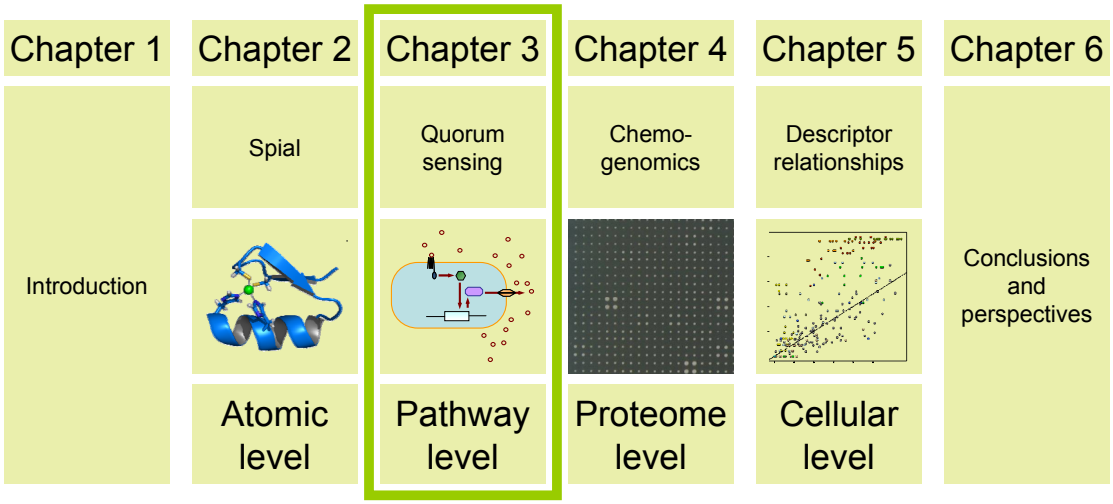
3 Quorum sensing



Contents of chapter 3

Summary of chapter 3

Cell-to-cell signalling is a prerequisite for the development of multicellular organisms such as animals and plants, but has also evolved in groups which would not usually be described as multicellular, such as bacteria and unicellular fungi. Quorum sensing is a process of cell-to-cell communication by which individual cells regulate their phenotype in response to the extracellular concentration of small molecules. This is achieved by the secretion of small molecules into the environment that bind sensory proteins and directly or indirectly affect transcription and translation. In this chapter, I use a combination of computational methods to identify quorum sensing systems in sequenced bacterial genomes. Furthermore, I establish a framework for the identification of transcription factors involved in the quorum sensing response in the yeast *Saccharomyces cerevisiae*.

Parts of this chapter appear in the following peer-reviewed articles:

Wuster, A. and M. Madan Babu (2008). “Conservation and evolutionary dynamics of the *agr* cell-to-cell communication system across firmicutes.” J Bacteriol **190**(2): 743-6.

Wuster, A. and M. Madan Babu (2008). “Chemical Molecules that Regulate Transcription and Facilitate Cell-to-cell Communication.” book chapter in Wiley Encyclopedia of Chemical Biology, Hoboken, New Jersey: John Wiley and Sons, Inc.

Wuster, A. and M. Madan Babu (2009). “Transcriptional control of the quorum sensing response in yeast.” Mol BioSyst, **DOI:** 10.1039/b913579k.

Introduction to quorum sensing

Quorum sensing (QS) is a way for individual cells to exchange information using small molecules (SMs) that bind sensory proteins and thus directly or indirectly affect transcription and translation. The binding threshold is assumed to be reached once the growing population, and hence the concentration of the secreted SM, attains a certain level. In the following, I use the term 'QS system' to mean a cell-to-cell communication system in unicellular organisms that functions via the secretion of communication molecules into the environment and their subsequent binding to sensor proteins. The term ‘communication molecule’ as I use it here refers to any SM that is secreted by one organism and changes the behaviour of another one. As I will

discuss later, it is not necessary that the communication molecule synthesis or detection machinery has evolved for this purpose, although this is often the case. Different systems can be distinguished by the different types of communication molecule they use, which are normally also associated with different types of signal synthesis, import and export, reception and response machinery (Pai and You 2009).

The study of cell-to-cell communication and its effects on transcription in unicellular organisms promises a variety of practical applications. One of them is the possibility of interfering with intercellular communication systems in pathogenic microbes, a process also referred to as quorum quenching. For this, it is first necessary to know which quorum systems are present in which genomes. To answer this is the main aim of this chapter. I approach the problem in three stages.

The first stage is a knowledge-based approach to discovering QS systems in a library of 384 fully sequenced microbial genomes. The second stage focuses on two specific QS systems, which are the *γ*-butyrolactone and the *agr* QS systems. I present evidence that the *agr* system is present across the bacterial phylum of firmicutes, including the human pathogen *Clostridium perfringens*. The third stage deals with the QS system of the yeast *Saccharomyces cerevisiae*. By employing an integrated *in silico* approach that makes use of large-scale genomics datasets, I identify key transcriptional regulators that control the differential expression of the genes affected by QS in *S. cerevisiae*.

For the remainder of this introduction, I will define QS more rigorously and introduce the QS systems that I will analyse in more detail later in this chapter.

What quorum sensing is

Whilst the current consensus seems to be that QS is a way of signalling between bacterial cells, there are a number of reasons why this may not always be the case. Firstly, signalling is the exchange of information between two willing partners (Mullard 2009). In other words, both the sender and the receiver of a communication molecule have to benefit from the process for it to qualify as signalling (table 3-1). This is not necessarily always the case in QS. It is often not clear whether the exchange of SMs has evolved for the purpose of signalling, or whether the SMs are, for example, metabolic side-products in one cell that just happen to have an influence on another cell.

Communication fitness effect to the...		**... sender**	
		Beneficial	Neutral/deleterious
... receiver	Beneficial	A. Signal	B. Cue/Spying
	Neutral/deleterious	C. Manipulation	D. ?

Table 3-1. Only the exchange of information that is beneficial to both the receiver and the sender should be called signal. Considering this, some SMs that have been called QS signals may not actually be signals.

Secondly, bacteria may secrete SMs that change their and other's behaviour without QS (in the sense of cell density sensing) occurring (figure 3-1). It has been shown that in addition to population density, diffusion barriers can also be sensed in the same way. For example, *Staphylococcus aureus* can induce QS-dependent genes when confined in a host endosome (Qazi, Counil et al. 2001). It has even been proposed that most of the pathways attributed to QS in fact are diffusion sensing pathways (Redfield 2002), although evidence for this is relatively scarce. As there seems to be no way to measure cell density that does not at the same time measure diffusion and spatial distribution, it may not be correct to assume that QS is always about cell density.

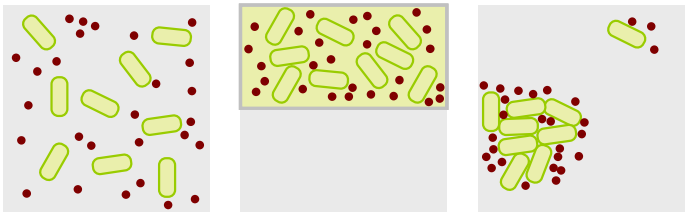


Figure 3-1. The secretion of SMs, based on their concentration, can be used to gather different types of information about the cell's environment. Only one of them is cell density (A). Two others are mass transfer or diffusion (B) and spatial distribution or clustering (C).

In some cases it may also be possible for bacteria that cluster together to have a different effect than the same bacteria spread over a greater volume. For example, bacteria of the genus *Bacillus* cause coagulation of human and mouse blood only when in clusters (Kastrup, Boedicker et al. 2008). This effect is different compared to non-clustered bacilli and happens without any apparent change in gene expression, or any detectable exchange of communication molecules.

In the following section, I will give an introduction to known QS systems that I will analyse using computational methods later in this chapter.

Cell-to-cell communication in bacteria

QS involves dedicated cellular systems for the production and detection of communication molecules, sometimes called quormones (quorum sensing pheromones). In bacterial species that employ QS, each cell secretes a basal amount of communication molecules at low cell density. As cell density increases, communication molecule concentration also increases, provided that the cells are not too far apart. Communication molecules bind to special receptors once their concentration exceeds a certain threshold. This in turn produces the physiological response.

In bacteria, QS regulated phenotypes include bioluminescence, exopolysaccharide production, virulence, conjugal plasmid transfer, antibiotic and exoenzyme production, biofilm formation, and growth inhibition (Lazdunski, Ventre et al. 2004). Types of communication molecules discussed in this section are acyl homoserine lactones (AHLs), AI-2 molecules, and modified oligopeptides (table 3-2). AHLs mostly affect transcription via a one-component signal transduction system, where the protein domain that binds the SMs 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 composed of a histidine kinase and a response regulator protein (Ulrich, Koonin et al. 2005). As we shall see in the example of *Vibrio harveyi,* in some instances the AI-2 signal transduction cascade can also be composed of more than two components.

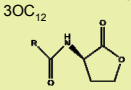
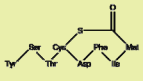
	A. Acyl homoserine lactone (AHL) system	B. Autoinducing peptide (AIP) system	C. AI-2 system	D. γ-butyrolactone system	E. Competence stimulating peptide (CSP) system
Example of bacterial species	*Vibrio fischeri*	*Staphylococcus aureus*	*Vibrio harveyi*	*Streptomyces griseus*	*Streptococcus pneumoniae*
Signalling molecule	Acyl homoserine lactone (AHL). R is acyl chain with 4 to 18 carbons, for example $3OC_{12}$ 	Autoinducing peptide (AIP, encoded by *agrD*) 	AI-2 (a furanosyl borate diester) 	γ-butyrolactone. R is an acyl side chain 	Competence stimulating peptide (CSP, encoded by *comC*) Glu-Met-Arg-Leu-Ser-Lys-Phe-Phe-Arg-Asp-Phe-Ile-Leu-Gln-Arg-Lys-Lys
Synthase	LuxI, synthesis from S-adenosyl methionine (SAM) and fatty acid carrier protein	Prepeptide modified by AgrB, which adds thiolactone ring	LuxS Synthesis from S-adenosyl methionine (SAM)	AfsA	Not applicable
Transporters (exporters and importer)	No transporter – AHL can diffuse through membranes	ABC exporter	No exporter for AI-2. no importer in *Vibrio harveyi*, but importer is present in *Escherichia coli* and *Salmonella* (Lsr ABC-type transporter)	No transporter known	ComAB
Sensor	LuxR	AgrC transmembrane sensor kinase binds extracellular AIP and phosphorylates cytoplasmic AgrA	LuxQ transmembrane sensor kinase binds periplasmic complex LuxP-AI-2, and phosphorylates cytoplasmic LuxU	ArpA	ComD transmembrane sensor kinase binds extracellular CSP and phosphorylates cytoplasmic ComE
Transcription regulator		AgrA	LuxO		ComE

Table 3-2. Components of the cell-to-cell communication systems discussed in the text.

Acyl homoserine lactones

QS via acyl homoserine lactones is the best characterised bacterial cell-to-cell communication system (table 3-2A). AHLs are often also referred to as autoinducer-1 (AI-1) type molecules. The term 'autoinducer' conveys that the synthesis of AHLs is regulated by positive feedback, as has been discovered in the marine bacterium *Vibrio fischeri* (Nealson, Platt et al. 1970). 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 a keto-group in position 3 (Whitehead, Barnard et al. 2001; Reading and Sperandio 2006). For example in *V. fischeri*, the AHL synthase LuxI produces $3OC_6$ homoserine lactone with a keto-group on the third of 6 carbons, whilst in *Agrobacterium tumefaciens* the LuxI homolog TraI produces the $3OC_8$ molecule with a keto group on the third of 8 carbons. In most of the studied systems AHLs are synthesised from S-adenosyl methionine and fatty acid carrier proteins by LuxI and its homologues. However, there are also alternative AHL synthases which are not homologous to LuxI. These include LuxM in *V. harveyi* and HdtS in *Pseudomonas fluorescens*. In many species, more than one AHL is synthesised by different LuxI homologues. For example, in *Rhizobium leguminosarum* six different AHLs synthesised by four different LuxI homologues have been identified (Lazdunski, Ventre et al. 2004). The protein structure of the LuxI homologue in *P. aeruginosa* has been resolved and a detailed mechanism of its function has been proposed (Gould, Schweizer et al. 2004).

Most AHLs cross membranes by diffusion and bind LuxR-like response regulators. LuxR-like response regulators simultaneously act as sensors and transcription factors (Choi and Greenberg 1991). This group of signal transduction systems, where the signal binding domain and the transcription-regulating DNA-binding domain are fused, are referred to as one-component signal transduction systems. They have been shown to be the most common sort of bacterial signal transduction systems (Madan Babu and Teichmann 2003; Aravind, Anantharaman et al. 2005; Ulrich, Koonin et al. 2005).

The N-terminal signal-binding domain of LuxR-like proteins has an *αβα*-sandwich GAF domain-like fold. This is particularly interesting as GAF domains are usually found in signalling and sensor proteins (Aravind and Ponting 1997). The structure of the *A. tumefaciens* LuxR homolog TraR (Zhang, Pappas et al. 2002) shows that, upon binding, AHL is deeply embedded in the protein. Contact with conserved hydrophobic and aromatic residues is established via a number of hydrogen bonds, which are either direct or via water. AHL binding is high-affinity, which means that bacteria are able to sense relatively small communication molecule concentrations. The specificity of the LuxR-like protein for AHL is determined by the acyl binding pocket of the LuxR homologue (Camilli and Bassler 2006), as AHLs differ only in their acyl chains. The C-terminal domain belongs to the DNA and RNA binding helix-turn-helix fold. LuxR-like proteins use this domain to bind their cognate DNA motifs such as the palindromic *lux* box in the case of *V. fischeri*, where it activates transcription of the *luxICDABE* operon. Because this operon encodes the AHL synthase LuxI, AHL synthesis is subject to positive feedback. However, in other species, LuxR orthologues can act as transcriptional repressors, such as is the case for *P. aeruginosa* RhlR and LasR, which have been shown by microarray analysis (Schuster, Lostroh et al. 2003; Clark, Acharya et al. 2009) to negatively regulate multiple genes.

Due to the positive feedback effect AHLs have on regulating the transcription of their own synthases, concentrations of AHLs can vary enormously between lower-density and higher-density cultures such as biofilms. *P. aeruginosa* $3OC_{12}$ homoserine lactone (with a keto-group on the third of the 12 carbons of the acyl chain) has been measured to have a concentration of 2-10 *μ*M in a standard lab culture and of up to 600 *μ*M in the vicinity of an *in vitro* biofilm. However, experimental sensitivity and

issues with AHL stability might have influenced the above results (Shiner, Rumbaugh et al. 2005).

Although the basic AHL QS system only consists of two proteins – LuxI and LuxR – there is a whole ensemble of proteins modulating it. These include the AiiA AHL lactonase of *Bacillus thuringiensis*, which inactivates AHL by hydrolysing the homoserine lactone ring. This is an instance of quorum quenching, more examples of which I will give later. Other LuxR homologues, such as *Escherichia coli* SdiA, recognise more than one AHL (Whitehead, Barnard et al. 2001; Reading and Sperandio 2006). This might be because SdiA is probably used to detect AHL produced by other species, as no LuxI homologue has been detected in *E. coli*. This may be a case of *E. coli* 'spying' on the communication molecules of other species (table 3-1).

Autoinducer-2

There is an overlap of AHL signalling with other QS systems, such as seen in *V. harveyi*. Apart from an AHL system, *V. harveyi* has a parallel system whose signalling molecules are referred to as autoinducer-2 (AI-2; table 3-2C). In *V. harveyi*, AI-2 is a furanosyl borate diester, whose precursor is 4,5-dihydroxy-2,3-pentanedione (DPD). DPD is synthesised by LuxS from S-adenosyl methionine (SAM). Thus, SAM is a precursor both in AHL synthesis and AI-2 synthesis. It is assumed that DPD spontaneously rearranges into AI-2 when borate is available (Chen, Schauder et al. 2002). In other species this rearrangement does not occur, and AI-2 has a different structure.

