ISSN 0026-2617, Microbiology, 2010, Vol. 79, No. 3, pp. 356–369. © Pleiades Publishing, Ltd., 2010.
Original Russian Text © T.P. Tourova, A.V. Korshunova, E.M. Mikhailova, D.Sh. Sokolova, A.B. Poltaraus, T.N. Nazina, 2010, published in Mikrobiologiya, 2010, Vol. 79, No. 3, pp. 376–389.

EXPERIMENTAL ARTICLES

Application of *gyrB* and *parE* Sequence Similarity Analyses for Differentiation of Species within the Genus *Geobacillus*

T. P. Tourova[a], A. V. Korshunova[b], E. M. Mikhailova[a], D. Sh. Sokolova[a], A. B. Poltaraus[c], and T. N. Nazina[a, 1]

[a] *Winogradsky Institute of Microbiology, Russian Academy of Sciences, pr. 60-letiya Oktyabrya 7, k. 2, Moscow, 117312 Russia*

[b] *Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, 119991 Russia*

[c] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, ul. Vavilova 32, Moscow, 119991, Russia*

Received August 12, 2009

Abstract—The primary structures of the genes encoding the β-subunits of a type II topoisomerase (gyrase, *gyrB*) and a type IV topoisomerase (*parE*) were determined for 15 strains of thermophilic bacteria of the genus *Geobacillus*. The obtained sequences were used for analysis of the phylogenetic similarity between members of this genus. Comparison of the phylogenetic trees of geobacilli constructed on the basis of the 16S rRNA, *gyrB*, and *parE* gene sequences demonstrated that the level of genetic distance between the sequences of the genes encoding the β-subunits of type II topoisomerases significantly exceeded the values obtained by comparative analysis of the 16S rRNA gene sequences of *Geobacillus* strains. It was shown that, unlike the 16S rRNA gene analysis, comparative analysis of the *gyrB* and *parE* gene sequences provided a more precise determination of the phylogenetic position of bacteria at the species level. The data obtained suggest the possibility of using the genes encoding the β-subunits of type II topoisomerases as phylogenetic markers for determination of the species structure of geobacilli.

Key words: Geobacillus, taxonomy, phylogeny, topoisomerase, gyrase, *gyrB, parE.*

DOI: 10.1134/S0026261710030124

In microbiology, the modern taxonomic process cannot be imagined without analysis of 16S rRNA gene sequences and DNA–DNA hybridization. Both approaches have caused revolutionary changes in bacterial taxonomy in the past [1, 2] and are still widely used. The sequence analysis of 16S rRNA genes is a generally accepted and easily applied technique for determination of the phylogenetic position of bacteria, which allows to obtain data not only on the microorganism under study, but also on the species composition of entire microbial communities without isolation of pure cultures. Data obtained by DNA–DNA hybridization remain one of the main genotypic criteria for species determination [1, 3].

There are, however, some limitations to the use of both approaches. The results of DNA–DNA hybridization as an indirect, approximate, method depend significantly on the procedure and reaction conditions, which may result in a significant experimental error. As a consequence, a discrepancy between the obtained results arises, which makes the boundary of the "gold standard" for bacterial species (70% DNA–DNA relatedness) a relative value which gives but a rough idea of the species affiliation of the studied strains.

The direct method of comparison of the nucleotide sequences of 16S rRNA genes is more advantageous in this regard; however, slow changes in rRNA sequences in the course of evolution that make it possible to carry out comparative analysis of microorganisms at all taxonomic levels, were found to be a significant shortcoming at the species level. Due to the evolutionary conservatism of this gene, different bacterial strains belonging to one species, or, in some cases, closely related species may have identical sequences, which makes this technique inapplicable as the sole and universal test for taxonomic studies [4]. Moreover, the presence of more than one ribosomal operon in the genomes of many bacteria, as well as the resultant potential variability of the compared genes, can significantly affect the results of their comparative analysis at the species level [5].

These circumstances call for a search for alternative approaches, which is directly recommended by the Taxonomy Committee [3]. The study of phylogenetic similarity between bacterial strains based on the results of comparative analysis of sequences of the genes encoding various metabolic functions has been proposed as an alternative approach. Comparative analysis of the nucleotide sequences of the genes belonging to the group of "housekeeping genes," i.e., genes that

[1] Corresponding author; e-mail: nazina@nmi.host.ru

determine the main metabolic processes, is the most promising method [6].

As phylogenetic markers, these genes have the following advantages similar to those of 16S rRNA genes: universal distribution, evolutionary conservatism, and, mostly, vertical inheritance. At the same time, comparative phylogenetic analysis of "housekeeping" genes may be more efficient at the lowest taxonomic levels for the following three reasons: (1) the level of conservatism varies for different genes; therefore, interspecific differences between the nucleotide sequences of some genes may be more pronounced than those between 16S rRNA gene sequences; (2) unlike 16S rRNA genes, these genes code for proteins and, thus, analysis of their sequences allows us to determine the frequency of synonymous substitutions for more precise determination of the taxonomic position; and (3) the majority of these genes are represented by single-copy genes.

At present, taxonomic studies involving the *recN* [7, 8], *dnaB*, *recA*, *rpoD*, *gyrB* [9, 10], and *rpoB* [11] genes, as well as some others, have been carried out for various groups of bacteria. However, there is no shared understanding of which genes are preferable for taxonomic studies with the highest resolution at low taxonomic levels.

The DNA genes encoding type II topoisomerases which control DNA topology are among the most promising alternative taxonomic (phylogenetic) markers. Bacterial genomes contain two homologous enzymes of this type, gyrase and topoisomerase IV, consisting of two subunits and containing 1200–1500 amino acids, which is statistically significant for phylogenetic analysis [12]. The gene encoding the gyrase β subunit (*gyrB*) is most widely used as a phylogenetic marker. Studies of some groups of bacteria revealed that the base pair substitution rate in the *gyrB* gene correlates with the results of DNA–DNA hybridization [13, 14] and that the horizontal transmission of the *gyrB* gene has the same rate as that of ribosomal genes [15]. Using similarity analysis of the *gyrB* gene sequences, strains belonging to several problematic genera, including *Pseudomonas* [9, 16], *Bacillus* [17], *Acinetobacter* [18], and *Gordonia* [19], were successfully separated.

The taxonomic description of species of the genus *Geobacillus* requires additional data, since the levels of 16S rRNA similarity between the members of this group are 99–100% [20, 21]. To date, analysis of the *recN* [7] and *rpoB* [11] gene sequences has been carried out to elucidate the phylogeny of this genus; however, the phylogenetic similarity between the *gyrB* gene sequences of *Geobacillus* species has not been analyzed yet.

The goal of the present work was to determine the sequences of *gyrB* (gyrase) and *parE* (topoisomerase IV) gene fragments of the representatives of the genus *Geobacillus* and carry out phylogenetic analysis of them in order to elucidate the level of correlation with the results of DNA–DNA hybridization and 16S rRNA genes analysis, as well as with the results of analysis of the previously studied housekeeping genes at the intra- and interspecific levels of relatedness.

MATERIALS AND METHODS

Bacterial strains and cultivation conditions. The *gyrB* and *parE* gene sequences of 17 *Geobacillus* strains were studied, including *Geobacillus stearothermophilus* DSM 22^T, *G. thermocatenulatus* $B1259^T$, *G. jurassicus* $DS1^T$ and DS2, *G. subterraneus* 34^T, K, and Sam, *G. uzenensis* U^T and X, and *G. gargensis* Ga^T, as well as *Geobacillus* sp. 3Feng, B-1024, vw3-1, 8m3, 1017, 46, and 49. Bacteria were grown at 55°C on a nutrient-rich medium (pH 7.0) containing yeast extract (2.5 g per 1 l of distilled water) and tryptone (5 g/l) [20].

DNA extraction and amplification of the studied genes. DNA was extracted from the biomass produced in midexponential phase cultures using the Diatom^{tm}DNAprep kit (Biokom, Russia) according to the manufacturer's recommendations with minor modifications. The purified DNA preparation was dissolved in 100 μl of bidistilled water and used as a template for the polymerase chain reaction.

PCR amplification of the 16S rRNA genes was carried out using the standard technique with universal primers [22].

