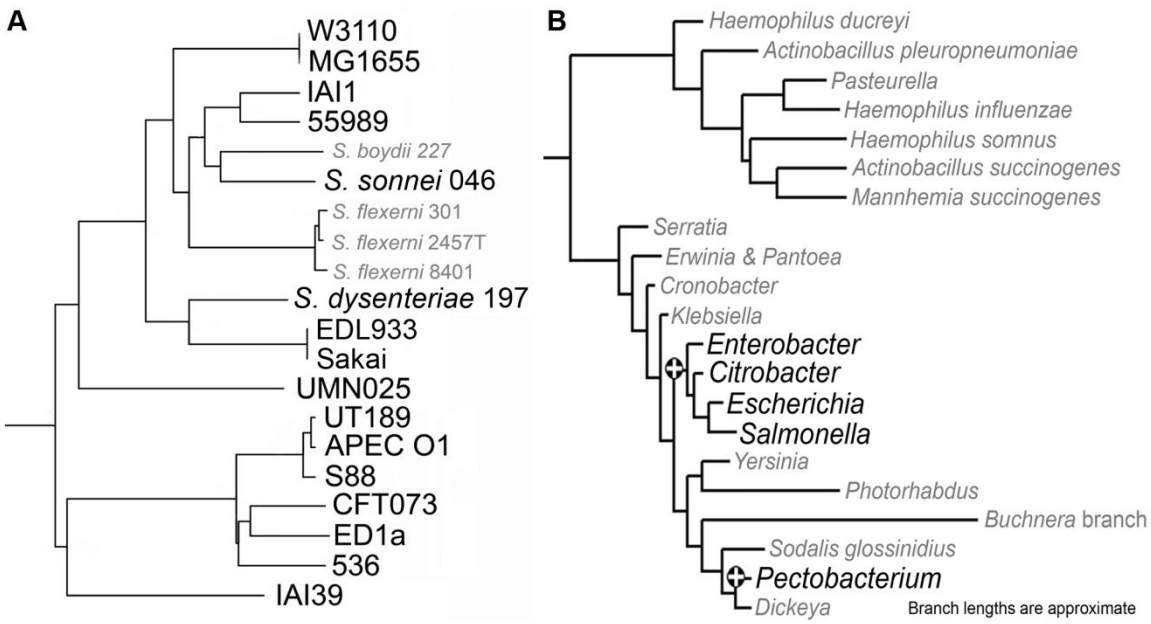


Figure 2-3. lac operon LMNAST hits overlaid onto phylogenetic distributions of different scopes. The larger, bolded leaves represent species that contain *lac* operon homologs, whereas the grayed italicized leaves were completely bereft. **A** An *E. coli* specific phylogenetic tree as adapted from [102], wherein all genes from the core E. coli were used to construct a consensus tree. Among the species/strains represented here, *lac* system homologs were absent from certain *Shigella.* Additionally, BW2952, SE11, IAI1, HS, E243227A, CFT073, K-12 DH10B, and 11 of 18 O:H serotyped strains contained truncated systems. Uniquely, CFT073 retains *lacA*, while missing *lacI* in an otherwise preserved *lac* system structure. **B** The phylogenetic dispersion of the *lac* system is mostly limited to *Escherichia* and proximal species, as seen in the 16s based tree adapted from [108]. Bolded leaves indicate the presence of the *lac* system in at least one strain.

between *lacI* and *lacZ* characters (e.g. in *E. cloacae*), likely increasing system expression sensitivity to *lacI* repression in these cases [93]. Surprisingly, in other instances, extended window searching revealed the only proximal structural change to be a missing *lacA* gene. This *lacA* degeneracy may be indicative of its relative functional unimportance compared to the other *lac* system members [109].

Some of the patterns described above can be inferred from coincidence heat charts (Figure 2-4). These matrices represent LMNAST results by the frequency of coincidence between gene characters within hits. The shade of an index represents the frequency of hits where the row gene coincides with the column gene, normalized against the total number of hits containing the row gene, which itself is denoted by (#). For example, in Figure 2-4, the left-most matrix is a representation of a theoretical set of homolog fragments (AB, BC, CD, ABC, BCD, and ABCD). This simple set was constructed to only reflect unbiased homologous recombination presumably resulting only in chromosomal rearrangements. In this set, B and C are extant in five inputs, while A and D are extant in three inputs. All three inputs containing A also contain B, two also contain C, and one also contains D. This is reflected in the shades of the grids in the top row.

The middle matrix represents the coincidence distribution amongst LMNAST *E. coli lac* hits. As an additional example, matrix element (2,1) is a rust color representing the 139 hits with a *lacI* character of the 155 also containing *lacZ*. Finally, the right-most matrix is the difference between the left and middle matrices. This particular analysis

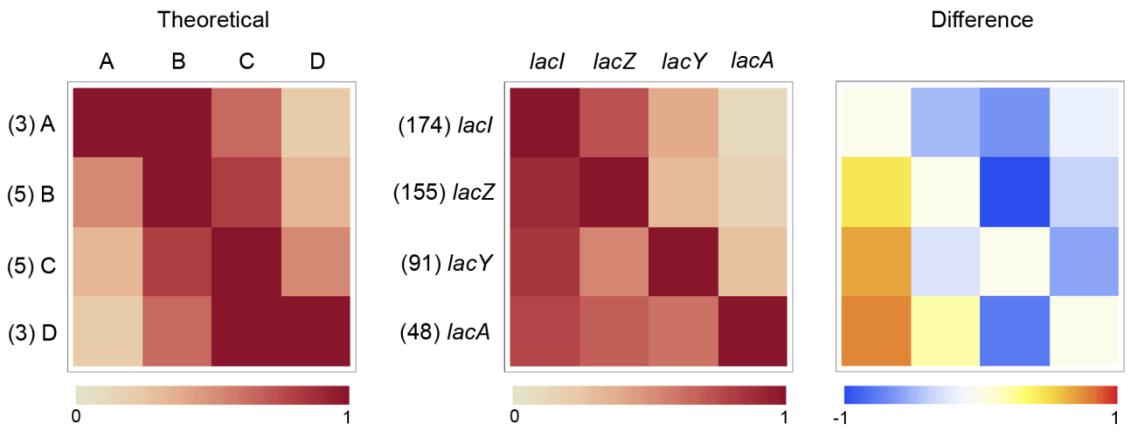


Figure 2-4. Coincidence heat map for lac operon LMNAST stringent search hits. Each shaded index represents the normalized frequency of hits containing the row gene that also contain the column gene out of the total number of hits containing the row gene, as denoted by (#). The matrix on the left is a representation of an unbiased set of evenly distributed homologs (AB, BC, CD, ABC, BCD, and ABCD). The middle matrix is the actual coincidence data. The matrix on the right is a heat map of the difference between the two. *LacI* is heavily over-represented according to the difference matrix. *LacY* occurred less than would be expected according to a random distribution.

suggests, for example, that *lacI* is relatively over-represented across all hits, and that nearly all other coincidences are under-represented; surprisingly, this includes coincidences involving the permease, *lacY*. Unlike *lacA*, *lacY* is believed necessary for lactose catabolism, possibly pointing to the use of a lower affinity transporter in such cases. Obversely, the over representation of *lacI* indicates a preference for the regulation of lactose catabolism.

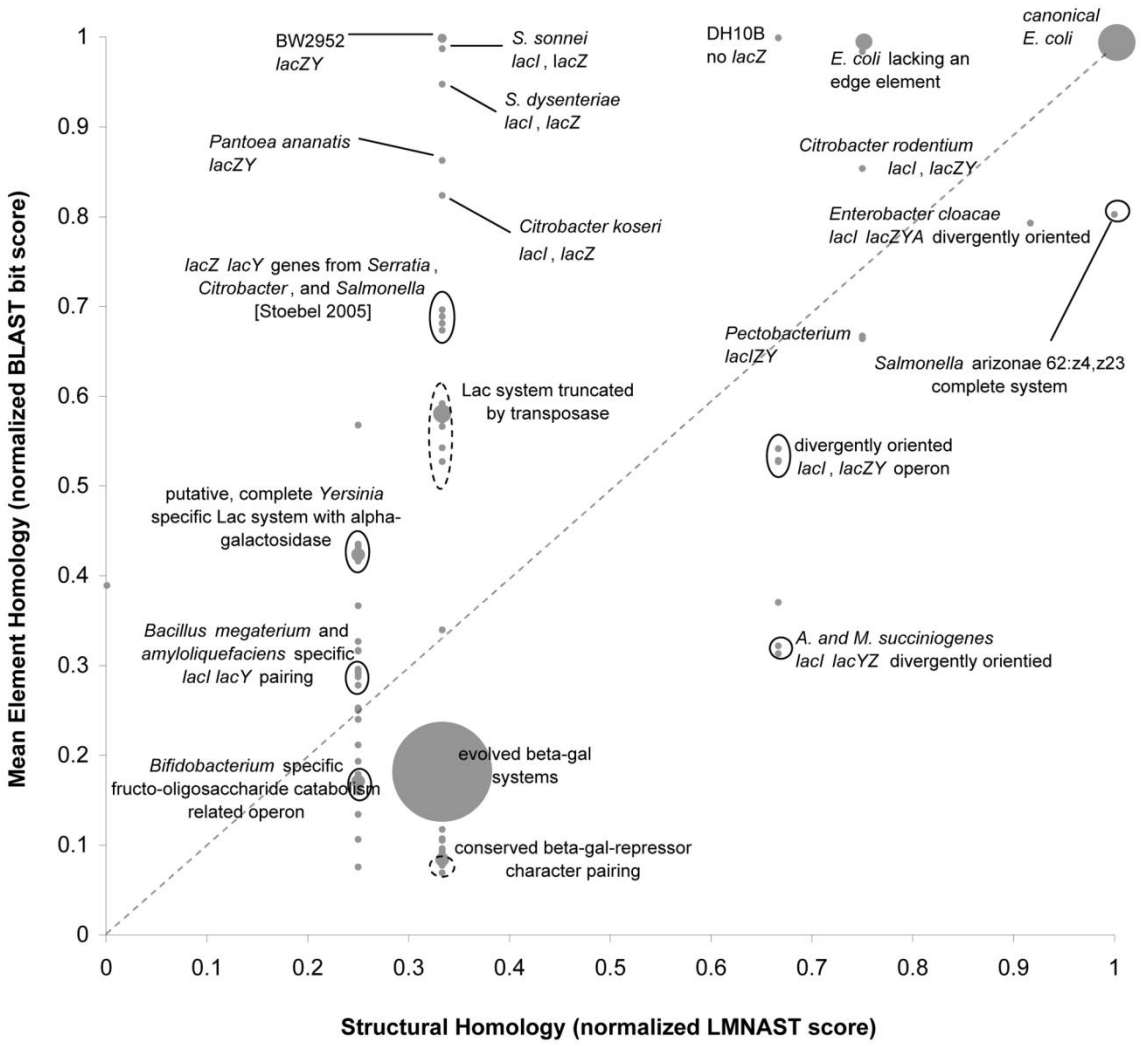
Of the strong criteria search results, 138 hits contained only two *lac* gene homologs (average meH 0.28). Two gene homologs represent the natural minimum of individual characters that a homologous system may contain. Such hits represented truncated systems, repurposed individual members, homoplasic convergence, or outright false positives. The majority of these hits fell within clusters of shared Genbank annotation in 2D similarity plots, which compare meHs (averaged BLAST homologies) against LMNAST homologies, or, put differently, average amino acid identities against the system's broader organizational identity. Generically then, purely vertical displacements imply perfect conservation across species through either vertical or, more likely, recent horizontal gene transfer accompanied by amelioration, while purely horizontal displacements indicate recent gene loss and/or rearrangement. For purposes of downstream analysis, it is interesting to speculate that the kinetics of the remaining genes are unaffected in cases of purely horizontal displacement. For systems subject to HGT, such liberties must necessarily be taken with less confidence.

In the case of the stringent *lac* search, similarity plots revealed a great deal of structural variability in the *lac* operon homologs of *E. coli* and near *E. coli* species (Figure 2-5). Nonetheless, the canonical *lac* operon (26) and the paralogous evolved

beta-galactosidase system (43) [110] are clearly the most dominant *lac* operon-homologs, perhaps partially reflecting the relative preponderance of fully sequenced *E. coli* strains.

Addressing the full breadth of two character homologs, 87 contained *lacZ* and *lacY* character types, all of which were adjacent, five of which were misdirected relative to one another. Numerous truncated systems had high meH but imperfect organizational similarity. This cohort was restricted to strains of *E. coli* and closely related *Shigella*, *Citrobacter*, and *Enterobacter* species, reflecting a generally confined phylogenetic breadth among LMNAST *lac* hits (Figure 2-3b), and reinforcing the idea of limited *lac* horizontal gene transfer (HGT) [111]. The remainder of the hits consisted of adjacent repurposed characters with functional valence around sugar metabolism.

This survey showed that LMNAST *E. coli lac* operon searches identified numerous ortholog and paralog instances. Relative disparities in gene preservation, gene loss, and structural rearrangements bearing signaling implications were delineated. While there was a significant degree of conformity to the standard genomic arrangement, the amount of diversity indicates that attention paid to related, non-canonical signaling units may be worthwhile.



835

836 ***Figure 2-5. 2D similarity plot of lac operon LMNAST stringent search hits overlaid***
837 ***with attributed annotation.*** Each gray dot represents the homology coordinate of a hit.
838 The size of the dot scales directly with the number of hits at the same coordinate. The
839 dashed line is a 1:1 line along which hits have the same degree of homology by both
840 BLAST and LMNAST measures. Seemingly vertical displacements may imply
841 horizontal gene transfer, while horizontal displacements may imply gene loss or
842 arrangement within the same or proximal species. Ovals indicate a clustering of similarly
843 annotated hits. Dashed ovals denote cases where only the majority of hits therein share

the labeled Genbank annotation. Here, the dominant features are the original structure and the evolved beta-galactosidase system. Very little HGT is apparent while gene loss and rearrangement are ostensibly more common.

2.4.2 *E. coli* K-12 W3110 Lsr System Query

Further testing of LMNAST was conducted with weak, stringent, and expanded window searches of the *E. coli* Lsr system. The query Lsr system consists of a kinase (LsrK: BAA15191), a repressor (LsrR: BAA15192), ABC transporter genes (LsrA: BAA15200, LsrC: BAA15201, LsrD: BAA15202, and LsrB: BAE76456), and AI-2-P processing genes (LsrF: BAE76457, LsrG: BAE76458). Along with AI-2, the Lsr system consists of multiple overlapping positive and negative feedback loops. Multimeric LsrR represses system expression emanating from the intergenic region. AI-2-P, itself catabolized by LsrF and LsrG, allosterically relieves that repression. Thus, both expression troughs and peaks are tightly regulated.[112] For the LMNAST search we used the Lsr genes spanning bp 1600331 to 1609003 of *E. coli* K12 substrain W3110 (Genbank nucleotide record AP009048). The number of hits returned using stringent criteria totaled 419.

Much like the *lac* operon, the Lsr system appeared subject to imperfect conservation. Certainly, many fully sequenced *E. coli* bore exact Lsr homologs (meH > 0.95). Exceptions were the truncated systems found in strains BL21 [113], REL606 [113], and E24377A, and the specific excision of Lsr systems from an otherwise preserved gene order in B2 type *E. coli* (Figures 2-6 and 2-S1) as revealed through expanded window searching.

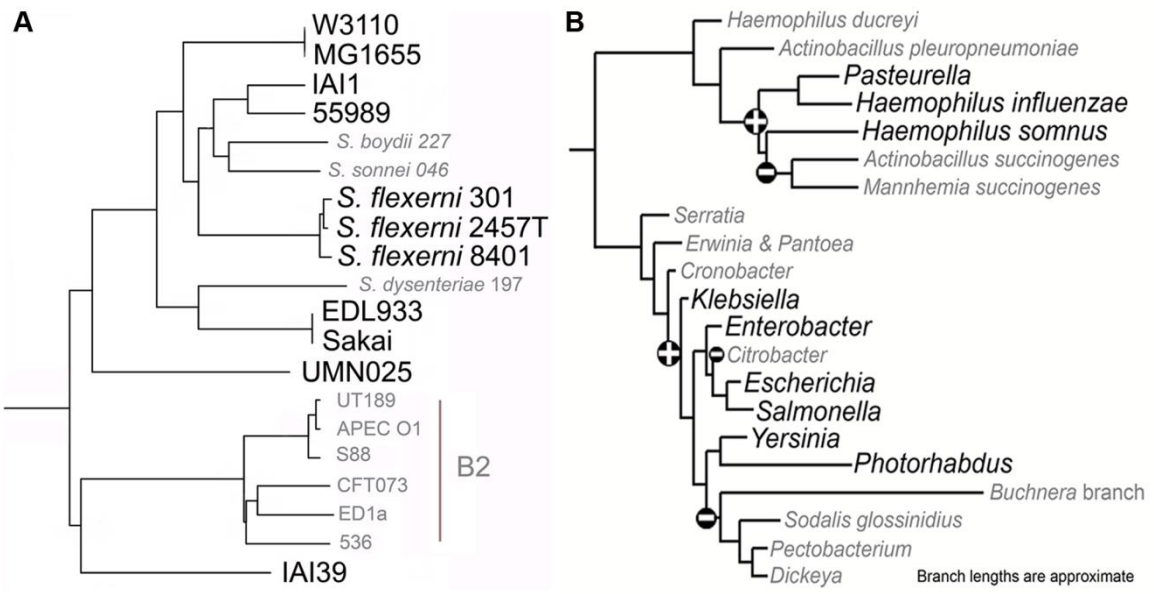


Figure 2-6. Lsr system LMNAST hits overlaid onto phylogenetic distributions of different scopes. A *E. coli* Lsr system LMNAST hits overlaid onto an *E. coli* specific phylogenetic distribution as developed in and adapted from [102]. The larger, bolded leaves contain Lsr system homologs, whereas the grayed italicized leaves do not. Lsr system loss is evident in B2 strains. **B** *E. coli* Lsr system LMNAST hits overlaid onto an Enterobacteriales and Pasteurellaceaes phylogenetic distribution adapted from [108]. The larger, bolded leaves contain Lsr system homologs, whereas the grayed italicized leaves do not. Loss and gain events are denoted by – and + respectively based on parsimony. Compared to the distribution of the *lac* operon, Lsr is more phylogenetically dispersed, but also a bit shallower.

Unlike the *lac* operon, numerous Lsr system homologs had perfect LMNAST homology but markedly reduced meH (Figure 2-8a). This is suggestive of amelioration following recent HGT events (which may itself be a reflection of a carefully tuned signal requiring the full complement and correct arrangement of Lsr elements). Indeed, Lsr system GC content varied in accordance with the background GC content, ranging from 0.35 to 0.71. Finer scale GC analysis revealed a single consistent and curious feature across all hits with meH greater than 0.3: a sharply spiking dip in fractional GC content near the intergenic region (Figure 2-S2). This dip is suggestive of a conserved DNA binding domain essential to the signal transduction process, which would also, however, be a regulatory feature outside the scope of LMNAST searches.

Imperfect LMNAST hits with meH greater than 0.3, deviated from the theoretical distribution according to a bias toward the conservation of *lsrB*, *F*, and *G*, relative to the *lsrA*, *C*, and *D* importer genes (Figure 2-7). This may be attributable to the fact that *lsrB*, *F*, and *G* likely pass cell signaling information downstream [73,74,97] , whereas loss of Lsr importer function might be partially redundant to a low affinity rbs pathway [61], the likely alternate AI-2 import pathway [49,55,112].

In contrast to high meH systems, many systems with low meH (less than 0.3) were involved in the metabolism of 5 carbon sugars, mainly ribitol and xylose, according to Genbank annotation (Figure 2-8b). Since AI-2 itself is mainly comprised of a 5 carbon ring, such homology is simultaneously intriguing and unsurprising. More generally among these low similarity hits, *lsrK* characters were commonly coincident with Lsr

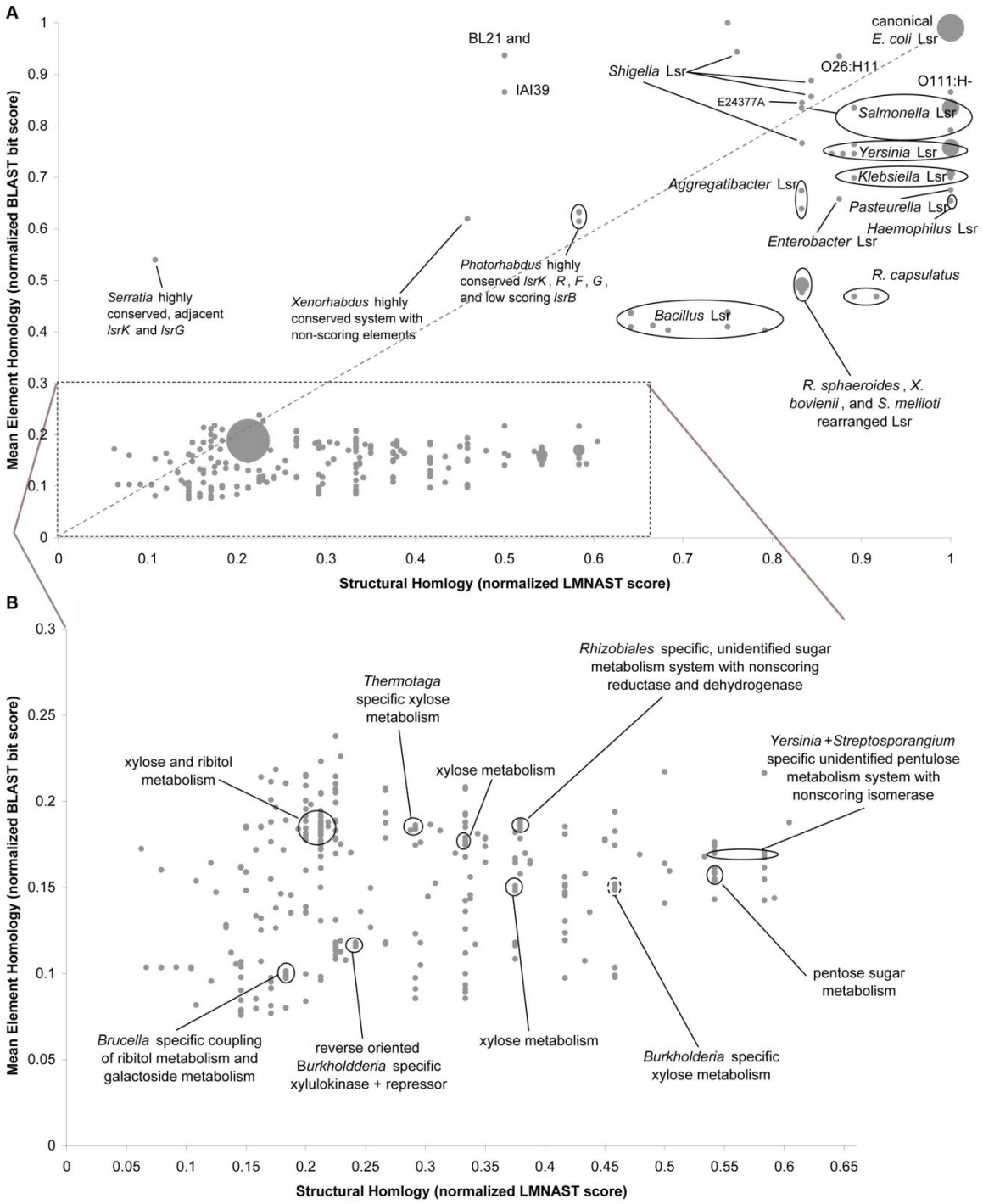


Figure 2-7. Annotated 2D similarity plot for Lsr system LMNAST weak search hits. *A.* HGT of homologous systems is evident among hits with perfect organizational homology but diminished mean element homology. A great number of hits have low similarity

along both axes. As seen in *B.*, these hits are mostly involved in the metabolism of 5 carbon carbohydrates according to their annotation. This is likely reflective of the fact that AI-2 is of similar structure to 5 carbon sugars.

939 importer characters (*lsrA*, *C*, *D*, and *B*), indicative of the functional link between these
940 characters. These various features were laid more strongly in relief when measured
941 against the proximal genetic background in an extended window search.

942 While a representation of hit variability preserving structural information can be
943 had from trackback plots (Figure 2-S1), additional salient results from stringent Lsr
944 extended window searching could be evinced from the more summary coincidence heat
945 maps (Figure 2-9). The matrices indicate that *lsrK* and *lsrA* genes were strongly
946 preserved among extended window hits. Also, if either *lsrF* or *lsrG* were present, the
947 remaining Lsr genes were likely present. The complete system rescission mentioned
948 before was hinted at, especially in rows 3 and 4, corresponding to the toxin/antitoxin
949 *hipAB* system. Intra-species variation of structural homology increased greatly when
950 using stringent rather than weak criteria (data not shown), mainly as a result of gene loss
951 to pseudo gene conversion, mostly among transporter genes—a bias most easily
952 explained as a matter of pure probability since there are more transporter genes than any
953 other type, and a fact whose functional significance is blunted by an alternative AI-2
954 import pathway.

