The Phylogeny of Bacillus cereus sensu lato

RICHARD T. OKINAKA and PAUL KEIM
Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, AZ 86011-4073

ABSTRACT The three main species of the Bacillus cereus sensu lato, B. cereus, B. thuringiensis, and B. anthracis, were recognized and established by the early 1900s because they each exhibited distinct phenotypic traits. B. thuringiensis isolates and their parasporal crystal proteins have long been established as a natural pesticide and insect pathogen. B. anthracis, the etiological agent for anthrax, was used by Robert Koch in the 19th century as a model to develop the germ theory of disease, and B. cereus, a common soil organism, is also an occasional opportunistic pathogen of humans. In addition to these three historical species designations, are three less-recognized and -understood species: B. mycoides, B. weihenstephanensis, and B. pseudomycoide. All of these “species” combined comprise the B. cereus sensu lato group. Despite these apparently clear phenotypic definitions, early molecular approaches to separate the first three by various DNA hybridization and 16S/23S ribosomal sequence analyses led to some “confusion” because there were limited differences to differentiate between these species (6). These and other results have led to frequent suggestions that a taxonomic change was warranted to reclassify this group to a single species (7, 8). But the pathogenic properties of B. anthracis and the biopesticide applications of B. thuringiensis appear to “have outweighed pure taxonomic considerations” and the separate species categories are still being maintained (9). B. cereus sensu lato represents a classic example of a now common bacterial species taxonomic quandary where relatively new molecular data must somehow be incorporated into a traditional hierarchical classification system (10).

Received: 17 October 2012, Accepted: 15 December 2015, Published: 12 February 2016
Editors: Patrick Eichenberger, New York University, New York, NY, and Adam Driks, Loyola University Medical Center, Maywood, IL
Correspondence: Richard Okinaka, Richard.Okinaka@NAU.EDU © 2016 American Society for Microbiology. All rights reserved.
AFLP APPROACH LEADS TO PHYLOGENETIC RESOLUTION OF B. ANTHRACIS AND B. CEREUS SENSU LATO

In the mid-1990s an amplified fragment length polymorphism (AFLP) method was developed to examine restriction fragment length polymorphisms (RFLPs) in whole genomes using restriction enzyme digestion coupled to PCR analysis and high-resolution electrophoresis (11). The advantage of AFLP was that the ends of the restricted fragments could be linked to specific primer sequences that then served as targets of PCR amplification for internal sequences that had not previously been described. This approach immediately offered greater genome coverage of RFLP sites and proved to be useful in resolving the monomorphic B. anthracis lineage and its close relatives, B. cereus, B. thuringiensis, and B. mycoides (12). This initial study uncovered 357 AFLP characters (polymorphic fragments) that could be used in cladistic and phenetic analyses to construct a phylogeny of the B. cereus group in addition to several distant relatives. In this instance, B. anthracis could clearly be distinguished from its two closest relatives, B. cereus/B. thuringiensis, and another member of this group, B. mycoides. This initial AFLP analysis not only provided the first evidence of significant diversity and a DNA-based phylogeny of 78 B. anthracis isolates, but it also provided an experimental approach to examine the phylogenetic relationship between diverse isolates contained in large B. cereus/B. thuringiensis collections (13, 14).

In the latter analysis, fluorescent AFLP analysis was performed on the DNA from an extensive collection of 332 diverse B. cereus, B. thuringiensis, and B. anthracis isolates (14). This analysis included 34 diverse Norwegian soil isolates (13, 15); 222 B. thuringiensis isolates representing 36 different serovars from a U.S. Department of Agriculture collection; 24 diverse B. anthracis isolates; 42 B. cereus isolates recovered either from contaminated food products or from clinical samples by the Food Research Institute, University of Wisconsin; and finally, 8 B. thuringiensis/B. cereus isolates from the American Type Culture Collection. These latter samples included several type strains that were useful in the comparison of these results with nearly simultaneous studies that were using multiple locus sequence typing (MLST) analysis to examine similar collections of isolates (15, 16).

