

Phylogenetic Analysis of *Pasteuria penetrans* by Use of Multiple Genetic Loci

Lauren Charles,¹ Ignazio Carbone,² Keith G. Davies,³ David Bird,¹ Mark Burke,¹
Brian R. Kerry,³ and Charles H. Opperman^{1*}

Center for the Biology of Nematode Parasitism, Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina 27606¹; Center for Integrated Fungal Research, Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina 27606²; and Nematode Interactions Unit, Rothamsted Research, Harpenden, Herts AL5 2JQ, United Kingdom³

Received 25 January 2005/Accepted 4 May 2005

Pasteuria penetrans is a gram-positive, endospore-forming eubacterium that apparently is a member of the *Bacillus-Clostridium* clade. It is an obligate parasite of root knot nematodes (*Meloidogyne* spp.) and preferentially grows on the developing ovaries, inhibiting reproduction. Root knot nematodes are devastating root pests of economically important crop plants and are difficult to control. Consequently, *P. penetrans* has long been recognized as a potential biocontrol agent for root knot nematodes, but the fastidious life cycle and the obligate nature of parasitism have inhibited progress on mass culture and deployment. We are currently sequencing the genome of the *Pasteuria* bacterium and have performed amino acid level analyses of 33 bacterial species (including *P. penetrans*) using concatenation of 40 housekeeping genes, with and without insertions/deletions (indels) removed, and using each gene individually. By application of maximum-likelihood, maximum-parsimony, and Bayesian methods to the resulting data sets, *P. penetrans* was found to cluster tightly, with a high level of confidence, in the *Bacillus* class of the gram-positive, low-G+C-content eubacteria. Strikingly, our analyses identified *P. penetrans* as ancestral to *Bacillus* spp. Additionally, all analyses revealed that *P. penetrans* is surprisingly more closely related to the saprophytic extremophile *Bacillus haladurans* and *Bacillus subtilis* than to the pathogenic species *Bacillus anthracis* and *Bacillus cereus*. Collectively, these findings strongly imply that *P. penetrans* is an ancient member of the *Bacillus* group. We suggest that *P. penetrans* may have evolved from an ancient symbiotic bacterial associate of nematodes, possibly as the root knot nematode evolved to be a highly specialized parasite of plants.

Pasteuria penetrans is an endospore-forming, gram-positive, obligate parasitic bacterium of the root knot nematodes (RKN), *Meloidogyne* spp. RKN have a very broad host range, which includes more than 2,000 plant species, and most cultivated crops are attacked by at least one species of *Meloidogyne* (33); this causes economic losses of more than \$50 billion per year. The problem in the subtropics and tropics is particularly severe, and many developing nations are seriously affected in terms of both food security and economics by RKN. Mature female RKN release hundreds of eggs into a proteinaceous matrix on the surface of the root. Following a first molt in the egg, a motile second-stage (J2) juvenile hatches in the soil and typically reinfects the same plant. The RKN J2 destructively penetrates the root, preferentially in the zone of elongation or at the site of lateral root emergence, and migrates intercellularly into the vascular cylinder, causing little or no injury. Once in the vascular cylinder, the nematode makes a commitment to establish a highly specialized feeding site, referred to as a giant cell. The relationship between an RKN and its host is both intimate and complex and involves dramatic changes both in the plant and in the nematode, leading to giant cell induction and gall formation. The *Meloidogyne* J2 stage is a nonfeeding, developmentally arrested, long-lived dispersal stage and can

survive in the soil for weeks or even months on stored lipid reserves, and this is the nematode stage exposed to *P. penetrans* spores in the soil.

The life cycle of this bacterium has coevolved with its host and begins when a J2 migrating through the soil becomes encumbered with endospores. The endospores do not germinate until the J2 has entered the plant root and established a feeding site. Sometime between the establishment of a feeding site and the second nematode molt, an endospore germinates and produces rhizoids which extend throughout the developing nematode. The rhizoids eventually produce bacterial rods that undergo rapid exponential growth, resulting in degeneration of the nematode's reproductive tract and inhibition of egg production. Sporogenesis is triggered in a manner similar to the manner observed for other bacilli and involves the initiation of a metal ion-sensitive, phosphorelay pathway (16) that results in the production of endospores. Because of both its high efficacy and its host specificity, *P. penetrans* represents a potentially ideal biological control agent for the economically important *Meloidogyne* crop pests (5, 26, 37, 43). However, the obligate nature of the bacterium's life style and its host specificity have made it difficult to develop *P. penetrans* into a commercial product. For these reasons, a genomic approach has recently been used to help understand the mechanisms of parasitism of *Pasteuria* spp. and the possible exploitation of their ecological niche (9).

There have been many attempts to classify this group of bacteria since it was first described as *Pasteuria ramosa* by

* Corresponding author. Mailing address: Center for the Biology of Nematode Parasitism, Department of Plant Pathology, North Carolina State University, Raleigh, NC 27606. Phone: (919) 515-6699. Fax: (919) 515-9500. E-mail: warthog@ncsu.edu.

Metchnikoff (23) in 1888 on *Daphnia*, a waterflea. In 1906, Cobb (6) studied the morphology of this parasite on the nematode *Dorylaimus bulbiferous* and claimed that it should be placed among the protozoans. This change was later accepted, and the bacterium was renamed *Duboscqia penetrans* by Thorne (40) in 1940. Since then, electron microscope techniques have shown that the bacterium is more *Bacillus*-like than protozoan, and hence, it was renamed again, as *Bacillus penetrans* (21). In 1985, Metchnikoff's work was rediscovered, and findings suggested that there were similarities to the original *P. ramosa*, which led to reversion of the genus back to *Pasteuria*, as described by Sayre and Starr (34). Because *P. penetrans* is a gram-positive, mycelial, endospore-forming bacterium, it was classified in the *Actinomycetales* (34).

Multilocus sequence typing exploits the variation that slowly accumulates in housekeeping genes in populations and is a selectively neutral method for characterizing bacteria (20). This approach has been shown to be extremely useful for characterizing pathogenic strains (44). Multilocus enzyme electrophoresis was never exploited to characterize *P. penetrans*, due in large part to the obligate parasitic nature of this organism and the difficulty in separating *Pasteuria* enzymes from host enzymes. However, the acquisition of genome sequence data for *P. penetrans* makes multilocus sequence typing a feasible approach and provides a method for inferring the organism's closest relatives, along with its phylogenetic history.

Automated sequencing techniques have been used with single genes to reevaluate the classification of this bacterium. Studies using 16S rRNA genes and rRNA have indicated that *Pasteuria* spp., including *P. penetrans*, belong to the *Clostridium-Bacillus-Streptococcus* branch of gram-positive eubacteria (2, 3, 28). This finding has recently been supported through the use of the *spo0A* gene, which placed *Pasteuria* with members of the supergenus *Bacillus* (42). However, phylogenetic analytical procedures that rely on only one gene at a time to classify a species have been found to be inaccurate in many cases, while the use of multilocus approaches have been shown to be more reliable (12, 14, 27, 30, 31, 35, 48).

