Diversifying Selection at the *Bacillus* Quorum-Sensing Locus and Determinants of Modification Specificity during Synthesis of the ComX Pheromone

M. Ansaldi† and D. Dubnau*

Public Health Research Institute and Department of Microbiology and Molecular Genetics, University of Medicine and Dentistry of New Jersey, Newark, New Jersey 07103

Received 3 September 2003/Accepted 8 October 2003

The competence quorum-sensing system of *Bacillus subtilis* consists of two-component regulatory proteins, ComP (histidine kinase) and the response regulator, ComA, an extracellular pheromone (ComX), and a protein that is needed for the proteolytic cleavage and modification of pre-ComX (ComQ). ComQ and pre-ComX are both necessary and sufficient for the production of active pheromone, which is released as an isoprenylated peptide. Laboratory strain 168 and a number of natural isolates of bacilli differ in the primary sequences of their pheromones as well as in the masses of their isoprenyl adducts. We have shown that ComX, ComQ, and the membrane-localized sensor domain of ComP are highly polymorphic in natural isolates of bacilli all closely related to the laboratory strain of *B. subtilis*. In this study, we used two statistical tests (the ratio of synonymous and nonsynonymous substitution rates and the Tajima D test) to demonstrate that these polymorphic sequences evolved by diversifying selection rather than by neutral drift. We show that the choice of isoprenoid derivative is determined by the C-terminal (mature) sequence of pre-ComX rather than by the ComQ protein. The implications of these findings for the evolution of the quorum-sensing system and for the protein-protein interactions involved in determining specificity are discussed.
marked diversity and that this variability extends to both ligand and receptor proteins in several cases (25). Thus, the Bacillus quorum-sensing proteins are among many involved in cell-cell communication that exhibit polymorphism. The variability in several of the eukaryotic reproductive genes has been shown to be due to positive Darwinian selection (25).

In this report, we used sequence analysis to show that the evolution of polymorphism at the quorum-sensing locus has been dominated by positive selective forces and consequently cannot be explained by neutral changes. We also explored the determinants of specificity during the modification of pre-ComX. We show that the nature of the isoprenoid modification is determined by the C-terminal domain of pre-ComX rather than by the putative isoprenyl transferase, ComQ. The implications of these findings for the evolution of the Bacillus quorum-sensing system are discussed.

MATERIALS AND METHODS

Strains and plasmid construction. The natural Bacillus isolates and sequences used in this study, as well as the various pheromone producer and tester strains, are described elsewhere (2, 12, 22, 27). The Bacillus tester strains RO-H-1 (BD2962), RO-B-2 (BD2983), and RO-E-2 (BD3020) were grown either in liquid competence medium supplemented with glucose (0.5%) (1), t-histidine, t-lysine, and t-methionine (50 μg/ml) or on tryptose blood agar base (Difco) supplemented with chloramphenicol (5 μg/ml), erythromycin (5 μg/ml), kanamycin (5 μg/ml), spectinomycin (5 μg/ml), or tetracycline (20 μg/ml) as appropriate. Escherichia coli XL1 Blue (Stratagene) was used for cloning the comQ and comX genes into the pET22(b) vector, and transformants were selected on tryptose blood agar (Difco) plates containing ampicillin (100 μg/ml). E. coli producer strains were constructed by introducing these plasmids into E. coli BL21(DE3), permitting isopropyl-β-d-thiogalactopyranoside (IPTG) induction of transcription from the T7 promoter of the pET22(b) derivatives. DNA manipulation, cloning, and molecular biological procedures were carried out with standard protocols.

For the comX exchange experiments, plasmid pED409 (comQROB-2, comXROB-2) was amplified with the divergent primers Bam-5’-CGGGATCCAGGAGTATACTTATGAAACAAGACATGATTG) and a Bam-5’-D-galactopyranoside (ONPG) as the substrate. Each assay was performed as described (27) with tester strains grown in competence medium in the presence of purified pheromones. Samples were collected at 2 h past the end of exponential growth and assayed for β-galactosidase activity with 2,4-dinitrophenyl-p-D-galactoside (ONPG) as the substrate. Each assay was repeated at least three times. Values represent the average of at least three determinations, which exhibited a variation from the mean of no more than 10%.

Analytical protein methods. N-terminal sequencing by Edman degradation and matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry were performed on purified pheromones at the Protein Core Facility, Columbia University.

