Out of the Abyss: Genome and Metagenome Mining Reveals Unexpected Environmental Distribution of Abyssomicins

Alba Iglesias¹, Adriel Latorre-Pérez², James E. M. Stach¹,³, Manuel Porcar²,⁴ and Javier Pascual²*

¹ School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, ² Darwin Bioprospecting Excellence S.L., Paterna, Spain, ³ Centre for Synthetic Biology and the Bioeconomy, Newcastle University, Newcastle upon Tyne, United Kingdom, ⁴ Institute for Integrative Systems Biology (I²SysBio), University of Valencia-CSIC, Paterna, Spain

Natural products have traditionally been discovered through the screening of culturable microbial isolates from diverse environments. The sequencing revolution allowed the identification of dozens of biosynthetic gene clusters (BGCs) within single bacterial genomes, either from cultured or uncultured strains. However, we are still far from fully exploiting the microbial reservoir, as most of the species are non-model organisms with complex regulatory systems that can be recalcitrant to engineering approaches. Genomic and metagenomic data produced by laboratories worldwide covering the range of natural and artificial environments on Earth, are an invaluable source of raw information from which natural product biosynthesis can be accessed. In the present work, we describe the environmental distribution and evolution of the abyssomicin BGC through the analysis of publicly available genomic and metagenomic data. Our results demonstrate that the selection of a pathway-specific enzyme to direct genome mining is an excellent strategy; we identified 74 new Diels–Alderase homologs and unveiled a surprising prevalence of the abyssomicin BGC within terrestrial habitats, mainly soil and plant-associated. We also identified five complete and 12 partial new abyssomicin BGCs and 23 new potential abyssomicin BGCs. Our results strongly support the potential of genome and metagenome mining as a key preliminary tool to inform bioprospecting strategies aimed at the identification of new bioactive compounds such as -but not restricted to- abyssomicins.

Keywords: abyssomicins, genome mining, metagenome mining, bioprospecting, biosynthetic gene cluster distribution and evolution

INTRODUCTION

Natural products are the main source of pharmaceutically interesting biomolecules. In particular, the search of microbial specialized metabolites has yielded a broad range of chemical structures with bioactivities, from antibiotics or antymycotics to immunosuppressants and anticancer compounds. Among those, compounds featuring tetronate moieties are attractive due to their versatile biological
activities. Most of these compounds are produced by bacteria from the phylum Actinobacteria and are built of a linear fatty acid or polyketide chain with a characteristic tetronic acid 4-hydroxy-2(5H)-furanone ring system.

Within the growing family of tetronates, compounds are classified taking into account the linearity or macrocyclization of the carbon backbone and the size of the central ring system (Vieweg et al., 2014). Spirotetronates are tetrone compounds in which two rings are linked to each other by a spiroatom, and include, amongst many others, the abyssomicins, chlorothricins, tetrocarcins, lobophorines, and quartromicins. This class of tetrone compounds shares important biosynthetic and structural features: a conjugated pair of carbon–carbon double bonds at the end of a linear polyketide chain, a characteristic exocyclic double bond on the tetrone ring system and a Diels–Alder reaction that forms the cyclohexene moiety and an additional macrocycle (Weixin et al., 2013; Vieweg et al., 2014).

The abyssomicins are an actively growing family of small spirotetrone natural products with a polyketide backbone and a C11 central ring system that has been widely studied due to the unique structural features and bioactivities that some of its members exhibit. Abyssomicin biosynthesis occurs in a variety of hosts isolated from different ecosystems. The first abyssomicins (B-D) were discovered in 2004 during the screening of 930 actinomycete extracts in a successful effort to find antibacterial compounds targeting folate biosynthesis. Those abyssomicins were fermentation products of the marine actinomycete Verrucosispora maris AB-18-0327, later reclassified as Micromonospora maris AB-18-0327 (Nouioui et al., 2018), isolated from sediments of the Sea of Japan (Riedlinger et al., 2004). Years later, other research groups found new abyssomicins produced by soil isolates of Streptomyces sp. HK0381, Streptomyces sp. CHI39, recently classified as Streptomyces abyssomicinicus CHI397 (Komaki et al., 2019), and Streptomyces sp. Ank 210, in Senegal, Mexico and Germany, respectively (Niu et al., 2007; Igarashi et al., 2010; Abdalla et al., 2011). After that, the production of abyssomicins was again reported in marine isolates: Verrucosispora sp. MS100128 (Wang et al., 2013), Streptomyces sp. RLUS1487 (León et al., 2015), and Verrucosispora sp. MS100047 (Huang et al., 2016). Finally, the last abyssomicins found were synthesized by the soil Streptomyces sp. LC-6-2 (Wang et al., 2017) and the marine Streptomyces koyangensis SCSIO 5802 (Song et al., 2017; Huang et al., 2018). During the review process of this paper, abyssomicin Y was discovered in fermentation extracts of the marine Verrucosispora sp. MS100137 (Zhang et al., 2020) (Supplementary Table S1).

Despite the limited number of bacterial strains identified so far as abyssomycin producers, this family of natural products presents a wide structural diversity. In fact, there are as many as 38 members classified as type I or type II abyssomicins, where the type I family includes abyssomicins B–E, G, H, J–L, and atrop-abyssomicin C, and type II abyssomicins are the enantiomeric counterparts of the type I compounds (Sadaka et al., 2018). The type II abyssomicins are further categorized by the degree of methylation and the presence of an inserted oxygen atom with the polyketide backbone. Type IIA abyssomicins have methyl substitutions at C4 and C12, type IIB have one methyl substitution at C12, and type IIC have one methyl substitution at C12 and an inserted oxygen atom in the macrocycle (Sadaka et al., 2018). This structural diversity has gifted this family of natural products with different clinically relevant biological activities. Atrop-abyssomicin C and abyssomicins C, 2 and J exhibit antimicrobial activity against Gram-positive bacteria, including MRSA, VRSA and different Mycobacteria strains (Sadaka et al., 2018). Abyssomicin 2 also possesses HIV inhibitory and reactivator properties and neabyssomicins A and C promote HIV-1 replication in a human lymphocyte model (León et al., 2015; Song et al., 2017). More recently, abyssomicins Y, D, L, and H were described as the first abyssomicins with significant inhibitory effects against influenza A virus (Zhang et al., 2020) (Supplementary Figure S1).

