Systematic and Applied Microbiology 38 (2015) 525–533

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

The lanthipeptides of *Bacillus methylotrophicus* and their association with genomic islands

L. Dias[a,1], T. Caetano[a,1], M. Pinheiro[b], S. Mendo[a,*]

[a] Department of Biology and CESAM, University of Aveiro, 3810-193 Aveiro, Portugal
[b] School of Medicine, University of St Andrews, Medical and Biological Sciences Building, North Haugh, St Andrews, United Kingdom

ARTICLE INFO

Article history:
Received 3 June 2015
Received in revised form 8 October 2015
Accepted 9 October 2015

Keywords:
Mersacidin
Amylolysin
Gene duplication
Codon usage
Lantibiotics

ABSTRACT

Bacillus methylotrophicus strains are known for their potential as plant-growth promoters and as microbial pesticides that effectively control plant diseases caused by bacteria and fungi. Over the past few years, a wide diversity of their secondary metabolites has been extensively characterized. Among these are the RiPPs lanthipeptides, which are an important and growing group of notable compounds. The increasing interest in *B. methylotrophicus* species, accompanied by the development of high throughput sequencing techniques, has resulted in a substantial number of full genomes being available. Here, an *in silico* analysis was performed on these genomes in order to survey the presence of lanthipeptide biosynthetic clusters. It was found that the pan genome of *B. methylotrophicus* only encoded the biosynthesis of mersacidin and amylolysin, which are lanthipeptides with antibacterial activity. However, the amylolysin gene cluster identified was comprised of more genetic elements than those previously described, and it had certain features of two-peptide lantibiotics. Additionally, it was also established that the association of lanthipeptides with genomic islands (GIs) was not confined to mersacidin. This was also found for the amylolysin cluster as well as other class I and class II lanthipeptides, supporting the idea that their production is probably related to functional adaptation.

Introduction

Natural products are secondary metabolites of living organisms including plants, bacteria and fungi. Among these natural products, antimicrobial peptides produced by bacteria are extremely important because they are a promising alternative for tackling antibiotic resistance [27]. Some of these compounds are lanthipeptides, which are ribosomally synthesized and post-translationally modified peptides (RiPPs) [3,41]. Lanthipeptides are a large and growing family of RiPPs characterized by the presence of the uncommon thioether-linkage-containing lanthionine (Lan) and methyllanthionine (MeLan), as well as by various unsaturated amino acids such as 2,3-didehydroalanine (Dha) and (*Z*)-2,3-didehydrobutyrine (Dhb) [3,41,53]. Lanthipeptides are encoded by the structural gene, generally designated *lanA*, and are synthesized as a linear and inactive precursor peptide. This precursor peptide can be physically divided into an N-terminal leader peptide, which is critical for the recognition of the post-translational modification (PTM) enzymes, and a C-terminal core peptide [3,52,60]. The PTMs are introduced in the core peptide, where the Ser and Thr residues are initially dehydrated to Dha and Dhb, respectively. Thereafter, the thiol group of Cys residues is added to Dha and Dhb via a Michael-type addition reaction, establishing the corresponding thioether bridges Lan and MeLan, respectively. The leader peptide is then removed by proteolytic cleavage, yielding the mature and active lanthipeptide [41,60]. The lanthipeptides are divided into four classes based on the nature of the enzymes that catalyse the dehydration reaction and the formation of Lan and MeLan [60]. In the genus *Bacillus*, only clusters encoding the biosynthesis of class I and class II lanthipeptides have been identified to date. In class I lanthipeptides, dehydration is carried out by a LanB dehydratase, and the formation of thioether rings is performed by a LanC cyclase, whereas in class II, both reactions are performed by the bifunctional enzyme LanM [32].

