https://dx.doi.org/10.4314/dujopas.v11i2a.31
ISSN (Print): 2476-8316
ISSN (Online): 2635-3490
Dutse Journal of Pure and Applied Sciences (DUJOPAS), *Vol. 11 No. 2a June 2025*

Genome Mining of *B. subtilis* Group for the Identification of Antibiotic and Secondary Metabolite Biosynthetic Gene Clusters

Tasiu Mahmud[1], Ibrahim Alhaji Sabo[2] Adamu Abdullahi Shehu[1] Yahaya Ubah Ya'u[3], Farida Isah El-hassan[1], Zakari Nuhu Lambu[3] Ibrahim Adamu Karfi[1]

[1]Department of Microbiology,
Faculty of Science
Aliko Dangote University of Science and Technology,
Wudil.

[2]Department of Microbiology,
Faculty of Pure and Applied Sciences,
Federal University Wukari,
Nigeria.

[3]Department of Science Laboratory Technology,
Faculty of Science,
Aliko Dangote University of Science and Technology,
Wudil.

Email: tasiu.microlabs@gmail.com

Abstract

Genome mining using bioinformatic tools has drastically increased the rate of discovery of specialized metabolite compared to traditional methods which rely on the isolation of individual compounds. Bacillus subtilis group devoted large portions of their genomes to the genes necessary for natural products biosynthesis. B.subtilis group possess ability to synthesize wide range of natural products known as specialized metabolite which played important roles in the physiology and metabolism of the bacteria and are also used as antibiotics, anticancer, enzymes, etc. The synthesis of these special metabolites depends on the abundance and distributions of biosynthetic gene clusters (BGCs) which can be identified in bacterial genome sequence. In the present study, 11genomic data of Bacillus subtilis group were retrieved from GeneBank database of the National Center for Biotechnology Information. The study employed antiSMASH v4.0 with default setting, the detection strictness was set to ''relaxed'' that can allow the identification of known and unknown gene clusters and their distribution in the genome. The results of the analysis identified and reported total of 97 biosynthetic gene clusters across the 11 genomes, the clusters were classified into 22 different classes of natural product including non-ribosomal peptide synthetase (NRPS) with high abundance and distribution of 29 across the 11 genomes. Similarly, the study revealed 32 (32.89%) unknown clusters across the 11genomic data for which no similar specialized products could be identified. The unknown clusters identified across the genomes of the Bacillus subtilis group might be predicted to encode for novel natural products and offer new avenue for drug discovery and development.

Keywords: Antibiotics, antiSMASH, Bacillus, Clusters, Genomes.

**Author for Correspondence*

INTRODUCTION

Genome mining using bioinformatic tools has drastically increased the rate of discovery of specialized metabolites compared to traditional methods which rely on the isolation of individual compounds (Iqbal, 2021). Nearly 40% of secondary metabolites cannot be chemically synthesized or may be produced in low concentration due to some problems arising from several parameters like medium composition, pH, temperature, strain improvement, agitation and aeration, subsequently, the production of secondary metabolites is generally repressed in logarithmic phase and depressed in stationary phase (Tiwari and Rana, 2015). An alternative method of identification of microbial secondary metabolites involves coupling bioinformatics and functional genomics to identify or locate biosynthetic gene clusters which are organized group of two or more genes on the bacterial chromosomes that together encode for enzymes responsible for biosynthesis of specialized metabolites (Hu *et al*., 2020). One biosynthetic gene cluster encode for enzyme that catalyze the formation of one or several similar natural compounds, however, bacterial genome usually contains 20-40 biosynthetic gene clusters (Martina *et al*., 2017; Put *et al*., 2024). *Bacillus subtilis* group possess ability to synthesize wide range of natural products known as specialized metabolites or secondary metabolite which play important roles in the physiology and metabolism of the bacteria and are also used as antibiotics, anticancer, antiparasitic, enzymes, enzymes inhibitors, surfactants and siderophore (Harwood *et al*., 2018). Yet, the antibiotics potentials of *B. subtilis* group has been unexplored, because, they exclusively devoted large portions of their genomes to the genes necessary for natural products biosynthesis. *B.subtilis* group consist of Eight closely related *Bacillus* bacterial species, including *B. subtilis subsp. Subtilis*, *B. subtilis subsp.spizizenii*, *B. firmus, B. licheniformis*, *B. mojavensis*, *B. vallismortis*, *B. clausii*, *B. lentus, B. sonorensis*, *B. amyloliquefaciens and B. atrophaeus* (Steinke *et al*., 2021).They are gram-positive, ubiquitors, spore-forming aerobic bacterial species and they have been well recognized for their ability to produce antibiotics and specialize metabolites (Nannan *et al*.,2021). The novelty of genome mining approach includes easy and efficient identification of biosynthetic gene clusters and their distribution and classification from genome data using bioinformatic tools that can annotate and mine the bacterial genome data, several tools have been developed, including ClustScan,CLUSEAN, NP.searcher, antiSMASH, PRISM,etc. However, antiSMASH version 4.0 were commonly used to enhance rapid drugs discovery and to fully realize the antibiotic biosynthetic potentials of *B.subtilis* group including genes that may be silent under laboratory condition, thus bridging the gap encountered in the culture screening technique which cannot revealed silence biosynthetic gene clusters, therefore, this study can offer new avenue for drug discovery (Yin *et al*., 2023; Chavali and Rhee 2018).

MATERIALS AND METHOD

Data collection

The methods of Grubbs *et al*.(2017) was adopted for the data collection, publicly available genome sequences of eleven bacterial strains representing the *B.subtilis* group were retrieved from the NCBI GeneBank database (https://www.ncbi.nlm.nih.gov/). The bacterial strains include *B. subtilis subsp. Subtilis strain SRCM101441*(>CP021507.1), *B.subtilis subsp.spizizenii strain TU-B-10* (>CP002905.1), *B. firmus strain T8* (>CP086235.1), *B. licheniformis strain BL1202* (>CP017247.1),*B. mojavensis strain UCMB5075* (>CP0514364.1), *B. vallismortis strain NBIF-001*(>CP020893.1), *B. clausii strain KSM-K16 DNA*-(>AP006627.1), *B. lentus strain NCTC4824* (>LS483476.1), *B. sonorensis strain PMC204* (>NZ_CP139190.1), *B. amyloliquefaciens strain HM618* (>NZ_CP029466.1) *and B. atrophaeus strain MBLB1156* (>CP120846.1).

Antibiotics and secondary metabolites biosynthetic genes clusters prediction in *B.subtilis* group

To predict or locate biosynthetic gene clusters in each genome, antiSMASH v4.0 was used.The detection strictness was set to ''relaxed'' and all other important parameters such as ClusterBlast, KnownClusterBlast, SubClusterBlast, MIBiG Cluster comparisons, Activesite Finder, RREFinder Cluster Pfam analysis, Pfam-based GO term annotation, TFBS analysis, TIGRFam analysis were enabled with default settings. The antiSMASH analysis was determined and the summary of the annotated BGCs and region of each identified cluster was generated, analyzed and classified (Grubbs *et al*., 2017; Williams *et al*., 2020).

Results

Table 1: Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B.subtilis subsp.spizizenii strain TU-B-10*

Cluster type	Cluster location	Most similar known cluster	% of similarity
NRPS	(472,594-537,985)nt	surfactin, NRP:lipopeptide	82
Terpene	(1,264,681-1,285,487)nt	-	-
transAT-PKS,PKS-like, T3PKS,NRPS	(1,867,609-1,982,455)	bacillaene,polyketide +NRP	100
NRPS,betalactone, transAT-PKTS	(2,040,306-2,117,528)nt	mycosubtilin, NRP+PK	100
Terpene	(2,193,170-2,215,068)nt	-	-
T3PKS	(2,263,453-2,304,601)nt	-	-
NRP-metallophore,NRPS	(3,202,527-3,254,309)nt	bacillibactin, NRP	100
Lanthipeptide-class-I	(3,404,804-3,431,029)nt	subtilin RiPP:lanthipeptide	100
CDPS	(3,550,889-3,571,628)nt	pulcherriminic acid ,other	100
Sactipeptide	(3,813,374-3,834,983)nt	subtilosin A,RiPP:thiopeptide	100
Other	(3,838,573-3879,991)nt	bacilysin , other	100
Epipeptide	(4,108,424-4,130,122)nt	-	-

Key: NRPS-Non-ribosomal peptide synthethase, T3PKS- Type I polyketide synthase, NRP-Non-ribosomal peptide, CDPS-Cyclic dipeptide synthases, PK-Polyketide, RiPP-Ribosomally synthesize post-translationally modified peptide, nt- Nucleotide.

