Spring 4-11-2000

Molecular Evolution of Insecticidal Spore-Forming Bacteria

John Pool
Western Michigan University, jpool@wisc.edu

THE CARL AND WINIFRED LEE HONORS COLLEGE

CERTIFICATE OF ORAL EXAMINATION

John Pool, having been admitted to the Carl and Winifred Lee Honors College in 1996 successfully presented the Lee Honors College Thesis on April 11, 2000.

The title of the paper is:

Molecular Evolution of Spore-forming Bacteria

Dr. DeWayne Shoemaker
Biological Sciences

Dr. John Geiser
Biological Sciences

Molecular Evolution of Insecticidal Spore-Forming Bacteria

Honors College Thesis

Student: John Pool

Lee Honors College

Western Michigan University

Thesis Committee:

Professor DeWayne Shoemaker (chair)

Professor John Geiser

April 11, 2000

Abstract

Molecular methods are increasingly being used to determine the phylogeny of microorganisms. This research was intended to determine phylogenetic relationships for bacteria of the species *Bacillus thuringiensis* and other members of the *Bacillus cereus* group. Each strain was analyzed by its *sasp-B* gene sequence to determine its species classification and relation to other strains studied. Results of this study indicate that according to the *sasp-B* gene tree, the species *Bacillus thuringiensis* is paraphyletic with respect to both *Bacillus cereus* and *Bacillus anthracis*. Some unexpected results and implications for species designations are also discussed.

Introduction and Background

Bacillus thuringiensis is a commercially important species of bacteria, most notably due to its insecticidal properties. *Bacillus thuringiensis* produces a toxin which is known to kill mosquitos, flies, beetles, and worms. *Bt* toxin has long been a useful insect control agent for organic farmers (Woodfin, 1997). More recently, crop plants such as corn and cotton have been genetically engineered to express *Bt* toxin (Wadman, 1997).

Recently, there has been considerable interest in the reported ability of one *Bacillus* strain to break down the explosives TNT (tri-nitro toluene) and GTN (glycerol tri-nitrate). This type of process, whereby a biological organism eliminates an environmental contaminant, is known as bioremediation. Thus, bacterial strains which effectively break down these explosives could be helpful

in decontaminating sites such as explosive manufacturing facilities and military target ranges (Leighton, personal communication).

Physiological traits were found to be uninformative with regard to determining whether the explosive-degrading strain belonged to *Bacillus cereus* or *Bacillus thuringiensis*. This is a common situation, as these two species are thought to be closely related (Leighton, personal communication).

It would be useful to know how the different *Bacillus* strains are related to each other, in an evolutionary sense, to predict which strains are likely to be useful for specific bioremediation or agricultural applications. Biologists have traditionally relied on observable phenotypic differences such as visual appearance under a microscope or antigens possessed to acquire such knowledge. However, these techniques may not resolve the true phylogenetic relationships between the strains being studied, since they do not address the genetic basis of evolution.

However, molecular methods, such as examination of the DNA sequence for a specific gene, can often provide this kind of phylogenetic information. Gene sequences from many different individuals are obtained, and from the differences in nucleotide sequence a phylogenetic tree can be constructed. This strategy is advantageous because it examines the genetic basis of evolution and often delivers clear, precise results to questions of phylogenetic relationships and species identification.

Scientists often analyze the nucleotide sequence of a housekeeping gene, such as *16S rRNA*, to examine the phylogeny of distantly related organisms. These genes are responsible for basic cell functions and are very highly conserved from

one species to another, even allowing for the relative classification of organisms of different kingdoms. However, it would be difficult to find any variability in a genes sequence like *16S rRNA* between closely related species or strains.

This problem can be addressed by examining a gene that has a much faster evolutionary clock than a housekeeping gene like *16S rRNA*. The *sasp-B* gene is present throughout the genus *Bacillus*, shows DNA sequence variation between closely related species (Sun and Setlow, 1987) and even between strains within some species, like *Bacillus thuringiensis*. The *sasp-B* gene was selected for use in this project for these reasons.

Sasp codes for a small, acid-soluble, spore protein. Each *Bacillus* strain has several genes that code for these proteins (Setlow, 1988). In this experiment the B-type gene is studied. Bacteria of the genus *Bacillus* are known for their ability to form spores. In this process, the DNA and other essentials are packaged into the spore, which forms inside the cell and develops a tough exterior. Spore formation allows these bacteria to survive environmental threats, emerging to form new vegetative cells when conditions improve.

Many *sasp* proteins probably play a role in protecting the DNA against damage from ultraviolet radiation during the spore stage, but *sasp-B* is known only to provide free amino acids for the new vegetative cell as the gene product is broken down by proteases (Setlow, 1988). It may be this lack of a more specific function, and hence low selective constraint, that allows for so much variation in all sections of the *sasp-B* gene except for the two protease binding sites (Sun and Setlow, 1987).

In summary, this experiment seeks to determine the phylogenetic relationships of members of *Bacillus thuringiensis* and other species of the *Bacillus cereus* group. The results of this experiment also address other questions. For example, are the species classifications *Bacillus thuringiensis* and *Bacillus cereus* phylogenetically justifiable groupings? And, which species did the GTN degrading strain belong to?

Experimental Method:

All strains examined were obtained from the Bacillus Stock Center or the American Tissue Culture Collection, and all were designated as *Bascillus thuringiensis* by these sources. Specific strain designations can not be offered until this data is published. Each strain was innoculated onto a nutrient medium and allowed to grow for at least 24 hours.

Genomic DNA Isolation

Genomic DNA was isolated from each strain studied. This involved the following procedure (El-Helow, unpublished protocol):

1. Transfer an isolated colony from an overnight TBAB plate to 10mL liquid medium such as LB or SPIM in a sterile 125mL sidearm flask.
2. Put in a shaker at 37°C until Klett reading is around 200 (will take a few hours).
3. Transfer 1.5-2.0mL into an eppendorf tube.
4. Centrifuge for 2min at maximum speed. Discard the supernatant. Store at -20°C until ready to proceed.

