



# Pan-Genome Portrait of *Bacillus mycooides* Provides Insights into the Species Ecology and Evolution

Krzysztof Fiedoruk,<sup>a</sup>  Justyna M. Drewnowska,<sup>b</sup> Jacques Mahillon,<sup>c</sup> Monika Zambrzycka,<sup>b</sup>  Izabela Swiecicka<sup>b,d</sup>

<sup>a</sup>Department of Microbiology, Medical University of Białystok, Białystok, Poland

<sup>b</sup>Department of Microbiology, Faculty of Biology, University of Białystok, Białystok, Poland

<sup>c</sup>Laboratory of Food and Environmental Microbiology, Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

<sup>d</sup>Laboratory of Applied Microbiology, Faculty of Biology, University of Białystok, Białystok, Poland

**ABSTRACT** *Bacillus mycooides* is poorly known despite its frequent occurrence in a wide variety of environments. To provide direct insight into its ecology and evolutionary history, a comparative investigation of the species pan-genome and the functional gene categorization of 35 isolates obtained from soil samples from northeastern Poland was performed. The pan-genome of these isolates is composed of 20,175 genes and is characterized by a strong predominance of adaptive genes (~83%), a significant amount of plasmid genes (~37%), and a great contribution of prophages and insertion sequences. The pan-genome structure and phylogenetic studies had suggested a wide genomic diversity among the isolates, but no correlation between lineages and the bacillus origin was found. Nevertheless, the two *B. mycooides* populations, one from Białowieża National Park, the last European natural primeval forest with soil classified as organic, and the second from mineral soil samples taken in a farm in Jasienówka, a place with strong anthropogenic pressure, differ significantly in the frequency of genes encoding proteins enabling bacillus adaptation to specific stress conditions and production of a set of compounds, thus facilitating their colonization of various ecological niches. Furthermore, differences in the prevalence of essential stress sigma factors might be an important trail of this process. Due to these numerous adaptive genes, *B. mycooides* is able to quickly adapt to changing environmental conditions.

**IMPORTANCE** This research allows deeper understanding of the genetic organization of natural bacterial populations, specifically, *Bacillus mycooides*, a psychrotrophic member of the *Bacillus cereus* group that is widely distributed worldwide, especially in areas with continental cold climates. These thorough analyses made it possible to describe, for the first time, the *B. mycooides* pan-genome, phylogenetic relationship within this species, and the mechanisms behind the species ecology and evolutionary history. Our study indicates a set of functional properties and adaptive genes, in particular, those encoding sigma factors, associated with *B. mycooides* acclimatization to specific ecological niches and changing environmental conditions.

**KEYWORDS** *Bacillus mycooides*, pan-genome, plasmids, phages, insertion sequences, environment, adaptation, sigma factors

The *Bacillus cereus sensu lato* encompasses endospore-forming aerobic rods present in a wide variety of habitats, including soil, water, and plant and animal tissues as well as food matrices (1–3). These Gram-positive bacilli display various forms of lifestyles, from saprophytic soil bacteria to intestinal symbionts of insects and mammals or to human pathogens (2, 4). Due to specific features such as endospore and/or toxin production, members of the *B. cereus* group have significant impacts on human health, agriculture, and food qualities. For instance, infection with *Bacillus anthracis*, the etiological agent of anthrax, can result in high mortality of humans and herbivores (5).

**Citation** Fiedoruk K, Drewnowska JM, Mahillon J, Zambrzycka M, Swiecicka I. 2021. Pan-genome portrait of *Bacillus mycooides* provides insights into the species ecology and evolution. *Microbiol Spectr* 9:e00311-21. <https://doi.org/10.1128/Spectrum.00311-21>.

**Editor** Jeffrey A. Gralnick, University of Minnesota

**Copyright** © 2021 Fiedoruk et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Izabela Swiecicka, [izabelas@uwb.edu.pl](mailto:izabelas@uwb.edu.pl).

**Received** 10 June 2021

**Accepted** 28 June 2021

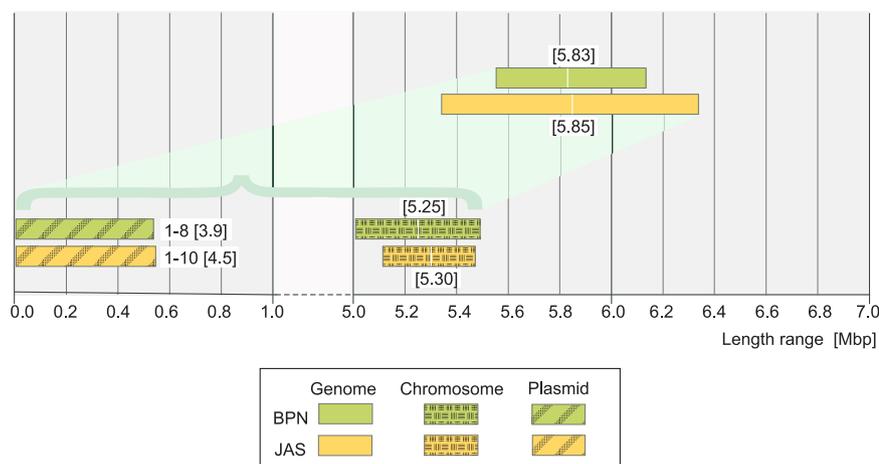
**Published** 21 July 2021

Also, some strains of *B. cereus sensu stricto* and *Bacillus cytotoxicus*, recognized as saprophytes or opportunistic foodborne pathogens, are occasionally involved in food poisoning and intestinal infections associated with a variety of toxins (6, 7). However, *Bacillus thuringiensis*, an insect pathogen, has remarkable biotechnological potential, as it is used as a biopesticide against numerous crop pests or to control vectors of animal diseases (8). Similarly, *Bacillus toyonensis* is exploited as a probiotic in animal feed (9). These bacteria still generate growing interest, for both scientific and economic reasons. This is in contrast with *Bacillus mycoides*, another member of the *B. cereus* group, which has so far remained much less studied, although it displays several interesting features, including adaptation to low temperature.

Due to recent reclassification within *B. cereus sensu lato* (10), *B. mycoides* encompasses psychrotrophic bacilli forming unusual rhizoidal colonies on agar plates and *Bacillus weihenstephanensis* forming regular colonies on solid media. The latter was distinguished from *B. cereus* based on its psychrotrophic character and the presence of specific signatures in the *cspA* and 16S rRNA genes (11). Yet, several studies have since suggested to merge these two species. For instance, Liu and coworkers (12) gathered *B. weihenstephanensis* and *B. mycoides* strains in one group with similarity of approximately 80% in the digital DNA-DNA hybridization (dDDH), while Drewnowska and Swiecicka (13) noted that isolates classified as these taxa fell into the same phylogenetic clade. Based on multifaceted analyses (e.g., dDDH, metabolic activities, physiological features, and chemotaxonomic traits), it was recommended to consider *B. weihenstephanensis* as a heterotypic synonym of *B. mycoides* (10). In fact, psychrotolerance and specific sequences in the 16S rRNA genes and *cspA* genes did not properly distinguish *B. weihenstephanensis* from some other *B. cereus sensu lato* (14). Based on this reclassification, in this work, we consider as *B. mycoides* both the isolates of rhizoidal colonies and those forming regular colonies, which have the ability to grow at temperatures ranging from 5 to 40°C. Adaptation to temperatures is one of the most important factors of bacterial diversification (15). *B. cereus sensu lato* comprises bacilli growing at temperatures ranging from psychrotrophic (*B. mycoides* and *B. wiedmannii*) through mesophilic (*B. anthracis*, *B. cereus sensu stricto*, and *B. thuringiensis*) to thermophilic (*B. cytotoxicus*) (4, 16). Moreover, thermotypes within the group adapted to different thermal niches are congruent with their ribosomal protein profiles (17) and the phylogenetic groups I to VII (7, 15). With regard to psychrotrophic species, while *B. wiedmannii* belongs to phylogroup II, *B. mycoides* (together with former *B. weihenstephanensis*) constitutes phylogroup VI in the scheme proposed by Drewnowska et al. (7) and Guinebrière et al. (15).

Strains of our *B. mycoides* collection have been isolated from soil samples of continental cold climate areas according to the Köppen-Geiger climate classification (7). Interestingly, it was noted that the *B. mycoides* isolates differed in their capacity to utilize chitin (18) and to synthesize water-soluble melanin (19). Nevertheless, little is known about the ecology and adaptation of *B. mycoides* to different habitats or the extent of species diversity, especially from a genomic standpoint. It seems that *B. mycoides* plays important roles in the functioning of natural environments, and some genes present in its genome can be transferred between members of the group via, *inter alia*, the conjugation processes, as observed between other members of *B. cereus sensu lato* (20, 21). Factually, *B. cereus sensu lato* is characterized by intensive genome plasticity. Plasmids, phages, and other genetic elements including insertion sequences (ISs) may constitute a significant part of their genomes (22–24). These genetic elements are, at the same time, involved in *B. cereus sensu lato* adaptation to diverse environments and species evolution mostly through mutation, recombination, or horizontal gene transfer (25, 26).

The development of next-generation sequencing (NGS) technologies has enriched public databases with numerous sequenced genomes, making it possible to expand the understanding of intra- and interspecies genome diversity and bacterial evolution. Especially, comparative genomic studies with the use the pan-genome concept allow



**FIG 1** Characteristics of genomes of *B. mycooides* isolated from soil samples collected in Białowieża National Park (BPN; green plots) and in a farm in Jasienówka (JAS; yellow plots), northeastern Poland. Average length of genomes and chromosomes are given in brackets. The range of the plasmid number and average number of plasmids per strain are shown next to the plasmid range size plots.

us to assess the genome modification and provide insights into bacterial evolution and adaptation to specific environmental niches (27–31). To obtain insight into *B. mycooides* biology and evolutionary history, we performed a deep comparative investigation of the *B. mycooides* pan-genome and the functional gene categorization of isolates originating from environments differing in nutrient status and anthropogenic pressure.

