Molecular evolution of the major outer-membrane protein gene (*oprF*) of *Pseudomonas*

Josselin Bodilis and Sylvie Barry

Correspondence

Josselin Bodilis
josselin.bodilis@univ-rouen.fr

LMDF (Laboratoire de Microbiologie Du Froid), UPRES 2123, ABISS (Atelier de Biologie, Informatique, Statistique et Sociolinguistique), Université de Rouen, 76821 Mont Saint Aignan, France

The major outer-membrane protein of *Pseudomonas*, OprF, is multifunctional. It is a non-specific porin that plays a role in maintenance of cell shape, in growth in a low-osmolarity environment, and in adhesion to various supports or molecules. OprF has been studied extensively for its utility as a vaccine component, its role in antimicrobial drug resistance, and its porin function. The authors have previously shown important differences between the OprF and 16S rDNA phylogenies: *Pseudomonas fluorescens* isolates split into two quite separate clusters, probably according to their ecological niche. In this study, the evolutionary history of the *oprF* gene was investigated further. The study of G + C content at the third codon position, synonymous codon usage (codon adaptation index, CAI) and genomic context showed no evidence of horizontal transfer or gene duplication. Similarly, a robust likelihood test of incongruence showed no significant incongruence between the *oprF* phylogeny and the species phylogeny. In addition, the ratio of nonsynonymous mutations to synonymous mutations (K_a/K_s) is high between the different clusters, especially between the two clusters containing *P. fluorescens* isolates, highlighting important modifications in evolutionary constraints during the history of the *oprF* gene. Since OprF is known as a pleiotropic protein, modifications in evolutionary constraints could have resulted from variations in cryptic functions, correlated with the ecological fingerprint. Finally, relaxed constraints and/or episodic positive evolution, especially for some *P. fluorescens* strains, could have led to a phylogeny reconstruction artifact.

Received 2 November 2005

Revised 9 December 2005

Accepted 15 December 2005

INTRODUCTION

OprF, the major outer-membrane protein in *Pseudomonas*, has been studied extensively due to its utility as a vaccine component, its role in antimicrobial drug resistance, and its porin function (Dé *et al.*, 1995; Jaouen *et al.*, 2004; Orange, 1994; Rawling *et al.*, 1995; Worgall *et al.*, 2005). This protein has only been found in this genus (but see Rediers *et al.*, 2004) and may be considered as a diagnostic protein for *Pseudomonas sensu stricto* (de Mot *et al.*, 1994; Vermeiren *et al.*, 1999; Aagot *et al.*, 2001). OprF shares C-terminal similarity with OmpA of *Escherichia coli* and thus is a member of the OmpA superfamily of porins (de Mot *et al.*, 1994).

OprF is an oligomer, possibly a trimer, that is associated with both LPS and peptidoglycan (Nikaido, 2003; Tamber & Hancock, 2004). The protein consists of three domains: the N-terminal domain, a central linker region, and the C-terminal domain. The N-terminal domain of OprF (the first 160 amino acids) forms an eight-stranded β -barrel

(Brinkman *et al.*, 2000). A central linker region (amino acids 161–209) contains a variable sequence of multiple proline-alanine repeats followed by a much-conserved sequence with two disulfide bonds that is absent in some *Pseudomonas* species (Vermeiren *et al.*, 1999; Bodilis *et al.*, 2004). The C-terminal region of OprF (amino acids 210–326) shares considerable similarity with that of OmpA (56 %), including the fact that its structure has not been completely elucidated (Pautsch & Schulz, 1998). Two folding models have been proposed (Nikaido, 2003; Tamber & Hancock, 2004). In the first model, OprF forms a 16- β -stranded membrane, since the C-terminal half is exposed and available to monoclonal antibodies (von Specht *et al.*, 1995; Rawling *et al.*, 1995). In the second model, the C-terminal half is a globular domain that lies in the periplasm since it is involved in peptidoglycan binding (Rawling *et al.*, 1998). Finally, to explain the variations in channel size observed in OprF (Bellido *et al.*, 1992), it has been suggested that the C-terminal domain can translocate between the periplasm and the surface (Nikaido, 2003; Tamber & Hancock, 2004).

In addition to having a complex and partially elucidated structure, OprF is a multifunctional protein. It is a non-specific porin permitting passive diffusion of small polar

Abbreviations: CAI, codon adaptation index; GC3, G + C content at the third codon position; ML, maximum likelihood.

nutrients (Orange, 1994). The OprF pore changes its channel size according to the growth conditions, and this could affect outer-membrane permeability in the three major species of the genus *Pseudomonas*: *P. fluorescens*, *P. putida* and *P. aeruginosa* (Dé *et al.*, 1997; Jaouen *et al.*, 2004). OprF is also involved in maintaining cell shape and in growth in a low-osmolarity environment (Woodruff & Hancock, 1989; Rawling *et al.*, 1998). Moreover, this protein probably plays a role in adhesion to roots in *P. fluorescens* strains isolated from the rhizosphere (de Mot *et al.*, 1992) and in adhesion to fibronectin in *P. fluorescens* (Rebière-Huët *et al.*, 1999). For *P. aeruginosa*, it has recently been reported that the binding of human interferon- γ to OprF results in the expression of the PA-I lectin, a quorum-sensing-dependent virulence determinant (Wu *et al.*, 2005).

In spite of the extensive studies of this protein and its potential taxonomic, ecological and medical applications, only two phylogenetic studies have been reported (Vermeiren *et al.*, 1999; Bodilis *et al.*, 2004). In the latter, we showed important differences in tree topologies between the OprF and 16S rDNA phylogenies obtained from 69 *Pseudomonas* isolates or reference strains (Bodilis *et al.*, 2004). The incongruence is that both the clusters, the one containing *P. fluorescens* strains and the one containing *P. putida* strains, each divides into two parts (termed o-clusters); no cross horizontal transfer was observed between the *fluorescens* o-clusters and the *putida* o-clusters (Fig. 1). Since the *fluorescens* 1 o-cluster contained most of the bulk soil (non-rhizospheric) isolates and the *fluorescens* 2 o-cluster contained all the clinical isolates and most of the rhizospheric isolates, this dichotomy of OprF seemed to be correlated with the ecological niche.

We previously proposed three explanations for the differences observed between the OprF and 16S rDNA phylogenies (Bodilis *et al.*, 2004): (i) an ancient horizontal transfer of *oprF* gene (after the *P. fluorescens*–*P. putida* speciation), (ii) an ancient duplication of the gene (before the speciation) followed by several losses, or (iii) artifacts in phylogeny reconstructions. These different possibilities and more generally the evolutionary history of the *oprF* gene have been investigated in this study.

METHODS

Database mining. Three different sets of sequences were used in this study, as follows.