A diagrammatic representation of a basal tree from this AFLP analysis is illustrated in Fig. 1. The key features of this tree illustrate that the 332 isolates are dispersed into 3 major clusters and into 10 distinct branches (labeled A to K) within the 3 clusters. More importantly, each of the three clusters contains representatives of both B. cereus and B. thuringiensis isolates. These results portray an important and consistent general theme, i.e., the increased overall resolution offered by AFLP does not separate B. cereus and B. thuringiensis isolates into distinct clusters. Instead representatives of both species are found scattered throughout the three main clusters defined by the AFLP analysis. These findings have major implications for the evolution of all the subgroups under the nomenclature umbrella for B. cereus sensu lato. It supports the notion that a B. cereus sensu lato genomic background has evolved in a primarily clonal fashion to form the major phylogenetic branches of this group. However, the species designations for this group are based primarily on the horizontal gene transfer of various plasmids, genomic islands, etc., that contained specialized factors that helped to define specific phenotypes that became the hallmark features of, e.g., B. anthracis and B. thuringiensis (8).

Figure 1 is a representation of the entire AFLP tree without illustrating significant branch resolution, but highly resolved individual branches (Fig. 3 to 8, Hill et al. [14]) within this tree reveal a potentially wider spectrum of evolutionary developments. Branches A and C, for example, consist mostly of B. thuringiensis isolates with only a limited number of related serovars, suggesting that a clonal expansion and fitness of specific B. thuringiensis isolates may have founded these clades. This is also reflected in the B. anthracis lineage, which is a monophyletic, clonal expansion of a relatively young branch that has sparse diversity and exchange within all the known B. anthracis isolates (12, 17). Conversely, other branches in the AFLP tree, e.g., clusters 2 and 3, appear to have several branches where a specific B. cereus sensu lato lineage may have given rise to both B. cereus and B. thuringiensis isolates as a result of horizontal transfer events that occurred later in the evolutionary time scale. A similar phenomenon appears to have occurred more recently in the AFLP F branch, where B. cereus isolates, not in the monophyletic clade that is B. anthracis, have acquired a pXO1 plasmid (B. cereus G9241 [18]) or both the pXO1 and pXO2 plasmids (B. cereus CI, biovar anthracis [19]). Various typing schemes indicate that these two isolates (B. cereus G9241 and B. cereus CI) are close relatives of the B. anthracis lineage and that both would belong to AFLP group 1 (Fig. 1).

MLST

MLST (http://pubmlst.org) was originally designed as a molecular typing method that could take advantage of
the portability and exchange of sequence data between laboratories (20). The basic idea was to identify seven or more “housekeeping genes” from closely related populations such as the B. cereus sensu lato group and to generate ∼500-bp sequences from each of these seven genes in each isolate of interest. The ∼3,500 bp of sequence for each isolate then defined its sequence type (ST) when the sequence profiles were established by comparative analysis to all other isolates in the database. The MLST database for B. cereus currently (as of

![AFLP-based phylogenetic tree of B. cereus sensu lato](https://doi.org/10.1128/microbiolspec.TBS-0012-2012.f1)
Three independent *B. cereus sensu lato* MLST studies were published in 2004 (15, 16, 21) and four other reports have appeared since (22–25). The singular consistent observation between these data sets and four independent AFLP reports (13, 14, 26, 27) is that there are three main clusters that appear to be conserved in all the studies (25). Taken together, these findings support the widely held view that the *B. cereus sensu lato* group, in general, has a basic clonal population structure (15, 16).

It should be reiterated that these conclusions are based on both MLST data, which target “housekeeping” genes that are presumed to be more conserved, and AFLP data, which are often driven by small repeated elements (i.e., variable number tandem repeats [17]) that can mutate and evolve more rapidly than single-nucleotide polymorphisms (SNPs).

The most recent study to date (27, 28) has a composite analysis of 2,213 isolates in the *B. cereus sensu lato* group assembled by various combinations of MLST, AFLP, and multienzyme electrophoresis (MEE [29]). “Altogether, the global analysis confirms and extends the results underlying the opportunistic nature of *B. cereus* group organisms, and the fact that isolates responsible for disease outbreaks and contamination of foodstuffs can originate from various genetic backgrounds” (27). Again, the data sets illustrate that *B. cereus*, *B. thuringiensis*, and *B. mycoides* isolates are dispersed throughout the basal *B. cereus sensu lato* phylogenetic tree. This does not, however, preclude the existence of clonal expansions where specific lineages may have become significant factors as clinical or insecticidal or environmental clusters. Examples include the highly conserved *B. anthracis* lineage, clonal clinical complexes causing periodontal or emetic disease in humans (23), and clonal complexes that might be specific for certain *B. thuringiensis* isolates and insecticidal toxins (e.g., branch C in reference 14).