The genus *Bacillus* comprises various producers of a diversity of specialized/secondary metabolites with biotechnological applications, which include food-safe products [1,11]. Over the last few years, the rapid development of high throughput sequencing techniques has made a large number of *Bacillus* genomes available (over 600 BioProjects, according to the NCBI). This information has contributed to the *in silico* identification of novel gene clusters encoding novel natural products [7,19,26]. Regarding the lanthipeptides, this approach has resulted in the characterization

* Corresponding author. Tel.: +351 234 370 350.
E-mail address: smendo@ua.pt (S. Mendo).
[1] The authors contributed equally to the manuscript.

http://dx.doi.org/10.1016/j.syapm.2015.10.002
0723-2020/

of cerecidins, haloduracin, lichenicidin and thuricidins, which are produced by strains of the *B. cereus* and the *B. subtilis* groups [5,38,56,58]. All these lanthipeptides have antibacterial activity and are therefore referred to as lantibiotics [47]. *Bacillus methylotrophicus* strains (formerly *Bacillus amyloliquefaciens* subsp. *plantarum* [20]) have been extensively investigated due to their potential application in various settings, including plant-growth promotion [40]. Given the relevance of these species, the number of full and draft genomes of *B. methylotrophicus* strains has increased dramatically in recent years. Presently, it is known that the *B. methylotrophicus* pan-genome encodes the biosynthesis of two class II lanthipeptides: mersacidin and amylolysin [2,6,28]. In the present study, the occurrence and characteristics of these two clusters among the different *B. methylotrophicus* strains were analyzed in more detail. Ref. [28] reported that the mersacidin cluster was part of a genomic island. Therefore, the association of genomic islands (GIs) with other lanthipeptide genes besides mersacidin was investigated.

Materials and methods

Identification and analysis of gene clusters encoding lanthipeptide biosynthesis

The full genomes of *B. methylotrophicus* (available at 23rd of September 2015) were screened using the antiSMASH 2.0 platform for the presence of lanthipeptide clusters [39]. The accession numbers of the genomes analyzed are: NC_017061; NC_017912; NC_022530; NC_020410; NC_022653; NC_009725; NC_020272; NC_023073; NC_022081; NZ_CP007244; NZ_CP006890; NC_019842; NC_016784; NC_022075; CP007165; NZ_CP009679; NZ_CP011346; NZ_CP011347; NZ_CP011278; NZ_CP010556 and NC_000964. Additionally, the bioinformatic tool ORF Finder was used to identify omitted genes on the gene clusters of lanthipeptides identified by antiSMASH. The prediction of each gene function was conducted by searching homologies with other protein sequences using the CDD database and the BLAST tool. The annotation of the lanthipeptide clusters was performed with Artemis 16.0.0 software [46]. Figures representing the linear comparison of the genomic loci were obtained with the Easyfig 2.1 application [49] and OmniGraffle Pro 5.4.4 software.

Phylogenetic analysis of B. methylotrophicus *strains*

The complete genomes of 21 species were downloaded from NCBI and aligned with Mauve (2.4.0) using the default settings [15]. Thereafter, the stripSubsetLCBs (locally collinear blocks) (http://darlinglab.org/mauve/snapshots/2015/2015-01-09/linux-x64/) script was used to extract core blocks, which created core alignments longer than 5000 bp and included all 21 genomes. To convert the core alignments to FASTA format, a script available from http://www.bioperl.org/wiki/Converting_alignment_files was used. The phylogenetic tree was constructed with the MEGA 6 software [51] using the Tamura–Nei model [50] with gamma correction (alpha value = 0.5) and a bootstrap of 1000 replicates.