(i) The 7-sequence sets (or 6-sequence set for the *cobA* gene) were retrieved from the finished genomes of *P. aeruginosa* PAO1, *P. putida* KT2440, *P. syringae* pv. *tomato* DC3000, *P. syringae* pv. *syringae* B728a, *P. fluorescens* Pf0-1 and *P. fluorescens* Pf-5 (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>), and the unfinished genome of *P. fluorescens* SBW25, sequenced with about an eightfold coverage (http://www.sanger.ac.uk/Projects/P_fluorescens/). The sets corresponded to the sequences of *oprF*, 16S rDNA, concatenated ribosomal genes, genes upstream of *oprF* (*ppsA*, *estX*, *menG*, *cmaX*, *crfX*, *cmpX*, *sigX*), concatenated upstream genes and *cobA* (only six sequences because of a deletion in strain Pf-5). For the set of ribosomal genes, 46 ribosomal

genes, ubiquitous in the seven sequenced *Pseudomonas* genomes, were aligned individually and concatenated, leading to 18 045 unambiguously aligned nucleotide positions.

(ii) The 20-sequence sets (for *oprF* and 16S rDNA genes) were composed of the 7-sequence sets plus the 13 sequences for which the accession numbers are given in Table 1.

(iii) The 69-sequence set (*oprF* gene) was composed in part of the sequences from a previous phylogenetic study (Bodilis *et al.*, 2004), minus the hybrid *oprF* of strains MFY71 and MFY140, and minus the *oprF* of strain MFY78 because it occurs at a flawed position between the *putida* 1 o-cluster and the *fluorescens* 1 o-cluster. The remainder of the 69-sequence set included the *oprF* gene of strains DC3000, B728a and Pf-5. This set included the set of 20 sequences and, therefore, the set of 7 sequences.

Phylogenetic analyses. For the three different sets of sequences, protein alignments (and nucleotide alignment for the 16S rDNA) were generated using CLUSTAL X version 1.81, with default parameters (Thompson *et al.*, 1997). These were truncated to the same size as the shortest sequence. The sequences in the 20- and 69-sequence sets were only partial sequences: amino acid residues 6–328 for OprF of strain PAO1 (there are 350 amino acids in the complete protein, with the peptide signal) and nucleotide positions 113–1363 for the 16S rDNA of strain PAO1 (there are 1536 nucleotides in the complete gene). Although *oprF* and 16S rDNA are complete in the sequenced genomes, the same partial sequences were used for the 7-sequence sets, to compare the results more easily. All alignments were optimized visually and all ambiguous positions and positions with gaps were removed for all subsequent analyses. In addition, positions corresponding to the central domain of OprF, which contained a low-complexity region (proline rich) and several gaps, were eliminated. Altogether, amino acid residues 1–5, 79–84, 125–126, 168–171, 187–233, 308 and 328–350 were removed for the OprF of the PAO1 strain. Finally, all nucleotide alignments were deduced from corrected protein alignments.

Phylogenetic neighbour-joining trees were computed using MEGA v2.0 (Saitou & Nei, 1987; Kumar *et al.*, 2001). Multiple substitution events were taken into account using Poisson correction, Kimura two-parameter correction (Kimura, 1980) and Jukes–Cantor correction (Jukes & Cantor, 1969), for protein sequences, nucleotide sequences and synonymous positions, respectively. Bootstrap confidence levels were determined by randomly resampling the alignment positions 1000 times.

Maximum-likelihood (ML) analyses were performed using PAUP software version 4.0b10 (D. L. Swofford, Smithsonian Institution, Sunderland, MA, USA). The GTR model was used with an among-site rate heterogeneity approximated by a discrete gamma distribution with four divisions (nCat parameter) plus a proportion of invariant sites (pInvar parameter). The best tree was sought by three successive iterations with TBR (tree bisection and reconnection) branch swapping: for each iteration, all parameters (GTR parameters, gamma shape and pInvar) were estimated on the best tree of the previous iteration and were fixed for the heuristic search. The tree used for parameters of the first iteration was a neighbour-joining tree with Kimura two-parameter correction.

The topologically constrained searches were performed in the same way as the unconstrained search except for the option ‘enforce constraints’. For example, for the topology constraint 3 (Fig. 2), the constraint was: (PAO1,KT2440,[SBW25,Pf0-1,Pf-5,(DC3000,B728a)]). While only the defined monophyletic groups were enforced (i.e. DC3000 + B728a and SBW25 + Pf0-1 + Pf-5 + DC3000 + B728a), all the other nodes were free during the search for the best tree. Topologies 4 and VIII correspond to the species phylogeny expected

Fig. 1. Phylogenetic relationships among the 20-sequence set of *oprF* genes (a), the 7-sequence set of *oprF* genes (b) and the 7-sequence set of 46 concatenated ribosomal genes (c) from *Pseudomonas* species. The unrooted dendrograms were generated, using the neighbour-joining algorithm, from evolutionary distances computed according to the Kimura two-parameter model (Kimura, 1980). Numbers on tree branches report bootstrap results (obtained for 1000 repetitions) from analysis of nucleotide sequences (below branch) and for analysis of protein sequences (above branch). Only bootstrap values $\geq 50\%$ are reported. The horizontal length of branches is proportional to the estimated number of substitutions.

(i.e. monophylies of species), for the 7- and 20-sequence sets, respectively. Topologies 8 and III correspond to the *OprF* phylogeny, for the 7- and 20-sequence sets, respectively. The other topologies tested correspond to intermediate and alternative topologies. The intermediate topologies, for example 1, 2 and 3 for topology 4, are less resolved. In other words, topology 4 is a nested topology constraint of the topology constraints 1, 2 and 3 (Fig. 2). The alternative topologies (9, 10, IX and X) are, *a priori*, unlikely.

The Shimodaira–Hasegawa test was used to determine whether constrained trees are significantly different ($P < 0.05$) from the unconstrained tree (Shimodaira & Hasegawa, 1999). For each tree, the likelihood was calculated with parameters estimated on the unconstrained tree. If a constrained tree had the same likelihood as the unconstrained tree (i.e. it is noted in the ‘Best topologies’ column of Table 2), we considered that the monophylies enforced by the topology constraint were present in the unconstrained tree. In other words, the

Table 1. Sequence accession numbers of the 20-sequence sets (16S rDNA and *oprF* gene)

