The Bacillus cereus group species and phylogeny : A brief overview of a complex subject

Vincent Sanchis-Borja

The *Bacillus cereus* group species and phylogeny

A brief overview of a complex subject

Vincent Sanchis-Borja

HuPlant (Cost Action 16110) training School

AgroParisTech
INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT
PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES

Micalis
Microbiologie de l'Alimentation au service de la Santé

INRA
SCIENCE & IMPACT

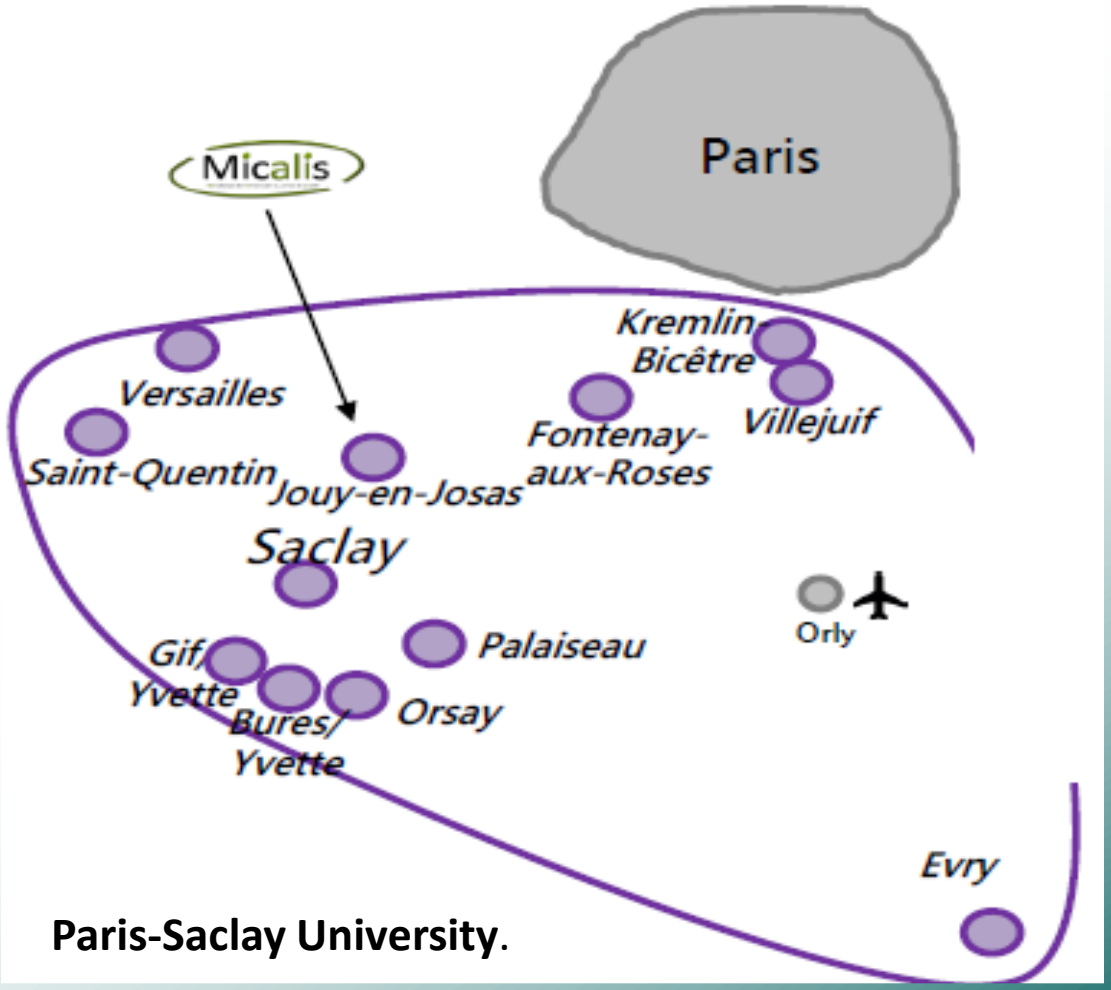
The Micalis Institute

Micalis Institute is a very large joint research unit, associating INRA, a national public institute for research in agronomy and AgroParisTech, a Graduate Institute of Technology for Life, Food and Environmental Sciences.

- Created on January 1st, 2010, by merging 8 pre-existing units

- All staff gathered in Jouy-en Josas since July 2016

- Part of the of the new Paris-Saclay University that brings together a group of 19 higher education institutions (research institutes, higher education institutions, “grandes écoles , research centers) alongside and competitiveness and business cluster



Paris-Saclay University.

The Micalis Institute

- It brings together over 350 persons including 125 researchers, engineers and professors, and more than 120 post-docs, PhD students, master students, and trainees.

- It is composed of 23 research teams organized in " Thematic departments" and hosts 4 technological platforms and 1 pre-industrial demonstrator.

- The aim of the Micalis institute is to develop innovative research in the field of "Food and Gut Microbiology for human Health".

- It is composed of 3 "thematic" or scientific departments:

1) Bacterial Adaptation and Pathogenesis (Emergence and control of food-borne opportunistic pathogen microorganisms)

2) Food and gut microbial ecosystem and food-microbiota-host functional interactions

3) Systems and synthetic microbiology.

Specific missions of the « Bacterial Adaptation and Pathogenesis » department

- 9 teams 67 INRA staff ~30 students/post-docs/interns-in-training

Scientific objectives

- Define the factors that facilitate establishment of a microbial population in food and in animal host (mammals - insects) environments.

- Identify and characterize the reprogramming mechanisms used by microbes in response to their environment

- Monitor and characterize the factors that lead to the emergence of microbial sub-populations or variants in food reservoirs and in the host mucosa.

Main practical missions :

- Targets to develop new methods to fight infections
- Fighting pathogens in their reservoir
- Fighting pathogens in foods and the processing chain
- Diagnostic tools
- Decision aid tools
- Participation in expertise to support public authorities, to support the development of standards and guidelines

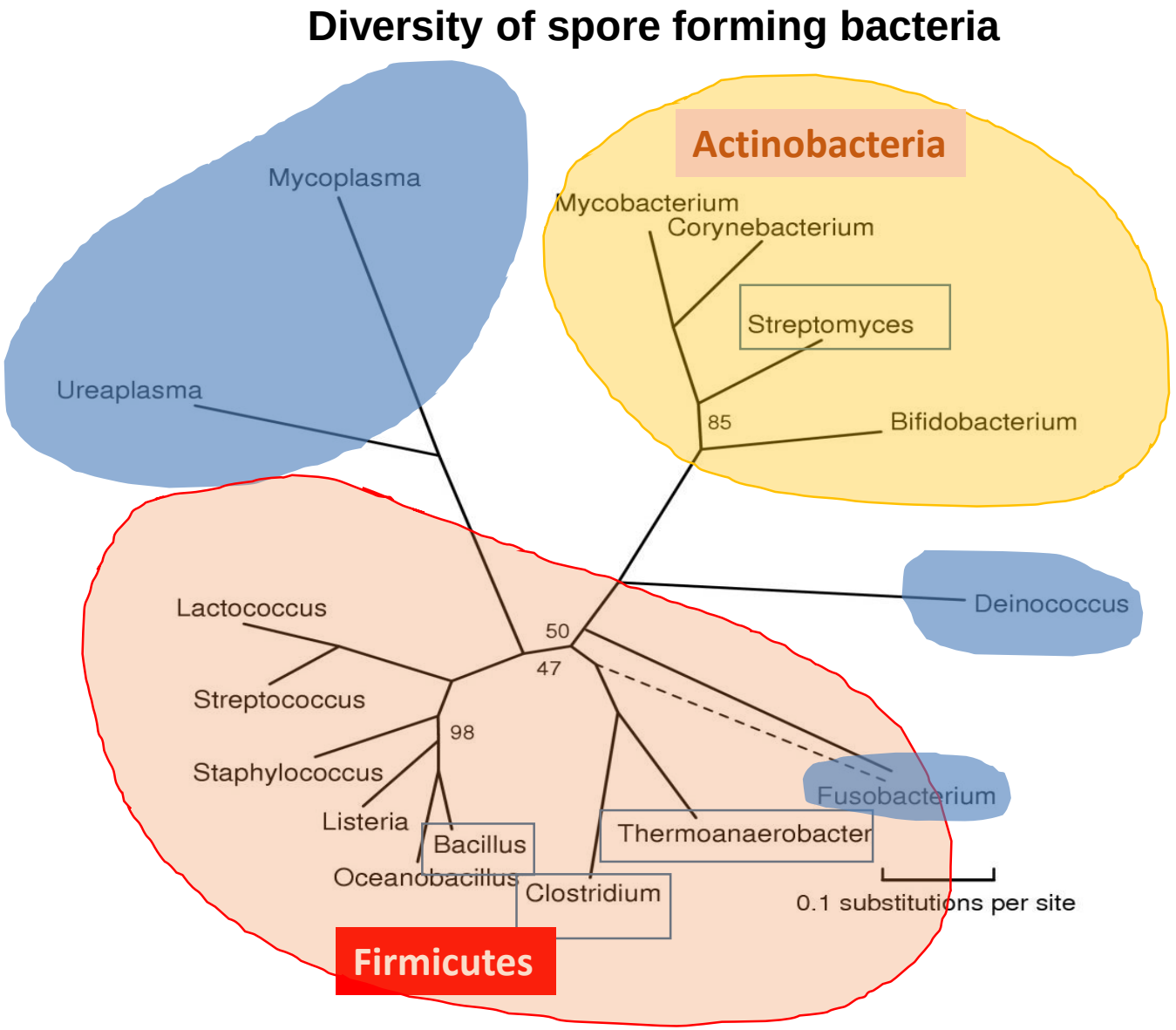
What is a bacterial species?

Species: A group of closely related and morphologically similar individuals, that, actually or potentially, interbreed. The concept of a bacterial species is less definitive than for higher organisms. A bacterial species may be regarded as a group of strains that share many features in common and exhibit a particular level of DNA homology and that differ significantly from another group of strains

- A bacterial (or genomic) species is currently defined from a phylogenetic perspective; it is a monophyletic set of strains having most of their genome in common.
- The modern polyphasic species distinction states that a group of strains belonging to one species must show more than 70% reassociation during DNA-DNA hybridization with good thermal stability of the reassociated DNA ($\Delta Tm < 5°C$), < 5 % mol G+C difference of total genomic DNA and > 97-98 % 16S rRNA identity or ANI value > 95%
- Applying the 95% ANI threshold to Primates will result in a single species that includes humans, chimpanzees and all monkeys.

General Characteristics of the Genus *Bacillus*

- Gram-positive, endospore-forming,
- Mostly saprophitic
- Aerobic and catalase positive
- Versatile in degrading complex macrom
- Primary habitat is soil
- 3 species of medical importance:
  - *Bacillus anthracis*
  - *Bacillus cereus*
  - *Bacillus cytotoxicus*



Modified from Bern and Goldberg BMC Evolutionary Biology 2005

Diversity of the *Bacillus cereus* group

- Heterogeneous group of genetically related spore forming low GC % *Bacillus* species

- 9 species :
  - *Bacillus anthracis (Ba) (1855)*
  - *Bacillus mycoides (Bm) (1886)*
  - *Bacillus cereus sensu stricto (Bc ss) (1887)*
  - *Bacillus thuringiensis (Bt) (1901)*
  - *Bacillus pseudomycoides (Bp) (1998)*
  - *Bacillus weihenstephanensis (Bw) (1998)*
  - *Bacillus cytotoxicus (B cyt) (2013)*
  - *Bacillus toyonensis (2013)*
  - *Bacillus wiedmanii (2015)*

- In addition, *B. gaemokensis, B. manliponensis, B. bingmayongensis* have been proposed as representing novel species but have not yet been validly published.

- The differentiation of species within *the B. cereus* group is particularly complex and the genetic proximity between members of the *B. cereus* group is so close that, from a strictly phylogenetic point of view, the classification of the species within it is highly controversial and debated.

The *Bacillus cereus sensu lato* group: six distinctive phenotypes

Bacillus anthracis : etiological agent of anthrax, a mammalian disease

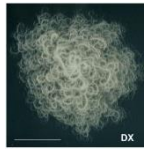
Bacillus cereus sensu stricto : opportunistic food pathogen