955 These initial *E. coli* searches motivated other orthologous Lsr system queries. Full
956 results for *E. coli*, *S.* Typhimurium, and *B. cereus* searches are available in Table 2-S2.
957 These additional searches identified other possible Lsr system homologs, HGT partners,
958 and non-canonical system-associated gene candidates. In Table 2-S2, we delineate
959 operon directionality and gene homology. It is interesting to note that system variants
960 exist among noted human pathogens: *Yersinia pestis*, *Bacillus anthracis*, and

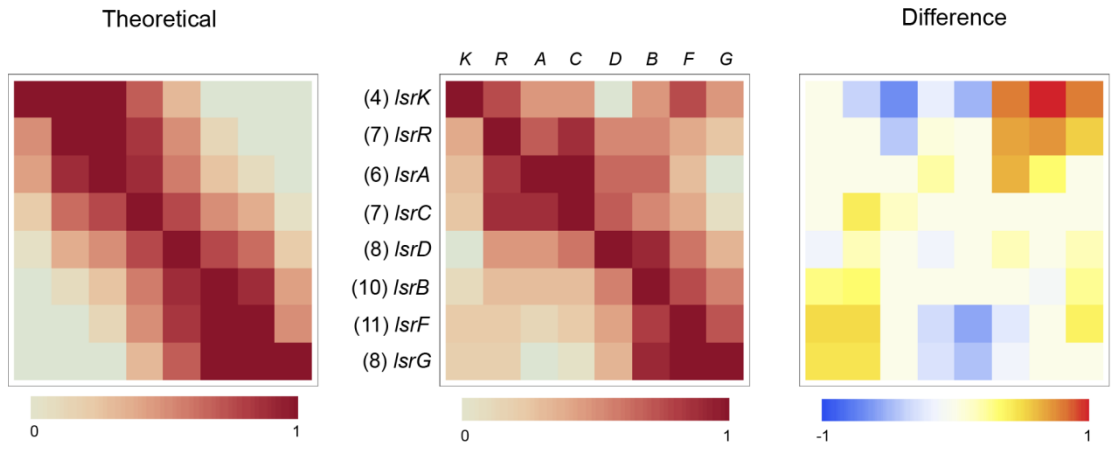


Figure 2-8. Coincidence matrix for E. coli Lsr system LMNAST stringent search hits. This coincidence matrix depicts the subset of hits with a mean element homology > 0.3 and also containing 4-6 gene characters. This subset was chosen for its intermediate degree of homology to the query Lsr system. *LsrF* and *lsrG* characters were found to be overrepresented among these hits coincident to hits also containing *lsrR* and l*srK* characters.

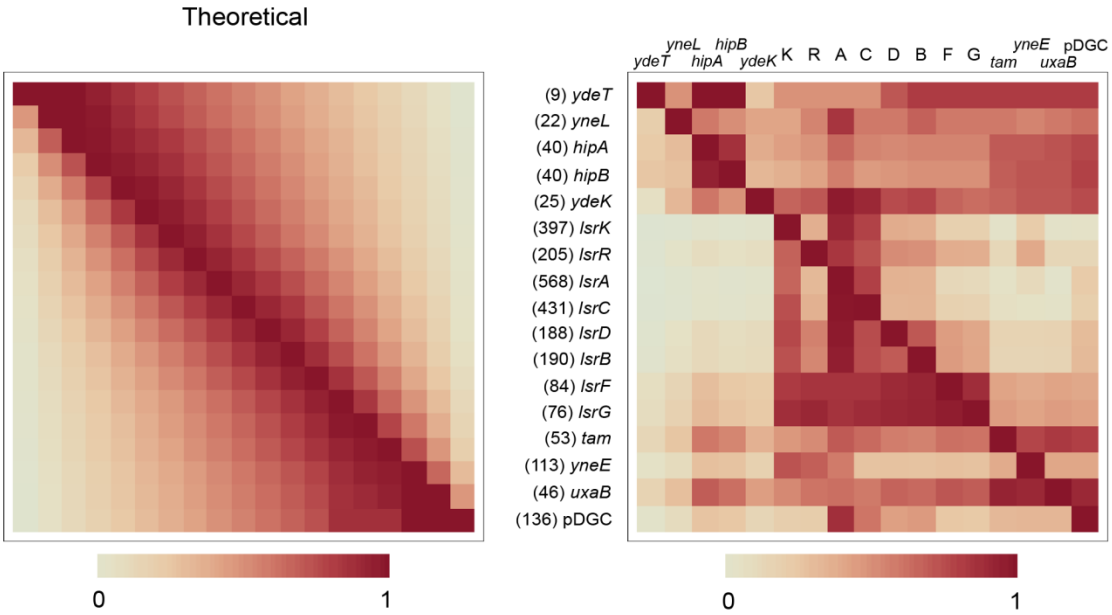


Figure 2-9. Coincidence matrix for E. coli Lsr system LMNAST extended window stringent search hits. Letters represent the respective lsr genes. 1-5 represent the five genes preceding the Lsr system: *ydeT*, *yneL*, *hipA*, *hipB*, and *ydeK-lipoprotein*. 14-17 represent the four genes after the Lsr system: *tam* (transaconitate methyltransferase), *yneE*, *uxaB*, and a predicted diguanylate cyclase. The figure suggests a) strong conservation of the association between Lsr system genes relative to its neighbors, b) hits in which the Lsr system has been excised entirely from its gene neighbors, and c) a weak coincidence between yneH-glutaminase characters and the Lsr system. The matrix also indicates that the overall prevalence of *lsrF* and *lsrG* characters is lower than other canonical Lsr characters, although the presence of *lsrF* and *lsrG* characters is a good predictor of the presence of other Lsr genes.

Haemophilus influenzae. In some instances, *lsrRK* are either absent (e.g. *E. coli* BL21) or are associated with altered intergenic regions suggesting altered regulatory control (e.g. *Yersinia pestis* Antiqua). In other cases transporter genes are distributed with altered bias due to position in the bidirectional operons (e.g. *Yersinia pseudotuberculosis* PB1/+). In some cases there is no *LsrFG* component (e.g. *Shigella flexerni* 2002017). LsrF and LsrG are AI-2-P processing enzymes that lower the cytoplasmic AI-2-P level, thereby contributing to the repression of AI-2 induced genes.

Given even only this modest degree of dispersion, it is nonetheless reasonable to suggest that the Lsr autoinduction system is, in fact, extant among scores of bacterial species and that because the organization of genes within the regulatory architecture is varied, the downstream phenotypic behaviors aligned with the AI-2 regulated QS genes is likewise variable. Thus, our results are in line with a general hypothesis that the AI-2 quorum sensing system is broadly distributed and that the specific needs of the bacteria in a given niche are met by disparate operon arrangements.

The overall phylogenetic distribution of the Lsr system mirrors that as developed by Pereira, et al. in the cluster they denote as Group I.[13] Here, however, details were fleshed out with different emphases. The Lsr LMNAST search captured the diversity of pseudo gene conversion, structural rearrangement, and additional hitchhiking genes associated with the Lsr system that exist in the present nr/nt database. Moreover, inferences regarding regulatory Lsr system signals could be made that might also map to phylogenies or possibly, with much more effort, related ecological niches.

2.4.3 Analysis of Lsr System Search Results

Results from the various LMNAST searches were reconciled by taking the highest scoring hit among overlaps within each nucleotide record. Lsr system homologs clustered mainly in gammaproteobacteria with the greatest density being among *E. coli* strains. Diffusely manifesting in more distantly related bacterial species, the Lsr system appears to have been subject to several HGT events. That is, the Lsr system is absent in numerous Enterobacteriaceae species, while HGT gain events happened at the root of the *Bacillus cereus* group, to *R. sphaeroides* and *R. capsulatus* separately, to *Sinorhizobium meliloti*, and to *Spirochaeta smaragdinae* (Figure 2-10). Curiously, while these bacteria occupy distinct ecological niches, they are all common to soil or water environments.

Multiple extended window searches indicated that *S. enterica* was the most proximal cluster for every Lsr system HGT candidate. The sharing of a novel Lsr system-associated "mannose-6-phosphate isomerase" (NP_460428) between *Bacillus cereus* group members, *S. smaragdine,* and *S. enterica*, further strengthened the suggestion of HGT partnership. The gene annotated as "mannose-6-phosphate isomerase" or "sugar phosphate isomerase," has recently been shown to be part of the LsrR regulon in *Salmonella*.[114] Although not part of the *E. coli* regulon, it was also associated with *S. smaragdinae* and *B. cereus* group orthologs. In keeping with a possible AI-2-P processing role, it was consistently adjacent to *lsrK*.

Among gammaproteobacteria, parsimony suggests that two gain events of the Lsr system occurred: one deeply rooted in enterobacteriales and one in a pasteurellaceaes ancestor. In the enterobacteriales branch, besides *Escherichia*, *Shigella*, and *Salmonella*,

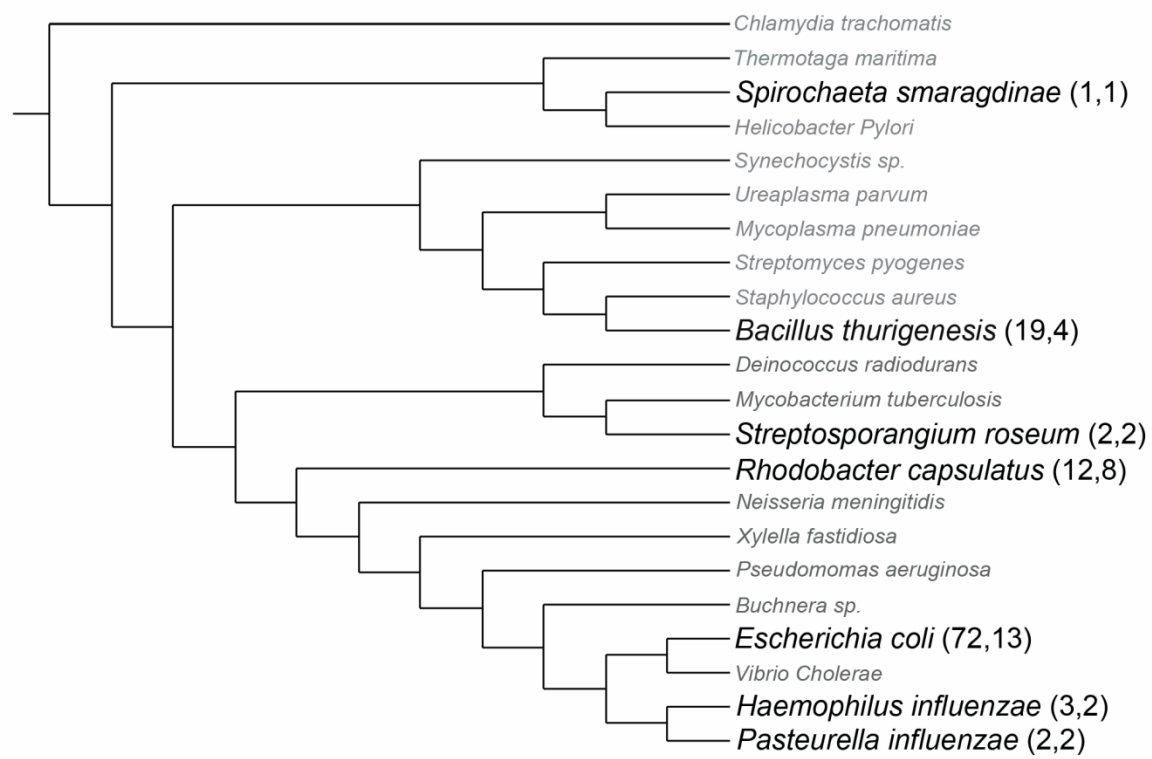


Figure 2-10. Phylogenetic distribution of Lsr at different phylogenetic scales using reconciled LMNAST results. *E. coli* Lsr system LMNAST hits overlaid onto a bacterial phylogenetic distribution as developed in and adapted from [52]. Each leaf bears a representative member from a larger unseen collapsed branch. The larger, bolded leaves contain Lsr system homologs within the collapsed branch, whereas the grayed italicized leaves do not. Parenthesized numbers indicate the number of strains and species with Lsr system LMNAST hits contained within the collapsed branch.

Lsr organizational homologs were found in *Enterobacter*, *Photorhabdus*, and *Xenorhadbus* species, although most of these instances lacked importer genes (*lsrACDB*).

While it is thought that regulatory proteins conserved across such long phylogenetic distances often regulate different targets [115], the regulation of community-related functions by different manifestations of the Lsr system (such as biofilm maturation checkpoints in *E. coli* [60] and possible biofilm dispersion in *B. cereus* [82]) suggests a convergent tendency to leverage a quorum/environment sensing capacity inherent to the Lsr system. Indirect influence over a broader regulon may be abetted by the involvement of AI-2, the Lsr system substrate, in metabolic pathways [51].

2.4.3.1 Putative Lsr system in Rhizobiales

A grouping of Rhizobiales common to plant symbioses had a conserved set of adjacent, low homology elements and nonscoring elements that consisted of unidirectionally expressed genes annotated as: ABC-type transporters (ribose, putative, putative), a DeoR family repressor (LsrR synonym), aldo-keto reductase (non-scoring), glycerol-3-phosphate dehydrogenase (nonscoring LsrF synonym), fructose-biphosphate aldolase (LsrF synonym), and xylulose kinase (LsrK synonym). Somewhat inescapably, synonymy is at least partly a function of precedence resulting in attribution bias. The indeterminacy of signaling similarity between these homologs and better characterized Lsr systems suggests a need for further research.

2.5 Concluding Remarks

LMNAST is a program that evaluates similarity or homology on the level of gene organization, conducting a search for patterns and prevalence constrained by a BLAST E-value filter. Program results overlaid onto phylogenetic data allow visual inspection of phylogenetic density and dispersion. 2D homology plots display system variability among LMNAST orthologs, and when overlaid with genera/species clustering, reveal the degree of system conservation within and across genera/species when organizational homology decreases and element homology is constant. Clustering also enables the identification of conserved system homologs. Organizational information is lost when using coincidence heat charts, but suggestions of the underlying structural variability remain nonetheless. This is particularly true for coincidence representations of extended window searches. For these searches, contextual associations with non-canonical genes also may emerge. Trackback plots illustrate both variety and structural information, albeit in a less dense format. These representations are especially useful in combination. It should be noted that the results are nearly entirely comprised of excerpts from fully sequenced genomes. Results are also biased by BLAST input, as characters with more element homologs (e.g. *lsrA*) appear more frequently in hits.

Generically, LMNAST identified query homologs with a variety of deletions, insertions, misordering, and misdirections. While nearly any source of mutagenesis may result in a frank mutation, affecting a system's organizational homology, homologous recombination, insertion sequences, transposable elements, and combinations thereof are likely to be of particular consequence for LMNAST searches. Deletions may be a result of pseudo gene conversion, of chromosomal rearrangements, or part and parcel of an

insertion event—if the insertion results in a gap sufficiently large as to disconnect hit elements from one another. In the case of such insertions, sufficiently weak criteria may be of use, with the caveat that decreased stringency increases the number of false positives. From a signaling perspective, depending on the impacted elements and the nature of the inserted sequence, gap presence could result in system discoordination; and the longer the gap the more probable and severe the discoordination, most likely to the detriment of system function.

As for the specific test queries examined herein, while the *lac* operon is well characterized in its canonical form, there nonetheless exists a great deal of frank variation from the textbook case. Of particular interest are homologous instances where structural rearrangement could influence self-regulation of component expression. Also of note are its multiple signaling component deletions. Such abbreviated modules are frequently repurposed in a related context. Complete *lac* operons were found among nearly all *E. coli* strains. Incomplete *lac* operons were found to be distributed only among closely related Enterobacteriaceae species comprised almost entirely of *Escherichia*, *Citrobacters*, *Enterobacters*, and *Serratias* as expected based on limited *lac* operon HGT [111]. This difference between the rates of decay for the two homology signals over phylogenetic space may be suggestive of distinct selection pressures guiding the two systems. Also identified through LMNAST were conserved, *E. coli*-specific evolved beta-galactosidase systems [110], demonstrating a capacity to find closely related systems.

On par, Lsr system hit structural similarity was less well correlated with meH than *lac* operon results, a phenomenon presumably associated with apparent Lsr system HGT. The Lsr system was phylogenetically dispersed more widely than the *lac* operon, even

while its distribution remained densest among gammaproteobacteria. Much like the *lac* operon, Lsr system structure was subject to significant variability. *lsrK* and *lsrR* characters were common to many hits. *lsrF* and *lsrG* were the least common; the inclusion of both elements nearly always coincided with the presence of all other Lsr characters as well. Lsr-contextually associated genes and novel putative Lsr systems were also elucidated.

The dispersion of Lsr to bacteria as far afield as the *S. smaragdinae*, the first *Spirochaeta* to be fully sequenced [116], is intriguing. It suggests that while the depth of Lsr dispersion may not be significant, that its exposed breadth will expand incrementally at a rate proportional to microbial genome sequencing. While the direct regulon of such HGT systems is expected to be limited [69,115], the proximity of the substrate to key metabolic pathways may allow the Lsr system to confer contextual phenotypic advantages by impacting downstream pathways with its capacity to recompartmentalize a metabolic intermediate. Moreover, the known regulatory requirements for functional integration of the Lsr system are minimal, consisting entirely of interaction with cAMP-CRP complex, which is deeply rooted in eubacteria. Gene organization differences between dispersed Lsr homologs, may indicate distinct signaling outcomes, in turn suggesting the appropriation of the Lsr system's inherent quorum capacity to drive distinct phenotypes suited to a given bacteria's needs within its particular niche.

Unlike the results for the *lac* operon, Lsr system results returned a large number of other-annotated, low homology systems. This speaks to both the inherent difficulty of extrapolating based on homology and the utility of the additional, complementary homology measure yielded by LMNAST searching. Overall, given the complexity of the

results, numerous aspects may be of interest. Some graphical tools of a complementary nature (e.g. 2D similarity plots and coincidence heat maps) have been used here for distillation and closer inspection. The extant variation of the queried modular systems, as captured by frank changes in gene organization, was revealed.

2.6 Supplemental Materials

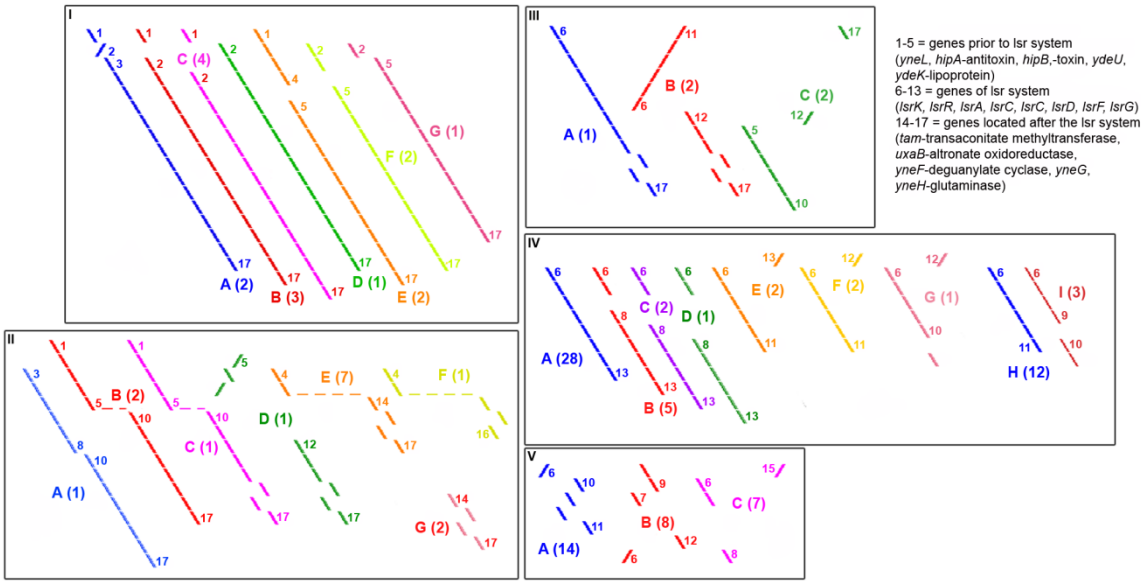


Figure 2-S1. Trackback plots for Lsr system LMNAST extended window stringent search hits. These diagrams describe the variety of LMNAST hits in greater detail. A straight diagonal line indicates complete agreement with the query. Rearrangements are represented by discontinuities. Relative redirection is indicated by a flipping of the diagonal orientation. Deletion is indicated by horizontal dashed gaps. Insertion is indicated by vertical gaps. The legend in the upper right hand corner indicates which numbers correspond to which genes. Trackback plots are organized into categories: A, B, C, D, and E according to the following categories: I. Prototype Lsr systems, II. Modified Lsr systems with pre and post-Lsr adjacent characters, III. Modified Lsr systems with post-Lsr adjacent characters, IV. Modified systems without Lsr adjacent characters, and V. Highly modified Lsr systems. For exact subgroup membership see Table S1.

1169

1170 ***Table 2-S1***

Group	Subgr	Species-Strain
I (pre-Lsr-post)	A	*Escherichia coli* ATCC 8739
		Escherichia coli DH1
	B	*Escherichia coli* O103:H2 str. 12009
		Escherichia coli O157:H7 str. EC4115
		Escherichia coli SE11
	C	*Escherichia coli* O111:H- str. 11128
		Escherichia coli O157:H7 str. TW14359
		Escherichia coli O26:H11 str. 11368
		Escherichia coli 55989 chromosome
	D	*Escherichia coli* IAI1
	E	*Escherichia coli BW2952*
		Escherichia coli str. K12 substr. DH10B
	F	*Escherichia coli* O157:H7 str. Sakai
		Escherichia coli O157:H7 EDL933
	G	*Escherichia coli* SMS-3-5
II (pre-modified Lsr-	A	*Escherichia coli* E24377A
	B	*Escherichia coli* B str. REL606
		Escherichia coli BL21(DE3)
	C	*Escherichia coli* IAI39
	D	*Shigella dysenteriae* Sd197
	E	*Escherichia coli* 536
		Escherichia coli APEC O1
		Escherichia coli CFT073
		Escherichia coli IHE3034
		Escherichia coli S88 chromosome
		Escherichia coli SE15
		Escherichia coli UTI89
	F	*Escherichia coli* ED1a chromosome
	G	*Escherichia coli* 042
		Shigella boydii CDC 3083-94
III (modified Lsr-post)	A	*Escherichia fergusonii* ATCC 35469
	B	*Shigella flexneri* 2a str. 2457T
		Shigella flexneri 2a str. 301
	C	*Sinorhizobium meliloti* 1021 plasmid pSymB
		Rhodobacter sphaeroides KD131 chromosome 2
IV (modified Lsr)	A	*Aggregatibacter aphrophilus* NJ8700, complete genome
		Enterobacter sp. 638
		Haemophilus influenzae PittEE
		Haemophilus somnus 129PT
		Haemophilus somnus 2336