Two-round PCR amplification of the *gyrB* gene was carried out, including (1) amplification with the degenerate primers Up-1-deA and Up-2R [16] and (2) reamplification of the PCR products with the primers Up1S–Up-2Sr, complementary to the 5' ends of the degenerate primers Up-1-deA and Up-2R [16], or with the direct gyrBs-G(gyr1-G) primer specially designed for geobacilli together with the Up-2Sr primer. For selective amplification of the *parE* gene, the primers top2F-geo–top1R-geo (top3R-geo) were designed. The reaction mixture (20 μl) contained approximately 10–50 ng of the DNA template. The primer sequences and PCR regimes are listed in Table 1.

Cloning and sequencing. The PCR products obtained as a result of amplification and reamplification were sequenced directly using the relevant primers or cloned into the pTZ57R/T vector using a PCR Cloning Kit (Fermentas, Lithuania) in order to obtain the *gyrB* gene sequences of the strains for which amplicon heterogeneity was revealed by direct sequencing. The selection of transformants, detection of an insert in the vector, and sequencing of the insert were carried out using the standard technique.

Phylogenetic analysis. Preliminary analysis of the obtained nucleotide sequences was performed using the NCBI BLAST software package (http://www.ncbi.nlm.nih.gov/blast). The nucleotide sequences were aligned with the sequences retrieved from the GenBank database using the CLUSTALX 2.0 software package (http://bisp.u-strasbg.fr/fr/Documentation/ClustalX/). The phylogenetic trees were constructed

Table 1. Primer sequences and PCR protocols used for amplification and sequencing of the gyrase β subunit (*gyrB*) and type IV topoisomerase (*parE*)

Target template	Reaction	Primer	Primer sequence (5'–3')	Reaction protocol (stage, temperature, and duration)		References
gyrB	PCR amplification	Up-1 (Up-1-deA)	GAAGTCATCATGACCGTTCTGCAY GCNGGNGGNAARTTYG(A)	(1) 95°C, 3 min; (2) 94°C, 1 min; 60°C, 1 min; 72°C, 1.5 min; 35 cycles; (3) 72°C, 2 min		[15] with modifications
		Up-2R	AGCAGGGTACGGATGTGCGAGCC RTCNACRTCNGCRTCNGTCAT			
	Reamplification and sequencing	Up-lS	GAAGTCATCATGACCGTTCTGCA	(1) 95°C, 3 min; (2) 94°C, 1 min; 50°C, 1 min; 72°C, 1.5 min; 25 cycles; (3) 72°C, 2 min		
		Up-2Sr	AGCAGGGTACGGATGTGCGAGCC			
	PCR, reamplification, and sequencing	gyrBs-G	ACGGACGAACGGGTTGA	(1) 95°C, 3 min; (2) 94°C, 1 min; 50°C, 1min; 72°C, 1 min; 25 cycles; (3) 72°C, 2 min		Present work
		gyrl-G	GACGGATGAACGGGTCGA			
		gyr2-G	TTGCTTGTGTAGCCGTCGTT			
parE	PCR, sequencing	toplR-geo	GCRTCKGTCATRATRATCACTTTGTC	PCR: (1) 95°C, 3 min; (2) 94°C, 1 min; 55°C, 1 min; 72°C, 1 min; 35 cycles; (3) 72°C, 2 min	Sequencing: (1) 95°C, 5 min; (2) 94°C, 0.5 min; 55°C, 15 s; 60°C, 4 min; 25 cycles	
		top3R-geo	TTYTKCGGATTGCGCSMYTG			
		top2F-geo	GGNAARTTYGGNCAAGGC			

with the neighbor-joining method using the TREECONW software package (http://bioc-www.uia.ac.be/u/yvdp/treeconw.html). The significance of the branching order (%) was determined by bootstrap analysis of 1000 alternative trees.

The 16S rRNA, *gyrB*, and *parE* gene sequences obtained in this work were deposited in GenBank under accession numbers: *gyrB*—GU459227–GU459239; GU323951, GU323952; *parE*—GU459240–GU459250; GU323953, GU323954, GU994005, GU994006; 16S rRNA—GU459251, GU459252, GU994000–GU994004.

RESULTS

Amplification of the *gyrB* and *parE* genes. Initially, the previously proposed degenerate primers Up-1–Up-2R [16] allowing amplification and direct sequencing of the PCR products obtained from a wide range of bacterial objects were used for amplification of the *gyrB* gene fragments. However, these primers were found to be inefficient for *Geobacillus* species. Comparison of the primer Up-1 with the *gyrB* gene sequences obtained from the total genomes of *Geobacillus kaustophilus* HTA 426 and *G. thermodenitrificans* NG80-2 allowed us to modify the primer by removal of the A nucleotide from the 3' end, since the presence of the G nucleotide at the same site is typical of *Geobacillus* species. As a result of the use of the modified primer Up-1-deA, a PCR product of the expected size, about 1100 nucleotides, was obtained for ten *Geobacillus* strains, including *G. stearothermophilus* DSM 22^T, *G. thermocatenulatus* B1259^T, *G. jurassicus* DS1^T, and *G. subterraneus* 34^T and Sam, as well as *Geobacillus* sp. 3Feng, 46, B-1024, 1017, and 8m3. However, direct sequencing of the obtained amplicons was successful only for *G. stearothermophilus* DSM 22^T and *G. jurassicus* DS1^T, as well as for the *Geobacillus* strains 1017, 8m3, B-1024, and vw3-1; the obtained sequence conformed to the *gyrB* gene only in the case of the first four strains, whereas, in the case of strains B-1024 and vw3-1, it conformed to the *parE* gene. As a result of cloning of the remaining amplicons for four *Geobacillus* strains, both the *gyrB* and *parE* gene sequences were obtained simultaneously. Thus, the degenerate primers used in this work were not always specific enough to amplify the *gyrB* gene alone, which had been previously demonstrated [15, 23].

Therefore, at the next stage of analysis, primers that selectively amplified the *gyrB* and *parE* genes were designed and applied using the *de novo* determined *gyrB* and *parE* sequences and those available from the GenBank database (Table 1). PCR products of the expected size (about 850 bp) were obtained for six *Geobacillus* strains, *G. jurassicus* DS2, *G. gargensis* Ga^T, *G. subterraneus* K, and *G. uzenensis* U^T and X, as well as for *Geobacillus* sp. 49 using the primer pairs gyrBs-G(gyr1-G)–Up-2r(Up2Sr). Direct sequencing confirmed that the obtained PCR products were fragments of the *gyrB* gene. Similarly, fragments of the *parE* gene of the expected size (about 1150 (900) bp) were obtained for *G. stearothermophilus* DSM 22^T, *G. jurassicus* DS1^T, *Geobacillus* sp. 8m3, *G. gargensis* Ga^T, *G. subterraneus* K, and *G. uzenensis* U^T and X, as well as for *Geobacillus* sp. 46, 49, and 1017 with the application of the primer pairs top2F-geo–top1R-geo (top3R-geo). However, none of the primer pairs used in this work allowed us to obtain a PCR product from the target *parE* gene of *Geobacillus* sp. 8m3.

Hence, the available systems of oligonucleotide primers for amplification of the *gyrB* and *parE* gene fragments cannot yet be considered universal for all representatives of the genus *Geobacillus*. This should be remembered when using these primers for molecular ecological studies of the diversity of *Geobacillus* species in natural ecosystems.

Phylogenetic analysis of the 16S rRNA gene sequences. In this work, we compared the 16S rRNA gene sequences of newly isolated *Geobacillus* strains and of the strains belonging to the species we previously described [20, 21, 24], as well as of the type strains of other validated *Geobacillus* species [25–29]. Moreover, we used the 16S rRNA gene sequences from the total genomes of five reference *Geobacillus* strains available in the GenBank database, including two strains belonging to *G. kaustophilus* and *G. thermodenitrificans* and three unidentified strains.

According to phylogenetic analysis of the 16S rRNA gene sequences, *Geobacillus* strains formed five main clusters (Fig. 1) with a 97.6–100% similarity within each cluster and a 93.5–97.3% similarity between them. The majority of *Geobacillus* species belonged to cluster I, and the level of interspecific similarity between the sequences of this cluster was extremely high (97.8–99.9%). In many cases, this level was comparable to the intraspecific one for *Geobacillus* strains (99.8–100%), which did not permit reliable species differentiation.