To resolve the placement of *P. penetrans* among bacteria, a deep phylogenetic analysis was performed with 40 housekeeping genes separately and concatenated. Further analyses were performed with a subset of 27 of these genes. Various phylogenetic methods, including Bayesian, maximum-likelihood, and maximum-parsimony methods, were used to further strengthen the validity of the analysis.

MATERIALS AND METHODS

Bacterial species from a supertree constructed by Daubin et al. (7) were searched in the National Center for Biotechnology Information database (45) to identify completely sequenced genomes. Completely sequenced genomes are necessary to ensure that the genes collected are orthologous. The species used, whose phyla are distributed throughout the superkingdom *Eubacteria*, were chosen to avoid taxon bias in the resulting trees (4).

In previous single-gene phylogenetic analyses, *P. penetrans* has been demonstrated to fall into the *Bacillus* clade (2, 3, 42), but the specific relationships are still unresolved. Kobayashi et al. (18) in 2003 identified the genes essential to *Bacillus subtilis*. Since these genes include independent, housekeeping genes that are known to evolve slowly, they have the highest probability of giving the most accurate and unbiased tree when they are concatenated (30). The genes chosen were blasted using a tblastn query (1) and Blossum62 matrix against *P. penetrans* unpublished sequence data with the following restrictions: a maximum e-value of 1e-20 and a minimum bit score of 100. The genes recovered were then evaluated

TABLE 1. Descriptions of genes used in phylogenetic analyses

Gene	Description
<i>accA</i> ^a	Fatty acid biosynthesis acetyl coenzyme A carboxylase (α subunit)
<i>adk</i>	Purine biosynthesis adenylate kinase
<i>argS</i>	tRNA synthetase arginyl-tRNA synthetase
<i>dapA</i> ^a	Diaminopimelate biosynthesis dihydrodipicolinate synthase
<i>dapB</i> ^a	Diaminopimelate biosynthesis dihydrodipicolinate reductase
<i>dnaB</i>	DNA replication-initiation of chromosome replication/membrane attachment
<i>dnaE</i>	DNA replication DNA polymerase III (α subunit)
<i>eno</i> ^{a,b}	Glycolysis enolase
<i>fnt</i>	tRNA ^{Met} modification methionyl-tRNA formyltransferase
<i>glmS</i>	Amino sugar metabolism L-glutamine-D-fructose-6-phosphate amidotransferase
<i>groEL</i> ^{a,b}	Protein folding class I heat shock protein (chaperonin)
<i>gyrA</i>	DNA packaging DNA gyrase (subunit A)
<i>gyrB</i>	DNA packaging DNA gyrase (subunit B)
<i>infA</i>	Translation initiation factor IF-1
<i>map</i> ^a	Protein modification methionine aminopeptidase
<i>metS</i>	tRNA synthetase methionyl-tRNA synthetase
<i>mraY</i>	Peptidoglycan biosynthesis
<i>murB</i>	Peptidoglycan biosynthesis UDP-N-acetylenolpyruvylglucosamine reductase
<i>murC</i> ^{a,b}	Peptidoglycan biosynthesis UDP-N-acetylmuramate-alanine ligase
<i>murD</i>	Peptidoglycan biosynthesis UDP-N-acetylmuramoylalanyl-D-glutamate ligase
<i>priA</i>	DNA replication primosomal replication factor Y
<i>racE</i> ^a	Peptidoglycan biosynthesis glutamate racemase
<i>mc</i>	RNA modification RNase III
<i>rplC</i>	Ribosomal protein L3 (BL3)
<i>rplF</i> ^a	Ribosomal protein L6 (BL8)
<i>rplJ</i> ^{a,b}	Ribosomal protein L10 (BL5)
<i>rplU</i> ^{a,b}	Ribosomal protein L21 (BL20)
<i>rpmJ</i>	Ribosomal protein L36 (ribosomal protein B)
<i>rpoB</i>	Transcription RNA polymerase (β subunit)
<i>rpoC</i>	Transcription RNA polymerase (β' subunit)
<i>rpsB</i>	Ribosomal protein S2
<i>rpsK</i>	Ribosomal protein S11 (BS11)
<i>rpsM</i> ^{a,b}	Ribosomal protein S13
<i>secA</i>	Secretion preprotein translocase subunit (ATPase)
<i>sigA</i>	Transcription RNA polymerase major σ factor
<i>tkt</i> ^{a,b}	Glycolysis transketolase
<i>topA</i>	DNA packaging DNA topoisomerase I
<i>trxA</i> ^{a,b}	Thioredoxin
<i>tsf</i>	Translation elongation factor
<i>tyrS</i>	tRNA synthetase tyrosyl-tRNA synthetase (major)

^a Gene excluded from the concatenation analyses of the 27 genes and 28 taxa.

^b The single-gene tree did not place *P. penetrans* among the low-G+C-content, gram-positive bacteria.

for evidence of horizontal gene transfer based on G+C content, codon usage, amino acid usage, and gene position (15); potential horizontal gene transfer candidates were discarded. The remaining genes were then blasted in the National Center for Biotechnology Information database against the bacterial species chosen using the criteria described above and BLAST tools. The BLAST results were used as a guide to identify full gene sequences in the corresponding species genome. A total of 40 genes which were found in the most species and contained the longest identical sequences, ranging from 50 to over 400 amino acids long, were chosen for analysis (Tables 1 and 2).

The multiple-sequence alignments were constructed using both the global alignment program ClustalW (39) and the local alignment program DiAlign (24). The best alignment was chosen, and hand-eye adjustments were made in GeneDoc (25) until the levels of sequence identity were at least 30%. This was done by pruning any region violating position homology in the multiple align-