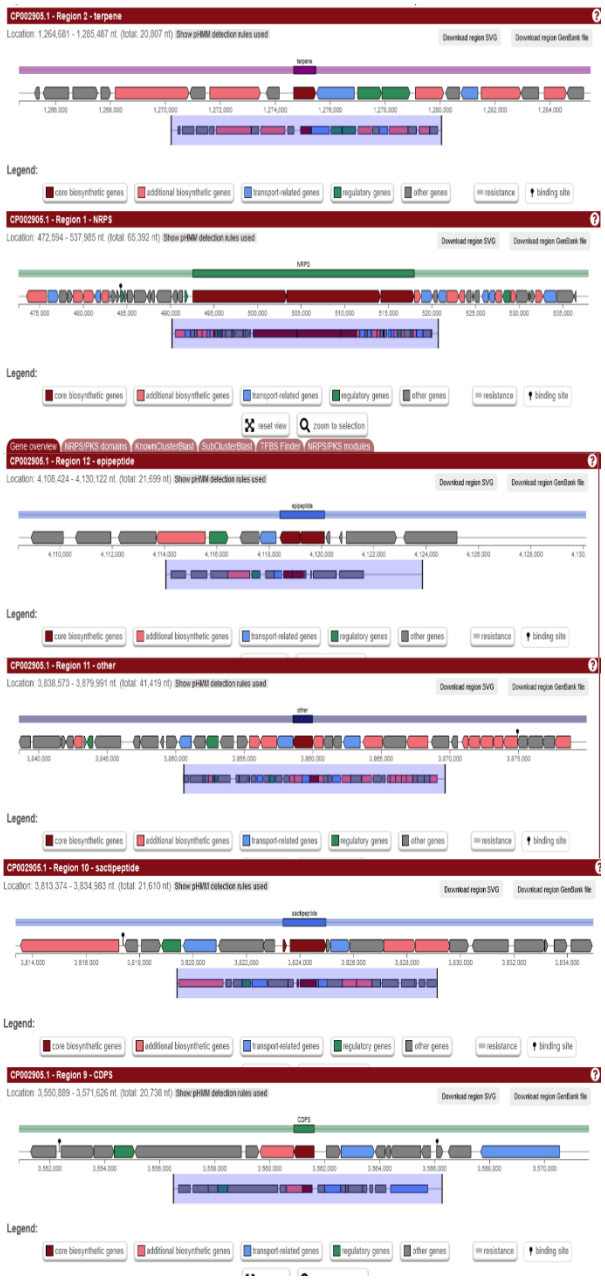

Fig1a.Antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) in *B.subtilis subsp.spizizenii subsp spizizenii strain TU-B-10 genomestrain TU-B-10* genome.

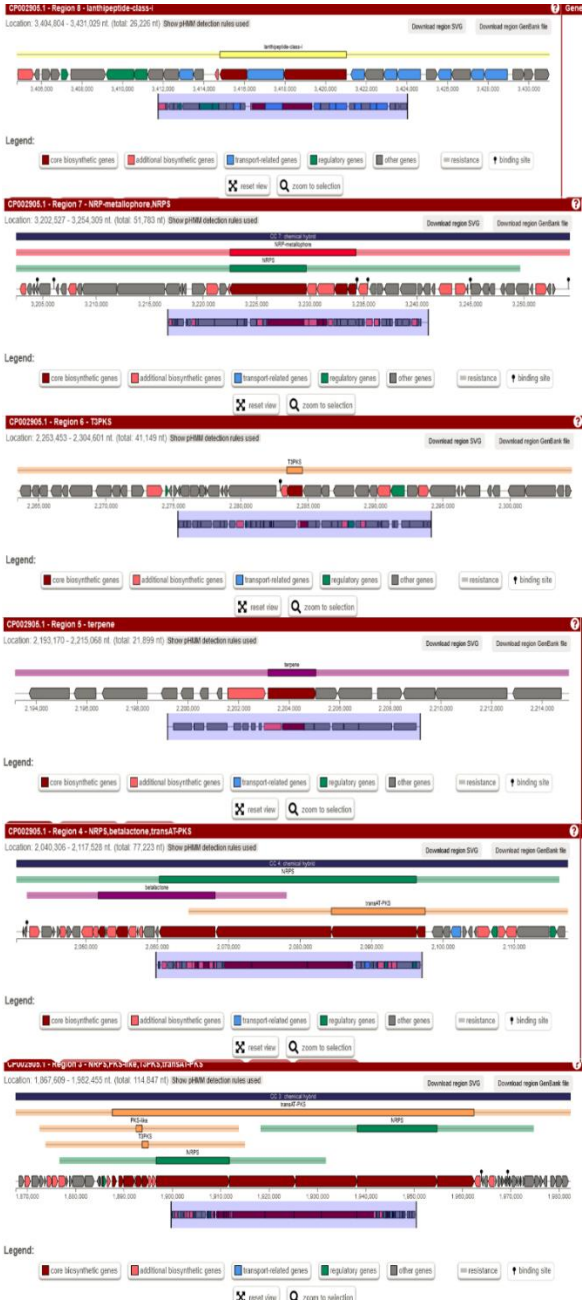

Fig.1b.Antibiotics and secondary metabolites biosynthetic gene clusters (BGCs)in *B.subtilis*

Table 2*:* Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B. subtilis subsp. Subtilis strain SRCM101441.*

Cluster type	Cluster location	Most similar known cluster	% of similarity
NRPS	(187,235-235,967)nt	locillomycin B and C, NRP+PK	21
NRPS	(389,937-455,328)nt	surfactin , NRP:lipopeptide	82
Terpene	(1,221,682-1,242479)nt	-	-
NRPS,Betalactone	(1,976,807-2,027,997)nt	fengycin, NRP	93
Terpene	(2,119,687-2,141,585)nt	-	-
T3PKS	(2,190,842-2,231,956)nt	1-carbapen-2-em-3-carboxylic acid,other	16

NRP-metallophore,NRPS	(3,135,824-3,187,590)nt	bacillibactin, NRP	100
Sactipeptide	(3,786,457-3,808,068)nt	subtilosin A, RiPP:thiopeptide	100

Key: NRPS-Non-ribosomal peptide synthethase, T3PKS- Type III polyketide synthase, NRP-Non-ribosomal peptide, PK-Polyketide, RiPP-Ribosomally synthesize post-translationally modified peptide, nt- Nucleotide.

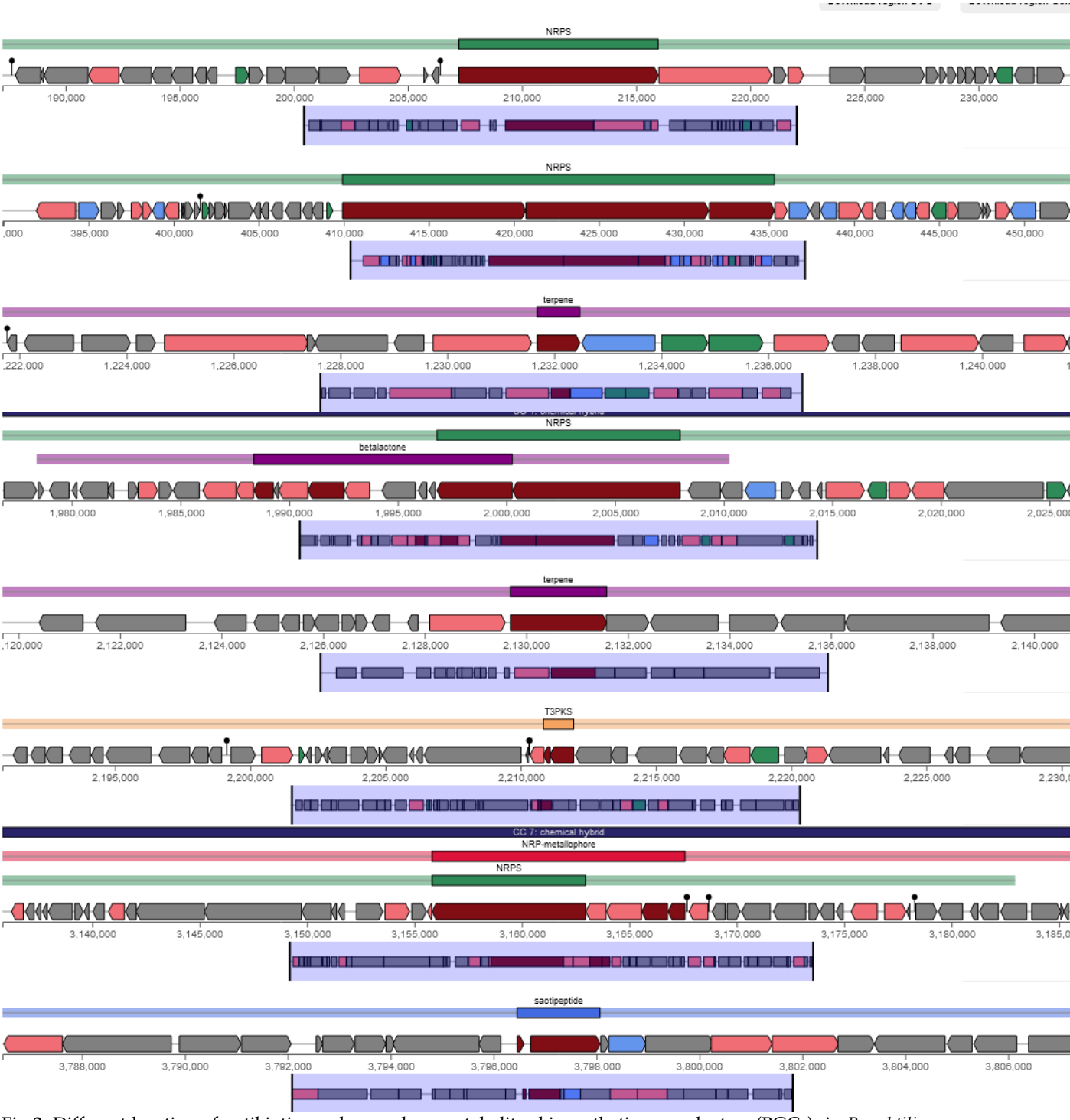

Fig 2. Different location of antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) in *B. subtilis subsp. Subtilis strain SRCM101441* genome.

