AGRO PRODUCTIVIDAD

Colegio de Postgraduados

Genome mining for bioprospecting of biosynthetic genes clusters for bacterial metabolites potentially useful in agroecological production

Alberus J. Yvens1; Verdel-Aranda, Karina2; Martinez-Hernandez Aida, Yañez-Morales, María de Jesús3, Lara-Reyna, Joel1

1Colegio de Postgraduados, Campus Campeche. Carretera Haltunchén-Edzná km 17.5, Sihochac, Champotón, Camp. México, C. P. 24400.
2Tecnológico Nacional de México, Instituto Tecnológico de Chiná, Campeche, México. Calle 11 S/N entre 22 y 28. Chiná, Campeche, México C. P. 24520.
1Colegio de Postgraduados, Campus Montecillo. Km. 36.5, México 136 5, Montecillo, Montecillo, Méx., C. P. 56230
*Correspondence: jlara@colpos.mx

ABSTRACT

Objective. To describe the relevance and some tools of genome mining to explore genetic and molecular determinants encoded in bacterial genomes to address agronomic problems.

Design/Methodology/Approach. Literature review of the importance of bacteria as a reservoir of biosynthetic gene clusters (BGC), involved in the production of metabolites with biological activity as anti-pathogens; and of genome mining as a tool to reveal this potential.

Results. Bioinformatic tools are useful for the exploration of bacterial genomes and have the potential to contribute to the resolution of agronomy problems. For example, the use of bacteria, their genes and metabolites for the control of phytopathogens that attack crops of global importance. Likewise, the limitations of the genome mining and their coupling with other experimental approaches to achieve bioprospecting of BGC or their related metabolites are summarized.

Limitations of the study/implications: Although the use of genome mining to explore the potential of bacteria is a very powerful approach, it will always be necessary the experimental corroboration at the laboratory level, to confirm the hypotheses generated by bioinformatics tools.

Findings/conclusions: Genome mining allows to take advantage of the large number of bacterial genomes currently sequenced, that are available in public databases to understand the genetic bases of their biological activities. As well as for the heterologous expression of biosynthetic genes, or the identification and purification of new metabolites. The foregoing with the objective of contributing with more effective and environmentally friendly solutions that address agronomic problems.

Keywords: bacterial genomes, bioprospection, genome mining, biosynthetic gene clusters (BGC), phytopathogens.

Citation: Alberus, J. Y., Verdel-Aranda, K., Martinez-Hernandez, A., Yañez-Morales, M. de J., & Lara-Reyna, J. (2022). Genome mining for bioprospecting of biosynthetic genes clusters for bacterial metabolites potentially useful in agroecological production. *Agro Productividad*. https://doi.org/10.32854/agrop.v15i9.2226

Academic Editors: Jorge Cadena Iñiguez and Libia Iris Trejo Téllez

Received: February 11, 2022.
Accepted: July 18, 2022.
Published on-line: October 06, 2022.

Agro Productividad, 15(9). September. 2022. pp: 77-87.

INTRODUCTION

It is well known that bacteria produce a wide variety of chemical compounds with different metabolic origins and different biological activities. A biosynthetic gene cluster (BGC) is a group of genes that, when expressed together, produce enzymes involved in

Image by Gerd Altmann at Pixabay

the biosynthesis of a bioactive metabolite, also called secondary (Medema *et al.*, 2015). BGCs are the toolbox available to bacteria to produce complex chemical structures that can be applied in different areas, such as solving some agronomic problems. There are many applications of bacteria, their metabolites or enzymes in the field of agronomy that have not yet been exploited in Mexico according to their potential. For example, the great losses caused each year by phytopathogens that attack crops of economic importance. This and other problems could be addressed through bioprospecting of bioactive microbial metabolites (BMMs).

The isolation and chemical identification of BMM until their application in a given product is a long run. However, current advances in accelerated genome sequencing, and the use of bioinformatics tools to "mine" information, accelerate and smooth processes, allowing the identification of which biosynthetic genes are potentially involved in some biological activity and what are the possible production routes of metabolites. Such proposals for possible genes subsequently have to be addressed from a biochemical point of view or using other complementary tools such as metabolomics or genetic engineering. This review recapitulates the main route for the systematic and exhaustive search of clusters of genes involved in the biosynthesis of bioactive compounds, emphasizing genome mining applied in the identification of possible BGCs and the bioactive metabolites they produce, for agronomic use in the control of phytopathogens.

Genomic and metabolic plasticity of bacteria

Bacterial genomes are very diverse. They contain a core genome, which includes genes that provide essential information and is conserved for most bacteria. They also contain a flexible part in which are the genes that give identity to each species. These genes provide additional characteristics that differentiate one bacterium from another both genetically and metabolically. This flexible part of the genome is where genomic plasticity occurs in bacteria, which is defined as the ability to incorporate genes or complete sets of genes (operons) into their genome (Bennett, 2004).