TABLE 2. Spreadsheet of species versus genes used in the analyses

Taxon	Genes ^a																	
	<i>accA</i>	<i>ack</i>	<i>argS</i>	<i>dapA</i>	<i>dapB</i>	<i>dnaB</i>	<i>dnaE</i>	<i>eno</i>	<i>fmt</i>	<i>glmS</i>	<i>groEL</i>	<i>gyrA</i>	<i>gyrB</i>	<i>infA</i>	<i>map</i>	<i>metS</i>	<i>maY</i>	<i>murB</i>
<i>Borrelia burgdorferi</i>	0	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
<i>Treponema pallidum</i>	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Chlamydomphila pneumoniae</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Chlamydia muridarum</i>	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1
<i>Chlamydia trachomatis</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Aquifex aeolicus</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Mycoplasma genitalium</i>	0	1	1	0	0	1	1	1	1	0	0	1	1	1	1	1	0	0
<i>Mycoplasma pneumoniae</i>	0	1	1	0	0	1	1	1	1	0	0	1	1	1	1	1	0	0
<i>Bacillus anthracis</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Bacillus cereus</i>	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1
<i>Bacillus halodurans</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Bacillus subtilis</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Pasteuria penetrans</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Staphylococcus aureus</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Lactococcus lactis</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Streptococcus pyogenes</i>	1	1	0	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1
<i>Mycobacterium tuberculosis</i>	0	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	0
<i>Mycobacterium leprae</i>	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Campylobacter jejuni</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Helicobacter pylori</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Neisseria meningitidis</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Xylella fastidiosa</i>	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
<i>Pseudomonas aeruginosa</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Buchnera aphidicola</i>	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Escherichia coli</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Pasteurella multocida</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Haemophilus influenzae</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Rickettsia prowazekii</i>	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1
Total for 28 taxa	20	28	25	19	23	27	28	27	28	24	21	28	28	28	28	28	26	25
<i>Sulfolobus solfataricus</i>	0	0	1	1	0	0	0	1	0	1	0	0	0	0	1	1	0	0
<i>Pyrococcus abyssi</i>	0	1	1	0	0	0	0	1	0	1	0	0	0	0	1	1	0	0
<i>Halobacterium</i> sp. strain NRC-1	0	1	0	0	0	0	0	1	0	1	0	1	1	0	1	1	0	0
<i>Archaeoglobus fulgidus</i>	0	1	1	1	1	0	0	1	0	0	0	1	1	1	1	1	0	0
<i>Methanococcus jannaschii</i>	0	0	1	1	1	0	0	1	0	1	0	0	0	1	1	1	0	0
Total for 33 taxa	20	31	29	22	25	27	28	32	28	28	21	30	30	30	33	33	26	25
Alignment size (amino acids)	178	181	129	63	116	142	197	267	283	123	349	219	151	36	122	415	82	166

^a 1, gene is present; 0, gene is absent or less than 30% identity. Boldface type indicates data that were used only in the analysis of 40 genes and 33 species.

ment and also excluding species with a consistently low level of sequence identity. These adjustments were the basis for delimiting the start and stop positions of the gene sequences in each alignment (Table 2). Since the species have a high degree of diversity, alignment was performed at the amino acid level.

Phylogenetic analyses were then conducted with these amino acid sequence alignments; the parameters used in each program are indicated below. The first analysis was performed with each gene separately using maximum-likelihood and maximum-parsimony methods. Then separate alignments were concatenated using SNAP Workbench (29) for a total of 6,036 amino acids and were analyzed again using the same programs. Two different subsets of this concatenation, both consisting of the same 28 taxa and 27 genes, were analyzed using Bayesian and maximum-likelihood methods, as well as maximum parsimony. These subsets were created to minimize the potential of phylogenetic artifacts, such as unequal taxon sampling bias, derived from the inclusion of species and genes with large amounts of missing data. By including only the species containing at least 27 genes, the taxon number was reduced to 28. The species that were removed included all the species in the phyla *Crenarchaeota* and *Euryarchaeota*. After taxa were removed, the multiple-sequence alignment of each gene was reexamined and kept if there were no regions in the alignment that disrupted positional homology (Table 2). The largest of these subsets included all 27 full gene sequences and a total of 4,236 amino acids. For the other subset all of the insertions/deletions (indels) were removed using SNAP Workbench, which left the most conservative data set with a total of 3,032 amino acids.

The phylogenetic analyses for all datasets were performed as described below. In all cases, a strict consensus tree was inferred to avoid placing emphasis on any one topology. Gaps were treated as missing data or ambiguous characters, and no molecular clock was assumed. Maximum-likelihood analysis using PAML 3.12

(47) inferred a single tree via stepwise addition. The discrete-gamma model and empirical amino acid substitution matrix of Jones et al. (17) were used to perform the analysis (46). All model parameters (e.g., amino acid substitution rates) were estimated empirically from the data. Maximum-parsimony searches in MEGA 2.1 (19) were performed using close-neighbor interchange, uniform weighting, and 5,000 bootstrap replicates. The close-neighbor interchange initial trees were selected by using random addition trees with 10 replications each and a search level of 3. A Bayesian analysis using MrBayes (32) was performed to allow the greatest heterogeneity among sites while starting parameter values similar to those used in the maximum-likelihood analyses were used. We used gamma-distributed rates across sites, with substitutions occurring according to a time-reversible model. A uniformly shaped prior distribution was chosen, which included all topologies equally probable, a priori, and unconstrained branch lengths. The rate matrix for the prior was fixed to the Jones-Taylor-Thornton model for consistency by maximum-likelihood analyses. One cold chain and three heated chains were used with the sample frequency set to 50. The Markov chain Monte Carlo analysis was run for 10,000 generations for each analysis due to the complexity of the protein data set.

RESULTS

All phylogenetic analyses performed with 40 housekeeping genes of 33 bacterial species consistently placed *P. penetrans* between *Bacillus halodurans* and *Staphylococcus aureus* in the

TABLE 2—Continued

Genes ^a																					Total for 40 genes	
<i>murC</i>	<i>murD</i>	<i>priA</i>	<i>racE</i>	<i>mc</i>	<i>rplC</i>	<i>rplF</i>	<i>rplJ</i>	<i>rplU</i>	<i>rpmI</i>	<i>rpoB</i>	<i>rpoC</i>	<i>rpsB</i>	<i>rpsK</i>	<i>rpsM</i>	<i>secA</i>	<i>sigA</i>	<i>tkf</i>	<i>topA</i>	<i>trxA</i>	<i>tsf</i>	<i>tyrS</i>	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	35
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	36
1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	36
1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	33
1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	39
1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	39
0	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	28
0	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1	1	27
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	37
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	37
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40
1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	35
26	25	26	21	28	28	28	24	28	28	28	28	28	28	28	28	28	20	27	26	28	28	
0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	1	13
0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	0	0	1	1	0	0	1	15
0	0	1	0	0	0	0	1	0	0	0	1	1	1	1	0	0	0	1	1	0	1	16
0	0	0	0	0	1	0	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	19
0	0	0	0	0	1	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	18
26	25	27	21	28	30	28	27	28	28	29	32	33	33	33	28	28	23	31	30	30	33	
160	97	126	163	165	108	42	44	100	37	345	68	197	118	56	239	187	192	195	48	66	64	

gram-positive, low-G+C-content bacilli with good support (Fig. 1 to 3).