In *V. harveyi*, in the absence of AI-2, the membrane-bound kinase LuxQ undergoes autophosphorylation on a conserved histidine residue (Neiditch, Federle et al. 2005; Neiditch, Federle et al. 2006). The phosphoryl group is then transferred from the histidine of LuxQ to an aspartate of the response regulator LuxU. Phospho-LuxU in turn phosphorylates LuxO. Phospho-LuxO together with the sigma factor σ^{54} then activates transcription of a set of small RNAs. These small RNAs, together with the RNA chaperone Hfq, contribute to the degradation of the mRNA of the LuxR transcription factor. LuxR is therefore the ultimate effector of the system in the presence of AI-2 in *V. harveyi*. It is important to distinguish this LuxR from the non-homologous LuxR protein that binds AHL in *V. fischeri* as discussed in the previous section. When AI-2 is present, it is bound by LuxP, which is a periplasmic binding protein. LuxP-AI-2 binding to the LuxQ kinase triggers a dephosphorylation cascade

by turning the kinases into phosphatases. LuxQ dephosphorylates LuxU, which in turn dephosphorylates LuxO. The result is that the mRNA of the transcription factor LuxR is no longer degraded by sRNAs, and it can regulate its target genes (Waters and Bassler 2006). Interestingly, LuxU can also be phosphorylated by two other mechanisms. The first mechanism involves $3OC_4$ homoserine lactone that binds to the LuxN sensor kinase, which in turn dephosphorylates LuxU. The second mechanism acts via the unidentified autoinducer CAI-1. Thus, AI-2, $3OC_4$ homoserine lactone and CAI-1 funnel their signals into one common system. Furthermore, sensing of the AHL $3OC_4$ homoserine lactone by a membrane-bound kinase in *V. harveyi* shows that AHLs can also be sensed by pathways which are dissimilar to the one-component pathway outlined in the previous section.

Processed oligopeptides

Many Gram-positive bacteria use processed oligopeptides as communication molecules. The precursor peptides are typically between 40 and 65 amino acid residues in length. These pre-peptides are processed (cleaved) in all known cases and the resulting peptide communication molecules are typically 5 to 34 residues in length. In many cases the peptides are also modified. The minimum components for peptide communication, apart from the peptide signal itself, are a membrane-bound histidine kinase, and a response regulator with an aspartate phosphorylation residue. These components constitute a two-component signal transduction system, as opposed to the one-component system observed in AHL-based communication.

Peptide communication systems include the competence stimulating peptide (CSP) of *Streptococcus pneumoniae* (Havarstein, Coomaraswamy et al. 1995) (table 3-2E), and the peptide antibiotic nisin in *Lactococcus lactis*, which positively regulates its own expression. Another peptide communication molecule discovered in *Staphylococcus aureus* is the autoinducing peptide (AIP; table 3-2C), which is derived from the precursor AgrD to which a thiolactone ring is added between the carboxyl residue and a Cysteine residue at position 5 by AgrB. The AIP is sensed by the AgrC sensor kinase, which may dimerise upon signal binding. The signal is then passed on to the AgrA response regulator, which activates transcription of the *agr* operon (Ji, Beavis et al. 1997; Reading and Sperandio 2006), therefore again leading to autoinduction (figure 3-3B). Both AgrD and ComX, another peptide communication molecule precursor found in *B. subtilis*, have an amphipathic motif in their N-terminal region, which might serve the purpose of recruiting the peptide to the membrane where it can be processed (Zhang, Lin et al. 2004). This is often done by a dedicated

ATP binding cassette (ABC) transporter, which in some cases recognises a so-called GG leader sequence (consensus sequence LSxxELxxIxGG) at the N-terminal side of the region encoding the actual signal (Dirix, Monsieurs et al. 2004).

Peptide signals may be more flexible than the small communication molecules discussed above, as they do not require a specialised synthase and can change to adapt to ecological niches by simple codon mutation (Morrison 2002). The observation that some genes encoding peptide signals are more variable than others might be an indication that this actually happens.

Other systems

Other bacterial QS systems include the *P. aeruginosa* quinolone signal (PQS; 2-heptyl-3-hydroxy-4-quinolone) synthesised by PqsH, which is quite hydrophobic and is exported out of the cell by a vesicular transport system analogous to the ones used by eukaryotes. It seems that PQS is directly facilitating vesicle formation, as *pqsH* mutants do not produce vesicles. However, the phenotype can be rescued by the addition of exogenous PQS (Mashburn and Whiteley 2005). The *pqs* genes have also been found in *Burkholderia*, although the production of PQS in these species could not be confirmed.

Fruiting body formation in the bacterium *Myxococcus xanthus* is facilitated by the diffusible A-signal, which seems to consist of six different amino acids generated by extracellular proteolysis, and the contact-dependent C-signal. The C-signal is a cell-surface protein (Kaiser 2004). Contact-dependent communication has also been observed in *E. coli*, where it is facilitated by CdiA and CdiB and leads to growth inhibition (Aoki, Pamma et al. 2005).

Another QS signal that I will discuss in more detail later is the *γ*-butyrolactone system found in Actinobacteria (table 3-2D).

The list discussed here is by no means complete. Other QS molecules include *Bacillus subtilis* ComX, *Streptomyces coelicolor* SapB, and the *Rhizobium meliloti* Nod factor (Shapiro 1998; Bassler and Losick 2006). It can be expected that even more bacterial QS systems await discovery, especially when considering the vast amounts of microbial diversity made available by metagenomic sequencing projects.

Quorum quenching

In the human pathogen *P. aeruginosa* it has been shown that disrupting its QS system diminishes virulence (Kastrup, Boedicker et al. 2008). This 'quorum quenching' takes the form of enzymes degrading the communication molecules. Two different sorts of enzymes doing this for AHLs have been described: AHLases (*e.g.* AiiA) hydrolyse the lactone ring, resulting in acyl homoserine, whilst AHL-acylases (*e.g.* AiiD) break the amide bond, cleaving homoserine lactone from the acyl side chain (Leadbetter and Greenberg 2000; Park, Hwang et al. 2006). Communication molecules can also be quenched by organisms which do not produce the molecules, presumably in order to gain an advantage over communicating bacterial species in the same ecological niche. For example, *Rhodococcus* is able to degrade AHL signals without having any known ability to produce them (Jafra, Przysowa et al. 2006).

Cross-genome interactions

Several instances of the exchange of SMs between bacteria and eukaryotes are known. This can either take the form of bacterial communication molecules being received by eukaryotic cells, or eukaryotic communication molecules being received by bacterial cells (Mullard 2009).

For example, QseC is a sensor protein that in enterohaemorrhagic *Escherichia coli* is activated by both AI-3 and the mammalian hormones adrenaline and noradrenaline. In both cases, the expression of virulence genes is up-regulated (Clarke, Hughes et al. 2006). It is however not clear if QseC in *E. coli* has evolved to respond to adrenaline and noradrenaline in bacteria, or whether it does so because there has been horizontal gene transfer between eukaryotes and bacteria (Iyer, Aravind et al. 2004). Another example of a eukaryote interfering with bacterial cell-to-cell communication is Apolipoprotein B (APOB), which smothers the AIP in *S. aureus* and prevents it from becoming virulent. Mice that lack APOB are more susceptible to *S. aureus* infection (Peterson, Mack et al. 2008). This may explain why one quarter of humans can permanently live with *S. aureus* in their noses.

	Beneficial to bacterium	**Beneficial to eukaryote**
Bacterium produces SM, eukaryote receives it	A. The pathogenic bacterium *Pseudomonas aeruginosa* produces C12, which inhibits the NF-κB pathway in the mammalian host, therefore interfering with the immune response (Kravchenko, Kaufmann et al. 2008).	B. The mobile zoospores of the green alga *Ulva intestinalis* preferentially settle on *Vibrio* biofilms. They may detect these biofilms by the AHL they produce (Tait, Joint et al. 2005; Joint, Tait et al. 2007).
Eukaryote produces SM, bacterium receives it	C. *Agrobacterium* recognises several SMs produced by plants. This includes phenolic compounds, which stimulate the transcription of genes needed for infection (Winans 1992).	D. The red alga *Delisea pulchra* produces a halogenated furanone, which inhibits QS in the bacterium *Pseudomonas aeruginosa*. This in turn promotes the disintegration of biofilms (Hentzer, Riedel et al. 2002; Hentzer, Wu et al. 2003).

Table 3-3. Examples of different types of information exchange between bacteria and eukaryotes. The exchange of a signal-like SM can be from the bacterium to the eukaryote or from the eukaryote to the bacterium. Cases like B and C could be thought of as 'spying' or 'interception', whilst A and D could be thought of as 'chemical warfare'.

Another way in which QS can be disrupted is by communication molecule analogues that have similar structures and functions as agonists or antagonists to the actual molecules. In *S. aureus*, where QS is mediated by AIPs, different strains use slightly different but homologous communication peptides. These tend to interfere with each other and inhibit the QS cascades of other strains (Bassler and Losick 2006).

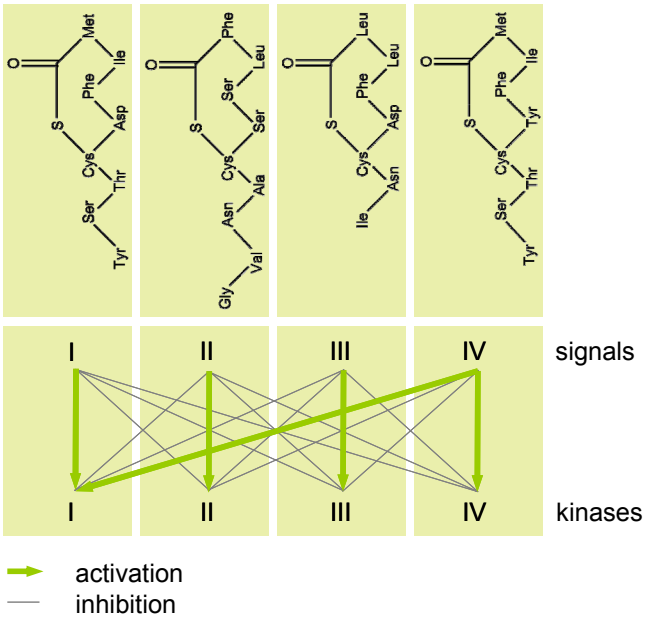


Figure 3-2. In the human pathogen *S. aureus*, the expression of virulence genes is controlled by autoinducing peptides (AIPs). Different strains of *S. aureus* use different AIPs (AIP-I to AIP-IV). The *agrD* genes encoding these AIPs are relatively divergent in terms of sequence. In a case of kin discrimination, each of those AIPs inhibits the sensor kinases of competing strains whilst activating the sensor kinases of its own strain (Morrison 2002; Novick and Geisinger 2008).

Cell-to-cell communication in yeast

In yeasts, several QS-like mechanisms have been reported. In all cases, the major phenotype affected by QS is the transition between the filamentous form and the solitary yeast form. For example, *Histoplasma capsulatum* is a parasitic yeast that can either exist in its filamentous form in soil or in its yeast form as a parasite of humans. Once it enters the host, its morphology switches to the yeast form which synthesises cell-wall polysaccharides such as *α*-(1,3)-glucan. It has been shown that the glucan concentration increases in a cell-density dependent fashion. A culture grown in fresh medium to which filtrate from a dense culture is added produces glucan, which suggests the existence of a factor that promotes glucan incorporation into the cell wall (see section *Tools and Techniques* for more details on this approach).

Farnesol

The existence of a number of different QS molecules has been reported for the human pathogen *Candida albicans*. At low densities, the cells develop germ tubes (filamentous protrusions) which are not normally observed at high cell densities, which suggests that the switch between unicellular yeast and filamentous form depends on cell density. A molecule which blocks the formation of these germ tubes at high cell densities has been identified as farnesol, which has only been observed to have this function in *C. albicans* [53]. Farsenol, as used for communication in *C. albicans*, consists of an OH group and a branched C_{15} side chain. Farnesol acts by affecting transcription. The product of *TUP1* is a transcriptional repressor regulating the yeast-to-filamentous transition by negatively regulating the transcription of hyphal-specific genes. *TUP1* expression has been shown to be increased by farnesol. In *tup1Δ/tup1Δ*, as well as other obligatorily filamentous mutants, farnesol production is increased (Nickerson, Atkin et al. 2006), indicating some sort of negative feedback.

Another molecule involved in the high-to-low cell density transition identified in *C. albicans* is tyrosol (Chen, Fujita et al. 2004). As opposed to farnesol it promotes cell growth and the development of germ tubes at low cell densities. Expression profiling of cultures under conditions of reduced tyrosol concentration showed reduced expression of proteins involved in DNA synthesis and cell cycle regulation (Sprague and Winans 2006). Other putative *C. albicans* QS molecules include the substance MARS of unknown identity and, with diminutive effect, farsenoic acid (Nickerson, Atkin et al. 2006).

Aromatic alcohol derivatives

An altogether more detailed picture of density-dependent cell-to-cell communication has been uncovered for *Saccharomyces cerevisiae*, where phenylethanol (a phenylalanine aromatic alcohol derivative) and tryptophol (a tryptophan aromatic alcohol derivative) have been implicated in QS (Chen and Fink 2006). As in the other examples of yeast QS discussed above, these molecules regulate the transition to the filamentous form. They synergistically affect the upregulation of *FLO11* via the cAMP-dependent kinase Tpk2p (a PKA subunit) and the transcription factor Flo8p. Flo11p, the product of *FLO11*, is the GPI anchored cell surface flocculin protein and is essential for filamentous growth. *S. cerevisiae* strains with deletions of either *TPK2* or *FLO8* do not form filaments in response to aromatic alcohols (Chen and Fink 2006). I present my research on the *S. cerevisiae* aromatic alcohol QS system in more detail later in this chapter.

Tools and Techniques

Chemical

The identification of cell-to-cell communication molecules is a non-trivial task. At an abstract level, a simple assay that allows us to ascertain if a certain phenotype is influenced by QS is to grow cells in culture in stationary phase for some time. During that period, potential communication molecules can accumulate. The cells are then filtered and/or centrifuged out of the growth medium and the remainder is purified. If the addition of this conditioned medium to fresh, exponentially growing cells induces the phenotype in question, one possible explanation is that QS molecules affecting the phenotype were present in the filtrate. High-resolution liquid chromatography/mass spectrometry (LC/MS) and nuclear magnetic resonance (NMR) spectroscopy can aid in the identification of the actual QS molecules. This is achieved by enabling the researcher to compare the spectra of synthetic molecules to molecules purified from the conditioned medium.

Genetic

Genetic techniques to investigate QS include the disruption of the genes in a pathway that produces QS molecules, knock-out of sensor and response regulator genes, addition of purified or synthetic QS molecules, or addition of quorum quenchers. The identity of synthetic and actual (purified) communication molecules can be verified by adding synthetic molecules to a mutant culture which cannot produce its own signals. If the phenotypic effects are the same, this indicates that the

synthetic and purified molecule share the same structure. It can however also happen that chemically similar but not identical SMs elicit similar effects. For example, the QS homoserine lactone $3OC_{12}$, as produced by the bacterium *P. aeruginosa*, has been shown to mimic the effects of farsenol in *C. albicans*, probably due to it being somewhat similar in its side chain structure.

Computational

QS can also be studied by computational methods. The structure of many QS molecules as well as that of their cognate synthases and receptor molecules has been solved by X-ray diffraction analysis or NMR. Gene expression microarray datasets under QS conditions are also publicly available. Integrating structural and gene expression data with the abundantly available genome sequence data will therefore be of great value for the discovery of new QS systems as well as for the integration of QS with intracellular signalling pathways, and later in this chapter I give examples of this in bacteria and in yeast.

A survey of quorum sensing protein domains

Until now, I have introduced different known QS systems and their components, including communication molecule synthase and receptor proteins. A knowledge of the diverse components of different QS systems allows the identification of new systems in fully sequenced genomes using a computational approach.

I used a hand-curated library of hidden Markov models (HMMs) of protein domains that are unique to QS (table 3-4) to identify new QS systems in 384 fully sequenced microbial genomes. The result was the discovery of a number of putative QS systems in genomes that were not previously known to possess such systems.

Results and discussion

I identified 585 QS protein domains, which are listed by genome in appendix 1. The most common domains were the AHL-binding Autoind_bind domain (177 instances) and the AI-2 synthase LuxS (129 instances).

