Article

An Insight Into the Intraspecific Variation of Biosynthetic Gene Clusters Between Strains of *Burkholderia thailandensis* spp.

João Luiz Baldim* and Marisi Gomes Soares*

Instituto de Química, Universidade Federal de Alfenas, 37130-000 Alfenas-MG, Brazil

The present study aimed to investigate the intraspecific variation of biosynthetic gene clusters (BGCs) in different strains of *Burkholderia thailandensis* in order to guide natural products (NPs) discovery process. Species from the genus *Burkholderia* are emerging as promising species due to their biosynthetic potential. Through genome-mining strategies, it was able to identify that *B. thailandensis* strains present major genome variation between chromosomes I and II and the standard. The positioning of BGCs also differs when comparing each chromosome. Classical pathways as well as terpene and bacteriocins were commonly identified to all of them and BGCs related to the production of nonribosomal peptides and polyketides compounds are often noticed. In addition, hybrids BGCs were identified as using large amount of replicon information. Among all species studied, the strain MSMB121 showed greater potential for biosynthesizing novel natural products and after phylogenetic analysis, the likelihood of recognizing sites of novelties was assigned.

Keywords: *Burkholderia thailandensis*, genome-mining, natural products discovery

Introduction

Species of the genus *Burkholderia* have emerged as owing promising biosynthetic capability for diverse natural products (NPs). Recently, a remarkable study about the potential of microorganisms in biosynthesizing NPs pointed that proteobacteria species present large number of biosynthetic gene clusters (BGCs).[1] This global analysis included species from the genus *Pseudomonas* spp. and *Burkholderia* spp. as containing the majority of BGCs counting for proteobacteria representatives. However, the great interest is related to *Burkholderia* spp., once *Pseudomonas* spp. are extensively studied. About the likelihood of producing novel NPs, *Burkholderia* genomes present, statistically, higher percentage of thiotemplate modular systems than those of bacilli, cyanobacteria, myxobacteria and fungi, and is only second to that of actinobacteria.[2] These modular systems are related to the production of many classes of pharmaceutical compounds, including polyketide synthase (PKS)- and nonribosomal peptide synthase (NRPS)-related products.[3]

According to the German Collection of Microorganisms and Cell Cultures (DSMZ) database, the genus *Burkholderia* comprises more than 90 species. These species inhabit the most diverse types of ecological niches such as soil, water, rhizosphere and plant surface.[4]

NPs from *Burkholderia* spp. are structurally and functionally diverse, comprising benzoquinone; lactone; and polyene compounds, lasso peptides, nonribosomal peptides, statins, and polyketides. These compounds present important biological activities. In addition, some of these small molecules from *Burkholderia* sp. have entered as drug candidates to preclinical evaluation.[5]

Burkholderia thailandensis (*B. thailandensis*) E264 presented a high level of similarity to the BGC BTH-II0204-207 from *Burkholderia pseudomallei* (*B. pseudomallei*) K96243 related to the production of betulinan/terferol analogues.[6] Experiments indicate that the compound isolated from this BGC, BTH-II0204-207:A (**1**), is potent PDE4 inhibitor.[7] It is worth to highlight that the first compound considered as PDE4 inhibitor to the treatment of chronic obstructive pulmonary disease was approved by US Food and Drug Administration (FDA) in 2011.[7,8] A specific BGC of *B. thailandensis* E264 also presents high level of homology to the BGC responsible to the production of class II lasso peptides of *E. coli*. Further studies with the strain E264 provided the discovery of a class II lasso peptide called capstruin (structure not shown) that presents antimicrobial activities.[9] Species as *B. mallei* ATCC 23344, *B. pseudomallei* K96243

*e-mail: marisigs@gmail.com; jotaelebaldim@gmail.com

and *B. thailandensis* E264 presented similar BGCs that encode hybrid PKS-NRPS pathways presenting unusual domains that provided malleilactone (**2**) and burkholderic acid (**3**).[10,11] These compounds presented, respectively, moderate activity against gram-positive bacteria and weak cytotoxicity.[10,11] Genome-guided approaches led to the isolation of an interesting class of polyene amides named thailandamides from *B. thailandensis* E264, related to a hybrid 17-module *trans*-AT PKS-NRPS pathway. Thailandamide lactone (**4**) presented moderate antiproliferative activity against human tumor cell lines.[12,13] The thailandepsins also isolated from *B. thailandensis* E264 are related to FK228 BGC in *Chromobacterium violaceum*.[14] FK228 is an FDA approved anticancer drug related to the treatment of refractory cutaneous and peripheral T cell lymphoma.[15] The thailandepsins A (**5**) and B (**6**) also possess growth inhibition characteristics to different cancer cell lines, as well as colon, melanoma and renal cancer cells.[16] However, the mechanism of action needs to be further studied.

