J Mol Evol (2008) 66:630–642
DOI 10.1007/s00239-008-9115-3

# Is 16S rDNA a Reliable Phylogenetic Marker to Characterize Relationships Below the Family Level in the Enterobacteriaceae?

**Marianna Naum · Eric W. Brown · Roberta J. Mason-Gamer**

Received: 20 November 2007 / Accepted: 28 April 2008 / Published online: 27 May 2008


**Abstract** The phylogenetic relationships of multiple enterobacterial species were reconstructed based on 16S rDNA gene sequences to evaluate the robustness of this housekeeping gene in the taxonomic placement of the enteric plant pathogens *Erwinia*, *Brenneria*, *Pectobacterium*, and *Pantoea*. Four data sets were compiled, two of which consisted of previously published data. The data sets were designed in order to evaluate how 16S rDNA gene phylogenies are affected by the use of different plant pathogen accessions and varying numbers of animal pathogen and outgroup sequences. DNA data matrices were analyzed using maximum likelihood (ML) algorithms, and character support was determined by ML bootstrap and Bayesian analyses. As additional animal pathogen sequences were added to the phylogenetic analyses, taxon placement changed. Further, the phylogenies varied in their placement of the plant pathogen species, and only the genus *Pantoea* was monophyletic in all four trees. Finally, bootstrap and Bayesian support values were low for most of the nodes, and all nonterminal branches collapsed in strict consensus trees. Inspection of 16S rDNA nucleotide alignments revealed several highly variable blocks punctuated by regions of conserved sequence. These data suggest that 16S rDNA, while effective for both species-level and family-level phylogenetic reconstruction, may underperform for genus-level phylogenetic analyses in the Enterobacteriaceae.

**Keywords** *Erwinia* · 16S rDNA · Housekeeping genes · Phylogenetic analysis · Enterobacteriaceae · GARLI

M. Naum · R. J. Mason-Gamer
Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA

M. Naum (✉)
Center for Food Safety and Applied Nutrition, FDA, 5100 Paint Branch Parkway, HFS-712, College Park, MD 20740, USA
e-mail: marianna.naum@fda.hhs.gov

E. W. Brown
Division of Microbiology, Center for Food Safety and Applied Nutrition, FDA, College Park, MD 20740, USA

## Introduction

Differentiation and classification of extant bacterial taxa have accelerated over the last two decades with the increased use of molecular phylogenetic data (Charlebois et al. 2003; Gould 1996; Martinez et al. 2004). DNA sequencing has transformed the classification of prokaryotes from broad groupings based on resemblances in morphological, biochemical, and physiological characteristics to detailed phylogenies resulting from sequences unique to bacterial species and even strains (Olsen and Woese 1993). DNA sequences from the 16S rDNA gene have been used to determine numerous prokaryotic phylogenies at all taxonomic levels (Christensen et al. 2004; Davidov and Jurkevitch 2004; Dorsch et al. 1992; Woese 1987). This universal applicability of 16S rDNA sequences has enabled microbiologists to confirm broad classifications, such as the $\gamma$-bacteria at the phylum level, as well as relationships of species that reside within the same family or genus. The wide application of 16S rDNA sequences stems from the highly conserved and variable components that compose the ribosomal subunits (Olsen and Woese 1993), making 16S rDNA one of the most widely exploited genes for use in microbial phylogenetic studies.

It is notable that 16S rDNA gene sequences have been useful in reconstructing the phylogenies of closely related species and strains in the family Enterobacteriaceae (Wertz

et al. 2003). The most familiar species of the family are human pathogens and include pathogenic *Escherichia coli*, *Shigella flexneri*, and *Yersinia enterocolitica*, all of which can cause severe gastrointestinal illnesses (Toth et al. 2006). Other species, such as *Erwinia amylovora*, the causative agent of fire blight in rosaceous plant hosts, are agriculturally significant (Krieg and Holt 1985; van der Zwet and Beer 1995). In a family as diverse as the Enterobacteriaceae, deciphering the evolutionary relationships among animal and plant pathogen members is crucial, especially since they both employ common and conserved virulence mechanisms to elicit infection on their specific host cells.

Historically, classification studies have focused solely on either animal (Moran et al. 2005; Wertz et al. 2003) or plant (Fessehaie et al. 2002; Kwon et al. 1997) pathogen representatives of the family. Hauben et al. (1998), however, completed a comprehensive phylogenetic analysis on a large number of enterobacterial animal and plant pathogen species based on 16S rDNA gene sequences. As a result of this study, the taxonomy of the plant pathogen genera was emended. A new generic scheme was proposed that partitioned the phytopathogenic species into four distinct clusters: *Erwinia*, *Brenneria*, *Pectobacterium*, and *Pantoea* (Hauben et al. 1998), with *Brenneria* being basal to both animal and plant pathogen clusters.

Studies using 16S rDNA to determine microbial phylogenies (Christensen et al. 2004; Davidov and Jurkevitch 2004; Dorsch et al. 1992; Hauben et al. 1998; Woese 1987) have positioned this gene as an ideal candidate for reconstructing robust phylogenies of *Erwinia*, *Brenneria*, *Pectobacterium*, and *Pantoea*. Such phylogenies, in association with other housekeeping genes, could serve as a reference point for identifying instances of horizontal gene transfer among virulence genes and other auxiliary sequences in the genomes of these species. In this paper we evaluate the robustness and reliability of 16S rDNA gene sequences in phylogeny reconstruction of plant pathogen enterobacterial species by assessing how the resulting phylogenies are affected by the use of different plant pathogen accessions and varying numbers of animal pathogen representatives. To investigate these effects, we compiled and explored four data sets, two of which are from previously published phylogenies (Hauben et al. 1998; Kwon et al. 1997). Each data set focused on the enteric plant pathogens *Erwinia*, *Brenneria*, *Pectobacterium*, and *Pantoea*, with the first two comprised of the same 27 plant pathogen sequences. The last two data sets consisted of sequences published by Kwon et al. (1997) and Hauben et al. (1998), which were downloaded from GenBank. All four data sets were analyzed using the same methods in order to eliminate disparity resulting from the use of different phylogenetic analysis methods. The animal pathogen and outgroup strains used are the same as those of Kwon et al. (1997) in the first data set and a subset of Hauben et al. (1998) in the second. The resulting trees were inconsistent in the taxonomic placement of both plant and animal pathogen representatives in the four data sets. Additionally, they all displayed poor maximum likelihood (ML) bootstrap and Bayesian support, indicating that 16S rDNA, while effective for both species-level and family-level phylogenetic reconstruction, may underperform for genus-level phylogenetic analyses.

## Materials and Methods

### Strains and Sampling

Data set A is comprised of 33 samples (Table 1) including 27 plant pathogen isolates, 5 representative animal pathogen species, and 1 outgroup taxon (Kwon et al. 1997). Data set B includes 49 samples (Table 2), consisting of the same 27 plant pathogen isolates as in data set A, 18 representative animal pathogen species, and 4 outgroup taxa (Hauben et al. 1998). Data set C consists of the 22 16S rDNA gene sequences previously published by Kwon et al. (1997) (Table 3), and data set D consists of the 67 sequences published by Hauben et al (1998) (Table 4). The four data sets were used to characterize relationships within *Erwinia*, *Brenneria*, *Pectobacterium*, and *Pantoea* and to evaluate the placement of the four phytopathogenic genera within the Enterobacteriaceae.

Strains of *Erwinia*, *Brenneria*, *Pectobacterium*, and *Pantoea* listed in data sets A and B were obtained from the laboratories of Drs. Eric Brown and Amy Charkowski as frozen cultures and LB slants, respectively. Strain authentication was confirmed based on 16S rDNA sequence prior to their use.