Bacillus thuringiensis : enthomopathogen, world wide use as a bio-pesticide

Bacillus toyonensis : probiotic

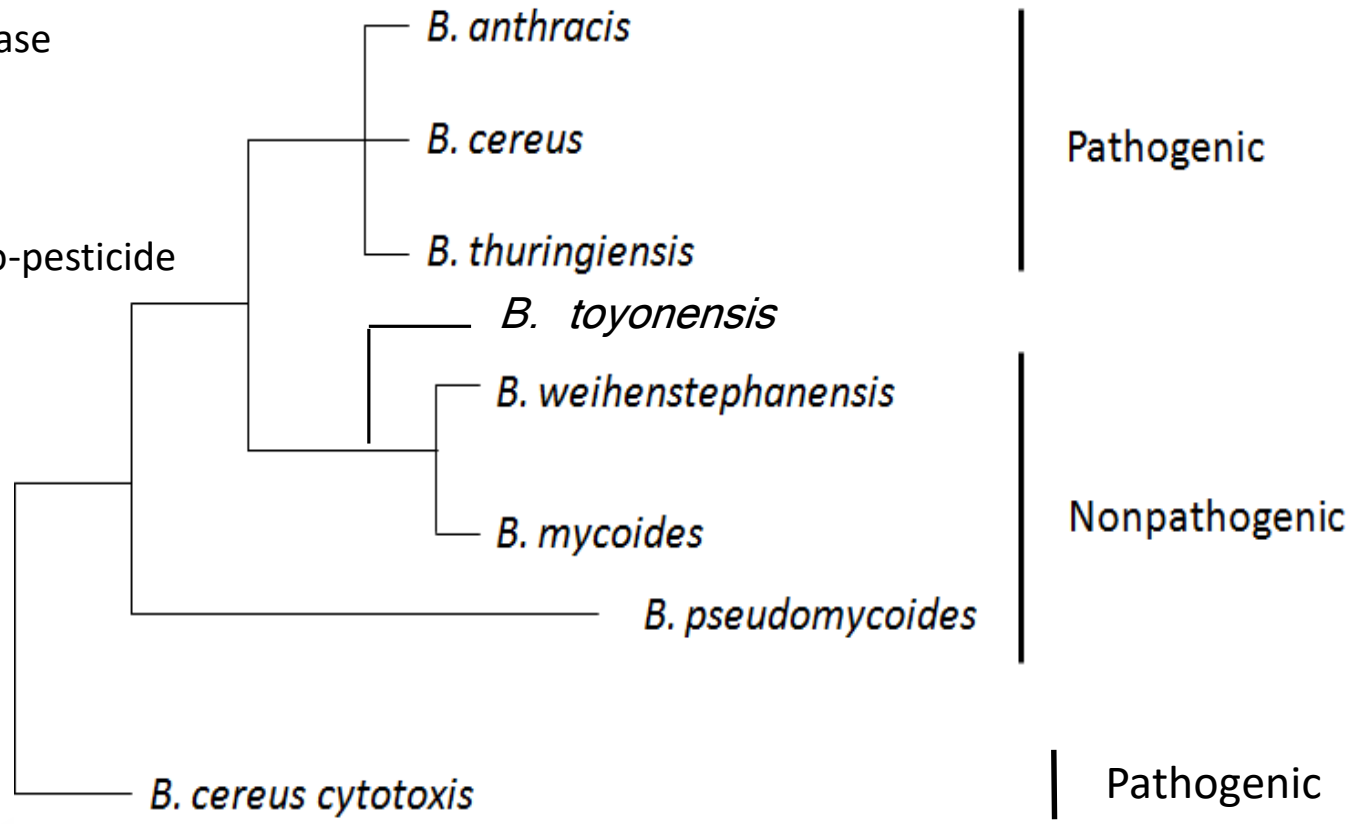
Bacillus weihenstephanensis : psychrotrophic (<6°C)

Bacillus mycoides : form rhizoidal colonies



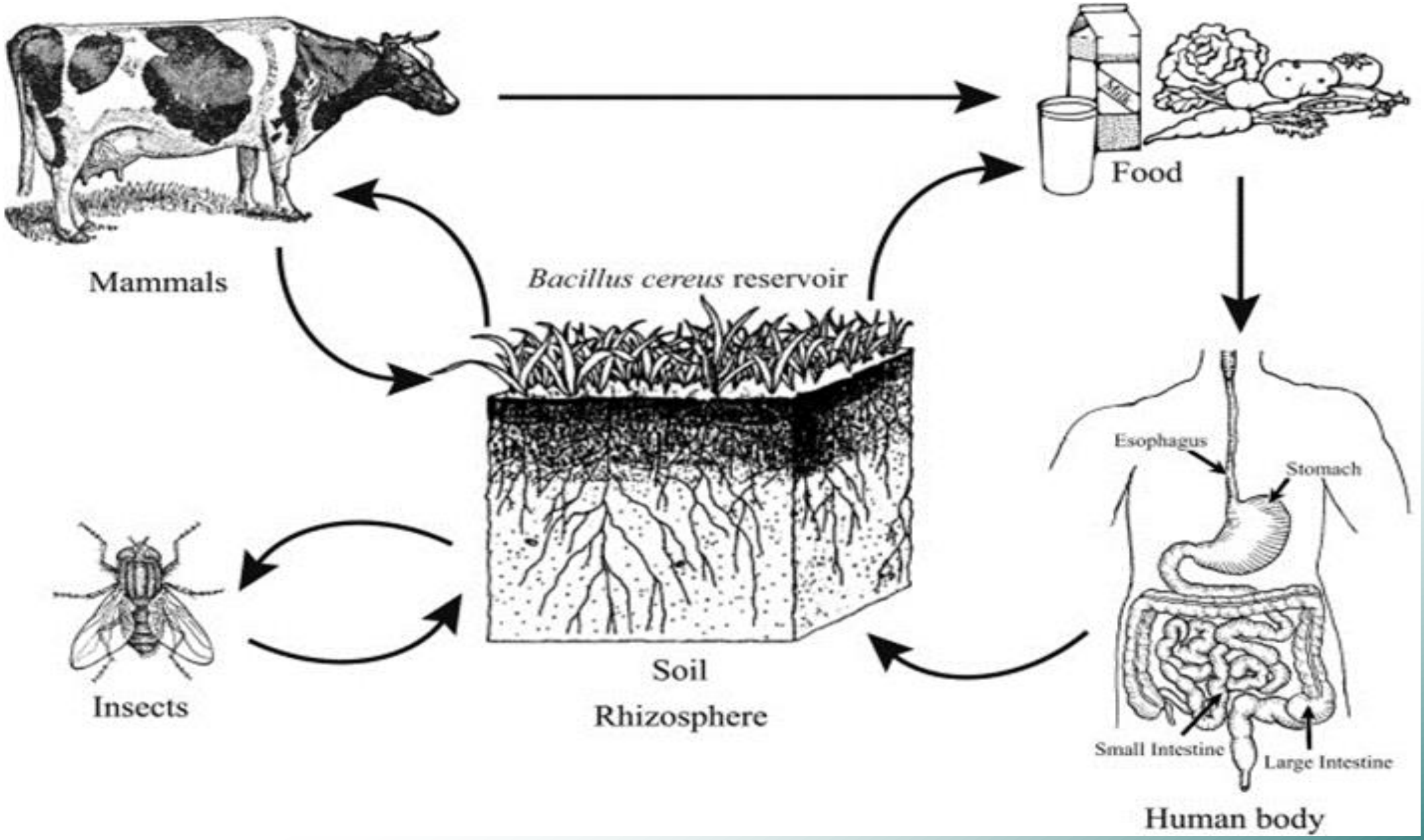
Bacillus pseudomycoides : form rhizoidal colonies

Bacillus cytotoxicus : severe food pathogen, thermophilic (>50°C)



Some species can growth at refrigeration temperatures others are moderately thermotolerant

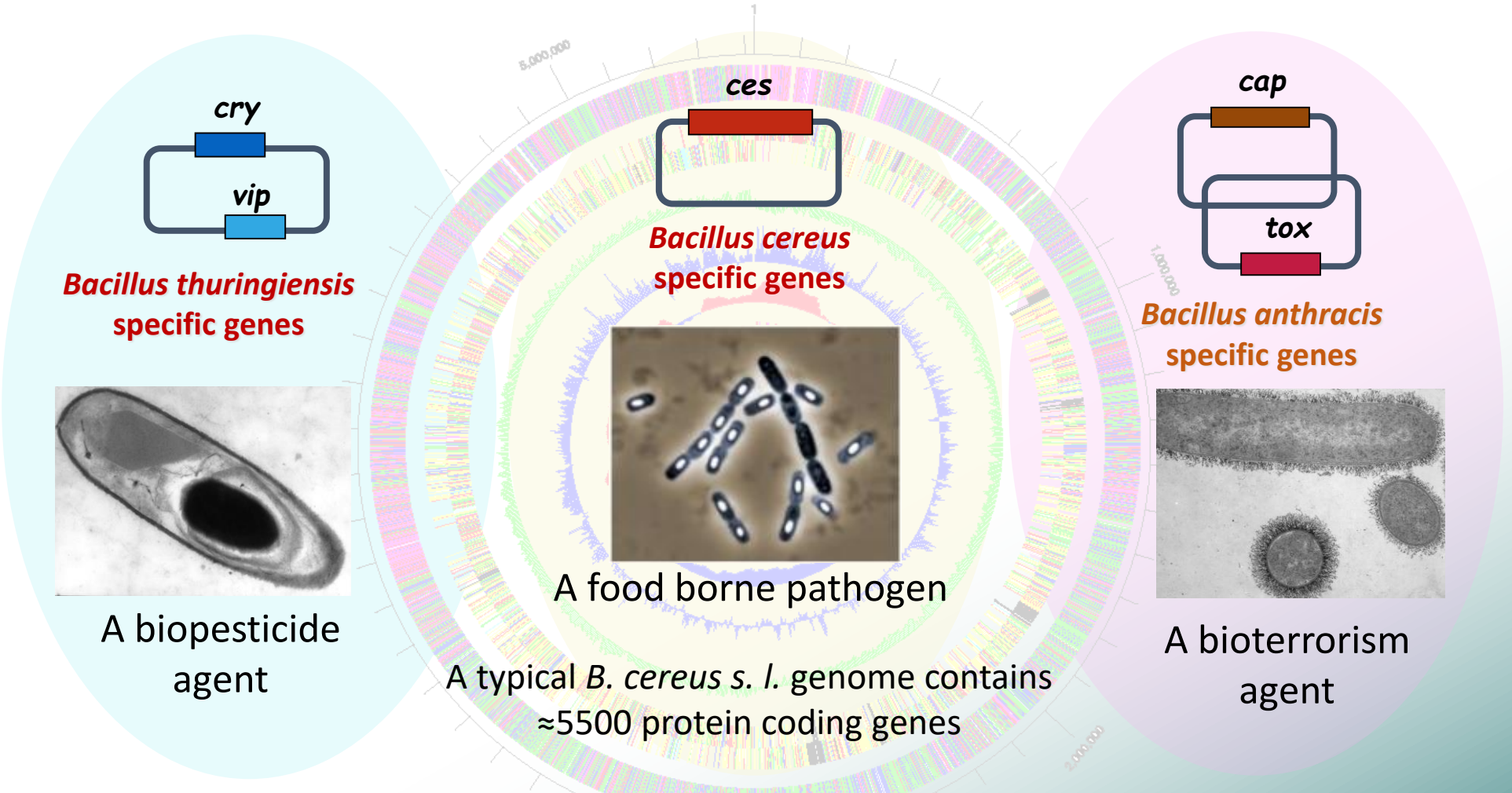
Bacillus cereus sensu lato ecology



The *Bacillus cereus sensu lato* group

Gram^{+}, sporulating, low GC% (35%) bacteria

Set of common genes represent ~ 75% of the genome (~ 4 Mb)



A biopesticide agent

A food borne pathogen

A bioterrorism agent

A typical *B. cereus s. l.* genome contains ≈5500 protein coding genes

Harbors a diverse range of plasmids that vary in number and in size (2–200kb)

The specific pathogen properties of these bacteria are due to plasmid

Bacillus cereus group diversity

- The three main species *B. anthracis*, *B. cereus sensu stricto* and *B. thuringiensis,* were recognized and established by the early 1900s because they each exhibited distinct phenotypic properties..
- Despite these apparently clear phenotypic definitions, early molecular approaches to separate them by various DNA hybridization and 16S/23S ribosomal sequence analyses led to some “confusion” because there were limited differences to differentiate between these species.
- *B. cereus sensu lato* represents a classic example of a now common bacterial species taxonomic dilemma.
- the pathogenic properties of *B. anthracis* and the biopesticide applications of *B. thuringiensis* appear to “have outweighed pure taxonomic considerations”

Bacillus anthracis

B. anthracis (anthrax bacillus) is the main pathogen of the group.

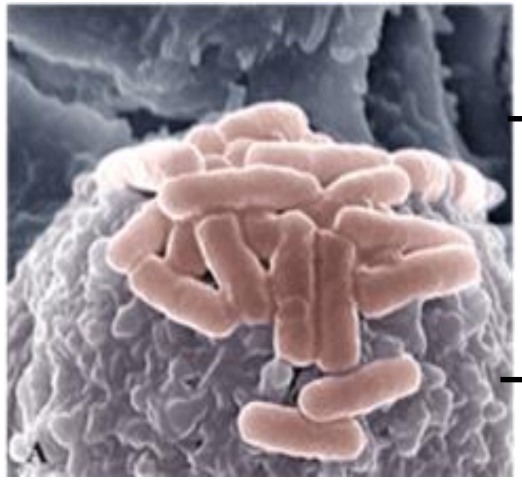
- primarily a pathogen of herbivores
- form spores which can remain viable in the environment for years.
- plasmid encoded virulence factors – polypeptide capsule made of gamma-D-glutamic acid and exotoxins
- Unlike *B. thuringiensis* and *B. cereus*, *B. anthra*cis is non-mobile, non-hemolytic and does not produce proteases and phospholipases
- 3 types of anthrax:
  - cutaneous – spores enter through skin, black sore-eschar; least dangerous
  - pulmonary – inhalation of spores
  - gastrointestinal – ingested spores

Bacillus cereus sensu stricto

- *Bacillus cereus sensu stricto* is a common airborne and found in the soil, water, vegetation and foodstuffs such as meat, milk, cereals and spices.
- Form spores which are found almost everywhere including hospitals, laboratories and food factories
- Grows in foods, spores survive cooking and reheating
- Frequent foodborne disease agent
- produces 2 types of enterotoxins - diarrhoeal enterotoxins and an emetic toxin; both are associated with food poisoning. *B. cereus* also produces a lecithinase (acts on the phospholipids on the cell membrane), and haemolysins (cereolysin and cytolysin).
- Ingestion of toxin-containing food causes nausea, vomiting, abdominal cramps and diarrhea; 24 hour duration
- Increasingly reported in immunosuppressed patients