**HOMOLOGOUS RECOMBINATION**

Asexual reproduction in bacteria defines clonal population structure where all the progeny are derived from a single parental cell. However, recombination within these populations can scramble and redistribute the DNA polymorphisms and cause a decomposition of this clonal pattern after many generations. The early MLST studies indicated that there was a strong clonal structure to the *B. cereus sensu lato* populations defined by specific lineages that contained fixed SNPs that appeared to be unique to each lineage. These same analyses indicated that the distribution of a limited number of SNPs was not in congruence with a strict clonal structure, i.e., there were a limited number of SNPs that were not fixed to a specific branch but were shared between different lineages (16). This would be evidence for genetic exchange by homologous or non-homologous exchange or recombinination between different lineages of the *B. cereus* subgroup.

From a historical perspective, a large body of evidence has helped to define homologous recombination and DNA excision repair as part of the general strategy that cells use to repair the two main forms of damage to DNA: double-strand breaks and nucleotide damage, respectively (30). The homologous recombination repair system is a *RecA* protein-driven process where the damaged DNA (a double-strand break) initiates a cascade of events including the eventual alignment and repair synthesis of a damaged region by using an undamaged template from a sister chromatid (30). This is an important “housekeeping” function because most bacteria, e.g., *Escherichia coli* and *Bacillus subtilis*, cannot tolerate a single unrepaired double-strand break in its chromosome. In humans, the disease ataxia telangiectasia, a deficiency in double-strand break repair, is short-lived, and patients with this disease are susceptible to many ionizing radiation-induced maladies (30).

While the “housekeeping” chores of the homologous exchange system to repair damaged DNA go unnoticed in large populations of bacteria, ancillary processing by these same proteins can also cause rare but effective homologous exchange between closely related bacteria. These events are rare because they first have to involve a form of “sexual” exchange of DNA between related bacteria via a number of processes, e.g., conjugation/ transformation, transduction, etc. In addition, this transferred DNA must either be incorporated into the new genome as “foreign DNA” via insertion sequences, “genomic islands,” etc., or as in homologous recombination as donor DNA that can recombine with a homologous template in the recipient cell. If the donor DNA is sufficiently different from the homologous stretch in a recipient cell, i.e., it contains several SNPs that are “foreign” to the recipient DNA, then this newly incorporated stretch of DNA can be detected by programs like eBurst (31–33) and ClonalFrame (34). This is the case for the *B. cereus sensu lato* group.

The housekeeping genes used in MLST were designed to be separated by sufficient distances to lessen the probability that more than one of the seven selected loci would be involved in any single recombination event.
(20), and this typing scheme also led to the use of allele designations for each unique sequence at each locus. For example, in the B. cereus MLST site (http://pubmlst.org/bcereus/) there are 240 different sequence alleles in the gfp fragment (the first 7 MLST sequences) that were discovered in 1,518 isolates. These allele designations led to the development of methods to establish genetic relationships utilizing cluster analysis of the seven alleles in each sequence type, e.g., eBurst (31–33).

A more recent inference model (ClonalFrame; git clone https://github.com/xavierdidelot/ClonalFrameML) indicates that MLST data can be used to determine the clonal relationship of bacteria while also providing the chromosomal position of homologous recombination that can potentially disrupt clonal patterns (34, 35). Unlike eBurst, ClonalFrame does not treat every allele designation with equal weight, and it recognizes that recombination events occur at a constant rate of substitutions to a contiguous region of sequence. When ClonalFrame encounters two strains with six of seven alleles having only one or two differences between each of these alleles and then a seventh allele having many nucleotide differences, it does not dismiss these two strains as being unrelated. Rather, it suggests that they have a clonal relationship with a homologous recombination in the seventh allele originating from an outside source.