RESULTS

Positive selection at the comXRP loci. Significant polymorphism occurs at the comXRP loci in natural isolates of Bacillus compared with the variability of the housekeeping genes rpoB and gyrA (2). In each case, this polymorphism was more pronounced in the N-terminal sensor domain of ComP than in the conserved transmitter domain. When these strains were grouped by similarities at the quorum-sensing locus, a strong correlation was noted with specificity in the quorum-sensing response. These data permitted the classification of the strains into pherotype groups, defined as groups of strains that cross activate for the quorum-sensing response.

To pursue the relationship between polymorphism at the quorum-sensing locus and the emergence of different phero-
types, we constructed a tree from the comQ and comX sequences, based on the incidence of synonymous changes, inferred from the modified Nei-Gojobori method (19). This method is used to compute the numbers of synonymous and nonsynonymous substitutions and the numbers of potential synonymous and nonsynonymous sites. Diversifying selection results in an excess of nonsynonymous substitutions compared to the frequency expected from random (neutral) evolution. The comQ and comX sequences were combined and analyzed together because comX is a small gene. This simplification is appropriate, since previous phylogenetic analysis suggested that comQ and comX have similar evolutionary histories (2). A tree showing the phylogeny of the combined comQ and comX sequences of the eight strains analyzed is shown in Fig. 1.

The extensive polymorphism exhibited by the comQ and comX genes may be due to positive selection. To test this hypothesis, we again used the modified Nei-Gojobori method to compute the frequencies of nonsynonymous mutations at nonsynonymous sites (dN) and the frequencies of synonymous substitutions at synonymous sites (dS), in pairwise comparisons of the eight sequences (19) (Table 1). Neutral theory predicts that the dN/dS ratio will equal 1, whereas positive selection will result in a ratio greater than 1.

Without the sequences comprising comX and comQ have been subjected to diversifying selection.

This analysis was extended to comP. The ComP protein consists of a relatively conserved cytoplasmic transmitter domain and a polytopic, membrane-localized sensor domain of about 400 amino acid residues. Based on work with comP chimeras, we established that the sensor moiety, which is also the polymorphic portion of ComP, determines the specificity of the quorum-sensing response (unpublished data). We therefore restricted our analysis to the sensor domain of the comP genes, consisting of the promoter-proximal 1,200 nucleotide residues. The dN and dS values obtained for the receiver domain of ComP were 558.2 and 159.4, respectively (Table 2). Pairwise comparisons with a Z test gave values less than 0.004 with a strong bias toward nonsynonymous substitutions.

TABLE 1. comQX synonymous and nonsynonymous substitutions

<table>
<thead>
<tr>
<th>Non-synonymous difference (dN) in strain</th>
<th>168</th>
<th>RO-C-2</th>
<th>RO-FF-1</th>
<th>RO-H-1</th>
<th>RO-B-2</th>
<th>B. licheniformis</th>
<th>RO-F-3</th>
<th>RO-E-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>168</td>
<td>168</td>
<td>41.833</td>
<td>104.750</td>
<td>103.167</td>
<td>101.083</td>
<td>119.917</td>
<td>121.083</td>
<td>96.833</td>
</tr>
<tr>
<td>RO-C-2</td>
<td>168</td>
<td>85.167</td>
<td>94.917</td>
<td>111.250</td>
<td>106.833</td>
<td>117.417</td>
<td>113.833</td>
<td>96.08</td>
</tr>
<tr>
<td>RO-FF-1</td>
<td>168</td>
<td>181.250</td>
<td>185.083</td>
<td>118.417</td>
<td>110.083</td>
<td>108.500</td>
<td>117.167</td>
<td>106.083</td>
</tr>
<tr>
<td>RO-H-1</td>
<td>236.833</td>
<td>231.750</td>
<td>260.583</td>
<td>238.33</td>
<td>23.833</td>
<td>113.500</td>
<td>108.083</td>
<td>105.250</td>
</tr>
<tr>
<td>RO-B-2</td>
<td>249.917</td>
<td>240.167</td>
<td>258.917</td>
<td>74.167</td>
<td>233.500</td>
<td>241.917</td>
<td>110.083</td>
<td>113.000</td>
</tr>
<tr>
<td>B. licheniformis</td>
<td>254.083</td>
<td>251.583</td>
<td>279.500</td>
<td>223.500</td>
<td>241.917</td>
<td>124.083</td>
<td>108.167</td>
<td></td>
</tr>
<tr>
<td>RO-F-3</td>
<td>254.917</td>
<td>261.167</td>
<td>261.833</td>
<td>245.917</td>
<td>239.667</td>
<td>252.917</td>
<td>252.917</td>
<td>90.417</td>
</tr>
<tr>
<td>RO-E-2</td>
<td>244.167</td>
<td>244.917</td>
<td>266.917</td>
<td>236.750</td>
<td>226.000</td>
<td>234.833</td>
<td>234.833</td>
<td>239.583</td>
</tr>
</tbody>
</table>

*ds and dN were computed with the modified Nei-Gojobori test and MEGA 2.0 software. The values represent the frequencies of differences at synonymous (right) and nonsynonymous (left) sites for each pair.