### Bacterial DNA Extraction and PCR Amplification

Total genomic DNA was extracted from the plant pathogen species listed in Table 1 using a silica-based matrix (Instagene; Bio-Rad Inc.) as described in the manufacturer's directions. Briefly, bacterial cells from 24-h cultures grown on LB agar at room temperature were resuspended in Instagene matrix and heated at 56°C for 30 min. The preps were then vortexed vigorously and boiled for 10 min. Suspensions were then pelleted and the supernatant containing the DNA was decanted into a clean microtube. The optimum amount of eluate to be used in amplification reactions was determined to be 2 µl, after an optimization PCR was carried out using 20, 10, 5, 2, and 1 µl of DNA. Oligonucleotide primers were designed for amplification of the 16S rDNA gene from conserved sequences flanking variable regions of the genes, using *Erwinia*, *Brenneria*,

**Table 1** Bacterial strains and accessions for published and novel partial 16S rDNA gene sequences comprising data set A

| Taxon | Strain no. | GenBank accession no. |
|---|---|---|
| *Erwinia* | | |
| *E. amylovora* | 278 | EU490590 |
| *E. tracheiphila* | 5845 | EU490598 |
| *E. rhapontici* | 1, 29283 | EU490593, EU490594 |
| *E. lupinicola* | 31 | EU490595 |
| *E. mallotivora* | 8645, 8646 | EU490591, EU490592 |
| *E. psidii* | 8423, 8429 | EU490596, EU490597 |
| *Brenneria* | | |
| *B. rubrifaciens* | 5950 | EU490604 |
| *B. nigrifluens* | 1391, 4789 | EU490601, EU490602 |
| *B. alni* | pvfi20 | EU490603 |
| *B. quercina* | 1846, 1895 | EU490599, EU490600 |
| *Pectobacterium* | | |
| *P. carotovora* | Eca6, 15713 | EU490611, EU490612 |
| *Ssp. carotovora* | SCRI1043 | EU490610 |
| *Ssp. atroseptica* | 1693 | EU490607 |
| *Ssp. brasiliensis* | WPP163, SCRI488 | EU490608, EU490609 |
| *Ssp. wasabiae* | EC16 | EU490613 |
| *P. chrysanthemi* | 29267, 29269 | EU490605, EU490606 |
| *P. cypripedii* | | |
| *Pantoea* | | |
| *P. agglomerans* | 33243 | U80202 |
| *P. stewartii* | 8199 | U80208 |
| *P. ananatis* | 19321 | U80209 |
| Animal pathogen strains | | |
| *Escherichia coli* | MG1655 | NC_000913 |
| *Yersinia pestis* | Evpst + c | Z75317 |
| *Klebsiella pneumoniae* | DSM30104 | X87276 |
| *Serratia marcescens* | ATCC 13880$^T$ | M59160 |
| *Hafnia alvei* | ATCC 13337$^T$ | M59155 |
| Outgroup | | |
| *Proteus vulgaris* | | J01874 |

**Table 2** Additional bacterial strain accessions for partial published 16S rDNA gene sequences used in conjunction with Table 1 to compose data set B

| Taxon | Strain no. | GenBank accession no. |
|---|---|---|
| *Shigella* | | |
| *S. flexneri* | | X80679 |
| *S. dysenteriae* | | X80680 |
| *S. boydii* | | X96965 |
| *S. sonnei* | | X96964 |
| *Hafnia* | | |
| *H. alvei* | ATCC 13337$^T$ | N59155 |
| *Escherichia* | | |
| *E. coli* | MG1655 | NC_000913 |
| *E. vulneris* | ATCC 33821$^T$ | X80734 |
| *E. hermanii* | | X80675 |
| *Salmonella* | | |
| *S. give* | | X80683 |
| *S. typhimurium* | | X80681 |
| *S. paratyphi* | | U88547 |
| *Yersinia* | | |
| *Y. pestis* | Evpst + c | Z75317 |
| *Y. mollaretii* | ER-2975 | X75280 |
| *Y. enterocolitica* | ER-26039-92 | Z49829 |
| *Klebsiella* | | |
| *K. pneumoniae* | DSM30104 | X87276 |
| *K. planticola* | DSM3069 | X93215 |
| *Serratia* | | |
| *S. marcesens* | ATCC 13880$^T$ | M59160 |
| *Citrobacter* | | |
| *C. freundii* | | M59291 |
| *Outgroups* | | |
| *Photorhabdus luminescens* | DSM 3368 | X82248 |
| *Xenorhabdus nematophilus* | DSM 3370 | X82251 |
| *Xenorhabdus bovienii* | DSM 4766 | X82252 |
| *Proteus vulgaris* | | J01874 |

and *Pectobacterium* 16S rDNA sequences from GenBank. The primers used were 16SF1 (5′-GCA GGC CTA ACA CAT GCA AGT CG-3′) and 16SR1 (5′-GCA ACC CAC TCC CAT GGT GTG ACG-3′). Amplification reactions were carried out with 2 µl of DNA template, 2.5 µl of 10× PCR buffer (Sigma), 1.5 µl of 25 mM $MgCl_2$, 2 µl of 2.5 mM dNTPs, 2.5 µl of each 10 µM primer stock, 0.1 µl of 5 U/µl *Taq* DNA polymerase (Sigma), and purified $H_2O$ for a total reaction volume of 25 µl. PCR (Ehrlich et al. 1991) was performed in a PTC-200 thermal cycler (MJ Research) under the following conditions: initial denaturation at 94°C for 2 min, 94°C for 50 s, 58°C for 50 s, and 72°C for 45 s (35 cycles), ending with incubation at 72°C for 8 min.

## Cloning and DNA Sequencing

PCR products were purified using QIAquick gel extraction kits (QIAGEN) according to the manufacturer's protocol. Briefly, completed PCR products were analyzed on 1% agarose gel and visualized with ethidium bromide and UV

**Table 3** Accessions for partial 16S rDNA gene sequences published by Kwon et al. (1997) making up data set C

| Taxon | Strain no. | GenBank accession no. |
|---|---|---|
| *Erwinia* | | |
| *E. amylovora* | ATCC 15580$^T$ | U80195 |
| *E. persicinus* | ATCC 35998$^T$ | U80205 |
| *E. rhapontici* | ATCC 29283$^T$ | U80206 |
| *Brenneria* | | |
| *B. rubrifaciens* | ATCC 29291$^T$ | U80207 |
| *B. nigrifluens* | ATCC 13028$^T$ | U80203 |
| *B. salicis* | ATCC 15712$^T$ | U80210 |
| *Pectobacterium* | | |
| *P. carotovora* | ATCC 15713$^T$ | U80197 |
| *Ssp. carotovora* | ATCC 43762$^T$ | U80198 |
| *Ssp. etavasculorum* | ATCC 43216$^T$ | U80199 |
| *Ssp. wasabiae* | ATCC 11663$^T$ | U80200 |
| *P. chrysanthemi* | | |
| *Pantoea* | | |
| *P. herbicola (P. agglomerans)* | ATCC 33243$^T$ | U80202 |
| *P. milletiae (P. agglomerans)* | ATCC 33261$^T$ | U80183 |
| *P. stewartii* | ATCC 8199$^T$ | U80208 |
| *P. uredovora (P. ananatis)* | ATCC 19321$^T$ | U80209 |
| *P. ananas* | ATCC 33243$^T$ | U80196 |
| Animal pathogen strains | | |
| *Escherichia coli* | | J01695 |
| *Yersinia intermedia* | ER 3854 | X75279 |
| *Klebsiella pneumoniae* | DSM30104 | X87276 |
| *Serratia marcesens* | ATCC 13880$^T$ | M59160 |
| *Hafnia alvei* | ATCC 13337$^T$ | M59155 |
| Outgroup | | |
| *Proteus vulgaris* | | J01874 |

light. Positive PCR products were gel excised and incubated in buffer QG at 50°C for 10 min. DNA was bound in QIAquick spin columns by microcentrifugation, and purified DNA was eluted in 20 µl of buffer. Cloning reactions were completed using pGem-T Easy cloning kits (Promega) according to the manufacturer's protocol except that the ligation and transformation reaction volumes were cut in half. The products were ligated to pGem-T vectors with $T_4$ DNA ligase at 4°C overnight. Transformation proceeded via conventional heat shock methods (Manniatis et al. 1982) into competent *E. coli* (JM109) cloning cells. Cloned inserts were amplified directly from colonies using the original PCR primers and amplification conditions. PCR reactions were cleaned using exonuclease I and shrimp alkaline phosphatase by heating to 37°C for 15 min and then to 80°C for an additional 15 min (Mason-Gamer 2004). Purified PCR products were sequenced bidirectionaly using ABI Big Dye Terminator v3.1 and v1.1 Cycle