Pfam domain	Communication molecule	Details	Number of domains identified
Autoind_synth	Acyl homoserine lactone (AHL)	LuxI-like AHL synthase	95
Autoind_bind		AHL-binding domain of LuxR-like response regulators	177
AgrD	Autoinducing peptide (AIP)	AIP precursor protein	18
AgrB		AIP precursor modifying enzyme	43
ComX	ComX	ComX precursor peptide	3
LuxS	Autoinducer-2 (AI-2)	AI-2 synthase	129
LuxP		Periplasmic protein that binds AI-2	7
LuxQ-periplasm		Periplasmic domain of the transmembrane protein LuxQ	5
AfsA	γ-butyrolactone	γ-butyrolactone synthase	6
ArpA		γ-butyrolactone receptor protein	24
ComD	Competence stimulating peptide (CSP)	CSP sensor kinase	74
ComC		CSP precursor peptide	2
ComE		CSP response regulator	2
Total			585

Table 3-4. List of the domains in the hand-curated hidden Markov model (HMM) library used in this analysis. For a domain to be included, it had to be unique to QS systems, *i.e.* be known to be involved in QS, and only in QS. See appendix 1 for the genomes in which each of the domains could be identified.

Each QS system tends to be limited to a single bacterial phylum. For example, the Autoind_bind and Autoind_synth domains associated with the AHL system are only found in Proteobacteria. The only exception seems to be the AI-2 system, whose synthase LuxS can be found in genera as diverse as the firmicute *B. subtilis* and the γ-proteobacterium *V. harveyi*. Due to this, AI-2 has been proposed to be involved in interspecies communication (Bassler and Losick 2006). However, homologues of LuxQ, which is involved in sensing AI-2, have only been found in the *Vibrio* genus. Therefore, it can be assumed that sensing of AI-2 occurs in a different way in other species. For example, in *E. coli* and *Salmonella typhimurium* AI-2 is imported into the cytoplasm by an ATP binding cassette (ABC) transporter. This transporter recognises the periplasmic protein LsrB, which has the same periplasmic binding fold as LuxP. AI-2 is then phosphorylated by the cytoplasmic kinase LsrK. It has been suggested that phospho-AI-2 subsequently interacts with the transcription factor LsrR (Xavier, Miller et al. 2007). An alternative explanation for the many genomes that possess an AI-2 synthase but no receptor proteins is that AI-2 is not involved in cell-to-cell communication in these species. For example, LuxS in *S. aureus* could not be shown to have any involvement in QS (Doherty, Holden et al. 2006). This is because LuxS is not solely devoted to AI-2 production but also has a function in the methionine

metabolic pathway (Rezzonico and Duffy 2008). Therefore, many of the LuxS homologues identified here may not produce communication molecules but simply metabolic side-products that are not meant to convey information between cells.

A *γ*-butyrolactone quorum sensing system in *Rhodococcus*?

One QS system that seems to be limited to Actinobacteria is the *γ*-butyrolactone system. It is similar to the AHL-based system in the sense that there is some structural similarity between *γ*-butyrolactone and AHL, but also in the sense that it is a one-component system where the communication molecule sensing protein is also the response regulator.

Streptomyces is the only genus where *γ*-butyrolactone has been shown to be a communication molecule (Waters and Bassler 2006). My survey of QS domains in microbial genomes shows that in the genome of the *Rhodococcus* strain RHA1, homologues for protein domains of both the *γ*-butyrolactone synthase and the receptor can be found. This suggests that *γ*-butyrolactone might play a role in this organism too. The case for these homologues acting as a QS system is supported by the close genomic proximity of the synthase genes and the response regulator genes, which is also observed in confirmed AHL QS systems.

Protein function	**Synthase**	**SM binding and effector domain**
Streptomyces name	AfsA	ArpA
Pfam domain architecture	AfsA	TetR_N - Pfam-B_4709
Homologues in *Streptomyces avermitilis* MA-4680	1	4
Homologues in *Streptomyces coelicolor* A3(2)	2	6
Homologues in *Rhodococcus* species RHA1	1	8

Table 3-5. The necessary components of a *γ*-butyrolactone QS system are a communication molecule synthase, AfsA, and a communication molecule response regulator, ArpA. Both components are present in the *Rhodococcus* strain RHA1.

AfsA homologues also exist in *Frankia* and *Gloeobacter*, and ArpA homologues exist in some Mycobacteria and in *Nocardia*. However, of the fully sequenced bacterial genomes included in this survey, there is none outside *Streptomyces* and *Rhodococcus* that encodes both ArpA and AfsA homologues. This means that whilst some species may be able to produce or sense *γ*-butyrolactone, there are none outside *Streptomyces* and *Rhodococcus* that can do both and therefore have a fully functional *γ*-butyrolactone QS system.

Although so far there has been no confirmation of the *γ*-butyrolactone QS system in *Rhodococcus* strain RHA1, there is further evidence that supports the case that such a system is present. Firstly, *Rhodococcus* produces *γ*-butyrolactone in high cell density cultures (Curragh, Flynn et al. 1994). Secondly, *Rhodococcus* has an unusually high number of AHL-inactivating lactonases. The number genes encoding AHL-inactivating Lactamase_B Pfam domains in the genomes of Actinobacteria included in this survey varies between one and 21. Only the *Rhodococcus* strain RHA1 is an outlier and has 28 genes encoding proteins with such domains (data not shown). These lactonases may play a role in the intracellular metabolism of lactone compounds such as γ-butyrolactone (Uroz, Chhabra et al. 2005). In this context, it is also interesting that *Rhodococcus* strain RHA1 has the only Autoind_synth domain coding sequence found outside the Proteobacteria.

In summary, the identification of all necessary components of a QS system in *Rhodococcus* strain RHA1 demonstrates that computational methods can give information about previously uncharacterised QS systems. Later in this chapter I use similar approaches to discover further QS systems in the bacterial phylum of firmicutes.

Materials and methods

All sequence data were obtained from the NCBI genome database (ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/). For most of this analysis I used a set of 384 fully sequenced microbial genomes. Through an in-depth survey of the literature available on QS, I identified protein domains that are unique to QS processes (table 3-4). For each of these domains, Hidden Markov Models (HMMs) were available on the protein families database Pfam (Finn, Mistry et al. 2006). I used these HMM models to search the 384 genomes for further instances of the same domains. These were then used to refine the HMM model and to re-search the genomes in an iterative process until saturation. Both the hmmbuild and hmmsearch programs that I used are part of the HMMER suite of programmes (http://hmmer.janelia.org/) (Durbin, Eddy et al. 1998).

Some of the methods I use in the next section are similar to the ones used in this one, but its focus is narrower. Instead of dealing with a wide range of bacterial genomes, it concentrates on firmicutes only, and instead of dealing with a wide range of QS systems, it deals with the *agr* system only.

The *agr* quorum sensing system in firmicutes

As already discussed above, the pathogenic firmicute *S. aureus* has a well-characterised cell-to-cell communication system whose signals are modified peptides (Ji, Beavis et al. 1997). This system is encoded in the *accessory gene regulator* (*agr*) locus (Recsei, Kreiswirth et al. 1986). The *agr* system is an important regulator of biofilm formation and virulence factor expression (Kleerebezem, Quadri et al. 1997), and consists of four genes encodeding a polycistronic transcript (figure 3-3A). The *agrD* and *agrB* genes encode the precursor signalling peptide and the enzyme that cleaves and modifies the precursor peptide, respectively. The resultant signal is called autoinducing peptide (AIP). The *agrC* gene encodes the transmembrane kinase that senses the AIP, and the *agrA* gene encodes the response regulator that is phosphorylated by AgrC and effects changes in gene expression, respectively (figure 3-3B).

The AIP signal has seven to nine residues, depending on the strain. The only consistently conserved residue of the AIP is a cysteine at the 5th position from the C-terminus (figure 3-4). This cysteine forms a thiolactone ring with the residue at the C-terminus (Mayville, Ji et al. 1999). The AIP is produced from the precursor peptide AgrD. The N- and C-terminal portions of AgrD are cleaved off by the membrane-associated enzyme AgrB, which also forms the thiolactone ring, hence producing the AIP. AgrB might also be involved in the export of the AIP across the cell membrane (Qiu, Pei et al. 2005).

However, overexpression of AgrB and AgrD does not result in increased AIP production. AgrB therefore does not seem to be the only processing enzyme required for the synthesis of AIP (Saenz, Augsburger et al. 2000). Once it is in the intercellular space, AIP diffuses until it binds the AgrC sensor kinase. This kinase can either be located in the membrane of the cell which produces AIP, or in the membrane of another *S. aureus* cell. The AgrC kinase is part of a two-component signal transduction system. The second component is the cytoplasmic response regulator AgrA. AgrA has two domains: The N-terminal domain, which is a receiver domain that becomes phosphorylated by AgrC, and the C-terminal domain, which is the DNA binding LytTR domain. Phosphorylation of the receiver domain leads to AgrA activation and DNA binding. AgrA positively regulates transcription of the *agr* operon. This means that a positive feedback effect takes place, with a high level of AIP leading to even higher *agr* expression. Apart from the *agrBDCA* promoter, another

promoter positively regulated by AgrA is the one controlling expression of RNAIII. The RNAIII promoter is located next to the *agrBDCA* promoter, but is transcribed in the opposite direction (figure 3-3A). Until recently, it was believed that RNAIII rather than AgrA is the immediate regulator of most of the genes regulated by the *agr* operon (Benito, Kolb et al. 2000). However, it has recently been shown that the genes regulated by the *agr* operon can be divided into RNAIII-dependent and RNAIII-independent, AgrA-dependent genes (Queck, Jameson-Lee et al. 2008). Although RNAIII is a regulatory RNA, its transcript also contains the δ-haemolysin gene *hld*. In some strains of *S. aureus*, the staphylococcal accessory regulator (SarA) protein is also necessary to activate transcription of *agrBDCA* and RNAIII (Cheung and Projan 1994). SarA is a relatively abundant protein that seems to be a global regulator modulating the expression of many genes.

Systems that are similar to the *agr* system either by homology or analogy have been identified in firmicutes outside *Staphylococcus*. In some cases they have different names. For example, the homologous *agr*-like system in *Enterococcus faecalis* is known as *fsr* system (Qin, Singh et al. 2000), and the one in *Lactobacillus plantarum* is called *lam* system (Diep, Havarstein et al. 1994). Here, I investigate the distribution of *agr*-like systems outside the Staphylococcaceae in order to understand the conservation and plasticity of their interaction with other cellular components and signal transduction and communication systems.

Results and discussion

In order to define *agr* homologues in taxa other than *Staphylococcus*, I built hidden Markov models (HMMs) of protein domains which are considered to be unique to the *agr* system. In a complementary approach, I identified *agr* homologues with PSI-BLAST searches. Protein domains which I considered to be unique to the *agr* system are the AgrB domain and the AgrD domain. To my knowledge, no function of AgrB apart from processing of AgrD has been reported so far. The remaining protein domains of the *agr* system are the HATPase_c Pfam domain of the histidine kinase and the Response_reg and LytTR domains of the response regulator. However, these domains are not unique to the *agr* system and were therefore not considered to be prognostic of it. I built HMMs of the AgrB and the AgrD domains, starting with the published Pfam domains and adding homologous sequences identified in the literature.

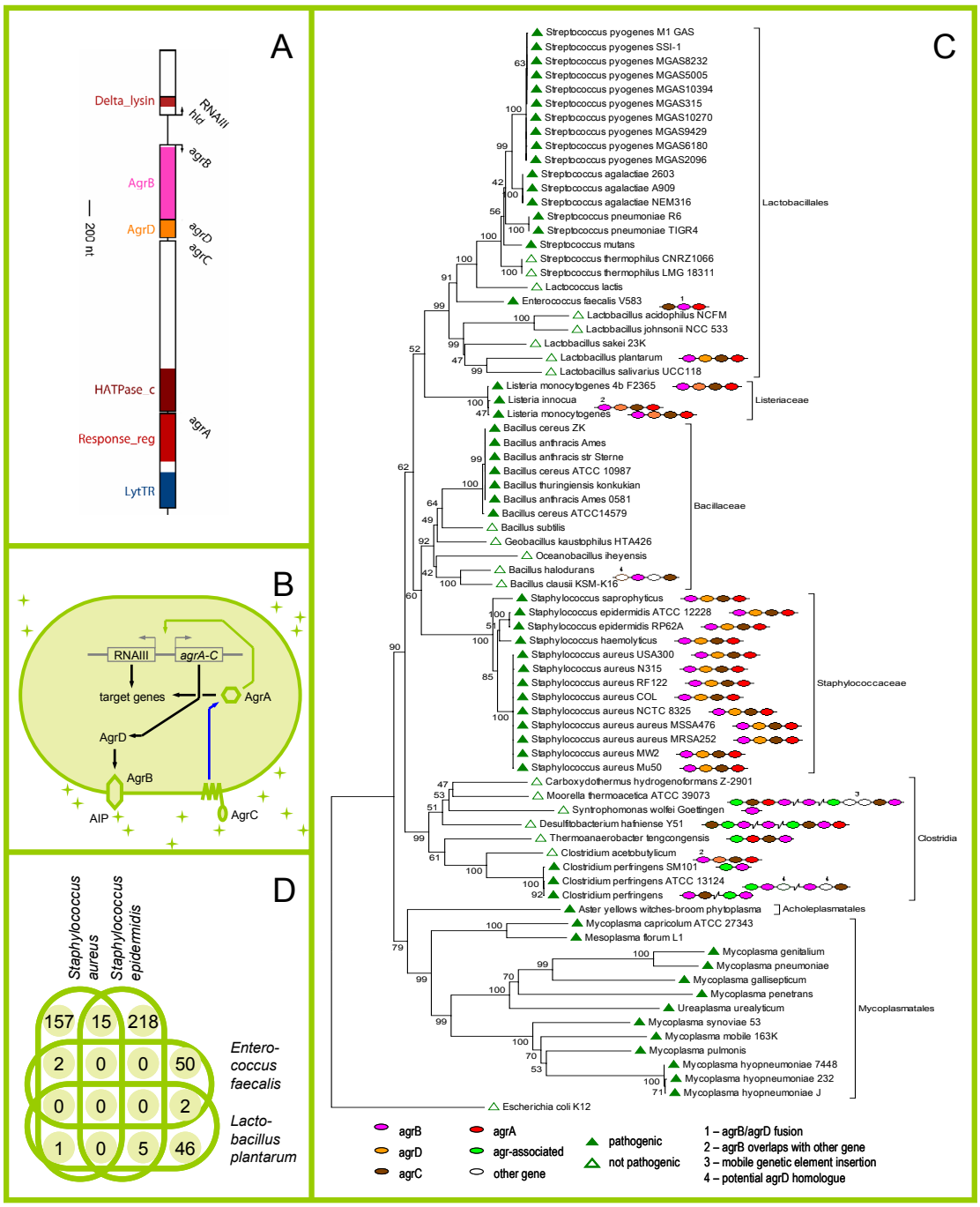


Figure 3-3. (A) Organisation of the *agr* locus in *S. aureus*, with Pfam domains highlighted in colour. (B) Cartoon representation of the basic interactions of the *agr* system in *S. aureus*. The interactions of the proteins encoded by the *agrBDCA* operon leads to a positive feedback loop and further *agrBDCA* expression. Arrow colours: phosphorylation in blue and transcriptional interactions in green. RNAIII and *hld*, which encodes *δ*-lysin, are controlled by the same promoter and share the same transcript. The second promoter controls *agrBDCA*. The *agrC* and *agrA* genes encode the transmembrane kinase which senses the AIP (green crosses), and a response regulator (hexagon) that is phosphorylated by AgrC and affects changes in gene expression, respectively. Phosphorylation of the AgrA receiver domain by AgrC leads to AgrA activation and DNA binding. AgrA positively regulates transcription of the *agr* operon. This results in the production of a high level of AIP leading to even higher agr expression. Apart from the *agrBDCA* promoter, another promoter positively regulated by AgrA is the one controlling expression of a small non-coding RNA, RNAIII. The RNAIII promoter is located next to the *agrBDCA* promoter, but is transcribed in the opposite direction. RNAIII together with AgrA is the immediate regulator of the genes regulated by the *agr* operon. **[Figure legend continues on next page]**

[Continuation of legend for figure 3-3] (C) Tree of the bacterial phylum of firmicutes, based on ribosomal sequences. Where applicable, the genomic organisation of the *agr* locus homologue is depicted to the right of the operational taxonomic unit. One instantly obvious finding is that homologues of the *agr* system are spread across firmicutes and are not limited to a single phylogenetic class. Furthermore, the genomic organisation of the *agr* system is conserved outside the family Staphylococcaceae. That is to say, in most cases *agrB* is linked to a two-component signal transduction system. While each genome contained only a single copy of the AgrD domain, some contained more than one copy of the AgrB domain: *Moorella thermoacetica* (3 copies), *Desulfitobacterium hafniense* Y51 (3 copies), *Clostridium perfringens* (2 copies), and *Clostridium perfringens* ATCC 13124 (2 copies). (D) Venn diagram showing numbers of differentially expressed genes and how many orthologous clusters are differentially expressed in more than one genome (see section *agr*-regulated genes for details).