Each of the comparisons gave P values less than 0.004 with a Z test, rejecting the neutral evolution hypothesis, except for the comparison between RO-H-1 and RO-B-2, which gave a P value of 0.461 (data not shown). Figure 1 shows that RO-H-1 and RO-B-2 comprised the most similar pair. The Fisher exact test and the nonmodified (original) Nei-Gojobori method (19) yielded very similar results, attesting to the robustness of these results.

Another test of neutral evolution is that of Tajima (26), which compares two measures of nucleotide diversity, one based on the number of segregating (variable) sites and the other on the average number of pairwise differences between sequences. Neutral evolution predicts that these two measures will be equal. However, diversifying selection will alter the population structure so that the second measure (based on the frequency of allelic differences) will produce a higher value. D, the difference between the two estimates, will therefore be positive. With this method, we obtained a D value of 2.09 with a total of 936 sites and 713 segregating sites (Table 2). Since the expected mean of D based on neutral evolution is 0 and the P value in this case for rejection of the neutral evolution hypothesis was less than 0.05, this result independently suggests that neutral evolution cannot explain the observed polymorphism. The large positive value of D strongly suggests that the sequences comprising comX and comQ have been subjected to diversifying selection.
TABLE 2. Statistic test summarya

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Analyzed sequences</th>
<th>comQX</th>
<th>comP</th>
<th>comP TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of sequences</td>
<td></td>
<td>8</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>No. of sites</td>
<td></td>
<td>1,113</td>
<td>1,200</td>
<td>642</td>
</tr>
<tr>
<td>No. of segregating sites</td>
<td></td>
<td>713</td>
<td>1115</td>
<td>235</td>
</tr>
<tr>
<td>Nucleotide diversity (Pr)</td>
<td></td>
<td>0.405</td>
<td>0.598</td>
<td>0.179</td>
</tr>
<tr>
<td>Nucleotide difference (avg)</td>
<td></td>
<td>379.53</td>
<td>717.70</td>
<td>108.30</td>
</tr>
<tr>
<td>Tajima’s test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D value</td>
<td></td>
<td>2.09</td>
<td>2.61</td>
<td>-0.34</td>
</tr>
<tr>
<td>P value</td>
<td></td>
<td>&lt;0.05</td>
<td>&lt;0.001</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>Modified Nei-Gojobori test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dN</td>
<td></td>
<td>230.9 (9.0)</td>
<td>558.2 (7.5)</td>
<td>54.6 (4.7)</td>
</tr>
<tr>
<td>dS</td>
<td></td>
<td>103.4 (4.3)</td>
<td>159.4 (4.0)</td>
<td>51.49 (3.8)</td>
</tr>
<tr>
<td>dN/dS</td>
<td></td>
<td>2.232</td>
<td>3.502</td>
<td>1.061</td>
</tr>
</tbody>
</table>

a SD, sensor domain; TD, transmitter domain. Tajima’s test was performed with DnaSP software. Values correspond to the mean of Tajima’s D value computed at each segregating site with exclusion of the sites with gaps. The modified Nei-Gojobori test was performed with MEGA 2.0 software. The values represent the overall mean of the differences at synonymous and nonsynonymous sites after pairwise distance evaluation. Standard errors (in parentheses) were computed by bootstrap replications (500 replicates).

suggest that diversifying selection rather than drift was responsible for the marked polymorphism at this locus.

When we applied this analysis to 634 residues, including 235 segregating sites, encoding the conserved transmitter domain of comP as a control, we obtained a D value of −0.3 with a P value of >0.05 (Table 2). For this domain, in contrast to the sensor domain, we cannot exclude the possibility that the limited polymorphism is due to neutral evolution. Taken as a whole, our analysis strongly supports a role for positive Darwinian selection in the coevolution of the comX and comQ genes and of the comP sensor domains of these Bacillus species.

Our previous work with 13 sequences suggested that comX, comQ, and comP coevolved, since similarity trees constructed for these three genes are completely congruent and differ from trees constructed for the housekeeping genes rpoB and gyrA (2). Since the quorum-sensing genes coevolved under positive selection, they presumably did so in order to maintain critical interactions between proteins, e.g., between ComQ and pre-ComX and between ComX and the sensor domain of ComP. We have begun to explore these interactions and the factors that determine the specificity of the posttranslational modifications of ComX.