Table 3: Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B. atrophaeus strain MBLB1156*,

Cluster type	Cluster location	Most similar known cluster	% of similarity
NRPS,lanthipeptide-class-i	(195,966-242,975)nt	subtilomycin, RiPP: lanthipeptide	100
NRPS	(366,670-432,081)nt	surfactin, NRP:lipopeptide	86
Terpene	(1,161,277-1,182,119)nt	--	
transAT-PKS,PKS-like,T3PKS,NRPS	(1,810,015-1,925,537)nt	bacilaene, PK+NRP	100
NRPS, Betalactone,transAT-PKS	(2,043,133-2,180,812)nt	fengycin, NRP	100
Terpene	(2,219,583-2,241,472)nt	--	
T3PKS	(2,315,112-2,356,209)nt	1-carbapen-2-em-3-carboxylic acid,other	16
NRP-metallophore,NRPS	(3,200,927-3,252,794)nt	bacillibactin NRP	100
Thiopeptide, LAP	(3,265,932-3,296,035)nt	-	
Sactipeptide	(3,766,725-3,788,334)nt	subtilosin A, RiPP:thiopeptide	87

Key: NRPS-Non-ribosomal peptide synthethase, T3PKS- Type III polyketide synthase, NRP-Non-ribosomal peptide, PK-Polyketide, RiPP-Ribosomally synthesize post-translationally modified peptide, nt- Nucleotide, LAP-linea azoline containing peptides.

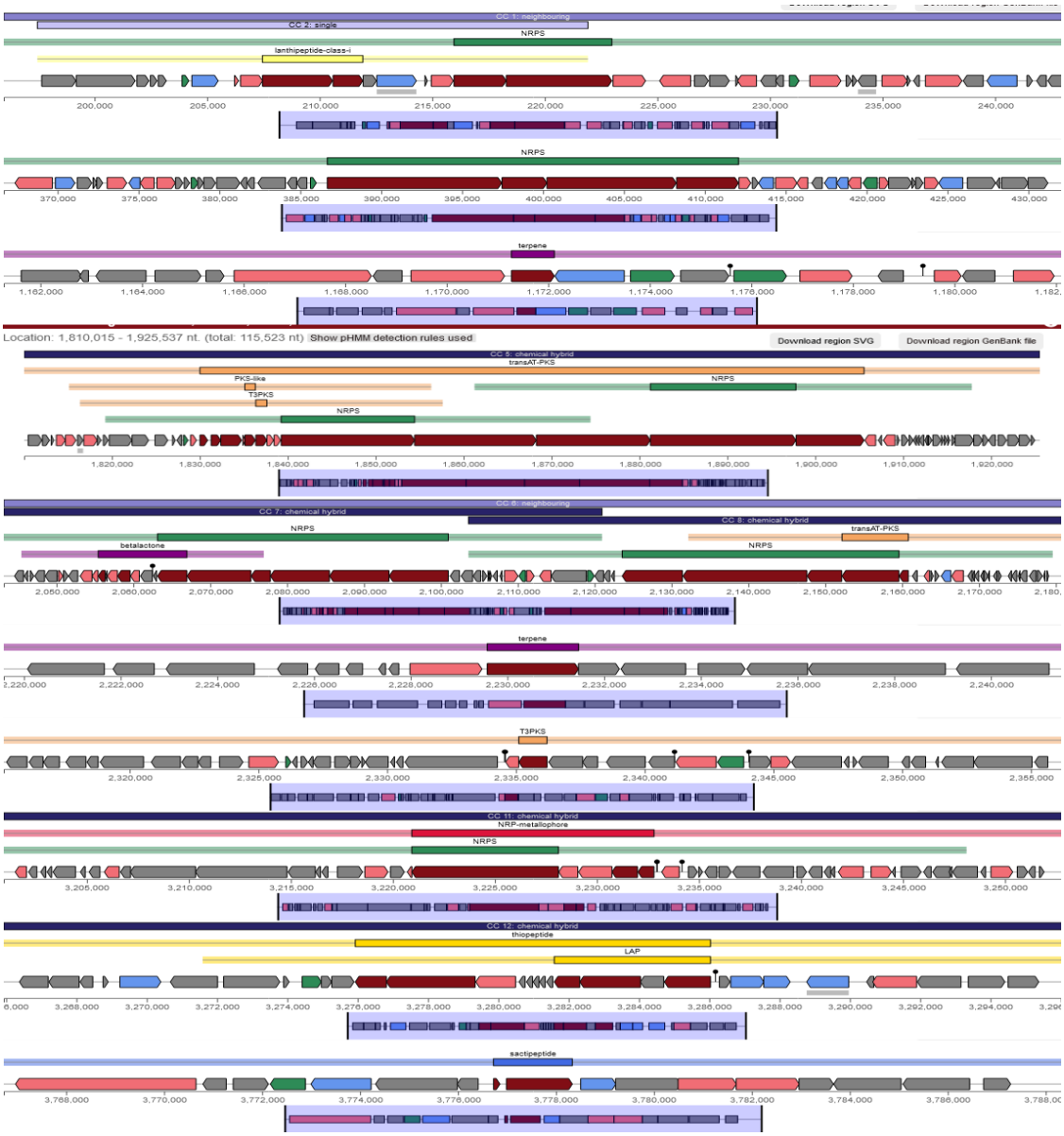

Fig 3. Different locations of antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) in *B. atrophaeus strain MBLB1156* genome.

Table 4: : Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B. clausii strain KSM-K16 DNA*

Cluster type	Cluster location	Most similar known cluster	% of similarity
Ectoine	(362,076-372,480)nt	ectoine, other	75
T3PKS	(1,066,322-1,107,422)nt	-	
		-	
NI-sideropho	(2,047,901-2,079,432)nt	-	
		-	
Lanthipeptide-class-i	(3,691,647-3,716,154)nt	-	
		-	

Key: T3PKS- Type III polyketide synthase, nt- Nucleotide.

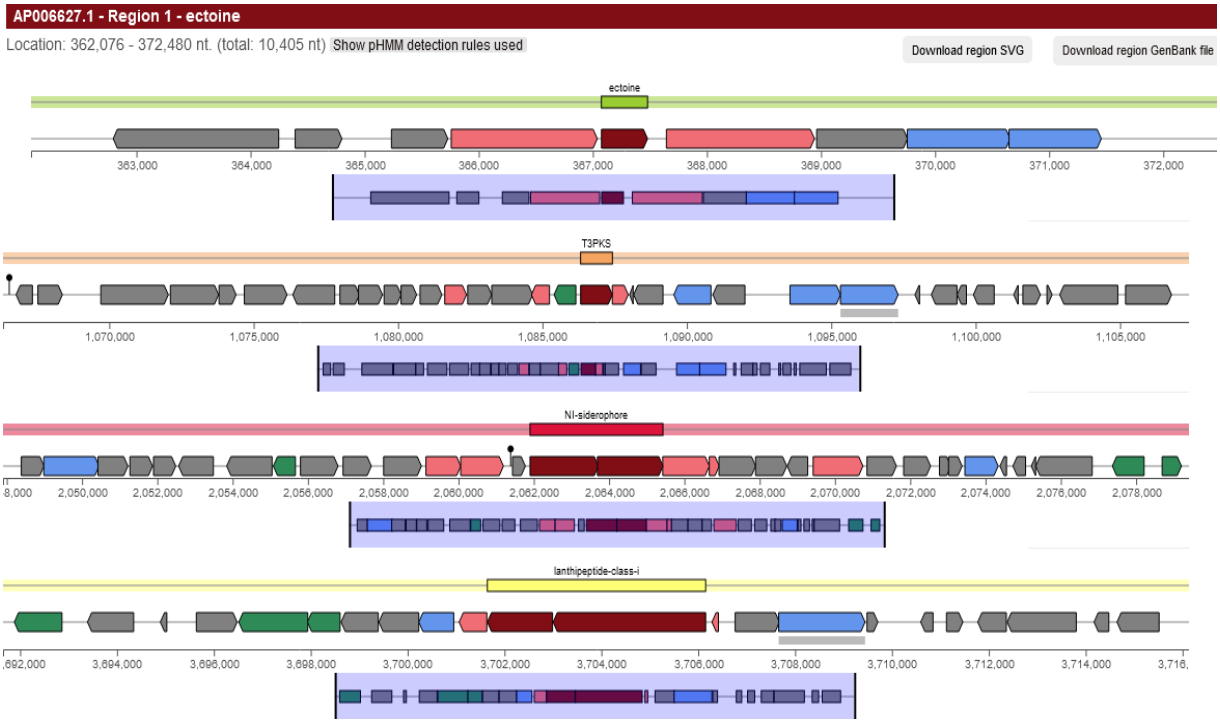

Fig 4.Different locations of antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) in *B. clausii strain KSM-K16* genome.

Table 5: Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B. lentus strain NCTC4824*

Cluster type	Cluster location	Most similar known cluster	% of similarity
Lassopeptide	(1,807,780-1,833,968)nt	paeninodin RiPP	100
T3PKS	(1,980,368-2,021,468)nt	-	
		-	
Terpene	(2,859,900-2,881,777)nt	surfactin, NRP:lipopeptide	8
NRPS	(4,108,399-4,152,124)nt	-	-

Key: NRPS-Non-ribosomal peptide synthethase, T3PKS- Type III polyketide synthase, NRP-Non-ribosomal peptide, RiPP-Ribosomally synthesize post-translationally modified peptide, nt- Nucleotide.

