Microbiology (1997), 143, 3071–3084 Printed in Great Britain

REVIEW ARTICLE

The bacterial flagellin gene as a biomarker for detection, population genetics and epidemiological analysis

Craig Winstanley[1]† and J. Alun W. Morgan[2]

Author for correspondence: Craig Winstanley. Tel: +44 1274 383561. Fax: +44 1274 309742. e-mail: C.Winstanley@Bradford.ac.uk

[1] Biosciences Group, School of Natural and Environmental Sciences, Coventry University, Priory Street, Coventry CV1 5FB, UK

[2] Horticulture Research International, Wellesbourne, Warwickshire CV35 9EF, UK

Keywords: flagellin, detection, population genetics, epidemiology, antigenic variation

Overview

Since the development of the necessary molecular biological tools, analysis of nucleic acid biomarkers for the detection and identification of bacteria has become widespread. rRNA has often been the target of choice for such studies, and consequently there is an enormous amount of sequence data available for rRNA molecules. In addition to providing species-specific markers, rRNA sequence comparisons have been used to reassess phylogenetic relationships (Olsen *et al.*, 1994). However, rRNA molecules, even the variable regions, are often indistinguishable when closely related species or strains of the same species are compared. This means that rRNA sequence analysis is of limited use when estimation of genetic variation amongst closely related bacterial populations is required, for studies of population genetics, for epidemiological analysis, or for the development of strain-specific detection methods.

Proteins located on the surface of a bacterial cell show a much greater rate of divergence in amino acid sequence than those located internally (Whittam, 1995). The genes encoding such proteins are therefore ideal candidates as potential biomarkers to assess intraspecies genetic variation. Increasingly, flagellin genes are being targeted for such approaches. This review aims to summarize the nature of flagellin gene and protein variation, and the exploitation of such variability for detection, population genetics and epidemiological analysis.

†**Present address:** Department of Biomedical Sciences, University of Bradford, Bradford, West Yorkshire BD7 1DP, UK.

Flagellar arrangement and filament composition

The flagellum is the organelle responsible for motility in the majority of bacterial species. Flagellar activity is coupled to a sensory apparatus in a control system that allows movement of a cell towards attractive environments and away from repellent ones. The basic structure of a bacterial flagellum can be sub-divided into three parts: (i) the basal body, which consists of a series of rings and a central rod, is anchored in the inner and outer membranes of the cell and provides the motor for the chemotactic apparatus; (ii) the hook is located external to the cell and provides the link between the basal body and the filament; and (iii) the flagellar filament, the largest portion of the flagellum, consists of repeating sub-units of the protein flagellin in a helical arrangement and often extends many times the length of the cell. Rotation of this filament (clockwise or counterclockwise), which is controlled by the motor, ultimately determines the direction of swimming of the cell. There are a number of excellent reviews concerning flagellar structure, assembly, energetics and genetics (Macnab, 1992, 1996; Khan, 1993; Aizawa, 1996).

The arrangement of flagella on the surface of bacteria varies greatly between organisms. There can be single polar, tufted polar (two to six flagella) or many different peritrichous (or lateral) configurations. Some flagella, for example the endoflagella of the spirochaetes, are actually found in the periplasmic space (Limberger *et al.*, 1992). Azospirilla, a group of diazotrophic soil bacteria with the potential to increase the yield of economically important cereals and grasses, can display a single polar flagellum and several lateral flagella (Moens *et al.*, 1995a). The filaments of the lateral flagella (encoded by genes designated *laf* and produced during growth on solid media) are thinner and have a shorter wavelength than the filament of the polar

Table 1. Examples of eubacterial flagellin genes for which complete gene and deduced protein sequences are available

Species	GenBank accession number*	Number of residues†
Agrobacterium tumefaciens (*flaA*)	X80701	306
Azospirillum brasilense (*laf1*) (lateral flagellum)	U26679	414
Aquifex pyrophilus	U17575	231
Bacillus subtilis‡	X56049	303
Bacillus thuringiensis (*flaB*)‡	X67139	280
Bartonella bacilliformis	L20677	375
Bordetella bronchiseptica	L13034	391
Borrelia burgdorferi	X16833	336
Burkholderia cepacia	AF011370	384
Campylobacter jejuni	J05635	576
Caulobacter crescentus	J01556	274
Escherichia coli	M14358	498
Helicobacter pylori (*flaA*)	X60746	510
Legionella pneumophila	X83232	475
Listeria monocytogenes	X65624	287
Proteus mirabilis (*fliC1*)	L07270	365
Pseudomonas aeruginosa	M57501	394
Rhizobium meliloti (*flaA*)	M24526	395
Roseburia cecicola	M20983	293
Salmonella typhimurium	D13689	495
Serpulina hyodysenteriae	X63513	285
Serratia marcescens	M27219	351
Shigella boydii (cryptic flagellin gene)	D26165	NA
Treponema phagedenis	M94015	286
Vibrio parahaemolyticus (*flaA*) (polar flagellum)	U12816	377
Wolinella succinogenes	M82917	518
Yersinia enterocolitica	L33467	358

NA, Not applicable; flagellin genes are cryptic. A hypothetical predicted amino acid sequence was included in the analysis presented in Figs 2 and 3.

* These sequences were used in the construction of the predicted amino acid sequence alignments presented in Figs 2 and 3.

† The number of amino acid residues is predicted from the nucleotide gene sequence. In a number of cases the N-terminal methionine is cleaved in the mature protein.

‡ Two representatives of the genus *Bacillus* were included because of the considerable divergence observed in their flagellin gene sequences.

flagellum (encoded by genes designated *fla* and produced during growth in liquid media). A similar mixed flagellation and surface-dependent induction of lateral flagella has also been reported for the marine bacterium and human pathogen, *Vibrio parahaemolyticus* (McCarter, 1995). A single polar flagellum propels the bacterium in liquid environments. Under conditions in which the polar system (Fla) is not suitable for movement, such as on surfaces or in highly viscous environments, synthesis of the second flagellar system (Laf) is induced. Unlike the polar flagellum, the multiple peritrichously arranged lateral flagella are unsheathed. Under certain circumstances, *V. parahaemolyticus* can simultaneously assemble the two distinct types of flagella. Whilst the polar flagellum is synthesized constitutively, lateral flagella are produced only in response to conditions in which the polar flagellum is unable to function, such as on surfaces (McCarter, 1995). Bacterial swarming, whereby vegetative bacteria differentiate into hyperflagellated, filamentous swarm cells capable of coordinated and rapid population migration, has been observed in a number of Gram-negative bacteria, notably *Proteus mirabilis* where swarming facilitates the pathogenicity of the organism (Allison *et al.*, 1992). Swarming motility is the subject of reviews by Allison & Hughes (1991), Harshey (1994) and Belas (1996).

The composition of the flagellar filament can also vary. In many bacteria there are simple flagella with filaments composed of repeats of a single flagellin subunit. There are, however, a number of bacteria with flagella polymerized from multiple flagellins. These include

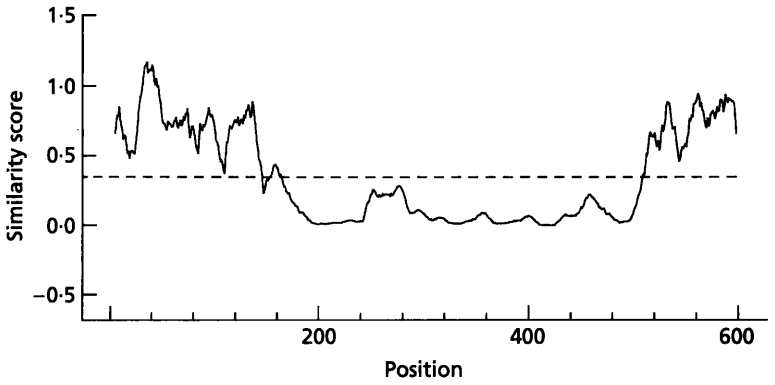

Fig. 1. Plot of flagellin sequence similarity against position on the protein. The plot was obtained by applying the program PlotSimilarity (Genetics Computer Group, University of Wisconsin, USA) to compare the 27 complete deduced protein sequences indicated in Table 1.

Caulobacter spp. (Driks *et al.*, 1989; Minnich *et al.*, 1988), *Rhizobium* spp. (Pleier & Schmitt, 1991), *Vibrio anguillarum* (McGee *et al.*, 1996), *Bacillus thuringiensis* subsp. *alesti* (Lövgren *et al.*, 1993) and many members of the domain Archaea (Jarrell *et al.*, 1996). In *Helicobacter* spp. (Josenhans *et al.*, 1995) and *Campylobacter* spp. (Guerry *et al.*, 1991), two distinct flagellin subunits (FlaA and FlaB) are required for full motility. In both cases the *flaA* gene, controlled by a σ^{28} promoter similar to promoters controlling flagellin genes in other bacteria, is expressed at higher levels than *flaB*, which is controlled by a more specialized σ^{54} promoter.

