UNIVERSITY OF NOVA GORICA
GRADUATE SCHOOL

COMPARATIVE GENOMICS OF BACTERIAL COMMUNICATION AND COOPERATION: FUNCTIONAL ANNOTATION OF N-ACYL HOMOSERINE LACTONE QUORUM SENSING GENES IN GRAM-NEGATIVE BACTERIA

DISSERTATION

Sonal Kumari

Mentor: Prof. Sándor Pongor

Nova Gorica, 2014

Abstract

Bacterial communities are a major form of life maintained by complex networks of intra- and interspecies communication, of which quorum sensing (QS) is the best studied example. QS is based on a small set of signal production and signal sensing genes that control bacterial behavior. Inspite of the importance of the phenomenon, genes of bacterial communication such as QS are poorly annotated in public databases. This PhD thesis deals with subsystem based functional annotation of QS genes in bacterial genomes, with special regard to N-acyl homoserine lactone (AHL) signaling. I carried out a comprehensive survey of AHL QS genes in Gram negative bacteria, using a combination of Hidden Markov Models and a classification of chromosomal arrangements that we termed local topologies. It was observed that the QS genes are found in a number of specific arrangements and describe 16 types of such local topologies in gram-negative bacteria. It was found – in a comprehensive survey of the *Burkholderia* genus – that they are linked with given types of the chemical signal and in some cases, with specific overlap patterns of the QS genes. We also found that the AHL QS genes form distinct subgroups within their known orthology types, and these subgroups are characterized by a distinct gene arrangement and a specific chemical signal (AHL) type. This suggests that the QS systems appeared before the current bacterial strains diverged from each other. I applied a similar approach to the ComQXPA system of *Bacillus subtilis* and found analogous rules such as the characteristic differences in gene overlaps the appearance of orthologous subgroups. 31 new occurrences were found. *com* locus was restricted to *B. subtilis* as previously thought, but it was also observed in species of different orders such as clostridiales, suggesting this QS system is phylogenetically widespread.

Izvleček

Bakterijske skupnosti so ena glavnih oblik življenja, ki jo bakterije vzdržujejo s pomočjo kompleksnih inter- in intravrstnih komunikacijskih omrežij. Ena izmed najbolj preučenih oblik komunikacije med bakterijami je zaznavanje quoruma ali s tujko quorum sensing (QS). QS temelji na manjšem setu genov, ki vsebujejo informacijo za produkcijo signalov in za njihovo zaznavanje. Kljub pomembnosti pojava so geni bakterijske komunikacije, kot so QS geni, slabo anotirani v javnih bazah podatkov. Pričujoča doktorska disertacija se ukvarja z anotacijo funkcij na podlagi podsistemov QS genov v bakterijskih genomih, s posebnim poudarkom na N-acil homoserin laktonskem (AHL) signaliziranju. S pomočjo Hidden Markovih modelov v kombinaciji s klasifikacijo kromosomskih ureditev, ki smo jih poimenovali lokalne topologije, sem izvedla obsežno raziskavo AHL QS genov pri po Gramu negativnih bakterijah. Ugotovljeno je bilo, da lahko QS gene najdemo v številnih specifičnih ureditvah in jih pri po Gramu negativnih bakterijah uvrščamo v 16 tipov tovrstnih lokalnih topologij. Obsežna raziskava rodu *Burkholderia* je razkrila, da so specifične topologije povezane s specifičnim tipom kemičnih signalov in v nekaterih primerih tudi s specifičnim vzorcem prekrivanja QS genov. Opazili smo tudi, da lahko AHL QS geni znotraj svojih ortolognih tipov tvorijo različne podskupine, ki jih karakterizirata določena ureditev genov in specifični tip AHL signala. To kaže, da so se QS sistemi v zgodovini pojavili predno so se razvile sodobne bakterijske vrste in se oddaljile ena od druge. Podoben pristop sem uporabila tudi pri sistemu ComQXPA, ki je značilen za bakterijo *Bacillus subtilis* in tudi pri tem sistemu odkrila podobna pravila, ko so karakteristične razlike v prekrivanju genov, ki se pojavljajo v ortolognih podskupinah. S pomočjo računalniških orodij smo odkrili 31 novih pojavitev *com* lokusa za katerega se je do sedaj domnevalo, da se nahaja le v bakteriji *Bacillus subtilis*. Pojavitve tega lokusa smo našli tudi v vrstah iz različnih redov, kot pri npr. redu Clostridiales, kar kaže, da je ta QS sistem filogenetsko zelo razširjen.

List Of Publications

This thesis is based on the following publications:

1. **Choudhary, K.S.**; Dogsa, I.; Marsetic, Z.; Hudaiberdiev, S.; Vera, R.; Pongor, S.; Mandic-Mulec, I. (2014). ComQXPA quorum sensing systems may not be unique to Bacillus subtilis: A census in prokaryotic genomes. *PlosOne, 9(5), e96122*.
2. **Choudhary, K.S.**, Hudaiberdiev, S., Gelencsér, Z., Coutinho, B.G., Venturi, V., Pongor, S. (2013). The organization of the quorum sensing luxI/R family genes in *Burkholderia*. *Int. J. Mol. Sci.*, *14*(7), 13727-13747.
3. Gelencsér, Z.; **Choudhary, K.S.**; Coutinho, B.G.; Hudaiberdiev, S.; Galbáts, B.; Venturi, V.; Pongor., S. (2012). Classifying the topology of AHL-driven quorum sensing circuits in proteobacterial genomes. *Sensors. 12(5), 5432-5444.*
4. Gelencsér, Z.; Galbáts, B.; Gonzalez, J.F.; **Choudhary, K.S.**; Hudaiberdiev, S.; Venturi, V.; Pongor, S. (2012). Chromosomal arrangement of AHL driven quorum sensing circuits in *Pseudomonas*. *ISRN Microbiology. 2012, 6.*

List of Abbreviations

ORF	Open Reading Frame
HMM	Hidden Markov Model
QS	Quorum Sensing
ACP	Acyl Carrier Protein
AHL	Acyl-homoserine lactone
HTH	Helix-Turn-Helix
AI-2	Autoinducer 2
HSL	Homoserine Lactone
DSF	Diffusible Signaling Factor
BDSF	*Burkholderia* Diffusible Signaling Factor
PQS	*Pseudomonas* Quinolone Signal
BCC	*Burkholderia cepacia* Complex
PBE	Plant Beneficial and Environmental
MGV	Microbial Genome Viewer

List of Tables

Table 1: Processes controlled by QS systems **13**

Table 2: List of protein sequence sets used for building HMM recognizers **40**

Table 3: Typical topological patterns found in complete bacterial genomes **51**

Table 4: Intervening genes in Simple Topologies **52**

Table 5: Typical chromosomal arrangements of AHL-driven quorum sensing circuits in *Pseudomonas* (complete and Draft genomes) **57**

Table 6: Examples of *Pseudomonas* species with of AHL-driven quorum sensing networks **59**

Table 7: Chemical structure of N-acylated homoserine lactone (AHL) signals used by LuxR in the known species of the genus *Burkholderia* **63**

Table 8: Typical local topologies of AHL-driven quorum sensing circuits in *Burkholderia* **66**

Table 9: Correlation between the chemical structure and the local gene arrangements **67**

Table 10: Quorum sensing (QS) topologies in the *Burkholderia cepacia* complex **69**

Table 11: QS topologies in *Burkholderia pseudomallei* group **75**

Table 12: QS topologies in Plant beneficial and environmental *Burkholderia* sps. **78**

Table 13: QS topologies in Plant pathogenic *Burkholderia* sps. **80**

Table 14: List of the species having comQXPA loci **82**

List Of Figures

Figure 1: Flowchart of Phylogenetic footprinting **7**

Figure 2: Functional association between genes linked by fusion method **8**

Figure 3: Gene Context View of hprK gene in phylum Fermicutes **9**

Figure 4: Bacterial communication types **12**

Figure 5: Hypothetical scheme for hexanoyl homoserine lactone synthesis catalyzed by LuxI in *V. fischeri* **15**

Figure 6: Pictorial representation of AHL-mediated Quorum sensing mechanism. **17**

Figure 7: Phylogenetic analysis of LuxI proteins. Subfamily tree highlighted in red are the ones involved in non-canonical AHL synthesis. **19**

Figure 8: Domain structure of the LuxR proteins. The codes denote the PFAM (PF) domain identifiers. **21**

Figure 9: Classes of LuxR proteins based on their interaction with AHL molecules and multimeric properties **23**

Figure 10: Regulatory outline of N-AHL-based QS signaling. Pointed arrows indicate activation, and the hammerhead arrow indicates inhibition **24**

Figure 11: AHL signals produced by different bacteria **26**

Figure 12: 16S rRNA sequences of the recognized species of the *Burkholderia* genus showing the three major clades viz. BCC complex, *pseudomallei* group and PBE group **31**

Figure 13: Mechanism of Two-component system **33**

Figure 14: Domain structure of the com proteins. The codes denote the PFAM (PF) domain identifiers. **35**

Figure 15: Different types of QS signals used by various bacteria. **37**

Figure 16: Workflow **44**

Figure 17: Desired Result **44**

Figure 18: Examples of file formats **47**

Figure 19: Example of a comparative context map. Visualization of the gene context of the glnR transcription regulator in 11 different species. **48**

Figure 20: Workflow pipeline **49**

Figure 21: Clustering of LuxI and their negative regulators RsaL **54**

Figure 22: Gene Overlap patterns in the two topology types **60**

Figure 23: Cladogram of LuxI proteins present in *Pseudomonas* genomes **62**

Figure 24: Comparison of Gene topologies present in *Pseudomonas* and *Burkholderia* 65

Figure 25: Clustering of LuxI protein sequences and perceived signals by LuxR homologues in complete *Burkholderia* genomes 68

Figure 26: QS regulatory circuits in *B. cenocepacia* J2315 71

Figure 27: Cladogram of the orthologs of ORF BCAM1871, encoding a protein with an HMG-CoA domain in *Burkholderia* sps. 72

Figure 28: Typical arrangements of QS genes in chromosome II of the BCC group. 73

Figure 29: An example of the chromosomal arrangement of QS genes in completely sequenced genome of *B. pseudomallei* strains 76

Figure 30: Schematic representation of complex regulatory circuit in B. *pseudomallei* K92643 76

Figure 31: Regulatory circuit of BraR/I system in plant-beneficial *Burkholderia* group 77

Figure 32: Chromosomal arrangement of QS genes (braI/R) and OriC in completely sequenced members of the plant-beneficial and environmental group 79

Figure 33: Regulatory circuit of xenI2/R2 and bxeR genes in plant-beneficial and environmental group 79

Figure 34: Cladogram of ComQ proteins 85

Figure 35: com loci in *B. subtilis* type and non- *B. subtilis* type clade, overlaps are shown in color. 86

Figure 36: Unusual comQXPA-like loci. 87

Figure 37: Transmembrane domain architecture in ComP proteins. 89

Acknowledgements

First and foremost I would like to express my sincere gratitude to my supervisor Prof. Sándor Pongor for his guidance throughout my PhD, for sharing his experiences, for his patience during my thesis writing and moreover for his immense enthusiasm for research which motivated me every time to learn something new with each project. I would also like to thank my tutor, Dr. Vittorio Venturi for his guidance and for helping me with bacteriology questions throughout this project. I am grateful to ICGEB and to the University of Nova Gorica for giving me the opportunity to develop my academic career in this wonderful institute, in the beautiful city of Trieste.

I would like to thank my coauthors, especially Zsolt Gelencser (Budapest), Prof. Ines Mandic-Mulec and Ms. Ziva Marsetic (Ljubljana).

I was privileged to have amazing labmates, Attila (past member), Sanjar and Roberto. I appreciate their help in every project of mine and would like to thank them for the wonderful atmosphere we had at work.

Many thanks to all my friends in Trieste for making my stay here a memorable one with wonderful experiences and being available whenever I needed them.

I would like to thank my parents for always encouraging me for pursuing research, for always being with me in every phase of my failures and success. I can't express my thanks just in words.

Finally, I give my special thanks to my husband Nikhil for his continuous support and care. I would not have dared to come this far without your support. Thank you for everything.

Contents

Abstract ..... **i**
Izvleček ..... **iii**
List Of Publications ..... **v**
List of Abbreviations ..... **vii**
List of Tables ..... **ix**
List Of Figures ..... **xi**
Acknowledgements ..... **xiii**
1 Introduction **1**

1.1 Definition: Genome Annotation **1**

1.2 Types of Genome Annotation **2**

1.2.1 Structural Annotation of Genome 2

1.2.2 Functional annotation of Proteins 3

1.3 Bacterial Communication And Quorum Sensing: **10**

1.4 Processes controlled by Quorum Sensing **13**

1.5 Quorum Sensing signals **14**

1.6 Acyl-homoserine lactones Quorum Sensing **15**

1.7 Acyl Homoserine Lactones **17**

1.8 Non canonical AHL synthesis **18**

1.9 AHL Quorum Sensing components **19**

1.9.1 AHL synthases 19

1.9.2 LuxR proteins 20

1.9.3 Solo LuxR-type proteins 21

1.9.4 Classes of LuxR proteins 22

1.10 Quorum Sensing regulation **24**

1.11 Signal Specificity **25**

1.12 AHL-mediated QS systems in Proteobacteria **26**

1.13 AHL-mediated QS systems in *Pseudomonas* **27**

1.14 AHL-mediated QS systems in *Burkholderia* sps. **29**

1.15 Autoinducer-2 mediated QS **32**

1.16 Quorum sensing in Gram - positive bacteria: Peptide Signaling... 32
1.17 Other systems ... 36

2 Objectives and Methods ... 39

2.1 Aims and Objectives: ... 39
2.2 Methods/Approaches ... 40

2.2.1 Data collection: ... 40
2.2.2 QS proteins detection: ... 40
2.2.3 Local arrangement or topology of QS genes: ... 43
2.2.4 Chromosomal arrangement of QS genes: ... 44
2.2.5 Multiple sequence alignment and Tree display: ... 45
2.2.6 Neighborhood detection: ... 46

3 AHL-mediated QS systems in Proteobacteria ... 50

3.1 Results and Discussion ... 50

3.1.1 Local Topological types in Proteobacteria ... 50
3.1.2 Taxonomic Distribution of Patterns of QS gene ... 53

3.2 Summary ... 55

4 AHL-mediated QS systems in *Pseudomonas* ... 56

4.1 Results and Discussion ... 56

4.1.1 Local Topology types in *Pseudomonas* ... 56
4.1.2 Chemical Signals produced ... 58
4.1.3 Gene Overlaps ... 60
4.1.4 Evolution of QS genes in *Pseudomonas* ... 61

4.2 Summary ... 62

5 Arrangement of QS circuits in *Burkholderia* ... 63

5.1 Results and Discussion ... 64

5.1.1 Local Topology types in *Burkholderia* 64
5.1.2 Evolution of QS genes in *Burkholderia* 66
5.1.3 *Burkholderia cepacia* complex (BCC)..... 68
5.1.4 *Burkholderia pseudomallei* group 73
5.1.5 Plant Beneficial and Environmental (PBE) group 77
5.1.6 Other Pathogenic *Burkholderia* sps..... 80

5.2 Summary 81

6 Peptide Quorum sensing: ComQXPA in *Bacillus subtilis* 82

6.1 Results and Discussion 82

6.1.1 Local arrangement of the comQXPA genes..... 85
6.1.2 Unusual com-like loci 87
6.1.3 ComP Sequence variability 88

6.2 Summary 89

7 Conclusion and Discussion 90
8 References 94

1 Introduction

1.1 Definition: Genome Annotation

Genome annotation is the biological interpretation of the DNA sequence (Koonin EV 2003). The term genome annotation describes two distinct processes: structural annotation and functional annotation. Structural annotation can simply be defined as a process which involves the correct identification and localization of distinct sequence elements such as genes, regulatory elements, transposons, repetitive elements and more. Functional annotation means the process of assigning functions to the genes and the biological process in which it takes part. Annotation requires some degree of automation apart from manual assignment. Automatic annotation tools perform the task by computer analysis, softwares are necessary to run routine tasks in a batch mode and organize the results from different programs in a human-readable form. After this point, annotation is mostly manual which requires extensive human expertise. However, both are co-related and work parallel to each other.

The **GeneQuiz** (http://www.sander.ebi.ac.uk/genequiz/) project was the first automatic system for genome analysis, which performed similarity searches and based on certain sets of predefined rules, analyzed the results and assigned functional annotations to them (Scharf et al. 1994).

Since then, many tools were developed. Some of them are: RAST - Rapid Annotation using Subsystem Technology (Aziz et al. 2008), JCVI Annotation service (Tanenbaum et al. 2010), JGI/DOE IMG annotation service (Markowitz et al. 2012), MAKER Web Annotation Service (Cantarel et al. 2008).

1.2 Types of Genome Annotation

As discussed above, genome annotation can be divided into two major categories: Structural annotation and Functional annotation.

1.2.1 Structural Annotation of Genome

Structural annotation is the identification of genomic elements. It mainly involves identification of ORFs, intron-exons, coding regions and location of regulatory motifs. Some of the best known tools for gene prediction are Glimmer (Delcher et al. 1999), GENSCAN (Burge & Karlin 1997) and GenMark (Borodovsky & McIninch 1993).

Generally, two strategies are involved in gene prediction: (I) *ab initio* or *de novo* methods, which predict genes on the basis of local sequence characteristics; (II) similarity-based methods, uses the process of sequence similarity to known genes.

1.2.1.1 Ab-initio methods

Ab-initio methods rely on mathematical models such as *signals, for instance* short DNA motifs or *coding statistics* (Yandell & Ence 2012). Signal detection denotes the detection of consensus motifs such as promoters, splice sites, start/stop codons. There are various approaches for this problem such as consensus string representations, regular expressions and weight matrices. Algorithms mainly used for this purpose today are Hidden markov model (HMM) and Neural Networks (NN).

Coding statistics employs compositional differences between exons and introns like ORF length, GC content and codon biasness to predict gene structures.

The great advantage of *ab initio* gene predictors for annotation is that, they do not need any external evidence such as protein alignments to identify a gene or to determine its intron–exon structure.

This method was first described in prokaryotic genomes. In prokaryotes, genes are defined by start and stop codons, these regions are known as open reading frames (ORF). However, this method has few limitations, for example, they report CDSs but fail to report untranslated regions (UTRs) or alternatively spliced transcripts. Also, this method relies on pre-compiled training sets having organisms specific genomic traits such codon frequency, GC content and exonic and intronic lengths. So, unless the query genomes are closely related to an organism for which precompiled parameter files are available, this method will not prove to be very useful. If these limitations are overcome, the accuracy of this method is close to 100 percent (Yandell & Ence 2012).

1.2.1.2 Homology based methods

Homology based gene finding exploits a similarity between unknown genomic regions and protein or DNA sequences in the available databases. Traditionally, the primary tools for finding similarity are BLAST (Altschul et al. 1990) and FASTA (Pearson & Lipman 1988) even though HMMER (Eddy 2008) is gaining importance in recent years. Species can be considered as homologs based on genomic patterns, i.e., number of nucleotides, gene order and percentage of similarity between two species. It is generally believed that almost 50% genes can be identified via similarity searching even though the success rate is highly dependent on previously collected data and can be limited by low database quality.

1.2.2 Functional annotation of Proteins

Functional annotation can be defined as linking a gene with a particular biological process and assigning a function to it. In addition, a gene can be functionally linked with other genes when it is co-expressed with the genes of known function. Alternatively, a gene can be linked with other genes through a shared genomic neighborhood (especially in prokaryotes) which is also known

as gene context view. Functional annotation can be attained by similarity based approaches as well as comparative analysis.

1.2.2.1 Similarity based approaches

The similarity approach denotes the use on inter-species sequence similarities based on the on the evolutionary relations between genes. It involves searching of databases that contain genes/proteins with known functions, collected from experimental results. This prediction relies on the assumption that the homologous proteins perform similar functions and therefore the experimentally determined function of one protein can be tentatively assigned to its homologs. Similarity search can be carried out over whole sequences (global alignment) or a part of sequence or domains (local alignment).

Global alignment:

An alignment that assumes that the two proteins are basically similar over the entire length of one another is known global pairwise alignment. The alignment attempts to match them to each other from end to end, even though parts of the alignment are not very convincing. The classical global alignment tool is Needleman-Wunsch algorithm (Needleman & Wunsch 1970).

Local alignment:

An alignment that searches for segments of the two sequences (e.g. domains or motifs) that match well is known as local pairwise alignment. There is no attempt to force entire sequences into an alignment, just those parts that appear to have good similarity, according to certain criterion. Local alignment tools include the classical Smith-Waterman algorithm (reference) as well as the heuristic program BLAST, which is the most frequently used bioinformatics program today (reference).

Both local and global alignment algorithms can be implemented to perform pairwise or multiple alignment.

Pairwise alignment is used to identify the regions of similarity between two protein's amino acid or nucleotide sequences. This approach is used for database searching when the results are ranked according to similarity score and the query is assigned the function of the strongest hit. In contrast, *Multiple alignment* is the alignment of three or more biological sequences of similar length. Multiple alignment tools include ClustalW (Thompson et al. 1994), Muscle (Edgar 2004).

In terms of functional annotation, an automated use of similarity searching, especially global alignment has several limitations. For instance, the new sequence may be homologous to genes/proteins with unknown function, and also that homologous genes may not perform the same function. This makes the prediction even more difficult because gene might be related to each other through duplication events. These errors may affect the ongoing annotation prediction. Local alignment in such prediction is very helpful in this respect especially since protein domains are indicative of function, especially in eukaryotes. Generally, domains are the structural and functional building blocks of proteins that can exist, evolve and function independently of the rest of the protein chain (Dhir et al. 2010). Identification of similar domains and domain contexts can help a great deal in functional annotation. Protein domain prediction falls into the category of supervised classification but the prediction methods differ widely in the form and amount of background knowledge used in the analysis. For example, when we use HMMs for domain identification, we assume knowledge on the protein classes we want to compare (Dhir et al. 2010). New protein sequences can also be studied with respect to functional annotation by searching domain sequence databases such as SBASE (Dhir et al. 2010). The most widely used tool for this purpose today use various consensus representations for describing protein families, functional motifs or domains. The Interpro site is a collection of a variety of tools that describe protein families and domains and important sites (Hunter et al. 2012; Quevillon et al. 2005). The underlying methods are mostly based on multiple alignment

of proteins with known functions using HMMs. A number of such secondary protein sequence databases like Pfam, SMART, PRINTS are built using HMM profiles (Quevillon et al. 2005).

1.2.2.2 Comparative analysis

Broadly speaking, comparative or context based analysis can be considered an extension of pair-wise sequence alignment. Some methods of context based analysis are phylogenetic footprinting (Lenhard et al. 2003), protein fusion (Yanai et al. 2001), and chromosomal gene neighbourhood analysis (Overbeek et al. 1999a).

Phylogenetic footprinting

Coding regions are more conserved among different species as compared to their non-coding counterparts (Lenhard et al. 2003), or in other words, it can be said that the coding regions are under greater evolutionary constraint than the nonfunctional elements, and, thus, encounter a slower rate of sequence change through time. This is exploited for alignment of genomic segments from closely related organisms to identify "phylogenetic footprints": regions with slower rates of sequence change than the background (Ganley & Kobayashi 2007). Tools using this approach are ROSETTA, SLAM and TWINSCAN (Korf et al. 2001; Alexandersson et al. 2003; Batzoglou et al. 2000).

While using this technique, care is taken to align the query sequence with the genomes which are not very divergent since more divergent species will have less sequence homology between orthologous genes. Therefore, it is necessary to select species that are related enough to detect homology, but divergent enough to maximize non-alignment "noise".

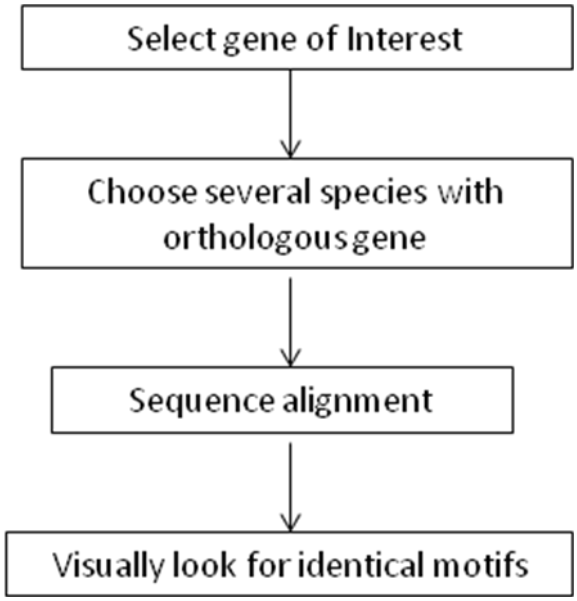


Figure 1: Flowchart of Phylogenetic footprinting

Protein fusion

Itai et. al highlighted the fact that two genes can be functionally associated if there is an instance in which both are fused to form a continuous sequence in another organism (Yanai et al. 2001). The continuous sequence is an indication of a link between fused genes. This idea assumes the fact that orthologous genes have same functions in both the organisms.

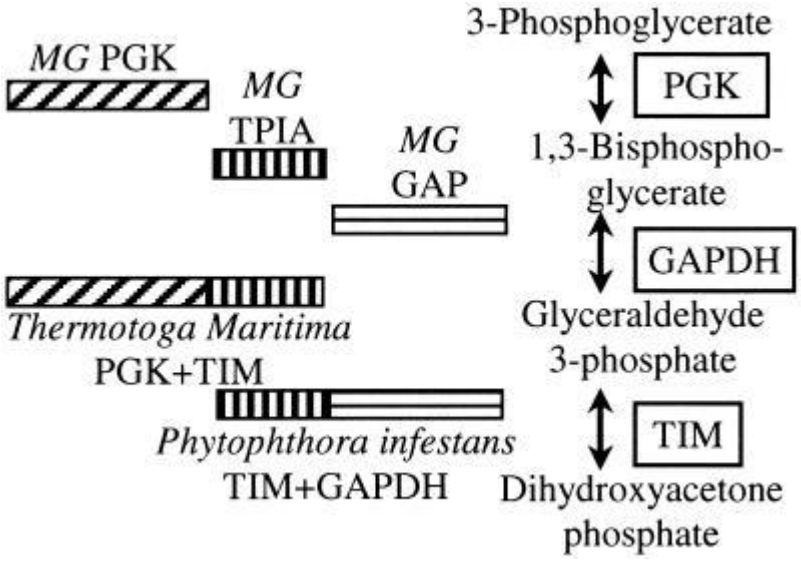


Figure 2: Functional association between genes linked by fusion method (adapted from (Yanai et al. 2001)).

Figure above shows two fusion genes in *T. maritima* and *P. infestans* functionally linking three genes in *Mycoplasma genitalium.*

Chromosomal gene neighborhood

It is very well known that in prokaryotes, genes with related functions are close to each other either as part of an operon or a functional neighbourhood. These form gene clusters and are often conserved in prokaryotes (Overbeek et al. 1999b; Overbeek et al. 1999a). This was concluded based on the following principal arguments: (i) Monte Carlo simulations, which showed that the possibility of finding identical strings of more than two genes by chance in more than two genomes is extremely low; (ii) most of these conserved strings that include characterized genes are either operons or are functionally linked genes and can be predicted to form operons; (iii) typical conserved gene strings matches the characteristic size of operons which is 2 to 4 genes; (iv) conserved gene strings that include genes from adjacent, independent operons are extremely rare; (v) nearly all conserved gene

strings contain genes that are transcribed in the same direction (Wolf et al. 2001; Koonin EV 2003). As a result, it can be presumed that the conserved gene strings are co-regulated, i.e. form operons, even if they contain additional promoters.

There are tools that allow functional prediction based on gene context view such as STRING (von Mering et al. 2007), PROLINKS (Bowers et al. 2004) and GeConT (Martinez-Guerrero et al. 2008). While STRING and PROLINKS concentrate on protein sequences, GeConT allows to visualize the genes and their genomic context. Yet, another tool, Microbial Genome Viewer (MGV) allows the visualization of gene in terms of their genomic context but also builds chromosome wheels and linear genome maps from genome annotation data (Kerkhoven et al. 2004).

The results produced by context-based methods are often very informative and also visually appealing.

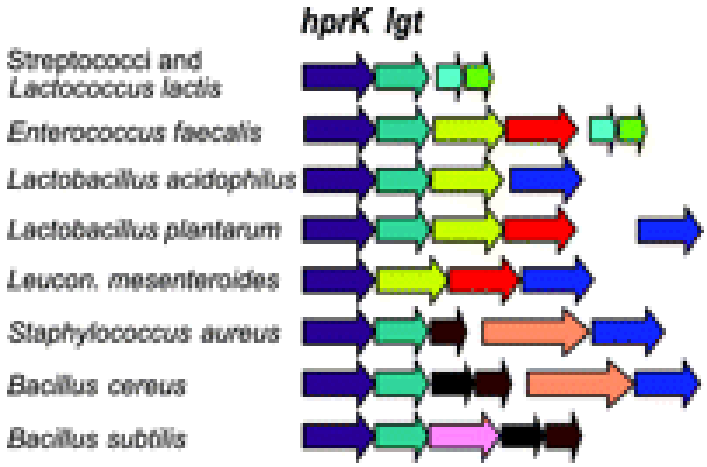


Figure 3: Gene Context View of hprK gene in phylum Fermicutes (adapted from Deutscher et. al. (Deutscher et al. 2006)**).**

Comparative gene finding include conserved functional relationships and provide information complementary to sequence homology (Plata et al. 2012).

In all the above listed methods, a single genome can be sequenced at once. However with dramatic increase in the quantity of sequencing data each year, it has become more and more difficult to maintain the accuracy of the annotation.

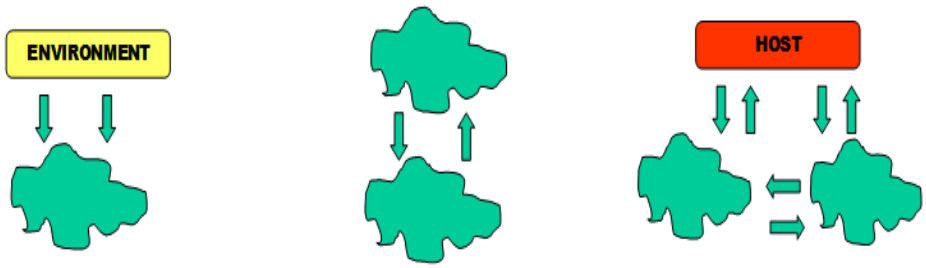
The quality of data highly depends of the method used i.e, manual or automatic. This may include laboratory experiments to analyze the data and further processed with manual or automatic annotations. Automated procedures are more vulnerable to errors. To improve the accuracy of such high-throughput annotations, Overbeek et. al suggested to annotate a single subsystem of collected genomes rather than annotation of all the genes in a single genome (Overbeek et al. 2005). According to Overbeek et. al a subsystem is a set of functional roles that together implement a specific biological process or structural complex. A pathway can be more generalized term for subsystem.