## RESULTS

**Genome features of *B. mycooides* from northeastern Poland.** The *B. mycooides* isolates under study were isolated from soil samples taken in northeastern Poland. Twenty-one originated from Białowieża National Park (BPN), the last European natural primeval forest under protection of World Heritage as a biosphere reserve and with soil classified as organic, while 14 strains were from mineral soil samples taken from a farm in Jasienówka (JAS), a place with strong anthropogenic pressure. Their genomes were sequenced using next-generation sequencing (NGS). The main features of the genomes and the National Center for Biotechnology Information (NCBI) database accession numbers are shown in Table S1 in the supplemental material, whereas detailed characteristics of individual isolates, such as (i) chromosomes, plasmids, and genome length, (ii) the number of coding DNA sequences (CDSs), (iii) the number of various types of RNAs, and (iv) GC content, are summarized in Table S2.

The average lengths of *B. mycooides* genomes were similar in both populations, but there was a slightly narrower range among the BPN isolates (from 5.55 to 6.13 Mb) than among the farm ones (from 5.34 to 6.34 Mb) (Fig. 1; Table S3). Both fit into the genome size range of the species reference strains, which varies from 4.88 to 6.50 Mb (NCBI). Plasmid lengths ranged from 3.78 to 538.90 kb and from 3.95 to 545.96 kb within the BPN and JAS populations, respectively (Fig. 1; Tables S2 and S3). The mean GC contents of the genomes were 33.8% (BPN isolates) and 33.6% (JAS isolates), while the number of CDSs varied from 5,463 to 6,466 among the BPN isolates and from 5,296 to 6,368 among the JAS isolates (Table S2).

***B. mycooides* pan-genome.** A pan-genome, considered all genes occurring in the analyzed data set, is generally divided into four categories of genes (core, soft core, shell, and cloud) based on their presence in the analyzed set of strains. The last two gene categories are considered the adaptive genes. As shown in Table 1 and Fig. 2, the *B. mycooides* pan-genome is complex, with a predominance of adaptive genes that account for up to 83.1%.

The *in silico* analysis detected 146 plasmids among the 35 *B. mycooides* isolates, contributing from 1.50% to 14.68% of their genomes (Fig. 1; Table S2). The presence of plasmids smaller than 100 kb was confirmed by their separation through agarose gel

**TABLE 1** Pan-genome and pan-plasmidome of *B. mycooides* isolates from northeastern Poland

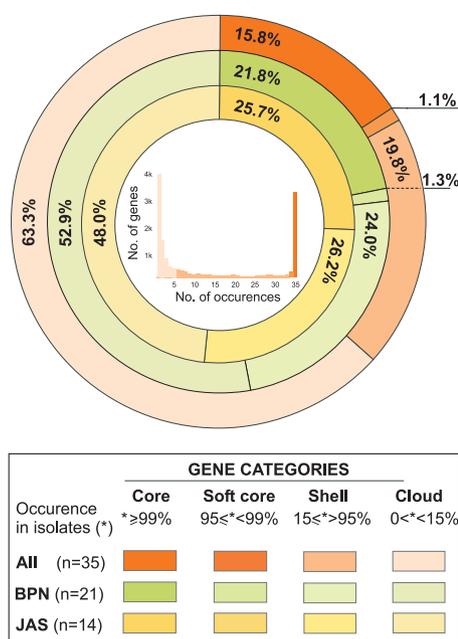
| Category        | No. of genes            |              |                       |                             |              |                       |
|-----------------|-------------------------|--------------|-----------------------|-----------------------------|--------------|-----------------------|
|                 | Pan-genome <sup>a</sup> |              |                       | Pan-plasmidome <sup>b</sup> |              |                       |
|                 | BPN (n = 21)            | JAS (n = 14) | All isolates (n = 35) | BPN (n = 21)                | JAS (n = 14) | All isolates (n = 35) |
| Core genes      | 3,409                   | 3,526        | 3,185                 | 4                           | 0            | 0                     |
| Soft-core genes | 201                     | 0            | 201                   | 0                           | 0            | 0                     |
| Shell genes     | 3,738                   | 3,597        | 4,008                 | 859                         | 877          | 827                   |
| Cloud genes     | 8,254                   | 6,608        | 12,781                | 4,070                       | 3,214        | 6,564                 |
| Total genes     | 15,602                  | 13,731       | 20,175                | 4,933                       | 4,091        | 7,391                 |

<sup>a</sup>Core genes are common to at least 99% of strains; soft-core genes are present in between 95% and 99% of strains; shell genes are present in between 15% and 95% of strains; cloud genes are present in less than 15% of strains; total genes, a collection of all genes present in the whole set of strains under study.

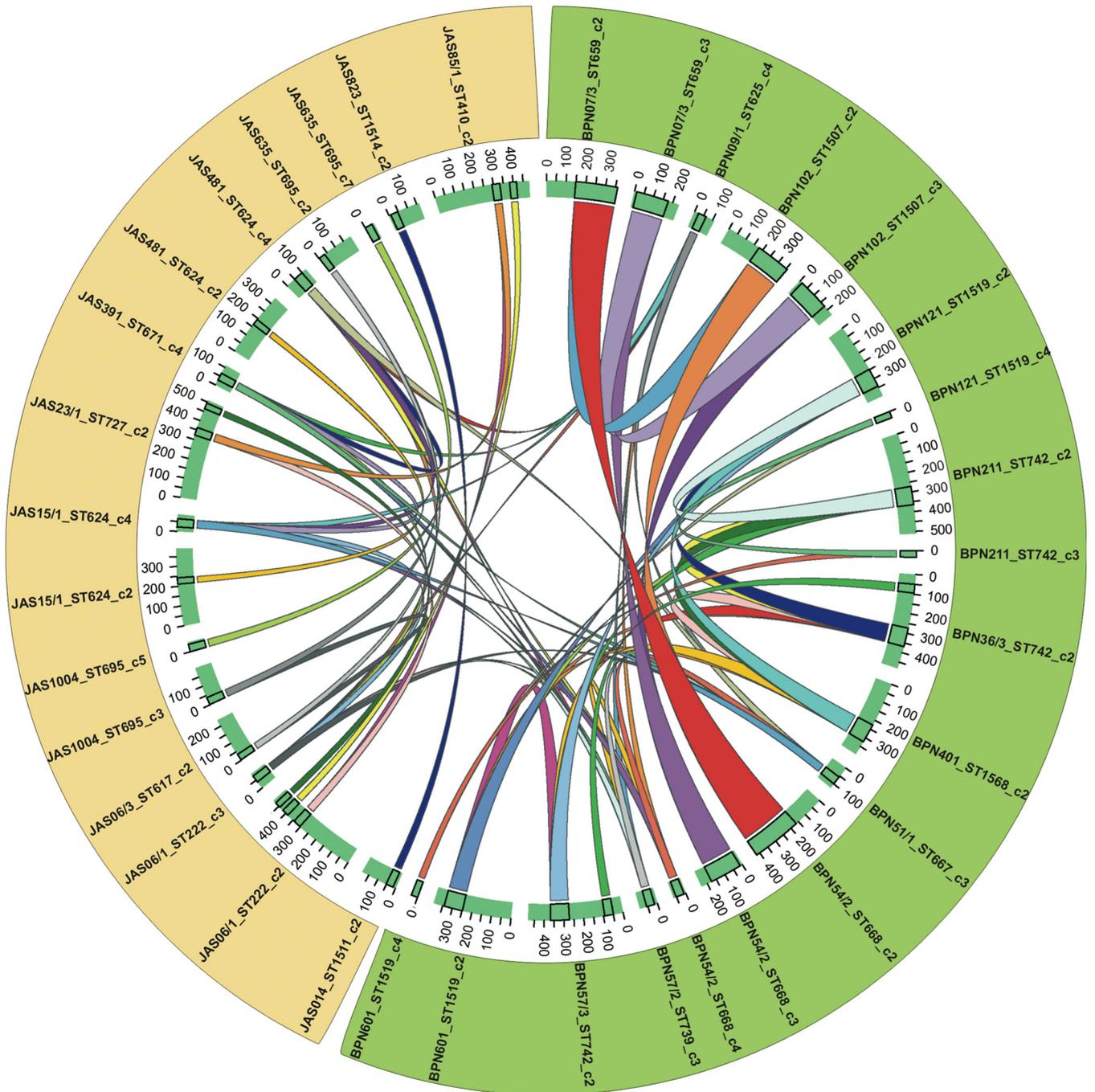
<sup>b</sup>Core genes, genes present on plasmids harbored by at least 99% of strains; soft-core genes, genes present on plasmids harbored by between 95% and 99% of strains; shell genes, genes present on plasmids harbored by between 15% and 95% of strains; cloud genes, genes present on plasmids harbored by less than 15% of strains; total genes, a collection of all genes present on plasmids harbored by the whole set of strains under study.

(Fig. S1). In the *B. mycooides* under study, 7,391 genes were found to be associated with plasmids ( $n = 146$ ). These genes were divided into the pan-plasmidome categories by extending the concept of the pan-genome (Table 1). Genes meeting the conditions of the core and soft-core pan-plasmidome were not found, yet 827 genes were classified into the shell category and 6,564 were classified as cloud genes. Interestingly, in the BPN plasmids, but not in the JAS ones, four genes were found to be common to all plasmids (Table 1). Nevertheless, a significant majority of the plasmid genes, 4,070 (82.5%) and 3,214 (78.6%) within BPN and JAS *B. mycooides*, respectively, occurred on plasmids harbored by less than 15% of the isolates under study. Plasmid comparison revealed the presence of common DNA fragments much more often between plasmids found among isolates of the same population, especially in those originating from Białowieża National Park (Fig. 3).

