

## Evidence that the *algI/algJ* Gene Cassette, Required for O Acetylation of *Pseudomonas aeruginosa* Alginate, Evolved by Lateral Gene Transfer

Michael J. Franklin,<sup>1,2\*</sup> Stephanie A. Douthit,<sup>1,2</sup> and Marcella A. McClure<sup>1,3</sup>

Department of Microbiology,<sup>1</sup> Center for Biofilm Engineering,<sup>2</sup> and Center for Computational Biology,<sup>3</sup> Montana State University, Bozeman, Montana 59717

Received 23 December 2003/Accepted 19 April 2004

*Pseudomonas aeruginosa* strains, isolated from chronically infected patients with cystic fibrosis, produce the O-acetylated extracellular polysaccharide, alginate, giving these strains a mucoid phenotype. O acetylation of alginate plays an important role in the ability of mucoid *P. aeruginosa* to form biofilms and to resist complement-mediated phagocytosis. The O-acetylation process is complex, requiring a protein with seven transmembrane domains (AlgI), a type II membrane protein (AlgJ), and a periplasmic protein (AlgF). The cellular localization of these proteins suggests a model wherein alginate is modified at the polymer level after the transport of O-acetyl groups to the periplasm. Here, we demonstrate that this mechanism for polysaccharide esterification may be common among bacteria, since AlgI homologs linked to type II membrane proteins are found in a variety of gram-positive and gram-negative bacteria. In some cases, genes for these homologs have been incorporated into polysaccharide biosynthetic operons other than for alginate biosynthesis. The phylogenies of AlgI do not correlate with the phylogeny of the host bacteria, based on 16S rRNA analysis. The *algI* homologs and the gene for their adjacent type II membrane protein present a mosaic pattern of gene arrangement, suggesting that individual components of the multigene cassette, as well as the entire cassette, evolved by lateral gene transfer. AlgJ and the other type II membrane proteins, although more diverged than AlgI, contain conserved motifs, including a motif surrounding a highly conserved histidine residue, which is required for alginate O-acetylation activity by AlgJ. The AlgI homologs also contain an ordered series of motifs that included conserved amino acid residues in the cytoplasmic domain CD-4; the transmembrane domains TM-C, TM-D, and TM-E; and the periplasmic domain PD-3. Site-directed mutagenesis studies were used to identify amino acids important for alginate O-acetylation activity, including those likely required for (i) the interaction of AlgI with the O-acetyl precursor in the cytoplasm, (ii) the export of the O-acetyl group across the cytoplasmic membrane, and (iii) the transfer of the O-acetyl group to a periplasmic protein or to alginate. These results indicate that AlgI belongs to a family of membrane proteins required for modification of polysaccharides and that a mechanism requiring an AlgI homolog and a type II membrane protein has evolved by lateral gene transfer for the esterification of many bacterial extracellular polysaccharides.

Many bacteria produce surface and extracellular polysaccharides that play a variety of roles in bacterial survival. Examples include polysaccharides that protect bacteria from desiccation, allow root nodulation (43), and form the intercellular matrices of biofilms (12, 50). Surface and extracellular polysaccharides often act as virulence factors by protecting the bacteria from host immune and nonimmune defenses (3, 63, 65). The structures of surface and extracellular polysaccharides vary among bacteria, and these structures are important for their functional role. Structural variations include differences in the sugar subunits and in the glycosidic linkages between the subunits. In addition, the sugar subunits are often modified with carboxylic acids or amino acids, and these modifications also affect the functional roles of the polymers. Examples of polysaccharide modifications include esterification of cellular wall lipoteichoic acid (wLTA) with D-alanyl groups (5). These modifications are important for the physical structure of certain gram-positive bacteria (11) and, in the case of *Staphylococcus*

*aureus*, help protect the bacteria from antimicrobial peptides (56). Succinoglycan of *Sinorhizobium meliloti* is modified with O-succinyl groups, and these groups are required for plant root tip nodulation (42). The presence of O-acetyl groups on lipopolysaccharides (LPSs) affects the serological properties of O antigens (49, 68).

Alginate is an extracellular polysaccharide produced by *Pseudomonas aeruginosa* (16). Alginate is an important virulence factor of *P. aeruginosa*, since it encapsulates strains of *P. aeruginosa* found in chronic pulmonary infections of patients with cystic fibrosis and protects the bacteria from host defenses (3, 63, 65). *P. aeruginosa* alginate is a polymer of  $\beta$ 1-4-linked mannuronic acid residues with randomly interspersed guluronic acid residues (10, 21, 29). Alginate is esterified with O-acetyl groups at the O-2 and/or O-3 of the mannuronate residues (13, 66). The presence of O-acetyl modifications affects the physical properties of the alginate, including its viscosity and its interactions with divalent cations (67). Alginate O acetylation is also necessary for the formation of thick three-dimensional biofilms by mucoid *P. aeruginosa* (50), and for resistance of *P. aeruginosa* to complement-mediated and opsonic antibody-mediated phagocytosis (57). Therefore, O-acetylated alginate is an important virulence factor of *P. aeruginosa*.

\* Corresponding author. Mailing address: Department of Microbiology, 109 Lewis Hall, Montana State University, Bozeman, MT 59717. Phone: (406) 994-2420. Fax: (406) 994-4926. E-mail: umbfm@montana.edu.

TABLE 1. Bacterial strains and plasmids examined in this study

| Strain or plasmid     | Genotype and/or phenotype <sup>a</sup>   | Mutagenic oligonucleotide sequence | Source or reference |
|-----------------------|--|------------------------------------|---------------------|
| <i>P. aeruginosa</i>  |  |                                    |                     |
| FRD1                  | Cystic fibrosis isolate; Alg <sup>+</sup>  |                                    | This laboratory     |
| FRD1177               | <i>algI7Δ</i> Alg <sup>+</sup>   |                                    | 24                  |
| FRD1176               | <i>algJ6Δ</i> alg <sup>+</sup>   |                                    | 24                  |
| Plasmids <sup>b</sup> |  |                                    |                     |
| pRK2013               | ColE1-Tra(RK2) <sup>+</sup> ; Km <sup>r</sup>  |                                    | 19                  |
| pALTER-Ex1            | Phagemid Tc <sup>r</sup>   |                                    | Promega             |
| pMF54                 | P <sub>trc</sub> expression vector pKK233-2 with <i>oriV<sub>SF</sub></i> <i>oriT lacI<sup>q</sup></i> ; Ap <sup>r</sup> |                                    | 21                  |
| pMF149                | pMF54 with <i>algI</i> on a 1.6-kb NcoI-HindIII fragment   |                                    | 24                  |
| pMF150                | pMF54 with <i>algJ</i> on a 1.1-kb NcoI-HindIII fragment   |                                    |                     |
| pMF267                | <i>algI</i> T130A  | GCATCTCCTTCTACGCCCTTCGAGTCGATCA    | This study          |
| pMF271                | <i>algI</i> W278F  | GATCACCGAGTTCTTCCGGCGCTGGCACAT     | This study          |
| pMF272                | <i>algI</i> R280A  | CCGAGTTCTGGCGGGCCTGGCACATCAGCC     | This study          |
| pMF275                | <i>algI</i> S284A  | GGCGCTGGCACATCGCCTGTGACCTGGC       | This study          |
| pMF276                | <i>algI</i> S286A  | GGCACATCAGCCTGGCGACCTGGCTGCGCG     | This study          |
| pMF278                | <i>algI</i> H322A  | TCGGCGGCGCTGTGGGCCGGCCCAACTCA      | This study          |
| pMF265                | <i>algI</i> K94A   | TGCTCGGGTACTTCGCGTACGCCAACTTCG     | This study          |
| pMF266                | <i>algI</i> S127A  | TGCCGACATCGGCATCGCCTTCTACACCTTCG   | This study          |
| pMF268                | <i>algI</i> P162A  | TTCGTCGCGATCTTCGCGCACCTGATCGCC     | This study          |
| pMF273                | <i>algI</i> W281F  | GTCTGGCGGCGCTTCCACATCAGCCTGTC      | This study          |
| pMF274                | <i>algI</i> H282A  | TCTGGCGGCGCTGGGCCATCAGCCTGTGCA     | This study          |
| pMF277                | <i>algI</i> W321F  | CCTGGCGGCGCTTCCACGGCCCAACTT        | This study          |
| pMF290                | <i>algI</i> D242A  | GCGCAGCTCTACTTCGCCTTCTCCGGCTACAG   | This study          |
| pMF291                | <i>algI</i> F243A  | AGCTCTACTTCGACGCCTCCGGCTACAGCGAC   | This study          |
| pMF292                | <i>algI</i> Y246F  | TCGACTTCTCCGGCTTCCAGCGACATGGCCATC  | This study          |
| pMF293                | <i>algI</i> R279A  | ATCACCGAGTTCTGGGCGCGCTGGCACATCAG   | This study          |
| pMF294                | <i>algI</i> S286G  | TGGCACATCAGCCTGGGGACCTGGCTGCGCGA   | This study          |
| pMF295                | <i>algI</i> P268A  | GAGAACTCAACCAGGCCTACATCAGCCAGTC    | This study          |
| pMF296                | <i>algI</i> Y292F  | CCTGGCTGCGCGACTTCCCTACATCAGCCTG    | This study          |
| pMF297                | <i>algI</i> Y292S  | CCTGGCTGCGCGACTCCCTTACATCAGCCTG    | This study          |
| pMF298                | <i>algI</i> G167A  | GCACTGATCGCCGCCCGGTGCTGCGCTTCAA    | This study          |
| pMF299                | <i>algI</i> P168A  | ACCTGATCGCCGCGCGGTGCTGCGCTTCA      | This study          |
| pMF339                | <i>algI</i> G323A  | CGGCTGTGGCACGCCGCAACTTCACTA        | This study          |
| pMF315                | <i>algJ</i> P135A  | GTCTGCGGTGATCGCGGCCAAGGCCCGCC      | This study          |
| pMF316                | <i>algJ</i> K137A  | CGGTGATCCCCGCCGCGGCCCGCTGTATC      | This study          |
| pMF318                | <i>algJ</i> H195A  | TGCCACCCGACACCGCCTGGTGGCTCGGGC     | This study          |
| pMF319                | <i>algJ</i> W196F  | GCACCCGACCCCACTTCTCGCCGCTCGGCGCG   | This study          |
| pSAD59                | <i>algJ</i> T192A  | CGGTGTTCTGCGCGCCGACCCCACTGGTTCG    | This study          |
| pSAD61                | <i>algJ</i> T192G  | GCGGTGTTCTGCGCGCCGACCCCACTGGTTCG   | This study          |
| pSAD63                | <i>algJ</i> D193A  | GTTCCTGCGCACCCGCCCACTGGTTCGCGCCG   | This study          |

<sup>a</sup> Abbreviations: Alg<sup>+</sup>, alginate overproduction; Ap<sup>r</sup>, ampicillin resistance; Km<sup>r</sup>, kanamycin resistance; Tra<sup>+</sup>, transfer by conjugation.

<sup>b</sup> Plasmids pMF267 to pMF339 are pMF149 containing the indicated mutation. Plasmids pMF315 to pSAD63 are pMF150 containing the indicated mutation.

In previous studies, we identified three genes—*algI*, *algJ*, and *algF*—that are required for the O acetylation of *P. aeruginosa* alginate (22, 23) and characterized their protein products (24). AlgF is a periplasmic protein. AlgJ is type II membrane protein, linked to the inner membrane by an uncleaved signal peptide with the remainder of the protein facing the periplasm. AlgI is an integral membrane protein with seven membrane-spanning helices. Each of these three components is required for alginate O acetylation, suggesting a complex model for O acetylation wherein esterification occurs either associated with the bacterial inner membrane or in the periplasm after polymannuronate polymerization. Although complex, this model may represent a common mechanism for the esterification of many bacterial extracellular polysaccharides since, as shown here, AlgI homologs are found in a broad diversity of bacteria.

To gain a better understanding of the alginate O acetylation mechanism and to help identify functional domains in the alginate O-acetylation complex, we performed homology searches of AlgF, AlgJ, and AlgI amino acid sequences. Few AlgF homologs were found. However, AlgI homologs were identified among a variety of bacteria not known to produce alginate,

suggesting that AlgI may belong to a family of proteins involved in the esterification of surface or extracellular polysaccharides. The *algI* genes from these homologs are often linked to genes for one or more putative type II membrane protein. There are three distinct subsets of these type II membrane proteins, with each class apparently not sharing a common ancestor. However, each set has a motif surrounding highly conserved aspartate and histidine residues. One subset of these proteins has motifs that are conserved with AlgJ, and we used this class of proteins to identify amino acids important for the activity of AlgJ. We also used the results of these sequence homology studies to characterize the evolution of AlgI and AlgJ and to identify conserved amino acid motifs within AlgI. The results indicate that the genes for these proteins evolved by lateral transfer and that these gene cassettes may incorporate into larger polysaccharide biosynthetic operon.

#### MATERIALS AND METHODS

**Bacterial strains, plasmids, and media.** Bacterial strains, plasmids, and mutagenic oligonucleotides used in the present study are listed in Table 1. *Escherichia coli* and *P. aeruginosa* were routinely cultured in L broth (10 g of tryptone,

5 g of yeast extract, and 5 g of NaCl/liter). *Pseudomonas* isolation agar (Difco) was used to select for *P. aeruginosa* after matings with *E. coli*. Antibiotics when used were at the following concentrations: ampicillin at 100 µg/ml and carbenicillin at 300 µg/ml.