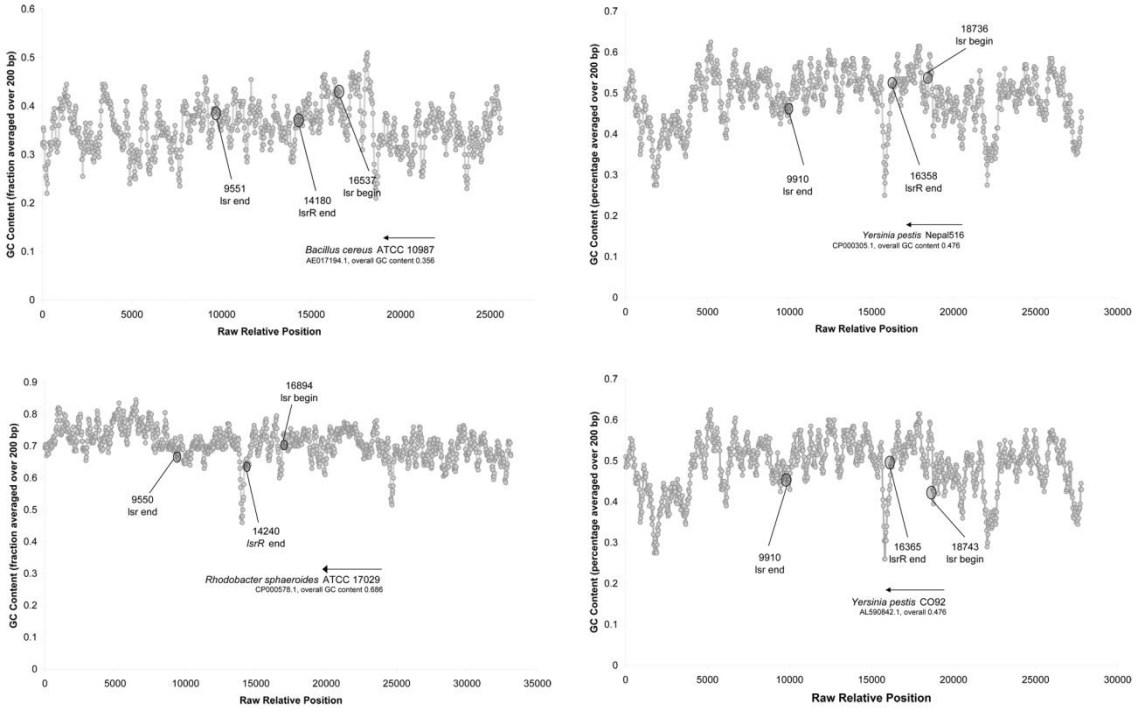
		Klebsiella pneumoniae 342
		Klebsiella pneumoniae subsp. pneumoniae MGH 78578
		Pasteurella multocida subsp. multocida str. Pm70
		Salmonella enterica subsp. enterica serovar Agona str. SL483
		Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67
		Salmonella enterica subsp. enterica serovar Enteritidis str. P125109
		Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91
		Salmonella enterica subsp. enterica serovar Heidelberg str. SL476
		Salmonella enterica subsp. enterica serovar Paratyphi A str. AKU_12601
		Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150
		Salmonella enterica subsp. enterica serovar Paratyphi B str. SPB7
		Salmonella enterica subsp. enterica serovar Paratyphi C str. RKS 4594
		Salmonella enterica subsp. enterica serovar Schwarzengrund str. CVM19633
		Salmonella enterica subsp. enterica serovar Typhi Ty2
		Salmonella enterica subsp. enterica serovar Typhimurium str. D23580
		Salmonella enterica subsp. enterica serovar Typhimurium str. LT2
		Yersinia enterocolitica subsp. enterocolitica 8081
		Yersinia pestis Angola
		Yersinia pestis CO92
		Yersinia pseudotuberculosis IP 31758
		Yersinia pseudotuberculosis IP32953
		Yersinia pseudotuberculosis PB1/+
		Yersinia pseudotuberculosis YPIII
	B	*Klebsiella pneumoniae* NTUH-K2044 DNA
		Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853
		Salmonella enterica subsp. enterica serovar Newport str. SL254
		Salmonella enterica subsp. enterica serovar Typhimurium str. 14082S
		Yersinia pestis biovar Microtus str. 91001
		Yersinia pestis Pestoides F
	C	*Yersinia pestis* Antiqua
		Yersinia pestis Nepal516
	D	*Yersinia pestis* KIM
	E	*Rhodobacter sphaeroides* 2.4.1 chromosome 2
		Rhodobacter sphaeroides ATCC 17029 chromosome 2
	F	*Bacillus anthracis* str. ‘Ames Ancestor’
		Bacillus anthracis str. Ames
	G	*Bacillus anthracis* str. A0248
	H	*Bacillus anthracis* str. CDC 684
		Bacillus anthracis str. Sterne
		Bacillus cereus ATCC 10987
		Bacillus cereus ATCC 14579
		Bacillus cereus B4264
		Bacillus cereus E33L
		Bacillus cereus G9842

		Bacillus cereus Q1
		Bacillus thuringiensis serovar konkukian str. 97-27
		Bacillus thuringiensis str. Al Hakam
		Bacillus weihenstephanensis KBAB4
		Shigella sonnei Ss046
	I	*Bacillus cereus* AH187
		Bacillus cereus AH820
		Bacillus cereus 03BB102
V (Non-continous)	A	*Escherichia coli* ED1a chromosome
		Serratia proteamaculans 568
		Shewanella halifaxensis HAW-EB4
		Yersinia pestis Angola
		Yersinia pestis Antiqua
		Yersinia pestis biovar Microtus str. 91001
		Yersinia pestis CO92
		Yersinia pestis KIM
		Yersinia pestis Nepal516
		Yersinia pestis Pestoides F
		Yersinia pseudotuberculosis IP31758
		Yersinia pseudotuberculosis IP32953
		Yersinia pseudotuberculosis PB1/+
		Yersinia pseudotuberculosis YPIII
	B	*Agrobacterium radiobacter* K84 chromosome 2
		Ochrobactrum anthropi ATCC 49188 chromosome 2
		Rhizobium leguminosarum bv. trifolii WSM1325 plasmid pR132501
		Rhizobium leguminosarum bv. viciae plasmid pRL12
		Rhizobium sp. NGR234 plasmid pNGR234b
		Silicibacter sp. TM1040
		Sinorhizobium medicae WSM419 plasmid pSMED01
		Sinorhizobium meliloti 1021 plasmid pSymB
	C	*Burkholderia cenocepacia* AU 1054 chromosome 1
		Burkholderia cenocepacia HI2424 chromosome 1
		Burkholderia cenocepacia J2315 chromosome 1
		Burkholderia cenocepacia MC0-3 chromosome 1
		Burkholderia multivorans ATCC 17616 DNA
		Burkholderia sp. 383 chromosome 1
		Burkholderia xenovorans LB400 chromosome 1

1171

1172

1173



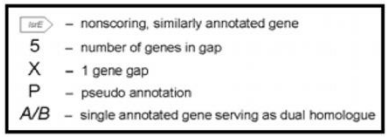
1174

1175 ***Figure 2-S2. GC content demonstrates consistent spiking dip at intergenic region.*** GC
1176 content graphs for *Bacillus cereus* ATCC10967, *Yersinia Pestis* Nepal516, *Rhodobacter*
1177 *sphaeroides* ATCC 17029, and *Yersinia pestis* CO92. Graphs are labeled with lsr system
1178 beginning (end of *lsrK*), *lsrR* gene intersection with the intergenic region (labeled lsrR
1179 end), and lsr system ending (end of *lsrG*). Arrow direction indicates the direction of
1180 *lsrACDBFG* expression. Proximal to the intergenic region is a conserved dip in GC
1181 content.

1182

1183

1184

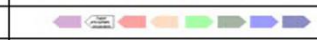
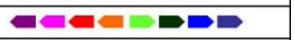
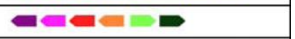


Search:	*E. coli K-12 W3110 (Weak)*	*S. Enterica Serovar Typhimurium LT2 (Strict)*	*B. Cereus ATCC 10987 (Strict)*
	IsrK IsrR IsrA IsrC IsrD IsrB IsrF IsrG	IsrK IsrR IsrA IsrC IsrD IsrB IsrF IsrG IsrE	IsrF IsrK IsrR IsrA IsrC IsrD IsrB
100%			
50%			
20%			
Hits:			
Aggregatibacter actinomycetemcomitans D11S-1		join	join
Aggregatibacter aphrophilus NJ8700		join	join
Enterobacter cloacae subsp. cloacae ATCC 13047			
Enterobacter sp. 638			
Escherichia coli 042			
Escherichia coli 55989			
Escherichia coli ATCC 8739			
Escherichia coli BW2952			
Escherichia coli DH1			
Escherichia coli HS			
Escherichia coli IAI1			
Escherichia coli O103:H2 str. 12009			
Escherichia coli O111:H- str. 11128			
Escherichia coli O157:H7 EDL933			
Escherichia coli O157:H7 str. EC4115			
Escherichia coli O157:H7 str. Sakai			
Escherichia coli O157:H7 str. TW14359			
Escherichia coli SE11			
Escherichia coli SMS-3-5			
Escherichia coli str. K12 substr. DH10B			
Escherichia coli str. K-12 substr. MG1655			
Escherichia coli str. K12 substr. W3110			
Escherichia coli UMN026			
Escherichia fergusonii ATCC 35469			
Haemophilus influenzae PittEE			
Haemophilus somnus 129PT		join	
Haemophilus somnus 2336			
Klebsiella pneumoniae 342			

Klebsiella pneumoniae NTUH-K2044
Klebsiella pneumoniae subsp. pneumoniae MGH 78578
Klebsiella variicola At-22
Pasteurella multocida subsp. multocida str. Pm70
Photorhabdus luminescens subsp. laumondii TTO1
Rhodobacter capsulatus SB1003
S. enterica serovar Typhi (Salmonella typhi) strain CT18
S. enterica subsp. enterica serovar Agona str. SL483
S. enterica subsp. enterica serovar Choleraesuis str. SC-B67
S. enterica subsp. enterica serovar Dublin str. CT_02021853
S. enterica subsp. enterica serovar Enteritidis str. P125109
S. enterica subsp. enterica serovar Gallinarum str. 287/91
S. enterica subsp. enterica serovar Heidelberg str. SL476
S. enterica subsp. enterica serovar Newport str. SL254
S. enterica subsp. enterica serovar Paratyphi A str. AKU_12601
S. enterica subsp. enterica serovar Paratyphi A str. ATCC 9150
S. enterica subsp. enterica serovar Paratyphi B str. SPB7
S. enterica subsp. enterica serovar Paratyphi C strain
S. enterica subsp. enterica serovar Schwarzengrund str. CVM19633
S. enterica subsp. enterica serovar Typhi Ty2
S. enterica subsp. enterica serovar Typhimurium str. 14028S
S. enterica subsp. enterica serovar Typhimurium str. D23580
S. enterica subsp. enterica serovar Typhimurium str. LT2
Shigella flexneri 2a str. 2457T
Shigella flexneri 2a str. 301
Yersinia enterocolitica subsp. enterocolitica 8081
Yersinia pestis Angola
Yersinia pestis Antiqua
Yersinia pestis biovar Microtus str. 91001
Yersinia pestis CO92
Yersinia pestis D106004
Yersinia pestis D182038
Yersinia pestis KIM
Yersinia pestis Nepal516
join
join
join

1186

Yersinia pestis Z176003
Yersinia pseudotuberculosis IP 31758
Yersinia pseudotuberculosis IP32953
Yersinia pseudotuberculosis PB1/+
Yersinia pseudotuberculosis YPIII
Agrobacterium radiobacter K84
Bacillus anthracis str. A0248
Bacillus anthracis str. 'Ames Ancestor'
Bacillus anthracis str. Ames
Enterobacter cloacae subsp. cloacae NCTC 9394
Escherichia coli E24377A
Escherichia coli O26:H11 str. 11368
Rhodobacter sphaeroides 2.4.1 chromosome 2
Rhodobacter sphaeroides ATCC 17029 chromosome 2
Rhodobacter sphaeroides KD131 chromosome 2
Rhodospirillum rubrum ATCC 11170
Shigella flexneri 2002017
Shigella flexneri 5 str. 8401
Sinorhizobium medicae WSM419 plasmid pSMED01
Sinorhizobium meliloti 1021 plasmid pSymB
Xenorhabdus bovienii SS-2004
Agrobacterium tumefaciens str. C58
Bacillus anthracis str. CDC 684
Bacillus anthracis str. Sterne
Bacillus cereus 03BB102
Bacillus cereus AH187
Bacillus cereus AH820
Bacillus cereus ATCC 10987
Bacillus cereus ATCC 14579
Bacillus cereus B4264
Bacillus cereus E33L
Bacillus cereus G9842
Bacillus cereus Q1

Bacillus thuringiensis BMB171			
Bacillus thuringiensis serovar konkukian str. 97-27			
Bacillus thuringiensis str. Al Hakam			
Bacillus weihenstephanensis KBAB4			
Burkholderia phymatum STM815			
Meiothermus ruber DSM 1279			
Nakamurella multipartita DSM 44233			

Agrobacterium radiobacter K84			
Escherichia coli B str. REL606			
Escherichia coli BL21(DE3)*			
Escherichia coli IAI39			
Photorhabdus asymbiotica ATCC 43949			
Rhizobium etli CFN 42 plasmid p42b			
Rhizobium leguminosarum bv. trifolii WSM1325 plasmid pR132501			
Rhizobium leguminosarum bv. viciae plasmid pRL12			
Rhizobium sp. NGR234 plasmid pNGR234b			
Silicibacter sp. TM1040			
Sinorhizobium meliloti 1021 plasmid pSymB			
Streptosporangium roseum DSM 43021			
Thermobispora bispora DSM 43833			
Verminephrobacter eiseniae EF01-2			
Xenorhabdus nematophila ATCC 19061			
Yersinia pseudotuberculosis PB1/+			

Salmonella enterica subsp. enterica serovar Typhimurium SL1344			
Bacillus cereus biovar anthracis str. CI			
Spirochaeta smaragdinae DSM 11293			
Yersinia pestis Pestoides F			

Table 2-S2. Results from three separate LMNAST searches for Lsr system homologs. The *E. coli* K-12 W3110 search shown was completed using weak criteria (a lower gap extension penalty and less rigid adherence to annotation), whereas the *B. Cereus* ATCC 10987 and *S. enterica* Typhimurium LT2 searches used stringent criteria (higher penalties for deviation from the query pattern and adherence to supplied annotation). "Join" indicates a difficulty in handling irregular annotation where the first gene annotated in a record spans the end and the beginning of the record. Arrow direction indicates the direction of transcription along the genome. Color is a stand-in for character type, and the degree of shading indicates degree of element homology, with the darkest shade representing 100% element homology.

Chapter 3: Quorum desynchronization leads to bimodality and patterned behaviors in microbial consortia

3.1 Abstract

Quorum sensing (QS) is a type of signaling used to coordinate behavior in bacterial consortia via secreted autoinducers. Here we focus on Lsr (LuxS regulated) based QS signaling that activates only a fraction of a cell population through the 'universal' QS signal autoinducer-2 (AI-2). Our modeling indicates that bimodality arises from desynchronized, QS-activated importation of AI-2, and is sensitive to cell-cell distance and cell density. According to our agent-based models, through this mechanism, we found Lsr QS drives spatial organization of cell signaling. That is, Lsr induced AI-2 internalization results in emergent "cluster-then-disperse" behavior when acting in concert with AI-2 chemoattraction, and speckled activation in surface attached bacterial colonies. This is contrasted against rapid and population-wide expansion of QS activity emerging from LuxIR based QS in a growing bacterial colony. Further, Lsr signaling also results in differential QS activation between Lsr/LuxS types in *in silico* mixed cultures. One particular finding was that signal-negative cells were also found to be effectively signal-blind in Lsr QS. More broadly, interstrain communication modulated QS activation patterns were elucidated, helping to frame the complex Lsr QS dynamics arising in sociomicrobiological settings.

3.2 Introduction

Quorum sensing (QS) is a bacterial response to self-secreted signaling molecules generically known as "autoinducers". While QS has been observed among individual bacteria in experimentally manipulated settings [4–6], QS is often described as informing the coordination of processes (such as virulence factor production and biofilm formation) that are metabolically burdensome and ineffectual for individual cells, yet beneficial at multicellular or population scales [80]. Population scale coordination typically arises from the decompartmentalization of the autoinducer that acts as a shared pool of extracellular signal available to coordinate individual cells that are sufficiently proximal. This regulatory strategy can reduce "noisy" inputs and other heterogeneity by focusing phenotypic outcomes and organizing population activity [117,118]. Coordination varies in degree and fashion, and depends upon the mechanistic underpinnings of the QS system involved [94,119,120]. Importantly, a given "quorum's" collective response to heterogeneity is likely to reflect the topology of the cognate QS signaling "module" [121,122]. Among the better studied QS systems are LuxIR (background **Figure 3-1**), which is widely utilized in synthetic biology because of its genetic simplicity, and Lsr (foreground **Figure 3-1**), which is widely distributed among gammproteobacteria [14]. LuxIR is induced by autoinducer-1 (AI-1) (an umbrella term encompassing a variety of species specific acyl-homoserine lactones (AHLs)) whereas Lsr is induced by autoinducer-2 (AI-2) (which is produced by *luxS* in nearly half of eubacteria [46]). LuxIR signaling is comprised of an intracellular positive feedback loop [21] that is tied to positive intercellular feedback at the population scale. This is contrasted against Lsr signaling the induction of which features

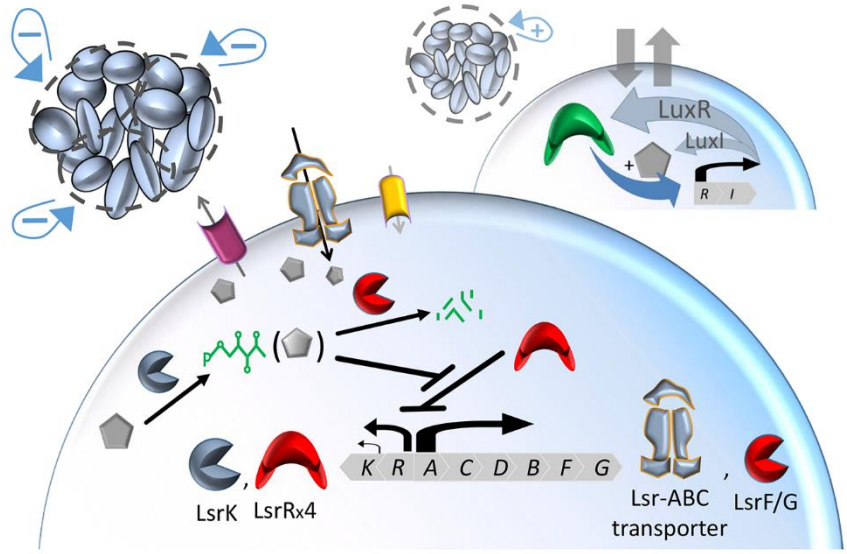


Figure 3-1. LuxIR and Lsr QS activity intracellularly and intercellularly. The LuxIR system (background) involves a simple intercellular feedforward loop that entangles AHL synthase production with LuxR regulatory production, which itself promotes the expression of both LuxI and LuxR when AHL concentration is sufficiently high. This loop is connected to intercellular feedforward activity due to passive diffusion of AHL across the cell membrane. This is in contrast to the Lsr system (foreground), where the feedforward loop of LsrACDB increasing AI-2 intracellularly, LsrK phosphorylating AI-2, and AI2-P derepressing LsrR which in turn stimulates LsrACDB and LsrK expression does not drive extracellular AI-2 concentrations higher, but instead depletes them, creating negative intercellular feedback.

the rapid depletion of extracellular AI-2 [55], intertwining intracellular positive feedback with extracellular negative feedback. Lsr activity is subject to further negative feedback by the enzymatic processing of phosphorylated AI-2 through LsrF and LsrG [73–75]. Among the small molecules produced by LsrF and LsrG under aerobic conditions are glycerol-3-phosphate [123] and phosphoglycolic acid [75], respectively, which serve as metabolic intermediates for AI-2 reassimilation into primary metabolism.

As described herein, we explored and contrasted QS operation in a variety of contexts through the use of finite element-agent based models. In part, our modeling efforts were motivated by indications that bimodal Lsr expression may arise in pure cultures, where cell populations developed into QS-activated and non-QS activated fractions [94]. While it has been suggested that a transient bimodality may arise as a consequence of intracellular signaling topology [95], we hypothesized that a more permanent population bifurcation might develop from hyperlocal competition for AI-2 between cells. This possibility was probed here by employing a population of ODEs wherein nongenetic heterogeneity was explicitly incorporated. The ramifications of such a hyperlocal competition were then considered in the contexts of cell motility and mixed consortia.

LuxIR and Lsr QS were contrasted and found to generate different spatial patterns of cell signaling *in silico*. We recapitulated an earlier experimental study wherein the sudden expansion of LuxIR QS activation from a cell colony center (or "Supernova") was observed [33]. This stood in clear contrast to speckled QS activation found in Lsr-mediated cell colonies. Further, the confluence of Lsr QS and AI-2 chemoattraction in *E. coli* [97] was examined, the combination of which produced a "cluster-then-disperse"

behavior that is consistent with concepts of sociomicrobiology [124] wherein "travel" to better locales may be one of the currencies of public goods [125].

Additionally, the consequences of Lsr QS circuitry were examined *in silico* in mixed cultures of different Lsr/LuxS cell types. Δ*luxS* and Δ*lsrFG* mutants were chosen for their prevalence and because they represent slower and faster QS activating populations, respectively. Coupling these mutants with their wildtype counterparts illustrated how Lsr dynamics manifest in the preferential activation of one Lsr/LuxS cell population over another. Just as activation patterns arising from Lsr signaling were distinct from those associated with LuxIR activity, interpopulation transactions between Lsr/LuxS subgroups revealed different incentives for social cheating than have previously been associated with generic QS signaling.

Methods 3.3

3.3.1 Modeled Cell Behaviors

During a given time step, based on cell growth characteristics and local concentration of substrate, cells divided, moved, and responded to autoinducer concentrations in their direct extracellular environment through an inflection of their QS dynamics as modeled by ODE trajectories. AI-2 was also exchanged between grids in an approximation of diffusion. Full model details and in particular those common to all simulations are provided in **SI Methods**.

3.3.1.1 *Chemotactic Swimming.* To interrogate QS activation patterning under different regimes of motion, the mode of cell motility was varied. To begin either swimming or colony growth simulations, cells or colony centers were placed randomly in the simulated environment. In simulations of chemotactic behavior, cells adjusted direction every 0.1 seconds. If the average concentration of AI-2 experienced by the cell over the previous 0.1 seconds was less than a minimum sensitivity or less than the average concentration of AI-2 from the previous 15 second period (reflecting the time required for accommodation [126]), the cell moved in a random direction and at a speed produced from a distribution with an average of 20 μm/sec, in accordance with results from cell tracking studies [unpublished]. If the short term average concentration was higher than the long term average, the cell continued moving in the same direction as it had the previous time step, with slight error.

3.3.1.2 *Colony Growth*. Expanding colony growth was loosely mimicked by cells moving according to space filling considerations. In the interior of the colony, daughter

cells pushed neighbors outward based on the shortest distance to the colony boundary available. At colony boundaries, cells moved into randomly chosen free space. All motions were constrained by simulation boundaries.

3.3.1.3 *LuxIR QS*. To simulate LuxIR behavior, cells produced AHL at high, baseline, or intermediate rates (10 μM/min to 1 μM/min) corresponding to active, inactive, or linearly activating QS, respectively. The rate of AHL production in the QS inactive state (baseline AHL production) was varied across the population with the same coefficient of variation that was applied to the parameter *basal* in Lsr simulations ($\sigma = 0.0225$, log-normal distributions). Diffusion across the membrane was modeled at a conductivity of 0.6 between the intracellular and extracellular spaces driven by concentration difference [127]. LuxIR expression was activated over a thirty minute period in a linear progression once an intracellular AHL threshold of 2.9 μM was exceeded, coarsely capturing inherent system cooperativity and the time lag associated with transcription and translation.

3.3.1.4 *Lsr QS*. The ODEs governing Lsr activity, associated parameters, and the rationale behind the design choices are described in **SI**. Lsr related ODEs relied heavily upon Michaelis-Menten kinetics and cooperative Hill terms. Parameters were chosen to be biologically realistic and to conform to the behavior of the system as gleaned from previously conducted experiments (see **SI**) [55,56,94,112].

3.3.2 Simulation Variants: Gene Deletions and Mixed Populations

As metabolic burdens [128] were not considered, gene deletions were modeled by setting the appropriate parameter to zero, reflecting ablated activity. The gene deletions examined consisted of *ΔluxS* and *ΔlsrFG*. These were modeled by setting parameter

values to zero such that $K_{synth} = 0$ and $kCat = 0$, respectively. These variants differed in their sensitivity to external AI-2. Simulations placing variant populations and wildtype cells in the same environment were run. The two populations were equivalently populated to initiate the simulation and divided according to the same heuristics. Chemotaxing motility was stripped from these simulations in order to focus on the effect from deletion of Lsr and auxiliary proteins.

In each of these cases, regardless of the cell type simulated, all bacteria and growing colonies were accounted for at every time step and in every grid point. Owing to the complex nature of the system, a finite difference method was used. Care was taken to minimize error propagation (studies of grid size, time step, growth parameters, etc.).

3.4 Results

3.4.1 ***Lsr Autoinduction in Pure Cultures***. We first modeled desynchronized Lsr QS activation. Specifically, the value of the model parameter *basal* (the rate of AI-2 transport through the low-flux AI-2 importer), was distributed among modeled cells in a log normal fashion. All other parameters were held constant so that the relative background AI-2 uptake rate and corresponding speed to Lsr QS activation were the only distinguishing features among the entire cell population. We observed that only a fraction of the total cell population became QS activated, generating a bimodal pattern. The exact on/off balance was influenced by the degree of variance in the *basal* distribution (**Figure 3-2A**). As discussed extensively in the **SI**, this bimodality ostensibly resulted from QS active cells depriving inactive cells of AI-2 with which to induce. Consider an *i*th cell with $basal_i << basal_{Mean}$ that would undergo QS activation if $basal_i = basal_j$ for all *i*,*j*, instead remaining inactive. When the coefficient of variance was zero, the entire population became activated almost simultaneously. Increasing the variance of the value for *basal* increased the population fraction remaining inactive, attributable to cells with higher values of *basal* from the leading edge of the distribution activating earlier and earlier as the variance of the distribution increased. The earlier the activation of these cells relative to the mean, the faster extracellular AI-2 was drawn down, thus preventing the activation of cells with lower *basal* values, which themselves required more and more time to accumulate sufficient AI-2P. This conceptual model serves as a basis to explain previously reported flow cytometry data, where only about half of the total population took on the QS-activated phenotype [94].