ClonalFrame was used to reconstruct the evolutionary history of 667 strains in the B. cereus sensu lato group from MLST data (9). This analysis again confirmed the presence of three major clonal clusters and also demonstrated a variety of genetic exchanges between and within these clusters including a high number of exchanges with sources external to any of the clades. As an example, Didelot et al. (9) describe a subset of 35 recombination events that were inferred by ClonalFrame involving the specific cluster 2 (16). To illustrate and simplify how ClonalFrame was able identify and define these events, we have examined a single clade in Priest’s original cluster 2 in detail. A maximum likelihood MLST phylogenetic tree for a clade defined as Sotto is illustrated in Fig. 2A, and it shows the relationship between seven sequence types (STs) in this clade. Note that ST-49, ST-55, and ST-9 are positioned at points that are distal to the remaining STs that otherwise form a tight cluster.

ClonalFrame identified a single gene sequence (ilv) to be responsible for creating the distal relationship between ST-49 and ST-55 from the cluster containing ST-12, ST-16, ST-23, and ST-56. Figure 2B illustrates a MLST maximum parsimony tree of six of the MLST genes (minus the ilv allele) in two subclades (Priest’s Tolworthi and Sotto). Notice how the removal of the ilv sequence has caused all of the ST types to form a tight Sotto cluster. In Fig. 2C, a maximum parsimony tree of the ilv locus by itself illustrates that the ilv allele in ST-55 is identical to isolates from a subclade, Tolworthi, which is located on a different branch on the same cluster 2. This is indicative of a homologous exchange between an isolate from the Sotto clade with an isolate from the Tolworthi clade in the region containing the ilv locus. The ilv allele for ST-49, on the other hand, was identical to the ilv allele for ST-13, a sequence type that is found in another cluster 2 subclade named Kurstaki (16).

This single subclade (Sotto) contains six sequence types (excluding ST-9), and, in 42 isolates, this group has two presumed examples of homologous exchange between (i) isolates from the Tolworthi and Sotto subclades and (ii) isolates from the Kurstaki and Sotto subclades that involve the regions surrounding the ilv locus. This example exemplifies the overall state of recombination in this subgroup of bacilli. ClonalFrame has demonstrated that, despite the use of only seven MLST fragments per genome, covering ~0.05% of the average B. cereus sensu lato genome, there is a considerable amount of recombination in the B. cereus sensu lato group. Didelot et al. (9) estimated that the B. cereus group is significantly less clonal than its close relative, Staphylococcus aureus, as measured by the relative impact of recombination and mutation (rfm) values ranging between 0.69 and 2.90 in B. cereus versus a value of 0.1 for S. aureus.

Two other significant conclusions included the observation that the pathogenic B. cereus strains are distributed throughout the first two clades of the phylogenetic tree and that increased or decreased rates of homologous recombination were not apparent within the pathogenic lineages of B. cereus. These results are consistent with the notion that B. cereus sensu lato is an “opportunistic” pathogen without specific predispositions that appear to be associated with pathogenicity.

THE PANGEOME OF B. CEREUS SENSU LATO

It has become evident that genetic content in individual isolates from any given bacterial species can vary considerably (36–39). The advent of next-generation sequencing technologies has caused an escalation in attempts to define the pangenome, or the whole genome complement or gene variation in a single clade or species. This was accomplished by the process of multiple genome comparisons of isolates within a given species
The initial concept of the pangenome was defined by two basic components: a “core” genome that consisted of genes shared by all strains in a species and a “dispensable” genome that consisted of genes that are found in some but not all of the strains. The core genes would conceptually provide the functions that define the basic biology of the species and the dispensable genes would define the diversity and impact selective advantages such as adaptation and antibiotic resistance.

In B. cereus sensu lato it is easy to visualize a “core” genome as consisting of a conserved framework that maintains functions that would allow an “opportunistic” or “fit” isolate to adapt to a new environment that is suitable for a lifestyle as a common soil organism or as a pathogen of mammals, humans, and insects. It is also appropriate to think of the “dispensable” genome as consisting, in part, of the plethora of small and large plasmids that are uniquely associated with the phenotypes that define...
each of these “species.” What is not obvious is the role of the core genome in the overall diversity of these species.