We focus on subsystem based annotation of bacterial genomes with respect to bacterial communication genes.

1.3 **Bacterial Communication And Quorum Sensing:**

Bacterial communities constitute a major form of life which is perhaps the largest contributor to the ecosphere in terms of materials transformed. These communities appear all over the natural habitats including the oceans and the soil, but they also coexist with other living organisms, they populate the gastrointestinal tract of metazoa, they fix nitrogen for plants, etc. Cells within these communities are involved in complex communication and sensing processes. The underlying processes are studied by various disciplines of biological research, One of their distinctive feature is the fact that they often induce social phenomena, i.e. behaviour patterns that are characteristic of communities, but not of an individual cell. The object of this thesis is quorum sensing, a fundamental molecular mechanism that involves production and sensing of signals that ultimately give rise to a density dependent change of cell behaviour. In order to put quorum sensing into a proper context, let's outline a common logical scheme for bacterial sensing and communication.

The sensory capacities of bacteria have been noticed for long times. Bacteria are capable of sensing nutrient and moving towards them, or sensing toxic substances and moving away from them. This process, called chemotaxis (Wadhams & Armitage 2004). With the advent of the genomic era it was noticed that sensory gene arrangements are common to many kinds of bacteria, and it was hypothesized that the versatility of a bacterium “a bacterial IQ) can be defined based on the number of the sensing genes (Galperin 2005). From the logical point of view, chemotaxis is an unidirectional sensing process wherein the bacterium passively senses a molecule of the environment (Figure 4). Bidirectional or mutual signaling occurs when cells produce and emit signals so they are capable to sense the signals of the other cells. Quorum sensing is a special case of bidirectional or mutual signaling. In quorum sensing, cells can perceive the same signaling molecules that they produce (autocrine signaling). Since the signal can diffuse away or decompose in the environment, the cells are capable of sensing their environment. And since the external concentration of the signal is dependent on bacterial density, the cells will indirectly sense the density of their own species. Finally, there are environments in which the mutual sensing and environmental sensing are both pronounced. For instance, bacteria of the gut microbiome are all under the strong environmental control of the host organism. On the other hand, they can stimulate the host to emit signals that will affect the other bacteria in the community. And inevitably, they will also sense and response to the molecules produced by members of the gut microbiome. This is a highly simplified view, since, for instance, environmental sensing or mutual signaling probably cannot exist in a pure form. Nevertheless, it highlights an obvious fact: environmental sensing, quorum sensing and host-microbe interactions are all built on related molecular mechanisms, receptors, regulators, some of which are conserved from unicellular organisms to humans.



Environmental sensing (e.g. Chemotaxis)

Mutual signaling and sensing (e.g. quorum sensing)

Combination of environmental and mutual sensing (e.g. gut flora)

Figure 4: Bacterial communication types

The goal of this thesis is the study of Quorum sensing (QS), a process in which bacteria regulate gene expression according to their local population density with the help of small signaling molecules (Fuqua et al. 1994). Quorum sensing involves the production, release, and detection of signal molecules called autoinducers. Extracellular autoinducer levels increase in proportion to increasing cell-population density. Upon reaching a minimal threshold level, signaling molecule binds to a sensor/regulator protein and these regulate the transcription of target genes (Miller & Bassler 2001; Fuqua et al. 1994).

Some bacterial processes (such as the invasion of a host organism or a natural habitat) cannot be carried by single cells alone but, rather, these processes only become successful when undertaken in synchrony by a group of cells. Bacteria can use quorum sensing to determine if there are a sufficient number of cells present to successfully initiate particular tasks. One such example of this is the production of bioluminescence by the marine bacterium *Vibrio fischeri*, a symbiont of the Hawaiian bobtail squid *Euprymna scolopes*.

When the population of *V. fischeri* reaches a certain density in the light organ of the squid, the concentration of the signaling molecule, exceeds a threshold extra-cellularly, triggering a cascade of cellular events that finally manifest in the production of light (Nealson & Hastings 1979; Nealson et al. 1970).

Originally, it was assumed that this phenomenon is unique to Vibrios, but it is clear that many bacteria use quorum sensing to recognize when they are alone versus when they are in a community, to distinguish self from non-self, and to control important collective activities.

1.4 **Processes controlled by Quorum Sensing**

Gram-positive and Gram-negative bacteria use quorum sensing communication circuits to regulate a variety of physiological activities. These processes include symbiosis, virulence, competence, conjugation, antibiotic production, motility, sporulation, and biofilm formation (Miller & Bassler 2001). Examples of processes controlled by quorum sensing are shown in Table 1

Table 1: Processes controlled by QS systems

Species	System	Examples of phenotype
B. glumae BGR1	*tofIR*	Polar flagellum (Kim et al. 2007)
B. glumae BGR1	*tofIR*	Swimming motility (Kim et al. 2007)
B. glumae BGR1	*tofIR*	Swarming motility (Kim et al. 2007)
B. thailandensis	*btaIR1/btaIR3*	Swarming motility (Ulrich et al. 2004)
B.thailandensis	*btaIR1/btaIR3*	Twitching motility (Ulrich et al. 2004)
B. cenocepacia J2315	*cciIR*	cepIR (O'Grady et al. 2009a)
B. vietnamiensis G4	*cepIR,bviIR*	BviIR (Malott & Sokol 2007)
B. pseudomallei 1710	*bpmlIR*	MprA protease (Valade et al. 2004)
B. glumae BGR1	*tofIR*	Toxoflavin (Kim et al. 2007)
B. thailandensis	*btaIR1/btaIR3*	beta-hemolytic (Ulrich et al. 2004)
B. subtilis	*comQXPA*	Surfactin (López et al.

		2009)
B. pseudomallei 1710	*bpmIR*	Melioidosis (Valade et al. 2004)
B. pseudomallei K96243	*bpsIR1, bpsIR2, bpsIR3*	Melioidosis (Holden et al. 2004; Kiratisin & Sanmee 2008; Ulrich 2004)
P. aerugenosa	*lasIR, rhlIR*	Cystic fibrosis (Singh et al. 2000)
B. subtilis	*comQXPA*	Competence (Dubnau 1991)
B. ambifaria	*bafIR*	Antifungal (Zhou et al. 2003)
B. thailandensis	*btaIR2*	Bactobolin production (Seyedsayamdost et al. 2010; Duerkop et al. 2009)
B. sp.383(considering B. lata 383)	*cepIR*	Pyrrolnitrin (Schmidt et al. 2009)
B. cenocepacia J2315	*cepIR*	Ornibactin (Lewenza & Sokol 2001)

1.5 **Quorum Sensing signals**

Different QS systems can be distinguished by the different types of signaling molecule they use, which are normally also associated with different types of signal synthesis, import and export, reception and response machinery (Pai & You 2009). Some QS related response include bioluminescence, exo-polysaccharide productions, virulence, antibiotic production, biofilm formation (Lazdunski et al. 2004).

Different types of signaling molecules used in QS systems are acyl homoserine lactones (AHLs), AI-2 molecules, and modified oligopeptides. AHLs mostly affect transcription via a one-component signal transduction system, where the receptor protein that binds the AHLs is fused to a DNA binding domain. Peptide communication molecules and AI-2, on the other hand, often affect transcription via two-component signal transduction systems which is composed of a histidine kinase and a response regulator protein (Ulrich et al. 2005).

1.6 **Acyl-homoserine lactones Quorum Sensing**

Acyl-homoserine lactone (acyl-HSL) mediated QS system is used by a large number of Gram negative bacterial species that interact with plant and animal hosts. This type of bacterial cell-to-cell communication was discovered in the context of microbial ecology, but it is now evident that acyl-HSL signaling is important in plant and animal (including human) diseases (Parsek & Greenberg 2000). AHL signaling was first described in luminous marine bacteria *Vibrio fischeri* (Eberhard et al. 1981; Nealson et al. 1970). AHL production is catalyzed by an AHL synthase (*luxI* gene product) belonging to the LuxI-family of proteins; this enzyme requires S-adenosylmethionine (SAM) and an acylated acyl carrier protein (ACP) from the fatty acid biosynthesis pathway, as the substrates (Schaefer et al. 1996).

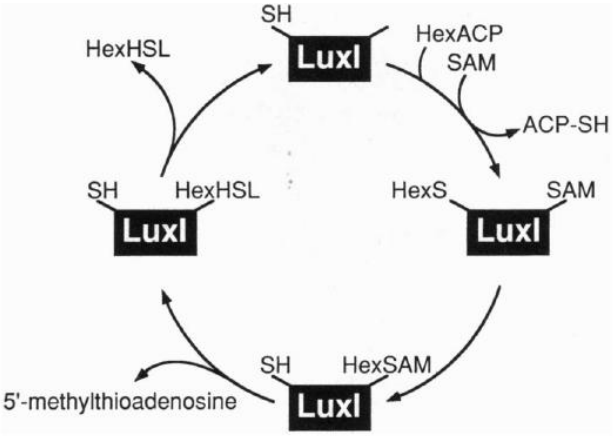


Figure 5: Hypothetical scheme for hexanoyl homoserine lactone synthesis catalyzed by LuxI in *V. fischeri*
[adapted from (Schaefer et al. 1996)**]**

Schaefer et al and Sitnikov et al. suggested a model, according to this model, the acyl group is transferred from the ACP to an active site cysteine in LuxI, SAM binds to the enzyme and an amide bond is formed between carbon 1 of the fatty acid and the amino group of SAM to form an acyl-SAM intermediate. The formation of acyl-SAM is then followed by lactone ring formation to yield

the acyl homoserine lactone and 5'- methylthioadenosine (Schaefer et al. 1996; Sitnikov et al. 1995).

Upon reaching a minimal threshold level, these signaling molecules enter the surrounding media, and subsequently penetrate into another nearby cell. Upon entering the cell, the autoinducer binds to a specific sensor/regulator protein which belongs to the LuxR-family of transcriptional regulators and forms a positive transcriptional complex. Mostly, AHLs diffuse through membranes and bind LuxR-like response regulators. LuxR-like response regulators simultaneously act as sensors and transcription factors (Choi & Greenberg 1991).

The N-terminal region of the LuxR protein has an AHL-binding domain which is reported to facilitate the formation of functional homodimers, allowing it to bind to DNAs. In this type of signal transduction systems where, signal binding domain and transcriptional-regulating DNA binding domain are fused, are referred to as one-component signal transduction system. This AHL-LuxR complex then regulate the transcription of target genes (Miller & Bassler 2001). Most commonly, one of the targets of the LuxR/AHL complex is the *luxI* gene thus creating a positive feedback loop which increases the production of AHL. This positive feedback loop is believed to be important for the timing of the QS response in accordance with the AHL concentration hence population density (Goryachev et al. 2006).

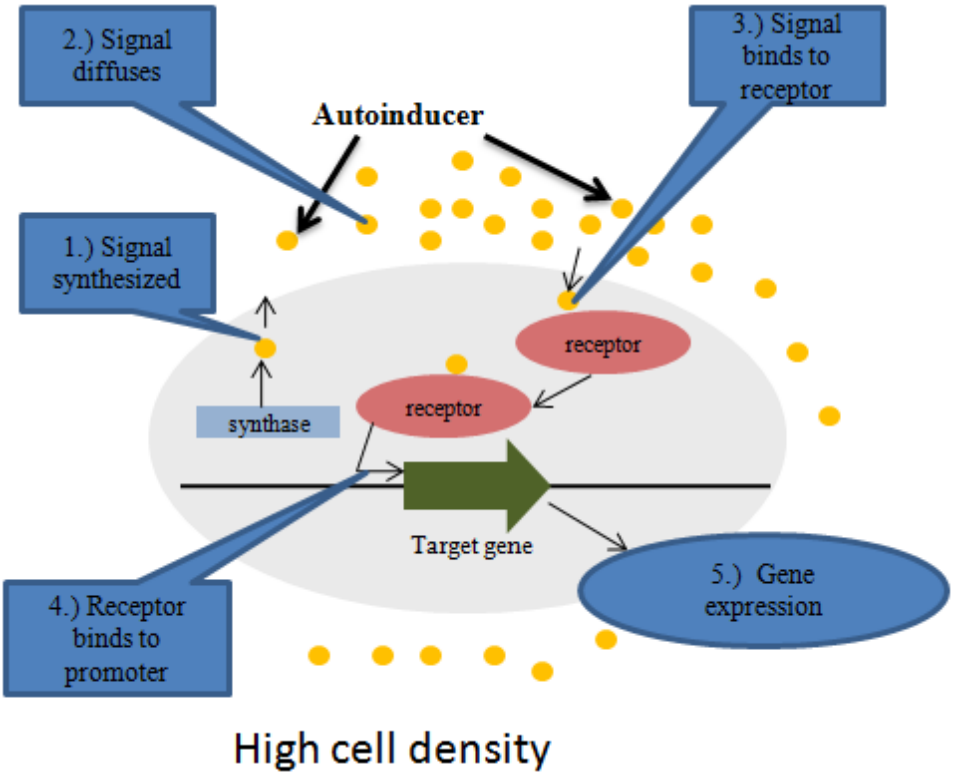


Figure 6: Pictorial representation of AHL-mediated Quorum sensing mechanism.

1.7 Acyl Homoserine Lactones

AHLs are composed of a homoserine lactone ring with an attached fatty acid chain, which can vary in length between 4 and 18 carbons, and may or may not have an oxo-group in position 3 (Whitehead et al. 2001; Reading & Sperandio 2006). The types of AHLs produced and the ways in which AHL production is regulated varies greatly among bacterial species. Some organisms possesses a single AHL synthase gene which is specific for production of a single AHL class (Watson et al. 2002; Churchill & Chen 2011). However, there are many species that have more than one AHL synthases producing different AHLs. For example *B. mallei* has two *luxI/R* pair homologs producing four different AHLs (Z Gelencsér, Choudhary, et al. 2012). Thus, it seems that the synthesis of AHLs depends on the presence and levels of different AHL synthases, which in turn have different specificities for AHL production.

1.8 **Non canonical AHL synthesis**

The most studied and common concept of AHL synthesis is that AHL synthases uses straight chain fatty acid from cellular pool (Schaefer et al. 1996) but there are studies for some non-canonical views about AHL synthesis. One such example is the AHL synthases of *Rhodopseudomonas palustris* which uses an aromatic fatty acid to produce aryl-HSL (Schaefer et al. 2008; Ahlgren et al. 2011).

The believe of AHL synthase using acyl-ACPs rather than acyl-coAs as a preferred substrate was also negated. Andrea Lindemann et. al (Lindemann et al. 2011) showed that BjaI from *B. japonicum* USDA110 is active with isovaleryl-CoA (IV-CoA), not with isovaleryl-acyl carrier protein (IV-ACP). ACP and CoA have similar chemical characteristics, and both form adducts with carboxylated substrates and present these substrates to enzymes. ACP proteins typically function in biosynthesis and in transferring fatty acid substrates, whereas CoA is found linked to more structurally diverse carboxylated substrates. CoA-modified compounds are intermediates in biosynthetic pathways as well as in biodegradation pathways (Lindemann et al. 2011). Figure below shows a cladogram of LuxI proteins, with the highlighted ones having non-canonical AHL systems.

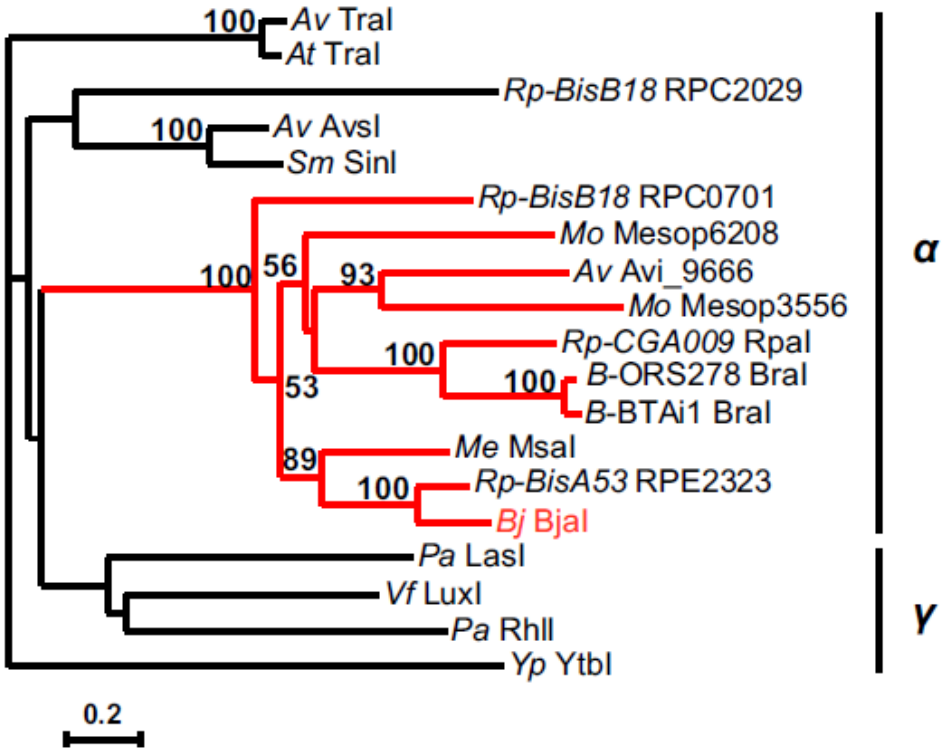


Figure 7: Phylogenetic analysis of LuxI proteins. Subfamily tree highlighted in red are the ones involved in non-canonical AHL synthesis (adapted from (Lindemann et al. 2011)**).**

1.9 AHL Quorum Sensing components

1.9.1 AHL synthases

LuxI proteins vary from 190 to 226 amino acids in length (Fuqua & Greenberg 2002) and the sequence conservation between any members of this family is 20% identity / 40% similarity (Churchill & Chen 2011). However, there is no simple correlation between the acyl-HSL synthesized and the level of sequence identity between LuxI proteins. Although, it has been noted that many of the LuxI-type proteins that direct the synthesis of 3-oxo-acyl-HSLs have a conserved threonine residue at position 143 (Fuqua & Greenberg 2002).

The LuxM and AinS proteins of *Vibrio harveyi* and *V. fischeri* respectively, are a second family of acyl-HSL synthases (Bassler et al. 1994; Milton et al. 2001). These protein show no sequence similarity to LuxI proteins, however, they use SAM and acyl-ACP as observed in LuxI. While, in contrast to LuxI proteins, AinS can use acyl-coA.

1.9.2 LuxR proteins

LuxR-type AHL receptors are approximately 250 amino acid residues in length which is folded into two functional modules. The N-terminal domain (NTD) holds the AHL-binding site, and the C-terminal domain (CTD) possesses the DNA-binding activity. All functional LuxR homologues also contain a helix–turn–helix (HTH) motif in their carboxyl terminus, which is essential for DNA binding (Fuqua & Greenberg 2002). The LuxR-type proteins can be of two types: transcription activators and transcription repressors (Churchill & Chen 2011). Transcription activators: the LuxR-type proteins binds to the DNA targets at positions that promote transcriptional activation of the controlled genes only in the presence of AHL for example: LuxR (*Vibrio fischeri*, in bio-luminescence) (Kaplan & Greenberg 1987), and LasR/RhlR (*Pseudomonas aeruginosa*, virulence factor expression and biofilm formation) (Ochsner et al. 1994; Latifi et al. 1995; Gambello & Iglewski 1991). Whereas transcription repressors like, EsaR (*Pantoea stewartii*, in exopolysaccharide production) (Beck von Bodman & Farrand 1995) binds to the DNA target of the promoter region in the absence of AHLs, and represses the controlled genes (Minogue et al. 2002). The presence of the AHL ligands relieves gene repression, most likely by altering the DNA binding of EsaR. EsaR can also act as an AHL regulated transcriptional activator, where the presence of AHL antagonizes its activation ability.

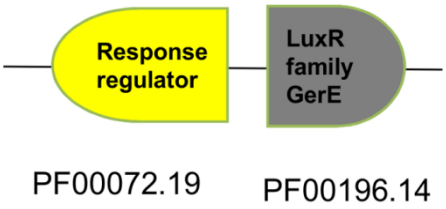


Figure 8: Domain structure of the LuxR proteins. The codes denote the PFAM (PF) domain identifiers.

1.9.3 Solo LuxR-type proteins

The number of LuxR and LuxI type proteins in bacterium is not always equal. It was observed that LuxR regulators are sometimes in excess as regards to AHL synthases, these are termed as orphan/solo LuxR as they are not associated with a cognate AHL synthases (Patankar & González 2009). In some of the cases, these orphan/solo LuxR have been found to respond to endogenously/exogenously produced AHLs or to molecules produced by eukaryotes. These solo LuxR therefore, might help in the expansion of QS network and be involved in interspecies and inter-kingdom signaling. The presence of LuxR solos in non-AHL-producing bacteria, on the other hand, could allow them to respond to AHLs produced by other bacteria, possibly allowing them to team up or maybe switch to a competitive behaviour towards their neighbours. It cannot be barred that some LuxR solos which lack conserved amino acid residues in the AHL domain are functioning in a ligand-independent way. Then again, these imperfect LuxR solos could be binding to signals which are not AHLs but are produced by nearby bacteria or eukaryotic cells. If they respond to compounds produced by eukaryotes, they are not necessarily involved in QS but rather in informing bacteria of their whereabouts (Subramoni & Venturi 2009; Subramoni et al. 2011). The best-characterized of these solo LuxR-type proteins are ExpR of *Sinorhizobium meliloti*, BisR of *Rhizobium leguminosarum bv. viciae* and QscR of *Pseudomonas aeruginosa*. These bacteria have multiple complete QS system. These solos are involved in regulatory circuits of QS systems, for example,

BisR in *Rhizobium leguminosarum bv. viciae. R.leguminosarum bv. viciae* has three QS networks, cin, rai and tra, with an additional LuxR-type protein, BisR. BisR negatively regulates cin QS system, thereby maintaining a low endogenous 3O-C14-HSL concentration (as CinI synthesizes 3O-C14-HSL) (Danino et al. 2003). Another example of an solo LuxR-type protein is QscR in *P. aeruginosa*. QscR which negatively regulates both the las and rhl QS system by binding LasR, RhlR and their cognate AHLs (Ledgham et al. 2003; Case et al. 2008). It is interesting to note that QscR has less specificity for AHL signals and is more sensitive to 3O-C10-HSL, an AHL not produced by *P. aeruginosa*, suggesting that QscR may function by responding to other AHLs produced by cohabitating microbes (Ledgham et al. 2003).

1.9.4 Classes of LuxR proteins

The LuxR proteins can be divided into different classes based on the interactions with their signal molecules (AHLs) and their multimeric properties (Stevens et al. 2011). Figure 9 shows different classes of this protein family.

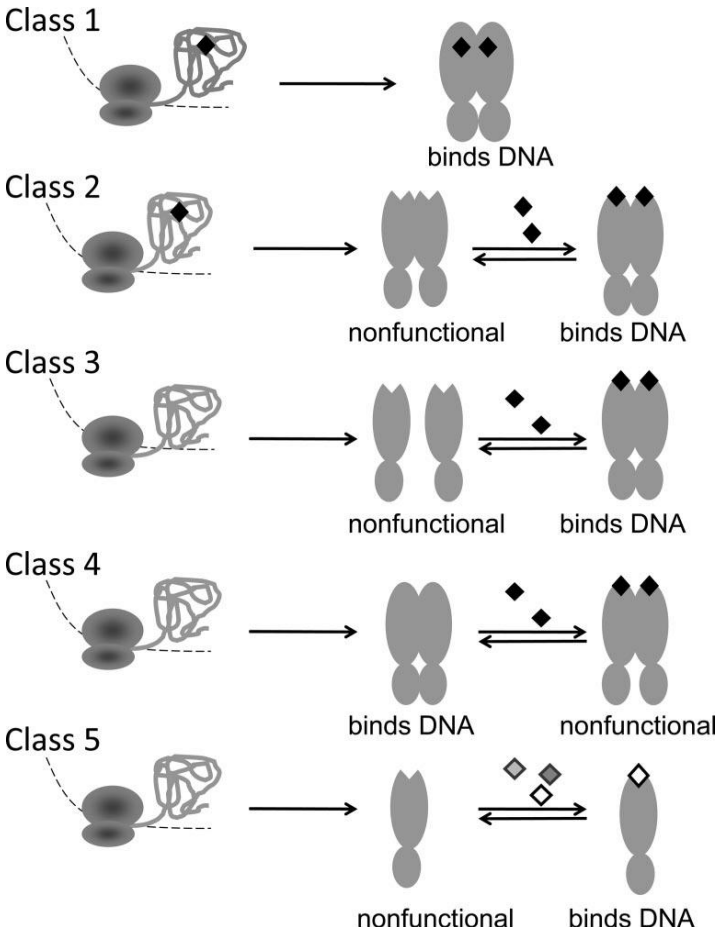


Figure 9: Classes of LuxR proteins based on their interaction with AHL molecules and multimeric properties

(Stevens et al. 2011).

In Class I regulators, LuxR binds to AHL co-translationally. In the absence of AHL, LuxR is rapidly degraded by proteases, such as the Clp and Lon. TraR is an example of class I regulators (Zhu & Winans 2001; Zhu & Winans 1999). Class 2 type of regulators such as LuxR in *V. fischeri* show enhanced stability in the presence of AHL, but the binding is readily reversible (Urbanowski et al. 2004). Class 3 regulators such as MrtR in *Mesorhizobium tianshanense* do not depend on AHL for protein folding, but the ligand is required for dimerization and activation (Sappington et al. 2011). Classes 1-3 include LuxR that are activators and become functional only after interacting with the AHL either co- or post-translationally. The activity of the class 1-3 proteins is enhanced in the presence of AHL. In contrast, class 4 homologues are mainly repressors such

as EsaR from *Pantoea stewartii*, and are capable of only functioning as dimeric DNA-binding proteins and regulating transcription in the absence of AHL (Minogue et al. 2005).

Class 5 proteins are solo LuxRs that do not dimerize in response to AHL, and the AHLs they recognize are produced from neighboring cells; there is no self-produced/native AHL. An example of class 5 proteins is SdiA (Patankar & González 2009).

1.10 **Quorum Sensing regulation**

N-AHL-based QS signaling is often referred to as autoinduction, requiring only a synthase and a sensor/regulator protein. However, in the absence of regulation of QS, autoinduction would increase signal levels without limit. A down-regulation loop, which turns on at higher signal concentrations is the simplest way to limit and stabilize the signal levels (Figure 10).

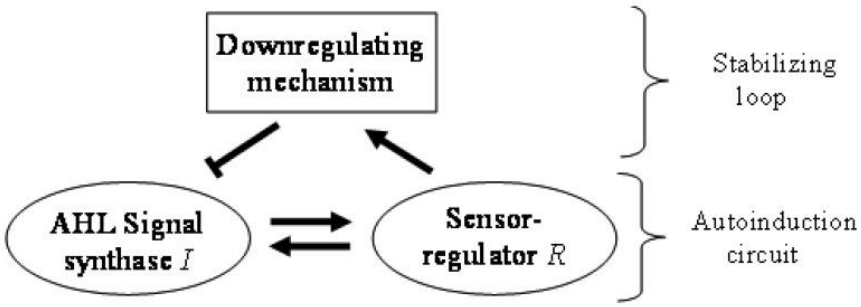


Figure 10: Regulatory outline of N-AHL-based QS signaling. Pointed arrows indicate activation, and the hammerhead arrow indicates inhibition
(adapted from Gelencser et al (Z Gelencsér, Choudhary, et al. 2012)**).**

There are a variety of mechanisms that can regulate QS systems. Some intergenic genes between *luxR* and *luxI* acts as negative regulator. The DNA-binding negative regulator RsaL acts as a homo-dimer by binding on the bi-directional *rsaL-luxI* promoter (Rampioni et al. 2012; Rampioni et al. 2006). Its negative effect is evident when its mutation leads to enhancement of surface

motility (twitching and swarming) and the production of secreted virulence factors (i.e., elastase, hemo- lysins, and hydrogen cyanide) with respect to the wild-type strain (Rampioni et al. 2012).

RsaM is another small regulatory protein believed to be acting in a similar way to RsaL in many *organisms* (Mattiuzzo et al. 2011; Venturi et al. 2011), its disruption results in the synthesis of larger quantities of AHLs. Both *rsaM* and *rsaL* are located in the intergenic regions between *luxR* and *luxI* homologs.

1.11 **Signal Specificity**

AHLs vary in their acyl chain length (from C4 to C18) shown in Figure 11. Recently, AHLs with methyl- branched acyl chains and multiple double bonds in the acyl chain has been discovered (Thiel et al. 2009). Some AHLs also vary in the degree of oxidation at the C3 position. The intrinsic selectivity of the AHL synthase is responsible for the preference for unsubstituted-, 3-oxo, or 3-hydroxy-acyl-ACPs. For example the AHL synthases RhlI and LasI produces an unsubstituted, C4-HSL and 3-oxo-C12-HSL respectively, at the same time in vivo from the same cellular pool of acyl-ACPs.

AHL synthase	Species	Signal
		3-oxo-AHLs
EsaI	*Pantoea sterwartii*	3-oxo-C6-HSL
LuxI	*Vibrio fischeri*	3-oxo-C6-HSL
YspI	*Yersinia pestis*	3-oxo-C6/C8-HSL
TraI	*Agrobacterium tumefaciens*	3-oxo-C8-HSL
VanI	*Vibrio anguillarum*	3-oxo-C10-HSL
LasI	*Pseudomonas aeruginosa*	3-oxo-C12-HSL
SinI	*Sinorhizobium meliloti*	3-oxo-C14-HSL
		3-oxo-9-cis-C16-HSL
		3-hydroxy-AHLs
	Xenorhabdus nematophilus	3-hydroxy-C4-HSL
PhzI	*Pseudomonas fluorescens*	3-hydroxy-C6/C8-HSL
CinI	*Rhizobium leguminosarum*	3-hydroxy-7-cis-C14-HSL
		Unsubstituted AHLs
RhlI	*Pseudomonas aeruginosa*	C4-HSL
CviI	*Chromobacterium violaceum*	C6-HSL
CepI	*Burkholderia cepacia*	C8-HSL
BmaI1	*Burkholderia mallei*	C8/C10-HSL
CerI	*Rhodobacter sphaeroides*	7-cis-C14-HSL
SinI	*Sinorhizobium melioti*	9-cis-C16-HSL
		C18-HSL

Figure 11: AHL signals produced by different bacteria
Modified from Churchill and Chen et. al (Churchill & Chen 2011)

However, there is no correlation between sequence composition and acyl-chain length of AHL synthases (Churchill & Chen 2011). The structures of EsaI and LasI suggest that the preference for acyl-ACPs with different acyl-chain lengths, may involve more complex sequence, and, there are differences at the level of tertiary structure in different AHL synthases.