Other interesting NPs from *B. thailandensis* MSMB43 are the thailanstatins A-B (**7-8**, Figure 1), belonging to the FR901464-family of microbial products that have a pyran ring heavily substituted with different groups in one end and an acetyl group at the other end. The biggest difference between thailanstatins and FR901464 is the lack on the hydroxyl group and the presence of a carboxyl moiety, resulting in higher stability to thailanstatins.[17] Thailanstatins inhibit mRNA splicing and are related to antiproliferative activities against human cancer cell lines.[17] Thailanstatins compounds also showed therapeutic application against glaucoma due to modulation of glucocorticoid receptor splicing process.[18]

At a glance, genome-mining strategies are taking important role in the field of NPs discovering. Computational biology and genomics are changing the approach of NPs research by understanding specifically how nature produces compounds.[19] In this sense, species of the genus *Burkholderia* are providing an extensive number of NPs after genome-guided strategies applied. *B. thailandensis* has presenting an increasing interest due to its biosynthetic capability.[20] Since the most studied strain is the *B. thailandensis* E264, all available genomic sequences related to this species were investigated in order to evaluate the potential of other *B. thailandensis* strains shedding lights to their biosynthetic potential, guiding NPs discovery process.
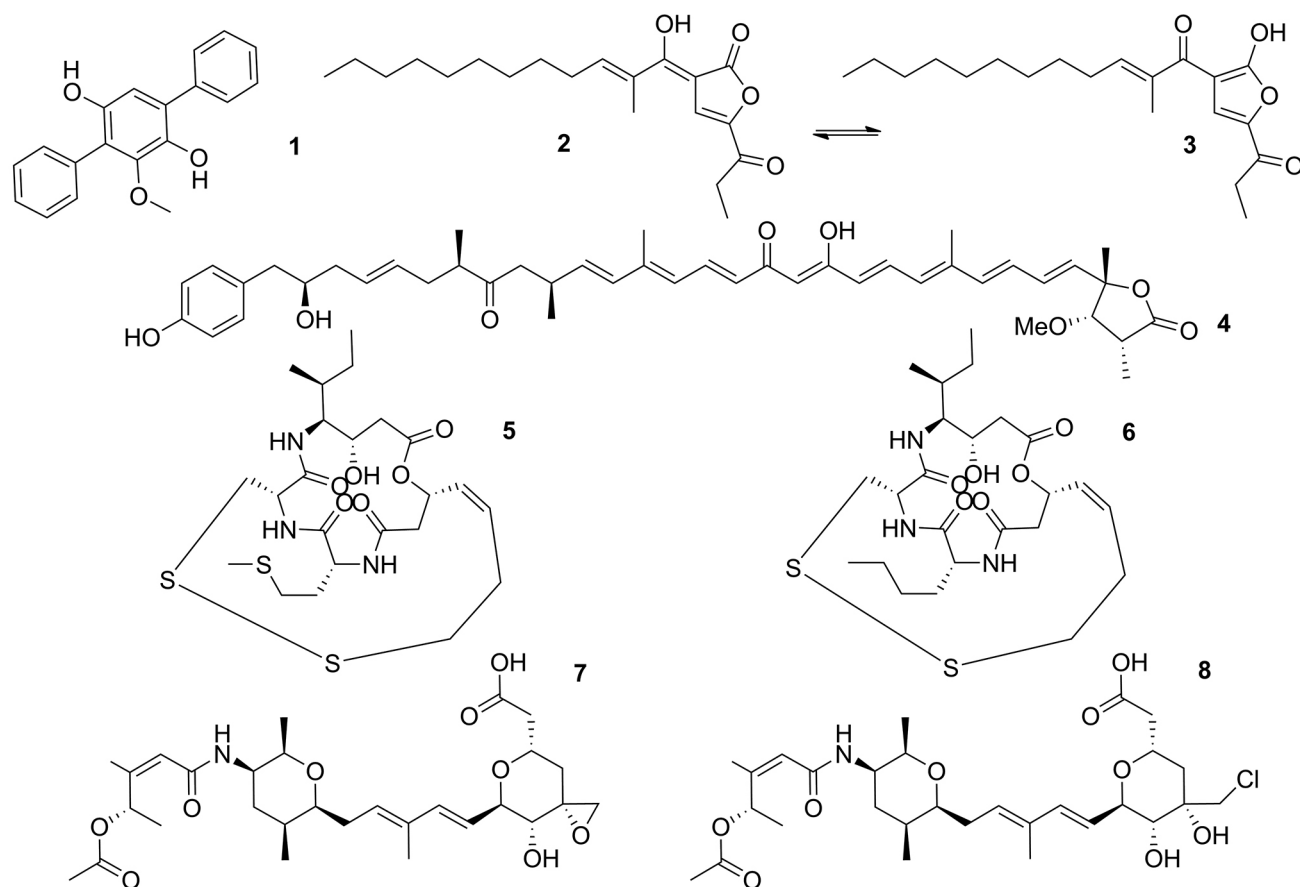

Figure 1. Natural products (NPs) isolated from *B. thailandensis* and their singular structures.