B. cereus impact on public health and food industry

Opportunistic pathogen capable of causing several types of infections

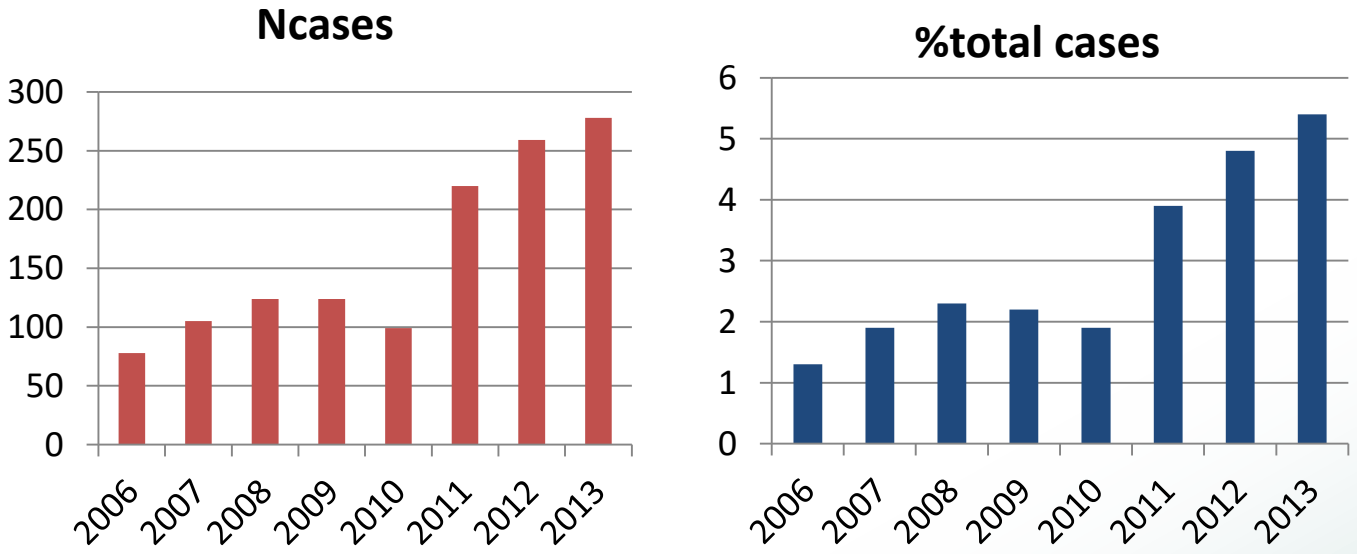


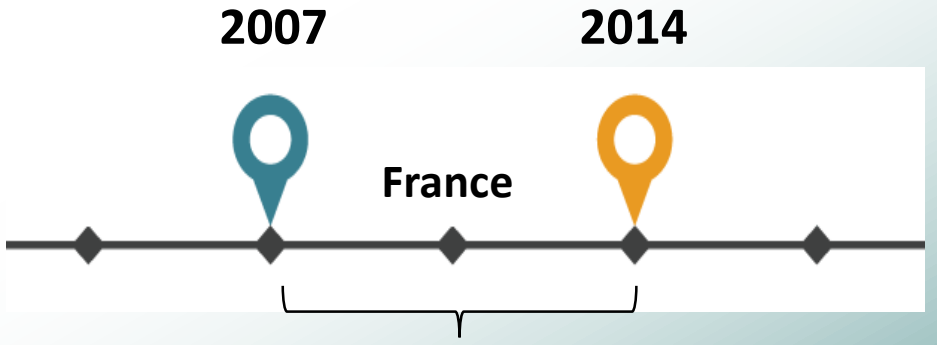
- **Food poisoning**
  - **Diarrheal syndrome**
  - **Emetic syndrome**
- **Severe infections (mostly immunocompromised patients)**
  - **Systemic infections:** Pneumonia, Meningitis, Endocarditis, Septicemia

Specific role in spoilage of heat processed foods and non heated foods

Foodborne pathogenic spore forming bacteria : *B. cereus* impact on public health

- In EU : 4th cause of foodborne outbreaks in 2013 (*Salmonella*, viruses, *S. aureus*, *B. cereus*, *Clostridium*)
- In France *:* 6th to 2nd cause of foodborne outbreaks 1996-2012
- In France : 2nd agent most frequently confirmed in foodborne illnesses : 22% of 2478 cases (INVS, 2014)





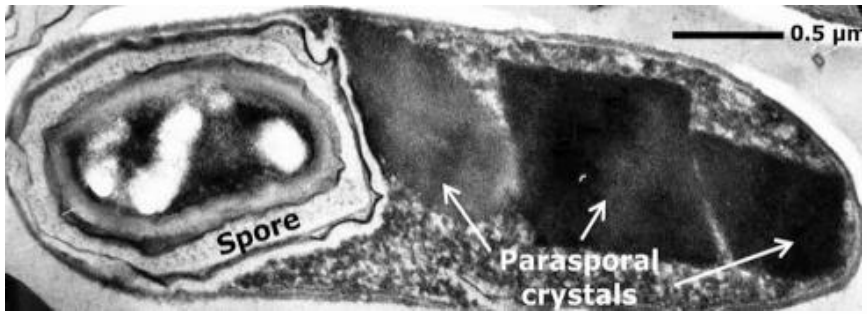
- 413 strong-evidence foodborne outbreaks
- 6 657 people affected
- 352 hospitalisations

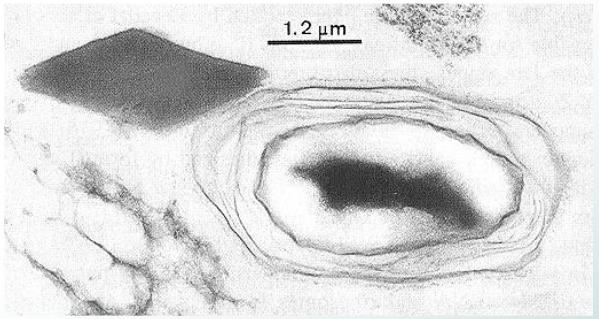
Growing impact in EU of *B. cereus* foodborne diseases (EFSA-ECDC)

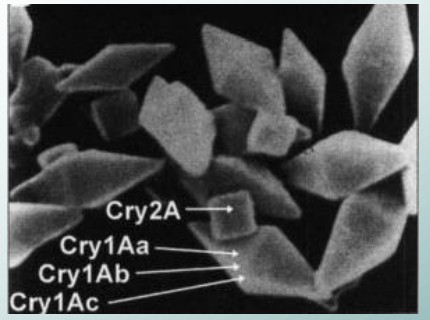
[EFSA BIOHAZ Panel, 2016]

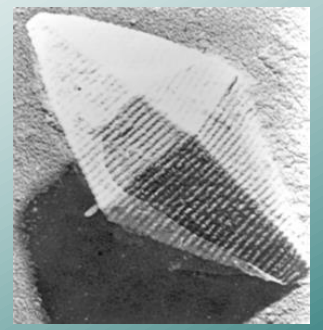
Bacillus thuringiensis

- *Bacillus thuringiensis* is a common, spore-forming soil bacterium.
- Characterized by the production of crystal inclusions containing proteins (Cry toxins) that are toxic to insects
- The toxicity of these crystal proteins against certain insects and their high specificity led to the development of bio-insecticides
- Commonly used in garden sprays & for commercial agriculture, including organic farming
- Extremely well-known toxin in terms of human health & environmental safety

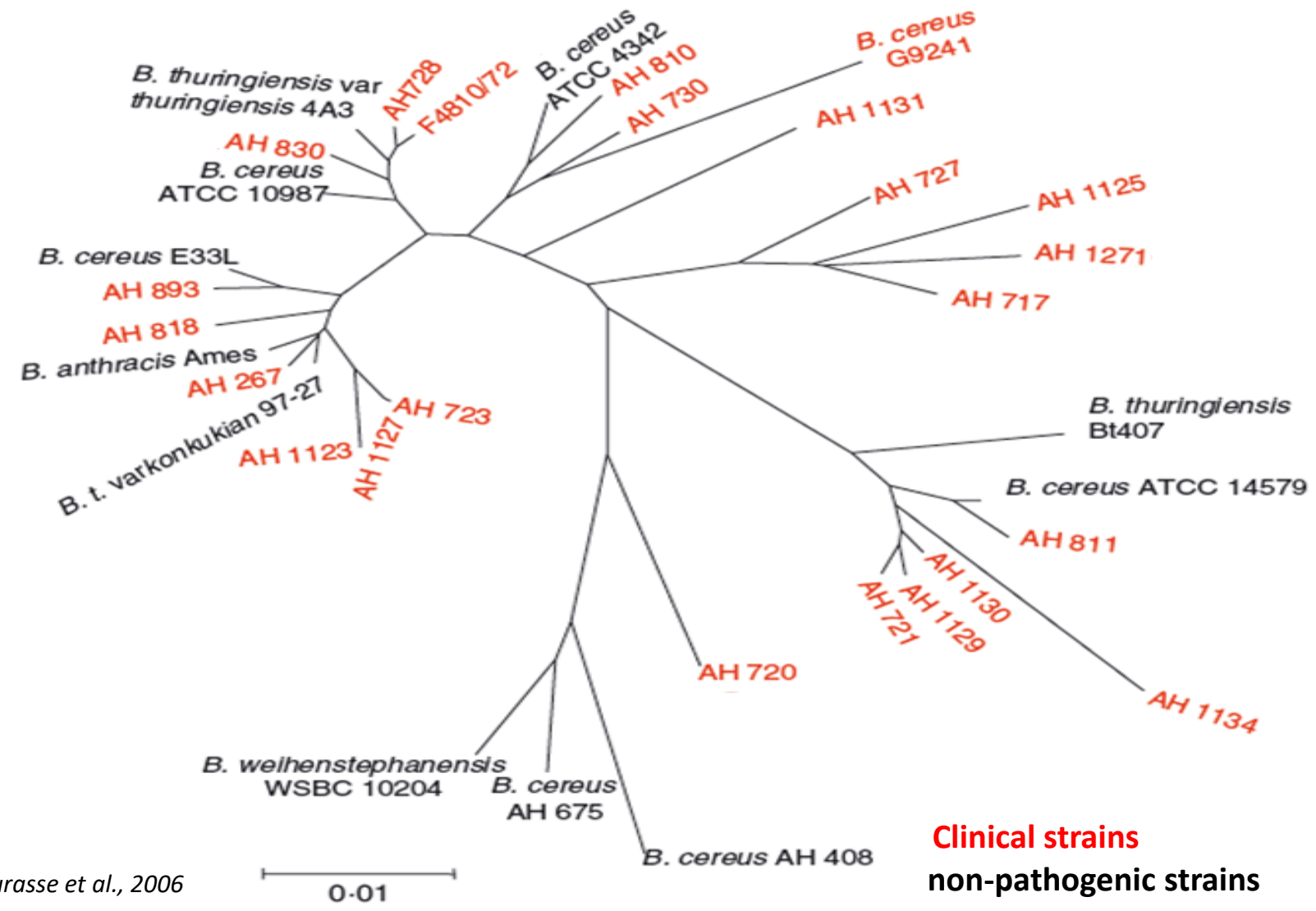








Phylogeny and pathogenicity of the *B. cereus* group



Tourasse et al., 2006

Clinical strains

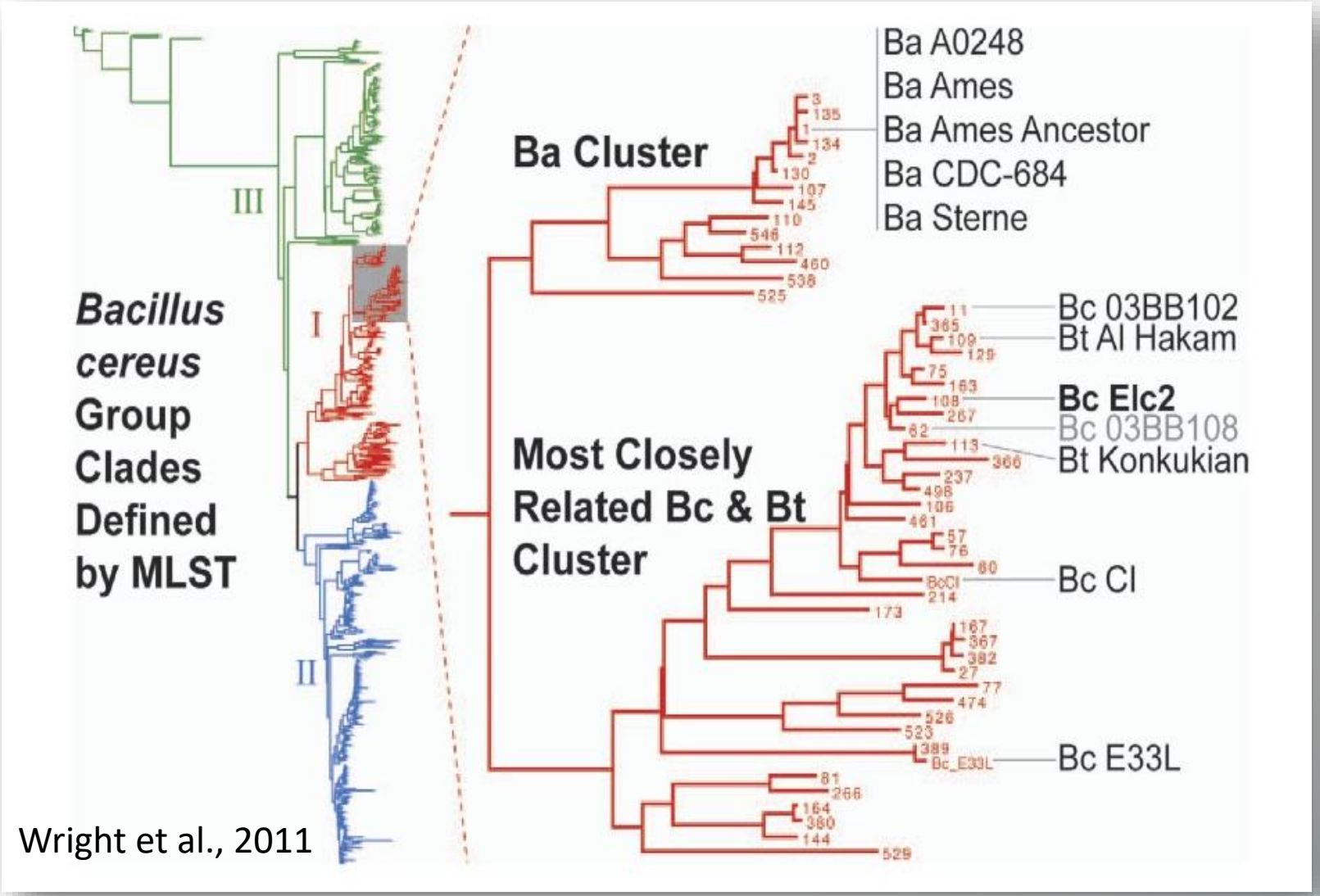
non-pathogenic strains

Phylogenetic tree of strains representative of the *B. cereus* group based on MLST analysis