Species	Strain†	Reference	16S rDNA accession no.	<i>oprF</i> gene accession no.
aeruginosa o-cluster				
<i>P. aeruginosa</i>	PAO1	See Methods		
<i>P. resinovorans</i>	LMG 2274 ^T	Vermeiren <i>et al.</i> (1999)	Z76668	AF117968
<i>P. stutzeri</i>	LMG 11199 ^T	Kerstens <i>et al.</i> (1996)	U26262	AF117971
<i>P. mendocina</i> *	ATCC 25411 ^T	Kerstens <i>et al.</i> (1996)	M59154	
<i>P. mendocina</i> *	LMG 1223 ^T	Vermeiren <i>et al.</i> (1999)		AF117966
fluorescens 1 o-cluster				
fluorescens 1 lineage				
<i>P. fluorescens</i>	Pf0-1	See Methods		
<i>P. fluorescens</i>	F113	Ramette <i>et al.</i> (2001)	AJ278814	AF115337
<i>P. fluorescens</i>	MFY30	Bodilis <i>et al.</i> (2004)	AY331341	AY331300
syringae lineage				
<i>P. syringae</i> pv. <i>tomato</i>	DC3000	See Methods		
<i>P. syringae</i> pv. <i>syringae</i>	B728a	See Methods		
<i>P. cichorii</i>	LMG 2162 ^T	Vermeiren <i>et al.</i> (1999)	Z76658	AF117963
fluorescens 2 o-cluster				
<i>P. fluorescens</i>	Pf-5	See Methods		
<i>P. fluorescens</i>	SBW25	See Methods		
<i>P. fluorescens</i>	MF0	Bodilis <i>et al.</i> (2004)	AY331340	U19743
<i>Pseudomonas</i> sp.	MFY143	Bodilis <i>et al.</i> (2004)	AY331368	AY331327
putida 1 o-cluster				
<i>P. putida</i>	KT2440	See Methods		
<i>P. putida</i>	MFY57	Bodilis <i>et al.</i> (2004)	AY331346	AY331305
<i>P. putida</i> *	LMG 2257 ^T	Vermeiren <i>et al.</i> (1999)		AF117967
<i>P. putida</i> *	DSM 291 ^T	Moore <i>et al.</i> (1996)	Z76667	
putida 2 o-cluster				
<i>Pseudomonas</i> sp.	MFY59	Bodilis <i>et al.</i> (2004)	AY331347	AY331306
<i>Pseudomonas</i> sp.	MFY69	Bodilis <i>et al.</i> (2004)	AY331352	AY331311
<i>Pseudomonas</i> sp.	MFY161	Bodilis <i>et al.</i> (2004)	AY331372	AY331331

*For *P. mendocina* and *P. putida*, the two strain types described were originally the same strain.

†Strains used for the 7-sequence sets are in bold.

topologies of the constrained and unconstrained trees were the same. If the likelihood of a constrained tree was different from that of the unconstrained tree, we considered that some of the monophylyies enforced by the topology constraint were not present in the unconstrained tree. By using the Shimodaira–Hasegawa test, the likelihood of this constrained tree was significantly different (i.e. it is noted in the ‘Topologies significantly different’ column of Table 2) or not significantly different (i.e. it is noted in the ‘Topologies not significantly different’ column of Table 2).

G+C content and codon usage. The G+C content at the third codon position (GC3) was determined, using MEGA v2.0 software, on the sequence sets corrected for phylogenetic analyses.

For the 7-sequence sets, codon usage was estimated using the codon adaptation index (CAI). CAI is a measure of the relative adaptation of the gene codon usage towards the codon usage of highly expressed genes for that organism (Sharp & Li, 1987). To estimate this index, we used as a reference pool for each organism (i.e. a reference codon usage table) the 46 concatenated ribosomal genes, since it is impossible for us to determine which genes are highly expressed (especially on the unfinished genome of *P. fluorescens* SBW25). Highly expressed ribosomal protein genes are commonly used in genome analyses where experimental expression data are limited or unavailable (Lafay *et al.*, 2000; Palacios &

Wernegreen, 2002; Herbeck *et al.*, 2003). The EMBOS programs *cusp* and *cai* were used to estimate the reference codon usage table of a strain and the CAI of a gene, respectively (Rice *et al.*, 2000).

Synonymous and nonsynonymous rates of nucleotide substitution. The synonymous rate (K_s) and nonsynonymous rate (K_a) were calculated using the modified Nei–Gojobori method implemented in the MEGA v2.0 software (Zhang *et al.*, 1998). The transition to transversion ratio was fixed at 2 and the Jukes–Cantor correction was used to account for multiple substitutions at the same site. The transition to transversion ratios estimated from *oprF* and ribosomal genes, were 1.2 and 2.3, respectively. The net mean synonymous or nonsynonymous distance between groups of taxa (d_A) was calculated using MEGA v2.0 software and was given by $d_A = d_{XY} - (d_X + d_Y)/2$, where d_{XY} is the mean distance between groups X and Y (i.e. the arithmetic mean of all pairwise distances between taxa in the inter-group comparisons), and d_X and d_Y are the mean within-group distances.

RESULTS

Since species differentiation based on the sequence of 16S rDNA alone is in many cases not possible, especially in the genus *Pseudomonas*, we attempted to confirm the ‘species

Fig. 2. Topology constraints used for the test of incongruence. (a) Topologies tested for the 7-sequence sets (or 6-sequence sets for the *cobA* gene). Topology 4 corresponds to the expected species phylogeny (i.e. monophilies of species) and is a nested topology of 1, 2 or 3. Topology 8 corresponds to the *OprF* phylogeny and is a nested topology of 2, 5, 6 or 7. Topologies 9 and 10 are alternative topologies. (b) Topologies tested for the 20-sequence sets. All topologies assume monophilies of o-clusters or lineages (defined in Fig. 1a) and are nested topologies of topology I. Topology VIII corresponds to the expected species phylogeny and is a nested topology of IV, V, VI or VII. Topology III corresponds to the *OprF* phylogeny and is a nested topology of topology II. Topologies IX and X are alternative topologies.

Table 2. Tests of incongruence and other parameters

The Shimodaira–Hasegawa test (Shimodaira & Hasegawa, 1999) was used to determine whether constrained trees are significantly different (i.e. incongruent) from the unconstrained tree.

Set of genes	No. of sequences	Length of sequences (bp)	GC3 content (mol%)	CAI	Pairwise identities (%) between amino acids	Mean K_s (K_a/K_s)	Topology constraints tested*		
							Best topologies†	Topologies not significantly different ($P > 0.05$)	Topologies significantly different ($P < 0.05$)
<i>oprF</i>	7	786	64.5–79.0	0.82–0.85	56.9–93.9	0.45 (0.54)	2, 5, 6, 7, 8	1, 3, 4	9, 10
<i>oprF</i>	20	786	59.9–79.0	–	52.3–100	0.47 (0.51)	I, II, III	IV, V, VI, VII, VIII , X	IX
16S rDNA	7	1233	–	–	–	–	2, 3	1, 4 , 5, 6, 9	7, 8 , 10
16S rDNA	20	1233	–	–	–	–	–	I, IV, V, VI, VII, VIII	II, III , IX, X
Concatenated ribosomal genes	7	18045	58.3–71.9	0.74–0.76	89.7–99.6	0.31 (0.12)	2, 3	1, 4	5, 6, 7, 8 , 9, 10
<i>ppsA</i>	7	2292	78.4–88.1	0.70–0.76	90.6–98.6	0.48 (0.09)	2, 3	1, 4	5, 6, 7, 8 , 9, 10
<i>estX</i>	7	990	69.7–83.9	0.46–0.61	56.2–100	0.79 (0.40)	2	1, 2, 3, 4 , 5, 6, 7, 8 , 9, 10	–
<i>menG</i>	7	489	71.8–85.8	0.59–0.70	81.5–93.0	0.65 (0.13)	1, 2, 9, 10	3, 4 , 5, 6, 7, 8	–
<i>cmaX</i>	7	996	71.1–83.8	0.45–0.57	60.4–94.9	0.71 (0.27)	1, 2	3, 4	5, 6, 7, 8 , 9, 10
<i>crfX</i>	7	207	63.7–82.6	0.46–0.61	72.4–100.0	0.71 (0.16)	1, 2, 9, 10	3, 4 , 5, 6, 7, 8	–
<i>cmpX</i>	7	825	67.3–85.8	0.43–0.57	80.3–98.9	0.71 (0.11)	1, 2	3, 4	5, 6, 7, 8 , 9, 10
<i>sigX</i>	7	591	55.9–75.7	0.35–0.47	88.8–98.5	0.85 (0.06)	–	1, 2, 3, 4 , 5, 6, 7, 8 , 9, 10	–
7, 8 , 9, 10	–	–	–	–	–	–	–	–	–
Concatenated upstream genes	7	6390	72.4–85.4	0.55–0.61	78.8–96.9	0.63 (0.18)	1, 2, 3, 4	9, 10	5, 6, 7, 8
Downstream gene (<i>cobA</i>)	6	690	71.8–82.2	0.40–0.61	74.8–96.1	0.80 (0.18)	1, 2, 3, 4	5, 6, 7, 8 , 9, 10	–