Experimental

Dataset

Genomic information related to *B. thailandensis* strains were downloaded from the National Center of Biotechnology Information (NCBI) genome database.[21] Chromosome I and II sequences were chosen in order to correlate the differentiation between them for all available sequences. A complete list of *B. thailandensis* strains containing accession numbers is available at Supplementary Information (Table 1). All sequence files were downloaded according to the available data at NCBI on November 2015. All analyses used Chr1 results of in Actinobacteria representative *Streptomyces coelicolor* A3(2) [*S. coelicolor* A3(2)] as a control in order to compare similarities and uniqueness of *B. thailandensis* strains (proteobacteria). *S. coelicolor* A3(2) is a well-studied antibiotic-producing bacterium (accession number NC_003888.3).

BGCs finder and classification

BGCs predictions were made using antiSMASH 3.0.[22] Investigations were made in order to search for, among other results, an overview in genomic information allowing detecting classes of compounds, level of genomic similarities to already known compounds, and their core structures, including the comparative gene cluster analysis. NRPS-PKS predicted results were elected according to the consensus between the classic comparisons made by the database using NRPSpredictor2,[23] Stachelhaus code[24] and Minowa.[25]

Network profile

Datasets were submitted to Cytoscape 3.3.0[26] to the creation of a statistical visual correlation between Chr1 and Chr2 of *B. thailandensis* strains and the standard. The algorithm used in order to correlate results of chromosomes was AllegroLayout 2.2.3[27] with Fruchterman-Reingold Layout. Networks and clusters were formed for each step of analyses: (*i*) Chr *vs*. type of compound; and (*ii*) Chr *vs*. homology. Visual statistics were built correlating results to their respective degrees.

Jaccard index calculation

Jaccard indexes (JI) were calculated in order to correlate the level of similarities between strains of *B. thailandensis* and compounds classes, structure and homology to known BGCs. Raw results were overlapped and clustered. The clustering settings followed hierarchical parameters, using values from JI scores and Euclidean distance as standard method. Analyses were proceeded using the software Gitools 2.2.3.[28]

Dendrogram similarity

Taxonomic analysis about diversification of *B. thailandensis* BGCs was proceeded using sequences identified by antiSMASH results. BGCs sequences were downloaded as .fasta format and named according to their respective species, chromosome and type of compound homology. Nucleotide sequences were aligned with the purpose of comparing the distribution of BGCs of *B. thailandensis* strains and their homology levels at the phylogenetic level. The dendrogram was built in Mega6.[29] The analysis involved 100 nucleotide sequences aligned by ClustalW using default parameters.[30] The distribution of BGCs classes was inferred using the neighbor joining method.[31] Comparisons between higher homologies levels and BGCs grouping were calculated with the purpose of differentiate BGCs characteristics according to their NPs most probable related structure.

Results and Discussion

Most bacteria have one or two circular replicons that encode set of genes for the most diverse functions, including the production of NPs. In the case of genes encoding information of the biosynthesis of NPs, these sets of genes are called BGCs.[32] These BGCs encode diverse types of information related to enzymes, regulatory proteins and transporters that are essential to the biosynthetic machinery of a given metabolite. BGCs data also allows to mine genomes and identify sets of genes that participate in a specific biosynthetic pathway by computational analysis predicting their final products.

B. thailandensis is a model microorganism of the genus *Burkholderia* for the investigation of NPs biosynthetic capability. Due to its potential, the strain *B. thailandensis* E264 is well studied and several research groups have shown different methods to obtain NPs from it. In this sense, other *B. thailandensis* strains are also promising, since they share a significant level of DNA similarity.

Comparisons between Chr1 and Chr2 of *B. thailandensis* strains along with *S. coelicolor* A3(2) and their related types of compounds are showed in Figure 2a. Network results showed that both chromosomes of *B. thailandensis* putatively produce unique chemistry when compared to the standard.

The most probable class of compounds from these species are correlated to NRPS, terpene, and T1PKS.

As usual, all chromosomes encode information to BGCs correlated to the production of bacteriocins, that are proteinaceous toxins naturally produced by bacteria in order to colonize the environment in which they occur.[33]

Chr1 and Chr2 of *B. thailandensis* do not correlate themselves to their class of compounds other than the four classes cited above and Chr2 presents major diversity of classes. Due to evolutionary questions related to the production of specific compounds, these classes of compounds are different between replicons (Figure 2b). In addition, independently of the four common classes of compounds correlated to all species, compounds linked to *S. coelicolor* A3(2) presented no direct correlation to Chr1 or Chr2 of *B. thailandensis* strains.

Core structures produced by *B. thailandensis* strains do not present similarities when comparing Chr1 and Chr2. Their NRPSs machinery seems to not assembly the same monomers due to different levels of evolution between these two replicons. JI levels of core structures are also very small (Figure 3). However, when investigating one chromosome at time, they present genomic information pointing to the production of similar structures. In the case of different species, as reported in the literature, different substituents could be present in the same class of compounds, leading to different NPs including improvements of their biological activities.[34,35] These results are well explained in Figure 4.