**AlGI, DltB, or genome accession numbers.** GenBank accession numbers were as follows: *P. aeruginosa* FRD1, AAB09781; *P. aeruginosa* PAO1, A83203 (71); *Azotobacter vinelandii*, AAC04568 (77); *Helicobacter pylori* 26695, AAD07902 (76); *Neisseria meningitidis* Z2491, CAB84711 (52); *N. meningitidis* MC58, AAF41650 (75); *Campylobacter jejuni*, CAB75247 (54); *Treponema pallidum*, AAC65540 (25); *Sinorhizobium meliloti*, CAC48986 (20); *Bordetella pertussis*, NC.002929, *Bordetella bronchiseptica*, NC.002927, *Bordetella parapertussis*, NC.002928 (53); *Nitrosomonas europaea*, NC.004757 (9); *Porphyromonas gingivalis* W83, NC.002950 (48); *Helicobacter hepaticus*, NC.004917 (72); *Lactococcus rhamnosus*, AAF09292 (14, 30); *Streptococcus pneumoniae*, AAK76229 (74); *Staphylococcus aureus* N315-BAB42033 (38); *Bacillus subtilis*, P39580 (27, 55); *Bacillus anthracis* Ames, NC.003997 (60); *Streptococcus pyogenes* M1 GAS, AAK34155 (18); *Streptococcus mutans*, AAC05775 (4); *Staphylococcus xylosum*, AAD01943 (56); *Clostridium tetani*, NC.004557 (6); *Clostridium perfringens*, NC.003366 (64); *Clostridium acetabutylicum*, NC.003030 (51); *Saccharomyces cerevisiae*, NP015135 (28); *Saccharomyces cerevisiae*, NP011431 (28); *Homo sapiens*, NP060664; *Arabidopsis thaliana*, BAB08549 (47); and *Plasmodium yoelii* yoelii-AABL000000000 (8).

Preliminary sequence data was obtained from The Institute for Genomic Research website (<http://www.tigr.org>) for the following organisms: *Desulfovibrio vulgaris*, *Treponema denticola*, *Listeria monocytogenes*, *Enterococcus faecalis*, and *Pseudomonas syringae*. Preliminary sequence data was obtained from the Joint Genome Institute for the following organisms: *Rhodopseudomonas palustris*, *Desulfotobacterium hafniense*, *Magnetospirillum magnetotacticum*, *Magnetococcus* sp. strain MC-1, and *Pseudomonas fluorescens*. Preliminary sequence data was obtained from OU-ACGT (<http://www.genome.ou.edu>) for *Neisseria gonorrhoeae*. Sequence data for *Candida albicans* was generated at the Stanford DNA Sequencing and Technology Center. Sequence data for *Clostridium difficile* (NC.002933), *Bacteroides fragilis*, and *Candida albicans* were produced by the Pathogen Sequencing Group at the Sanger Centre (<ftp://ftp.sanger.ac.uk/pub/pathogens/cd>).

**DNA manipulations.** General DNA manipulations were performed as described previously (2). Restriction endonucleases were purchased from New England Biolabs. Site-directed mutagenesis experiments were performed by first ligating the NcoI-HindIII fragments containing *algI* or *algJ* from plasmid pMF149 and pMF150 (24) into phagemid pALTER-EX1 (Promega), producing plasmids pMF244 and pSAD3. Single-stranded DNAs of pMF244 and pSAD3 were isolated, and site-directed mutagenesis was performed by using the Altered Sites mutagenesis protocol (Promega). Mutagenic oligonucleotides (shown in Table 1) were synthesized by Integrated DNA Technologies. Site-directed mutations were verified by DNA sequence analysis. After mutagenesis, the NcoI-HindIII fragments containing *algI* or *algJ* with point mutations were ligated into the NcoI-HindIII site of the *P. aeruginosa* P<sub>trc</sub> expression vector pMF54 (23). Triparental matings were used to mobilize the plasmids from *E. coli* into *P. aeruginosa* *algI* deletion strain FRD1177 *algI*Δ or into *algJ* deletion strain FRD1176 *algJ*Δ (24) by using the conjugative helper plasmid pRK2013 (19). Specific plasmid constructs and oligonucleotide sequences are shown in Table 1.

**Assays for alginate.** Alginates were collected from culture supernatants of mucoid *P. aeruginosa* strains grown for 24 h at 37°C in L broth supplemented with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) and carbenicillin. Alginates were purified from culture supernatants by precipitation once with 2% cetyl pyridinium chloride and twice with isopropanol (23). The concentration of alginate in solution was determined by the carbazole method of Knutson and Jeanes (36). Briefly, a solution of purified alginate (30 µl) was mixed with 1.0 ml of borate-sulfuric acid reagent (10 mM H<sub>3</sub>BO<sub>3</sub> in concentrated H<sub>2</sub>SO<sub>4</sub>), and 30 µl of carbazole reagent (0.1% in ethanol) was added. The mixture was heated to 55°C for 30 min, and the alginate concentration was determined spectrophotometrically at 530 nm by using *Macrocystis pyrifera* alginate (Sigma) as a standard.

**Assays for O acetylation of alginate.** The chemical method described previously (22) was used to measure alginate O acetylation. Briefly, 500 µl of an alginate solution was incubated with 500 µl of alkaline hydroxylamine (0.35 M NH<sub>2</sub>OH, 0.75 M NaOH) for 10 min at 25°C. The reaction mixture was acidified with 500 µl of 1.0 M perchloric acid, followed by the addition of 500 µl of 70 mM ferric perchlorate in 0.5 M perchloric acid. The concentration of O-acetyl groups was determined spectrophotometrically at 500 nm based on a standard curve by using ethyl acetate as the substrate. Alginate O acetylation was also determined by Fourier transform infrared spectroscopy (FTIR) as described previously (24).

**Immunoblot analysis.** Whole cells were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis with 4% stacking gels and 12% resolving gels

(39). After sodium dodecyl sulfate-polyacrylamide gel electrophoresis, proteins were electroblotted onto nitrocellulose membranes (1). The membranes were probed with affinity-purified AlgI antibodies. Goat anti-rabbit immunoglobulin G conjugated to horseradish peroxidase was used as the secondary antibody. Antibody binding was detected by chemiluminescent analysis (2).

**Computational analysis.** Sequence homology searches were performed by using the BLAST algorithm (1) at the National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov>) and on the unfinished microbial genomes at The Institute for Genomic Research (<http://www.tigr.org>) and at the Sanger Centre for Genome Research (<http://www.sanger.ac.uk>). Pairwise percent similarities of proteins was calculated by a BLAST analysis of two sequences program at NCBI (73) using the default Blosum 62 matrix. Multiple sequence alignments were performed by using CLUSTAL X with the default parameters and the Gonnet series matrix (32, 34). Phylogenetic analysis of aligned protein sequences was analyzed by using the neighbor-joining bootstrap analysis (17) in the CLUSTAL X program package. The TMHMM program (<http://www.cbs.dtu.dk/services/TMHMM/TMHMM2.0b.guide.html>) (37) within the simple modular architecture research tool (SMART; <http://smart.embl-heidelberg.de/>) (58, 62) was used to predict transmembrane (TM) domains.

## RESULTS

**Identification of AlgI homologs in non-alginate-producing bacteria.** In *P. aeruginosa*, *algI* is the ninth gene on the alginate biosynthetic operon (Fig. 1). Deletions of *algI* resulted in strains that produce alginate that are not O acetylated (23). Since AlgI is an integral membrane protein, it may be required for the transport of the O-acetyl group across the cytoplasmic membrane for the acylation of alginate at the polymer level (24). Sequence homology searches with BLASTP and BLAST of unfinished microbial genomes revealed AlgI homologs that are widespread among a variety of bacteria not known to produce alginate. Included among these homologs are the DltB proteins of *Bacillus subtilis* and many other gram-positive bacteria. DltB is encoded on the *dlt* operon that includes the following: *dltA* > *dltB* > *dltC* > *dltD* > *dltE* (Fig. 1). DltB proteins of gram-positive bacteria, such as *Lactobacillus rhamnosus* (30), *Bacillus subtilis* (55), and *Streptococcus mutans* (4), are required for the O alanylation of wLTA and may be involved in transport of the O-alanyl groups across the cytoplasmic membrane (35). The DltB homologs found in the genome projects of the other gram-positive bacteria, e.g., *Staphylococcus aureus* (38) and *Streptococcus pyogenes* (18), may also play roles in O-alanyl transport since these DltB homologs are also encoded on *dlt* operons. Other genes of the *dlt* operon include *dltA*, which encodes the D-alanyl carrier protein synthetase (30), and *dltC*, which encodes the D-alanyl carrier protein (35). DltD is a type II membrane protein that is probably required for transfer of alanyl groups from D-alanyl carrier protein to LTA (15).

In addition to DltB of the gram-positive bacteria, AlgI homologs were identified among bacteria from a variety of phylogenetic groups (Fig. 1). Included among the AlgI proteins from gram-negative bacteria are homologs from *Pseudomonas syringae*, which contains two genes for AlgI. One homolog is associated with an alginate biosynthetic operon, similar to *P. aeruginosa*. The other is linked to genes with similarity to the cellulose biosynthesis genes of *Gluconacetobacter xylinus* (46) (Fig. 1). This gene arrangement is similar to the homologs from *Pseudomonas fluorescens* and is involved in the O acetylation of cellulose (69, 70). The sulfate-reducing bacteria *Desulfovibrio vulgaris* and *Desulfovibrio desulfuricans* also contain *algI* homologs. In *Desulfovibrio vulgaris*,

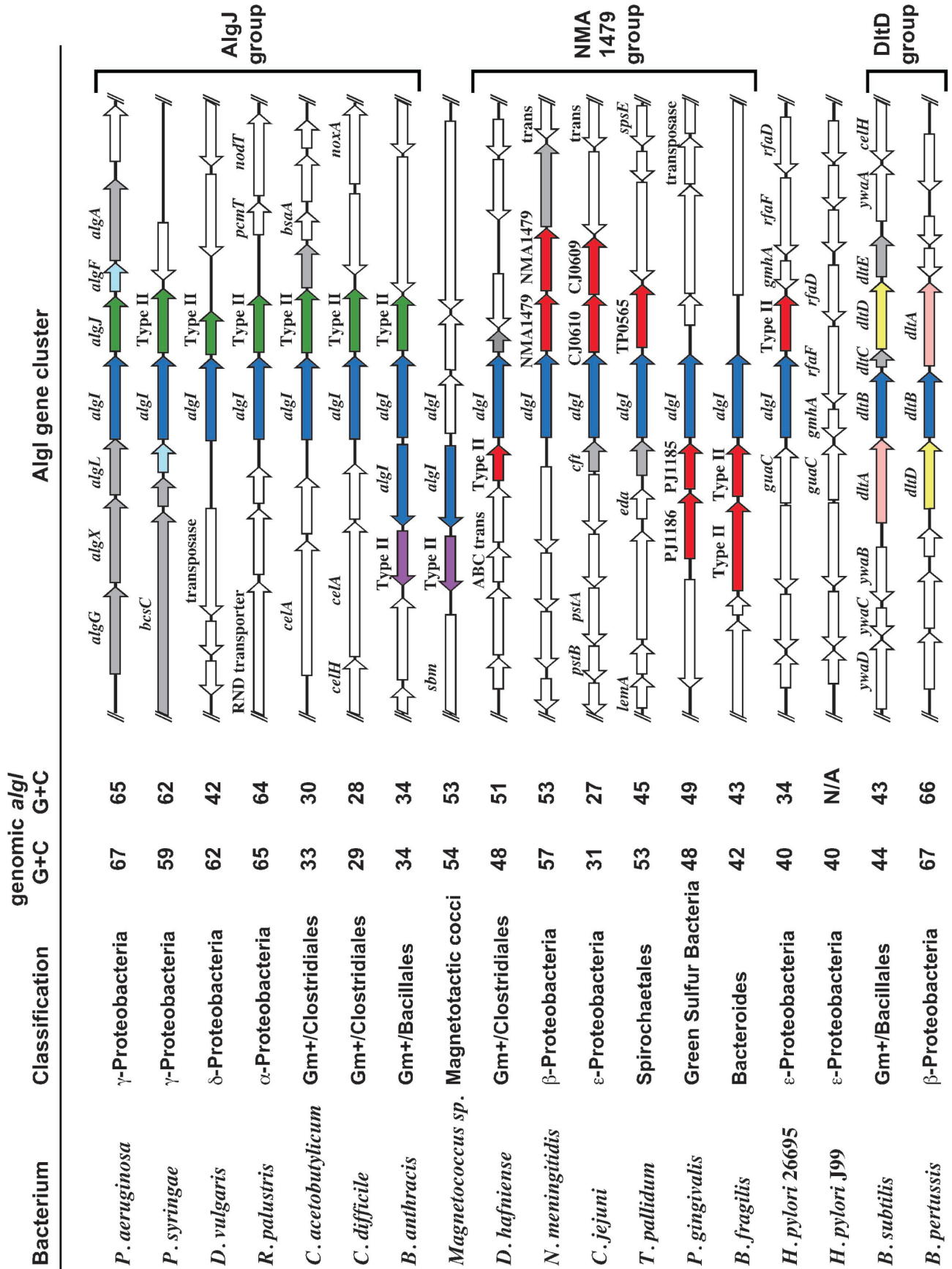


FIG. 1. Position of *algI* or *dltB* homologs within operons or putative operons. The blue arrows represent the *algI* or *dltB* homolog. All colored arrows indicate genes either demonstrated or likely to be on the same operon as *algI* or *dltB*. The green arrows show the genes for the type II membrane proteins with homology to AlgJ. The red arrows indicate genes for type II membrane proteins with conserved amino acid motifs to the *N. meningitidis* NMA1479 protein. The yellow arrows indicate genes with homology to DltD of *Bacillus subtilis*. Shown in purple are genes for type II membrane homologs of *Bacillus anthracis* and *Magnetococcus* sp. Also shown are the *algF* homologs (light blue), showing differing gene order for the alginate biosynthetic operon of *P. aeruginosa* and the putative cellulose biosynthetic operon of *P. syringae*. Shown in pink is *dltA*; the inverted gene order for *Bordetella pertussis* is also shown. The G+C contents of the genomic DNA was obtained from the Codon Usage Database (<http://www.kazusa.or.jp/codon/>) and compared to the G+C content of the *algI* homologs.

the homolog is flanked by a transposase gene and by several genes for bacteriophage proteins. Other gram-negative bacteria that contain AlgI homologs include the dental pathogens *Porphyromonas gingivalis* (which also contains adjacent transposase genes), *T. denticola*, the gastric pathogen *Campylobacter jejuni*, and *H. pylori*, as well as *N. meningitidis*, *N. gonorrhoeae*, and *T. pallidum*. Two strains of *H. pylori* have been sequenced, but only one strain, 26695, contains an *algI* homolog (Fig. 1). The *algI* homolog from *Bordetella* spp., gram-negative bacteria in the beta subdivision, are more similar to DltB than to AlgI (described below).

