Phylogeny of the *Enterobacteriaceae* based on genes encoding elongation factor Tu and F-ATPase β -subunit

Sonia Paradis,^{1,2,4} Maurice Boissinot,^{1,2} Nancy Paquette,⁴ Simon D. Bélanger,¹ Eric A. Martel,¹ Dominique K. Boudreau,¹ François J. Picard,¹ Marc Ouellette,^{1,2} Paul H. Roy^{1,3} and Michel G. Bergeron^{1,2}

Correspondence

Michel G. Bergeron
Michel.G.Bergeron@
crchul.ulaval.ca

¹Centre de recherche en infectiologie de l'Université Laval, Centre hospitalier universitaire de Québec (pavillon CHUL), Sainte-Foy, Québec, Canada G1V 4G2

^{2,3}Division de microbiologie, faculté de Médecine² and département de biochimie et microbiologie, faculté des Sciences et Génie³, Université Laval, Sainte-Foy, Québec, Canada G1K 7P4

⁴Infectio Diagnostic (I.D.I.) Inc., Sainte-Foy, Québec, Canada G1V 2K8

The phylogeny of enterobacterial species commonly found in clinical samples was analysed by comparing partial sequences of their elongation factor Tu gene (*tuf*) and of their F-ATPase β -subunit gene (*atpD*). An 884 bp fragment for *tuf* and an 884 or 871 bp fragment for *atpD* were sequenced for 96 strains representing 78 species from 31 enterobacterial genera. The *atpD* sequence analysis exhibited an indel specific to *Pantoea* and *Tatumella* species, showing, for the first time, a tight phylogenetic affiliation between these two genera. Comprehensive *tuf* and *atpD* phylogenetic trees were constructed and are in agreement with each other. Monophyletic genera are *Cedecea*, *Edwardsiella*, *Proteus*, *Providencia*, *Salmonella*, *Serratia*, *Raoultella* and *Yersinia*. Analogous trees based on 16S rRNA gene sequences available from databases were also reconstructed. The *tuf* and *atpD* phylogenies are in agreement with the 16S rRNA gene sequence analysis, and distance comparisons revealed that the *tuf* and *atpD* genes provide better discrimination for pairs of species belonging to the family *Enterobacteriaceae*. In conclusion, phylogeny based on *tuf* and *atpD* conserved genes allows discrimination between species of the *Enterobacteriaceae*.

INTRODUCTION

Members of the family *Enterobacteriaceae* are facultatively anaerobic, Gram-negative rods that are catalase-positive and oxidase-negative (Brenner, 1984). They are found in soil, water and plants, and also in animals ranging from insects to humans. Many enterobacteria are opportunistic pathogens. In fact, members of this family are responsible for about 50 % of nosocomial infections in the US (Brenner, 1984). Therefore, this family is of considerable clinical importance.

Published online ahead of print on 27 May 2005 as DOI 10.1099/ijms.0.63539-0.

The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA, *tuf* and *atpD* gene sequences obtained in this study are listed in Table 1.

Further trees based on *tuf*, *atpD* and 16S rRNA gene sequences, and scatterplots comparing pairwise distance between taxa, are available as supplementary figures in IJSEM Online.

The major classification studies on the family *Enterobacteriaceae* were based on phenotypic traits (Brenner *et al.*, 1980, 1999; Dickey & Zumoff, 1988; Farmer *et al.*, 1980, 1985a, b) such as biochemical reactions and physiological characteristics. However, phenotypically distinct strains may be closely related by genotypic criteria and may belong to the same genospecies (Bercovier *et al.*, 1980; Hartl & Dykhuizen, 1984). Also, phenotypically close strains (bio-groups) may belong to different genospecies, like *Klebsiella pneumoniae* and *Enterobacter aerogenes* (Brenner, 1984), for example. Consequently, identification and classification of certain species may be ambiguous with techniques based on phenotypic tests (Janda *et al.*, 1999; Kitch *et al.*, 1994; Sharma *et al.*, 1990).

More advances in the classification of members of the family *Enterobacteriaceae* have come from DNA–DNA hybridization studies (Brenner *et al.*, 1980, 1986, 1993; Farmer *et al.*, 1980, 1985a; Izard *et al.*, 1981; Steigerwalt *et al.*, 1976).

Furthermore, the phylogenetic significance of bacterial classification based on 16S rRNA gene sequences has been recognized by many workers (Stackebrandt & Goebel, 1994; Wayne *et al.*, 1987). However, members of the family *Enterobacteriaceae* have not been subjected to extensive phylogenetic analysis of the 16S rRNA gene (Spröer *et al.*, 1999). In fact, this gene was not thought to solve taxonomic problems concerning closely related species because of its very high degree of conservation (Brenner, 1992; Spröer *et al.*, 1999). Another drawback of the 16S rRNA gene is that it is found in several copies within the genome (seven in *Escherichia coli* and *Salmonella typhimurium*) (Hill & Harnish, 1981). Because of sequence divergence between the gene copies, direct sequencing of PCR products is seldom suitable for achieving a representative sequence (Cilia *et al.*, 1996; Hill & Harnish, 1981). Other genes, such as *gap* and *ompA* (Lawrence *et al.*, 1991), *rpoB* (Mollet *et al.*, 1997) and *infB* (Hedegaard *et al.*, 1999), have been used to resolve the phylogeny of enterobacteria. However, none of these studies covered an extensive number of species.

tuf and *atpD* are the genes encoding elongation factor Tu and the F-ATPase β -subunit, respectively. Elongation factor Tu is involved in peptide chain formation (Ludwig *et al.*, 1990). The two copies of the *tuf* gene (*tufA* and *tufB*) found in enterobacteria (Sela *et al.*, 1989) share high levels of identity (99%) in *Salmonella typhimurium* and in *Escherichia coli*. A recombination phenomenon could explain sequence homogenization between the two copies (Abdulkarim & Hughes, 1996; Grunberg-Manago, 1996). F-ATPase is present on the plasma membranes of eubacteria (Nelson & Taiz, 1989). It works mainly in ATP synthesis (Nelson & Taiz, 1989), and the β -subunit contains the catalytic site of the enzyme. Elongation factor Tu and F-ATPase have been highly conserved throughout evolution and show functional constancy (Amann *et al.*, 1988a; Ludwig *et al.*, 1990). Phylogenies based on protein sequences from elongation factor Tu and the F-ATPase β -subunit have shown good agreement with each other and with the rRNA gene sequence data (Ludwig *et al.*, 1993). These phylogenies were reconstructed, respectively, from 36 species belonging to 32 bacterial genera and from 29 species belonging to 27 bacterial genera.

We elected to sequence 884 bp fragments of *tuf* and *atpD* from 96 clinically relevant enterobacterial strains representing 78 species from 31 genera. These DNA sequences were used to create phylogenetic trees that were compared with 16S rRNA gene sequence trees generated using sequence data available in public databases. These trees revealed good agreement with each other and demonstrated the high resolution of *tuf* and *atpD* phylogenies at the species level.

METHODS

Bacterial strains and genomic material. All bacterial strains used in this study were obtained from the American Type Culture Collection (ATCC), Manassas, VA, USA, or the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ),

Braunschweig, Germany. Whenever possible, type strains were chosen. Identification of all strains was confirmed by classical biochemical tests using the automated MicroScan WalkAway-96 system equipped with a Negative BP Combo Panel Type 15 (Dade Behring Canada). Genomic DNA was purified using the G NOME DNA kit (Bio 101). Genomic DNA from *Yersinia pestis* was kindly provided by Dr Robert R. Brubaker of Michigan State University. The strains used in this study are described in Table 1.