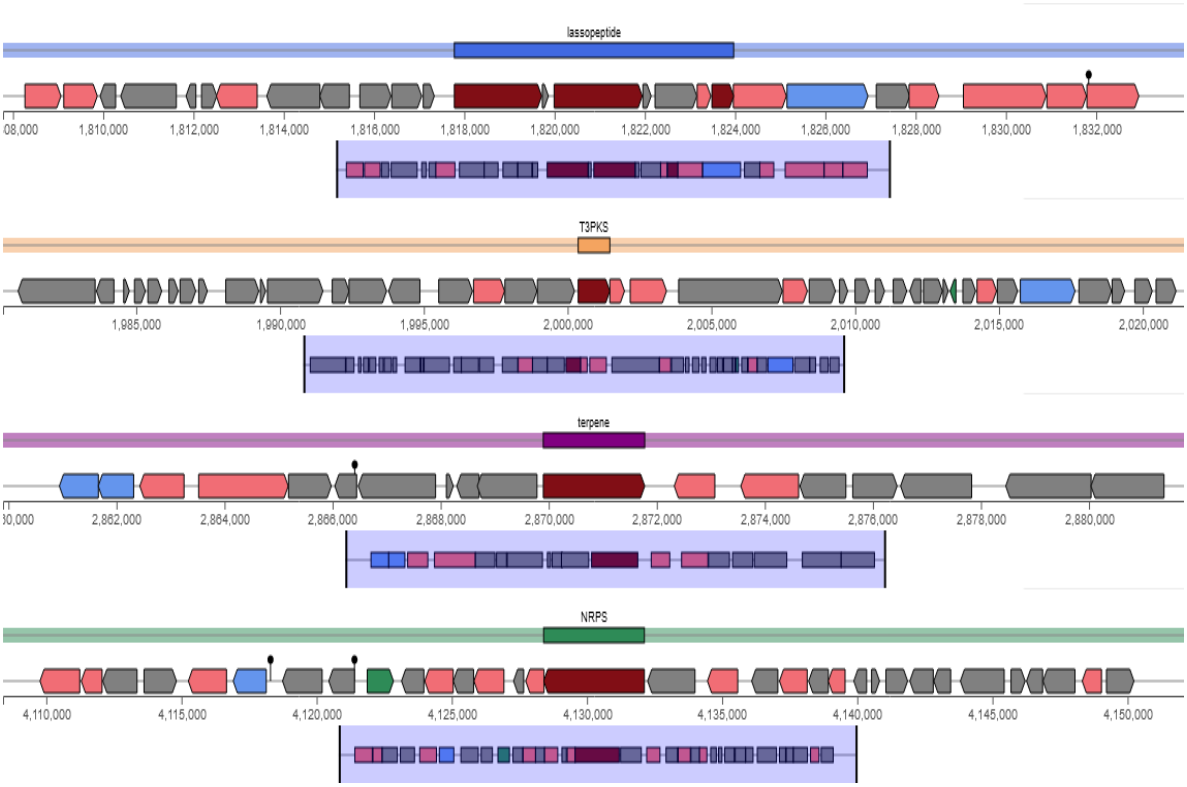

Fig 5. Different locations of antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) in *B. lentus strain NCTC4824*

Table 6*:* Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B. licheniformis strain BL1202* genome.

Cluster type	Cluster location	Most similar known cluster	% of similarity
NRPS	(375,652-441,092)nt	lichenysin, NRP	100
Thiopeptide,LAP	(1,151,366-1,192,681)nt	butirosin A/butirosin B, saccharide	7
NI-siderophore	(1,300,662-1,334,128)nt	schizokinen, other	60
Betalactone	(2,203,433-2,231,947)nt	fengycin, NRP	53
T3PKS	(2,402,539-2,448,555)nt	--	
Terpene	(2,402,539-2,351,731)nt	--	
CDPS	(3,619,929-3,640,678)nt	pulcherriminic acid, other	66
NRP-metallophore,NRPS	(3,917,462-3,969,206)nt	bacillibactin/bacillibactin E and F,NRP	100
Lassopeptide,RRE-containing	(4,071,649-4,094,382)nt	--	
Lanthipeptide-class-i	(4,158,272-4,185,233)nt	lichenicidin VK21A1,VK21A2, RiPP:lanthipeptide	100

Key: NRPS-Non-ribosomal peptide synthethase, T3PKS- type III polyketide synthase, NRP-Non-ribosomal peptide, PK-Polyketide, RiPP-Ribosomally synthesize post-translationally modified peptide, nt- Nucleotide,LAP-linea azoline containing peptides, CDPS-Cyclic dipeptides synthase.

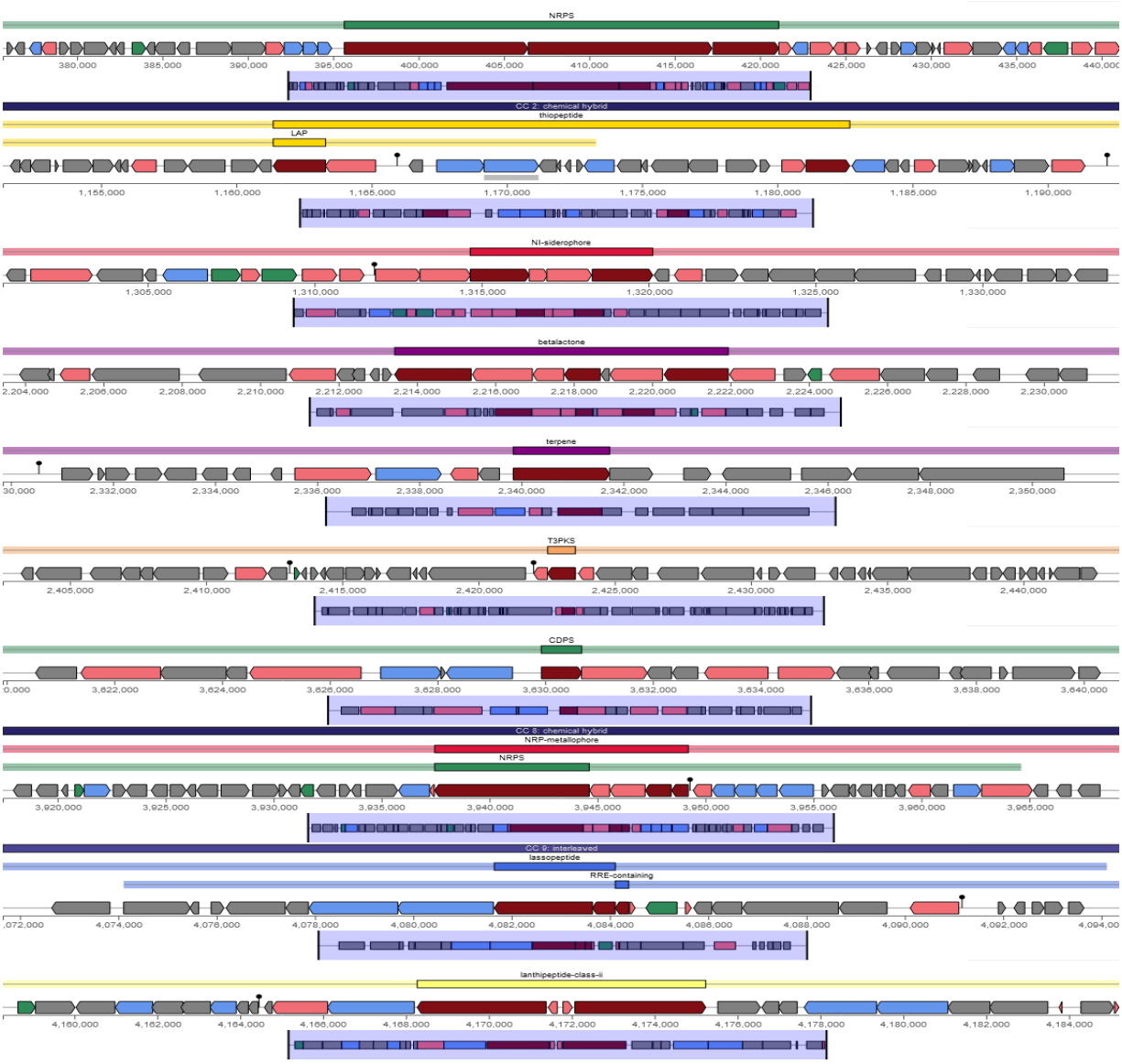

Fig 6. Different locations of antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) in from *B. licheniformis strain BL1202* genome.

Table 7:Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B. mojavensis strain UCMB5075*

Cluster type	Cluster location	Most similar known cluster	% of similarity
NRPS	(353,380-418,975)nt	surfactin,NRP:lipopeptide	78
Terpene	(1,153,601-1,174,415)nt	-	
		-	
NRPS,Betalactone	(1,869,837-1,947,616)nt	fengycin, NRP	100
Terpene	(2,025,801-2,047,699)nt	--	
T3PKS	(2,094,862-2,135,959)nt	laterocidine, NRP	5
NRP-metallophore,NRPS	(3,046,989-3,098,763)nt	bacillibactin, NRP	100
Lanthipeptide-class-i	(3,241,140-3,267,365)nt	subtitilin, RiPP:lanthipeptide	100
CDPS	(3,376,685-3,397,431)nt	pulcherriminic acid, other	100
Sactipeptide	(3,640,722-3,662,334)nt	subtilosin A RiPP:thiopeptide	100
Other	(3,665,123-3,706,540)nt	bacilysin, other	100

Key: NRPS-Non-ribosomal peptide synthethase, T3PKS- Type III polyketide synthase, NRP-Non-ribosomal peptide, PK-Polyketide, RiPP-Ribosomally synthesize post-translationally modified peptide, nt- Nucleotide, CDPS- Cyclic dipeptide synthase.