1.12 AHL-mediated QS systems in Proteobacteria

Proteobacteria was first established by Carl Woese in 1987 and he called it "purple bacteria and their relatives" (Woese 1987). Because of the great diversity of forms found in this group, the Proteobacteria are named after Proteus, a Greek god of the sea, capable of assuming many different shapes. Proteobacteria are gram-negative and based on rRNA sequences can be divided into classes of alphaproteobacteria, betaproteobacteria, gammaproteobacteria, deltaproteobacteria, zetaproteobacteria and epsilonproteobacteria.

Case et al. analyzed 517 genomes for the presence of *luxR* and *luxI* genes and found that such genes are present exclusively in Proteobacteria (Case et al.

2008). Although, AHL-based quorum sensing is broadly distributed among Proteobacteria, still the *V. fischeri* system has remained an important model, particularly through studies of its symbiosis with the Hawaiian bobtail squid, *Euprymna scolopes*.

Many organisms use AHL-mediated QS system to colonise host and cause pathogenesis (Parsek & Greenberg 2000). Similarly, symbiotic microbes often utilize AHL quorum sensing to develop symbiotic root nodules on leguminous plants, example: *Rhizobia* (Oldroyd et al. 2005).

Apart from Vibrios, AHL systems are also highly studied in *Pseudomonas* and *Burkholderia*, because of their roles in many animal and plant pathogenesis.

1.13 **AHL-mediated QS systems in *Pseudomonas***

Pseudomonas is a genus of Gram-negative aerobic gamma-proteobacteria that can inhabit several environmental niches. These include important and dangerous pathogens such as *P. syringae*, a plant pathogen and *P. aeruginosa*, a human opportunistic pathogen. There are also some which have the ability to colonize plant-related niches, such as the rhizosphere (e.g. *P. aeruginosa*, *P. fluorescens*, *P. putida*, *P. aureofaciens* and *P. chlororaphis*), where they can act as plant-beneficial bacteria by antagonizing plant-deleterious microorganisms and through the production of traits that directly influence plant disease resistance and growth (Venturi 2006). The plant-beneficial Pseudomonads are found not only in rhizosphere but have also been isolated from different environments including soil and fresh water. Most attention from scientists is, however, focused on the opportunistic human pathogen *P. aeruginosa*.

AHL QS is particularly interesting in the pseudomonads due to the presence, diversity, and complexity of regulatory circuits present in various species. In the case of *P. aeruginosa*, AHL QS seems conserved and ubiquitous, being composed of a complex hierarchy of two LuxI/R pairs and a series of regulators (Venturi 2006; Schuster & Greenberg 2006). In fact, it has been estimated that quorum sensing regulates up to 3% of *P. aeruginosa* genes. On

the other hand, most strains of *P. fluorescens* and *P. putida* do not possess an AHL QS system (Elasri et al. 2001; Steindler et al. 2008).

P. aeruginosa is the most extensively studied organism for AHL QS system among Pseudomonads. *P. aeruginosa* harbors three QS systems: two LuxI/LuxR-type QS circuits that function in series to control expression of virulence factors as well as a third, non-LuxI/LuxR-type system called the *Pseudomonas* quinolone signal (PQS) system (Rutherford & Bassler 2012).

In the *las* system, LasI synthesizes *N*-(3-oxo-dodecanoyl)-L-homoserine lactone (3-oxo-C12-AHL), which interacts with LasR and activates transcription of target genes including those encoding virulence factors such as elastase, proteases, and exotoxin A (Gambello et al. 1993; Schuster et al. 2003; Gambello & Iglewski 1991). One of the LasR–3OC12HSL targets is *lasI*, which establishes an autoinducing feed-forward loop (Seed et al. 1995). Another target of regulation by LasR–3OC12HSL is a second *luxI* homolog called *rhlI* (Latifi et al. 1996). In the *rhl* system, RhlI synthesizes *N*-(butanoyl)-l-homoserine lactone (C4-AHL), which interacts with the cognate regulator RhlR and activates target gene promoters. There is inter-related regulation between the two systems. These regulate the production of multiple virulence factors including elastase, alkaline protease, exotoxin A, rhamnolipids, pyocyanin, lectins, superoxide dismutases and biofilm formation (Smith & Iglewski 2003).

AHL based QS system has also been identified in beneficial pseudomonads. *P. aureofaciens* 30-84 is a soil-borne bacterium that colonizes the wheat rhizosphere. This strain produces three phenazine antibiotics which inhibits the causative agent *Gaeumannomyces graminis var. tritici* and hence suppresses the disease in wheat plant. Phenazines also enhances the competition of *P. aureofaciens* 30-84 within the wheat rhizosphere with other invading organisms. Expression of the phenazine biosynthetic operon is controlled by the phzR/phzI N-acyl-homoserine lactone (AHL) QS system (Pierson et al. 1994; Wood et al. 1997). A second AHL QS system, CsaI/R, was also found in *P. aureofaciens* 30-84 (Zhang & Pierson 2001). This system although, does not regulate phenazine production, helps in rhizosphere colonization and in regulating the biosynthesis of cell-surface components.

Other important beneficial strains which involve AHL-QS system for rhizosphere colonisation are *P. chlororaphis* strain PCL1391, and *P. putida* strains WCS358 and IsoF.

1.14 **AHL-mediated QS systems in *Burkholderia* sps.**

The genus *Burkholderia* belong to beta sub-division of proteobacteria and contain Gram negative species that are ubiquitous in the environment and may cause a number of diseases in plants, animals and humans. Conversely, *Burkholderia* appear to be involved in nitrogen fixation in the rhizosphere and various species have the ability to metabolize pesticides and polychlorinated biphenyls (PCBs), making them attractive from an agricultural perspective. Quorum sensing (QS) systems appear to be crucial regulators of virulence and other phenotypic traits in *Burkholderia*. Furthermore, QS systems appear to be crucial in governing overall colonization and niche invasion. It is therefore essential that we have a more complete understanding of quorum sensing systems in *Burkholderia* and in particular the genetic basis upon which it operates.

One of the challenges faced by researchers is that *Burkholderia* are very diverse from taxonomic and genetic perspectives. For many years, *Burkholderia* were recognized as members of the non-fluorescent pseudomonads. However, later they were found to be taxonomically heterogeneous and on the basis of rRNA-DNA hybridization experiments five species homology groups were devised (Palleroni et al. 1973). The genus *Pseudomonas* was subsequently divided into five well-defined rRNA homology groups. Subsequently, polyphasic taxonomy analyses, including 16S rRNA sequence analysis, DNA-DNA hybridization and fatty acid analysis, led to the establishment of the *Burkholderia* genus to accommodate seven species of the *Pseudomonas* rRNA group II (*P. cepacia, P. caryophylli, P. gladioli, P. mallei, P. pseudomallei, P. solanacearum* and *P. picketti)* (Gillis et al. 1995; Yabuuchi et al. 1992). *Burkholderia* can be divided into three major clades. These include the *B. cepacia* complex (BCC), the "*pseudomallei*" group and a

recently described clade that contains non-pathogenic plant-beneficial *Burkholderia* species **(**Figure 12**)**.

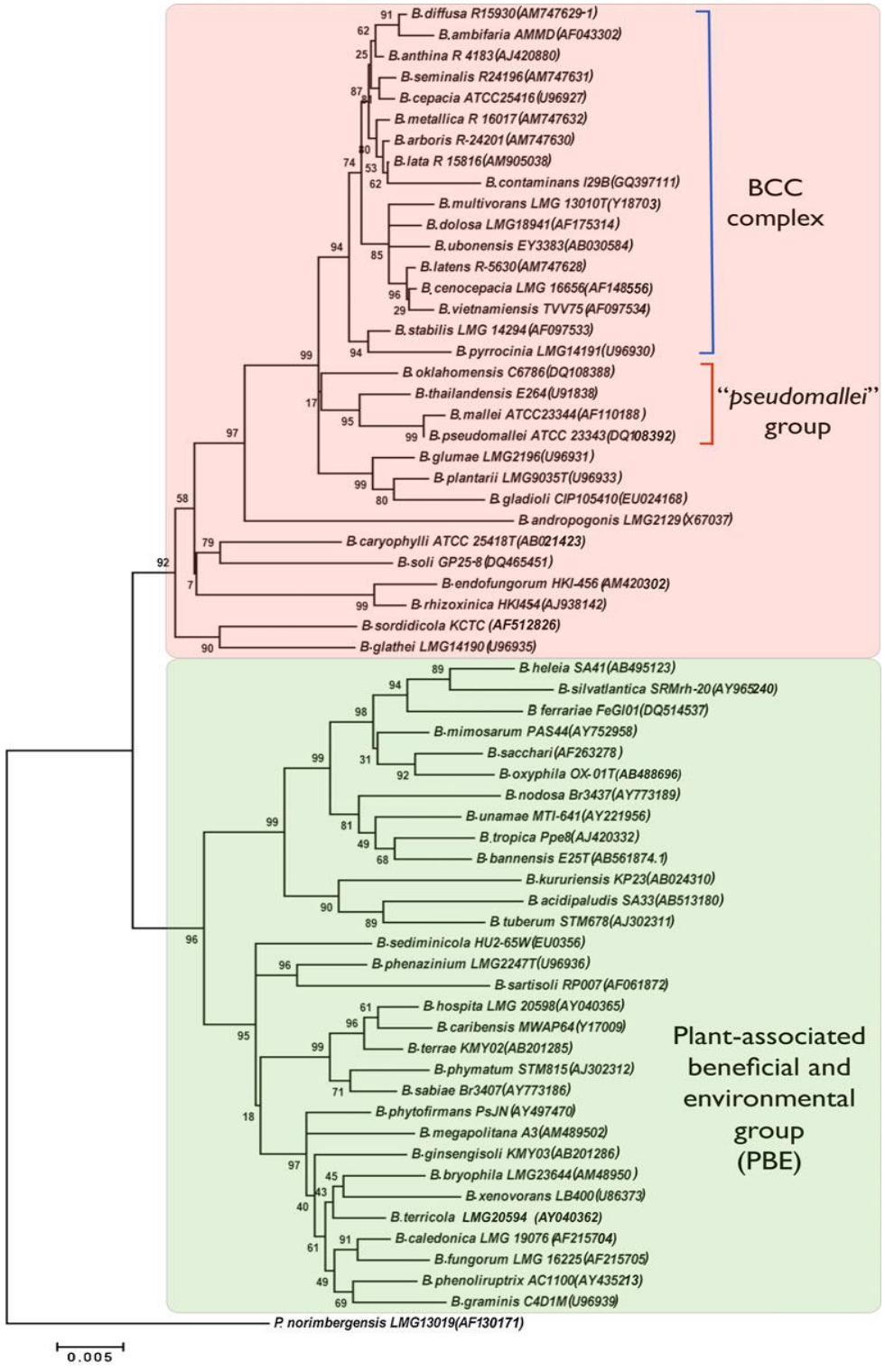


Figure 12: 16S rRNA sequences of the recognized species of the *Burkholderia* genus showing the three major clades viz. BCC complex, *pseudomallei* group and PBE group

adapted from (Suárez-Moreno et al. 2012)

Despite the apparent genetic diversity within the *Burkholderia* genus a common characteristic is that they many appear to utilize quorum sensing (QS) as part of their colonization and invasion strategies.

1.15 **Autoinducer-2 mediated QS**

Apart from AHL-mediated QS, some organisms have a parallel QS system whose signaling molecules are furanosyl borate diester, and are referred to as autoinducer-2. For ex., in *V. harveyi* 4,5-dihydroxy-2,3-pentanedione (DPD) is the precursor of autoinducer-2 and is synthesized by LuxS from S-adenosyl methionine (SAM). Thus, SAM is a precursor both in AHL synthesis and AI-2 synthesis (Winzer et al. 2002). AI-2 binds to LuxP and forms a AI-2-LuxP complex. This complex binds to luxQ kinase which triggers a dephosphorylation cascade by turning the kinases into phosphatases. LuxQ dephosphorylates LuxU, which in turn dephosphorylates LuxO. This then activates the transcription of the target genes (Waters & Bassler 2006).

1.16 **Quorum sensing in Gram - positive bacteria: Peptide Signaling**

In contrast to gram-negative bacteria, gram-positive bacteria use peptides as communication molecules. Peptide signaling requires a membrane-bound histidine kinase, and a response regulator. These components constitute a two-component signal transduction system (Kleerebezem et al. 1997). Similar to the mechanisms by which gram-negative bacteria use LuxI/R systems, each gram-positive bacterium uses a signal that is specific for its cognate receptors (Waters & Bassler 2005).

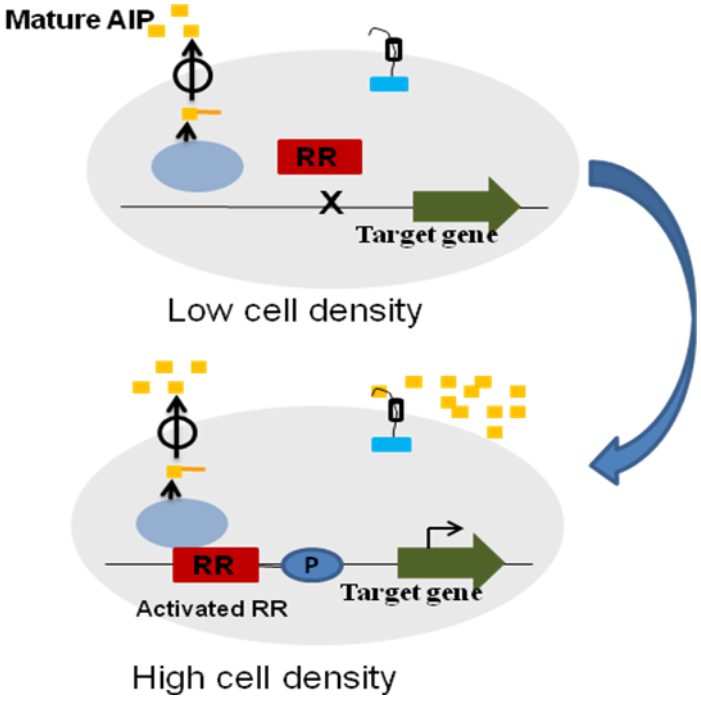


Figure 13: Mechanism of Two-component system
Modified from Muhammad Faisal Siddiqui et. al (Siddiqui et al. 2012)

However, unlike AHLs, these peptides are not freely diffusible and need a special oligopeptide transporter for transportation across the cell membrane (Waters & Bassler 2005). Among the Gram-positive peptide signals, there is a surprising amount of chemical variation. The simplest of these signals are the unmodified, linear peptides that are secreted by many Gram-positives. Some of the best known examples of this signal class are the *Bacillus* peptide signals, the competence peptide of *Streptococcus pneumoniae*, and the *Enterococcus* pheromones. An emerging class of peptide signals are the cyclic lactones and thiolactones produced by *Staphylococcus*, *Enterococcus*, and many other genera of Gram-positives. Finally, some peptide signals are the subject of more complex posttranslational modification, such as the *Bacillus subtilis* competence pheromones. Among the peptide signals, there is surprising diversity and complexity within the biosynthetic pathways (Thoendel & Horswill 2010).

A very well known example of peptide communication is the regulation of genetic competence in *Bacillus subtilis* that uses the ComX signal (a 6 amino acid peptide) (Okada et al. 2005). The comX system is a typical quorum sensing system as comX influences the behaviour of the producing cell population in a cell-density dependent manner (Thoendel & Horswill 2010). The functioning of the com system in *B. subtilis* relies on four proteins, comQ, comX, comP, and comA that are encoded by four adjacent reading frames situated on the same strand of the *B. subtilis* chromosome. The *com* locus functions as a classical autoinductive quorum sensing system in which the peptide signal is released by the cell, the external concentration of the signal is sensed by the releasing cell, and as a result, the production of the signal increases and other genes are also upregulated. Below is the description of the four com proteins.

ComQ

ComQ is a protein of 286 to 309 amino acids in length (average length 299). ComQXPA QS system requires an enzyme for postranslational modification of the ComX prepeptide. ComQ functions as an isoprenyl transferase and has the ability to attach the isoprenyl units to the tryptophan residue of ComX signal peptide. ComQ is reported to be a single-domain protein (e.g. PFAM: PF00348.12), with some databases (UNIPROT) indicating a putative membrane-binding segment of the protein.

ComX

The *comX* gene codes for a 52 to 73 residue-long precursor protein. The sequence of ComX is extremely variable (PFAM identifier: PF05952), but it contains a tryptophan within the 5 to 10 amino acids of the C terminus (Ansaldi et al. 2002).

ComP

ComP consists of two domains: histidine kinase domain (PFAM: PF07730) and an ATP-binding domain (PFAM: PF02518). The rest of the protein consists of a series of transmembrane helices separated by intra and extracellular loops,

The ComX binding site is located on the extracellular part of its N-terminal membrane-associated domain (Piazza et al. 1999).

ComA

ComA is a protein of 196 to 245 residues in length (average length 214 residues). The structure of ComA is similar to transcriptional response regulator. The N-terminal domain (PFAM: PF00072.19) carries the phosphorylation domain involving Asp residues. The C-terminal encompasses DNA-binding domain (PFAM: PF00196.14) (Weinrauch et al. 1990; Roggiani & Dubnau 1993).

The domain structure of the four proteins is shown in Figure 14.

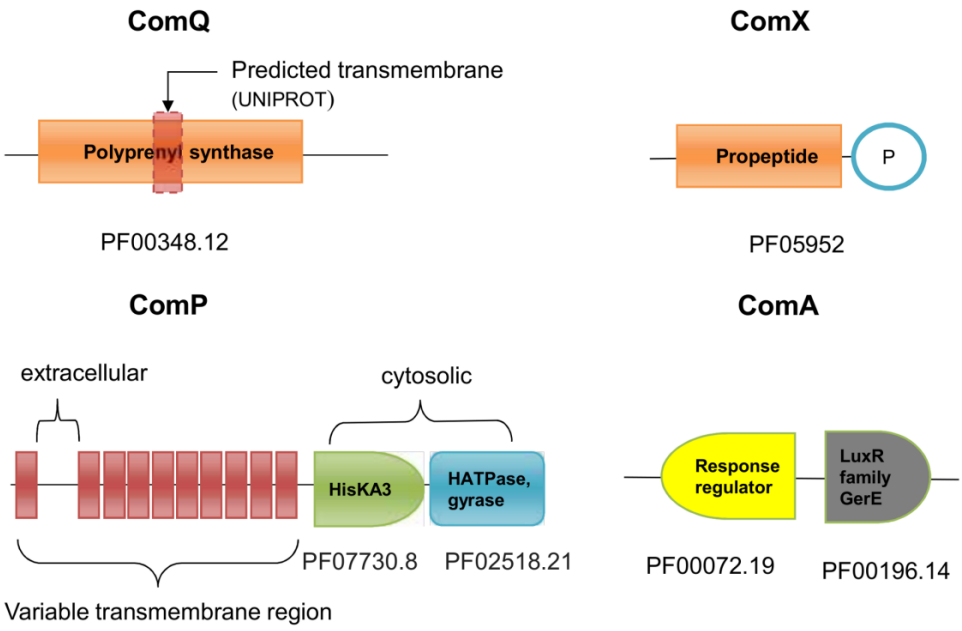


Figure 14: Domain structure of the com proteins. The codes denote the PFAM (PF) domain identifiers.

Other peptide communication systems include the competence stimulating peptide (CSP) of *Streptococcus mutans* (Petersen et al. 2006), the peptide antibiotic nisin in *Lactococcus lactis* (Kuipers et al. 1998). and autoinducing peptide AIP in *Staphylococcus aureus* (Thoendel et al. 2011).

1.17 **Other systems**

Apart from AHL, AI-2 and peptide QS system, few other QS systems have been observed. One of them include Quinolone signaling in *P. aeruginosa* in which signaling molecule is 2-heptyl-3-hydroxy-4-quinolone (HHQ), PQS, synthesized by PqsH. It is hydrophobic in nature and requires a transport system to be exported out of cell (Diggle et al. 2007). However, 2-alkyl-4-quinolone (AHQ) signaling is more restricted and has so far been detected only in *P. aeruginosa* and certain *Burkholderia* (Hodgkinson et al. 2010; Diggle et al. 2006).

PQS biosynthesis is directed in 2 steps. The PqsABCD proteins catalyze the condensation of β-keto fatty acids with anthranilate to yield AHQs, prominent among which is 2-heptyl-4(1*H*)-quinolone (HHQ) (Bredenbruch et al. 2005; Déziel et al. 2004; Pesci et al. 1999). Synthesis of PQS is primarily under the control of the AHL-dependent QS system (Wade et al. 2005), although it can also be produced independently of it (Diggle et al. 2003; Xiao et al. 2006). PQS is not the only alkyl-quinolone made by *P. aeruginosa* (Déziel et al. 2004; Pesci et al. 1999). Others include HHQ, 2-nonyl-4-quinolone (NHQ), and 2-heptyl-4-quinolone *N*-oxide (HHQNO). The signaling function(s), if any, of NHQ and HHQNO is not clear.

Some diffusible signal factors (DSF) mediated QS has been identified in *Xanthomonas* spp. It consists of three major QS components: RpfF, an enoyl-CoA hydratase catalyzing the synthesis of the signal molecule; RpfC and RpfG are the two-component system for the perception and transduction of the extracellular DSF family signals (Guo et al. 2012). DSF signals which are mainly produced by *Xanthomonas campestris* has recently been discovered in human pathogen *Burkholderia cenocepacia complex* and hence, it is known as **B**urkholderia **D**iffusible **S**ignaling **F**actor **(BDSF)**. Chemically, BDSF is cis-2-dodecenoic acid and is structurally similar to DSF of *Xanthomonas campestris* differing only in a methyl substitution at C11. In contrast to DSF signaling in *Xanthomonas*, BDSF employ one-component system for signal transduction.

RpfFBc is responsible for production of diffusible signal and RpfR acts as sensor/receptor (Deng et al. 2012).

Figure below shows different QS signals exploited by bacterial communities for communication.

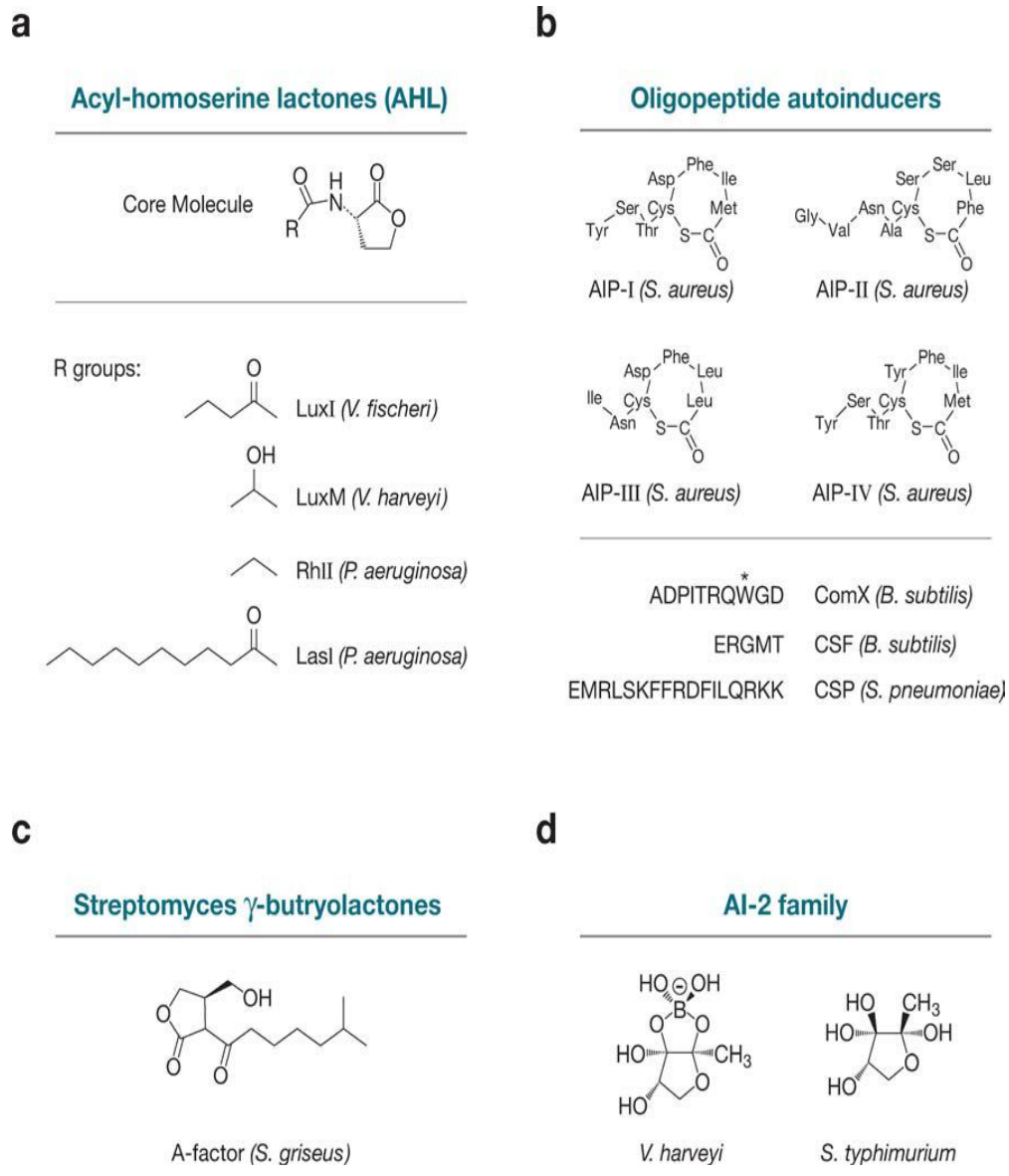


Figure 15: Different types of QS signals used by various bacteria. Modified from Waters and Bassler et. al. (Waters & Bassler 2005)

My main goal is to functionally annotate QS genes, check the arrangement of these genes on sequence and chromosomal level and make it available on an online repository, starting with AHL systems then expanding to other systems.

2 Objectives and Methods

2.1 Aims and Objectives:

The discovery of chemical communication among bacteria in the 1990s has fundamentally changed the traditional view that pictures bacteria as single-celled organisms living in isolation. In the last fifteen years, it has become increasingly evident that bacteria have the potential to establish highly complex communities. In fact, most bacteria are able to monitor their population density by producing and detecting small molecular weight signaling compounds (also called autoinducers) in a process which has been termed "quorum sensing" (QS) (Fuqua et al. 1994). In bacterial colonies, individual bacteria communicate, cooperate but also compete for resources using a small set of chemical signals. The basic regulatory mechanisms of QS are well understood but the fact that QS regulating proteins can respond to a variety of external signals has opened new vistas. Current knowledge of regulator mechanisms is often confined to single bacterial species and the relation between mechanisms and genomic architecture is poorly understood. The general goal of the project was to initiate systematic genomic studies for mapping, identifying and classifying QS-related genomic elements (genes, regulons) in the growing body of bacterial genomes, in order to better understand the evolution of QS communication and cooperation.

Aim 1: Systematic survey of regulatory mechanisms of *N*-acyl homoserine lactone (AHL) QS in gram negative bacteria.

Aim 2: Building of Hidden Markov Model recognizers (HMMs) for AHL QS, relying on the ICGEB collection of protein domains and functionally annotated sequences.

Aim 3: Screening of the available bacterial genomes for homologs of AHL QS genes and detection of chromosomal topology in the pertinent groups, together with their re-annotation.

2.2 Methods/Approaches

2.2.1 Data collection:

Systematic survey of regulatory mechanisms of AHL QS in gram negative bacteria was done by using all electronically available literature databases and public resources. It was very important to have background knowledge of mechanisms in question. Protein fasta sequences of complete genomes and draft genomes containing were downloaded from NCBI FTP site ftp://ftp.ncbi.nih.gov/genomes/Bacteria/ and ftp://ftp.ncbi.nih.gov/genomes/Bacteria_DRAFT/ respectively. However, there are some individual Genbank entries in NCBI RefSeq database, which are not included in the NCBI FTP site, these were also collected from http://www.ncbi.nlm.nih.gov/RefSeq/.

2.2.2 QS proteins detection:

For mapping and identifying QS proteins, Hidden Markov Model (HMM) recognizers were used. These HMM recognisers were build with the HMMER program, HMMER 3.0 http://hmmer.janelia.org/, based on four core set of protein sequences of LuxR, LuxI, RsaL and RsaM proteins shown inTable 2.

Table 2: List of protein sequence sets used for building HMM recognizers

(Z Gelencsér, Choudhary, et al. 2012; Choudhary et al. 2013; Gelencsér, Galbáts, et al. 2012).