PCR primers. The eubacterial *tuf* and *atpD* gene sequences available from public databases were analysed using the GCG package (version 8.0) (Accelrys). On the basis of multiple sequence alignments, two highly conserved regions were chosen for each gene, and PCR primers were derived from these regions with the help of OLIGO primer analysis software (version 5.0) (National Biosciences). A second 5' primer was designed to amplify *atpD* for a few enterobacteria in which it was difficult to amplify the gene with the first primer set. When required, the primers contained inosines or degeneracies to account for variable positions. Oligonucleotide primers were synthesized with a model 394 DNA/RNA synthesizer (PE Applied Biosystems). The PCR primers used in this study are listed in Table 2.

DNA sequencing. An 884 bp portion of the *tuf* gene and an 884 bp portion (or alternatively an 871 bp portion for a few enterobacterial strains) of the *atpD* gene were sequenced for all of the enterobacteria listed in Table 1. Amplifications were performed with 4 ng genomic DNA. The 40 μ l PCR mixtures used to generate PCR products for sequencing contained 1.0 μ M each primer, 200 μ M each dNTP (Pharmacia Biotech), 10 mM Tris/HCl (pH 9.0 at 25 °C), 50 mM KCl, 0.1% (w/v) Triton X-100, 2.5 mM MgCl₂, 0.05 mM BSA and 1.0 U *Taq* DNA polymerase (Promega) combined with TaqStart (Clontech Laboratories). The PCR mixtures were subjected to thermal cycling (3 min at 95 °C and then 35 cycles of 1 min at 95 °C, 1 min at 55 °C for *tuf* or 50 °C for *atpD*, and 1 min at 72 °C, with a 7 min final extension at 72 °C) using a PTC-200 DNA Engine thermocycler (MJ Research). PCR products of the predicted sizes were recovered from a methylene-blue-stained agarose gel as described previously (Ke *et al.*, 2000).

Both strands of the purified amplicons were sequenced using the ABI Prism BigDye Terminator cycle sequencing ready reaction kit (PE Applied Biosystems) on an automated DNA sequencer (model 377; PE Applied Biosystems). Amplicons from two independent PCR amplifications were sequenced for each strain to ensure the absence of sequencing errors attributable to nucleotide misincorporations by the *Taq* DNA polymerase. Sequence assembly was performed with the aid of SEQUENCHER 3.0 software (Gene Codes).

DNA sequences from 16S rRNA genes were obtained mostly from public databases. 16S rRNA gene sequences for *Escherichia fergusonii* and *Escherichia vulneris* were obtained using published primers (Lane, 1991). The strains used, and their descriptions, are shown in Table 1.

Phylogenetic and distance analysis. Multiple sequence alignments were performed using PileUp from the GCG package (version 10.0) and checked by eye with the editor SeqLab to edit sequences when necessary and to identify regions containing gaps, indels or ambiguities to be excluded from the phylogenetic analysis. *Haemophilus influenzae*, *Pasteurella multocida* subsp. *multocida*, *Shewanella putrefaciens* and *Vibrio cholerae* were used as an outgroup because they do not belong to the family *Enterobacteriaceae* but are phylogenetically close to that family. Bootstrap subsets (750 or 1000 sets) and phylogenetic trees were generated with the neighbour-joining algorithm from Dr David Swofford's PAUP (Phylogenetic Analysis Using Parsimony) software, versions 4.0b4a and 4.0b6 (Sinauer Associates). The distance model used was Kimura two-parameter (Kimura, 1980).

Table 1. Strains analysed

Strains used in this study for sequencing of partial *tuf*, *atpD* and 16S rRNA genes are listed. Strains used in other studies for sequencing of the 16S rRNA gene are also shown; strain numbers on the same row represent the same strain although strain numbers may vary in the publications.

Taxon	Strain used		GenBank/EMBL/DDBJ accession numbers		
	In this study	By others	16S rRNA gene	<i>atpD</i> gene	<i>tuf</i> gene
<i>Budvicia aquatica</i>	DSM 5075 ^T	DSM 5075 ^T	AJ233407	AX110912	AX111110
<i>Buttiauxella agrestis</i>	DSM 4586 ^T	DSM 4586 ^T	AJ233400	AX110913	AX111105
<i>Cedecea davisae</i>	DSM 4568 ^T			AX109523	AX109284
<i>Cedecea lapagei</i>	DSM 4587 ^T			AX109524	AX109286
<i>Cedecea neteri</i>	ATCC 33855 ^T			AX109525	AX109285
<i>Citrobacter amalonaticus</i>	ATCC 25405 ^T	CDC 9020-77 ^T	AF025370	AX109527	AX109291
<i>Citrobacter braakii</i>	ATCC 43162			AX109528	AX109292
<i>Citrobacter braakii</i>		CDC 080-58 ^T	AF025368		
<i>Citrobacter farmeri</i>	ATCC 51112 ^T	CDC 2991-81 ^T	AF025371	AX109530	AX109294
<i>Citrobacter freundii</i>	ATCC 8090 ^T	DSM 30039 ^T	AJ233408	AX109531	AX109295
<i>Citrobacter koseri</i>	ATCC 27156 ^T			AX109529	AX109293
<i>Citrobacter sedlakii</i>	ATCC 51115 ^T	CDC 4696-86 ^T	AF025364	AX109533	AX109296
<i>Citrobacter werkmanii</i>	ATCC 51114 ^T	CDC 0876-58 ^T	AF025373	AX109534	AX109297
<i>Citrobacter youngae</i>	ATCC 29935 ^T			AX109535	AX109298
<i>Edwardsiella hoshinae</i>	ATCC 33379 ^T			AX109544	AX109313
<i>Edwardsiella tarda</i>	DSM 30052 ^T			AX109545	AX109314
<i>Edwardsiella tarda</i>		CDC 4411-68	AF015259		
<i>Enterobacter aerogenes</i>	ATCC 13048 ^T	JCM 1235 ^T	AB004750	AX110938	AX109316
<i>Enterobacter amnigenus</i>	ATCC 33072 ^T	JCM 1237 ^T	AB004749	AX109548	AX109318
<i>Enterobacter asburiae</i>	ATCC 35953 ^T	JCM 6051 ^T	AB004744	AX109549	AX109319
<i>Enterobacter cancerogenus</i>	ATCC 35317 ^T			AX109550	AX109320
<i>Enterobacter cloacae</i>	ATCC 13047 ^T			AX109551	AX109321
<i>Enterobacter gergoviae</i>	ATCC 33028 ^T	JCM 1234 ^T	AB004748	AX109552	AX109322
<i>Enterobacter hormaechei</i>	ATCC 49162 ^T			AX109553	AX109323
<i>Enterobacter sakazakii</i>	ATCC 29544 ^T	JCM 1233 ^T	AB004746	AX109554	AX109324
<i>Erwinia amylovora</i>	ATCC 14976			AX110919	AX111147
<i>Erwinia amylovora</i>		DSM 30165 ^T	AJ233410		
<i>Escherichia coli</i> 1	ATCC 11775 ^T	ATCC 11775 ^T	X80725	AX110211	AX110241
<i>Escherichia coli</i> 2	ATCC 25922	ATCC 25922	X80724	AX110212	AX110242
<i>Escherichia coli</i> 3	ATCC 35401			AX110210	AX110239
<i>Escherichia coli</i> 4	ATCC 43895	ATCC 43895	Z83205	AX110209	AX110240
<i>Escherichia fergusonii</i>	ATCC 35469 ^T		AF530475	AX109562	AX109346
<i>Escherichia hermannii</i>	ATCC 33650 ^T			AX109563	AX109347
<i>Escherichia vulneris</i>	ATCC 33821 ^T		AF530476	AX109564	AX109348
<i>Ewingella americana</i>	ATCC 33852 ^T			AX109566	AX109351
<i>Ewingella americana</i>		NCPPB 3905	X88848		
<i>Hafnia alvei</i>	ATCC 13337 ^T	ATCC 13337 ^T	M59155	AX109578	AX109364
<i>Haemophilus influenzae</i>	ATCC 9833			AY134489	AY134483
<i>Haemophilus influenzae</i>		ATCC 33391 ^T	M35019		
<i>Klebsiella oxytoca</i>	ATCC 13182 ^T	ATCC 13182 ^T	U78183	AX110922	AX111106
<i>Klebsiella pneumoniae</i>					
subsp. <i>pneumoniae</i>	ATCC 13883 ^T	DSM 30104 ^T	AJ233420	AX109584	AX111528
subsp. <i>ozaenae</i>	ATCC 11296 ^T	ATCC 11296 ^T	Y17654	AX109580	AX109369
subsp. <i>rhinoscleromatis</i>	ATCC 13884 ^T			AX110012	AX109371
<i>Kluyvera ascorbata</i>	DSM 4611 ^T			AX109585	AX109372
<i>Kluyvera ascorbata</i>		ATCC 14236	Y07650		
<i>Kluyvera cryocrescens</i>	DSM 4588 ^T			AX109586	AX109373
<i>Kluyvera georgiana</i>	DSM 9409 ^T	DSM 9409 ^T		AX109587	AX109374
<i>Leclercia adecarboxylata</i>	ATCC 23216 ^T			AX109518	AX109377