Multilocus sequence typing (MLST), amplified fragment length polymorphism (AFLP) have been used to extensively analyze *B. cereus* sensu lato

MLST cannot clearly differentiate pathogenic from benign strains

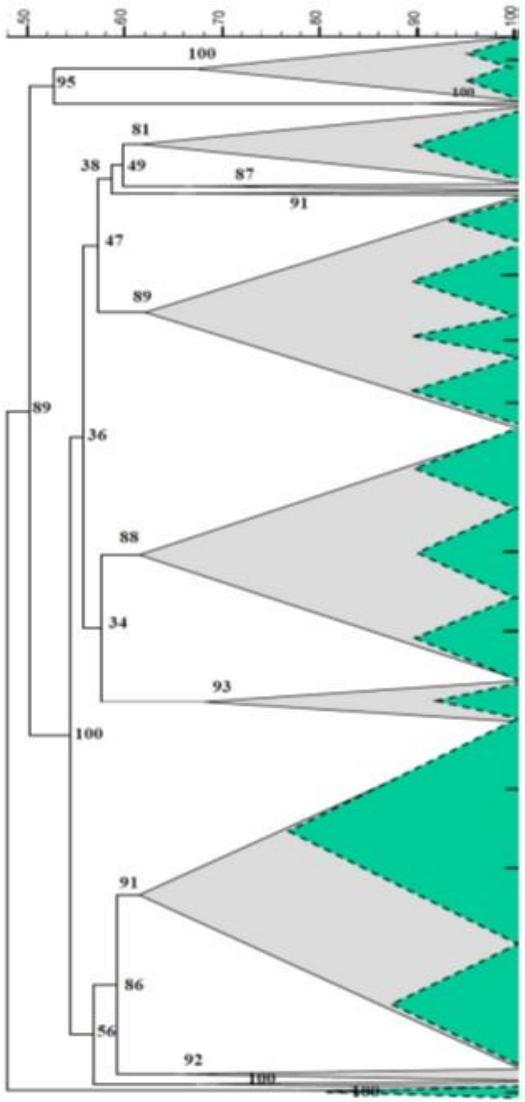
MLST analysis reveals 3 well phylogenetically separate clades



Genetic relationships among *Bacillus cereus* group strains based on multilocus sequence type data obtained for 7 loci

- *B. anthracis* = clonal species
- Strains are really genetically closed
- Highly resolutive methods are necessary to discriminate B. anthracis strains:

An ecotype classification approach using fAFLP and phenotypic properties divides the phylogeny of the *B. cereus* group into seven major groups

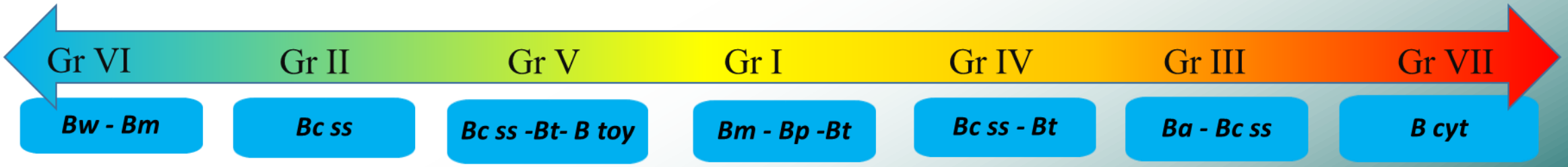


Phylogenetic group	Species	Temperature of growth	Thermotype
I	*B. pseudomycoides*	10°C to 43°C	mesophilic
II	*B. cereus* *B. thuringiensis*	7°C to 40°C	psychrotolerant
III	*B. cereus* *B. thuringiensis* *B. anthracis*	15°C to 45°C	mesophilic
IV	*B. cereus* *B. thuringiensis*	10°C to 45°C	mesophilic
V	*B. cereus* *B. thuringiensis*	8°C to 40°C	intermediate
VI	*B. mycoides* *B. weihenstephanensis* *B. thuringiensis*	5°C to 37°C	psychrotrophic
VII	B. cytotoxicus	20°C to 55°C	thermotolerant

Guinebretière *et al.*, 2008

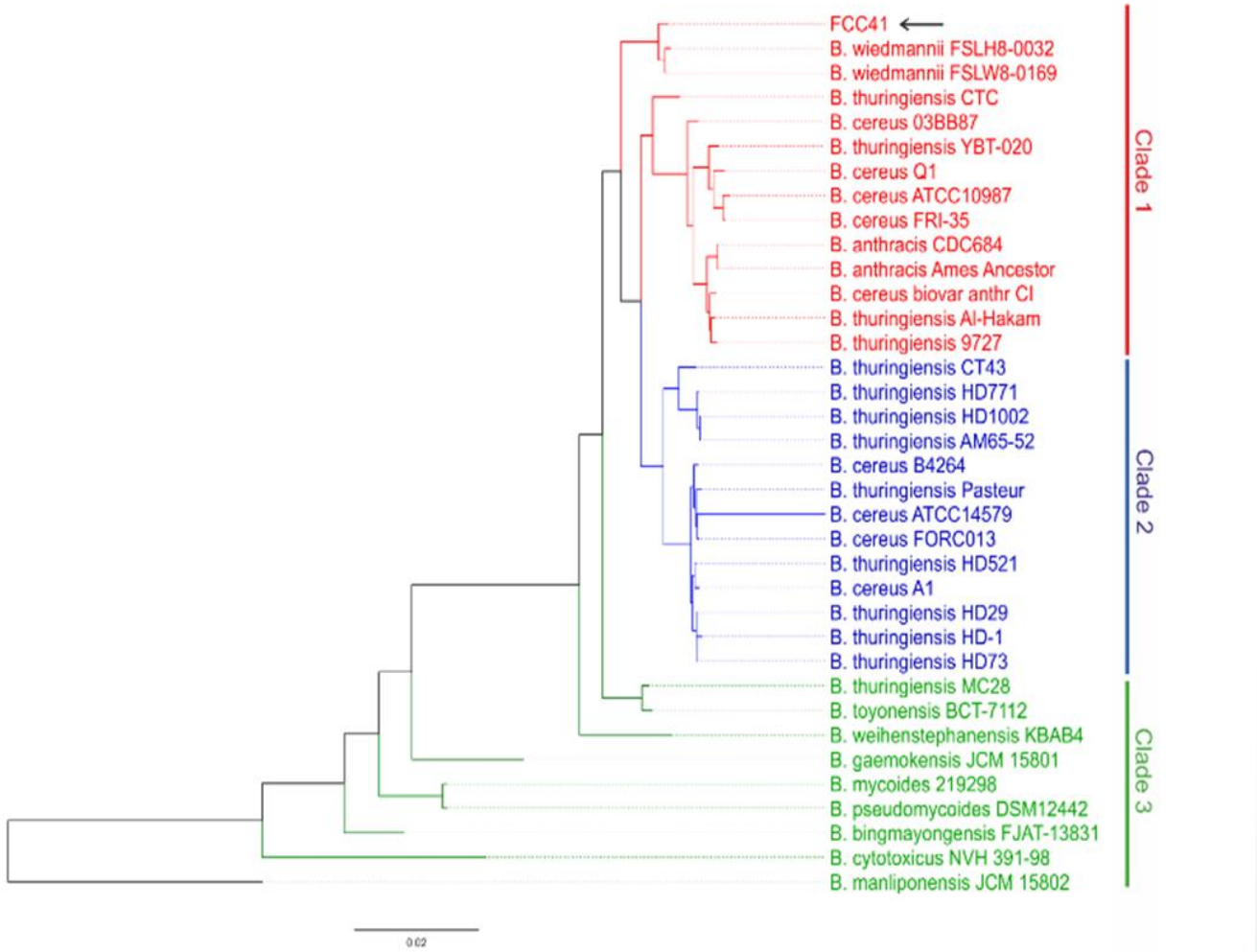
An ecotype classification approach using fAFLP and phenotypic properties divides the phylogeny of the *B. cereus* group into seven major groups

	Phylogenetic group	% strains with positive growth at :												
		4 °C	5 °C	7 °C	8 °C	10 °C	15 °C	20 °C	35 °C	37 °C	40 °C	43 °C	45 °C	50 °C	55 °C
Thermotolerant	VII	0	0	0	0	0	0	100	100	100	100	100	100	100	0
mesophilic	III	0	0	0	0	0	100	100	100	100	100	100	85	0	0
	IV	0	0	0	0	100	100	100	100	100	100	83	58	0	0
	I	0	0	0	0	75	100	100	100	100	100	25	0	0	0
	V	0	0	0	14	100	100	100	100	100	100	0	0	0	0
Psychrotolerant	II	0	0	73	87	100	100	100	100	100	100	0	0	0	0
	VI	0	40	100	100	100	100	100	100	86	0	0	0	0	0



Guinebretière *et al.*, 2008

Delineating bacterial species Genome BLAST distance phylogeny (GBDP) and average nucleotide identity (ANI)

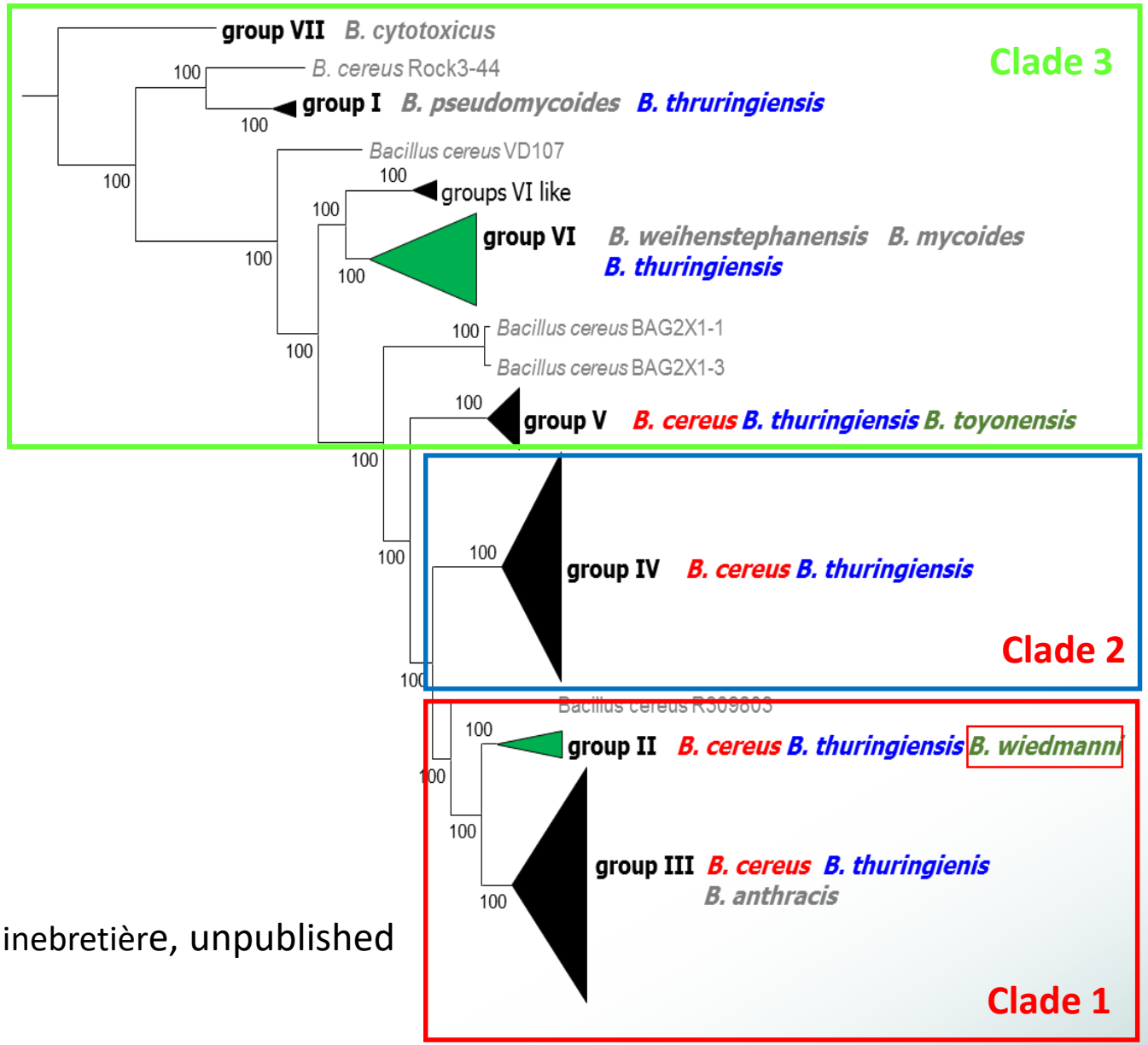


Core genome (bcgTree) phylogenetic tree obtained using 107 essential housekeeping genes and 35 B. cereus s. l. genomes