In the similarity network achieved with the plasmid ATLAS (pATLAS) platform, two large clusters comprising plasmids harbored by isolates from both locations were observed (Fig. 4). One cluster was composed of 31 small plasmids (<50 kb), and the



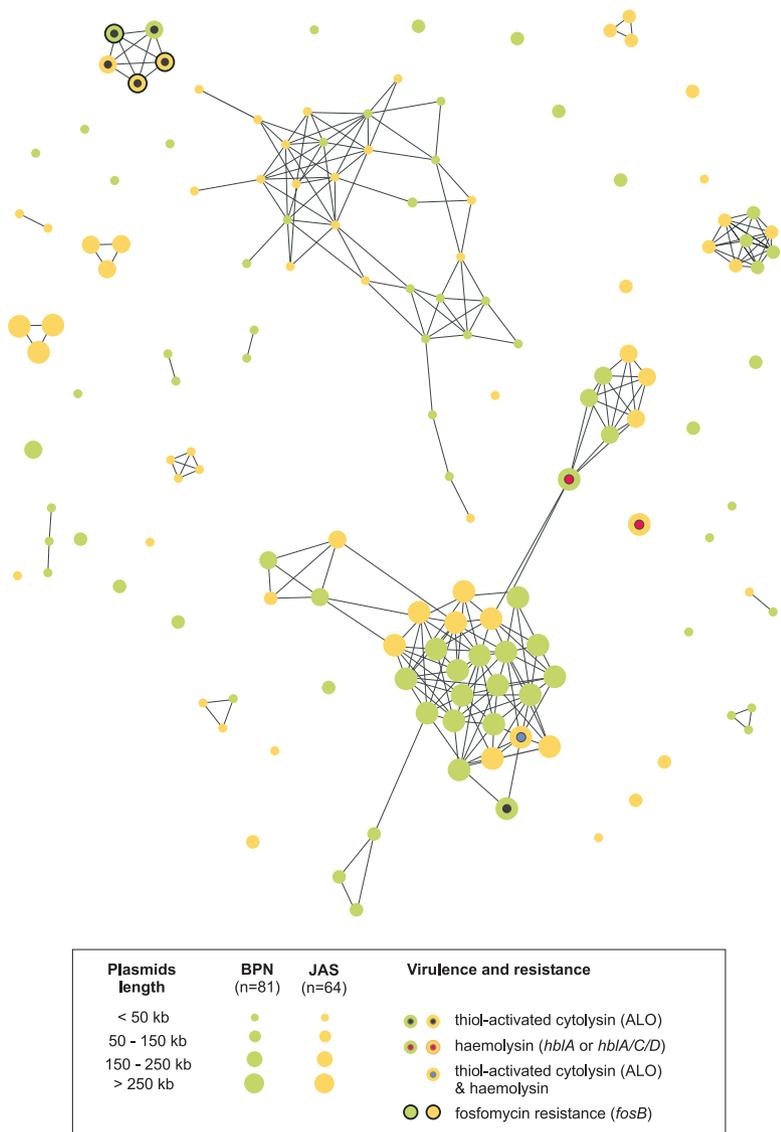
**FIG 2** Percentages of genome genes in each pan-genome category for all *B. mycooides* under study (orange ring), isolates from Białowieża National Park (BPN; green ring), and isolates from a farm in Jasienówka (yellow ring). The distribution of gene occurrences in *B. mycooides* genomes is shown in the graph in the center of the figure, where genes present in one genome only are indicated by the left peak, and those present in all genomes are indicated by the right peak.



**FIG 3** Nucleotide sequence-based comparison of 30-kb syntenic DNA fragments of plasmids harbored by *B. mycoides* under study. Plasmids present in the isolates from Białowieża National Park (BPN) are marked in green, while those from the Jasienówka farm (JAS) are marked in yellow. ST, sequence type; c2, c3, c4, and c5, contigs indicating plasmids in particular strains (for details, see Table S2 in the supplemental material).

second one was composed of 38 plasmids mostly of a size >250 kb, which formed four subclusters. The other plasmids ( $n = 76$ ) were singletons or clustered in small assemblages, localized separately from the above-mentioned large clusters. Note that nine *B. mycoides* plasmids were found to be potentially involved in toxinogenesis (encoding thiol-activated cytolysin and/or hemolysin BL) and fosfomycin resistance. Five of these plasmids, all encoding thiol-activated cytolysin, formed a single cluster (Fig. 4).

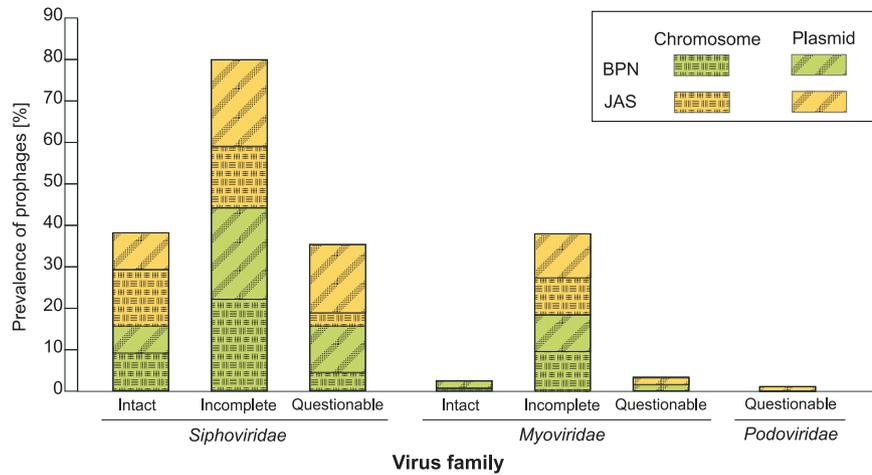
Besides plasmids, prophages are of the highest significance in lateral gene exchange in bacteria. Altogether in the genomes of 35 *B. mycoides* isolates under study, we found 37 putative, 35 questionable, and 108 incomplete prophages (Table



**FIG 4** Nucleotide sequence-based similarity network of *B. mycooides* plasmids from soil samples from Białowieża National Park (green circles) and a farm in Jasienówka (yellow circles) achieved with the plasmid ATLAS (pATLAS) program.

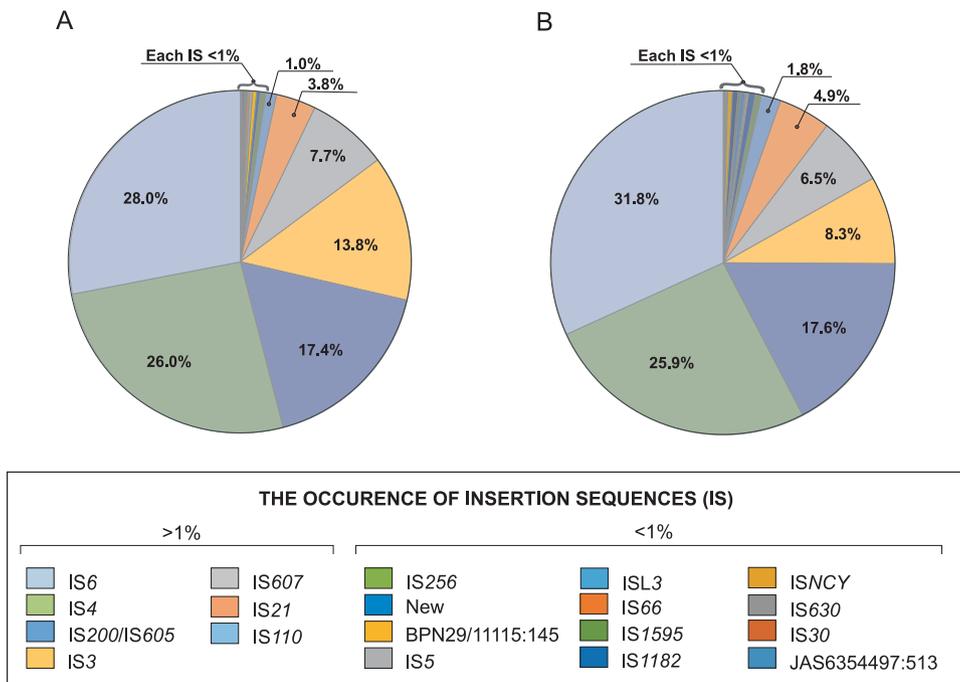
S4). An important proportion of these elements (59%) were located on plasmids. The prophages were classified to *Siphoviridae*, *Myoviridae*, and *Podoviridae*, with the predominance of the first family in both *B. mycooides* populations (Fig. 5). Detailed characteristics of prophages present in the *B. mycooides* isolates, e.g., prophage location, region length, total number of proteins encoded, GC content, and the most common phage corresponding to the sequence, are given in Table S4.

To improve on the understanding of the *B. mycooides* genome organization, we performed a thorough *in silico* analysis of the IS occurrence in Polish isolates of the species. Altogether, 1,381 and 921 potential ISs were found within the BPN and JAS populations, respectively, with the number in individual isolates ranging from 44 (BPN29/1) to 119 (BPN03/1) among BPN isolates and from 37 (JAS94/5) to 119 (JAS83/3) among JAS isolates (Table S5). ISs present in *B. mycooides* isolated from soil samples taken in Białowieża National Park were classified into 16 families (Fig. 6). With the exception of IS30, all of these families were also noted in the genomes of JAS isolates. In *B. mycooides* from both locations, members of the IS6 (~30%), IS4 (~26%), and IS200\_IS605 (~17%) predominated (Fig. 6).