Table 1. cont.

Taxon	Strain used		GenBank/EMBL/DDBJ accession numbers		
	In this study	By others	16S rRNA gene	<i>atpD</i> gene	<i>tuf</i> gene
<i>Leminorella grimontii</i>	DSM 5078 ^T	DSM 5078 ^T	AJ233421	AX109590	AX109380
<i>Moellerella wisconsinensis</i>	DSM 5076 ^T			AX109593	AX109386
<i>Morganella morganii</i>					
subsp. <i>morganii</i>	ATCC 25830 ^T			AX109596	AX109388
subsp. <i>sibonii</i>	ATCC 51206			AY134486	AY134480
<i>Obesumbacterium proteus</i>	DSM 2777 ^T	DSM 2777 ^T	AJ233422	AX110924	AX111109
<i>Pantoea agglomerans</i>	ATCC 27155 ^T	DSM 3493 ^T	AJ233423	AX109597	AX109401
<i>Pantoea agglomerans</i>	ATCC 27989			AX109547	AX109317
<i>Pantoea dispersa</i>	ATCC 14589 ^T			AX109598	AX109402
<i>Pasteurella multocida</i> subsp. <i>multocida</i>	NCTC 10322 ^T	NCTC 10322 ^T	M35018	AX109599	AX109403
<i>Plesiomonas shigelloides</i>	ATCC 14029 ^T	ATCC 14029 ^T	X74688	AX110926	AX111107
<i>Pragia fontium</i>	DSM 5563 ^T	DSM 5563 ^T	AJ233424	AX109600	AX109409
<i>Proteus hauseri</i>	ATCC 13315	DSM 30118	AJ233425	AX109602	AX109415
<i>Proteus mirabilis</i>	ATCC 25933			AX109601	AX110793
<i>Proteus penneri</i>	ATCC 33519 ^T			AX110026	AX109414
<i>Proteus vulgaris</i>	ATCC 6361			AY134488	AY134482
<i>Providencia alcalifaciens</i>	ATCC 9886 ^T			AX109603	AX109416
<i>Providencia rettgeri</i>	ATCC 29944 ^T			AY134487	AY134481
<i>Providencia rustigianii</i>	ATCC 33673 ^T			AX109605	AX109418
<i>Providencia stuartii</i>	ATCC 33672			AX109606	AX109419
<i>Rahnella aquatilis</i>	DSM 4594 ^T	DSM 4594 ^T	AJ233426	AX109608	AX109424
<i>Raoultella ornithinolytica</i>	DSM 7464 ^T			AY134485	AY134479
<i>Raoultella ornithinolytica</i>		CIP 103.364	U78182		
<i>Raoultella planticola</i>	ATCC 33531 ^T	JCM 7251 ^T	AB004755	AX109583	AX109368
<i>Salmonella bongori</i>	ATCC 43975 ^T			AY134484	AY134478
<i>Salmonella bongori</i>		JEO 4162	AF029226		
<i>Salmonella choleraesuis</i>					
subsp. <i>arizonae</i>	ATCC 13314 ^T			AX109609	AX109425
subsp. <i>choleraesuis</i>					
serovar Choleraesuis	ATCC 7001			AX109610	AX109426
serovar Enteritidis*	DSM 9898 ^T			AX110027	AX109998
serovar Enteritidis*		SE22	U90318		
serovar Paratyphi A	ATCC 9150			AX109614	AX109879
serovar Paratyphi B	ATCC 8759			AX109615	AX110000
serovar Typhi*	ATCC 10749			AX109617	AX109432
serovar Typhi*		ATCC 19430 ^T	Z47544		
serovar Typhimurium*	ATCC 14028			AX109618	AX111148
serovar Typhimurium*		ATCC 13311 ^T	X80681		
serovar Virchow	ATCC 51955			AX109619	AX110001
subsp. <i>diarizonae</i>	ATCC 43973 ^T			AX109611	AX109427
subsp. <i>houtenae</i>	DSM 9221 ^T			AX109612	AX109429
subsp. <i>indica</i>	ATCC 43976 ^T			AX109613	AX109430
subsp. <i>salamae</i>	DSM 9220 ^T			AX109616	AX109431
<i>Serratia ficaria</i>	DSM 4569 ^T	DSM 4569 ^T	AJ233428	AX109620	AX109880
<i>Serratia fonticola</i>	DSM 4576 ^T	DSM 4576 ^T	AJ233429	AX109621	AX109433
<i>Serratia grimesii</i>	DSM 30063 ^T	DSM 30063 ^T	AJ233430	AX109622	AX110002
<i>Serratia liquefaciens</i>	ATCC 27592 ^T			AX109623	AX109434
<i>Serratia marcescens</i>	ATCC 13880 ^T	DSM 30121 ^T	AJ233431	AX109624	AX109435
<i>Serratia odorifera</i>	ATCC 33077 ^T	DSM 4582 ^T	AJ233432	AX109625	AX109436
<i>Serratia plymuthica</i>	DSM 4540 ^T	DSM 4540 ^T	AJ233433	AX109626	AX109437
<i>Serratia rubidaea</i>	DSM 4480 ^T	DSM 4480 ^T	AJ233436	AX109627	AX109438
<i>Shewanella putrefaciens</i>	ATCC 8071 ^T	ATCC 8071 ^T	X82133	AX110927	AX111108
<i>Shigella boydii</i>	ATCC 9207	ATCC 9207	X96965	AX109629	AX109439

Table 1. cont.