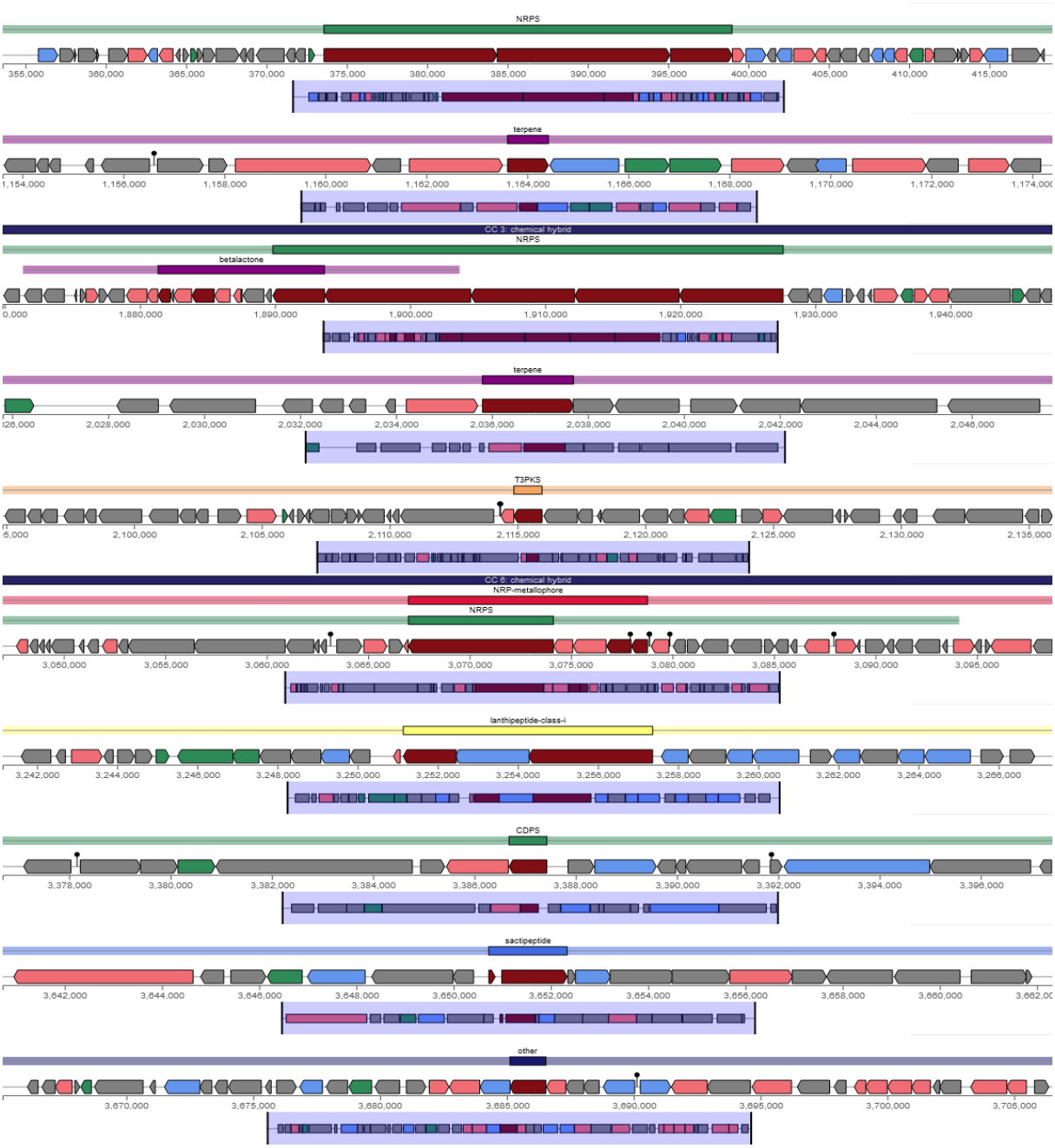

Fig 7. Different location of antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) in *B. mojavensis strain UCMB5075* genome.

Table 8: Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B. sonorensis strain PMC204*

Cluster type	Cluster location	Most similar known cluster	% of similarity
Cyclic-lactone-autoinducer	(157,146-177,907)nt	amyloliquecidin GF610, RiPP	60
Lanthipeptide-class-i	(219,876-244-289)nt	entianim, RiPP:lanthipeptide	18
NRPS	(411,405-476,840)nt	lichenysin, NRP	100
Thiopeptide,LAP	(1,226,423-1,267,610)nt	butirosin A/butirosin B, saccharide	7
NI-siderophore	(1,372,053-1,405,515)nt	schizokinen, other	60
Opine-like-metallophore	(1,998,385-2,020,525)nt	bacillpaline, other	100
NRPS,T1PKS	(2,033,033-2,131,409)nt	plipastatin, NRP	30
Betalactone	(2,472,027-2,500,534)nt	fengycin, NRP	53
Terpene	(2,614,793-2,636,682)nt	-	
		-	

T3PKS	(2,683,490-2,724,593)nt	--	
NRPS	(2,859,447-2,942,464)nt	bacitravcin, NRP	100
RRE-containing, Lassopeptide	(3,830,346-3,852,767)nt	--	
NRP-metallophore,NRPS	(4,283,204-4,335,008)nt	bacillibactin,bacillibactin E and F NRP	100

Key: NRPS-Non-ribosomal peptide synthethase, T3PKS- Type III polyketide synthase, NRP-Non-ribosomal peptide, RiPP-Ribosomally synthesize post-translationally modified peptide, nt- Nucleotide,LAP- linea azoline containing peptides, RRE-Ribosomally-synthesized recognition element.

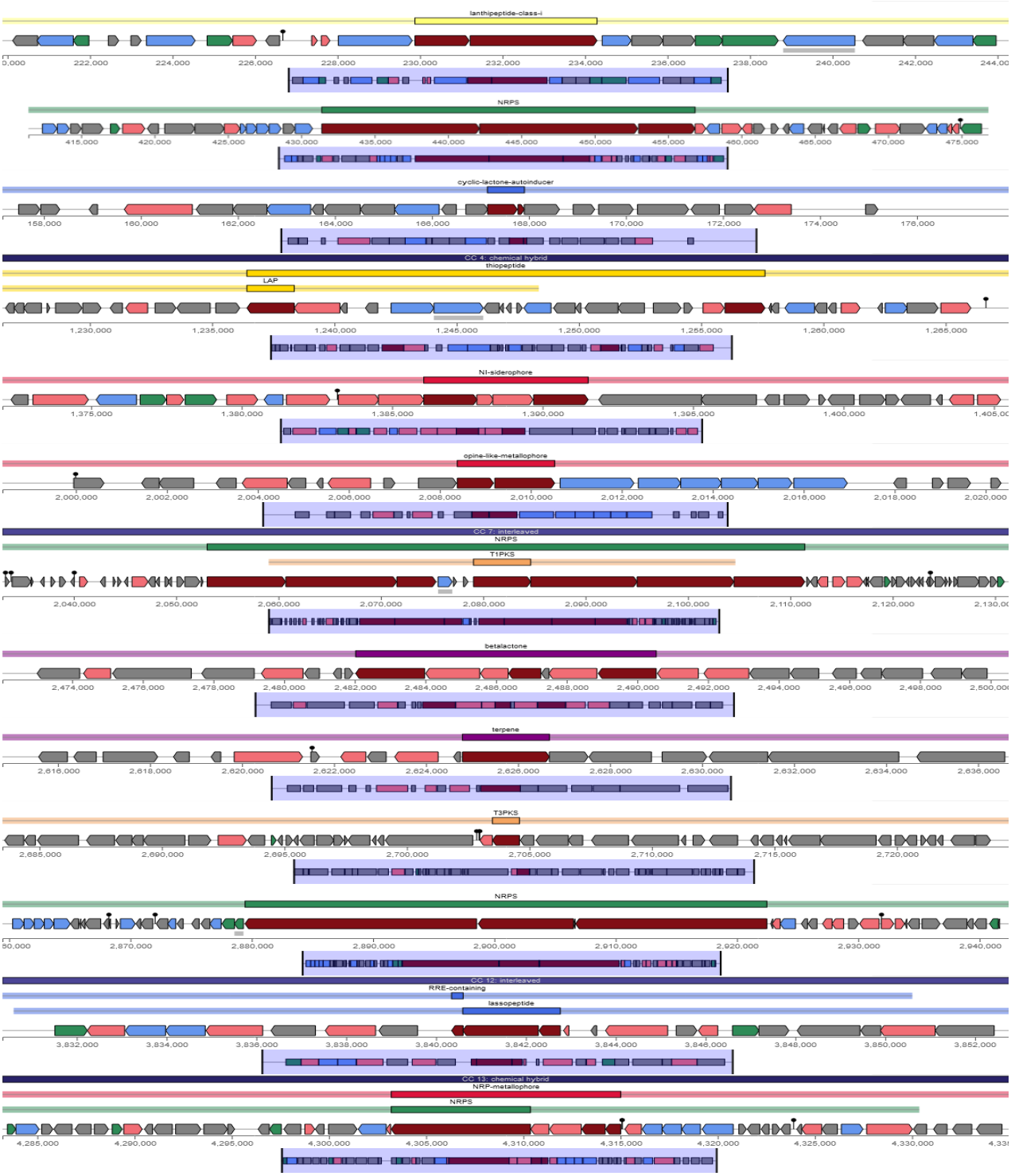

Fig 8. Different location of Antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) in from *B. sonorensis strain PMC204* genome.