This hindered identification of some novel and reference *Geobacillus* strains. Strain 3Feng, which is closely related to the type strain *G. thermoglucosidasius* (100% 16S rRNA gene sequence similarity) and the reference strain WCH70, as well as strains B-1024 and vw3-1, which are closely related to the type strain *G. toebii* (99.0–99.9%), could be preliminarily identified as representatives of these species. Strain 8m3 could be preliminarily affiliated with the species *G. pallidus*, with the type strain of which it forms cluster IV (99.8% 16S rRNA gene sequence similarity). All other strains belonged to cluster I. The reference strain G11MC16 was affiliated with the species *G. thermodenitrificans* (99.6% 16S rRNA gene sequence similarity). Strain 46 was affiliated with the species *G. stearothermophilus* (99.9% 16S rRNA gene sequence similarity), which confirmed our earlier conclusion [21]. At the same time, the phylogenetic positions of strains 49, and 1017, and the reference strain Y412MC61 were uncertain, since the level of

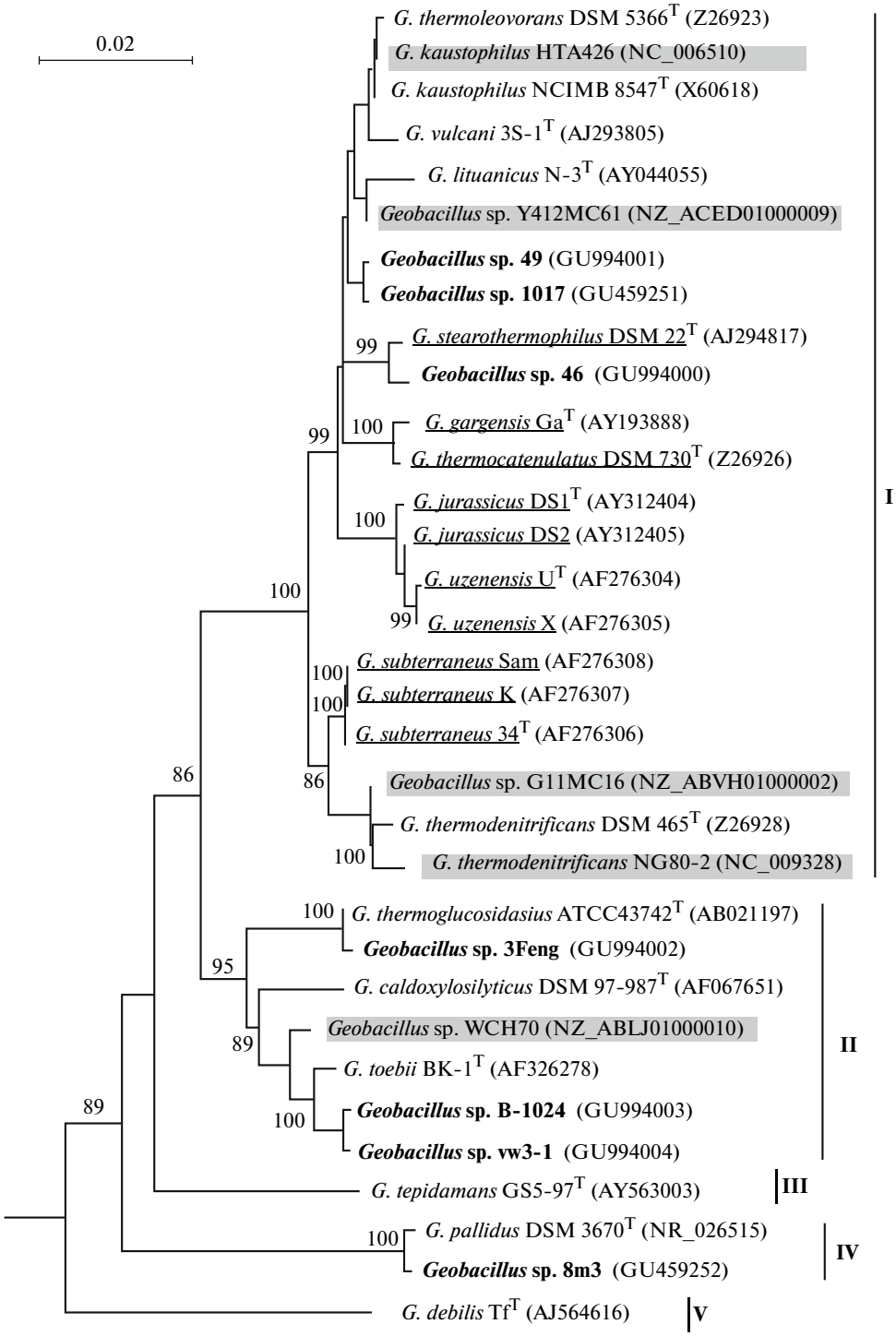

Fig. 1. Phylogenetic tree of the genus *Geobacillus* constructed on the basis of the 16S rRNA gene sequences. The tree was constructed using the neighbor-joining method, with the *B. subtilis* sequence as an outgroup. The strains the 16S rRNA sequences of which were determined in this work are shown in bold. The strains the *gyrB* and *parE* sequences of which were determined in this work are underlined. The reference strains for which data on the sequence structure of the total genome are available are shaded gray. The numerals show the significance of the branching order as determined by bootstrap analysis (only bootstrap values above 85% were considered as significant). The bar shows the evolutionary distance corresponding to two substitutions per 100 nucleotides.

16S rRNA similarity between these species was high and approximately equal to the similarity level between them and the type strains of several closely related *Geobacillus* species within this cluster, which was similarly high.

Phylogenetic analysis of the *gyrB* and *parE* gene sequences. The compared *gyrB* and *parE* gene sequences included the sequences obtained in this study and the reference ones from five completely sequenced *Geobacillus* genomes. In the course of the initial phylogenetic analysis, the *gyrB* gene sequences were compared with the analogous sequences of other genera of the family *Bacillaceae* (*Bacillus, Paenibacillus, Alicylobacillus, Amphibacillus, Anoxybacillus, Halobacillus, Virgibacillus*, and *Sulfobacillus*) available in the GenBank database. Similarly to the results of the 16S rRNA gene analysis, *Geobacillus* strains formed a separate phylogenetic group in the constructed phylogenetic tree (data not presented), which confirmed the taxonomic status of these strains conforming to the genus level.

The topologies of the phylogenetic trees constructed on the basis of phylogenetic analysis of the *gyrB* and *parE* gene sequences of *Geobacillus* strains under study corresponded completely to the topology of the "ribosomal" tree. In the *gyrB* tree, clusters I, II, and IV were observed (Fig. 2a), the level of similarity within these clusters was 74.4–100%, and the similarity level between the clusters was 71.6–80.9%. In the *parE* tree, clusters I and II were observed (Fig. 2b); the level of similarity within these clusters was 85.5–100%, while the level of similarity between the clusters was 75.2–79.8%. Due to greater phylogenetic divergence of the sequences within the main clusters as compared to the results of the 16S rRNA gene analysis, the relationships between strains and species were more pronounced, which was most noticeable in the case of the most taxonomically problematic cluster I.

In the *gyrB* tree, this cluster was reliably divided (100% bootstrap support) into five subclusters (IA–IE); the level of similarity within these subclusters (94.5–100%) was considerably higher than the level of similarity between the subclusters (84.4–90.6%). However, these subclusters did not show exact correspondence with the species composition of this cluster. For instance, subclass IB included the *gyrB* gene sequences of the *G. stearothermophilus* type strain and those of the closely related strain 46, subclass ID included the nucleotide sequences of the reference strain G11MC16 and *G. thermodenitrificans* NG80-2, and subclass IE included the sequences of three *G. subterraneus* strains. At the same time, subcluster IA included the *gyrB* gene sequences of the *G. gargensis* and *G. thermocatenulatus* type strains, as well as of the reference strain Y412MC61, *G. kaustophilus* HTA426, and strains 49 and 1017; subcluster IC included the *gyrB* gene sequences of the *G. uzenensis* and *G. jurassicus* strains. In the *parE* tree, cluster I had the same structure as in the *gyrB* tree (the level of similarity between sequences within the subclusters was 95.6–100%; the level of similarity between the subclusters was 86.3–93.3%), with the only exception. In this tree, subcluster IA was divided into two independent clusters, IA-1 containing the *parE* gene sequences of the *G. gargensis* and *G. thermocatenulatus* type strains and IA-2 containing the *parE* gene sequences of the reference strains Y412MC61 and *G. kaustophilus* HTA426, as well as of the novel strain 1017.