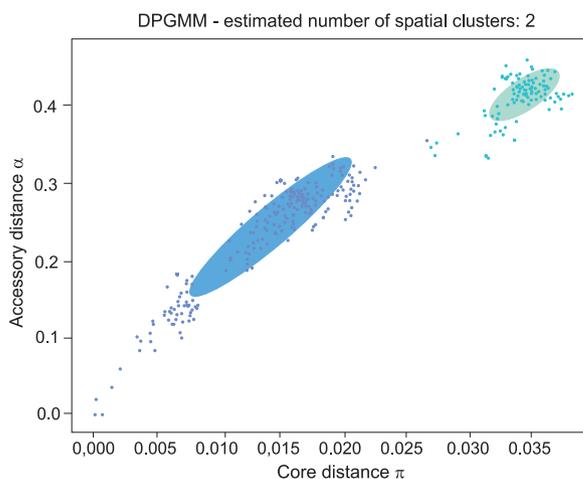


**FIG 5** Prevalences of prophages among *B. mycooides* populations from northeastern Poland. The prediction of whether the region contains a prophage was achieved with the PHASTER web server (accessed June 2020) based on a completeness score as follows: an intact phage, score >90%; an incomplete phage, score 70% to 90%; a questionable phage, score <70%.

**Phylogenetic studies of *B. mycooides*.** In a phylogenetic tree constructed using a maximum likelihood method from the aligned single nucleotide polymorphism (SNP) positions in chromosomes, isolates under study grouped together with three *B. mycooides* reference strains, whereas the other nine references (eight representing *B. cereus sensu lato* and *Bacillus subtilis* 168) clustered outside *B. mycooides* (Fig. S2). These results strongly indicate the monophyletic nature of the *Bacilli* examined. Interestingly, subclusters within *B. mycooides* included isolates from both populations. In the average nucleotide identity (ANI) analysis, performed in order to clarify the genetic relatedness between *B. mycooides*, the monophyletic nature of the isolates and the *B. mycooides* references was confirmed (Fig. S3). The ANI values ranged from 94.89% (BPN601 versus JAS014) to 99.98% (BPN121 versus BPN601) among isolates and from 94.90% (DSM11821



**FIG 6** Occurrences of insertion sequences (ISs) in genomes of *B. mycooides* originating from Białowieża National Park (A) and a farm in Jasienówka (B).



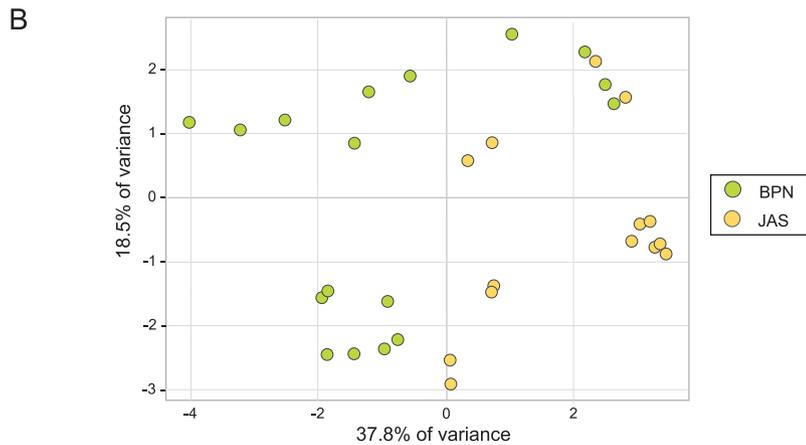
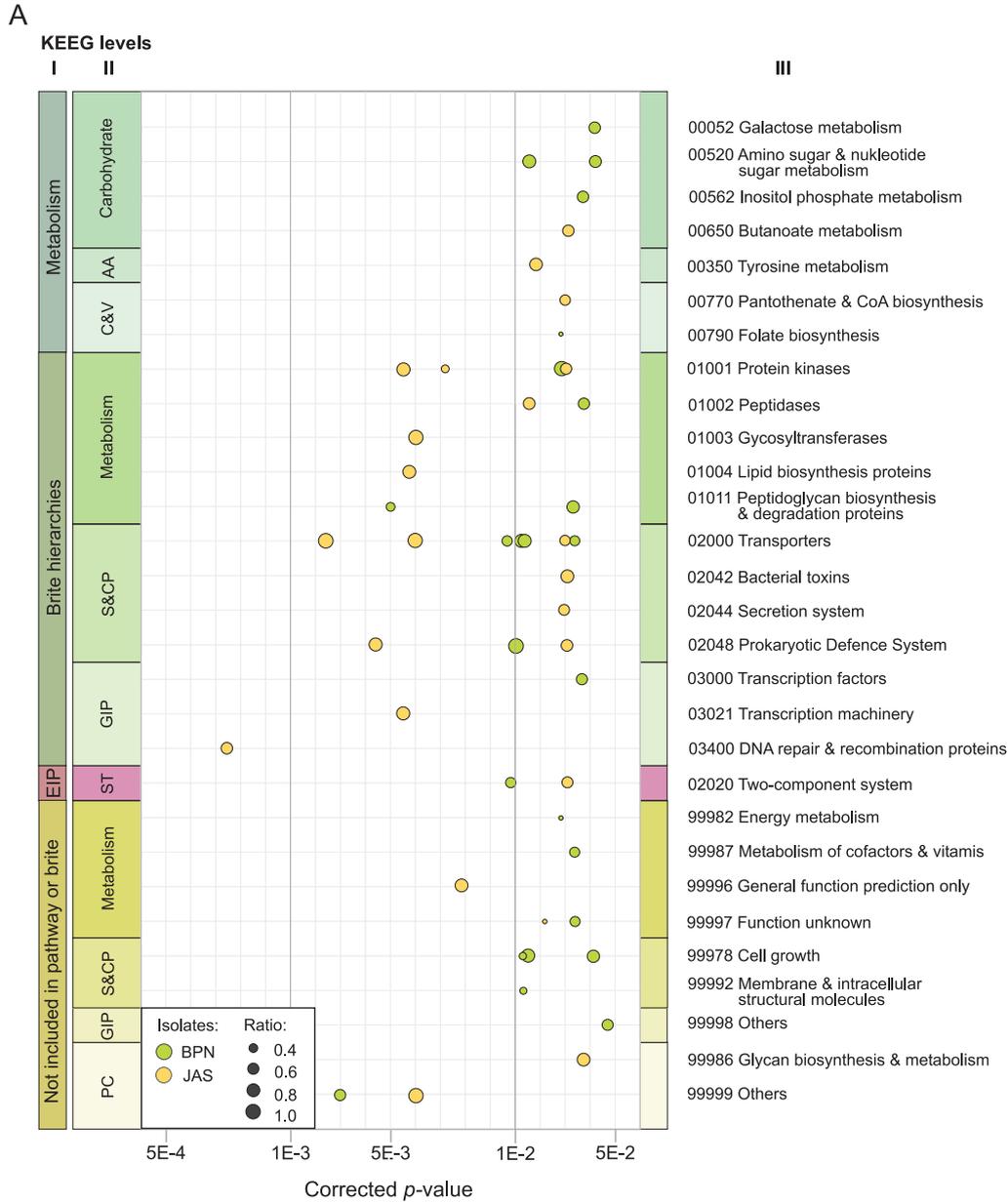
**FIG 7** The distribution of core ( $\pi$ , x axis) and accessory ( $\alpha$ , y axis) pairwise genome distances of *B. mycoides* under study. The distances were calculated using Dirichlet process Gaussian mixture Model (DPGMM) in the PopPUNK software.

versus JAS014) to 99.51% (ATCC 6462 versus JAS48/1) when the *B. mycoides* reference strains were considered. As expected, the topologies of both the phylogenetic tree (Fig. S2) and the ANI tree (Fig. S3) were similar, i.e., the isolates closely related in the phylogenetic tree fell into the same ANI assemblages, such as in the case of BPN121, BPN211, BPN36/3, BPN573, BPN601, and BPN37/2.

The distribution of core and accessory pairwise chromosome distances with PopPUNK showed the presence of two clusters within *B. mycoides*, each containing both BPN and JAS isolates (Fig. 7; Table S2). The 28 bacilli forming the first cluster (a blue ellipse in Fig. 7) are more similar in both their core ( $\alpha$ ) and accessory ( $\pi$ ) genomes ( $\alpha$  ranged from 0.15 to 0.32;  $\pi$  ranged from 0.006 to 0.020), compared to the second one (a turquoise ellipse) comprising seven isolates ( $\alpha$  ranged from 0.38 to 0.48;  $\pi$  ranged from 0.032 to 0.039).

**Functional characterization of *B. mycoides* genes.** The classification against Kyoto Encyclopedia of Genes and Genomes (KEGG) with the use of the BlastKOALA algorithm (32), summarized in Table S7, was achieved for 106,028 (61.8%) of 171,469 and 6,445 (45.5%) of 14,196 chromosomal and plasmid annotated proteins, respectively. In *B. mycoides* from both locations, the major portions of genes were associated with metabolism, with a predominance of carbohydrate metabolism, amino acid metabolism, and metabolism of cofactors and vitamins among both the chromosomal and plasmid genes. The second largest number of orthologs pertained to the environmental information processing category. In both populations, this group encompassed around one-fourth of the chromosomal genes and even higher proportions of the plasmid genes (specifically, 27% and 32% of plasmid genes in the BPN and JAS bacilli, respectively). The genes grouped in the genetic information processing made the third largest group, with a predominance of the translation category among the chromosomal genes and replication/repair among the plasmid genes.