These basic concepts have recently been examined within the B. cereus sensu lato subgroup by analyses of 58 diverse genomes (42). This study included the generation of high-redundancy whole-genome sequences by 454 pyrosequencing (43) of 45 B. cereus sensu lato strains containing an array of isolates based on geographical, phenotypic, and phylogenetic diversity. These data sets were combined with 13 previously sequenced genomes to establish the pangenome for B. cereus sensu lato. This included only a single B. anthracis genome, the Ames ancestor (44), to avoid overrepresenting this highly conserved and closed pangenome (38, 41).

Zwick et al. (42) defined the pangenome of B. cereus sensu lato using the expanded terminology of Lapierre and Gogarten (45). They found a typical bimodal distribution for 22,975 gene clusters in the 58 genomes (Fig. 3). They also defined gene families found in six or fewer genomes as “accessory” genes, gene families found in 49 or more genomes as “extended core,” and those in between as “character genes.” The core genome (genes found in every genome) consisted of 1,754 genes and the extended core took this number for B. cereus up to 3,904. As mentioned earlier, the overall analysis of the whole genome (accessory, character, and extended core genes) of these 58 diverse B. cereus sensu lato isolates defines an “open” pangenome for this group of closely related microorganisms (41, 61). This analysis included the B. anthracis lineage that represents a classic example of a “closed” genome and may eventually describe several other lineages that might be deemed as closed genomes as part of specific clonal human or insect pathogenic lineages (27).

A phylogenetic tree constructed by using a distance-based approach and concatenated chromosomal core proteins for the 58 genomes is illustrated in Fig. 4. This whole-genome phylogeny agreed with previous MLST and AFLP studies indicating that the B. cereus sensu lato group is separated into three major clusters or clades (14–16). This core gene data analysis also was consistent with the notion that, individually, the isolates defined as B. cereus, B. thuringiensis, and B. mycoides are not confined to discrete clades. To avoid further confusion, the authors chose the designations used by Priest et al. (16) to indicate that clade 1 contains the well-defined B. anthracis lineage and clade 2 has a large presence of B. thuringiensis isolates, and hence the subclades named after serotypes Tolworthi, Kurstaki, Sotto, and Thurini- giensis. The article by Tourasse et al. (27) provides a comprehensive analysis to establish the relationships between the clade and cluster designations in AFLP, MLST, and MEE in the various studies on the B. cereus sensu lato group.

The analysis of the whole-genome sequences of 58 diverse B. cereus sensu lato isolates defines an “open” pangenome for this group of closely related microorganisms. The data garnered from the “core genomes” of these isolates confirm that the group has a clonal phylogenetic structure and that isolates designated as either B. thuringiensis or B. cereus are scattered throughout the “core genome” tree. These core data represent unparalleled DNA signatures, the whole-genome sequences from 58 diverse B. cereus sensu lato isolates, and they support the idea that this group consists of a relatively conserved genomic background that could be treated as a single phylogenetic entity.

**FIGURE 3** A graph of the distribution of gene families across B. cereus sensu lato genomes redrawn from Zwick et al. (41). This figure is based on the definition of the extended core as genes encoding proteins present in 49 or more genomes and accessory genes as those present in <6 genomes. The class between these extremes defined the character gene set. doi:10.1128/microbiolspec.TBS-0012-2012.f3
FIGURE 4 Whole-genome phylogeny of *B. cereus* sensu lato. This tree was redrawn based on data sets of concatenated, conserved protein sequences by using a neighbor-joining algorithm (41). Note that the relative distribution of the isolates based on a conserved whole-genome phylogeny is essentially the same as those observed in numerous MLST and AFLP studies and separated into three major clades. doi:10.1128/microbiolspec.TBS-0012-2012.f4
Clade 3 was in a group that Priest et al. (16) called “others” and included a B. mycoides isolate and two B. weihenstephanensis isolates. The “core gene” whole-genome phylogeny expanded this clade to 15 isolates and demonstrated that clade 3 is in reality a polyphyletic grouping containing several new clades. Recent descriptions of psychrotolerant strains appear to fall into categories that include additional B. mycoides and B. weihenstephanensis strains and analysis of a selection of these genomes may help to define these new clades in line with this pangenome analysis (26–28).