The topologies of the *gyrB* and *parE* trees confirmed the affiliation of strain 46 to the species *G. stearothermophilus* and of the reference strain G11MC16, to *G. thermodenitrificans*, and affiliation of strains 1017 and 49, as well of the reference strain Y412MC61, to *G. kaustophilus*.

DISCUSSION

Since their description [20], bacteria of the genus *Geobacillus* have aroused considerable scientific interest due to both the role that they play in natural ecosystems and their potential importance to biotechnology and bioremediation. Over the past several years, several new strains belonging to this genus have been described, including *G. jurassicus* [21], *G. gargensis* [24], *G. lituanicus* [25], *G. vulcani* [26], *G. tepidamans* [27], *G. debilis, G. pallidus* [28], and *G. toebii* [29]. The extent of phylogenetic divergence deduced from the 16S rRNA gene sequences of various species belonging to the genus *Geobacillus* does not exceed 7% and reaches 0% for some closely related species. In the phylogenetic tree, the 16S rRNA gene sequences of some species form separate, distinct branches (subclusters); however, the majority of 16S rRNA gene sequences form a single cluster of closely related sequences (cluster I), which makes identification of new strains very difficult. In such cases, the results of DNA–DNA hybridization are used, which, however, do not give an unambiguous answer. For example, the obtained data on the interspecific level of DNA–DNA hybridization between *G. thermocatenulatus, G. kaustophilus*, and *G. thermoleovorans* (Table 2) [20, 21, 24] contradicted the data presented in [30]. According to these authors, the level of DNA–DNA hybridization between these species was higher (more than 75%); therefore, it was proposed that these strains should be clustered into one genomovar, the species *G. thermoleovorans*. In similar experiments carried out by other authors [26], the levels of DNA–DNA hybridization between these species were close to those obtained in the present study (40–61%). A discrepancy in the results of DNA–DNA hybridization obtained by different authors may be attributed to an experimental error resulting from the differences in reaction conditions and the absence of adequate control values at the inter- and intraspecific levels. In addition, the results of DNA–DNA hybridization of newly isolated *Geobacillus* strains obtained by different authors indicate the interspecific level of their relatedness (Fig. 3), which

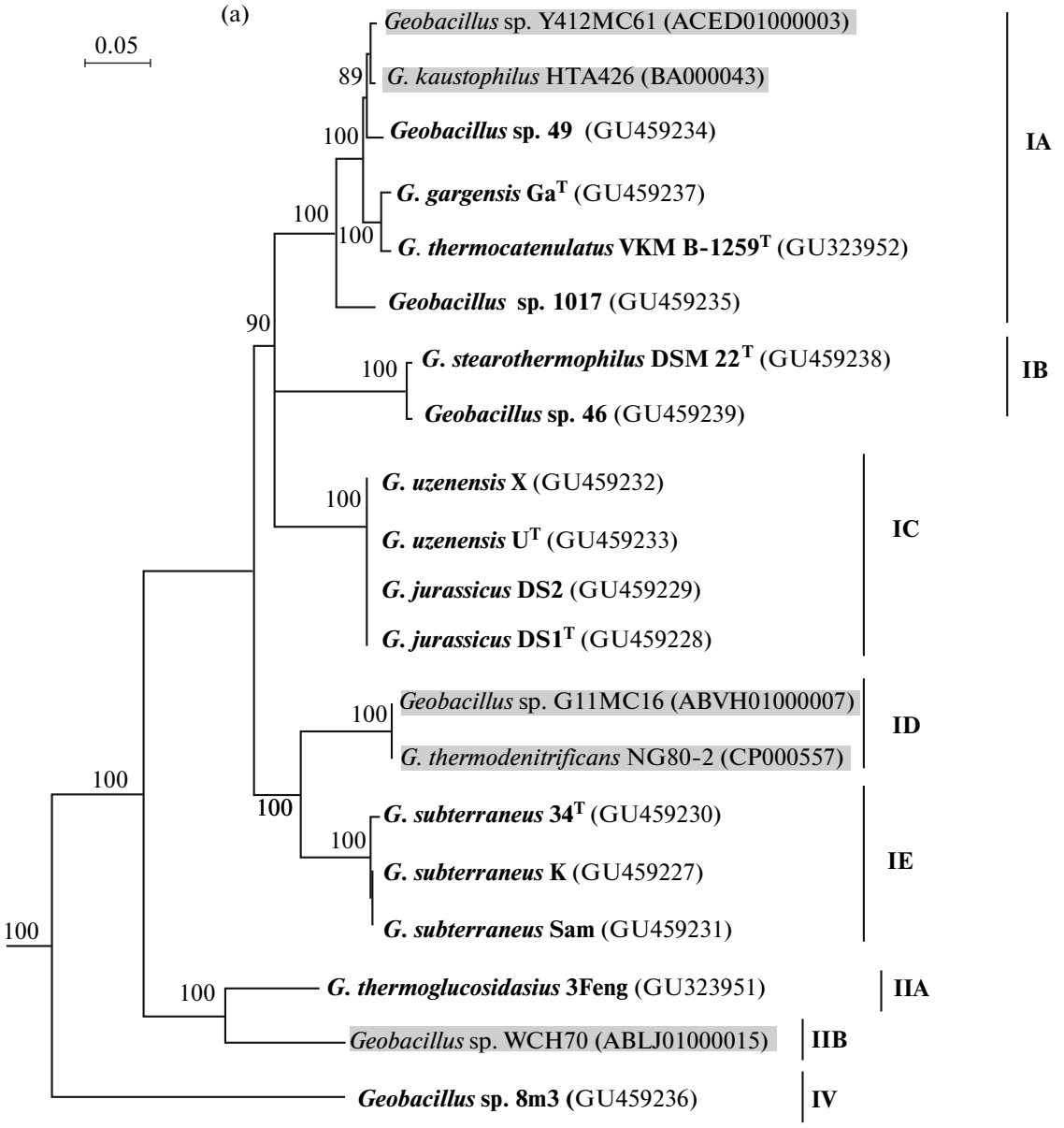

Fig. 2. Phylogenetic tree of the genus *Geobacillus* constructed on the basis of analysis of the "housekeeping" genes *gyrB* (a), *parE* (b), *recN* (c), and *rpoB* (d). The tree was constructed using the neighbor-joining method, with the *B. subtilis* sequence as an outgroup. In the trees (a) and (b), the strains the *gyrB* and *parE* sequences of which were determined in this work are shown in bold. In the trees (c) and (d), these strains are underlined. The reference strains for which data on the sequence structure of the total genome are available are shaded gray. The numerals show the significance of the branching order as determined by bootstrap analysis (only bootstrap values above 85% were considered as significant). The bar shows the evolutionary distance corresponding to five substitutions per 100 nucleotides.

helped to describe new species belonging to the phylogenetic cluster I, including *G. uzenensis, G. subterraneus* [20], *G. jurassicus* [21], *G. gargensis* [24], *G. lituanicus* [25], and *G. vulcani* [26]. It should be noted that clusters I and II can be distinctly differentiated in the dendrogram based on the results of DNA–DNA hybridization, as in the case when the results of comparative analysis of the nucleotide sequences of both ribosomal and protein-encoding genes were used (Table 2).