**Table 4** Accessions for partial 16S rDNA gene sequences published by Hauben et al. (1998) making up data set D

| Taxon | Strain no. | GenBank accession no. |
|---|---|---|
| *Erwinia* | | |
| *E. amylovora* | LMG 2024$^T$ | Z96088 |
| *E. persicinus* | LMG 11254$^T$, LMG 2691 | Z96086, AJ001190 |
| *E. rhapontici* | LMG 2688$^T$ | Z96087 |
| *E. psidii* | LMG 7034 | Z96085 |
| *E. tracheiphila* | LMG 2906$^T$ | Y13250 |
| *E. mallotivora* | LMG 2708$^T$ | Z96084 |
| *Brenneria* | | |
| *B. rubrifaciens* | LMG 2709$^T$ | Z96098 |
| *B. nigrifluens* | LMG 2694$^T$ | Z96095 |
| *B. alni* | ICMP 12657$^T$ | AJ223468 |
| *B. paradisiacal* | LMG 2542$^T$ | Z96096 |
| *B. salicis* | LMG 2698$^T$ | Z96097 |
| *B. quercina* | LMG 2724$^T$ | AJ223469 |
| *Pectobacterium* | | |
| *P. cypripedii* | LMG 2657$^T$ | Z96094 |
| *P. carotovora* | LMG 2404$^T$ | Z96089 |
| *Ssp. carotovora* | LMG 2466$^T$ | Z96091 |
| *Ssp. betavasculorum* | ICMP 9121$^T$ | AJ223408 |
| *Ssp. wasabiae* | LMG 2386$^T$ | Z96090 |
| *Ssp. atroseptica* | LMG 17566$^T$ | AJ223407 |
| *Ssp. odorifera* | LMG 2804$^T$ | Z96093 |
| *P. chrysanthemi* | LMG 17936$^T$ | AJ223409 |
| *P. cacticidum* | | |
| *Pantoea* | | |
| *P. agglomerans* | LMG 2565 | Z96082 |
| *P. agglomerans 2* | LMG 2660 | Z96083 |
| *P. stewartii ssp. stewartii* | LMG 2715$^T$ | Z96080 |
| *P. stewartii ssp. indologenes* | LMG 2632$^T$ | Y13251 |
| *P. ananatis* | LMG 2665$^T$ | Z96081 |
| *P. ananas* | ATCC 33243$^T$ | U80196 |
| *Shigella* | | |
| *S. flexneri* | | X80679 |
| *S. dysenteriae* | | X80680 |
| *S. boydii* | | X96965 |
| *S. sonnei* | | X96964 |
| *Escherichia* | | |
| *E. coli* | MG1655 | NC_000913 |
| *E. vulneris* | ATCC 33821$^T$ | X80734 |
| *E. hermanii* | | X80675 |
| *Salmonella* | | |
| *S. give* | | X80683 |
| *S. sofia* | | X80677 |
| *S. shomron* | | X80678 |

**Table 4** continued

| Taxon | Strain no. | GenBank accession no. |
|---|---|---|
| *S. typhimurium* | | X80681 |
| *S. paratyphi* | | U88547 |
| *Enterobacter* | | |
| *E. dissolvens* | LMG 2683$^T$ | Z96079 |
| *E. cancerogenus* | LMG 2693$^T$ | Z96078 |
| *E. nimipressuralis* | LMG 10245$^T$ | Z96077 |
| *E. sp. HC B* | | U39556 |
| *Klebsiella* | | |
| *K. pneumoniae* | DSM30104 | X87276 |
| *K. planticola* | DSM3069$^T$ | X93215 |
| *Yersinia* | | |
| *Y. pestis* | Evpst + c | Z75317 |
| *Y. mollaretii* | ER-2975 | X75280 |
| *Y. bercovieri* | ER-2937 | X75281 |
| *Y. rohdei* | ER-2935 | X75276 |
| *Y. kristensenii* | ER-2819 | X75278 |
| *Y. aldovae* | ATCC 35236 | X75277 |
| *Y. intermedia* | ER-3854 | X75274 |
| *Y. ruckeri* | ATCC 29473 | X75275 |
| *Y. enterocolitica* | ER-26039-92 | Z49829 |
| *Serratia* | | |
| *S. marcesens* | ATCC 13880$^T$ | M59160 |
| *Citrobacter* | | |
| *C. freundii* | | M59291 |
| *Hafnia* | | |
| *H. alvei* | | M59155 |
| *Rahnella* | | |
| *R. aquatilis* | ATCC 33989 | X79939 |
| *Ewingella* | | |
| *E. americana* | | X88848 |
| *Pleisiomonas* | | |
| *P. shigelloides* | ATCC 14029$^T$ | X74688 |
| *Outgroups* | | |
| *Photorhabdus luminescens* | DSM 3368 | X82248 |
| *Xenorhabdus nematophilus* | DSM 3370 | X82251 |
| *Xenorhabdus bovienii* | DSM 4766 | X82252 |
| *Xenorhabdus bediingii* | DSM 4767 | X82254 |
| *Xenorhabdus poinnarii* | DSM 4768 | X82253 |
| *Xenorhabdus japonicus* | IAM 14265$^T$ | D78008 |
| *Arsenophonus nasoniae* | ATCC 49151 | M90801 |
| *Proteus vulgaris* | | J01874 |

Sequence kits (Applied Biosystems) according to the manufacturer's directions, except that final reaction volumes were halved (10 µl) and the amount of BigDye was quartered (2 µl); reactions were run on an ABI 377 DNA sequencer according to the manufacturer's directions. DNA sequences were edited and combined in Sequencer v.4.1 (Gene Codes Corp.). Sequences have been deposited in GenBank under accession numbers EU490590 to EU490613.

### Sequence Alignment and Phylogenetic Analyses

DNA sequences for all four data sets were aligned using CLUSTAL W v1.5 (Thompson et al. 1992), followed by manual adjustments in MacClade v.4.0 (Maddison and Maddison 2000). Regions totaling approximately 100 bp were excluded from the beginning and end of each data set to avoid gaps due to missing data. Phylogenetic analyses for all data sets were performed using PAUP* (Swofford 2001) and GARLI (Zwickl 2006). The aligned sequences were first analyzed in PAUP* using maximum parsimony (MP) under heuristic search methods using tree bisection-reconnection (TBR) branch-swapping and 100 random taxon addition replicates. The shortest trees were then used as the starting topology for the evaluation of 16 nested models of sequence evolution (Frati et al. 1997; Sullivan et al. 1997; Swofford et al. 1996). Parameter space was searched for the best tree with simultaneous estimation for model parameters using a ML search conducted in GARLI. As recommended by Zwickl (2006), multiple runs were performed ($\sim$20) to ensure that results are consistent as the algorithm is stochastic. The log likelihood values of each run were retained in order to compare the individual runs.

Branch support was determined by 100 ML bootstrap iterations in GARLI (Felsenstein 1985) and with Bayesian posterior probability (MrBayes v.3.1.2) approximation of 1 million generations discarding 25% (2500) of the tree samples as recommended in the MrBayes v.3.1.2 manual (Huelsenbeck and Ronquist 2001). For Bayesian character support methods, parameters of sequence evolution estimated from the final ML tree were used (Table 5). Only support values with a bootstrap score of 50 or better and a Bayesian probability of 0.95 or higher appear in Figs. 1–3.