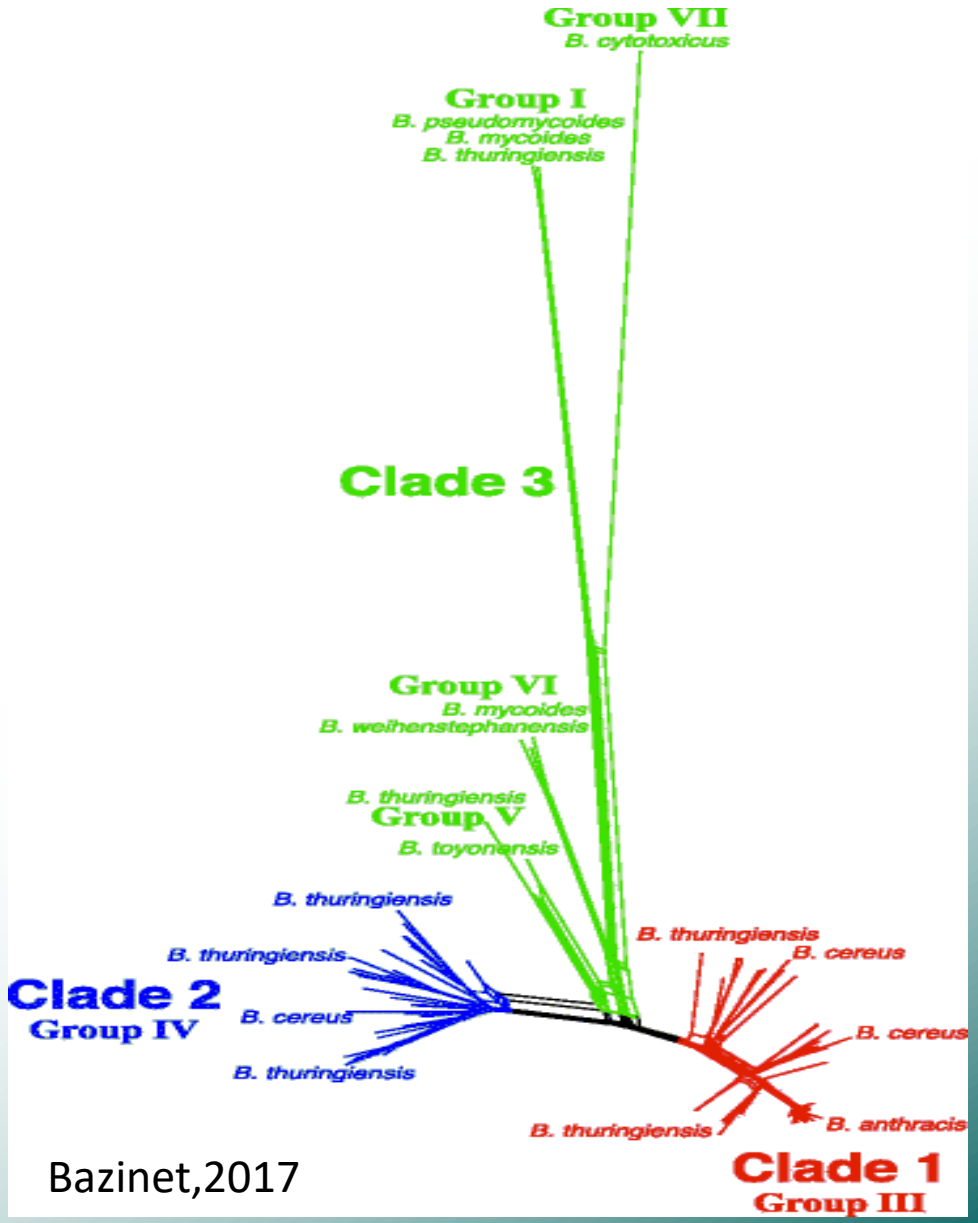
- several *B. cereus* strains in clade 1 have recently been re-designated *B. wiedmanii*

- Clade 3 is in reality a polyphyletic grouping and contains a greater diversity of species including *B. cereus*, *B. cytotoxicus*, *B. mycoides*, *B. thuringiensis*, *B. toyonensis*, and *B.weihenstephanensis* as well as *B. gaemokensis, B. manliponensis* and *B. bingmayongensis,* that have been recently proposed as representing novel species .

The phylogenetic and ecotypic structure distributed into seven major phylogenetic subdivisions is coherent with the 3 clades characterized in the group

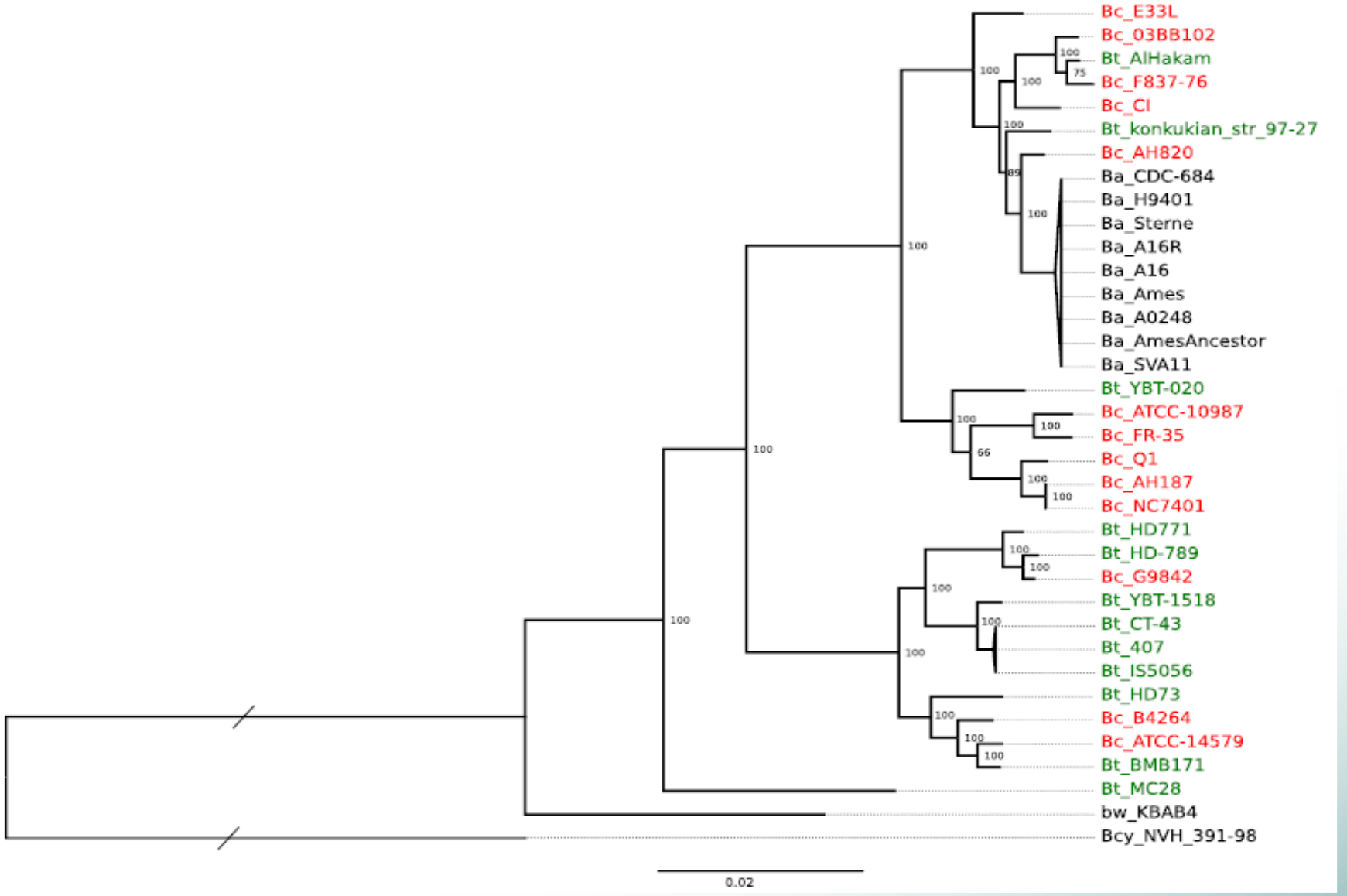


Guinebretière, unpublished



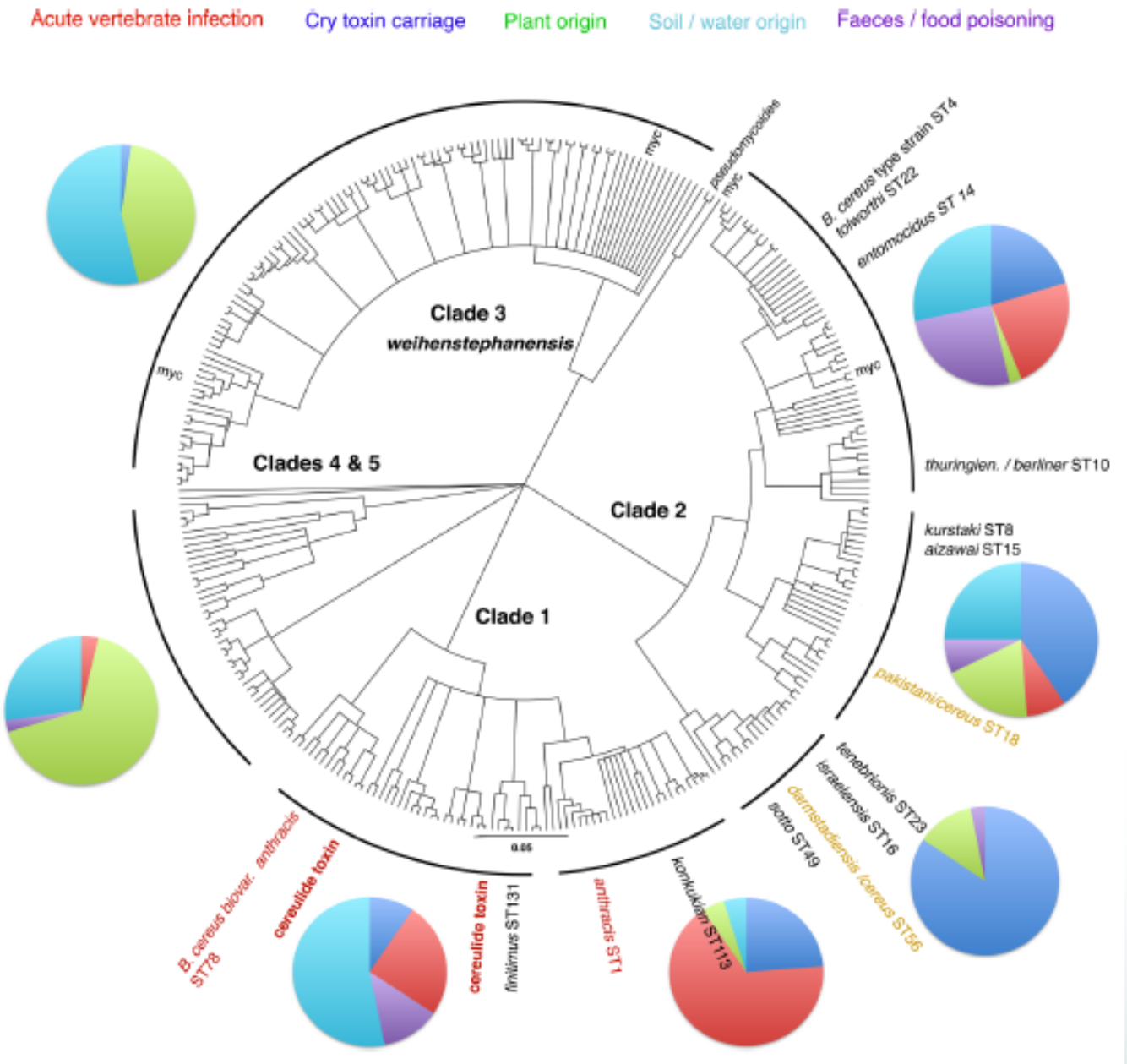
Bazinet,2017

Maximum-likelihood phylogenetic reconstruction of 36 *B. cereus sensu lato* strains using a concatenated 2 125 single copy proteins orthologs shared by all strains



presence or absence of the virulence plasmids and the *cry* gene is largely uncorrelated with the phylogenetic position of the host bacteria.

Host and habitat association across distinct clades in the *B. cereus* group



- The major MLST clades have different patterns of host association or varying ability to cause food poisoning

- 72% of the 548 isolates in the *anthracis* clade (clade 1) have been associated with vertebrate infections

- 29% of the 866 isolates in clade 2 are associated with vertebrate infections

Raymond and Federici, 2017

Delineating bacterial species using digital DNA:DNA hybridization (dDDH)

- Liu et al in 2015 claimed between 19-20 new « putative » species using *in silico* DNA–DNA hybridization (DDH) conducted using the Genome-to-Genome Distance Calculator and a species cut-off of 70 %.

- A first difficulty using this type of method, based on theoretical calculations DDH, is that it is very sensitive to the theoretical delimitation line

- Another difficulty is how to eliminate everything that is wrong, including errors in the genome (such as genomes which are not very reliably sequenced or assembled, or incomplete genomes, etc.).