Taxon	Strain used		GenBank/EMBL/DDBJ accession numbers		
	In this study	By others	16S rRNA gene	<i>atpD</i> gene	<i>tuf</i> gene
<i>Shigella dysenteriae</i>	ATCC 11835			AX109630	AX109440
<i>Shigella dysenteriae</i>		ATCC 13313 ^T	X96966		
<i>Shigella flexneri</i>	ATCC 12022	ATCC 12022	X96963	AX109631	AX109441
<i>Shigella sonnei</i>	ATCC 29930 ^T			AX109632	AX109442
<i>Shigella sonnei</i>		ATCC 25931	X96964		
<i>Tatumella pyseos</i>	DSM 5000 ^T	DSM 5000 ^T	AJ233437	AX109657	AX109499
<i>Trabulsiella guamensis</i>	ATCC 49490 ^T			AX109658	AX109500
<i>Yersinia enterocolitica</i>	ATCC 9610 ^T	ATCC 9610 ^T	M59292	AX109660	AX109502
<i>Yersinia frederiksenii</i>	ATCC 33641 ^T			AX109661	AX109503
<i>Yersinia intermedia</i>	ATCC 29909 ^T			AX109662	AX109504
<i>Yersinia pestis</i>	KIM D27			AX110028	AX109505
<i>Yersinia pestis</i>		ATCC 19428 ^T	X75274		
<i>Yersinia pseudotuberculosis</i>	ATCC 29833 ^T			AX109663	AX109506
<i>Yersinia rohdei</i>	ATCC 43380 ^T	ER-2935 ^T	X75276	AX109664	AX109507
<i>Yokenella regensburgei</i>	ATCC 35313 ^T			AX109665	AX109508
<i>Vibrio cholerae</i>	ATCC 25870			AX109941	AX109942
<i>Vibrio cholerae</i>		ATCC 14035 ^T	X74695		

*Phylogenetic serovars considered as species in the Approved Lists (Skerman *et al.*, 1980).

Distance Matrices Parsing and Plotting (DiMPP, a software tool freely available at <http://www.cri.crchul.ulaval.ca/dimpp/>) was used to obtain scatterplots for pairwise gene comparison into the genetic distance space. These distance plots were analysed to determine visually how well each taxonomic level (in this case species, genera and families) is resolved by each of the two compared genes.

Bootstrap and partition homogeneity test. To determine the number of bootstrap replications needed for the phylogenetic analyses, phylogenetic reconstructions were first repeated with exactly the same parameters at least twice with 100 bootstrap replications. If the consensus trees gave different topologies, the number of bootstrap replications was increased before repeating the phylogenetic reconstructions again (at least twice). The smallest number of bootstrap replications giving a stable consensus topology was chosen: for the *tuf* and *atpD* consensus trees, the smallest number of bootstrap replications required was 750. This number of bootstrap replications

was also used for the *tuf*, *atpD* and 16S rRNA gene sequence consensus trees (available as Supplementary Fig. S1 in IJSEM Online). We repeated the same procedure for the *tuf-atpD* tree. This latter tree was stable with 1000 replications. The comparison of consensus trees reconstructed with different numbers of bootstrap replications showed that the instability of consensus topologies is observed at nodes that exhibit bootstrap values around 50% (data not shown). This comparison revealed that this instability is not decreased with longer sequences. This could be explained by the fact that the submission of longer sequences brings a larger number of possible sequences randomly generated by the bootstrap calculation. Alternatively, these discrepancies could be attributed to incongruent phylogenetic signals between *atpD* and *tuf*. Indeed, a partition homogeneity test (ILD test in PAUP with 100 replicates) showed a *P* value of 0.01, suggesting an apparent conflict between the *tuf* and *atpD* phylogenies.

Table 2. PCR primers used for sequencing

The nucleotide positions given are for *Escherichia coli* *tuf* and *atpD* sequences (GenBank accession numbers AE000410 and V00267, respectively). Numbering starts from the first base of the initiation codon.

Primer	Sequence (5′–3′)	Position	Amplicon length (bp)
<i>tuf</i>			
T1	AAYATGATACIGGIGCIGCICARATGGA	271–299	884
T2	CCIACIGTICKICCRCCYTCRCG	1132–1154	
<i>atpD</i>			
A1	RTIATIGGIGCIGTIRTIGAYGT	25–47	884
A2	TCRTCIGCIGGIACRTAIAIYIGCYTG	883–908	
A3	TIRTIGAYGTCGARTTCCCTCARG	38–61	871
A2	TCRTCIGCIGGIACRTAIAIYIGCYTG	883–908	

RESULTS AND DISCUSSION

Sequence data

A PCR product of the expected size of 884 bp was obtained for *tuf* and one of 884 or 871 bp for *atpD* from all bacterial strains tested. After subtracting for biased primer regions and ambiguous single-strand data, 765 bp for *tuf* and 732 bp for *atpD* were subjected to phylogenetic analysis. The sequences obtained in this study are comparable to enterobacterial sequences from other studies available in public databases (Abdulkarim *et al.*, 1991; Amann *et al.*, 1988b; Blattner *et al.*, 1997; Christensen & Olsen, 1998; Hudson *et al.*, 1981; Perna *et al.*, 2001; Saraste *et al.*, 1981). However, some degree of polymorphism was observed. Zero to three and zero to nine differences in *tuf* and *atpD* sequences were found between *Escherichia coli* strains sequenced in this study and *Escherichia coli* K-12 MG1655 (Blattner *et al.*, 1997). This polymorphism is comparable to that found between *Escherichia coli* K-12 MG1655 and *Escherichia coli* EDL933 (serovar O157:H7) (Perna *et al.*, 2001), for which four and six differences are encountered, respectively. The *atpD* sequence was appended to the *tuf* sequence for every strain. Indeed, it is preferable to join two or more genes in order to submit more biological information for phylogenetic analysis when their evolution is similar for the taxa under study. The *tuf*–*atpD* dual gene alignment used for phylogenetic inference was 1414 bp long. All of the 16S rRNA gene sequences listed in Table 1, obtained from 58 strains representing 53 species belonging to 28 genera, were aligned and 1300 bp were subjected to phylogenetic analysis. Gaps were excluded to perform *tuf*, *atpD*, *tuf*–*atpD* and 16S rRNA gene sequence analyses.

Signature sequences

Multiple sequence alignments revealed no indels for *tuf*, whereas *atpD* had three distinct regions with indels. The region between positions 105 and 121 of *atpD* of *Escherichia coli* (GenBank accession no. V00267) (Saraste *et al.*, 1981) exhibited three different combinations involving one or two amino acid indels: one combined *Budvicia aquatica*, *Pragia fontium* and *Leminorella grimontii*, another was unique to *Plesiomonas shigelloides* and a third was found in species not belonging to the *Enterobacteriaceae*, including *Shewanella*

putrefaciens, *Haemophilus influenzae* and *Pasteurella multocida*, which were used as an outgroup. The lack of conservation of this 105–121 region suggests that parallelism, convergence or back-substitution events could have occurred. Therefore, further analyses will be required to determine the phylogenetic significance of these indels.

A 5 aa insertion located between positions 327 and 328 of *atpD* of *Escherichia coli* was observed for the type strains of *Pantoea agglomerans*, *Pantoea dispersa* and *Tatumella ptyseos*. This indel can be considered as a signature sequence for *Pantoea* species and *Tatumella ptyseos* (Fig. 1). In fact, the presence of a conserved indel of defined length and sequence which is flanked by conserved regions could suggest a common ancestor, particularly when members of a given taxon share this indel (Gupta, 1998). To our knowledge, this is the first demonstration to suggest a close common ancestor for the genera *Pantoea* and *Tatumella*. Also, this 5 aa indel could represent a useful marker for helping to resolve *Pantoea* classification. The transfer of *Enterobacter agglomerans* to *Pantoea agglomerans* was proposed by Gavini *et al.* (1989). However, rapid phenotypic identification systems are unable to distinguish unequivocally between the different species belonging to the *Erwinia herbicola*–*Enterobacter agglomerans* complex (Gavini *et al.*, 1989). The groups within this complex could be individualized by DNA hybridization but the heterogeneity of the complex limits phenotypic identification. Interestingly, *atpD* sequence data were obtained from a second *Pantoea agglomerans* strain in addition to the type strain. It was found that *Pantoea agglomerans* ATCC 27989 does not possess the 5 aa indel, suggesting that this strain may be misclassified and most likely does not belong to the genus *Pantoea* (Fig. 1). Strain ATCC 27989 was deposited as *Enterobacter agglomerans* biogroup 7, and, although we could not find a reference justifying the name change for this particular strain, it should be noted that strains of biogroup 7 can be found in at least three different DNA relatedness groups (Brenner *et al.*, 1984).