The periplasmic flagellar filaments of many spirochaetes contain between two and six multiple proteins. The structure of the periplasmic flagella (PFs) of spirochaetes is distinct from that of the flagella of other motile bacteria. The PFs are composed of multiple protein species referred to as FlaA and FlaB proteins. FlaA proteins, which are unlike other flagellins in sequence, form the PF sheath, whereas FlaB proteins, which are homologous to other bacterial flagellins in sequence, comprise the core. Within a given species there are one or two different FlaA proteins and three or four different FlaB proteins (Norris *et al.*, 1993). In *Treponema pallidum*, it has been shown that *flaA* gene sequences encode a hydrophobic signal peptide that is cleaved off in the mature protein. FlaB proteins, like other eubacterial flagellins, lack a signal peptide sequence and are secreted with an intact N-terminus (Norris *et al.*, 1993). It is believed that FlaA proteins are transported across the cytoplasmic membrane by standard mechanisms for polypeptides, whereas FlaB proteins are transported through the core of the periplasmic flagella and added to the end, as has been described for other bacteria (Namba *et al.*, 1989). The *flaA* promoter is typical of those recognized by σ^{70} transcriptional factors rather than the σ^{28} promoters associated with *flaB* and motility genes from other bacteria. It has been reported that the flagella of *Borrelia* spp. differ from those of other spirochaetes in that they are comprised of a single flagellin protein (FlaB) and are not surrounded by an outer sheath layer (Charon *et al.*, 1992). The deduced promoter sequences of *Borrelia* flagellin genes do not resemble either σ^{28} or σ^{70} promoter regions, indicating that expression of the *fla* genes of *Borrelia* spp. is controlled by an alternative σ factor (Noppa *et al.*, 1995). Recent evidence suggests that proteins homologous to FlaA are also present and expressed in *Borrelia burgdorferi* (Ge & Charon, 1997). It is not clear why such proteins have not been found associated with purified periplasmic flagella in *Borrelia* spp.

Caulobacter crescentus has one of the most widely studied flagellar systems because of the morphological events involved in the life cycle of the organism, which includes both swarmer cells (with flagella) and stalked cells (without flagella). There are three related flagellin proteins in the *Caul. crescentus* filament (25, 27·5 and 29 kDa in size), none of which are absolutely essential for motility, although their assembly may be required for normal flagellar function (Minnich *et al.*, 1988). The fully functional filament is comprised of four regions, each varying in the relative contribution of the component flagellins (Driks *et al.*, 1989; Minnich *et al.*, 1988). In *Caul. crescentus*, some flagellar filaments are surrounded by a set of three windings that form a sheath. It has been suggested that this sheath is composed of protofilaments of flagellin wound around the filament (Trachtenberg & DeRosier, 1992). Flagellar sheaths resembling loosely associated periflagellar membranes are found in *Bdellovibrio bacteriovorus* (Tomashow & Rittenberg, 1985), various *Vibrio* spp. (Sjoblad *et al.*, 1983) and *Helicobacter pylori* (Geis *et al.*, 1989), where a sheath protein has been identified (Luke & Penn, 1995). In *Vibrio cholerae* (Fuerst & Perry, 1988) and *H. pylori* (Doig & Trust, 1994), there is evidence from the study of lipopolysaccharide antigens indicating that outer membrane components are present in the flagellar sheath. It should be noted that these sheaths are different from those of spirochaete flagella, which are an integral part of the flagellum.

Archael flagellins have been found to differ in a number of respects from other bacterial flagellins. The primary sequence of archaeal flagellins shares no homology with eubacterial flagellins. In fact, archaeal flagellin gene sequences have revealed greater similarity to type IV pilins than to bacterial flagellins (Jarrell *et al.*, 1996). There are also notable differences in flagellar assembly. In archaeobacteria, proflagellins with attached leader

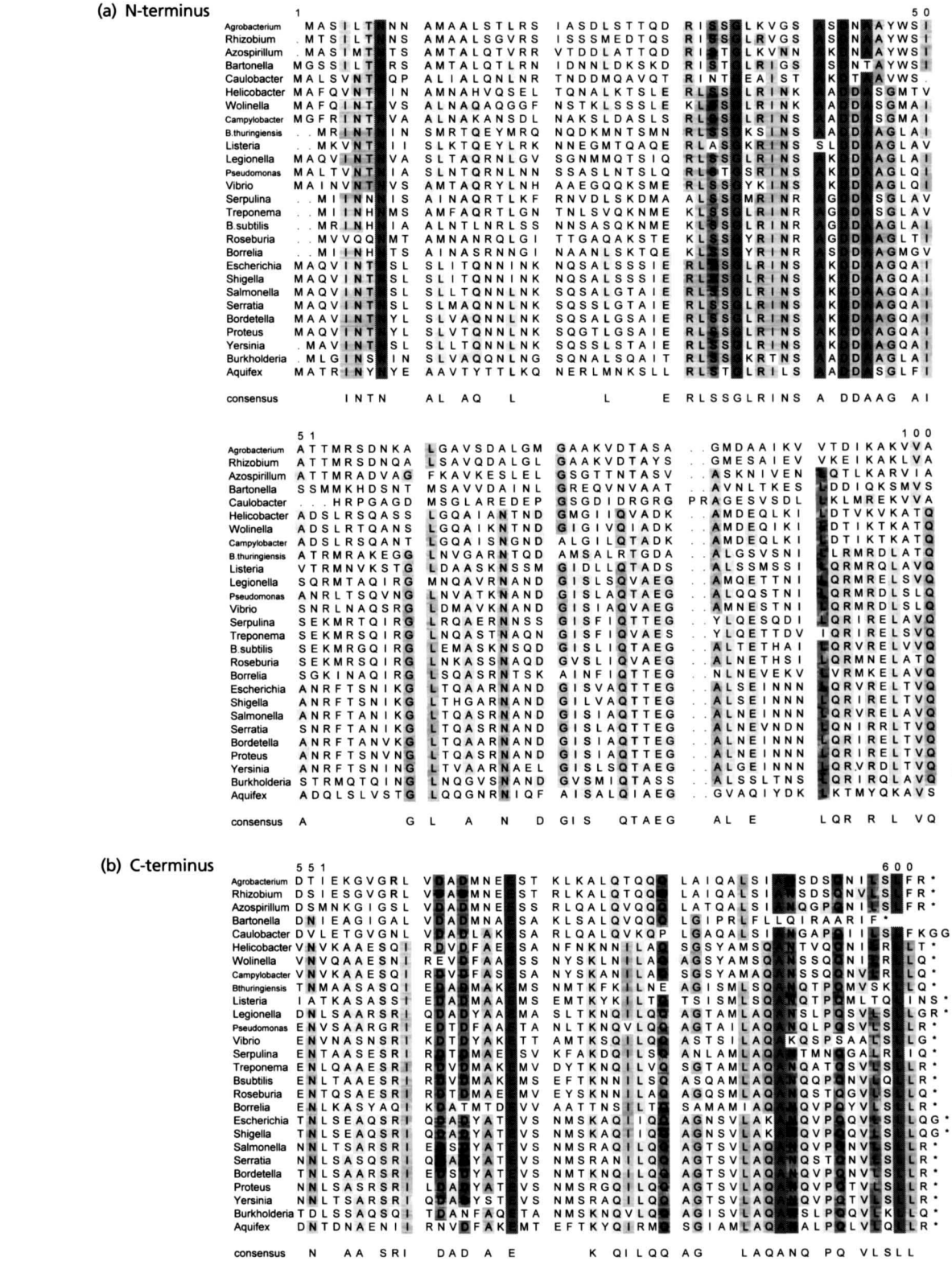

Fig. 2. Alignment of N-terminal and C-terminal conserved regions of bacterial flagellin genes. Alignment of 27 complete deduced flagellin protein sequences representing 26 different genera was performed using the program PILEUP (Genetics Computer Group, University of Wisconsin, USA). The sequences used in the analysis are indicated in Table 1. N-terminal (a) and C-terminal (b) conserved regions are shown. Positions at which 15 or more residues are conserved are indicated by progressively darker shading. The four different levels of shading represent 15–19, 20–23, 24–25 and 26–27 conserved residues, respectively.

peptides are transported to the cytoplasmic membrane in the vicinity of the polar cap structure where the proflagellin is cleaved prior to transport of the flagellin across the cytoplasmic membrane and incorporation of flagellin subunits at the base of the growing filament. This contrasts with eubacterial flagellin assembly which occurs at the distal tip of the filament following transport up through the hollow growing flagellum, and does not require the chaperone assistance of leader peptides. Further details concerning the unusual properties of archaeal flagella can be found in the review of Jarrell *et al.* (1996).