Among the mechanisms that contribute to genomic plasticity are point mutations that alter the expression of genes or the metabolic activity of encoded enzymes; mechanisms of genetic exchange such as recombination between genomes, or conjugation (plasmid exchange) between bacteria. As well as the horizontal transfer of genes promoted by bacteriophages, mobility of genome fragments by transposons, repairs and integration of DNA regions of different sizes (Dobrindt and Hacker, 2001; Sela *et al.*, 2018). Genomic plasticity is considered one of the keys to the evolution of bacteria (Sela *et al.*, 2018); the infinite metabolic possibilities that result from it make bacteria a source of numerous bioactive metabolites that can be of beneficial use with different applications.

Main types of biosynthetic gene clusters

There are different classes of biosynthetic gene clusters, classified according to the type of metabolite in whose biosynthesis participate. In addition to their origin, bacterial bioactive metabolites are grouped into ribosomal and non-ribosomal products. Here we address the main groups that belong to the non-ribosomal type; non-ribosomal peptide

synthetases and polyketide synthases which are two large sets of biosynthesis enzymes that have been studied with greater emphasis due to their frequent prevalence in bacterial genomes.

The BGC of non-ribosomal peptide synthetases (NRPS) have as their core, multi-modular enzymes that condense amino acids in a linear way through the formation of thioester-type peptide bonds (Figure 1, right), forming peptides different from those synthesized in the ribosome since, in addition to the 20 amino acids, they integrate more than 500 non-proteinogenic amino acids, heterocyclic elements and glycosylated residues, through biosynthetic pathways from other origins (Walsh and Fischbach, 2010).

Polyketide synthases (PKS) are an extensive group of enzymes classified into three types, whose activity is the modular condensation of monomers of acyl-CoA, malonyl and methyl malonyl, important intermediates in the biosynthesis of many organic compounds (Fischbach and Walsh, 2006).

NRPS genes such as PKS contain functional domains that once the enzyme is expressed become modules for substrate activation and for initiation, elongation, modification and termination of the product that can be a bioactive metabolite. Additionally, they may have transport proteins, transferases, and other auxiliary enzymes (Figure 1, left).