5. Resuspend the cell pellet in 563μL of 1x Tris-EDTA (TE) buffer, 3μL of 100mg/mL lysozyme, 3μL of 20mg/mL proteinase K, and 1μL of 100mg/mL RNase A. Vortex until solution is homogenous.
6. Incubate at 37°C for one hour, which should lead to a reduction in turbidity. If not, incubate for additional time.
7. Add 30μL of 10% SDS detergent.
8. Incubate at 60-65°C for 30 minutes.
9. Add 100μL of 5M NaCl. Invert slowly several times to mix.
10. Centrifuge for 2 minutes at maximum speed. Transfer a clear volume of supernatant to a new tube.
11. Add 60% volume isopropanol.

12a. If DNA strands are visible, transfer them, using a Pasteur pipette melted into a hook at the end, to another tube containing 50μL of 1x TE buffer. Otherwise:

12b. If no DNA is visible, centrifuge for 10 minutes at maximum speed. Rinse with 95% ethanol and repeat centrifugation. Set upside down on a paper towel for five minutes to dry. Add 30μL of 1x TE buffer to dissolve.

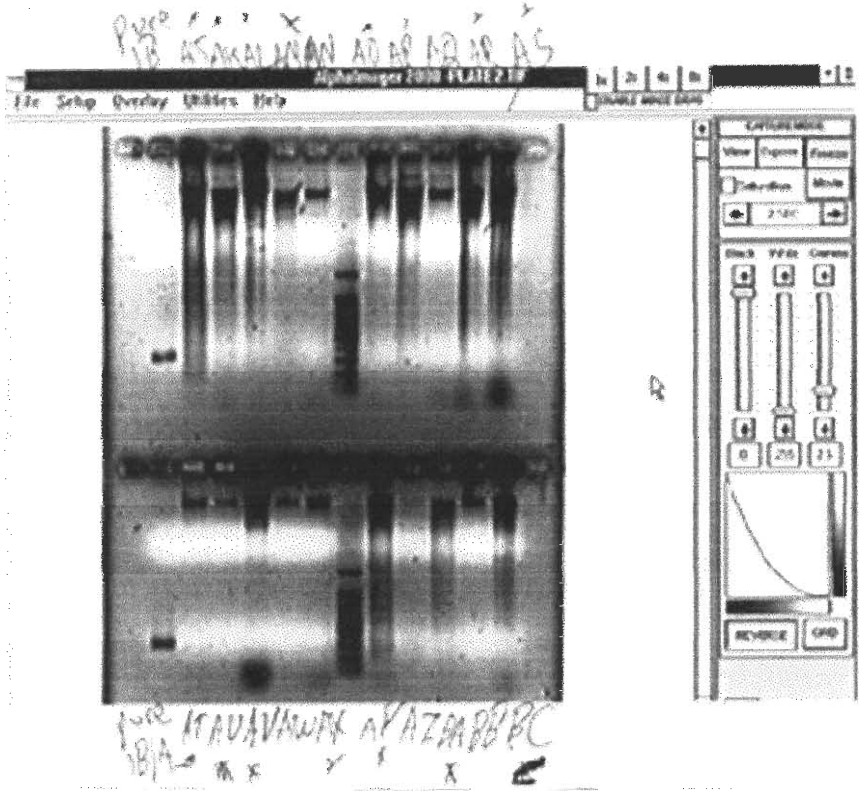
13. Store at -20°C.

Electrophoresis and Visualization

The presence and approximate concentration of genomic DNA was then determined by gel electrophoresis. A 1.5% agarose gel was used in this procedure. The gel was melted in a microwave and poured into a gel try with a comb to solidify. A 2μL sample from each strain's isolated genomic DNA was added to loading buffer and inserted into a well in the gel. A current of 90 volts was then applied to the gel for 45 minutes. Next, the gel was immersed in an

ethidium bromide solution for 30 minutes. Then, the gel was viewed under a UV imager, and a picture was produced and printed out. Darker bands indicated that a higher yield of genomic DNA was obtained.