The BGCs associated to the Chr1 of *B. thailandensis* species generally encode information for NRPSs to assembly monomers such as alanine-arginine; cysteine-threonine and ornithine-aspartate-serine; while Chr2 assembles cysteine-cysteine, valine-glycine, and malate-aspartate most of the time. Chr2 of *B. thailandensis* strains also present higher level of genomic information related to hybrid pathways.

Lately, the genus *Burkholderia* has shown to produce differentiated NPs, even in known classes of compounds.[36] Due to this characteristic, the use of *S. coelicolor* A3(2) as standard was successful in order to confirm uniqueness of *B. thailandensis* strains. The BGCs of *B. thailandensis* showed the lowest values of JI when compared to those of *S. coelicolor* A3(2). These values suggest that *B. thailandensis* species are highly different to the standard in the potential of biosynthesizing NPs.

In addition, core structures identified from Chr1 and Chr2 of *B. thailandensis* seems to not be originated from high levels of nucleotide homology to cluster themselves with similar scores, implying that they are independent in how they work in order to biosynthesize NPs.
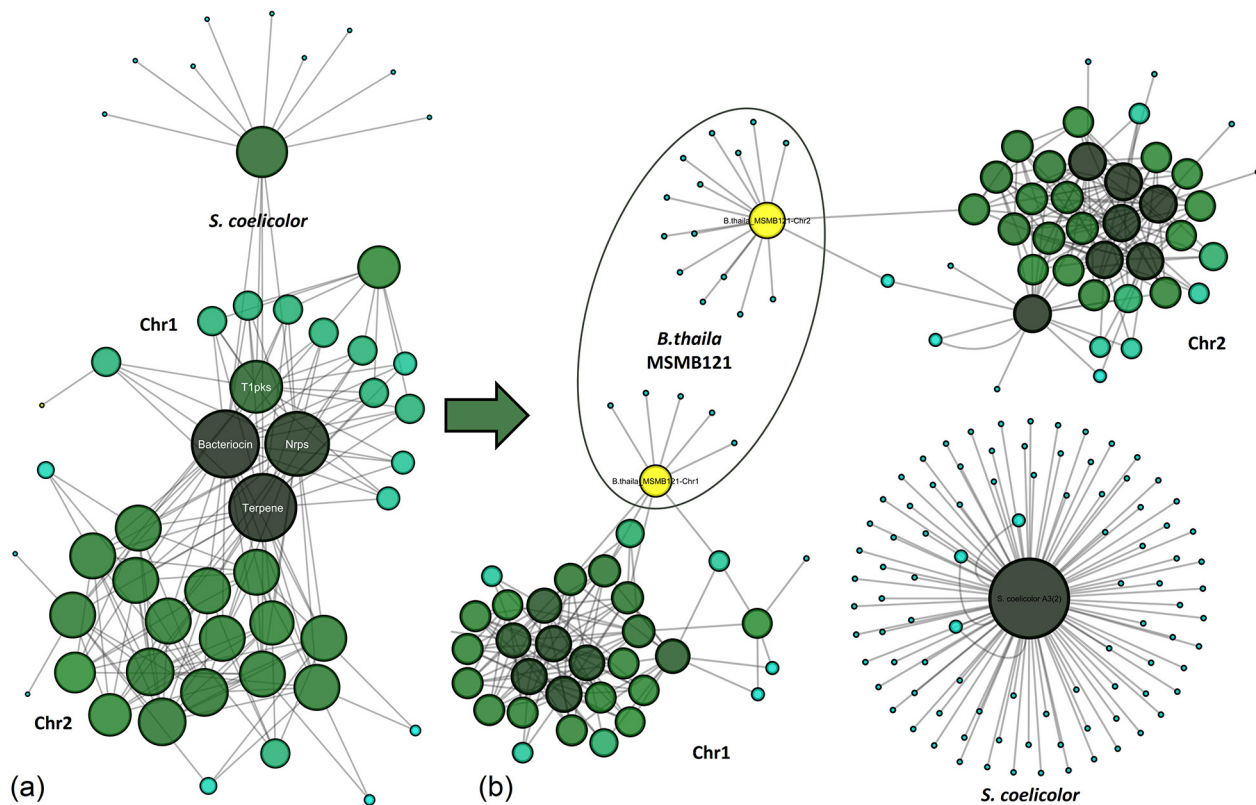