The *agr* system is widespread among firmicutes

In the 384 genomes used in this analysis, 33 instances of genes encoding the AgrB domain and 18 instances of those encoding the AgrD domain were found, all of them in firmicutes. All identified cases of the AgrD domain were found in close genomic proximity to a gene encoding an AgrB domain, which means that there are no orphan *agrD* homologues. Each genome contained only a single copy of the AgrD domain. However, four genomes contained more than one copy of the AgrB domain. These were *Moorella thermoacetica* (3 copies), *Desulfitobacterium hafniense* Y51 (3 copies), *Clostridium perfringens* (2 copies), and *Clostridium perfringens* ATCC 13124 (2 copies). Each of the AgrB loci is visualised in figure 3-3C. All 18 loci which contained both the AgrD and the AgrB homologue were considered to be true QS loci. 4 other loci were also considered to be true QS loci because they encoded a histidine kinase and a response regulator linked to the AgrB homologue. In the case of *Enterococcus faecalis* V583, the *agrD* homologue *fsrD* is transcribed together and in frame with *fsrB*, but is translated independently of it (Nakayama, Chen et al. 2006). In some cases *agrB* was linked to the kinase and response regulator genes of a two component system, but not to a putative *agrD* homologue. In these cases the *agr* system could still be functional if a protein that has the function of AgrD, but has no detectable sequence homology to it, is present. This indeed seems to happen for a number of putative *agr* systems, where I found a small ORF encoded in the same position downstream of *agrB*, where *agrD* is normally found in *Staphylococcus*. Alternatively, a short gene such as *agrD* might still be present but could have escaped present ORF annotation software.

The genomic context of the AgrB homologues was mapped onto a phylogenetic tree built from ribosomal protein sequences (figure 3-3C). One instantly obvious finding from figure 3-3C is that homologues of the *agr* system are spread across firmicutes and are not limited to a single phylogenetic class. Furthermore, the genomic organisation of the *agr* system is conserved outside the family Staphylococcaceae.

That is to say, in most cases *agrB* is linked to a two-component signal transduction system.

These findings pose the question of whether the *agr* QS system is an evolutionarily ancient system which was present in the common ancestor of firmicutes, or whether it is a more recent innovation and has been transferred between groups by horizontal gene transfer. To investigate this, I built two more phylogenetic trees (not shown). The first tree was built from an alignment of AgrB homologues. The second tree was built from an alignment of AgrB, AgrD, AgrC, and AgrA homologues, whose sequences were concatenated. The topology of these trees should agree with the topology of the tree built from ribosomal sequences if *agr* is inherited only vertically. Although these trees generally agree with each other and the ribosomal tree, the *agr* locus of *Clostridium acetobutylicum* clusters with the Listeriaceae. *C. acetobutylicum* however is a member of the Clostridia, therefore making its *agr* system a possible candidate for horizontal gene transfer. To my knowledge, this is the first reported instance of a possible case of horizontal gene transfer between the Clostridia and the Listeriaceae. Horizontal gene transfer is also supported by AgrD of *C. acetobutylicum*, which shows remarkable similarity to AgrD of the Listeriaceae (figure 3-4). Other members of the genus *Clostridium* do not have this high level of sequence similarity to the Listeriaceae. I did not find evidence for *C. acetobutylicum* and *Listeria* species sharing a common niche to explain this observation.

```
Staphylococcus saprophyticus ATCC 15305 -MNVIKSISKSISKSISNYFAKVFASIGSISTINPCFG-YTDESEIPK-ELTDLYE-- 72494665
Staphylococcus epidermidis ATCC 12228   --------MENIFNLFIKFFTTILEFIGTVAGDSVCAS-YFDEPEVPE-ELTKLYE-- 27316101
Staphylococcus epidermidis RP62A        --------MENIFNLFIKFFTTILEFIGTVAGDSVCAS-YFDEPEVPE-ELTKLYE-- 57638027
Staphylococcus haemolyticus JCSC1435    --------MTVLVDLIIKLFTFLLQSIGTIASFTPCTT-YFDEPEVPE-ELTNAK--- 68446720
Staphylococcus aureus MRSA252           --------MKKLLNKVIELLVDFFNSIGYRAAYINCDF-LLDEAEVPK-ELTQLHE-- 49242392
Staphylococcus aureus MSSA476           --------MKKLLNKVIELLVDFFNSIGYRAAYINCDF-LLDEAEVPK-ELTQLHE-- 49245275
Staphylococcus aureus MW2               --------MKKLLNKVIELLVDFFNSIGYRAAYINCDF-LLDEAEVPK-ELTQLHE-- 21205132
Staphylococcus aureus COL               --------MNTLFNLFFDFITGILKNIGNIAAYSTCDF-IMDEVEVPK-ELTQLHE-- 57284895
Staphylococcus aureus NCTC 8325         --------MNTLFNLFFDFITGILKNIGNIAAYSTCDF-IMDEVEVPK-ELTQLHE-- 87203491
Staphylococcus aureus USA300            --------MNTLFNLFFDFITGILKNIGNIAAYSTCDF-IMDEVEVPK-ELTQLHE-- 87126405
Staphylococcus aureus Mu50              --------MNTLVNMFFDFIIKLAKAIGIVGGVNACSS-LFDEPKVPA-ELTNLYD-K 14247810
Staphylococcus aureus N315              --------MNTLVNMFFDFIIKLAKAIGIVGGVNACSS-LFDEPKVPA-ELTNLYD-K 13701831
Staphylococcus aureus RF122             --------MNTLVNMFFDFIIKLAKAIGIVGGVNACSS-LFDEPKVPA-ELTNLYD-K 82657170

Lactobacillus plantarum WCFS1           -----------MKQKMYEAIAHLFKYVGAKQLVMCCVG-IWFETKIPD-ELRK----- 28272731

Listeria innocua                        ----MKNMNKSVGKFLSRKLEEQSMKVADSSMSKACFM-FVYEPKSPFVKMQEKNENK 16412463
Listeria monocytogenes                  ----MKNMNKSVGKFLSRKLEEQSMKVADSSMSKACFM-FVYEPKSPFVKMQEKNENK 16409408
Listeria monocytogenes str. 4b F2365    -------MNKSVGKFLSRKLEEQSMKVADSSMSKACFM-FVYEPKSPFVKMQEKNENK 46879544

Clostridium acetobutylicum ATCC 824     -MNLKEQLNKVNDKFIKG-LGKASMKIGEQANGK-CVLVTLYEPKMPE-ELLKENIDK 15022901

Clostridium perfringens SM101           --------MKKLNKNLLTLFAALTTVIATTVATSACIW-FTHQPEEPK-SLRDE---- 110682078
Clostridium perfringens ATCC 13124      --------MKKLNKNLLTLFAALTTVVATTVATSACLW-FTHQPEEPK-SLRDE---- 110674548
Clostridium perfringens ATCC 13124      MRKSILSMLALLTTLTINASTTSFATIGLEEM-----------PESMKKSR------- 110673345

Enterococcus faecalis V583              ------------------------DGVGTKPRLNQNSPNIFGQWMGQTEKPKKNIE-K 29343827
Bacillus halodurans C-125               -------------MKRTAKVISKATLGISKAFVNASSP-LIYAPKIPNGLKKQK---- 10176097
```

Figure 3-4. Alignment of potential AgrD homologues. GI numbers are given in the column to the right. Colour codes: red, experimentally determined signalling peptides. Yellow background, residues strongly conserved across the homologues. Green background, residues conserved between Listeria and *Clostridium acetobutylicum*. In bold, potential AgrD homologues that were identified by genomic context (linkage to AgrB) rather than by homology.

New *agr*-like systems

In three genomes I was able to identify a putative *agr* system where, to my knowledge, none had been reported before. These genomes are *Moorella thermoacetica* ATCC 39070, *Desulfitobacterium hafniense* Y51, and *Thermoanaerobacter tengcongensis*. All three species are members of the taxonomic class Clostridia. In all three genomes, *agrB* occurs as a genomic neighbour of a gene encoding a response regulator and a gene encoding a histidine kinase, which makes a function of the locus in cell-to-cell communication plausible (Bassler and Losick 2006). I propose that these genomes might contain a functional peptide QS circuit. In the following, I discuss the implications of QS in these organisms. Like all Clostridia, these species are anaerobic, and *Moorella* and *Thermoanaerobacter* are thermophilic (Xue, Xu et al. 2001; Drake and Daniel 2004). All three species are non-pathogenic. The genus *Desulfitobacterium* consists of bacteria that survive in a wide range of environments. Desulfitobacteria are metabolically diverse and use a variety of compounds as electron acceptors, including metals, nitrate, sulphite, and halogenated compounds. This makes the Desulfitobacteria a candidate for anaerobic bioremediation processes. In anaerobic fixed-film reactors *D. hafniense* has been shown to occur in a biofilm together with sulphate reducing bacteria (Villemur, Lanthier et al. 2006). If QS via the *agr* locus indeed occurs in *D. hafniense*, it can be assumed that it would modulate biofilm formation as it does in other species (Dunman, Murphy et al. 2001; Bourgogne, Hilsenbeck et al. 2006; Cassat, Dunman et al. 2006).

In the genomes that contain an AgrB homologue that is not linked to a complete two-component system, *agr* mediated QS might also be functional. This would require the two-component sensing system to be encoded somewhere else in the genome rather than being linked to *agrB*. Genomes which have an *agrB* homologue that is not linked to a two-component system are those of *Bacillus halodurans*, *Syntrophomonas wolfei*, and three *C. perfringens* strains. However, in the *Clostridium perfringens* strain ATCC 13124 and strain 13 one of the two paralogues of AgrB co-occurs with a histidine sensor kinase. Peptide-mediated QS in the *Clostridium perfringens* species would be of particular interest, as this species is the causative agent of the human disease *gas gangrene* (Present, Meislin et al. 1990). The possibility of *agr*-like systems in the genus *Clostridium* has recently been explored (Burrell 2006). Burrell examined uncultured *Clostridium* species isolated from landfill leachate reactor biomass by cloning PCR products. Some homologues of *agrC* were identified using

primers from the *agrC* homologue identified in *C. thermocellum*. However it was not established if these *agrC* homologues are functional components of a QS circuit.

Multiple paralogous copies of AgrB in Clostridia genomes

Four of the genomes I examined in the Clostridia have more than one copy of AgrB. These were *M. thermoacetica* ATCC 39073 (3 copies), *D. hafniense* Y51 (3 copies), *C. perfringens* (2 copies), and *C. perfringens* ATCC 13124 (2 copies). The phylogenetic evidence suggests that the two *C. perfringens* strains diverged after a duplication event which created two *agrB* paralogues. This duplication event in the common ancestor of the two strains must have been ancient, as the two copies are phylogenetically distant. The same applies to the duplication events that have given rise to multiple copies of AgrB in *D. hafniense* and *M. thermoacetica*. In all genomes only one paralogue of AgrB is linked with a histidine kinase as well as a response regulator, whilst the other paralogues seem to be orphans. This raises the question what the function, if any, of the paralogous copies might be. I did not find any genes which were consistently linked to the orphans or were predicted to be in the same operon. Something which might explain the presence of multiple paralogous AgrB homologues in the same genomes is that peptide 'signals' do not only serve the purpose of cell-to-cell signalling, but are also involved in environmental sensing. It can be speculated that the structure of peptide signals can be altered by the extracellular conditions, subsequently passing this information on to two-component systems (Morrison 2002). If this is the case, selection pressures other than signalling efficiency and reliability act on peptide signalling systems. This provides a possible explanation why for example *M. thermoacetica* or *D. hafniense* possess multiple AgrB homologues: In an instance of subfunctionalisation they might use one or more of the paralogues for environment sensing rather than cell-to-cell signalling. That all genomes with multiple copies of *agrB* are found in the Clostridia could be because many species in this class are versatile and can persist in a wide range of environments, therefore having an increased need for environmental sensing. It would not be necessary that the two-component system sensing the peptide is encoded in close genomic proximity to the processing enzyme, AgrB. An alternative hypothesis is that AgrB produces a peptide which interferes with QS in related but competing strains. This would explain why one of the *agrB* genes appears to be an orphan without a linked two-component system. Another explanation would be that additional AgrB paralogues are used to increase signal production.

A novel gene associated with the *agr* system in Clostridia

In five of the eight genomes of the class Clostridia which have AgrB, a conserved unidentified gene is linked with the *agrB* locus (green in figure 3-3C). I speculate that this gene might, in some way, be involved in the function of the *agr*-like system in Clostridia. No homologues of this gene could be identified outside Clostridia, and no functional annotation is available. This result could be verified by PSI-BLAST searches against the non-redundant NCBI protein sequence database. Analysis of this short gene (typically 150-170 residues) using the TMHMM script (http://www.cbs.dtu.dk/services/TMHMM/) gave strong support for the existence of 5 conserved transmembrane regions, making it likely that the gene product locates to the membrane. No conserved signal peptide cleavage sites could be found using the SignalP server (http://www.cbs.dtu.dk/services/SignalP/). A homologue of the protein in *Syntrophomonas wolfei* Goettingen is fused to a histidine kinase. However, I could not detect any homology of the gene to other C-terminal regions of histidine kinases using PSI-BLAST. The N-terminus of the protein was predicted to be located on the extracellular side, whilst the C-terminus was predicted to be located in the cytoplasm. The first extracellular loop of the protein contains the highly conserved residues Arg, Leu, Gly, and His (figure 3-5).

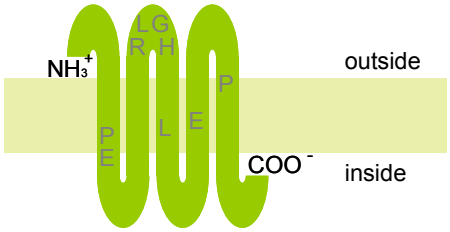


Figure 3-5. Predicted transmembrane topology of a protein which is encoded in close proximity to the *agr* locus in many Clostridia. Highly conserved residues in grey.