Figure 1. Image from gel electrophoresis following DNA isolation



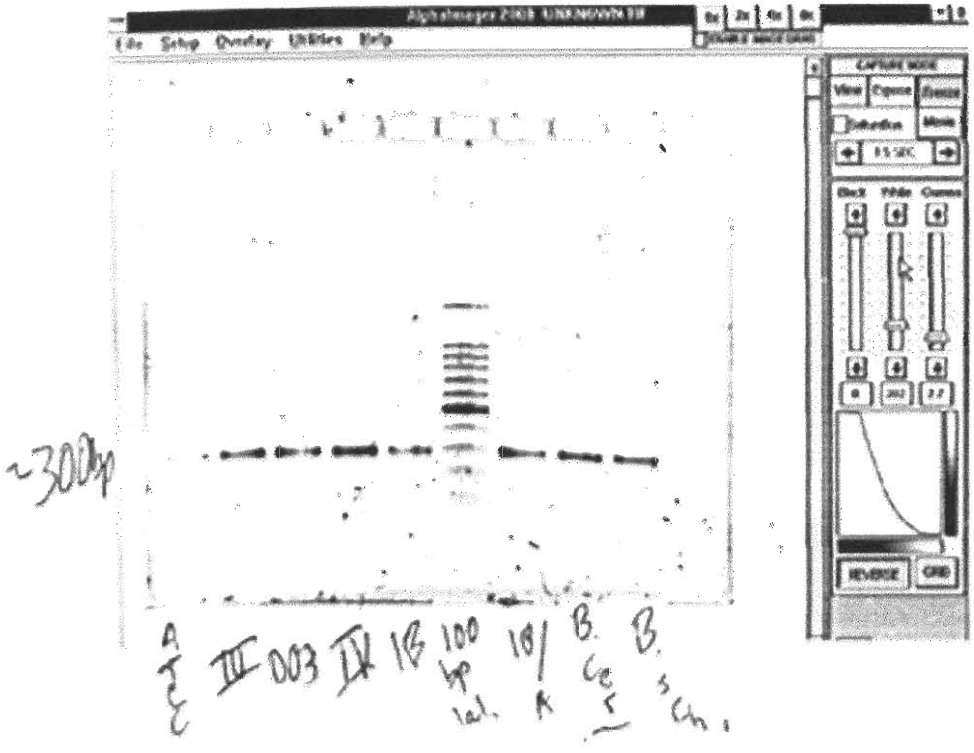
Polymerase Chain Reaction

Next, a strain's isolated genomic DNA was used in a PCR (Polymerase Chain Reaction) procedure to amplify the *sasp-B* gene. PCR involved a cycle of three steps repeated 35 times, resulting in millions of copies of the *sasp-B* DNA sequence. The first stage, 60 seconds at 94°C, denatured the double stranded DNA. The second stage, 60 seconds at 55°C, allowed primers to anneal. During third stage, 60 seconds at 72°C, the new DNA sequence could be extended from

the primer site. The procedure also included a final 7 minute extension phase at 72°C. PCR reaction mix included 36μL deionized water, 5μL TRIS-KCl buffer, 2.8μL magnesium chloride, 3μL deoxynucleotides (A, C, G, & T), 0.2μL "*Taq*" DNA polymerase, and 1μL each of forward and reverse primers, in addition to 1μL of genomic DNA.

For each strain, an appropriate pair of primers had to be found for PCR. The primers, which work outside of the *sasp-B* coding region on each side, were designed for the non-coding region, which often differs considerably between strains. Trial and error was therefore required to find a pair of forward primer and reverse primer that produced significant amplification of the *sasp-B* gene for any given strain. Presence of PCR product was verified by gel electrophoresis, using the same method as described above.

Figure 2. Image from gel electrophoresis following PCR



Following a successful PCR reaction, PCR products were purified using the QIAquick PCR Purification Kit from Qiagen, according to the following protocol:

1. Add 5 volumes of Buffer PB to 1 volume of the PCR reaction and mix.
2. Place a QIAquick spin column in a provided 2mL collection tube.
3. To bind DNA, apply the sample to the QIAquick column and centrifuge for 30 - 60 seconds at 5000 rpm.
4. Discard flow-through. Place QIAquick column back into the same tube.
5. To wash, add 0.75mL Buffer PE to column and centrifuge for 30 – 60 seconds at 5000 rpm.
6. Discard flow-through. Place QIAquick column back in the same tube. Centrifuge column for an additional 1 minute at maximum speed.
7. Place QIAquick column in a clean 1.5mL microfuge tube.

8. To elute DNA, add 50μL Buffer EB (10mM Tris-HCl, pH 8.5) or H_2O to the center of the QIAquick column and centrifuge for 1 minute at maximum speed. For increased DNA concentration, add a lesser volume of EB and let stand for one minute before centrifugation.

DNA Sequencing and Sequence Analysis

Purified PCR product was then sent to the University of California at Berkeley DNA Sequencing Facility in two 1.5ml eppendorf tubes: one containing the PCR - amplified DNA and forward primer, the other containing PCR - amplified DNA and the reverse primer. Sequencing was performed using an automated ABI Model 377 sequencer. The sequencing laboratory returned the *sasp-B* DNA sequence results via text e-mail, and in the form of a chromatograph, like the two seen below (one forward, one reverse).