AlgI homologs, not linked to *dlt* operons, are also present in gram-positive bacteria. Bacteria containing both a DltB homolog (encoded on the *dlt* operon) and an AlgI homolog, include the *Clostridium* spp. *Clostridium difficile* and *Clostridium acetobutylicum*, which contain *algI* adjacent to cellulose biosynthetic genes (Fig. 1). *Bacillus anthracis* contains open reading frames for three DltB/AlgI homologs. In addition to the *dltB* gene located within a *dlt* operon, *Bacillus anthracis* contains two open reading frames for proteins with greater similarity to AlgI of *P. aeruginosa* than to DltB. The two AlgI homologs of *Bacillus anthracis* are adjacent to each other but arranged in opposite transcriptional orientation (Fig. 1). One of the *algI* homologs in *Bacillus anthracis* was previously observed by Mesnage et al. and shown to be near *csaA* and *csaB*, which are required for the addition of a pyruvyl group to a peptidoglycan-associated polysaccharide (45).

Overall, AlgI homologs were identified in more than 50 bacterial species from most of the taxa that have representative organisms undergoing genome sequencing projects. The bacteria contain from one to as many as ten copies of the *algI* gene (*Magnetospirillum magnetotacticum* contains four copies, and *Leptospira interrogans* contains ten copies), suggesting that this gene is widely distributed among bacterial phyla and that it is evolutionarily conserved. Homologs of AlgI are also found in several eukaryotes (*Plasmodium yoelii yoelii*, *Saccharomyces cerevisiae*, *Homo sapiens*, and *Arabidopsis thaliana*).

***algI* homologs are often linked to genes for type II membrane proteins that have conserved amino acid motifs.** In *P. aeruginosa* *algJ* lies immediately downstream of *algI*. AlgJ is classified as a type II membrane protein as described by Pugsley (59), since the protein is anchored in the bacterial inner membrane by an uncleaved signal peptide, whereas most of the protein resides in the periplasm (24). Sequence homology searches revealed few proteins with significant sequence identity to AlgJ (other than AlgJ and AlgX from other pseudomonads). However, most of the *algI* homologs are linked to genes for putative type II membrane proteins (Fig. 1). These predicted protein products show characteristic signal peptides with no signal peptidase cleavage site. Although they demonstrate very little primary sequence identity over the length of the proteins, multiple sequence alignments revealed conserved motifs among these proteins. These type II membrane proteins fall into three groups that do not appear to have a common ancestor for the three groups. The AlgJ group includes AlgJ-like proteins from *P. aeruginosa*, *P. syringae*, *Desulfovibrio vulgaris*, the *Clostridium* spp., and *Bacillus anthracis* (shown in green in Fig. 1). These proteins contain the conserved motifs P[X]K and RTD[X]HW (Fig. 2A). The second group of type II membrane proteins, the NMA1479 group, includes proteins

**A) AlgJ group**

|                          |      |          |      |            |
|--------------------------|------|----------|------|------------|
|                          |      | * *      |      | O** **     |
| <i>P. aeruginosa</i>     | 130- | LAVIPAKA | 191- | RTDTHWSPLG |
| <i>P. syringae</i>       |      | VAVVPDKS |      | RTDTHWSECG |
| <i>N. europaeae</i>      |      | IMIGPNKG |      | KDTHWNALG  |
| <i>D. vulgaris</i>       |      | ILIGPNKS |      | RTDTHWNLLG |
| <i>R. palustris-1</i>    |      | FVVVPDKH |      | KDDSHWTDWG |
| <i>C. tetani</i>         |      | FMLVPTSI |      | KTDHHTTNG  |
| <i>C. perfringens</i>    |      | TMIVPNSY |      | KTDHHTSYG  |
| <i>B. anthracis-1</i>    |      | FALPPSKT |      | KTDHWNMDG  |
| <i>C. acetobutylicum</i> |      | YIMPPDKY |      | RTDHHWQIPT |
| <i>C. difficile</i>      |      | YLMTPDKY |      | KTDHHTKVET |
| <i>R. palustris-2</i>    |      | YIVTPSKV |      | RGGIHWNELA |
| <i>R. palustris-3</i>    |      | FVVTPSKL |      | RDGAHWNDLG |

**B) NMA1479 group**

|                             |  |           |  |             |
|-----------------------------|--|-----------|--|-------------|
| <i>B. fragilis-2</i>        |  | IGDISHIRG |  | PDHVHYLPEG  |
| <i>P. gingivalis-1185</i>   |  | LGDSHLSG  |  | TDHIHYTVGG  |
| <i>T. pallidum-0565</i>     |  | CGDSQMRV  |  | KDGIHYTIEG  |
| <i>H. hepaticus-1083</i>    |  | IGDSMMQG  |  | DDGVHFTARG  |
| <i>C. jejuni-610</i>        |  | IGDSLMOG  |  | DDGVHFTPSG  |
| <i>P. fluorescens-2</i>     |  | VGDSLMOG  |  | DDGIHFTTTG  |
| <i>N. meningitidis-1479</i> |  | AGDSLMOG  |  | KDGIHFTAEG  |
| <i>N. gonorrhoeae-2</i>     |  | AGDSLMOG  |  | KDGIHFTAEG  |
| <i>D. hafniense</i>         |  | IGDSL-TG  |  | EDGIHFQKEF  |
| <i>B. fragilis-1</i>        |  | FGDSFIEA  |  | YDYTHINFRG  |
| <i>P. gingivalis-1186</i>   |  | LGDSFIEG  |  | RDYTHLAHRG  |
| <i>H. hepaticus-1082</i>    |  | FGDSHIAG  |  | AD-VHLLPNG  |
| <i>C. jejuni-0609</i>       |  | YGDSHMAA  |  | ND-VHLLTIKG |
| <i>P. fluorescens-1</i>     |  | LGDSHTAA  |  | NDLIHLTADG  |
| <i>N. meningitidis-1480</i> |  | IGDSHTAG  |  | KDGVHFSAKG  |
| <i>N. gonorrhoeae-1</i>     |  | IGDSHTAG  |  | KDGVHFSAQG  |

**C) DltD group**

|                      |  |           |  |             |
|----------------------|--|-----------|--|-------------|
| <i>C. difficile</i>  |  | LGSSSELSS |  | RDVMHLGTLKG |
| <i>S. aureus</i>     |  | YGSSSELGK |  | SDAVHIGWKG  |
| <i>B. anthracis</i>  |  | YGSSSEFAR |  | KDHMLGWKG   |
| <i>B. subtilis</i>   |  | YGSSSELSR |  | KDTHIGWKG   |
| <i>L. rhamnosus</i>  |  | IGSSSELSM |  | EDTHIGWRG   |
| <i>S. pyogenes</i>   |  | FGSSEWSR  |  | QDTHLGNWG   |
| <i>S. mutans</i>     |  | FGSSEWQR  |  | QDTHLGNWG   |
| <i>S. pneumoniae</i> |  | FGSSEWLR  |  | KDTHLGNWLG  |
| <i>B. pertussis</i>  |  | LGSSSELSS |  | RDDMHLAELG  |

FIG. 2. Sequence alignments of conserved motifs for the three subsets of type II membrane proteins that are genetically linked to *algI* or *dltB*. (A) AlgJ group; (B) NMA1479 group; (C) DltD group. The white letters highlighted in black show the conserved motifs surrounding an aspartate and an histidine residue, which are found in all three groups. The gray highlighted letters show conserved amino acids within each group. Asterisks indicate amino acids of AlgJ from *P. aeruginosa* that reduce alginate O acetylation at least threefold when mutated. The circle indicates an amino acid that when mutated does not affect alginate O acetylation.

from *N. meningitidis*, *Desulfotobacterium hafniense*, *Campylobacter jejuni*, *T. pallidum*, *Porphyromonas gingivalis*, and *Bacteroides fragilis* (shown in red in Fig. 1). This group of proteins share no overall sequence identity with the AlgJ proteins. They contain the conserved motifs GDS[3X]G and D[2X]HY[3X]G (Fig. 2B). Unlike the *algI* homologs, the genes for the NMA1479 group sometimes lie upstream of *algI*, and in some cases two NMA1479 homologs with conserved motifs are adjacent to *algI*. A third group of type II membrane proteins includes the DltD proteins. These proteins contain conserved motifs that are similar to the NMA1479 proteins GSSE and D[2X]HLG[2X]G (Fig. 2C). All three groups of the type II

membrane proteins contain a motif with a conserved histidine residue, followed by a large nonpolar amino acid, and a conserved aspartate residue either two or three amino acids preceding the histidine (Fig. 2).

*Bacillus anthracis* contains two *algI* homologs and also has two *algI* linked genes for type II membrane proteins (Fig. 1). One of the type II proteins falls within the AlgJ group, whereas the other type II protein is related to a protein from *Magnetococcus* spp. and has no identity with either the AlgJ, NMA1479, or DltD groups.

**Phylogenetic analysis reveals clades of AlgI homologs that do not reflect organism phylogeny.** The AlgI and DltB homologs were aligned by using the CLUSTAL X program, and bootstrap analysis was used to characterize the phylogenetic relationship of the proteins. To avoid potential biases due to the TM domains, the alignments and bootstrap analyses were performed on the entire protein sequences and on the sequence of the highly conserved cytoplasmic portion of the protein (described below). Both analyses gave similar results, and the analysis for the complete protein sequences is shown in Fig. 3. The bacterial proteins form two distinct clades: those most closely related to DltB and those related to AlgI of *P. aeruginosa*. The DltB group contains all DltB proteins from gram-positive bacteria that are encoded on *dlt* operons. The homologs from the *Bordetella* spp. are most closely related to the DltB clade, but these proteins have similarities intermediate to AlgI and DltB. For the gram-positive bacteria, the DltB phylogeny reflects the organism phylogeny based on 16S rRNA sequence analysis.

The second clade contains proteins related to AlgI of the alginate-producing bacteria *P. aeruginosa* and *Azotobacter vinelandii* and includes homologs from the non-alginate-producing gram-negative and gram-positive bacteria. In the case of the AlgI clade, the phylogeny of the AlgI proteins greatly differs from the organism phylogeny. For example, AlgI from *P. aeruginosa* (a high G+C content  $\gamma$ -proteobacterium) is more closely related to the AlgI of the low-G+C gram-positive bacteria, *Clostridium difficile* and *Clostridium acetobutylicum*, than to one of the AlgI homologs from another  $\gamma$ -proteobacteria, *P. fluorescens*. AlgI homologs from the  $\beta$ -proteobacteria, rather than forming one distinct clade, are scattered throughout the phylogenetic tree. The two AlgI homologs from *Bacillus anthracis*, although adjacent to each other on the genome, show greater similarity to other members of the AlgI clade (42% identity and 57% similarity of *Bacillus anthracis* AlgI-2 to *Magnetococcus* sp. AlgI) than to each other (35% identity and 53% similarity), suggesting that one copy of this gene was obtained from an outside source rather than by gene duplication. An AlgI homolog from a eukaryote—*Plasmodium yoelii yoelii*—clustered with the bacterial AlgI proteins. The results indicate that the AlgI homologs were obtained after bacterial speciation and likely incorporated into the genomes of these organisms by lateral gene transfer (LGT).