RNAIII and the *agr* system

Like transcription of the *agrBDCA* operon, AgrA also positively regulates transcription of RNAIII in *S. aureus*. RNAIII has 541 nucleotides and forms a secondary structure containing 14 hairpin loops (Benito, Kolb et al. 2000). RNAIII in *S. aureus*, but not in *S. saprophyticus*, also contains the coding region of the δ-toxin gene *hld* (Sakinc, Kulczak et al. 2004). In *S. aureus*, RNAIII together with AgrA is the major regulator of genes controlled by the *agr* system (figure 3-3B). It can be speculated that this is because RNAIII allows quicker post-transcriptional regulation compared to slower transcriptional regulation. A similar principle has been observed in the AI-2 cell-to-cell communication system of *V. harveyi*, where small RNAs play an important role as effectors of the QS response (Lenz, Mok et al. 2004). In *S. aureus*, RNAIII has been shown to increase expression of the α-toxin gene *hla* by binding to the 5'

untranslated region (UTR) of its mRNA. This is achieved by a high level of sequence complementarity between the UTR and the 5' end of RNAIII (Morfeldt, Taylor et al. 1995). I investigated how widespread this sort of regulation might be amongst firmicutes by scanning whole genomic nucleotide sequences for RNAIII. I searched the nucleotide sequences of complete genomes which encode a copy of AgrB for RNAIII homologues. This resulted in highly significant hits for all *S. aureus* and *S. epidermidis* strains. Putative hits were also found for *S. haemolyticus* and *S. saprophyticus,* albeit at a lower significance level. In all cases, RNAIII was encoded in close genomic proximity to the *agr* locus, with its promoter pointing in the opposite direction, as depicted in Figure 3-3A. Putative RNAIII in *Staphylococcus* was also found when no *hld* gene was predicted. The expression of RNAIII has been confirmed experimentally in *S. epidermidis* (Tegmark, Morfeldt et al. 1998) and *S. saprophyticus* (Sakinc, Kulczak et al. 2004). To my knowledge, RNAIII function has not yet been examined in *S. haemolyticus*. It should be noted that the fact that I did not identify RNAIII outside Staphylococcaceae does not necessarily mean that *agr* does not exert its effect via noncoding RNAs in these species. However, I conclude that RNAIII is an innovation which was made in the common ancestor of the Staphylococcaceae.

agr-regulated genes

Differences in gene expression levels between wildtype and *agr* knockout cells have previously been compared in *Lactobacillus plantarum* WCFS1 (Sturme, Nakayama et al. 2005), *Enterococcus faecalis* OG1RF (Bourgogne, Hilsenbeck et al. 2006), *S. aureus* UAMS-1 (Cassat, Dunman et al. 2006), and *S. epidermidis* (Yao, Vuong et al. 2006). According to my criteria (see Materials and Methods), 175 genes are differentially expressed in *S. aureus* UAMS-1, 54 in *L. plantarum*, as well as 54 in *E. faecalis*, and 238 in *S. epidermidis*. After defining clusters of orthologous genes, I asked if there are any orthologues which are differentially expressed in all four genomes. I did not find any cluster that is differentially regulated in all four genomes. There are two clusters that are upregulated by *agr* in both *E. faecalis* and *L. plantarum*. These two clusters correspond to AgrB and AgrC orthologues. There are also two clusters which are upregulated in *S. aureus* N315 and *E. faecalis*. These clusters correspond to a sugar ABC transporter and a prephenate dehydratase, which is an enzyme involved in phenylalanine synthesis (MetCyc database (Caspi, Foerster et al. 2006)). There is one cluster that is upregulated in both *S. aureus* N315 and *L. plantarum*, corresponding to a short chain dehydrogenase/reductase.

Therefore, there is no significant propensity for orthologous genes to be regulated by the *agr* system in the four different species.

Genes that are involved in biofilm formation, such as those encoding exopolysaccharide biosynthesis proteins, are downregulated in *S. aureus* and *E. faecalis*. Some exopolysaccharide biosynthesis genes, such as those of the *cps2* cluster, are upregulated in *L. plantarum*. This seems to contradict the observation that deletion of *lamA*, the *agrA* homologue, has the phenotypic effect of reduced adherence to a glass substratum compared to the wildtype. However the polysaccharides might be membrane associated and capsular rather than secreted. This in turn could prevent adhesion factors from increasing the cell's propensity to form biofilms, as has been observed in *S. aureus* (Sturme, Nakayama et al. 2005). However, it is not clear why *agr* would promote biofilm formation in the pathogen *E. faecalis* but not in *S. aureus* N315, which is also a pathogen. The production of virulence proteins is up-regulated by the *agr*-like system in both virulent species. A clear conclusion is that the *agr* system is not a regulator of virulence factors only and is also active in non-pathogenic firmicutes.

It should be noted that the platforms for *L. plantarum* and *E. faecalis* were dual-channel cohybridisation arrays, while the platform for *Staphylococcus aureus* UAMS-1 was an Affymetrix GeneChip. The absolute ratios by which genes were up-or downregulated with respect to the wildtype are therefore not directly comparable. However, the identity of differentially expressed genes should be platform independent, and therefore comparable across experiments.

Concluding remarks

The number of completely sequenced microbial genomes is increasing steadily. In addition, an enormous amount of sequence data is generated from metagenomics projects. As it is very unlikely that all bacterial cell-to-cell communication systems have already been described, new methods for predicting these communication systems are required. Here I have applied an integrated approach which is capable of identifying previously unknown systems.

Materials and methods

All sequence data used was obtained from the NCBI genome database. For most of this analysis I used a set of 384 fully sequenced genomes.

Homologue identification

In order to define *agr* homologues in taxa other than *Staphylococcus*, I built HMMs of protein domains that I considered to be unique to the *agr* system. Protein domains which I considered to be unique to the *agr* system are the AgrB domain and the AgrD domain, each found in proteins with the respective name in *Staphylococcus*. The remaining protein domains of the *agr* system are the HATPase_c domain of the histidine kinase and the Response_reg and LytTR domains of the response regulator. However, these domains are not unique to the *agr* system and were therefore not taken to be indicative of it. As before, I built HMMs of the AgrB and the AgrD domains, starting with published Pfam models (Finn, Mistry et al. 2006) and adding homologous sequences identified in the literature. These sequences were aligned using the MUSCLE multiple alignment programme (Edgar 2004) and manually corrected. Following this, HMMs were created from the multiple sequence alignments using the hmmbuild programme. These HMMs were then used to search the microbial genomes for further homologues using the hmmsearch programme. These homologues were then aligned again and used to build another model. This process was repeated until further iterations did not identify new homologues. Both hmmbuild and hmmsearch are part of the HMMER suite of programmes (http://hmmer.janelia.org/; (Durbin, Eddy et al. 1998).

To validate my findings, I also did PSI-BLAST searches (Altschul, Madden et al. 1997) against the nonredundant NCBI protein sequence database, which is more exhaustive than the database of fully sequenced genomes used here. All searches were iterated until saturation. The results of the PSI-BLAST searches are available on the supplementary material website.

Tree building

A guide tree of firmicutes was constructed using ribosomal proteins. *Escherichia coli* K12 was used as an outgroup to root the tree. The ribosomal proteins used were L1, L3, S2, and S5. The corresponding genes were determined by Ciccarlli *et al.* to be very likely to occur only once in each genome and to be unlikely to be affected by horizontal gene transfer (Ciccarelli, Doerks et al. 2006). HMMs for these proteins were downloaded from Pfam and used to search the firmicute genomes for homologues. Each genome was found to contain no more than one copy of each protein. The four proteins were concatenated in each genome and aligned using MUSCLE. The alignment was inspected by eye and gaps were removed.

The tree in Figure 3-3C was constructed using the Neighbour Joining algorithm. A Poisson correction model was used as a model of amino acid substitution. The tree building procedure was repeated using a PAM matrix and a Jones-Taylor-Thornton matrix as substitution models, but the topology of the resulting tree was identical in all cases. Furthermore, I calculated bootstrap values by replicating the tree-building procedure 1000 times. To evaluate the robustness of the method I constructed the same tree using Maximum Parsimony and Minimum Evolution and compared the results with the NJ tree. The topology within the seven groups Lactobacillales, Listeriaceae, Bacillaceae, Staphylococcaceae, Clostridia, Acholeplasmatales, and Mycoplasmatales was largely the same with these methods, although the relationship between the groups was not.

For all tree building procedures MEGA 3.1 (Kumar, Tamura et al. 2004) was used.

RNAIII identification

Noncoding RNAs can tentatively be identified by using an approach that combines sequence homology with secondary structure information. For this, I used the programme INFERNAL, which predicts noncoding RNAs by using covariance models that take into account both sequence homology and RNA secondary structure. The model of RNAIII used by INFERNAL was obtained from the Rfam database (Griffiths-Jones, Bateman et al. 2003). Whole genomic DNA sequences of those genomes that have a predicted *agr* system were searched.

Gene expression data analysis

Gene expression data was collected from four previously published experiments (Dunman, Murphy et al. 2001; Sturme, Nakayama et al. 2005; Bourgogne, Hilsenbeck et al. 2006; Cassat, Dunman et al. 2006; Yao, Vuong et al. 2006). Genes that were expressed at least three-fold differently in the *agr* mutants compared to the wildtype were identified. For the two cohybridisation arrays this was done by using the data as published in the respective papers. For the GeneChip experiment (Cassat, Dunman et al. 2006), data was downloaded from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/projects/geo/), where its identifier is GSE5466. The GeneChip used for these hybridisations was the Saur2a array, which uses genomic sequences and ORF predictions from the *Staphylococcus aureus* strains N315, MRSA252, MSSA476, NCTC8325, COL, and Mu50, as well as N315 intergenic regions which are longer than 50 nucleotides. For simplicity, strain N315

was used as a reference strain in this analysis. Because this was not done by Dunman *et al.*, the list of genes which are differentially expressed differ from the one reported here. Furthermore, only genes which showed a three-fold difference in expression level in the *agr* mutant were included. For all three experiments, expression data from the transition between late exponential growth phase and early stationary phase was used.

Detection of orthologues

Orthologues between the genomes *S. aureus N315*, *L. plantarum*, *E. faecalis*, and *S. epidermidis* were detected using the Inparanoid programme (Remm, Storm et al. 2001).

The *Saccharomyces cerevisiae* quorum sensing system

Outside of bacteria, QS has also been reported in fungi. Although in fungi the chemical identity of the communication molecules varies, in all reported cases the major phenotype reported is the transition between the filamentous growth form and the solitary yeast form (Chen, Fujita et al. 2004; Chen and Fink 2006; Nickerson, Atkin et al. 2006; Sprague and Winans 2006). In the model organism *S. cerevisiae*, the aromatic alcohols phenylethanol and tryptophol, derived from the amino acids phenylalanine and tryptophan, serve as communication molecules in low nitrogen conditions (Chen and Fink 2006; Hogan 2006). Although the molecular machinery for synthesising phenylethanol and tryptophol is known, it is not yet clear how *S. cerevisiae* senses and transduces the signal. What we do know is that phenylethanol and tryptophol synergistically affect the up-regulation of *FLO11* via the cAMP-dependent PKA subunit Tpk2p and the transcription factor Flo8p (Chen and Fink 2006) (figure 3-6). Flo11p, the product of *FLO11*, is a glycosylphosphatidylinositol (GPI) anchored cell surface flocculin protein and is essential for filamentous growth (Barrales, Jimenez et al. 2008; Fischer, Valerius et al. 2008; Vinod, Sengupta et al. 2008). *S. cerevisiae* strains with deletions of either *TPK2* or *FLO8* do not form filaments in response to the aromatic alcohols (Chen and Fink 2006).

The production of phenylethanol and tryptophol is cell-density dependent. *ARO9* and *ARO10* are two genes required for their synthesis. The expression of these aromatic aminotransferases is induced by tryptophol via the Aro80p transcription factor. Since tryptophol induces the enzymes required for its own synthesis, this results in a positive feedback loop.

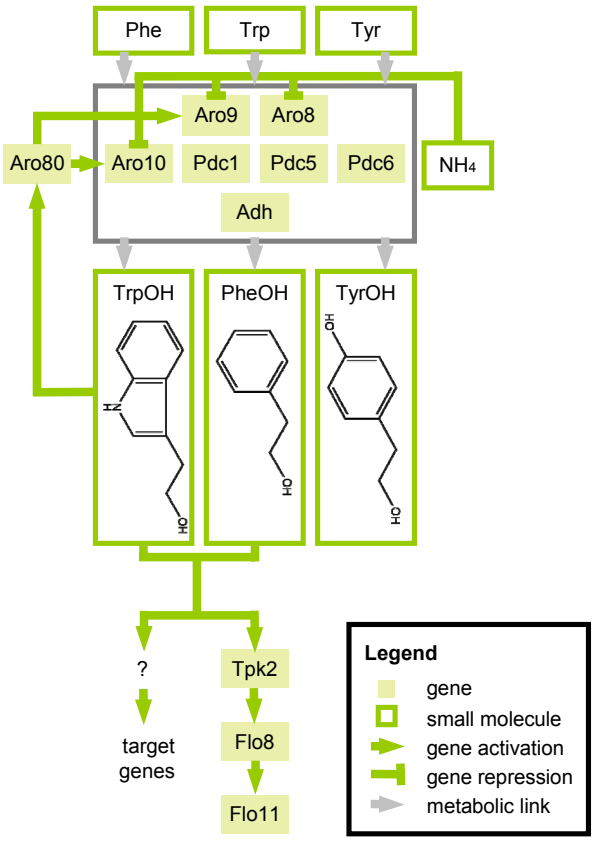


Figure 3-6. Until recently, very few genes have been implicated to be part of the aromatic alcohol QS pathway in yeast. Aromatic aminotransferases (Aro8, Aro9 and Aro10), pyruvate decarboxylases (Pdc1, Pdc5, and Pdc6) and alcohol dehydrogenases (Adh) synthesise the aromatic alcohols tryptophol (TrpOH), phenylethanol (PheOH) and tyrosol (TyrOH) from the amino acids tryptophan (Trp), phenylalanine (Phe) and tyrosine (Tyr), respectively. Aro9 and Aro10 are regulated by the transcription factor Aro80, whose activity in turn is regulated by TrpOH, thereby establishing a positive feedback loop. TrpOH and PheOH also change the expression level of around 200 other target genes, but the precise pathway by which these changes are affected was previously unknown.

Therefore, cells at high population density produce more aromatic alcohols per cell than cells at low population density. The tryptophol and phenylethanol synthesis pathways also overlap with the nitrogen sensing pathway. Ammonia represses filamentous growth as well as the expression of the above mentioned *ARO9* and *ARO10* genes. However, elements of the mitogen activated protein kinase-protein kinase A (MAPK-PKA) pathway, which is required in *S. cerevisiae* to transmit information about nitrogen shortage, do not seem to be affected by aromatic alcohol communication molecules (Chen and Fink 2006).

In *S. cerevisiae*, aromatic alcohols affect the transcript abundance of hundreds of different genes (Chen and Fink 2006). Despite this strong effect, FLO11 is the only differentially regulated gene for which the specific transcription factor that links aromatic alcohols with altered expression is known (Chen and Fink 2006). Considering the importance of *S. cerevisiae* as a model organism, it is important to gain a better understanding of the interaction of aromatic alcohol response with the

rest of the transcriptional network of the cell and the evolutionary conservation of its components in related species. In this part of my thesis, I address these questions by integrating present knowledge about the function of QS in *S. cerevisiae* with transcription regulatory networks, fungal genome sequences, and comparative genomics data (figure 3-7). I expect that the approach I present here can also be applied to examining the effects of SMs other than the aromatic alcohols of this study.

Results and discussion

Aromatic alcohols cause the differential expression of hundreds of genes by two-fold or more (Chen and Fink 2006). The genes that are differentially expressed upon addition of the three different aromatic alcohols tryptophol, phenylethanol, tyrosol, or a combination of all three, which I refer to as 3OH, overlap significantly (figure 3-8). Phenylethanol has the strongest effect and causes the differential expression of 412 genes. Tryptophol causes the differential expression of 264 genes, tyrosol of 251 genes and 3OH of 246 genes (table 3-6).

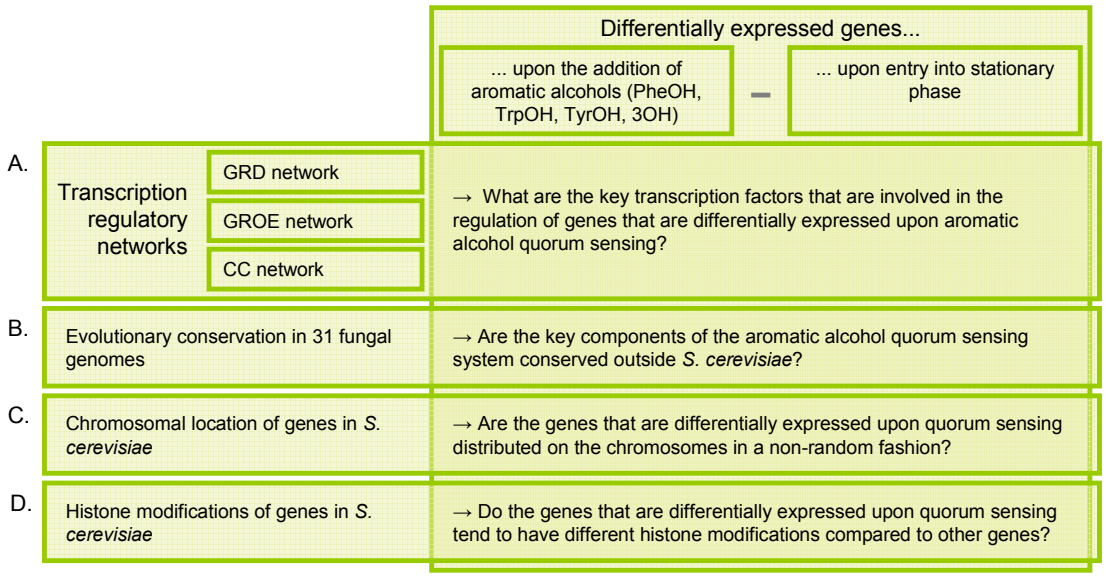


Figure 3-7. I determined sets of genes that are differentially expressed upon the addition of the aromatic alcohols tryptophol (TrpOH), phenylethanol (PheOH) and tyrosol (TyrOH), or a combination of all three (3OH), but not upon entry into stationary phase. (A) By integrating this data with three transcription regulatory networks obtained in different ways (GRD: changes in gene expression upon transcription factor deletion, GROE: changes in gene expression upon transcription factor overexpression, CC: ChIP-chip), I determined the transcription factors that are most likely to be involved in the differential regulation of these genes. (B) By examining the evolutionary conservation of a subset of the differentially expressed genes, plus other genes that have previously been implicated in the QS response, I found that many of the key components of the QS system are only conserved in the genus *Saccharomyces*. (C) By examining the chromosomal location of the differentially expressed genes, I found that they tend to be distributed in a non-random fashion. (D) In order to determine whether genes that are differentially expressed upon the addition of aromatic alcohols have different histone modifications than genes that are not, I compared the histone modifications of those genes.