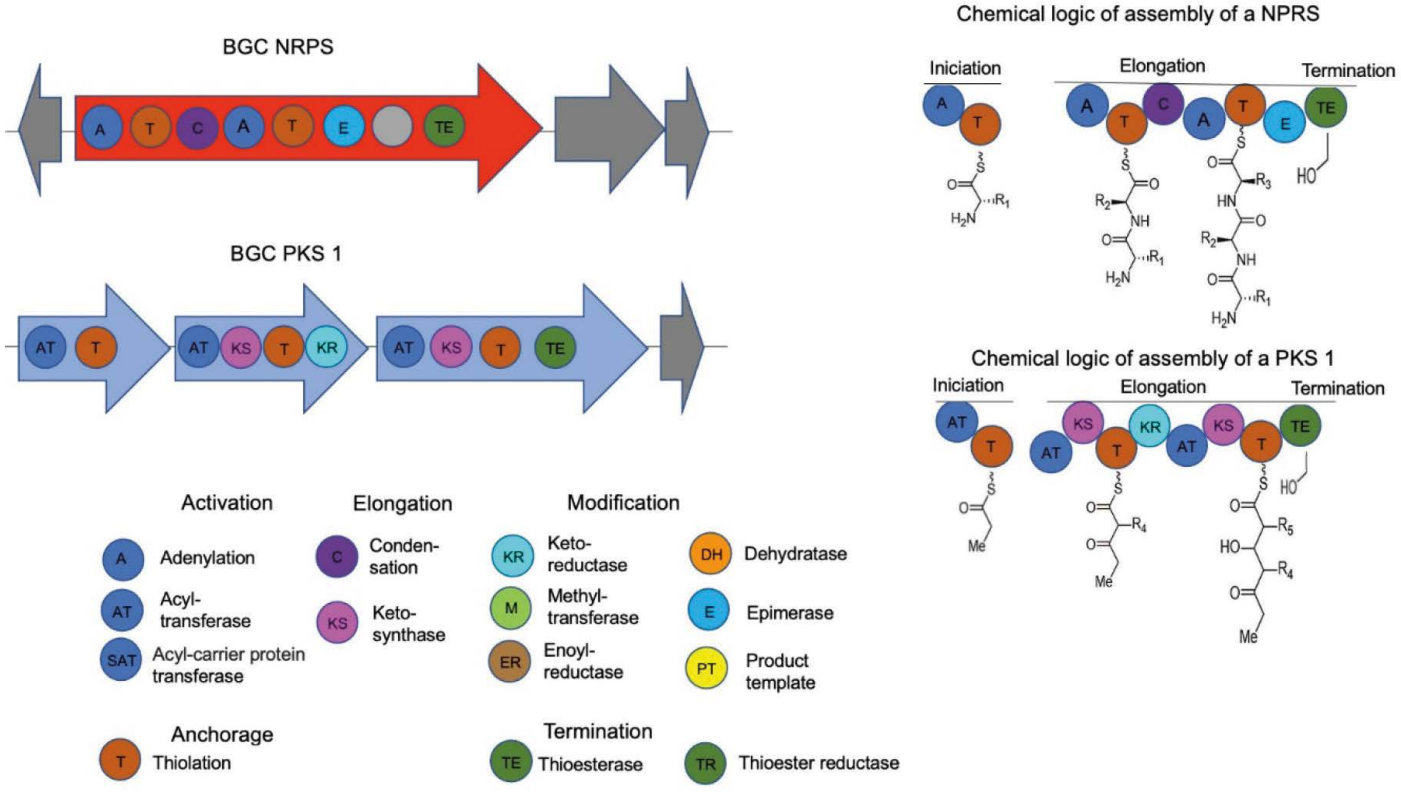


Figure 1. Example of bacterial biosynthetic clusters coding for an NRPS or a PKS 1 and the assembly of the metabolites they produce. Arrows represent the genes of a BGC that encode biosynthetic enzymes (red or blue), as well as genes that encode transport proteins, auxiliary enzymes, etc., (gray). On the right is the assembly logic (biosynthesis) of non-ribosomal peptides and polyketides. As well as the most common functional domains or modules (circles) of NRPS and PKS and their function (below).

Traditional strategy for the identification of BGC of bioactive metabolites in microorganisms

The traditional method for identifying biosynthetic gene clusters of secondary metabolites in bacteria (Figure 2, top) is based on the isolation of the metabolite guided by the bioactivity of supernatants from bacterial cultures. Also, on the use of chemical methods, such as mass spectrometry and nuclear magnetic resonance, to deduce the chemical structure of the bioactive metabolite.

Subsequently, the corresponding biosynthetic genes are inactivated by induced mutagenesis or randomized genetic deletion, followed by the detection and isolation of non-producing clones (Bachmann, Lanen, & Baltz, 2014; Read *et al.*, 2020). These mutant strains that do not produce the metabolite are analyzed to identify their genes affected; therefore, they are responsible for the biosynthesis of the metabolite. Despite the great advantage of directly isolating and characterizing the active product, this approach has as its main limitation that the expression of BGC is controlled by factors found in the natural habitat of bacteria. This means that, in laboratory conditions those genes may not be expressed, which generates the so-called cryptic genes (Clardy *et al.*, 2006; Lee *et al.*, 2020).

Genome mining

Genome mining is the set of steps and tools, totally dependent on bioinformatics, that allows to identify bacterial clusters from genome sequences, by comparing against

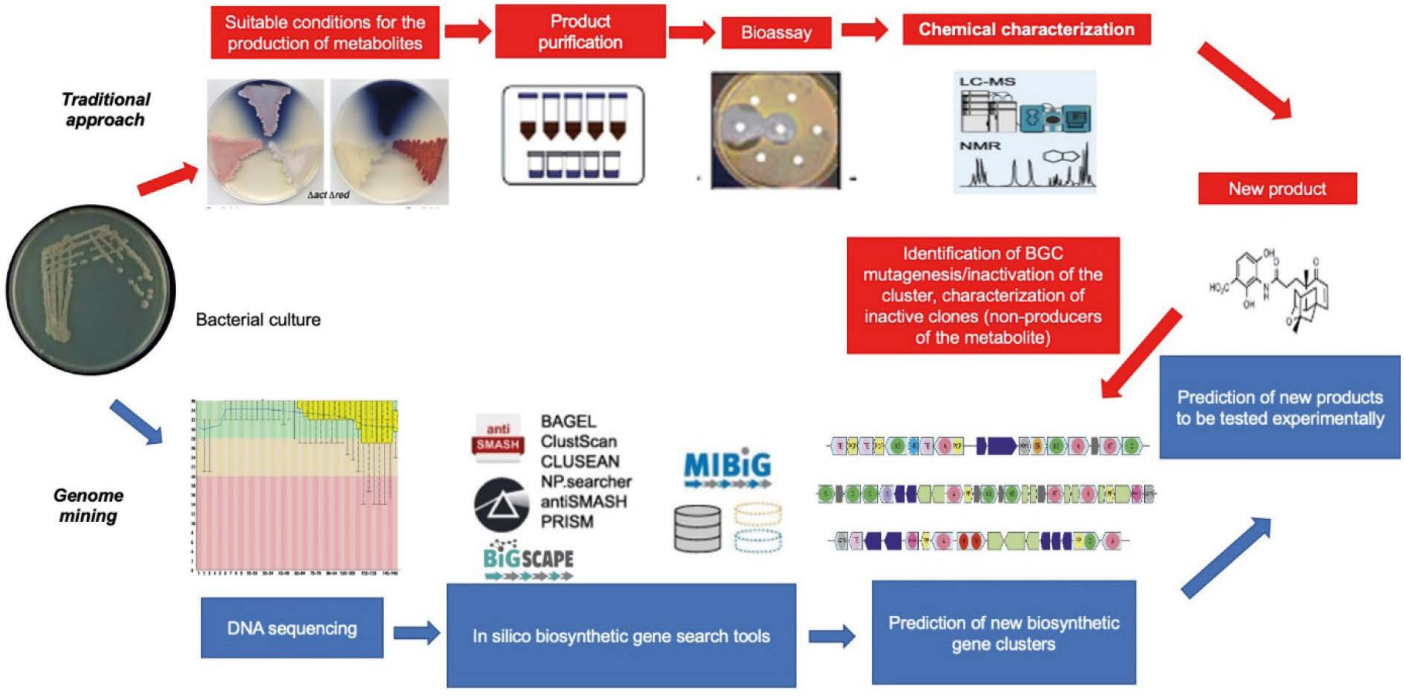


Figure 2. Comparison of the traditional strategy and the genome mining approach for the seeking of new bioactive metabolites and the BGCs that synthesize them.

previously described clusters. The existence of conserved regions between clusters of various bacteria greatly facilitates the genome mining; although bioactive metabolites can be highly diverse in their chemical structure, the biosynthetic machinery that includes assembly enzymes, belongs to the same families of highly conserved enzymes, especially PKS and NRPS (Lee *et al.*, 2020).