The initial maximum-likelihood and maximum-parsimony analyses performed for each gene separately gave different results. For the majority of the trees, *P. penetrans* fell into the low-G+C-content, gram-positive clade, as expected. There were, however, seven instances where the placement of the bacterium in the phylogeny was seemingly random (Table 1). For example, *P. penetrans* was found to be most closely related to proteobacteria for the *groEL* gene, with a bootstrap value of 99% (Fig. 4), and to the high-G+C-content, gram-positive bacteria for the *eno* gene, with a bootstrap value of 90% (Fig. 5). There were also a few clades with strong support in one gene tree that were incongruent with clades supported by data for other gene loci. The genes with the most consistent phylogenetic placement of species, with some minor incongruencies within clades, coincide with the subset of genes chosen for the concatenation analyses. When the maximum ln likelihood of the individual gene trees (Table 3) was examined, it was apparent that trees inferred from alignments that were similar in length had log likelihood values that were similar even though the overall tree topologies and the specific placement of species differed.

The maximum-likelihood tree inferred from the concatenation of all 40 genes was concordant with 33 of the 40 individual gene trees (Fig. 2). The tree for the combined data set resolved accepted clades of species within the same phylum. Each clade was further split into smaller clades that corroborated widely accepted phylogenetic and taxonomic relationships (7, 45). The placement of *P. penetrans* in this phylogeny agrees with the initial hypothesis that it is a member of the *Bacillus* class and resolves its exact placement between *B. halodurans* and *S. aureus*.

The next series of phylogenetic analyses were performed with 27 housekeeping genes and 28 bacterial species. The genes were concatenated and analyzed using maximum-likelihood and maximum-parsimony methods (Fig. 3). The phylogenetic inferences from each of these methods were consistent with the placement of *Pasteuria* ancestral to *Bacillus* and the *Lactococcus/Streptococcus* clade. As shown in Fig. 3, maximum parsimony did not provide as much support for the phylogenetic relationships in the inferred tree as the parametric methods provided (Fig. 1 and 2).

The most conservative analysis was performed with the set of 27 genes with all indels removed. This data set was then examined using Bayesian, maximum-likelihood, and maxi-

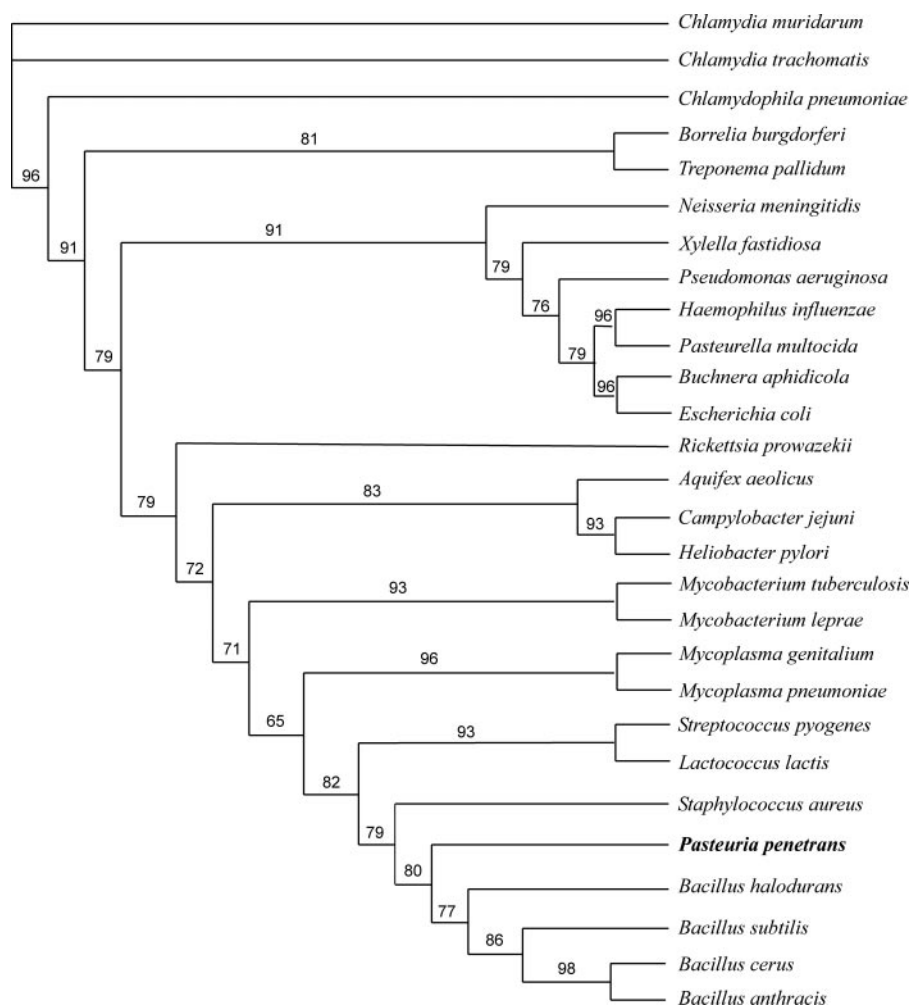


FIG. 1. Bayesian analysis of the 27 independent and concatenated genes with all indels removed for 28 bacterial species. The Bayesian posterior probabilities are indicated below the branches; the Markov chain Monte Carlo analysis was run for 10,000 generations.

mum-parsimony methods. All three phylogenies placed *P. penetrans* in the *Bacillus* clade, between *B. halodurans* and *S. aureus* (Fig. 1). The posterior probability was 80 and the bootstrap value was 85%. Importantly, the results of each of these analyses mirrored the results obtained with the full gene sequences and the initial data set for 40 genes and 33 species.

DISCUSSION

The results of the concatenated species tree analysis place *P. penetrans* in the *Bacillus* clade, between *B. halodurans* and *S. aureus* (Fig. 1 and 2). This result suggests that *Pasteuria* is ancestral to other *Bacillus* spp. Within this group, *P. penetrans* is more closely related to the saprophytic extremophile *B. halodurans* and its close saprophytic relative *B. subtilis* than to the pathogenic species *Bacillus anthracis* and *Bacillus cereus*. Unfortunately, at the time of this analysis the genome of *Bacillus thuringiensis* was not available. These results were independent of which phylogenetic method was used and whether the data set contained 40 or 27 genes with or without indels.

For most of the phylogenetic trees based on single genes, the inferred clades are also found in the concatenated tree, with

slight variations in the exact clade positions. However, for some of the genes, especially those with low numbers of species (Table 2), the position of *P. penetrans* varies (Table 1 and Fig. 4 and 5). 31 The incongruence observed here is a typical result for single-gene phylogenies (30). This is due to a number of biological and methodical factors that are related to gene evolution and phylogenetic analyses, such as the rate of evolution, variable sites, gene size, taxon bias, base composition, and the number of parsimony-informative sites. A contributing factor to the differences seen in these single-gene phylogenies, which was also why these genes were selected, is the fact that they are mostly unlinked (Table 1). Such independence between genes can lead to differences in evolutionary changes, as noted above, causing the placement of the species to differ in each tree. These results strengthen the argument that gene trees do not always accurately indicate species evolutionary relationships even when slowly evolving orthologs are used (4, 12, 27, 30). Further analyses are needed to reveal the cause of incongruencies observed in the individual trees. For now, our results indicate that single-gene trees may not be precise enough to infer accurate species evolutionary relationships for characterization of *Pasteuria* isolates.