LuxI homologues				
YP_002649215.1,	YP_003261728.1,	YP_003262850.1,	YP_003296640.1,	YP_003331715.1,
YP_003608088.1,	YP_004230809.1,	YP_003366470.1,	YP_004012993.1,	YP_004106681.1,
YP_004106954.1,	YP_004108425.1,	YP_003910269.1,	YP_003520250.1,	YP_003530770.1,
YP_003538486.1,	YP_003546445.1,	YP_003558209.1,	YP_003566926.1,	YP_003568278.1,
YP_003576501.1,	YP_003729883.1,	YP_003930460.1,	YP_003734012.1,	YP_003749682.1,
YP_003750860.1,	YP_003744153.1,	YP_003740503.1,	YP_003740954.1,	YP_003847234.1,

YP_003885141.1, YP_004088230.1, YP_004115279.1, NP_521405.1, NP_522340.1, NP_767703.1, NP_106262.1, NP_106661.1, NP_109412.1, NP_385945.1, YP_002965845.1, YP_002966879.1, YP_002346031.1, YP_002347420.1, YP_002426405.1, YP_428477.1, YP_105963.1, YP_106161.1, YP_110894.1, YP_111576.1, YP_001005892.1, YP_554693.1, YP_555669.1, YP_165635.1, YP_167511.1, NP_669050.1, NP_670673.1, YP_528965.1, YP_234707.1, NP_250123.1, NP_252166.1, YP_002232872.1, YP_002234481.1, YP_768958.1, YP_048233.1, NP_793636.1, YP_789671.1, YP_791820.1, NP_903761.1, NP_993604.1, NP_994737.1, YP_674865.1, YP_371808.1, YP_001114940.1, YP_001117676.1, YP_439001.1, YP_439708.1, YP_273860.1, YP_508562.1, YP_071011.1, YP_071751.1, YP_206882.1, YP_914595.1, YP_002551489.1, YP_002549360.1, YP_002541324.1, YP_659946.1, YP_317245.1, YP_776005.1, YP_617566.1, YP_617628.1, YP_838353.1, YP_623506.1, YP_470411.1, YP_473057.1, YP_001024425.1, YP_001025818.1, YP_001077901.1, YP_001078152.1, YP_989942.1, YP_001062290.1, YP_001063210.1, YP_335777.1, YP_337633.1, YP_972130.1, YP_484039.1, YP_486927.1, YP_567542.1, YP_569311.1, YP_530592.1, YP_531903.1, YP_781244.1, YP_001231849.1, YP_001604809.1, YP_001606209.1, YP_001399709.1, YP_001400525.1, YP_001220569.1, YP_001241094.1, YP_001242901.1, YP_001075256.1, YP_001076162.1, YP_002537871.1, YP_001327237.1, YP_453964.1, YP_681952.1, YP_650194.1, YP_651865.1, YP_647981.1, YP_649109.1, YP_002220095.1, YP_855089.1, YP_001161918.1, YP_001163229.1, YP_001347034.1, YP_001349251.1, YP_001143471.1, YP_001583944.1, YP_001860597.1, YP_001811255.1, YP_001531662.1, YP_001534185.1, YP_001761364.1, YP_001476305.1, YP_001888022.1, YP_001893789.1, YP_001083198.1, YP_001844795.1, YP_001777918.1, YP_001779189.1, YP_001641952.1, YP_001772211.1, YP_001758390.1, YP_001776814.1, YP_001783295.1, YP_002158590.1, YP_001948920.1, YP_002423669.1, YP_001927659.1, YP_001203094.1, YP_002128524.1, YP_002976728.1, YP_001989358.1, YP_001991324.1, YP_002282165.1, YP_002495630.1, YP_002496260.1, YP_002497058.1, YP_001906897.1, YP_001908005.1, YP_001832057.1, YP_002826208.1, YP_002317565.1, YP_001979200.1, YP_001985290.1, YP_001719546.1, YP_001720402.1, YP_001873009.1, YP_001873806.1, YP_002265246.1, YP_002327281.1, YP_002439140.1, YP_002441565.1, YP_002923740.1, YP_003019698.1, YP_003002473.1, YP_002955226.1, YP_002909043.1, YP_002934276.1, YP_002360442.1, YP_002947663.1, YP_003964946.1, YP_003941574.1, YP_001603070.1, YP_003070966.1, YP_001715479.1, YP_004144716.1, YP_004145051.1, NP_945673.1
LuxR homologues
YP_002649216.1, YP_003261727.1, YP_003262848.1, YP_003296639.1, YP_003331714.1, YP_003608086.1, YP_004230807.1, YP_003366469.1, YP_004012994.1, YP_004106680.1, YP_004106955.1, YP_004108424.1, YP_003910271.1, YP_003520251.1, YP_003530769.1, YP_003538485.1, YP_003546444.1, YP_003558208.1, YP_003566925.1, YP_003568279.1, YP_003576500.1, YP_003729882.1, YP_003930459.1, YP_003734010.1, YP_003749681.1, YP_003750859.1, YP_003744152.1, YP_003740504.1, YP_003740953.1, YP_003847232.1, YP_003885142.1, YP_004088229.1, YP_004115278.1, NP_521406.1, NP_522339.1, NP_767702.1, NP_106261.1, NP_106660.1, NP_109411.1, NP_385944.1, YP_002965846.1, YP_002966880.1, YP_002346032.1, YP_002347421.1, YP_002426403.1, YP_428476.1, YP_105961.1, YP_106160.1, YP_110896.1, YP_111575.1, YP_001005891.1, YP_554691.1, YP_555670.1, YP_165634.1, YP_167510.1, NP_669049.1, NP_670674.1, YP_528967.1, YP_234708.1, NP_250121.1, NP_252167.1, YP_002232873.1, YP_002234479.1, YP_768957.1, YP_048234.1, NP_793635.1, YP_789670.1, YP_791822.1, NP_903760.1, NP_993605.1, NP_994736.1, YP_674864.1, YP_371810.1, YP_001114942.1, YP_001117674.1, YP_439002.1, YP_439706.1, YP_273861.1, YP_508561.1, YP_071012.1, YP_071752.1, YP_206883.1, YP_914594.1, YP_002551488.1, YP_002549361.1, YP_002541325.1, YP_659944.1, YP_317246.1, YP_776003.1, YP_617565.1, YP_617627.1, YP_838351.1, YP_623508.1, YP_470410.1, YP_473056.1, YP_001024423.1, YP_001025820.1, YP_001077903.1, YP_001078154.1, YP_989940.1, YP_001062292.1, YP_001063209.1, YP_335776.1, YP_337635.1, YP_972129.1, YP_484040.1, YP_486928.1, YP_567541.1, YP_569310.1, YP_530593.1, YP_531902.1, YP_781245.1, YP_001231850.1, YP_001604810.1, YP_001606210.1, YP_001399708.1, YP_001400524.1, YP_001220570.1, YP_001241092.1, YP_001242900.1, YP_001075258.1, YP_001076161.1, YP_002537872.1, YP_001327236.1, YP_453965.1, YP_681951.1, YP_650193.1, YP_651866.1, YP_647982.1, YP_649110.1, YP_002220093.1, YP_855090.1, YP_001161917.1, YP_001163230.1, YP_001347033.1, YP_001349253.1, YP_001143472.1, YP_001583946.1, YP_001860599.1, YP_001811253.1, YP_001531661.1, YP_001534186.1, YP_001761363.1, YP_001476304.1, YP_001888024.1, YP_001893790.1, YP_001083200.1, YP_001844797.1, YP_001777917.1, YP_001779191.1, YP_001641953.1, YP_001772212.1, YP_001758389.1, YP_001776815.1, YP_001783296.1, YP_002158591.1, YP_001948918.1, YP_002423670.1, YP_001927660.1, YP_001203095.1, YP_002128523.1, YP_002976727.1, YP_001989359.1, YP_001991323.1, YP_002282164.1, YP_002495629.1, YP_002496262.1, YP_002497059.1, YP_001906896.1, YP_001908006.1, YP_001832058.1, YP_002826207.1, YP_002317567.1, YP_001979199.1, YP_001985289.1, YP_001719545.1, YP_001720401.1, YP_001873010.1, YP_001873807.1, YP_002265247.1, YP_002327279.1, YP_002439139.1, YP_002441567.1,

YP_002923741.1, YP_003019697.1, YP_003002472.1, YP_002955225.1, YP_002909041.1, YP_002934275.1, YP_002360441.1, YP_002947664.1, YP_003964947.1, YP_003941575.1, YP_001603072.1, YP_003070967.1, YP_001715477.1, YP_004144717.1, YP_004145052.1, NP_945674.1,
RsaL homologue
NP_250122.1, YP_001349252.1, YP_001860598.1, YP_001888023.1, YP_002441566.1, YP_002794907.1, YP_003608087.1, YP_003847233.1, YP_003910270.1, YP_554692.1, YP_791821.1
RsaM homologues:
YP_439707.1, YP_001062653.1, YP_776004.1, YP_001117675.1

These core set of protein sequences were retrieved from functionally annotated sequences (SBASE, (Vlahovicek et al. 2005)), the Cluster of Orthologous Groups database COG (Tatusov et al. 2003), and the UNIPROT clusters (Apweiler et al. 2004). Bioinformatics programs like BLAST (Altschul et al. 1990), Smith-Waterman (Smith & Waterman 1981) were also used to collect homologs of these proteins, using the bacterial genome section of the NCBI collection. The CLUSTAL program (Thompson et al. 1994) (accessed via the EBI Web portal, http://www.ebi.ac.uk/Tools/msa/clustalw2/) was used for constructing the multiple sequence alignments which were then processed by the HMMBUILD (Eddy 1998) program to build HMM recognizers. These HMM recognizers were in turn used to scan the protein sequence data. HMMER aims to be significantly *more* accurate and *more* able to detect remote homologs because of the strength of its underlying mathematical models. Hits stronger than E- value of $1e^{-10}$ were taken as potential homologues of QS genes. We observed that hits stronger than E-value 10^{-10} were consistent with data found in NCBI for fully annotated genomes. Prediction of QS genes in draft genomes were a bit tricky. Some draft genomes have no annotation. Potential ORFs of such draft genomes were predicted by Glimmer (Delcher et al. 1999) and the translated ORFs were then subjected to HMMsearch.

As the protein encoded by *luxR* and *luxI* have spurious similarities to a number of different protein families, we manually checked the length, sequence coverage and residue conservation of each similarity.

2.2.3 Local arrangement or topology of QS genes:

The QS-genes were then analyzed with an automated process to reveal the recurring arrangement patterns of QS genes with respect to one another to denote whether the genes are convergent, divergent, synthase upstream, receptor upstream, or synthase and receptor separated by other genes, *etc*. This was cross verified by manual check. Gene topology is a broad term that can include the arrangement of genes within chromosomes, with respect to the replication origin or other chromosomal elements. In this work the words "topological arrangement" or briefly "topology" was used to denote the arrangement within a close neighborhood of the QS regulatory genes. The distance threshold for the two genes to be in a topology was 3,000 bp. PROSITE-like syntax was used (Falquet et al. 2002) to denote a topology. The *luxR, luxI, rsaL* and *rsaM* genes are abbreviated as *R, I, L* and *M*, respectively, and X is used for all other genes. An arrow above each gene symbol then shows the direction of transcription. With this notation, for example $\vec{R}\vec{I}$ denotes adjacent *luxR* and *luxI* genes transcribed in the same direction. $\vec{R}X(>5)\vec{I}$ denotes *luxR* and *luxI* genes transcribed in the same direction with more than five genes between the *luxR* and *luxI*, without specifying the direction of transcription of the X gene. The process of QS detection and defining local topology was automatised by using Galaxy workflow system hosted through local server. Tools needed for QS detection and determining topology were integrated into Galaxy. Workflow was built from several linked steps that consumed inputs, processed and converted data and produced desired result i.e. local topology of QS genes. Figure 16 below shows the workflow designed for QS gene detection and determining local topology and desired output obtained from Galaxy.

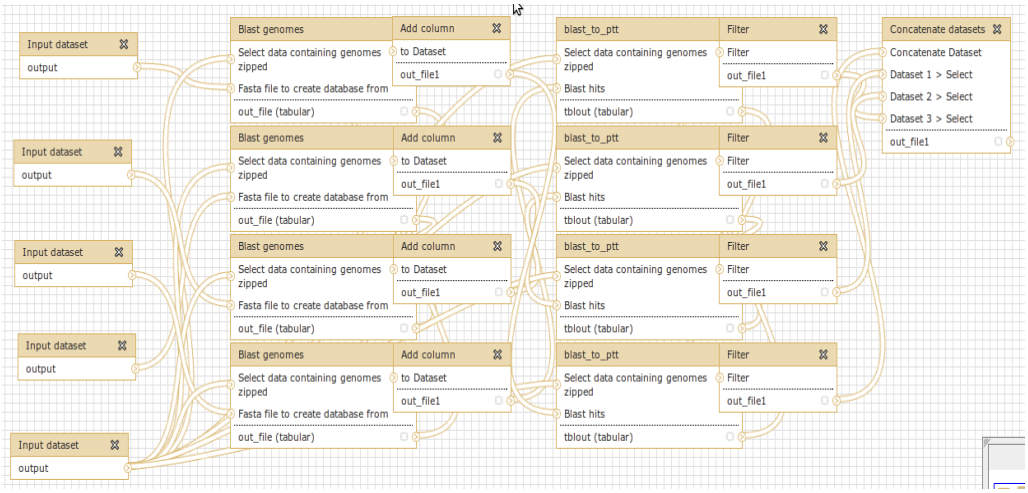


Figure 16: Workflow

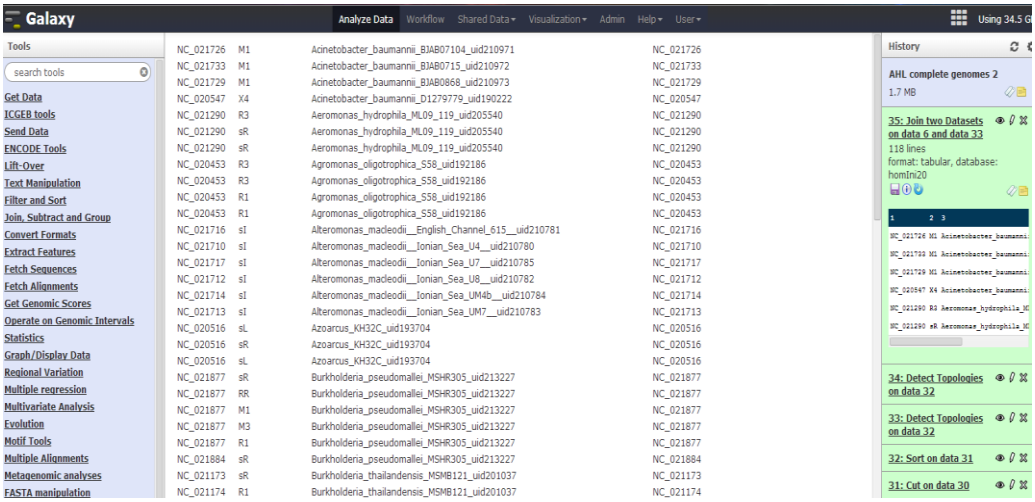


Figure 17: Desired Result

2.2.4 Chromosomal arrangement of QS genes:

Chromosomal topology is the term used to define topological patterns of QS genes with respect to OriC. In order to depict the Chromosomal Topology visually, a Python script was written. This script uses topologies and their coordinates in the genome as input parameters and generates circular genome diagrams. The script scales the coordinates to 2π degrees, in order to

draw into a circle. In addition to co-ordinates of QS genes, the coordinates of the OriC were included to serve as a reference point in terms of visualizing the distance of topologies.

OriC was calculated from GenSkew online software tool (http://genskew.csb.univie.ac.at/) which computes and plots nucleotide skew data. The software GenSkew computes the normal and the cumulative skew of two selectable nucleotides for a given sequence. The result is displayed in two different graphs. The user can choose the pair of nucleotides the skew should be calculated for. In such a skew, the relation of the nucleotides is given by a formula. The particular values are computed by the formula:

Skew = (nucleotide1 - nucleotide2) / (nucleotide1 + nucleotide2).

For the computation the genome is separated into parts, so called "windows". For every window the value of the skew is calculated with the formula. The normal graph displays every single value related to the particular position in sequence. The cumulative graph adds up the values for all previous windows up to the certain position. The windows are shifted along the sequence by a certain stepsize. To improve the result the stepsize can be reduced. The global minimum and maximum are displayed in the cumulative graph. The minimum and maximum of a GC-skew can be used to predict the origin of replication (minimum) and the terminus location (maximum) in prokaryotic genomes respectively.

Since the coordinates of *luxI*, *luxR*, *rsaL* and *rsaM* are very near to each other and are difficult to distinguish; therefore, coordinates of *luxR* genes were chosen in each case.

2.2.5 Multiple sequence alignment and Tree display:

Multiple sequence alignment for the selected proteins were performed by CLUSTAL program (accessed via the EBI Webportal, http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Thompson et al. 1994). CLUSTAL calculates trees using the neighbour-joining method. Trees were

visualized using the MEGA5 program package installed from http://www.megasoftware.net (Tamura et al. 2011).

2.2.6 Neighborhood detection:

Neighborhood genes provide useful information on evolutionary functions of QS genes. Also, in prokaryotes, genes with related functions are close to each other either as part of operon or functional neighborhood. These form gene clusters and are often conserved in prokaryotes. Checking the conservation of neighborhoods can further help in functional annotation of QS genes. Gene neighborhoods were checked by Microbial Genome Viewer (http://mgv2.cmbi.ru.nl/genome/index.html) (Kerkhoven et al. 2004). Microbial genome Viewer (MGV) allows users to visualize the genomic context of any set of genes/proteins in a comparative fashion. Within MGV2, such comparative views are called Custom Comparative Maps. Currently, sets of proteins/genes can be supplied in three ways; by uploading a file containing a list of NCBI GI codes, by uploading a list of locus tags or by uploading a phylogenetic tree in newick format (note: the nodes of the tree must contain NCBI GI codes). Figure 18 shows example file format needed for genome context visualization.

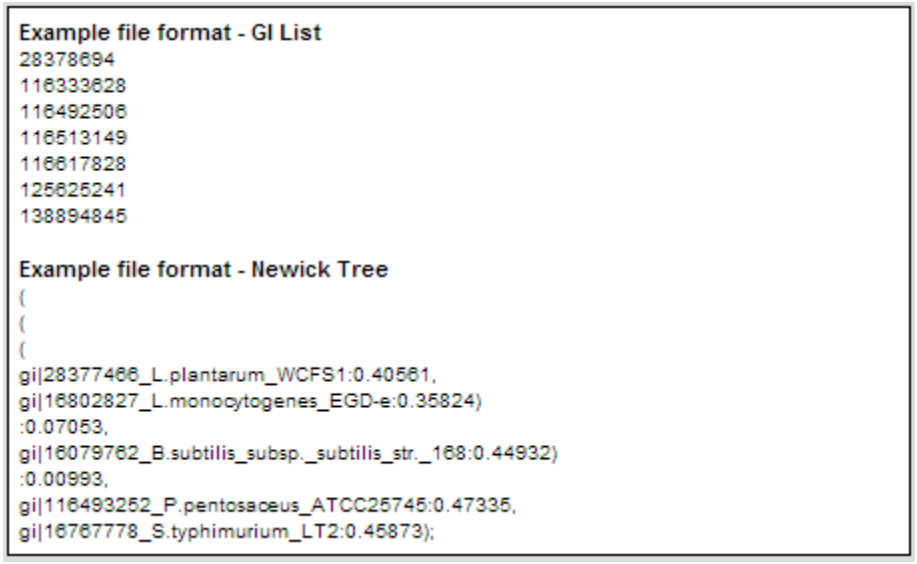


Figure 18: Examples of file formats

The concurrent visualization of a great diversity of gene-related data such as COG categories, PFAM domains, regulatory elements, sub-cellular location and transcriptomic expression ensures a high information density.

Figure 19 shows an example of output by MGV.

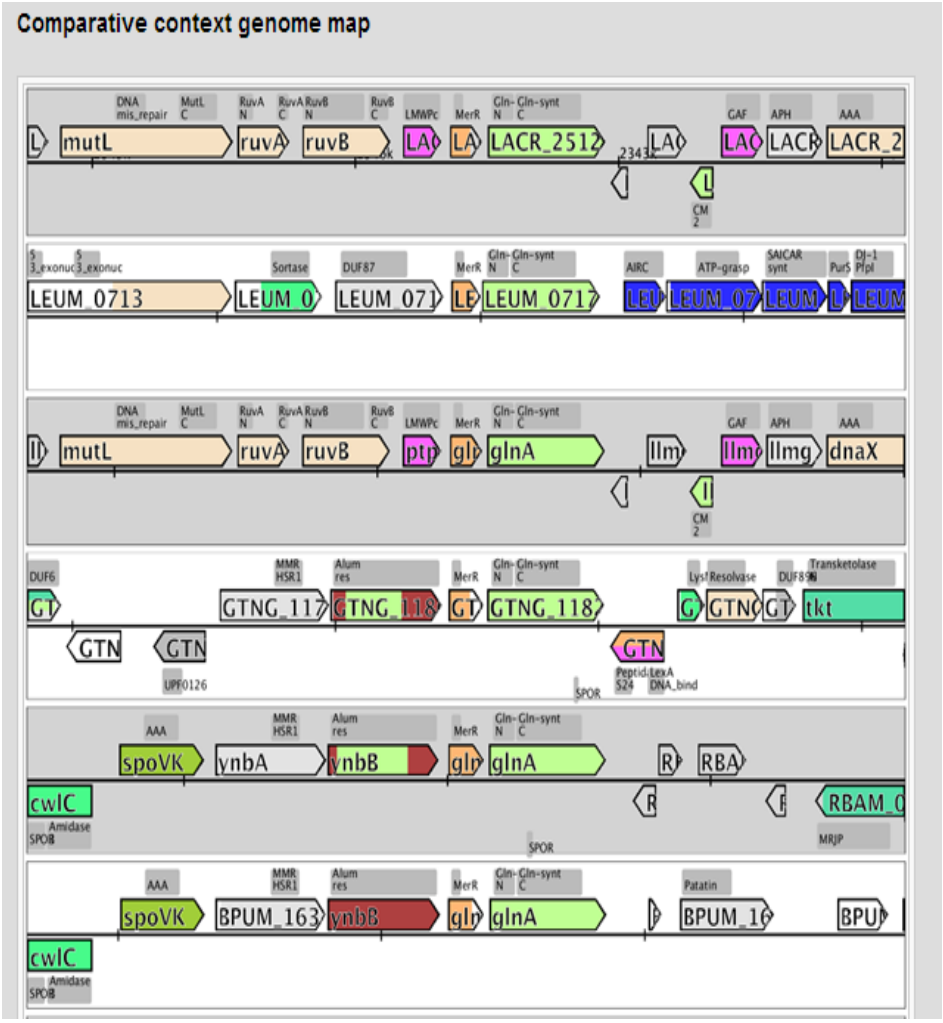


Figure 19: Example of a comparative context map. Visualization of the gene context of the glnR transcription regulator in 11 different species.

Images are generated in scalable vector graphics (SVG) format, which is suitable for creating high-quality scalable images and dynamic Web representations. All images can also be exported to PNG-format (Kerkhoven et al. 2004).

Figure 20 shows a simplified workflow summarizing the methods used above.

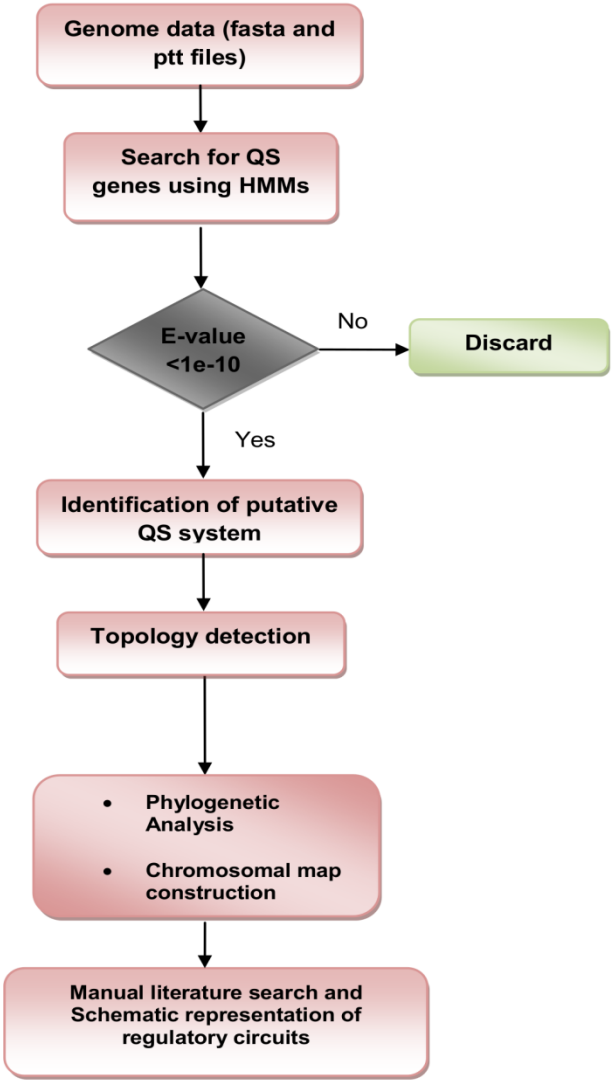


Figure 20: Workflow pipeline

3 AHL-mediated QS systems in Proteobacteria

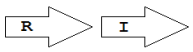
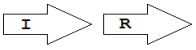
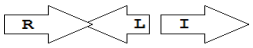
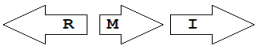
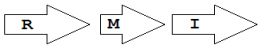
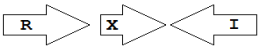
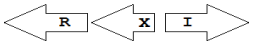
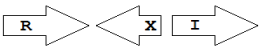
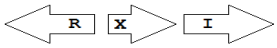
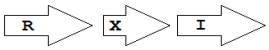
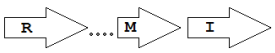
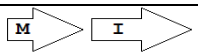
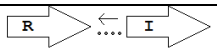
1403 complete bacterial genomes having around 4.8 million genes were analyzed for the presence of QS genes (NCBI databases downloaded in August 2012). This analysis was based on the reading frames and showed that 226 complete genomes contained QS genes that were close to other QS genes (*i.e.*, according to a distance threshold of 3,000 bp). All these belong to proteobacterial genomes. Of the 4.8 million genes we analyzed, 674 are *luxR* genes (unannotated: 33), 294 are *luxI* genes (unannotated: 13), 44 are rsa*L* genes (unannotated: 16) and 37 are rsa*M* (all were unannotated). The results of our search have been deposited to an open online repository from where the arrangements and the individual sequences can be retrieved, (http://bacteria.itk.ppke.hu/QStopologies/) (Zsolt Gelencsér, Choudhary, et al. 2012).

3.1 Results and Discussion

3.1.1 Local Topological types in Proteobacteria

The gene arrangement patterns were classified into two large groups: a.) simple topologies and b.) complex topologies. Simple topologies consist of *luxI/luxR* genes that are either vicinal or are separated by a single gene. Such arrangements are characterized by typical patterns of transcriptional orientation that are conserved in many proteobacteria. On the other hand, complex topologies consist of *luxR/luxI* genes separated by a larger and more variable number of intervening genes. These topologies are characteristic of *Agrobacterium* and *Rhizobium* species. Together, these topologies were grouped into 16 distinct categories. Table 3 below shows the summary of the topological categories found in the survey.

Table 3: Typical topological patterns found in complete bacterial genomes

ID	Pattern	Gene topology	Occurrence in complete Proteobacterial genomes				
			Total	alpha	beta	gamma	delta
Simple Topologies							
R1	$\vec{R}\vec{I}$		101	71	14	16	0
R2	$\vec{R}\overleftarrow{I}$		53	2	2	46	3
R3	$\overleftarrow{R}\vec{I}$		11	1	3	7	0
R4	$\vec{I}\vec{R}$		2	2	0	0	0
L1	$\vec{R}\overleftarrow{L}\vec{I}$		16	0	7	9	0
M1	$\overleftarrow{R}\vec{M}\vec{I}$		30	0	20	10	0
M2	$\vec{R}\vec{M}\vec{I}$		1	0	1	0	0
X1	$\vec{R}\vec{X}\overleftarrow{I}$		1	0	0	1	0
X2	$\overleftarrow{R}\overleftarrow{X}\vec{I}$		2	2	0	0	0
X3	$\vec{R}\overleftarrow{X}\vec{I}$		4	0	2	2	0
X4	$\overleftarrow{R}\vec{X}\vec{I}$		1	1	0	0	0
X5	$\vec{R}\vec{X}\vec{I}$		2	1	1	0	0
Complex Topologies							
M3	$\vec{R}X(2-11)\vec{M}\vec{I}$		6	0	6	0	0
M4	$\vec{R}\vec{M}X(<7)\vec{I}$		1	0	1	0	0
M31	$\vec{M}\vec{I}$		2	0	2	0	0
X6	$\vec{R}\overleftarrow{X}(7)\vec{I}$		1	1	0	0	0
X7	$\vec{I}X(>7)\vec{R}$		4	4	0	0	0

Within simple topologies, the majority of cases are tandem $\vec{R}\vec{I}$ topologies and convergent $\vec{R}\overleftarrow{I}$ topologies. A sub-group of simple topologies (Table 3) had a

single intervening gene between *luxR* and *luxI*. Table 4 below shows the intervening genes in simple topology and their probable function along with the organisms in which they are found. It is apparent that the intervening genes have a few well determined functions.

Table 4: Intervening genes in Simple Topologies

Gene-types	**No. found in complete bacterial genomes**	**Potential role**	**Example of genomes**
rsaL	11	Negative regulator	*P. putida, P. fuscovaginae, P. aeruginosa* LESB58, *P. aeruginosa* PAO1
rsaM	29	Negative regulator	*B. pseudomallei* K96243
mupX	1	Negative regulator	*P. fluorescens* NCIMB 10586
Integrases/ transposases	2	DNA mobilization	*B. vietnamiensis* G4 *Methylobacterium nodulans* ORS 2060
LuxR like regulator	1	Unknown	*Gluconacetobacter diazotrophicus* PAl 5
Unknown function(X)	5	Unknown	*B. mallei* NCTC 10247, *Saccharophagus degradans* 2-40

Out of 48 such single intervening genes, 11 code for RsaL and 29 code for RsaM proteins. RsaL and RsaM both are known to negatively regulate quorum sensing. RsaL (which is predominant in pseudomonad genomes (de Kievit et al. 1999)) in *P. aeruginosa* prevents expression of the *R* gene by binding to DNA next to the *lux*-box (Mattiuzzo et al. 2011). Similarly, RsaM has been reported to negatively regulate QS in *P. fuscovaginae* (Mattiuzzo et al. 2011). *rsaL* (L in our notation) form $\vec{R}\vec{L}\vec{I}$ and *rsaM* (M in our notation) form mainly $\vec{R}\vec{M}\vec{I}$ arrangement (M1), but also appear in M1, M2 and M3; Of the rest of the intervening genes, X genes found in RXI topology, *P. fluorescens* NCIMB10586 was seen to have *mupX*, which is an amidase-hydrolase, and is able to digest/degrade the AHL signal of the same species (El-Sayed et al.

2001). MupX, can therefore also be considered as a negative QS regulator. Apart from MupX, two of the X genes are involved in DNA-mobilization (an integrase and a transposase), and the rest are hypothetical proteins of unknown function.

In contrast to simple topologies, the group of longer, complex topological patterns showed greater variety although the number of occurrences were considerably less. The groups of *Agrobacterium* and various *Rhizobia* often contain complex topologies. An interesting example in this group was the MI pattern of *B. ambifaria*, where two well distinguishable QS genes (*rsaM* and *luxI*) appear in their usual tandem topology, but without an annotated or computationally identifiable *luxR* homolog in their vicinity.