One indicator of recent LGT events is that the DNA G+C content of the gene of interest may differ from the G+C content of the surrounding genomic DNA, since ancient genes tend to ameliorate to the genomic G+C content over time (40). In most of the cases examined here, the G+C contents of the *algI* homologs reflect the genomic G+C content for the particular organism (as shown in Fig. 1). At the DNA level, the

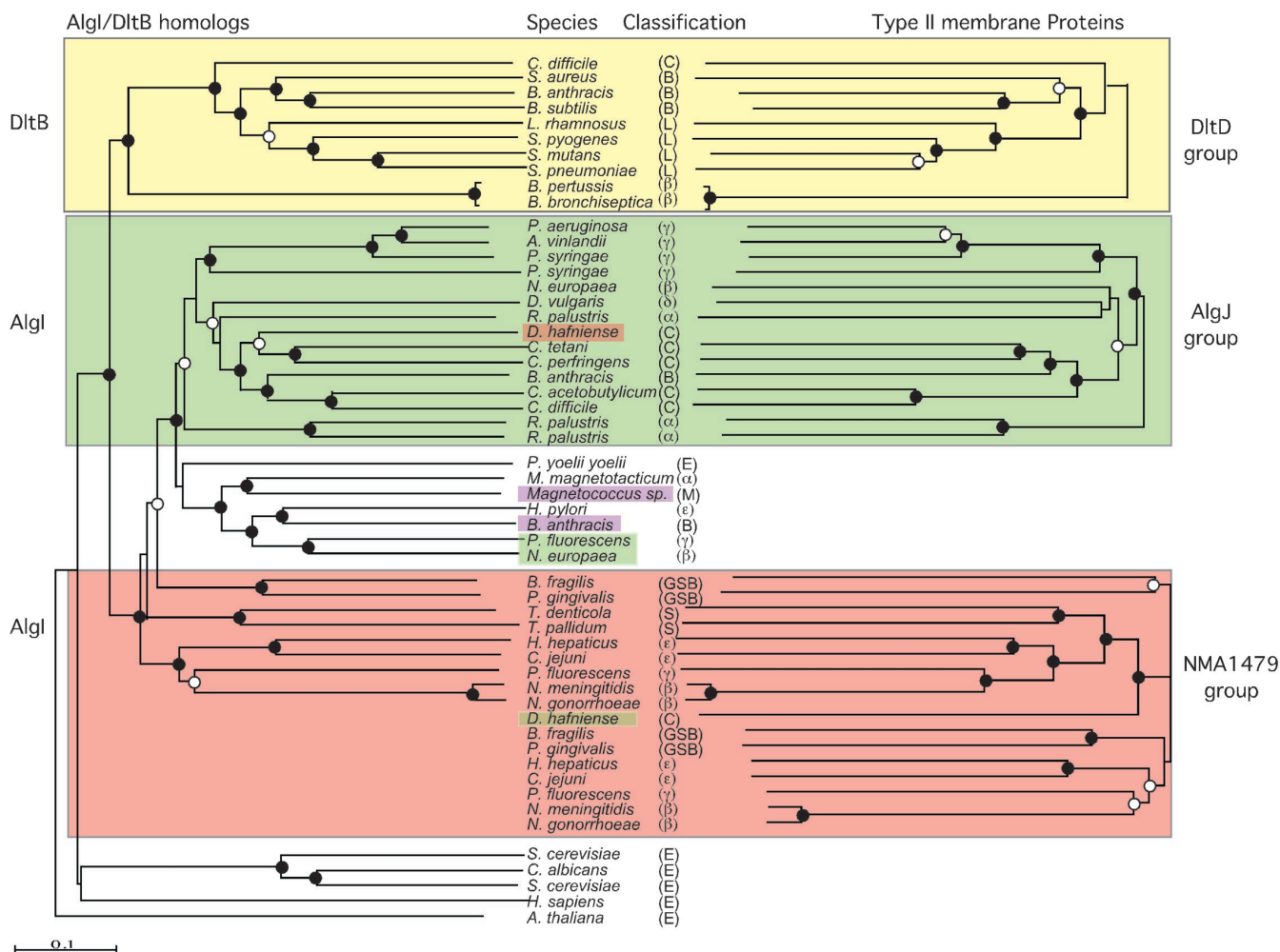


FIG. 3. Phylogenetic analysis of AlGI and DltB homologs (shown on the left-hand side) and of the type II membrane proteins encoded by genes adjacent to *algI* or *dltB* (right-hand side). Trees were constructed by using neighbor-joining and bootstrapping analysis of aligned sequences. Filled circles indicate branch points with bootstrap support of >90%. Open circles indicate branch points with bootstrap values of >70%. Branch points without circles had bootstrap values of 50 to 75%. The yellow shaded region shows the phylogeny of the DltB proteins and their genetically linked DltD protein. The green-shaded region shows the phylogeny of a subset of AlGI proteins (left) and their linked AlgJ homologs (right). The red-shaded region shows a subset of AlGI homologs and their linked homologs to NMA1479 of *N. meningitidis*. The pink-shaded region shows one of the *Bacillus anthracis* AlGI and the *Magnetococcus sp.* AlGI homologs that are linked to genes for type II membrane proteins related to each other but not related to the type II membrane proteins of the other three groups. *Desulfotubercium hafniense* encodes an AlGI homolog closely related to the *P. aeruginosa* AlGI clade but a type II membrane protein related to *N. meningitidis* NMA1479. Also shown are AlGI/DltB homolog eukaryotes. Classifications: C, Clostridiales; B, Bacillales; L, Lactobacillales; β, β-proteobacteria; γ, γ-proteobacteria; δ, δ-proteobacteria; α, α-proteobacteria; M, Magnetococcus sp.; GSB, green sulfur bacteria; S, Spirochaetales; E, eukaryotes.

*algI* genes were highly diverged, ranging from 27 to 66% G+C for the different organisms, indicating that these genes are ancient in origin and have highly varied codon usages. In some cases, however, the G+C content of the *algI* gene differs from the host DNA, indicating more recent LGT events. For example, *algI* from *Desulfovibrio vulgaris* has 20% difference in G+C content from the genomic DNA (Fig. 1). The *T. pallidum* and the *H. pylori* *algI* gene G+C contents differ from the host bacterium by 8%, also indicating that these genes may have been obtained by recent LGT.

**The *algI* homologs and the adjacent gene(s) for the type II membrane protein are topologically congruent gene cassettes.** The genes for the type II membrane protein are linked to *algI* and contain conserved motifs. Therefore, we used a phylogenetic approach to determine the relationship between these

linked genes. Since the three groups of type II membrane proteins (shown in Fig. 2) do not appear to have a common ancestor, we performed the phylogenetic analysis on the individual groups. For the DltD group, the phylogeny of DltB reflects both the 16S rRNA and the DltB phylogenies (shown in yellow in Fig. 3), suggesting that the *dlt* operon had a common ancestor prior to speciation. The other type II membrane proteins (the AlgJ group and the NMA1479 group) do not reflect the organism 16S rRNA-based phylogeny (Fig. 3). However, these phylogenies are topologically congruent with that of the AlGI proteins (Fig. 3). For example, the AlgJ group of proteins reflects the clade of AlGI proteins most closely related to the *P. aeruginosa* AlGI (shown in green in Fig. 3). Of particular interest, one of the type II membrane proteins from *Bacillus anthracis* clusters with this AlgJ group, and its AlGI

homolog is adjacent to this protein in the AlgI phylogeny. However, the other *Bacillus anthracis* type II membrane protein is not related to AlgJ but is related to the type II membrane protein from *Magnetococcus* sp. This is the same phylogenetic arrangement observed for the AlgI proteins *Magnetococcus* sp. and *Bacillus anthracis*.

The NMA1479 group is also topologically congruent with the AlgI phylogenies (shown in peach in Fig. 3). In some cases two genes for type II membrane proteins are adjacent to the *algI* gene. When two genes are adjacent to *algI*, the second set of proteins forms a separate clade, and this second clade also reflects the phylogeny of the adjacent AlgI homologs.

Exceptions to this congruence include proteins from *Desulfotibacterium hafniense* (Fig. 3). In this case, the AlgI protein clusters with proteins linked to an AlgJ homolog. However, the type II membrane protein is more closely related to the NMA1479 group (shown in peach in Fig. 3). The *P. fluorescens* and *Nitrosomonas europaea* AlgI proteins cluster with the *Bacillus anthracis*/*Magnetococcus* branch, whereas the linked type II membrane proteins are related to AlgJ.

**Conserved motif in AlgJ are required for alginate O-acetylation activity.** The multiple sequence alignments of the type II membrane proteins indicate that these proteins are highly diverged. However, they contain conserved amino acid motifs. Therefore, we performed site-directed mutagenesis studies to determine whether these conserved motifs are required for enzyme activity of AlgJ in *P. aeruginosa*. After mutagenesis, the mutant *algJ* genes were ligated into the *P. aeruginosa* expression vector pMF54 and introduced into *P. aeruginosa* FRD1155 *algJ6Δ*, a strain that does not O acetylate alginate (24). Alginates from the resulting strains were purified and analyzed for O acetylation by the colorimetric method and by FTIR. All strains were verified for levels of *algJ* expression by using immunoblots with AlgJ antibodies (24). All strains shown here demonstrated approximately equal levels of AlgJ protein (data not shown), although we cannot rule out the possibility of protein misfolding, based on these immunoblot analyses.

The ester linkage for the O-acetyl group absorbs infrared radiation at  $1,730\text{ cm}^{-1}$  and  $1,250\text{ cm}^{-1}$ , as seen in the spectrum of alginate from wild-type strain *P. aeruginosa* FRD1 (Fig. 4A). The colorimetric assay indicated that this strain had 0.79 mol of O-acetyl/mol of alginate. *P. aeruginosa* FRD1176 (*algJ6Δ*) containing the control vector pMF54 showed no absorbance at  $1,730$  and  $1,250\text{ cm}^{-1}$  and therefore no esterification of the alginate (Fig. 4B). The colorimetric data were consistent with the FTIR data. Within the 191-RTD[X]HW motif of AlgJ, the conserved histidine residue H195, when converted to alanine, completely abolished alginate O-acetylation activity (Fig. 4C). Conversion of the conserved tryptophan W196 to phenylalanine reduced alginate O-acetylation fourfold (Fig. 4D), and the D193A mutation abolished alginate O acetylation (Fig. 4E). Conversion of the conserved arginine R191 to alanine had no effect on alginate O-acetylation levels (not shown). Since one of the *Rhodospseudomonas palustris* strains had a glycine at the equivalent position of the AlgJ T192, we converted the T192 to both a glycine (Fig. 4E) and an alanine (Fig. 4F) and tested alginate O-acetylation activity. The alanine mutation reduced alginate O-acetylation activity approximately twofold, whereas the glycine mutation reduced activity fivefold, indicating that although glycine was found in

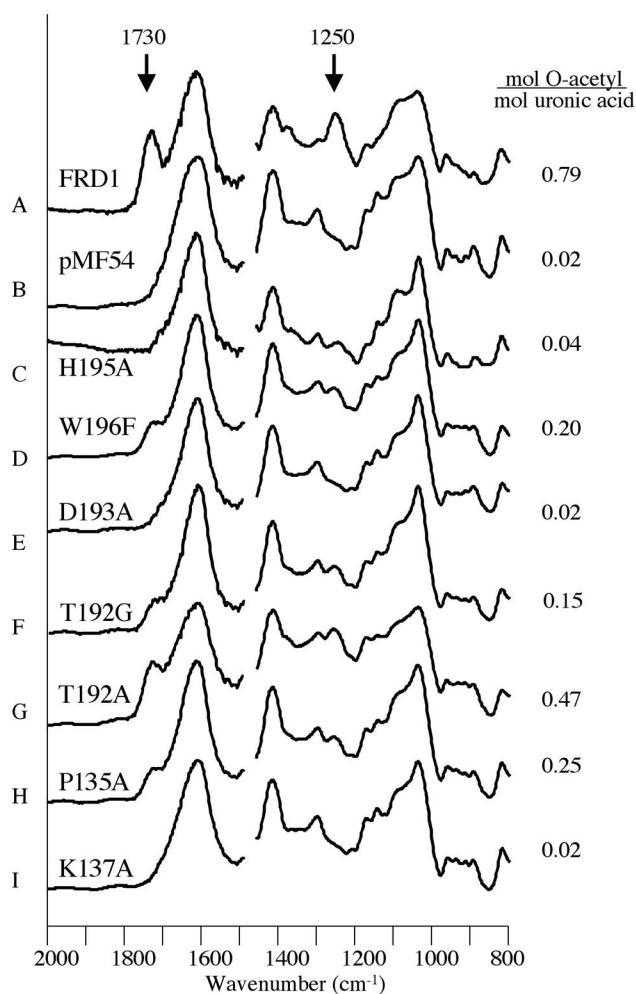


FIG. 4. FTIR spectra of alginate purified from mucoid *P. aeruginosa* strains showing presence or absence of ester linkages to O-acetyl groups. Alginates from wild-type strain FRD1 (A), FRD1176 *algJ6Δ* with control plasmid pMF54 (B), and FRD1176 (*algJ6Δ*) (C to I) with plasmids containing point mutations as indicated. The molar ratios of O-acetyl groups to uronic acid residues were determined by the colorimetric methods described in Materials and Methods. The data represent the averages for three independent strains containing the designated plasmid.

the *Rhodospseudomonas palustris* AlgJ homolog, this substitution was not fully tolerated in AlgJ of *P. aeruginosa*. Mutations were also used to determine the role of the 135-P[X]K motif on AlgJ activity. The P135A mutation reduced alginate O-acetylation threefold, whereas the K137A mutation abolished alginate O acetylation (Fig. 4H and I).