Previous works also elucidated the complete abyssomicin BGC present in the genome of M. maris AB-18-032 and proposed a model for the biosynthesis of atrop-abyssomicin C, the atrop-isomer of abyssomicin C and main product synthesized by M. maris AB-18-032 (Nicolaou and Harrison, 2006, 2007; Keller et al., 2007; Gottardi et al., 2011). This abyssomicin biosynthetic gene cluster (aby) comprises 25 genes, distributed along 56 kb in M. maris AB-18-032 genome. The cluster consists of several transcriptional regulators (abyR, abyH, abyI, and abyC), an ABC exporter system (abyF1-F4), a drug resistance transporter (abyD), a cytochrome P450 system (abyte, abyW, and abyZ), a cytochrome P450 gene (abyte), a monoxygenase (abyte), a type II thioesterase (abyte), a Diels–Alderase (abyte), the PKS I genes (abyte1, abyB2, and abyB3) and five genes (abyteA1-A5) involved in the assembly of the tetronic acid moiety (Figure 1A and Supplementary Table S10) (Gottardi et al., 2011; Byrne et al., 2016). The partially sequenced cluster of the isolate Verrucosispora sp. MS100047 is 99% similar to aby BGC (Figure 1B and Supplementary Table S13).

After that, the discovery of abyssomicins M–X as fermentation products of Streptomyces sp. LC-6-2 led to the description of a new abyssomicin BGC (abs). This cluster consists of 30 genes disposed along 62 kb and presents homologs to most of the genes within aby BGC (Supplementary Table S12) but displays also two unique regulators (absC1 and absC2) and a set of four new tailoring genes (absG1, absG2, absI, and absJ) (Wang et al., 2017) (Figure 1C). A third cluster responsible for neoabyssomicin/abyssomicin biosynthesis (abm) was identified in S. koyangensis SCSIO 5802. Composed of 28 genes distributed along 63 kb, it presents five genes (abmK, abmL, abmM, abmN, and abmE2) with no apparent homologous counterparts in the aby cluster and two more (abmF and abmG) that appear to be in abs BGC but not in aby BGC (Figure 1D and Supplementary Table S11) (Tu et al., 2018). The latest abyssomicin BGC (abi) was found in S. abyssomicinicus CHI39 and is almost identical to abm BGC (Figure 1E and Supplementary Table S14) (Komaki et al., 2019).

The environmental diversity of the abyssomicin-producing isolates suggests that abyssomicin biosynthesis could be ubiquitously distributed in nature, and bioprospecting could focus on those environments heavily colonized by Actinobacteria of the genus Micromonospora and Streptomyces. There are few studies concerning the driving forces behind the transmissibility...
In the abyssomicin family, the chemical diversity found is likely to have arisen after horizontal transfer of the abyssomicin gene cluster into new hosts with subsequent domain swapping and point mutations (Ridley et al., 2008). As domain swapping is predicted to occur both within and between BGCs (Jenke-Kodama et al., 2006), the host background (genomic context) will influence structural diversification. Moreover, the enzymes involved in the synthesis of the tetronate (AbyA1-A5) and the spiro-tetronate-forming Diels–Alderase (AbyU) are capable of accepting structurally diverse substrates (Ye et al., 2014; Grabarczyk et al., 2015; Abugrain et al., 2017). Thus, identification of abyssomicin BGCs in different genomic contexts is a reasonable strategy to identify structurally novel abyssomicins.

In the present work, in order to investigate the environmental colonization of abyssomicin-producing bacteria as well as the structural diversity of abyssomicin BGCs, we have systematically explored the distribution of abyssomicin BGC and its evolution through the analysis of publicly available genomic and metagenomic data, targeting the Diels–Alderase (AbyU/AbsU/AbmU) that catalyzes the intramolecular [4 + 2] cycloaddition reaction of the linear abyssomicin polyketide precursor.

### MATERIALS AND METHODS

#### Diels–Alderase Directed Metagenome Mining

A total of 3027 metagenomes available in the JGI metagenomes database (IMG; accessed February–April 2019) were analyzed for AbyU/AbsU/AbmU homologs presence using the site option to carry out BLASTp (default parameters). The sequences of AbyU (Micromonospora maris AB-18-032), AbsU (Streptomyces sp. LC-6-2), and AbmU (Streptomyces koyangensis SCSIO 5802) used as query can be found in Supplementary Table S3.

Habitats frequently populated by Micromonospora and Streptomyces species were selected, primarily from soil and aquatic environments but also from other less common environments, including fresh-water, artificial and host-associated environments (Supplementary Figure S2). The detailed classification of the metagenomic samples from aquatic, terrestrial, engineered and host-associated environments mined for AbyU, AbsU, and AbmU can be found in Supplementary Tables S5–S8. For complete details on the metagenomes analyzed 1https://img.jgi.doe.gov/
and the Diels–Alderase positive metagenomes please refer to Supplementary Material 2.

In order to investigate possible taxonomic biases between the Diels–Alderase positive and negative metagenomes, the relative abundance of the domain Bacteria and the phylum Actinobacteria of 50 Diels–Alderase positive metagenomes were compared against 50 aquatic and 50 terrestrial Diels–Alderase negative metagenomes, randomly selected from the 3027 pool. The taxonomic assignments of both the assembled and unassembled metagenomes’ reads were carried out using the IMG/JGI site option “Phylogenetic Distribution of Genes – Distribution by BLAST percent identities” and are presented here in form of relative abundance. The Mann–Whitney U test was used to investigate whether the relative abundance of Bacteria and Actinobacteria was significatively different between the aquatic, terrestrial and Diels–Alderase positive metagenomes (Nachar, 2008). In order to investigate bias in the sequencing depth, the metagenome size (bp) of those same 50 Diels–Alderase positive metagenomes was compared against the 50 aquatic and 50 terrestrial Diels–Alderase negative metagenomes. The Mann–Whitney U was equally applied to identify significant differences in the sequencing depth.