## Results

The 16S rDNA alignments of data sets A and B consisted of 33 and 49 taxa, respectively, and 1282 characters. The alignment of data set C (Kwon et al. 1997) consisted of 22 taxa and 1456 characters, while the alignment of data set D (Hauben et al. 1998) consisted of 67 taxa and 1472 characters. For all four alignments, longer sequences were truncated at the beginning and end in order to eliminate gaps from the termini of the data sets. The range of variation observed in log likelihood scores between runs was very small, and the tree and model parameters

**Table 5** Maximum likelihood model parameters used in the phylogenetic analyses

| Gene | Model | Base frequencies | Rate matrix | I + Γ |
|---|---|---|---|---|
| 16S (our data set; Fig. 1a) | GTR + I + Γ | A, C, G, T (0.246346, 0.233858, 0.317755, 0.202042) | AC, AG, AT, CG, CT, GT (1.51592, 3.33956, 1.71604, 1.21563, 6.34090, 1.00000) | 0.724256, 0.559444 |
| 16S (our data set; Fig. 1b) | GTR + I + Γ | A, C, G, T (0.241436, 0.229258, 0.318765, 0.210539) | AC, AG, AT, CG, CT, GT (0.973277, 2.46077, 1.60564, 1.29081, 3.98952, 1.00000) | 0.529117, 0.352613 |
| 16S (Kwon et al. data set; Fig. 2) | GTR + I + Γ | A, C, G, T (0.246049, 0.229034, 0.317297, 0.207621) | AC, AG, AT, CG, CT, GT (0.84157, 4.41299, 1.14315, 1.54295, 6.62153, 1.00000) | 0.702197, 408220 |
| 16S (Hauben et al. data set; Fig. 3) | GTR + I + Γ | A, C, G, T (0.242669, 0.226653, 0.313996, 0.216682) | AC, AG, AT, CG, CT, GT (1.11692, 3.90284, 1.84386, 2.10014, 5.10788, 1.00000) | 0.692247, 0.468159 |

corresponding to the best score were used. The best ML trees are presented in Figs. 1–3.

In Fig. 1a, *Pantoea* forms a monophyletic clade, whereas *Erwinia*, *Brenneria*, and *Pectobacterium* are all polyphyletic. The *Pantoea* clade is derived from within a group of erwinias, with *Erwinia lupinicola* strain 1 as the basal taxon of the *Erwinia–Pantoea* clade. *Erwinia rhapontici* 2, *Pectobacterium chrysanthemi* 2, and *Pectobacterium carotovorum* ssp. *carotovora* 2 form a clade that is sister to the *Erwinia–Pantoea* clade. Finally, the two *E. psidii* strains and *E. rhapontici* 1 group with the *Brenneria* and additional *Pectobacterium* taxa.

In Fig. 1b, a markedly different phylogeny is revealed from the data set containing the 18 animal pathogen taxa. The *E. rhapontici* 2, *P. chrysanthemi* 2, and *P. carotovorum* ssp. *carotovora* 2 clade is no longer sister group to the *Erwinia–Pantoea* clade, however, *Pantoea* is still derived from within the erwinias. *Pantoea*, which remains monophyletic, has *E. lupinicola* 1 as its closest relative, and together they are sister group to *E. amylovora* 1, *E. mallotivora* 1 and 2, and *E. tracheiphila* 1. Taxa are also rearranged in the large plant pathogen clade of *Erwinia*, *Brenneria*, and *Pectobacterium* species, with some examples including *Brenneria alni* 1, *E. psidii* 1, and *P. carotovorum* ssp. *carotovora* 1, though there is little support for conflict in this large clade. Thus, from these two trees, which include the same 27 plant pathogen sequences and only differ in the number of animal pathogen representatives and outgroup taxa, it is apparent that placement of the enterobacterial phytopathogenic species is affected by the number of animal pathogen taxa included in the phylogenetic analyses.

Figures 2 and 3 also depict different phylogenetic patterns, though many nodes lack support. In both trees, *Erwinia*, *Brenneria*, and *Pantoea* each form monophyletic clades, whereas *Pectobacterium* is polyphyletic. Additionally, *Erwinia* and *Pantoea* species consistently group as each other's closest relatives, as do *Brenneria* and *Pectobacterium*. However, the placement of *Pectobacterium cypripedii* varies in the two phylogenies, with *P. cypripedii* serving as the closest relative to the *Erwinia* clade in Fig. 2 and to the *Brenneria* clade in Fig. 3. The two trees also differ from the previously published phylogenies of both Kwon et al. (1997) and Hauben et al. (1998). However, because support values for most of the nodes are very low in both the previously published and the reanalyzed trees, these differences can likely be attributed to poorly supported and unstable nodes, rather than to differences among analytical methods.

It is noteworthy that none of the phylogenies presented in this study are consistent with the current placement of the plant pathogen genera within the Enterobacteriaceae. The trees resulting from the first two data sets consisting of *Erwinia*, *Brenneria*, *Pectobacterium*, and *Pantoea* sequences generated in our laboratory support multiple independent origins for the phytopathogenic enterobacterial species (Fig. 1). The third tree, derived from a reanalysis of the Kwon et al. (1997) data (data set C), supports a single origin for the phytopathogens (Fig. 2), whereas the fourth tree, derived from a reanalysis of the Hauben et al. (1998) data (data set D), supports two independent origins, one for *Erwinia* and *Pantoea* and one for *Brenneria* and *Pectobacterium* (Fig. 3). Moreover, it is important to note that even though the same plant pathogen species are represented in all four trees, the strains used to obtain the 16S rDNA gene sequences in data sets A and B are different from those used by Kwon et al. (1997) or by Hauben et al. (1998). Nonetheless, these analyses indicate that 16S rDNA gene sequences may be limited in their ability to reconstruct the phylogeny of enterorbacterial plant pathogens at the genus level.

To identify the reasons that 16S rDNA gene trees are so different from one another, we examined the data set alignments, which consisted of large highly conserved regions interspersed by small highly variable regions (Fig. 4). From our 16S rDNA alignment of the 1282 bp in data set B, we were able to identify approximately six hypervariable domains. The conserved parts of the gene sequences were nearly identical for all the enterobacterial taxa, whereas the variable parts seemed to consist of a random array of nucleotides. The observed rate heterogeneity is likely to reflect evolutionary constraints related to