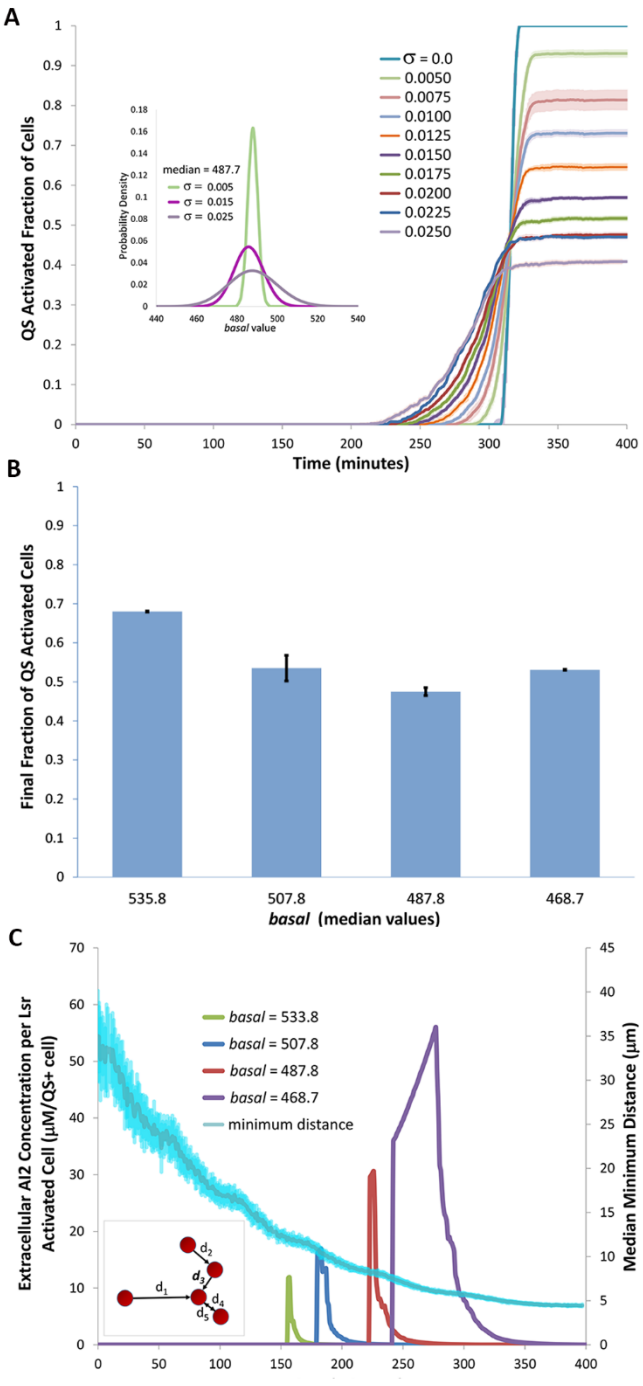


Figure 3-2. Lsr autoinduction of pure cultures leads to bimodal phenotype. **A** The fraction of the population QS activated over time was influenced by the standard deviation of the natural logarithm, σ, of the log-normal distribution for the parameter *basal*, here for σ values ranging from 0 to 0.025, run in triplicate. Dark lines represent the average values, whereas lighter surrounding lines represent the standard deviation. Activation was the slowest but most abrupt for the cell population with a variance of zero, quickly achieving complete activation. Inset is the distribution of the value of *basal* for three different values of σ. **B** Inset is the plateaued value of the population fraction that was Lsr active for populations of different median *basal* values. In the main figure, the overall extracellular concentration of AI-2 per Lsr induced cell is represented on the primary ordinate axis. Each curve represents an average from three different simulations of non-taxis swimming cells. The median minimal distance between cells averaged for 20 simulations of non-taxis swimming cells is represented on the secondary ordinate axis. At the right top is a pictorial example of the minimum distance between cells (where d_{min} = min (d_1, d_2, d_3,... d_n, n=6).

1436

1437

1438

We also note that the distribution of the bifurcation was possibly influenced by the related but competing factors of intercellular distance and cellular concentration. This was suggested by a general correlation between decreasing *basal* and decreasing QS activation as seen in the inset of **Figure 3-2B**. As indicated by the timing of spikes in **Figure 3-2B**, reducing *basal* resulted in delayed activation. This delayed activation corresponded to diminished intercellular distance due to growth (light blue line) calculated as the minimum distance between neighboring cells (a cartoon of which is depicted on the top right corner of **Figure 3-2B** for an *i*th cell). Increased cell proximity may have amplified the negative feedback associated with QS activation, as the local depletion of AI-2 by QS active cells was able to delay or prevent the activation of more cells. Further evidence of this cell-cell distance effect is provided in the **SI**. The putative influence of intercellular distance appeared to be constrained, however. That is, when cell concentration was too high, more QS activated cells were required to draw down local AI-2 concentrations as indicated by the relatively broad peak of the last curve depicting AI2 concentration divided by the number of QS activated cells. Subsequently, depletion of AI-2 from the environment was delayed, allowing more cells to become QS activated than would have otherwise. This competing effect was ostensibly dominant for populations with the lowest median value of *basal* in **Figure 3-2B**, as the QS activated fraction increased instead of continuing to decrease.

3.4.2 ***Pattern formation in QS systems: LuxIR vs Lsr***. In **Figure 3-3A**, we show that a "supernova-like" LuxIR QS activation pattern emerged from near the colony center, loosely recapitulating experimental results seen with LuxIR engineered cells [33,129]. For this simulation, cell motility was limited to gliding along the surface in a space filling

capacity purely as a function of cell outgrowth (cell division resulting in occupation of otherwise empty neighboring site by daughter cell) without directional bias. Once initiated, LuxIR activation (red color) spread quickly outward from its point of origin often near the center of colony growth (due to positive feedback processes of cell growth and autoinduced AHL production), engulfing inactive neighbors in a building wave of high AHL concentration, coordinating population expression in what appears as a traveling wave front of activated cells.

In **Figure 3-3B**, we depict simulated behavior for the Lsr system with identical gliding motility rules as for the LuxIR simulations. Instead of a blossoming wave of QS activation (**Figure 3-3A**), scattered patches of activation appeared, presumably influenced by the same Lsr driven AI-2 recompartmentalization dynamics associated with bimodal activation. Undergirding the distinction between these activation patterns were differences in the interplay between non-genetic heterogeneity and intercellular QS feedback, as discussed in **SI**, where activation heterogeneity was also quantified. Thus, in the case of Lsr, QS activation was not apparently localized, nor was it uniform. Instead, fractional QS activation was observed among the population whole.

3.3.3 ***Cell Motility – Lsr QS based pattern emergence***. Swimming was modeled either as an unbiased process or was governed by heuristics approximating AI-2 chemoattraction, described in the methods. Early in the simulations, prior to QS activation, chemotaxing cells assembled into clusters [130], attracted to each other by AI-2 molecules. Largely, cell populations gradually coalesced along simulation boundaries as our boundary conditions preclude diffusion through these boundaries. Since clusters

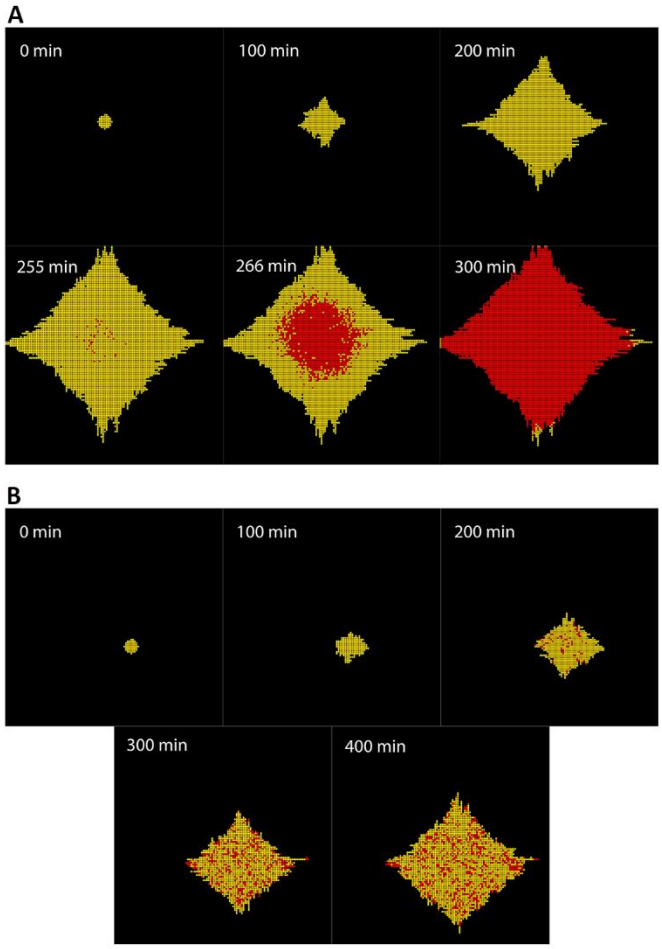


Figure 3-3. QS dynamics coupled with gliding during colony growth. **A** Images from a representative simulation for LuxIR/AHL dynamics coupled with colony growth. QS active cells are in red, whereas inactive cells are in yellow. **B** Images from a representative simulation of Lsr/AI-2 dynamics coupled with colony growth. QS active cells are in red, whereas uninduced cells are yellow.

were formations of relatively high local cell density and correspondingly high AI-2 concentration, cells within clusters activated more quickly than unclustered cells. Once Lsr activity was sufficiently pervasive, clusters were transformed into sites of the most rapid AI-2 depletion, dissipating the AI-2 gradient and the impetus for clustering, the ultimate result being cluster dispersal. This emergent behavior, depicted here for the first time in a model, generates an interesting overall "collect-disperse" pattern. This pattern is exemplified in **Figure 3-4**. At zero minutes a randomized distribution of an initial cell population was observed. At 100 minutes, cell clusters began to coalesce, as seen at the bottom and left edge and near the top left corner of the environment. At 200 minutes, clusters were more pronounced, and at 244 minutes, QS activated cells began to emerge (red). At 266 minutes more cells were QS active and the clusters became more dispersed. By 300 minutes clusters had become entirely dispersed. Through many repeated simulations, the exact placement and number of clusters were found to be inconsistent. Quantification of this clustering (by cell-cell distance) is available in **SI**. Interestingly, as different motility modes implied different cell-cell proximity, motility appeared to feedback onto fractional activation, which is also discussed in **SI**.

3.4.4 ***Mixed Population Simulations***. Additional consequences of desynchronized Lsr based AI-2 recompartmentalization were identified through the study of competition for AI-2 among *in silico* mixed cultures of a wildtype population with varied derivative mutant populations. Here, initial conditions were the same as in previous simulations, except that half the modeled cell population had modulated Lsr/LuxS activity, reflecting different genotypes. **Figure 3-5A** represents the results from a wildtype pure culture as a

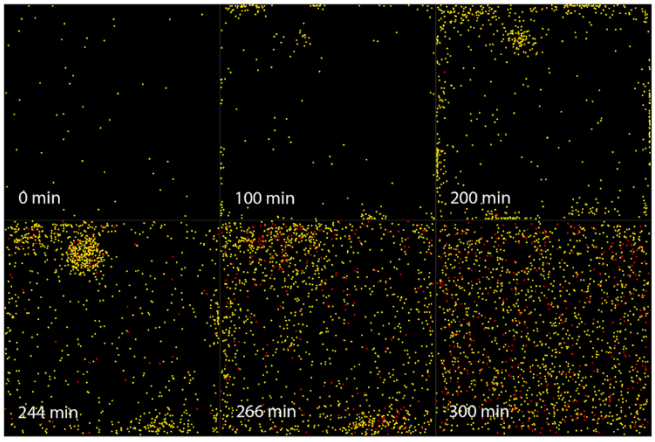


Figure 3-4. Cluster-disperse pattern from combination of Lsr and chemotaxis. Images from a representative simulation of Lsr/AI-2 dynamics coupled with chemoattraction to AI-2. Autoinduced cells appear in red, whereas uninduced cells appear in yellow.

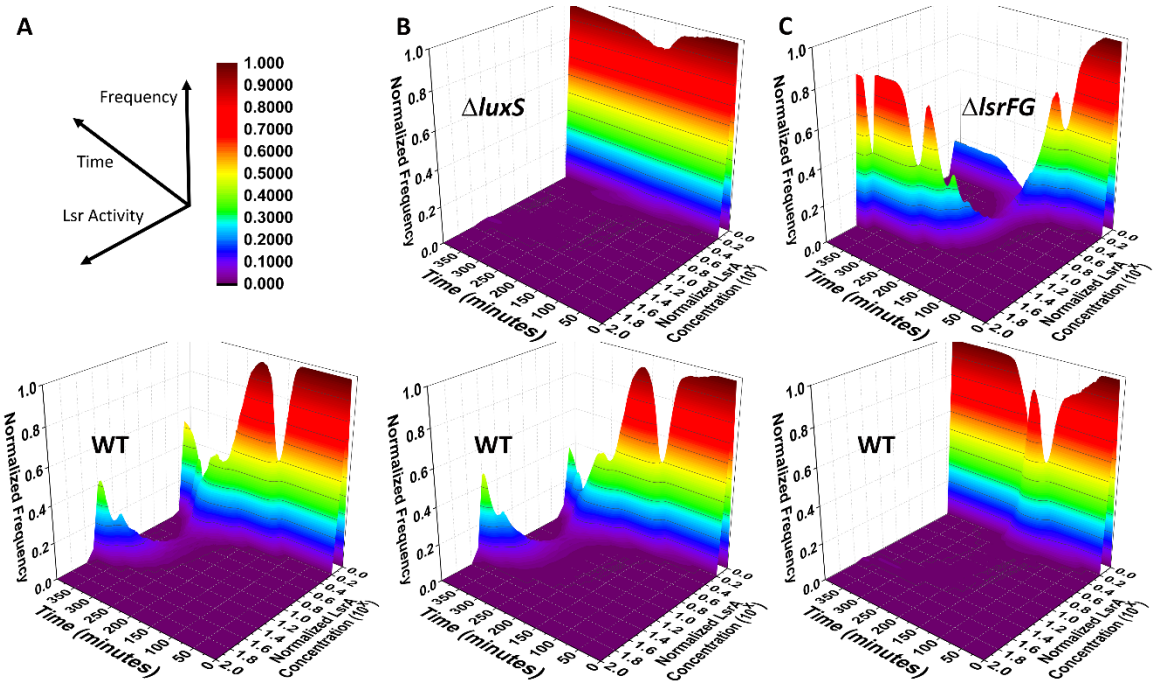


Figure 3-5. Mixed culture simulations. Each graph may be considered as being composed of a series of histograms over time. The vertical axis represents frequency, the right axis (towards the reader) represents increasing log transformed concentration of transporter protein, and the left hand axis (away from the reader) represents increasing time in minutes. **A** wildtype cells alone. **B** Δ*luxS* populations mixed with wildtype cells. **C** Δ*lsrFG* cells mixed with wildtype cells.

reference (lower panel). Initially, all cells were QS inactive (represented here as normalized to 0.0). Around 200 minutes, subpopulations began to branch off into high expressers, representing QS activation in these cells. In **Figure 3-5B** where Δ*luxS* (above) and wildtype (below) cells were cultured together *in silico*, wildtype cells became dominantly activated over the course of the simulation, with Δ*luxS* cells remaining largely Lsr inactive. In fact, even by the end of the simulation, Δ*luxS* population Lsr expression levels were only nominally above their baseline.

As in **Figure 3-5C**, mixing of wildtype cells (below) with Δ*lsrFG* cells (above) also resulted in uneven activation between populations, but for ostensibly different reasons. Here, Δ*lsrFG* cells were dominantly activated and became so early on, due to the early activation of Δ*lsrFG* cells, which accumulate AI2-P more readily than wildtype cells [131]. This then deprived wildtype cells of AI-2, resulting in a smaller activated fraction than when wildtype cells were cultured by themselves.

3.5 Discussion

Bimodal expression arising from clonal origins is a common phenomenon, frequently associated with pattern formation and differentiation in multicellular organisms [132,133]. As is commonly the case and here also, population bifurcation is expected to arise as a consequence of nonlinear responses to nongenetic heterogeneity manifested in bimodal protein expression [134,135]. Specifically, our results suggest that bimodal expression arises when the intercellular negative feedback associated with Lsr activated AI-2 recompartmentalization is desynchronized across a population (here represented by varied rates of AI-2 uptake and subsequent Lsr mediated gene expression). This bimodal expression may represent role diversification. Bacterial diversification is frequently framed within the context of bet hedging [136] or as a graded response to environmental conditions, an example being the different conditions that at the margins of biofilms compared to those within the bulk [137]. Although Lsr signaling does influence biofilm development in *E. coli* [60], whether population diversification through Lsr QS represents bet-hedging or some other transient specialization within the context of a population-wide transition to a sessile lifestyle remains a point of significant interest. Regardless, *in silico* experiments here clearly frame Lsr QS activity within a context of phenotype diversification. We further explored how this diversification operates spatially and between strains with different Lsr/LuxS activity. Spatial self-arrangement of Lsr activity was evaluated and contrasted against that of LuxIR. LuxIR dynamics produced a sudden rapidly expanding activation from the colony center. With an autoinducer that requires active transport between cell compartments, the Lsr system provided a stark contrast by producing speckled activation. This distinction may be attributed to the coupling of positive intracellular and extracellular

feedbacks to one another in LuxIR QS, whereas intracellular positive feedback drives an extracellular negative feedback in Lsr QS (recall **Figure 3-1**).

Lsr QS populations subject to AI-2 chemoattraction displayed an interesting collect-disperse behavior, which informed a complete cycle of nutritional resource exploitation: seek, find, deplete, and seek again. To our best knowledge, this is the first time that such behavior has been characterized by any mathematical model or otherwise described in the context of the specific QS mechanisms investigated here. Since extracellular AI-2 could be viewed as a potential marker for the exponential growth of a broad swath of bacterial species (its production is a direct consequence of primary carbon metabolism) [27,46], motility toward such active growth is likely to be most fruitful for a generalist like *E. coli*, potentially reinforcing chemotactic tendencies toward other substrates. As catabolite repression is epistatic to QS switching/Lsr activation [55], which is itself correlated with late exponential/early stationary phase growth [54], the rapid depletion of AI-2 from local extracellular space leading to cell redispersion could be seen as part and parcel of a larger coordinated behavior of a population seeking out new energy sources to exploit. Alternatively, as Lsr signaling appears to mediate both the influence of AI-2 on biofilm development [60] and its consumption as a secondary metabolite [46], partial population activation in this context may represent a mechanism for the alignment of secondary metabolism consumption with the metabolic burden of producing public goods. As Lsr activation loosely corresponds with growth phase transitions, one specific instance of this concept is the possibility that chemotaxis to self-secreted AI-2 draws a population toward surfaces at which preparations for a sessile

lifestyle are initiated by the QS activated population fraction while the remaining cells continue to seek out other fertile grounds for colonization.

Aside from phenotype diversification within a population, the ability to move toward a favorable growth environment and absorb signaling molecules from the extracellular space could have multiple motivations. For example, influx of AI-2 via Lsr activation could be used to limit QS sensing processes in competitors — processes involved in the transition to quiescence or sessility including: increased antibiotic resistance in local bacteria [138]; increased predation defense [139,140], by preventing other AI-2 consumers from competing for a similar niche; exploiting available secondary metabolites [46], or some combination thereof. Additionally, as an established population, competition for a secondary metabolite might be an expensive proposition — especially if the metabolite or its breakdown products moonlight as epistatic signals for the transition to stationary phase phenotypes. However, denying potential competitor populations secondary metabolite and/or limiting the prevalence of potential chemotactic cues could be a worthwhile tradeoff, particularly if only a select fraction of the population is required for the job.

Whichever if any of these possibilities obtain for any particular circumstance, they all fall under the broad umbrella of hoarding amid the relative phylogenetic abundance of AI-2 in the microbial world [27]. To more carefully consider how changes to the Lsr system itself might inform hoarding dynamics in consortia, mixed population simulations were performed.

In mixed simulations, we found that *luxS* mutants remained largely inactive when paired with wildtype cells. This is attributable to a zero baseline of intracellular AI-2 in such cells, thus requiring a much higher extracellular threshold of AI-2 prior to activation. These populations represent a type of QS cheat termed "signal-negative" [141]. Such cells are metabolically unburdened by the production of autoinducer even though they participate in the production of social goods and enjoy their benefits. In this particular case, the Δ*luxS* population was also unable to compete for AI-2 against signal-producing wildtype cells, and remained largely QS inactive. Thus, in mixed cultures, *luxS* negative cells were effectively signal blind as well. As long as QS products are public goods, *luxS* mutants are likely to be doubly non-cooperative. If the benefit of activation is direct or the resulting product is not a public good for the consortia, such a defect would likely minimize signal-negative cheating. This incentive for cells to avoid signal-negative cheating is compounded by any metabolic burdens arising from an incomplete methyl cycle [46], creating an incentive for cooperation even more direct than those previously considered for generic QS processes [142]. Alternatively, *luxS* negative-Lsr positive cells could play a niche role within bacterial consortium, consuming AI-2 made by competing species [89].

LsrF and *lsrG* double mutants were chosen to study in part because they are the most commonly missing elements among Lsr homologs [14]. They were also found to be more sensitive to AI-2 in synthetic biology contexts [131], internalizing and accumulating AI-2 more rapidly than wildtype cells. Δ*lsrFG* cells thereby suppressed wildtype activation. In general, *lsrFG* mutants present a sort of paradox. Even though these populations may be slightly advantaged by a faster activation and corresponding AI-2

internalization, they are unable to utilize this increased pool metabolically, resulting merely in a higher AI2-P intracellular steady state. This may be warranted if the only role of AI2-P is to effect multiple downstream phenotypes via pleiotropic LsrR derepression. However, the degree to which LsrR derepression upregulates non-Lsr genes remains unknown [143].

Alternatively, as representatives of other strains or species, interactions of lower AI-2 sensitivity strains (Δ*luxS)* and higher AI-2 sensitivity (Δ*lsrFG*) strains with wildtype cells suggest that consortial Lsr activation for cell types with different Lsr/AI-2 sensitivity groups is likely to be highest for any particular subset when all constituent cells share a similar or lower Lsr/AI-2 sensitivity. This adds another layer of complexity to the efficiency sensing view of QS, where bacteria use QS machineries to evaluate whether dependent secreted public goods will serve their intended purpose, insofar as Lsr QS may allow interrogation of the environment for other cells with varied Lsr/LuxS affinity types. This potential outcome is one of many stemming from the intercellular negative feedback of the Lsr system operating even amid simple consortia as studied here.

3.6 Concluding Remarks

While the simulations contained herein attempt to model quorum sensing contextually, the extent to which our *in silico* experiments correctly predict emergent behaviors must be considered [125].

As a general matter, the reported results rely upon the validity of assumptions regarding the sensitivity of QS and chemotactic behaviors to AI-2. As a rule we have tried to make assumptions that are supported by the literature. For example, the minimal concentration at which AI-2 triggers chemotactic behaviors appears to be lower than that at which QS can be triggered for wildtype Lsr systems [131]. Even if both behaviors were triggered at similar concentrations, clustering phenomena still likely obtain due to time lags associated with transcription, translation, and rate-limited transport through low flux pathways.

Additionally, as previously mentioned, QS phenomenon apply only where extracellular transport is dominated by diffusive processes, when convection is minimal; otherwise, autoinducer dilution prevents activation [144]. Therefore, QS has often been reframed as diffusion or efficiency sensing [145]. Regardless of these semantics and connected interpretations, these phenomenon are likely to prevail within a protected volume or cavity with restricted access to flow conditions.

Of further complication is the fact that the greater the diffusivity of the AI-2, the less pronounced chemotactic phenotypes are likely to be, as gradients rapidly dissipate. Also, the greater the diffusivity of AHL, the slower the colony will be to activate. While diffusivity coefficient values were specific to AI-2 and AHL, they were estimates based

on Wilke-Chang correlation calculations[146]. However, while increasing diffusivity would blunt chemotactic behaviors, bimodality would be exacerbated as the effective distance between cells became smaller. An additional concern regarding the extracellular transport of chemical species is that bacterial movement might be reasonably expected to result in superdiffusive conditions directly proximal to the cell membrane. Due to the low Reynold's number at these scales, however, this superdiffusivity should not extend into the surrounding bulk.

Moreover, when asserted noise is tuned lower, Lsr bifurcation is also likely to dissipate. While the heterogeneity chosen is probably conservative, studies of population heterogeneity that track a significant number of single cell expression histories were unable to be identified.

Finally, from a purely qualitative perspective, swimming populations generally appear to slow upon entering late-exponential or early-stationary phase growth. That is to say that the average speed of propulsion becomes markedly distributed, in a manner probably associated with growth phase heterogeneity. Commonly, in such circumstances, macroscopic self-propulsion comes to a halt altogether for some cells. As QS activity is often associated with growth phase transitions, it is beguiling to imagine a scenario where chemotaxis draws a population toward surfaces at which sessile behaviors are initiated by the less mobile population upon QS activation, while cells that still can respond chemotactically are less likely to be QS activated and continue to seek out more fertile grounds for colonization.