Diels–Alderase Directed Genome Mining and Identification of Putative Abyssomicin BGCs
BLASTp of AbyU, AbsU, and AbmU were carried out against the non-redundant protein sequences database (NCBI; accessed April 2019). The Diels–Alderase containing genomes were then submitted to antiSMASH (Blin et al., 2019) (accessed April 2019; default parameters used) for BGC mining. The location of the Diels–Alderase homolog within the genome was used to verify BGC presence in antiSMASH. When a BGC was found by antiSMASH in the desired location, ORF, protein size and proposed annotation were collated and BLASTp of every protein was carried out against the non-redundant protein sequences database (NCBI) to obtain the closest homolog (Supplementary Tables S15–S85). BLASTp was used to verify/redefine the BGCs limits established by antiSMASH. In cases where antiSMASH did not identify any BGC, reconstruction of the Diels–Alderase homolog nearby genomic region was done manually from the corresponding genome in NCBI. All recovered BGCs were classified based on their completeness and novelty (Table 1).

Evolutionary Analysis
All the proteins identified through genome mining that produced significative alignments (E-value < 10−6) against AbyU, AbsU, and AbmU were aligned and the Neighbor-Joining algorithm was used to create a phylogenetic tree using MAFIT (accessed May 2019) (Katoh et al., 2002). The RefSeq annotated genomes of the microorganisms harboring those proteins were used to create a phylogenomic tree using UBCG (default parameters) (Na et al., 2018). The phylogenetic and phylogenomic trees were visualized and annotated with iTOL (Letunic and Bork, 2019).

A manual synteny analysis was carried out for all the newly discovered abyssomicin and potential abyssomicin BGCs (both total and partial), which were classified accordingly as described below (Table 2). The presence of mobile elements within the Diels–Alderase positive mined genomes was studied using Island Viewer 4 (Bertelli et al., 2017).

RESULTS
Habitat Distribution of the Diels–Alderase Positive Metagenomes
In order to study the habitat distribution of the bacteria harboring an abyssomicin BGC, and considering that the Diels–Alderase AbyU could be used as an abyssomicin-biosynthesis specific marker, we mined 3027 publicly available metagenomes for the presence of AbyU and its already known homologs AbsU and AbmU (Supplementary Tables S3, S4). 27% of the analyzed metagenomes had aquatic origin, 31% belonged to soil samples, 22% were plant-associated and the remaining 20% covered human-built environments and different host-associated

**TABLE 1 | Classification of the recovered BGCs found through Diels–Alderase directed genome mining.**

| BGC                     | The Diels–Alderase homolog is                                                                 |
|-------------------------|---------------------------------------------------------------------------------------------|
| Abyssomicin, total      | Part of an abyssomicin BGC and it is possible to recover the sequence and structure of the entire BGC. |
| Abyssomicin, partial    | Part of an abyssomicin BGC that is likely to be complete but due to the sequencing technology used there are some incomplete genes, frame shifts, gaps or the cluster is on a contig edge. |
| Potential abyssomicin, total | Part of a BGC whose product may potentially be an abyssomicin according to antiSMASH and it is possible to recover the sequence and structure of the entire BGC. |
| Potential abyssomicin, partial | Part of a BGC whose product may potentially be an abyssomicin according to antiSMASH but there are some genes missing or incomplete, frame shifts, gaps or the cluster was on a contig edge. |
| Potential BGC, total    | Surrounded by genes that could form a BGC altogether, but it is unclear which could be its product. |
| Potential BGC, partial  | Surrounded by genes that could form a BGC altogether, but it is unclear which could be its product and there were some incomplete genes, frame shifts, gaps or the cluster was on a contig edge. |
| Not a BGC               | Not likely to be part of any BGC.                                                           |
| Not enough data         | In a contig whose length makes it not possible to gain any knowledge.                       |
| Quartromycin, total     | In a quartromycin BGC and the sequence of the cluster is complete.                          |
| Quartromycin, partial   | In a quartromycin BGC but the sequence of the corresponding PKS is incomplete.              |
| Potential tetronomycin, total | Part of a potential tetronomycin BGC.                                                     |
| Potential chlorothricin, partial | In a chlorothricin BGC but the sequence of the corresponding PKS is incomplete.            |
TABLE 2 | Features shared by the potential abyssomicin BGCs described in this study.

| Cluster type | Shared motifs |
|--------------|---------------|
| 1a           | Upstream the PKS genes: Entire abyA1-A5 operon next to/nearby abyU which in most cases is located next to abyK, abyH and abyI. Downstream the PKS genes: Synteny is maintained from abyC to abyV with punctual rearrangements. |
| 1b           | Same conserved blocks as type 1a clusters but all genes are upstream the PKS genes. Extra copy of abyV downstream the PKS. |
| 2a           | Upstream the PKS genes: Presence of abyA1-A5 operon except abyA2 followed by abyN. Downstream the PKS genes: Synteny is maintained from abyA2 to abyV with punctual rearrangements. |
| 2b           | Same conserved blocks as type 2a clusters but all genes are upstream the PKS genes except abyA2. |
| 3            | Entire abyA1-A5 operon followed by abyT. |
| 4            | ABC transport system divided by the presence of abyM, abyZ, abyA3 and abyA2 are located upstream the first PKS gene. The PKS operon harbors between abyB1 and abyB2 a set of genes that includes at least abyA1 and abyU. |
| 5            | No synteny between themselves nor with other cluster types. |

microbiomes (Supplementary Figure S2). Our results showed that the three Diels–Alderase homologs share a similar habitat distribution, 31% of the AbyU positive metagenomes were from soil, 68% were plant-associated and 1% Arthropoda-associated (Figure 2B); 55% of the AbsU-positive had soil origin and 45% were plant-associated (Figure 2C) and AbmU displayed a similar distribution to AbyU with the only difference being its additional presence in an artificial bioreactor environment (Figure 2D). Surprisingly, however, none of the AbyU, AbsU, or AbmU positive metagenomes had aquatic origin.