Although the proportion of the KEGG functional level 2 categories among *B. mycoides* populations originating from Białowieża National Park and Jasienówka farm were similar (Table S7), significant differences between these two populations were found within the KEGG level 3 categories (Fig. 8). For instance, 52.4% of BPN isolates (versus 14.3% of JAS isolates) carried a gene coding the mycobilin-producing heme oxygenase (EC 1.14.99.57), and 61.2% of them (versus 21.4% of JAS isolates) possessed an operon responsible for *myo*-inositol utilization. Meanwhile, most JAS isolates (85.7%) and only 33.3% of BPN isolates possessed the *sigB* operon (Fig. S4), and the majority of them (71.4%) possessed loci carrying genes associated with lantibiotic synthesis; this feature seems to be less common within the BPN population (19.0%). On the other



**FIG 8** Significant differences in the presence of the KEGG level 3 categories in *B. mycooides* from Białowieża National Park (BPN) and the Jasienówka farm (JAS). (A) Differences assessed based on Fisher exact test ( $P < 0.05$ ). (B) Principal- (Continued on next page)

hand, an operon encoding alternative  $\sigma^{70}$  family RNA polymerase sigma factor (denoted *sigM-3*) seems to be unique for BPN isolates (71.4%) (Fig. S5). In addition, 64.3% of the JAS isolates (versus 4.8% of BPN isolates) carried the DNase V gene (EC 3.1.21.7).

## DISCUSSION

Prokaryotes are characterized by immense changeability, mostly associated with (i) episodic mutations, (ii) transposition, (iii) horizontal gene transfer (HGT), especially through transformation, transduction, and conjugation, and (iv) gene loss events, which are not as noticeable as gene acquisition via HGT. The balance between these processes reveals a definite shape of gene occurrence in prokaryote genomes. As a result, huge variations in gene content can be observed within species or even within clones or lineages (33). While the core and the soft-core genes encode proteins associated with fundamental biological processes and phenotypic features, adaptive genes significantly contribute to the species diversity by encoding supplementary biochemical pathways and additional functions, mostly essential for selective advantages to different niches (27, 34). The predominance of adaptive genes (83.1%) in the *B. mycooides* pan-genome allows these bacilli to inhabit various ecological niches and to function in a variety of environments.

The differentiation of bacterial pan-genomes is closely related to the species life-style (29) as well as to genome features, such as the ability to integrate exogenous DNA (34, 35). For instance, in obligate intracellular pathogens, such as *Piscirickettsia salmonis* causing salmonid rickettsial septicemia (36) and *Chlamydia* spp. involved in pneumonia infections (37), the core genes constitute approximately 50% of their pan-genomes. In contrast, members of the *B. cereus* group display smaller core genomes, e.g., 31% for *B. cereus* (38), 39% for *B. thuringiensis* (38), and 16% for *B. mycooides* (this study), indicating a much higher potential for adaptation. Interestingly though, *B. anthracis*, the most pathogenic member of the *B. cereus* group, also recognized as a monophyletic species (5), has a core genome that reaches 64% of its pan-genome (38).

*B. cereus sensu lato* plasmids can make up a large part of genomes and play crucial roles in many processes linked to secondary metabolism, virulence, and environmental adaptation (22, 25, 39, 40). However, whereas plasmids in *B. cereus sensu lato* species known for pathogenicity, such as *B. anthracis* or *B. thuringiensis*, have been extensively studied (8, 24, 26, 40), knowledge about these genetic elements in the other members of the group, such as those harbored by *B. mycooides*, remains limited. Our study showed a richness of genes associated with *B. mycooides* plasmids, as all of them represented the adaptive pan-plasmidome category. Interestingly, the presence of common DNA fragments among plasmids found from isolates of the same population suggests that horizontal gene transfer (HGT) may occur frequently among *Bacilli* classified to the species. Recently, Hu et al. (21) evidenced that in soil, *B. thuringiensis* plasmids with conjugal and mobilizing capability provoke gene transfer within the *B. cereus* group populations and, in consequence, potentially modify the population structure of these bacilli. Thus, a similar process cannot be excluded with regard to *B. mycooides*, especially in soil rich in nutrients, as it was noted among the Białowieża National Park samples. Similarly, the *B. mycooides* plasmids were found to be potentially involved in intestinal infection and antibiotic resistance. Moreover, as for other *B. cereus sensu lato* members, *B. mycooides* harbors prophages on both chromosomes and plasmids. It is worth noting that these prophages contained a significant number of CDSs (3,685 and 2,360 within BPN and JAS isolates, respectively), which may be of great importance for the functioning of these bacilli in diverse ecological niches and their evolution.

### FIG 8 Legend (Continued)

component analysis (PCA) of the statistically significant differences in the presence of the level 3 KEGG categories in the two populations. Ratio indicates the quotient of the number of strains carrying a gene to the number of strains in the population. Isolates belonging to each population are identified in green (BPN) and yellow (JAS).

Differentiation among species of the *B. cereus* group mostly results from lateral gene transfer, intragenomic rearrangement within genomes, and the ability to integrate in genomes as exogenous DNA. These processes are mostly linked to insertion sequences (ISs), which are common in prokaryotic genomes and have the capacity to import many genes into bacteria over short times (25, 26, 41). ISs are considered the main actors in evolutionary dynamics within bacteria due to their involvement in various DNA rearrangements, including insertions and deletions (so-called indel events), duplications, and inversions (32). For instance, indel events were responsible for the disruption of the *sigM-3* operon in JAS isolates (Fig. S5). As with other bacterial species, ISs have likely affected genome plasticity displayed by *B. cereus sensu lato* (25). In the *B. mycooides* genomes under study, IS6, IS4, and IS200\_IS605 predominated, which were also indicated by Fayad et al. (25) in genomes of four *B. mycooides* strains and two *B. weihenstephanensis* strains. These authors observed variability of the IS distribution among *B. cereus sensu lato* species and concluded that some families such IS3, IS4, and IS200/IS605 are ubiquitous among members of the group, whereas others such as IS1, IS30, and IS982 were seldom, which is in line with our observations of the 35 *B. mycooides* isolates. It is noteworthy that IS3, IS4, and IS6 are also dominant in genomes of *B. thuringiensis* serovar israelensis (26). Such a numerous representation of ISs in *B. cereus sensu lato* gives these bacilli additional possibilities of dynamic changes of their genomes, successfully adapting these bacteria to the occupied environments. Indeed, colocalization of disrupted *ISBth5* element (IS4 family) possibly explains the uneven distribution of the *sigB* operon between BPN and JAS populations (see Fig. S5 in the supplemental material).

Sigma factors, in particular, the extracytoplasmic function (ECF) family, owing to their fine-tune regulation of stress response genes, are considered key elements of adaptation to specific environmental changes and factors shaping bacterial evolution (42). For instance, duplication (or loss) of sigma factor genes followed by their further divergence and specialization in controlling of a smaller number of genes, e.g., those critical for the adaptation and survival in new ecological niches, was proposed by Schmidt et al. (43) as an important driving force behind the evolution of *B. cereus sensu lato* members. More importantly, such subfunctionalization of sigma factors may subsequently lead to divergence/specialization of the regulated genes. Therefore, it will be interesting to investigate whether the differences in SigB and SigM-3 prevalence between BPN and JAS isolates is correlated with differential regulation. Although *B. cereus* SigB was recognized as an important heat shock protein (HSP) from group II (i.e., with maximum induction after 15 min, which also involves factors activated by other stresses, such as salt or ethanol [44]), the function of SigM-3 remains to be elucidated. Nevertheless, from the evolutionary perspective, SigM-3 may represent a primal adaptation to the unique environment of the natural primeval forest, since it represents a different type of sigma factor (Fig. S6) and seems to be less redundant in JAS isolates.

In the PopPUNK analysis, the *B. mycooides* strains under study were classified into two clusters. Interestingly, in the “blue” cluster (with lower values of the  $a$  and  $\pi$  parameters) the average IS count was smaller than for the bacteria of the “turquoise” group (with higher values of the  $a$  and  $\pi$  parameters). Although the low number of bacilli studied does not allow any final conclusions to be drawn, these results suggest that among mobile genetic elements (MGEs), ISs play an important role in genetic variation of *B. mycooides*, as was also earlier observed for other members of the group, especially *B. thuringiensis* and *B. cereus sensu stricto* (25, 45).

The ANI analysis brought strong evidence that the isolates under study belong to *B. mycooides* according to a scheme proposed by Carroll et al. (46). Yet, in both the *B. mycooides* phylogenetic tree and the ANI heat map, the isolates displaying rhizoidal growth and those forming regular colonies on solid media occurred in the same clades within the phylogenetic tree, proving additional legitimacy of the *B. mycooides* and *B. weihenstephanensis* reclassification into one species within the *B. cereus* group (10). Moreover, although some BPN and JAS *B. mycooides* isolates showed the tendency to

group together, there was no a clear distinction between bacilli originating from the two locations, i.e., Białowieża National Park and the Jasienówka farm.

Next-generation sequencing of bacterial isolates and the search for functional categories of *B. mycooides* genes allowed us, on one hand, to portray the ortholog content of these bacilli and, on the other hand, to report functional differences among the pan-genomes of isolates originating from diverse environments (33). In the KEGG classification accomplished in this study, the major portions of genes were associated with metabolism, environmental information processing, and genetic information processing. Interestingly, a large proportion of plasmid genes in *B. mycooides* was sorted into the replication/repair category. Indeed, a substantial portion of genes from this category was earlier observed in the accessory genomes of other *B. cereus sensu lato* isolates and in the *B. subtilis* group by Kim et al. (38). The authors suggested that these genes are often engaged in mobilome-related processes (38), which has a significant role in the evolution of this species (25).