3.7 Supplemental Information

3.7.1 Text and Discussion

The focus of the main text, the Lsr system, is one of two known signaling transduction systems that responds to the 'universal' quorum sensing signal, autoinducer-2 (AI-2). In *E. coli* [60], *Salmonella* [77], and *A. acetinomycetemcomitans* [67], Lsr based QS drives or at least influences cooperative behaviors among multicellular consortia of bacteria. Experiments from Tsao, et al. [94] used a two plasmid system to amplify Lsr based expression by transcribing T7RNAP under pLsrR control on a single copy plasmid and linking GFP expression to a T7 promoter on a mid-to-high-copy pET200 plasmid. By one interpretation, this created a reporter system for Lsr QS activation [144,147]. Interestingly, pure cultures of reporter transfected-wildtype bacteria developed a fluorescence distribution similar to that of mixed cultures. This suggested that one fraction of the wildtype population became QS active, while the other remained off in a bimodal fashion, belying the general association of QS with whole population coordination.

Previous studies have modeled the prototypical LuxIR QS system as producing a strong switching behavior given sufficiently high cell density [127], usually attributed to the positive feedback inherent in the system topology [148]. The Lsr system too, has been shown to generate behavior reflecting a highly sensitive switch [112]. Modeling of Lsr activity in distinct cells within a population has not previously demonstrated lasting bimodality of system expression [95], however.

Unlike the LuxIR system, Lsr topology resembles that of sugar importer systems like the *lac* operon, which has been shown to produce bimodal activity when exposed to a nonmetabolizable inducer [149]. While thiomethyl galactoside and isopropyl-β-D-thiogalactopyranoside served as a constant stimulus in those experiments, sufficient production of AI-2 throughout the time course of interest could serve as an analog in the case of the Lsr system.

Using parallel ODEs to represent individual cells drawing from the same extracellular AI-2 pool, our model suggested that one subpopulation could prevent another from activating when paramet+ers were sufficiently perturbed. This analysis was extended to demonstrate that drawing parameter values from log normal distributions with different variances resulted in a range of partially activated populations. Additionally, our model suggested that noise from spatially associated stochasticity was by itself insufficient to generate bimodal expression patterns.

3.7.2 Methods

3.7.2.1 *Lsr ODE Model*

3.7.2.1.1 *Generalities and scope.* The model described herein was developed to simulate Lsr system behavior for bacteria in a batch reactor between lag and early stationary growth phases. At the endpoint, AI-2 should be functionally depleted from the supernatant [54,55,57], with extracellular concentrations peaking between 4-6 hours [150]. The developed equations relied on Hill and Michaelis-Menten like expressions to encapsulate reaction rate behaviors. In order to investigate population bifurcation, separate sets of

ODEs were run simultaneously, modeling multiple cells sharing the same extracellular space.

Lsr system components were modeled as translated from polycistronic mRNA species *lsrRK* and *lsrACDBFG*, the transcription of which was considered a function of LsrR and intracellular AI-2 concentrations. The proteins LsrK and LsrR were modeled as distinct species, LsrF and LsrG were modeled as a single entity. LsrA, LsrC, LsrD, and LsrB were folded into a single type of entity, OP, as they form a complex ABC type importer. Additionally, "LsrFG", for while their products are distinct they both use AI2-P as their substrate and neither's products are known to feed back onto Lsr activity. Thus LsrF and LsrG were functionally equivalent for our specific purposes [54]. Of the proteins outside the divergent Lsr operons that influence Lsr activity in *E. coli* (*luxS*, *ydgG*, *crp*, and *pts*), none are themselves modulated by Lsr activity [55,58,63]. Moreover, the concentration of these species have not been shown to markedly vary within the time frame of interest. These proteins were therefore considered constant and their activity rates were simplified to a Michaelis-Menten like dependence on the substrate alone.

3.7.2.1.2 *mRNA expression.* Production of mRNA species was modeled with a modified Hill equation derived from an expression for the fraction of free DNA to total DNA (free DNA + repressed DNA) as a function of LsrR and AI2P concentrations. Whereas active LsrR was modeled as a tetramer [151], AI2P cooperativity was asserted. Cooperativity for AI2P derepression was assumed to be on the order of cooperativity for protein/DNA binding. A lower bound for such cooperativity might be 1.38-2.72, which was measured for a nonspecific interaction between DNA and a multidomain protein in *Mycobacterium smegmatis* [152]. An upper bound for such interactions is believed to be around 10 [153]. As

each LsrR has a distinct AI2P binding domain, we chose 4 as a baseline level of cooperativity. The trajectory of the system using this cooperativity was similar to that for degrees of cooperativity greater than 4. At tested cooperativities less than 4, the drop in activation sharpness was sufficient to noticeably dull the rate of Ai-2 recompartmentalization from the extracellular space. The transcription rates for the two different polycistronic mRNA species were set as equivalent, as the data on the relative strength of expression toward LsrA and LsrR is inconsistent [56,70].

3.7.2.1.3 *Protein synthesis.* Lsr proteins expressed from the same polycistronic mRNA species were modeled as having the same translation rate. While different ribosome binding sites along the polycistrons are likely to have different affinities for ribosomes, in the absence of data, the co-regulation implied by operon structure and protein function was given overriding consideration.

3.7.2.1.4 *mRNA and protein degradation.* As an additional simplification, both polycistronic mRNA species were modeled with the same rate of degradation, with a half-life on the order of 10 minutes, which was an order of magnitude faster than that used for proteins. While Lsr proteins are expected to have different vulnerabilities to proteolytic degradation, the result is either higher or lower quasi-steady state levels. The effects of such differences could be accounted for by either lowering or increasing corresponding unconstrained enzymatic rates.

3.7.2.1.5 *Cell growth.* Cell growth was modeled as Monod growth. The maximum rate and Monod saturation rate constants of 0.032/min and 75 μM, respectively, were had from a fit to OD_{600} measurements from a batch growth at 37°C. The growth rate also

informed the dilution of cellular components. As a generality, the rate of cell growth strongly influenced simulation results, as increasing cell density accelerated extracellular AI-2 accumulation. That is, faster growth decreased the time to autoinduction.

3.7.2.1.6 *AI-2 transport.* Instantaneous concentration and dilution associated with transport between extracellular and intracellular spaces was treated by including a dilution/concentration term of 10^{12} as seen in equation (10). This dilution factor accounted for the difference between the femtoliter intracellular environment and the milliliter term associated with cell concentration.

Influx into the periplasm through porins was modeled as a diffusion process with a Michaelis-Menten form, rate limited at high concentrations, but otherwise proportional to the concentration difference between extracellular and periplasmic compartments. Influx through Lsr ABC type complexes was modeled as a function of importer complex concentration and periplasmic AI-2. The transporter's component proteins were treated as a single species. Transporter complex was assumed to form with a cooperativity of 4, as a reflection of the fact that four independent components are involved in complex formation, namely LsrA, LsrB, LsrC, and LsrD. Although the nucleotide binding component, LsrA, operates as a dimer, any additional degree of freedom was considered to be eliminated by the fact that the LsrA dimer pair is fused together. Further, this transporter complex was asserted to act upon periplasmic AI-2 according to a Michaelis-Menten like dynamic. Michaelis-Menten like dynamics were also asserted for the operation of the alternative importer as well as for AI-2 export through YdgG. That is, these processes were assumed to be constrained by a maximal velocity above a saturating

AI-2 concentration, and that at lower AI-2 concentrations activity was approximately linear with respect to AI-2 itself.

3.7.2.1.7 *AI-2 degradation and synthesis.* The rate of extracellular AI-2 degradation was assumed to be minimal based on a lack of attenuated AI-2 activity in bioassays after incubation of *in vitro* synthesized samples overnight at 37^{o}C. The rate of cytoplasmic AI-2 degradation was set significantly higher to account for experiments measuring cytoplasmic AI-2 in *E. coli* without functional Lsr activity, wherein cytoplasmic AI-2 concentrations dropped significantly once the stationary phase had been well established [150]. While the time frame of this marked decrease fell outside the scope of *in silico* experiments herein, it nonetheless suggests the existence of mechanisms that degrade AI-2 intracellularly, independent of Lsr system expression. Periplasmic degradation of AI-2 was modeled as intermediate to the extracellular and cytosolic rates, and did not bear strongly on the activity of the Lsr system, since the absolute moles in periplasmic pools were limited compared to extracellular and intracellular species.

The same experiments that provide evidence for this degradation also suggest that the rate of synthesis is not constant [150], even if the expression of *luxS* and *pfs* are not strongly varied across time [56]. Nonetheless, we model AI-2 synthesis as constant, which over the course of the time scale of interest we assume to be operationally approximate.

3.7.2.1.8 *Equations*

The specific form of the equations used and the parameters values used herein were as follows:

1868

1869 (1) $$\frac{d[AI_{peri}]}{dt} = V_{\max porin}\frac{[AI_{out}]-[AI_{peri}]}{Km_{porin}+[AI_{out}]+[AI_{peri}]} - (V_{in}[OP]^{coop}+basal)\frac{[AI_{peri}]}{Km_{ABC}+[AI_{peri}]} - (\gamma_{AIperi}+\mu(t))[AI_{peri}],$$

1870 (2) $$\frac{d[AI_{in}]}{dt} = K_{synth} + (V_{in}[OP]^{coop}+basal)\frac{[AI_{peri}]}{Km_{ABC}+[AI_{peri}]} - k_{Phos}[LsrK]\frac{[AI_{in}]}{Km_{Phos}+[AI_{in}]} - V_{ydgG}\frac{[AI_{in}]}{K_{export}+[AI_{in}]} - (\gamma_{AIin}+\mu(t))[AI_{in}],$$

1871 (3) $$\frac{d[AI2P]}{dt} = k_{Phos}[LsrK]\frac{[AI_{in}]}{Km_{Phos}+[AI_{in}]} - k_{Cat}[LsrFG]\frac{[AI2P]}{Km_{Cat}+[AI2P]} - (\gamma_{AI2P}+\mu(t))[AI2P],$$

1872 (4) $$\frac{d[OP]}{dt} = k_P[mRNA2] - (\gamma_T - \mu(t))[OP],$$

1873 (5) $$\frac{d[LsrFG]}{dt} = k_P[mRNA2] - (\gamma_{FG} - \mu(t))[LsrFG],$$

1874 (6) $$\frac{d[LsrR]}{dt} = k_B[mRNA1] - (\gamma_R - \mu(t))[LsrR],$$

1875 (7) $$\frac{d[LsrK]}{dt} = k_B[mRNA1] - (\gamma_K - \mu(t))[LsrK],$$

1876 (8) $$\frac{d[mRNA1]}{dt} = V_{mR1}\frac{1}{1+\frac{k_2[LsrR]^{coop2}}{r_2+[AI2P]^{coop3}}} - (\gamma_{MR1}+\mu(t))[mRNA1],$$

1877 (9) $$\frac{d[mRNA2]}{dt} = V_{mR2}\frac{1}{1+\frac{k_2[LsrR]^{coop2}}{r_2+[AI2P]^{coop3}}} - (\gamma_{MR2}+\mu(t))[mRNA2],$$

1878 (10) $$\frac{d[AI_{out}]}{dt} = dilutionF[cells](V_{max\ porin}\frac{[AI_{peri}]-[AI_{out}]}{Km_{porin}+[AI_{out}]+[AI_{peri}]} - V_{ydgG}\frac{[AI_{in}]}{K_{export}+[AI_{in}]})$$

1879 (11) $$\frac{d[cells]}{dt} = [cells]\frac{yield[Substr]}{K_{mid}+[Substr]},$$

1880 (12) $$\frac{d[Substr]}{dt} = -take[cells]\frac{yield[Substr]}{K_{mid}+[Substr]}$$

, where [*Substr*] represents the concentration of substrate, [AI_{peri}] represents the concentration of periplasmic AI-2, [AI_{in}] represents the concentration of cytoplasmic AI-2, [AI_{out}] represents the concentration of extracellular AI-2, [*AI2P*] represents phosphorylated AI-2, [*LsrR*] represents LsrR concentration, [*LsrK*] represents LsrK concentration, [*LsrFG*] represents LsrF and LsrG concentration, [*mRNA2*] represents *lsrRK* concentration, [*mRNA1*] represents *lsrACDBFG* concentration, and [*OP*] represents transporter protein concentration.

3.7.2.1.9 *Parameter Values.* A table of parameter values can be found in **Table 3-S1**. The maximal rate of transport through the alternative importer, as governed by *basal*, was set lower than the rate of export through YdgG, allowing extracellular AI-2 accumulation at baseline levels of Lsr activity. The remaining parameters were set such that this initial AI-2 flux out was overcome by Lsr activation.

Specifically, according to BB170 bioassays of culture supernatant, AI-2 appears to peak between 4-5 hours in a batch culture. FRET-LuxP assays and bioassays suggest that the peak concentration is less than 80 μM and greater than 40 μM, respectively [55,150]. Once extracellular accumulation stalls, AI-2 appears to be drawn down to low concentrations of AI-2 in less than an hour and then to concentrations below bioassay sensitivity within the next hour.

Asserting that transcription is directly coupled to translation for LacZ, transcription was fit to Miller assays of LacZ expression from pLsrA and pLsrR promoters, indicating that transcription begins evolving prior to 4 hours after culture initiation, reaching a several fold higher level of expression prior to 6 hours [56].

Table S1. Estimated parameter values

Parameter	Value	Parameter	Value	Parameter	Value
$V_{max\,porin}$	8000 μM/min	γ_{AIin}	0.15 /min	$r2$	$7x10^{-5}$ μM^4
Km_{porin}	1 μM	$kCat$	50 /min	$coop2$	4
V_{in}	$2x10^{12}$ /(min*μM^3)	Km_{cat}	1.4 μM	$coop3$	6
$coop$	4	k_P	18.8 /min	γ_{MR1}	0.693 /min
$basal$	500 μM/min	γ_T	0.01 /min	γ_{MR2}	0.693 /min
Km_{ABC}	0.5 μM	γ_{FG}	0.01 /min	$dilutionF$	10^{-12}
γ_{AIperi}	0.015 /min	k_B	18.8 /min	γ_{AI2P}	0.3 /min
K_{synth}	220 μM/min	γ_R	0.01 /min	$yield$	0.032 /min
k_{phos}	80 /min	γ_K	0.01 /min	K_{mid}	2250 μM
Km_{phos}	1.4 μM	V_{mR1}	$2x10^{-5}$ μM/min	$take$	$4.5x10^{-7}$ [Substr]/[cell]
V_{ydgG}	1250 μM/min	V_{mR2}	$2x10^{-5}$ μM/min		
K_{export}	0.5 μM	$k2$	$3.5x10^8$		

While the selected parameter set in combination with the given ODE's produced a much faster depletion of extracellular AI-2 than was observed experimentally, the fraction of QS induced cells evolved over an extended period of time [94], and it was this evolving population's averages to which the model was fit (**Figure 3-S1**). In order to generate an evolving fractional Lsr autoinduction, a tractable number of cells was modeled within a finite difference agent based scheme. Each cell's processes were modeled by its own set of ODEs and the cell population was slightly desynchronized in order to generate the fractional induction seen by Tsao, et al [94].

A parameter search was carried out to identify parameter sets satisfying the above criteria. Among other measures of fit, the time to Lsr autoinduction was sensitive to changes in all parameters effecting transport due to their direct bearing upon the balance of accumulating cytoplasmic AI-2 and corresponding AI2-P species. The parameter space satisfying the available data including extracellular AI-2 concentration, was nonetheless broad with similar sensitivities and behaviors over a range of values.

3.7.2.1.10 *Initial Values.* Initial values were set according to approximate steady state values for a system without lacking AI-2 production, where the Lsr system was uninduced: [cells] = 3e7, [*Substr*] = 1500 μM, [AI_{peri}] = 0.015 μM, [AI_{in}] = 0.2 μM, [AI_{out}] = 0.02 μM, [*AI2P*] = 0.03 μM, [*LsrR*] = 0.0012 μM, [*LsrK*] = 0.0012 μM, [*LsrFG*] = 0.0012 μM, [*OP*] = 0.0012 μM, and [*mRNA2*] = [*mRNA1*] = 0.

3.7.2.1.11 *Numerical solutions.* The equations and parameters were solved using NDSolve in Mathematica 8.0 utilizing "StiffnessSwitching" methods. The solution required the "StiffnessSwitching" option, without which, the solution was unstable.

3.7.2.2 *Finite difference-agent based model*

3.7.2.2.1 *Modeled Environment.* The environment was defined as a 500 x 500 x 6 μm volume, and was divided into 2 x 2 x 6 μm elements. Cells either exported or imported AI-2 from their intracellular space into or from the finite difference element in which their cell centers were found. AI-2 also diffused between finite difference elements as modeled by a forward in time-central in space scheme with an assumed diffusion coefficient of $5x10^{-7}$ cm^2/s. The boundaries of the simulation were modeled as impermeable.

3.7.2.2.2 *Adaptation of equations and solutions.* With few exceptions, the previously described ODEs were repurposed without modification in the finite difference/agent based model. Growth was one exception, as division became a discontinuous stochastic event. Monod growth dynamics informed the median of a log normal distribution of doubling time with a σ of 0.05 where the median rate was updated every time step. The exchange of AI-2 between the environment and a cell was localized to the space in which the cell center was found at the beginning of the time step. This allowed a synchronized update of the final grid concentration at the end of the time step. Furthermore, the AI-2 dilution/concentration factor was adjusted from 10^{12} to 24 to account for the difference between the milliliter volume associated with cell concentration and the implied volume of grid elements.

3.7.2.2.3 *Cell ODE Numerical Solution Method.* The numerical method used in the agent based modeling was a second order Runga-Kutta with an *ad hoc* allowance for stiffness. In order to achieve efficient calculation, we used an explicit method with the exception of periplasmic AI-2 after transporter concentration exceeded 4 times its initial concentration. This threshold was chosen based upon NDSolve interpolation, as a point at which the Lsr system including transporter expression and subsequent concentration had already begun its transition to an active state. After this threshold was surpassed, porins were assumed to be the rate limiting element in the transport of AI-2 from the extracellular space to the cytosol, and modeling was made to reflect this by substituting terms describing ABC-type transporter activity with terms describing porin activity; furthermore, periplasmic AI-2 was held at zero, reflecting the numerical solution from the ODE system described in previously.

3.7.2.2.4 *Cell Division.* Cell division was governed by individual counters that incremented at each time step. Once a cell's counter exceeded its doubling time, division occurred. At the time zero, cell counters were set randomly between zero and the maximum allowed value. Doubling time was varied between cells according to a mean growth rate based on Monod kinetics (with parameters found in **Table S1**) with a log normal distribution (variance of 0.05) in order to desynchronize cell doubling. Upon division, both mother and daughter cells acquired new growth rates from a log normal distribution, while bearing duplicate properties including initial position, and their age counters were reset to zero. Newly initialized cells were assigned a basal rate of AI-2 flux through alternative importer pathways (the parameter, *basal*) or a rate of AHL synthesis, for Lsr or LuxIR simulations respectively.

3.7.2.2.5 *AI-2 Diffusion.* AI-2 diffusion was modeled using a central difference approximation. While both truncation and roundoff error arise from this process, the overall behavior of the simulation was not expected to be dramatically impacted, as clustering behavior also obtained when elements were four times the size. The diffusion coefficient used (~$5x10^{-7}$ cm^2/s) was approximated using the Wilke-Chang correlation [146].

3.7.2.2.6 *Time interval ordering.* The order of calculations within a time step was as follows. First, the average growth rate was determined from a Monod growth dynamic solved by a second order Runga-Kutta method as a function of substrate concentration and *E. coli* density. Second, each *E. coli* divided or did not divide, marginally accommodated its AI-2 chemotactic threshold (if appropriate), and moved (according to the particular scheme employed). For the specific purposes of the simulations reported here, cells moved randomly in space each time step at an average rate of 20 μm/sec, coming to an average distance of 1.3 μm per time step. Correspondingly, cells rarely moved more than one grid element at a time for any given time step. Finally, each grid was subject to calculations approximating diffusion. That is, AI-2 was allowed to move according to molecular diffusivity and distance/time calculations. After all diffusion processes were calculated over the entire environment, the exchange of AI-2 between bacteria and the grids in which the bacteria were present was calculated according to Lsr dynamics governed by previously described ODEs. The sum of these changes to AI-2 concentration were then applied to each grid and each cell. Time step and grid size were chosen such that moderate changes to these measures resulted in qualitatively indistinct outcomes.

All random assignments were drawn according to the Mersenne Twister algorithm [154]. Seeds for the algorithm were changed every simulation.

3.7.3 Results

3.7.3.1 Numerical solutions to ODEs. Given the ODEs and parameter values used herein, the system evolved from a state of low expression to a state of high expression (QS activated), as can be seen in **Figure 3-S1**, where the trajectories of different state variables are represented, normalized to their maximum values. This corresponds to the known pattern of Lsr activity development. At a particular AI2-P-to-LsrR threshold, system expression is de-repressed, leading to the increased expression of transporter proteins, which itself leads to an influx of extracellular AI-2 and a sustained increase in the expression of *lsr* mRNA species.

In the model, heightened activity appeared to persist even after the depletion of extracellular AI-2, by virtue of an altered flux balance from net AI-2 exportation to one of net importation of processing. That Lsr autoinduction leads to a shifts in the compartmentalization of AI-2 species from one dominated by extracellular species to one dominated by cytosolic species is one interpretation of **Figure 3-S2**.

3.7.3.2 System Sensitivity to basal rate of AI-2 uptake. As expected, system behavior was sensitive to parameters that markedly affected the rate of AI2-P accumulation when Lsr system expression was low. Among these was the parameter *basal* (**Figure 3-S3**). With increasing $basal$ value, the rate of AI2-P accretion increased, accelerating the time to Lsr activation.