In order to find an explanation to the absence of Diels–Alderase positive metagenomes in aquatic environments, we investigated possible taxonomic and sequencing depth biases between Diels–Alderase positive and negative metagenomes from aquatic origin. Specifically, we compared the relative abundance of assembled and unassembled reads belonging to the domain Bacteria and the phylum Actinobacteria in 50 randomly selected Diels–Alderase positive metagenomes from different environments against 50 aquatic Diels–Alderase negative metagenomes. The Mann–Whitney U test showed that the relative abundance of reads of the domain Bacteria and the phylum Actinobacteria was higher in Diels–Alderase positive metagenomes than in aquatic Diels–Alderase negative metagenomes (Supplementary Figure S3). Similarly, the relative abundance of Bacteria and Actinobacteria was lower in terrestrial Diels–Alderase negative metagenomes than in Diels–Alderase positive metagenomes (Supplementary Figure S3). On the other hand, we compared the sequencing depth, of those same 50 randomly selected Diels–Alderase positive metagenomes against the 50 aquatic and 50 terrestrial Diels–Alderase negative metagenomes. The Mann–Whitney U test showed that the sequencing depth of the Diels–Alderase positive metagenomes was significatively higher than the sequencing depth of the aquatic and terrestrial Diels–Alderase negative metagenomes (Supplementary Figure S4).

Diels–Alderase Directed Genome Mining and Diversity of Abyssomicin BGCs

In order to gain a better overview over how abyssomicin-producing bacteria are environmentally distributed and the structural diversity of abyssomicin BGCs in nature, both partial and complete genomes available in public databases were mined. In a BLASTp of AbyU, AbsU, and AbmU against the RefSeq NR database, 74 Diels–Alderase homologs from 66 different genomes were identified (Supplementary Table S9).

All the 66 Diels–Alderase positive genomes belonged to culturable bacterial strains. The habitat distribution of these isolates was, overall, similar to that found by metagenome mining. Specifically, about one third of the strains were recovered from soil, one third from plant-associated environments, and the remaining were recovered from mammals, annelids and lichens (Supplementary Figure S5). Unlike the metagenome mining results, some Diels–Alderase positive bacterial species were recovered from marine environments.

The bacterial genomes were analyzed in order to locate those Diels–Alderase homologs and study whether they were part of a potential abyssomicin BGC. This way, it was possible to identify and annotate five total and 12 partial new abyssomicin BGCs and 23 new potential abyssomicin BGCs never described until now and with similar but not identical architectures to aby, abs, and abm clusters (Figure 3). Eleven of the Diels–Alderase homologs could be located in potential BGCs, three more were found in genomic regions apparently unrelated to any BGC and 11 were located in short contigs from which it was impossible to infer any information (Supplementary Figure S7). Finally, two Diels–Alderase homologs were found in two different quartromycin BGCs and another two in potential tetronomycin and chlorothricin BGCs.

From the newly identified Diels–Alderase homologs it was possible to recover 40 total or partial new clusters potentially involved in the biosynthesis of abyssomicins (Supplementary Figures S8–S11 and Supplementary Tables S15–S85). These clusters were further classified according to their synteny in order to analyze their structural diversity. The analysis was carried out manually, as the modular nature of BGCs made the application of general synteny analysis tools impossible. Considering the diversity of biosynthetic genes and their disposition, abyssomicin and potential abyssomicin BGCs were classified into seven cluster types (Table 2). There were four genomes containing type 1a clusters and ten genomes displaying type 1b clusters from the genera Micromonospora, Actinokineospora, Frankia, Herbidospora, and Streptomyces (Supplementary Figure S8). There were seven clusters classified as type 2a and two clusters classified as type 2b. In this case, type 2a clusters were found in Streptomyces, Actinokineospora, and Micromonospora and type
2b only in *Frankia* (Supplementary Figure S9). Five clusters were classified as type 3, all belonging to *Streptomyces* and three clusters were type 4 clusters found in *Streptomyces* and *Streptacidiphilus* (Supplementary Figure S10). Finally, there were 13 clusters that did not present enough similarity to any of the cluster types described above. These clusters were found in *Frankia, Actinokinespora, Lentzea, Kutneria, Micromonospora, Streptomyces, Saccharothrix,* and *Actinocrispum* and did not share any outstanding synteny pattern amongst themselves (Supplementary Figure S11) neither with the five potential tetronomycin, chlorothricin, or quartromycin BGCs that were also found from the Diels–Alderase directed genome mining (Supplementary Figure S12). The genomes that harbored a Diels–Alderase that was not part of an abyssomicin or potential abyssomicin BGC were not considered for this classification.

## Evolutionary History of Abyssomicin BGCs

Most of the Diels–Alderase positive bacteria were taxonomically identified as belonging to the phylum *Actinobacteria* and most of them to the genus *Streptomyces* (37 isolates), followed by seven *Frankia*, three *Herbidospora*, three *Actinomadura* and three *Micromonospora* strains (Figure 4). As was expected, all the genera formed monophyletic clusters, corroborating their correct taxonomic assignment. The abyssomicin BGCs were only
identified in several species of some actinobacterial genera but not in all, suggesting that the abyssomicin BGCs may be acquired through horizontal gene transfer (HGT) events. This hypothesis was reinforced by the fact that the phylogenetic history of the Diels–Alderase (Figure 5) does not follow the same evolutionary history as of the species tree (Figure 4).