The results of comparative analysis of the sequences of the genes encoding type II topoisomerase, gyrase, and type IV topoisomerase demonstrate that these genes can be used along with 16S rRNA genes as phylogenetic markers for differentiation of the species of the genus *Geobacillus*. For the

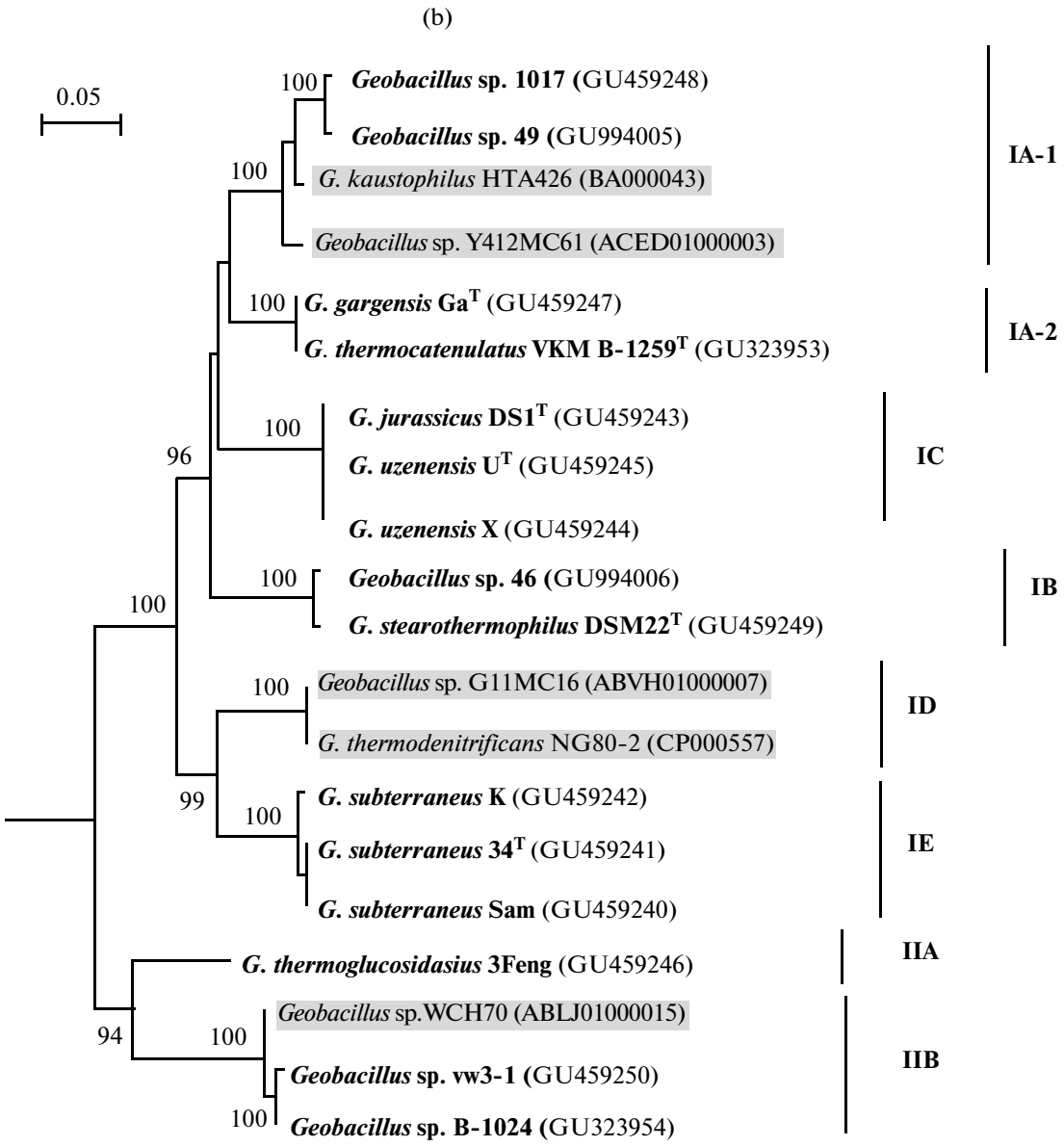

Fig. 2. Contd.

majority of *Geobacillus* species, the extent of phylogenetic divergence of the *gyrB* and *parE* gene sequences was significantly higher than that determined on the basis of the results of comparative analysis of 16S rRNA gene sequences and reached 30 and 25% (never lower than 10%) for the *gyrB* and *parE* genes, respectively. Although the topologies of the constructed phylogenetic trees correlated with the topology of the 16S rRNA tree, they were much more complex, making it possible to distinguish separate branches (subclusters) correlating well with the species structure of the genus *Geobacillus*. The level of genetic distance within the subclusters represented mainly by strains belonging to one species varied from 0 to 5% (Table 2). Hence, unlike the results of analysis of the 16S rRNA gene sequences, comparative analysis of type II topoisomerase sequences allows more precise determination of the inter- and intraspecific levels of relatedness between *Geobacillus* species (Table 2).

Similar patterns have been previously observed during the phylogenetic analysis of other "housekeeping" genes of geobacilli, *recN* [7] and *rpoB* [11]. The topologies of the phylogenetic trees of *Geobacillus* species based on the comparative analysis of these genes (Figs. 2c and 2d) correlated well with the phylogenetic trees based on the comparative analysis of the genes encoding type II topoisomerases. However, comparative analysis of the latter genes allows more precise

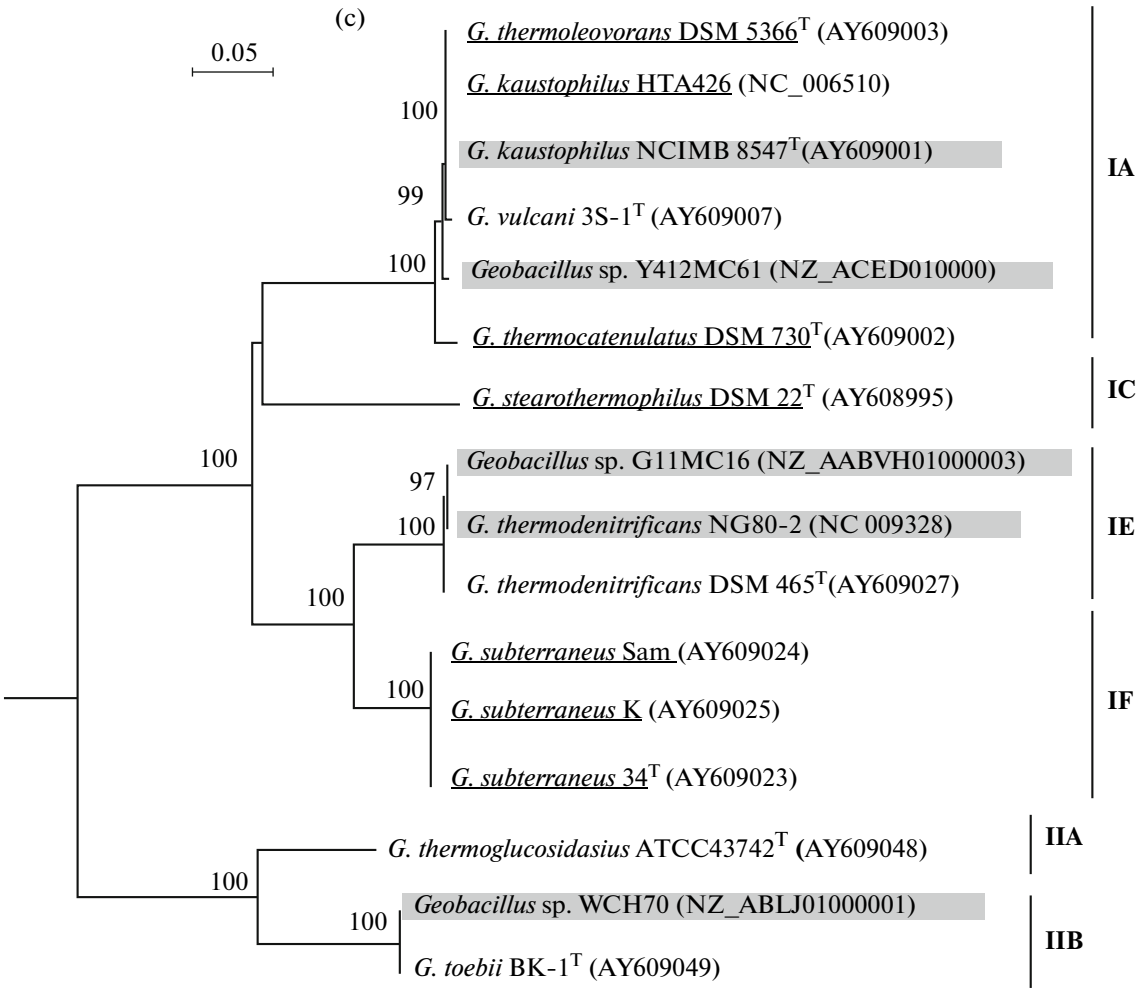

Fig. 2. Contd.

determination of the species composition of the genus *Geobacillus*. For instance, analysis of the *rpoB* gene does not allow differentiation between *G. stearothermophilus, G. kaustophilus, G. thermoleovorans, G. thermocatenulatus, G. lituanicus*, and *G. gargensis* (subclass IAB). Analysis of the *recN* and *gyrB* genes allowed differentiation of only one species, *G. stearothermophilus* (subclass IB), within this group. At the same time, analysis of the *parE* gene allowed us to distinguish reliably the species *G. thermocatenulatus* (subclass IA-1) and *G. kaustophilus* (subclass IA-2), which supported the results of DNA–DNA hybridization that indicate the independent taxonomic status of these species (Table 2). Thus, among the presently studied genes, the *parE* genes can give the highest resolution when determining the taxonomic status of *Geobacillus* species.