Figure 3: Forward Chromatograph

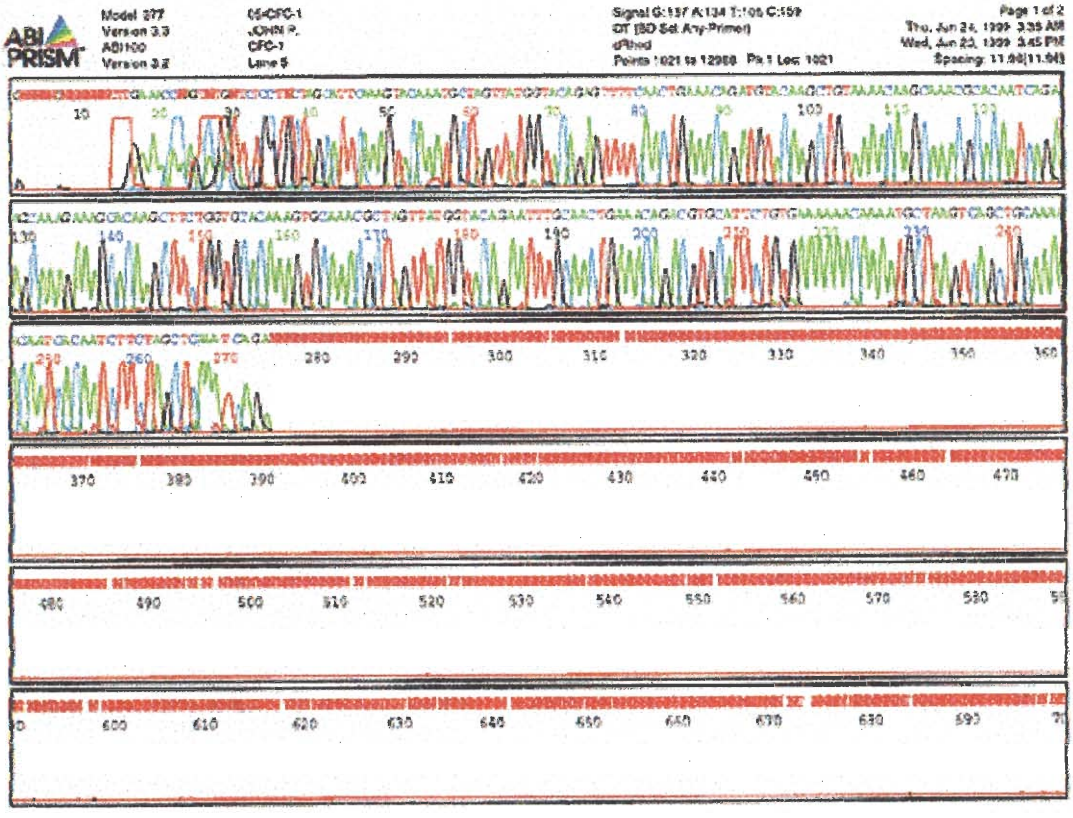
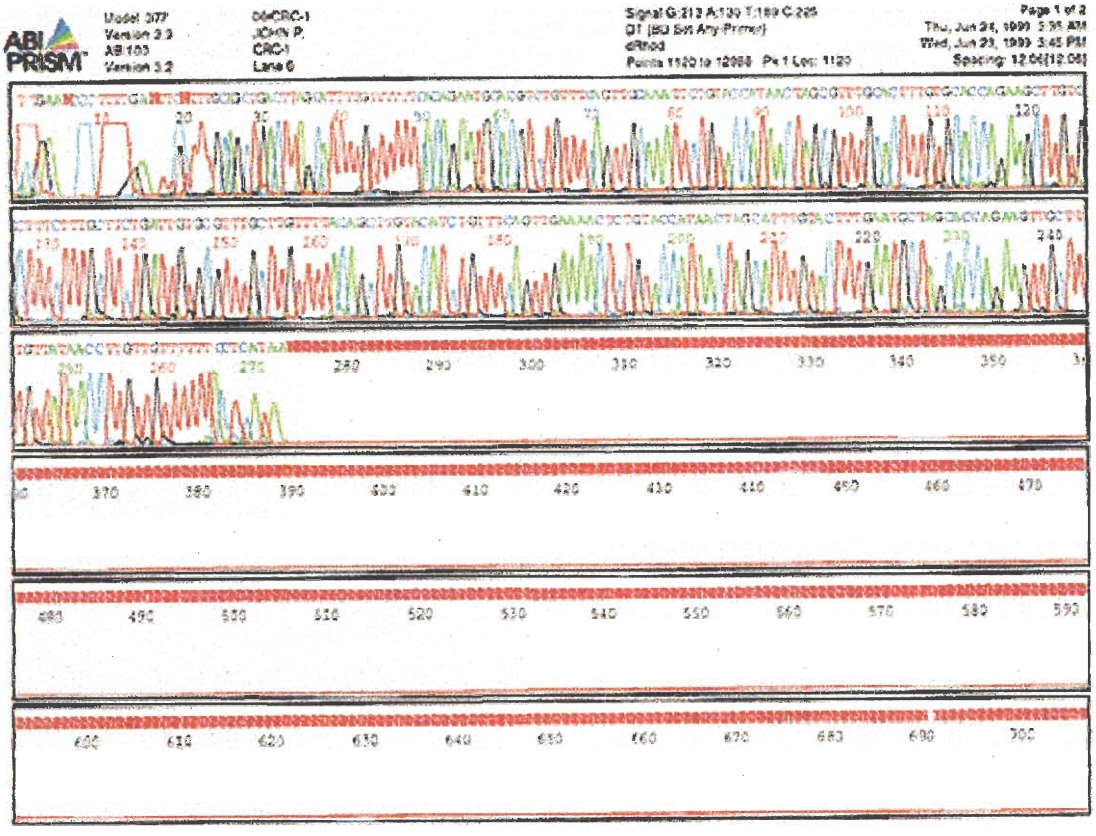


Figure 4: Reverse Chromatograph



The *sasp-B* DNA sequence for each strain was analyzed with DNAstar software and the Biology Workbench web site. An alignment of forward and complementary reverse sequencing results was made on DNAstar. The start and stop codons of the *sasp-B* gene were identified. A consensus gene sequence was then determined by resolving the nucleotide differences between forward and reverse sequencing results. This was done by examining the accuracy of each forward or reverse result in the surrounding area of the sequence, and by referring to the decisiveness of the result shown by the chromatograph – taller peaks of a single color are more accurate, while several shorter peaks indicate a less accurate result. The consensus *sasp-B* DNA sequence for each strain was then transferred to Biology Workbench.