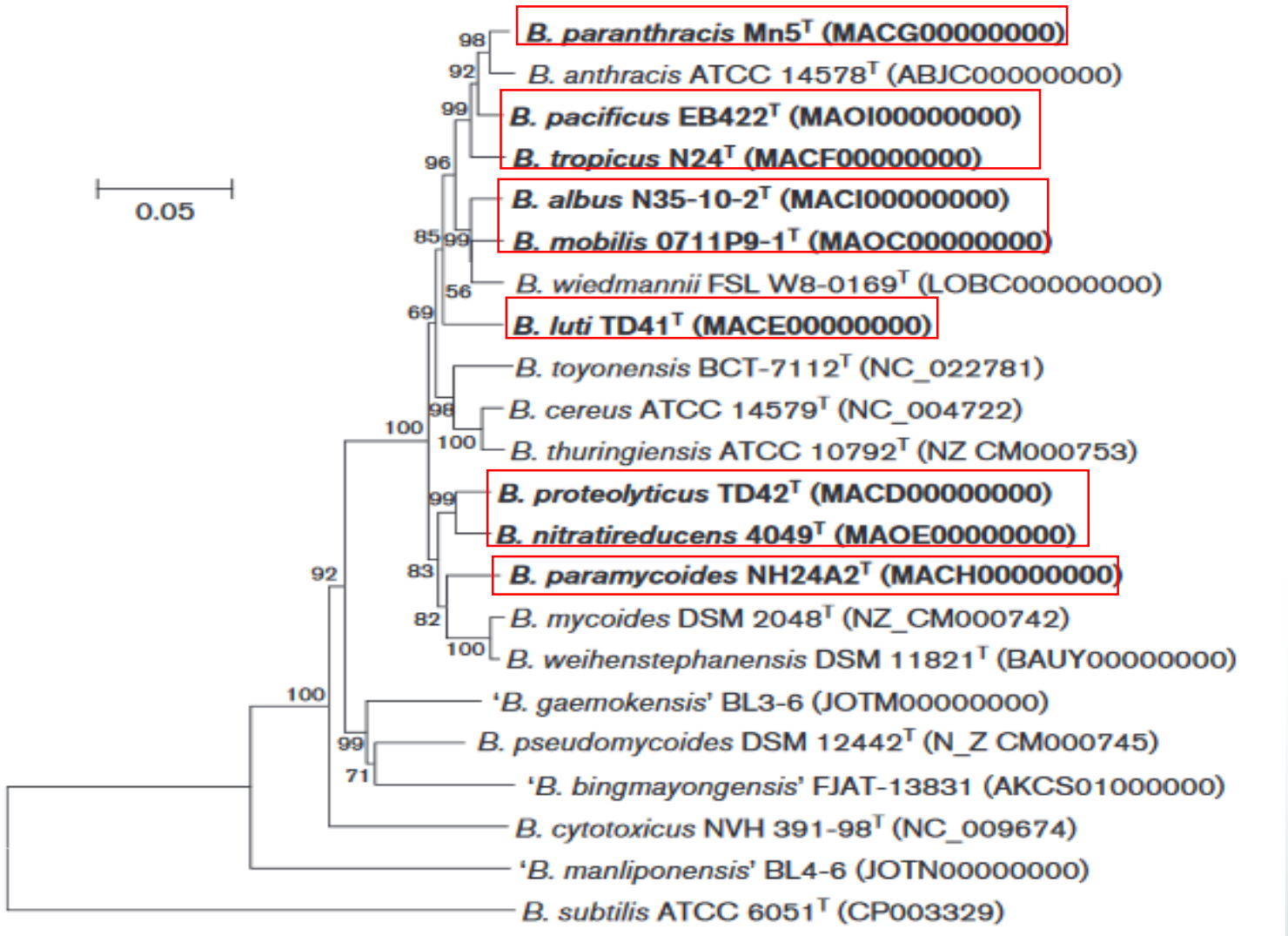
- And also how to take into account variations in genome size related to the presence of more or less large plasmids, and also the actual variation in the size of chromosome

- So, this approach can produce many potential new species ….. but very small phylogenetic groups could also be seen as subgroups within *B. cereus* group genomes.

- **My opinion is that using data on the complete genomes to calculate/predict values of DNA-DNA hybridization, to delimit species *B. cereus* group, is to take the problem upside down and does not seem a very relevant and sound approach.**

Proposal of nine novel species of the *Bacillus cereus* group

Liu et al., Int J Syst Evol Microbiol, 2017



Phylogenetic positions of the nine novel strains and other type strains of the *B. cereus* group based on seven concatenated housekeeping gene sequences

The digital DNA-DNA hybridization (dDDH) and average nucleotide identity (ANI) values between the nine strains were, respectively, below the 70 and 96 % threshold values for species definition.

These nine « new » species are *: Bacillus paranthracis, Bacillus pacificus, Bacillus tropicus, Bacillus albus, Bacillus mobilis, Bacillus luti, Bacillus proteolyticus, Bacillus nitratireducens* and *Bacillus paramycoides*

Take home messages

- Bacterial species are defined based on arbitrary thresholds for phenotypic and/or genetic similarity that have no universal biological relevance. Species definitions within *B. cereus s. l.* are not currently based on phylogenetic relatedness, but rather on phenotypes such as virulence, physiology and morphology.

- The 16S rRNA gene is unable to effectively distinguish between the closely related species within the *B. cereus* group and is therefore not suitable for the classification of the *B. cereus*

- Phylogenetic analyses (MLST, AFLP and WGS studies) indicate that there are three main clades that appear to be conserved, irrespective of the data source or analysis methodology used.

- AFLP and MLST with seven concatenated housekeeping gene sequences can reliably classify *B. cereus* group isolates into major phylogenetic groups (I to VII) almost as accurately as whole-genome comparisons.

- Therefore, the basic phylogenetic structure of *B. cereus s. l.* can be accurately computed by relatively quick phylogenetic analyses based solely on the distribution of accessory genes.

- Delineation of bacterial species using whole-genome sequence-based Genome BLAST Distance Phylogeny (GBDP) approach and/or *in silico* DNA–DNA hybridization (DDH) suggests that the *B. cereus* group can be divided into many additional clusters, and accordingly several novel putative species have been proposed.

- However, the species definition within the *B. cereus s. l.* is only useful when associated with phenotypes that are relevant for humans (e.g. for doctors / veterinary / agronomists) such as causing a specific disease in humans or animals and if combined with knowledge about the environmental conditions and pathogenicity.

Important questions that remain to be answered

In the case of the *B. cereus* group

- Can we unequivocally distinguish the different species that make up this bacterial group?

- Can a correlation be established between a given genotype and a pathotype?

- Are there genes, potentially involved in virulence that are specific to the strains that are particularly virulent for humans?

- Can we design specific and unequivocal tools/markers to distinguish between the different members or strains of the *B. cereus* group and distinguish between “harmless” strains and those that are toxic to humans?

- Can phylogeny be a good indicator of infection or food-poisoning risk for vertebrates

- Can we estimate the intensity of genetic exchanges within the *B. cereus* group and measure their impact on the emergence of pathogenic clones?

Fine-scale analysis of the genetic structure of sympatric populations of *Bacillus thuringiensis* and *Bacillus cereus*

- Genetic exchange has been little studied in natural populations of bacteria of the spore-forming *Bacillus cereus* group. Such data are required notably because several bacteria from this group are involved in human pathogenicity whereas others are extensively used in agricultural pest control.

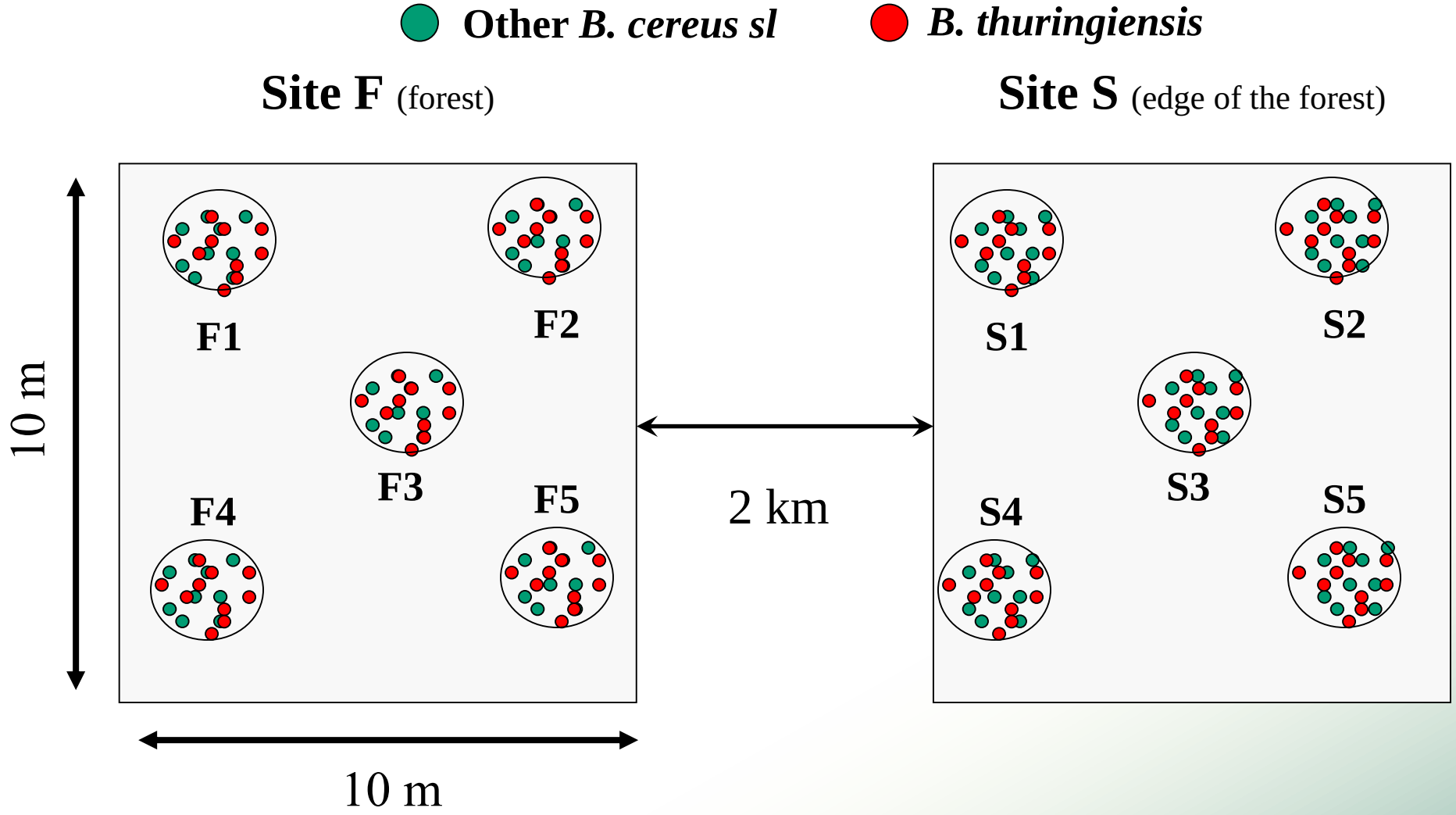
- The biological consequences of recombination (the generation of novel genotypes) are dependent on ecological factors; different strains of a species must be present within the same niche for genetic exchange to have any chance of leading to genetic variation

- Recombination in bacteria is qualitatively different from that in higher organisms, as it involves the replacement of a small region of the chromosome (perhaps a few kilobases) with the corresponding region from another isolate of the species or, in some cases, from a closely related species

- Thus, local populations must be studied to determine the extent of genetic exchange within and between species in natural populations of bacteria

Our goal was to determine whether or not sympatric populations of *Bt* and *Bc* were genetically differentiated and to estimate the intensity of genetic exchanges within the *B. cereus* group