However, analysis of this and other genes did not permit differentiation between the species pairs *G. thermocatenulatus–G. gargensis* (subclass IA-1) and *G. jurassicus–G. uzenensis* (subclass IC) exhibiting 99.5–100 and 99.7–100% pairwise similarity of their sequences for the *gyrB* and *parE* genes, respectively (which corresponds to the intraspecific level) (Table 2). Since the results of DNA–DNA hybridization were particularly important for the description of new species in these pairs, their taxonomic status has raised doubts from the perspective of the newly obtained data. The taxonomic status of the species *G. lituanicus* [25] and *G. vulcani* [24, 26], which have been described based mainly on the results of DNA–DNA hybridization, also remains unconfirmed. It is possible that further studies of the *gyrB* and *parE* genes of various members of the subclaster IA containing closely related species are required to come to a more accurate conclusion.

This study provided for a more precise determination of the taxonomic status of unidentified reference *Geobacillus* strains, as well as of novel strains isolated from oil strata. Based on the intraspecific level of sim-

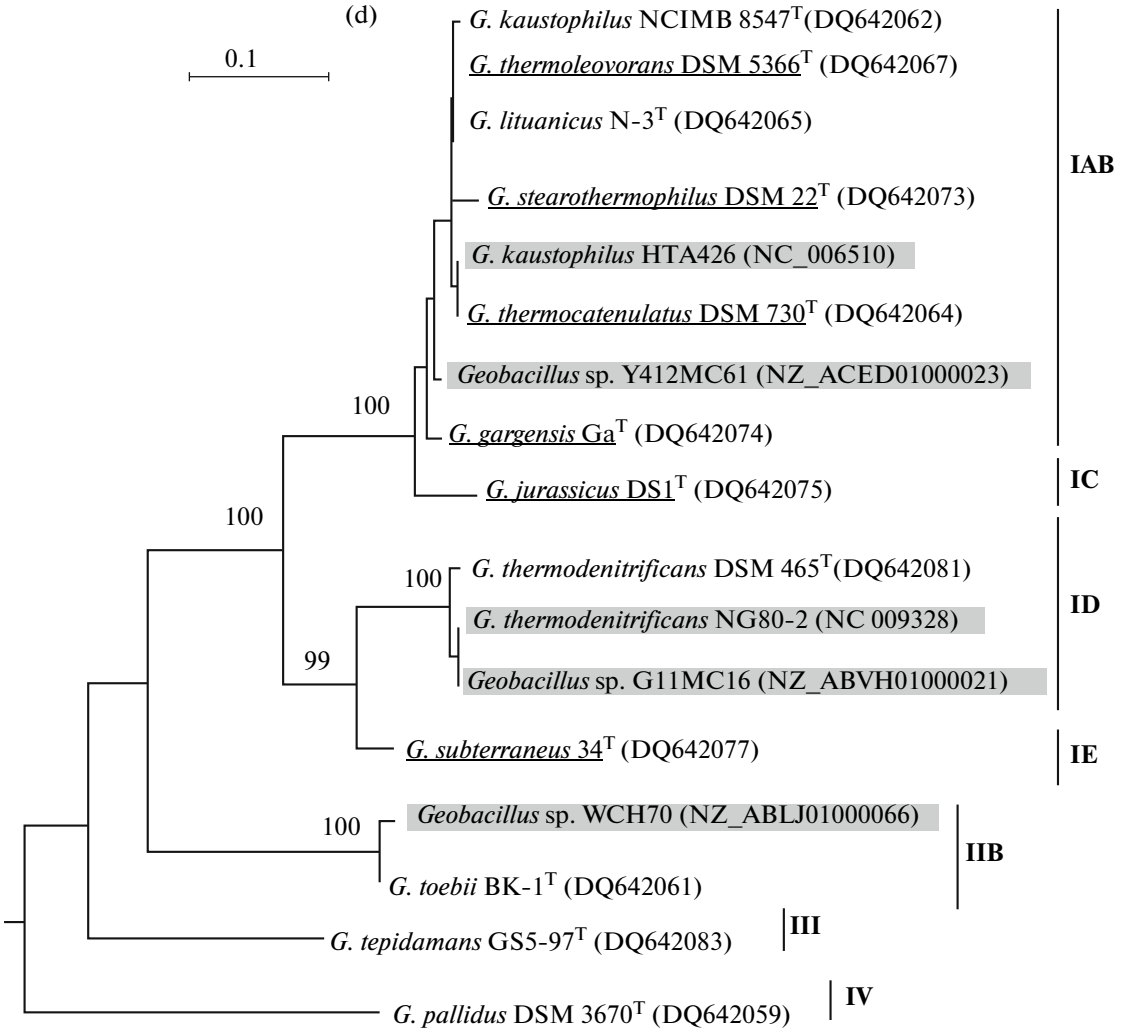

Fig. 2. Contd.

ilarity between the *gyrB* and *parE* gene sequences (99.6 and 99.3%, respectively), strain 46 may be affiliated with the species *G. stearothermophilus*. Since the *gyrB* and *parE* gene sequences of some type strains remain to be determined, the results of the identification of other studied strains are considered tentative. It should be noted that, in the case of strain 8m3, which was tentatively identified as a representative of the species *G. pallidus*, only the *gyrB* gene sequence with a deep divergence from the *gyrB* gene sequences of other *Geobacillus* strains (30%) was obtained. We failed to obtain the *parE* gene sequence of this strain, probably due to some changes at the primer binding site. It has been previously suggested [31] that *G. pallidus* should be excluded from the genus *Geobacillus* and assigned to *Aeribacillus pallidus* gen. nov., comb. nov. Although this proposal has not been validated yet, it does not contradict our results.

Changes in the *gyrB* gene sequence at the binding site of the primer Up2R can be detected using the results of analysis of the total genome of the reference strain WCH70 presumably belonging to *G. toebii*. Since, according to the level of similarity between the *parE* gene sequences (99.3–100%), the new strains B-1024 and vw3-1 belong to the same species, it may be suggested that similar changes occurred in the *gyrB* sequences of these strains, which prevented us from amplifying their *gyrB* gene.

On the basis of the results of analysis of the type II topoisomerase sequences, we propose that the *gyrB* and *parE* genes be used for determining the species composition of the genus *Geobacillus* as phylogenetic markers with the highest resolution. However, their application is limited due to the fact that the universality of the available primer systems is not sufficient, which can be improved by increasing the number of studied *Geobacillus* strains.

Table 2. Comparison of the results of DNA–DNA hybridization and the similarity level between the *gyrB* and *parE* gene sequences of *Geobacillus* species at the intra- and interspecific levels (%)

Strains and species	5366^T	22^T	466	7263^T	B-1259^T	Ga^T	DS1^T	DS2	U^T	X	34^T	Sam	K	Vw3-1	B-1024	3Feng
G. thermoleovorans DSM 5366^T	100															
G. stearothermophilus 22^T	51	100	84.8 87.7	85.5 89.9	85.8 91.8	86.9 91.4	87.0 88.6	87.6	87.8 89.0	88.0 88.6	85.4 86.9	85.8 87.9	85.3 88.2	80.5	80.5	77.9 82.6
G. thermodenitrificans DSM 466	31	32 48*	100	85.3 86.9	85.2 87.5	85.4 87.1	84.8 87.9	85.6	85.6 87.8	85.5 87.6	90.7 91.4	90.7 91.5	90.5 91.2	82.7	82.3	77.4 83.5
G. kaustophilus DSM 7263^T	54 84**	44 61*	41 40*	100	97.9 92.9	97.7 92.5	89.4 90.8	88.8	88.7 90.9	88.3 90.9	88.0 86.7	86.2 87.5	86.9 87.5	82.3	82.0	79.3 83.8
G. thermocatenulatus B-1259^T	51 73**	37	47	47	100	99.2 100	79.2 91.0	88.6	88.6 91.2	78.4 90.9	86.6 87.9	85.6 88.7	86.5 88.3	82.0	81.5	78.2 83.6
G. gargensis Ga^T	38	48	44	46	43	100	88.8 91.0	88.8	88.8 91.0	88.8 90.8	86.0 87.5	85.9 87.7	85.9 87.9	81.7	81.6	79.7 83.7
G. jurassicus DS1^T	53	44	33	53	51		100	100	100 99.7	100 99.6	86.0 87.2	85.3 87.6	86.2 87.6	81.0	81.2	79.9 84.0
G. jurassicus DS2							82	100	100	100	86.0	86.2	86.2			81.0
G. uzenensis U^T	45	38	45	45	54		49		100	100 99.8	86.0 87.7	86.2 88.4	86.2 88.0	80.8	80.8	79.8 84.0
G. uzenensis X	48	33	43		51		52		80	100	85.9 87.4	86.0 88.0	86.0 87.5	80.8	80.9	80.5 83.9
G. subterraneus 34^T	48	53	45	55	49	34	53		49	40	100	99.9 99.7	99.9 99.0	80.7	80.8	78.4 82.9
G. subterraneus Sam	45	37	47		48				42	44	91	100	100 99.4	81.9	81.1	78.5 83.5
G. subterraneus K	41	39	44		44				32	37	96	93	100	80.7	81.0	78.3 82.8
Geobacillus sp. vw3-1														100	99.8	87.4
Geobacillus sp. B-1024		16			27	21			15		19			98	100	88.5
Geobacillus sp. 3Feng				12										48		100