**An ordered series of motifs is required for AlgI activity: characterization of TM domains TM-C, TM-D, and TM-E.** The sequence alignment information was used to identify an ordered series of motifs in the AlgI homologs: conserved regions of the protein that are probably important for enzyme function (44). In our previous study using alkaline phosphatase protein fusions, we predicted that AlgI is a membrane protein with seven TM domains (24). Alignment of the seven TM domains of AlgI/DltB homologs demonstrated conserved hydrophobic amino acid substitution in TM-A, TM-B, TM-F, and TM-G but no amino acid identity. However, amino acid iden-

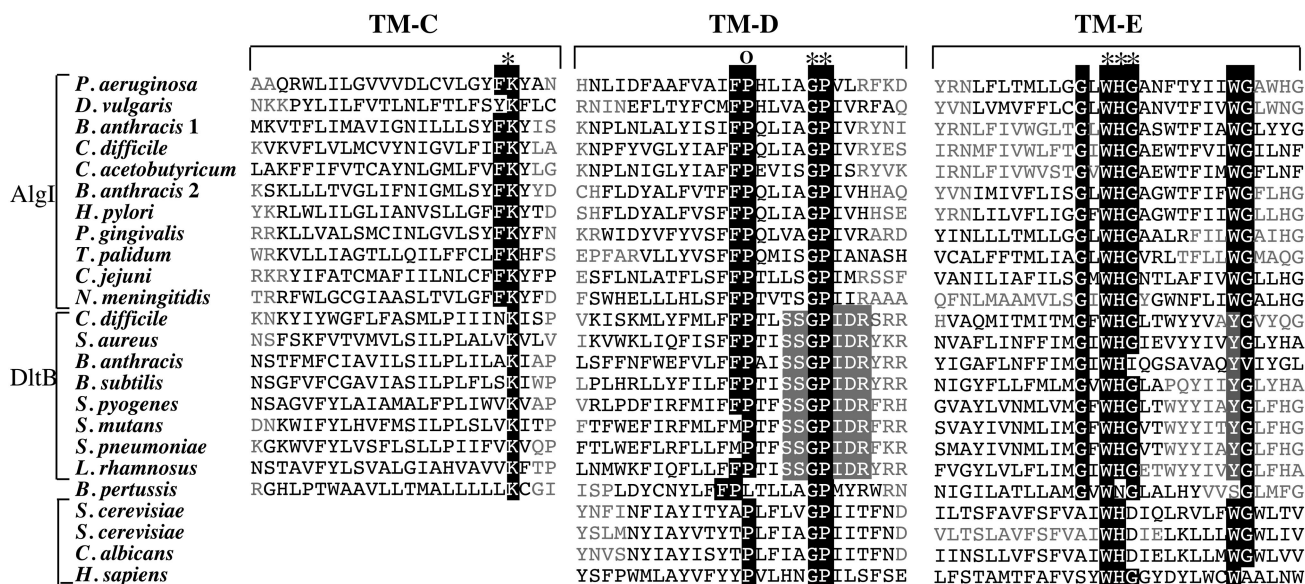


FIG. 5. Sequence alignment of the TM domains TM-C, TM-D, and TM-E. Black letters within the sequence alignment indicate the TM regions predicted by TMHMM program (37). The white letters shaded in black are amino acids conserved throughout the AlgI/DltB homologs. The white letters shaded in gray are amino acids conserved throughout the DltB homologs. Asterisks indicate amino acids subjected to site-directed mutagenesis, where the mutation resulted in at least a fourfold decrease in alginate O acetylation. The circles indicate an amino acid change that had little effect on alginate O acetylation.

tity is present in TM-C, TM-D, and TM-E of the AlgI/DltB homologs, including amino acids with polar side chains (Fig. 5). TM-C contains a conserved lysine residue in approximately the same position within the predicted membrane spanning region of each of the bacterial proteins. TM-D contains the conserved motif FP[4X]GP in the AlgI homologs, P[2X]SSGP in the DltB homologs, and P[4X]GP in the eukaryotic homologs. TM-E has the conserved motif G[1X]WHG[7X]WG in the AlgI homologs, G[1X]WHG[7X]YG and A[1X]WNG[7X]SG in the DltB homologs, and A[1X]WHD[7X]WG in the eukaryotic homologs (Fig. 5).

To determine whether conserved amino acids in the TM-C, TM-D, and TM-E domains are required for AlgI activity, we performed site-directed mutagenesis and determined the in vivo O-acetylation activity in *P. aeruginosa* FRD1177 (*algI7Δ*) (24). To help ensure that the mutations did not affect membrane topology, all noncyclic amino acids were converted to alanines, and the conserved tryptophans were converted to phenylalanine. Alginates from the resulting strains were purified from these strains and assayed for O-acetylation (Table 2). In TM-C, mutation of the conserved lysine resulted in greater than 10-fold decrease in alginate O acetylation. In TM-D, the mutations P162A in the conserved 162-P[4X]PG motif had little effect on alginate O acetylation, whereas the mutations P167A and G168A abolished alginate O acetylation. In the 279-G[1X]WHG[7X]WG motif of TM-E, the W281F and G283A mutations resulted in a three- and fourfold decreases in O acetylation, respectively. No alginate O acetylation was detected in the strain containing a mutation in histidine (H282A).

**Cytoplasmic domain 4 contains conserved amino acid motifs required for alginate O acetylation.** Our previous *phoA* fusion data predicted a cytoplasmic domain (CP-4) that is highly conserved among the AlgI homologs. Figure 6 shows this re-

gion with conserved amino acids of the AlgI homologs and of the DltB homologs highlighted. In the AlgI homologs this region of the protein contains three conserved motifs with the sequences [L/I][F/Y][X]DFSGYXD, NF[2X]P, and FW[X]RWHISLS[5X]Y[L/I]Y[2X]LGG. Similar motifs were also observed in the DltB homologs, and part of this cytoplasmic regions was also conserved among the AlgI homologs from eukaryotes (Fig. 6). Mutagenesis studies here demonstrate that D242A, F243A, and Y246F in the 239-[L/I][F/Y][X]DFS GYSD motif abolish alginate O acetylation (Table 2). Although conserved, mutation P268A, in the 264-NF[2X]P motif, did not affect alginate O acetylation in vivo. Mutations in the 277-FW[X]RWHISLS[5X]Y[L/I]Y[2X]LGG motif affected alginate O acetylation. The W278F, R279A, and W281F mutations resulted in five- to tenfold reduction in alginate O acetylation. Mutation S283A and S285A reduced alginate O acetylation three- to fourfold. Mutation R280A, although in a highly conserved amino acid, had little effect on activity. Since the AlgI homologs from *Bacillus anthracis* and from *Clostridium difficile* contained glycine residues instead of serines at the position corresponding to S285 of *P. aeruginosa* AlgI, we changed this amino acid to a glycine and to an alanine. Conversion S285A reduced alginate O acetylation threefold, whereas S285G had no effect on O-acetylation activity (Table 2), indicating that glycine could substitute for serine in this position. Since the DltB proteins contained phenylalanine at position Y292 of the *P. aeruginosa* AlgI, we mutated this tyrosine to both a phenylalanine and a serine. The Y292F mutation resulted in a twofold decrease in alginate O acetylation, and the Y292S mutation resulted in a threefold decrease in alginate O-acetylation activity, demonstrating that these amino acids could substitute for tyrosine in *P. aeruginosa* but with reduced in vivo activity.

TABLE 2. Assay for in vivo O acetylation of alginate by mutant AlgI

| Strain  | Plasmid <sup>a</sup> | Domain <sup>b</sup> | Sequence <sup>c</sup>   | O acetylation (mol of O-acetyl group/mol of uronic acid) <sup>d</sup> |
|---------|----------------------|---------------------|---|---|
| FRD1    | wt                   | TM-C                | QRWLILGVVV <sup>U</sup> DL <sup>C</sup> VLGYFKYA              | 0.70  |
| FRD1177 | pMF54                |                     |   | 0.01  |
|         | pMF149(wt)           | TM-C                | QRWLILGVVV <sup>U</sup> DL <sup>C</sup> VLGYFKYA              | 0.59  |
|         | pMF265               | TM-C                | QRWLILGVVV <sup>U</sup> DL <sup>C</sup> VLGYF <sup>A</sup> YA | 0.05  |
|         | pMF149(wt)           | TM-D                | NLIDFAAFVAIFPHLIA <sup>G</sup> PVL                            | 0.59  |
|         | pMF268               | TM-D                | NLIDFAAFVAIF <sup>A</sup> H <sup>L</sup> IAGPVL               | 0.46  |
|         | pMF298               | TM-D                | NLIDFAAFVAIFPHLIA <sup>A</sup> PVL                            | 0.01  |
|         | pMF299               | TM-D                | NLIDFAAFVAIFPHLIA <sup>G</sup> AVL                            | 0.02  |
|         | pMF149(wt)           | TM-E                | LFLTMLLGG <sup>L</sup> LWHGANFTYII                            | 0.59  |
|         | pMF277               | TM-E                | LFLTMLLGG <sup>L</sup> LWHGANFTYII                            | 0.19  |
|         | pMF278               | TM-E                | LFLTMLLGG <sup>L</sup> WAGANFTYII                             | 0.04  |
|         | pMF339               | TM-E                | LFLTMLLGG <sup>L</sup> LWHGANFTYII                            | 0.13  |
|         | pMF149(wt)           | PP-3                | THILLPIGISFYTFESISY   | 0.59  |
|         | pMF266               | PP-3                | THILLPIG <sup>L</sup> IFYTFESISY                              | 0.16  |
|         | pMF267               | PP-3                | THILLPIGISFY <sup>A</sup> FESISY                              | 0.36  |
|         | pMF149(wt)           | CP-4                | LYP <sup>D</sup> DFSGYS <sup>D</sup> MAIGLGLMM                | 0.59  |
|         | pMF290               | CP-4                | LYP <sup>A</sup> DFSGYS <sup>D</sup> MAIGLGLMM                | 0.01  |
|         | pMF291               | CP-4                | LYP <sup>D</sup> ASGYS <sup>D</sup> MAIGLGLMM                 | 0.01  |
|         | pMF292               | CP-4                | LYP <sup>D</sup> DFSG <sup>E</sup> S <sup>D</sup> MAIGLGLMM   | 0.03  |
|         | pMF149               | CP-4                | MGFRFMENFNQ <sup>P</sup> YISQSIT                              | 0.59  |
|         | pMF295               | CP-4                | MGFRFMENFNQ <sup>A</sup> YISQSIT                              | 0.63  |
|         | pMF149(wt)           | CP-4                | FWR <sup>R</sup> RWHISLSTWLRDYL <sup>Y</sup> I                | 0.59  |
|         | pMF271               | CP-4                | F <sup>F</sup> RRRWHISLSTWLRDYL <sup>Y</sup> I                | 0.06  |
|         | pMF293               | CP-4                | F <sup>W</sup> RRWHISLSTWLRDYL <sup>Y</sup> I                 | 0.10  |
|         | pMF272               | CP-4                | F <sup>W</sup> RRWHISLSTWLRDYL <sup>Y</sup> I                 | 0.43  |
|         | pMF273               | CP-4                | F <sup>W</sup> RRWHISLSTWLRDYL <sup>Y</sup> I                 | 0.11  |
|         | pMF274               | CP-4                | F <sup>W</sup> RRWHISLSTWLRDYL <sup>Y</sup> I                 | 0.02  |
|         | pMF275               | CP-4                | F <sup>W</sup> RRWHISLSTWLRDYL <sup>Y</sup> I                 | 0.40  |
|         | pMF276               | CP-4                | F <sup>W</sup> RRWHISLSTWLRDYL <sup>Y</sup> I                 | 0.22  |
|         | pMF294               | CP-4                | F <sup>W</sup> RRWHISLSTWLRDYL <sup>Y</sup> I                 | 0.45  |
|         | pMF296               | CP-4                | F <sup>W</sup> RRWHISLSTWLRD <sup>E</sup> LYI                 | 0.30  |
|         | pMF297               | CP-4                | F <sup>W</sup> RRWHISLSTWLRD <sup>S</sup> LYI                 | 0.21  |

<sup>a</sup> wt, Wild type.

<sup>b</sup> That is, the domain containing the conserved motif. TM, transmembrane; PP, periplasm; CP, cytoplasm.

<sup>c</sup> The wild-type sequence is shown, with underlined amino acids showing changes made by site-directed mutagenesis.

<sup>d</sup> Units represent moles of O-acetyl groups, as determined by the assay of Franklin et al. (24) divided by the moles of uronic acid residues, as determined by the assay of Knutson and Jeanes (36). The data shown represent the averages of three to four samples.

**Periplasmic domain 3 contains conserved amino acids required for alginate O acetylation.** Periplasmic domain 3 (PP-3) has the conserved motif, 123-P[X]GIS[2X]TF in the AlgI homologs (Table 2) and F[X]GISY[X]TF in the DltB homologs. A site-directed mutation S127A reduced alginate O acetylation fourfold. Although a threonine was conserved at position T130 of the *P. aeruginosa* AlgI, throughout most of the AlgI and DltB homologs it was not observed in the other alginate-producing bacterium, *Azotobacter vinelandii* and, as predicted, a T130A mutation in the *P. aeruginosa* AlgI had no effect on alginate O acetylation.