Flagellin gene and protein structure

The flagellin gene sequence database is large and expanding. Table 1 demonstrates the diversity of bacterial genera from which the flagellin gene and deduced protein sequences have been obtained. Flagellins have a distinctive domainal structure (Fig. 1), comprising conserved N- and C-terminal regions, and a central domain that may vary considerably in both amino acid sequence and size (Joys, 1988; Wilson & Beveridge, 1993). Wei & Joys (1985) reported the observation that the terminal regions of *Salmonella* flagellin genes are far more highly conserved than required to maintain the integrity of the flagellin amino acid sequence, suggesting a regulatory role for terminal nucleotide sequences. The central domain appears far less constrained and is widely understood to be responsible for flagellar antigenic variability. Homma *et al.* (1987) proposed the hairpin model, suggesting that flagellin monomers fold into a hairpin-like conformation, with the conserved N- and C-terminal domains prevalent on the inside and responsible for defining the basic filament structure, whilst the central, variable domain is exposed on the surface. Data derived from mutation and X-ray diffraction studies give credence to this type of model (Wilson & Beveridge, 1993). There is certainly ample evidence that the central domain can be altered radically without any adverse effects on the assembly of the flagellin into a filament. Newton *et al.* (1989) were able to insert a cholera toxin epitope into the flagellin of *Salmonella typhimurium*. The modified flagellin was assembled into flagella and the cells remained motile. Lu *et al.* (1995) reported the development of a system for studying protein–protein interactions involving the use of the bacterial flagellum to display random peptide libraries on the surface of *Escherichia coli*. A dispensable region in the central domain of the flagellin gene (*fliC*) was replaced with the entire coding sequence for *E. coli* thioredoxin (*trxA*). The resulting protein was efficiently exported and assembled into partially functional flagella. Random peptide libraries could then be displayed by insertion into a particular region of the thioredoxin. Kuwajima (1988) reported that, for *E. coli* K-12, the minimum size for a functional flagellin comprises 193 N-terminal and 117 C-terminal residues, of the normal complement of 497 residues.

There is some evidence that the flagellin central region may itself have structural constraints. *Wolinella succinogenes* possesses a single polar flagellum which contains in the central antigenic domain two conserved regions also present in *Campylobacter* spp. and *H. pylori* (Schuster *et al.*, 1994). There has also been a report identifying sequence motifs in the centre of *Pseudomonas putida* strain PaW8 flagellin which exhibit significant sequence homology to flagellins from *Campylobacter coli* and *Campylobacter jejuni* (Jank *et al.*, 1995). This homology was identified in a region of *Ps. putida* PaW8 flagellin not represented in the flagellin of another *Ps. putida* strain, PRS2000. At 81 kDa, the estimated molecular mass of *Ps. putida* PaW8 flagellin exceeds that of *Ps. putida* PRS2000 flagellin by 31 kDa (Winstanley *et al.*, 1994). Such findings may well be atypical, and restricted only to certain organisms, but it is possible that further analysis may reveal general structural requirements for the flagellin central domain.

There are two flagellin forms (encoded by *flaA* and *flaB*) incorporated into the flagella of *Ba. thuringiensis* subsp. *alesti* (Lövgren *et al.*, 1993). The *flaB* gene is flanked by long (355 bp) direct repeats that protrude into the coding region in the N- and C-terminal regions. DNA sequences related to *flaB* have been found in multiple copies in some *Ba. thuringiensis* subspecies, such as *kurstaki* (Lövgren *et al.*, 1993). It may be that the repeat sequences are involved in recombination events leading to amplification of flagellin genes in some *Ba. thuringiensis* isolates, and that this amplification leads to a selective advantage (Lövgren *et al.*, 1993).

Cryptic flagellin genes have been found in two species of *Shigella* (Tominaga *et al.*, 1994). The *fliC* gene from *Shigella flexneri* was cloned and introduced into a mutant *E. coli* Δ*fliC* strain. The flagellin was produced and assembled into normal-type flagella. Nucleotide sequence comparisons suggest that loss of motility in *Shigella* is evolutionarily a recent event.

In general, flagellin proteins lack cysteine residues and contain little or no tyrosine, tryptophan and histidine. There are some exceptions to this. For example, there is a single cysteine residue in *Roseburia cecicola* and three cysteine residues have been found in two different *Ps. putida* flagellins (Winstanley *et al.*, 1994), although there is no evidence indicating involvement in disulphide bridge formation. Alignment of N- and C-terminal flagellin deduced amino acid sequences from representatives of 26 eubacterial genera is shown in Fig. 2. Although such alignments demonstrate the location of conserved regions, it is clear that there are relatively few positions at which the residues are conserved in all flagellins sequenced to date. Proteins can vary greatly in amino acid sequence whilst remaining more or less functionally equivalent, and there are likely to be different structural constraints depending on whether the flagellar filament is free-standing or required to be bundle-forming. Filaments interacting directly with the external environment of the cell could also be expected to differ from those of sheathed flagella.

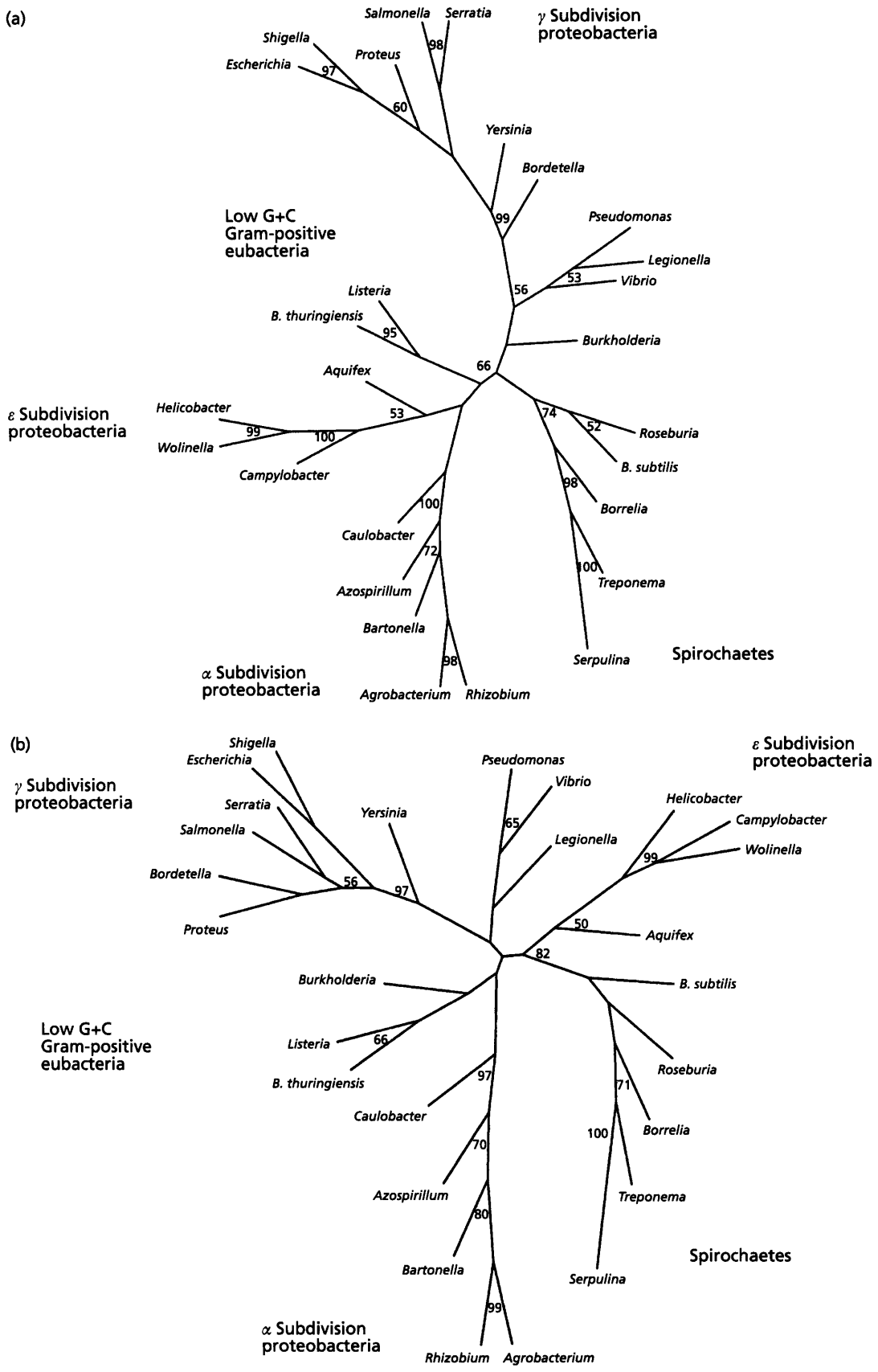

Fig. 3. For legend see facing page.