3.1.2 Taxonomic Distribution of Patterns of QS gene

Our analysis showed that tandem $\vec{R}\vec{I}$ topology was more common in α-proteobacteria, the convergent $\vec{R}\overleftarrow{I}$ pattern was more evident in γ-proteobacteria, while the $\vec{R}\overleftarrow{L}\vec{I}$ pattern and the $\overleftarrow{R}\vec{M}\vec{I}$ pattern was found apparently only in β and γ proteobacteria and not in α-proteobacteria. This suggested that the arrangement patterns of QS genes are separated among the taxonomic classes. The question arised whether the known QS proteins, such as LuxI and LuxR, cluster simply according to the known taxonomy or according to the topological pattern. Cladograms of 154 LuxI and LuxR protein sequences showed a clear tendency that the proteins cluster according to the topological patterns. The $\vec{R}\vec{I}$, $\vec{R}\overleftarrow{I}$, $\vec{R}\overleftarrow{L}\vec{I}$ and $\overleftarrow{R}\vec{M}\vec{I}$ patterns form clearly distinguishable groups both in the LuxR and in the LuxI trees. Also, it was very rare that proteins from same genomes formed close neighborhoods. In more detail, of the LuxI-type proteins that occur in genomes with multiple QS circuits, 38% had a nearest sequence similarity neighbor from the same family, 35% from the same genus, 26% from the same species, and approximately 1% from the same genome. The frequency with which *luxR* genes occur was similar: 37%, 36%, 27%, and 1%, respectively. This indicates that the patterns may have formed before the modern strains diverged from each other.

An example in Figure 21 shows that LuxI proteins present in $\vec{R}\vec{L}\vec{I}$ patterns clearly separate from those present in $\vec{R}\vec{I}$ patterns within the same genome and cluster together with the respective genes of $\vec{R}\vec{L}\vec{I}$ of another (β or γ) class (Figure 21(A)). At the same time, the clustering of the RsaL proteins (Figure 21(B)) is identical to the clustering of their accompanying LuxI genes (Figure 21(A)).

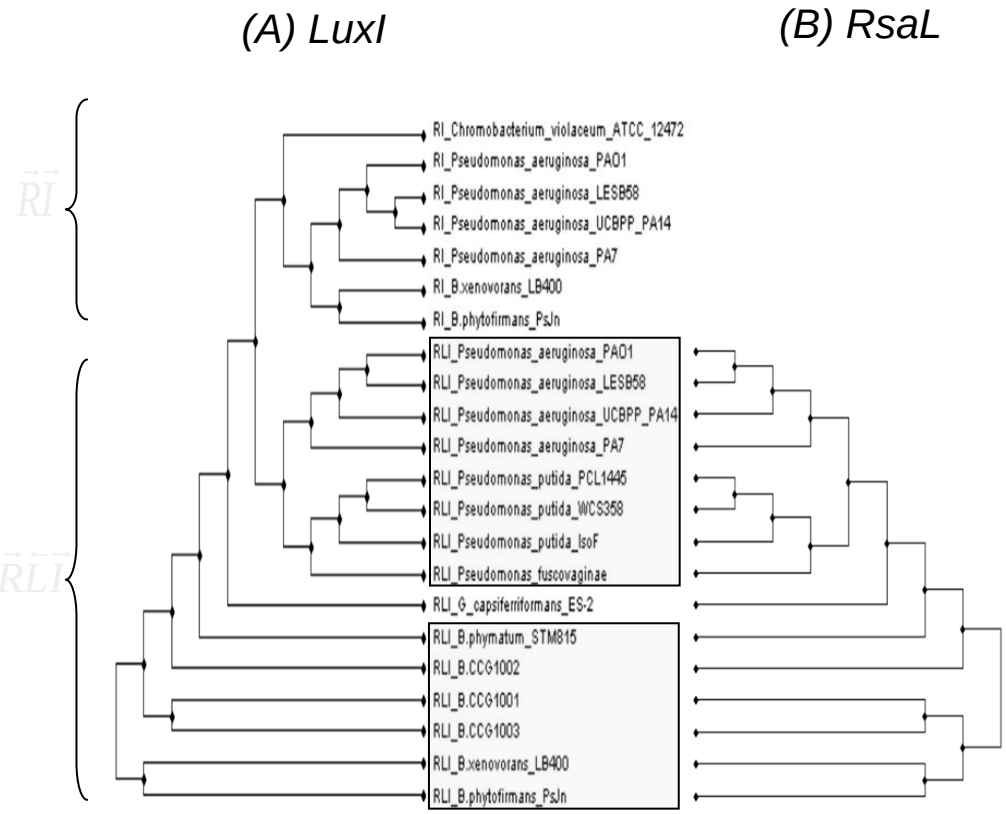


Figure 21: Clustering of LuxI and their negative regulators RsaL
(A) Shows tree of LuxI proteins present RI and RLI pattern, and (B) shows RsaL proteins in RLI pattern. The sequences were taken from genomes harboring rsaL genes located between *luxR* and *luxI* genes. The clustering was carried out by the Phylip program package (Ropelewski et al. 2010)**. The numerical value at each node indicates the bootstrap value supporting every split in the lineage (out of 1,000 bootstrap replicates).**

In other words, QS proteins seem to cluster according to gene topology at various taxonomic levels, which suggests that the *luxR*, *luxI* as well as the

intervening genes may have evolved together. The clustering of QS genes suggests that the topological units might act as regulatory modules that evolve together.

3.2 **Summary**

It can be concluded that there are more diverse topological patterns in QS genes than previously indicated (Hirakawa et al. 2011; Ahlgren et al. 2011). It is apparent that QS proteins with similar topological pattern cluster together at various taxonomic levels, which suggests that the *luxR*, *luxI* as well as the intervening genes may have evolved together. In other words, the QS genes within a distinct topology type appear to be orthologs that diverged from the rest of their homologs before the modern bacterial species diverged from each other.

This work was a preliminary, computational census of topological patterns found in complete proteobacterial genomes. The annotated topologies are deposited online with links to sequences and genome annotations at http://bacteria.itk.ppke.hu/QStopologies/.

4 AHL-mediated QS systems in *Pseudomonas*

Out of 226 proteobacterial complete genomes and 426 draft genomes which contained QS genes, we selected the *Pseudomonas* genomes as the first detailed case study. Out of 55 complete and 164 draft genomes of *Pseudomonas,*13 and 45 genomes respectively contained defined QS topologies. A few selected Genbank entries were also investigated for the presence of AHL QS genes.

4.1 Results and Discussion

4.1.1 Local Topology types in *Pseudomonas*

Out of 16 distinct topology types found in Proteobacteria, *Pseudomonas* had only six (Table 5 below), all belonging to simple gene arrangements. More specifically, the QS genes found in *Pseudomonas* are arranged either like RI, where the two genes are vicinal or like *RXI,* which has at least one additional gene between the two *LuxI* and *LuxR* family genes (Zsolt Gelencsér, Galbáts, et al. 2012).

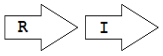
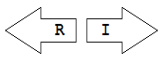
The RI Topology. This kind of arrangement is either be tandem (unidirectional), convergent or divergent. *Pseudomonas* contains only tandem and convergent topology.

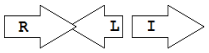
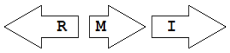
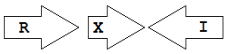
The RXI Topology. In these topologies, one gene is found between the *R* and the *I* genes. The *X* gene in *Pseudomonas* were most frequently *L* (RsaL, found in *P. aeruginosa*, *P. putida,* and *P. fuscovaginae* species). RsaL was shown to belong to the tetrahelical superclass of H-T-H proteins (Rampioni et al. 2007). Members of this family are widespread repressors in bacteria and bind to DNA as dimers. It was found that homologues of *RsaL* frequently occur outside QS circuits in various bacterial genomes (data not shown). In *P. fuscovaginae,* RsaL binds to DNA next to the lux box and prevents the expression of the *R* gene (Mattiuzzo et al. 2011). In contrast, *RsaM* (*M*) is a protein of unknown

structure that seems to occur only in the context of QS circuits. In *P. fuscovaginae M* was found to negatively regulate QS (Mattiuzzo et al. 2011). *P. fuscovaginae* is the only *Pseudomonas* which was found to contain *M* gene, which is at the same time, the only pseudomonad found so far to contain both *L* and *M* genes. *P. fluorescens* NCIMB 10586 contains a gene coding for an enzyme *mupX* in the X position that was shown to degrade the AHL signal produced by the same species thereby decreasing the QS response (Hothersall et al. 2011).

The arrangement of *RI* and *RXI* topologies are shown in the Table 5.

Table 5: Typical chromosomal arrangements of AHL-driven quorum sensing circuits in *Pseudomonas* (complete and Draft genomes)

ID	Gene topology	Occurrence in *Pseudomonas*
R1		*P. aeruginosa* (Strains: LESB58, PA7, PAO1, UCBPP-PA14, M18, DK2, NCGM2, A506, 152504,2192, 39016, ATCC 14886, ATCC 25324, ATCC 700888, CI27, CIG1, E2, MPAO1 P1, MPAO1 P2, NCMG1179, PAb1, PACS2, PADK2, PAO579) *P. chlororaphis* (Strains: sp. *aureofaciens* 30 84, O6) *Pseudomonas* sp. GM17, *Pseudomonas* sp. M47T1
R2		*P. syringae* (Strains: pv. *phaseolicola* 1448A, pv. *syringae* B728a, pv. *tomato* str. DC3000, *aceris* M302273, *aesculi* 0893 23, *aesculi* 2250, *aesculi* NCPPB3681, *avellanae* ISPaVe013, *avellanae* ISPaVe037, *glycinea* B076, *glycinea* race 4, *lachrymans* M301315, *mori* 301020, *morsprunorum* M302280, *tabaci* ATCC 11528, *tomato* K40, *tomato* Max13, *tomato* NCPPB 1108) *P. chlororaphis* (Strains: PCL1391, *aureofaciens* 30 84, O6), *P. fluorescens* 2-79, *Pseudomonas* sp. GM17, *P. savastanoi* NCPPB 3335
R3		*Pseudomonas* sp. GM18,*Pseudomonas* sp. GM30

L1		*P. aeruginosa* (Strains: LESB58, PA7, PAO1, UCBPP-PA14, M18, PAO579, PADK2 CF510, PAb1, NCMG1179, MPAO1_P2, MPAO1_P1,E2, CIG1, CI27, C3719, ATCC_700888, ATCC_25324, ATCC_14886, 39016, 2192, 152504, 138244, 2_1_26), *P. putida* (Strains: WCS358, IsoF, PCL1445) *P. fuscovaginae* UPB0736
M1		*P. fuscovaginae* UPB0736
X1		*P. fluorescens* NCIMB 10586

4.1.2 Chemical Signals produced

We mined publicly available databases and literatures for AHL chemical signals produced by *Pseudomonas* species. Strains of *P. aeruginosa* have two systems: las system and rhl system. In the las system, the *lasI* gene product directs the synthesis of N-(3-oxo-dodecanoyl)-L-homoserine lactone (3-oxo-C12-AHL), which interacts with LasR and activates target promoters. In the rhl system, rhlI directs the synthesis of N-(butanoyl)-L-homoserine lactone (C4-AHL), which interacts with the cognate regulator RhlR and activates target gene promoters (Venturi 2006). *Pseudomonas syringae* strain B728a synthesizes, N-(3-oxo-hexanoyl)-L-homoserine lactone (3-oxo-C6-AHL) through ahlI, which then interacts with AhlR. PhzI of rhizosphere-colonizing *P. chlororaphis* strain PCL1391 directs the synthesis of N-hexanoyl-L-homoserine lactone (C6-AHL), which interacts with cognate regulator PhzR. luxI homologs of strains of *P. putida*, ppuI produces 3-oxo-C12-HSL. However, there are strains of *Pseudomonas* in which no AHL is detected, like in *P. fluorescens* NCIMB 10586 (Laue et al. 2000) and *P. syringae* pv. *phaseolicola* (Dumenyo & Mukherjee 1998).

Table 6 shows the AHL driven QS circuits in few of the *Pseudomonas* species and the signals produced.

Table 6: Examples of *Pseudomonas* species with of AHL-driven quorum sensing networks

***Pseudomonas* species**	**QS circuit**	**Chemical Signal Produced**	**Pattern id (from Table 1)**
P. aeruginosa LESB58 (Winstanley et al. 2009)	rhlR/rhlI	C4-HSL	R1
	lasR/rsaL/lasI	3OC12-HSL	L1
P. aeruginosa PA7 (Roy et al. 2010)	rhlR/rhlI	C4-HSL	R1
	lasR/rsaL/lasI	3OC12-HSL	L1
P. aeruginosa PAO1 (Stover et al. 2000)	rhlR/rhlI	C4-HSL	R1
	lasR/rsaL/lasI	3OC12-HSL	L1
P. aeruginosa UCBPP-PA14 (Lee et al. 2006)	rhlR/rhlI	C4-HSL	R1
	lasR/rsaL/lasI	3OC12-HSL	L1
P. fuscovaginae UPB0736 (Mattiuzzo et al. 2011)	sR/rsaM/sI	C10-HSL, C12-HSL	M1
	vR/rsaL/vI	3-OC10-HSL, 3-OC12-HSL	L1
P. syringae pv. *phaseolicola* 1448A (Joardar et al. 2005)	AhlR/AhlI	No AHL detected	R2
P. syringae pv. *syringae* B728a (Feil et al. 2005)	psyr_1622/psyr_1621	3OC6-HSL	R2
P. syringae pv. *tomato* str. DC3000 (Buell et al. 2003)	psyR/psyI	-	R2
P. chlororaphis PCL1391 (Khan et al. 2007; Wood & Pierson 1996)	phzR/PhzI	C6-HSL	R2
P. fluorescens 2-79 (Mavrodi et al. 1998)	phzR/PhzI	3-OH-C6-HSL, 3-OH-C8-HSL, 3-OH-C10-HSL	R2
P. fluorescens NCIMB 10586 (El-Sayed et al. 2003)	mupR/mupX/mupI	No AHL detected	X1
P. putida WCS358 (Bertani & Venturi	uR/rsaL/uI	3OC12-HSL	L1

2004)			
P. putida IsoF (Steidle et al. 2002)	ppuR/rsaL/ppuI	3OC12-HSL	L1
P. putida PCL1445 (Dubern et al. 2006)	ppuR/rsaL/ppuI	3OC12-HSL	L1

4.1.3 Gene Overlaps

Conserved gene overlaps were seen to be very common in simple topologies. In our analysis, two topologies are found to contain this feature, namely $\vec{R}\overleftarrow{L}\vec{I}$ (L1) and the $\vec{R}\overleftarrow{I}$ (R2). The gene overlap pattern is shown in Figure 22.

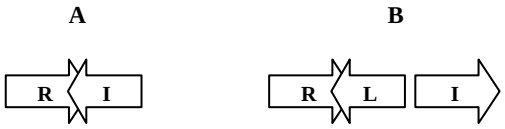


Figure 22: Gene Overlap patterns in the two topology types A.) R2 and B.) L1

The L1 type QS circuits contained an overlap between convergently transcribed *luxR* and *rsaL* gene. The length of the overlap was seen to vary in different species. The R2 type of arrangements in *P. syringae* contained overlapping *R* and *I* genes (2 to 68 bp) while *P. fluorescens* genomes was seen to have no overlaps in QS genes. The L1 type QS circuit of *P. aeruginosa* contained an overlap of 10 bp and the same overlap was 20 bp long in *P. fuscovaginae*. In contrast, *P. putida* had a L1 circuit where the R and L are close (4 bp apart) but not overlapping (Zsolt Gelencsér, Galbáts, et al. 2012). Tsai and Winans noted that the overlapping R2-like arrangement is common to QS circuits in which *R* proteins are able to fold, dimerize, bind DNA, and regulate transcription in the absence of AHLs; moreover, these proteins are antagonized by their cognate AHLs (Tsai & Winans 2010). It was also suggested by some authors that the expression of one member of a convergent or overlapping gene pair might antagonize the expression of the second

member resulting in activation or repression of different functions or phenotypes (Ahlgren et al. 2011).

4.1.4 Evolution of QS genes in *Pseudomonas*

The cladograms of LuxR and LuxI proteins showed a seemingly complex classification scheme which can however be explained by the local topology of their AHL systems. To be precise, the I and R genes seemed to cluster according to their local topology, rather than according to the species. For instance, an I gene with a given topology within a species was seen to be more similar to an I gene of identical topology within another species, than to an I gene of the same species, but having a different local topology. It thus appeared that local topology within the chromosome may have evolved before the separation of various *Pseudomonas* clades (Figure 23). In other words, we see a similar tendency as before, namely the QS genes within a distinct topology type appear to be orthologs that diverged from the rest of their homologs before the modern bacterial species diverged from each other.

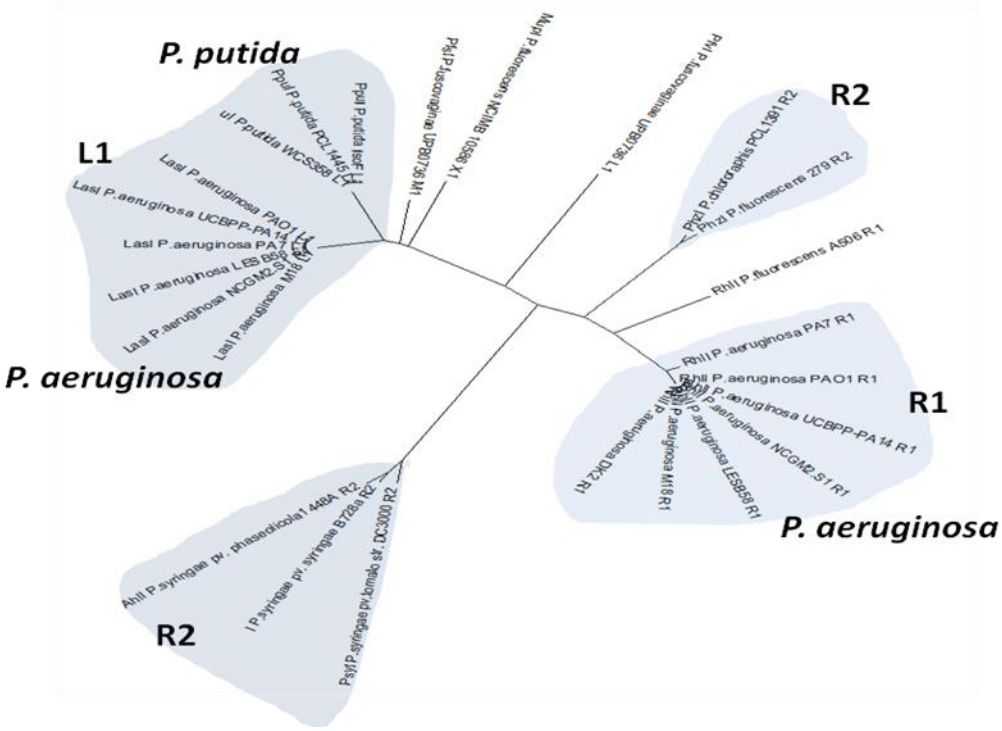


Figure 23: Cladogram of LuxI proteins present in *Pseudomonas* genomes

4.2 **Summary**

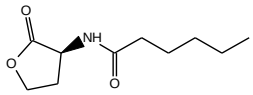
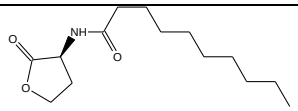
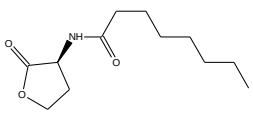
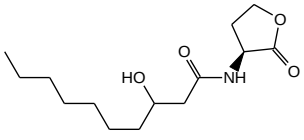
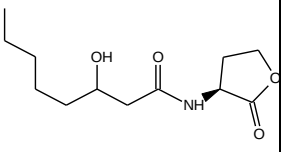
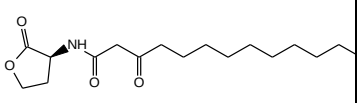
We observed that the chromosomal arrangements found in QS genes seem more varied than expected so the search for common regulatory principles remains an important task for future research. It was hypothesized that the negative regulators of QS that are often found between the canonical *luxR*/ and *luxI*-family genes may be crucial for stabilizing the output of QS circuits. It was also observed that the I and R genes seemed to cluster according to their local topology (which strictly correlates with the chemical nature of the signal), rather than according to the species suggesting that the local topology within the chromosome may have evolved before the separation of various *Pseudomonas* clades.

5 Arrangement of QS circuits in *Burkholderia*

The second detailed case study is the genus *Burkholderia*, another widely studied organisms for its AHL QS circuitry (Eberl 2006). A special characteristics of this genus is the fact that it contains both pathogenic and beneficial species groups which makes it an interesting group to study.

Based on 16S rRNA analysis, the genus *Burkholderia* is classified into three main taxonomical clusters/clades, *i.e*. the *B. cepacia* complex (BCC), the *B. pseudomallei* group (consisting of *B. mallei* and *B. pseudomallei*) and the recently defined plant-beneficial (PBE) group (Suárez-Moreno et al. 2012). Of the near 20 AHL signals described so far (mainly C4 to C18-homoserine lactones (HSL) as well as HSLs with the C3 position substituted or unsubstituted by an oxo, or hydroxyl group), *Burkholderia* responds to six types of AHLs (Table below). Information regarding the chemical signals produced by LuxI homologs of *Burkholderia* was collected by available literatures (Table 7).

Table 7: Chemical structure of N-acylated homoserine lactone (AHL) signals used by LuxR in the known species of the genus *Burkholderia*

Symbol	Structure	Symbol	Structure
N-hexanoyl-*L*-Homoserine lactone**(C6-HSL)**		*N*-decanoyl-*L*-Homoserine lactone (**C10-HSL**)	
N-octanoyl-*L*-Homoserine lactone (**C8-HSL**)		*N*-3-hydroxydecanoyl Homoserine lactone **(OHC10-HSL)**	
N-3-hydroxy-octanoyl Homoserine lactone **(OHC8-HSL)**		*N*-3-oxo-tetradecanoyl-*L*-homoserine lactone (**OC14-HSL**)	

5.1 Results and Discussion

5.1.1 Local Topology types in *Burkholderia*

Out of the 16 local topological arrangements described in proteobacteria, 10 were seen to occur in *Burkholderia* (comprising of completely sequenced genomes, draft genomes and individual Genbank entries), including a new topology type named here as M4 ($\vec{R}\vec{M}X(<7)\vec{I}$), which had previously not been detected in other Proteobacteria. Out of 39 completely sequenced *Burkholderia* genomes, 34 (87%) were found to have AHL QS systems that belonged to eight distinct local topology types. In contrast, out of 44 completely sequenced genomes of *Pseudomonads*, only 11 (25%) had AHL QS genes that belonged to three topology types. Two of the three topology types, R1 ($\vec{R}\vec{I}$) and L1 ($\vec{R}\vec{L}\vec{I}$) are also present in *Burkholderia* (Figure 24).

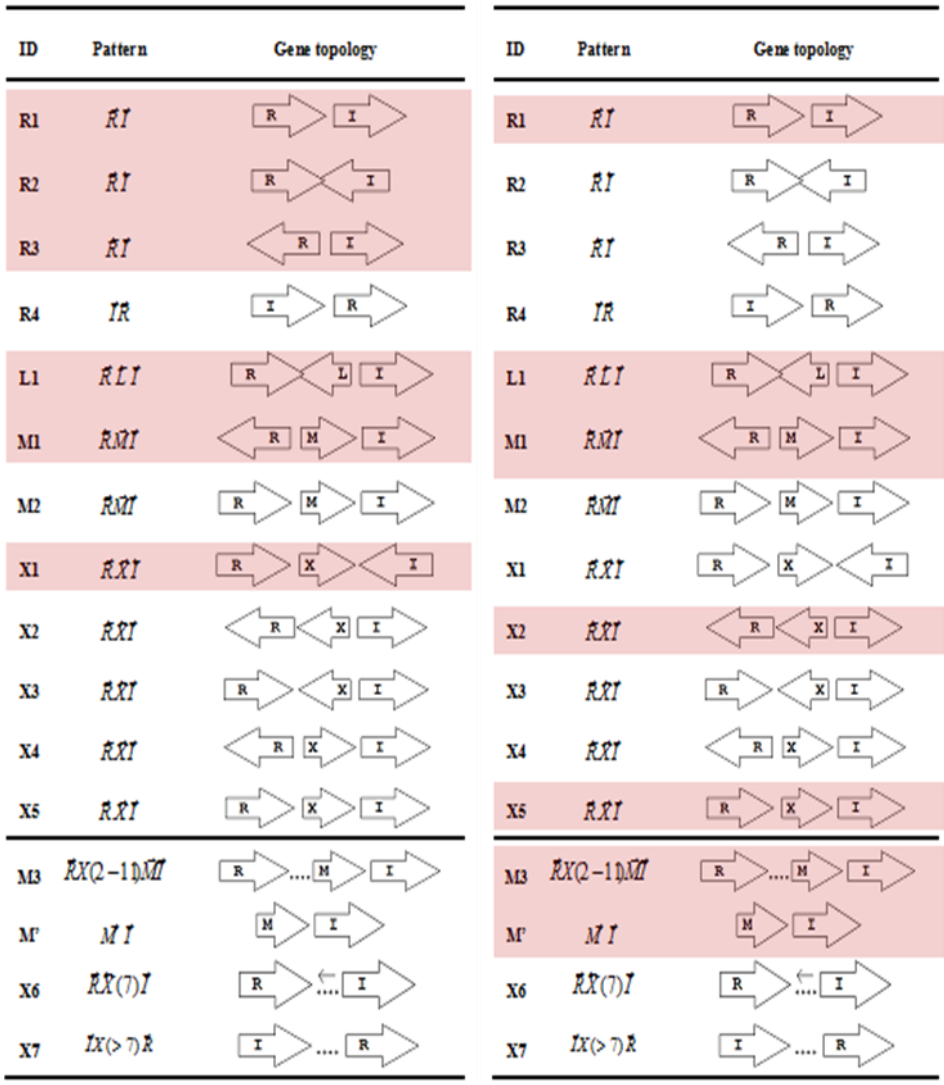
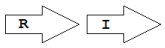
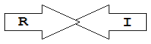
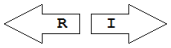
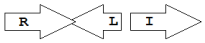
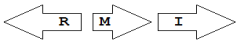
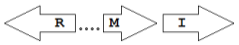
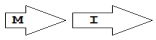
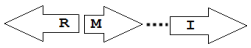
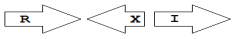
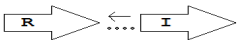


Figure 24: Comparison of Gene topologies present in *Pseudomonas* and *Burkholderia*

While the arrangements of AHL genes of pseudomonads are apparently simpler and have less intervening sequences between the R and I genes, the Burkholderia genus, especially the B. pseudomallei group, has a number of more complex arrangements (Table 8) (Choudhary et al. 2013).

Table 8: Typical local topologies of AHL-driven quorum sensing circuits in *Burkholderia*

ID	Gene topology	Occurrence in *Burkholderia*			
		BCC group	***B. pseudomallei* group**	**Plant-beneficial / environmental group**	**Other pathogenic *Burkholderia* species**
RI topologies					
R1		3	17	4	0
R2		0	0	0	1
R3		8	2	0	3
RXI topologies					
L1		0	0	13	0
M1		24	46	0	3
M3		0	30	0	0
M31		5	5	0	0
M4		1	0	0	0
X3		0	24	1	0
X5		1	0	0	0
X6		0	1	0	0

5.1.2 Evolution of QS genes in *Burkholderia*

Similar to *Pseudomonas*, the cladograms of LuxR and LuxI proteins showed that the I and R genes cluster according to their local topology, rather than according to the taxonomy. For instance, an I gene with a given topology

within a species (e.g., $\vec{R}\vec{I}$ in *B. pseudomallei* 1710b) was seen to be consistently more similar to an I gene of identical topology within another species ($\vec{R}\vec{I}$ in *B. thailandensis* E264), than to an I gene of the same species, but having a different local topology (e.g., $\vec{R}\vec{M}\vec{I}$ in *B. pseudomallei*). It thus appeared that local topology within the chromosome may have evolved before the separation of various *Burkholderia* clades (Figure 25).

The reported chemical structure of the AHL signals was also seen to be in good correlation with the local topology of the AHL systems, *i.e*., the same chemical signal was produced by AHL systems belonging to identical or related local topologies, or to a few topology types (Table 9) (Choudhary et al. 2013).

Table 9: Correlation between the chemical structure and the local gene arrangements

Symbol	**Occurrences in *Burkholderia***			
	BCC group	***B. pseudomallei* group**	**Plant-beneficial / environmental group**	**Other pathogenic *Burkholderia* species**
C6-HSL	2 (R1)	0	0	0
C8-HSL	14 (M1)	13 (M1)	0	2 (M1: *B. glumae* and R3: *B. plantarii*)
OHC8-HSL	0	11 (R1 and X3: *B. mallei*; R1: *B. thailandensis* and M3 of *B. pseudomallei*)	3 (R1)	0
C10-HSL	1 (X5)	0	0	0
OHC10-HSL	0	5 (R1 of *B. pseudomallei* and M3 of *B. thailandensis*)	0	0
OC14-HSL	0	0	9 (L1)	0

The agreement across genera was less equivocal (details not shown). For instance, an AHL system of L1 topology in *B. xenovorans* is regulated by the signal OC14-HSL, while an analogous AHL system in *P. aeruginosa* is regulated by a different signal, OC12-HSL. In comparison, the R1 topology is connected to a variety of signals in at least 96 proteobacterial species in which it has been detected so far (Z Gelencsér, Choudhary, et al. 2012). This is more clear with the clustering of LuxI proteins (Figure 25) (Choudhary et al. 2013).

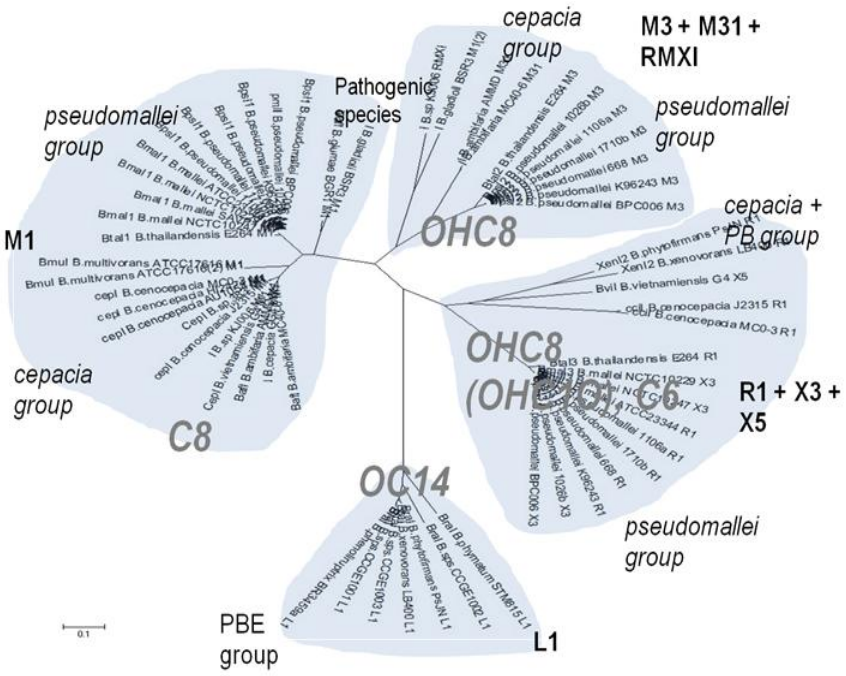


Figure 25: Clustering of LuxI protein sequences and perceived signals by LuxR homologues in complete *Burkholderia* genomes

Interestingly, each of the three clades of *Burkholderia* has different types of QS genes arrangement. In the section below, each of three clades are discussed individually.