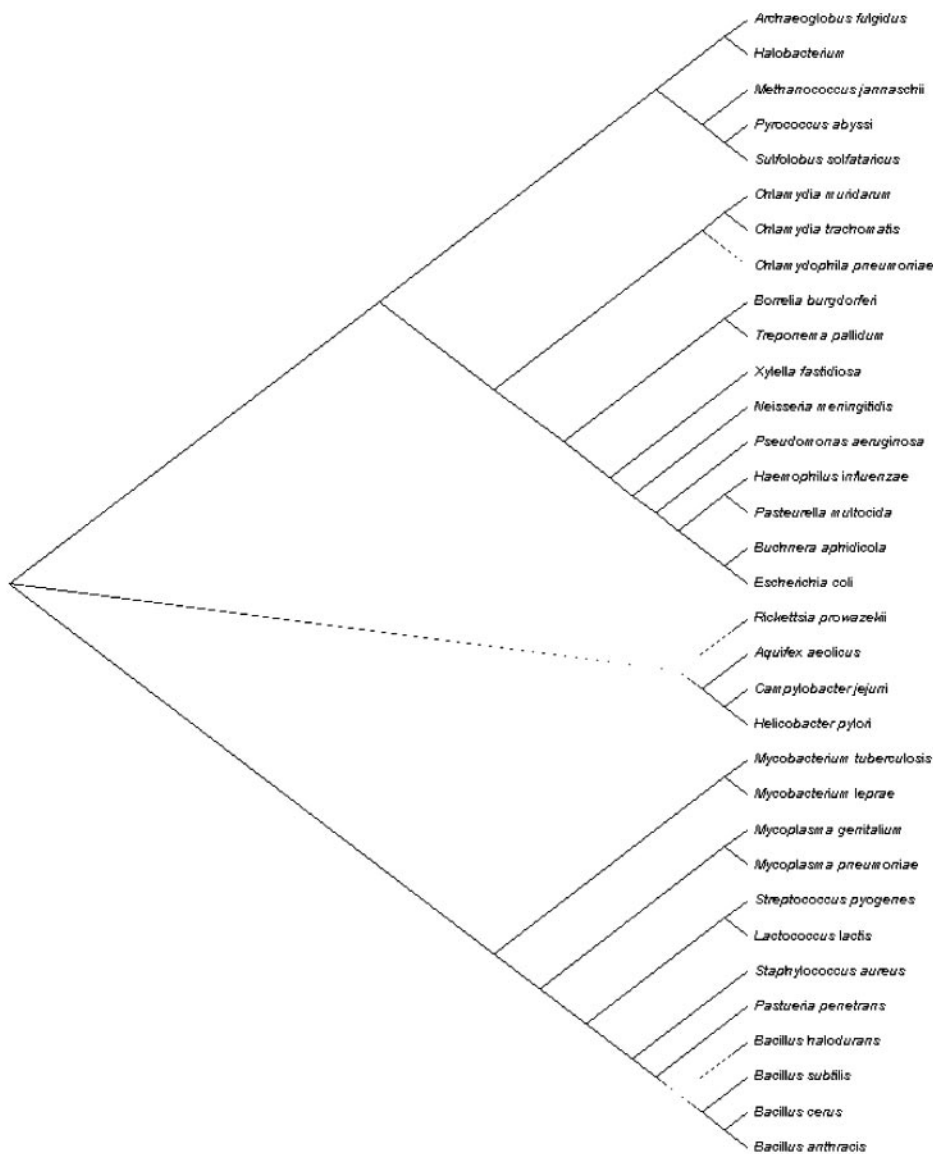


FIG. 2. Maximum-likelihood analysis of the 40 concatenated housekeeping genes for 33 bacterial species.

The earliest DNA studies to characterize *Pasteuria* all relied on using universal bacterial primers based on the 16S rRNA subunit and sequencing the PCR-amplified products. These studies all showed that *Pasteuria* is closely related to endospore-producing bacilli (2, 3). The formation of endospores is one of the most complex developmental processes in prokaryotes and involves hundreds of sporulation-specific genes (36). Genomic analysis has shown that many of the genes in the sporulation pathway of these bacteria are conserved, and several recent studies have started to use some of these genes to construct phylogenies. The phylogenetic relationship obtained using *sigE* and *sigF* showed that *Pasteuria* is ancestral to other bacilli and taxonomically in the middle of the bacilli, respectively (28). In a separate study using *spo0A*, *P. penetrans* and *P. ramosa* were both shown to be *Pasteuria* is rooted deeply in the supergenus *Bacillus*. As might be expected, phylogenies based

on single genes are likely to produce different results. The results presented here indicate that individual gene trees, although each tree was robustly supported, showed great variation in their placement of species.

We have shown that a concatenation of several genes results in a robust phylogenetic tree, regardless of which analysis method is used. By combining amino acid sequence alignments, the phylogenetic signal is increased and noise is minimized, creating a more accurate phylogeny (4). Maximum likelihood is known to take into account many of the underlying issues concerning amino acid sequence evolution, such as composition bias and rate heterogeneity. This method has been reported to be better than maximum-parsimony analysis (41) and Bayesian methods (31), but in these analyses, there were no apparent differences in the deep phylogenies inferred. Unfortunately, due to the computational difficulty with such a