*B. mycooides* from Białowieża National Park and Jasienówka farm differed significantly in the presence of genes encoding specific proteins enhancing these bacilli in their natural environments. For instance, BPN isolates can use inositol as a carbon source due to the presence of the operon responsible for stepwise reactions of *myo*-inositol. Indeed, this sugar cyclic polyalcohol is common in soil and plants (47, 48), which may suggest that the *myo*-inositol degradation operon gives environmental advantages to occupy niches where this compound is present as a source of carbon (47). Along the same lines, most JAS isolates (i) can synthesize lantibiotic, as it was also observed among other members of the *B. cereus* group (49), and (ii) possess DNase V (EC 3.1.21.7), providing an alternative repair mechanism for DNA damage caused by, *inter alia*, environmental factors such as nitric oxide free radicals (50). Moreover, similarly to other representatives of the *Bacillus* genus, such as the *B. subtilis* group (51), *B. mycooides* and other members of *B. cereus sensu lato* produce a set of compounds (52, 53) facilitating the colonization of various ecological niches.

In conclusion, in this study, we mainly focused on the identification of the *B. mycooides* pan-genome and its functional properties associated with adaptation of these bacilli to specific ecological niches. Thanks to this study, the number of genomes available for this species increased from 74 to 109, constituting a valuable data set for comparative analyses undertaken for both the *B. mycooides* species and the whole *B. cereus* group. Soil, the primary niche for *B. mycooides* and its relatives in the *B. cereus* group (2, 13), is a highly heterogenous environment, in which the content of nutrients, particle size, humidity, and chemical properties may change drastically in short space and time frames (21, 54). The results of the present study illustrate how *B. mycooides* and kin have and are well adapted to these ever-changing environments.

## MATERIALS AND METHODS

***B. mycooides* isolates.** Altogether, we studied 35 *B. mycooides* isolates from soil samples taken in highly diverse environments, e.g., in the natural reserve of Białowieża National Park ( $n = 21$ ; BPN isolates) and a farm in Jasienówka ( $n = 14$ ; JAS isolates), both located in northeastern Poland (see Table S1 in the supplemental material). Bacterial isolation, multilocus sequence typing (MLST) (<https://pubmlst.org/organisms/bacillus-cereus>), and soil sample examination were performed during a previous project (13). For this study, isolates were taken randomly from a set of *B. mycooides* preliminary characterized by MLST as mentioned above (13). With a few exceptions, the isolates represented different sequence types (STs): among the 21 BPN isolates, 17 had different STs, and the 14 JAS isolates belonged to 12 STs (for details see Table S1).

**Next-generation sequencing.** Genomic DNA of the *B. mycooides* isolates was extracted from an overnight culture of an isolate in Luria-Bertani (LB) broth using a DNeasy blood and tissue kit (Qiagen GmbH, Hilden, Germany) and the manufacturer's protocol for Gram-positive bacteria. The DNA quality was assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) and gel electrophoresis, while the DNA quantity was measured using both a Qubit 2.0 fluorometer with a Qubit double-stranded DNA (dsDNA) HS assay kit (Thermo Fisher Scientific) and a 2200 TapeStation instrument with a genomic DNA ScreenTape assay (Agilent Technologies Inc., Santa Clara, CA, USA).

Each isolate was sequenced using both Illumina (Illumina Inc., San Diego, CA, USA) and Oxford Nanopore Technologies (ONT; Oxford, United Kingdom) sequencing platforms. For Illumina, genomic libraries were prepared using a Nextera XT Index kit (Illumina Inc.) according to the manufacturer's

protocol (document no. 15031942 v02) and quantified by capillary electrophoresis, applying the Agilent high-sensitivity D5000 ScreenTape system (Agilent Technologies Inc.). Libraries were sequenced on a MiSeq instrument (Illumina) using v2 reagents with 2 by 250-bp paired-end reads. For Oxford Nanopore (ONT), genomic libraries were prepared using a ligation sequencing kit (SQK-LSK109) strictly according to the ONT protocol “Genomic DNA by ligation (SQK-LSK109)” (GDE\_0963\_v109\_rev5\_14Aug2019) with recommended third-party consumables, i.e., NEBNext formalin-fixed, paraffin-embedded (FFPE) repair mix and NEBNext end repair/dA-tailing module (New England BioLabs, Ipswich, MA, USA). Then, the libraries were sequenced with the MinION Mk 1B for 48 h using flow cell FC-106D-R9.4.1. ONT base calling was performed with ONT Albacore sequencing pipeline software (ver. 1.2.6).

**Genome assembly.** Unicycler (ver. 0.4.0) software with default options under the normal mode was used for hybrid genome assembly and circularization of Illumina and ONT reads. The assembly was preceded by cleaning the Illumina and ONT reads from adapters using Trimmomatic (ver. 0.36) (parameters: PE -phred33 ILLUMINACLIP: NexteraPE-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) (55) and Porechop (ver 0.2.0, <https://github.com/rwick/Porechop>) software, respectively, using the program default parameters. In addition, the Filtlong tool (ver. 0.1.1, <https://github.com/rwick/Filtlong>) was used for filtering ONT reads by quality and selecting  $\geq 1,000$ -bp reads for the assembly (parameters: min\_length, 1000; keep\_percent, 90; target\_bases, 500,000,000; trim-split, 100; mean\_q\_weightm 10).

**Pan-genome and pan-plasmidome assignment.** The *B. mycooides* pan-genome size and the gene categories were obtained with Roary software ver. 3.13.0, applying default options except identity for the blastp algorithm, which was set to 90% instead of 95% (56). Genes were divided into four categories as recommended by Inglin et al. (34) and Kaas et al. (57), e.g., core genes (common to at least 99% of strains), soft-core genes (present in between 95% and 99% of strains), shell genes (present in between 15% and 94% of strains), and cloud genes (present in less than 15% of strains). The Roary software ver. 3.13.0 with default options was also used in the *B. mycooides* pan-plasmidome determination, considering all plasmid genes present in the analyzed isolates and dividing them into four pan-plasmidome categories by extending the concept of the pan-genome (for details, see Table 1). Pan-genomes and pan-plasmidomes were computed for all *B. mycooides* isolates under study jointly ( $n = 35$ ) and separately for the BPN ( $n = 21$ ) and JAS ( $n = 14$ ) bacilli. The pan-plasmidome was considered a subset of the *B. mycooides* pan-genome.

Similarity of plasmids was assessed based on DNA 30-kb synteny fragments using the Sibelia program ver. 3.0.7 (58). The *B. mycooides* nucleotide sequence-based plasmid similarity network was constructed with the plasmid ATLAS (pATLAS) software (<http://www.patlas.site>) using the program default options and databases (59).

Antibiotic resistance genes and virulence determinants were searched using ABRicate v0.8 (<https://github.com/tseemann/abricate>), with the ResFinder, CARD ver. 2.0.3, and VFDB databases. Prophages present in *B. mycooides* under study were indicated with the use of PHASTER web server (accessed June 2020) (60), while insertion sequences were determined using ISEScan ver. 1.7.1 (61).

**Phylogenetic analysis.** The phylogenetic analysis was prepared with the use of nucleotide sequences of *B. mycooides* isolates under study. Furthermore, the nucleotide sequences of *B. mycooides* KBAB4, *B. mycooides* DSM 11821, *B. mycooides* ATCC 6462 (the species type strain), *B. cereus* ATCC 14589, *B. cereus* ATCC 10987, *B. thuringiensis* ATCC 10792, *B. thuringiensis* HD1, *B. anthracis* Ames, *B. wiedmannii* FSL W8-0169, *B. toyonensis* BCT-7112, *B. cytotoxicus* NVH391098, and *B. subtilis* 168 were included.

Construction of the tree reflecting the phylogenetic relationship of *B. mycooides* isolates was achieved using Reference Sequence Alignment-Based Phylogeny (RealPhy) program ver. 112 (62) with default options and RAXML ver. 8.2.10 software as the tree building tool with a bootstrap of 1,000 replicates (parameters: -m GTRGAMMA -p 12345 -s polymorphisms\_move.phy -o *B. subtilis* strain 168 -N 1000). The gene content similarity between strains was assessed using an average nucleotide identity (ANI) analysis that generated a gene content correlation matrix visualized as a heat map.

The distance between the core and accessory genes of pairs of the isolates under study was achieved with Population Partitioning Using Nucleotide *k*-mers (PopPUNK) program (<https://web.poppunk.net/>) based on *k*-mer differences (63). Default options were applied in all analyses.

**Gene functional annotation.** Genes of all *B. mycooides* strains under study were assigned to Kyoto Encyclopedia of Genes and Genomes (KEGG) functional categories using eggnoG-mapper (ver. 5.0) software (64) along with the BLAST of KEGG Orthology And Links Annotation (BlastKOALA) web server (65). In addition, the enrichM tool (<https://github.com/geronimp/enrichM>) was applied to compare differences in occurrence of KEGG categories between *B. mycooides* from the Białowieża National Park and from the Jasienówka farm using the Fisher exact test and principal-component analysis (PCA). The differences were interpreted as statistically significant at a *P* value of  $< 0.05$ .

**Plasmid isolation.** To verify the *in silico* analyses, plasmids of up to 100 kb in *B. mycooides* isolates were purified using the HiSpeed plasmid maxi kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's protocol, and separated in 1% Prona Plus agarose gel (Laboratories Conda, Spain) as described by Swiecicka et al. (8). Following the visualization with a Midori stain, the plasmid sizes were assessed by comparison with those in *B. thuringiensis* strain IS5056 (66).