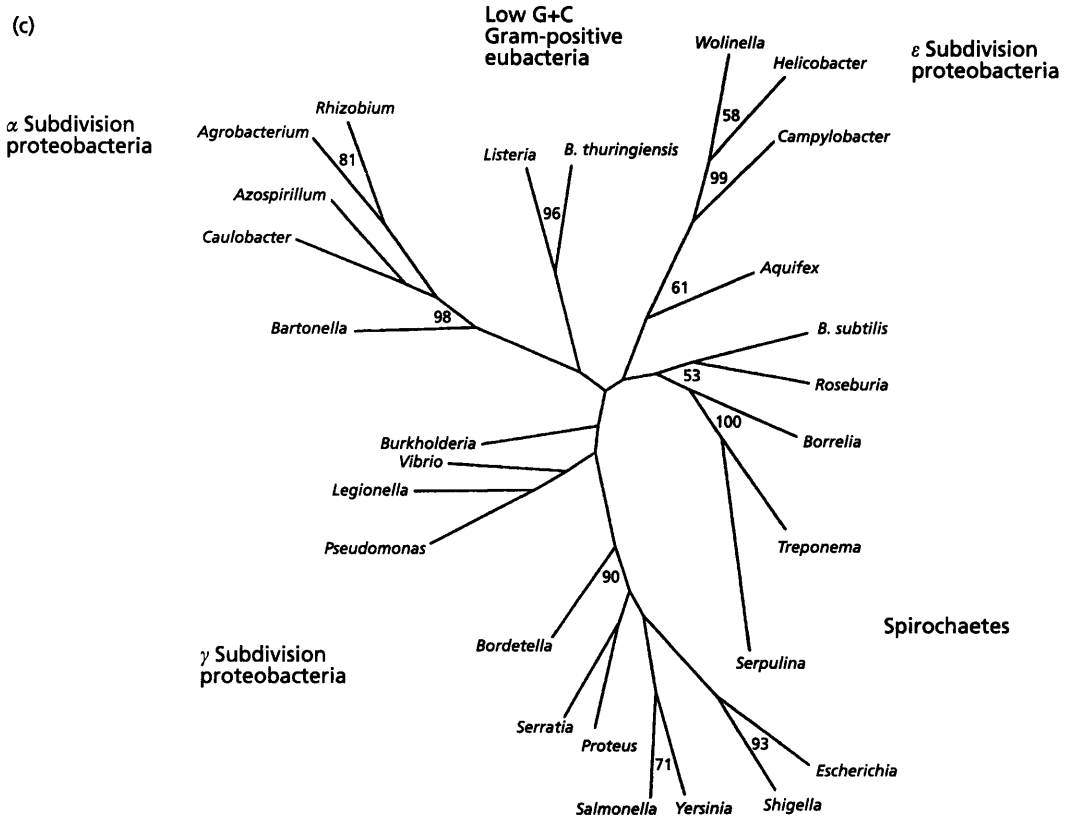

Fig. 3. Flagellin sequence similarity trees. Alignment of 27 deduced flagellin protein sequences, as used in Fig. 2 and indicated in Table 1, was performed using the program PILEUP (Genetics Computer Group, University of Wisconsin, USA). Trees, representing alignment using the complete flagellin sequence (a), the first 100 N-terminal amino acids (b) and the final 100 C-terminal amino acids (c), were constructed using the PHYLIP program. Bootstrap values, indicated on the trees, were derived using 100 alternatives (values under 50% are omitted).

In the N-terminal domain, the greatest area of conservation is found between positions 30 and 51 (Fig. 2), with a smaller conserved area much closer to the N-terminus (positions 5–8; Fig. 2). After 50 amino acids, the sequence becomes more variable, although there are some weaker conserved areas between positions 51 and 100. In the C-terminal domain, the greatest area of conservation is found within the final 16–18 amino acid residues. There is also a second shorter conserved area located 18–20 residues further upstream (Fig. 2).

Because of the considerable variation in the size of the central region of flagellins, sequence comparisons are difficult, although this can be overcome by using comparisons based on terminal sequences only (DeShazer *et al.*, 1997). When either complete flagellin amino acid sequences, the first 100 N-terminal residues, or the final 100 C-terminal residues are aligned, separate groupings can be discriminated (Fig. 3). Although there are some anomalies, phylogenetic trees based on flagellin sequences are largely consistent with similar trees based on rRNA comparisons (Olsen *et al.*, 1994). *Rhizobium*, *Bartonella*, *Caulobacter*, *Agrobacterium* and *Azospirillum* form a clear grouping of α subdivision proteobacteria (Moens *et al.*, 1995a). Although there is considerable variation within this group, the *Bartonella* group clusters separately from all other bacterial genera (Fig. 3).

Aquifex pyrophilus, a hyperthermophilic Gram-negative rod that is not readily grouped by rRNA sequence analysis (Olsen *et al.*, 1994), also has a distinct flagellin amino acid sequence. The flagellin protein of this organism is stable at temperatures of 85 °C and beyond, temperatures at which the flagellar filaments of mesophilic organisms break down into their constituent flagellin monomers. The thermostable flagellin is characterized by increases in aromatic and proline residues, as well as a significant increase in hydrophobic residues. Behammer *et al.* (1995) have suggested that these compositional features enable the formation of more compact flagellin monomers and a more stable contact between neighbouring subunits in the polymer.

There are some differences observed for the relationships of the groupings with each other, depending on whether the whole protein sequence or a 100 amino acid terminal sequence is used, but the groupings themselves are consistent. *Campylobacter*, *Helicobacter* and *Wolinella* form a cluster of ε subdivision proteobacteria, and *Ba.*

thuringiensis forms a grouping of low G+C Gram-positive eubacteria with *Listeria* (Fig. 3). The enterobacteria *Escherichia*, *Shigella*, *Salmonella*, *Serratia*, *Proteus* and *Yersinia*, members of the γ subdivision proteobacteria, cluster together with *Bordetella*, a member of the β subdivision. *Pseudomonas*, *Legionella* and *Vibrio* form a second grouping of γ subdivision proteobacteria. *Burkholderia*, a member of the β subdivision, does not cluster closely with any of the other genera, suggesting that its flagellin may be more representative of the β subdivision than the flagellin of *Bordetella bronchiseptica*. Additional flagellin sequences from other genera of the β subdivision proteobacteria are required to confirm this.

A cluster of spirochaetes (*Borrelia*, *Treponema* and *Serpulina*) is observed close to *R. cecicola* (Gram-negative cell wall with Gram-positive 16S rRNA sequence) and *Bacillus subtilis*, which does not cluster with *Ba. thuringiensis* (Fig. 3).

Serotype variation and antigenicity

The role of flagella and motility in bacterial virulence was the subject of a review by Penn & Luke (1992). The flagellins of pathogenic bacteria are known to be both highly immunogenic and variable in sequence. In humans, the early immune response in Lyme disease is primarily directed against the *Bo. burgdorferi* flagellin. For this reason, peptide fragments derived from flagellin may be applied in specific immunological tests to identify the disease (Rasiah *et al.*, 1992).

One of the best known examples of an organism able to vary its flagella is *S. typhimurium*, where flagellin phase variation relies on the presence of two alternatively expressed flagellin genes, *fliC* (phase 1) and *fljB* (phase 2) present in the same strain (Silverman & Simon, 1980). Switching between the phase 1 and phase 2 flagellins occurs by a recombination event involving an invertible element. Because flagellin gene sequences have been analysed in many different bacteria (Wilson & Beveridge, 1993; Table 1), detailed analysis of how flagellin proteins can vary antigenically due to genetic events such as recombination and horizontal gene transfer has been carried out. The evolutionary mechanisms generating serotype diversity have been studied in a number of bacteria including *Salmonella* where additional variation can be generated by recombination and lateral transfer. Smith *et al.* (1990) sequenced the central, antigen-determining region of the phase 1 flagellin gene (*fliC*) in strains of several *Salmonella* serovars. Based on the results obtained, the authors suggested that the major evolutionary mechanisms generating new serovars were lateral transfer and recombinational events.

In *Clostridium chauvoei* flagella are associated with virulence. Phase variation in motility and flagellation occurs in this organism, with spontaneous non-motile variants arising at an unusually high rate (10^{-4} per generation; Tamura *et al.*, 1995). Tamura *et al.* (1992) carried out an antigenic analysis of *Clos. chauvoei* flagella with protective and non-protective mAbs. They observed that the protective mAbs did not react with CNBr-cleaved peptides of flagellin whereas a non-protective mAb, Mo-114, did. The authors concluded that protective mAbs may recognize conformational epitopes derived from the secondary structure of flagellin.