Interestingly, abyssomicin BGCs are usually associated with genomic islands (GI; Supplementary Figures S8–S11 and Supplementary Tables S86–S101) which may allow its HGT among taxa. Specifically, GI were identified in the abyssomicin BGC of some Streptomyces, Frankia, Herbidospora, Micromonospora, and Actinokineospora, nearby it (±10 kb upstream or downstream BGC) or nearby the Diels–Alderase (±10 kb upstream or downstream) (Figure 5). Albeit the vast majority this HGT events take place among members of the phylum Actinobacteria, two Proteobacteria, namely Pantoena sp. A4 and Photobacterium ganghwaense JCM 12487, harbored a Diels–Alderase. The Diels–Alderase genes of both Proteobacteria strains were phylogenetically related among them and in turn related to other Streptomyces strains (Figure 5). The closest neighbor to both proteobacterial strains was Streptomyces caatingaensis. Thereby, these intra- and inter- phyla HGT events may be explained by the presence of mobile elements such as transposases and integrases flanking or within the BGCs (Supplementary Tables S86–S101). Moreover, several Diels–Alderase paralogs were found in the mined genomes of Actinokineospora auranticolor YU 961-1, Frankia sp. Cc1.17, Streptomyces sp. NL15-2K, Streptomyces sp. NNRRL F-525, Streptomyces sp. NNRRL S-31, and Streptomyces syringae NNRRL B-16468 (Figure 5).

On the other hand, evolutionary pressure has shaped the abyssomicin BGCs, widening the functional and structural diversity of this secondary metabolite. In fact, the presence of tailoring genes is variable among species as well as the Diels–Alderase gene location within the BGCs (Supplementary Figures S8–S11 and Supplementary Tables S15–S85). However, the synteny of abyssomicin BGCs lacks phylogenetic signal and hence the abyssomicin BGC classification that we propose in the present study could not be used to trace its evolutionary history.

**DISCUSSION**

**Habitat Distribution of the Diels–Alderase Hosts Discovered Through Metagenome and Genome Mining**

To date, only ten cultured bacterial strains have been reported to produce abyssomicins (Supplementary Table S1). From these strains, 38 abyssomicins with differences at structural and bioactivity levels have been characterized (Sadaka et al., 2018). With the aim of studying the distribution of those microorganisms capable of producing new abyssomicin molecules, we have analyzed *in silico* an extensive diversity of metagenomes and genomes. AbyU is the natural Diels–Alderase present in abyssomicin BGC that catalyzes the formation of the heterocyclic ring system that characterizes this family of natural products. Very few enzymes in nature can catalyze this reaction and despite being capable of accepting structurally diverse substrates, sequence conservation with the closest known spirotetronate cyclases is minimal (Byrne et al., 2016). We selected this enzyme to lead the mining as it is essential in abyssomicin biosynthesis.

Here, we mined 3027 metagenomes for the presence of AbyU, AbsU, and AbmU, and our results showed that Diels–Alderase positive microorganisms have a strikingly diverse environmental distribution, being mainly present in soil and plant-associated microbiomes but totally absent in aquatic habitats (Figure 2). Since the few isolates reported in the literature to produce abyssomicins were equally distributed between aquatic and soil environments (Figure 2A and Supplementary Table S1), our results were totally unexpected. After examining the taxonomic composition of 50 aquatic Diels–Alderase negative and 50 Diels–Alderase positive metagenomes from different environments (Supplementary Figure S3), we could conclude that the Diels–Alderase positive metagenomes have a higher relative abundance of *Bacteria* and *Actinobacteria* than Diels–Alderase negative metagenomes from aquatic environments (Supplementary Figure S3). Furthermore, those Diels–Alderase negative metagenomes from aquatic environments showed, in general, a lower sequencing depth than the Diels–Alderase positive metagenomes (Supplementary Figure S4). Therefore, the fact that metagenomes of aquatic origin have a lower sequencing depth and that the abundance of *Bacteria* and *Actinobacteria* is lower could make it less likely to sequence Diels–Alderase homolog genes when shotgun sequencing aquatic metagenomes. On the contrary, by using the appropriate culture techniques, those low abundant abyssomicin-producing *Actinobacteria* could be enriched from aquatic environments (Supplementary Table S2).

Interestingly, we observed that all the abyssomicin-producing strains isolated from aquatic environments so far come specifically from marine sediments (Supplementary Table S2). This led us to consider that the abyssomicins could play a key role in the biology or ecology of bacteria inhabiting benthic regions. Moreover, it is tempting to hypothesize that abyssomicin-producing bacteria may be involved in symbioses with higher organisms, which has been seen before for other different antibiotic-producing strains that play an important role as defensive symbionts both in marine and terrestrial ecosystems (Gunatilaka, 2006; Seipke et al., 2012; Adnani et al., 2017). The abyssomicins could also act as signal molecule in plant-bacteria communication or as precursors involved in plant growth and development, as reported before in the *Frankia* and *Micromonospora* genera through, for example, the formation of nitrogen fixing actinomycodes (Trujillo et al., 2010; Sellstedt and Richau, 2013). Further investigations will be needed in order to unravel the biased habitat distribution of Diels–Alderase positive bacteria.

Altogether, we identified 74 Diels–Alderase homologs present in 66 different genomes (Supplementary Table S9) from which it was possible to identify and annotate five total and 12 partial new abyssomicin BGCs and 23 new potential abyssomicin BGCs. Indeed, all these 40 abyssomicin and potential abyssomicin
producers are culturable strains whose habitat distribution follows the same patterns found through the metagenome mining as none of them was recovered from aquatic samples (Supplementary Figure S6). In our case, 60.6% of the Diels–Alderase positive genomes displayed an abyssomicin or potential abyssomicin BGC. In the remaining genomes in which the Diels–Alderase was not located in any BGC, we could not predict its metabolic function. Previous studies reported other Diels–Alderase involved in the synthesis of other natural products, with the exception of riboflavin synthases that are involved in primary metabolism (Lichman et al., 2019).

Therefore, based on the genome- and metagenome mining, we can conclude that the potential abyssomicin producers have a cosmopolitan distribution albeit their presence in aquatic habitat is limited. This strongly suggests that abyssomicin bioprospecting efforts should not be focused on aquatic environments but rather on soil and plant-associated ones. Also, two Diels–Alderase homologs were found in two different quartromicin BGCs and another two in potential tetronominycin and chlorothricin BGCs.