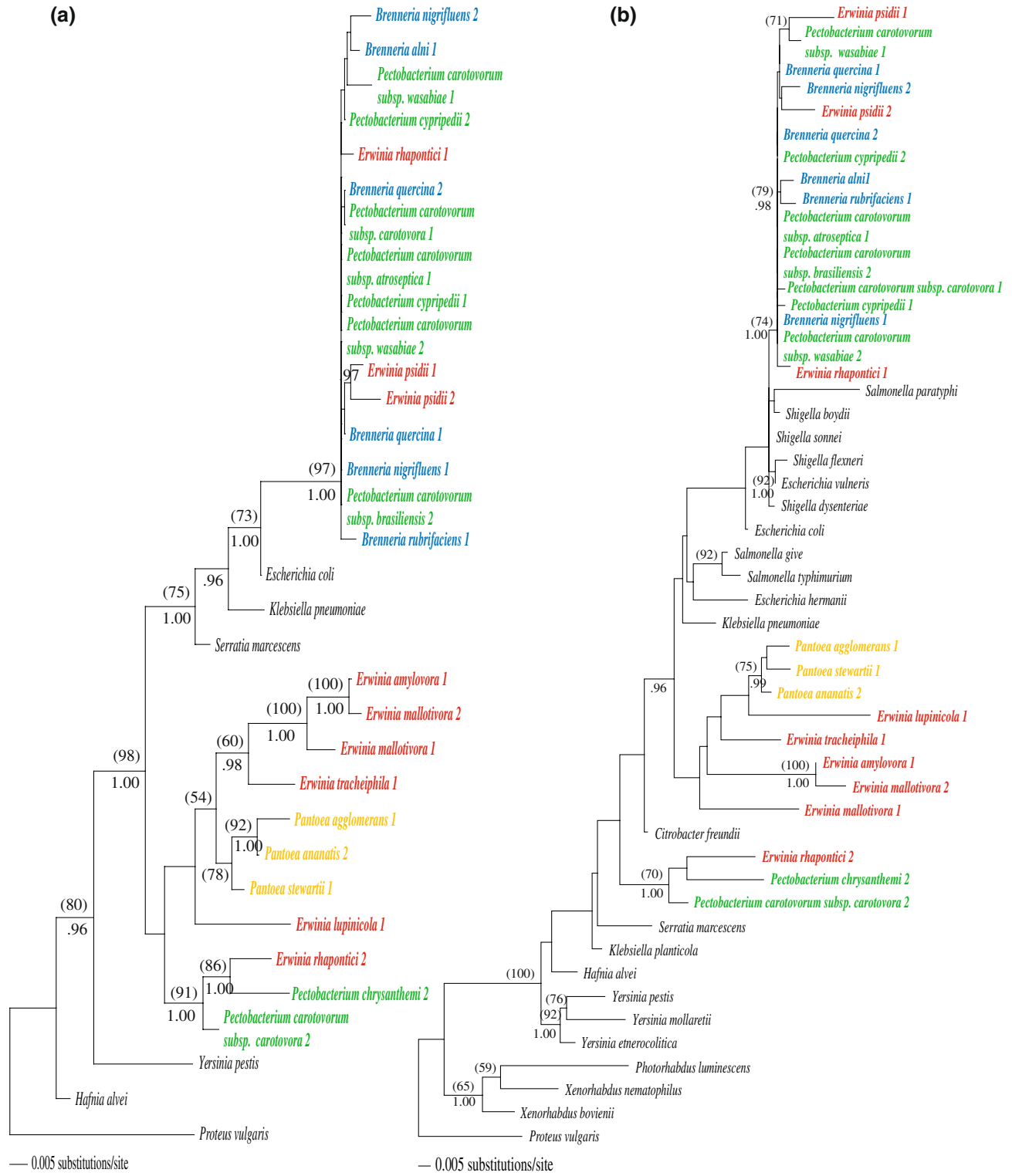


**Fig. 1** Maximum likelihood (ML) analysis of 16S rDNA sequences from plant and animal pathogen enterobacterial species. (**a**) Likelihood tree comprising 16S rDNA gene sequences from 27 plant pathogen and 5 animal pathogen enterobacterial strains. The log likelihood range for this analysis was $-\ln L$: 5086.66695 to 5086.66606. (**b**) Likelihood tree comprising 16S rDNA gene sequences from 27 plant pathogen and 18 animal pathogen enterobacterial representatives. The log likelihood range for this analysis was $-\ln L$: 5887.44756 to 5887.48735. ML solutions were generated in GARLI (reference) and visualized in PAUP* (reference). ML bootstrap nodal support was generated in PAUP* and subsequent values are reported for each node in parentheses. Bayesian nodal support values are presented for all nodes without parentheses

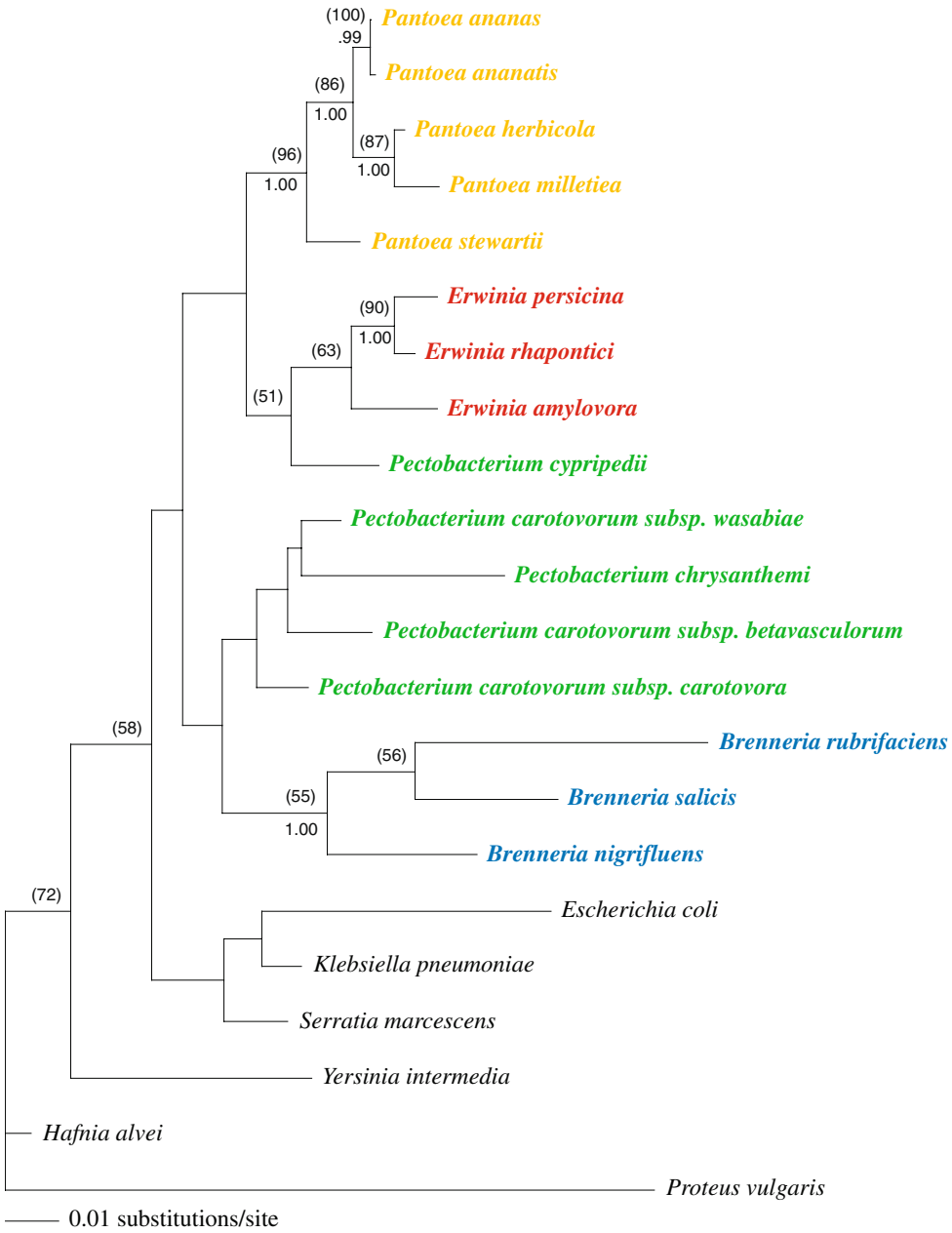


**Fig. 2** Phylogenetic analysis of the Kwon et al. (1996) 16S rDNA enterobacterial sequence data set. The resultant maximum likelihood (ML) tree for this data set was produced in GARLI and viewed in PAUP*. The log likelihood range for this analysis was $-\ln L$: 10118.59778 to 10118.73990. ML bootstrap nodal support was generated in PAUP* and subsequent values are reported for each node in parentheses. Bayesian nodal support values are presented for all nodes without parentheses

rRNA function. Some of the stem-and-loop structures contributing to the formation of the ribosomal subunit interact with other components and these structures must be conserved. Therefore, as is the case for *Streptomyces* (Ueda et al. 1999), portions of the 16S rDNA gene are subject to different evolutionary pressures.

To evaluate the resolution provided by the conserved portions of the gene sequences, we edited data set B, eliminating all variable regions and analyzing only the conserved regions of the gene sequences. Of the original alignment comprising a total of 1282 bp, 67 bp was excluded from the phylogenetic analysis. These included positions 338–362, 523–533, 885–908, and 1017–1027. The resulting ML tree (Fig. 5) resolved minimal nodes differentiating the outgroup and ingroup taxa, as well as the *Yersinia*, the *Pantoea*, and some *Erwinia* species, while most terminal nodes were unresolved and unsupported, resulting in a widespread polytomy. For example, the largest portion of the tree containing *Brenneria quercina* 2, *Pectobacterium cypripedii* 1, *Escherichia coli*, *Escherichia vulneris*, *Shigella dysenteriae*, *Shigella sonnei*, and others failed to resolve relationships among this diverse group of plant and animal pathogen representatives. As a result, different regions of the 16S rDNA gene sequence provide different phylogenetic signals, and the topology of the trees changes according to the number of taxa included.