In *Pseudomonas aeruginosa* there are two major groups of flagellar antigenic types (a and b). Whilst type b strains have flagellins with a molecular mass of 53 kDa, type a flagellins are more heterogeneous with molecular masses ranging from 45 kDa to 52 kDa (Winstanley *et al.*, 1996). Lagacé *et al.* (1995) reported IgG responses to both flagellin types in patients with cystic fibrosis (CF), with type b flagellins provoking a greater response. The authors suggested that the more restricted response to type a antigen in CF sera may be due to the heterogeneous nature of type a flagella. It has been reported that many *Ps. aeruginosa* isolates obtained from chronically colonized CF patients are non-motile and lack flagellin expression (Mahenthiralingham *et al.*, 1994). This loss of motility was associated with resistance to ingestion by macrophages, a property which may enable the organisms to persist in the respiratory tract of CF patients, once colonization is established. However, analysis of sera suggests that at least a proportion of bacteria continue to express flagellar antigens in chronically colonized CF patients (Lagacé *et al.*, 1995).

In *Campylobacter* infection the flagellum is the immunodominant antigen (Nachamkin & Yang, 1989). *Camp. coli* has been shown to undergo reversible flagellar antigenic variation between two antigenic types (type 1, T1; and type 2, T2). The flagellar filament of *Campylobacter* spp. is composed of two highly related flagellins, FlaA and FlaB. Both flagellins are incorporated into a complex flagellar filament in both antigenic types, yet the flagella produced by T1 cells can be distinguished from flagella produced by T2 cells on the basis of differences in the apparent molecular mass of component flagellin proteins (Alm *et al.*, 1992). The antigenic specificities exhibited by T1 and T2 cells are thought to result from post-translational modification (Alm *et al.*, 1992). Doig *et al.* (1996) have identified a sero-specific glycosyl moiety implicated in post-translational modification in *Campylobacter*. Genes (*ptmA* and *ptmB*) required for modification of *Camp. coli* flagellin have been identified and characterized (Guerry *et al.*, 1996). In a rabbit model, a *ptmA* mutant exhibited a reduced ability to elicit protection against subsequent colonization, although the ability to colonize the rabbit intestinal tract was unaffected. This suggests that surface-exposed modifications of flagellin may be more important than primary amino acid sequence in eliciting protection against *Campylobacter* (Guerry *et al.*, 1996).

Glycosylation has also been reported for the polar flagellum of *Azospirillum brasilense* (Moens *et al.*, 1995b), although it is much more common to find

glycosylated flagellins in the Archaea (Jarrell *et al.*, 1996). Post-translational modification of flagellins has also been observed in other bacteria. In the flagellin of *Ps. aeruginosa*, phosphorylated tyrosines have been identified (Kelly-Wintenberg *et al.*, 1993).

Recombination or lateral gene transfer events have also been implicated in antigenic variation in *Camp. jejuni* (Wassenaar *et al.*, 1995). The two tandem genes encoding the flagellar filament of *Camp. jejuni*, *flaA* and *flaB*, are highly similar and therefore subject to recombination. Such recombination was demonstrated by inserting an antibiotic resistance marker into the *flaA* gene and subsequently isolating a recombinant in which the antibiotic resistance gene had been repositioned into *flaB* (Wassenaar *et al.*, 1995). In the same study, recombinational events following the uptake of exogenous DNA by naturally competent *Camp. jejuni* were also demonstrated.

Bahrani *et al.* (1991) reported a strong immune response to the flagella of *Pr. mirabilis*. The rate of synthesis of the flagellin protein (FlaA) of *Pr. mirabilis* increases dramatically during cell differentiation. Whilst swimmer cells have only a few flagella, the elongated swarmer cells are covered by thousands of newly synthesized flagella. The *flaA* locus of *Pr. mirabilis* consists of two tandemly linked and nearly identical flagellin genes, *flaA* and *flaB*, although only *flaA* is expressed to produce flagellin protein for filament assembly (Belas, 1994). Expression of *flaA* is increased during differentiation, whilst *flaB* remains silent throughout. Although putative σ^{28} promoters have been identified for both genes, it has been suggested that the *flaB* promoter, which contains two mismatches from the σ^{28} consensus recognition sequence, is not recognized by the σ^{28} polymerase. Antigenically distinct FlaA proteins have been observed following the reversion of FlaA$^-$ mutants (Belas, 1994). The reversion process was associated with deletions occurring in the *flaA* locus.

Flagellin-based detection/identification of bacteria

Because of their domainal structure, flagellin genes are ideal candidates for PCR amplification. This feature has led to the development of a number of applications based on PCR amplification of bacterial flagellin genes. These include methods for the detection or identification of specific bacteria. Oyofo & Rollins (1993) targeted a region of the *Campylobacter flaA* gene for detection of *Camp. coli* and *Camp. jejuni* in environmental water samples. Detection was possible with a high degree of specificity and sensitivity even in water samples where culture of the organism was not possible. PCR amplification of *flaA* and *flaB* gene sequences has been used for the specific detection of *Camp. coli* and *Camp. jejuni* in chicken faecal samples both before and after selective enrichment (Rasmussen *et al.*, 1996). A detection limit of 1–20 cells was achieved using this approach.

A comparison of the *Borrelia hermsii* and *Bo. burgdorferi* flagellin gene sequences enabled the design of species-specific oligonucleotide primers for use in PCR-based detection (Picken, 1992). Not only was a *Bo. burgdorferi*-specific probe developed, but further sequence analysis of amplified products led to the design of oligonucleotides for use in differentiating three *Bo. burgdorferi* groups. Flagellin gene sequence divergence was suggested as a useful variable target for the differentiation of closely related *Borrelia* species (Picken, 1992).

Way *et al.* (1993) reported a multiplex PCR approach to the detection of *Salmonella* spp. Primers specific for the *H-li* region of the *S. typhimurium* H1 flagellin gene (*fliC*), and for a region bordered by *hin* (which encodes an invertase enzyme involved in phase variation) and H2 flagellin gene (*fliB*) sequences, were included in the reaction mixture. The authors reported the specific detection of *Salmonella typhi*, *S. typhimurium*, *Salmonella paratyphi* A and *Salmonella enteritidis*. The primers were tested on environmental samples such as soil and water to demonstrate the detection of *Salmonella* spp. without the need to culture the organism. A method for the rapid identification of *Listeria* spp., based on PCR amplification of the *flaA* gene using oligonucleotide primers designed to genus-specific regions of the gene, has also been developed (Gray & Kroll, 1995).

Denning *et al.* (1997) reported the use of flagellin gene target sequences in a PCR-based approach to the detection of *Pseudomonas fluorescens* SBW25 in a field release experiment. The stability of the gene was assessed by RFLP analysis of flagellin genes amplified from isolates obtained during the release. The flagellin gene proved to be a stable marker for strain identification throughout the period of the release. Because of the location of flagella on the outside of bacterial cells, flagellins are also useful targets for immunological detection. Morgan *et al.* (1991) reported a strategy of immunocapture to detect a specific *Ps. putida* strain in lake water. The surfaces of magnetic polystyrene beads were coated with a flagellin-specific antibody (MLV1). Antibody-coated beads could then be introduced into water samples. Bead–cell complexes were subsequently recovered in a magnetic field.

Application of flagellin gene variation in systematics/population genetics

Detailed analysis of variation in the *fliC* nucleotide sequence has been carried out in *Salmonella* spp. (Selander *et al.*, 1994). The complete *fliC* sequence for strains of 15 *Salmonella enterica* serovars of subspecies I, II, IV and VII that express combinations of six phase 1 flagellar antigenic factors of the g series (f, g, m, s, t and z_{51}) have been reported (Masten & Joys, 1993; Selander *et al.*, 1994). For the purposes of generating evolutionary trees, the flagellin gene sequences were separated into three regions comprising the two terminal regions (C1 and C2) and the central variable region (V). A comparison of phylogenetic trees generated by (i) multilocus enzyme electrophoresis (MLEE), (ii) the

nucleotide sequence of the combined C1 and C2 regions of *fliC*, and (iii) the nucleotide sequence of the variable region (V) of *fliC*, was made (Selander *et al.*, 1994). There was clear evidence of highly conserved flagellin gene sequences being present in strains exhibiting divergence in overall chromosomal genetic character. The occurrence of the same flagellin serotype in distantly related strains (indicated by MLEE) was attributed to horizontal exchange. The sequence analysis provided clear evidence that parts or all of the epitope-determining V region have been exchanged within and between subspecies. This is in stark contrast to studies of *Salmonella* population genetics using markers such as genes encoding metabolic enzymes and other housekeeping proteins (Selander *et al.*, 1994), which suggest that the population structure of *S. enterica* is essentially clonal.

It has been suggested that the extensive flagellar variation exhibited in *S. enterica*, whereby the *fliC* genes of salmonellae have evolved a mosaic structure made up from recombined segments, may be an adaptive response to enable the reinfection of a host (Brunham *et al.*, 1993). An alternative explanation is based on the absence of functional constraint, particularly in the central, variable region of flagellins, with the consequence that amino acid substitutions are able to accumulate free from counter-selection (Joys, 1988). A more comprehensive study of flagellar variation amongst environmental bacteria, free from any necessity for antigenic variation, may provide the answer, although even without the struggle to avoid the immune system, there may be a requirement for variability of surface structures. It has been observed that sensitivity to flagellotropic phage may be serotype-dependent (Iino, 1977).