A 7 aa insertion located between positions 603 and 604 of the *atpD* gene of *Escherichia coli* was observed in the *Vibrio cholerae* sequence obtained in this study (data not shown). More *Vibrio* sequences will be required to evaluate the significance of this indel.

	301	311	321		331
<i>Escherichia coli</i>	GAGATCGGTG	AAGAAGAGCG	TTGGGCG		ATTCACCGC
<i>Pantoea agglomerans</i> ATCC 27989	GACATCGGTG	AAGAAGAGCG	TTGGGCG		ATCCACCGT
<i>Pantoea agglomerans</i> ATCC 27155 ^T	GAGCTGAAAG	AAGAAGATGG	CAGCGCAGTA	GAGATCGCCT	CTATTCACCGC
<i>Pantoea dispersa</i>	GACCTGAAAG	AAGAAGACGG	CAGCGCTGTA	GAGGTTTCCT	CTATTCATCGC
<i>Tatumella ptyseos</i>	GACCTGAAGA	ACGAAGATGG	TAGCAATGTT	GAGGTGAAC	CTATTCACCGT
<i>Escherichia coli</i>	E--I--G--E	--E--E--R-	-W--A----	-----	-I--H--R--
<i>Pantoea agglomerans</i> ATCC 27989	D--I--G--E	--E--E--R-	-W--A----	-----	-I--H--R--
<i>Pantoea agglomerans</i> ATCC 27155 ^T	E--L--K--E	--E--D--G-	-S--A--V--	E--I--A--S	-I--H--R--
<i>Pantoea dispersa</i>	D--L--K--E	--E--D--G-	-S--A--V--	E--V--S--S	-I--H--R--
<i>Tatumella ptyseos</i>	D--L--K--N	--E--D--G-	-S--N--V--	E--V--N--S	-I--H--R--

Fig. 1. *Pantoea* and *Tatumella* species-specific signature indel in *atpD*. The nucleotide positions given are for the *Escherichia coli atpD* sequence (GenBank accession no. V00267). Numbering starts from the first base of the initiation codon.

Phylogenetic trees based on partial *tuf*, *atpD* and 16S rRNA gene sequences of members of the *Enterobacteriaceae*

Bootstrap consensus trees reconstructed from *tuf*, *atpD* and *tuf-atpD* sequences are shown in Fig. 2(a), (b) and (c), respectively. The phylogenetic trees generated from partial *tuf* and *atpD* sequences are similar overall, but they show minor differences in branching. The *atpD* tree shows more monophyletic groups corresponding to species that belong to the same genus than does the *tuf* tree. Monophyletic genera observed on the *atpD* consensus tree are *Cedecea*, *Edwardsiella*, *Proteus*, *Providencia*, *Salmonella*, *Serratia*, *Raoultella* and *Yersinia*. Since *atpD* is more divergent than *tuf*, the former could allow better resolution for tree reconstruction. Whatever the gene used for tree reconstruction, some genera are not monophyletic, e.g. *Escherichia*, *Klebsiella* and *Enterobacter*. These results support previous phylogenies based on the genes *gap* and *ompA* (Lawrence *et al.*, 1991), *rpoB* (Drancourt *et al.*, 2001; Mollet *et al.*, 1997) and *infB* (Hedegaard *et al.*, 1999) and on DNA–DNA hybridization studies (Brenner *et al.*, 1986; Farmer *et al.*, 1985a).

There were few minor conflicts in branching between the *tuf* gene and the *atpD* gene. These differences could reflect small sequence differences, which could impact branching of genetically close taxa. This is the case for (i) *Enterobacter aerogenes* and *Raoultella* species, (ii) *Escherichia hermannii* and *Escherichia vulneris*, (iii) *Escherichia coli*, *Escherichia fergusonii* and *Shigella* species, (iv) serovars and subspecies of the same genospecies and (v) species of the same genus.

Four slightly more important discrepancies between *tuf* and *atpD* phylogenies are more difficult to explain. (i) In terms of the *tuf* gene, *Erwinia amylovora* is closer to *Pantoea* species than to *Tatumella ptyseos*. Phylogeny based on 16S rRNA gene sequences (Spröer *et al.*, 1999) confirms this branching. Nevertheless, this result is not congruent with the *atpD* phylogeny or with the indel (Fig. 1) shared only by the type strains of *Pantoea* species and *Tatumella ptyseos*. Moreover, bootstrap values better support the *atpD* branching. Therefore, *atpD* phylogeny could be more reliable for branching between these three genera. (ii) Branching of *Leminorella grimonii* with *Edwardsiella* species with the *tuf* gene is supported neither by *atpD* phylogeny nor by 16S rRNA gene sequence phylogeny (Spröer *et al.*, 1999), suggesting that the *tuf* gene could have evolved at a slower pace in the genus *Leminorella*. (iii) *tuf* phylogeny reveals a closer relationship between *Trabulsiella guamensis* and *Citrobacter farmeri*, while *atpD* shows more distant branching. In fact, the distance between these species is much smaller with the *tuf* gene and corresponds to distances obtained between two taxa of the same genus. (iv) *Moellerella wisconsensis* is closer to the genera *Proteus* and *Providencia* according to *atpD* gene analysis than according to *tuf* gene analysis. 16S rRNA gene sequences were not available for *Trabulsiella guamensis* or for *Moellerella*

wisconsensis. Perhaps further phylogenetic studies based on other genes could help to resolve these ambiguities.

Even though the *Pantoea* and *Tatumella* species-specific indel was excluded for phylogenetic analysis, type strains of *Pantoea agglomerans* and *Pantoea dispersa* grouped together and were distant from *Pantoea agglomerans* ATCC 27989, adding further evidence that careful analysis is required for the identification of species belonging to the heterogeneous *Erwinia herbicola*–*Enterobacter agglomerans* complex. In fact, with respect to the *tuf* and *atpD* genes, *Pantoea agglomerans* strain ATCC 27989 exhibits branch lengths similar to those for *Enterobacter* species. No comparisons of 16S rRNA gene sequences could be realized, because of the unavailability of the 16S rRNA gene sequence for *Pantoea agglomerans* strain ATCC 27989. Therefore, until further reclassification of this genus, we suggest that this strain should remain a member of the genus *Enterobacter*.

tuf and *atpD* trees exhibit very short genetic distances between taxa belonging to the same genetic species, including species segregated on the basis of clinical considerations. For example, *Escherichia coli* and *Shigella* species were confirmed to be of the same genetic species by hybridization studies (Brenner *et al.*, 1972a, b, 1982b), as well as by phylogenies based on 16S rRNA genes (Wang *et al.*, 1997) and *rpoB* genes (Mollet *et al.*, 1997). Hybridization studies (Bercovier *et al.*, 1980) and phylogeny based on 16S rRNA gene sequences (Ibrahim *et al.*, 1994) also demonstrated that *Yersinia pestis* and *Yersinia pseudotuberculosis* are of the same genetic species. Five genospecies analysed in this study are represented by at least two members: *E. coli*–*Shigella* species, *Yersinia pestis* and *Yersinia pseudotuberculosis*, *Klebsiella pneumoniae* subspecies, *Morganella morganii* subspecies and *Salmonella choleraesuis* subspecies. *Salmonella choleraesuis* is a less tightly knit species than the other four genospecies. In fact, strains from *Salmonella choleraesuis* show DNA–DNA hybridization levels of 57–99 % between subspecies and these hybridization levels are more than 76 % within each subspecies (Le Minor *et al.*, 1982). The genetic definition of a species generally would include strains with approximately 70 % or greater DNA–DNA relatedness (Wayne *et al.*, 1987). Therefore, *Salmonella choleraesuis* is a genetically broad species in accordance with DNA–DNA hybridization analyses and our phylogenetic results.