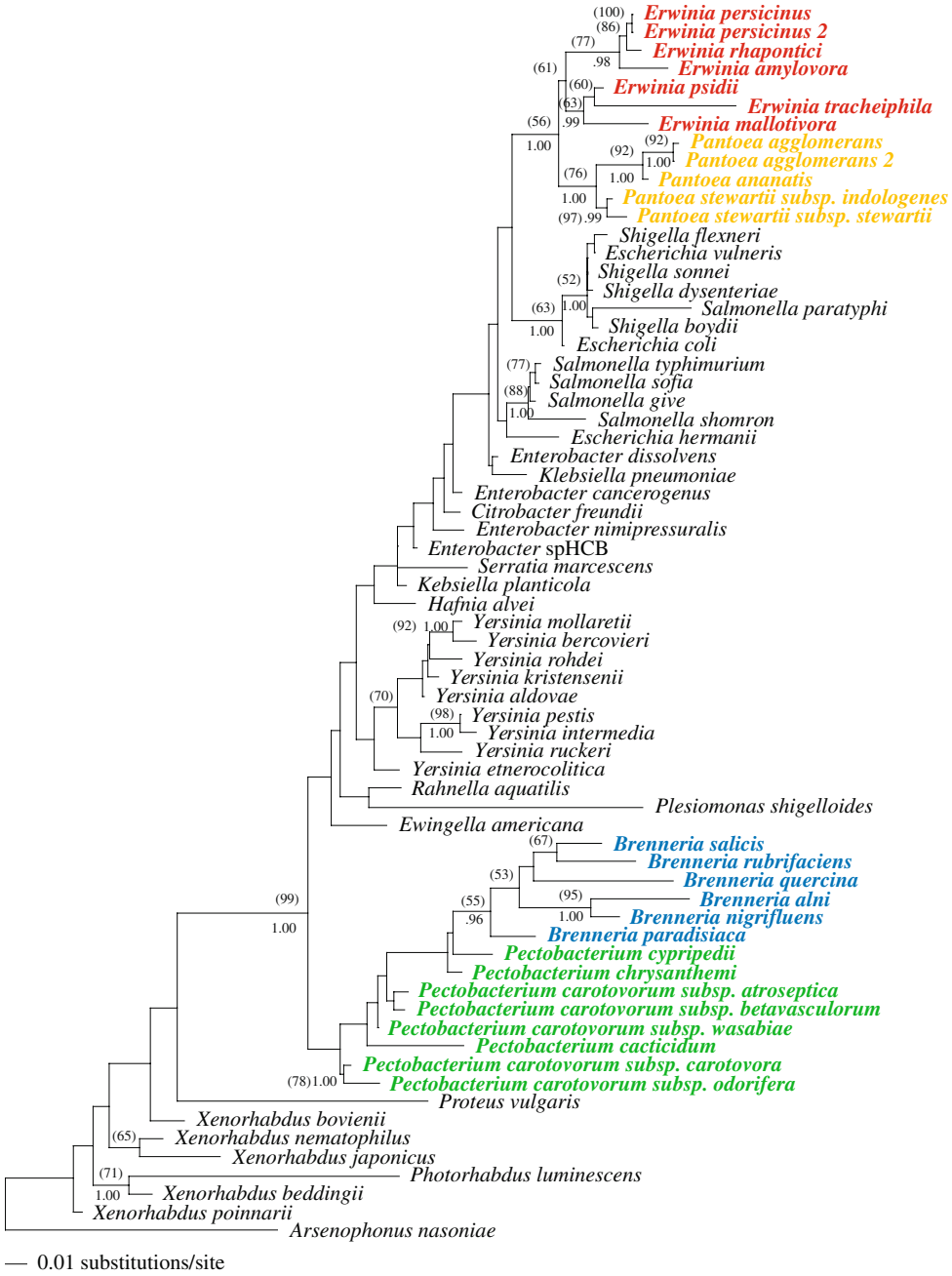


**Fig. 3** Phylogenetic analysis of the Hauben et al. (1997) 16S rDNA enterobacterial sequence data set. The resultant maximum likelihood (ML) tree for this data set was produced in GARLI and viewed in PAUP*. The log likelihood range for this analysis was $-\ln L$: 5149.07892 to 5149.07908. ML bootstrap nodal support was generated in PAUP* and subsequent values are reported for each node in parentheses. Bayesian nodal support values are presented for all nodes without parentheses

It was observed that the deeper nodes in the four phylogenies (Figs. 1–3) were not well supported by either the ML bootstrap or the Bayesian values. In the first two trees, the deeper nodes are denoted by fairly low bootstrap values but high Bayesian values (Fig. 1), whereas in Figs. 2 and 3 most support values, including many of those more derived in the tree, are below 50 and 0.95 respectively, lending no support to the backbone of the two phylogenies. It is interesting that low bootstrap support values were also present in both the Kwon et al. (1997) and the Hauben et al. (1998) original published phylogenies. Thus, we cannot confidently determine whether the plant pathogen genera within the Enterobacteriaceae are monophyletic as in Fig. 2 or polyphyletic as in Figs. 1 and 3.

## Discussion

Despite known variation between previously published 16S rDNA gene phylogenies of enterobacterial species (Hauben et al. 1998; Kwon et al. 1997; Martinez et al. 2004; Sproer et al. 1999; Young 2001), it was anticipated that use of the same methods for alignment and analysis of the four data sets would result in robust and consistent phylogenies.

Unexpectedly, however, the striking inconsistency present in the 16S rDNA gene trees resulting from our analyses (Figs. 1–3) has signaled a potential difficulty in correctly revealing the taxonomic history of the enterobacterial plant pathogens for this gene.

The striking difference in sequence conservation observed among various segments of the 16S rDNA sequences analyzed here warrants further discussion. It was apparent that conserved regions of the gene offer little taxonomic resolution below the family level, because they are not variable among closely related enterobacterial taxa. These same regions do, however, differ among more distantly related taxa, such as the Enterobacteriaceae and Pseudomonadaceae (Toth et al. 2006), and have been useful in defining higher-level bacterial groupings (Olsen and Woese 1993). The variable 16S rDNA domains, on the other hand, appear potentially unique to species, and have been used to discriminate among closely related taxa (Louws et al. 1999). Even among species-level analyses, however, these variable regions may accrue substantial background noise (homoplasy), as additional taxa are included in analyses. Consequently, in our analyses few stable characters can be found supporting the different groups of taxa, resulting in inconsistent and poorly supported phylogenies. This was confirmed through our analysis of the conserved 16S rDNA sequence regions (Fig. 5).

The potentially confounding phylogenetic signal emanating from the variable portions of 16S rDNA warrants a further cautionary note when undertaking short repetitive sequencing and genotyping of populations. This is particularly true for methods such as pyrosequencing and primer extension assays where only a few nucleotide substitutions are culled for diagnostic identification or diversity assessment (Ahmadian et al. 2000; Jordan et al. 2005; Monstein et al. 2001). For assays targeting 16S rDNA sequences, the end result of any short sequencing approach may be confounded, particularly if the informative SNP(s) resides within the variable domains noted here. Identification or diversity assessments that rely on such targets, especially in the absence of any surrounding sequence information from the 16S rDNA gene itself or other conserved genes, may do more to obscure our understanding of strain relationships than to accurately delineate evolutionary relatedness.