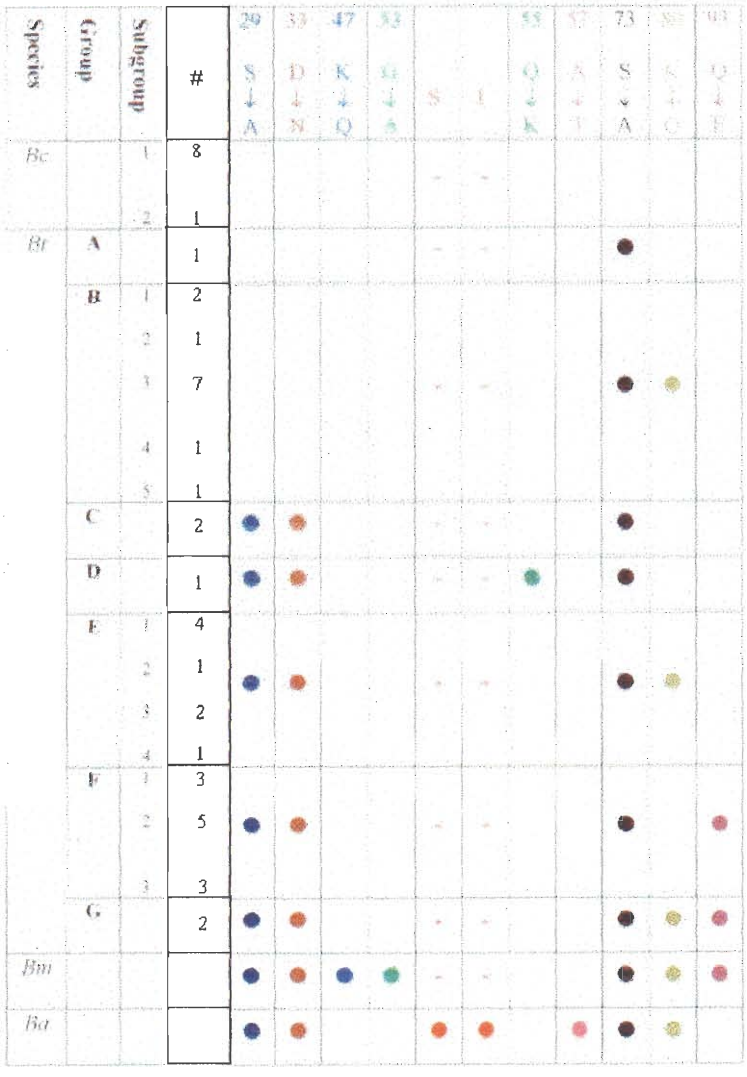
The translated amino acid sequence was then found using Biology Workbench, and the results were compared to the results from other strains to determine the strain's species and group. Species designation was determined by whether the strain's *sasp-B* amino acid sequence matched the standard amino acid sequence of a known member of a species, such as *Bacillus cereus*. However, a species can include more than one *sasp-B* amino acid sequence. In this case, strains were sorted into groups, each of which has a distinct *sasp-B* amino acid sequence. Nucleotide sequence was then used to determine the strain's subgroup, for instances in which a change in the nucleotide sequence did not lead to a change in the amino acid sequence of *sasp-B*.

Finally, sequence data obtained from the above methods were used to construct a phylogenetic tree. Sequence data from several groups and subgroups of *Bacillus thuringiensis*, along with representative sequences of *Bacillus cereus, anthracis,* and *mycoides* were used to make the tree. A majority rule, unweighted parsimony tree rooted to *Bascillus mycoides* as the outgroup was generated using PAUP software.

Results:

All 66 strains examined were arranged by their amino acid differences from the standard *Bacillus cereus* amino acid sequence.

Table 1. Amino acid alterations in *sasp-B* of *B. cereus* group



Explanation of Table 1: This table describes the differences in the amino acid sequence of the *sasp-B* gene among the *Bacillus* strains we examined.

Species refers to the species to which a strain of bacteria belongs. Nearly all of the strains examined belonged to *Bacillus cereus* or *Bacillus thuringiensis* (the lone exception will be mentioned later). The species *Bacillus anthracis* and *Bacillus Mycoides* are listed mainly for reference.

Group / Subgroup - the strains are sorted into lettered groups based on their amino acid sequence for the *sasp-B* gene. The numbers refer to subgroup, which is based on the nucleotide sequence. Sometimes strains in the same group can have differences in nucleotide base sequence that don't change the amino acid sequence.

Number (#) – Refers to the number of strains sequenced which belonged to this subgroup (or group, where multiple subgroups do not exist).

Amino acid differences - The numbers at the top of the right six columns tell us where along the *sasp-B* gene an amino acid substitution is located. Letters represent standard one letter abbreviations for amino acids. The top letter is the amino acid present in *Bacillus cereus*, the bottom letter is the amino acid it has been changed to in the strain being examined.

The following figure illustrates the amino acid differences between Bacillus cereus and each of the Bacillus thuringiensis groups. Each letter represents one amino acid.

Table 2: Amino acid sequence alignment of *Bacillus thuringiensis* groups.

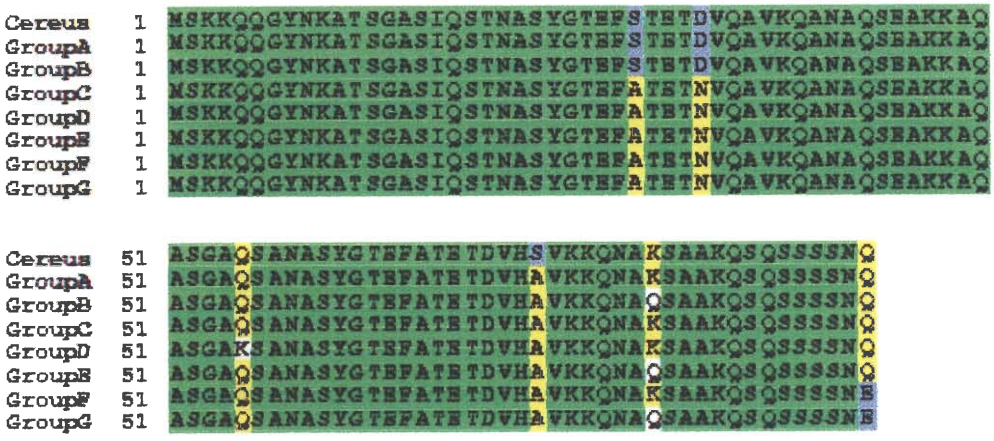


Table 2: Alignment of *sasp-B* amino acid sequences for *Bacillus Cereus* and *Bacillus thuringiensis* groups A through E. Single letter abbreviations are used for the amino acids. The numbers 1 and 51 refer to the amino acid position in the *sasp-B* gene where each row of the table starts.

Key of single letter amino acid abbreviations:

A - Alanine	I – Isoleucine	R - Arginine
C - Cysteine	K – Lysine	S - Serine
D - Aspartic Acid	L – Leucine	T - Threonine
E - Glutamic Acid	M - Methionine	V - Valine
F - Phenylalanine	N – Asparagine	W - Tryptophan
G - Glycine	P - Proline	Y - Tyrosine
H - Histidine	Q - Glutamine	

Two strains, here designated as "W" and "X" (more specific strain identifications can not be given until this study has been published), gave unusual results following PCR amplification of the *sasp-B* gene. Gel electrophoresis following

PCR revealed two distinct bands, when certain pairs of PCR primers were used, for both of these strains. One DNA fragment appeared at the normal located for a *Bacillus cereus* group *sasp-B* gene, while the other fragment was longer.