atpD phylogeny revealed *Salmonella choleraesuis* subspecies divisions consistent with the actual taxonomy. This result was also observed by Christensen & Olsen (1998). On the other hand, *Salmonella choleraesuis* subspecies are not resolved as well by *tuf* phylogeny. *atpD* and *tuf* phylogenies suggest that *Salmonella bongori* is another *Salmonella choleraesuis* subspecies. This observation is corroborated by 16S rRNA (Supplementary Fig. S1) and 23S rRNA gene sequence phylogeny (Christensen *et al.*, 1998), is qualified by DNA hybridization values (Le Minor *et al.*, 1982) and is contradicted by multilocus enzyme electrophoresis (Reeves *et al.*, 1989). In fact, the DNA–DNA hybridization level between *Salmonella bongori* and *Salmonella choleraesuis*

strains ranges from only 51 % up to 64 %, while intraspecies DNA–DNA hybridization levels for *Salmonella bongori* strains are above 91 % (Le Minor *et al.*, 1982). Le Minor *et al.* (1982) observed that *Salmonella bongori* could be considered as a novel species. Finally, Reeves *et al.* (1989) proposed the novel combination *Salmonella bongori* comb. nov. It

had been previously observed that recently diverged species might not be recognizable on the basis of conserved sequences even if DNA hybridization established them as being different species (Fox *et al.*, 1992). Therefore, *Salmonella bongori* and *Salmonella choleraesuis* could be considered as distinct, though recently diverged, species.

The phylogenetic relationships between *Salmonella*, *Escherichia coli* and *Citrobacter freundii* are not well defined. 16S and 23S rRNA gene sequence data reveal a closer relationship between *Salmonella* and *Escherichia coli* than between *Salmonella* and *Citrobacter freundii* (Christensen & Olsen, 1998; Spröer *et al.*, 1999), while DNA–DNA hybridization studies (Selander *et al.*, 1996) and *infB* phylogeny (Hedegaard *et al.*, 1999) showed that *Salmonella* is more closely related to

Citrobacter freundii than to *Escherichia coli*. In that regard, the *tuf* and *atpD* phylogenies are coherent with 16S and 23S rRNA gene sequence analysis, showing a closer relationship between the genus *Salmonella* and *Escherichia coli* than between the genera *Salmonella* and *Citrobacter*.

According to the *tuf* and *atpD* phylogenies (Supplementary Fig. S1a, b), *Escherichia fergusonii* is very close to the

Fig. 2. Trees based on sequence data from (a) *tuf*, (b) *atpD* and (c) *tuf-atpD*. The phylogenetic analysis was performed with the neighbour-joining method, calculated using the Kimura two-parameter method. Values on each branch indicate the occurrence (%) of the branching order in 750 bootstrapped trees for (a) and (b), and in 1000 bootstrapped trees for (c). *Haemophilus influenzae*, *Pasteurella multocida* subsp. *multocida*, *Shewanella putrefaciens* and *Vibrio cholerae* were used as an outgroup. Strain names and sequence accession numbers are listed in Table 1. Similar trees including only those strains for which 16S rRNA gene sequences were available are shown in Supplementary Fig. S1 in IJSEM Online.

Escherichia coli–*Shigella* genetic species. This observation is corroborated by the 16S rRNA gene sequence phylogeny (Supplementary Fig. S1c) (McLaughlin *et al.*, 2000) but not by the DNA hybridization values. In fact, the DNA–DNA hybridization level between *Escherichia fergusonii* and *Escherichia coli*–*Shigella* is only 49–63 % (Farmer *et al.*, 1985a). Therefore, *Escherichia fergusonii* could be a recently diverged species, such as is the case for *Salmonella bongori*.

To simplify the comparisons, phylogenetic trees for *tuf* and *atpD* (Supplementary Fig. S1a, b) were reconstructed using sequences corresponding to taxa for which 16S rRNA gene sequences were available in the GenBank/EMBL databases. To complete this study, we determined the 16S rRNA gene sequences of *Escherichia fergusonii* and *Escherichia vulneris* (Supplementary Fig. S1c). The *tuf* and *atpD* trees were similar to those generated using additional taxa (shown in Fig. 2). The tree for 16S rRNA gene sequences gave a poorer resolution power at the species and genus levels than did the *tuf* and *atpD* trees. Indeed, the 16S rRNA gene sequence tree exhibited more multifurcation (polytomies) than did the *tuf* and *atpD* trees.

Notwithstanding the apparent incongruence of *tuf* and *atpD*, the phylogeny based on *tuf*–*atpD* appears to improve some bootstrap values, and, in some cases, to resolve a few of the polytomies. Indeed, according to that consensus tree (Fig. 2c), *Budvicia aquatica* and *Pragia fontium* are resolved from the species belonging to the genus *Yersinia*. Also, *Plesiomonas shigelloides* is branched deeper than the group *Hafnia alvei*–*Obesumbacterium proteus* and *Morganella morganii* subspecies. Moreover, the branch with *Lemnorella grimontii* and species of the genus *Edwardsiella* appears as a sister group of the *Cedecea*–*Klebsiella*–*Enterobacter*–*Escherichia*–*Salmonella*–*Citrobacter* group. This latter group has been defined as the ‘core’ of the family *Enterobacteriaceae* (Brenner *et al.*, 1982a). Finally, the *Citrobacter koseri*–*Citrobacter sedlakii* group and *Pantoea agglomerans* ATCC 27989 branch between the *Escherichia coli*–*Shigella*–*Escherichia fergusonii*–*Salmonella* group and the other enterobacteria belonging to the ‘core’.

Distance analysis with DiMPP showed that, for each pair of strains compared with each other, *tuf* and *atpD* distances were sufficient to allow clear discrimination between different species, whereas 16S rRNA gene sequences often exhibited much shorter distances between species (see Supplementary Fig. S2 available in IJSEM Online). Other studies confirm that sequence analysis of 16S rRNA genes is not an appropriate method for delineation at lower taxonomic levels; for example, sequence heterogeneities among 16S rRNA operons can affect phylogenetic analysis at the species level (Cilia *et al.*, 1996; Clayton *et al.*, 1995). Moreover, the low evolutionary rate of this gene can cause failure in the distinction of closely related taxa (Palys *et al.*, 1997). However, the majority of phenotypically close enterobacterial species could be easily discriminated genotypically using *tuf* or *atpD* gene sequences.

Conclusion

In this study, the phylogenetic affiliations of 96 enterobacterial strains representing 78 species from 31 genera were revealed by analyses based on *tuf* and *atpD* genes. These genes exhibit phylogenies consistent with the 16S rRNA gene sequence phylogeny. For example, they show that the family *Enterobacteriaceae* is monophyletic. However, *tuf* and *atpD* distances provide a higher discriminating power at the species level. In fact, *tuf* and *atpD* provide better discrimination between different genospecies, such that primers and molecular probes could be designed for diagnostic purposes. Therefore, they represent good target genes for distinguishing phenotypically close enterobacteria belonging to different genetic species, e.g. *Klebsiella pneumoniae* and *Enterobacter aerogenes*. Preliminary studies support these observations, and diagnostic tests based on *tuf* and *atpD* gene sequence data for identifying enterobacteria are currently under development in our laboratory.