Conservation of amino acid sequence was not observed throughout the AlgI homologs for other regions of the protein, and therefore other protein domains were not analyzed by site-directed mutagenesis here. Based on the bioinformatic and site-directed mutagenesis results, we propose a model for AlgI of *P. aeruginosa* that is shown in Fig. 7. The positions of the ordered series of motifs and key amino acid required for AlgI activity are indicated.

## DISCUSSION

Bacteria produce a variety of extracellular polysaccharides that enhance their survival in natural environments. The polysaccharides are often modified with functional groups, and these modifications affect the chemical and physical properties of the polymers and influence the ability of the bacteria to survive in their ecological niches. One example of a survival advantage imparted by polymer modification is the growth of bacteria in biofilms. Biofilms are bacteria (and other organisms) associated with surfaces, often encapsulated with extracellular polysaccharide. Two reports indicate that polysaccharide O acetylation is important for biofilm formation (50, 70). In *P. aeruginosa*, alginate O acetylation is essential for mucoid strains to produce thick three-dimensional biofilms that extend from the substratum. In *P. fluorescens*, O acetylation of cellulose is required for the bacteria to form a pellicle (biofilm at the air-water interface). In both of these cases, the O acetylation of the polymer apparently provides the structural integrity of the matrix material, allowing biofilm formation.

One mechanism for polysaccharide modification with acyl groups requires a gene cassette containing *algI* and one or more genes for type II membrane protein, such as *algJ* of *P. aeruginosa*. These cassettes appear to have evolved by LGT and in some cases have been incorporated into larger polysaccharide biosynthetic operons. In *P. aeruginosa*, *P. fluorescens*, and *P. syringae*, the *algI/algJ* cassettes were incorporated into the alginate and cellulose biosynthetic operon, where, in combination with *algF*, they modify their respective polysaccharides with O-acetyl groups. In *P. aeruginosa*, the *algI/algJ* cassette was incorporated upstream of *algF*, whereas in *P. syringae* and *P. fluorescens* the cassettes were incorporated downstream of an *algF* homologs (as shown in Fig. 1 and in references 69 and 70), yielding differing gene arrangements for these operons and providing additional evidence that *algI/algJ* were obtained by LGT.

Several lines of evidence indicate that the *algI* gene cassettes evolved by LGT. First, the phylogeny of AlgI proteins does not reflect the organism phylogeny. The AlgI proteins from individual groups of bacteria, such as the  $\gamma$ - or  $\beta$ -proteobacteria do not form unique clades but rather are scattered throughout the AlgI phylogenetic tree. These AlgI proteins from the proteobacteria often cluster with AlgI proteins from distantly related bacteria, including members of the *Clostridiaceae* and *Bacillaceae*. In addition, one AlgI homolog from a eukaryote clusters with the bacterial homologs. The phylogenetic relationship of the linked gene for the type II membrane protein also does not reflect the organism phylogeny (with the exception of the DltD homologs). However, the phylogeny of these proteins is congruent with the AlgI phylogeny, suggesting that these two (or three) gene cassettes were transferred by LGT as units. Second, although *algI* gene cassettes appears to be ancient in origin (the genes are widely diverged at the DNA level), in some cases they appear to have been incorporated into bacterial genomes recently. In several cases, the G+C content of the *algI* cassette does not reflect the G+C content of the surrounding genomic DNA. This is particularly apparent with *Desulfovibrio vulgaris*, *T. pallidum* and *H. pylori*. In *H. pylori*, the *algI* homolog (and its linked gene) was only found in 33% of *H. pylori* strains tested in a microarray study (61),

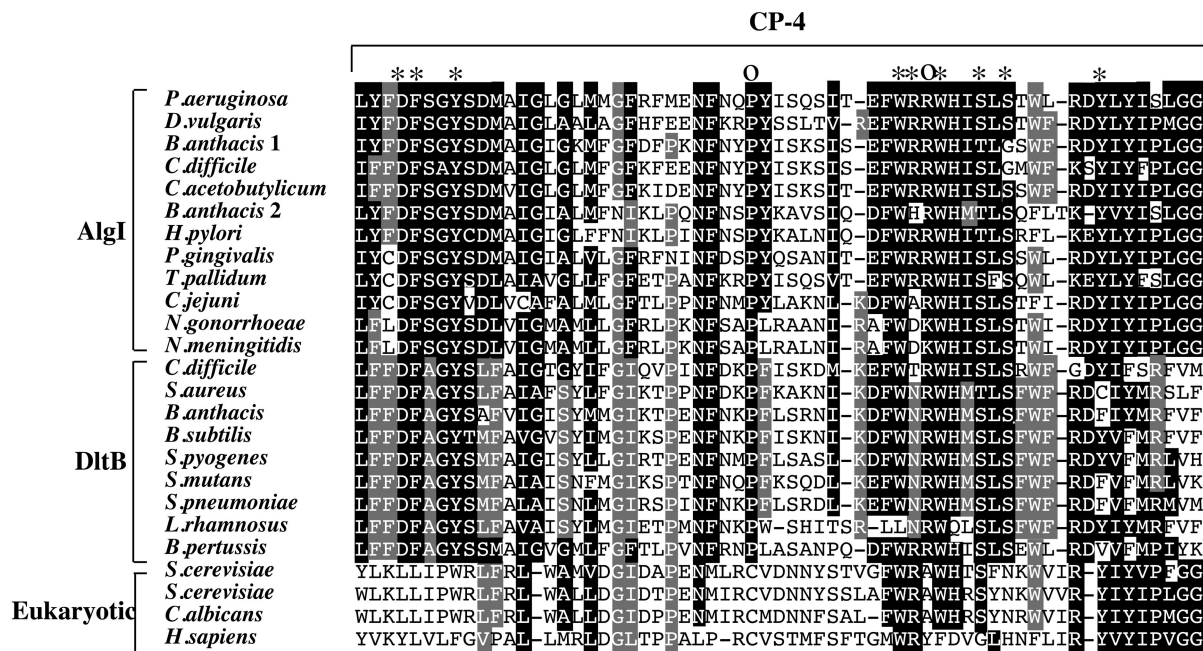


FIG. 6. Sequence alignment showing ordered motifs in the cytoplasmic domain 4 (CP-4). The symbols are similar to those in Fig. 5, with black shaded areas showing amino acids with identity through the AlgI homologs and extending into the DltB homologs, and the gray highlights showing amino acids with identity through the DltB homologs and extending into the AlgI homologs. Asterisks indicate amino acids subjected to site-directed mutagenesis, where the mutation resulted in at least a fourfold decrease in alginate O acetylation. Circles indicate amino acid change that had little effect on alginate O acetylation.

suggesting that *algI* was either obtained from an outside source in these isolate or that this gene cassette was deleted prior to divergence of the strains lacking *algI*. Third, the *algI* genes are often proximate to phage or transposase genes or gene remnants. This is apparent in *Desulfovibrio vulgaris*, *Porphyromonas gingivalis*, *Clostridium botulinum*, and *Bacteroides fragilis*. A very recent gene transfer event involving *algI* occurred in *Sinorhizobium meliloti*, where an ~6.5-kb segment of DNA, including *algI*, appears to have been duplicated between the chromosomal DNA and the pSymB megaplasmid, giving this organism two copies of *algI* and its adjacent downstream gene (7, 20, 26).

Lawrence and Roth proposed a mechanism for operon evolution in bacteria via LGT where proximate genes would have a greater chance of transfer as a unit than genes spaced widely apart (41). This would ultimately allow the assembly of genes with related function into operons and may result in a mosaic organization of genes on the chromosome (41). LGT in the assembly of operons may also lead to a mosaic pattern of genes within operons, if genes with similar functions were assembled at different times independently. This appears to be the case for the assembly of the *algI* homologs and the linked genes for the type II membrane proteins. In the case of the AlgJ group, the gene for the type II membrane protein (AlgJ homologs shown in green in Fig. 1) is always downstream of *algI*. Therefore, these genes likely assembled into a cassette, and then the cassette was transferred by LGT and/or diverged over time through speciation. In this case, the *algI/algJ* gene cassette would have a common ancestor, but the ancestry would not necessarily reflect the organism phylogeny, as is the case shown here. This would account for the topological congruency of the AlgI and AlgJ phylogenies and for the similarity of *algI* and *algJ*

G+C content within each cassette. On the other hand, the genes for the NMA1479 group (shown in red in Fig. 1) appear to have been assembled on at least two occasions independently. In one case, the genes for the type II membrane proteins lie downstream of the *algI* gene (*N. meningitidis*, *Campylobacter jejuni*, and *T. pallidum*). In the other case, the NMA1479-like genes are arranged upstream of *algI* (*Porphyromonas gingivalis* and *Bacteroides fragilis*). The differing gene arrangement of these linked genes likely represents operon assembly at a different time, followed by LGT or divergence. This LGT mechanism for evolution of the gene cassette would also account for the difference in gene arrangement seen in the *dlt* operons observed in the *Bordetella* sp. compared to the *dlt* operons of the gram-positive bacteria, which show inverse gene order compared to each other.

The gene linkages seen here helped identify motifs in AlgJ required for enzyme activity. AlgJ from *P. aeruginosa* demonstrates little sequence identity to other proteins, and few hits were identified for AlgJ by using BLAST searches. However, characterization of the genes adjacent to the *algI* homologs helped identify a set of proteins with conserved motifs to AlgJ. The results suggest that these proteins, although distantly related, are homologs to AlgJ. First, all *algJ* homologs are genetically linked to *algI*. Second, all have putative uncleaved signal peptides. Third, all have an amino acid motif surrounding a conserved histidine residue. In addition, the phylogeny of the AlgJ homologs reflects that of the AlgI homologs. Therefore, although highly diverged compared to the AlgI proteins, the AlgJ homologs appear to have a common ancestor. Sequence alignments of these homologs allowed the identification of two conserved amino acid motifs in these AlgJ ho-

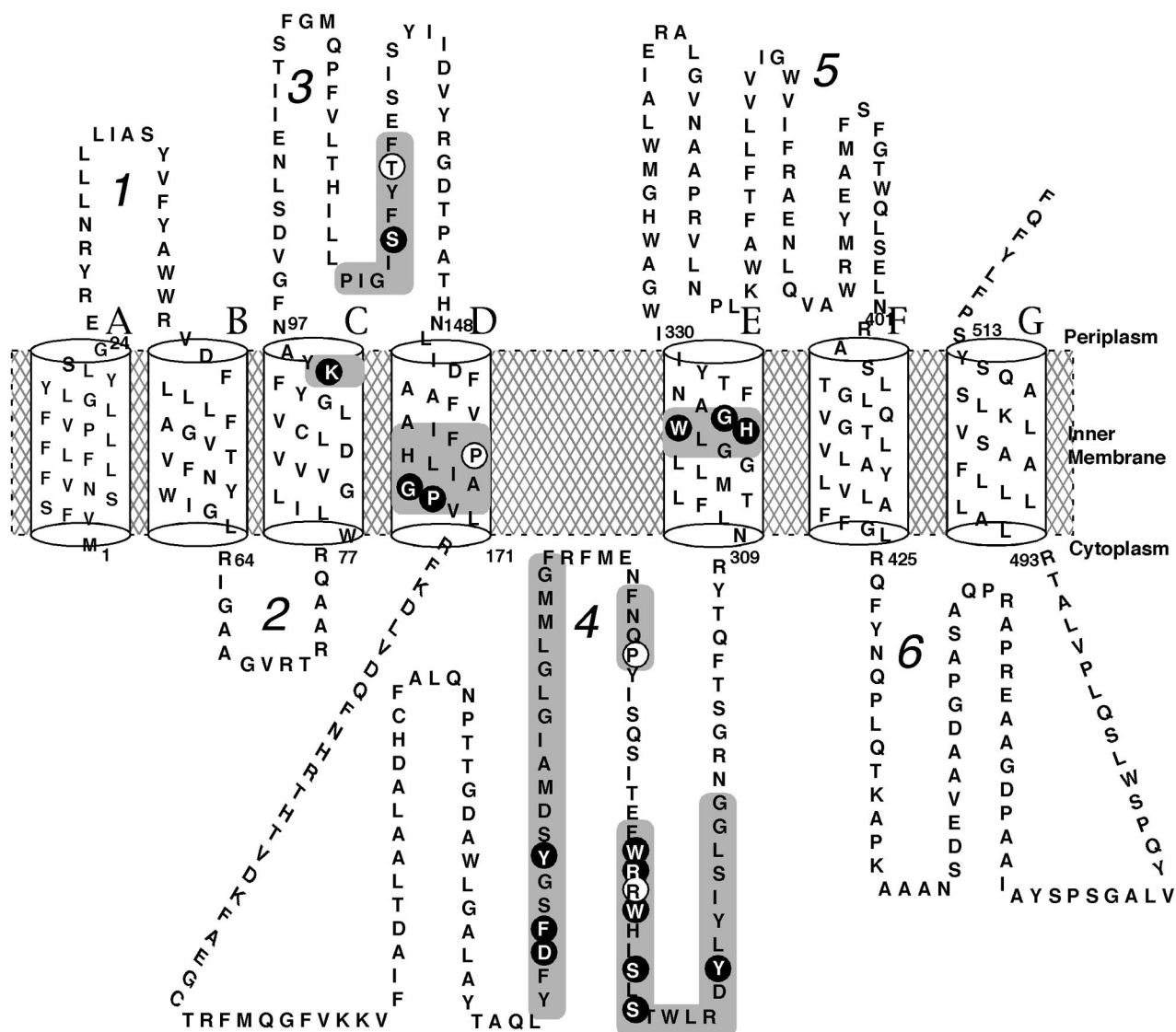


FIG. 7. Model of AlgI from *P. aeruginosa*, showing the highly conserved ordered series of motifs (shaded gray). Filled circles indicate the sites where amino acid substitutions resulted in at least a fourfold decrease in alginate O-acetylation activity. Open circles indicate conserved amino acid that when mutated do not affect alginate O acetylation.

mologs, and site-directed mutagenesis studies demonstrated that these motifs are important for enzyme activity. In particular, the conserved histidine may provide an active center for the transfer of acyl groups to the polysaccharide, since a similar motif is found in all three of the groups of type II membrane proteins: the AlgJ group, the NMA1479 group, and the DltD group.