Amino acid sequence of mature ComX determines the nature of the posttranslational modification. Each pheromone precursor must be processed and modified by its cognate ComQ enzyme. We previously showed that the isoprenoid modifications on ComX pheromones generally differ among the pherotypes (2). Based on analysis by mass spectrometry, there are three different modifications, which correspond in mass to geranyl or farnesyl groups and to an unknown isoprenoid derivative. We therefore asked whether the nature of the modification was determined by the sequence of pre-ComX or by ComQ. To accomplish this, we performed a comQ-comX exchange experiment between the genes of two phylogenetically related quorum-sensing loci, RO-H-1 and RO-B-2 (Fig. 1), which produce pheromones with different modifications (120 and 136 Da, respectively). Both combinations (comQRO-H1 comXRRO-B2 and comQRO-B2 comXRRO-H1) were cloned into the pET22(b) vector and expressed in E. coli. Culture supernatants were subjected to the pheromone purification protocol, and fractions were tested for their abilities to induce srfA-lacZ expression in RO-H-1 and RO-B-2 tester strains (Table 3).

The comQRO-B2 comXRRO-H1 combination produced no active pheromone, and no peak with absorbance at 214 nm corresponding to inactive pheromone eluted from the C18 column used for purification (not shown). In contrast, the comQRO-H1 comXRRO-B2 combination produced an active pheromone exhibiting the RO-B-2 specificity (Table 3), although the in vivo activity was about half that obtained with the native combination (comQRO-B2 comXRRO-B2), probably indicating a reduced efficiency in the maturation process. Mass spectrometry analysis of the purified comQRO-H1 comXRRO-B2 pheromone revealed that it carried the RO-B-2 modification of 136 Da (data not shown), even though processed by a modifying enzyme that ordinarily introduces a modification with a mass of 120 Da. Thus, the nature of the posttranslational modification was determined by the pre-ComX sequence itself rather than by the modifying enzyme.

Based on these results, two other comX genes (comXRO-C2 and comXRO-E2) were individually combined with comQRO-H1 in the same vector (see Materials and Methods). The comQRO-H1 comXRRO-C2 combination produced no active pheromone detectable by biological activity (data not shown), whereas the comQRO-H1 comXRRO-E2 combination produced some active pheromone with the RO-E-2 specificity, although again with reduced yield (Table 3). Clearly, productive heterologous ComQ-pre-ComX interactions occur with reduced efficiency or not at all, consistent with the idea that in each case, comQ and comX coevolved to ensure optimal protein-protein interactions.

In the ComX precursors, the N-terminal part is more conserved than the extreme C-terminal part, which corresponds to the mature peptide (Fig. 2). It is thus probable that the ComQ proteins interact at least in part with the precursor part of ComX, in order to cleave and modify pre-ComX. To address this hypothesis and to determine whether the extreme C-ter-
The quorum-sensing locus has an average G+C content markedly lower than that of the B. subtilis chromosome (2) and was probably introduced by horizontal transmission into a common ancestor of the strains that we have studied. The results presented in this report suggest that since that event, the quorum-sensing locus has been subjected to strong positive selection, resulting in the evolution of new pherotype specificities and of dramatic sequence polymorphism. At least three protein-protein interactions occur for the quorum-sensing system to work: ComQ with pre-ComX, ComX with the sensor domain of ComP, and ComP with ComA. The last of these interactions is between the conserved transmitter domain of ComP and the conserved receiver domain of ComA, and these interactions will probably not differ significantly among the pherotypes. The other two interactions are pherotype specific.

We have previously noted that similarity trees constructed for ComQ, pre-ComX, and the N-terminal domain of ComP are completely congruent. These trees do not correspond with those constructed by using housekeeping genes (2) or with the independently derived phylogeny of the cognate strains (22). The results indicate that *comQ*, *comX*, and *comP* coevolved, as expected from their protein-protein interactions, and did so independently of the rest of the genome. It appears likely that a change in one component which alters response specificity will be followed by changes in the other components in order to optimize the protein interactions. For instance, we may envisage that a mutational alteration in *comP* that decreases its interaction with a foreign (noncognate) ComX molecule might increase fitness by preventing inappropriate quorum-sensing responses. Since foreign pheromones sometimes interfere with the response to a cognate pheromone (2), change may protect against such interference. Such changes may carry a price, since they may also decrease responsiveness to the cognate ComX molecule. A change in *comX* which improves the fit between ComX and the altered ComP might mitigate this effect, but at the further price of decreased interaction between ComQ and pre-ComX. A mutational change in *comQ* might then result in a temporarily “optimized” system. Note that our evidence (Table 3) suggests that the efficiency of pheromone production depends on interactions between ComQ and both the N-terminal domain and the mature pheromone moieties of pre-ComX.