In summary, this study shows that *tuf*, *atpD* and a *tuf*–*atpD* combination represent highly valuable phylogenetic tools offering discriminatory power superior to that of 16S rRNA gene sequences for distinguishing between species. Moreover, extensive evolutionary distance comparisons using a group of conserved genes should help to better define a genetic basis for classification into genera and families. This would be of great value for revisiting the taxonomy of bacterial species.

ACKNOWLEDGEMENTS

We thank Pascal Lapierre for the design of *tuf* sequencing primers. S. P. received scholarships from Fondation Dr George Phénix (Outremont, Québec, Canada) and from le Fonds de recherche en santé du Québec. This research project was supported by grant PA-15586 from the Canadian Institutes of Health Research and by Infectio Diagnostic (I.D.I) Inc., Ste-Foy, Québec, Canada.

REFERENCES

- Abdulkarim, F. & Hughes, D. (1996). Homologous recombination between the *tuf* genes of *Salmonella typhimurium*. *J Mol Biol* **260**, 506–522.
- Abdulkarim, F., Tuohy, T. M., Buckingham, R. H. & Hughes, D. (1991). Missense substitutions lethal to essential functions of EF-Tu. *Biochimie* **73**, 1457–1464.
- Amann, R., Ludwig, W. & Schleifer, K. H. (1988a). β -Subunit of ATP-synthase: a useful marker for studying the phylogenetic relationship of eubacteria. *J Gen Microbiol* **134**, 2815–2821.
- Amann, R., Sostak, P., Ludwig, W. & Schleifer, K. H. (1988b). Cloning and sequencing of genes encoding the beta subunits of the ATP-synthases from *Enterobacter aerogenes* and *Flavobacterium ferrugineum*. *FEMS Microbiol Lett* **50**, 101–106.
- Bercovier, H., Mollaret, H. H., Alonso, J. M., Brault, J., Fanning, G. R., Steigerwalt, A. G. & Brenner, D. J. (1980). Intra- and interspecies relatedness of *Yersinia pestis* by DNA hybridization and its relationship to *Yersinia pseudotuberculosis*. *Curr Microbiol* **4**, 225–229.

- Blattner, F. R., Plunkett, G., III, Bloch, C. A. & 14 other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474.
- Brenner, D. J. (1984). Facultatively anaerobic gram-negative rods. Family I. *Enterobacteriaceae*. In *Bergey's Manual of Systematic Bacteriology*, vol. 1, pp. 408–420. Edited by N. R. Krieg & J. G. Holt. Baltimore: Williams & Wilkins.
- Brenner, D. J. (1992). Additional genera of the *Enterobacteriaceae*. In *The Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications*, pp. 2922–2937. Edited by A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K. H. Schleifer. New York: Springer.
- Brenner, D. J., Fanning, G. R., Skerman, F. J. & Falkow, S. (1972a). Polynucleotide sequence divergence among strains of *Escherichia coli* and closely related organisms. *J Bacteriol* **109**, 953–965.
- Brenner, D. J., Fanning, G. R., Steigerwalt, A. G., Orskov, I. & Orskov, F. (1972b). Polynucleotide sequence relatedness among three groups of pathogenic *Escherichia coli* strains. *Infect Immun* **6**, 308–315.
- Brenner, D. J., Richard, C., Steigerwalt, A. G., Asbury, M. A. & Mandel, M. (1980). *Enterobacter gergoviae* sp. nov.: a new species of *Enterobacteriaceae* found in clinical specimens and the environment. *Int J Syst Bacteriol* **30**, 1–6.
- Brenner, D. J., McWhorter, A. C., Knutson, J. K. & Steigerwalt, A. G. (1982a). *Escherichia vulneris*: a new species of *Enterobacteriaceae* associated with human wounds. *J Clin Microbiol* **15**, 1133–1140.
- Brenner, D. J., Steigerwalt, A. G., Wathen, H. G., Gross, R. J. & Rowe, B. (1982b). Confirmation of aerogenic strains of *Shigella boydii* 13 and further study of *Shigella* serotypes by DNA relatedness. *J Clin Microbiol* **16**, 432–436.
- Brenner, D. J., Fanning, G. R., Leete Knutson, J. K., Steigerwalt, A. G. & Krichevsky, M. I. (1984). Attempts to classify Herbicola group-*Enterobacter agglomerans* strains by deoxyribonucleic acid hybridization and phenotypic tests. *Int J Syst Bacteriol* **34**, 45–55.
- Brenner, D. J., McWhorter, A. C., Kai, A., Steigerwalt, A. G. & Farmer, J. J., III (1986). *Enterobacter asburiae* sp. nov., a new species found in clinical specimens, and reassignment of *Erwinia dissolvens* and *Erwinia nimipressuralis* to the genus *Enterobacter* as *Enterobacter dissolvens* comb. nov. and *Enterobacter nimipressuralis* comb. nov. *J Clin Microbiol* **23**, 1114–1120.
- Brenner, D. J., Grimont, P. A., Steigerwalt, A. G., Fanning, G. R., Ageron, E. & Riddle, C. F. (1993). Classification of citrobacteria by DNA hybridization: designation of *Citrobacter farmeri* sp. nov., *Citrobacter youngae* sp. nov., *Citrobacter braakii* sp. nov., *Citrobacter werkmanii* sp. nov., *Citrobacter sedlakii* sp. nov., and three unnamed *Citrobacter* genomospecies. *Int J Syst Bacteriol* **43**, 645–658.
- Brenner, D. J., O'Hara, C. M., Grimont, P. A. & 7 other authors (1999). Biochemical identification of *Citrobacter* species defined by DNA hybridization and description of *Citrobacter gillenii* sp. nov. (formerly *Citrobacter* genomospecies 10) and *Citrobacter murlinae* sp. nov. (formerly *Citrobacter* genomospecies 11). *J Clin Microbiol* **37**, 2619–2624.
- Christensen, H. & Olsen, J. E. (1998). Phylogenetic relationships of *Salmonella* based on DNA sequence comparison of *atpD* encoding the beta subunit of ATP synthase. *FEMS Microbiol Lett* **161**, 89–96.
- Christensen, H., Nordentoft, S. & Olsen, J. E. (1998). Phylogenetic relationships of *Salmonella* based on rRNA sequences. *Int J Syst Bacteriol* **48**, 605–610.
- Cilia, V., Lafay, B. & Christen, R. (1996). Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol Biol Evol* **13**, 451–461.
- Clayton, R. A., Sutton, G., Hinkle, P. S., Jr, Bult, C. & Fields, C. (1995). Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. *Int J Syst Bacteriol* **45**, 595–599.
- Dickey, R. S. & Zumoff, C. H. (1988). Emended description of *Enterobacter cancerogenus* comb. nov. (formerly *Erwinia cancerogena*). *Int J Syst Bacteriol* **38**, 371–374.
- Drancourt, M., Bollet, C., Carta, A. & Rousselier, P. (2001). Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. *Int J Syst Evol Microbiol* **51**, 925–932.
- Farmer, J. J., III, Asbury, M. A., Hickman, F. W., Brenner, D. J. & the *Enterobacteriaceae* Study Group (1980). *Enterobacter sakazakii*: a new species of “*Enterobacteriaceae*” isolated from clinical specimens. *Int J Syst Bacteriol* **30**, 569–584.
- Farmer, J. J., III, Fanning, G. R., Davis, B. R., O'Hara, C. M., Riddle, C., Hickman-Brenner, F. W., Asbury, M. A., Lowery, V. A., III & Brenner, D. J. (1985a). *Escherichia fergusonii* and *Enterobacter taylora*, two new species of *Enterobacteriaceae* isolated from clinical specimens. *J Clin Microbiol* **21**, 77–81.
- Farmer, J. J., III, Davis, B. R., Hickman-Brenner, F. W. & 12 other authors (1985b). Biochemical identification of new species and biogroups of *Enterobacteriaceae* isolated from clinical specimens. *J Clin Microbiol* **21**, 46–76.
- Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P., Jr (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**, 166–170.
- Gavini, F., Mergaert, J., Beji, A., Mielcarek, C., Izard, D., Kersters, K. & De Ley, J. (1989). Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. *Int J Syst Bacteriol* **39**, 337–345.
- Grunberg-Manago, M. (1996). Regulation of the expression of aminoacyl-tRNA synthetases and translation factors. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, pp. 1432–1457. Edited by F. C. Neidhardt, R. I. Curtiss, J. L. Ingraham & 7 other editors. Washington, DC: American Society for Microbiology.
- Gupta, R. S. (1998). Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* **62**, 1435–1491.
- Hartl, D. L. & Dykhuizen, D. E. (1984). The population genetics of *Escherichia coli*. *Annu Rev Genet* **18**, 31–68.
- Hedegaard, J., Steffensen, S. A., Nørskov-Lauritsen, N., Mortensen, K. K. & Sperling-Petersen, H. U. (1999). Identification of *Enterobacteriaceae* by partial sequencing of the gene encoding translation initiation factor 2. *Int J Syst Bacteriol* **49**, 1531–1538.
- Hill, C. W. & Harnish, B. W. (1981). Inversions between ribosomal RNA genes of *Escherichia coli*. *Proc Natl Acad Sci U S A* **78**, 7069–7072.
- Hudson, L., Rossi, J. & Landy, A. (1981). Dual function transcripts specifying tRNA and mRNA. *Nature* **294**, 422–427.
- Ibrahim, A., Goebel, B. M., Liesack, W., Griffiths, M. & Stackebrandt, E. (1994). The phylogeny of the genus *Yersinia* based on 16S rDNA sequences. *FEMS Microbiol Lett* **114**, 173–178.
- Izard, D., Gavini, F., Trinel, P. A. & Leclerc, H. (1981). Deoxyribonucleic acid relatedness between *Enterobacter cloacae* and *Enterobacter amnigenus* sp. nov. *Int J Syst Bacteriol* **31**, 35–42.
- Janda, J. M., Abbott, S. L. & Albert, M. J. (1999). Prototypal diarrheagenic strains of *Hafnia alvei* are actually members of the genus *Escherichia*. *J Clin Microbiol* **37**, 2399–2401.
- Ke, D., Boissinot, M., Huletsky, A., Picard, F. J., Frenette, J., Ouellette, M., Roy, P. H. & Bergeron, M. G. (2000). Evidence for horizontal gene transfer in evolution of elongation factor Tu in enterococci. *J Bacteriol* **182**, 6913–6920.

- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.
- Kitch, T. T., Jacobs, M. R. & Appelbaum, P. C. (1994). Evaluation of RapID onE system for identification of 379 strains in the family *Enterobacteriaceae* and oxidase-negative, gram-negative nonfermenters. *J Clin Microbiol* **32**, 931–934.
- Lane, D. J. (1991). 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*, pp. 115–175. Edited by E. Stackebrandt & M. Goodfellow. New York: Wiley.
- Lawrence, J. G., Ochman, H. & Hartl, D. L. (1991). Molecular and evolutionary relationships among enteric bacteria. *J Gen Microbiol* **137**, 1911–1921.
- Le Minor, L., Veron, M. & Popoff, M. (1982). The taxonomy of *Salmonella*. *Ann Microbiol* **133**, 223–243 (in French).
- Ludwig, W., Weizenegger, M., Betzl, D., Leidel, E., Lenz, T., Ludvigsen, A., Mollenhoff, D., Wenzig, P. & Schleifer, K. H. (1990). Complete nucleotide sequences of seven eubacterial genes coding for the elongation factor Tu: functional, structural and phylogenetic evaluations. *Arch Microbiol* **153**, 241–247.
- Ludwig, W., Neumaier, J., Klugbauer, N. & 9 other authors (1993). Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes. *Antonie van Leeuwenhoek* **64**, 285–305.
- McLaughlin, I. J., Valentine, J. & Dodge, D. E. (2000). Intraspecies taxonomy of multiple clinical isolates from members of the family *Enterobacteriaceae*. In *Abstracts of the 100th General Meeting of the American Society for Microbiology*, Los Angeles, CA, USA, 24 May 2000, abstract no. R13, p. 629. Washington, DC: American Society for Microbiology.
- Mollet, C., Drancourt, M. & Raoult, D. (1997). *rpoB* sequence analysis as a novel basis for bacterial identification. *Mol Microbiol* **26**, 1005–1011.
- Nelson, N. & Taiz, L. (1989). The evolution of H⁺-ATPases. *Trends Biochem Sci* **14**, 113–116.
- Palys, T., Nakamura, L. K. & Cohan, F. M. (1997). Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Syst Bacteriol* **47**, 1145–1156.
- Perna, N. T., Plunkett, G., III, Burland, V. & 25 other authors (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529–533.
- Reeves, M. W., Evins, G. M., Heiba, A. A., Plikaytis, B. D. & Farmer, J. J., III (1989). Clonal nature of *Salmonella typhi* and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. nov. *J Clin Microbiol* **27**, 313–320.
- Saraste, M., Gay, N. J., Eberle, A., Runswick, M. J. & Walker, J. E. (1981). The *atp* operon: nucleotide sequence of the genes for the gamma, beta, and epsilon subunits of *Escherichia coli* ATP synthase. *Nucleic Acids Res* **9**, 5287–5296.
- Sela, S., Yogev, D., Razin, S. & Bercovier, H. (1989). Duplication of the *tuf* gene: a new insight into the phylogeny of eubacteria. *J Bacteriol* **171**, 581–584.
- Selander, R. K., Li, J. & Nelson, K. (1996). Evolutionary genetics of *Salmonella enterica*. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, pp. 2691–2707. Edited by F. C. Neidhardt, R. I. Curtiss, J. L. Ingraham & 7 other editors. Washington, DC: American Society for Microbiology.
- Sharma, N. K., Doyle, P. W., Gerbasi, S. A. & Jessop, J. H. (1990). Identification of *Yersinia* species by the API 20E. *J Clin Microbiol* **28**, 1443–1444.
- Skerman, V. B. D., McGowan, V. & Sneath, P. H. A. (1980). Approved lists of bacterial names. *Int J Syst Bacteriol* **30**, 225–420.
- Spröer, C., Mendrock, U., Swiderski, J., Lang, E. & Stackebrandt, E. (1999). The phylogenetic position of *Serratia*, *Buttiauxella* and some other genera of the family *Enterobacteriaceae*. *Int J Syst Bacteriol* **49**, 1433–1438.
- Stackebrandt, E. & Goebel, B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**, 846–849.
- Steigerwalt, A. G., Fanning, G. R., Fife-Asbury, M. A. & Brenner, D. J. (1976). DNA relatedness among species of *Enterobacter* and *Serratia*. *Can J Microbiol* **22**, 121–137.
- Wang, R. F., Cao, W. W. & Cerniglia, C. E. (1997). Phylogenetic analysis and identification of *Shigella* spp. by molecular probes. *Mol Cell Probes* **11**, 427–432.
- Wayne, L. G., Brenner, D. J., Colwell, R. R. & 9 other authors (1987). International Committee on Systematic Bacteriology. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**, 463–464.