71 genes are differentially regulated by all four of these treatments (figure 3-8). Though tyrosol has not been implicated as a QS molecule, its effect on transcription

with 251 differentially expressed genes is almost as strong as that of tryptophol with 264 differentially expressed genes and comparable to the 412 genes that are differentially expressed upon treatment with phenylethanol. The overlap between the differentially expressed genes between the two treatments is 141. The overlap between genes differentially expressed upon treatment with tyrosol and treatment with phenylethanol is 146 (figure 3-8).

Attributing differential expression to transcription factors

Filamentation and the cellular response to aromatic alcohol communication molecules involves a number of cellular changes, including the altered subcellular location of signalling proteins (Bharucha, Ma et al. 2008) and changes in the transcript levels of select genes. I hypothesised that differential expression of genes upon addition of aromatic alcohols is due to the differential activity of only a few transcription factors. In order to establish their identity, I developed a method that integrates transcription regulatory networks with gene expression data.

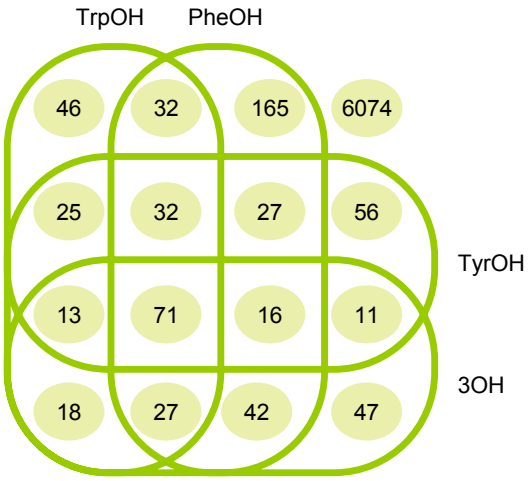


Figure 3-8. Overlap between the genes differentially expressed after the addition of different aromatic alcohols. Of a total of 6702 *S. cerevisiae* genes, 628 are differentially regulated in response to at least one of the treatments. 264 genes are differentially regulated in response to tryptophol (TrpOH), 412 in response to phenylethanol (PheOH), 251 in response to tyrosol (TyrOH), and 245 in response to a combination of all three aromatic alcohols (3OH). 71 genes are differentially expressed in response to all four treatments.

I independently from each other integrated three different transcription regulatory networks with the set of differentially expressed genes. For each transcription factor *f*, I calculated the value of d_f, which corresponds to how much genes controlled by *f* in a particular network are overrepresented in the differentially expressed genes upon treatment (see Materials and Methods). It should be noted that the transcription regulatory networks were not created under QS conditions and that chromatin remodelling under different conditions may have affected the topology of the

networks by blocking or unblocking transcription factor access to gene regulatory regions. Despite these effects, my analysis clearly identifies key transcription factors that are likely to have a role in this process.

Chen and Fink have demonstrated that the transcription factor Flo8p has differential activity in the presence of phenylethanol and tryptophol (Chen and Fink 2006). Nevertheless, *FLO8* is not differentially expressed after aromatic alcohol treatment. Neither does Flo8p control a higher proportion of differentially regulated genes than expected according to any transcriptional network upon phenylethanol treatment (figure 3-9 and appendix 2). Upon tryptophol treatment, Flo8p regulates a higher proportion of differentially regulated genes only according to the GRD network (see Material and Methods). Taken together, these results show that Flo8p is not solely responsible for the large changes in transcript levels observed upon aromatic alcohol treatment. The same applies to Aro80p, which is the only other transcription factor that has previously been reported to differentially regulate genes after the addition of an aromatic alcohol (Chen and Fink 2006). Therefore, other transcription factors must be responsible for the regulation of the differentially expressed genes.

My results suggest that the transcription factors Msn2p, Mig1p, Rgm1p, Sip4p, and Cat8p regulate genes that are differentially expressed upon phenylethanol treatment according to more than one transcriptional network. Only Cat8p and Mig1p regulate the differentially expressed genes according to all three methods, what makes it likely that these genes are important for aromatic alcohol communication. Cat8p has previously been reported to be involved in the regulation of gluconeogenic genes and most of the enzymes of the glyoxylate cycle (Biddick, Law et al. 2008). Like Cat8p, Sip4p is also involved in the regulation of enzymes involved in gluconeogenesis (Schüller 2003). Mig1p is also involved in the regulation of enzymes that participate in the response to glucose repression by influencing activators of respiration and has also been shown to regulate *CAT8* (Schüller 2003). Because filamentous growth is promoted by QS, and because filamentation may aid in survival in nutrient limiting conditions by increasing access to nutrients (Hogan 2006), these findings are consistent. It has also been hypothesised that pleiotropic drug resistance (PDR) transporters are involved in QS in *S. cerevisiae* by exporting communication molecules (Hlavacek, Kucerova et al. 2009). My findings support this hypothesis, as the transcription factor Pdr1p has a d_f value of greater than one for genes differentially regulated upon the addition of tryptophol and tyrosol (figure 3-9). As Pdr1p also regulates the expression of the PDR transporter, this could mean that in

an additional feedback loop QS molecules activate Pdr1p, which activates expression of *PDR5*, which may in turn lead to the export of further QS molecules. This is also supported by the finding that disruption of *PDR5* leads to decreased filamentation (Jin, Dobry et al. 2008).

Quorum sensing and entry into stationary phase

In order to identify genes that are differentially expressed upon entry into stationary phase, microarray data from two different previously published datasets was obtained (Gasch, Spellman et al. 2000; Andalis, Storchova et al. 2004). I defined a gene as being differentially expressed upon entry into stationary phase if it was differentially expressed in any experiment involving entry into stationary phase in the dataset by Gasch *et al*. (Gasch, Spellman et al. 2000), as well as being differentially expressed in any experiment involving entry into stationary phase in the dataset by Andalis *et al*. (Andalis, Storchova et al. 2004). Between 53% and 64% of the genes whose transcript levels are altered in response to aromatic alcohols also have altered transcript levels upon entry into stationary phase (table 3-6). This shows that the two responses overlap but are distinct.

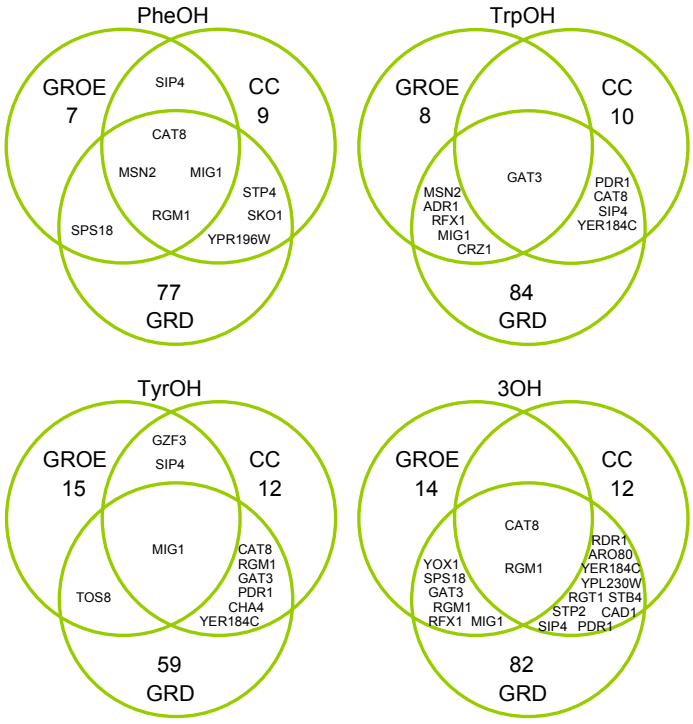


Figure 3-9. The Venn diagrams show the overlap between the transcription factors (TFs) identified using the three transcriptional networks (CC, GROE, GRD). The data refers to transcription factors that are, according to each network, strongly involved ($d_f > 1$) in the regulation of genes differentially expressed by the addition of aromatic alcohols. Only transcription factors which are predicted to have such a strong involvement according to at least two transcriptional networks are shown by name. The numbers of the other transcription factors that are overrepresented according to only one network are shown in the corresponding sections. Please refer to appendix 2 for details of statistical significance and a list of other TFs identified using the individual networks.

Treatment	*S. cerevisiae*			*C. albicans* orthologues of differentially expressed *S. cerevisiae* genes
	All differentially expressed genes	Genes that are not differentially expressed upon entry into stationary phase	Proportional decrease	
phenylethanol	412	147	64%	125
tryptophol	264	114	57%	63
tyrosol	251	118	53%	62
3OH (combination)	246	99	60%	61

Table 3-6. In order to identify genes that are differentially expressed by aromatic alcohol treatment but that are not also differentially expressed during entry into stationary phase, I removed all stationary phase-specific genes from the dataset (column 2). The number of *C. albicans* genes whose orthologues are differentially expressed in *S. cerevisiae* is given in the last column.

Genomic organization and nucleosome modification of the differentially expressed genes

Next, I investigated whether the genes that are differentially expressed upon the addition of aromatic alcohol communication molecules are distributed across the chromosomes randomly. Non-random distribution of these genes may allow quicker access of transcription factors to accessible regions of chromatin whenever this is required (Janga, Collado-Vides et al. 2008). To assess whether genes differentially expressed after aromatic alcohol addition are distributed across chromosomes in a specific way, I created random sets of genes and compared their chromosomal locations to the actually observed locations. I found that genes that are differentially expressed when aromatic alcohols are added are not distributed amongst chromosomes randomly. Genes that are differentially expressed upon the addition of phenylethanol tend to reside on chromosome 15 more often than expected by chance (figure 3-10). This also holds when genes that are differentially expressed upon entry into stationary phase are removed as described in the methods. No highly significant effect was found for tyrosol, which is in agreement with the finding that only tryptophol and phenylethanol are likely to be QS molecules (Chen and Fink 2006).

I also found that genes that are differentially expressed in response to phenylethanol are clustered within chromosomes. 48.3% of genes that are differentially regulated in response to phenylethanol are located within 10 genes of another gene (Pal and Hurst 2003; Janga, Collado-Vides et al. 2008) regulated by this aromatic alcohol (PI = 0.483, p = 0.03; figure 3-10). This drops to 26.1% when filtering removing genes that are also differentially regulated upon entry into stationary phase (PI = 0.261, p = 0.02). This suggests that genes regulated by the communication molecule

phenylethanol are not randomly distributed, but show preference for certain yeast chromosomes and regions on those chromosomes.

Figure 3-10. Genomic distribution of the genes differentially expressed under QS conditions. The chromosomal position of genes that are differentially expressed after the addition of the two QS molecules phenylethanol (left column) and tryptophol (right column) is not random. The first row shows all genes that are differentially expressed after the addition of the aromatic alcohols, and the second row shows a subset of those genes that is not also differentially expressed upon entry into stationary phase. Each gene is represented by a triangle; genes on the Watson strand are on the top and genes on the Crick strand are on the bottom of the chromosome. The figure shows that the genes that are differentially expressed in the presence of phenylethanol are distributed in a non-random way even when removing stationary phase genes. Genes differentially expressed in the presence of tryptophol are also distributed in a non-random way, but the distribution is not statistically significant when the stationary phase genes are removed.

Altogether, this implies that genes that are differentially expressed upon the addition of the aromatic alcohol QS molecules tryptophol and phenylethanol are preferentially located in specific chromosomal territories. In order to investigate this further, I examined the nucleosome modifications of those differentially expressed genes by integrating my data with a large-scale histone modification dataset (Pokholok, Harbison et al. 2005). My results show that the histone modifications H3K4Me3 (trimethylation of Lysine 4 on histone H3) and H4Ac (Acetylation of histone H4) are

significantly ($p < 0.01$) less prevalent in genes that are differentially expressed upon the addition of tryptophol (figure 3-11).

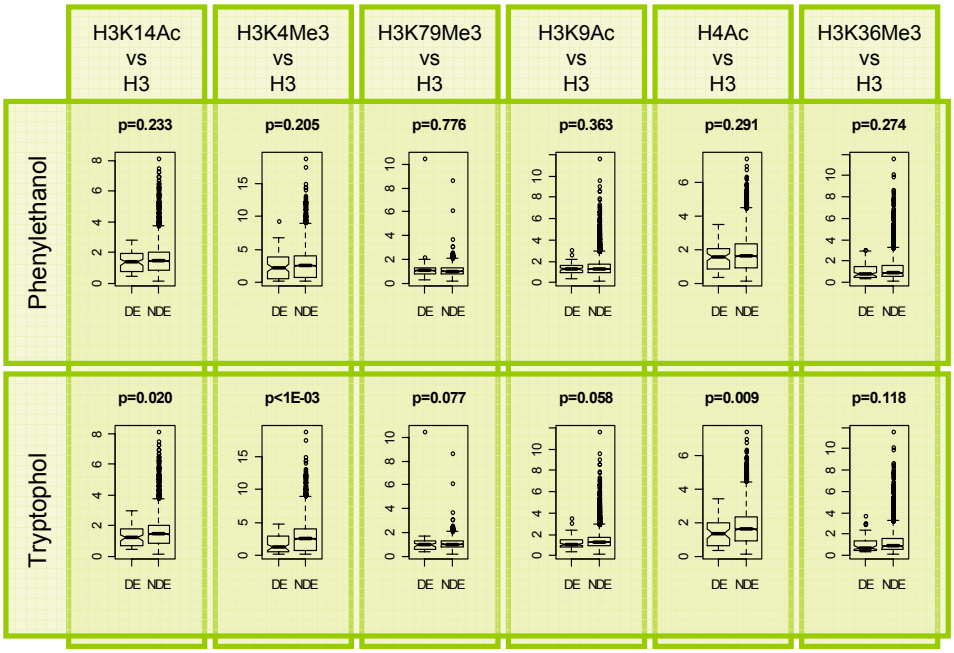


Figure 3-11. Six different histone modifications in genes that are differentially expressed upon the addition of the aromatic alcohols phenylethanol and tryptophol (DE) and those that are not (NDE). As before, genes that are differentially expressed upon entry into stationary phase have been subtracted. The *p*-values between the two sets were computed using a Wilcoxon signed-rank test. Only the histone modifications H3K4Me3 (trimethylation of Lysine 4 on histone H3) and H4Ac (Acetylation of histone H4) are significantly ($p < 0.01$) less prevalent in genes that are differentially expressed upon the addition of tryptophol. No histone modifications have a significantly different prevalence in genes that are differentially expressed upon the addition of phenylethanol.