*Topology constraints are described in Fig. 2 and Methods. Topologies in bold: topologies 4 and VIII correspond to the species phylogeny expected (i.e. monophyly of species), for the 7- and 20-sequence sets, respectively; topologies 8 and III correspond to the *OprF* phylogeny, for the 7- and 20-sequence sets, respectively.

†The constrained trees and the unconstrained tree had the same likelihood (i.e. the same topology).

phylogeny' (i.e. monophylies of all species) in this genus. We identified 46 ribosomal protein genes (termed ribosomal genes) that are ubiquitous in seven sequenced *Pseudomonas* genomes (Table 1) and that have no duplicated copies. The ribosomal sequences were aligned individually and concatenated, leading to 18 045 unambiguously aligned nucleotide positions. The distance tree (Kimura two-parameter correction) of our concatenated gene set and bootstrap values are shown in Fig. 1(c). Alternatively, an analysis with deduced protein sequences was conducted; bootstrap values are also reported on the tree. While the monophyly of *P. fluorescens* is only supported with nucleotide sequences, the monophylies of *P. syringae* and *P. syringae* plus *P. fluorescens* are supported by maximum bootstrap values (100 %) with both protein and nucleotide sequences.

Similar analyses with the *oprF* gene, however, led to a different topology. Two *P. fluorescens* strains (SBW25 and Pf-5) are separated from the third *P. fluorescens* strain (Pf0-1) and the two *P. syringae* strains by a node with a bootstrap value of 99 % (Fig. 1b). Analyses on a broader set of sequences (Table 1) were also performed and led to similar results, with the *P. fluorescens* strains splitting into two separate groups, previously termed *fluorescens* 1 o-cluster and *fluorescens* 2 o-cluster by Bodilis *et al.* (2004) (Fig. 1a). In addition, two putida o-clusters, also previously described, were found. The separation of the two *fluorescens* o-clusters by two nodes is not strongly supported (< 90 %), especially in nucleotide analysis.

Test of incongruence

To evaluate the incongruence between *oprF* and species phylogeny, we conducted a maximum-likelihood (ML) analysis using the GTR (general time reverse) model of evolution with an among-site rate heterogeneity approximated by a discrete gamma distribution and a proportion of invariant sites (see Methods). Different ML searches for the best tree were performed with or without constraint topologies (Fig. 2, Table 2). For each set of sequences, the resulting trees were compared, using the Shimodaira-Hasegawa test, to determine whether constrained trees were significantly different from the unconstrained tree (Shimodaira & Hasegawa, 1999). If a constrained tree had the same likelihood as the unconstrained tree, we considered that the monophylies enforced by the topology constraint were present in the unconstrained tree (see Methods).

For concatenated ribosomal genes, since the unconstrained tree showed monophylies of *P. syringae* and *P. syringae* plus *P. fluorescens* (topology constraints 2 and 3), the *OprF* topology (topology constraint 8) was very significantly rejected ($P < 0.001$), as were its intermediate topologies (5, 6 and 7). Although not significantly rejected ($P = 0.376$), the monophyly of *P. fluorescens* (topology constraint 1) was not found on the unconstrained tree (Table 2).

For the 16S rDNA gene (7- and 20-sequence sets), the unconstrained tree also showed monophylies of *P. syringae*

and *P. syringae* plus *P. fluorescens*, in addition to being significantly inconsistent with *OprF* topology ($P < 0.05$). For the 20-sequence set, since constrained topologies assume a monophyly for all o-clusters and lineages (defined in Fig. 1), the topology VII (monophyly of *P. syringae* plus *P. fluorescens*) is not the best topology. For the two sets of sequences, the monophyly of *P. fluorescens* was not significantly rejected ($P = 0.213$ and $P = 0.109$ for the 7- and 20-sequence sets, respectively).

As expected, through similar ML analyses performed on the *oprF* gene with 7- or 20-sequence sets, we obtained the *OprF* topology (constrained topology 8 or III, respectively). However, for the two sets of sequences, the species topology (4 or VIII) is not significantly rejected ($P = 0.071$ and $P = 0.163$ for the set of 7 and 20 sequences, respectively). Therefore, there is no significant incongruence between the *OprF* and species phylogenies.

Synonymous codon usage

Alternative synonymous codons are generally not used with the same frequency. In *E. coli*, synonymous codons that are recognized more efficiently and accurately by the most abundant tRNA are preferred, especially in highly expressed genes (Sharp & Li, 1987). The strength of preference (i.e. bias of codon usage) is expected to be correlated with the level of gene expression. Moreover, a weak bias of codon usage may result from either a low expression of the gene, or a recent horizontal transfer from an organism that has a different synonymous codon usage. Since the *OprF* protein is the major surface protein of *Pseudomonas*, the *oprF* gene should be highly expressed. For this gene, the bias of codon usage is expected to be high, unless the gene was recently transferred.

To estimate deviations in codon usage, the codon adaptation index (CAI) was calculated for *oprF* gene. CAI is a measure of the relative adaptiveness of the codon usage of a gene towards the codon usage of highly expressed genes for that organism: the higher the index value, the greater the codon usage bias (Sharp & Li, 1987). As a reference for highly expressed genes, we used the 46 concatenated ribosomal genes for each organism (see Methods).