AlgI is a member of a family of uncharacterized proteins found in a diversity of bacteria and in some eukaryotes (33). The AlgI homologs contain an ordered series of motifs that are important for the alginate O-acetylation activity in *P. aeruginosa*. In light of the divergence of the *algI* homologs at the DNA level, the peptide motifs within this family of proteins are well conserved, indicating that this series of motifs was maintained through evolution and that these motifs play roles in the functions of the AlgI homologs, as is the case shown here for AlgI of *P. aeruginosa*.

Based on our multiple alignments and on membrane topology predictions, the AlgI/DltB proteins have four TM domains that likely play a structural role (TM-A, TM-B, TM-F, and TM-G). These TM domains contain amino acids with hydrophobic side chains but little amino acid identity throughout the alignment. On the other hand, three of the TMs (TM-C, TM-D, and TM-E) have amino acid identity, including amino acids not generally associated with TM domains. For example, a charged lysine residue is present in TM-C, and proline and glycine residues, which may interrupt helical structures, are found in TM-D. Replacement of these residues with alanines reduced or abolished alginate O acetylation, indicating the importance of these amino acids in AlgI activity. In addition, replacement of conserved tryptophan, histidine, and glycine residues in the GXWHG motif of TM-E also reduced or abolished alginate O acetylation. Since alginate is likely O acetylated at the polymer level in the periplasm (24), the motifs

within these TMs are likely required for the transfer of the O-acetyl groups from their cytoplasmic precursor across the cell inner membrane to the periplasm.

Conserved motifs are also present in periplasmic domain 3 and in cytoplasmic domain 4 of the AlgI homologs. Based on the mutagenesis results, we propose the model for the activity of AlgI in *P. aeruginosa* shown in Fig. 7. Conserved motifs in cytoplasmic domain 4 are likely involved in interaction with alginate O-acetylation precursor. D-Alanyl carrier protein acts as the precursor for D-alanylation of wLTA in the gram-positive bacteria (14, 31). Since AlgI is homologous to DltB, acetyl-acyl carrier protein may be the precursor for alginate O acetylation. The conserved motifs in three of the TM domains are likely involved in transport of the O-acetyl group across the cytoplasmic membrane. Alternatively, these TMs may play a role in export of acetyl-acyl carrier protein across the membrane, as proposed for export of the D-alanyl carrier protein carrier protein across the membrane in the gram-positive bacteria (35). The conserved motif important for AlgI activity in periplasmic domain 3 may be required for transfer of the O-acetyl group to the alginate polymer or for interaction with periplasmic proteins, AlgJ, or AlgF.

Since only pseudomonads and closely related species are known to produce alginate, the AlgI homologs from these other species are likely required for functions other than alginate O acetylation. However, the additional AlgI homologs probably play similar roles in polysaccharide modification. For example, it has been demonstrated that an AlgI homolog from *P. fluorescens* is required for esterification of cellulose with O-acetyl groups (69). Cellulose synthase genes are also found adjacent to *algI* homologs in *P. syringae*, *Clostridium acetobutylicum*, and *Clostridium difficile*. Since homologs to the *bcs* genes of *Glucosacetobacter xylinus* were recently identified in *E. coli* and in *Salmonella enterica* serovar Typhimurium (78), cellulose production may be widely distributed in bacteria. Many of the other *algI* homologs are located adjacent to genes for polysaccharide biosynthesis. The *Bacillus anthracis* AlgI homologs are near the *csaA* and *csaB* genes, which are required for pyruvylation of peptidoglycan-associate polysaccharide (45). In *H. pylori* the *algI* homolog is adjacent to *rfaD* and *rfaE*, required for LPS biosynthesis, although not on the same operon. It is not yet known whether LPS is esterified in *H. pylori* strains that contain an *algI* homolog. Modification of LPS with O-acetyl groups has been observed in *P. aeruginosa* (49). A membrane protein, not related to AlgI, was shown to be involved in this modification, resulting in the conversion of the O-antigen serotype from O5 to O16. The genes for this serotype conversion, including the gene for the membrane protein, are carried on a bacteriophage, demonstrating a role for LGT in structural modification of the LPS of *P. aeruginosa*.

#### ACKNOWLEDGMENTS

We thank I. King Jordan and David Nivens for assistance with this research. We thank The Institute for Genomic Research, The Sanger Centre, OU-ACGT, and the Stanford DNA Sequencing and Technology Center for providing preliminary sequences.

This study was supported by Public Health Service grants AI-46588 (M.J.F.) and AI-28309 (M.A.M.) from the National Institute of Allergy and Infectious Diseases, by grant P20/RR-16455-01 from the National Center for Research Resources, and by a Research Career Development Award to M.A.M.

#### REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.). 1993. *Current protocols in molecular biology*, vol. 2. Greene Publishing Associates, Inc./John Wiley & Sons, Inc., New York, N.Y.
- Baltimore, R. S., and M. Mitchell. 1982. Immunologic investigations of mucoid strains of *Pseudomonas aeruginosa*: comparison of susceptibility to opsonic antibody in mucoid and nonmucoid strains. *J. Infect. Dis.* 141:238–247.
- Boyd, D. A., D. G. Cvitkovitch, A. S. Bleiweis, M. Y. Kiriukhin, D. V. Debabov, F. C. Neuhaus, and I. R. Hamilton. 2000. Defects in D-alanyl-lipoteichoic acid synthesis in *Streptococcus mutans* results in acid sensitivity. *J. Bacteriol.* 182:6055–6065.
- Brautigan, V. M., W. C. Childs III, and F. C. Neuhaus. 1981. Biosynthesis of D-alanyl-lipoteichoic acid in *Lactobacillus casei*: D-alanyl-lipophilic compounds as intermediates. *J. Bacteriol.* 146:239–250.
- Bruggemann, H., S. Baumer, W. F. Fricke, A. Weizer, H. Liesegang, I. Decker, C. Herzberg, R. Martinez-Arias, R. Merkl, A. Henne, and G. Gottschalk. 2003. The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. *Proc. Natl. Acad. Sci. USA* 100:1316–1321.
- Capela, D., F. Barloy-Hubler, J. Gouzy, G. Bothe, F. Ampe, J. Batut, P. Boistard, A. Becker, M. Boutry, E. Cadieu, S. Dreano, S. Gloux, T. Godrie, A. Goffeau, D. Kahn, E. Kiss, V. Lelaure, D. Masuy, T. Pohl, D. Portetelle, A. Puhler, B. Purnelle, U. Rampsperger, C. Renard, P. Thebault, M. Vandenberg, S. Weidner, and F. Galibert. 2001. Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proc. Natl. Acad. Sci. USA* 98:9877–9882.
- Carlton, J. M., S. V. Angiuoli, B. B. Suh, T. W. Kooij, M. Perete, J. C. Silva, M. D. Ermolaeva, J. E. Allen, J. D. Selengut, H. L. Koo, J. D. Peterson, M. Pop, D. S. Kosack, M. F. Shumway, S. L. Bidwell, S. J. Shallom, S. E. van Aken, S. B. Riedmuller, T. V. Feldblyum, J. K. Cho, J. Quackenbush, M. Sedegah, A. Shoabi, L. M. Cummings, L. Florens, J. R. Yates, J. D. Raine, R. E. Sinden, M. A. Harris, D. A. Cunningham, P. R. Preiser, L. W. Bergman, A. B. Vaidya, L. H. van Lin, C. J. Janse, A. P. Waters, H. O. Smith, O. R. White, S. L. Salzberg, J. C. Venter, C. M. Fraser, S. L. Hoffman, M. J. Gardner, and D. J. Carucci. 2002. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* 419:512–519.
- Chain, P., J. Lamerdin, F. Larimer, W. Regala, V. Lao, M. Land, L. Hauser, A. Hooper, M. Klotz, J. Norton, L. Sayavedra-Soto, D. Arciero, N. Hommes, M. Whittaker, and D. Arp. 2003. Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *J. Bacteriol.* 185:2759–2773.
- Chitnis, C. E., and D. E. Ohman. 1990. Cloning of *Pseudomonas aeruginosa algG*, which controls alginate structure. *J. Bacteriol.* 172:2894–2900.
- Clemans, D. L., P. E. Kolenbrander, D. V. Debabov, Q. Zhang, R. D. Lunsford, H. Sakone, C. J. Whittaker, M. P. Heaton, and F. C. Neuhaus. 1999. Insertional inactivation of genes responsible for the D-alanylation of lipoteichoic acid in *Streptococcus gordonii* DL1 (Challis) affects intrageneric coaggregations. *Infect. Immun.* 67:2464–2474.
- Danese, P. N., L. A. Pratt, and R. Kolter. 2000. Exopolysaccharide production is required for development of *Escherichia coli* K-12 biofilm architecture. *J. Bacteriol.* 182:3593–3596.
- Davidson, J. W., C. J. Lawson, and I. W. Sutherland. 1977. Localization of O-acetyl groups in bacterial alginate. *J. Gen. Microbiol.* 98:603–606.
- Debabov, D. V., M. P. Heaton, Q. Zhang, K. D. Stewart, R. H. Lambalot, and F. C. Neuhaus. 1996. The D-alanyl carrier protein in *Lactobacillus casei*: cloning, sequencing, and expression of *dltC*. *J. Bacteriol.* 178:3869–3876.
- Debabov, D. V., M. Y. Kiriukhin, and F. C. Neuhaus. 2000. Biosynthesis of lipoteichoic acid in *Lactobacillus rhamnosus*: role of DltD in D-alanylation. *J. Bacteriol.* 182:2855–2864.
- Evans, L. R., and A. Linker. 1973. Production and characterization of the slime polysaccharide of *Pseudomonas aeruginosa*. *J. Bacteriol.* 116:915–924.
- Felsenstein, J. 1992. Estimating effective population size from samples of sequences: a bootstrap Monte Carlo integration method. *Genet. Res.* 60:209–220.
- Ferretti, J. J., W. M. McShan, D. Ajdic, D. J. Savić, G. Savić, K. Lyon, C. Primeaux, S. Sezate, A. N. Suvorov, S. Kenton, H. S. Lai, S. P. Lin, Y. Qian, H. G. Jia, F. Z. Najjar, Q. Ren, H. Zhu, L. Song, J. White, X. Yuan, S. W. Clifton, B. A. Roe, and R. McLaughlin. 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc. Natl. Acad. Sci. USA* 98:4658–4663.
- Figurski, D., and D. R. Helinski. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc. Natl. Acad. Sci. USA* 76:1648–1652.
- Finan, T. M., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F. J. Vorholter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding, and A. Puhler. 2001. The complete sequence of the 1,683-kb pSymB megaplasmid