Extensive diversity has also been observed in the *H. pylori* flagellin genes *flaA* and *flaB* (Forbes *et al.*, 1995). Studies on allelic variation suggest that such diversity is due to the reassortment of flagellin gene sequences between strains. Observations in *Helicobacter* indicate that a relatively small number of individual sequence mutations recombine together in random combinations to generate greater allelic diversity. If the only source of divergence was due to mutational accumulation, then the combinations of mutations in the different alleles would be expected to be in linkage disequilibrium. A comparatively small number of mutations reassorted into numerous and varied, novel combinations, suggests the involvement of DNA transfer between strains. *Helicobacter* is known to be naturally transformable under laboratory conditions and may have a non-clonal population structure exhibiting extreme diversity. Although the evidence from flagellin genes suggests that this may be the case (Forbes *et al.*, 1995), the experience of flagellar variation in *Salmonella* indicates that such conclusions cannot be drawn safely from flagellin gene variation alone.

Flagellin gene sequences have been applied to phylogenetic studies in *Bo. burgdorferi* (Fukunaga & Koreki, 1996), where 11 isolates of *Bo. burgdorferi sensu lato* were aligned with representatives of the three species recognized to be associated with Lyme disease: *Bo. burgdorferi sensu stricto*, *Borrelia garinii* and *Borrelia afzelii*. The study was used to assign isolates previously designated as ribotypes IV, V and VI, and not previously proven to be associated with any of the three recognized species involved in Lyme disease, to the species *Bo. garinii*. The data were obtained by PCR amplification of flagellin genes, using oligonucleotide primers designed to N- and C-terminal conserved regions. Amplified products were subsequently cloned and sequenced prior to alignment and construction of phylogenetic trees.

Regions of the gene encoding flagellin (*flaA*) were included in a sequencing study along with genes encoding the invasion-associated protein (*iap*), listeriolysin O (*hly*) and 23S rRNA for a range of *Listeria monocytogenes* isolates (Rasmussen *et al.*, 1995). No differences were found in rRNA sequences, but based on the nucleotide sequence differences observed in the other three target genes, *L. monocytogenes* strains could be divided into three distinct sequence types, exhibiting a strong degree of conservation within each type. There was complete agreement between the assignments based on the three different genes, including *flaA*, and the groupings were further confirmed by serotyping, PFGE and MLEE.

Flagellin gene variation as a target for epidemiological studies

By comparing the flagellin gene sequences from a number of related strains it is frequently possible to design oligonucleotide primers specific for N-terminal and C-terminal conserved regions. Using these primers the central, variable region of flagellin genes can be amplified by the PCR. Variability between amplified products can subsequently be assessed by RFLP analysis or DNA sequencing. RFLP patterns need to be interpreted with some caution since single-base mutations can account for differences. The epidemiological significance of such small variations is questionable. These limitations are often overcome by employing a number of different restriction enzymes to indicate the presence of allelic differences. Each restriction enzyme effectively samples for a different selection of mutations within the allele. A combination of several restriction enzymes can lead to reasonably comprehensive analysis of sequence variation throughout an allele. Although it is relatively secure to infer differences between strains, incontrovertible proof of the relatedness of strains is not easily obtained by a PCR/RFLP approach without recourse to the use of multiple enzymes, nucleotide sequencing or a combination of typing methods.

In a recent study it was demonstrated that the length of flagellin gene amplified product generated by the PCR can be used to separate the vast majority of *Ps. aeruginosa* clinical isolates into two homogeneous groups, differing in product size (1·0 kb and 1·25 kb) and

corresponding to the major antigenic types (a and b) (Winstanley *et al.*, 1996). In this study, flagellin gene sequences from 64 clinical isolates were amplified and subjected to RFLP analysis by using six restriction enzymes to digest the amplified products. By using RFLP the isolates were assigned to one of 13 groups. The method was rapid, reproducible and universally applicable to all isolates, in contrast to serotyping which failed to satisfactorily resolve 49% of the strains tested (Winstanley *et al.*, 1996). The flagellin gene RFLP method was applied in conjunction with macrorestriction analysis, using PFGE, to demonstrate the transmissibility of a β-lactam-resistant strain of *Ps. aeruginosa* in a CF clinic (Cheng *et al.*, 1996). The two molecular methods were equally effective at differentiating *Ps. aeruginosa* isolates from CF patients.

A combined PCR/RFLP method has also been applied to study the distribution and polymorphism of flagellin genes from North American isolates of *Camp. coli* and *Camp. jejuni* (Alm *et al.*, 1993). Primers specific for *flaA*, encoding the major flagellin, were used to obtain amplified products that were subsequently digested with *Pst*I and *Eco*RI to demonstrate conserved *flaA* genes within the majority of serotype groups. Alternative primers, designed to be specific for *flaB*, were employed to confirm the presence of this second flagellin gene in all isolates. In most serogroups, no variation between *flaA* and *flaB* gene sequences was detectable by restriction analysis. In a similar study, Mohran *et al.* (1996) found much greater variability of flagellin genes within the same serotype group for isolates obtained in Egypt. Nachamkin *et al.* (1993) also reported variability within serotype groups after using an enzyme with a 5 base recognition sequence to digest flagellin gene amplified products. Thomas *et al.* (1997) used flagellin gene polymorphism analysis to monitor diversity amongst *Camp. jejuni* isolates in a commercial poultry flock during the course of a 7 week infection. Over the period of the study five RFLP profiles were detected, with three dominant and two at low frequency, demonstrating that multiple genotypes could exist within the flock.

Kilger & Grimont (1993), employing an approach involving PCR amplification of the phase 1 flagellin gene (*fliC*), demonstrated that restriction profiles of flagellin gene amplified products can be used to differentiate several *Salmonella* flagellar types. The method involves a considerably less time-consuming process than conventional serotyping, with the additional advantage that strains of the serotype *gallinarum-pullorum*, which carry a cryptic gene for flagellar type g,m, can be distinguished from non-motile variants of serotype *typhi*.

Forbes *et al.* (1995) reported that allelic variation in the *H. pylori* flagellin genes (*flaA* and *flaB*) can be exploited in a PCR/RFLP approach to typing this organism. Regional differences were noted in the relative frequencies of *flaA* RFLP patterns, with the most common pattern in a particular geographical region often being rare in another region. The authors reported that epidemiological information based on the use of only one or two enzymes should be interpreted with caution, since many *flaA* alleles could only be distinguished by employing a combination of four restriction enzymes. The number of differentiable alleles was largely a function of the number of different enzymes used.

Conclusions

Because of the non-essential nature of much of the flagellin protein, bacterial flagellin genes and proteins exhibit considerable intra-species variation that can be utilized for specific detection or identification of species or strains. This characteristic has been exploited for a number of years in serotyping analysis targeting flagellar antigens. Molecular biological approaches offer the potential for rapid and reproducible analysis of diversity. In particular, the structure of flagellin genes, with terminal conserved regions, make them ideal candidates for PCR amplification, which can be combined with RFLP or sequence analysis to target variation in the central region.

The quantity of sequence data available makes flagellar variation a marker with widespread potential uses for the detection or identification of motile bacteria and for studies of the epidemiology or population genetics of a range of organisms.

Acknowledgements

The authors would like to acknowledge the support of The Wellcome Trust and the Biotechnology and Biological Sciences Research Council.

References

Aizawa, S.-I. (1996). Flagellar assembly in *Salmonella typhimurium*. *Mol Microbiol* **19**, 1–5.

Allison, C. & Hughes, C. (1991). Bacterial swarming: an example of prokaryotic differentiation and multicellular behaviour. *Sci Prog* **75**, 403–422.

Allison, C., Coleman, N., Jones, P. L. & Hughes, C. (1992). The ability of *Proteus mirabilis* to invade human urothelial cells is coupled to motility and swarming differentiation. *Infect Immun* **60**, 4740–4746.

Alm, R. A., Guerry, P., Power, M. E. & Trust, T. J. (1992). Variation in antigenicity and molecular weight of *Campylobacter coli* VC167 flagellin in different backgrounds. *J Bacteriol* **174**, 4230–4238.

Alm, R. A., Guerry, P. & Trust, T. J. (1993). Distribution and polymorphism of the flagellin genes from isolates of *Campylobacter coli* and *Campylobacter jejuni*. *J Bacteriol* **175**, 3051–3057.

Bahrani, F. K., Johnson, D. E., Robbins, D. & Mobley, H. L. T. (1991). *Proteus mirabilis* flagella and MR/P fimbriae: isolation, purification, N-terminal analysis, and serum antibody response following experimental urinary tract infection. *Infect Immun* **59**, 3574–3580.