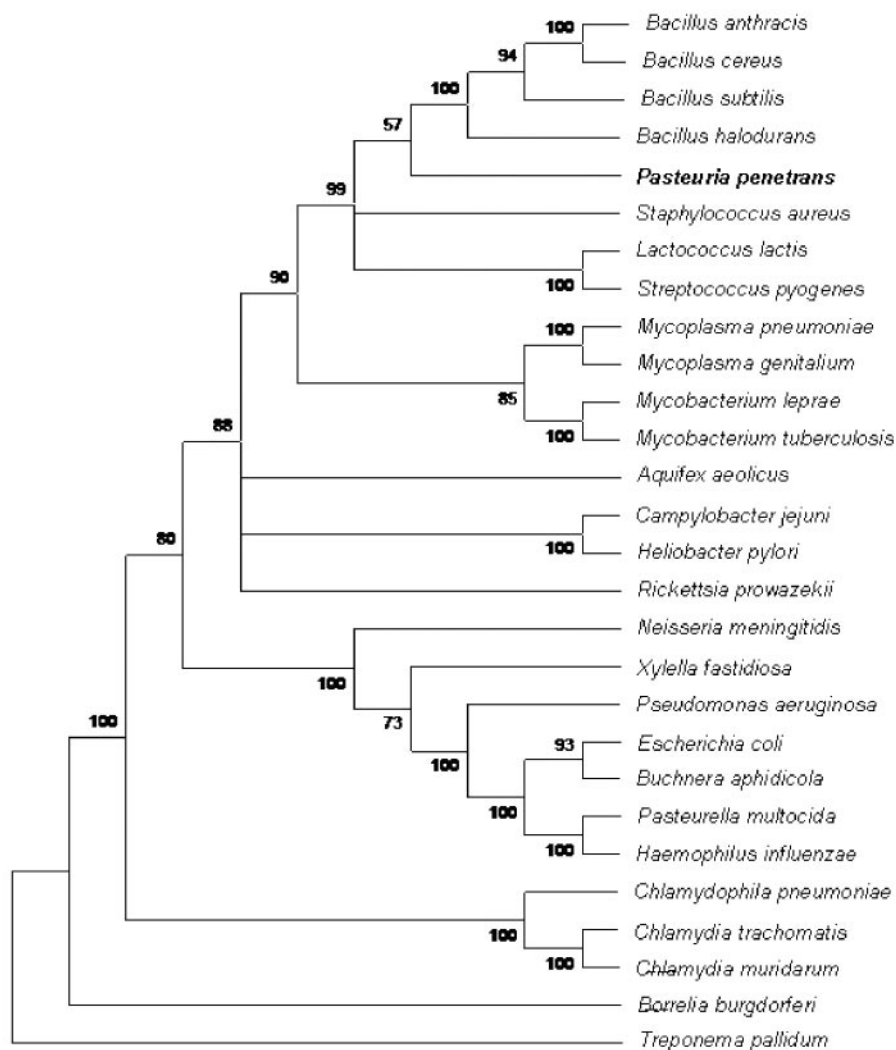


FIG. 3. Maximum-parsimony analysis for 28 taxa and a concatenation of 27 full-length genes. Only bootstrap values greater than 50% are shown.

large protein data set, Bayesian inference was employed only for the smallest, least ambiguous data set. The results of this analysis did, however, mirror those of the maximum-likelihood and maximum-parsimony analyses. Furthermore, different inference methods were used to reduce the potential of violating the assumptions of a particular method, which may result in an erroneous tree. For example, Bayesian posterior probabilities are thought to overestimate the reliability of a tree, whereas bootstrap values are believed to be underestimates (38). Our inference of the same tree from a few robust phylogenetic methods increases the credibility of the consensus tree, as demonstrated here.

Not only did the different programs result in the same tree, but the same tree was also inferred when 40 genes and 33 taxa or 27 genes, with or without indels, and 28 taxa were used. In the smaller analysis, the species that were removed contained the fewest genes used in the analysis. Incidentally, these taxa all belong in the phyla *Crenarchaeota* and *Euryarchaeota*. The taxon selection, in this case, had no influence on branch resolution, as found in the metazoan tree by Rokas and coworkers

(31). Likewise, the number of genes (40 or 27 genes) had no effect on the exact branch placement of the species in question. Therefore, in this case both the larger and smaller data sets were adequate to resolve the tree correctly, as opposed to the single-gene analysis (30). Since the number of amino acids in an alignment drastically affects the amount of time needed for the analysis, using the smaller data set greatly improves the efficiency of the phylogenetic inference method. Including areas of ambiguity (e.g., indels) also makes the analyses increasingly difficult. Since identical results were obtained with the data sets with and without indels, no extra information was gained from the indels, and therefore, they were not necessary. Consequently, there was no loss of information or resolution when indels, genes with low species counts, and species with low gene counts were removed. There was only a gain of efficiency and computability in the data set.

Our resolution of the *P. penetrans* phylogenetic position should enable us to exploit comparative genomics to study the relationship of this organism to its closest relatives in the *Bacillus* clade and to discover the genetic basis for parasitism of

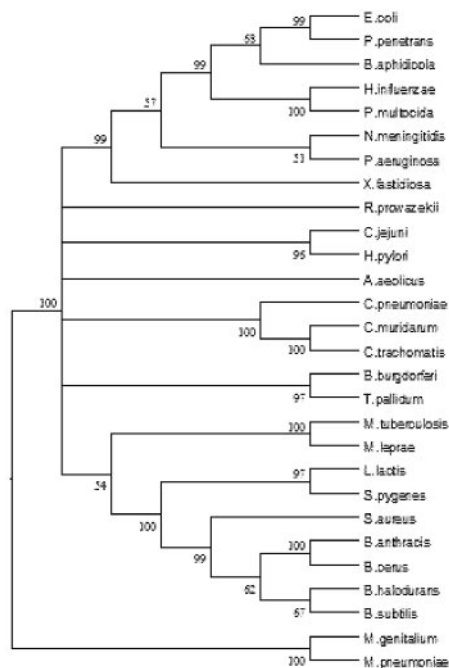


FIG. 4. Gene tree for *groEL* inferred using maximum parsimony, placing *P. penetrans* in the proteobacterial clade with 99% bootstrap support.

the root knot nematode. The analysis undertaken so far has been based on genes extracted from a single population of *Pasteuria* (RES147). It is well documented that while some populations of *P. penetrans* attach only to a particular species

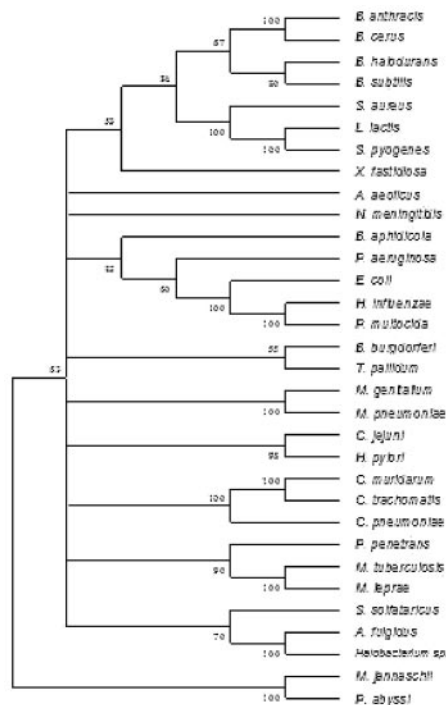


FIG. 5. Gene tree for *eno* inferred using maximum parsimony, placing *P. penetrans* in the high-G+C-content, gram-positive clade with 90% bootstrap support.