The genes of ribosomal proteins had high a CAI value (from 0.74 in strain PAO1, to 0.76 in strain KT2440) but CAI values for the *oprF* gene were even higher (from 0.82 in strain Pf-5, to 0.85 in strain KT2440) (Table 2). Therefore, there is a strong codon usage bias in the *oprF* gene, as expected for a highly expressed and not recently transferred gene. However, the absence of a recent horizontal transfer event cannot be deduced from this result: when the CAI value was calculated for the *oprF* gene of each strain, using the highly expressed genes of the other six *Pseudomonas* species as reference, the index value remained high (from 0.79 to 0.85), except when the reference genes from the PAO1 strain were used (from 0.61 to 0.74). These results suggest that, for the *oprF* gene, the CAI index is not sensitive

Fig. 3. Comparisons of evolutionary distance, difference in GC3 content and synonymous distance between concatenated ribosomal and *oprF* genes. Pairwise comparisons were performed from the 7-sequence sets. For each graph, the 21 plots correspond to the 21 different pairwise comparisons from 7 strains. (a) Evolutionary distance (Kimura two-parameter model, K2P) as a function of difference in GC3 content between concatenated ribosomal genes. (b) Difference in GC3 content between *oprF* genes as a function of difference in GC3 content between concatenated ribosomal genes. (c) Evolutionary distance (Kimura two-parameter model, K2P) as a function of synonymous distance (modified Nei–Gojobori model) between concatenated ribosomal genes. (d) Synonymous distance between *oprF* genes as a function of synonymous distance between concatenated ribosomal genes.

enough to highlight any differences between synonymous codon usage in *P. putida*, *P. syringae* and *P. fluorescens*.

GC3 composition (GC content at the third codon position)

Bacteria vary widely in their genomic G+C content, especially at third positions of codons, suggesting variable mutational biases among organisms (Muto & Osawa, 1987). In the same way as with codon usage bias, this feature can be used to detect a recent horizontal transfer from an organism that has a different GC3 content.

The GC3 contents were variable between sequences of concatenated ribosomal genes (from 58.3 mol% in strain SBW25, to 71.9 mol% in strain PAO1) (Table 2). Interestingly, evolutionary distance is strongly linked to difference in GC3

content in *Pseudomonas* (Fig. 3a). Moreover, while some GC3 contents were higher in *oprF* genes (from 59.9 mol% in strain MFY143, to 79 mol% in PAO1), the correlation between difference in GC3 content between concatenated ribosomal genes and difference in GC3 content between *oprF* genes is quite good ($r^2=0.82$) (Fig. 3b), suggesting an absence of recent horizontal transfer between distant *Pseudomonas* organisms. In fact, the GC3 values shown in Table 3 were quite informative. The GC3 value ranged from 59.9 to 71.0 mol% in the *fluorescens* 1 and 2 o-clusters, and from 72.5 to 78.3 mol% in the *putida* 1 and 2 o-clusters.

Genomic context

Since genes that diverged by vertical evolution should be found at homologous positions in closely related organisms, it was interesting to observe the genomic context of *oprF* in

Table 3. Synonymous and nonsynonymous rates and other parameters in the *oprF* gene

o-cluster(s)*	No. of sequences	GC3 content (mol%)	Pairwise identities (%) between amino acids	Mean K_s (K_a/K_s) within o-clusters	Net mean K_s (K_a/K_s) between o-clusters†			
					aeruginosa	fluorescens 1	fluorescens 2	putida 1
All o-clusters	20	59.9–79.0	52.3–100.0	0.47 (0.51)	–	–	–	–
aeruginosa	4	72.1–79.0	71.0–80.2	0.46 (0.44)	–	–	–	–
fluorescens 1	6	65.3–67.9	86.3–95.8	0.25 (0.28)	0.22 (0.73)	–	–	–
fluorescens 2	4	59.9–67.2	71.0–99.6	0.41 (0.37)	0.25 (0.87)	0.16 (1.29)	–	–
putida 1	3	72.5–73.6	98.1–100.0	0.08 (0.12)	0.21 (0.90)	0.25 (0.31)	0.28 (0.86)	–
putida 2	3	75.2–77.5	97.7–100.0	0.11 (0.07)	0.19 (1.08)	0.29 (0.50)	0.26 (0.76)	0.21 (0.64)
All o-clusters	69	59.9–79.4	51.9–100.0	0.43 (0.53)	–	–	–	–
aeruginosa	8	69.9–79.4	71.0–100.0	0.45 (0.35)	–	–	–	–
fluorescens 1	31	63.4–68.7	83.6–100.0	0.19 (0.28)	0.25 (0.81)	–	–	–
fluorescens 2	21	59.9–71.0	67.2–100.0	0.34 (0.37)	0.30 (0.87)	0.22 (1.05)	–	–
putida 1	4	72.5–73.6	95.4–100.0	0.14 (0.12)	0.20 (1.08)	0.22 (0.37)	0.29 (0.89)	–
putida 2	5	74.4–78.3	96.1–100.0	0.11 (0.10)	0.22 (1.08)	0.27 (0.59)	0.30 (0.72)	0.18 (0.73)

*Clusters in bold correspond to the 20-sequence set.

†Numbers in bold correspond to K_a/K_s values higher than 1.

the seven sequenced *Pseudomonas* genomes. The regions upstream (about 7 kb) and downstream (about 14 kb) of the *oprF* gene are schematically represented in Fig. 4.

The *oprF* gene has a similar position on the leading strand, at about 2 Mb from the origin of replication, for six genomes (the *oprF* position is still unknown on the unfinished SBW25 genome). While some deletions or insertions are detected downstream from the *oprF* gene, the genes upstream are conserved at identical homologous positions. Moreover, the position of the gene just downstream from the *oprF* gene (*cobA*) is conserved on all genomes, except in the Pf-5 genome, where a region of about 6–7 kb, between the *oprF* and *pyrD* genes, is deleted (Fig. 4). Analyses of incongruence were also performed with the genes upstream and downstream from *oprF* (Table 2). The unconstrained tree showed monophyly of *P. fluorescens* or *P. syringae* plus *P. fluorescens* (topology constraints 1 or 3), for *cobA* and for five out of the seven genes upstream from *oprF*. There is even a significant difference from OprF topology (topology constraint 8) with three genes upstream from *oprF* (*ppsA*, *cmaX* and *cmpX*). Moreover, with the concatenation of the seven upstream genes, the OprF topology (topology constraint 8) was very significantly rejected ($P < 0.001$), as well as intermediate topologies (5, 6 and 7), while the species topology (4) was the best.

Search for an adaptive selection

In comparison with upstream and downstream sequences, pairwise distances between protein sequences showed quite high divergence in the OprF protein, e.g. only 56.9% identity between the protein sequences of strains PAO1 and SBW25 (Table 2). In fact, the real divergence is even higher, since the most variable positions of the alignment were removed for analyses (see Methods). Even so, two genes (*estX* and *cmaX*) have deduced protein sequences with a

similar divergence to that of *oprF*. Since such a divergence in the *oprF* gene suggests an unusual evolutionary history, we attempted to detect positive selection through comparison of synonymous (K_s) and nonsynonymous (K_a) substitution rates.

Since saturation of synonymous sites is an important problem in the analysis of selective constraints when nucleotide sequences are compared, we initially analysed the correlation between the evolutionary distance and the synonymous distance on the concatenated ribosomal genes. If synonymous sites are saturated, we might expect a curve with a plateau. In fact, since we obtained a very strong linear correlation ($r^2 = 0.99$), no saturation exists at synonymous positions between the ribosomal genes of *Pseudomonas* (Fig. 3c). Moreover, although the regression coefficient is quite weak ($r^2 = 0.68$), the correlation between synonymous distances of *oprF* and ribosomal genes clearly showed that there was no saturation. We therefore conclude that K_s and K_a can be compared to infer modification of evolutionary constraints in the *oprF* gene.