The program ClonalFrame was used to revisit the patterns of homologous recombination within the 7.4 Mb core genomes in the 58 B. cereus sensu lato genomes. The ratio for the overall effects of homologous recombination and mutation (r/m) was 2.91, which is somewhat higher than the results obtained from ClonalFrame analysis of the 7 MLST fragments of 667 B. cereus sensu lato isolates (9). Nevertheless, these results are consistent with the idea that the r/m values for this group of bacteria are intermediate in relationship to a highly recombinating population such as Helicobacter pylori (r/m = 13.6 [46]) and Burkholderia species (r/m ~ 2.5 [47]) versus a highly clonal population such as S. aureus (r/m = 0.1 [9]).

The analysis of the pangenome of B. cereus sensu lato confirms the relatively conserved clonal structure and a potential role for homologous recombination in the evolution of this group of bacilli. These are generalizations that had already been indicated since the “confusing” phylogenetic mixtures of B. cereus, B. thuringiensis, and B. mycoides were first acknowledged by DNA hybridization and 16S RNA studies (6, 48, 49) and then again supported by more recent AFLP, MEE, and MLST investigations (13–16, 21–27). However, Zwick et al. (42) have also used pangenome analyses to generate additional “global” insights concerning the evolution of this group of “opportunist” pathogens. Pangenomes contain gene inventories that allow us to determine precise differences between pathogenic and nonpathogenic strains of a single species (39, 50). In many instances, these differences are measured by the loss and/or gain of sharply defined cassettes of genes. Two prime examples in the B. cereus sensu lato are (i) B. anthracis, which acquired the two plasmids (pX01 and pX02) containing the tripartite toxin gene complex and a capsule synthesis gene set, respectively, and (ii) B. thuringiensis, whose properties are mainly attributed to plasmids containing specific insecticidal toxin genes. From a more global perspective, gene acquisition and gene loss are now perceived as major underlying factors in the emergence and evolution of bacterial pathogens (51).

Several early whole-genome analyses of different microbes have revealed that bacteria have tendencies toward trimming and streamlining their genomes on the basis of functional needs (51–54). This “decay” of genomes appears to be fostered by selective pressures dictated along evolutionary timelines. Ochman and Davalos (51) have illustrated these evolutionary forces by the comparison of the relative genome sizes of free-living species versus the genome sizes of related facultative and obligate pathogens. These analyses suggest that free-living bacteria maintain relatively large genome sizes (5 to 10 Mb), because most of their genome functions are required for survival in diverse environments. Conversely, related facultative and obligate pathogens rely on host functions and nutrients and are maintained in smaller population sizes that eventually lead to accumulation of nonsynonymous mutations in extraneous genes and the eventual loss of unnecessary elements within these genomes. The net effect is that recent and facultative pathogens have genomes of intermediate size (2 to 5 Mb), while obligate symbiotic or pathogenic bacteria have relatively small genomes (0.5 to 1.5 Mb).

By these criteria, the B. cereus sensu lato subgroup, in general, classified as a soil organism, has a number of recently emerged pathogens of humans, insects, and other species. Zwick et al. indicate that there were very few character genes that could be used to describe any one clade within the B. cereus sensu lato. In addition, there was very little evidence that core genes were specifically being deleted from any single clade. A very specific analysis seeking evidence for gene cassettes directed toward a predisposition and adaptation toward pathogenicity included the comparison of the genomes in the clade that included three isolates that contained the pX01 plasmid: B. anthracis, B. cereus G2941, and B. cereus CI (18, 19). These whole-genome comparisons to other isolates in this clade did not reveal any genes that were unique to the three isolates except for those acquired from the pX01 and/or pX02 plasmids. These results support the notion that these three genomes are not uniquely adapted to a lifestyle that would include the invasion and a natural fitness for growth in mammals.

Although not directly involved with phylogenetic issues, there is one emerging concept that does impact how particular isolates might readily adapt to different niches. This idea indicates that specific phenotypic differences can be caused by altered gene expression for a variety of trans-acting factors (55), rather that the loss
or gain of specific coding regions. Prime examples are the effects attributed to the plcR and SigP-RsiP regulons in *B. anthracis* (56, 57). PlcR is a regulon whose expression regulates multiple genes related to virulence in *B. cereus* but is not expressed in *B. anthracis*. Similarly, mutations in the SigP-RsiP sigma factor regulon have recently been shown to affect the expression of the beta-lactamase genes that affect the sensitivity of *B. anthracis* to penicillin (57).