Table 9:Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B. amyloliquefaciens strain HM618*

Cluster type	Cluster location	Most similar known cluster	% of similarity
Thiopeptide,LAP	(349,288-379,020)nt	-	
		-	
NRPS	(390,422-455,829)nt	surfactin, NRP:Lipopeptide	91
PKS-like	(1,042,096-1,083,316)nt	butirosin A and B, saccharide	7
Terpene	(1,166,140-1,186,880)nt	-	
	-		
NRPS	(1,569,241-1,636,972)nt	paenibacterin, NRP	60
TransAT-PKS,T3PKS, NRPS	(1,784,795-1894,906)nt	bacillaene, polyketide +NRP	100
Terpene	(2,120,851-2,142,734)nt	-	
	-		
T3PKS	(2,210,528-2,251,626)nt	-	
	-		
TransAT-PKS	(2,367,511-2,473,689)nt	difficidin, polyketide	100
NRP-metallaphore,NRPS,RiPP-like	(3,087,878-3,139,670)nt	bacillibactin, NRP	100
Other	(3,671,371-3,712,789)nt	bacilysin, other	100

Key: NRPS-Non-ribosomal peptide synthethase, T3PKS- Type III polyketide synthase, NRP-Non-ribosomal peptide, PKS-Polyketide synthase, RiPP-Ribosomally synthesize post-translationally modified peptide, nt-Nucleotide, LAP- linea azoline containing peptides.

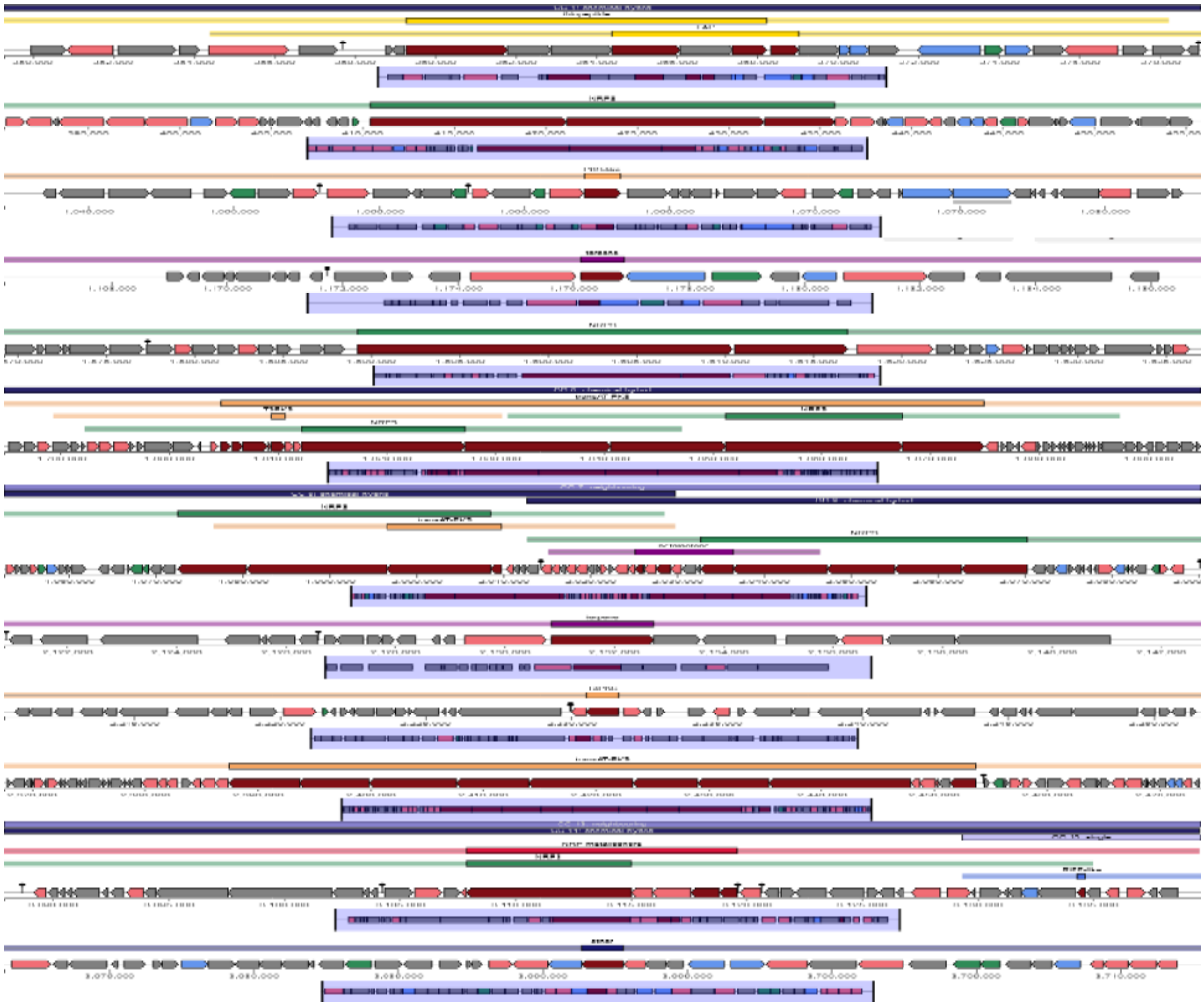

Fig 9. Different location of Antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) from *B. amyloliquefaciens strain HM618* genome.

Table 10: : Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B. firmus strain T8*

Cluster type	Cluster location	Most similar known cluster	% of similarity
T3PKS	(1,582,388-1,623,470)nt	-	-
LAP, Terpene	(2,029,968-2,055,805)nt	-	-
NI-siderophore	(4,089,425-4,120,937)nt	thailanstatin A, NRP+polyketide	10

Key: T3PKS- Type III polyketide synthase, NRP-Non-ribosomal peptide, nt- Nucleotide,LAP- linea azoline containing peptides.

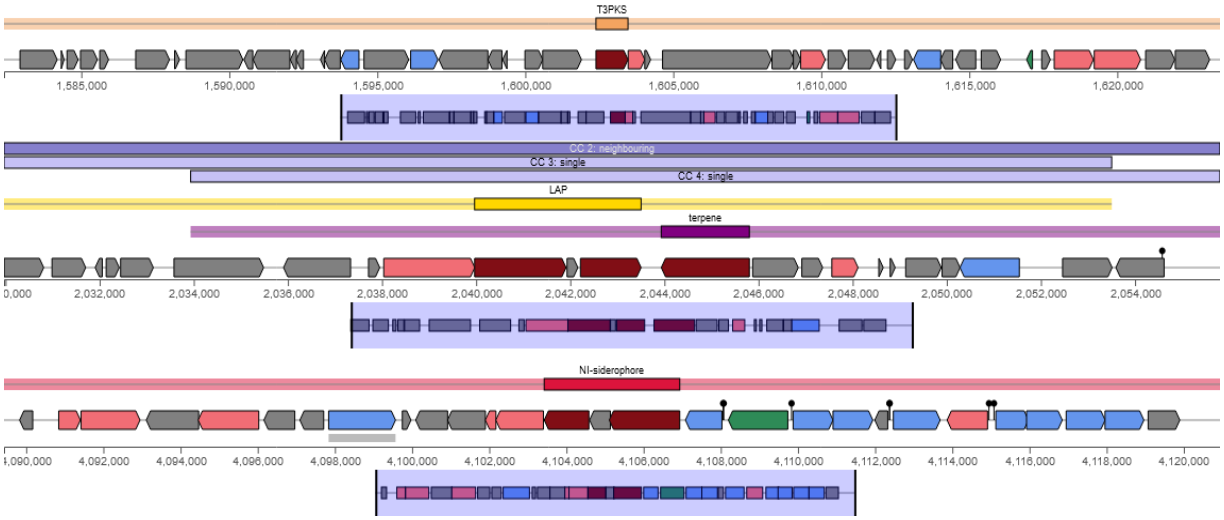

Fig 10. Different location of Antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) in *B. firmus strain T8* genome.

Table 11:Identified antibiotics and secondary metabolites region using strictness 'relaxed' from *B. vallismortis strain NBIF-001*,

Cluster type	Cluster location	Most similar known cluster	% of similarity
NRPS	(322,200-387,607)nt	surfactin, NRP:lipopeptide	82
PKS-like	(924,056-965,300)nt	butirosin A and B, saccharide	7
Terpene	(1,047,344-1,068,084)nt	-	-
Lanthipeptides-class-ii	(1,188,577-1,217,465)nt	-	-
transAT-PKS	(1,384,028-1,1,472,259)nt	macrolactin H, polyketide	100
transAT-PKS,T3PKS,NRPS	(1,690,950-1,801,043)	bacillaene, polyketide+NRP	100
NRPS,transAT-PKS, Betalactone	(1,865,674-2,003,475)nt	fengycin, NRP	100
Terpene	(2,028,702-2,050,585)nt	-	-
T3PKS	(2,113,903-2,155,003)nt	-	-
transAT-PKS	(2,269,989-2,376,171)nt	difficidin, polyketide	100
NRP-metallophore,NRPS,RiPP-like	(3,000,874-3,052,665)nt	bacillibactin, NRP	100
Other	(3,588,975-3,630,393)nt	bacilysin, other	100

Key: NRPS-Non-ribosomal peptide synthethase, T3PKS- Type III polyketide synthase, NRP-Non-ribosomal peptide, PKS-Polyketide synthase, RiPP-Ribosomally synthesize post-translationally modified peptide, nt-Nucleotide.