Substantial topological differences were noted between the trees reported here. While observed differences between the two sets of trees (Figs. 1a and 2, Figs. 1b and 3) could result from homoplasy in the highly variable regions of the 16S rDNA gene sequences, intraspecific variation may have also been a contributor to these differences. Here, we provide support for the presence of both homoplasy and intraspecific variation from our analyses. The trees presented in Fig. 1 consist of the same plant pathogen sequences, and only vary in the number of animal pathogen and outgroup taxa. As a result, observed differences between Fig. 1a and 1b cannot be attributed to intraspecific variation among *Erwinia*, *Brenneria*, *Pectobacterium*, and *Pantoea* strains, but are most probably the result of increased character conflict introduced by the additional sequences. Alternatively, in Figs. 1a and 2, we kept the number of animal pathogen representatives and outgroup taxa constant, and used sequences from different plant pathogen accessions, still representing most of the same species. Hence, the differences between these two

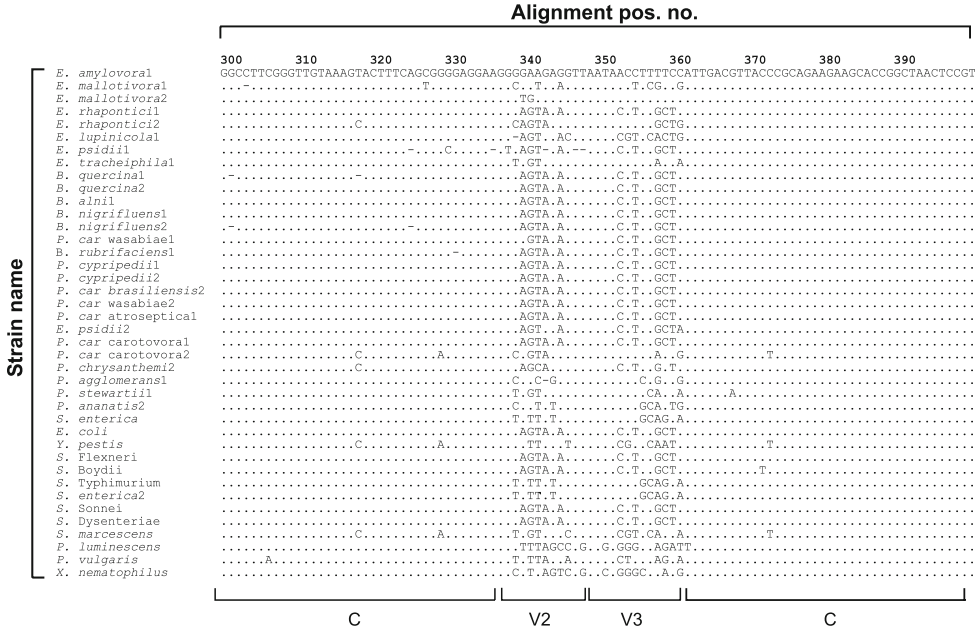


**Fig. 4** A portion of the 16S rDNA nucleotide sequence alignment for 40 enterobacterial species/strains analyzed here depicting two of the six hypervariable sequence regions (designated 'V2' and 'V3') and adjacent conserved nucleotide segments (designated 'C'). The region of the alignment shown here was extracted from the full-length sequence alignment of the 16S rDNA data set used to construct Fig. 1b and exemplifies the conserved and variable regions scattered throughout the 16S rDNA sequence alignment. The nucleotide positions are numbered at the top of the alignment and correspond to the full-length alignment analyzed in Fig. 1b

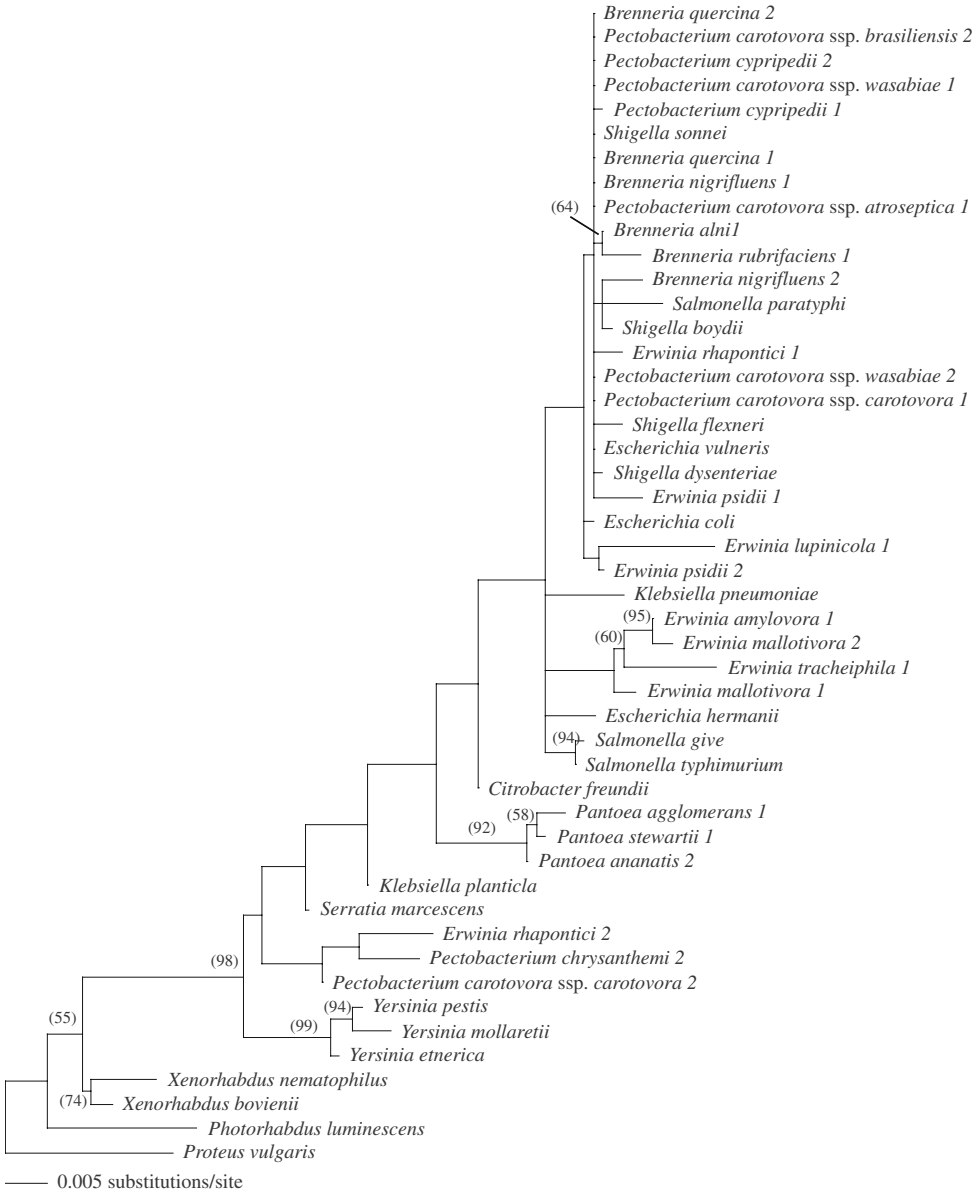


**Fig. 5** Maximum likelihood (ML) analysis of the conserved 16S rDNA sequence regions exclusively. The 49 taxa included here are the same as those analyzed in Fig. 1b. The likelihood tree shown here was produced in GARLI and viewed in PAUP* and illustrates the loss of phylogenetic resolution when hypervariable domains are omitted from the alignment. ML bootstrap nodal support was generated in PAUP* and subsequent values are reported for each node in parentheses

trees are not due to an increase in animal pathogen representatives but, rather, appear to result from intraspecific variation (e.g., homoplasy between members of the same species in the hypervariable regions of the gene). This suggests that the phylogenetic positions of species might be affected by sequence variation among strains. As a result, caution should be taken when drawing conclusions about taxonomic groups from reconstructions that fail to account for intraspecific variation.