The DNA fragment in each band was sequenced separately to determine whether or not both fragments from strains W and X represented a *sasp-B* gene. The summarized results are as follows:

W short band: *sasp-B* gene from *Bacillus thuringiensis* group B (two amino acid alterations from standard *Bacillus cereus* sequence)

W long band: resembles *sasp-B* from *Bacillus thuringiensis* group E , but with one new mutation (six alterations from *Bacillus cereus*), and a 111 base pair inserted sequence.

X short band: *sasp-B* gene from *Bacillus mycoides* (seven alterations from *Bacillus cereus*)

X long band: identical coding sequence to W long band.

Analysis of the longer *sasp-B*-like sequence from strains W and X revealed that it resembles a *Bacillus thuringiensis sasp-B* gene with a 111 base pair insert, and is not listed in any major gene sequence database. Examination of this insert showed that it was similar in sequence to two other 111 base pair sequences within the longer *sasp-B*-like sequence. These three similar repeated sequences were labeled "subgenes" A, B, and C. Subgene B was labeled as the insert, according to the nucleotide sequence alignment against the standard *sasp-B* sequence of *Bacillus cereus*, made by Biology Workbench's CLUSTALW application. The three subgenes were found to be no more than 85% – 90% similar to each other.

Further investigation revealed that each *Bacillus cereus* group *sasp-B* gene contains two of these subgenes. They are located one immediately after the other, comprising most of the gene, with some non-repeated sequence at the beginning and the ending of the coding region. Strains W and X have three subgenes instead of two in the long DNA fragment, making the total sequence of these fragments exactly 111 base pairs longer than a typical *sasp-B* gene. A comparison of these *sasp-B* gene structures would look like this:

Bacillus cereus	W and X long fragment
Initial coding	Initial coding
Subgene A	Subgene A
Subgene C	Subgene B
End coding	Subgene C
	End coding

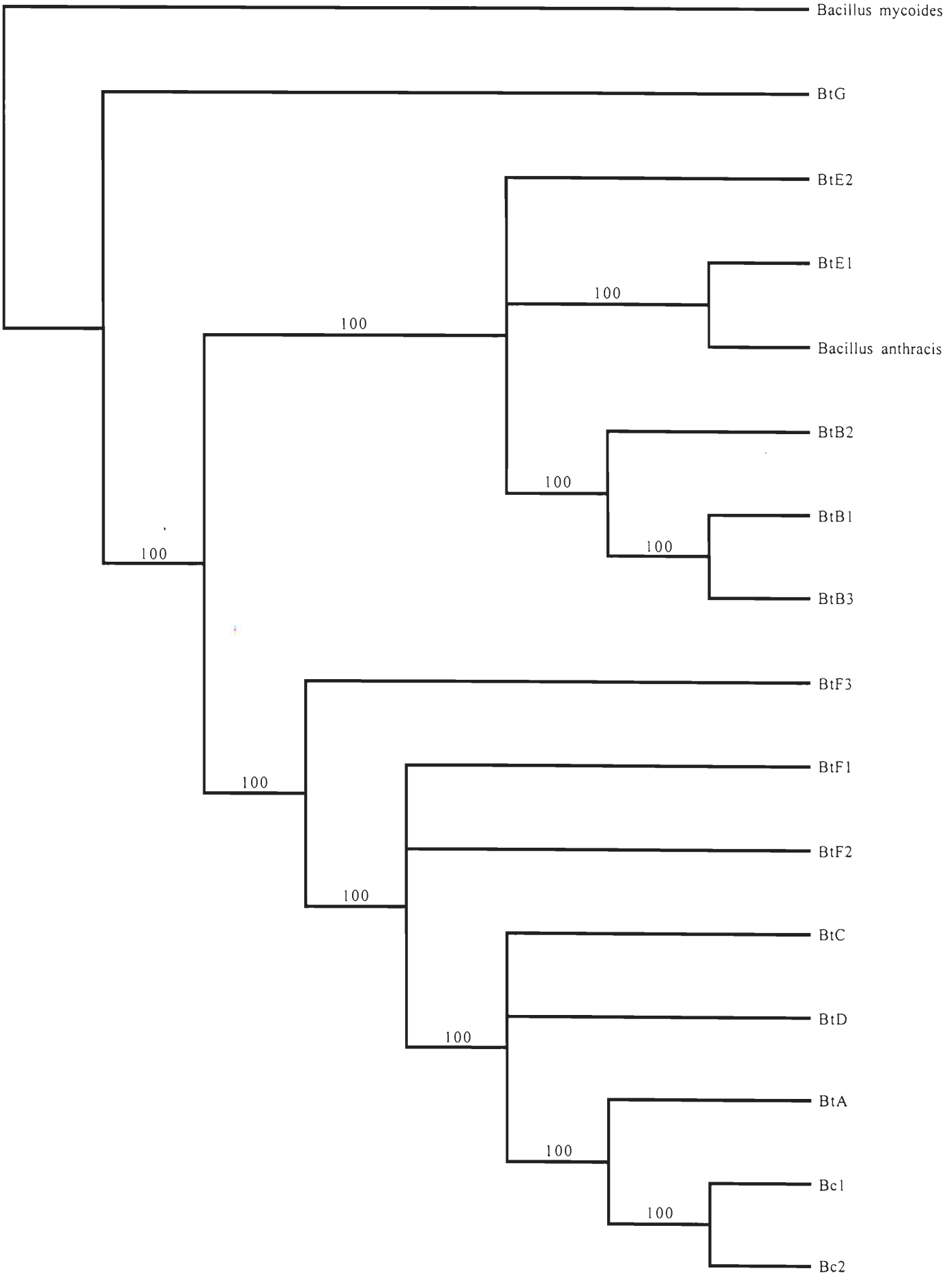
Figure 5: A phylogenetic tree obtained from *sasp-B* DNA sequence data is shown on the following page. The key to the phylogenetic tree is as follows:

BtE2 = *Bacillus thuringiensis*, group E, subgroup 2

Bc1 = *Bacillus cereus*, subgroup 1

The numbers along branches refer to bootstrap values, a measure of confidence in that portion of the phylogenetic tree

Majority rule



As shown by the phylogenetic tree, *Bacillus thuringiensis* is paraphyletic with respect to both *Bacillus cereus* and *Bacillus anthracis* at the *sasp-B* locus. This means that some strains of *Bacillus thuringiensis* are more closely related to *Bascillus cereus* or *Bascillus anthracis* than they are to other strains of *Bacillus thuringiensis*. Although some branchings remain unresolved, the resolution of these branchings is not relevant to the paraphyletic status of *Bacillus thuringiensis* with respect to *Bacillus cereus* and *Bacillus anthracis*.

Discussion:

Interpretation of sequencing results

As illustrated in Table 1, there is considerable variation in amino acid sequence among the *sasp-B* genes of *Bacillus thuringiensis* strains. This would seem to support the idea that this species encompasses a relatively wide range of bacteria, in comparison with the more narrowly defined *Bacillus cereus* and *Bacillus anthracis*.

All of the sequenced strains had been listed as *Bacillus thuringiensis*, probably based on their ability to kill insects. Yet, *sasp-B* sequence analysis showed that 9 of the 66 strains examined actually belonged to the species *Bacillus cereus*. It is quite likely that these strains were wrongly categorized as *Bacillus thuringiensis* based on a plasmid they were carrying that allowed them to kill insects, since insecticidal ability is notable characteristic of this species..

These data demonstrate an advantage of molecular techniques (gene sequencing) over phenotypic observation (sorting by insecticidal ability). Phenotypic observation can sort organisms into groups that share certain characteristics. But by using molecular sequence data, we can be assured that every difference we see is a direct product of that organism's evolutionary history, rather than some aspect of phenotypic plasticity or the ephemeral presence of a plasmid, for example. Although DNA sequence analysis is not immune to homoplasy, strong selection, and horizontal transfer of DNA, it remains a vital tool in determining phylogenetic relationships.

Anomalous results: strains W and X

The fact that strains W and X, which contain normal-sized *sasp-B* from different species (*Bacillus thuringiensis* and *Bacillus mycoides*), share an identical longer *sasp-B* gene is quite puzzling. The shared, longer *sasp-B* gene closely resembles neither of the normal-sized *sasp-B* genes from strains W and X. And these two normal sized *sasp-B* genes do not resemble each other, appearing to come from different species.

How then did the long *sasp-B* gene arise and come to be in both of these organisms? To arise in one of these strains from the normal *sasp-B* gene would have required a duplication of the entire gene, another duplication for the 111 base pair inserted sequence, plus several amino acid alterations – since even without the insert, the long *sasp-B* gene is three amino acids different from the short *sasp-B* gene of strain W, and four amino acids different from the short *sasp-B* of strain X.

Since the long *sasp-B* gene does not appear to have arisen from either of the normal *sasp-B* genes involved, another possibility would be that both organisms acquired the long *sasp-B* gene through horizontal gene flow. Then the obvious question would be, where did this gene come from originally? None of the databases on Biology Workbench were able to match the amino acid sequence of the long *sasp-B* gene to a known protein from any bacterium. The long *sasp-B* gene appears to be a coding sequence that has never before been catalogued.

Another possibility is that the long *sasp-B* gene is representative of a more ancient *Bacillus* genotype. Perhaps strains W and X are among the ancestors of modern *Bacillus thuringiensis* and *Bacillus mycoides*, and the long sequence is the type of *sasp-B* gene which was possessed by an ancestral strain. It seems plausible that in such an ancestral strain, which possessed only the long *sasp-B* gene, a partial duplication of this gene occurred, or a duplication followed by a deletion occurred, producing a bacterium that possessed *sasp-B* genes of two different lengths. A problem with this hypothesis might be: if this arose from duplication, why would the long fragment be so highly conserved, while several changes occurred in the short fragment to produce the differences seen between the W short and X short *sasp-B* genes. Possibly the long *sasp-B* gene was needed for a specific purpose, and was therefore highly conserved, while the redundant and functionless short *sasp-B* gene could undergo change more rapidly. Then, once each strain evolutionarily modified its short *sasp-B* to the point where it could perform its current function – perhaps the same function as originally served by the long *sasp-B* gene – each strain was free to delete the long *sasp-B* gene from its genome, with favorable evolutionary consequences. Another serious problem with this theory is that it requires a number of independent deletions of the long *sasp-B* gene to produce the varieties of

Bacillus we observe today. It is unlikely that all strains in this study posessed the long *sasp-B* sequence, since the *sasp* gene family has been well characterized in *Bacillus*, and no similar sequence is present in major databases. Additionally, a considerable number of primer combinations failed to amplify any similar DNA sequence in any other strain. Therefore, explaining the absence of *sasp-B* from other *Bacillus* strains is a problem for this hypothesis.

Perhaps the most likely explanation for the results obtained for strains W and X is that they each carry the long *sasp-B* gene on a plasmid. The capacity for genetic transformation between many *Bacillus* species has been established (Zawadzki, Roberts, and Cohan, 1995). This would seem to suggest that two different *Bacillus* species such as the *Bacillus thuringiensis* and *Bacillus mycoides* involved in this example might contain some of the same plasmids. In particular, *Cry* plasmids contain a large number of genes and may be present in *Bacillus* strains. It seems far more probable that strains W and X share a similar plasmid, rather than relying on the unlikely scenario that the long *sasp-B* represents an ancestral genotype. More research is needed to determine the exact nature of the long *sasp-B* sequence.