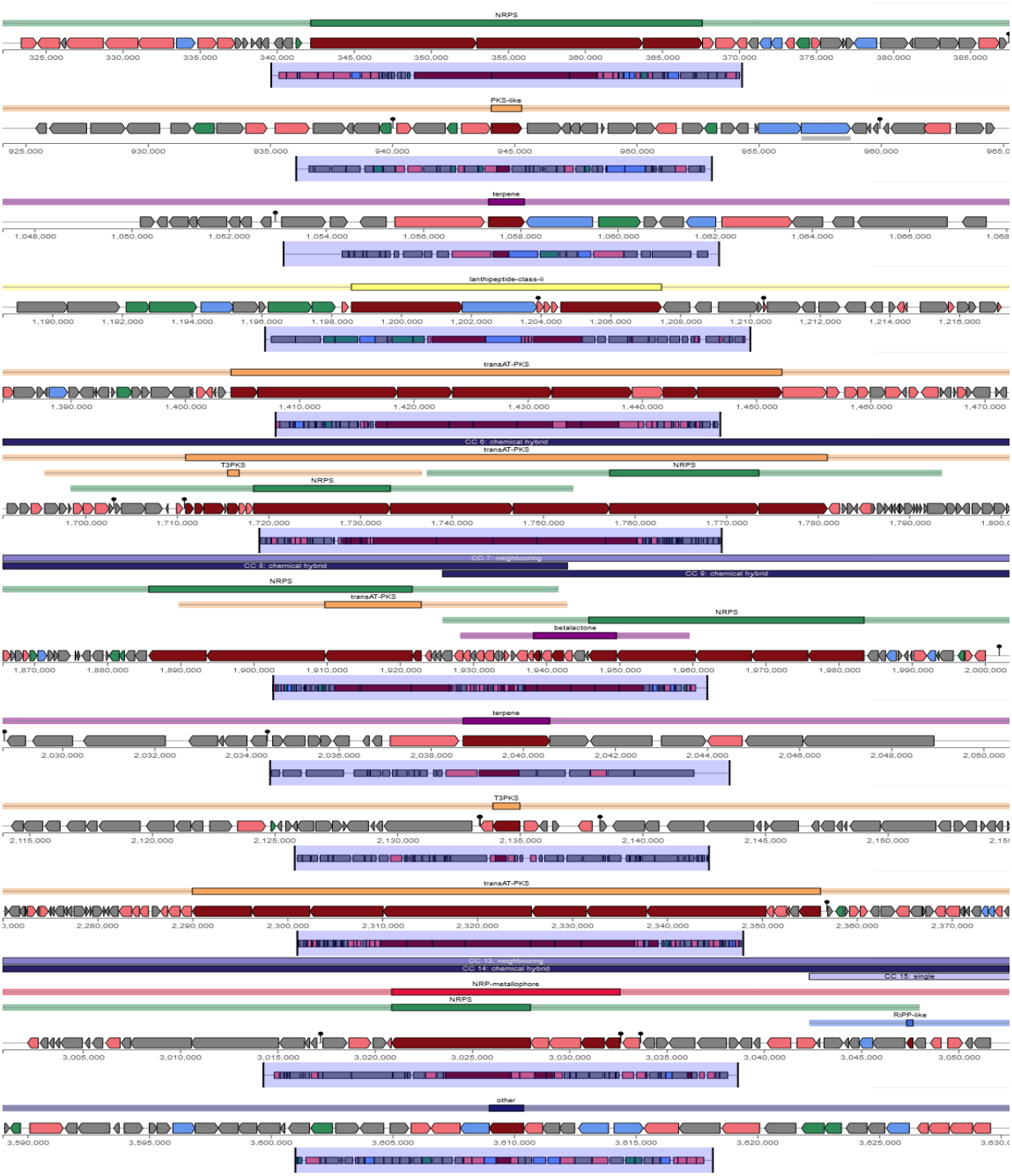

Fig 11. Different location of Antibiotics and secondary metabolites biosynthetic gene clusters (BGCs) in *B. vallismortis strain NBIF-001* genome.

Discussion

According to the genomic data retrieved from gene bank, the average genome size of the *B.subtilis* group was 4.11mb, the strain with highest genome size were *Bacillus licheniformis strain BL1202* (4.27mb), while *Bacillus vallismortis strain NBIF-001* has the lowest genome size of 3.80mb .The results for the identification of biosynthetic gene clusters (BGCs) were presented in Table 1-11 and Figure 1-11. A total of 97 antibiotics and secondary metabolites biosynthetic gene clusters were identified across the 11 genomes of *B.subtilis* group, which is equivalent to an average of 8.82 per genome. These clusters were classified into 22 different types and they are predicted for non-ribosomal peptide synthethase (NRPS), Terpenes, type

III polyketides synthase (T3PKS), type I polyketide synthase,transAT-PKS,betalactone,NRP-metallophore,Lanthipeptide-classI,Sactipeptide,Thiopeptide, linea azoline containing peptides (LAP), Ni-siderophore, Lassopeptide, RRE-containing, ribosomally-synthesized post-translationally modified peptides (RiPP), RiPP-like,CDPS, epipeptide, Cyclic lactone autoinducer,Opine-like metallophore, ectoine and some clusters described as 'Other'. Similarly, biosynthetic gene clusters such as NRPS, RiPP, T3PKS, T1PKS and terpenes that were identified in this study have also been reported by previous studies, because closely related members of firmicutes are known to abundantly produce various NRPS and PKS gene clusters (Caulier *et al*., 2019; Srivastava *et al*., 2022).

Many of the identified BGCs in this study appeared to be unknown and potentially new clusters that may produce new compounds, these clusters were predicted for NRPS, terpenes, T3PKS, Lanthipeptide-class-I,Thiopeptide, LAP, Ni siderophore, Lassopeptide, RRE-containing, Cyclic lactone autoinducer, this finding may be as a result of environmental factors and multiple stress which may affect the BGCs composition (Zhang *et al*., 2024) . These findings were in line with another study which also reported unknown similar clusters in the genome of B. *subtilis* group (Wu *et al*.,2023).

NRPS clusters were the most abundant across the 11 *Bacillus subtilis* genomes, however, there is uniform distribution of NRPSs clusters in the genomes of *B. subtilis subsp. subtilis strain, B.subtilis subsp. spizizenii, B. atrophaeus strain, B. vallismortis strain and B.amyloliquefaciens strains.* Previous genome mining studies have indicated that 31% of *Bacillus* Species harbors NRPS gene clusters (Srivastava *et al*., 2022).

However, NRPS cluster appears to be absent in the genomes of *B.firmus* and *B.clausii* only, therefore, NRPS gene clusters were conserved in the genomes of *Bacillus subtilis* group. The presence of single module NRPS and multimodule NRPSs genes clusters with high abundance across the genomes were similar with previous studies of Grubbs *et al*. (2017). Majority of the NRPSs BGCs were known clusters (not new) and are predicted to produce surfactin (lipopeptide), lychenysin, paenbacterin, bacilaene, fengycin, subtilomycin, locillomycin B,locillomycin C, bacitracin and bacillibactin, however, there is only one NRPS cluster that appears to be new from genome of *B. lentus strain NCTC4824,* these findings suggested that the new identified clusters could be source of novel antibiotic or secondary metabolites that may increase more space for natural products and discovery of drugs (Steinke *et al*., 2020).However, the similarity percentage of the known clusters were slightly lower when compared with that of study conducted by Yin *et al*.(2023), this is due to environmental factors which may affect BGCs distribution and composition.

Terpenes BGCs were the most conserved BGCs across the 11 genomes with an average of 1.45 per genome, this study identified 16 terpenes gene clusters which encoded for the synthesis of chains of activated isoprene units (C_5H_8) that make up diverse class of natural products. The distribution of terpenes BGCs is uniform across *B. subtilis* group, except in *B. clausii KSM-K16, B. lentus strain NCTC4824 and B.sonorensis strain PMC204*.Majority of the terpenes BGCs were unknown (new gene clusters), this findings revealed the potentials of *B.subtilis* group for producing novel natural products, however, the known identified terpenes clusters are predicted to produce lipopeptides (Harwood *et al*., 2018).

Type III polyketides synthase (T3PKS) gene cluster was identified in nearly all the 11 *Bacillus* strains with an average of 1.36 per genome. T3PKS represent the second BGC with highest number of unknowns BGCs, this indicate the possibility of producing novel antibiotics and

secondary metabolites which could be developed into useful drugs. In this study, only one T1PKS gene cluster were identified from *B. sonorensis strain PMC204*, however, T2PKS gene cluster appears to be absent across the 11 genomes. The known T3PKS gene clusters are predicted to produce bacillaene and 1-carbapen-em-3-carboxilic acid. Recently, Yin *et al*. (2023) analyzed the distribution of BGCs in *Bacillus* species and found that T1PKS, T2PKS and T3PKS were among the most abundant classes of BGCs.

transAT-PKS gene cluster appears to be third most abundance BGC in the *B.subtilis* group, however, its limited to only genomes of *B. subtilis subsp. spizizenii TU-B-10, B. atrophaeus strain MBLB1156, B. vallismortis strain NBIF-001* and *B. amyloliquefaciens strain HM618*.The transAT-PKS gene clusters was predicted to produce bacillaene,difficidin and fengycin. Similarly, NRP-metallophore gene clusters were also identified with uniform distribution in all the strains except, *B.firmus strain T8*, *B. lentus strain NCTC4824* and *B. clausii KSM-K16*.The predicted natural specialized product from NRP-metallophore gene clusters were only bacillibactin, this is due to the fact that BGC can either encode for the formation of single or multiple compound (Grubbs *et al*., 2017).