Nucleotide base composition and codon usage analysis

The G + C content, the GC_3 (frequency of GC nucleotides present at the third position of synonymous codons), the Codon Adaptation Index (CAI) and the expected CAI (eCAI) values for each gene were obtained from the CAIcal server (http://genomes.urv.cat/CAIcal/) [43,44], using the adequate codon usage table available from the codon usage database (http://www.kazusa.or.jp/codon/). The housekeeping genes of the *B. methylotrophicus* YAU B9601-Y2 and IT-45 strains included in the analysis were: *dnaG*, *frr*, *gyrB*, *infB*, *infC*, *lepA*, *nusA*, *pgk*, *pheS*, *pyrG*, *rplB*, *rplC*, *rplK*, *rplL*, *rplT*, *rpsB*, *rpsC*, *rpsE*, *rpsM*, *smpB* and *tsf*. The information on the GC content of the complete genomes was obtained from http://www2.unil.ch/comparativegenometrics [45]. The housekeeping genes used for %GC comparison were: *gyrB*, *nusA*, *pgk* and *pheS*.

Prediction of genomic islands (GIs)

The presence of GIs was predicted using the IslandViewer 3 online tool (http://www.pathogenomics.sfu.ca/islandviewer/browse/) [17]. IslandViewer 3 is a "three-in-one" web-accessible computational resource that integrates SIGI-HMM [55], IslandPath-DIMOB [29] and IslandPick [35] methods for prediction. SIGI-HMM and IslandPath-DIMOB are sequence-based methods: SIGI-HMM measures codon usage and IslandPath-DIMOB measures dinucleotide bias in order to identify possible GIs. IslandPick uses a comparative genomic method to develop a stringent data set of GIs and non-GIs, which requires the availability of several phylogenetically related genomes to enable the prediction.

Results

Identification of clusters encoding the biosynthesis of lanthipeptides in Bacillus methylotrophicus

The identification of gene clusters encoding genes associated with the biosynthesis of lanthipeptides in *B. methylotrophicus* was performed using the antiSMASH platform on the full genomes available. The analysis demonstrated that these clusters were present in five closely related strains (Fig. 1). IT-45, LFB112, L-H15, L-S60 and NJN-6 genomes possessed the amylolysin cluster (*amy*; Fig. 2) that was first characterized in the *B. methylotrophicus* GA1 strain [25]. The YAU B9601-Y2 strain (equivalent to Y2) encoded the mersacidin (*mrs*; Fig. 3) cluster, as previously described by Ref. [28]. Amylolysin and mersacidin are class II lanthipeptides with antibacterial activity and are thus referred to as lantibiotics. It is known that some strains have incomplete clusters that are often not detected by *in silico* screenings. For instance, *B. methylotrophicus* $FZB42^T$ (formerly *B. amyloliquefaciens* subsp. *plantarum*) contained the protein-encoding genes involved in the self-protection mechanism of mersacidin, but the complete gene cluster was not present (Fig. 2). Hence, we investigated the occurrence of incomplete amylolysin and mersacidin clusters in all the other *B. methylotrophicus* genomes. It was found that JS25R, NAU-B3, CC178 and UCMB5036 had the same mersacidin genes as $FZB42^T$ (Fig. 3). However, JS25R and NAU-B3 were more closely related to YAU B9601-Y2 (that had the full *mrs* cluster) than to strains $FZB42^T$, CC178 and UCMB5036 (Fig. 1). In a previous study, the *mrs* cluster was identified in the YAU B9601-Y2 genome and was associated with a genomic island (GI) [28]. In this current study, we investigated if this association was also observed for amylolysin, the other *B. methylotrophicus* lanthipeptide. Therefore, the genomic context of these clusters, their GC content and codon usage pattern, was further compared and analyzed.

Genomic context of the amylolysin gene cluster

The amylolysin gene cluster of strain GA-1 (Fig. 2) was composed of the following: the structural gene *amyA*, the gene encoding the AmyM modification enzyme, the gene encoding the bifunctional protein AmyT, two self-protection genes (*amyFE*), two regulator genes (*amyKR*), and a gene with unknown function (*amyX*) [25]. The *amy* cluster was detected in the genomes of *B. methylotrophicus* strains IT-45, LFB112, L-H15, L-S60 and NJN-6. The analysis of its flanking regions revealed the presence of another gene encoding a modification enzyme (herein named *amyM2*), located downstream