The pathogenic firmicute species *Staphylococcus aureus* has a well-characterized cell-to-cell communication system whose signals are modified peptides (6). This system is encoded in the accessory gene regulator (*agr*) locus (13). The *agr* cell-to-cell signaling system is an important regulator of biofilm formation and virulence factor expression (7). The *agr* locus consists of four genes, which are *agrB*, *agrD*, *agrC*, and *agrA* (Fig. 1A). The *agrD* and *agrB* genes encode the precursor signaling peptide and the enzyme cleaving and modifying the precursor peptide, respectively (Fig. 1B). The resultant signal is called autoinducing peptide (AIP). Systems which are similar to the *agr* system either by homology or by analogy have been identified in firmicutes outside *Staphylococcus*. In some cases they have different names. For example, the *agr*-like system in *Enterococcus faecalis* is known as the *fsr* system (12) and the one in *Lactobacillus plantarum* is called the *lam* system (3). Here we investigate the distribution of *agr*-like systems outside *Staphylococcaceae* in order to understand the conservation and plasticity of their interaction with other cellular components.

In order to define *agr* homologues in taxa other than *Staphylococcus*, we built hidden Markov models of protein domains which are considered to be unique to the *agr* system, namely, the *AgrB* and *AgrD* domains. To our knowledge, no function of *AgrB* apart from processing of *AgrD* has been reported so far. In a complementary approach, we identified different names. For example, the *agr* homologue was considered to be true quorum-sensing loci. Four other loci were also considered to be true quorum-sensing 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 (10).

To investigate whether the *agr* system is an evolutionarily ancient system, we built two phylogenetic trees by use of different methods. 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 (see the supplemental material). 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* is, however, a member of the class *Clostridia*, therefore making its *agr* system a candidate for horizontal gene transfer. This is also supported by *AgrD* of *Clostridium acetobutylicum* showing remarkable similarity to *AgrD* of *Listeriaeae* (Fig. 2). To our knowledge, this would be the first reported instance of horizontal gene transfer between *Clostridia* and *Listeriaeae*.

In three genomes we were able to identify a putative *agr* system where, to our knowledge, none had been reported before. These genomes are *Moorella thermoacetica* ATCC 39070, *Desulfitobacterium hafniense* Y51, and *Thermoanaerobacter tengcongensis*. We propose that these genomes might contain a functional peptide quorum-sensing circuit. Although we could not detect genes which take the role of *agrD* in these species, this could be due to errors in the prediction of small genes or to such genes being nonhomologous and located elsewhere on the genome or fused with other genes, as in the case of *Enterococcus faecalis* (10). Alternatively, this might represent instances of a system that can recognize peptides secreted by other organisms (cross talk). In the genomes, such as those of *Bacillus halodurans*, *Syringophomonas wolfii*, and the three *Clostridium perfringens* strains, which contain an *AgrB* homologue that is not linked to a complete two-component system, *agr*-mediated quorum sensing might also be functional. This would require the two-component sensing system to be encoded elsewhere in the genome rather than being linked to *agrB*. However, in the *Clostridium perfringens* strain ATCC
FIG. 1. (A) Organization of the agr locus in S. aureus, with Pfam domains highlighted in color. nt, nucleotides. (B) Cartoon of the basic interactions of the agr system in S. aureus. The interactions of the proteins encoded by the agrBDCA operon lead to a positive-feedback loop and further agrBDCA expression. Blue arrows, phosphorylation; green arrows, transcriptional interactions. RNAIII and hld, which encodes δ-lysin,
13124 and strain 13, one of the two paralogues of AgrB cooccurs with a histidine sensor kinase. Peptide-mediated quorum sensing in the \textit{Clostridium perfringens} species would be of particular interest as this species is the causative agent of the human disease gas gangrene (11). Interestingly, we also identified an uncharacterized gene frequently present in close genomic proximity to the \textit{agr} locus which encodes a transmembrane protein with conserved residues (Fig. 1C; see also Fig. S1 in the supplemental material).

In the \textit{Clostridium perfringens} strains as well as in \textit{Bacillus halodurans}, we were able to identify short open reading frames which might have the function of \textit{agrD} without having detectable sequence similarity to it (marked with the number 4 in Fig. 1C and highlighted in bold in Fig. 2). Although they lack significant sequence similarity to known AgrD homologues, they possess key sequence features connected to AIP function: a central cysteine residue which is required for the formation of the thiolactone ring (9) and a Pro-X-X-Pro motif which might serve as a recognition site for AgrB (16) (Fig. 2).