Phylogenetic relationships and species implications

Examination of the phylogenetic tree produced for the *sasp-B* sequence data indicates a paraphyletic relationship between *Bacillus thuringiensis* and *Bacillus cereus*, and between *Bacillus thuringiensis* and *Bacillus anthracis*. The gene tree has a high degree of support for both of these phylogenetic relationships. From a natural history perspective, it seems plausible that early *Bacillus cereus* and *Bacillus anthracis* each evolved from a separate *thuringiensis*-like ancestor.

In light of these results, the status of these three traditionally defined species comes into question. Whether these species designations are to be retained depends on the species definition that is considered. Of course some species definitions, such as the biological and recognition species concepts, will not be applicable to a study of asexual organisms. Other species definitions, such as the ecological species concept, would require information not acquired by this project (de Quieroz, 1998).

Among those species definitions that can be applied to this situation, there are differences in regard to the status of the traditionally defined species in this study. Some species definitions, such as the cohesion and phenetic species concepts, focus at least in part on phenotypic cohesion or closeness (de Quieroz, 1998). These definitions could potentially acknowledge *Bacillus anthracis*, *cereus*, and *thuringiensis* as acceptable species. Presumably, these species were originally demarcated by aspects of their phenotypes, such as visual appearance (under a microscope), insecticidal ability, or antigens.

However, the application of some phylogenetic and genealogical species definitions could give quite different results. Many such definitions require a condition of reciprocal monophyly between species. This is certainly not the case in the paraphyletic relationships indicated by the *sasp-B* gene trees. These species definitions would demand that the species designations *Bacillus cereus* and *Bacillus anthracis* either be eliminated (referring to the whole group as *Bacillus thuringiensis*) or greatly expanded (into currently *Bacillus thuringiensis* lineages) to form reciprocally monophyletic groups.

Although this analysis leaves us without a definite categorization scheme for these species of the *Bacillus cereus*, some additional perspective can be gained. First, different species criteria may be appropriate for different applications. In a strictly phylogenetic sense, it might make sense to require at least a certain degree of reciprocal monophyly. If an ecological study was being done, other criteria would probably be more appropriate. And second, it has been proposed that species may go through stages, fulfilling criteria along the way. In this fashion, a species might become a cohesion species at one point, and later become a genealogical species at a later time (Harrison, 1998).

Future direction and implications of this research

While the stage of the experiment reached to date does provide much useful information, the research as a whole is not over yet. The next major step will be cloning of the *sasp-B* gene. Each *sasp-B* gene that represents a unique coding sequence (each subgroup) will be cloned into competent *Escheria coli* cells using a plasmid vector. Special primer binding sites on the plasmid will allow the entire coding region to be sequenced again, giving us more accurate data around the beginning and ending of the coding sequence. It will also provide a convenient way to store the gene. Additionally, Southern blot hybridization can be performed to confirm the presence of the long *sasp-B* sequence in strains W and X.

This experiment has the potential to be helpful in several areas of scientific research. For example, having an accurate phylogenetic tree for the *Bacillus cereus* group could allow scientists to more quickly identify strains that might be useful in bioremediation projects, such as GTN and TNT. Additionally, since *B. t.*

toxin has widespread agricultural uses, knowing the phylogeny of the species could help in research ranging from organic farming to genetic engineering.

Conclusion:

Sequence analysis of the *sasp-B* appears to be an effective method to understand phylogenetic relationships among strains of bacteria belonging to the *Bacillus cereus* group. According to the *sasp-B* gene tree, *Bacillus thuringiensis* is paraphyletic with respect to both *Bacillus cereus* and *Bacillus anthracis*.

Acknowledgements:

I would like to thank:

Professor Ehab El-Helow, Professor Terrance Leighton, Professor Stephen Ekunwe, Sasha Shafikhani, and everyone else at the Leighton Lab for their help in designing this experiment and teaching me the laboratory techniques involved.

Laurel Egenberger and Tiffany Dressen of CSEE, the Lawrence Berkeley National Laboratory, and the U.S. Department of Energy for giving me this great opportunity to learn.

Professor DeWayne Shoemaker for serving on my thesis committee and helping me with the phylogenetic tree, Professor John Geiser for serving on my thesis

committee, and the Lee Honors College of Western Michigan University for providing thesis guidelines and examples.

Sources Cited:

de Quieroz, K. 1998. "The general lineage concept of species, species criteria, and the process of speciation" In D. J. Howard and S. H. Berlocher, eds. *Endless Forms*. Oxford, UK: Oxford University Press, pp. 57-75.

El-Helow, E. unpublished DNA isolation protocol, 1999.

Harrison, R. H. 1998 "Linking Evolutionary Pattern and Process". In D. J. Howard and S. H. Berlocher, eds. *Endless Forms*. Oxford, UK: Oxford University Press, pp. 19-31.

Leighton, T. personal communication, 1999.

Setlow, P. 1988. "Small, acid-soluble spore proteins of *Bacillus* species: Structure, synthesis, genetics, function, and degradation". Ann. Rev. Microbiol. 42:319-338

Shaw, K. L. 1998. "Species and the Diversity of Natural Groups". In D. J. Howard and S. H. Berlocher, eds. *Endless Forms*. Oxford, UK: Oxford University Press, pp. 44-56.

Sun, D. and Setlow, P. "Cloning and nucleotide sequencing of genes for a second type of small, acid-soluble spore proteins of *Bacillus cereus*, *Bacillus stearothermophilus*, and "*Thermoactinomyces thalpophilus*". Journal of Bacteriology. 3088-3093.

Wadman M. 1997. Dispute over insect resistance to crops. Nature 388:817.

Woodfin M. 1997. Bugging out. Sierra 82:27.

Zawadzski, P., Roberts, M. S., and Cohan, F. M. "The log-linear relationship Between sexual isolation and sequence divergence in *Bacillus* Transformation is robust". Genetics 140:917-932