Consistent with this trend, Zwick et al. (42) did not detect significant evidence for excessive gene loss and the accumulation of nonsynonymous mutations in clades 1 and 2, which house the main human and insect pathogenic strains. These results suggest the absence of the early signs for genome decay that is frequently associated with facultative pathogens and includes phenomena such as the accumulation of pseudo-genes in superficial genomic elements (51). *B. cereus* is often described as an “opportunist” pathogen and, while this designation is usually reserved for infections involving nonpathogenic organisms becoming pathogenic in compromised individuals, this terminology appears appropriate when considered alongside the analysis of the pan-genome of this subgroup. Zwick et al. (42) chose to describe the clade 1 and 2 isolates as “hopeful monsters” (58, 59) because these typically soil isolates appear to contain genomic complements that would allow them to exist as human or insect pathogens “in waiting” – waiting for the appropriate “toxin”-containing plasmid and/or an environmental niche that might allow a specific lineage to rise to prominence.

Curiously, genome “decay” is apparent in clade 3 of *B. cereus sensu lato* as measured by gene loss, accumulation of nonsynonymous mutations, and higher rates of homologous recombination relative to mutation. This clade was not well represented in earlier studies, and more recent data suggest that this group may contain several clades with greater phylogenetic and phenotypic diversity than observed with clades 1 and 2 (26–28, 42). An unusual isolate, NVH391-98, a severe foodborne pathogen whose genome is severely reduced in size (5.2 to 5.5 Mb in *B. cereus* versus 4.0 Mb in NVH391-98), has been used as an outgroup to root a *B. cereus sensu lato* 16S RNA tree (60) and a whole-genome core tree (42). The NVH391-98 genome does have collinear aspects with respect to the *B. cereus sensu lato* genome but this distant relative, while suitable as an outgroup, is currently being proposed as a separate species, *B. cytotoxins*. Nevertheless, this rooting does point to the diverse clade 3 (potentially clades 4, 5, 6, and 7) as an ancestral branch of the *B. cereus sensu lato* and suggests that NVH391-98 may share a common ancestor with certain isolates in clade 3 that are also exhibiting signs of genome decay in their core genome (e.g., see top of Fig. 1B, Zwick et al. [42]).

**CONCLUSIONS**

Despite evidence for homologous recombination, the phylogeny of *B. cereus sensu lato* as a whole remains one of clonal expansion. This has held true from conclusions based on earliest 16S RNA analysis through a more recent history that includes MEE, AFLP, MLST, and whole-genome sequence analyses. More importantly, the central debate about nomenclature and the *B. cereus sensu lato* group remains unchanged. These numerous studies have solidified the idea that, while the phenotype of each individual isolate may be determined by a variety of physiological tests and marker analyses, the phylogeny of the species as a whole has a conserved, clonal structure. This produces a structure that has mixtures of *B. cereus*, *B. thuringiensis*, *B. mycoides*, and others, scattered throughout the phylogenetic tree. Recent pan-genome analyses have revealed that two of the first three “clades” of *B. cereus sensu lato* have pathogenic lineages that have NOT begun to show signs of genome “decay” that is more “typical” of facultative and obligate pathogenic lineages in other species. These findings suggest that potentially “facultative pathogens” and “free-living” *B. cereus* isolates have similar, relatively unaltered, basal genomic structure and size; and this, in turn, suggests that the movement of different kinds of pathogenic elements remains an ongoing phenomenon in this species. These notions support the idea that the *B. cereus sensu lato* group is in part “hopeful monsters” that can be transformed into new pathogenic lineages under the right set of circumstances. A potential example of this phenomenon, in progress, may be *B. cereus* isolates from the *B. anthracis*-containing clade 1 that have acquired pXO1 (BcG9241) and/or pXO2 (*B. cereus* or BA CI) and may be new, emerging lineages that can cause anthrax-like diseases.

**REFERENCES**