Behammer, W., Shao, Z., Mages, W., Rachel, R., Stetter, K.-O. & Schmitt, R. (1995). Flagellar structure and hyperthermophily: analysis of a single flagellin gene and its product in *Aquifex pyrophilus*. *J Bacteriol* **177**, 6630–6637.

Belas, R. (1994). Expression of multiple flagellin-encoding genes of *Proteus mirabilis*. *J Bacteriol* **176**, 7169–7181.

Belas, R. (1996). *Proteus mirabilis* and other swarming bacteria. In *Bacteria as Multicellular Organisms*, pp. 183–219. Edited by J. A. Shapiro & M. Dworkin. Oxford: Oxford University Press.

Brunham, R. C., Plummer, F. A. & Stephens, R. S. (1993). Bacterial antigenic variation, host immune response, and pathogen–host coevolution. *Infect Immun* **61**, 2273–2276.

Charon, N. W., Greenberg, E. P., Koopman, M. B. H. & Limberger, R. J. (1992). Spirochete chemotaxis, motility, and the structure of the spirochetal periplasmic flagella. *Res Microbiol* **143**, 597–603.

Cheng, K., Smyth, R. L., Govan, J. R. W., Doherty, C., Winstanley, C., Denning, N., Heaf, D. P., van Seane, R. & Hart, C. A. (1996). Spread of a β-lactam-resistant *Pseudomonas aeruginosa* in a cystic fibrosis clinic. *Lancet* **348**, 639–642.

Denning, N., Morgan, J. A. W., Whipps, J. M., Saunders, J. R. & Winstanley, C. (1997). The flagellin gene as a stable marker for detection of *Pseudomonas fluorescens* SBW25. *Lett Appl Microbiol* **24**, 198–202.

DeShazer, D., Brett, P. J., Carlyon, R. & Woods, D. E. (1997). Mutagenesis of *Burkholderia pseudomallei* with Tn5-OT182: isolation of motility mutants and molecular characterization of the flagellin structural gene. *J Bacteriol* **179**, 2116–2125.

Doig, P. & Trust, T. J. (1994). Identification of surface-exposed outer-membrane antigens of *Helicobacter pylori*. *Infect Immun* **62**, 4526–4533.

Doig, P., Kinsella, N., Guerry, P. & Trust, T. J. (1996). Characterization of a post-translational modification of *Campylobacter* flagellin: identification of a sero-specific glycosyl moiety. *Mol Microbiol* **19**, 379–387.

Driks, A., Bryan, R., Shapiro, L. & DeRosier, D. J. (1989). The organization of the *Caulobacter crescentus* flagellar filament. *J Mol Biol* **206**, 627–636.

Forbes, K. J., Fang, Z. & Pennington, T. H. (1995). Allelic variation in the *Helicobacter pylori* flagellin genes *fla*A and *fla*B: its consequences for strain typing schemes and population structure. *Epidemiol Infect* **114**, 257–266.

Fuerst, J. A. & Perry, J. W. (1988). Demonstration of lipopolysaccharide on sheathed flagella of *Vibrio cholerae* O:1 by protein A–gold immunoelectron microscopy. *J Bacteriol* **170**, 1488–1494.

Fukunaga, M. & Koreki, Y. (1996). A phylogenetic analysis of *Borrelia burgdorferi sensu lato* isolates associated with Lyme disease in Japan by flagellin gene sequence determination. *Int J Syst Bacteriol* **46**, 416–421.

Ge, Y. & Charon, N. W. (1997). An unexpected *flaA* homolog is present and expressed in *Borrelia burgdorferi*. *J Bacteriol* **179**, 552–556.

Geis, G., Leying, H., Suerbaum, S., Mai, U. & Opferkuch, W. (1989). Ultrastructure and chemical analysis of *Campylobacter pylori* flagella. *J Clin Microbiol* **27**, 436–441.

Gray, D. I. & Kroll, R. G. (1995). Polymerase chain reaction amplification of the *flaA* gene for the rapid identification of *Listeria* spp. *Lett Appl Microbiol* **20**, 65–68.

Guerry, P., Alm, R. A., Power, M. E., Logan, S. M. & Trust, T. J. (1991). Role of two flagellin genes in *Campylobacter* motility. *J Bacteriol* **173**, 4757–4764.

Guerry, P., Doig, P., Alm, R. A., Burr, D. H., Kinsella, N. & Trust, T. J. (1996). Identification and characterization of genes required for post-translational modification of *Campylobacter coli* VC167 flagellin. *Mol Microbiol* **19**, 369–378.

Harshey, R. M. (1994). Bees aren't the only ones: swarming in Gram-negative bacteria. *Mol Microbiol* **13**, 389–394.

Homma, M., Fujita, H., Yamaguchi, S. & Iino, T. (1987). Regions of *Salmonella typhimurium* flagellin essential for its polymerization and excretion. *J Bacteriol* **169**, 291–296.

Iino, T. (1977). Genetics of structure and function of bacterial flagella. *Annu Rev Genet* **11**, 161–182.

Jank, B., Doblhoff-Dier, O., van der Plas, J. & Habermann, B. (1995). Sequence motifs in a flagellin of *Pseudomonas putida*. *Microbiology* **141**, 1491–1492.

Jarrell, K. F., Bayley, D. P. & Kostyukova, A. S. (1996). The archaeal flagellum: a unique motility structure. *J Bacteriol* **178**, 5057–5064.

Josenhans, C., Labigne, A. & Suerbaum, S. (1995). Comparative ultrastructural and functional studies of *Helicobacter pylori* and *Helicobacter mustelae* flagellin mutants: both flagellin subunits, FlaA and FlaB, are necessary for full motility in *Helicobacter* species. *J Bacteriol* **177**, 3010–3020.

Joys, T. M. (1988). The flagellar filament protein. *Can J Microbiol* **34**, 452–458.

Kelly-Wintenberg, K., South, S. L. & Montie, T. C. (1993). Tyrosine phosphate in a- and b-type flagellins of *Pseudomonas aeruginosa*. *J Bacteriol* **172**, 2458–2461.

Khan, S. (1993). Gene to ultrastructure: the case of the flagellar basal body. *J Bacteriol* **175**, 2169–2174.

Kilger, G. & Grimont, P. A. D. (1993). Differentiation of *Salmonella* phase 1 flagellar antigen types by restriction of the amplified *fliC* gene. *J Clin Microbiol* **31**, 1108–1110.

Kuwajima, G. (1988). Construction of a minimum-size functional flagellin of *Escherichia coli*. *J Bacteriol* **170**, 3305–3309.

Lagacé, J., Peloquin, L., Kermani, P. & Montie, T. C. (1995). IgG subclass responses to *Pseudomonas aeruginosa* a- and b-type flagellins in patients with cystic fibrosis: a prospective study. *J Med Microbiol* **43**, 270–276.

Limberger, R. J., Slivienski, L. L., Yelton, D. B. & Charon, N. W. (1992). Molecular genetic analysis of a class B periplasmic-flagellum gene of *Treponema phagedenis*. *J Bacteriol* **174**, 6404–6410.

Lövgren, A., Zhang, M.-Y., Engström, Å. & Landén, R. (1993). Identification of two expressed flagellin genes in the insect pathogen *Bacillus thuringiensis* subsp. *alesti*. *J Gen Microbiol* **139**, 21–30.

Lu, Z., Murray, K. S., van Cleave, V., LaVallie, E. R., Stahl, M. L. & McCoy, J. M. (1995). Expression of thioredoxin reductase random peptide libraries on the *Escherichia coli* cell surface as functional fusions to flagellin: a system designed for exploring protein–protein interactions. *Bio/Technology* **13**, 366–372.

Luke, C. J. & Penn, C. W. (1995). Identification of a 29 kDa flagellar sheath protein in *Helicobacter pylori* using a murine monoclonal antibody. *Microbiology* **141**, 597–604.

McCarter, L. L. (1995). Genetic and molecular characterization of the polar flagellum of *Vibrio parahaemolyticus*. *J Bacteriol* **177**, 1595–1609.

McGee, K., Horstedt, P. & Milton, D. L. (1996). Identification and characterization of additional flagellin genes from *Vibrio anguillarum*. *J Bacteriol* **178**, 5188–5198.

Macnab, R. M. (1992). Genetics and biogenesis of bacterial flagella. *Annu Rev Genet* **26**, 131–158.

Macnab, R. M. (1996). Flagella and motility. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn, pp.

123–145. Edited by F. C. Neidhardt and others. Washington, DC: American Society for Microbiology.

Mahenthiralingham, E., Campbell, M. E. & Speert, D. P. (1994). Nonmotility and phagocytic resistance of *Pseudomonas aeruginosa* isolates from chronically colonized patients with cystic fibrosis. *Infect Immun* **62**, 596–605.

Masten, B. J. & Joys, T. M. (1993). Molecular analysis of the *Salmonella g* flagellar antigen complex. *J Bacteriol* **175**, 5359–5365.