TABLE 3. Number of amino acids in ascending order for each single-gene alignment and corresponding maximum ln likelihood obtained from maximum-likelihood analysis

Gene	No. of amino acids in alignment	Maximum ln likelihood
<i>infA</i>	36	-810.2
<i>rpmJ</i>	37	-890.1
<i>rplF^a</i>	42	-954.0
<i>rplJ^{a,b}</i>	44	-1,778.7
<i>trxA^{a,b}</i>	48	-2,159.7
<i>rpsM^{a,b}</i>	56	-2,159.7
<i>dapA^a</i>	63	-2,545.9
<i>tyrS</i>	64	-2,584.8
<i>tsf</i>	66	-2,266.5
<i>rpoC</i>	68	-1,869.1
<i>mraY</i>	82	-2,573.5
<i>murD</i>	97	-3,905.0
<i>rplU</i>	100	-3,560.0
<i>rplC</i>	108	-3,843.1
<i>dapB^a</i>	116	-4,849.8
<i>rpsK</i>	118	-3,396.3
<i>map^a</i>	122	-5,479.6
<i>glmS</i>	123	-4,015.1
<i>priA</i>	126	-4,093.3
<i>argS</i>	129	-5,692.1
<i>dnaB</i>	142	-2,848.0
<i>gyrB</i>	151	-4,704.0
<i>murC^{a,b}</i>	160	-5,643.3
<i>racE^a</i>	163	-5,122.1
<i>mec</i>	165	-6,251.7
<i>murB</i>	166	-6,095.1
<i>accA^a</i>	178	-4,105.2
<i>adk</i>	181	-8,024.8
<i>sigA</i>	187	-4,839.9
<i>tkt^{a,b}</i>	192	-6,785.3
<i>topA</i>	195	-7,248.6
<i>dnaE</i>	197	-6,409.0
<i>rpsB</i>	197	-6,054.5
<i>gyrA</i>	219	-5,470.8
<i>secA</i>	239	-5,468.1
<i>eno^{a,b}</i>	267	-10,165.7
<i>fnt</i>	283	-12,315.5
<i>rpoB</i>	345	-8,998.4
<i>groEL^{a,b}</i>	349	-7,764.8
<i>metS</i>	415	-18,388.6
No indels		
27G/28S	3,032	-22,783.2
27G/28S	4,236	-117,243.2
40G/33S	6,036	-217,221.9

^a Gene not included in the 27-gene, 28-taxon concatenated gene set.
^b Gene that misplaced *Pasteuria* in the single-gene tree analyses.
^c The no indel data are the data for the concatenated data sets. 27G, 27 genes; 28S, 28 species; 40G, 40 genes; 33S, 33 species.

of RKN (8, 11, 13, 22, 37), other populations are specific for a particular population of a species of the nematode (8, 10, 37). This intraspecific biological variation and its possible role in horizontal gene transfer may be revealed through genetic comparison to other closely related pathogenetic bacteria identified in this study. It is also interesting that the previous suggestion that *P. penetrans* may be an ancestral member of the bacilli and not a recently evolved organism is strongly supported by the results presented here. Most significantly, the parasitic organism *P. penetrans* is more closely related to the saprophytic bacilli than to the animal pathogens.

ACKNOWLEDGMENTS

This work was supported by the North Carolina Agricultural Research Service and by Rothamsted Research, Ltd.