While K_s is quite variable, even in the same genomic context as the *oprF* gene (from 0.45 in the *oprF* gene, to 0.85 in the *sigX* gene), the K_a/K_s ratio is even more variable (from 0.06 in the *sigX* gene to 0.54 in the *oprF* gene), suggesting variable evolutionary constraints between genes (Table 2). As expected, the two genes that have similar divergence to *oprF* in their deduced protein sequences (the *estX* and *cmaX* genes), also have a relatively high K_a/K_s ratio (0.40 and 0.27, respectively). Although the K_s value of *oprF* is slightly less than the K_s value of those two genes, the K_a value of the three genes is almost the same (but that does not suggest identical evolutionary constraints).

Since a K_a/K_s ratio greater than 1 is considered to be strong evidence of positive selection for amino acid replacements,

Fig. 4. Schematic representation of the genomic region around the *oprF* gene in *Pseudomonas*. ORFs localized above of the line are in the leading strand. This representation is arbitrary for SBW25, since the genome is unfinished. To simplify the representation, the names of ORFs were standardized to names from the annotation of the PAO1 genome.

a ratio less than 1 is generally taken as evidence that the proteins have evolved under negative or purifying selection. Significantly elevated K_a/K_s ratios have been shown for very few whole proteins (Endo *et al.*, 1996). However, K_s and K_a values are computed as an average over both amino acid sites and time. This approach is very conservative, since it

does not detect either episodic positive selections or positive selections that only occur for some codons (Messier & Stewart, 1997; Yang *et al.*, 2000; Yang & Nielsen, 2002). In order to check whether the *oprF* gene had caused episodes of positive selection on some lineages, we estimated

the K_a/K_s ratio in the different o-clusters. Interestingly, while all K_a/K_s values calculated from within-cluster comparisons are lower than those calculated from overall comparisons, between-cluster comparisons show higher ratios, some of them being greater than 1 (Table 3). Because the higher ratio is too near to 1 (1.29; for the comparison between fluorescens o-clusters 1 and 2), we cannot conclude that positive selection occurred; relaxed evolutionary constraints would probably lead to a similar result. Even so, our results clearly show that selective constraints between o-clusters, especially between fluorescens o-clusters, have varied over time.

DISCUSSION

We previously showed a difference in tree topologies between the *OprF* and 16S rDNA phylogenies from 69 *Pseudomonas* isolates or reference strains (Bodilis *et al.*, 2004). The differences corresponded essentially to a separation of the cluster containing *P. fluorescens* strains and the one containing *P. putida* strains, while no horizontal transfer was observed between clusters containing *P. fluorescens* and *P. putida* strains (termed o-clusters). To explain this phylogeny, we proposed three hypotheses: (i) an ancient horizontal transfer (after the *P. fluorescens*–*P. putida* speciation), (ii) an ancient duplication (before the speciation) followed by several losses, or (iii) an artifact in phylogeny reconstruction. These different possibilities have been investigated in this study.

The gene encoding 16S rDNA has been widely used as a molecular sequence to reconstruct species phylogenies because it was assumed that intraspecific variation and horizontal transfer was low for this gene (Woese, 1987). However, in some groups of bacteria, particularly in the genus *Pseudomonas*, the inferred phylogenies based on 16S rDNA lack resolution at the intrageneric level due a slow rate of evolution (Moore *et al.*, 1996; Anzai *et al.*, 2000). Concatenation of 46 genes encoding ribosomal proteins provided a very robust systematic analysis of the relationships between *P. aeruginosa*, *P. putida*, *P. syringae* and *P. fluorescens* (Fig. 1). Monophylies of *P. syringae* and *P. syringae* plus *P. fluorescens*, which are supported by maximum bootstrap values with both protein and nucleotide sequences, were also supported by high bootstrap values in studies based on other genes: *gyrB-rpoD* concatenation, *atpD-carA-recA*-16S rDNA concatenation and the *rpoB* gene (Yamamoto *et al.*, 2000; Hilario *et al.*, 2004; Ait Tayeb *et al.*, 2005). Robust tests of incongruence have confirmed this topology on concatenated ribosomal genes and on the 16S rDNA gene. On the other hand, the monophyly of *P. fluorescens* is only supported with nucleotide sequences (but with a maximum value) and not found by ML analysis, although it was not significantly rejected. By contrast, since the distance tree of the *oprF* gene (or *OprF* protein) shows a separation of *P. fluorescens* by a strongly supported node (99%), ML analysis does not significantly support the incongruence between this topology and the species

topology. Therefore, an artifact in the reconstruction of *oprF* phylogeny cannot be rejected.

To search for clues about horizontal transfer or duplication, we investigated the synonymous codon usage, GC3 content and genomic context of the *oprF* gene. Altogether, these features clearly showed an absence of ancient duplication or recent horizontal transfer. Ancient horizontal transfer followed by modification of the GC3 content towards the new genomic features cannot be formally excluded but it is unlikely because the chromosome context around the *oprF* gene is conserved and only a double crossing-over could explain a horizontal transfer of the *oprF* gene. In strain Pf-5, an interesting deletion, just downstream of *oprF*, could correspond to a replacement of the gene, although this feature is not present in the second strain of the fluorescens 2 o-cluster (SBW25 strain). This conservation of the genomic context has even already been shown in additional *Pseudomonas* strains, including a third strain of the fluorescens 2 o-cluster, strain OE28.3 (Brinkman *et al.*, 1999). In this study, Brinkman *et al.* (1999) showed an influence of the *sigX* gene (a putative ECF sigma factor gene) on expression regulation of the *oprF* gene. Moreover, it is unlikely that the GC3 of *oprF* gene has greatly changed after horizontal transfer because there is no saturation at synonymous positions in the *oprF* gene and there is similar synonymous usage in *Pseudomonas* (except in strain PAO1). Finally, the K_s values between fluorescens o-clusters were closer or identical (for the 7- and 20-sequence sets, respectively) than between fluorescens and putida o-clusters (Table 3). This feature is well illustrated by trees built from synonymous distances (Fig. 5). Although the species topology was not found with the *oprF* gene, there was a monophyly of *P. fluorescens* plus *P. syringae*, well supported with the 20-sequence set (Fig. 5a). For the concatenated ribosomal genes, a phylogeny with synonymous distance shows a monophyly of *P. fluorescens* supported by a maximum bootstrap value (Fig. 5c); the same is true for the phylogeny with evolutionary distance (with nucleotide sequences) (Fig. 1c). The *P. fluorescens* intracluster resolution was even slightly better with synonymous distance than with evolutionary distance.