Minnich, S. A., Ohta, N., Taylor, N. & Newton, A. (1988). Role of the 25-, 27-, and 29-kilodalton flagellins in *Caulobacter crescentus* cell motility: method for construction of deletion and Tn*5* insertion mutants by gene replacement. *J Bacteriol* **170**, 3953–3960.

Moens, S., Michiels, K., Keijers, V., van Leuven, F. & Vanderleyden, J. (1995a). Cloning, sequencing and phenotypic analysis of *laf1*, encoding the flagellin of the lateral flagella of *Azospirillum brasilense* Sp7. *J Bacteriol* **177**, 5419–5426.

Moens, S., Michiels, K. & Vanderleyden, J. (1995b). Glycosylation of the flagellin of the polar flagellum of *Azospirillum brasilense*, a Gram-negative nitrogen-fixing bacterium. *Microbiology* **141**, 2651–2657.

Mohran, Z. S., Guerry, P., Lior, H., Murphy, J. R., Elgendy, A. M., Mikhail, M. M. & Oyofo, B. A. (1996). Restriction fragment length polymorphism of flagellin genes of *Campylobacter jejuni* and/or *Campylobacter coli* isolates from Egypt. *J Clin Microbiol* **34**, 1216–1219.

Morgan, J. A. W., Winstanley, C., Pickup, R. W. & Saunders, J. R. (1991). Rapid immunocapture of *Pseudomonas putida* cells from lake water by using bacterial flagella. *Appl Environ Microbiol* **57**, 503–509.

Nachamkin, I. & Yang, X. H. (1989). Human antibody response to *Campylobacter jejuni* flagellin protein and a synthetic N-terminal flagellin peptide. *J Clin Microbiol* **27**, 2195–2198.

Nachamkin, I., Bohachick, K. & Patton, C. M. (1993). Flagellin gene typing of *Campylobacter jejuni* by restriction fragment length polymorphism analysis. *J Clin Microbiol* **31**, 1531–1536.

Namba, K., Yamashita, I. & Vonderviszt, F. (1989). Structure of the core and central channel of bacterial flagella. *Nature* **342**, 648–654.

Newton, S. M. C., Jacob, C. O. & Stocker, B. A. D. (1989). Immune response to cholera toxin epitope inserted in *Salmonella* flagellin. *Science* **244**, 70–72.

Noppa, L., Burman, N., Sadziene, A., Barbour, A. G. & Bergström, S. (1995). Expression of the flagellin gene in *Borrelia* is controlled by an alternative σ factor. *Microbiology* **141**, 85–93.

Norris, S. J. & the *Treponema pallidum* polypeptide research group (1993). Polypeptides of *Treponema pallidum*: progress toward understanding their structural, function, and immunologic roles. *Microbiol Rev* **57**, 750–779.

Olsen, G. J., Woese, C. R. & Overbeek, R. (1994). The winds of (evolutionary) change: breathing new life into microbiology. *J Bacteriol* **176**, 1–6.

Oyofo, B. A. & Rollins, D. M. (1993). Efficacy of filter types for detecting *Campylobacter jejuni* and *Campylobacter coli* in environmental water samples by polymerase chain reaction. *Appl Environ Microbiol* **59**, 4090–4095.

Penn, C. W. & Luke, C. J. (1992). Bacterial flagellar diversity and significance in pathogenesis. *FEMS Microbiol Lett* **100**, 331–336.

Picken, R. N. (1992). Polymerase chain reaction primers and probes derived from flagellin gene sequences for specific detection of the agents of Lyme disease and North American relapsing fever. *J Clin Microbiol* **30**, 99–114.

Pleier, E. & Schmitt, R. (1991). Expression of two *Rhizobium meliloti* flagellin genes and their contribution to the complex filament structure. *J Bacteriol* **173**, 2077–2085.

Rasiah, C., Schiltz, E., Reichert, J. & Vogt, A. (1992). Purification and characterization of a tryptic peptide of *Borrelia burgdorferi* flagellin, which reduces cross-reactivity in immunoblots and ELISA. *J Gen Microbiol* **138**, 147–154.

Rasmussen, O. F., Skouboe, P., Dons, L., Rossen, L. & Olsen, J. E. (1995). *Listeria monocytogenes* exists in at least three evolutionary lines: evidence from flagellin, invasive associated protein and listeriolysin O genes. *Microbiology* **141**, 2053–2061.

Rasmussen, H. N., Olsen, J. E., Jorgensen, K. & Rasmussen, O. F. (1996). Detection of *Campylobacter jejuni* and *Camp. coli* in chicken faecal samples by PCR. *Lett Appl Microbiol* **23**, 363–366.

Schuster, S. C., Bauer, M., Kellermann, J., Lottspeich, F. & Baeuerlein, E. (1994). Nucleotide sequence of the *Wolinella succinogenes* flagellin, which contains in the antigenic domain two conserved regions also present in *Campylobacter* spp. and *Helicobacter pylori*. *J Bacteriol* **176**, 5151–5155.

Selander, R. K., Li, J., Boyd, E. F., Wang, F.-S. & Nelson, K. (1994). DNA sequence analysis of the genetic structure of populations of *Salmonella enterica* and *Escherichia coli*. In *Bacterial Diversity and Systematics*, pp. 17–49. Edited by F. G. Priest, A. Ramos-Cormenzana & B. J. Tindall. New York: Plenum.

Silverman, M. & Simon, M. (1980). Phase variation: genetic analysis of switching mutants. *Cell* **19**, 845–854.

Smith, N. H., Beltran, P. & Selander, R. K. (1990). Recombination of *Salmonella* phase 1 flagellin genes generates new serovars. *J Bacteriol* **172**, 2209–2216.

Sjoblad, R. D., Emala, C. W. & Doetsch, R. N. (1983). Bacterial flagellar sheaths: structures in search of a function. *Cell Motil* **3**, 93–103.

Tamura, Y., Kijima, M., Ohishi, K., Takahashi, T., Suzuki, S. & Nakamura, M. (1992). Antigenic analysis of *Clostridium chauvoei* flagella with protective and non-protective monoclonal antibodies. *J Gen Microbiol* **138**, 537–542.

Tamura, Y., Kijima-Tanaka, M., Aoki, A., Ogikubo, Y. & Takahashi, T. (1995). Reversible expression of motility and flagella in *Clostridium chauvoei* and their relationship to virulence. *Microbiology* **141**, 605–610.

Thomas, L. M., Long, K. A., Good, R. T., Panaccio, M. & Widders, P. R. (1997). Genotypic diversity among *Campylobacter jejuni* isolates in a commercial broiler flock. *Appl Environ Microbiol* **63**, 1874–1877.

Tomashow, L. S. & Rittenberg, S. C. (1985). Isolation and composition of sheathed flagella from *Bdellovibrio bacteriovorus* 109J. *J Bacteriol* **163**, 1047–1054.

Tominaga, A., Mahmoud, M. A.-H., Mukaihara, T. & Enomoto, M. (1994). Molecular characterization of intact, but cryptic, flagellin genes in the genus *Shigella*. *Mol Microbiol* **12**, 277–285.

Trachtenberg, S. & DeRosier, D. J. (1992). A three-start helical sheath on the flagellar filament of *Caulobacter crescentus*. *J Bacteriol* **174**, 6198–6206.

Wassenaar, T. M., Fry, B. N. & van der Zeijst, B. A. M. (1995). Variation of the flagellin gene locus of *Campylobacter jejuni* by recombination and horizontal gene transfer. *Microbiology* **141**, 95–101.

Way, J. S., Josephson, K. L., Pillai, S. D., Abbaszadegan, M., Gerba, C. P. & Pepper, I. L. (1993). Specific detection of *Salmonella*

spp. by multiplex polymerase chain reaction. *Appl Environ Microbiol* **59**, 1473–1479.

Wei, L.-N. & Joys, T. M. (1985). Covalent structure of three phase-1 flagellar filament proteins of *Salmonella*. *J Mol Biol* **186**, 791–803.

Whittam, T. S. (1995). Genetic population structure and pathogenicity in enteric bacteria. In *Population Genetics of Bacteria* (Society for General Microbiology Symposium **52**), pp. 217–245. Edited by S. Baumberg, J. P. W. Young, E. M. H. Wellington & J. R. Saunders. Cambridge: Cambridge University Press.

Wilson, D. R. & Beveridge, T. J. (1993). Bacterial flagellar filaments and their component flagellins. *Can J Microbiol* **39**, 451–472.

Winstanley, C., Morgan, J. A. W., Pickup, R. W. & Saunders, J. R. (1994). Molecular cloning of two *Pseudomonas* flagellin genes and basal body structural genes. *Microbiology* **140**, 2019–2031.

Winstanley, C., Coulson, M. A., Wepner, B., Morgan, J. A. W. & Hart, C. A. (1996). Flagellin gene and protein variation amongst clinical isolates of *Pseudomonas aeruginosa*. *Microbiology* **142**, 2145–2151.