Bacillus sonorensis strain PMC204 contain 13 BGCs which is the highest when compared with BGCs found in the genomes of other members, this finding might be as a result of larger genome size in *Bacillus sonorensis strain PMC204*.However, *Bacillus firmus strain T8* were found with only 3 BGCs (Table 10) which is the lowest when compared with BGCs found in the other genomes of the *Bacillus strains*. Cyclic-lactone-autoinducer and opine-like metallophore gene clusters were also found in the genome of *Bacillus sonorensis strain PMC204* only and they are predicted to produce amyloliquecidin GF610 and bacillopaline respectively, this may also be due to the genome size and composition (Yin *et al*., 2023).

The presence of betalactone BGCs has been established by these studies, all the betalactone BGCs showed similarity with previous BGCs and they are predicted to produce only fengycin. Similarly, gene clusters related to fengycin synthesis in *Bacillus subtilis* has been reported (Su *et al*., 2020).

however, there are no any unknown cluster that may lead to the formation of new compounds, this finding were consistent with another previous studies by Kaspar *et al*.(2019).Lanthipeptide-class-I biosynthetic gene cluster appears to be present in all the *Bacillus* genomes except *B.subtilis subsp. subtilis strain, B.firmus strain T8, B. amyloliquefaciens strain HM618* and *B. lentus strain NCTC4824.*

Similarly, the distribution of lanthipeptides-class-1 BGCs is uniform with only one unknown found in the genome of *B.clausii KSM-K16 DNA*, however, the known lanthipeptide-class-I BGCs are predicted to produce known compounds including lichenicidin VK21A2, entianin, subtilin and lanthipeptides.

Sectipeptides and thiopeptides biosynthetic gene clusters appears to be present in only 4 strains each and they are predicted to produce subtilosin A and subtilosin B. Meanwhile, there is presence of only one unknown cluster of thiopeptide in the genome of *B. atrophaeus strain MBLB1156*, this suggest potential novelty of the antibiotic and secondary metabolites that may be associated with this cluster, this also indicate the potentiality of *B.atrophaeus strain MBLB1156* for producing new natural compounds.NI-siderophore and -linea azoline containing peptides (LAP) biosynthetic gene clusters also exist with less abundance, but there are 2 unknown LAP and 1 unknown NI-siderophore BGCs that may be considered as

potential source of natural compounds. However, the known clusters of NI-siderophore and LAP are predicted to produce schizokenem, thailastatin A, butirosin A and butirosin B (Wu *et al*., 2023).The presence of bacteriocins biosynthetic gene clusters (RiPP) has been established by this studies, *B.vallismortis strain NBIF-001* and *B.amyloliquefaciens strain HM618* were the only strains with RiPP biosynthetic gene cluster which is predicted to produce only bacillibactin, this result were in agreement with previous finding which reported the antimicrobial potential of *B.amyloliquefaciens* species (Caulier *et al*.,2019).
There are nearly 5 clusters identified as 'other' by the antiSMASH software, these gene clusters were predicted for only bacilysin. Similarly, CDPS biosynthetic gene cluster exist and it's predicted to produce pulcherriminic acid.

CONCLUSION

The biosynthetic gene clusters for antibiotic and secondary metabolites identified in this study could be helpful for future research, the results which were presented in Tables and Figures can be utilized for exploring the potentials of *B. subtilis* group for producing novel antibiotics and secondary metabolites. All the *Bacillus* strains analyzed contain at least two new gene cluster that can be source of new antibiotics and secondary metabolites which could be develop in to useful drugs against rapid emergence of broad-spectrum antibiotic resistance. Therefore, further research is strongly recommended using other genome mining resource such as NRPS gene search, RiPPquest, PKS gene search, BLAST, CLUSEAN, BAGEL, ClustScan, NP.searcher, PRISM etc.

REFERENCES

Caulier, S., Nannan, C., Gillis, A., Licciardi, F., Bragard, et al., 2019. Overview of the antimicrobial compounds produced by members of the *Bacillus subtilis* group. *Front. in Microbiol*. *10*(FEB), 1–19. Doi:10.3389/fmicb.2019.00302

Chavali, A. K and Rhee, S. Y., 2018.Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Briefings in Bioinformatics* 19 (5), 1022–1034 Doi: 10.1093/bib/bbx020

for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*. *Computational and*

Front. Microbiol. 14:1239837. Doi: 10.3389/fmicb.2023.1239837

Grubbs, K.J; Bleich R.M; Santa K.C; Allen S.E, Farag S, et al., 2017. Large-scale bioinformatics analysis of *Bacillus* genomes uncovers conserved roles of natural products in bacterial physiology.*Mol. Biol. And physiol.* 2 (6) :1-16 e00040-17mSystems . Doi:10.1128/mSystems.00040-17.

Harwood C, R; Jean-Marie M; Susanne P and Jos´e A,.2018. Secondary metabolite production and the safety of industrially important members of *the Bacillus subtilis group*. *FEMS Microbiol. Reviews, Vol. 42, No. 6* Federation of European microbiological societies.

Hu, D; Sun, C; Jin, T; Fan G; Mok K. M,. et al. 20209.Exploring the Potential of Antibiotic Production From Rare Actinobacteria by Whole-Genome Sequencing and Guided MS/MS Analysis. *Front. Microbiol.*; 11:1540. Doi: 10.3389/fmicb.2020.01540

Iqbal, S; Ullah, N and Janjua, H.A,.2021.In Vitro Evaluation and Genome Mining of *Bacillus subtilis Strain RS10* Reveals Its Biocontrol and Plant Growth-Promoting Potential. Agricult. 11, 1273. Doi:10.3390/agriculture11121273

Kaspar, F; Neubauer, P; Gimpel, M,. 2019.Bioactive secondary metabolites from Bacillus subtilis: a comprehensive review.*J Nat. prod.* 82:2038-2053.

Lee, N., Hwang, S., Kim, J., Cho, S., Palsson, B., et al. 2020. Mini review: Genome mining approaches. *Structural Biotechnol. J. 18*, 1548-1556. Doi:10.1016/j.csbj.2020.06.024

Martina, A; Marius, S; Evi S; Nadine Z and Peter S,. 2017. Mining Bacterial Genomes for Secondary Metabolite Gene Clusters Antibiotics: *Methods Mol. Biol*. 1520: 9-2. DOI:10.1007/978-1-4939-6634-9

Nannan, C; Quynh, V.H; Annika, G; Simon, C.T; Thanh, T.N,. 2021.Bacilysin within the *Bacillus subtilis* group: gene prevalence *versus* antagonistic activity against Gram-negative foodborne pathogens *J. of Biotechnol*. 327 ; 28–35 . Doi:10.1016/j.jbiotec.2020.12.017

Put, H; Gerstmans, H; Vande, C.H; Faarten, F; Michiels, J,.2024. *Bacillus subtilis* as a host for natural product discovery and engineering of biosynthetic gene clusters. *Nat.Prod.Rep*. 3-7. DOI:10.1039/d3np00065f.

Srivastava, A.K; Srivastava, R; Bharati, A.P; Singh, A.K; Sharma, A; et al.2022. Analysis of Biosynthetic Gene Clusters, Secretory, and Antimicrobial Peptides Reveals Environmental Suitability of Exiguobacterium profundum PHM11. *Front. Microbiol*. 12:785458. Doi: 10.3389/fmicb.2021.785458

Steinke, K; Mohite O.S; Weber, T; and Kovacs A.T; et al.2021. Phylogenetic distribution of secondary metabolites in the Bacillus subtilis species complex. mSystems 6:e00057-21. Doi:10.1128/mSystems.00057-21.

Su, Z; Chen, X; Liu, X; Guo, Q; Li, S; et al. 2020. Genome mining and UHPLC–QTOF–MS/MS to identify the potential antimicrobial compounds and determine the specificity of biosynthetic gene clusters in Bacillus subtilis NCD-2. *BMC Genomics*, 21:767.Doi:10.1186/s12864-020-07160-2

Tiwari, R and Rana C.S;.2015. Plant secondary metabolites: A review. *Intern. J. Of Engineering Res. and General Sci.*; 3(5):661-670.

Williams, A.N; Sorout, N; Cameron, A.J and Stavrinides J ;.2020.The Integration of Genome Mining, Comparative Genomics, and Functional Genetics for Biosynthetic Gene Cluster Identification. *Front. Genet*. 11:600116.Doi: 10.3389/fgene.2020.600116

Wu, D; Fu, L; Cao, Y; Dong, N and Li, D;.2023. Genomic insights into antimicrobial potential and optimization of fermentation conditions of pig-derived *Bacillus subtilis* BS21. *Structural Biotechnol. J. 18*, 1548-1556. Doi:10.1016/j.csbj.2020.06.024

Yin, Q-J; Ying, T-T; Zhou, Z-Y; Hu, G-A; Yang C-L; et al.2023 .Species-specificity of the secondary biosynthetic potential in *Bacillus*. *Front. Microbiol.* 14:1271418. Doi: 10.3389/fmicb.2023.1271418

Zhang, Z; Zhang, L; Lihan, Z; Chu, H; Zhou, J; et al.2024.Diversity and distribution of biosynthetic gene clusters in agricultural soil microbiomes. *J Environ. Microbiol. Res*, 9:4, 01263-23. Doi:10.1128/msystems.01263-23