Next, we asked whether the genes which are transcriptionally regulated by \textit{agr} are conserved between different species with a known \textit{agr} quorum-sensing system. Differences in gene expression levels between wild-type and \textit{agr} knockout strains have previously been obtained for \textit{Lactobacillus plantarum} WCFS1 (14), \textit{Enterococcus faecalis} OG1RF (1), \textit{Staphylococcus aureus} NCTC 8325 (4), \textit{Staphylococcus aureus} UAMS-1 (2), and \textit{Staphylococcus epidermidis} (15). We used these data to compare genes which are differentially expressed in these \textit{agr} mutants (see Fig. S3 in the supplemental material). According to our criteria (see the supplemental material), 175 genes were differentially expressed in \textit{Staphylococcus aureus}, 54 in \textit{Lactobacillus plantarum}, 54 in \textit{Enterococcus faecalis}, and 238 in \textit{Staphylococcus epidermidis}. After defining clusters of orthologous genes, we asked if there were any orthologues which were controlled by the same promoter and share the same transcript. The second promoter controls \textit{agrBDCA}. The \textit{agrC} and \textit{agrA} genes encode the transmembrane kinase which senses the AIP (cyan circles) and a response regulator (red hexagon) which 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 \textit{agr} operon. This results in the production of a high level of AIP, leading to even higher \textit{agr} expression.

Apart from the \textit{agrBDCA} promoter, another promoter positively regulated by AgrA is the one controlling expression of a small noncoding RNA, RNAIII. The RNAIII promoter is located next to the \textit{agrBDCA} promoter but is transcribed in the opposite direction. RNAIII rather than AgrA is the immediate regulator of most of the genes regulated by the \textit{agr} operon. (C) Tree of firmicutes based on ribosomal sequences with bootstrap values (1,000 iterations). Where applicable, the genomic organization of the \textit{agr} locus homologue is depicted to the right of the operational taxonomic unit. One instantly obvious finding is that homologues of the \textit{agr} system are spread across firmicutes and are not limited to a single phylogenetic class. Furthermore, the genomic organization of the \textit{agr} system is conserved outside the family \textit{Staphylococcaceae}, that is to say, in most cases \textit{agrB} is linked to a two-component signal transduction system. While each genome contained only a single copy of the \textit{AgrD} domain, some contained more than one copy of the \textit{AgrB} domain: \textit{M. thermoacetica} (three copies), \textit{D. hafniense} Y51 (three copies), \textit{C. perfringens} (two copies), and \textit{C. perfringens} ATCC 13124 (two copies). (D) Venn diagram showing numbers of differentially expressed genes and how many orthologous clusters are differentially expressed in more than one genome.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{alignment.png}
\caption{Alignment of potential AgrD homologues. GenBank identifiers are given in the column to the right. Red, experimentally determined signaling peptides; yellow background, residues strongly conserved across the homologues; green background, residues conserved between \textit{Listeria} and \textit{Clostridium acetobutylicum}. Boldface type shows potential AgrD homologues which were identified by genomic context (linkage to AgrB) rather than by homology. The conserved cysteine, which is required for the formation of the thiolactone ring, and the Pro-X-X-Pro motif are highlighted at the top of the alignment (see text for details).}
\end{figure}
differentially expressed in all three genomes. We did not find any cluster which was differentially regulated in all three genomes (Fig. 1D). Therefore, it appears that there is no significant propensity for orthologous genes to be regulated by the agr system in the three different species. Using the Infernal program (5) to detect small noncoding RNAs, we found that regulation of genes by agr via the global regulator RNAIII (Fig. 1B) occurred only in the Staphylococcus species. This suggests that RNAIII is an evolutionary innovation unique to Staphylococcaceae and that the downstream regulators under the control of agr can evolve rapidly.

In agreement with previous observations, genes which are involved in biofilm formation, such as those encoding exopolysaccharide biosynthesis proteins, are downregulated by agr in Staphylococcus aureus and Enterococcus faecalis (8). Some exopolysaccharide biosynthesis genes, such as those of the cps2 cluster, are upregulated in Lactobacillus 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 results for the wild type. The production of virulence proteins is upregulated by the agr-like system in Enterococcus faecalis and Staphylococcus aureus N315, which are both virulent. A clear conclusion is that the agr system is not a universal regulator of virulence factors only and is also active in nonpathogenic firmicutes. Furthermore, we conclude that although the functions which are regulated, including biofilm formation, are similar, the individual genes are not. This suggests that the regulatory network controlled by agr is flexible and can adapt rapidly to changing environments. Therefore, the agr system might be of use in the emerging field of synthetic biology.

A.W. thanks the Medical Research Council for funding. A.W. and M.M.B. thank Ingmar Schafer, Aswin Seshasayee, and Subhajyoti De for critically reading the manuscript.

REFERENCES