Figure 2. Network correlating replicons of *B. thailandensis* strains according to their classes of compounds and level of similarity to already known compounds. (a) Type of compounds and their connection between Chr1, Chr2 and *S. coelicolor* A3(2); and (b) homology between species investigated highlighting the most dissimilar strain. *S. coelicolor* A3(2), the standard strain show non correlation to *B. thailandensis* strains homologies.
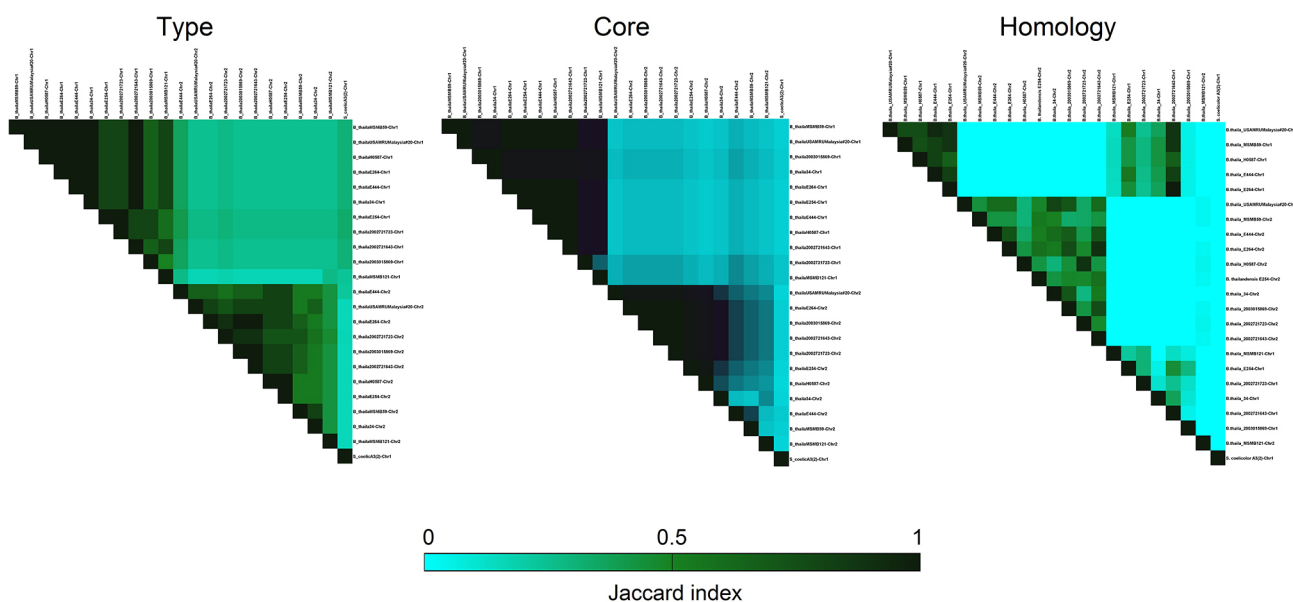
Figure 3. Overlapping Chr1 and Chr2 at three different levels of comparisons. Type, core and homology correlations are showed in Jaccard index scores. Results are calculated based on comparisons with *S. coelicolor* A3(2) strain.

One interesting detail about all analysis is that the strain MSMB121 presents higher levels of differentiation to all *Burkholderia* strains studied in this work. Chr1 and Chr2 of MSMB121 presented, respectively, 4 and 14 unique levels of homology to all other BGCs. These 18 levels of similarity do not present connection to other species in this work, suggesting that novel NPs from this strain could have structural moieties similar to their respective known NPs. Preliminarily, these observations suggest that *B. thailandensis* MSMB121 holds greater chances of biosynthesize novel NPs than others *B. thailandensis* strains. For comparisons, the second position in this analysis is *B. thailandensis* MSB59, containing 4 exclusive levels of similarity summing results of both chromosomes. Different chromosomal levels of similarity could imply directly in monomers flexibility leading to improvements in biosynthetic steps ending in different NPs.[34,35] This could be explained observing different clusters related to the production of thailanstatins. Their different levels of similarity are related to thailanstatins-like compounds with different substituents or side chains.

Hierarchical homology analysis showed that the Chr2 of *B. thailandensis* MSMB121 and *S. coelicolor* A3(2) are grouped in the same sub-branch. This strain is the most similar to the standard between all *Burkholderia* strains studied in this work. Genomic comparisons between the standard and the strain MSMB121 would provide useful information about their NPs potential. Details are available at Supplementary Information (Figure S1).

Results of hierarchical analyses using Euclidean distance further investigating NRPs- and PKs-related structures showed that the largest group expressing higher levels of similarity to the production of NPs is composed by the strains E444, E264, and H0587 (details are available at Supplementary Information, Figure S2). Their core structure is based on NRPS-related BGCs assembling alanine and arginine most of the time and presenting 4% of nucleotide homology to the BGC related to azinomycin B. On the other hand, the strains grouped by their similarities in the Chr2 are the strains 2002721723, E264, 2002721643, and 2003015869. These strains are strongly correlated to the BGC that encodes information for the production of malleobactin (homology greater than 90%), pyochelin (homology of 100% for all strains) and bactobolin (homology of 100% for all). These results confirmed that Chr1 and Chr2 of *B. thailandensis* strains are independent in the way they encode information for the biosynthesis of NPs and their genomes seems to be quite dissimilar to the standard. Since genome-mining results showed that *B. thailandensis* strains differ in the level of similarity of their BGCs, all sequences classified as PKS- and NRPS-related compounds were further investigated at the genomic level.