5.1.3 *Burkholderia cepacia* complex (BCC)

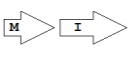
BCC is composed of at least 17 species, including *Burkholderia cepacia*, *B. multivorans*, *B. cenocepacia*, *B. vietnamiensis*, *B. ambifaria*, *B. stabilis*, *B. dolosa*, *B. anthina* and *B. pyrrocina* (Venturi et al. 2004). The AHL QS in the BCC group consists of *luxI/R* homologs known as *cepI* and *cepR*.

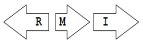
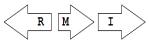
CepI synthesizes two AHL signals, the C8-HSL (*N*-octanoyl-L-homoserine lactone) and C6-HSL (N-hexanoyl-L-homoserine lactone) in greater and lesser amounts respectively. CepR binds and consequently responds to the abundant cognate C8-HSL (Venturi et al. 2004).

The arrangement of *cepI* and *cepR* was seen to be common in all BCC members. The *cepI* and *cepR* genes are seen to be divergently transcribed and most commonly contained *rsaM* homolog in the intergenic region. This kind of arrangement was termed as M1 ($\overleftarrow{RM}\overrightarrow{I}$) (Zsolt Gelencsér, Choudhary, et al. 2012). The intergenic gene *rsaM* is believed to be a negative regulator of the AHL QS system (Mattiuzzo et al. 2011).

Recently another QS system, called *cciI/R*, was identified in *B. cenocepacia* having pathogenicity islands such as *B. cenocepacia* K56-2, *B. cenocepacia* J2315 and *B. cenocepacia* MC0-3. This system produces primarily C6-HSL and minor amounts of C8-HSL (O'Grady et al. 2009a; Malott et al. 2005). *cciI* and *cciR* are transcribed in the same direction and accordingly the topology was named as R1. All *B. cenocepacia* members also contained a solo/orphan *luxR* homolog named as cepR2 (O'Grady et al. 2012; Subramoni & Venturi 2009). QS topologies in BCC members are shown in Table 10. Table includes only *luxR/I* homolog pairs (Choudhary et al. 2013).

Table 10: Quorum sensing (QS) topologies in the *Burkholderia cepacia* complex

Species	**QS system**	**ID**	**Gene Topology**	**Chr**	**Major AHL**	**Comments**
B. ambifaria	bafR/-/bafI	M1		Chr 2	C8	All *B. ambifaria* strains
	-/-	M3 1		Chr 3		Some strains such as *B. ambifaria* AMMD, *B.ambifaria* MC40-6
	-/-	R1		NA		*B. ambifaria* IOP40
B. cenocepacia	cepR/- /cepI	M1		Chr 2	C8	All *B. cenocepacia* strains
	cciR/cciI	R1		Chr 2	C6	Some strains such as *B. cenocepacia* K56-

						2, *B. cenocepacia* J2315 and *B. cenocepacia* MC0-3
B. multivorans	bmuR/-/bmuI	M1		Chr 2	C8	*B. multivorans* ATCC 17616
B. multivorans ATCC 17616*	sdiA/-/bmuI	M1		Chr 2	C8	*B. multivorans* ATCC 17616*
B. vietnamiensis	cepR/-/cepI	M1		Chr 2	C8	*B. vietnamiensis* G4
	bviR/-/bviI	X5		Chr 3	C10	*B. vietnamiensis* G4
	-/-	R2		NA		Present in *B. vietnamiensis* C2822, *B. vietnamiensis* G4 etc.
B. cepacia	cepI/cepR	R2		NA		Present in *B. cepacia* ATCC 25416, *B. cepacia* DBO1, *B. cepacia* K56-2

* same strain but different bioproject.

Key : NA= the genome is draft or individual genbank entry and the information about chromosome number is not available

It was observed that the QS genes are mainly present on chromosome 2, except for X5 topology of *B. vietnamiensis* and M31 topology of *B. ambifaria* which is present on chromosome 3.

Importantly, the QS systems in *B. cenocepacia* are interrelated with each other; CepR regulates *cepI* and *cciIR* genes, however, CciR negatively regulates the expression of *cepI* thus allowing for a negative regulatory feedback loop on the *cepI/R* system (Malott et al. 2005). The solo/orphan *cepR2* which lacks associated cognate *luxI*-AHL synthase gene, negatively regulates itself and is also negatively regulated by CciR (Malott et al. 2009). Figure 26 below shows the regulatory pattern observed in BCC members.

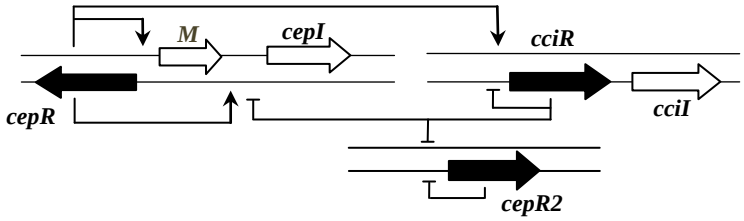


Figure 26: QS regulatory circuits in *B. cenocepacia* J2315
(O'Grady et al. 2009b).

5.1.3.1 Neighborhood genes

Studies on *B. cenocepacia* have shown that a neighboring gene downstream from *cepI* can also influence the production of AHL (O'Grady et al. 2012). This was identified as ORF BCAM1871, it is co-transcribed with *cepI* and is regulated by CepR. Orthologs of the ORF BCAM1871 contain a domain of 3-hydroxy-3-methyl-glutaryl-CoA reductase family (HMG-CoA reductase). Its genomic location is conserved in all BCC members. Previous studies suggest that BCAM1871 alone cannot activate AHL synthesis, but it enhances the rate of AHL production (O'Grady et al. 2012). Apart from the BCC group, the orthologs of ORF BCAM1871 was also seen to be conserved in *B. mallei* and *B. pseudomallei*, it flanks downstream to *luxI* homologs, and the corresponding systems are flanked upstream by an Mg transporter gene. (*i.e.*, the M1 topology (RMI) is complemented by two conserved flanking genes in these species) (Figure 27).

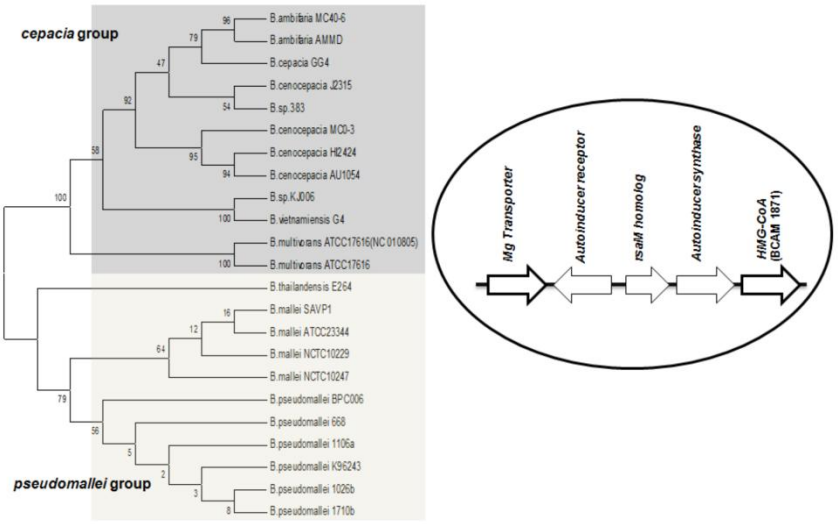


Figure 27: Cladogram of the orthologs of ORF BCAM1871, encoding a protein with an HMG-CoA domain in *Burkholderia* sps.

The orthologs are well separated into *cepacia* and *pseudomallei* groups. The RMI ($\vec{RMI}$) motive is flanked by conserved genes on both sides in members of both *cepacia* and *pseudomallei* groups. The numbers on the tree branches indicate bootstrap values (%).

5.1.3.2 Chromosomal arrangement of QS genes

Circular map representations revealed that the chromosomal arrangements of the AHL QS genes in *BCC* members are relatively well conserved with respect to the potential OriC (Figure 28).

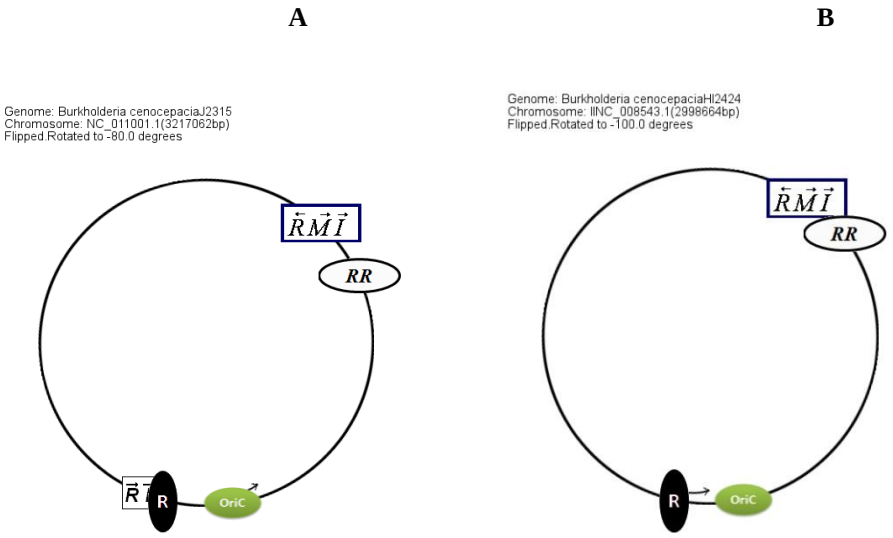


Figure 28: Typical arrangements of QS genes in chromosome II of the BCC group. (A) Arrangement of QS genes in BCC members with cciR/I genes (for example B. cenocepacia J2315 and B. cenocepacia MC0-3). (B) Arrangement of QS genes in BCC members having just one pair of *luxR/I* homologs (*cepR/I*) (for example *B. cenocepacia* AU1054, *B. cenocepacia* HI2424, *B. cepacia* GG4 etc). OriC denotes the origin of replication, solo *luxR* genes (i.e. those without adjacent I genes) are denoted by black and white ovals, respectively, with the latter indicating two adjacent *luxR* homologues.

5.1.4 *Burkholderia pseudomallei* group

The *Burkholderia pseudomallei* group consists of *B. mallei* and *B. pseudomallei* subgroups. Both are characterized by multiple AHL QS systems as well as additional *luxR* homologs (Larsen & Johnson 2009).

Burkholderia mallei

The genome of *B. mallei* contains two *luxI* and four *luxR* homologs arranged as two complete *luxI/R* AHL QS systems and two orphan/solo *luxR* homologs (Duerkop et al. 2008). The complete QS systems in *B. mallei* are called BmaI/R, and their *luxR* homologues *bmaR1* and *bmaR3* respond to signals produced by adjacent *luxI* homolog genes *bmaI1* and *bmaI3*, respectively (Duerkop et al. 2007). BmaI1/R1 shares sequence similarity with BpsI1/R1 in *B. pseudomallei*, producing and responding to C8-HSL (Duerkop

et al. 2007). *bmaI1-bmaR1* are transcribed in opposite orientation to each other and are separated by an intergenic ORF which is a homolog of *rsaM*. The second LuxI/R homolog pair of *B. mallei* is BmaR3/I3. The *bmaR3-bmaI3* pair is transcribed in the same direction and do not contain ORFs in the intergenic region. BmaI3 produces multiple AHL molecules and includes abundant levels of *N*-3-hydroxy-octanoyl-HSL (3OHC8-HSL). In addition to 3OHC8-HSL, BmaI3 produces *N*-3-hydroxy-hexanoyl-HSL (3OHC6-HSL), and *N*-3-hydroxy-decanoyl-HSL (3OHC10-HSL) in minor amounts. BmaR3 responds to the most abundant product 3OHC8-HSL (Duerkop et al. 2008). The role of orphan/solo LuxR homologs, BmaR4 and BmaR5 are currently unknown. In this genus, all the QS genes are present on chromosome 2.

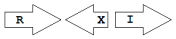
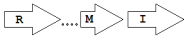
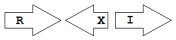
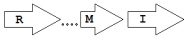
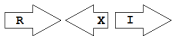
Burkholderia pseudomallei

The *B. pseudomallei* genome has been identified as having three *luxI/R* pair homologs, namely *bpsI1/R1*, *bpsI2/R2*, and *bpsI3/R3*, and two orphan or solo *luxR* regulator homologs, designated as *bpsR4* and *bpsR5*. With the exception of *bpsR5,* which is present on chromosome 1, most of the QS genes are present on chromosome 2 (Kiratisin & Sanmee 2008).

The major AHL produced by BpsI1 is C8-HSL (*N*-octanoyl-L-homoserine lactone), whereas BpsI2 and BpsI3 are associated predominantly with *N*-3-hydroxy-octanoyl homoserine lactone (OHC8-HSL) and *N*-3-hydroxy-decanoyl homoserine lactone (OHC10-HSL) respectively (Gamage et al. 2011). *bpsI1* is divergently transcribed to *bpsR1,* representing the M1 topology, and is separated by an intergenic region which contains an ORF homologous to *rsaM*. *bpsI2* is transcribed in the same direction as *bpsR2*, representing the M3 topology, and is separated by two to seven genes, one of which is a *rsaM* homolog. *bpsI3* is also transcribed in the same direction as *bpsR3* but is not separated by any of the intergenic ORFs and represents the R1 topology.

The local topological patterns and location of QS genes in members of *B. pseudomallei* group are shown in Table **11** below.

Table 11: QS topologies in *Burkholderia pseudomallei* group

Species	QS System	ID	Gene Topology	Chr.	Major AHL	Comments
B. mallei	*bmaR3/bmaI3*	R1		Chr2	OCH8	*B. mallei* ATCC 23344
	bmaR1/-/bmaI1	M1		Chr2	C8	Present in all *B. mallei* strains
	bmaR3/-/bmaI3	X3		Chr2	OCH8	Present in some strains such as *B. mallei* NCTC 10229 and *B. mallei* NCTC 10247, *B. mallei* FMH etc.
B. pseudomallei	*bpsR3/bpsI3*	R1		Chr2	OHC10	Present in all *B. pseudomallei* strains
	bpsR1/-/bpsI1	M1		Chr2	C8	Present in all *B. pseudomallei* strains
	bpsR2/-/bpsI2	M3		Chr2	OHC8	Present in all *B. pseudomallei* strains
	-/-/-	X3		Chr2		Present in some strains such as *B. pseudomallei* 1026b, *B. pseudomallei* MSHR346, *B. pseudomallei* 112 etc.
	bpsI/bpsR	R2		NA		*B. pseudomallei* 844
B. thailandensis	*btaR3/btaI3*	R1		Chr2	OHC8	Present in some strains of *B. thailandensis*
	btaR1/-/btaI1	M1		Chr2	C8	Present in all *B. thailandensis* strains
	btaR2/-/btaI2	M3		Chr2	OHC8, OHC10	Present in all *B. thailandensis* strains
	-/-/-	X3		NA		Present in some strains such as *B. thailandensis* Bt4, *B. thailandensis* TXDOH

5.1.4.1 Chromosomal arrangement of QS genes

The location of the QS genes within the circular chromosome 2 also shows similarities between genomes even though the conservation is stronger in *B. pseudomallei* than in *B. mallei* (**Figure 29**).

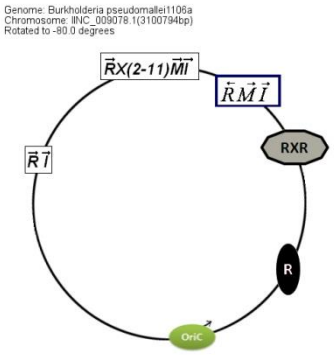


Figure 29: An example of the chromosomal arrangement of QS genes in completely sequenced genome of *B. pseudomallei* strains
(for example *B. pseudomallei* 1026b, *B. pseudomallei* 1026a etc). RXR: two solo luxR homologs are separated by a hypothetical gene. The X gene missing in some *B. pseudomallei*.

QS regulation of *B. pseudomallei* is rather complex and hierarchical. BpsR1 regulates *bpsI1* in the presence of a cognate AHL in a positive feedback loop. BpsR5 also partially activates *bpsR1*. *bpsI2* is constitutively expressed and its expression is enhanced by each of the BpsR - especially by BpsR1 and BpsR3 - in the presence of 3-oxo-C8HSL. *bpsI3* is also constitutively expressed but its expression is repressed in the presence of any BpsR protein, although the repression is less in the case of BpsR3 (Kiratisin & Sanmee 2008)(Figure 30)

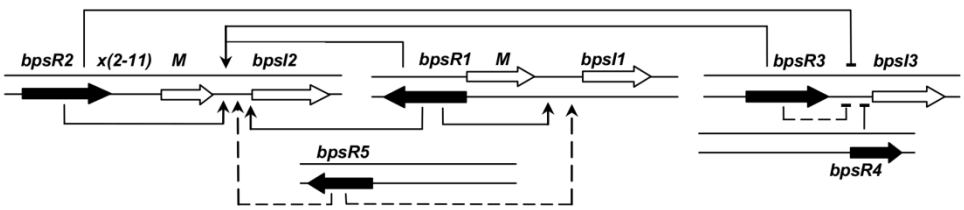


Figure 30: Schematic representation of complex regulatory circuit in B. *pseudomallei* K92643 (Kiratisin & Sanmee 2008; Choudhary et al. 2013)**. Dashed lines indicate partial regulation.**

5.1.5 Plant Beneficial and Environmental (PBE) group

This group of the *Burkholderia* genus currently includes 29 non-pathogenic species which are most often associated with plants (Suárez-Moreno et al. 2010). Only few members of this group have been identified to possess the AHL QS systems. One system similar to the LasI/R system of *P. aeruginosa* is BraI/R, and is highly conserved among all species of this group. The *braR* and *braI* genes are transcribed in the same direction and are under strict negative regulation by the RsaL repressor that is present between the *braR* and *braI* genes (Suárez-Moreno et al. 2008) (Figure 31). This arrangement was termed the L1 topology ($\vec{R}\overleftarrow{L}\vec{I}$).

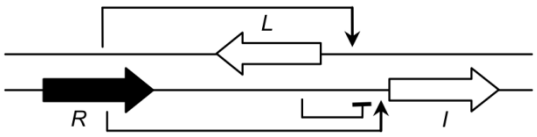
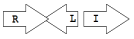


Figure 31: Regulatory circuit of BraR/I system in plant-beneficial *Burkholderia* group (for example *B. xenovorans* LB400) (Suárez-Moreno et al. 2010)

BraI produces multiple AHLs, including 3-oxo-C6-HSL, 3-oxo-C8-HSL, 3-oxo-C10-HSL, 3-oxo-C12-HSL and 3-oxo-C14-HSL. BraR responds best to 3-oxo-C14-HSL meaning that is likely to be the cognate AHL for BraI/R systems (Suárez-Moreno et al. 2010).

Some members of the PBE group (mostly from the *B. xenovorans*, *B. graminis* and *B. phytofirmans* subclade) have an additional AHL system that resembles LuxI/R pairs in other *Burkholderia* species. This additional AHL system is named as XenI2/R2 and produces and responds to 3-hydroxy-C8-HSL (Table 12).

Table 12: QS topologies in Plant beneficial and environmental *Burkholderia* sps.

Species	QS system	ID	Gene Topology	Chr. No.	Major AHL	Comments
Plant beneficial *Burkholderia* sps.	*braR/rsaL /braI*	L1		Chr2	OC14	All Plant beneficial strains have this system. Ex: *B. phymatum STM815*, *B. phytofirmans PsJN*, *B. xenovorans LB400*, *B. graminis C4D1M*, *B. unamae* etc.
Plant beneficial *Burkholderia* sps.	*xenR2/xen I2*	R1		Chr1(*B. phytofirmans PsJN)*, Chr 3 (*B. xenovorans* LB400)	OHC8	Present in some plant beneficial strains. Ex: *B. phytofirmans PsJN*, *B. xenovorans LB400*, *B. graminis C4D1M*

5.1.5.1 Chromosomal arrangement of QS genes

In members of the PBE group, *braI/R* genes are always present on chromosome 2. The additional QS system XenI2/R2 is present on chromosome 1 and chromosome 3 in *B. phytofirmans* and *B. xenovorans*, respectively. The position of QS genes in the PBE group within the chromosomes was conserved with respect to the *oriC* (Figure 32).

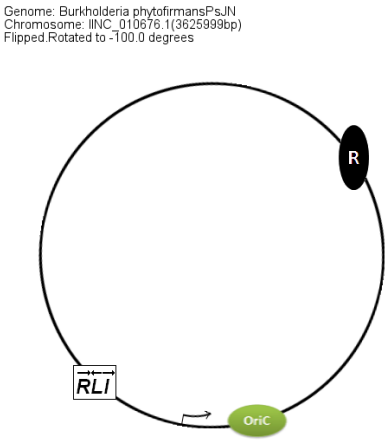


Figure 32: Chromosomal arrangement of QS genes (braI/R) and OriC in completely sequenced members of the plant-beneficial and environmental group
(for ex. *B. phytofirmans* PsJN, *B. xenovorans* LB400, etc)

In contrast to members of the *BCC* group, it seems that there is no hierarchical connection between the BraI/R and XenI2/R2 systems. However, an orphan LuxR homolog called BxeR is present in *B. xenovorans* LB400 and other strains from this cluster (*B. graminis* C4D1M, *B. terricola* LMG30594, *B. phytofirmans* PsJN and *B. phenoliruptrix* AC1100) and negatively regulates *xenI2/R2* (Suárez-Moreno et al. 2010) (Figure 33).

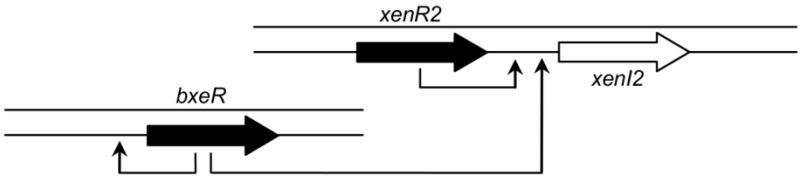
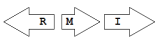
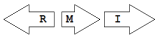


Figure 33: Regulatory circuit of xenI2/R2 and bxeR genes in plant-beneficial and environmental group
(for example *B. xenovorans* LB400) (Suárez-Moreno et al. 2010)**.**

5.1.6 Other Pathogenic *Burkholderia* sps.

There are some human and plant pathogens, including phytopathogenic *B. glumae, B. gladioli* and *B. plantarii* that do not cluster with either the *BCC* or with the *B. pseudomallei* groups. Quorum sensing is known to play an important role in the pathogenicity of these bacteria (Solis et al. 2006). Typical arrangements are shown in Table 13. *B. glumae* causes disease in rice seedlings and *B. gladioli* is both a rice pathogen as well as an opportunistic pathogen in humans, and is phylogenetically close to *B. glumae* (Devescovi et al. 2007). Another plant pathogen is *B. plantarii* which also requires the AHL QS system for its pathogenicity (Chun et al. 2009). Interestingly, two different strains of *B. glumae* has different topology of AHL QS genes. *luxR* and *luxI* homologs in *B. glumae* BGR1 are divergently transcribed and are separated by an intergenic gene *tofM*, which is an *rsaM* homolog (Chen et al. 2012). In contrast, the LuxR homolog in *B. glumae* ATCC 33617 is 110 amino acids longer and is believed to be non-functional (Devescovi et al. 2007). This *luxR* homolog is also divergently transcribed with respect to its *luxI* homolog but is not separated by any intergenic gene (Devescovi et al. 2007). The topology of *B. gladioli* is similar to *B. glumae* BGR1. In *B. plantarii*, *luxR* and *luxI* homologs are divergently transcribed and are adjacent to each other.

Table 13: QS topologies in Plant pathogenic *Burkholderia* sps.

Species	QS system	ID	Gene Topology	Chr. No.	Major AHL	Comments
B. glumae	*tofR/tofM/tofI*	M1		Chr 2	C8	*B. glumae BGR1*
	tofR/tofI	R3		NA	Non functional	*B. glumae* ATCC33617 and *B. glumae* AU6208
B. gladioli	-/-/-	M1		Chr2		*B. gladioli* BSR3
	-/-/-	M1		plasmid		*B. gladioli* BSR3

B. plantarii	*plaR/plaI*	R3		NA	C8	*B. plantarii* ATCC43733

Regulation of the QS system in these species is canonical in that a positive feedback loop is present. The repressive action of TofM in *B. glumae* BGR1 remains unclear and needs to be further studied (Chen et al. 2012).

5.2 **Summary**

The ability to produce AHL signals is very common in the genus *Burkholderia*. The genes of *Burkholderia* QS systems follow at least 11 typical topological arrangements (10 are common to topological arrangements in other Proteobacteria). Some species, like *B. mallei* and *B. pseudomallei* and certain members of the BCC group, contain multiple AHL QS systems that usually do not operate independently from each other but form a complex hierarchical regulatory network. QS genes in *Burkholderia* are usually located on chromosome 2 where most genes related to virulence and secretion systems are found (Whitlock et al. 2007).

The chemical structure of the AHL signals, the local topology of the QS genes and the location of QS systems within the chromosomes show a degree of conservation throughout the entire genus. The fact that some AHL QS systems, which produce OHC8 and OHC10, can be associated with different local topologies in different species possibly indicates that QS topologies represent orthologous subgroups that diverged from each other before the *Burkholderia* genus separated into the current taxonomical groups. It can be speculated that the AHL signals appeared before the appearance of the genus and the subsequent rearrangements of the ancestral QS systems lead to the local and chromosomal topologies seen in current species.

6 Peptide Quorum sensing: ComQXPA in *Bacillus subtilis*

The third case study is the ComQXPA system of *Bacillus subtilis*, which is perhaps the best studied such system in Gram positive bacteria. Using the methodology developed for AHL based QS systems, I searched bacterial genomes for candidate loci similar to the *B. subtilis* ComQXPA locus, by combining Hidden Markov Model (HMM) recognizers with filtering criteria based on the structural and organizational properties of ComQXPA QS systems.

6.1 Results and Discussion

We were interested to know whether or not ComQXPA architectures occur outside the *B. subtilis* species. To investigate this question, all prokaryotic genomes (2620 complete and 6970 draft genomes having 644474 annotated and 505155 unannotated contigs, a total of over 4.7 million protein sequences) were scanned for ComQXPA-related proteins, using Hidden Markov recognizers. The search revealed that in addition to the 21 occurrences explicitly mentioned in the literature and/or in the databases, there are 39 new occurrences in some of which one, more or all of the functions were indicated as hypothetical. The complete list of the species is shown in Table 14

Table 14: List of the species having comQXPA loci

Genome	Chr/Contig	Strand	From	To	Clade[1]	Overlap type[2]
Bacillus HYC 10	NZ_AMSH01000014	-	16598	20776	S	B
Bacillus M 26	NZ_AJWW01000025	-	45674	49787	S	B
Bacillus 5B6	NZ_AJST01000001	-	2805760	2809860	S	A
Bacillus B14905	NZ_AAXV01000005	-	5998	9953	N	D
Bacillus BT1B CT2	NZ_GL635727	-	403211	407296	S	A
B. subtilis SMY	NZ_CM000490	-	3251855	3255958	S	B
B. subtilis NCIB 3610	NZ_CM000488	-	3251815	3255918	S	B
B. subtilis MB732	NZ_AOTY01000001	-	74262	78365	S	B
B. subtilis 168	NZ_CM000487	-	3251761	3255864	S	B

B. subtilis SC 8	NZ_AGFW01000001	+	1041642	1045792	S	A
B. subtilis JH642	NZ_CM000489	-	3224835	3228938	S	B
B. subtilis spizizenii ATCC 6633	NZ_ADGS01000012	+	47108	51202	S	A
B. subtilis inaquosorum KCTC 13429	NZ_AMXN01000007	+	189370	193553	S	A
B. cereus VD102	NZ_JH792240	I	462923	467112	N	A
B. cereus BAG4X12_1	NZ_JH791951	I	4689952	4694141	N	A
B. cereus MSX A1	NZ_JH792105	I	452254	456443	N	A
B. isronensis B3W22	NZ_AMCK01000012	-	30729	34664	N	E
B. sonorensis L12	NZ_AOFM01000009	-	41264	45367	S	B
B. pumilus ATCC 7061	NZ_ABRX01000002	-	211408	215508	S	A
B. licheniformis WX 02	NZ_JH636050	-	3362792	3366900	S	A
B. amyloliquefaciens plantarum M27	NZ_AMPK01000004	+	732759	736905	S	B
B. atrophaeus C89	NZ_AJRJ01000035	+	32850	36985	S	B
B. azotoformans LMG 9581	NZ_AJLR01000035	-	2395	6619	N	A
B. mojavensis RO H 1	NZ_JH600279	-	76612	80677	S	A
B. vallismortis DV1 F 3	NZ_JH600244	-	112511	116637	S	A
Geobacillus G11MC16	NZ_ABVH01000020	-	15553	19626	N	A
Anoxybacillus flavithermus TNO 09 006	NZ_KB205935	+	581038	585047	N	A
Lysinibacillus fusiformis ZC1	NZ_ADJR01000053	I	105302	109400	N	A
Lysinibacillus fusiformis ZC1	NZ_ADJR01000053	+	179719	183681	N	D
Paenibacillus curdlanolyticus YK9	NZ_AEDD01000016	-	19156	23592	N	A
Desulfosporosinus youngiae DSM 17734	NZ_CM001441	-	3509306	3513580	N	D
*Bacillus*_JS	NC_017743	-	3117558	3121632	S	A
B. subtilis RO NN_1	NC_017195	-	3043640	3047706	S	A
B. subtilis natto BEST195	NC_017196	-	3015448	3019587	S	A
B. subtilis BSn5	NC_014976	-	1260169	1264309	S	B
B. subtilis 168	NC_000964	-	3252804	3256907	S	B
B. subtilis BSP1	NC_019896	+	970082	974185	S	B
B. subtilis QB928	NC_018520	-	3184243	3188346	S	B
B. subtilis subsp. subtilis str. BAB-1	NC_020832	-	3027997	3032071	S	A
B. subtilis spizizenii W23	NC_014479	-	3036239	3040333	S	A
B. subtilis spizizenii TU B 10	NC_016047	-	3194765	3198905	S	B
B. pumilus SAFR 032	NC_009848	-	2835144	2839257	S	B
B. amyloliquefaciens FZB42	NC_009725	-	2994495	2998536	S	B
B. amyloliquefaciens DSM7	NC_014551	-	3026887	3031011	S	A
B. amyloliquefaciens LL3	NC_017190	-	3040520	3044644	S	A
B. amyloliquefaciens IT-45	NC_020272	+	905489	909592	S	A
B. amyloliquefaciens plantarum AS43 3	NC_019842	-	3027197	3031314	S	A
B. amyloliquefaciens subsp. *plantarum* UCMB5113	NC_022081	-	2987612	2991741	S	C
B. amyloliquefaciens subsp. plantarum UCMB5033	NC_022075	-	3144440	3148608	S	A

B. amyloliquefaciens subsp. *plantarum* UCMB5036	NC_020410	-	2974358	2978508	S	C
B. atrophaeus 1942	NC_014639	-	2685134	2689269	S	B
Syntrophobotulus glycolicus DSM 8271	NC_015172	-	806051	810307	N	A
Desulfosporosinus meridiei DSM 13257	NC_018515	+	2120062	2124318	N	A
Lysinibacillus sphaericus C3 41	NC_010382	-	4400921	4404847	N	A
Lysinibacillus sphaericus C3 41	NC_010382	-	4235148	4239109	N	D
B. licheniformis 9945A	NC_021362	-	3397421	3401535	S	B
Geobacillus C56 T3	NC_014206	+	359832	363950	N	A
Geobacillus thermoglucosidasius C56 YS93	NC_015660	-	3536986	3541061	N	A
Anoxybacillus flavithermus WK1	NC_011567	+	589585	593594	N	D
B. coagulans 36D1	NC_016023	+	2267386	2271490	N	A

[1]Clades are indicated as S (*Bacillus Subtilis* type) or N (non-B. subtilis type)

[2]Overlap types (A-E) are indicated in Figure 3

[3]Evidence is indicated with the number of reference, unless a superscript (4 or 5) indicates it otherwise.