The presence of those four Diels–Alderase homologs within BGCs belonging to other natural products is well justified, as quartromicin, tetronominycin, and chlorothricin share the same tetronate cycloaddition as the abyssomicins (Vieweg et al., 2014). Moreover, 11 of the Diels–Alderase homologs detected in the mined genomes were in potential non-abyssomicin BGCs, three more were found in genomic regions a priori unrelated to any BGC and 11 appeared in short contigs from which it was impossible to infer any information. In this case, only 10 of the 66 genomes analyzed were completely sequenced and only seven isolates were sequenced with third generation sequencing technologies (Supplementary Table S9).

The identification of the Diels–Alderase homologs location within the genomes and the recovery of potential BGCs was influenced by the quality of the sequencing technology used and the assembly level achieved by each previous individual study. Some factors such as the high G + C content of actinomycete genomes affect the sequencing reactions and the assembly process (Nakamura et al., 2011), however, the biggest
challenge appears to be the recovery of the highly conserved and modular sequences of polyketide synthases (PKS) characterized by displaying highly similar intragenic and intergenic tandem repeats at nucleotide level, which in many cases are longer than the read-length of the sequencing technology used (Gomez-Escribano et al., 2016). Moreover, large PKS clusters can often be distributed along several contigs, and it has been demonstrated that sequencing errors can introduce false frameshifts into the large PKS sequences (Blazić et al., 2012). Finally, the presence of Diels–Alderase homologs outside abyssomicin BGCs, could be explained by the presence of transposases flanking Diels–Alderase homologs allowing their genetic recombination along the genome (Supplementary Tables S86–S101). Specifically, the Diels–Alderase homologs of Streptomyces caatingaensis CMAA 1322 and Streptomyces armeniacus ATCC 15676 were not part of an abyssomicin BGC but showed transposases on both sides (Figure 5 and Supplementary Tables S90, S92).

Evolutionary History of Abyssomicin BGC

It is well-known that Actinobacteria are characterized by their ability to produce a wide variety of specialized metabolites and, despite the problem of re-discovering already known molecules, bacteria from the phyla Actinobacteria are still one of the most prolific sources of chemical diversity (Genilloud, 2017). The presence of abyssomicin BGCs is limited to the phylum Actinobacteria, mainly representatives of the genus Streptomyces and Frankia. The constraint of the abyssomicin BGC to some specific strains suggests that speciation was not the primary driver for dissemination of this cluster (Figure 4). Instead, HGT may have played an important role in the transmission of abyssomicin BGCs, which may have jumped among taxa through mobile elements (Ziemert et al., 2014; Hall et al., 2017). Indeed, many integrases and transposases were found surrounding or inside the abyssomicin BGCs (Supplementary Tables S86–S101).
Many BGCs in Actinobacteria evolve through HGT events, but only a few studies have demonstrated it (Choudoir et al., 2018). For example, in a genome mining study on 75 Salinispora strains, 124 pathways involved in the synthesis of PKS and non-ribosomal peptide synthetase (NRPS) natural products were identified and showed that HGT events were responsible for the majority of pathways, which occurred in only one or two strains, as acquired pathways were incorporated into genomic islands (Ziemert et al., 2014). In another example, the secondary metabolite clusters on the chromosomes of Streptomyces avermitilis ATCC31267 were found to contain many transposase genes in the regions near both ends of the clusters, suggesting these transposases might have been involved in the transfer of these clusters (Omura et al., 2001). Similarly, it was demonstrated that the rifamycin BGC in Salinispora arenicola CNS-205 had been acquired through HGT directly from Amycolatopsis mediterranei S699 by genomic island movement (Penn et al., 2009).

Although HGT events are more frequent among phylogenetically close taxa, in this case within the phylum Actinobacteria, HGT events can take place among different phyla. In the present study, we could identify a possible HGT event of Diels–Alderase from a representative of the genus Streptomyces to two strains of the phylum Proteobacteria, namely Pantoea sp. A4 and Photobacterium ganghwense JCM 12487 (Figure 5). The transmission of functional BGCs among phyla was also reported by other authors (Zeng et al., 2014). Unfortunately, neither transposases nor integrases were identified nearby the Diels–Alderasers of Pantoea sp. A4 and Photobacterium ganghwense JCM 12487, which could have explained the HGT event.

The acquisition of an abyssomicin BGC by a bacterial strain could increase its evolutionary fitness and therefore enhance its competitiveness against other members of the community. In fact, the biological activity of abyssomicins includes antimicrobial activities against Gram-positive bacteria and Mycobacteria (Riedlinger et al., 2004; Freundlich et al., 2010). Other biological activities discovered so far are antitumor properties, latent human immunodeficiency virus (HIV) reactivator, anti-HIV and HIV replication inducer properties (Sadaka et al., 2018). The wide diversity of abyssomicin BGCs that we have found through genome mining suggests that a plethora of abyssomicin-like molecules remain undiscovered.

CONCLUSION

The aim of this study was to shed some light into the structural diversity, habitat distribution, and evolutionary history of abyssomicin BGC. Through metagenome and genome mining, we discovered that the habitat distribution of microorganisms harboring a Diels–Alderase is restricted to that of the phylum Actinobacteria, with mainly representatives of the genus Streptomyces and Frankia, which are primarily present in soil and plant-associated environments. Surprisingly, we did not find any Diels–Alderase positive bacterium in aquatic environments although six out of ten reported abyssomicin producers were isolated from marine sediments. Therefore, all the strains that present abyssomicin BGCs have been observed to be associated to organic or inorganic solid substrates. Based on the habitat distribution of Diels–Alderase positive bacteria, we hypothesize that microorganisms producing abyssomicin-like molecules could play key ecological roles in the corresponding microbial communities.