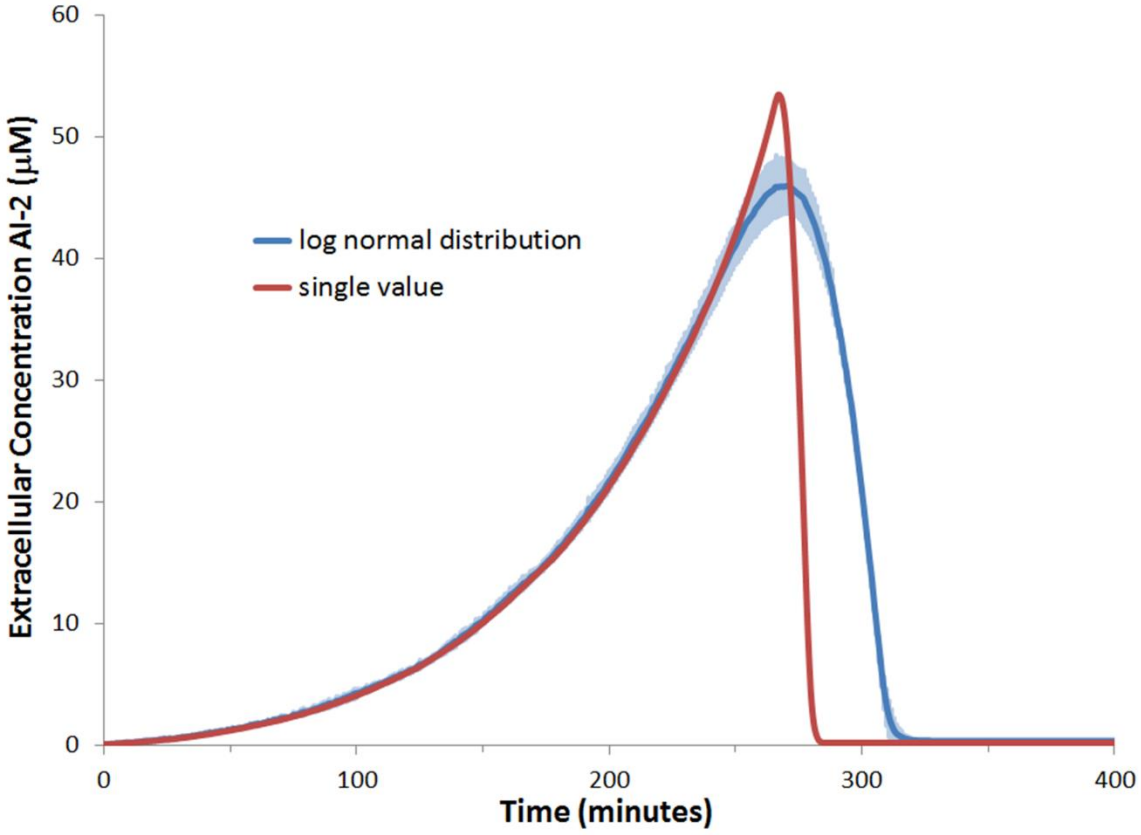


Figure 3-S1. Comparison of solution for population with a single basal value versus a population with a unimodally distributed value of *basal*. Juxtaposition of the solution for extracellular AI-2 for a simulation of cells with a single basal value versus the average solution of extracellular AI-2 for a simulation of cells with a log normal distribution of the parameter *basal*

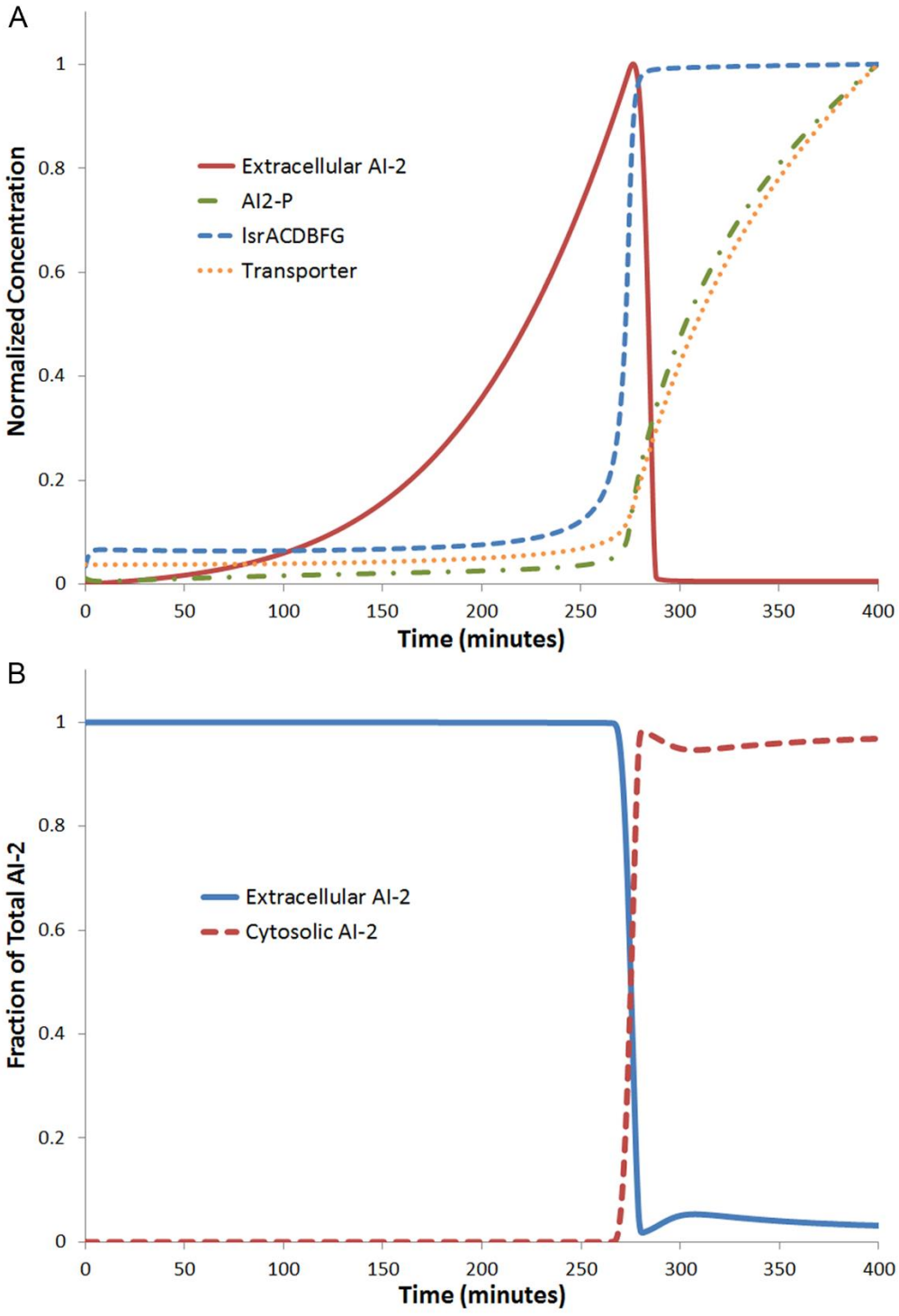


Figure 3-S2. Numerical solution to developed ODE equations. A Selected interpolated trajectories of the solution to the ODEs from 3.3.1.7 with the parameters from 3.3.1.8. **B** Mole fraction of total AI-2 species located either extracellularly or cytosolically. The denominator was the mole sum of extracellular, periplasmic, and cytosolic AI-2.

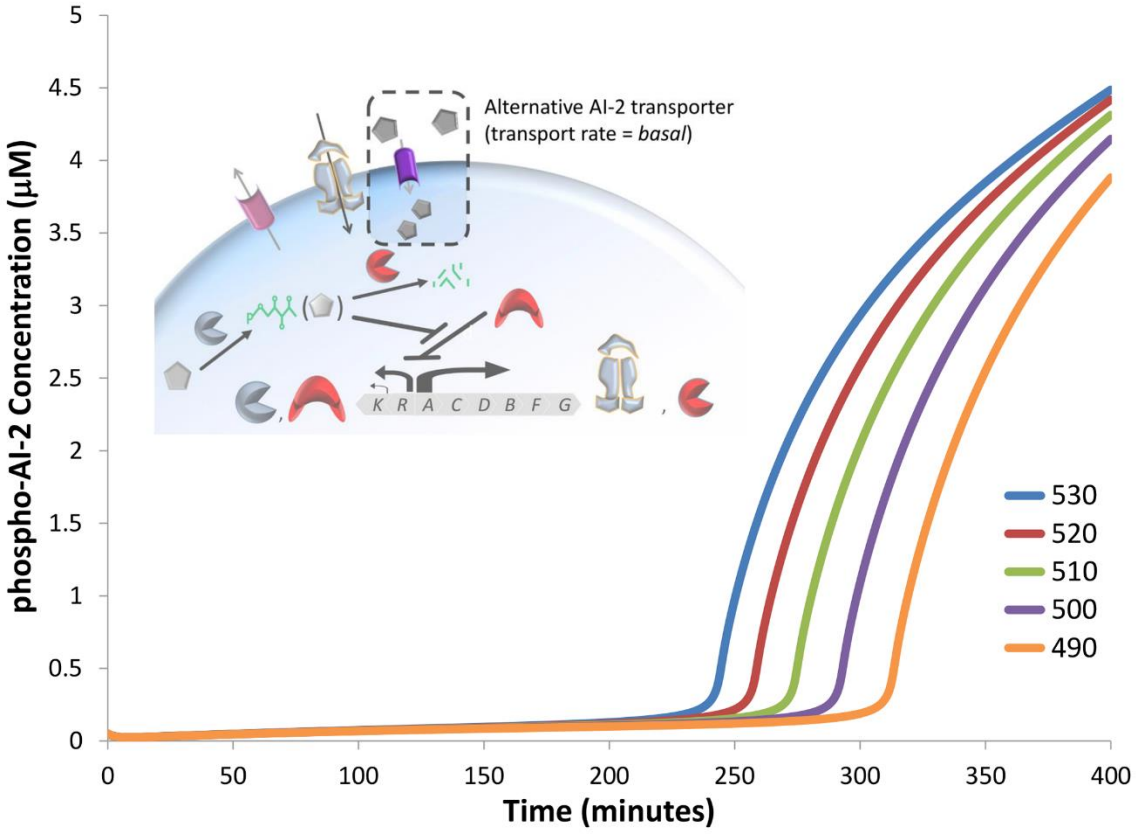
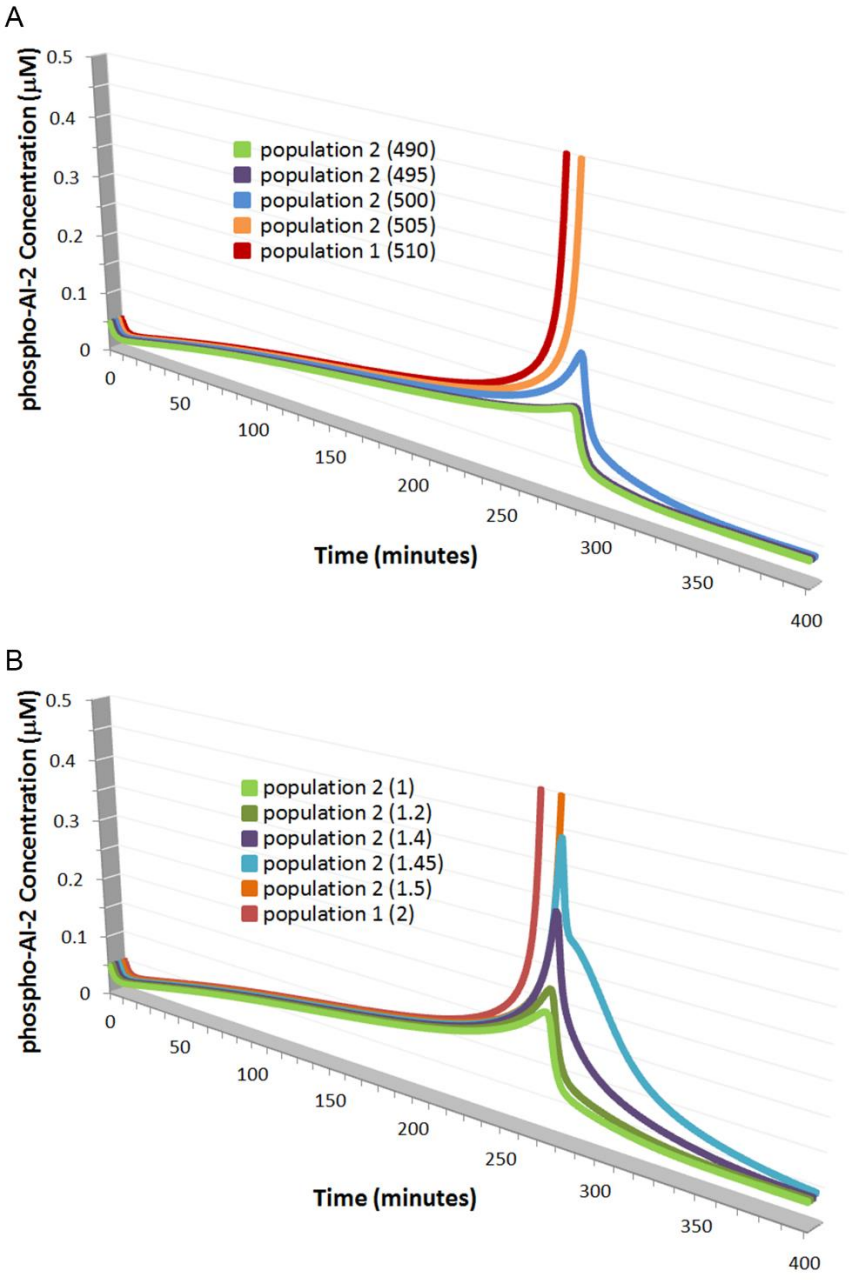


Figure 3-S3. Parameter sensitivity to the parameter basal as seen by changes in the time until system activation. System trajectories given different values of the parameter *basal*.

Specifically, an ~8% increase in the rate of AI-2 influx through the alternative pathway from the base value resulted in a ~21% reduction in the time to activation. The values shown in this figure represent a range of basal that corresponds roughly to that of ribose uptake via its low affinity mechanism. Or, correspondingly, that *basal* level denoted 530 corresponds to 10% of the maximum achieved when the uptake mechanism is switched on. All these values (490-530) represent metabolically realistic values of small molecule uptake.

3.7.3.3 Two sets of Lsr with different rates of basal AI-2 uptake. More importantly, when two sets of ODEs drew from the same pool of extracellular AI-2, one system of ODEs could prevent the other from activating given sufficient parameter perturbation, as seen in **Figure 3-S4**. As the perturbation dwindled, the two ODEs converged to the same high expression state. That is, two sets of ODEs, representing two different populations, were arranged to use the same extracellular AI-2. Solutions to ODE's in **Figure 3-S4A** shared the same parameter values except for that of *basal*, which reflects the rate of AI-2 flux into the cell from the periplasmic space through the alternative pathway. The value of *basal* was set to 510 for one population and was adjusted continually lower for a second, gradually slower activating, population. When the difference between the two values of *basal* was sufficiently large, the faster activating population depleted extracellular AI-2 prior to the slower population achieving a sufficient intracellular threshold to trigger Lsr induction, thus preventing QS activation.



2069

2070 **Figure 3-S4. Solution to dual ODE system where second population had varied**
2071 **values of parameters related to the transport of AI-2.** Population 1's parameter values
2072 were held constant. **A** Over separate simulations, population 2's parameter, *basal* was
2073 varied at comparatively lower values. The parameter values are in parentheses. The
2074 trajectory of AI2-P accumulation for each of these cases is represented here. **B** Over
2075 separate simulations, population 2's parameter, V_{in} was varied at comparatively lower
2076 values. The trajectory of AI2-P accumulation for each of these cases is represented here.
2077 The values in parentheses reflect $V_{in}/10^{12}$.

2078

While *basal* represented a specific activity, the phenomenon of a faster Lsr activating group preventing a slower group from inducing proved a more general property of the model, holding when other parameters to which Lsr induction was sensitive were varied instead. This can be seen in **Figure 3-S4B**, where the variation in the assigned value of the parameter V_{in}, representing the rate of AI-2 transport through the Lsr ABC-type transporter, was substituted for variation in the value of *basal*. A decreasing series of values for V_{in} , all smaller than that of population 1, were assigned to population 2. Again, once the difference in V_{in} between populations became sufficient, population 1 was able to prevent Lsr activation in population 2.

3.7.3.4 Full population of cells with Lsr in a finite difference environment. While coupling two systems of ODEs to the same pool of extracellular AI-2 allowed us to interrogate possible Lsr signal bifurcation of two different populations, we extended this analysis to include a full population of cells assigned *basal* values from a log-normal distribution instead of an effective binomial one. This analysis was conducted using mixed finite difference-agent based simulations, as described in the methods.

As discussed in the main text, Lsr induction was bimodal, with the exact balance dependent on the variance of the *basal* distribution. This likely arose due to cells with higher values of *basal* from the leading edge of the distribution activating earlier and earlier as the variance of the distribution increased. The earlier the activation of these cells relative to the mean, the faster extracellular AI-2 species were drawn down, thus preventing the activation of cells with lower *basal* values, which themselves on average were liable to activate later and later.

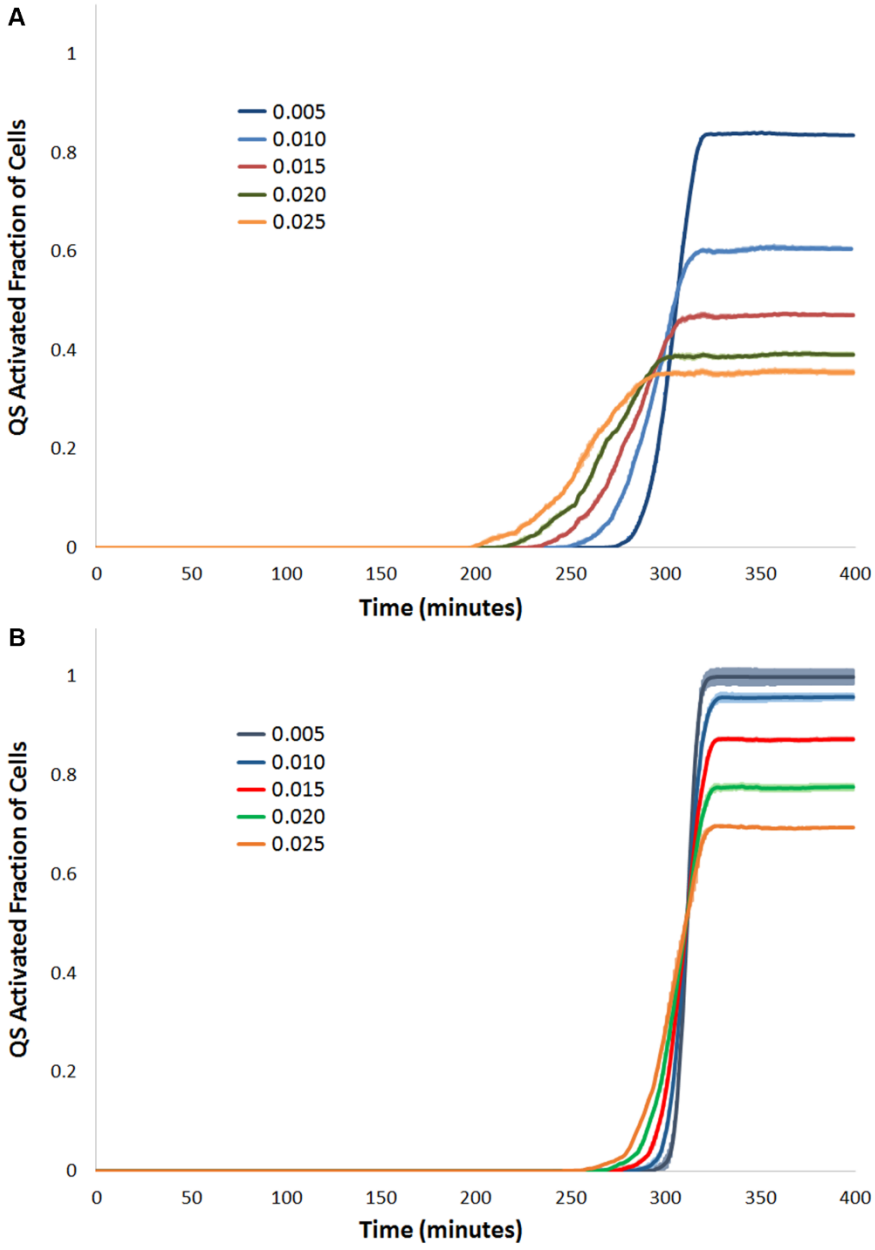


Fig. S5 - Fraction of cell population QS activated decreases as the variation of the parameters K_{synth} (A) and V_{ydgG} (B) increases. The influence of standard deviation of the natural logarithm, σ, of the log-normal distributions on the fraction of the population that was QS activated is represented here for standard deviation values ranging from 0 to 0.025, each run in triplicate. Standard deviation of the natural logarithm for the log-normal distribution was used because the coefficient of variance is solely dependent on this measure. Dark lines represent the average values, whereas lighter surrounding lines represent the standard deviation.

Apart from *basal*, separately varying K_{synth} and V_{ydgG} , the rate of AI-2 synthesis and AI-2 export, respectively, achieved similar changes to the bimodal nature of Lsr activation (**Figure 3-S5**). Although activation levels plateaued at different fractions of the population when the same standard deviations were applied to the distribution of these parameters, increased variation in both K_{synth} and V_{ydgG} consistently decreased the fraction of the population that was ultimately activated. This serves to further indicate that population level bimodal expression of the Lsr system may be a function of any heterogeneity that desynchronizes AI-2 recompartmentalization.

3.7.3.5 Minimal role of spatial heterogeneity of AI-2. In addition to evaluating the role of heterogeneous expression at the population scale, whether spatially associated stochasticity might influence bimodal expression was also inferred, mainly from comparison of simulations using a standard finite difference scheme against simulations where the entire environment was defined by a single element. Treating the environment homogenous made all AI-2 simultaneously available to all cells, whereas cells only interacted with AI-2 in their own element using a standard finite difference approach. As shown in **Figure 3-S6**, the approaches yielded highly similar AI2-P trajectories when cell motility was undirected. This was the case for all state variables modeled. Furthermore, in standard finite element environments, when governed by a single parameter set, cell populations became wholly activated over a very small window (**Figure 3-S5**, $\sigma = 0$). If heterogeneity arising from spatial stochasticity influenced the bimodal phenotype, population activation would be expected to be incomplete. The absence of such an effect implied that spatial stochasticity did not play a marked role in shaping bimodal response.

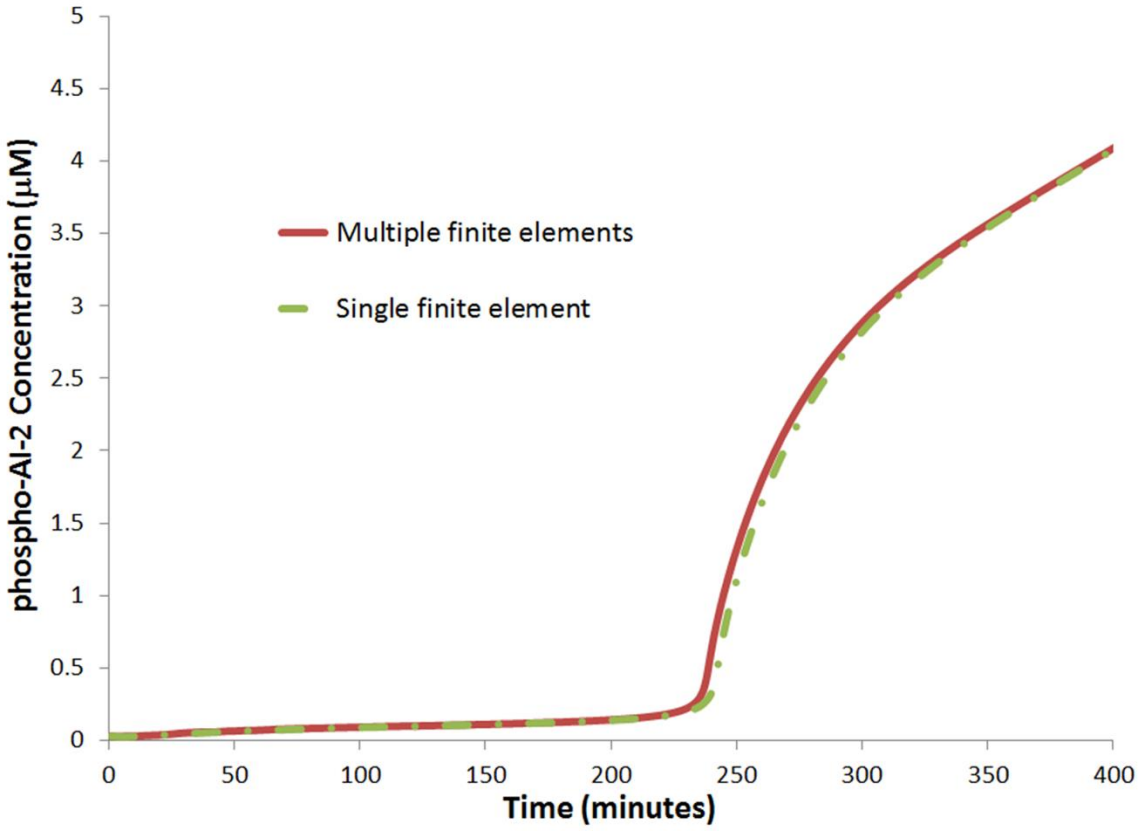


Figure 3-S6. Comparison of results from single versus multiple finite difference elements to define environment. The average trajectory of AI2-P for cells with the same parameter sets in simulations where the environment was defined as either a single finite difference element or by the standard array of elements as defined in the methods. Modeling with a single finite difference element eliminates spatial noise as a source of difference between cells. The addition of noise through the full implementation of finite difference elements, adds spatially associated noise to the simulation. This did not result in a significant change in the average trajectory of AI2-P.

3.7.3.6 Agreement between numerical ODE and finite difference agent based solutions. As a general comment, the agreement between numerical ODE solutions and the finite difference-agent based approach was inexact. In particular, the time to activation was offset between the two solutions as seen in **Figure 3-S7A**. Nonetheless, the solution trajectories were similar and an evaluation of the time to activation as a function of *basal* indicated that parameter sensitivities between the solution approaches were congruous, as seen in **Figure 3-S7B**.