[4]experimentally tested on a different strain of the same subspecies/species

[5] assumption based on sequence similarity

I = Hypothetical inverted ComX

The similarity cladograms of all four ComQXPA sequences showed two large clades. The larger clade were termed "*B. subtilis*-like" contained *B. subtilis* and a few other species from the *Bacillus subtilis-lichenoformis* group. The smaller and visibly more varied clade was termed as "non-*B. subtilis*-like" since it contained no *B. subtilis* sequences. This clade contained a few species from the *Bacillus* genus (*B. isronensis, B. coagulans, B. azotoformans and B. cereus*), but also species from *Lysinbacillus, Geobacillus, Anoxybacillus, Desulfosporosinus* and more distantly related organisms such as *Clostridiales*. Figure 34 shows the ComQ tree as an example.

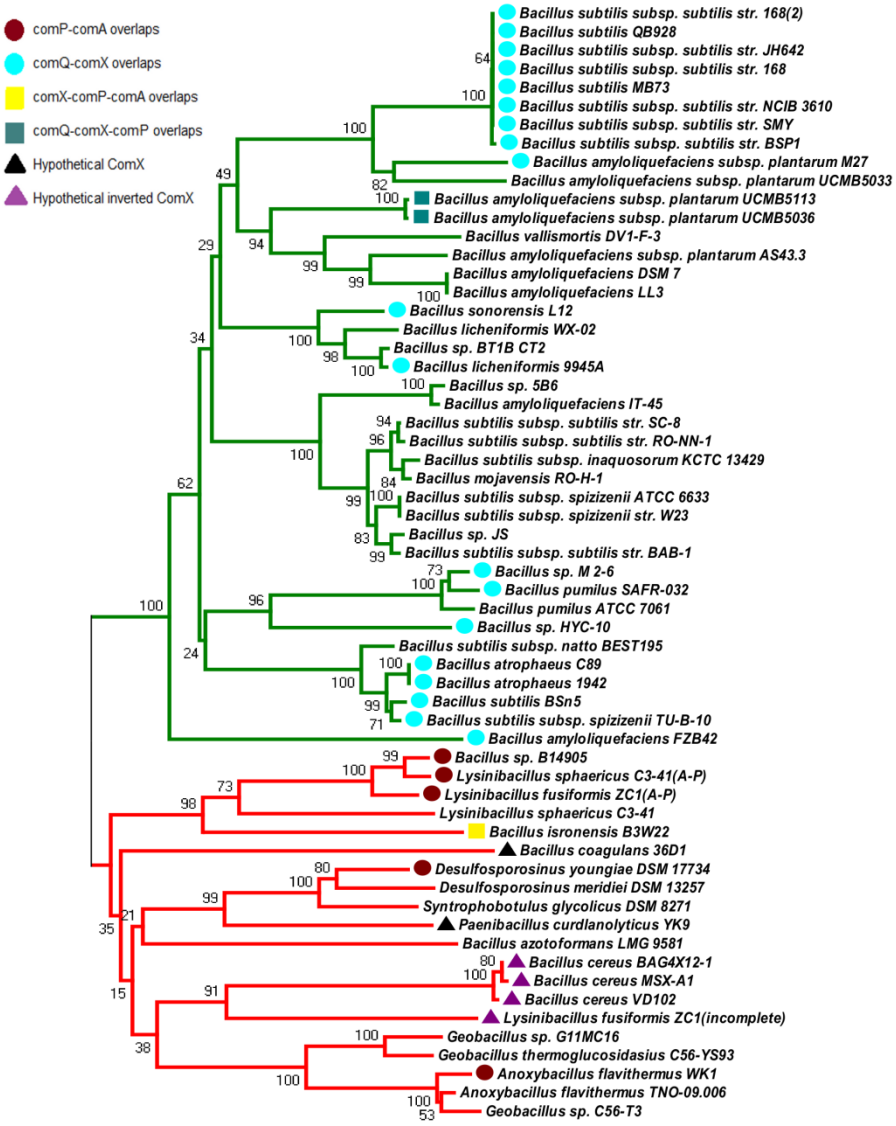


Figure 34: Cladogram of ComQ proteins

6.1.1 Local arrangement of the comQXPA genes

The chromosomal arrangements of the loci are schematically shown in Figure 35 where the reading frames are represented as arrows pointing to the direction of transcription. Definite pattern of overlaps were observed within the loci. The types of overlaps were named A through E, the species are noted in Table 14

The majority of the loci contains no overlapping reading frames (A in Figure 35). However, there are overlapping reading frames in ComQXPA loci that are spread in the two clades. i) In the *B. subtilis*-like clade the overlap types are dominated by an apparent mutation of the *comQ* stop codon that results in a 13-18 amino acids long C-terminal extension, giving rise to the *comQ-comX* overlaps. In two loci within the same clade (*B. amyloliquefaciens subsp. plantarum* UCMB5113, *B. amyloliquefaciens subsp. plantarum* UCMB5036), there is a second mutation causing a 14 amino acids N-terminal extension to ComP, leading to the comQ-comX-comP overlaps. ii) In the non-*B. subtilis*-like clade the most frequent overlaps (6 out of 14 species in the clade) are between ComP and ComA. Five out of these is caused by a C-terminal extension of ComP, however in *Anoxybacillus flavithermus* WK1 the overlap is due to a 22 amino acids N-terminal extension of ComA. In one locus (*B. isronensis* B3W22) there is an additional mutation that gives rise to a 3 amino acids N-terminal extension of ComP, leading to a *comX-comP-comA* overlap.

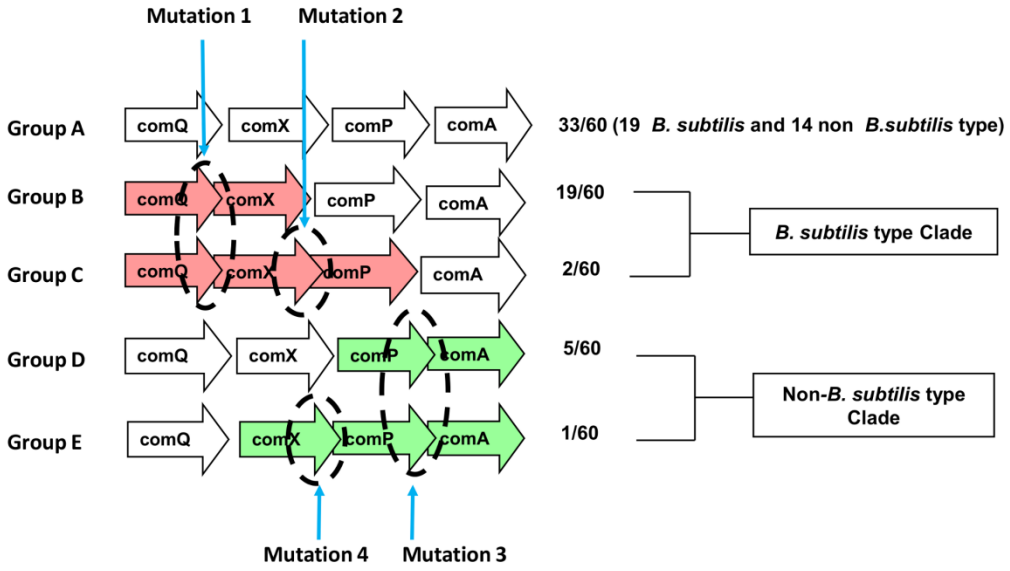


Figure 35: com loci in *B. subtilis* type and non- *B. subtilis* type clade, overlaps are shown in color. For the definition of the clades, see pylogenetic analysis. The numbers indicate the frequency of the type/the number total occurrences (i.e. 60).

In general, overlapping mutations are not rare in bacteria (Johnson & Chisholm 2004) most likely because bacterial genes are most frequently (>70%) located on one strand (Fukuda et al. 2003). However, the fact that we see two coherent sets of mutations (Mutations 1-2, and Mutations 3-4, respectively) that are confined to two separate clades, makes us believe that these mutations may have some logical reason. In fact, the estimated probability of finding these mutations in two separate groups by chance is low ($p < 0.001$). This makes us speculate that gene overlap patterns coevolved with *comQXPA* loci and may have adaptive fitness advantages. Overlapping genes are often found in regulatory operons and indeed, the primary role of *comQXPA* loci is to control gene expression (Comella & Grossman 2005). It was thus speculated that the expression of *comQXPA* genes may be different in the two clades. In other words, the fact that different kinds of mutations are established in the two clades made us hypothesize that the com loci of two clades may differ in the way how the transcription/translation of the genes is coupled – a statement that would be worth to test experimentally in the future

6.1.2 Unusual com-like loci

A few com-like arrangements were found outside the *Bacillu*s genus in which the *comX* sequence was of the right length and in the right position between *comQ* and *comP*, but showed no appreciable homology with the known *comX* sequences, except the tryptophan residue in the C-terminus. We termed these reading frames as hypothetical *comX* (Figure 36).

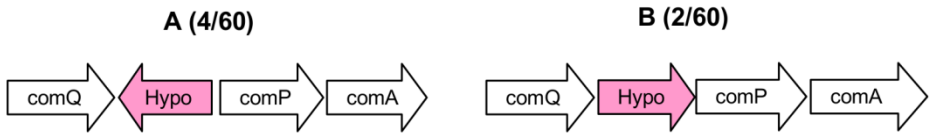


Figure 36: Unusual comQXPA-like loci.
(A) Non-canonical unusual com system is present *in B. cereus* VD102, *B. cereus* BAG4X12 1, *B. cereus* MSX A1 and *L. fusiformis* ZC1. Note that *Lysinibacillus* is the only genus whose two species have two com loci each: *L. sphaericus* C3 41 has two canonical loci while in *L. fusiformis* ZC1 one

locus is canonical and the other one is non-canonical, shown here. (B) Canonical unusual *com* system present in *Paenibacillus curdlanolyticus* YK9 and *B. coagulans* 36D1.

In three strains of *B. cereus* (VD102, BAG4X12 1, MSX A1) and in *L. fusiformis* ZC1 the hypothetical *comX* gene was seen to be on the opposite strand (Figure 36A). *Lysinibacillus* is the only genus whose two species have two *comQXPA* loci each: *L. sphaericus C3 41* has two canonical loci (all genes in the same strand), while in *L. fusiformis* ZC1 one locus is canonical and the other one has the hypothetical *comX* gene on the opposite strand, shown in Figure 36A. The other arrangement (Figure 36B) is canonical in terms that all genes are present in the same strand. Two species *Paenibacillus curdlanolyticus* YK9 and *B. coagulans* 36D1 have this kind of arrangement. However, it cannot be commented that these *com*-like loci are functional and that they function in the same way as in *B. subtilis*. Nevertheless, the fact that they are at least in part conserved in relatively distinct species makes them interesting subjects for further experiments.

6.1.3 ComP Sequence variability

A substantial variability was detected in the N-terminal domain of the ComP protein that is predicted to interact with ComX (Piazza et al. 1999). The variability of the N-terminal domain is especially prominent in comparison to the conserved cytoplasmic domain. In general, ComP protein contains 10 transmembrane domains (48 out of 60 occurrences). However there are proteins in both clades in which the N-terminal transmembrane domain is missing (9 out of 60 occurrences) meaning that the loop region of the intact protein becomes N-terminal (Figure 37).

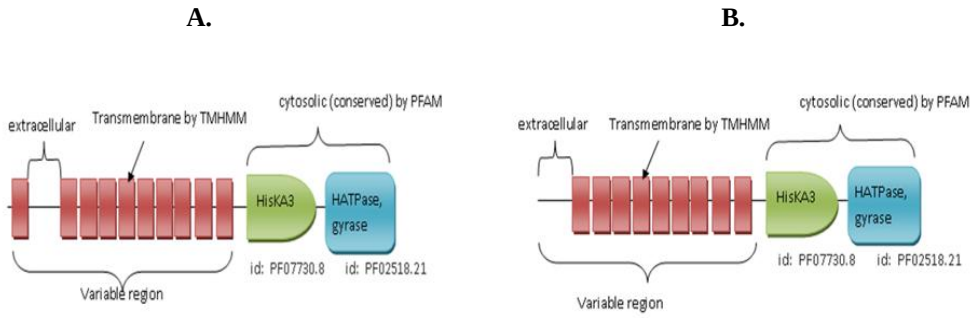


Figure 37: Transmembrane domain architecture in ComP proteins.
A.) 10 Transmembrane helices present in 48 proteins. B.) 9 Transmembrane helices present in 12 proteins.

6.2 **Summary**

In this work, a census of *comQXPA* locus proteins involved in the quorum sensing regulation of competence and other late growth adaptive traits in *B. subtilis* was presented. 31 new occurrences were found, many of them outside the *Bacillus* genus, with some from different orders (for instance order: Clostridiales, family: *Peptococcaceae*, genera: *Desulfosporosinus*, *Syntrophobotulus*). The local arrangement of the genes was quite conserved in all *com*-like occurrences. The only variability of gene topology was found in *com*-related operons whose function may however not be necessarily linked to quorum sensing, or may be based on a different type of peptide signal. Two conserved classes of overlap patterns in *B. subtilis*-like and non-*B subtilis* like *comQXPA* loci, respectively, was found which may be due to hitherto unknown transcriptional/translational differences.

In summary, a wholesale scan of current databases showed a number of novel occurrences of *com* QS regulatory locus originally identified in *B. subtilis*. It was found that this locus is phylogenetically more widespread than previously thought, and its organization has some commonalities with unrelated QS systems.

7 Conclusion and Discussion

In coming years we will all have access to over a thousand sequenced genomes. This data will grow to become the central resource in modern biology. Annotating this collection has always been the core challenge of modern bioinformatics. Overbeek et. al described a new approach to annotate based on idea of subsystems that promises to dramatically improve the quality and utility of annotations (Overbeek et al. 2005). This approach was used in this work to functionally annotate QS genes.

QS is a fundamental regulatory mechanism used by many bacteria to control collective traits that allow bacteria to exploit particular niches. For example, QS enables access of symbionts to nutrient-rich environments in hosts (Ruby 1996). Bacterial populations use QS to control biofilm formation, which provides members of the population superior access nutrients and thus enables them to out-compete non-biofilm-producing neighbors (Nadell & Bassler 2011). Finally, bacteria that make their living by exploiting eukaryotic hosts have coupled production of the virulence factors necessary for a pathogenic lifestyle to their ability to detect changes in cell population density.

In this thesis, a census of QS proteins is presented along with its local and chromosomal arrangements. The *luxIR* quorum-sensing genes are characterized by their discontinuous and diverse distribution among gram-negative bacteria along with diverse local gene arrangements. It can be concluded that the topology pattern among QS genes seem to follow a few basic rules—such as the conservation of genes between *luxR* and *luxI* genes, repeating patterns in certain taxa, *etc*.—but their variety is apparently greater than previously indicated (Ahlgren et al. 2011; Hirakawa et al. 2011). The majority of gene arrangement types are the simple topology made up of two genes *R* and *I*. Although, there are combinations where these two genes are separated by a single intervening gene. Interestingly, these intervening genes show a commonalities in terms of function. The most obvious feature of the various circuit topologies is the potential negative regulatory effect of *R* on *I*

which goes in parallel with the well known positive regulatory effect. In other words, *R* seems both to activate and to inhibit the *I* genes in a number of cases. In *RXI* circuits, *R* activates an *X* gene that decreases the effect of *I*. In the *R*2 circuits, the negative effect follows from the overlap between the convergently transcribed *R* and *I* genes (Tsai & Winans 2010). Regulatory circuits in which an element can both activate and inhibit another element are termed incoherent feed forward loops or *IFFL*s (Milo et al. 2002; Alon 2007). In contrast to simple feed forward arrangements, *IFFL*s can exhibit a number of complex behavior patterns (for a review see (Kim et al. 2008)). Perhaps the most important of these is the stabilization of the output signals: while simple feed forward circuits have no inherent limits on their output, IFFL networks have bounded output which ensures robustness against fluctuations in the input signal levels. Most often, QS regulatory circuits are simply referred to as autoinduction loops which, at least in theory, should increase their output without limits. The examples shown in this survey suggest that a stabilizing, negative regulatory pathway is present in many QS systems. It was found experimentally that deletion of *RsaL* or *RsaM* leads to a dramatic increase in AHL production, but the resulting mutants are less virulent than the wild type (Mattiuzzo et al. 2011), which shows, on the other hand, that the negative regulatory path may in fact be a crucial stabilizing element within the QS circuits. There were few instances of the presence of DNA-mobilization (an integrase and a transposase) genes in between *luxR* and *luxI* genes. The function of these DNA-mobilization genes are not known. Previously, there was an experimental evidence of *luxIR*-type quorum-sensing genes located in a mobile transposon in *Serratia* sp. strain ATCC 39006. This strongly suggest that these mobilization units play an important role in the transfer of quorum-sensing units between different bacterial genera and species (Wei et al. 2006). In *B. subtilis*, Auchtung et al. showed that excision and transfer of ICE*Bs1*, a mobile genetic element are regulated by a Phr peptide, the quorum-sensing signal encoded by the same transposon (Auchtung et al. 2005). These studies highlight the potential of horizontal transfer of quorum-sensing genes among different hosts. Seen from this evolutionary perspective we speculate that the appearance of a novel QS system may cause a major change in the lifestyle of a bacterial species, for instance it allows the species to colonize a new nutrient-

source. Consequently, there will be selection pressure against losing the new system by genetic rearrangements. On the other hand, continuously present genetic rearrangements may add new elements to the system, such as the regulation of new genes.

A point of interest is to compare the general features of AHL in gram negative bacteria to *comQXPA* loci in gram positive bacteria. At first sight, the AHL system is very different in many respects. Firstly, its signal molecules, *N*-acyl homoserine lactones, are secondary metabolites, as opposed to the post-translationally modified, ribosomally synthesized ComX. Secondly, the core AHL system is simpler as it contains only two proteins, the signal synthetase and the response regulator. Most AHL systems contain well-defined negative feedback regulatory components (co-expressed repressors, RNA interference etc. (Z Gelencsér, Choudhary, et al. 2012; Choudhary et al. 2013). Recently, it was shown that repression is also part of the ComQXPA locus; ComQ was found to provide negative feedback that modulates the QS response of the signal producer (Oslizlo et al. 2014). In addition, the overall architecture of the com-specific response regulator ComA and all AHL LuxR proteins bear similarities in as much as the C-terminal DNA-binding domain belongs to the same domain type (LuxR family GerE), and the signal-mediating domain is at the N terminal in both proteins. The underlying signaling mechanisms are naturally different since LuxR binds the AHL autoinducer molecule while the N-terminal domain of ComA is phosphorylated by the histidine kinase receptor ComP. Both systems contain conserved gene overlaps (40% of the systems in AHL and 45% in com). The AHL system is present in 252 species, 101 genera, 39 families, 22 orders and 5 classes, all in Proteobacteria (Choudhary et al. 2013; Z Gelencsér, Galbáts, et al. 2012; Z Gelencsér, Choudhary, et al. 2012), while the comQXPA system is present in 28 species, 8 genera, 4 families, 3 orders and 3 classes, all in Firmicutes. The biggest difference is in the variability of gene organization. In the AHL systems, we have a large variety of local gene arrangements, while in the *com* system, the gene order is highly conserved. Both systems seem to evolve autonomously. In terms of sequence similarity, AHL proteins in different topological arrangements are typically closer to their QS-linked homologs in other species (Choudhary et al. 2013; Z

Gelencsér, Galbáts, et al. 2012; Z Gelencsér, Choudhary, et al. 2012), than the homologous proteins in the same genome. The same tendency holds for the com proteins, even though the analysis could only be done for ComP proteins which has clear-cut non-QS homologues in most bacterial genomes. QS-linked ComP proteins cluster together with other ComPs from different genomes rather than clustering with other histidine kinases of the same genome. This shows that ComP has diverged from the other, non-QS linked histidine kinases before the modern bacterial species appeared, i.e. the emergence of comQXPA QS systems is not a recent evolutionary event.

The clustering of QS genes suggest that the topological units might act as regulatory modules that evolve together. QS elements in all the cases seemed to evolve earlier than the respective taxonomic clades. The present classification suggests that the patterns seem to be relatively conserved and their distribution among the various taxonomic groups is not random. This does not mean that one can make broad statements about the functional reasons of this apparent conservation. However, we believe that as more detailed data become available, the question, whether the topological patterns are associated with the distinct biological functions can be answered more easily. As QS circuits are involved in activating a large variety of genes in various bacteria, it can be speculated that there may be no simple correspondence between topology and the regulated functions. While we agree that these questions should be tackled by experiment, some supporting evidence can be gathered by comparing the gene neighborhood of QS circuits. We hypothesize that the known principles of gene expression (co-transcription, repression by proteins or RNA) must be combined into a finite set of topological patterns that allow stable positive autoregulation and control of linked genes.

As a general conclusion it can be pointed out that within the previously known orthologous groups of QS genes, genes linked to a given chemical signal form a distinct orthological subgroup characterized by common local topology within the chromosome.

8 References

Ahlgren, N.A. et al., 2011. Aryl-homoserine lactone quorum sensing in stem-nodulating photosynthetic bradyrhizobia. *Proceedings of the National Academy of Sciences of the United States of America*, 108(17), pp.7183–8.

Alexandersson, M., Cawley, S. & Pachter, L., 2003. SLAM: cross-species gene finding and alignment with a generalized pair hidden Markov model. *Genome research*, 13(3), pp.496–502.

Alon, U., 2007. Network motifs: theory and experimental approaches. *Nature reviews. Genetics*, 8(6), pp.450–61.

Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3), pp.403–10.

Ansaldi, M. et al., 2002. Specific activation of the Bacillus quorum-sensing systems by isoprenylated pheromone variants. *Molecular microbiology*, 44(6), pp.1561–73.

Apweiler, R. et al., 2004. UniProt: the Universal Protein knowledgebase. *Nucleic acids research*, 32(Database issue), pp.D115–9.

Auchtung, J.M. et al., 2005. Regulation of a Bacillus subtilis mobile genetic element by intercellular signaling and the global DNA damage response. *Proceedings of the National Academy of Sciences of the United States of America*, 102(35), pp.12554–9.

Aziz, R.K. et al., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9(1), p.75.

Bassler, B.L., Wright, M. & Silverman, M.R., 1994. Multiple signalling systems controlling expression of luminescence in Vibrio harveyi: sequence and function of genes encoding a second sensory pathway. *Molecular microbiology*, 13(2), pp.273–86.

Batzoglou, S. et al., 2000. Human and mouse gene structure: comparative analysis and application to exon prediction. *Genome research*, 10(7), pp.950–8.

Beck von Bodman, S. & Farrand, S.K., 1995. Capsular polysaccharide biosynthesis and pathogenicity in Erwinia stewartii require induction by

an N-acylhomoserine lactone autoinducer. *Journal of bacteriology*, 177(17), pp.5000–8.

Bertani, I. & Venturi, V., 2004. Regulation of the N-acyl homoserine lactone-dependent quorum-sensing system in rhizosphere Pseudomonas putida WCS358 and cross-talk with the stationary-phase RpoS sigma factor and the global regulator GacA. *Applied and environmental microbiology*, 70(9), pp.5493–502.

Borodovsky, M. & McIninch, J., 1993. GENMARK: Parallel gene recognition for both DNA strands. *Computers & Chemistry*, 17(2), pp.123–133.

Bowers, P.M. et al., 2004. Prolinks: a database of protein functional linkages derived from coevolution. *Genome biology*, 5(5), p.R35.

Bredenbruch, F. et al., 2005. Biosynthetic pathway of Pseudomonas aeruginosa 4-hydroxy-2-alkylquinolines. *Journal of bacteriology*, 187(11), pp.3630–5.

Buell, C.R. et al., 2003. The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000. *Proceedings of the National Academy of Sciences of the United States of America*, 100(18), pp.10181–6.

Burge, C. & Karlin, S., 1997. Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology*, 268(1), pp.78–94.

Cantarel, B.L. et al., 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research*, 18(1), pp.188–96.

Case, R.J., Labbate, M. & Kjelleberg, S., 2008. AHL-driven quorum-sensing circuits: their frequency and function among the Proteobacteria. *The ISME journal*, 2(4), pp.345–9.

Chen, R. et al., 2012. Dissection of quorum-sensing genes in Burkholderia glumae reveals non-canonical regulation and the new regulatory gene tofM for toxoflavin production. *PloS one*, 7(12), p.e52150.

Choi, S.H. & Greenberg, E.P., 1991. The C-terminal region of the Vibrio fischeri LuxR protein contains an inducer-independent lux gene activating domain. *Proceedings of the National Academy of Sciences of the United States of America*, 88(24), pp.11115–9.

Choudhary, K.S. et al., 2013. The Organization of the Quorum Sensing luxI/R Family Genes in Burkholderia. *International journal of molecular sciences*, 14(7), pp.13727–47.

Chun, H. et al., 2009. The quorum sensing-dependent gene katG of Burkholderia glumae is important for protection from visible light. *Journal of bacteriology*, 191(13), pp.4152–7.

Churchill, M.E.A. & Chen, L., 2011. Structural basis of acyl-homoserine lactone-dependent signaling. *Chemical reviews*, 111(1), pp.68–85.

Comella, N. & Grossman, A.D., 2005. Conservation of genes and processes controlled by the quorum response in bacteria: characterization of genes controlled by the quorum-sensing transcription factor ComA in Bacillus subtilis. *Molecular microbiology*, 57(4), pp.1159–74.

Danino, V.E. et al., 2003. Recipient-induced transfer of the symbiotic plasmid pRL1JI in Rhizobium leguminosarum bv. viciae is regulated by a quorum-sensing relay. *Molecular microbiology*, 50(2), pp.511–25.

Delcher, A.L. et al., 1999. Improved microbial gene identification with GLIMMER. *Nucleic acids research*, 27(23), pp.4636–41.

Deng, Y. et al., 2012. Cis-2-dodecenoic acid receptor RpfR links quorum-sensing signal perception with regulation of virulence through cyclic dimeric guanosine monophosphate turnover. *Proceedings of the National Academy of Sciences of the United States of America*, 109(38), pp.15479–84.

Deutscher, J., Francke, C. & Postma, P.W., 2006. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiology and molecular biology reviews : MMBR*, 70(4), pp.939–1031.

Devescovi, G. et al., 2007. Involvement of a quorum-sensing-regulated lipase secreted by a clinical isolate of Burkholderia glumae in severe disease symptoms in rice. *Applied and environmental microbiology*, 73(15), pp.4950–8.

Déziel, E. et al., 2004. Analysis of Pseudomonas aeruginosa 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication. *Proceedings of the National Academy of Sciences of the United States of America*, 101(5), pp.1339–44.

Dhir, S. et al., 2010. Detecting atypical examples of known domain types by sequence similarity searching: the SBASE domain library approach. *Curr Protein Pept Sci*, 11(7), pp.538–549.

Diggle, S.P. et al., 2006. Functional genetic analysis reveals a 2-Alkyl-4-quinolone signaling system in the human pathogen Burkholderia pseudomallei and related bacteria. *Chemistry & biology*, 13(7), pp.701–10.

Diggle, S.P. et al., 2007. The Pseudomonas aeruginosa 4-quinolone signal molecules HHQ and PQS play multifunctional roles in quorum sensing and iron entrapment. *Chemistry & biology*, 14(1), pp.87–96.

Diggle, S.P. et al., 2003. The Pseudomonas aeruginosa quinolone signal molecule overcomes the cell density-dependency of the quorum sensing hierarchy, regulates rhl-dependent genes at the onset of stationary phase and can be produced in the absence of LasR. *Molecular microbiology*, 50(1), pp.29–43.

Dubern, J.-F., Lugtenberg, B.J.J. & Bloemberg, G. V, 2006. The ppuI-rsaL-ppuR quorum-sensing system regulates biofilm formation of Pseudomonas putida PCL1445 by controlling biosynthesis of the cyclic lipopeptides putisolvins I and II. *Journal of bacteriology*, 188(8), pp.2898–906.

Dubnau, D., 1991. Genetic competence in Bacillus subtilis. *Microbiological reviews*, 55(3), pp.395–424.

Duerkop, B. a et al., 2009. Quorum-sensing control of antibiotic synthesis in Burkholderia thailandensis. *Journal of bacteriology*, 191(12), pp.3909–3918.

Duerkop, B.A. et al., 2008. The Burkholderia mallei BmaR3-BmaI3 quorum-sensing system produces and responds to N-3-hydroxy-octanoyl homoserine lactone. *Journal of bacteriology*, 190(14), pp.5137–5141.

Duerkop, B.A., Ulrich, R.L. & Greenberg, E.P., 2007. Octanoyl-homoserine lactone is the cognate signal for Burkholderia mallei BmaR1-BmaI1 quorum sensing. *Journal of bacteriology*, 189(14), pp.5034–40.