- from the N2-fixing endosymbiont *Sinorhizobium meliloti*. Proc. Natl. Acad. Sci. USA 98:9889–9894.
21. Franklin, M. J., C. E. Chitnis, P. Gacesa, A. Sonesson, D. C. White, and D. E. Ohman. 1994. *Pseudomonas aeruginosa* AlgG is a polymer level alginate C5-mannuronan epimerase. J. Bacteriol. 176:1821–1830.
  22. Franklin, M. J., and D. E. Ohman. 1993. Identification of *algF* in the alginate biosynthetic gene cluster of *Pseudomonas aeruginosa* which is required for alginate acetylation. J. Bacteriol. 175:5057–5065.
  23. Franklin, M. J., and D. E. Ohman. 1996. Identification of *algI* and *algJ* in the *Pseudomonas aeruginosa* alginate biosynthetic gene cluster which are required for alginate O acetylation. J. Bacteriol. 178:2186–2195.
  24. Franklin, M. J., and D. E. Ohman. 2002. Mutant analysis and cellular localization of the AlgI, AlgJ, and AlgF proteins required for O acetylation of alginate in *Pseudomonas aeruginosa*. J. Bacteriol. 184:3000–3007.
  25. Fraser, C. M., S. J. Norris, G. M. Weinstock, O. White, G. G. Sutton, R. Dodson, M. Gwinn, E. K. Hickey, R. Clayton, K. A. Ketchum, E. Sodergren, J. M. Hardham, M. P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J. K. Howell, M. Chidambaram, T. Utterback, L. McDonald, P. Artiach, C. Bowman, M. D. Cotton, J. C. Venter, et al. 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science 281:375–388.
  26. Galibert, F., T. M. Finan, S. R. Long, A. Puhler, P. Abola, F. Ampe, F. Barloy-Hubler, M. J. Barnett, A. Becker, P. Boistard, G. Bothe, M. Boutry, L. Bowser, J. Buhrmester, E. Cadieu, D. Capela, P. Chain, A. Cowie, R. W. Davis, S. Dreano, N. A. Federspiel, R. F. Fisher, S. Gloux, T. Godrie, A. Goffeau, B. Golding, J. Gouzy, M. Gurjal, I. Hernandez-Lucas, A. Hong, L. Huizar, R. W. Hyman, T. Jones, D. Kahn, M. L. Kahn, S. Kalman, D. H. Keating, E. Kiss, C. Komp, V. Lelaure, D. Masuy, C. Palm, M. C. Peck, T. M. Pohl, D. Portetelle, B. Purnelle, U. Ramsperger, R. Surzycki, P. Thebault, M. Vandebol, F. J. Vorholter, S. Weidner, D. H. Wells, K. Wong, K. C. Yeh, and J. Batut. 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293:668–672.
  27. Glaser, P., F. Kunst, M.-P. Arnaud, M.-P. Coudart, W. Gonzales, M.-F. Hullo, M. Ionescu, B. Lubochinsky, L. Marcelino, I. Moszer, E. Presecan, M. Santana, E. Schneider, J. Schweizer, A. Vertes, G. Rapoport, and A. Danchin. 1993. *Bacillus subtilis* genome project: cloning and sequencing of the 97-kb region from 325° to 333°. Mol. Microbiol. 10:371–384.
  28. Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. 1996. Life with 6000 genes. Science 274: 546:563–567.
  29. Haug, A., and B. Larsen. 1969. Biosynthesis of alginate: epimerisation of D-mannuronic to L-guluronic acid residues in the polymer chain. Biochim. Biophys. Acta 192:557–559.
  30. Heaton, M. P., and F. C. Neuhaus. 1992. Biosynthesis of D-alanyl-lipoteichoic acid: cloning, nucleotide sequence, and expression of the *Lactobacillus casei* gene for the D-alanine-activating enzyme. J. Bacteriol. 174:4707–4717.
  31. Heaton, M. P., and F. C. Neuhaus. 1994. Role of the D-alanyl carrier protein in the biosynthesis of D-alanyl-lipoteichoic acid. J. Bacteriol. 176:681–690.
  32. Higgins, D. G., and P. M. Sharp. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244.
  33. Hofmann, K. 2000. A superfamily of membrane-bound O-acyltransferases with implications for *wnt* signaling. Trends Biochem. Sci. 25:111–112.
  34. Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson. 1998. Multiple sequence alignment with CLUSTAL X. Trends Biochem. Sci. 23:403–405.
  35. Kiriukhin, M. Y., and F. C. Neuhaus. 2001. D-alanylation of lipoteichoic acid: role of the D-alanyl carrier protein in acylation. J. Bacteriol. 183:2051–2058.
  36. Knutson, C. A., and A. Jeanes. 1968. A new modification of the carboxyl analysis: application to heteropolysaccharides. Anal. Biochem. 24:470–481.
  37. Krogh, A., B. Larsson, G. von Heijne, and E. L. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305:567–580.
  38. Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi, and K. Hiramatsu. 2001. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. Lancet 357:1225–1240.
  39. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.
  40. Lawrence, J. G., and H. Ochman. 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. 44:383–397.
  41. Lawrence, J. G., and J. R. Roth. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 143:1843–1860.
  42. Leigh, J. A., J. W. Reed, J. F. Hanks, A. M. Hirsch, and G. C. Walker. 1987. *Rhizobium meliloti* mutants that fail to succinylate their calcofluor-binding exopolysaccharide are defective in nodule invasion. Cell 51:579–587.
  43. Leigh, J. A., E. R. Signer, and G. C. Walker. 1985. Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. Proc. Natl. Acad. Sci. USA 82:6231–6235.
  44. McClure, M. A. 1991. Evolution of retroposons by acquisition or deletion of retrovirus-like genes. Mol. Biol. Evol. 8:835–856.
  45. Mesnage, S., T. Fontaine, T. Mignot, M. Delepierre, M. Mock, and F. A. 2000. Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO J. 19:4473–4484.
  46. Nakai, T., A. Moriya, N. Tonouchi, T. Tsuchida, F. Yoshinaga, S. Horinouchi, Y. Sone, H. Mori, F. Sakai, and T. Hayashi. 1998. Control of expression by the cellulose synthase (*bcsA*) promoter region from *Acetobacter xylinum* BPR 2001. Gene 213:93–100.
  47. Nakamura, Y., S. Sato, E. Asamizu, T. Kaneko, H. Kotani, N. Miyajima, and S. Tabata. 1998. Structural analysis of *Arabidopsis thaliana* chromosome 5. VII. Sequence features of the regions of 1,013,767 bp covered by sixteen physically assigned P1 and TAC clones. DNA Res. 5:297–308.
  48. Nelson, K. E., R. D. Fleischmann, R. T. DeBoy, I. T. Paulsen, D. E. Fouts, J. A. Eisen, S. C. Daugherty, R. J. Dodson, A. S. Durkin, M. Gwinn, D. H. Haft, J. F. Kolonay, W. C. Nelson, T. Mason, L. Tallon, J. Gray, D. Granger, H. Tettelin, H. Dong, J. L. Galvin, M. J. Duncan, F. E. Dewhirst, and C. M. Fraser. 2003. Complete genome sequence of the oral pathogenic bacterium *Porphyromonas gingivalis* strain W83. J. Bacteriol. 185:5591–5601.
  49. Newton, G. J., C. Daniels, L. L. Burrows, A. M. Kropinski, A. J. Clarke, and J. S. Lam. 2001. Three-component-mediated serotype conversion in *Pseudomonas aeruginosa* by bacteriophage D3. Mol. Microbiol. 39:1237–1247.
  50. Nivens, D. E., D. E. Ohman, J. Williams, and M. J. Franklin. 2001. Role of alginate and its O acetylation in formation of *Pseudomonas aeruginosa* microcolonies and biofilms. J. Bacteriol. 183:1047–1057.
  51. Nolling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, and D. R. Smith. 2001. Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*. J. Bacteriol. 183:4823–4838.
  52. Parkhill, J., M. Achtman, K. D. James, S. D. Bentley, C. Churcher, S. R. Klee, G. Morelli, D. Basham, D. Brown, T. Chillingworth, R. M. Davies, P. Davis, K. Devlin, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, S. Leather, S. Moule, K. Mungall, M. A. Quail, M. A. Rajandream, K. M. Rutherford, M. Simmonds, J. Skelton, S. Whitehead, B. G. Spratt, and B. G. Barrell. 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature 404:502–506.
  53. Parkhill, J., M. Sebahia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabinowitz, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, and D. J. Maskell. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis*, and *Bordetella bronchiseptica*. Nat. Genet. 35:32–40.
  54. Parkhill, J., B. W. Wren, K. Mungall, J. M. Ketley, C. Churcher, D. Basham, T. Chillingworth, R. M. Davies, T. Feltwell, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Moule, M. J. Pallen, C. W. Penn, M. A. Quail, M. A. Rajandream, K. M. Rutherford, A. H. van Vliet, S. Whitehead, and B. G. Barrell. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature 403:665–668.
  55. Perego, M., P. Glaser, A. Minutello, M. A. Strauch, K. Leopold, and W. Fischer. 1995. Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in *Bacillus subtilis*. Identification of genes and regulation. J. Biol. Chem. 270:15598–15606.
  56. Peschel, A., M. Otto, R. W. Jack, H. Kalbacher, G. Jung, and F. Gotz. 1999. Inactivation of the *dlt* operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. J. Biol. Chem. 274: 8405–8410.
  57. Pier, G. B., F. Coleman, M. Grout, M. Franklin, and D. E. Ohman. 2001. Role of alginate O acetylation in resistance of mucoid *Pseudomonas aeruginosa* to opsonic phagocytosis. Infect. Immun. 69:1895–1901.
  58. Ponting, C. P., J. Schultz, F. Milpetz, and P. Bork. 1999. SMART: identification and annotation of domains from signalling and extracellular protein sequences. Nucleic Acids Res. 27:229–232.
  59. Pugsley, A. P. 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol. Rev. 57:50–108.
  60. Read, T. D., S. N. Peterson, N. Tourasse, L. W. Baillie, I. T. Paulsen, K. E. Nelson, H. Tettelin, D. E. Fouts, J. A. Eisen, S. R. Gill, E. K. Holtzapple, O. A. Okstad, E. Helgason, J. Rilstone, M. Wu, J. F. Kolonay, M. J. Beanan, R. J. Dodson, L. M. Brinkac, M. Gwinn, R. T. DeBoy, R. Madpu, S. C. Daugherty, A. S. Durkin, D. H. Haft, W. C. Nelson, J. D. Peterson, M. Pop, H. M. Khouri, D. Radune, J. L. Benton, Y. Mahamoud, L. Jiang, I. R. Hance, J. F. Weidman, K. J. Berry, R. D. Plaut, A. M. Wolf, K. L. Watkins, W. C.

- Nierman, A. Hazen, R. Cline, C. Redmond, J. E. Thwaite, O. White, S. L. Salzberg, B. Thomason, A. M. Friedlander, T. M. Koehler, P. C. Hanna, A. B. Kolsto, and C. M. Fraser. 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **423**:81–86.
61. Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow. 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc. Natl. Acad. Sci. USA* **97**:14668–14673.
62. Schultz, J., F. Milpetz, P. Bork, and C. P. Ponting. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **95**:5857–5864.
63. Schwarzmann, S., and J. R. Boring III. 1971. Antiphagocytic effect of slime from a mucoid strain of *Pseudomonas aeruginosa*. *Infect. Immun.* **3**:762–767.
64. Shimizu, T., K. Ohtani, H. Hirakawa, K. Ohshima, A. Yamashita, T. Shiba, N. Ogasawara, M. Hattori, S. Kuhara, and H. Hayashi. 2002. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc. Natl. Acad. Sci. USA* **99**:996–1001.
65. Simpson, J. A., S. E. Smith, and R. T. Dean. 1988. Alginate inhibition of the uptake of *Pseudomonas aeruginosa* by macrophages. *J. Gen. Microbiol.* **134**:29–36.
66. Skjåk-Bræk, G., H. Grasdalen, and B. Larsen. 1986. Monomer sequence and acetylation pattern in some bacterial alginates. *Carbohydr. Res.* **154**:239–250.
67. Skjåk-Bræk, G., F. Zanetti, and S. Paoletti. 1989. Effect of acetylation on some solution and gelling properties of alginates. *Carbohydr. Res.* **185**:131–138.
68. Slauch, J. M., M. J. Mahan, P. Michetti, M. R. Neutra, and J. J. Mekalanos. 1995. Acetylation (O-factor 5) affects the structural and immunological properties of *Salmonella typhimurium* lipopolysaccharide O antigen. *Infect. Immun.* **63**:437–441.
69. Spiers, A., S. Kahn, J. Bohannon, M. Travisano, and P. Rainey. 2002. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* **163**:33–46.
70. Spiers, A. J., J. Bohannon, S. M. Gehrig, and P. B. Rainey. 2003. Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. *Mol. Microbiol.* **50**:15–27.
71. Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrook-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, and I. T. Paulsen. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**:959–964.
72. Suerbaum, S., C. Josenhans, T. Sterzenbach, B. Drescher, P. Brandt, M. Bell, M. Droge, B. Fartmann, H. P. Fischer, Z. Ge, A. Horster, R. Holland, K. Klein, J. Konig, L. Macko, G. L. Mendz, G. Nyakatura, D. B. Schauer, Z. Shen, J. Weber, M. Frosch, and J. G. Fox. 2003. The complete genome sequence of the carcinogenic bacterium *Helicobacter hepaticus*. *Proc. Natl. Acad. Sci. USA* **100**:7901–7906.
73. Tatusova, T. A., and T. L. Madden. 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**:247–250.
74. Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**:498–506.
75. Tettelin, H., N. J. Saunders, J. Heidelberg, A. C. Jeffries, K. E. Nelson, J. A. Eisen, K. A. Ketchum, D. W. Hood, J. F. Peden, R. J. Dodson, W. C. Nelson, M. L. Gwinn, R. DeBoy, J. D. Peterson, E. K. Hickey, D. H. Haft, S. L. Salzberg, O. White, R. D. Fleischmann, B. A. Dougherty, T. Mason, A. Ciecko, D. S. Parksey, E. Blair, H. Citti, E. B. Clark, M. D. Cotton, T. R. Utterback, H. Khouri, H. Qin, J. Vamathevan, J. Gill, V. Scarlato, V. Masignani, M. Pizza, G. Grandi, L. Sun, H. O. Smith, C. M. Fraser, E. R. Moxon, R. Rappuoli, and J. C. Venter. 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**:1809–1815.
76. Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, J. C. Venter, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**:539–547.
77. Vazquez, A., S. Moreno, J. Guzman, A. Alvarado, and G. Espin. 1999. Transcriptional organization of the *Azotobacter vinelandii* *algX* genes: characterization of *algF* mutants. *Gene* **232**:217–222.
78. Zogaj, X., M. Nitz, M. Rohde, W. Bokranz, and U. Romling. 2001. The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. *Mol. Microbiol.* **39**:1452–1463.