3.7.3.7 Heterogeneity of local Lsr and LuxIR QS activation in growing colonies. Smoothing of heterogenous input by LuxIR QS was not shared by Lsr QS. This is implied in **Figure 3-S8A**, which shows the distribution of key parameters from each simulation type. For LuxIR simulations (left), the uninduced rate of AHL synthesis was varied among cells according to the depicted distribution. The higher the rate of synthesis for uninduced cells, the faster the QS activation. For Lsr simulations (right), the basal rate of AI-2 influx was varied. Here too, the higher the rate of basal AI-2 influx, the faster the QS activation. The red line in both distributions represents the average parameter value for the first cell to activate from twenty simulations. For both Lsr and LuxIR simulations, the first cells to activate were on average all from the higher end of the imposed heterogeneity. For LuxIR QS (left), however, the first cells to activate were not always those with the highest rate of basal AHL synthesis, as indicated by the wide standard deviation. This is clearer when compared to Lsr QS (right), where the first activators were exclusively found at the tip of distribution for *basal* values (reflected in the limited variance). Essentially, while LuxIR/AHL activation smoothed out heterogeneity associated with the rate of AHL production, Lsr/AI-2 dynamics were

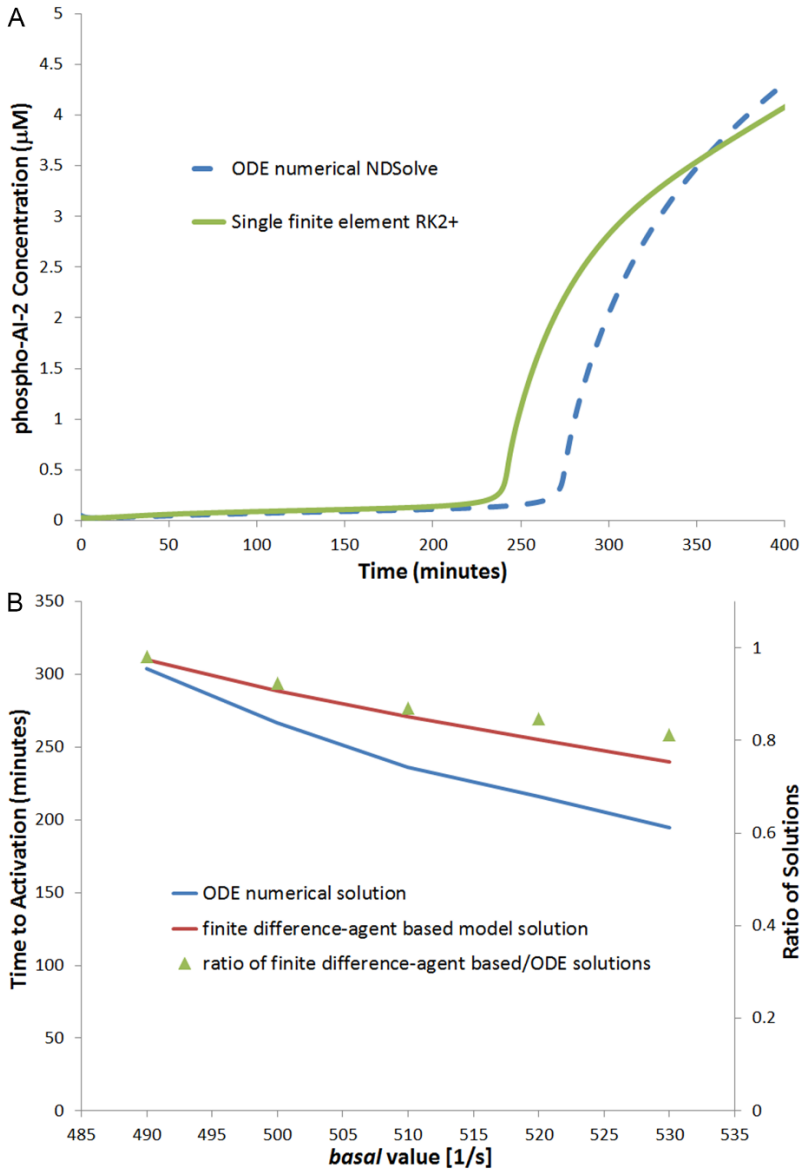


Figure 3-S7. Congruence of solution from finite difference-agent based modeling versus implicit solution of pure ODE's. A AI2-P trajectory from implicit numerical methods and the average AI2-P concentration from the finite difference-agent based approach. Here, cells from the finite-difference-agent based solution all held the same parameter values as that from the pure ODE solution. In the pure ODE approach, cells were modeled as a dependent variable. Ideally, the two solutions would bear identical traces. **B** The rate to activation was assessed by fitting the function, *f(t)*, from 12-152 minutes to a first order linear regression, *g(t)*. The first time point at which $f(t)-g(t)>2g(t)$ was considered the point of activation. The time to activation for each value of *basal* was calculated and the bearing on the solution by the modeling and numerical method used was evaluated by direct comparison along the primary axis and according to the ratio of activation times for the finite difference-agent based solution to the pure ODE solution on the secondary axis.

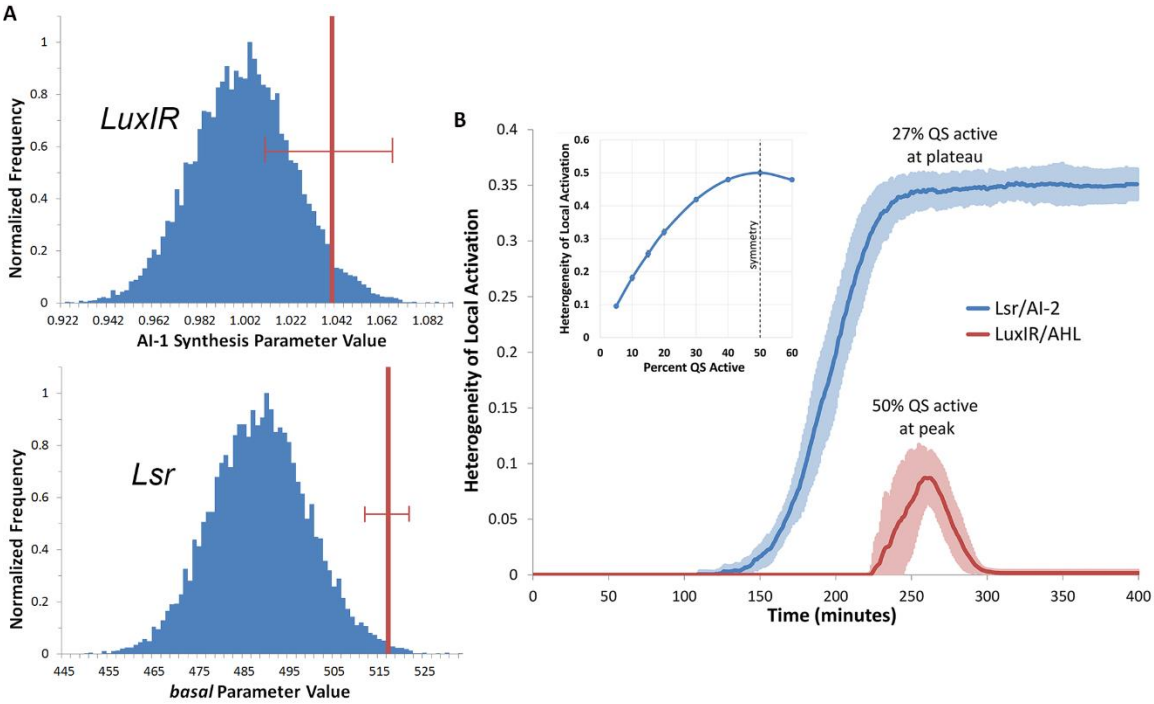


Figure 3-S8. Measures of the difference between LuxIR and Lsr activation in the context of colony growth. **A** Histograms of randomly generated parameter values, each with an event count of 10,000. The average of the parameter values associated with the first cell to QS activate in each simulation and its standard deviation from among 20 simulations are overlaid on the histogram in red. In LuxIR simulations, while the selection is biased, a significant number of cells with less than the median parameter value indicating that LuxIR has a smoothing effect on heterogeneity. **B** The dark lines represent the average local heterogeneity of 20 simulations, while the lighter, surrounding shades represent the standard deviation of those values. Inset is local heterogeneity of QS activation, where active cells were placed randomly within the colony as a function of percent of cells that were QS active for that colony (n=100).

unable to smooth similar non-genetic heterogeneity, instead producing population desynchronization. This difference in response to non-genetic heterogeneity was presumably a reflection of topological differences between QS signaling modules.

This distinction between LuxIR and Lsr QS responses was also reflected in the spatial heterogeneity of activation. Local heterogeneity of QS activity was measured by using a heuristic that scored highest when every cells' neighbors, up or down, left or right, were of the opposite QS state. That is, a value of one indicated a perfect checkerboard pattern of alternating activation, whereas a score of zero indicated that all cells were of the same QS state. That score was then averaged over the entire population to arrive at the score reported in **Figure 3-S8B**. For perspective, inset is a graph of the same measure for colonies whose QS active cells were randomly distributed, with the ordinate axis reflecting the likelihood that any one cells was QS activated. As the measure was only concerned with changes between QS state, scoring was symmetric about 0.5.

At time zero, all simulations began with a score of zero (QS unactivated). Lsr/AI-2 simulations were run with a median basal of 487.8 and a coefficient of variance of 0.052 ($\sigma = 0.0225$). For such an Lsr/AI-2 population, the first non-zero local QS heterogeneity values emerged near 120 minutes. The average of twenty simulations (dark blue line) and the standard deviation (surrounding light blue band) are depicted. As more cells became activated, local QS heterogeneity increased as both outlying and inner cells were induced. Local QS heterogeneity reached a plateau near 0.34 as QS activation began to abate. Given that the local QS heterogeneity for a cell colony wherein the same fraction of QS active cells were completely randomly distributed was 0.4 (inset), the local

QS heterogeneity for Lsr activation was high, while retaining some degree of non-randomness. This stood in contrast to the LuxIR case. LuxIR activation began near the colony center at 225 minutes, and increased rapidly, reflecting the rapid expansion of activation. However, spatial heterogeneity scoring of LuxIR QS activity peaked near 0.1 despite a ~50% QS activation rate at that point. Cell colonies where 50% of the cells were QS active but where those cells were randomly placed within the colony averaged a heterogeneity score near 0.5. Low spatial heterogeneity associated with LuxIR/AHL activation is attributed to two primary factors: QS activity originates from a single centralized location and QS "On"/"Off" distinctions are quickly obliterated as QS active cells turn on their QS inactive neighbors.

3.7.3.8 Evaluation of clustering when Lsr QS is coupled to AI-2 chemoattraction.

Cluster and dispersal patterns were observable from inspection of **Figure 3-S9,** where the distance between cells across simulated time is shown for different swimming modes. Here, fully functioning cells (green, Lsr + chemotaxis) were compared to non-chemotaxing populations (blue, randomly moving) or populations lacking the ability to recompartmentalize AI-2 (purple, *lsr* operon negative and non-specific uptake minus; V_{in} = 0, *basal* = 0). These alternative populations represented groups of cells that did not cluster or clustered but did not disperse, respectively. Initially, the median minimal distances for all population types were identical. The median minimal distance between cells decreased for all populations as a function of growth, as expected. However, the median minimal distance between cells for the two AI-2 chemotaxing populations decreased more rapidly than their non-taxis counterpart due to clustering. Cells unable to recompartmentalize (or uptake) AI-2 had a higher net flux out (they synthesize by do not

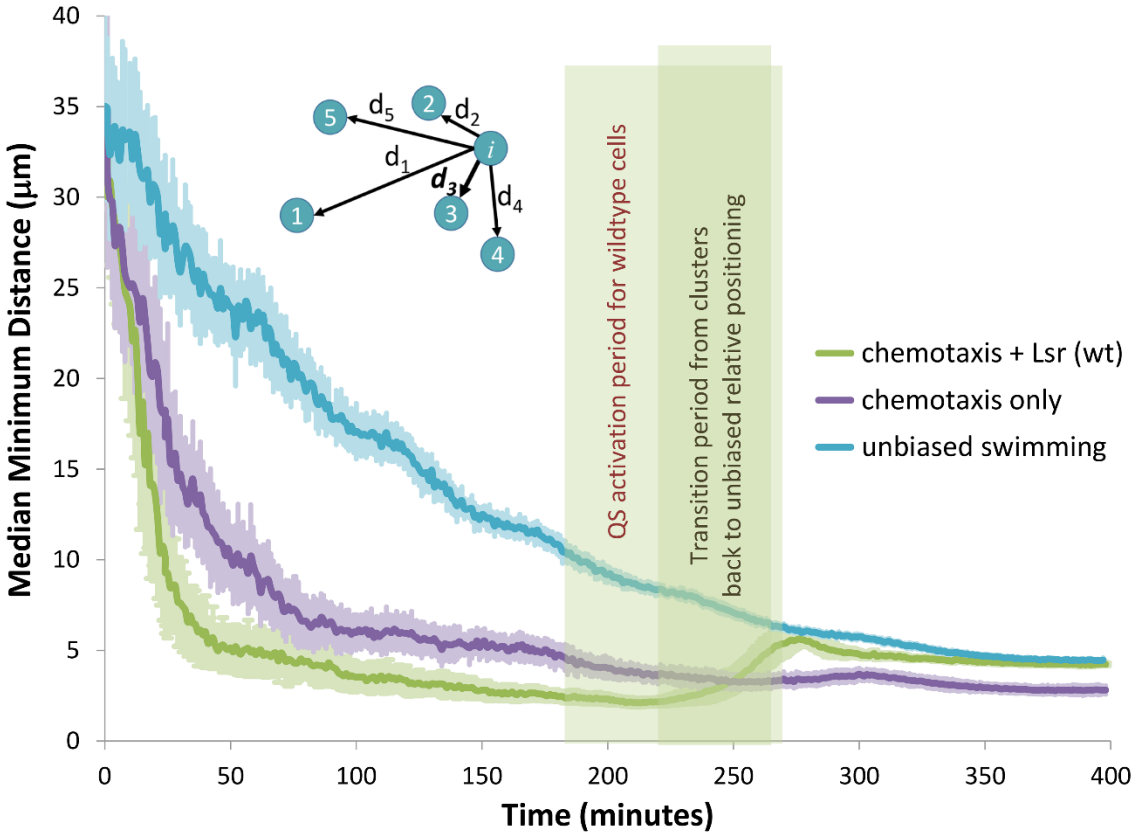


Figure 3-S9. Clustering of cells with lsr activity and AI-2 chemoattraction as measured by cell-cell distance. The median minimal distance between cells for (i) a population with both Lsr activity and AI-2 chemoattraction, (ii) a population lacking Lsr recompartmentalization ability, and (iii) another population lacking AI-2 chemoattraction but with recompartmentalization. Each condition was simulated twenty times, the average median minimal distance between cells is represented by the darker line, while the standard deviation of those values is represented by the surrounding lighter regions. The darker green box indicates the time over which the switch in median minimum distance between cells from that of an AI-2 chemoattracted population to one of a population not chemoattracted to AI-2 occurred for the wildtype population, while the lighter green region indicates the time over which Lsr induction began for the wildtype population. The bump around 250 minutes is indicative of this change. Inset is a pictorial example of the minimum distance between cells.

take in AI-2) than wildtype cells, achieving a higher extracellular concentration of AI-2 over a longer distance. This may account for the slower or looser clustering of such cells compared to the wildtype phenotype. For wildtype AI-2 chemotaxing cells (Lsr^+), the median minimal distance between cells began increasing around 225 minutes a short while after the population began to QS activate around 180 minutes. The delay was likely a function of the time required to begin drawing down the AI-2 concentration in these clusters, while the increasing distance between cells reflected the dispersion phenomena, which is apparent by 300 minutes. We note that the dispersion phenomena here is underrepresented locally because the calculated value of the intercell distance is averaged over the entire population, including all clusters and dispersed cells. In the end, the distance between wildtype cells matched the distance between cells lacking AI-2 chemoattraction, indicating that the clusters had fully dispersed. While this is an exciting outcome that could have broad ramifications, we know of no directly aligned observations.

3.7.3.9 Motility mode feedback onto population activation as a function of cell-cell distance. In general, we found that among the populations simulated here, non-taxis swimming populations were the slowest to QS activate. We note, however, that these same cells ultimately achieved the largest proportion of stably QS activated cells. This inverse correlation persisted across motility types (**Figure 3-S10A**). Cells in simulations of growing colonies activated the fastest but also experienced the smallest fraction of stably activated cells, whereas chemotaxing populations experienced intermediate levels of both speed to activation and of the final proportion activated. In other words, as a generalization, the higher the cell density the earlier the activation. However, higher

density also appeared to produce stronger negative extracellular feedback from QS activation as a smaller fraction of the population was ultimately activated. (**Figure 3-S10B**)

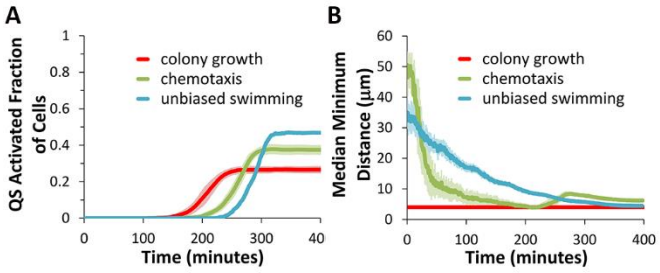


Figure 3-S10. Measures of the difference between different modes of motility when coupled with Lsr/AI-2 dynamics. A The fraction of the population that was QS activated over time, in simulations of different motility, with average values (n = 20) set in a darker thinner line and a lighter surrounding shade representing the standard deviation. **B** The median minimum distance between cells for populations influenced by different combinations of motility and AI-2 uptake. Dark lines are an average value (n = 20), while the surrounding lighter shades reflect the corresponding standard deviation. For example, cells undergoing colony growth had a predefined, regular distance between them, thus a single value prevailed across the entire time course and variability was zero.

Chapter 4: Conclusions

Living within consortia is likely the primary mode of existence for bacteria. Therein, they must frequently negotiate numerous interactions with other species. While AI-2 is likely to be only one molecule within a vast sea by which such interactions are mediated, due to the prevalence of LuxS [46] and YdgG [59] homologs, it is one that is likely to be found in a plethora of ecological contexts. This prevalence alone makes it a likely candidate for QS and QS-like operations, the actual downstream functions of which are expected to be manifold. In an ecological context, QS and QS-like operations are likely to mediate the most types of cell-cell interactions, consisting of cooperation among and between species, coercion of one species's action by another, and as cues from one bacteria to another without the intent of cooperation. Representing coercion, signal blind mutants within an otherwise cooperating community cause signal responding bacteria to QS. When the QS response involves the production of public secreted goods, signal blind mutants benefit without the metabolic expenditure and are coercing their signal responding counterparts into, at the very least, premature QS. Alternatively, purely as an indicator of cell density, autoinducers could potentially serve as a cue.

The widespread phylogenetic signal of the Lsr system as described in the second chapter suggests that the Lsr system plays a functional role in numerous bacteria. The exact nature of that role is possibly varied from species to species, especially considering that it is unlikely that the affinity of LsrR for the intergenic region coevolved in an exact manner. As the Lsr system is known to affect biofilm development, it is believed to influence cooperative behaviors at least to this extent [60]. In the third chapter, mathematically modeling the form native to *E. coli*, we lend further credence to the idea

that the Lsr system leads bacteria to serve a dual role, as both a cooperator and coercer as a result of bimodal expression arising from population heterogeneity.

We also demonstrated through modeling, that bimodality results in multiple emergent phenomenon depending on the mode of motility it is paired with, which itself fed back onto the bimodal phenotype *in silico*. As multiple homologs of the Lsr system exist, it is highly likely that three categories of Lsr homolog exist relative to *E. coli*'s system: more sensitive, equally sensitive, and less sensitive to AI-2. In terms of speed to activation these might roughly translate to: faster, equal, and slower to activate respectively. Here, we interrogated how bacteria containing such homologous systems interact with a population possessing a base *E. coli* Lsr system, and suggest *lsrFG* mutants, double *lsrFG luxS* mutants, and *luxS* mutants as representatives of each respective category. We showed how these competing populations could either largely inhibit, act in concert with, or cause wholesale activation of the wildtype population, respectively. For a *luxS* mutant population, the cost of coercion is the loss of the activated methyl cycle, indicating that at least for this case there is a built in disadvantage to free ridership.

Placing the Lsr system within the larger context of other QS architectures, our modeling strongly indicates that consequent to bimodal activation, the patterns of expression arising from the Lsr system are in stark contrast to those associated with LuxIR QS. Moreover, if placed in the same environment as the AI-2 activated TCRS, LuxPQ, our studies indicate that Lsr activation would likely curtail LuxPQ based signaling coordination, at least up to a point. Indeed, in isolation the Lsr system operates more closely to sugar importation systems than other QS systems. Clearly, the

distinction between sugar systems and the Lsr system is the context of self-production and self-secretion. Based on previous examination of QS [117,118], we believe that this secretion is likely to result in greater coordination compared to a population with the same average rate of AI2-P accretion but without YdgG or other means of AI-2 export. Aside from this consideration, our homology search indicates that the Lsr system phylogenetic signal is much less monophyletic than that for the *lac* system, even if it is more widespread.

References

1. Greenberg, E., Hastings, J. & Ulitzur, S. Induction of luciferase synthesis in *Beneckea harveyi* by other marine bacteria. *Arch. Microbiol.* **120,** 87–91 (1979).
2. Eberhard, A. Inhibition and activation of bacterial luciferase synthesis. *J. Bacteriol.* **109,** 1101–1105 (1972).
3. Fuqua, W. C., Winans, S. C. & Greenberg, E. P. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J. Bacteriol.* **176,** 269 (1994).
4. Hagen, S. J., Son, M., Weiss, J. T. & Young, J. H. Bacterium in a box: sensing of quorum and environment by the LuxI/LuxR gene regulatory circuit. *J. Biol. Phys.* **36,** 317–327 (2010).
5. Carnes, E. C. *et al.* Confinement-induced quorum sensing of individual *Staphylococcus aureus* bacteria. *Nat. Chem. Biol.* **6,** 41–45 (2009).
6. Boedicker, J. Q., Vincent, M. E. & Ismagilov, R. F. Microfluidic Confinement of Single Cells of Bacteria in Small Volumes Initiates High-Density Behavior of Quorum Sensing and Growth and Reveals Its Variability. *Angew. Chem. Int. Ed.* **48,** 5908–5911 (2009).
7. Alberghini, S. *et al.* Consequences of relative cellular positioning on quorum sensing and bacterial cell-to-cell communication. *FEMS Microbiol. Lett.* **292,** 149–161 (2009).
8. Hense, B. A. *et al.* Does efficiency sensing unify diffusion and quorum sensing? *Nat. Rev. Microbiol.* **5,** 230–239 (2007).
9. Redfield, R. J. Is quorum sensing a side effect of diffusion sensing? *Trends Microbiol.* **10,** 365–370 (2002).
10. Diggle, S. P., Gardner, A., West, S. A. & Griffin, A. S. Evolutionary theory of bacterial quorum sensing: when is a signal not a signal? *Philos. Trans. R. Soc. B Biol. Sci.* **362,** 1241–1249 (2007).
11. Mascher, T., Helmann, J. D. & Unden, G. Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol. Mol. Biol. Rev.* **70,** 910–938 (2006).
12. Case, R. J., Labbate, M. & Kjelleberg, S. AHL-driven quorum-sensing circuits: their frequency and function among the Proteobacteria. *ISME J.* **2,** 345 (2008).
13. Pereira, C. S., de Regt, A. K., Brito, P. H., Miller, S. T. & Xavier, K. B. Identification of functional LsrB-like autoinducer-2 receptors. *J. Bacteriol.* **191,** 6975–6987 (2009).
14. Quan, D. N. & Bentley, W. E. Gene network homology in prokaryotes using a similarity search approach: Queries of quorum sensing signal transduction. *PLoS Comput. Biol.* **8,** e1002637 (2012).
15. Bassler, B. L. Small talk: cell-to-cell communication in bacteria. *Cell* **109,** 421–424 (2002).
16. Anetzberger, C. *et al.* Autoinducers act as biological timers in Vibrio harveyi. *PloS One* **7,** e48310 (2012).
17. Tu, K. C., Long, T., Svenningsen, S. L., Wingreen, N. S. & Bassler, B. L. Negative Feedback Loops Involving Small Regulatory RNAs Precisely Control the *Vibrio harveyi* Quorum-Sensing Response. *Mol. Cell* **37,** 567–579 (2010).
18. Tu, K. C., Waters, C. M., Svenningsen, S. L. & Bassler, B. L. A small-RNA-mediated negative feedback loop controls quorum-sensing dynamics in Vibrio harveyi. *Mol. Microbiol.* **70,** 896–907 (2008).
19. Winson, M. K. *et al.* Construction and analysis of luxCDABE-based plasmid sensors for investigating N-acyl homoserine lactone-mediated quorum sensing. *FEMS Microbiol. Lett.* **163,** 185–192 (1998).
20. Swift, S. *et al.* Quorum sensing in *Aeromonas hydrophila* and *Aeromonas salmonicida*: identification of the LuxRI homologs AhyRI and AsaRI and their cognate N-acylhomoserine lactone signal molecules. *J. Bacteriol.* **179,** 5271–5281 (1997).