We have also shown important modifications in evolutionary constraints during the history of the *oprF* gene, since between-cluster K_a/K_s ratios were high, especially between fluorescens o-clusters and between aeruginosa and putida o-clusters (Table 3). Because the higher ratio is too near to 1, we cannot exclude that only relaxed evolutionary constraints had occurred. There now exist more sensitive ML tests for adaptive evolution that investigate individual codons in a gene and/or individual branches in a phylogeny (Yang *et al.*, 2000; Yang & Nielsen, 2002). However, those tests require a known topology of the phylogenetic tree, while the observed *oprF* phylogeny could result from an artifact.

In fact, modifications in evolutionary constraints were evident because the central region of the protein contains a

Fig. 5. Phylogenetic relationships among the 20-sequence set of *oprF* genes (a), the 7-sequence set of *oprF* genes (b) and the 7-sequence set of 46 concatenated ribosomal genes (c) from *Pseudomonas* species. The unrooted dendrograms were generated, using the neighbour-joining algorithm, from synonymous distances computed according to the modified Nei–Gojobori model (Zhang *et al.*, 1998). Numbers on tree branches (above or below) report bootstrap results obtained for 1000 repetitions. Only bootstrap values $\geq 50\%$ are reported. The horizontal length of branches is proportional to the estimated number of substitutions.

much-conserved surface-exposed motif, in all *Pseudomonas* species except in the fluorescens 2 o-cluster and in two species of the aeruginosa o-cluster (*P. mendocina* and *P. stutzeri*) (Vermeiren *et al.*, 1999; Bodilis *et al.*, 2004). This motif has four cysteine residues probably implicated in the formation of two disulfide bonds. According to the tree

topology (*OprF* or species topologies), we think that the absence of this cysteine-rich motif results from at least two independent deletions. The role of this motif is unknown, although similarities with the calcium-binding repeats of the thrombospondin protein (a eukaryotic extracellular matrix protein) have previously been shown (de Mot &

Vanderleyden, 1994). Interestingly, there is a proline-rich variable region situated next to the conserved motif, on the N-terminal side; this region is slightly longer when the conserved motif is absent. Since all the central region of *oprF* was removed in this study, modifications in evolutionary constraints must also have occurred in other regions of the gene.

Altogether, these results support the hypothesis of trans-species polymorphism (i.e. presence of alleles in different species that are more similar to each other than they are to alleles in the same species) in the *oprF* gene.

In a previous study, we suggested that the polymorphism of OprF could be correlated with the ecological niche, since the fluorescens 1 o-cluster contained most of the bulk soil (non-rhizospheric) isolates and the fluorescens 2 o-cluster contained all the clinical isolates and most of the rhizospheric isolates (Bodilis *et al.*, 2004). In order to estimate the proportion of *oprF* type 2 (in the cultivable and uncultivable fractions) from environmental DNA, we have developed a ratio PCR (J. Bodilis and others, unpublished). By using this molecular tool, we found that the proportion of *oprF* type 2 (without the cysteine-rich motif) from cultivated fields (wheat or flax) was significantly different between the rhizosphere samples and the adjacent bulk soil samples (J. Bodilis and others, unpublished). The physiological properties that are linked with *oprF* type 2 fitness are still unknown, but could concern adhesion (Bodilis *et al.*, 2004) or resistance to a variation of osmotic pressure, although the cysteine-rich motif is not essential for channel size modulations and pore features, as shown by Jaouen *et al.* (2004). Since OprF is known to be a pleiotropic protein, modifications in evolutionary constraints could result from variations in cryptic functions. More ecological studies and experiments with mutants in gnotobiotic systems should further improve our knowledge of the multifunction of OprF.

Finally, it was interesting to find OprF without the cysteine-rich motif in the two more divergent o-clusters (*aeruginosa* and *fluorescens* 2); this divergence cannot be explained by the simultaneous presence of the two OprF types in the *fluorescens* 2 o-cluster (Table 3). Nonsynonymous mutations must therefore have accelerated in these two evolutionary lineages though relaxed constraints and/or though episodic positive evolution. Either of these phenomena could have led to a phylogeny reconstruction artifact, by attracting the two longest branches of the tree towards each other.

ACKNOWLEDGEMENTS

The authors are grateful to Philippe Lopez for early critical comments and to Dilys Moscato for language assistance.

REFERENCES

- Aagot, N., Nybroe, O., Nielsen, P. & Johnsen, K. (2001). An altered *Pseudomonas* diversity is recovered from soil by using nutrient-poor *Pseudomonas*-selective soil extract media. *Appl Environ Microbiol* **67**, 5233–5239.
- Ait Tayeb, L., Ageron, E., Grimont, F. & Grimont, P. A. D. (2005). Molecular phylogeny of the genus *Pseudomonas* based on *rpoB* sequences and application for the identification of isolates. *Res Microbiol* **156**, 763–773.
- Anzai, Y., Kim, H., Park, J., Wakabayashi, H. & Oyaizu, H. (2000). Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *Int J Syst Evol Microbiol* **50**, 1563–1589.
- Bellido, F., Martin, N., Siehnel, R. & Hancock, R. (1992). Reevaluation, using intact cells, of the exclusion limit and role of porin OprF in *Pseudomonas aeruginosa* outer membrane permeability. *J Bacteriol* **174**, 5196–5203.
- Bodilis, J., Calbrix, R., Guerillon, J., Merieau, A., Pawlak, B., Orange, N. & Barry, S. (2004). Phylogenetic relationships between environmental and clinical isolates of *Pseudomonas fluorescens* and related species deduced from 16S rRNA gene and OprF protein sequences. *Syst Appl Microbiol* **27**, 93–108.
- Brinkman, F. S., Schoofs, G., Hancock, R. E. & de Mot, R. (1999). Influence of a putative ECF sigma factor on expression of the major outer membrane protein, OprF, in *Pseudomonas aeruginosa* and *Pseudomonas fluorescens*. *J Bacteriol* **181**, 4746–4754.
- Brinkman, F. S., Bains, M. & Hancock, R. E. (2000). The amino terminus of *Pseudomonas aeruginosa* outer membrane protein OprF forms channels in lipid bilayer membranes: correlation with a three-dimensional model. *J Bacteriol* **182**, 5251–5255.
- Dé, E., de Mot, R., Orange, N., Saint, N. & Molle, G. (1995). Channel-forming properties and structural homology of major outer membrane proteins from *Pseudomonas fluorescens* MFO and OE 28.3. *FEMS Microbiol Let* **127**, 267–272.
- Dé, E., Orange, N., Saint, N., Guerillon, J., de Mot, R. & Molle, G. (1997). Growth temperature dependence of channel size of the major outer-membrane protein (OprF) in psychrotrophic *Pseudomonas fluorescens* strains. *Microbiology* **143**, 1029–1035.
- de Mot, R. & Vanderleyden, J. (1994). A conserved surface-exposed domain in major outer membrane proteins of pathogenic *Pseudomonas* and *Branhamella* species shares sequence homology with the calcium-binding repeats of the eukaryotic extracellular matrix protein thrombospondin. *Mol Microbiol* **13**, 379–380.
- de Mot, R., Proost, P., van Damme, J. & Vanderleyden, J. (1992). Homology of the root adhesin of *Pseudomonas fluorescens* OE 28.3 with porin F of *P. aeruginosa* and *P. syringae*. *Mol Gen Genet* **231**, 489–493.
- de Mot, R., Schoofs, G., Roelandt, A., Declerck, P., Proost, P., Van Damme, J. & Vanderleyden, J. (1994). Molecular characterization of the major outer-membrane protein OprF from plant root-colonizing *Pseudomonas fluorescens*. *Microbiology* **140**, 1377–1387.
- Endo, T., Ikeo, K. & Gojobori, T. (1996). Large-scale search for genes on which positive selection may operate. *Mol Biol Evol* **13**, 685–690.
- Herbeck, J. T., Wall, D. P. & Wernegreen, J. J. (2003). Gene expression level influences amino acid usage, but not codon usage, in the tsetse fly endosymbiont *Wigglesworthia*. *Microbiology* **149**, 2585–2596.
- Hilario, E., Buckley, T. & Young, J. (2004). Improved resolution on the phylogenetic relationships among *Pseudomonas* by the combined analysis of *atpD*, *carA*, *recA* and 16S rDNA. *Antonie van Leeuwenhoek* **86**, 51–64.
- Jaouen, T., De, E., Chevalier, S. & Orange, N. (2004). Pore size dependence on growth temperature is a common characteristic of the major outer membrane protein OprF in psychrotrophic and mesophilic *Pseudomonas* species. *Appl Environ Microbiol* **70**, 6665–6669.