In general, the 16S rDNA data failed to recover the basal and intermediate nodes. Subsequently, our analyses were unable either to characterize the relationships among the plant pathogen species or to position the genera within the Enterobacteriaceae. As noted previously, many of the deep nodes are poorly supported and unstable. Poorly supported nodes were evident in the previously published phylogenies of enterobacterial taxa as well (Hauben et al. 1998; Kwon et al. 1997; Moran et al. 2005; Sproer et al. 1999), yet genera, species, and subspecies are revised continually, largely on the basis of these unstable nodes resulting from only a few phylogenetically informative sequence changes.

Taken together, these data indicate that current enterobacterial 16S rDNA phylogenies of the plant pathogen genera *Erwinia*, *Brenneria*, *Pectobacterium*, and *Pantoea* may be confounded, thus weakening their utility as taxonomic anchoring points for additional studies such as ascertaining the lateral gene transfer pathways of virulence. The observed inconsistencies among these 16S rDNA phylogenies result from a change in the number of homoplasious sites present, which could be explained by one of

two mechanisms. The variable and conserved domains of the 16S rDNA gene are subject to different functional constraints, and as a result, a more relaxed tolerance of mutations in the hypervariable regions would result in an increased number of homoplasious sites. Alternatively, horizontal gene transfer of short segments of DNA into these regions, which do not affect gene function, would also serve to confound phylogenetic signal. These pitfalls in part have prompted a move toward combined sequence analyses of multiple loci when undertaking taxonomic studies. It should be stressed that while 16S rDNA analysis is useful for identifying bacterial strains and reconstructing relationships among large and disparate taxonomic groups, it may retain less utility when attempting reconstructions among more closely related bacterial taxa. Additionally, the observed tree differences, attributed to intraspecific variation, reinforce the notion that using a single bacterial strain to represent a species may not be practical when reconstructing microbial phylogenies, particularly where taxonomical inferences are drawn and species or genera are emended. The quest for a robust and informative DNA-sequence-based phylogeny of *Erwinia*, *Brenneria*, *Pectobacterium*, and *Pantoea* will likely require examination and analysis of other less universally conserved housekeeping genes.

**Acknowledgments** The authors thank A. Charkowski of the University of Wisconsin, Madison, for very generously providing the *Pectobacterium* strains used in this study, and Hank Howe, Kieth Lampel, Diane McCarthy, and Michael Jorgensen for valuable comments on early versions of the manuscript. The research was supported by NSF Grant DEB 0426194 to R. J. Mason-Gamer and a UIC provosts award to M. Naum.

## References

Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P, Uhlen M, Lundeberg L (2000) Single-nucleotide polymorphism analysis by pyrosequencing. Anal Biochem 280(1):103–110

Charlebois RL, Beiko RG, Ragan MA (2003) Microbial phylogenomics: branching out. Nature 421:217

Christensen H, Kuhnert P, Olsen JE, Bisgaard M (2004) Comparative phylogenies of the housekeeping genes *atpD*, *infB*, and *rpoB* and the 16S rRNA gene within the Pasteurellaceae. Int J Sys Evol Microbiol 54:1601–1609

Davidov Y, Jurkevitch E (2004) Diversity and evolution of *Bdellovibrio*-and-like organisms (BALOs), reclassification of *Bacteriovax starrii* as *Peredibacter starrii* gen. nov., comb. nov., and description of the *Bacteriovorax-Peredibacter* clade as Bacteriovoracaceae fam. nov. Int J Sys Evol Microbiol 54:1439–1452

Dorsch M, Lane D, Stackenbrandt E (1992) Towards a phylogeny of the genus *Vibrio* based on 16S rRNA sequences. Int J Syst Bacteriol 42:58–63

Ehrlich HA, Gelford D, Snisky JJ (1991) Recent advances in the polymerase chain reaction. Science 252:1643

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fessehaie A, De Boer S, Levesque C (2002) Molecular characterization of DNA encoding 16S–23S rRNA intergenic spaces regions and 16S rRNA of pectolytic *Erwinia* species. Can J Microbiol 48:387–398

Frati F, Simon C, Sullivan J, Swofford DL (1997) Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. J Mol Evol 44:145–158

Gould SJ (1996) Planet of bacteria. In: Full house: the spread of excellence from Plato to Darwin. Harmony Books, New York

Hauben L, Moore ERB, Vauterin L, Steenackers L, Mergaert J, Verdonck L, Swings J (1998) Phylogenetic position of phytopathogens within the Enterobacteriaceae. Syst Appl Microbiol 21:384–397

Huelsenbeck PH, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Jordan JA, Butchko AR, Durso MB (2005) Use of pyrosequencing of 16S rRNA fragments to differentiate between bacteria responsible for neonatal sepsis. J Mol Diagn 7(1):105–110

Krieg NR, Holt JG (1985) Bergey's manual of systematic bacteriology. Williams and Wilkins, New York

Kwon SW, Go SJ, Kang HW, Ryu JC, Jo JK (1997) Phylogenetic analysis of *Erwinia* species based on 16S rRNA gene sequences. Int J Sys Bacteriol 47:1061–1067

Louws FJ, Rademaker JLW, de Bruijn FJ (1999) The three Ds of PCR-based genomic analysis of phytobacteria: diversity, detection, and disease diagnosis. Annu Rev Phytopathol 37:81–125

Maddison WP, Maddison DR (2000) MacClade. Sinauer, Sunderland, MA

Manniatis T, Fritsch E, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Springs Harbor Laboratory, Cold Spring Harbor, NY

Martinez J, Martinez L, Rosenblueth M, Silva J, Martinez-Romero E (2004) How are gene sequence analyses modifying bacterial taxonomy? The case of Klebsiella. Int Microbiol 7:261–268

Mason-Gamer RJ (2004) Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. Syst Biol 53:25–37

Monstein HJ, Nikpour-Badr S, Jonasson J (2001) Rapid molucular identification and subtyping of *Helicobacter pylori* by pyrosequencing of the 16S rDNA variable V1 and V3 regions. FEMS Microbiol Lett 199(1):103–107

Moran NA, Russel JA, Koga R, Fukatsu T (2005) Evolutionary relationships of three new species of Enterobacteriaceae living as symbionts of aphids and other insects. Appl Env Microbiol 71:3302–3310

Olsen GJ, Woese CR (1993) Ribosomal RNA: a key to phylogeny. FASEB J 7:113–123

Sproer C, Mendrock U, Swiderski J, Lang E, Stackenbrandt E (1999) The phylogenetic position of *Serratia*, *Buttiauxella* and some other genera of the family Enterobacteriaceae. Int J Sys Bacteriol 49:1433–1438

Sullivan J, Markert JA, Kilpatrick CW (1997) Phylogeography and molecular systematics of the *Peromyscus aztecus* species group (Rodentia: *Muridae*) inferred using parsimony and likelihood. Syst Biol 46:426–440

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. Sinauer, Sunderland, MA

Swofford DL (2001) PAUP*. Phylogenetic analysis using parsimony (*and other methods), v. 4b10. Sinauer, Sunderland, MA

Thompson JD, Higgins D, Gibson TJ (1992) CLUSTAL version W: a novel multiple sequence alignment program. Nucleic Acids Res 22:4673–4680

Toth IK, Pritchard L, Birch PRJ (2006) Comparative genomics reveals what makes an enterobacterial plant pathogen. Annu Rev Phytopathol 44:305–336

Ueda K, Seki T, Kudo T, Yoshida T, Kataoka M (1999) Two distinct mechanisms cause heterogeneity of 16S rRNA. J Bacteriol 181:78–82

van der Zwet T, Beer SV (1995) Fire blight—its nature, prevention, and control: a practical guide to integrated disease management. U.S. Dept Agr Agr Inform Bull 631:83

Wertz JE, Goldstone G, Gordon DM, Riley MA (2003) A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. J Evol Biol 16:1236–1248

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Young JM (2001) Implications of alternative classifications and horizontal gene transfer for bacterial taxonomy. Int J Sys Evol Microbiol 51:945–953

Zwickl D (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Biological Sciences. University of Texas, Austin