21. Ng, W.-L. & Bassler, B. L. Bacterial quorum-sensing network architectures. *Annu. Rev. Genet.* **43,** 197–222 (2009).
22. Waters, C. M. & Bassler, B. L. Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* **21,** 319–346 (2005).
23. Cao, J.-G. & Meighen, E. Purification and structural identification of an autoinducer for the luminescence system of Vibrio harveyi. *J. Biol. Chem.* **264,** 21670–21676 (1989).
24. Milton, D. L. *et al.* The LuxM Homologue VanM from *Vibrio anguillarum* directs the Synthesis of N-(3-Hydroxyhexanoyl) homoserine Lactone and N-Hexanoylhomoserine Lactone. *J. Bacteriol.* **183,** 3537–3547 (2001).
25. Hanzelka, B. L. *et al.* Acylhomoserine lactone synthase activity of the *Vibrio fischeri* AinS protein. *J. Bacteriol.* **181,** 5766–5770 (1999).
26. Makino, K. *et al.* Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of V cholerae. *The Lancet* **361,** 743–749 (2003).
27. Sun, J., Daniel, R., Wagner-Döbler, I. & Zeng, A.-P. 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 (2004).
28. Schauder, S., Shokat, K., Surette, M. G. & Bassler, B. L. The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. *Mol. Microbiol.* **41,** 463–476 (2001).
29. Chen, X. *et al.* Structural identification of a bacterial quorum-sensing signal containing boron. *Nature* **415,** 545–549 (2002).
30. Tait, K., Hutchison, Z., Thompson, F. L. & Munn, C. B. Quorum sensing signal production and inhibition by coral-associated *Vibrios*. *Environ. Microbiol. Rep.* **2,** 145–150 (2010).
31. Ng, W. *et al.* Signal production and detection specificity in Vibrio CqsA/CqsS quorum-sensing systems. *Mol. Microbiol.* **79,** 1407–1417 (2011).
32. Gray, K. M. & Garey, J. R. The evolution of bacterial LuxI and LuxR quorum sensing regulators. *Microbiology* **147,** 2379–2387 (2001).
33. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463,** 326–330 (2010).
34. Bulter, T. *et al.* Design of artificial cell–cell communication using gene and metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2299–2304 (2004).
35. Dai, Y., Toley, B. J., Swofford, C. A. & Forbes, N. S. Construction of an inducible cell-communication system that amplifies *Salmonella* gene expression in tumor tissue. *Biotechnol. Bioeng.* **110,** 1769–1781 (2013).
36. Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical/wires/'. *Nature* **469,** 212–215 (2011).
37. Saeidi, N. *et al.* Engineering microbes to sense and eradicate *Pseudomonas aeruginosa*, a human pathogen. *Mol. Syst. Biol.* **7,** (2011).
38. Danino, V. E., Wilkinson, A., Edwards, A. & Downie, J. A. Recipient-induced transfer of the symbiotic plasmid pRL1JI in *Rhizobium leguminosarum* bv. viciae is regulated by a quorum-sensing relay. *Mol. Microbiol.* **50,** 511–525 (2003).
39. Ahmer, B. M., van Reeuwijk, J., Timmers, C. D., Valentine, P. J. & Heffron, F. *Salmonella* typhimurium encodes an SdiA homolog, a putative quorum sensor of the LuxR family, that regulates genes on the virulence plasmid. *J. Bacteriol.* **180,** 1185–1193 (1998).
40. Yao, Y. *et al.* Structure of the *Escherichia coli* Quorum Sensing Protein SdiA: Activation of the Folding Switch by Acyl Homoserine Lactones. *J. Mol. Biol.* **355,** 262–273 (2006).
41. Houdt, R., Aertsen, A., Moons, P., Vanoirbeek, K. & Michiels, C. W. N-acyl-l-homoserine lactone signal interception by *Escherichia coli*. *FEMS Microbiol. Lett.* **256,** 83–89 (2006).
42. Michael, B., Smith, J. N., Swift, S., Heffron, F. & Ahmer, B. M. SdiA of *Salmonella enterica* is a LuxR homolog that detects mixed microbial communities. *J. Bacteriol.* **183,** 5733–5742 (2001).

43. Smith, J. N. & Ahmer, B. M. Detection of other microbial species by *Salmonella*: expression of the SdiA regulon. *J. Bacteriol.* **185,** 1357–1366 (2003).
44. Sharma, V. K., Bearson, S. M. & Bearson, B. L. Evaluation of the effects of sdiA, a luxR homologue, on adherence and motility of *Escherichia coli* O157: H7. *Microbiology* **156,** 1303–1312 (2010).
45. Dyszel, J. L. *et al. E. coli* K-12 and EHEC genes regulated by SdiA. *PLoS One* **5,** e8946 (2010).
46. Winzer, K. *et al.* LuxS: its role in central metabolism and the in vitro synthesis of 4-hydroxy-5-methyl-3 (2H)-furanone. *Microbiology* **148,** 909–922 (2002).
47. Bodor, A., Elxnat, B., Thiel, V., Schulz, S. & Wagner-Döbler, I. Potential for luxS related signalling in marine bacteria and production of autoinducer-2 in the genus *Shewanella*. *BMC Microbiol.* **8,** 13 (2008).
48. Tavender, T. J., Halliday, N. M., Hardie, K. R. & Winzer, K. LuxS-independent formation of AI-2 from ribulose-5-phosphate. *BMC Microbiol.* **8,** 98 (2008).
49. Li, J. *et al.* A stochastic model of *Escherichia coli* AI-2 quorum signal circuit reveals alternative synthesis pathways. *Mol. Syst. Biol.* **2,** (2006).
50. Meijler, M. M. *et al.* Synthesis and biological validation of a ubiquitous quorum-sensing molecule. *Angew. Chem. Int. Ed.* **43,** 2106–2108 (2004).
51. Vendeville, A., Winzer, K., Heurlier, K., Tang, C. M. & Hardie, K. R. Making'sense'of metabolism: autoinducer-2, LuxS and pathogenic bacteria. *Nat. Rev. Microbiol.* **3,** 383–396 (2005).
52. Daubin, V., Gouy, M. & Perriere, G. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res.* **12,** 1080–1090 (2002).
53. Taga, M. E., Semmelhack, J. L. & Bassler, B. L. 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 (2001).
54. Taga, M. E., Miller, S. T. & Bassler, B. L. Lsr-mediated transport and processing of AI-2 in *Salmonella* typhimurium. *Mol. Microbiol.* **50,** 1411–1427 (2003).
55. Wang, L., Hashimoto, Y., Tsao, C.-Y., Valdes, J. J. & Bentley, W. E. Cyclic AMP (cAMP) and cAMP receptor protein influence both synthesis and uptake of extracellular autoinducer 2 in *Escherichia coli*. *J. Bacteriol.* **187,** 2066–2076 (2005).
56. Wang, L., Li, J., March, J. C., Valdes, J. J. & Bentley, W. E. luxS-dependent gene regulation in *Escherichia coli* K-12 revealed by genomic expression profiling. *J. Bacteriol.* **187,** 8350–8360 (2005).
57. Xavier, K. B. & Bassler, B. L. Regulation of uptake and processing of the quorum-sensing autoinducer AI-2 in *Escherichia coli*. *J. Bacteriol.* **187,** 238–248 (2005).
58. Herzberg, M., Kaye, I. K., Peti, W. & Wood, T. K. YdgG (TqsA) controls biofilm formation in *Escherichia coli* K-12 through autoinducer 2 transport. *J. Bacteriol.* **188,** 587–598 (2006).
59. Rettner, R. E. & Saier Jr, M. H. The autoinducer-2 exporter superfamily. *J. Mol. Microbiol. Biotechnol.* **18,** 195–205 (2010).
60. Li, J. *et al.* Quorum sensing in *Escherichia coli* is signaled by AI-2/LsrR: effects on small RNA and biofilm architecture. *J. Bacteriol.* **189,** 6011–6020 (2007).
61. Shao, H., James, D., Lamont, R. J. & Demuth, D. R. Differential interaction of *Aggregatibacter* (*Actinobacillus*) *actinomycetemcomitans* LsrB and RbsB proteins with autoinducer 2. *J. Bacteriol.* **189,** 5559–5565 (2007).
62. Rader, B. A. *et al. Helicobacter pylori* perceives the quorum-sensing molecule AI-2 as a chemorepellent via the chemoreceptor TlpB. *Microbiology* **157,** 2445–2455 (2011).
63. Pereira, C. S. *et al.* Phosphoenolpyruvate phosphotransferase system regulates detection and processing of the quorum sensing signal autoinducer-2. *Mol. Microbiol.* **84,** 93–104 (2012).

64. Thompson, J. Galactose transport systems in *Streptococcus lactis*. *J. Bacteriol.* **144,** 683–691 (1980).
65. Henderson, P. J. Proton-linked sugar transport systems in bacteria. *J. Bioenerg. Biomembr.* **22,** 525–569 (1990).
66. Chaudhuri, B. N., Ko, J., Park, C., Jones, T. A. & Mowbray, S. L. Structure of d-allose binding protein from *Escherichia coli* bound to d-allose at 1.8 Å resolution. *J. Mol. Biol.* **286,** 1519–1531 (1999).
67. Shao, H., Lamont, R. J. & Demuth, D. R. Autoinducer 2 is required for biofilm growth of *Aggregatibacter* (*Actinobacillus*) *actinomycetemcomitans*. *Infect. Immun.* **75,** 4211–4218 (2007).
68. Armbruster, C. E. *et al.* RbsB (NTHI_0632) mediates quorum signal uptake in nontypeable Haemophilus influenzae strain 86-028NP. *Mol. Microbiol.* **82,** 836–850 (2011).
69. Thijs, I. M. *et al.* The AI-2-dependent regulator LsrR has a limited regulon in *Salmonella* Typhimurium. *Cell Res.* **20,** 966–969 (2010).
70. Byrd, C. M. Local and global gene regulation analysis of the autoinducer-2 mediated quorum sensing mechanism in *Escherichia coli*. (2011).
71. Pesavento, C. *et al.* Inverse regulatory coordination of motility and curli-mediated adhesion in *Escherichia coli*. *Genes Dev.* **22,** 2434–2446 (2008).
72. Park, J. T., Raychaudhuri, D., Li, H., Normark, S. & Mengin-Lecreulx, D. MppA, a Periplasmic Binding Protein Essential for Import of the Bacterial Cell Wall Peptidel-Alanyl-γ-d-Glutamyl-meso-Diaminopimelate. *J. Bacteriol.* **180,** 1215–1223 (1998).
73. Diaz, Z., Xavier, K. B. & Miller, S. T. The crystal structure of the *Escherichia coli* autoinducer-2 processing protein LsrF. *PloS One* **4,** e6820 (2009).
74. Marques, J. C. *et al.* Processing the Interspecies Quorum-sensing Signal Autoinducer-2 (AI-2) Characterization phospho-(S)-4, 5-dihydroxy-2, 3-pentanedione isomerization by LsrG protein. *J. Biol. Chem.* **286,** 18331–18343 (2011).
75. Xavier, K. B. *et al.* Phosphorylation and processing of the quorum-sensing molecule autoinducer-2 in enteric bacteria. *ACS Chem. Biol.* **2,** 128–136 (2007).
76. Pellicer, M. T., Nuñez, M. F., Aguilar, J., Badia, J. & Baldoma, L. Role of 2-phosphoglycolate phosphatase of *Escherichia coli* in metabolism of the 2-phosphoglycolate formed in DNA repair. *J. Bacteriol.* **185,** 5815–5821 (2003).
77. Choi, J. *et al.* LsrR-mediated quorum sensing controls invasiveness of *Salmonella* typhimurium by regulating SPI-1 and flagella genes. *PloS One* **7,** e37059 (2012).
78. Niu, C. *et al.* LuxS influences *Escherichia coli* biofilm formation through autoinducer-2-dependent and autoinducer-2-independent modalities. *FEMS Microbiol. Ecol.* **83,** 778–791 (2013).
79. Torres-Escobar, A., Juárez-Rodríguez, M. D., Lamont, R. J. & Demuth, D. R. Transcriptional regulation of *Aggregatibacter actinomycetemcomitans* lsrACDBFG and lsrRK operons and their role in biofilm formation. *J. Bacteriol.* **195,** 56–65 (2013).
80. Darch, S. E., West, S. A., Winzer, K. & Diggle, S. P. Density-dependent fitness benefits in quorum-sensing bacterial populations. *Proc. Natl. Acad. Sci.* **109,** 8259–8263 (2012).
81. Krin, E. *et al.* Pleiotropic role of quorum-sensing autoinducer 2 in *Photorhabdus luminescens*. *Appl. Environ. Microbiol.* **72,** 6439–6451 (2006).
82. Auger, S., Krin, E., Aymerich, S. & Gohar, M. Autoinducer 2 affects biofilm formation by *Bacillus cereus*. *Appl. Environ. Microbiol.* **72,** 937–941 (2006).
83. Li, L. *et al.* Analysis on *Actinobacillus pleuropneumoniae* LuxS regulated genes reveals pleiotropic roles of LuxS/AI-2 on biofilm formation, adhesion ability and iron metabolism. *Microb. Pathog.* **50,** 293–302 (2011).
84. Vidal, J. E., Ludewick, H. P., Kunkel, R. M., Zähner, D. & Klugman, K. P. The LuxS-dependent quorum-sensing system regulates early biofilm formation by *Streptococcus pneumoniae* strain D39. *Infect. Immun.* **79,** 4050–4060 (2011).

85. Li, M., Villaruz, A. E., Vadyvaloo, V., Sturdevant, D. E. & Otto, M. AI-2-dependent gene regulation in *Staphylococcus epidermidis*. *BMC Microbiol.* **8,** 4 (2008).
86. Ahmed, N. A., Petersen, F. C. & Scheie, A. A. AI-2/LuxS is involved in increased biofilm formation by *Streptococcus intermedius* in the presence of antibiotics. *Antimicrob. Agents Chemother.* **53,** 4258–4263 (2009).
87. Von Lackum, K. *et al.* Functionality of *Borrelia burgdorferi* LuxS: The Lyme disease spirochete produces and responds to the pheromone autoinducer-2 and lacks a complete activated-methyl cycle. *Int. J. Med. Microbiol.* **296,** 92–102 (2006).
88. Rader, B. A., Campagna, S. R., Semmelhack, M. F., Bassler, B. L. & Guillemin, K. The quorum-sensing molecule autoinducer 2 regulates motility and flagellar morphogenesis in *Helicobacter pylori*. *J. Bacteriol.* **189,** 6109–6117 (2007).
89. Pereira, C. S., McAuley, J. R., Taga, M. E., Xavier, K. B. & Miller, S. T. *Sinorhizobium meliloti*, a bacterium lacking the autoinducer-2 (AI-2) synthase, responds to AI-2 supplied by other bacteria. *Mol. Microbiol.* **70,** 1223–1235 (2008).
90. Duan, K., Dammel, C., Stein, J., Rabin, H. & Surette, M. G. Modulation of *Pseudomonas aeruginosa* gene expression by host microflora through interspecies communication. *Mol. Microbiol.* **50,** 1477–1491 (2003).
91. Geier, H., Mostowy, S., Cangelosi, G. A., Behr, M. A. & Ford, T. E. Autoinducer-2 triggers the oxidative stress response in *Mycobacterium avium*, leading to biofilm formation. *Appl. Environ. Microbiol.* **74,** 1798–1804 (2008).
92. Strassmann, J. E., Gilbert, O. M. & Queller, D. C. Kin discrimination and cooperation in microbes. *Annu. Rev. Microbiol.* **65,** 349–367 (2011).
93. Wu, K. & Rao, C. V. The role of configuration and coupling in autoregulatory gene circuits. *Mol. Microbiol.* **75,** 513–527 (2010).
94. Tsao, C.-Y., Hooshangi, S., Wu, H.-C., Valdes, J. J. & Bentley, W. E. Autonomous induction of recombinant proteins by minimally rewiring native quorum sensing regulon of *E. coli*. *Metab. Eng.* **12,** 291–297 (2010).
95. Gonzalez Barrios, A. F. & Achenie, L. E. *Escherichia coli* autoinducer-2 uptake network does not display hysteretic behavior but AI-2 synthesis rate controls transient bifurcation. *Biosystems* **99,** 17–26 (2010).
96. Englert, D. L., Manson, M. D. & Jayaraman, A. Flow-based microfluidic device for quantifying bacterial chemotaxis in stable, competing gradients. *Appl. Environ. Microbiol.* **75,** 4557–4564 (2009).
97. Hegde, M. *et al.* Chemotaxis to the quorum-sensing signal AI-2 requires the Tsr chemoreceptor and the periplasmic LsrB AI-2-binding protein. *J. Bacteriol.* **193,** 768–773 (2011).
98. Zahedmanesh, H. & Lally, C. A multiscale mechanobiological modelling framework using agent-based models and finite element analysis: application to vascular tissue engineering. *Biomech. Model. Mechanobiol.* **11,** 363–377 (2012).
99. Suyama, M. & Bork, P. Evolution of prokaryotic gene order: genome rearrangements in closely related species. *Trends Genet.* **17,** 10–13 (2001).
100. Moreno-Hagelsieb, G. & Collado-Vides, J. A powerful non-homology method for the prediction of operons in prokaryotes. *Bioinformatics* **18,** S329–S336 (2002).
101. Overbeek, R., Fonstein, M., D'souza, M., Pusch, G. D. & Maltsev, N. The use of gene clusters to infer functional coupling. *Proc. Natl. Acad. Sci.* **96,** 2896–2901 (1999).
102. Touchon, M. *et al.* Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* **5,** e1000344 (2009).
103. Fang, G., Rocha, E. P. & Danchin, A. Persistence drives gene clustering in bacterial genomes. *BMC Genomics* **9,** 4 (2008).
104. Harrington, E. D., Jensen, L. J. & Bork, P. Predicting biological networks from genomic data. *FEBS Lett.* **582,** 1251–1258 (2008).

105. Rogozin, I. B., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Computational approaches for the analysis of gene neighbourhoods in prokaryotic genomes. *Brief. Bioinform.* **5,** 131–149 (2004).
106. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
107. Stajich, J. E. *et al.* The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* **12,** 1611–1618 (2002).
108. Williams, K. P. *et al.* Phylogeny of gammaproteobacteria. *J. Bacteriol.* **192,** 2305–2314 (2010).
109. Roderick, S. L. The *lac* operon galactoside acetyltransferase. *C. R. Biol.* **328,** 568–575 (2005).
110. Arraj, J. A. & Campbell, J. H. Isolation and characterization of the newly evolved ebg beta-galactosidase of *Escherichia coli* K-12. *J. Bacteriol.* **124,** 849–856 (1975).
111. Stoebel, D. M. Lack of evidence for horizontal transfer of the lac operon into *Escherichia coli*. *Mol. Biol. Evol.* **22,** 683–690 (2005).
112. Hooshangi, S. & Bentley, W. E. LsrR quorum sensing 'switch' is revealed by a bottom-up approach. *PLoS Comput. Biol.* **7,** e1002172 (2011).
113. Studier, F. W., Daegelen, P., Lenski, R. E., Maslov, S. & Kim, J. F. Understanding the Differences between Genome Sequences of *Escherichia coli* B Strains REL606 and BL21 (DE3) and Comparison of the *E. coli* B and K-12 Genomes. *J. Mol. Biol.* **394,** 653–680 (2009).
114. Xue, T., Zhao, L., Sun, H., Zhou, X. & Sun, B. LsrR-binding site recognition and regulatory characteristics in *Escherichia coli* AI-2 quorum sensing. *Cell Res.* **19,** 1258–1268 (2009).
115. Price, M. N., Dehal, P. S. & Arkin, A. P. Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol* **9,** R4 (2008).
116. Mavromatis, K. *et al.* Complete genome sequence of *Spirochaeta smaragdinae* type strain (SEBR 4228T). *Stand. Genomic Sci.* **3,** 136 (2010).
117. Tanouchi, Y., Tu, D., Kim, J. & You, L. Noise reduction by diffusional dissipation in a minimal quorum sensing motif. *PLoS Comput. Biol.* **4,** e1000167 (2008).
118. Tabareau, N., Slotine, J.-J. & Pham, Q.-C. How synchronization protects from noise. *PLoS Comput. Biol.* **6,** e1000637 (2010).
119. Anetzberger, C., Pirch, T. & Jung, K. Heterogeneity in quorum sensing-regulated bioluminescence of Vibrio harveyi. *Mol. Microbiol.* **73,** 267–277 (2009).
120. Sayut, D. J., Kambam, P. K. R. & Sun, L. Noise and kinetics of LuxR positive feedback loops. *Biochem. Biophys. Res. Commun.* **363,** 667–673 (2007).
121. Kittisopikul, M. & Süel, G. M. Biological role of noise encoded in a genetic network motif. *Proc. Natl. Acad. Sci.* **107,** 13300–13305 (2010).
122. Fenley, A. T., Banik, S. K. & Kulkarni, R. V. Computational modeling of differences in the quorum sensing induced luminescence phenotypes of *Vibrio harveyi* and *Vibrio cholerae*. *J. Theor. Biol.* **274,** 145–153 (2011).
123. Marques, J. C. *et al.* LsrF, a coenzyme A-dependent thiolase, catalyzes the terminal step in processing the quorum sensing signal autoinducer-2. *Proc. Natl. Acad. Sci.* **111,** 14235–14240 (2014).
124. Parsek, M. R. & Greenberg, E. Sociomicrobiology: the connections between quorum sensing and biofilms. *Trends Microbiol.* **13,** 27–33 (2005).
125. Defoirdt, T. Can bacteria actively search to join groups? *ISME J.-Int. Soc. Microb. Ecol.* **5,** 569 (2011).
126. Sourjik, V. & Armitage, J. P. Spatial organization in bacterial chemotaxis. *EMBO J.* **29,** 2724–2733 (2010).
127. Dockery, J. D. & Keener, J. P. A mathematical model for quorum sensing in *Pseudomonas aeruginosa*. *Bull. Math. Biol.* **63,** 95–116 (2001).

128. Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H. & Kompala, D. S. Plasmid-encoded protein: the principal factor in the ‘metabolic burden’ associated with recombinant bacteria. *Biotechnol. Bioeng.* **35,** 668–681 (1990).
129. Melke, P., Sahlin, P., Levchenko, A. & Jönsson, H. A cell-based model for quorum sensing in heterogeneous bacterial colonies. *PLoS Comput. Biol.* **6,** e1000819 (2010).
130. Mittal, N., Budrene, E. O., Brenner, M. P. & van Oudenaarden, A. Motility of *Escherichia coli* cells in clusters formed by chemotactic aggregation. *Proc. Natl. Acad. Sci.* **100,** 13259–13263 (2003).
131. Wu, H. *et al.* Autonomous bacterial localization and gene expression based on nearby cell receptor density. *Mol. Syst. Biol.* **9,** (2013).
132. Albert, R. & Othmer, H. G. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* **223,** 1–18 (2003).
133. Lawrence, P. A., Casal, J. & Struhl, G. hedgehog and engrailed: pattern formation and polarity in the Drosophila abdomen. *Development* **126,** 2431–2439 (1999).
134. Birtwistle, M. R. *et al.* Emergence of bimodal cell population responses from the interplay between analog single-cell signaling and protein expression noise. *BMC Syst. Biol.* **6,** 109 (2012).
135. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* (2013).
136. Veening, J.-W., Smits, W. K. & Kuipers, O. P. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* **62,** 193–210 (2008).
137. De Beer, D. & Stoodley, P. Relation between the structure of an aerobic biofilm and transport phenomena. *Water Sci. Technol.* **32,** 11–18 (1995).
138. Coulthurst, S. J., Kurz, C. L. & Salmond, G. P. luxS mutants of *Serratia* defective in autoinducer-2-dependent ‘quorum sensing’show strain-dependent impacts on virulence and production of carbapenem and prodigiosin. *Microbiology* **150,** 1901–1910 (2004).
139. Høyland-Kroghsbo, N. M., Mærkedahl, R. B. & Svenningsen, S. L. A quorum-sensing-induced bacteriophage defense mechanism. *MBio* **4,** e00362–12 (2013).
140. Sun, S., Kjelleberg, S. & McDougald, D. Relative contributions of *Vibrio* polysaccharide and quorum sensing to the resistance of Vibrio cholerae to predation by heterotrophic protists. *PloS One* **8,** e56338 (2013).
141. Diggle, S. P., Griffin, A. S., Campbell, G. S. & West, S. A. Cooperation and conflict in quorum-sensing bacterial populations. *Nature* **450,** 411–414 (2007).
142. Dandekar, A. A., Chugani, S. & Greenberg, E. P. Bacterial quorum sensing and metabolic incentives to cooperate. *Science* **338,** 264–266 (2012).
143. Byrd, C. M. & Bentley, W. E. Quieting cross talk - the quorum sensing regulator LsrR as a possible target for fighting bacterial infections. *Cell Res* **19,** 1229–1230
144. Luo, X. *et al.* Biofabrication of stratified biofilm mimics for observation and control of bacterial signaling. *Biomaterials* **33,** 5136–5143 (2012).
145. Goryachev, A. B. Understanding Bacterial Cell− Cell Communication with Computational Modeling. *Chem Rev* **111,** 238–250 (2011).
146. Wilke, C. & Chang, P. Correlation of diffusion coefficients in dilute solutions. *AIChE J.* **1,** 264–270 (1955).
147. Terrell, J. L. *et al.* Integrated biofabrication for electro-addressed in-film bioprocessing. *Biotechnol. J.* **7,** 428–439 (2012).
148. Becskei, A., Séraphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20,** 2528–2535 (2001).
149. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & Van Oudenaarden, A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **427,** 737–740 (2004).

150. Zhu, J. & Pei, D. A LuxP-based fluorescent sensor for bacterial autoinducer II. *ACS Chem. Biol.* **3,** 110–119 (2008).
151. Wu, M., Tao, Y., Liu, X. & Zang, J. Structural basis for phosphorylated autoinducer-2 modulation of the oligomerization state of the global transcription regulator LsrR from *Escherichia coli*. *J. Biol. Chem.* **288,** 15878–15887 (2013).
152. Ganguly, A., Rajdev, P., Williams, S. M. & Chatterji, D. Nonspecific Interaction between DNA and Protein allows for Cooperativity: A Case Study with Mycobacterium DNA Binding Protein. *J. Phys. Chem. B* **116,** 621–632 (2011).
153. Von Dassow, G., Meir, E., Munro, E. M. & Odell, G. M. The segment polarity network is a robust developmental module. *Nature* **406,** 188–192 (2000).
154. Matsumoto, M. & Nishimura, T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul. TOMACS* **8,** 3–30 (1998).