Moreover, the vast structural diversity of abyssomicin BGCs that we have found could reflect its horizontal evolutionary history, and we predict that a plethora of abyssomicins remain unknown to date. Additionally, Diels–Alderase enzymes are of great value in synthetic chemistry, as the $[4 + 2]$ cycloaddition reaction they catalyze could facilitate the development of environmentally friendly synthetic routes to a wide variety of useful compounds. Finally, the discovery of Diels–Alderase homologs, could hold great potential as part of the synthetic biology toolbox to generate libraries of novel non-natural biomolecules. Taken together, the results of the present work reveal the interest of a new bioprospecting strategy to identify natural products such as abyssomicins out of their currently assumed environmental distribution.

AUTHOR’S NOTE

This manuscript has been released as a Pre-Print at bioRxiv (Iglesias et al., 2019).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

AI, AL-P, JS, MP, and JP conceived and designed this study. AI and AL-P performed the analyses. AI, AL-P, and JP analyzed the data. AI, MP, and JP wrote this manuscript.

FUNDING

This study was funded by the European Union through the BioRoboost project, H2020-NMBP-TR-IND-2018-2020/BIOTEC-01-2018 (CSA), Project ID 210491758 is acknowledged. AI is a recipient of a Newcastle University SAgE Doctoral Training Award (reference EJU/160317378). AL-P is a recipient of a Doctorado Industrial fellowship from the Spanish Ministerio de Ciencia, Innovación y Universidades (reference DI-17-09613). This work was also funded in the framework of the MIPLACE project (ref: PCI2019-111845-2, Programación Conjunta Internacional 2019, AEI).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00645/full#supplementary-material
REFERENCES
Abdalla, M. A., Yadav, P. P., Dittrich, B., Schüffler, A., and Laatsch, H. (2011). Ent-Homobasomycins A and B, two new spirretoxane metabolites from Streptomyces sp. Ank 218. Org. Lett. 13, 2156–2159. doi: 10.1021/ol100307y
Abgrain, M. E., Brunstedt, C. J., Osborn, A. R., Philipp, B., and Mahmud, T. (2017). A highly promiscuous 6-ketoacyl-ACP synthase (KAS) III-like protein is involved in Paenibacillus biosynthesis. ACS Chem. Biol. 12, 362–366. doi: 10.1021/acschembio.6b01043
Adnani, N., Rajski, S. R., and Bungi, T. S. (2017). Symbiosis-inspired approaches to antibiotic discovery. Nat. Prod. Rep. 34, 784–814. doi: 10.1039/c7np00009j
Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoad, G., Winson, G. L., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 45, W30–W35.
Blažič, M., Starcevic, A., Lisfi, M., Baranasic, D., Goranović, D., Fujs, ä, et al. (2019). Out of Freundlich, J. S., Lalgondar, M., Wei, J.-R., Swanson, S., Sorensen, E. J., Rubin, E. J., et al. (2011). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 45, W30–W35.
Gomez-Escribano, J. P., Alt, S., and Bibb, M. J. (2016). Next generation sequencing of secondary metabolite biosynthetic gene clusters coincides with lineage of natural product biosynthesis in bacteria. J. Nat. Prod. 79, 12215–12220. doi: 10.1073/pnas.211433198
Huang, H., Song, Y., Li, X., Wang, X., Ling, C., Qin, X., et al. (2018). Abyssomicin C, polycyclic macrolactones from the deep-sea derived Verrucosispora sp. Ank 218. Appl. Environ. Microbiol. 84, 7175–7188. doi: 10.1128/AEM.00318-18
Jenke-Kodama, H., Börner, T., and Dittmann, E. (2006). Natural biocombinatorics in the polyketide synthesis genes of the Actinobacterium Streptomyces avermitilis. PLoS Comput. Biol. 2:e132. doi: 10.1371/journal.pcbi.0020132
Kato, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3599–3605. doi: 10.1093/nar/gkf346
Keller, S., Schadt, H. S., Ortel, L. and Süssmuth, R. D. (2007). Action of atrop-abyssomicin C as an inhibitor of 4-amino-4-deoxychorismate synthase PabB. Angew. Chem. Int. Ed. 46, 8284–8286. doi: 10.1002/anie.200701836
Komi, H., Sakurai, K., Hosoyama, A., Kimura, A., Trujillo, M. E., Igarashi, Y., et al. (2019). Diversity of PKS and NRPS gene clusters between Streptomyces abyssomicinus sp. nov. and its taxonomic neighbor. J. Antibiot. 73, 141–151. doi: 10.7164/jab/00012
León, B., Navarro, G., Dickey, B. J., Stepan, G., Tsai, A., Jones, G. S., et al. (2015). Abyssomicin 2 reactivates latent HIV-1 by a PKC- and HDAC independent mechanism. Org. Lett. 17, 262–265. doi: 10.1021/ol503349y
Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47, W256–W259.
Lichman, B. R., O’Connor, S. E., and Kries, H. (2019). Biocatalytic strategies towards [4+2] cycloaditions. Chem. A Eur. J. 25, 6864–6877. doi: 10.1002/chem.201805412
Na, S. I., Kim, Y. O., Yoon, S. H., Ha, S., Baek, I., and Chun, J. (2018). UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. J. Microbiol. 56, 281–285.
Nachar, N. (2008). The man-whitney u: a test for assessing whether two independent samples come from the same distribution. Tutor. Quant. Methods Psychol. 4, 13–20. doi: 10.20982/tqmp.04.1.p013
Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., et al. (2011). Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. 39:e90. doi: 10.1093/nar/gkr344
Nicolaou, K. C., and Harrison, S. T. (2006). Total synthesis of abyssomicin C and atrop-abyssomicin C. Synthesis 45, 3256–3260. doi: 10.1002/anie.200601116
Nicolaou, K. C., and Harrison, S. T. (2007). Total synthesis of abyssomicin C, atrop-abyssomicin C, and abyssomicin D: implications for natural origins of atrop-abyssomicin C. J. Am. Chem. Soc. 129, 429–440. doi: 10.1021/ja067803p
Niu, X. M., Li, S. H., Goels, H., Schollmeyer, D., Hilliget, M., Grabley, S., et al. (2007). Abyssomicin C, a highly functionalized polycyclic metabolite from Streptomyces species. J. Nat. Prod. 70, 2319–2324. doi: 10.1021/nn7005999
Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J. P., Woyke, T., Kyrpides, N. C., et al. (2018). Genome-based taxonomic classification of the phylum Actinobacteria. Front. Microbiol. 9:2407. doi: 10.3389/fmicb.2018.02407
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., et al. (2001). Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites. Proc. Natl. Acad. Sci. U.S.A. 98, 12215–12220. doi: 10.1073/pnas.211433198
Penn, K., Jenkins, C., Nett, M., Udway, D. W., Gontang, E. A., McGlinchey, R. P., et al. (2009). Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. ISME J. 3, 1193–1203. doi: 10.1038/ismej.2009.58
Ridley, C. P., Ho, Y. L., and Khosla, C. (2008). Evolution of polyketide synthases in bacteria. Proc. Natl. Acad. Sci. U.S.A. 105, 4955–4960. doi: 10.1073/pnas.0710107105
Riedlinger, J., Reike, A., Zähner, H., Krisher, B., Bull, A. T., Maldonado, L. A., et al. (2004). Abyssomicins, inhibitors of the para-aminobenzoic acid pathway produced by the marine Verrucosporispora strain AB-18-032. J. Antibiot. 57, 271–279. doi: 10.1093/antibiotics.57.271
Sadaka, C., Ellsworth, E., Robert Hansen, P., Ewin, R., Damborg, P., and Watts, J. L. (2018). Review on abyssomicins: inhibitors of the chorismate pathway and folate biosynthesis. Molecules 23:1371. doi: 10.3390/molecules23061371
Seipke, R. F., Kaltenborn, M., and Hutchings, M. I. (2012). Streptomyces as symbionts: an emerging and widespread theme? EMBnet Microbiol. Rev. 36, 862–876. doi: 10.1111/j.1574-6976.2011.00313.x
Sellstedt, A., and Richau, K. H. (2013). Aspects of nitrogen-fixing Actinobacteria, in particular free-living and symbiotic Frankia. FEMS Microbiol. Lett. 342, 179–186.
Song, Y., Li, Q., Qin, F., Sun, C., Liang, H., Wei, X., et al. (2017). Neoabyssomicins A–C, polycyclic macrolactones from the deep-sea derived Streptomyces. Front. Microbiol. 8, 1–12.
Iglesias et al. Unexpected Environmental Distribution of Abyssomicins