Note: The results of DNA–DNA hybridization are given below the diagonal. Our results [20, 21, 24]; * data obtained by Caccamo et al. [26]; ** data obtained by Sunna et al. [30]. The levels of similarity between the *gyrB* and *parE* sequences are given above the diagonal (above and below the line, respectively). For *G. kaustophilus*, the results of its comparison with the strain *G. kaustophilus* HTA426 are presented; for *G. thermodenitrificans*, the results of its comparison with the strain *G. thermodenitrificans* NG80-2 are presented. The intraspecific level is designated by gray background and bold lines; the differences in the results of analysis of the *gyrB* and *parE* gene sequences at the interspecific level are designated by the bold dotted line. Empty boxes indicate that there are no data available.

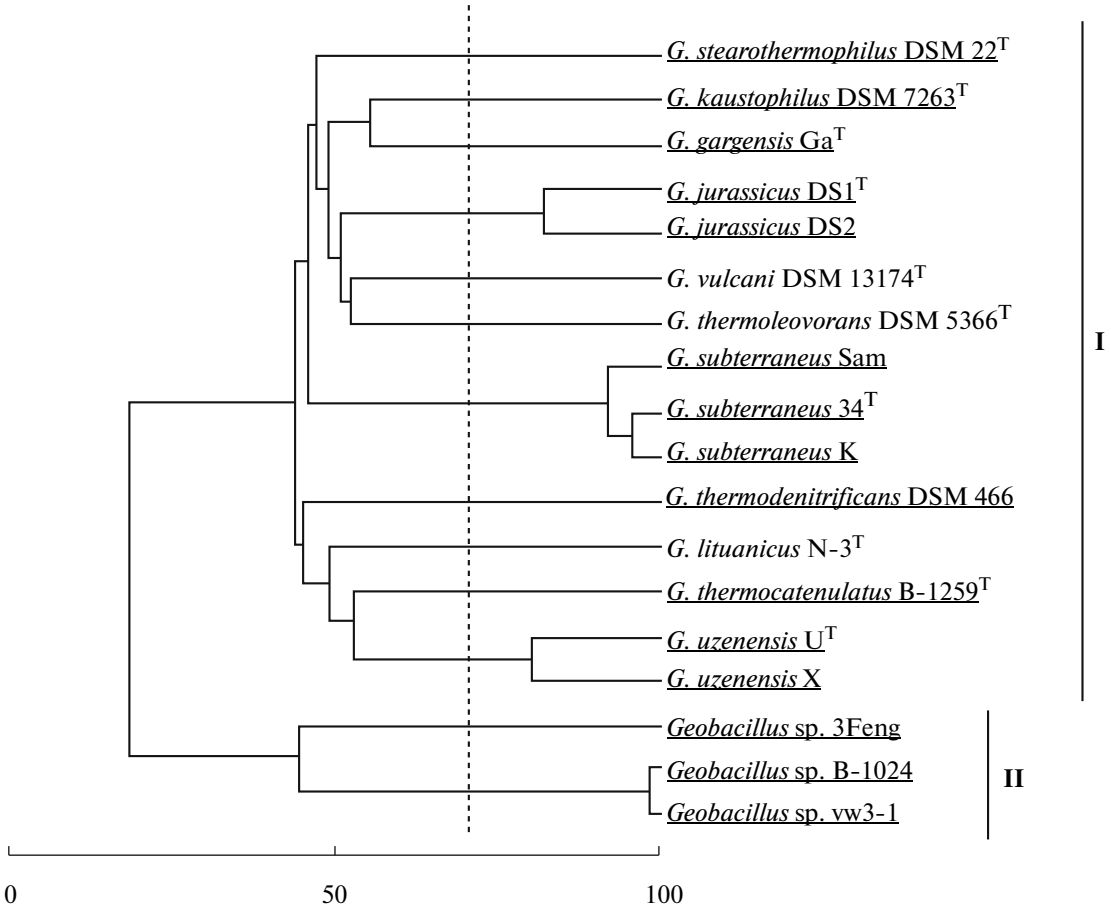

Fig. 3. Dendrogram based on the results of DNA–DNA hybridization representing the relatedness among the members of the genus *Geobacillus*. The strains for which data on the sequence structure of the *gyrB* and *parE* genes are available are underlined. The dendrogram was constructed on the basis of the average similarity matrix, for which missing values were calculated using the average values for a cluster. The trees were based on the average similarity matrix and constructed using the unweighted pair group method with arithmetic mean (UPGMA) implemented in the PHYLIP software package [32]. The bar shows the levels of DNA–DNA hybridization (%). The dotted line shows the intraspecific level of DNA–DNA homology.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (projects nos. 06-04-49128 and 05-04-39029), by the Ministry of Education and Science of the Russian Federation (grants NSh nos. 4174.2008.4, and 02.552.11.7073) and by the programs nos. 14 and 23 and "Molecular and Cell Biology" of the Presidium of the Russian Academy of Sciences.

REFERENCES

1. Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.O., Moore, L.H., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., and Trüper, H.G., International Committee on Systematic Bacteriology. Report of the ad hoc Committee on Reconciliation of Approaches to Bacterial Systematic, *Int. J. Syst. Bacteriol.*, 1987, vol. 37, no. 4, pp. 463–464.
2. Woese, C.R., Bacterial Evolution, *Microbiol. Rev.*, 1987, vol. 51, pp. 221–271.
3. Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A., Kämpfer, P., Maiden, M.C., Nesme, X., Rossello-Mora, R., Swings, G., Trüper, H.G., Vauterin, L., Ward, A.C., and Whitman, W.B., Report of the ad hoc Committee for the Re-Evaluation of the Species Definition in Bacteriology, *Int. J. Syst. Evol. Microbiol.*, 2002, vol. 52, pp. 1043–1047.
4. Ludwig, W., Nucleic Acid Techniques in Bacterial Systematics and Identification, *Int. J. Food. Microbiol.*, 2007, vol. 120, pp. 225–236.
5. Tourova, T.P., Copy Number of Ribosomal Operons in Prokaryotes and Its Effect on Phylogenetic Analyses, *Mikrobiologiya*, 2003, vol. 72, no. 4, pp. 437–452 [*Microbiology* (Engl. Transl.), vol. 72, no. 4, pp. 389–402].
6. Santos, S.R. and Ochman, H., Identification and Phylogenetic Sorting of Bacterial Lineages with Universally

Conserved Genes and Proteins, *Environ. Microbiol.*, 2004, vol. 6, pp. 754–759.