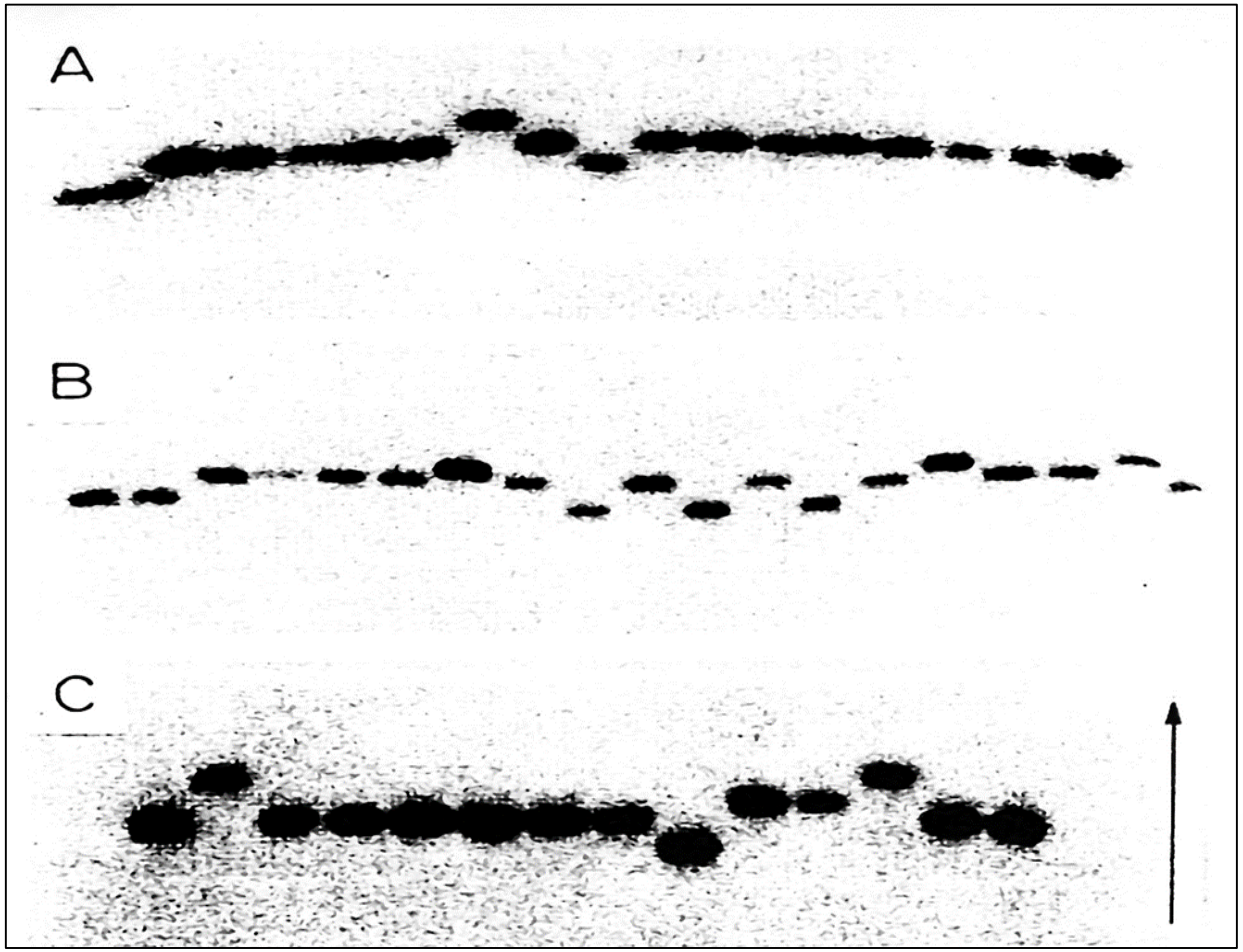
Sampling design and procedure

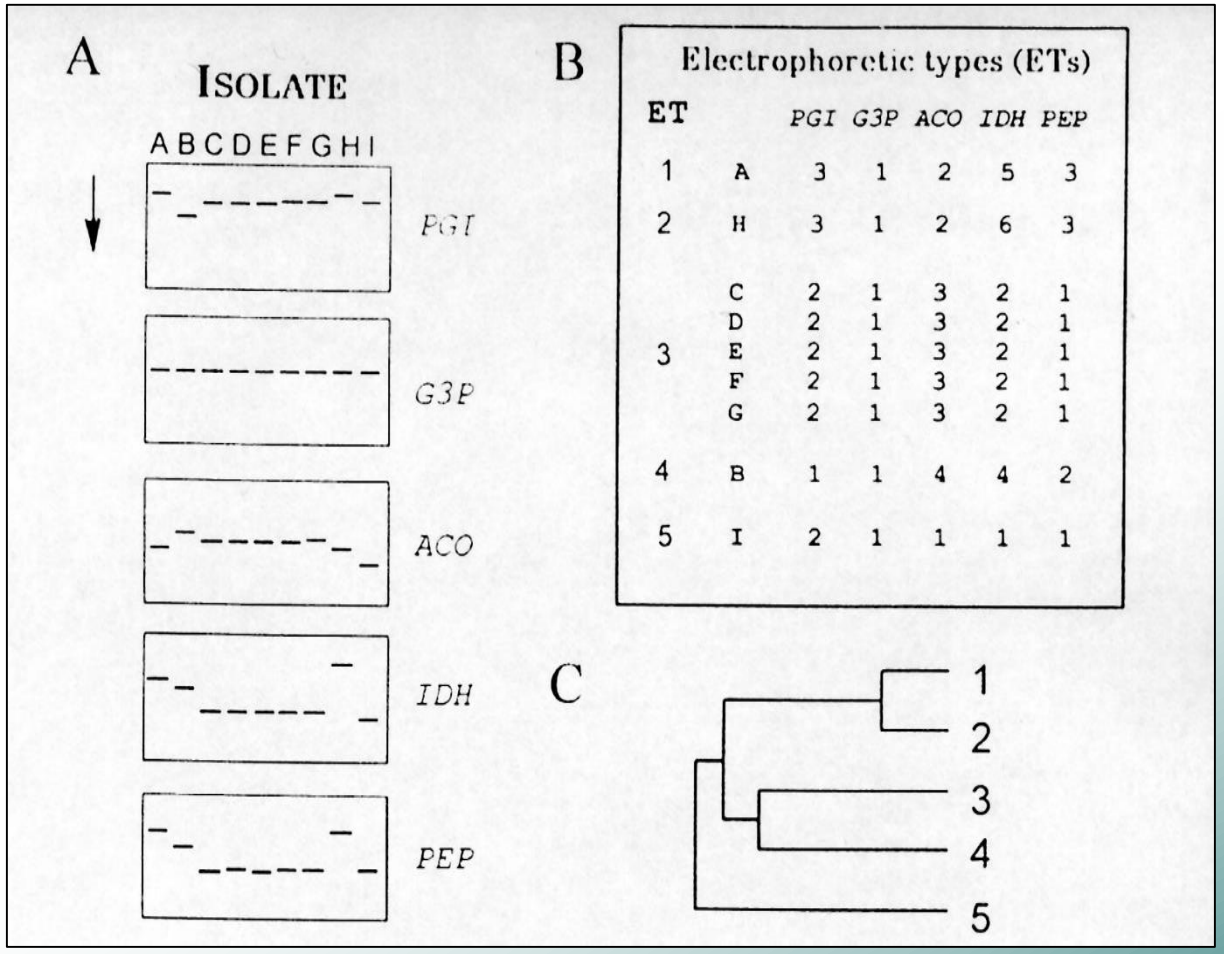


- 10 strains of each *Bacillus* species in each soil sample except in two samples where only 9 strains of Bt were recovered
- 100 strains of *Bc* and 98 of *Bt*

Multilocus enzyme analysis (MLEE)

Horizontal starch gel electrophoresis for allozyme analysis

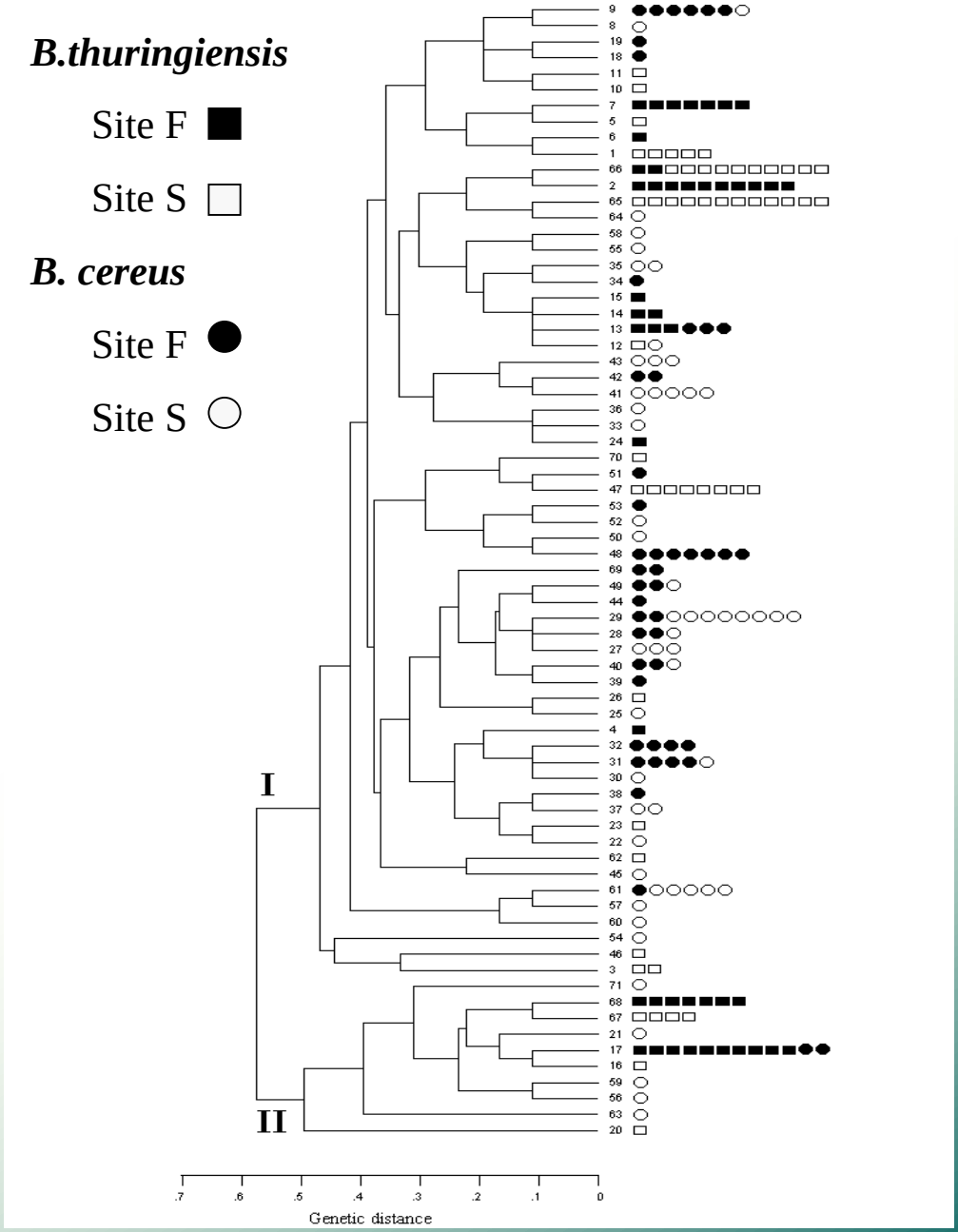




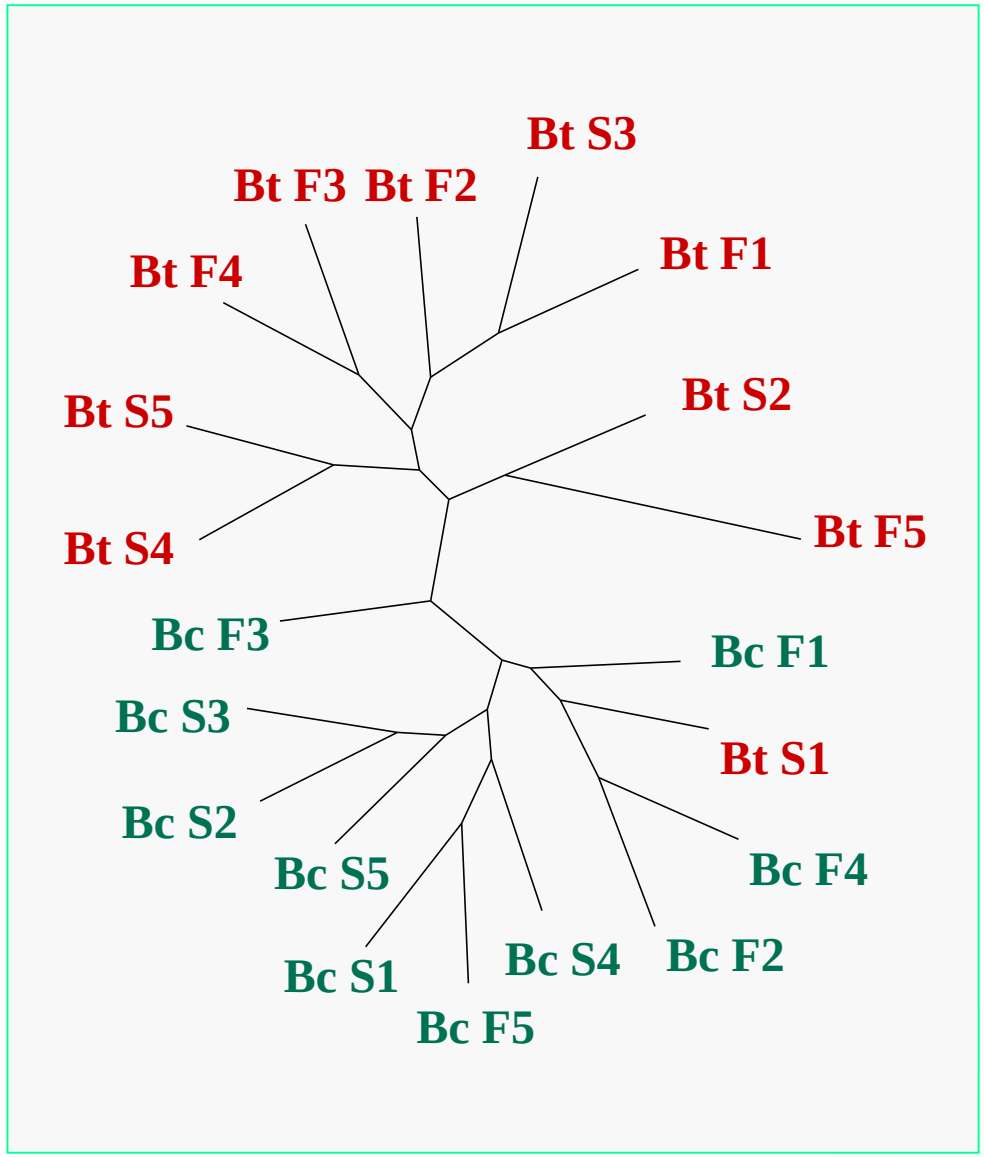
-This procedure detects those differences in the amino acid sequences of 9 metabolic enzymes that alter their electrophoretic mobility on starch gels.

Genetic diversity of local soil populations of *Bt* and *Bc*

- 7 out of the 9 Enzymatic loci were polymorphic
- At polymorphic loci, the number of alleles ranged from 2 to 6 alleles with a mean of 3.44 alleles per locus
- 71 different ETs in two main clusters
- At a genetic distance of 0.50, two clusters of ETs were identified, cluster I and cluster II, which included 61 and 10 ETs respectively



Genetic distances between soil samples of *Bt* and *Bc*

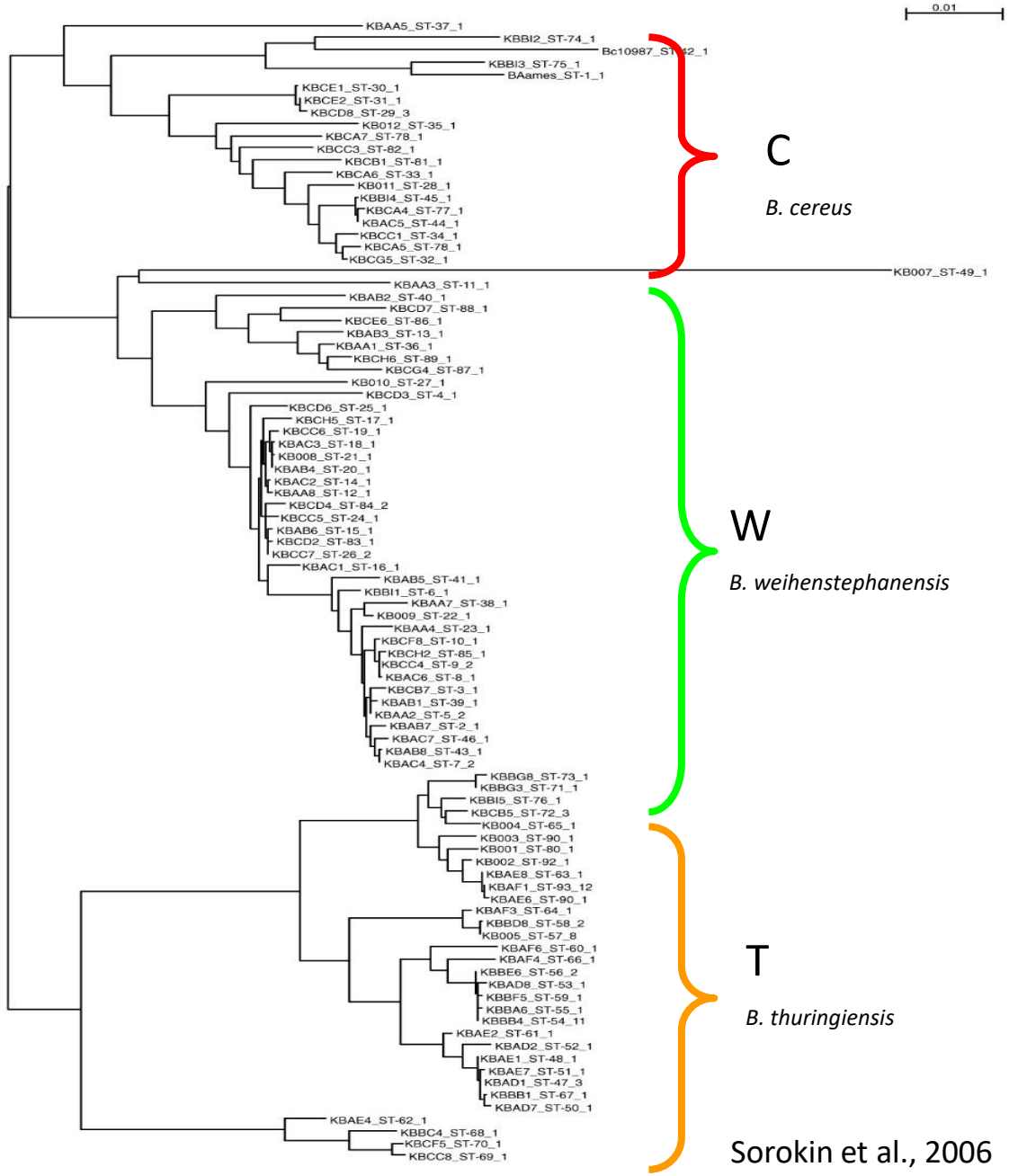


Samples of a given *Bacillus* species (*B. cereus* or *B. thuringiensis*) are genetically more similar to each other than to samples of the other species, even if isolated from the same soil sample.