This new approach is radically different from the traditional approach and allows new BGCs to be identified more quickly (Figure 2, bottom). In recent years; as a result of new sequencing technologies and the easy and inexpensive access to genome sequence data, many bioinformatics tools have been developed based on identifying the highly conserved regions of the BGC, for the annotation or labeling of coding regions in the genomes (Ren *et al.*, 2020; Cibichakravarthy and Jose, 2021). The application of these technologies has revealed that the biosynthetic potential of bacteria has been underestimated. For example, following the traditional strategy, only four BCCs were discovered and characterized in the bacterium *Streptomyces coelicolor* in 30 years; but it is currently known that members of this genus possess approximately 30 clusters of biosynthetic genes, including many whose products have not yet been identified (Bentley *et al.*, 2002; Lee *et al.*, 2020).

Bioinformatics tools most commonly used in genome mining

Below are the general route, hotspots, and bioinformatics tools most commonly used for genome mining in the search for BGC.

1. **Evaluating the quality of genomes**. The completeness and quality of a sequenced genome is a fundamental part of mining genomes in search for BGC. It is known that a large number of genomes that are sequenced and reported in the databases are not complete. For example, 36% of the genomes of the genus *Streptomyces* show poor completeness (Studholme *et al.*, 2016). Likewise, the quality of the sequenced nucleotide bases is of great importance for the prediction of BGC since high-quality genomes improve the prediction of coding sequences (CDS), which in turn improves the prediction of the large genes that usually compose the core machinery of BGC.
2. **Identification of genes and their function (Annotation of genomes)**. For the identification of genes encoding proteins, rRNA and tRNA, etc., a wide variety of tools are used. One of the most used is RAST (Rapid Annotation using Subsystem) that identifies genes, assigns functions, predicts groups of genes associated with abstract functional roles that are represented in the genome (subsystems), uses this information to reconstruct the metabolic network and makes the result easily manipulated for the user. In addition, the annotated genome can be explored in an environment that supports comparative analysis with other different genomes (Aziz *et al.*, 2008).
3. **Use of BGC databases**. Currently the largest reservoir of BGC is in the Atlas of biosynthetic gene clusters (Integrated Microbial Genomes IMG-ABC, Joint Genome Institute JGI) (Hadjithomas *et al.*, 2015). This gene bank contained more than 960 000 groups of putative genes as part of some BGC in 2016. Only a small fraction of these BGCs make it to the final product description. Recently, the

Minimum Information for Biosynthetic Gene clusters (MIBiG) initiative has carried out a manual re-annotation of approximately 1300 BGCs representing an important highly curated reference dataset (Medema *et al.*, 2015). A significant number of programs have been developed to conduct searches for biosynthetic gene clusters within genomes including BAGEL (de Jong *et al.*, 2006); ClustScan (Starcevic *et al.*, 2008); CLUSEAN (Weber *et al.*, 2009); NP.searcher (Li *et al.*, 2009); PRISM (Skinninder *et al.*, 2017); and antiSMASH (Blin *et al.*, 2019). Most of these programs rely on searching for highly conserved BGC sequences to map their location. The Secondary Metabolite Bioinformatics Portal is http://www.secondarymetabolites.org a platform that updated and exhaustively groups the specialized software for the search for new BGCs. In this platform is the antiSMASH program that has been the most widely used in the prediction of gene clusters and which we described below.

4. **BGC prediction**. Since its initial launch in 2011, antiSMASH has become the most widely used tool for the search for clusters of secondary and specialized biosynthetic genes in fungi and bacteria; including those producing polyketides, non-ribosomal peptides, terpenes, aminoglycosides, aminocoumarins, indolocarbazoles, antibiotics, bacteriocins, nucleosides, beta-lactams, butyrolactones, siderophores, melanins and others (Medema *et al.*, 2011). In addition to using the characteristic gene of each cluster type using HMM (Hidden Markov Models) profiles for gene cluster identification, antiSMASH uses an algorithmic approach to extend the gene pool by 5, 10 or 20 kb on both sides; therefore, poorly spaced clusters can be merged into "superclusters" (Blin *et al.*, 2021). AntiSMASH also provides additional domain analysis options of the NRPS/PKS; as well as annotation, prediction of the central chemical structure of polyketides and non-ribosomal peptides, comparative analysis of gene groups (ClusterBlast) and also the analysis of families of enzymes involved in secondary metabolism (Blin *et al.*, 2021). There are other tools based on phylogenetic reconstructions such as EvoMining (Cruz-Morales *et al.*, 2016) and NaPDoS (Ziemet *et al.*, 2012); as well as in similarity networks such as Genome Neighborhood Networks (GNNs) (Zao *et al.*, 2014). In both cases there is the limitation that they require extensive prior knowledge of homologous enzymes.

