

2 Overview of *Bacillus* Genomics

2.1 General characteristics of *Bacillus* genomes

Bacillus species are known for their diverse metabolic capabilities and adaptability to various environments. The genomes of *Bacillus* species typically range in size from approximately 5 to 6.5 million base pairs (bp) and have a G+C content around 35%-36% (Liu et al., 2017). For instance, the genome of *Bacillus thuringiensis* strain BM-BT15426 is 5 246 329 bp long with a G+C content of 35.40% (Liu et al., 2017), while *Bacillus thuringiensis* HM-311 has a genome size of 6,019,481 bp and a G+C content of 35.85%. These genomes often contain numerous plasmids, which carry genes responsible for various functions, including virulence factors and antibiotic resistance (Reyaz et al., 2019; Zuo et al., 2020).

2.2 Sequencing technologies and methods

The sequencing of *Bacillus* genomes has evolved significantly with advancements in sequencing technologies. Early sequencing efforts involved labor-intensive methods, but modern techniques such as PacBio RS II and Illumina HiSeq 4000 platforms have streamlined the process, allowing for high-throughput and accurate sequencing (Liu et al., 2017; Zuo et al., 2020). For example, the genome of *Bacillus thuringiensis* strain BM-BT15426 was sequenced using the PacBio RS II sequencer, which provides long-read sequencing capabilities, and assembled de novo using the HGAP pipeline (Liu et al., 2017). Similarly, *Bacillus thuringiensis* HM-311 was sequenced using both PacBio RS II and Illumina HiSeq 4000 platforms, combining long-read and short-read sequencing to achieve a comprehensive genome assembly (Zuo et al., 2020).

2.3. Genome annotation and analysis

Genome annotation is a critical step in understanding the functional capabilities of *Bacillus* species. Automated annotation tools such as the RAST server and NCBI's annotation pipeline are commonly used to predict coding sequences, identify genes, and assign functions based on homology (Quan et al., 2016; Reyaz et al., 2019). For instance, the genome of *Bacillus thuringiensis* X022 was annotated using the RAST server, which identified genes coding for insecticidal proteins and metabolic pathways (Quan et al., 2016). Similarly, the genome of *Bacillus thuringiensis* isolate T414 was annotated to reveal the presence of parasporal crystal protein genes and various virulence factors (Reyaz et al., 2019).

Comparative genomics further enhances our understanding by identifying core and pan-genomes, which distinguish conserved genes from those unique to specific strains. This approach has revealed significant functional differences among *Bacillus* species, such as variations in carbohydrate utilization and signal transduction pathways (Alcaraz et al., 2010). Additionally, comparative analysis of genomic and proteomic data can provide insights into gene expression and regulation, as demonstrated in studies of *Bacillus thuringiensis* strains (Rang et al., 2015; Quan et al., 2016).

3 Genomic Features of Bt

3.1 Genome structure and organization

Bacillus thuringiensis (Bt) exhibits a complex genome structure characterized by a circular chromosome and multiple plasmids. For instance, the Bt GR007 strain has a circular chromosome and three megaplasms, with the two largest megaplasms containing multiple pesticidal protein genes (Pacheco et al., 2021). Similarly, the Bt isolate T414 contains a chromosome and 15 different plasmids, which harbor various insecticidal genes (Reyaz et al., 2019). The genome of Bt X022 consists of one circular chromosomal DNA and seven plasmids, which include genes coding for several Cry proteins and a vegetative insecticidal protein (Quan et al., 2016).

3.2 Toxin gene clusters

3.2.1 Types of toxins

Bt produces a wide variety of insecticidal proteins, including *Cry*, *Cyt*, and *Vip* toxins. The GR007 strain, for example, contains 10 *cry* genes, two *vip* genes, and two binary toxin genes (Pacheco et al., 2021). The T414 strain harbors parasporal crystal protein genes (*cryIAa*, *cryIAb*, *cryIAc*, *cryIIAa*, *cry2Aa*, *cry2Ab*, and *cytI*) and a vegetative insecticidal protein gene (*vip3Aa*) (Reyaz et al., 2019). Additionally, the X022 strain contains genes coding for *CryIAc*, *CryIIa*, *Cry2Ab*, and *Vip3A* proteins (Quan et al., 2016).