BGCs of both chromosomes of *B. thailandensis* strains, as well as *S. coelicolor* A3(2), were aligned according to their nucleotide sequences. Divergent sequences expressing same core structures hold interesting features for the production of novel compounds due to the possibility of eliciting silent bacterial gene clusters after investigation of their biosynthetic power, resistance and metabolic profiles.[37]

In the dendrogram, the differences in branch length of a specific subtree referring to a NP indicate the likelihood

of biosynthesizing unique compounds to those already isolated, due to the variable levels of homology between BGCs. These small singularities in each group are identified as possessing different sum of branch length explaining nucleotide modifications in all BGCs.

In the case of the subtree related to malleilactone (Figure 4), there are two groups of BGCs, one (starting at position 1) containing two and other (starting at position 58) containing nine different BGCs. To the biggest subtree (position 58), there are four sequences presenting high level of DNA homology placed in the same sub-branch. On the other hand, the five other BGCs present dissimilar alignment. These evidences suggest that enzymes related to these BGCs could lead to different malleilactone-related compounds in the biosynthetic steps. The same might occur to all other subtrees, tracking *B. thailandensis* strains potential in biosynthesize novel NPs.

The alignment of *S. coelicolor* BGCs are placed individually. When correlated to *B. thailandensis* strains, they are placed along with BGCs related to the production of azinomycin. The highest level of homology encountered

to *B. thailandensis* BGCs containing information to the production of azinomycin was 4%, suggesting that the information present in these clusters mostly lead to the biosynthesis of other compounds.

These dissimilar branch length explains how compounds of the same class are biosynthesized by similar BGCs and the occurrence of different side chains or substituents. In some cases, there are the possibility of moieties presenting similar characteristics (polar or non-polar amino acids, for example) to be placed according to enzymatic steps involved in the biosynthesis of NPs.[38] In the case of *B. thailandensis* MSM121, the range of substitution per site, related to each BGC is largest than other strains. After aligned, the group of BGCs related to malleilactone biosynthesis (position 58) presents the strain MSMB121 as possessing the larger differentiation compared to others in this subtree, followed by the strains H0587 and USAMRUMalasya*20. This could be observed in all *S. coelicolor* A3(2) BGCs, that present lager genomic differentiation compared to all *Burkholderia* strains. As genomes are strictly connected to biosynthetic pathways
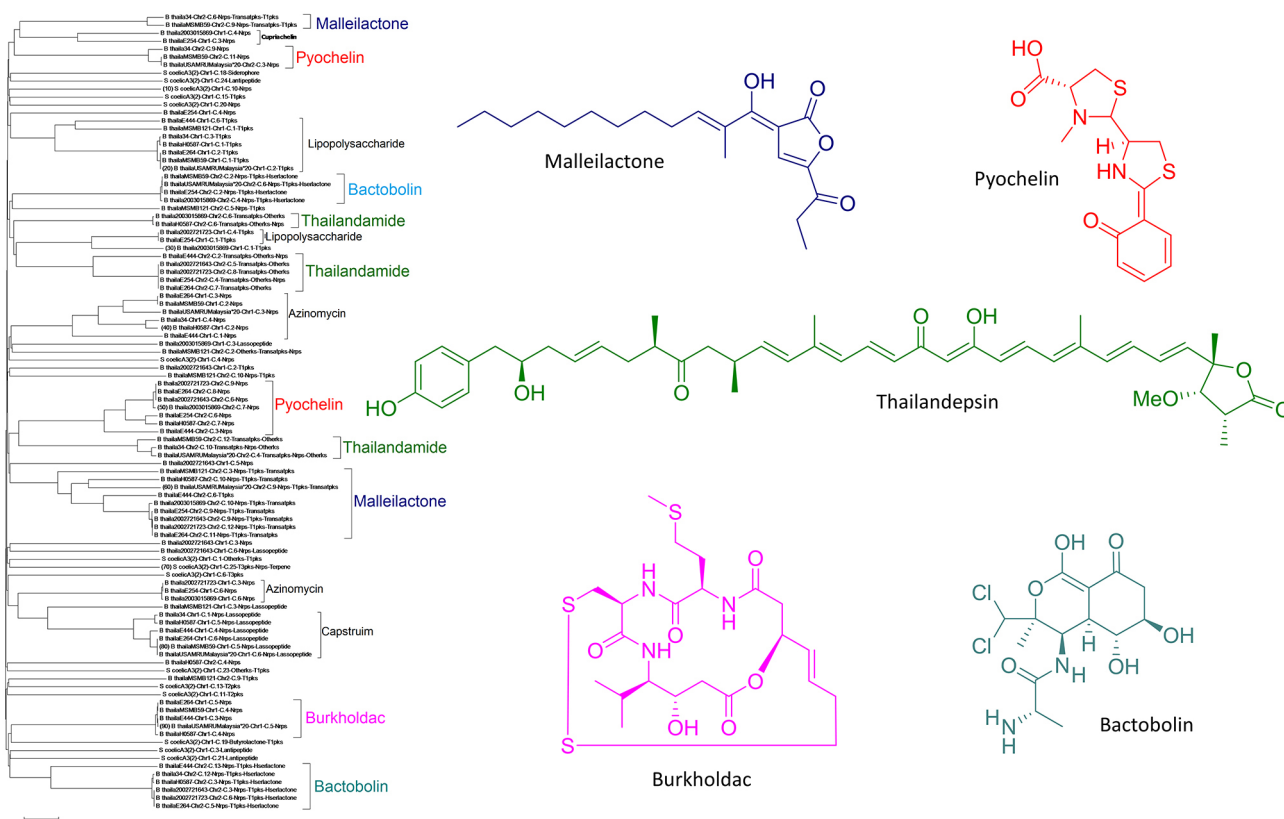