**Data availability.** The data sets supporting this study have been deposited in the National Center for Biotechnology Information (NCBI) database under accession numbers: CP035953 to CP035961 (BPN03/1), CP035962 to CP035964 (BPN07/3), CP066847 to CP066850 (BPN08/1), CP035965 to CP035972 (BPN09/1), CP072065 to CP072068 (BPN29/1), CP035994 to CP035996 (BPN36/2), CP035997 to CP035998 (BPN36/3), CP035999 to CP036003 (BPN37/1), CP036004 to CP036008 (BPN37/2), CP036009 to CP036013 (BPN43/2), CP036017 to CP036022 (BPN51/1), CP036023 to CP036025 (BPN52/2), CP036026 to CP036033 (BPN54/2), CP036034 to CP036039 (BPN57/2), CP036042 to CP036045 (BPN58/4), CP035973 to CP035977

(BPN102), CP035978 to CP035983 (BPN121), CP035984 to CP035986 (BPN211), CP031071 to CP031077 (BPN401), CP036040 to CP036041 (BPN573), CP036046 to CP036051 (BPN601), CP072061 to CP072064 (JAS06/1), CP036064 to CP036069 (JAS06/3), CP072057 to CP072060 (JAS12/5), CP071811 to CP071818 (JAS15/1), CP036099 to CP036100 (JAS23/1), CP036137 to CP036144 (JAS83/3), CP072055 to CP072056 (JAS85/1), CP036145 to CP036148 (JAS95/4), CP071808 to CP071810 (JAS014), CP036102 to CP036108 (JAS391), CP036115 to CP036120 (JAS481), CP036121 to CP036131 (JAS635), CP036132 to CP036136 (JAS823), CP036074 to CP036084 (JAS1004). For details see, Table S1.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 1.8 MB.

## ACKNOWLEDGMENTS

K.F. participated in the study concept, shared in the designation of the research experiments, took part in the NGS sequencing, and performed all bioinformatic analyses. J.M.D. participated in the performing the NGS sequencing, submitted data to the NCBI, and prepared final versions of the figures. J.M. participated in the study concept and reviewed the manuscript. M.Z. performed the KEGG designation. I.S. prepared the grant and obtained funding for the research, developed the study concept, wrote the manuscript, supervised the study, and provided the project administration.

We thank Piotr Iwaniuk, our Ph.D. student, for *B. mycooides* plasmid isolation.

The project was financed by the grant of National Science Centre in Poland (no. 2015/19/B/NZ8/03944 to I. Swiecicka). Infrastructure used in the study is maintained from the Specific Scientific Equipment grant of the Ministry of Education and Science in Poland (decision no. 8636/E-342/SPUB/2018 to I. Swiecicka). MiSeq sequencer Illumina Inc. equips the Center of Synthesis and Analysis BioNanoTechno of University of Bialystok founded by the European Union as a part of the Operational Program Development of Eastern Poland 2007-2013 (project POPW.01.03.00-20-034/09-00).

We declare no conflict of interest.

## REFERENCES

- Swiecicka I, De Vos P. 2003. Properties of *Bacillus thuringiensis* isolated from bank voles. *J Appl Microbiol* 94:60–64. <https://doi.org/10.1046/j.1365-2672.2003.01790.x>.
- Ceuppens S, Boon N, Uyttendaele M. 2013. Diversity of *Bacillus cereus* group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles. *FEMS Microbiol Ecol* 84:433–450. <https://doi.org/10.1111/1574-6941.12110>.
- Vidic J, Chaix C, Manzano M, Heyndrickx M. 2020. Food sensing: detection of *Bacillus cereus* spores in dairy products. *Biosensors* 10:15. <https://doi.org/10.3390/bios10030015>.
- Swiecicka I. 2008. Natural occurrence of *Bacillus thuringiensis* and *Bacillus cereus* in eukaryotic organisms: a case for symbiosis. *Biocontrol Sci Technol* 18:221–239. <https://doi.org/10.1080/09583150801942334>.
- Mock M, Fouet A. 2001. Anthrax. *Annu Rev Microbiol* 55:647–671. <https://doi.org/10.1146/annurev.micro.55.1.647>.
- Stenfors Arnesen LP, Fagerlund A, Granum PE. 2008. From soil to gut: *Bacillus cereus* and its food poisoning toxins. *FEMS Microbiol Rev* 32:579–606. <https://doi.org/10.1111/j.1574-6976.2008.00112.x>.
- Drewnowska JM, Stefanska N, Czerniecka M, Zambrowski G, Swiecicka I. 2020. Potential enterotoxicity of phylogenetically diverse *Bacillus cereus sensu lato* soil isolates from different geographical locations. *Appl Environ Microbiol* 86:e03032–19. <https://doi.org/10.1128/AEM.03032-19>.
- Swiecicka I, Bideshi DK, Federici BA. 2008. Novel isolate of *Bacillus thuringiensis* subsp. *thuringiensis* that produces a quasicuboidal crystal of Cry1Ab21 toxic to larvae of *Trichoplusia ni*. *Appl Environ Microbiol* 74:923–930. <https://doi.org/10.1128/AEM.01955-07>.
- Jiménez G, Blanch AR, Tamames J, Rosselló-Mora R. 2013. Complete genome sequence of *Bacillus toyonensis* BCT-7112<sup>T</sup>, the active ingredient of the feed additive preparation Toyocerin. *Genome Announc* 1:e01080-13. <https://doi.org/10.1128/genomeA.01080-13>.
- Liu Y, Lai Q, Shao Z. 2018. Genome analysis-based reclassification of *Bacillus weihenstephanensis* as a later heterotypic synonym of *Bacillus mycooides*. *Int J Syst Evol Microbiol* 68:106–112. <https://doi.org/10.1099/ijsem.0.002466>.
- Lechner S, Mayr R, Francis K, Prüss BM, Kaplan T, Wiessner-Gunkel E, Stewart GS, Scherer S. 1998. *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. *Int J Syst Bacteriol* 48:1373–1382. <https://doi.org/10.1099/00207713-48-4-1373>.
- Liu Y, Lai Q, Göker M, Meier-Kolthoff JP, Wang M, Sun Y, Wang L, Shao Z. 2015. Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Sci Rep* 5:14082. <https://doi.org/10.1038/srep14082>.
- Drewnowska JM, Swiecicka I. 2013. Eco-genetic structure of *Bacillus cereus sensu lato* populations from different environments in Northeastern Poland. *PLoS One* 8:e80175. <https://doi.org/10.1371/journal.pone.0080175>.
- Soufiane B, Côté J-C. 2013. *Bacillus weihenstephanensis* characteristics are present in *Bacillus cereus* and *Bacillus mycooides* strains. *FEMS Microbiol Lett* 341:127–137. <https://doi.org/10.1111/1574-6968.12106>.
- Guinebretière M-H, Thompson FL, Sorokin A, Normand P, Dawyndt P, Ehling-Schulz M, Svensson B, Sanchis V, Nguyen-The C, Heyndrickx M, De Vos P. 2008. Ecological diversification in the *Bacillus cereus* group. *Environ Microbiol* 10:851–865. <https://doi.org/10.1111/j.1462-2920.2007.01495.x>.
- Swiecicka I, Bartoszewicz M, Kasulyte-Creasey D, Drewnowska JM, Murawska E, Yernazarova A, Lukaszuk E, Mahillon J. 2013. Diversity of thermal ecotypes and potential pathotypes of *Bacillus thuringiensis* soil isolates. *FEMS Microbiol Ecol* 85:262–272. <https://doi.org/10.1111/1574-6941.12116>.
- Fiedoruk K, Drewnowska JM, Daniluk T, Leszczynska K, Iwaniuk P, Swiecicka I. 2017. Ribosomal background of the *Bacillus cereus* group thermotypes. *Sci Rep* 7:46430. <https://doi.org/10.1038/srep46430>.
- Drewnowska JM, Fiodor A, Barboza-Corona JE, Swiecicka I. 2020. Chitinolytic activity of phylogenetically diverse *Bacillus cereus sensu lato* from natural environments. *Syst Appl Microbiol* 43:126075. <https://doi.org/10.1016/j.syapm.2020.126075>.
- Drewnowska JM, Zambrzycka M, Kalska-Szostko B, Fiedoruk K, Swiecicka I. 2015. Melanin-like pigment synthesis by soil *Bacillus weihenstephanensis*