koyangensis SCSIO 5802. Tetrahedron 73, 5366–5372. doi: 10.1016/j.tet.2017.07.034

Trujillo, M. E., Alonso-Vega, P., Rodriguez, R., Carro, L., Cerda, E., Alonso, P., et al. (2010). The genus *Micromonospora* is widespread in legume root nodules: the example of *Lupinus angustifolius*. *ISME J.*, 4, 1265–1281. doi: 10.1038/isme.2010.55

Tu, J., Li, S., Chen, J., Song, Y., Fu, S., Ju, J., et al. (2018). Characterization and heterologous expression of the neoabyssomicin/abyssomicin biosynthetic gene cluster from *Streptomyces koyangensis* SCSIO 5802. *Microb. Cell Fact.*, 17, 1–14.

Vieweg, L., Reichau, S., Schobert, R., Leadlay, P. F., and Süssmuth, R. D. (2014). Recent advances in the field of bioactive tetronates. *Nat. Prod. Rep.*, 31, 1554–1584. doi: 10.1039/c4np00015c

Wang, Q., Song, F., Xiao, X., Huang, P., Li, L., Monte, A., et al. (2013). Abyssomicins from the South China Sea deep-sea sediment *Verrucosispora* sp.: natural thioether michael addition adducts as antitubercular prodrugs. *Angew. Chem. Int. Ed.*, 52, 1231–1234. doi: 10.1002/anie.201208801

Wang, X., Elshahawi, S. I., Cai, W., Zhang, Y., Ponomareva, L. V., Chen, X., et al. (2017). Bi- and tetracyclic spirotetronates from the coal mine fire isolate *Streptomyces* sp. LC-6-2. *J. Nat. Prod.*, 80, 1141–1149. doi: 10.1021/acs.jnatprod.7b00108

Weixin, T., Manghong, Z., Deng, Z., and Yuhui, S. (2013). Biosynthesis of tetronate antibiotics: a growing family of natural products with broad biological activities. *Sci. China Chem.*, 56, 1364–1371. doi: 10.1007/s11426-013-4921-x

Ye, Z., Musiol, E. M., Weber, T., and Williams, G. J. (2014). Reprogramming acyl carrier protein interactions of an acyl-CoA promiscuous trans-acyltransferase. *Chem. Biol.*, 21, 636–646. doi: 10.1016/j.chembiol.2014.02.019

Zeng, Y., Feng, F., Medová, H., Dean, J., and Koblížek, M. (2014). Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc. Natl. Acad. Sci. U.S.A.*, 111, 7795–7800. doi: 10.1073/pnas.140029511

Zhang, J., Li, B., Qin, Y., Karthik, L., Zhu, G., Hou, C., et al. (2020). A new abyssomicin polyketide with anti-influenza A virus activity from a marine-derived *Verrucosispora* sp. *MS100137. Appl. Microbiol. Biotechnol.*, 104, 1–11. doi: 10.1007/s00253-019-10217-2

Ziemert, N., Lechner, A., Wietz, M., Millán-Aguínaga, N., Chavarria, K. L., and Jensen, P. R. (2014). Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc. Natl. Acad. Sci. U.S.A.*, 111:E1130-9.

**Conflict of Interest:** AL-P and JP were employed at the time of research by the company Darwin Bioprospecting Excellence S.L. MP is founder of the company Darwin Bioprospecting Excellence S.L.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Iglesias, Latorre-Pérez, Stach, Porcar and Pascual. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.