Dumenyo, C. & Mukherjee, A., 1998. Genetic and physiological evidence for the production of N-acyl homoserine lactones by Pseudomonas syringae pv. syringae and other fluorescent plant pathogenic Pseudomonas species. *European journal of plant pathology*, (180789), pp.569–582.

Eberhard, A. et al., 1981. Structural identification of autoinducer of Photobacterium fischeri luciferase. *Biochemistry*, 20(9), pp.2444–9.

Eberl, L., 2006. Quorum sensing in the genus Burkholderia. *International journal of medical microbiology : IJMM*, 296(2-3), pp.103–110.

Eddy, S.R., 2008. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS computational biology*, 4(5), p.e1000069.

Eddy, S.R., 1998. Profile hidden Markov models. *Bioinformatics (Oxford, England)*, 14(9), pp.755–63.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), pp.1792–7.

Elasri, M. et al., 2001. Acyl-homoserine lactone production is more common among plant-associated Pseudomonas spp. than among soilborne Pseudomonas spp. *Applied and environmental microbiology*, 67(3), pp.1198–209.

El-Sayed, A.K. et al., 2003. Characterization of the mupirocin biosynthesis gene cluster from Pseudomonas fluorescens NCIMB 10586. *Chemistry & biology*, 10(5), pp.419–30.

El-Sayed, A.K., Hothersall, J. & Thomas, C.M., 2001. Quorum-sensing-dependent regulation of biosynthesis of the polyketide antibiotic mupirocin in Pseudomonas fluorescens NCIMB 10586. *Microbiology (Reading, England)*, 147(Pt 8), pp.2127–39.

Falquet, L. et al., 2002. The PROSITE database, its status in 2002. *Nucleic acids research*, 30(1), pp.235–8.

Feil, H. et al., 2005. Comparison of the complete genome sequences of Pseudomonas syringae pv. syringae B728a and pv. tomato DC3000. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31), pp.11064–9.

Fukuda, Y., Nakayama, Y. & Tomita, M., 2003. On dynamics of overlapping genes in bacterial genomes. *Gene*, 323, pp.181–7.

Fuqua, C. & Greenberg, E.P., 2002. Listening in on bacteria: acyl-homoserine lactone signalling. *Nature reviews. Molecular cell biology*, 3(9), pp.685–95.

Fuqua, W.C., Winans, S.C. & Greenberg, E.P., 1994. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *Journal of bacteriology*, 176(2), pp.269–75.

Galperin, M.Y., 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC microbiology*, 5, p.35.

Gamage, A.M. et al., 2011. N-Octanoylhomoserine lactone signalling mediated by the BpsI-BpsR quorum sensing system plays a major role in biofilm formation of Burkholderia pseudomallei. *Microbiology*, 157(Pt 4), pp.1176–86.

Gambello, M.J. & Iglewski, B.H., 1991. Cloning and characterization of the Pseudomonas aeruginosa lasR gene, a transcriptional activator of elastase expression. *J. Bacteriol.*, 173(9), pp.3000–3009.

Gambello, M.J., Kaye, S. & Iglewski, B.H., 1993. LasR of Pseudomonas aeruginosa is a transcriptional activator of the alkaline protease gene (apr) and an enhancer of exotoxin A expression. *Infect. Immun.*, 61(4), pp.1180–1184.

Ganley, A.R.D. & Kobayashi, T., 2007. Phylogenetic footprinting to find functional DNA elements. *Methods in molecular biology (Clifton, N.J.)*, 395, pp.367–80.

Gelencsér, Z., Galbáts, B., et al., 2012. Chromosomal arrangement of AHL.driven quorum sensing circuits in Pseudomonas. *ISRN Microbiology*, 2012, p.6.

Gelencsér, Z., Galbáts, B., et al., 2012. Chromosomal Arrangement of AHL-Driven Quorum Sensing Circuits in Pseudomonas. *ISRN Microbiology*, 2012, pp.1–6.

Gelencsér, Z., Choudhary, K.S., et al., 2012. Classifying the topology of AHL-driven quorum sensing circuits in proteobacterial genomes. *Sensors*, 12(5), pp.5432–5444.

Gelencsér, Z., Choudhary, K.S., et al., 2012. Classifying the topology of AHL-driven quorum sensing circuits in proteobacterial genomes. *Sensors (Basel, Switzerland)*, 12(5), pp.5432–5444.

Gillis, M. et al., 1995. Polyphasic Taxonomy in the Genus Burkholderia Leading to an Emended Description of the Genus and Proposition of Burkholderia vietnamiensis sp. nov. for N2-Fixing Isolates from Rice in

Vietnam. *International Journal of Systematic Bacteriology*, 45(2), pp.274–289.

Goryachev, A.B., Toh, D.J. & Lee, T., 2006. Systems analysis of a quorum sensing network: design constraints imposed by the functional requirements, network topology and kinetic constants. *Bio Systems*, 83(2-3), pp.178–87.

Guo, Y. et al., 2012. Diffusible signal factor-mediated quorum sensing plays a central role in coordinating gene expression of Xanthomonas citri subsp. citri. *Molecular plant-microbe interactions : MPMI*, 25(2), pp.165–79.

Hirakawa, H. et al., 2011. Activity of the Rhodopseudomonas palustris p-coumaroyl-homoserine lactone-responsive transcription factor RpaR. *Journal of bacteriology*, 193(10), pp.2598–607.

Hodgkinson, J. et al., 2010. Structure-activity analysis of the Pseudomonas quinolone signal molecule. *Journal of bacteriology*, 192(14), pp.3833–7.

Holden, M.T.G. et al., 2004. Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. *Proceedings of the National Academy of Sciences of the United States of America*, 101(39), pp.14240–5.

Hothersall, J. et al., 2011. Manipulation of quorum sensing regulation in Pseudomonas fluorescens NCIMB 10586 to increase mupirocin production. *Applied microbiology and biotechnology*, 90(3), pp.1017–26.

Hunter, S. et al., 2012. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic acids research*, 40(Database issue), pp.D306–12.

Joardar, V. et al., 2005. Whole-genome sequence analysis of Pseudomonas syringae pv. phaseolicola 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *Journal of bacteriology*, 187(18), pp.6488–98.

Johnson, Z.I. & Chisholm, S.W., 2004. Properties of overlapping genes are conserved across microbial genomes. *Genome research*, 14(11), pp.2268–72.

Kaplan, H.B. & Greenberg, E.P., 1987. Overproduction and purification of the luxR gene product: Transcriptional activator of the Vibrio fischeri luminescence system. *Proceedings of the National Academy of Sciences of the United States of America*, 84(19), pp.6639–43.

Kerkhoven, R. et al., 2004. Visualization for genomics: the Microbial Genome Viewer. *Bioinformatics (Oxford, England)*, 20(11), pp.1812–4.

Khan, S.R. et al., 2007. N-(3-hydroxyhexanoyl)-l-homoserine lactone is the biologically relevant quormone that regulates the phz operon of Pseudomonas chlororaphis strain 30-84. *Applied and environmental microbiology*, 73(22), pp.7443–55.

De Kievit, T. et al., 1999. RsaL, a novel repressor of virulence gene expression in Pseudomonas aeruginosa. *Journal of bacteriology*, 181(7), pp.2175–84.

Kim, D., Kwon, Y.-K. & Cho, K.-H., 2008. The biphasic behavior of incoherent feed-forward loops in biomolecular regulatory networks. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 30(11-12), pp.1204–11.

Kim, J. et al., 2007. Regulation of polar flagellum genes is mediated by quorum sensing and FlhDC in Burkholderia glumae. *Molecular microbiology*, 64(1), pp.165–79.

Kiratisin, P. & Sanmee, S., 2008. Roles and interactions of Burkholderia pseudomallei BpsIR quorum-sensing system determinants. *Journal of bacteriology*, 190(21), pp.7291–7.

Kleerebezem, M. et al., 1997. Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. *Molecular Microbiology*, 24(5), pp.895–904.

Koonin EV, G.M., 2003. Genome Annotation and Analysis. In *Sequence - Evolution - Function: Computational Approaches in Comparative Genomics*. Boston: Kluwer Academic.

Korf, I. et al., 2001. Integrating genomic homology into gene structure prediction. *Bioinformatics (Oxford, England)*, 17 Suppl 1, pp.S140–8.

Kuipers, O.P. et al., 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. *Journal of Biotechnology*, 64(1), pp.15–21.

Larsen, J.C. & Johnson, N.H., 2009. Pathogenesis of Burkholderia pseudomallei and Burkholderia mallei. *Military medicine*, 174(6), pp.647–651.

Latifi, A. et al., 1996. A hierarchical quorum-sensing cascade in Pseudomonas aeruginosa links the transcriptional activators LasR and RhlR (VsmR) to

expression of the stationary-phase sigma factor RpoS. *Molecular Microbiology*, 21(6), pp.1137–1146.

Latifi, A. et al., 1995. Multiple homologues of LuxR and LuxI control expression of virulence determinants and secondary metabolites through quorum sensing in Pseudomonas aeruginosa PAO1. *Molecular microbiology*, 17(2), pp.333–43.

Laue, B.E. et al., 2000. The biocontrol strain Pseudomonas fluorescens F113 produces the Rhizobium small bacteriocin, N-(3-hydroxy-7-cis-tetradecenoyl)homoserine lactone, via HdtS, a putative novel N-acylhomoserine lactone synthase. *Microbiology (Reading, England)*, 146 (Pt 1, pp.2469–80.

Lazdunski, A.M., Ventre, I. & Sturgis, J.N., 2004. Regulatory circuits and communication in Gram-negative bacteria. *Nature reviews. Microbiology*, 2(7), pp.581–92.

Ledgham, F. et al., 2003. Interactions of the quorum sensing regulator QscR: interaction with itself and the other regulators of Pseudomonas aeruginosa LasR and RhlR. *Molecular microbiology*, 48(1), pp.199–210.

Lee, D.G. et al., 2006. Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial. *Genome biology*, 7(10), p.R90.

Lenhard, B. et al., 2003. Identification of conserved regulatory elements by comparative genome analysis. *Journal of biology*, 2(2), p.13.

Lewenza, S. & Sokol, P.A., 2001. Regulation of ornibactin biosynthesis and N-acyl-L-homoserine lactone production by CepR in Burkholderia cepacia. *Journal of bacteriology*, 183(7), pp.2212–8.

Lindemann, A. et al., 2011. Isovaleryl-homoserine lactone, an unusual branched-chain quorum-sensing signal from the soybean symbiont Bradyrhizobium japonicum. *Proceedings of the National Academy of Sciences of the United States of America*, 108(40), pp.16765–70.

López, D. et al., 2009. Paracrine signaling in a bacterium. *Genes & development*, 23(14), pp.1631–8.

Malott, R.J. et al., 2009. A Burkholderia cenocepacia orphan LuxR homolog is involved in quorum-sensing regulation. *Journal of bacteriology*, 191(8), pp.2447–60.

Malott, R.J. et al., 2005. Characterization of the cciIR quorum-sensing system in Burkholderia cenocepacia. *Infection and immunity*, 73(8), pp.4982–92.

Malott, R.J. & Sokol, P.A., 2007. Expression of the bviIR and cepIR quorum-sensing systems of Burkholderia vietnamiensis. *Journal of bacteriology*, 189(8), pp.3006–16.

Markowitz, V.M. et al., 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic acids research*, 40(Database issue), pp.D115–22.

Martinez-Guerrero, C.E. et al., 2008. GeConT 2: gene context analysis for orthologous proteins, conserved domains and metabolic pathways. *Nucleic acids research*, 36(Web Server issue), pp.W176–80.

Mattiuzzo, M. et al., 2011. The plant pathogen Pseudomonas fuscovaginae contains two conserved quorum sensing systems involved in virulence and negatively regulated by RsaL and the novel regulator RsaM. *Environmental microbiology*, 13(1), pp.145–62.

Mavrodi, D. V et al., 1998. A seven-gene locus for synthesis of phenazine-1-carboxylic acid by Pseudomonas fluorescens 2-79. *Journal of bacteriology*, 180(9), pp.2541–8.

Von Mering, C. et al., 2007. STRING 7--recent developments in the integration and prediction of protein interactions. *Nucleic acids research*, 35(Database issue), pp.D358–62.

Miller, M.B. & Bassler, B.L., 2001. Quorum sensing in bacteria. *Annual review of microbiology*, 55, pp.165–99.

Milo, R. et al., 2002. Network motifs: simple building blocks of complex networks. *Science (New York, N.Y.)*, 298(5594), pp.824–7.

Milton, D.L. et al., 2001. The LuxM homologue VanM from Vibrio anguillarum directs the synthesis of N-(3-hydroxyhexanoyl)homoserine lactone and N-hexanoylhomoserine lactone. *Journal of bacteriology*, 183(12), pp.3537–47.

Minogue, T.D. et al., 2002. The autoregulatory role of EsaR, a quorum-sensing regulator in Pantoea stewartii ssp. stewartii: evidence for a repressor function. *Molecular microbiology*, 44(6), pp.1625–35.

Minogue, T.D. et al., 2005. The cell density-dependent expression of stewartan exopolysaccharide in Pantoea stewartii ssp. stewartii is a function of

EsaR-mediated repression of the rcsA gene. *Molecular microbiology*, 56(1), pp.189–203.

Nadell, C.D. & Bassler, B.L., 2011. A fitness trade-off between local competition and dispersal in Vibrio cholerae biofilms. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34), pp.14181–5.

Nealson, K.H. & Hastings, J.W., 1979. Bacterial bioluminescence: its control and ecological significance. *Microbiological reviews*, 43(4), pp.496–518.

Nealson, K.H., Platt, T. & Hastings, J.W., 1970. Cellular control of the synthesis and activity of the bacterial luminescent system. *Journal of bacteriology*, 104(1), pp.313–22.

Needleman, S.B. & Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3), pp.443–53.

O'Grady, E.P. et al., 2009a. Reciprocal regulation by the CepIR and CciIR quorum sensing systems in Burkholderia cenocepacia. *BMC genomics*, 10(1), p.441.

O'Grady, E.P. et al., 2009b. Reciprocal regulation by the CepIR and CciIR quorum sensing systems in Burkholderia cenocepacia. *BMC genomics*, 10(1), p.441.

O'Grady, E.P., Viteri, D.F. & Sokol, P. a, 2012. A unique regulator contributes to quorum sensing and virulence in Burkholderia cenocepacia. *PloS one*, 7(5), p.e37611.

Ochsner, U.A. et al., 1994. Isolation and characterization of a regulatory gene affecting rhamnolipid biosurfactant synthesis in Pseudomonas aeruginosa. *Journal of bacteriology*, 176(7), pp.2044–54.

Okada, M. et al., 2005. Structure of the Bacillus subtilis quorum-sensing peptide pheromone ComX. *Nature chemical biology*, 1(1), pp.23–4.

Oldroyd, G.E.D., Harrison, M.J. & Udvardi, M., 2005. Peace talks and trade deals. Keys to long-term harmony in legume-microbe symbioses. *Plant physiology*, 137(4), pp.1205–10.

Oslizlo, A. et al., 2014. Private link between signal and response in Bacillus subtilis quorum sensing. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, pp.1–6.

Overbeek, R. et al., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic acids research*, 33(17), pp.5691–702.

Overbeek, R. et al., 1999a. The use of gene clusters to infer functional coupling. *Proceedings of the National Academy of Sciences of the United States of America*, 96(6), pp.2896–901.

Overbeek, R. et al., 1999b. Use of contiguity on the chromosome to predict functional coupling. *In silico biology*, 1(2), pp.93–108.

Pai, A. & You, L., 2009. Optimal tuning of bacterial sensing potential. *Molecular systems biology*, 5, p.286.

Palleroni, N.J. et al., 1973. Nucleic Acid Homologies in the Genus Pseudomonas. *International Journal of Systematic Bacteriology*, 23(4), pp.333–339.

Parsek, M.R. & Greenberg, E.P., 2000. Acyl-homoserine lactone quorum sensing in gram-negative bacteria: a signaling mechanism involved in associations with higher organisms. *Proceedings of the National Academy of Sciences of the United States of America*, 97(16), pp.8789–93.

Patankar, A. V & González, J.E., 2009. Orphan LuxR regulators of quorum sensing. *FEMS microbiology reviews*, 33(4), pp.739–56.

Pearson, W.R. & Lipman, D.J., 1988. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, 85(8), pp.2444–8.

Pesci, E.C. et al., 1999. Quinolone signaling in the cell-to-cell communication system of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America*, 96(20), pp.11229–34.

Petersen, F.C., Fimland, G. & Scheie, A.A., 2006. Purification and functional studies of a potent modified quorum-sensing peptide and a two-peptide bacteriocin in Streptococcus mutans. *Molecular microbiology*, 61(5), pp.1322–34.

Piazza, F., Tortosa, P. & Dubnau, D., 1999. Mutational analysis and membrane topology of ComP, a quorum-sensing histidine kinase of Bacillus subtilis controlling competence development. *Journal of bacteriology*, 181(15), pp.4540–8.

Pierson, L.S., Keppenne, V.D. & Wood, D.W., 1994. Phenazine antibiotic biosynthesis in Pseudomonas aureofaciens 30-84 is regulated by PhzR in response to cell density. *Journal of bacteriology*, 176(13), pp.3966–74.

Plata, G. et al., 2012. Global probabilistic annotation of metabolic networks enables enzyme discovery. *Nature chemical biology*, 8(10), pp.848–54.

Quevillon, E. et al., 2005. InterProScan: protein domains identifier. *Nucleic acids research*, 33(Web Server issue), pp.W116–20.

Rampioni, G. et al., 2012. Functional characterization of the quorum sensing regulator RsaL in the plant-beneficial strain Pseudomonas putida WCS358. *Applied and environmental microbiology*, 78(3), pp.726–34.

Rampioni, G. et al., 2007. The Pseudomonas quorum-sensing regulator RsaL belongs to the tetrahelical superclass of H-T-H proteins. *Journal of bacteriology*, 189(5), pp.1922–30.

Rampioni, G. et al., 2006. The quorum-sensing negative regulator RsaL of Pseudomonas aeruginosa binds to the lasI promoter. *Journal of bacteriology*, 188(2), pp.815–9.

Reading, N.C. & Sperandio, V., 2006. Quorum sensing: the many languages of bacteria. *FEMS microbiology letters*, 254(1), pp.1–11.

Roggiani, M. & Dubnau, D., 1993. ComA, a phosphorylated response regulator protein of Bacillus subtilis, binds to the promoter region of srfA. *Journal of bacteriology*, 175(10), pp.3182–7.

Ropelewski, A.J., Nicholas, H.B. & Gonzalez Mendez, R.R., 2010. MPI-PHYLIP: parallelizing computationally intensive phylogenetic analysis routines for the analysis of large protein families. I. K. Jordan, ed. *PloS one*, 5(11), p.e13999.

Roy, P.H. et al., 2010. Complete genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7. *PloS one*, 5(1), p.e8842.

Ruby, E.G., 1996. Lessons from a cooperative, bacterial-animal association: the Vibrio fischeri-Euprymna scolopes light organ symbiosis. *Annual review of microbiology*, 50, pp.591–624.

Rutherford, S.T. & Bassler, B.L., 2012. Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harbor perspectives in medicine*, 2(11), p.a012427–.

Sappington, K.J. et al., 2011. Reversible signal binding by the Pseudomonas aeruginosa quorum-sensing signal receptor LasR. *mBio*, 2(1), pp.e00011–11.

Schaefer, A.L. et al., 2008. A new class of homoserine lactone quorum-sensing signals. *Nature*, 454(7204), pp.595–9.

Schaefer, A.L. et al., 1996. Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified Vibrio fischeri LuxI protein. *Proceedings of the National Academy of Sciences of the United States of America*, 93(18), pp.9505–9.

Scharf, M. et al., 1994. GeneQuiz: a workbench for sequence analysis. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 2, pp.348–53.

Schmidt, S. et al., 2009. Production of the antifungal compound pyrrolnitrin is quorum sensing-regulated in members of the Burkholderia cepacia complex. *Environmental microbiology*, 11(6), pp.1422–37.

Schuster, M. et al., 2003. Identification, Timing, and Signal Specificity of Pseudomonas aeruginosa Quorum-Controlled Genes: a Transcriptome Analysis. *Journal of Bacteriology*, 185(7), pp.2066–2079.

Schuster, M. & Greenberg, E.P., 2006. A network of networks: quorum-sensing gene regulation in Pseudomonas aeruginosa. *International journal of medical microbiology : IJMM*, 296(2-3), pp.73–81.

Seed, P., Passador, L. & Iglewski, B., 1995. Activation of the Pseudomonas aeruginosa lasI gene by LasR and the Pseudomonas autoinducer PAI: an autoinduction regulatory hierarchy. *J. Bacteriol.*, 177(3), pp.654–659.

Seyedsayamdost, M.R. et al., 2010. Quorum-sensing-regulated bactobolin production by Burkholderia thailandensis E264. *Organic letters*, 12(4), pp.716–9.

Siddiqui, M.F. et al., 2012. Targeting N-acyl-homoserine-lactones to mitigate membrane biofouling based on quorum sensing using a biofouling reducer. *Journal of biotechnology*, 161(3), pp.190–7.

Singh, P.K. et al., 2000. Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature*, 407(6805), pp.762–4.

Sitnikov, D.M., Schineller, J.B. & Baldwin, T.O., 1995. Transcriptional regulation of bioluminesence genes from Vibrio fischeri. *Molecular microbiology*, 17(5), pp.801–12.

Smith, R.S. & Iglewski, B.H., 2003. P. aeruginosa quorum-sensing systems and virulence. *Current opinion in microbiology*, 6(1), pp.56–60.

Smith, T.F. & Waterman, M.S., 1981. Identification of common molecular subsequences. *Journal of molecular biology*, 147(1), pp.195–7.

Solis, R. et al., 2006. Involvement of quorum sensing and RpoS in rice seedling blight caused by Burkholderia plantarii. *FEMS microbiology letters*, 259(1), pp.106–12.

Steidle, A. et al., 2002. Identification and characterization of an N-acylhomoserine lactone-dependent quorum-sensing system in Pseudomonas putida strain IsoF. *Applied and environmental microbiology*, 68(12), pp.6371–82.

Steindler, L. et al., 2008. The presence, type and role of N-acyl homoserine lactone quorum sensing in fluorescent Pseudomonas originally isolated from rice rhizospheres are unpredictable. *FEMS microbiology letters*, 288(1), pp.102–11.

Stevens, A.M. et al., 2011. Mechanisms and synthetic modulators of AHL-dependent gene regulation. *Chemical reviews*, 111(1), pp.4–27.

Stover, C.K. et al., 2000. Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature*, 406(6799), pp.959–64.

Suárez-Moreno, Z.R. et al., 2012. Common features of environmental and potentially beneficial plant-associated Burkholderia. *Microbial ecology*, 63(2), pp.249–66.

Suárez-Moreno, Z.R. et al., 2010. Commonalities and differences in regulation of N-acyl homoserine lactone quorum sensing in the beneficial plant-associated burkholderia species cluster. *Applied and environmental microbiology*, 76(13), pp.4302–17.

Suárez-Moreno, Z.R., Caballero-Mellado, J. & Venturi, V., 2008. The new group of non-pathogenic plant-associated nitrogen-fixing Burkholderia spp. shares a conserved quorum-sensing system, which is tightly regulated by the RsaL repressor. *Microbiology (Reading, England)*, 154(Pt 7), pp.2048–59.

Subramoni, S. et al., 2011. Bacterial subfamily of LuxR regulators that respond to plant compounds. *Applied and environmental microbiology*, 77(13), pp.4579–88.

Subramoni, S. & Venturi, V., 2009. LuxR-family "solos": bachelor sensors/regulators of signalling molecules. *Microbiology*, 155(Pt 5), pp.1377–85.

Tamura, K. et al., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, 28(10), pp.2731–9.

Tanenbaum, D.M. et al., 2010. The JCVI standard operating procedure for annotating prokaryotic metagenomic shotgun sequencing data. *Standards in genomic sciences*, 2(2), pp.229–37.

Tatusov, R.L. et al., 2003. The COG database: an updated version includes eukaryotes. *BMC bioinformatics*, 4(1), p.41.

Thiel, V. et al., 2009. New structural variants of homoserine lactones in bacteria. *Chembiochem : a European journal of chemical biology*, 10(11), pp.1861–8.

Thoendel, M. et al., 2011. Peptide signaling in the staphylococci. *Chemical reviews*, 111(1), pp.117–51.

Thoendel, M. & Horswill, A.R., 2010. *Biosynthesis of peptide signals in gram-positive bacteria.* 1st ed., Elsevier Inc.

Thompson, J.D., Higgins, D.G. & Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), pp.4673–80.

Tsai, C.-S. & Winans, S.C., 2010. LuxR-type quorum-sensing regulators that are detached from common scents. *Molecular microbiology*, 77(5), pp.1072–82.

Ulrich, L.E., Koonin, E. V & Zhulin, I.B., 2005. One-component systems dominate signal transduction in prokaryotes. *Trends in microbiology*, 13(2), pp.52–6.

Ulrich, R.L. et al., 2004. Mutational analysis and biochemical characterization of the Burkholderia thailandensis DW503 quorum-sensing network. *Journal of bacteriology*, 186(13), pp.4350–60.

Ulrich, R.L., 2004. Role of quorum sensing in the pathogenicity of Burkholderia pseudomallei. *Journal of Medical Microbiology*, 53(11), pp.1053–1064.

Urbanowski, M.L., Lostroh, C.P. & Greenberg, E.P., 2004. Reversible acyl-homoserine lactone binding to purified Vibrio fischeri LuxR protein. *Journal of bacteriology*, 186(3), pp.631–7.

Valade, E. et al., 2004. The PmlI-PmlR quorum-sensing system in Burkholderia pseudomallei plays a key role in virulence and modulates production of the MprA protease. *Journal of bacteriology*, 186(8), pp.2288–94.

Venturi, V. et al., 2004. Quorum sensing in the Burkholderia cepacia complex. *Research in microbiology*, 155(4), pp.238–44.

Venturi, V., 2006. Regulation of quorum sensing in Pseudomonas. *FEMS microbiology reviews*, 30(2), pp.274–91.

Venturi, V. et al., 2011. The virtue of temperance: built-in negative regulators of quorum sensing in Pseudomonas. *Molecular microbiology*, 82(5), pp.1060–70.

Vlahovicek, K. et al., 2005. The SBASE domain sequence resource, release 12: prediction of protein domain-architecture using support vector machines. *Nucleic Acids Res*, 33(Database issue), pp.D223–5.

Wade, D.S. et al., 2005. Regulation of Pseudomonas quinolone signal synthesis in Pseudomonas aeruginosa. *Journal of bacteriology*, 187(13), pp.4372–80.

Wadhams, G.H. & Armitage, J.P., 2004. Making sense of it all: bacterial chemotaxis. *Nature reviews. Molecular cell biology*, 5(12), pp.1024–37.

Waters, C.M. & Bassler, B.L., 2005. Quorum sensing: cell-to-cell communication in bacteria. *Annual review of cell and developmental biology*, 21, pp.319–46.

Waters, C.M. & Bassler, B.L., 2006. The Vibrio harveyi quorum-sensing system uses shared regulatory components to discriminate between multiple autoinducers. *Genes & development*, 20(19), pp.2754–67.

Watson, W.T. et al., 2002. Structural basis and specificity of acyl-homoserine lactone signal production in bacterial quorum sensing. *Molecular cell*, 9(3), pp.685–94.

Wei, J.-R. et al., 2006. A mobile quorum-sensing system in Serratia marcescens. *Journal of bacteriology*, 188(4), pp.1518–25.

Weinrauch, Y. et al., 1990. A Bacillus subtilis regulatory gene product for genetic competence and sporulation resembles sensor protein members of the bacterial two-component signal-transduction systems. *Genes & development*, 4(5), pp.860–72.

Whitehead, N.A. et al., 2001. Quorum-sensing in Gram-negative bacteria. *FEMS microbiology reviews*, 25(4), pp.365–404.

Whitlock, G.C., Estes, D.M. & Torres, A.G., 2007. Glanders: off to the races with Burkholderia mallei. *FEMS microbiology letters*, 277(2), pp.115–22.

Winstanley, C. et al., 2009. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of Pseudomonas aeruginosa. *Genome research*, 19(1), pp.12–23.

Winzer, K. et al., 2002. LuxS: its role in central metabolism and the in vitro synthesis of 4-hydroxy-5-methyl-3(2H)-furanone. *Microbiology*, 148(4), pp.909–922.

Woese, C.R., 1987. Bacterial evolution. *Microbiological reviews*, 51(2), pp.221–71.

Wolf, Y.I. et al., 2001. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome research*, 11(3), pp.356–72.

Wood, D.W. et al., 1997. N-acyl-homoserine lactone-mediated regulation of phenazine gene expression by Pseudomonas aureofaciens 30-84 in the wheat rhizosphere. *Journal of bacteriology*, 179(24), pp.7663–70.

Wood, D.W. & Pierson, L.S., 1996. The phzI gene of Pseudomonas aureofaciens 30-84 is responsible for the production of a diffusible signal required for phenazine antibiotic production. *Gene*, 168(1), pp.49–53.

Xiao, G., He, J. & Rahme, L.G., 2006. Mutation analysis of the Pseudomonas aeruginosa mvfR and pqsABCDE gene promoters demonstrates complex quorum-sensing circuitry. *Microbiology (Reading, England)*, 152(Pt 6), pp.1679–86.

Yabuuchi, E. et al., 1992. Proposal of Burkholderia gen. nov. and transfer of seven species of the genus Pseudomonas homology group II to the new

genus, with the type species Burkholderia cepacia (Palleroni and Holmes 1981) comb. nov. *Microbiology and immunology*, 36(12), pp.1251–75.

Yanai, I., Derti, A. & DeLisi, C., 2001. Genes linked by fusion events are generally of the same functional category: a systematic analysis of 30 microbial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 98(14), pp.7940–5.

Yandell, M. & Ence, D., 2012. A beginner's guide to eukaryotic genome annotation. *Nature reviews. Genetics*, 13(5), pp.329–42.

Zhang, Z. & Pierson, L.S., 2001. A second quorum-sensing system regulates cell surface properties but not phenazine antibiotic production in Pseudomonas aureofaciens. *Applied and environmental microbiology*, 67(9), pp.4305–15.

Zhou, H. et al., 2003. AHL-Deficient Mutants of Burkholderia ambifaria BC-F Have Decreased Antifungal Activity. *Current Microbiology*, 47(3), pp.174–179.

Zhu, J. & Winans, S.C., 1999. Autoinducer binding by the quorum-sensing regulator TraR increases affinity for target promoters in vitro and decreases TraR turnover rates in whole cells. *Proceedings of the National Academy of Sciences of the United States of America*, 96(9), pp.4832–7.

Zhu, J. & Winans, S.C., 2001. The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. *Proceedings of the National Academy of Sciences of the United States of America*, 98(4), pp.1507–12.