Conservation of quorum sensing components in fungal genomes

To investigate if the QS system is conserved in other fungal species, I identified the orthologues of key *S. cerevisiae* genes involved in QS in 31 fungal species (figure 3-12 and table 3-6). This is a similar method to the one I used to identify instances of the *agr* QS system in bacteria earlier in this chapter (Wuster and Madan Babu 2008). The results show that the genes involved in aromatic alcohol synthesis are present in most yeast species. This might be because they occupy key positions in the amino acid metabolic network (Caspi, Foerster et al. 2008). The transcription factor Aro80p, which has been implicated in positive feedback between aromatic alcohols and the expression of their synthesis proteins, seems to be conserved only within the genera *Saccharomyces* and *Candida*. Most of the other transcription factors and signal transducers regulated by aromatic alcohols are conserved only in the genus *Saccharomyces* (figure 3-12).

From my analysis of the chromosomal location of the differentially expressed genes, I observed that these genes are distributed in a non-random way across the 16 different chromosomes in *S. cerevisiae*. This could be an effect of the differentially expressed genes being regulated by the same system. In order to assess whether the orthologues of differentially expressed genes are also distributed in a non-random manner in other species, I investigated the chromosomal preference and clustering of those genes in *Candida albicans*. I found that the *C. albicans* orthologues are not clustered.

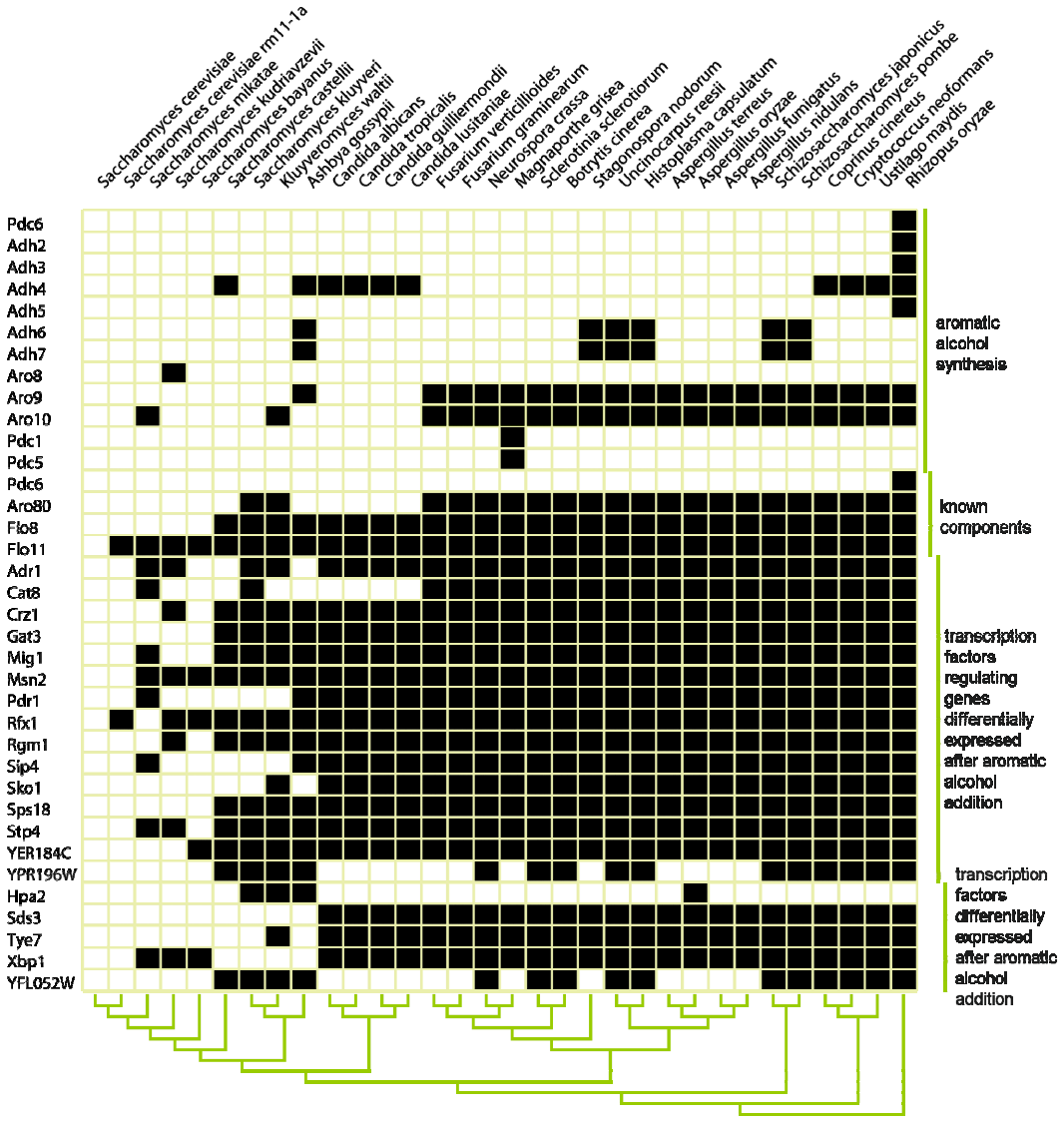


Figure 3-12. The pattern of presence and absence of *S. cerevisiae* QS genes in other fungal species supports the notion that aromatic alcohol QS is a recent evolutionary innovation. The columns of the matrix refer to fully sequenced fungal genomes, and the rows refer to genes involved in QS. If a gene is present in a genome, the corresponding square is white, and if it is absent, it is black. The genes are grouped according to how they are involved in QS (aromatic alcohol synthesis, known components of QS signal transduction pathway, transcription factors that regulate genes differentially expressed by phenylethanol and tryptophol, transcription factors regulated by phenylethanol and tryptophol). The tree at the bottom of the figure shows how the fungal species are related to each other according to Fitzpatrick *et al*. (Fitzpatrick, Logue et al. 2006).

When also taking into account experimental data, which shows that filamentation of *C. albicans* is not stimulated by tryptophol or phenylethanol (Chen and Fink 2006), this could suggest that these genes do not act in the same pathway (Janga, Collado-Vides et al. 2008).

Together with my finding that many of the genes that are involved in QS pathways in *S. cerevisiae* are not present in *C. albicans* and other more distantly related fungi, I suggest that it is unlikely that QS by aromatic alcohols is conserved in the same form outside the genus *Saccharomyces*.

Taken together, my results show that although QS using aromatic alcohols is a relatively recent evolutionary innovation in *S. cerevisiae*, and that the genes that are differentially expressed are not randomly distributed. Instead they show a clear chromosomal preference and tend to be clustered on the chromosomal arms. Through the integrated analysis of large-scale genomic datasets, I propose several transcription factors previously not known to participate in QS, such as Cat8p and Mig1p, to be involved in the process. My observation that the key elements in the circuit are conserved only in the genus *Saccharomyces*, but not in other genomes, suggests that the system is unlikely to operate in other species in the same form.

Materials and methods

Gene expression data

In order to identify genes that are differentially expressed in the presence of aromatic alcohols, I used microarray data of yeast cultures treated with the aromatic alcohols tryptophol, phenylethanol, tyrosol, or a combination of all three, as published recently by Chen and Fink (Chen and Fink 2006). Genes were considered to be differentially expressed if their intensity on the microarray differed 2-fold or more on treatment with the aromatic alcohol compared to the control condition. To identify genes that are differentially expressed only under aromatic alcohol communication conditions, it is important to identify and remove genes that are differentially expressed upon entry into stationary phase. To this end, genes that are differentially expressed upon entry into stationary phase were identified using gene expression data from two previously published experiments by Gasch *et al*. (Gasch, Spellman et al. 2000) and by Andalis *et al*. (Andalis, Storchova et al. 2004). In both experiments, the expression level of genes in different conditions, including entry into stationary phase, was determined using DNA microarrays. Genes that are uniquely differentially expressed upon

aromatic alcohol communication, but not upon entry into stationary phase, were thus identified by integrating these datasets. I defined a gene as being differentially expressed upon entry into stationary phase if it was differentially expressed in any experiment involving entry into stationary phase in the dataset by Gasch *et al*., as well as being differentially expressed in any experiment involving entry into stationary phase in the dataset by Andalis *et al*.

Identification of key transcription factors

In order to identify the transcription factors (TFs) that regulate the genes that are differentially expressed by the addition of aromatic alcohols, I integrated previously published gene expression data with transcriptional regulatory interactions that were reconstructed using three experimental methods that exploit different principles. One of them is based on ChIP-chip data (Lee, Rinaldi et al. 2002; Harbison, Gordon et al. 2004; Balaji, Madan Babu et al. 2006), which identifies direct *in vivo* DNA binding events for TFs. The other two are genetic methods that identify differentially expressed genes upon over-expression(Chua, Morris et al. 2006) or deletion (Hu, Killion et al. 2007) of TFs using microarrays. I refer to the networks obtained in this way as the CC (for Chip-Chip; 144 TFs regulating 4347 targets via 12230 regulatory interactions), GRD (for Genetic Reconstruction upon gene Deletion; 157 TFs regulating 1992 targets via 5589 regulatory interactions) and GROE (for Genetic Reconstruction upon Over-Expression; 55 TFs regulating 1951 targets via 5802 regulatory interactions) network (Balaji, Iyer et al. 2008). In the CC network, nodes represent TFs or target genes (TGs) and edges represent direct binding of the TF in the promoter region of the TG. Likewise, in the GRD and GROE networks, nodes represent TFs or TGs and a TF is linked to a TG if it is differentially expressed upon deletion or over-expression of the TF, respectively. For each of the three networks independently, I calculated the score d_f, for each transcription factor f, where

$$d_f = \log_2((t_{diffex} / t_{all}) / (a_{diffex} / a_{all}))$$

t_{diffex} is the number of differentially regulated genes controlled by f according to the specific network, t_{all} is the number of all genes regulated by f in the same network. a_{diffex} is the number of all differentially expressed genes upon treatment, and a_{all} is the number of all genes in the network. In other words, this value corresponds to how much genes controlled by the TF f in a particular network are overrepresented in the differentially expressed genes upon treatment. In this way, transcription factors that are likely to be responsible for regulating the genes that are differentially expressed

upon the addition of the aromatic alcohols can be identified. In my analysis, I chose TFs that were more than two-fold overrepresented ($d_f > 1$) as those which are likely to regulate the differentially expressed genes. I also calculated the statistical significance by estimating a p-value that each TF controls a higher proportion of the differentially expressed genes than what would be expected using a hypergeometric distribution as a null model (appendix 2).

Genomic distribution of differentially expressed genes

The chromosomal location of genes was obtained from http://www.yeastgenome.org/ for *S. cerevisiae* and from http://www.candidagenome.org/ for *C. albicans*. In order to assess whether an input set of genes is distributed randomly across the chromosomes or whether there is a preference for certain chromosomes, I performed two different types of statistical test. The first one is a χ^2 test that has as a null hypothesis that the differentially expressed genes are distributed across the 16 yeast chromosomes in an unbiased manner. The second test computes the expected distribution of genes on the 16 chromosomes by repeatedly (1000 times) and randomly picking the same number of genes as the set of differentially expressed genes. This is compared to the observed distribution of the differentially expressed genes to obtain statistical significance for chromosomal preference (p-values and Z-scores). In order to assess whether the distribution of genes within chromosomes is random or clustered, I calculated the proximity index (PI) for the differentially expressed set of genes. The PI is defined as the ratio of proximal genes to the set of differentially expressed genes. Two genes are defined to be proximal if they are separated by less than 10 genes on the chromosome (Pal and Hurst 2003; Janga, Collado-Vides et al. 2008). To assess statistical significance, the PI value was calculated from random sets of genes (1000 in this case) and compared to the observed value.

Orthologue detection

In order to assess whether the genes involved in aromatic alcohol communication are well-conserved within fungi, I identified their orthologues across 31 different fungal genomes (Cornell, Alam et al. 2007). The complete genome sequences were downloaded from http://fungal.genome.duke.edu/ and the orthologues were identified using OrthoMCL with the default parameters (Gao, Teplitski et al. 2003). The OrthoMCL output was filtered and visualised using an in-house Perl script (figure 3-12).

Epigenetic modifications

In order to determine whether genes that are differentially expressed upon the addition of the QS aromatic alcohols tryptophol and phenylethanol have different histone modifications compared to those that are not differentially expressed, I integrated my data with a large-scale nucleosome modification dataset (Pokholok, Harbison et al. 2005). As in many cases each *S. cerevisiae* gene is covered by more than one tile, in those cases I chose only to take into account the modification of the 5'-most tile. The *p*-value that the nucleosome modification in genes that are differentially expressed genes and those that are not are identical was computed using the Wilcoxon signed-rank test as implemented in the statistical programming package R.

References

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.

Andalis, A. A., Z. Storchova, et al. (2004). "Defects arising from whole-genome duplications in Saccharomyces cerevisiae." Genetics **167**(3): 1109-21.

Aoki, S. K., R. Pamma, et al. (2005). "Contact-dependent inhibition of growth in Escherichia coli." Science **309**(5738): 1245-8.

Aravind, L., V. Anantharaman, et al. (2005). "The many faces of the helix-turn-helix domain: transcription regulation and beyond." FEMS Microbiol Rev **29**(2): 231-62.

Aravind, L. and C. P. Ponting (1997). "The GAF domain: an evolutionary link between diverse phototransducing proteins." Trends Biochem Sci **22**(12): 458-9.

Balaji, S., L. M. Iyer, et al. (2008). "Comparison of transcription regulatory interactions inferred from high-throughput methods: what do they reveal?" Trends Genet **24**(7): 319-23.

Balaji, S., M. Madan Babu, et al. (2006). "Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast." J Mol Biol **360**(1): 213-27.

Barrales, R. R., J. Jimenez, et al. (2008). "Identification of novel activation mechanisms for FLO11 regulation in Saccharomyces cerevisiae." Genetics **178**(1): 145-56.

Bassler, B. L. and R. Losick (2006). "Bacterially speaking." Cell **125**(2): 237-46.

Benito, Y., F. A. Kolb, et al. (2000). "Probing the structure of RNAIII, the Staphylococcus aureus agr regulatory RNA, and identification of the RNA domain involved in repression of protein A expression." Rna **6**(5): 668-79.

Bharucha, N., J. Ma, et al. (2008). "Analysis of the yeast kinome reveals a network of regulated protein localization during filamentous growth." Mol Biol Cell **19**(7): 2708-17.

Biddick, R. K., G. L. Law, et al. (2008). "Adr1 and cat8 mediate coactivator recruitment and chromatin remodeling at glucose-regulated genes." PLoS ONE **3**(1): e1436.

Bourgogne, A., S. G. Hilsenbeck, et al. (2006). "Comparison of OG1RF and an isogenic fsrB deletion mutant by transcriptional analysis: the Fsr system of Enterococcus faecalis is more than the activator of gelatinase and serine protease." J Bacteriol **188**(8): 2875-84.

Burrell, P. C. (2006). "The detection of environmental autoinducing peptide quorum-sensing genes from an uncultured Clostridium sp. in landfill leachate reactor biomass." Lett Appl Microbiol **43**(4): 455-60.

Camilli, A. and B. L. Bassler (2006). "Bacterial small-molecule signaling pathways." Science **311**(5764): 1113-6.

Caspi, R., H. Foerster, et al. (2006). "MetaCyc: a multiorganism database of metabolic pathways and enzymes." Nucleic Acids Res **34**(Database issue): D511-6.

Caspi, R., H. Foerster, et al. (2008). "The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases." Nucleic Acids Res **36**(Database issue): D623-31.

Cassat, J., P. M. Dunman, et al. (2006). "Transcriptional profiling of a Staphylococcus aureus clinical isolate and its isogenic agr and sarA mutants reveals global differences in comparison to the laboratory strain RN6390." Microbiology **152**(Pt 10): 3075-90.