5. **Integrating Genome Mining with Metabolomics and Targeted Search.** The integration of different strategies can be crucial for the discovery and study of bacterial secondary metabolites. Metabolomics is an analytical profiling technique used to measure and compare a large number of metabolites present in biological samples (Manchester and Anand, 2017). Once combining genomics and metabolomics, a new non-ribosomal lipopeptide Stendomycin (antifungal) was discovered from *Streptomyces hygroscopicus* (Kersten *et al.*, 2011). In another study, bioactive metabolites of *Streptomyces roseosporus* were identified by integrating correlation networks between metabolite fragmentation patterns by mass spectrometry and metabolomics, leading to the discovery of Stenothrycin, a metabolite with antimicrobial activity against Gram positive and negative bacteria (Liu *et al.*, 2014). Additionally, the search for metabolites directed by function, by tracking their biological activity, can make the

search for secondary metabolites more successful and give it a more applied meaning as we may see below.

Bioactive compounds of bacteria applied in agronomy: the case of biological control against phytopathogens

An example of BGCs useful to promote agroecological production are those that code for the biosynthesis of metabolites applicable to the control of phytopathogens. Over several years various microorganisms including different bacteria have proved to have the ability to suppress diseases in economically important crops (Table 1). This antagonistic action of bacteria against phytopathogens is given among other factors by a variety of secondary metabolites including iron-chelating siderophores, antibiotics, volatile biocides, as well as lytic enzymes and degradation enzymes (Compant *et al.*, 2005; Pal and Gardener, 2006; Colla *et al.*, 2018; Vurukonda, Giovanardi and Stefani, 2018; Köhl, Kolnaar and Ravensberg, 2019). A recent example of the use of genome mining is presented in the work of Siupka *et al.* (2020), who found multiple clusters of biosynthetic genes in a strain of *Streptomyces* isolated from chimney ash, which showed antagonistic activity against pathogenic strains of *Fusarium avenaceum*, *Aspergillus niger*, *Nigrospora oryzae* ssp. *roseF7*, and *Curvularia coatesieae* ssp. *junF9*. To give one more example, in our laboratory a confrontation test was carried out with phytopathogenic fungi to discriminate within a strain collection of more than 300 bacteria; starting only from antagonistic bacteria to those phytopathogens at the genome sequence and perform mining in search of the BGC that carry out the activity for which they were selected. Once the BGC has been identified, they can be heterologously expressed in controlled systems such as bioreactors, for the isolation or purification of metabolites and their eventual application (unpublished study data).

CONCLUSIONS

Genome mining is an important tool for the quick and easy prediction of biosynthetic gene clusters, based on the genome data of bacteria. This approach easily and quickly opens a door for the possible resolution of problems in the agricultural sector. However, bioprospecting of metabolites derived from these genes requires a subsequent and extensive laboratory analysis. The activation of biosynthetic clusters of secondary metabolites that are silenced; purification of the final metabolic products; and successful elucidation of their chemical structure should be done.

The information available in the databases is scarcely exploited in applications and developments directed towards agriculture. The education of human resources, trained in the management of these bioinformatics tools and their integration into task groups is essential; towards the bioprospecting of the microbial resources identifiable in Mexico, thus accelerated through genome mining. On the knowledge of this genetic richness, it may be possible to contribute to the resolution of some agricultural problems in the country, in order to offer the use of natural products and their mass production as alternatives, through technological developments based on biosynthetic gene clusters.

Table 1. Examples of some of the antibiotics produced by bacteria antagonistic to fungi; bacteria that produce them, as well as the target pathogen, and the diseases they cause are shown.