What fitness advantages are conferred by the quorum-sensing system in general and by the acquisition of new pherotypes? The quorum-sensing system of *B. subtilis* 168 has been identified as the first step in competence regulation (15). ComA phosphorylation leads to the synthesis of ComK, the transcriptional activator of the late competence genes. However, ComA is not devoted only to competence development, as shown by the computer search for ComA binding sites (13) and by microarray analysis (21). For example, ComA activates transcription of the sublancin lantibiotic precursor gene, the surfactin operon, the gene for peptidase, and several genes of unknown function. ComK itself induces the expression of many genes involved in the uptake and utilization of novel substrates, in growth arrest, and in DNA repair, which together comprise what has been described as the K-state (4, 8, 20).

It is not known which of the downstream genes of the quorum-sensing regulon might contribute to the fitness advantages of quorum-sensing specificity. However, the widespread presence of these quorum-sensing genes among the bacilli suggests that they do contribute to fitness, and the present demonstration of positive selection for specificity-determining change at the quorum-sensing locus suggests strongly that specificity itself contributes some advantage. What might this advantage be? The very existence of quorum sensing suggests that a given

![FIG. 2. Pre-ComX sequence alignment. An arrow indicates the site chosen for the chimeric junction. In cases where the cleavage site is known, the mature peptide sequence is underlined.](http://jb.asm.org/DownloadedFrom.html)
strain has evolved to respond when its population density is high. Presumably it will maximize its fitness when it responds to its own population density and not that of nearby but related strains. A promiscuous response would make a strain subject to fluctuations in the presence of related bacilli. If we consider the development of competence as a fitness-enhancing factor, specificity would be equivalent to a sexual isolation mechanism. The presence of related pheromones would raise the specter of genetic invasion, and therefore selection would drive changes in the quorum-sensing genes, as described above.

The polymorphism exhibited by the Bacillus quorum-sensing genes can be placed in a wider context. The competence pheromones of the streptococci exhibit a similar diversity (9), as do the pheromones of the staphylococcal Agr system (10). In those systems too, the trees built from sequence comparisons of the pheromones and of the receptors are congruent, indicating probable coevolution of the corresponding genes. Particularly interesting is the generalization that proteins mediating sexual reproduction in many eukaryotic species, including protozoa, fungi, insects, and numerous additional invertebrates as well as mammals, exhibit more divergence than do nonreproductive proteins, and several such systems have been shown to exhibit diversifying selection (reviewed in reference 25).

In fact, both the ciliate Euplotes and certain basidiomycete fungi secrete peptide pheromones that mediate mating. In Euplotes raikovi the 36- to 39-residue mature pheromones from seven mating types exhibit only seven conserved amino acids, and six of these are cysteines that form three disulfide bonds (14). In the fungal system, the sex pheromones bind to cell surface receptors and induce the expression of mating genes via a phosphorylation cascade (5), much as in the case of competence induction in the bacilli. It is remarkable that these fungal sex pheromones are also processed from preproteins and are farnesylated near their C termini, as is also true in the yeast Saccharomyces cerevisiae (5). If we regard the quorum-sensing systems of Bacillus and of Streptococcus spp. as primarily directed toward the regulation of competence, it is tempting to place these systems in the wider context of cell-cell communication mechanisms that regulate sexual behavior.

We do not know if the different isoprenoid modifications contribute to specific recognition by the ComP receptor. However, it is likely that they do, since the modifications generally differ between pherotypes (2) and unmodified peptides are completely inactive (2, 15). Our results suggest that the ComQ protein recognizes both the N-terminal and C-terminal parts of the ComX precursor. Perhaps the active modifying enzyme is a heteromultimer containing both pre-ComX and ComQ. This hypothesis is supported by the success of the ComX exchange experiment in the case of the closely related strains RO-H-1 and RO-B-2, while for RO-H-1 and RO-E-2, which are less phylogenetically related, modification of the precursor portion of pre-ComX improved the response (Table 3). The gene exchange experiments also revealed that the mature part of ComX was sufficient to determine the choice of modification (Table 3).

The studies reported here provide the basis for further exploration of the protein interactions that mediate competence induction and transmembrane signaling in the ComP/ComA two-component system as well as investigations of the evolution of quorum sensing in bacteria.

ACKNOWLEDGMENTS

We thank all members of our laboratory for frequent discussions. We also thank A. Sorokin for providing the B. licheniformis sequenced.

This work was supported by NIH grant GM 57720.

REFERENCES