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Anderson, J. M., J. F. Preston, D. W. Dickson, T. E. Hewlett, and J. E. Maruniak. 1999. Phylogenetic analysis of *Pasteuria penetrans* by 16S rRNA gene cloning and sequencing. *J. Nematol.* **31**:319–325.
- Atibalentja, N., G. R. Noel, and L. L. Domier. 2000. Phylogenetic position of the North American isolate of *Pasteuria* that parasitizes the soybean cyst nematode, *Heterodera glycines*, as inferred from 16S rDNA sequence analysis. *Int. J. Syst. Evol. Microbiol.* **50**:605–613.
- Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**:972–976.
- Chen, Z. X., D. W. Dickson, R. McSorley, D. J. Mitchell, and T. E. Hewlett. 1996. Suppression of *Meloidogyne arenaria* race 1 by soil applications of endospores of *Pasteuria penetrans*. *J. Nematol.* **28**:159–168.
- Cobb, N. A. 1906. Fungus maladies of the sugar cane, with notes on associated insects and nematodes, 2nd ed. Hawaiian Sugar Planters Association bulletin no. 5. Hawaiian Sugar Planters Association, Honolulu, Hawaii.
- Daubin, V., M. Gouy, and G. Perrière. 2001. Bacterial molecular phylogeny using supertree approach. *Genome Informatics* **12**:155–164.
- Davies, K. G., and M. Redden. 1997. Diversity and partial characterization of putative virulence determinants in *Pasteuria penetrans*, the hyperparasite of root-knot nematodes. *J. Appl. Microbiol.* **83**:227–235.
- Davies, K. G., M. Fargette, G. Balla, A. Daudi, R. Duponnois, S. R. Gowen, T. Mateille, M. S. Phillips, A. Sawadogo, C. Trivino, E. Vouyoukalou, and D. L. Trudgill. 2001. Cuticle heterogeneity as exhibited by *Pasteuria* spore attachment is not linked to the phylogeny of parthenogenetic root-knot nematodes (*Meloidogyne* spp.). *Parasitology* **122**:111–120.
- Davies, K. G., M. Redden, and T. K. Pearson. 1994. Endospore heterogeneity in *Pasteuria penetrans* related to attachment to plant-parasitic nematodes. *Lett. Appl. Microbiol.* **19**:370–373.
- Davies, K. G., M. P. Robinson, and V. Laird. 1992. Proteins on the surface of spores of *Pasteuria penetrans* and their involvement in attachment to the cuticle of second-stage juveniles of *Meloidogyne incognita*. *J. Invertebr. Pathol.* **59**:18–23.
- Doyle, J. J. 1992. Gene trees and species tree: molecular systematics as one-character taxonomy. *Syst. Bot.* **17**:144–163.
- Duponnois, R., M. Fargette, S. Fould, J. Thioulouse, and K. G. Davies. 2000. Diversity of the bacterial hyperparasite *Pasteuria penetrans* in relation to root-knot nematodes (*Meloidogyne* spp.) control on *Acacia holosericea*. *Nematology* **2**:235–442.
- Fox, G. E., J. D. Wisotzkey, and P. Jurtshuk, 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.
- García-Vallvé, S., A. Romeu, and J. Palau. 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* **10**:1719–1725.
- Grimshaw, C. E., S. Huang, C. G. Hanstein, M. A. Strauch, D. Burbuly, L. Wang, J. A. Hoch, and J. M. Whiteley. 1998. Synergistic kinetic interactions between components of the phosphorelay controlling sporulation in *Bacillus subtilis*. *Biochemistry* **37**:1365–1375.
- Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**:275–282.
- Kobayashi, K., S. D. Ehrlich, A. Albertini, G. Amati, K. K. Andersen, M. Arnanan, K. Asai, S. Ashikaga, S. Aymerich, P. Bessieres, F. Boland, S. C. Brignell, S. Bron, K. Bunai, J. Chapuis, L. C. Christiansen, A. Danchin, M. Debarbouille, E. Dervyn, E. Deuring, K. Devine, S. K. Devine, O. Dreesen, J. Errington, S. Fillinger, S. J. Foster, Y. Fujita, A. Galizzi, R. Gardan, C. Eschevins, T. Fukushima, K. Haga, C. R. Harwood, M. Hecker, D. Hosoya, M. F. Hullo, H. Kakeshita, D. Karamata, Y. Kasahara, F. Kawamura, K. Koga, P. Koski, R. Kuwana, D. Imamura, M. Ishimaru, S. Ishikawa, I. Ishio, D. Le Coq, A. Masson, C. Mauel, R. Meima, R. P. Mellado, A. Moir, S. Moriya, E. Nagakawa, H. Nanamiya, S. Nakai, P. Nygaard, M. Ogura, T. Ohanan, M. O'Reilly, M. O'Rourke, Z. Pragai, H. M. Pooley, G. Rapoport, J. P. Rawlins, L. A. Rivas, C. Rivolta, A. Sadaie, Y. Sadaie, M. Sarvas, T. Sato, H. H. Saxild, E. Scanlan, W. Schumann, J. F. Seegers, J. Sekiguchi, A. Sekowska, S. J. Seror, M. Simon, P. Stragier, R. Studer, H. Takamatsu, T. Tanaka, M. Takeuchi, H. B. Thomaidis, V. Vagner, J. M. van Dijk, K. Watabe, A. Wipat, H. Yamamoto, M. Yamamoto, Y. Yamamoto, K. Yamane, K. Yata, K. Yoshida, H. Yoshikawa, U. Zuber, and N. Ogasawara. 2003. Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. USA* **100**:4678–4683.
- Kumar, S., K. Tamura, I. Jakobsen, and M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Arizona State University, Tempe.
- Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**:3140–3145.
- Mankau, R. 1975. *Bacillus penetrans* n. comb. causing a virulent disease of plant-parasitic nematodes. *J. Invertebr. Pathol.* **26**:333–339.
- Mendoza de Gíves, P., K. G. Davies, M. Morgan, and J. M. Behnke. 1999. Attachment tests of *Pasteuria penetrans* to the cuticle of plant and animal parasitic nematodes, free living nematodes and *srf* mutants of *Caenorhabditis elegans*. *J. Helminthol.* **73**:67–71.
- Metchnikoff, E. 1888. *Pasteuria ramosa*, un représentant des bactéries à divisions longitudinale. *Ann. Inst. Pasteur (Paris)* **2**:165–170.
- Morgenstern, B. 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**:211–218.
- Nicholas, K. B., and H. B. Nicholas, Jr. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author.
- Oostendorp, M., D. W. Dickson, and D. J. Mitchell. 1991. Population development of *Pasteuria penetrans* on *Meloidogyne arenaria*. *J. Nematol.* **23**:58–64.
- Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5**:568–583.
- Preston, J. F., D. W. Dickson, J. E. Maruniak, J. A. Brito, L. M. Schmidt, and R. M. Giblin-Davis. 2003. *Pasteuria* spp. Systematics and phylogeny of these bacterial parasites of phytopathogenic nematodes. *J. Nematol.* **35**:198–207.
- Price, E. W., and I. Carbone. 2005. SNAP: workbench management tool for evolutionary population genetic analysis. *Bioinformatics* **21**:402–404.
- Rokas, A., N. King, J. Finnerty, and S. Carroll. 2003. Conflicting phylogenetic signals at the base of the metazoan tree. *Evol. Dev.* **5**:346–359.
- Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**:798–804.
- Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574.
- Sasser, J. N. 1980. Root-knot nematodes: a global menace to crop production. *Plant Dis.* **64**:36–41.
- Sayre, R. M., and M. P. Starr. 1989. Genus *Pasteuria* Metchnikoff, 1888, p. 2601–2615. In S. T. Williams, M. E. Sharpe, and J. G. Holt (ed.), *Bergey's manual of systematic bacteriology*, vol. 4. Williams and Wilkins, Baltimore, Md.
- Scholl, E. H., J. L. Thorne, J. P. McCarter, and D. M. Bird. 2003. Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. *Genome Biol* **4**:R39.
- Stagier, P., and R. Losick. 1996. Molecular genetics of *Bacillus subtilis*. *Annu. Rev. Genet.* **30**:297–341.
- Stirling, G. R. 1984. Biological control of *Meloidogyne javanica* with *Bacillus penetrans*. *Phytopathology* **74**:55–60.
- Suzuki, Y., G. V. Glazko, and M. Nei. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. USA* **99**:16138–16143.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- Thorne, G. 1940. *Duboscqia penetrans* n. sp. (Sporozoa: Microsporidia, Nosematidae), a parasite of the nematode *Pratylenchus pratensis* (de Man) Filipjev. *Proc. Helminthol. Soc. Wash.* **7**:51–53.
- Thorne, J. 2000. Models of protein sequence evolution and their applications. *Curr. Opin. Genet. Dev.* **10**:602–605.
- Trotter, J. R., and A. H. Bishop. 2003. Phylogenetic analysis and confirmation of the endospore-forming nature of *Pasteuria penetrans* based on the *spo0A* gene. *FEMS Microbiol. Lett.* **29**:249–256.
- Trudgill, D. L., G. Bala, V. C. Blok, A. Daudi, K. G. Davies, M. Fargette, S. R. Gowen, J. D. Madulu, T. Mateille, W. Mwageni, C. Netscher, M. S. Phillips, S. Abdoussalam, G. C. Trivino, and E. Vouyoukalou. 2000. The importance of tropical root-knot nematodes (*Meloidogyne* spp.) and factors affecting the utility of *Pasteuria penetrans* as a biocontrol agent. *Nematology* **2**:823–845.
- Urwin, R., and M. C. J. Maiden. 2003. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol.* **11**:479–487.
- Wheeler, D. L., D. M. Church, A. E. Lash, D. D. Leipe, T. L. Madden, J. U. Pontius, G. D. Schuler, L. M. Schriml, T. A. Tatusova, L. Wagner, and B. A. Rapp. 2002. Database resources of the National Center for Biotechnology Information: 2002 update. *Nucleic Acids Res.* **30**:13–16.
- Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**:306–314.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biol. Sci.* **13**:555–556.
- Young, J. M. 2001. Implications of alternative classifications and horizontal gene transfer for bacterial taxonomy. *Int. J. Syst. Evol. Microbiol.* **51**:945–953.