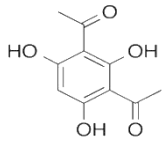
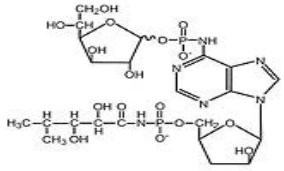
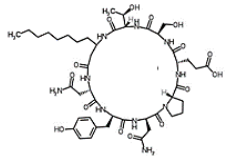
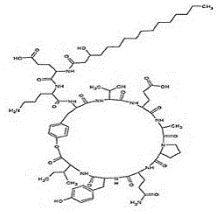
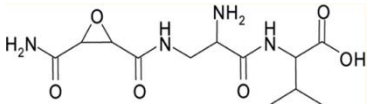
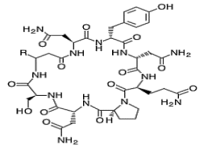
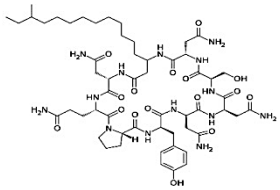
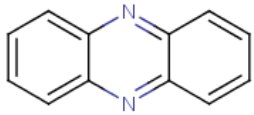
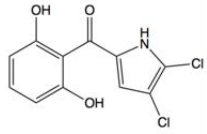
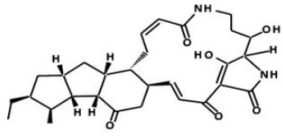
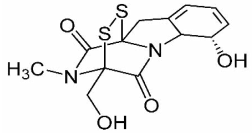
Antibiotic	Source / Target pathogen / Illness	Chemical structure	Ref.
2,4-diacetylphloroglucinol	*Pseudomonas fluorescens* F113 / *Pythium* spp. / Damping off		Shanahan *et al*., 1992
Agrocin 84	*Agrobacterium radiobacter* / *Agrobacterium tumefaciens* / Crown gall		Kerr, 1980
Bacilomycin D	*Bacillus subtilis* AU195 / *Aspergillus flavus* / Contamination of aflatoxin		Moyne *et al*., 2001
Fengycin	*Bacillus amyloliquefaciens* FZB42 / *Fusarium oxysporum* / Rot		Koumoutsi *et al*., 2004
Herbicolin	*Pantoea agglomerans* C9-1 / *Erwinia amylovora* / Fire blight		Wright *et al*., 2001
Iturin A	*B. subtilis* QST713 / *Botrytis cinérea* y *R. solani* / Damping off		Paulitz y Belanger 2001
Mycosubtilin	*B. subtilis* BBG100 *Pythium* / *Gaeumannomyces graminis* var. *Tritici* / Damping off		Leclere *et al*., 2005
Phenazin	*P. fluorescens* 2-79 and 30-84 / *Gaeumannomyces graminis* var. *tritici* / Foot or crown rot		Thomashow *et al*., 1990

Table. 1. Continues...

Antibiotic	Source / Target pathogen / Illness	Chemical structure	Ref.
Pyoluteorin	*P. fluorescens* Pf-5 / *Pythium ultimum* y *R. solani* / Damping off		Howell y Stipanovic 1980
Xanthobaccin A	*Lysobacter* sp. strain SB-K88 / *Aphanomyces cochlioides* / Damping off		Islam *et al.*, 2010
Gliotoxin	*Trichoderma virens* / *Rhizoctonia solani* / Root rot		Wilhite *et al.*, 2001

ACKNOWLEDGEMENTS

To the National Council for Science and Technology in Mexico (CONACYT) for the fund granting the No. CVU: 958433. To the staff and technicians of the College of Postgraduates Campus Campeche for the services and assistance provided.

REFERENCES

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., & Pusch, G. D. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics, 9*(1), 75. https://doi.org/10.1186/1471-2164-9-75

Bachmann, B. O., Van Lanen, S. G., & Baltz, R. H. (2013). Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *Journal of Industrial Microbiology & Biotechnology, 41*(2), 175–184. https://doi.org/10.1007/s10295-013-1389-9

Bennett, P. M. (2004). Genome Plasticity. *Genomics, Proteomics, and Clinical Bacteriology*, 71–113. https://doi.org/10.1385/1-59259-763-7:071

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M. ., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., & Hornsby, T. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature, 417*(6885), 141–147. https://doi.org/10.1038/417141a

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M., & Weber, T. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research, 49*(W1), W29–W35. https://doi.org/10.1093/nar/gkab335

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., & Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research, 47*(W1), W81–W87. https://doi.org/10.1093/nar/gkz310

Cibichakravarthy, B., & Jose, P. A. (2021). Biosynthetic Potential of Streptomyces Rationalizes Genome-Based Bioprospecting. *Antibiotics, 10*(7), 873. https://doi.org/10.3390/antibiotics10070873

Clardy, J., Fischbach, M. A. and Walsh, C. T. (2006) 'New antibiotics from bacterial natural products', *Nature Biotechnology 2006* 24:12, 24(12), pp. 1541–1550. doi: 10.1038/nbt1266.

Colla, G. et al. (2018) 'Plant Growth-Promoting Rhizobacteria: Context, Mechanisms of Action, and Roadmap to Commercialization of Biostimulants for Sustainable Agriculture'. doi: 10.3389/fpls.2018.01473.

Compant, S., Duffy, B., Nowak, J., Clement, C., & Barka, E. A. (2005). Use of Plant Growth-Promoting Bacteria for Biocontrol of Plant Diseases: Principles, Mechanisms of Action, and Future Prospects. *Applied and Environmental Microbiology, 71*(9), 4951–4959. https://doi.org/10.1128/aem.71.9.4951-4959.2005

Cruz-Morales, P., Kopp, J. F., Martínez-Guerrero, C., Yáñez-Guerra, L. A., Selem-Mojica, N., Ramos-Aboites, H., Feldmann, J., & Barona-Gómez, F. (2016). Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomycetes. *Genome Biology and Evolution, 8*(6), 1906–1916. https://doi.org/10.1093/gbe/evw125

de Jong, A., van Hijum, S. A. F. T., Bijlsma, J. J. E., Kok, J., & Kuipers, O. P. (2006). BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Research, 34*(Web Server), W273–W279. https://doi.org/10.1093/nar/gkl237

Dobrindt, U., & Hacker, J. (2001). Whole genome plasticity in pathogenic bacteria. *Current Opinion in Microbiology, 4*(5), 550–557. https://doi.org/10.1016/s1369-5274(00)00250-2