7. Zeigler, D.R., Application of a *recN* Sequence Similarity Analysis to the Identification of Species within the Bacterial Genus *Geobacillus, Int. J. Syst. Evol. Microbiol.*, 2005, vol. 55, pp. 1171–1179.
8. Zeigler, D.R., Gene Sequences Useful for Predicting Relatedness of Whole Genomes in Bacteria, *Int. J. Syst. Evol. Microbiol.*, 2003, vol. 53, pp. 1893–1900.
9. Yamamoto, S. and Harayama, S., Phylogenetic Relationships of *Pseudomonas putida* Strains Deduced from the Nucleotide Sequences of *gyrB*, *rpoD* and 16S rRNA Genes, *Int. J. Syst Bacteriol.*, 1998, vol. 48, pp. 813–819.
10. Weng, F.Y., Chiou, C.S., Lin, P.H.P., and Yang, S.S., Application of *recA* and *rpoB* Sequence Analysis on Phylogeny and Molecular Identification of *Geobacillus* Species, *J. Appl. Microbiol.*, 2009, vol. 107, pp. 452–464.
11. Meintanis C., Chalkou K.I., Kormas K.Ar., Lymperopoulou D.S., Katsifas E.A., Hatzinikolaou D.G., Karagouni A.D. Application of *rpoB* Sequence Similarity Analysis, REP-PCR and BOX-PCR for the Differentiation of Species within the Genus *Geobacillus, Lett. Appl. Microbiol.*, 2008, vol. 46, pp. 395–401.
12. Huang, W.M., Bacterial Diversity Based on Type II DNA Topoisomerase Genes, *Ann. Rev. Genetics*, 1996, vol. 30, pp. 79–107.
13. Suzuki, M., Narayama, S., and Yamamoto, S., Phylogenetic Analysis and Taxonomic Study of Marine *Cytophaga*-Like Bacteria: Proposal for *Tenacibaculum* gen. nov. with *Tenacibaculum maritimum* comb. nov. and *Tenacibaculum ovolyticum* comb. nov., and Description of *Tenacibaculum mesophilum* sp. nov. and *Tenacibaculum amylolyticum* sp. nov., *Int. J. Syst. Evol. Microbiol.*, 2001, vol. 51, pp. 1639–1652.
14. Wang, L.T., Lee, F.L., Tai, C.J., and Kasai, H., Comparison of *gyrB* Gene Sequences, 16S rRNA Gene Sequences and DNA-DNA Hybridization in the *Bacillus subtilis* Group, *Int. J. Syst. Evol. Microbiol.*, 2007, vol. 57, pp. 1846–1850.
15. Kasai, H., Watanabe, K., Gasteiger, E., Bairoch, A., Isono, K., Yamamoto, S., and Harayama, S., Construction of the *gyrB* Database for the Identification and classification of Bacteria, in *Genome Informatics*, Miyano, S. and Takagi, T., Eds., Tokyo: Universal Academic Press, 1998, pp. 13–21.
16. Yamamoto, S. and Harayama, S., PCR Amplification and Direct Sequencing of *gyrB* Genes with Universal Primers and Their Application to the Detection and Taxonomic Analysis of *Pseudomonas putida* Strains, *Appl. Environ. Microbiol.,* 1995, vol. 61, pp. 1104–1109.
17. Yamada, S., Ohashi, E., Agata, N., and Venkateswaran, K., Cloning and Nucleotide Sequence Analysis of *gyrB* of *Bacillus cereus, B. thuringiensis, B. mycoides*, and *B. anthracis* and Their Application to the Detection of *B. cereus* in Rice, *Appl. Environ. Microbiol.*, 1999, vol. 65, pp. 1483–1490.
18. Yamamoto, S. and Harayama, S., Phylogenetic Analysis of *Acinetobacter* Strains Based on the Nucleotide Sequences of *gyrB* Genes and on the Amino Acid Sequences of Their Products, *Int. J. Syst. Bacteriol.,* 1996, vol. 46, pp. 506–511.
19. Kang, Y., Takeda, K., Yazawa, K., and Mikami, Y., Phylogenetic Studies of *Gordonia* Species Based on *gyrB* and *secA*1 Gene Analyses, *Mycopathologia*, 2008, vol. 167, pp. 95–105.
20. Nazina, T.N., Tourova, T.P., Poltaraus, A.B., Novikova, E.V., Grigoryan, A.A., Ivanova, A.E., Lysenko, A.M., Petrunyaka, V.V., Osipov, G.A., Belyaev, S.S., and Ivanov, M.B., Taxonomic Study of Aerobic Thermophilic Bacilli: Descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from Petroleum Reservoirs and Transfer of *Bacillus stearothermophilus, Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the New Combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*, *Int. J. Syst. Evol. Microbiol.*, 2001, vol. 51, pp. 433–446.
21. Nazina, T.N., Sokolova, D.Sh., Grigoryan, A.A., Shestakova, N.M., Mikhailova, E.M., Poltaraus, A.B., Tourova, T.P., Lysenko, A.M., Osipov, G.A., and Belyaev, S.S., *Geobacillus jurassicus* sp. nov., a New Thermophilic Bacterium Isolated from a High-Temperature Petroleum Reservoir, and the Validation of the *Geobacillus* Species, *Syst. Appl. Microbiol.*, 2005, vol. 28, pp. 43–53.
22. Lane, D.J, 16S/23S RNA Sequencing, in *Nucleic Acid Techniques in Bacterial Systematics*, Stackebrandt, E. and Goodfellow, M., Eds., New York: Wiley, 1991, pp. 115–175.
23. Watanabe, K., Nelson, N., Harayama, S., and Kasai, H., ICB Database: the *gyrB* Database for Identification and Classification of Bacteria, *Nucleic Acids Res.*, 2001, vol. 29, pp. 344–345.
24. Nazina, T.N., Lebedeva, E.V., Poltaraus, A.B., Tourova, T.P., Grigoryan, A.A., Sokolova, D.Sh., Lysenko, A.M., and Osipov, G.A., *Geobacillus gargensis* sp. nov., a Novel Thermophile from a Hot Spring, and the Reclassification of *Bacillus vulcani* as *Geobacillus vulcani* comb. nov., *Int. J. Syst. Evol. Microbiol.*, 2004, vol. 54, pp. 2019–2024.
25. Kuisiene, N., Raugalas, J., and Chitavichius, D., *Geobacillus lituanicus* sp. nov., *Int. J. Syst. Evol. Microbiol.*, 2004, vol. 54, pp. 1991–1995.
26. Caccamo, D., Gugliandolo, C., Stackebrandt, E., and Maugeri, T.L., *Bacillus vulcani* sp. nov., a Novel Thermophilic Species Isolated from a Shallow Marine Hydrothermal Vent, *Int. J. Syst. Evol. Microbiol.*, 2000, vol. 50, pp. 2009–2012.
27. Schäffer, C., Franck, W.L., Scheberl, A., Kosma, P., McDermott, T.R., and Messner, P., Classification of Isolates from Locations in Austria and Yellowstone National Park as *Geobacillus tepidamans* sp. nov., *Int. J. Syst. Evol. Microbiol.*, 2004, vol. 54, pp. 2361–2368.
28. Banat, I.B., Marchant, R., and Rahman, T.J., *Geobacillus debilis* sp. nov., a Novel Obligately Thermophilic Bacterium Isolated from a Cool Soil Environment, and Reassignment of *Bacillus pallidus* to *Geobacillus pallidus* comb. nov., *Int. J. Syst. Evol. Microbiol.*, 2004, vol. 54, pp. 2197–2201.
29. Sung, M.H., Kim, H., Bae, J.W., Rhee, S.K., Jeon, C.O., Kim, K., Kim, J.J., Hong, S.P., Lee, S.G., Yoon, J.H., Park, Y.H., and Baek, D.H., *Geobacillus*

toebii sp. nov., a Novel Thermophilic Bacterium Isolated from Hay Compost, *Int. J. Syst. Evol. Microbiol.*, 2002, vol. 52, pp. 2251–2255.

30. Sunna, A., Tokajian, S., Burghardt, J., Rainey, F., Antranikian, G., and Hashwa, F., Identification of *Bacillus kaustophilus, Bacillus thermocatenulatus* and *Bacillus* Strain HSR as Members of *Bacillus thermoleovorans, Syst. Appl. Microbiol.*, 1997, vol. 20, pp. 232–237.
31. Minana-Galbis, D., Pinzón, D.L., Loren, J.G., Manresa, A., and Oliart-Ros, R.M., Reclassification of *Geobacillus pallidus* (Scholz et al. 1988) Banat et al. 2004 as *Aeribacillus pallidus* gen nov., comb. nov., *Int. J. Syst. Evol. Microbiol.*, 2009. doi: 10.1099/ijs.0.003699-0.
32. Felsenstein, J., PHYLIP, Phylogenetic Inference Package (Version 3.2), *Cladistics*, 1989, vol. 5, pp. 164–166.