Chen, H. and G. R. Fink (2006). "Feedback control of morphogenesis in fungi by aromatic alcohols." Genes Dev **20**(9): 1150-61.
Chen, H., M. Fujita, et al. (2004). "Tyrosol is a quorum-sensing molecule in Candida albicans." Proc Natl Acad Sci U S A **101**(14): 5048-52.
Chen, X., S. Schauder, et al. (2002). "Structural identification of a bacterial quorum-sensing signal containing boron." Nature **415**(6871): 545-9.
Cheung, A. L. and S. J. Projan (1994). "Cloning and sequencing of sarA of Staphylococcus aureus, a gene required for the expression of agr." J Bacteriol **176**(13): 4168-72.
Choi, S. H. and E. P. Greenberg (1991). "The C-terminal region of the Vibrio fischeri LuxR protein contains an inducer-independent lux gene activating domain." Proc Natl Acad Sci U S A **88**(24): 11115-9.
Chua, G., Q. D. Morris, et al. (2006). "Identifying transcription factor functions and targets by phenotypic activation." Proc Natl Acad Sci U S A **103**(32): 12045-50.
Ciccarelli, F. D., T. Doerks, et al. (2006). "Toward automatic reconstruction of a highly resolved tree of life." Science **311**(5765): 1283-7.
Clark, M. A., R. A. Acharya, et al. (2009). "Design, synthesis and selection of DNA-encoded small-molecule libraries." Nat Chem Biol.
Clarke, M. B., D. T. Hughes, et al. (2006). "The QseC sensor kinase: a bacterial adrenergic receptor." Proc Natl Acad Sci U S A **103**(27): 10420-5.
Cornell, M. J., I. Alam, et al. (2007). "Comparative genome analysis across a kingdom of eukaryotic organisms: specialization and diversification in the fungi." Genome Res **17**(12): 1809-22.
Curragh, H., O. Flynn, et al. (1994). "Haloalkane degradation and assimilation by Rhodococcus rhodochrous NCIMB 13064." Microbiology **140 (Pt 6)**: 1433-42.
Diep, D. B., L. S. Havarstein, et al. (1994). "The gene encoding plantaricin A, a bacteriocin from Lactobacillus plantarum C11, is located on the same transcription unit as an agr-like regulatory system." Appl Environ Microbiol **60**(1): 160-6.
Dirix, G., P. Monsieurs, et al. (2004). "Screening genomes of Gram-positive bacteria for double-glycine-motif-containing peptides." Microbiology **150**(Pt 5): 1121-6.
Doherty, N., M. T. Holden, et al. (2006). "Functional analysis of luxS in Staphylococcus aureus reveals a role in metabolism but not quorum sensing." J Bacteriol **188**(8): 2885-97.
Drake, H. L. and S. L. Daniel (2004). "Physiology of the thermophilic acetogen Moorella thermoacetica." Res Microbiol **155**(10): 869-83.
Dunman, P. M., E. Murphy, et al. (2001). "Transcription profiling-based identification of Staphylococcus aureus genes regulated by the agr and/or sarA loci." J Bacteriol **183**(24): 7341-53.
Durbin, R., S. Eddy, et al. (1998). Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge, Cambridge University Press.
Edgar, R. C. (2004). "MUSCLE: a multiple sequence alignment method with reduced time and space complexity." BMC Bioinformatics **5**: 113.
Finn, R. D., J. Mistry, et al. (2006). "Pfam: clans, web tools and services." Nucleic Acids Res **34**(Database issue): D247-51.
Fischer, C., O. Valerius, et al. (2008). "Posttranscriptional regulation of FLO11 upon amino acid starvation in Saccharomyces cerevisiae." FEMS Yeast Res **8**(2): 225-36.
Fitzpatrick, D. A., M. E. Logue, et al. (2006). "A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis." BMC Evol Biol **6**: 99.
Gao, M., M. Teplitski, et al. (2003). "Production of substances by Medicago truncatula that affect bacterial quorum sensing." Mol Plant Microbe Interact **16**(9): 827-34.

Gasch, A. P., P. T. Spellman, et al. (2000). "Genomic expression programs in the response of yeast cells to environmental changes." Mol Biol Cell **11**(12): 4241-57.

Gould, T. A., H. P. Schweizer, et al. (2004). "Structure of the Pseudomonas aeruginosa acyl-homoserinelactone synthase LasI." Mol Microbiol **53**(4): 1135-46.

Griffiths-Jones, S., A. Bateman, et al. (2003). "Rfam: an RNA family database." Nucleic Acids Res **31**(1): 439-41.

Harbison, C. T., D. B. Gordon, et al. (2004). "Transcriptional regulatory code of a eukaryotic genome." Nature **431**(7004): 99-104.

Havarstein, L. S., G. Coomaraswamy, et al. (1995). "An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae." Proc Natl Acad Sci U S A **92**(24): 11140-4.

Hentzer, M., K. Riedel, et al. (2002). "Inhibition of quorum sensing in Pseudomonas aeruginosa biofilm bacteria by a halogenated furanone compound." Microbiology **148**(Pt 1): 87-102.

Hentzer, M., H. Wu, et al. (2003). "Attenuation of Pseudomonas aeruginosa virulence by quorum sensing inhibitors." Embo J **22**(15): 3803-15.

Hlavacek, O., H. Kucerova, et al. (2009). "Putative role for ABC multidrug exporters in yeast quorum sensing." FEBS Lett **583**(7): 1107-13.

Hogan, D. A. (2006). "Quorum sensing: alcohols in a social situation." Curr Biol **16**(12): R457-8.

Hu, Z., P. J. Killion, et al. (2007). "Genetic reconstruction of a functional transcriptional regulatory network." Nat Genet **39**(5): 683-7.

Iyer, L. M., L. Aravind, et al. (2004). "Evolution of cell-cell signaling in animals: did late horizontal gene transfer from bacteria have a role?" Trends Genet **20**(7): 292-9.

Jafra, S., J. Przysowa, et al. (2006). "Detection and characterization of bacteria from the potato rhizosphere degrading N-acyl-homoserine lactone." Can J Microbiol **52**(10): 1006-1015.

Janga, S. C., J. Collado-Vides, et al. (2008). "Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes." Proc Natl Acad Sci U S A **105**(41): 15761-6.

Ji, G., R. Beavis, et al. (1997). "Bacterial interference caused by autoinducing peptide variants." Science **276**(5321): 2027-30.

Jin, R., C. J. Dobry, et al. (2008). "Large-scale analysis of yeast filamentous growth by systematic gene disruption and overexpression." Mol Biol Cell **19**(1): 284-96.

Joint, I., K. Tait, et al. (2007). "Cross-kingdom signalling: exploitation of bacterial quorum sensing molecules by the green seaweed Ulva." Philos Trans R Soc Lond B Biol Sci **362**(1483): 1223--33.

Kaiser, D. (2004). "Signaling in myxobacteria." Annu Rev Microbiol **58**: 75-98.

Kastrup, C. J., J. Q. Boedicker, et al. (2008). "Spatial localization of bacteria controls coagulation of human blood by 'quorum acting'." Nat Chem Biol **4**(12): 742-50.

Kleerebezem, M., L. E. Quadri, et al. (1997). "Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria." Mol Microbiol **24**(5): 895-904.

Kravchenko, V. V., G. F. Kaufmann, et al. (2008). "Modulation of gene expression via disruption of NF-kappaB signaling by a bacterial small molecule." Science **321**(5886): 259-63.

Kumar, S., K. Tamura, et al. (2004). "MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment." Brief Bioinform **5**(2): 150-63.

Lazdunski, A. M., I. Ventre, et al. (2004). "Regulatory circuits and communication in Gram-negative bacteria." Nat Rev Microbiol **2**(7): 581-92.

Leadbetter, J. R. and E. P. Greenberg (2000). "Metabolism of acyl-homoserine lactone quorum-sensing signals by Variovorax paradoxus." J Bacteriol **182**(24): 6921-6.

Lee, T. I., N. J. Rinaldi, et al. (2002). "Transcriptional regulatory networks in Saccharomyces cerevisiae." Science **298**(5594): 799-804.

Lenz, D. H., K. C. Mok, et al. (2004). "The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae." Cell **118**(1): 69-82.

Madan Babu, M. and S. A. Teichmann (2003). "Evolution of transcription factors and the gene regulatory network in Escherichia coli." Nucleic Acids Res **31**(4): 1234-44.

Mashburn, L. M. and M. Whiteley (2005). "Membrane vesicles traffic signals and facilitate group activities in a prokaryote." Nature **437**(7057): 422-5.

Mayville, P., G. Ji, et al. (1999). "Structure-activity analysis of synthetic autoinducing thiolactone peptides from Staphylococcus aureus responsible for virulence." Proc Natl Acad Sci U S A **96**(4): 1218-23.

Morfeldt, E., D. Taylor, et al. (1995). "Activation of alpha-toxin translation in Staphylococcus aureus by the trans-encoded antisense RNA, RNAIII." Embo J **14**(18): 4569-77.

Morrison, D. A. (2002). Is anybody here? Cooperative bacterial regulation via peptide signals between Gram-positive bacteria. Signals, Switches, Regulons, and Cascades. D. A. Hodgson and C. M. Thomas. Cambridge, Cambridge University Press**:** 231-249.

Mullard, A. (2009). "Microbiology: Tinker, bacteria, eukaryote, spy." Nature **459**(7244): 159-61.

Nakayama, J., S. Chen, et al. (2006). "Revised model for Enterococcus faecalis fsr quorum-sensing system: the small open reading frame fsrD encodes the gelatinase biosynthesis-activating pheromone propeptide corresponding to staphylococcal agrd." J Bacteriol **188**(23): 8321-6.

Nealson, K. H., T. Platt, et al. (1970). "Cellular control of the synthesis and activity of the bacterial luminescent system." J Bacteriol **104**(1): 313-22.

Neiditch, M. B., M. J. Federle, et al. (2005). "Regulation of LuxPQ receptor activity by the quorum-sensing signal autoinducer-2." Mol Cell **18**(5): 507-18.

Neiditch, M. B., M. J. Federle, et al. (2006). "Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing." Cell **126**(6): 1095-108.

Nickerson, K. W., A. L. Atkin, et al. (2006). "Quorum sensing in dimorphic fungi: farnesol and beyond." Appl Environ Microbiol **72**(6): 3805-13.

Novick, R. P. and E. Geisinger (2008). "Quorum sensing in staphylococci." Annu Rev Genet **42**: 541-64.

Pai, A. and L. You (2009). "Optimal tuning of bacterial sensing potential." Mol Syst Biol **5**: 286.

Pal, C. and L. D. Hurst (2003). "Evidence for co-evolution of gene order and recombination rate." Nat Genet **33**(3): 392-5.

Park, S. Y., B. J. Hwang, et al. (2006). "N-acylhomoserine lactonase producing Rhodococcus spp. with different AHL-degrading activities." FEMS Microbiol Lett **261**(1): 102-8.

Peterson, M. M., J. L. Mack, et al. (2008). "Apolipoprotein B Is an innate barrier against invasive Staphylococcus aureus infection." Cell Host Microbe **4**(6): 555-66.

Pokholok, D. K., C. T. Harbison, et al. (2005). "Genome-wide map of nucleosome acetylation and methylation in yeast." Cell **122**(4): 517-27.

Present, D. A., R. Meislin, et al. (1990). "Gas gangrene. A review." Orthop Rev **19**(4): 333-41.

Qazi, S. N., E. Counil, et al. (2001). "agr expression precedes escape of internalized Staphylococcus aureus from the host endosome." Infect Immun **69**(11): 7074-82.

Qin, X., K. V. Singh, et al. (2000). "Effects of Enterococcus faecalis fsr genes on production of gelatinase and a serine protease and virulence." Infect Immun **68**(5): 2579-86.

Qiu, R., W. Pei, et al. (2005). "Identification of the putative staphylococcal AgrB catalytic residues involving the proteolytic cleavage of AgrD to generate autoinducing peptide." J Biol Chem **280**(17): 16695-704.

Queck, S. Y., M. Jameson-Lee, et al. (2008). "RNAIII-independent target gene control by the agr quorum-sensing system: insight into the evolution of virulence regulation in Staphylococcus aureus." Mol Cell **32**(1): 150-8.

Reading, N. C. and V. Sperandio (2006). "Quorum sensing: the many languages of bacteria." FEMS Microbiol Lett **254**(1): 1-11.

Recsei, P., B. Kreiswirth, et al. (1986). "Regulation of exoprotein gene expression in Staphylococcus aureus by agar." Mol Gen Genet **202**(1): 58-61.

Redfield, R. J. (2002). "Is quorum sensing a side effect of diffusion sensing?" Trends Microbiol **10**(8): 365-70.

Remm, M., C. E. Storm, et al. (2001). "Automatic clustering of orthologs and in-paralogs from pairwise species comparisons." J Mol Biol **314**(5): 1041-52.

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

Saenz, H. L., V. Augsburger, et al. (2000). "Inducible expression and cellular location of AgrB, a protein involved in the maturation of the staphylococcal quorum-sensing pheromone." Arch Microbiol **174**(6): 452-5.

Sakinc, T., P. Kulczak, et al. (2004). "Cloning of an agr homologue of Staphylococcus saprophyticus." FEMS Microbiol Lett **237**(1): 157-61.

Schüller, H. J. (2003). "Transcriptional control of nonfermentative metabolism in the yeast Saccharomyces cerevisiae." Curr Genet **43**(3): 139-60.

Schuster, M., C. P. Lostroh, et al. (2003). "Identification, timing, and signal specificity of Pseudomonas aeruginosa quorum-controlled genes: a transcriptome analysis." J Bacteriol **185**(7): 2066-79.

Shapiro, J. A. (1998). "Thinking about bacterial populations as multicellular organisms." Annu Rev Microbiol **52**: 81-104.

Shiner, E. K., K. P. Rumbaugh, et al. (2005). "Inter-kingdom signaling: deciphering the language of acyl homoserine lactones." FEMS Microbiol Rev **29**(5): 935-47.

Sprague, G. F., Jr. and S. C. Winans (2006). "Eukaryotes learn how to count: quorum sensing by yeast." Genes Dev **20**(9): 1045-9.

Sturme, M. H., J. Nakayama, et al. (2005). "An agr-like two-component regulatory system in Lactobacillus plantarum is involved in production of a novel cyclic peptide and regulation of adherence." J Bacteriol **187**(15): 5224-35.

Tait, K., I. Joint, et al. (2005). "Disruption of quorum sensing in seawater abolishes attraction of zoospores of the green alga Ulva to bacterial biofilms." Environ Microbiol **7**(2): 229-40.

Tegmark, K., E. Morfeldt, et al. (1998). "Regulation of agr-dependent virulence genes in Staphylococcus aureus by RNAIII from coagulase-negative staphylococci." J Bacteriol **180**(12): 3181-6.

Ulrich, L. E., E. V. Koonin, et al. (2005). "One-component systems dominate signal transduction in prokaryotes." Trends Microbiol **13**(2): 52-6.

Uroz, S., S. R. Chhabra, et al. (2005). "N-Acylhomoserine lactone quorum-sensing molecules are modified and degraded by Rhodococcus erythropolis W2 by both amidolytic and novel oxidoreductase activities." Microbiology **151**(Pt 10): 3313-22.

Villemur, R., M. Lanthier, et al. (2006). "The Desulfitobacterium genus." FEMS Microbiol Rev **30**(5): 706-33.

Vinod, P. K., N. Sengupta, et al. (2008). "Integration of global signaling pathways, cAMP-PKA, MAPK and TOR in the regulation of FLO11." PLoS ONE **3**(2): e1663.

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

Whitehead, N. A., A. M. Barnard, et al. (2001). "Quorum-sensing in Gram-negative bacteria." FEMS Microbiol Rev **25**(4): 365-404.

Winans, S. C. (1992). "Two-way chemical signaling in Agrobacterium-plant interactions." Microbiol Rev **56**(1): 12-31.

Wuster, A. and M. Madan Babu (2008). "Conservation and evolutionary dynamics of the agr cell-to-cell communication system across firmicutes." J Bacteriol **190**(2): 743-6.

Xavier, K. B., S. T. Miller, et al. (2007). "Phosphorylation and Processing of the Quorum-Sensing Molecule Autoinducer-2 in Enteric Bacteria." ACS Chem Biol.

Xue, Y., Y. Xu, et al. (2001). "Thermoanaerobacter tengcongensis sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China." Int J Syst Evol Microbiol **51**(Pt 4): 1335-41.

Yao, Y., C. Vuong, et al. (2006). "Characterization of the Staphylococcus epidermidis accessory-gene regulator response: quorum-sensing regulation of resistance to human innate host defense." J Infect Dis **193**(6): 841-8.

Zhang, L., J. Lin, et al. (2004). "Membrane anchoring of the AgrD N-terminal amphipathic region is required for its processing to produce a quorum-sensing pheromone in Staphylococcus aureus." J Biol Chem **279**(19): 19448-56.

Zhang, R. G., T. Pappas, et al. (2002). "Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA." Nature **417**(6892): 971-4.