Fischbach, M. A., & Walsh, C. T. (2006). Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chemical Reviews, 106*(8), 3468–3496. https://doi.org/10.1021/cr0503097

Hadjithomas, M., Chen, I-Min. A., Chu, K., Ratner, A., Palaniappan, K., Szeto, E., Huang, J., Reddy, T. B. K., Cimermančič, P., Fischbach, M. A., Ivanova, N. N., Markowitz, V. M., Kyrpides, N. C., & Pati, A. (2015). IMG-ABC: A Knowledge Base to Fuel Discovery of Biosynthetic Gene Clusters and Novel Secondary Metabolites. *MBio, 6*(4). https://doi.org/10.1128/mbio.00932-15

Howell, C. R. (1980). Suppression of *Pythium ultimum*-Induced Damping-Off of Cotton Seedlings by *Pseudomonas fluorescens* and its Antibiotic, Pyoluteorin. *Phytopathology, 70*(8), 712. https://doi.org/10.1094/phyto-70-712

Islam, Md. T., Hashidoko, Y., Deora, A., Ito, T., & Tahara, S. (2005). Suppression of Damping-Off Disease in Host Plants by the Rhizoplane Bacterium Lysobacter sp. Strain SB-K88 Is Linked to Plant Colonization and Antibiosis against Soilborne Peronosporomycetes. *Applied and Environmental Microbiology, 71*(7), 3786–3796. https://doi.org/10.1128/aem.71.7.3786-3796.2005

Kersten, R. D., Yang, Y.-L., Xu, Y., Cimermancic, P., Nam, S.-J., Fenical, W., Fischbach, M. A., Moore, B. S., & Dorrestein, P. C. (2011). A mass spectrometry–guided genome mining approach for natural product peptidogenomics. *Nature Chemical Biology, 7*(11), 794–802. https://doi.org/10.1038/nchembio.684

Köhl, J., Kolnaar, R., & Ravensberg, W. J. (2019). Mode of Action of Microbial Biological Control Agents Against Plant Diseases: Relevance Beyond Efficacy. *Frontiers in Plant Science*, 10. https://doi.org/10.3389/fpls.2019.00845

Koumoutsi, A., Chen, X.-H.., Henne, A., Liesegang, H., Hitzeroth, G., Franke, P., Vater, J., & Borriss, R. (2004). Structural and Functional Characterization of Gene Clusters Directing Nonribosomal Synthesis of Bioactive Cyclic Lipopeptides in Bacillus amyloliquefaciens Strain FZB42. *Journal of Bacteriology, 186*(4), 1084–1096. https://doi.org/10.1128/jb.186.4.1084-1096.2004

LeclèreV., BéchetM., Adam, A., Guez J.-S., Wathelet, B., Ongena, M., Thonart, P., GancelF., Chollet-ImbertM., & Jacques, P. (2005). Mycosubtilin Overproduction by *Bacillus subtilis* BBG100 Enhances the Organism's Antagonistic and Biocontrol Activities. *Applied and Environmental Microbiology, 71*(8), 4577–4584. https://doi.org/10.1128/aem.71.8.4577-4584.2005

Lee, N., Hwang, S., Kim, J., Cho, S., Palsson, B., & Cho, B.-K. (2020). Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in Streptomyces. *Computational and Structural Biotechnology Journal*, 18, 1548–1556. https://doi.org/10.1016/j.csbj.2020.06.024

Li, M. H., Ung, P. M., Zajkowski, J., Garneau-Tsodikova, S., & Sherman, D. H. (2009). Automated genome mining for natural products. *BMC Bioinformatics, 10*(1), 185. https://doi.org/10.1186/1471-2105-10-185

Liu, W.-T., Lamsa, A., Wong, W. R., Boudreau, P. D., Kersten, R., Peng, Y., Moree, W. J., Duggan, B. M., Moore, B. S., Gerwick, W. H., Linington, R. G., Pogliano, K., & Dorrestein, P. C. (2014). MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in *Streptomyces roseosporus*. *The Journal of Antibiotics, 67*(1), 99–104. https://doi.org/10.1038/ja.2013.99

Manchester M. & Anisha A. (2017) Advances in Virus Research Chapter Two - Metabolomics: Strategies to Define the Role of Metabolism in Virus Infection and Pathogenesis, 98, 57-81. https://doi.org/10.1016/bs.aivir.2017.02.001.

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., & Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research, 39*(suppl_2), W339–W346. https://doi.org/10.1093/nar/gkr466

Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Chooi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., & Hadjithomas, M. (2015). Minimum Information about a Biosynthetic Gene cluster. *Nature Chemical Biology, 11*(9), 625–631. https://doi.org/10.1038/nchembio.1890

Moyne, A.-L.., Shelby, R., Cleveland, T. E., & Tuzun, S. (2001). Bacillomycin D: an iturin with antifungal activity against Aspergillus flavus. *Journal of Applied Microbiology, 90*(4), 622–629. https://doi.org/10.1046/j.1365-2672.2001.01290.x

Pal, K. K., & McSpadden Gardener, B. (2006). Biological Control of Plant Pathogens. *The Plant Health Instructor, 1*(PHI-A-2006-1117-02). https://doi.org/10.1094/phi-a-2006-1117-02

Paulitz, T. C. and Bélanger, R. R. (2001) 'Biological control in greenhouse systems', *Annual review of phytopathology*, 39, pp. 103–133. doi: 10.1146/ANNUREV.PHYTO.39.1.103.