- Jukes, T. & Cantor, C. (1969). *Evolution of Protein Molecules. Mammalian Protein Metabolism*. New York: Academic Press.
- Kerstens, K., Ludwig, W., Vancanneyt, M., de Vos, P., Gillis, M. & Schleifer, K.-H. (1996). Recent changes in the classification of the pseudomonads: an overview. *Syst Appl Microbiol* **19**, 465–477.
- Kimura, M. (1980). A simple method for estimation of evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.
- Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.
- Lafay, B., Atherton, J. C. & Sharp, P. M. (2000). Absence of translationally selected synonymous codon usage bias in *Helicobacter pylori*. *Microbiology* **146**, 851–860.
- Messier, W. & Stewart, C.-B. (1997). Episodic adaptative evolution of primate lysozymes. *Nature* **385**, 151–154.
- Moore, E. R. B., Mau, M., Arnscheidt, A., Böttger, E., Hutson, R., Collins, M., Van de Peer, Y., Wachter, R. & Timmis, K. N. (1996). The determination and comparison of the 16S rRNA gene sequences of species of the genus *Pseudomonas* (*sensu stricto*) and estimation of the natural intragenetic relationships. *Syst Appl Microbiol* **19**, 478–492.
- Muto, A. & Osawa, S. (1987). The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci U S A* **84**, 166–169.
- Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* **67**, 593–656.
- Orange, N. (1994). Growth temperature regulates the induction of beta-lactamase in *Pseudomonas fluorescens* through modulation of the outer membrane permeation of a beta-lactam-inducing antibiotic. *Microbiology* **140**, 3125–3130.
- Palacios, C. & Wernegreen, J. J. (2002). A strong effect of AT mutational bias on amino acid usage in *Buchnera* is mitigated at high-expression genes. *Mol Biol Evol* **19**, 1575–1584.
- Pautsch, A. & Schulz, G. (1998). Structure of the outer membrane protein A transmembrane domain. *Nat Struct Biol* **5**, 1013–1017.
- Ramette, A., Moenne-Loccoz, Y. & Defago, G. (2001). Polymorphism of the polyketide synthase gene *phlD* in biocontrol fluorescent pseudomonads producing 2,4-diacetylphloroglucinol and comparison of PhlD with plant polyketide synthases. *Mol Plant Microbe Interact* **14**, 639–652.
- Rawling, E., Martin, N. & Hancock, R. (1995). Epitope mapping of the *Pseudomonas aeruginosa* major outer membrane porin protein OprF. *Infect Immun* **63**, 38–42.
- Rawling, E. G., Brinkman, F. S. & Hancock, R. E. (1998). Roles of the carboxy-terminal half of *Pseudomonas aeruginosa* major outer membrane protein OprF in cell shape, growth in low-osmolarity medium, and peptidoglycan association. *J Bacteriol* **180**, 3556–3562.
- Rebière-Huët, J., Di Martino, P., Gallet, O. & Hulen, C. (1999). Interaction of *Pseudomonas aeruginosa* outer membrane proteins with plasmatic fibronectin. A root for new bacterial adhesins. *C R Acad Sci Paris* **322**, 1071–1080.
- Rediers, H., Vanderleyden, J. & de Mot, R. (2004). *Azotobacter vinelandii*: a *Pseudomonas* in disguise? *Microbiology* **150**, 1117–1119.
- Rice, P., Longden, I. & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276–277.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic tree. *Mol Biol Evol* **4**, 406–425.
- Sharp, P. & Li, W. (1987). The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**, 1281–1295.
- Shimodaira, H. & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**, 1114–1116.
- Tamper, S. & Hancock, R. (2004). The outer membranes of pseudomonads. In *Pseudomonas*, pp. 575–601. Edited by J. L. Ramos. New York: Plenum.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
- Vermeiren, H., Willems, A., Schoofs, G., De Mot, R., Keijers, V., Hai, W. & Vanderleyden, J. (1999). The rice inoculant strain *Alcaligenes faecalis* A15 is a nitrogen-fixing *Pseudomonas stutzeri*. *Syst Appl Microbiol* **22**, 215–224.
- von Specht, B., Knapp, B., Muth, G., Broker, M., Hungerer, K., Diehl, K., Massarrat, K., Seemann, A. & Domdey, H. (1995). Protection of immunocompromised mice against lethal infection with *Pseudomonas aeruginosa* by active or passive immunization with recombinant *P. aeruginosa* outer membrane protein F and outer membrane protein I fusion proteins. *Infect Immun* **63**, 1855–1862.
- Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221–271.
- Woodruff, W. & Hancock, R. (1989). *Pseudomonas aeruginosa* outer membrane protein F: structural role and relationship to the *Escherichia coli* OmpA protein. *J Bacteriol* **171**, 3304–3309.
- Worgall, S., Krause, A., Rivara, M. & 8 other authors (2005). Protection against *P. aeruginosa* with an adenovirus vector containing an OprF epitope in the capsid. *J Clin Invest* **115**, 1281–1289.
- Wu, L., Estrada, O., Zaborina, O. & 12 other authors (2005). Recognition of host immune activation by *Pseudomonas aeruginosa*. *Science* **309**, 774–777.
- Yamamoto, S., Kasai, H., Arnold, D. L., Jackson, R. W., Vivian, A. & Harayama, S. (2000). Phylogeny of the genus *Pseudomonas*: intragenetic structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiology* **146**, 2385–2394.
- Yang, Z. & Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* **19**, 908–917.
- Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A.-M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449.
- Zhang, J., Rosenberg, H. F. & Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A* **95**, 3708–3713.