- isolates from Northeastern Poland. PLoS One 10:e0125428. <https://doi.org/10.1371/journal.pone.0125428>.
20. Makart L, Commans F, Gillis A, Mahillon J. 2017. Horizontal transfer of chromosomal markers mediated by the large conjugative plasmid pXO16 from *Bacillus thuringiensis* serovar israelensis. Plasmid 91:76–81. <https://doi.org/10.1016/j.plasmid.2017.04.001>.
  21. Hu X, Huang D, Ogalo J, Geng P, Yuan Z, Xiong H, Wan X, Sun J. 2020. Application of *Bacillus thuringiensis* strains with conjugal and mobilizing capability drives gene transmissibility within *Bacillus cereus* group populations in confined habitats. BMC Microbiol 20:363. <https://doi.org/10.1186/s12866-020-02047-4>.
  22. Swiecicka I, Mahillon J. 2005. The clonal structure of *Bacillus thuringiensis* isolates from north-east Poland does not correlate with their *cry* gene diversity. Environ Microbiol 7:34–39. <https://doi.org/10.1111/j.1462-2920.2004.00662.x>.
  23. Van der Auwera G, Mahillon J. 2008. Transcriptional analysis of the conjugative plasmid pAW63 from *Bacillus thuringiensis*. Plasmid 60:190–199. <https://doi.org/10.1016/j.plasmid.2008.07.003>.
  24. Fiedoruk K, Daniluk T, Mahillon J, Leszczynska K, Swiecicka I. 2017. Genetic environment of *cry1* genes indicates their common origin. Genome Biol Evol 9:2265–2275. <https://doi.org/10.1093/gbe/evx165>.
  25. Fayad N, Kallassy Awad M, Mahillon J. 2019. Diversity of *Bacillus cereus sensu lato* mobile. BMC Genomics 20:436. <https://doi.org/10.1186/s12864-019-5764-4>.
  26. Gillis A, Fayad N, Makart L, Bolotin A, Sorokin A, Kallassy M, Mahillon J. 2018. Role of plasmid plasticity and mobile genetic elements in the entomopathogen *Bacillus thuringiensis* serovar israelensis. FEMS Microbiol Rev 42:829–856. <https://doi.org/10.1093/femsre/fuy034>.
  27. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-genome. Curr Opin Genet Dev 15:589–594. <https://doi.org/10.1016/j.gde.2005.09.006>.
  28. Yu J, Zhao J, Song Y, Zhang H, Sun Z. 2018. Comparative genomics of the herbivore gut symbiont *Lactobacillus reuteri* reveals genetic diversity and lifestyle adaptation. Front Microbiol 9:1151. <https://doi.org/10.3389/fmicb.2018.011151>.
  29. Azarian T, Huang I-T, Hanage WP. 2020. Structure and dynamics of bacterial populations: pangenome ecology, p 115–128. In Tettelin H, Medini D (ed), The pangenome: diversity, dynamics and evolution of genomes. Springer International Publishing, Cham, Switzerland.
  30. Laing CR, Whiteside MD, Gannon VPJ. 2017. Pan-genome analyses of the species *Salmonella enterica*, and identification of genomic markers predictive for species, subspecies, and serovar. Front Microbiol 8:1345. <https://doi.org/10.3389/fmicb.2017.01345>.
  31. Zhang X, Liu Z, Wei G, Yang F, Liu X. 2018. *In silico* genome-wide analysis reveals the potential links between core genome of *Acidithiobacillus thiooxidans* and its autotrophic lifestyle. Front Microbiol 9:1255. <https://doi.org/10.3389/fmicb.2018.01255>.
  32. Romaniuk K, Golec P, Dziewit L. 2018. Insight into the diversity and possible role of plasmids in the adaptation of psychrotolerant and metalotolerant *Arthrobacter* spp. to extreme Antarctic environments. Front Microbiol 9:3144. <https://doi.org/10.3389/fmicb.2018.03144>.
  33. Koskella B, Vos M. 2015. Adaptation in natural microbial populations. Annu Rev Ecol Syst 46:503–522. <https://doi.org/10.1146/annurev-ecolsys-112414-054458>.
  34. Inglin RC, Meile L, Stevens MJA. 2018. Clustering of pan- and core-genome of *Lactobacillus* provides novel evolutionary insights for differentiation. BMC Genomics 19:284. <https://doi.org/10.1186/s12864-018-4601-5>.
  35. Bazinet AL. 2017. Pan-genome and phylogeny of *Bacillus cereus sensu lato*. BMC Evol Biol 17:176. <https://doi.org/10.1186/s12862-017-1020-1>.
  36. Nourdin-Galindo G, Sánchez P, Molina CF, Espinoza-Rojas DA, Oliver C, Ruiz P, Vargas-Chacoff L, Cárcamo JG, Figueroa JE, Mancilla M, Maracaja-Coutinho V, Yañez AJ. 2017. Comparative pan-genome analysis of *Piscirickettsia salmonis* reveals genomic divergences within genogroups. Front Cell Infect Microbiol 7:459. <https://doi.org/10.3389/fcimb.2017.00459>.
  37. Collingro A, Tischler P, Weinmaier T, Penz T, Heinz E, Brunham RC, Read TD, Bavoi PM, Sachse K, Kahane S, Friedman MG, Rattei T, Myers GSA, Horn M. 2011. Unity in variety—the pan-genome of the Chlamydiae. Mol Biol Evol 28:3253–3270. <https://doi.org/10.1093/molbev/msr161>.
  38. Kim Y, Koh I, Young Lim M, Chung W-H, Rho M. 2017. Pan-genome analysis of *Bacillus* for microbiome profiling. Sci Rep 7:10984. <https://doi.org/10.1038/s41598-017-11385-9>.
  39. Hoton FM, Andrup L, Swiecicka I, Mahillon J. 2005. The cereulide genetic determinants of emetic *Bacillus cereus* are plasmid-borne. Microbiology (Reading) 151:2121–2124. <https://doi.org/10.1099/mic.0.28069-0>.
  40. Murawska E, Fiedoruk K, Swiecicka I. 2014. Modular genetic architecture of the toxigenic plasmid pIS56-63 harboring *cry1Ab21* in *Bacillus thuringiensis* subsp. *thuringiensis* strain IS5056. Pol J Microbiol 63:147–156. <https://doi.org/10.33073/pjm-2014-020>.
  41. Mahillon J, Chandler M. 1998. Insertion sequences. Microbiol Mol Biol Rev 62:725–774. <https://doi.org/10.1128/MMBR.62.3.725-774.1998>.
  42. Pinto D, da Fonseca RR. 2020. Evolution of the extracytoplasmic function  $\sigma$  factor protein family. NAR Genom Bioinform 2:lqz026. <https://doi.org/10.1093/nargab/lqz026>.
  43. Schmidt TR, Scott EJ, Dyer DW. 2011. Whole-genome phylogenies of the family Bacillaceae and expansion of the sigma factor gene family in the *Bacillus cereus* species-group. BMC Genomics 12:430. <https://doi.org/10.1186/1471-2164-12-430>.
  44. Van Schaik W, Tempelaars MH, Wouters JA, de Vos WM, Abee T. 2004. The alternative sigma factor  $\sigma^B$  of *Bacillus cereus*: response to stress and role in heat adaptation. J Bacteriol 186:316–325. <https://doi.org/10.1128/JB.186.2.316-325.2004>.
  45. Fayad N, Kambris Z, El Chamy L, Mahillon J, Kallassy Awad M. 2021. A novel antidipterian *Bacillus thuringiensis* strain: unusual Cry toxin genes in a high dynamic plasmid environment. Appl Environ Microbiol 87:e02294–20. <https://doi.org/10.1128/AEM.02294-20>.
  46. Carroll LM, Wiedmann M, Kovac J. 2020. Proposal of a taxonomic nomenclature for the *Bacillus cereus* group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes. mBio 11:e00034–20. <https://doi.org/10.1128/mBio.00034-20>.
  47. Yoshida K-i, Yamaguchi M, Morinaga T, Kinehara M, Ikeuchi M, Ashida H, Fujita Y. 2008. *myo*-Inositol catabolism in *Bacillus subtilis*. J Biol Chem 283:10415–10424. <https://doi.org/10.1074/jbc.M708043200>.
  48. Gonzalez-Uarquin F, Rodehutsord M, Huber K. 2020. *myo*-Inositol: its metabolism and potential implication for poultry nutrition – a review. Poult Sci 99:893–905. <https://doi.org/10.1016/j.psj.2019.10.014>.
  49. Xin B, Zheng J, Xu Z, Song X, Ruan L, Peng D, Sun M. 2015. The *Bacillus cereus* group is an excellent reservoir of novel lanthipeptides. Appl Environ Microbiol 81:1765–1774. <https://doi.org/10.1128/AEM.03758-14>.
  50. Wu J, Samara NL, Kuraoka I, Yang W. 2019. Evolution of inosine-specific endonuclease V from bacterial DNase to eukaryotic RNase. Mol Cell 76:44–56. <https://doi.org/10.1016/j.molcel.2019.06.046>.
  51. Caulier S, Nannan C, Gillis A, Licciardi F, Bragard C, Mahillon J. 2019. Overview of the antimicrobial compounds produced by members of the *Bacillus subtilis* group. Front Microbiol 10:302. <https://doi.org/10.3389/fmicb.2019.00302>.
  52. de la Fuente-Salcido N, Guadalupe Alanís-Guzmán M, Bideshi DK, Salcedo-Hernández R, Bautista-Justo M, Barboza-Corona JE. 2008. Enhanced synthesis and antimicrobial activities of bacteriocins produced by Mexican strains of *Bacillus thuringiensis*. Arch Microbiol 190:633–640. <https://doi.org/10.1007/s00203-008-0414-2>.
  53. Abriouel H, Franz CMAP, Omar NB, Gálvez A. 2011. Diversity and applications of *Bacillus* bacteriocins. FEMS Microbiol Rev 35:201–232. <https://doi.org/10.1111/j.1574-6976.2010.00244.x>.
  54. Raymond B, Wyres KL, Sheppard SK, Ellis RJ, Bonsall MB. 2010. Environmental factors determining the epidemiology and population genetic structure of the *Bacillus cereus* group in the field. PLoS Pathog 6:e1000905. <https://doi.org/10.1371/journal.ppat.1000905>.
  55. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
  56. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>.
  57. Kaas RS, Fries C, Ussery DW, Aarestrup FM. 2012. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. BMC Genomics 13:577. <https://doi.org/10.1186/1471-2164-13-577>.
  58. Minkin I, Patel A, Kolmogorov M, Vyahhi N, Pham S. 2013. Sibelia: a scalable and comprehensive synteny block generation tool for closely related microbial genomes. p 215–229. In Darling A, Stoye J (ed), Lecture Notes in Computer Science vol 8126. Springer, Berlin, Germany.
  59. Jesus TF, Ribeiro-Gonçalves B, Silva DN, Bortolaia V, Ramirez M, Carriço JA. 2019. Plasmid ATLAS: plasmid visual analytics and identification in high-throughput sequencing data. Nucleic Acids Res 47:D188–D194. <https://doi.org/10.1093/nar/gky1073>.

60. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–W21. <https://doi.org/10.1093/nar/gkw387>.
61. Xie Z, Tang H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33:3340–3347. <https://doi.org/10.1093/bioinformatics/btx433>.
62. Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Mol Biol Evol* 31:1077–1088. <https://doi.org/10.1093/molbev/msu088>.
63. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, Corander J, Bentley SD, Croucher NJ. 2019. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* 29:304–316. <https://doi.org/10.1101/gr.241455.118>.
64. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2 502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>.
65. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>.
66. Murawska E, Fiedoruk K, Bideshi DK, Swiecicka I. 2013. Complete genome sequence of *Bacillus thuringiensis* subsp. *thuringiensis* IS50056, an isolate highly toxic to *Trichoplusia ni*. *Genome Announc* 21:e0010813. <https://doi.org/10.1128/genomeA.00108-13>.