Raaijmakers, J. M., & Mazzola, M. (2012). Diversity and Natural Functions of Antibiotics Produced by Beneficial and Plant Pathogenic Bacteria. *Annual Review of Phytopathology, 50*(1), 403–424. https://doi.org/10.1146/annurev-phyto-081211-172908

Ren, H., Shi, C., & Zhao, H. (2020). Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways. *IScience, 23*(1), 100795. https://doi.org/10.1016/j.isci.2019.100795

Sela, I., Wolf, Y. I., & Koonin, E. V. (2018). Genome plasticity, a key factor of evolution in prokaryotes. *BioRxiv*. https://doi.org/10.1101/357400

Shanahan, P., O'Sullivan, D. J., Simpson, P., Glennon, J. D., & O'Gara, F. (1992). Isolation of 2,4-Diacetylphloroglucinol from a Fluorescent Pseudomonad and Investigation of Physiological Parameters Influencing Its Production. *Applied and Environmental Microbiology, 58*(1), 353–358. https://doi.org/10.1128/aem.58.1.353-358.1992

Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology, 7*(2), 285–295. https://doi.org/10.1016/s1367-5931(03)00020-6

Siupka, P., Piński, A., Babicka, D., & Piotrowska-Seget, Z. (2020). Genome Mining Revealed a High Biosynthetic Potential for Antifungal *Streptomyces* sp. S-2 Isolated from Black Soot. *International Journal of Molecular Sciences, 21*(7), 2558. https://doi.org/10.3390/ijms21072558

Skinnider, M. A., Merwin, N. J., Johnston, C. W., & Magarvey, N. A. (2017). PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Research, 45*(W1), W49–W54. https://doi.org/10.1093/nar/gkx320

Smith, K. P. (1993). Suppression of Cottony Leak of Cucumber with *Bacillus cereus* Strain UW85. *Plant Disease, 77*(2), 139. https://doi.org/10.1094/pd-77-0139

Starcevic, A., Zucko, J., Simunkovic, J., Long, P. F., Cullum, J., & Hranueli, D. (2008). ClustScan : an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Research, 36*(21), 6882–6892. https://doi.org/10.1093/nar/gkn685

Studholme, D. J. (2016). Genome Update. Let the consumer beware: Streptomyces genome sequence quality. *Microbial Biotechnology, 9*(1), 3–7. https://doi.org/10.1111/1751-7915.12344

Thomashow, L. S., Weller, D. M., Bonsall, R. F., & Pierson, L. S. (1990). Production of the Antibiotic Phenazine-1-Carboxylic Acid by Fluorescent Pseudomonas Species in the Rhizosphere of Wheat. *Applied and Environmental Microbiology, 56*(4), 908–912. https://doi.org/10.1128/aem.56.4.908-912.1990

Vurukonda, S. S. K. P., Giovanardi, D., & Stefani, E. (2018). Plant Growth Promoting and Biocontrol Activity of *Streptomyces* spp. as Endophytes. *International Journal of Molecular Sciences, 19*(4), 952. https://doi.org/10.3390/ijms19040952

Walsh, C. T., & Fischbach, M. A. (2010). Natural Products Version 2.0: Connecting Genes to Molecules. *Journal of the American Chemical Society, 132*(8), 2469–2493. https://doi.org/10.1021/ja909118a

Weber, T., Rausch, C., Lopez, P., Hoof, I., Gaykova, V., Huson, D. H., & Wohlleben, W. (2009). CLUSEAN: A computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *Journal of Biotechnology, 140*(1-2), 13–17. https://doi.org/10.1016/j.jbiotec.2009.01.007

Wilhite, S. E., Lumsden, R. D., & Straney, D. C. (2001). Peptide Synthetase Gene in *Trichoderma virens*. *Applied and Environmental Microbiology, 67*(11), 5055–5062. https://doi.org/10.1128/aem.67.11.5055-5062.2001

Wright, S. A. I., Zumoff, C. H., Schneider, L., & Beer, S. V. (2001). *Pantoea agglomerans* Strain EH318 Produces Two Antibiotics That Inhibit *Erwinia amylovora In Vitro*. *Applied and Environmental Microbiology, 67*(1), 284–292. https://doi.org/10.1128/aem.67.1.284-292.2001

Zhao, S., Sakai, A., Zhang, X., Vetting, M. W., Kumar, R., Hillerich, B., San Francisco, B., Solbiati, J., Steves, A., Brown, S., Akiva, E., Barber, A., Seidel, R. D., Babbitt, P. C., Almo, S. C., Gerlt, J. A., & Jacobson, M. P. (2014). Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks. *ELife*, 3. https://doi.org/10.7554/elife.03275

Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E., & Jensen, P. R. (2012). The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS ONE, 7*(3), e34064. https://doi.org/10.1371/journal.pone.0034064