Figure 4. Distribution of *B. thailandensis* and *S. coelicolor* biosynthetic gene clusters (BGCs) related to nonribosomal peptide synthase (NRPS)- and polyketide synthase (PKS)-related compounds. Subtrees related to a specific compound were highlighted with brackets and each main compound was colorized according to its subtree. Differentiation in sum length relates the likelihood of different compounds based on the structure of the main compound. The evolutionary history was inferred using the neighbor-joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences *per* site. The analysis involved 100 amino acids sequences. All positions containing gaps and missing data were eliminated. There were a total of 923 positions in the final dataset. Alignment was conducted in MEGA6.[29]

and the production of NPs, small homologies (or higher levels of substitution *per* sites) in genomic information leads to the production of different NPs. Thus, results of network and phylogeny are direct correlated presenting no linkage to any level of similarity to BGCs related to the standard strain. These observations explain the reason of NPs from *Burkholderia* strains be very dissimilar to the standard suggesting that *B. thailandensis* is a good reservoir of novel NPs.

Conclusions

Biosynthetic pathways are being strictly correlated to their genomic information in order to understand how diverse mechanisms are encoded, including the biosynthesis of NPs. Genomic-guided strategies are part of a new way of understanding how NPs are produced and how different they could be when investigating different strains of a given species based on their chromosomal composition. In addition, experiments adopted in this work shed light to how NPs discovery processes are supported by stat-of-the-art techniques in order to give information that in the past was impossible to achieve making the process of discovery highly rationale. After analyzes of all *B. thailandensis* strains, it was possible to infer that there are great chances of isolating novel NPs using specific culture media due to their biosynthetic capability and the likelihood of side chains modifications of known NPs due to their phylogeny. Finally, the differentiation in their alignment revealed that similar NPs belonging to known classes often occur, increasing the potential of this species in biosynthesizing novel compounds.

Supplementary Information

Supplementary data containing dendrograms and accession numbers are available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

References

1. Cimermancic, P.; Medema, M. H. H.; Claesen, J.; Kurita, K.; Brown, L. C. W.; Mavrommatis, K.; Pati, A.; Godfrey, P. A. A.; Koehrsen, M.; Clardy, J.; Birren, B. W. W.; Takano, E.; Sali, A.; Linington, R. G. G.; Fischbach, M. A. A.; Brown, L. C. W.; Mavrommatis, K.; Pati, A.; Godfrey, P. A. A.; Koehrsen, M.; Clardy, J.; Birren, B. W. W.; Takano, E.; Sali, A.; Linington, R. G. G.; Fischbach, M. A. A.; *Cell* **2014**, *158*, 412.

2. Minowa, Y.; Araki, M.; Kanehisa, M.; *J. Mol. Biol.* **2007**, *368*, 1500.

3. Diminic, J.; Zucko, J.; Ruzic, I. T.; Gacesa, R.; Hranueli, D.; Long, P. F.; Cullum, J.; Starcevic, A.; *J. Ind. Microbiol. Biotechnol.* **2013**, *40*, 653.

4. Ludovic, V.; Groleau, M.-C.; Dekimpe, V.; Déziel, E.; *J. Microbiol. Biotechnol.* **2007**, *17*, 1407.

5. Wilson, A. J.; Cheng, Y.-Q.; Khabele, D.; *J. Ovarian Res.* **2012**, *5*, 12.

6. El-Elimat, T.; Figueroa, M.; Raja, H. A.; Graf, T. N.; Adcock, A. F.; Kroll, D. J.; Day, C. S.; Wani, M. C.; Pearce, C. J.; Oberlies, N. H.; *J. Nat. Prod.* **2013**, *76*, 382.