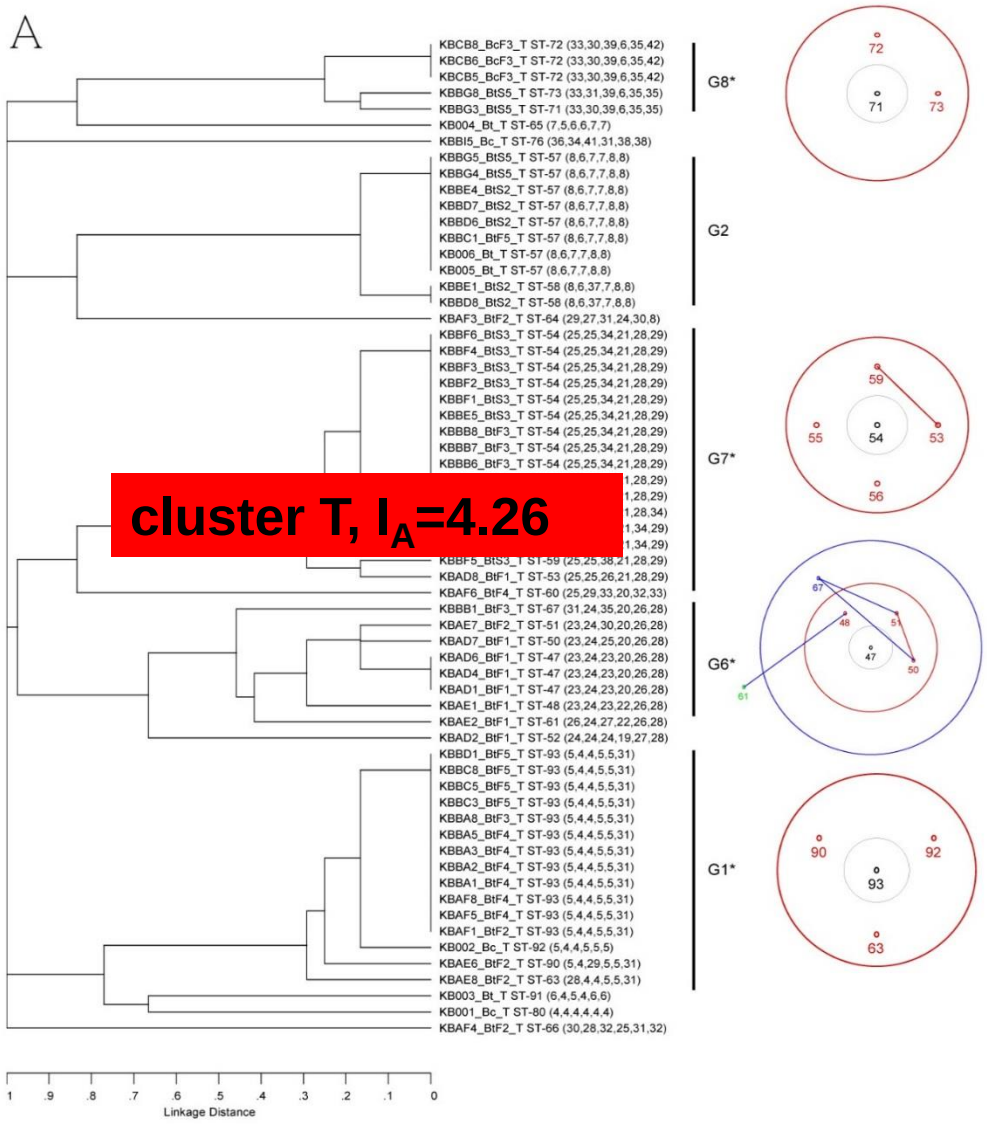
MLST studies of the *B. cereus* group strains collection isolated from the soil

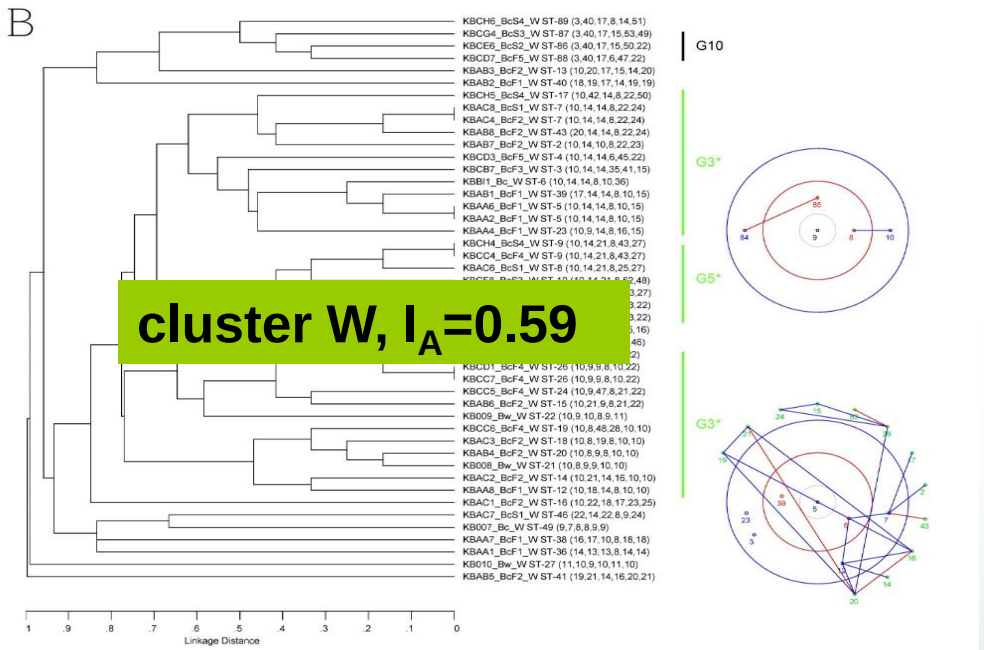
- This collection was constituted in a specific way to contain the strains isolated from the same soil samples and also from the samples collected in geographically distinct locations.
- 115 strains of the group *B. cereus* isolated from the forest soil near Versailles (“Versailles Collection”)
- Sequencing of six genes randomly distributed over the chromosome (*clpC*, *dinB*, *gdpD*, *panC*, *purF* and *yhfL*)
- MLSA scheme based on a concatenation and comparison of these 6 housekeeping genes sequences reveals 3 well phylogenetically separate clades designated C, W, T.

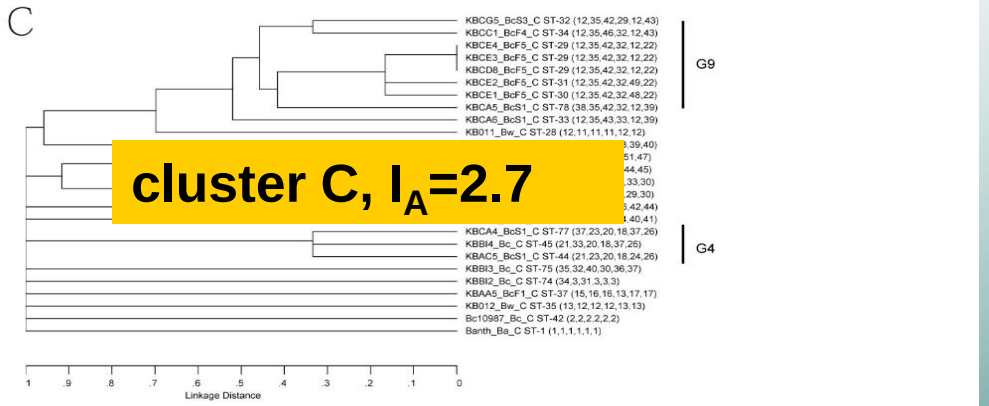
⇒ Most strains of the group W are psychrotrophs



T, W and C clusters fine structure comparison



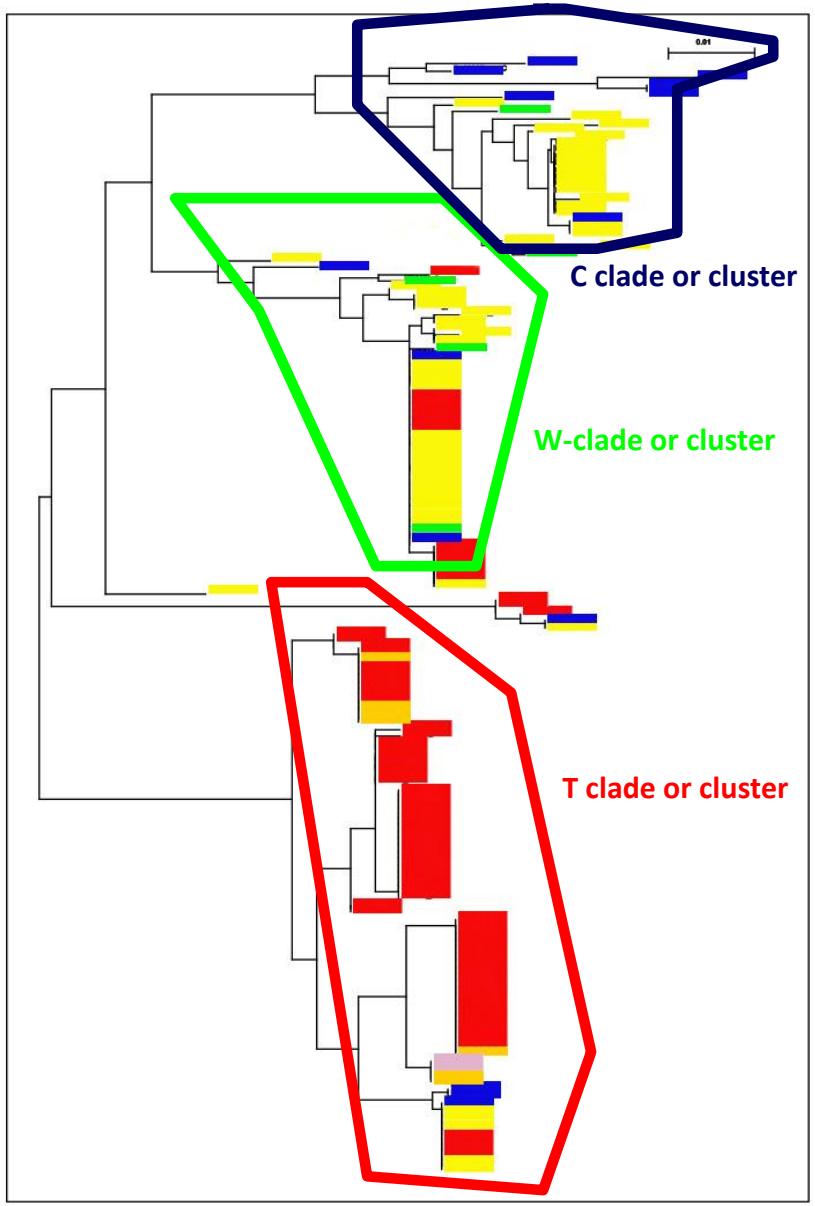




whole collection, I_A=2.917

Software START (Sequence Type Analysis and Recombinational Tests), Jolley et al, 2001

Results for *panC* gene clade and cluster contents



B. thuringiensis strains from Versailles forest (producing crystal toxin)

B.cereus strains from Versailles forest (non-producing crystal toxin)

proved human pathogens

proved insect pathogens or close type strains

B.weihenstephanensis strains (psychrotolerant, isolated in Germany)

approved animal probiotic strain

CONCLUSIONS :

1. At least three strain clades or clusters (species?) can be identified

2. *B. cereus* (no crystal toxin producing) and *B. thuringiensis* (producing crystal toxin) are clearly distinguishable, although the toxin transfer is possible

3. Human pathogenic strains are mostly in the *cereus* clade or cluster

Thank you for your attention

and your interest

Index of association (Maynard Smith *et al*. 1993)

A useful statistical test to provide a benchmark for the degree of clonality within a population is the index of association, a multi-locus measure of ***linkage disequilibrium*** (the non-random association of alleles).

$$I_A = (V_O/V_E) - 1$$

- V_O is the observed variance in the distribution of allelic mismatches in all pair-wise comparisons of the allelic profiles
- V_E the expected variance in a freely recombining population (linkage equilibrium)

In a recombining population, two isolates randomly recovered from the environment would be extremely unlikely to have identical, or even closely related, allelic profiles by MLEE or MLST, and the population is said at ***linkage equilibrium***

For populations at linkage equilibrium, V_O equals V_E and I_A has an expected value of zero

- A value of I_A significantly different from zero indicates a clonal structure
- A value of I_A not significantly different from zero indicates sexuality