7. Biggins, J. B.; Liu, X.; Feng, Z.; Brady, S. F.; *J. Am. Chem. Soc.* **2011**, *133*, 1638.

8. Taegtmeyer, A. B.; Leuppi, J. D.; Kullak-Ublick, G. A.; *Swiss Med. Wkly.* **2012**, *142*, 1.

9. Knappe, T. A.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. A.; *J. Am. Chem. Soc.* **2008**, *130*, 11446.

10. Franke, J.; Ishida, K.; Hertweck, C.; *Angew. Chem., Int. Ed.* **2012**, *51*, 11611.

11. Biggins, J. B.; Ternei, M. A.; Brady, S. F.; *J. Am. Chem. Soc.* **2012**, *134*, 13192.

12. Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J.; *Nat. Biotechnol.* **2008**, *26*, 225.

13. Ishida, K.; Lincke, T.; Behnken, S.; Hertweck, C.; *J. Am. Chem. Soc.* **2010**, *132*, 13966.

14. Cheng, Y. Q.; Yang, M.; Matter, A. M.; *Appl. Environ. Microbiol.* **2007**, *73*, 3460.

15. VanderMolen, K. M.; McCulloch, W.; Pearce, C. J.; Oberlies, N. H.; *J. Antibiot.* **2011**, *64*, 525.

16. Wang, C.; Henkes, L. M.; Doughty, L. B.; He, M.; Wang, D.; Meyer-Almes, F. J.; Cheng, Y. Q.; *J. Nat. Prod.* **2011**, *74*, 2031.

17. Liu, X.; Biswas, S.; Berg, M. G.; Antapli, C. M.; Xie, F.; Wang, Q.; Tang, M. C.; Tang, G. L.; Zhang, L.; Dreyfuss, G.; Cheng, Y. Q.; *J. Nat. Prod.* **2013**, *76*, 685.

18. Jain, A.; Liu, X.; Wordinger, R. J.; Yorio, T.; Cheng, Y.-Q.; Clark, A. F.; *Invest. Ophthalmol. Visual Sci.* **2013**, *54*, 3137.

19. Medema, M. H.; Cimermancic, P.; Sali, A.; Takano, E.; Fischbach, M. A.; *PLoS Comput. Biol.* **2014**, *10*, e1004016.

20. Zhuo, Y.; Liu, L.; Wang, Q.; Liu, X.; Ren, B.; Liu, M.; Ni, P.; Cheng, Y.-Q.; Zhang, L.; *J. Bacteriol.* **2012**, *194*, 4749.

21. http://www.ncbi.nlm.nih.gov/genome accessed in March 2016.

22. Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H.; *Nucleic Acids Res.* **2015**, *43*, W237.

23. Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.;

Kohlbacher, O.; *Nucleic Acids Res. Suppl.* **2011**, *39 (suppl. 2)*, W362; Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H.; *Nucleic Acids Res.* **2005**, *33*, 5799.

24. Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A.; *Chem. Biol.* **1999**, *6*, 493.

25. Minowa, Y.; Araki, M.; Kanehisa, M.; *J. Mol. Biol.* **2007**, *368*, 1500.

26. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T.; *Genome Res.* **2003**, *13*, 2498.

27. Yoon, J. S.; *AllegroLayout 2.2.3*; *OpenCL-Accelerated Interactive Force-Directed Layouts*; U.S. National Institute of General Medical Sciences (NIGMS), USA, 2015.

28. Perez-Llammas, C.; Lopez-Bigas, N.; *PLoS ONE* **2011**, *6*, e19541.

29. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S.; *Mol. Biol. Evol.* **2013**, *30*, 2725.

30. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; Mcgettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G.; *Bioinformatics* **2007**, *23*, 2947.

31. Saitou, N.; Nei, M.; *Mol. Biol. Evol.* **1987**, *4*, 406.

32. Medema, M. H.; Fischbach, M. A.; *Nat. Chem. Biol.* **2015**, *11*, 639.

33. Bakkal, S.; Robinson, S. M.; Ordonez, C. L.; Waltz, D. A.; Riley, M. A.; *Microbiology (London, U. K.)* **2010**, *156*, 2058.

34. Schaffer, M. L.; Otten, L. G.; *J. Mol. Catal. B: Enzym.* **2009**, *59*, 140.

35. Xie, Y.; Cai, Q.; Ren, H.; Wang, L.; Xu, H.; Hong, B.; Wu, L.; Chen, R.; *J. Nat. Prod.* **2014**, *77*, 1744.

36. Pidot, S. J.; Coyne, S.; Kloss, F.; Hertweck, C.; *Int. J. Med. Microbiol.* **2014**, *304*, 14.

37. Seyedsayamdost, M. R.; *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 7266.

38. Yang, L.; Ibrahim, A.; Johnston, C. W.; Skinnider, M. A.; Ma, B.; Magarvey, N. A.; *Chem. Biol.* **2015**, *22*, 1259.