ORIGINAL RESEARCH

A putative antimicrobial peptide from Hymenoptera in the megaplasmid pSCL4 of *Streptomyces clavuligerus* ATCC 27064 reveals a singular case of horizontal gene transfer with potential applications

Sebastián Ayala-Ruano1 | Daniela Santander-Gordón2 | Eduardo Tejera2,3 | Yunierkis Perez-Castillo3,4 | Vinicio Armijos-Jaramillo2,3

Abstract

*Streptomyces clavuligerus* is a Gram-positive bacterium that is a high producer of secondary metabolites with industrial applications. The production of antibiotics such as clavulanic acid or cephamycin has been extensively studied in this species; nevertheless, other aspects, such as evolution or ecology, have received less attention. Furthermore, genes that arise from ancient events of lateral transfer have been demonstrated to be implicated in important functions of host species. This approximation discovered relevant genes that genomic analyses overlooked. Thus, we studied the impact of horizontal gene transfer in the *S. clavuligerus* genome. To perform this task, we applied whole-genome analysis to identify a laterally transferred sequence from different domains. The most relevant result was a putative antimicrobial peptide (AMP) with a clear origin in the Hymenoptera order of insects. Next, we determined that two copies of these genes were present in the megaplasmid pSCL4 but absent in the *S. clavuligerus* ATCC 27064 chromosome. Additionally, we found that these sequences were exclusive to the ATCC 27064 strain (and so were not present in any other bacteria) and we also verified the expression of the genes using RNAseq data. Next, we used several AMP predictors to validate the original annotation extracted from Hymenoptera sequences and explored the possibility that these proteins had post-translational modifications using peptidase cleavage prediction. We suggest that Hymenoptera AMP-like proteins of *S. clavuligerus* ATCC 27064 may be useful for both species adaptation and as an antimicrobial molecule with industrial applications.

KEYWORDS

antimicrobial peptide, horizontal gene transfer, Hymenoptera, pSCL4, *Streptomyces clavuligerus* ATCC 27064
1 | INTRODUCTION

Streptomyces is a genus of Gram-positive bacteria characterized by its ability to produce antitumors, antihypertensives, immunosuppressives and, above all, antibiotics (Procópio, da Silva, Martins, de Azevedo, & de Araújo, 2012). Streptomyces clavuligerus is one of the most important species within the genus for its industrial and clinical applications. This microorganism was isolated from South American soil samples and first described by Higgsens and Kastner (1971). In addition to other members of the Streptomyces genus, it exhibits a very complex life cycle that includes multicellular, spore-bearing hyphae and semi-dormant spores that can survive for long periods in the soil (Bobek, Šmidová, & Čihák, 2017). Moreover, it was described that S. clavuligerus produces more than 20 secondary metabolites, including cephamycin C, holomycin, deacetoxycephalosporin C, penicillin N, and tunicamycin, among others (Table 1).

Streptomyces clavuligerus ATCC 27064 contains a linear chromosome (6.7 Mb), three linear short plasmids, pSCL1 (10.5 kb), pSCL2 (149.4 kb), and pSCL3 (442.2 kb), and a linear megaplasmid named pSCL4 (1.8 Mb) (Song et al., 2010). These plasmids can recombine with the chromosome, enhancing the genetic interchange of secondary metabolites (Medema et al., 2010). Interestingly, S. catleyae NRRL 3841 and S. lipmannii NRRL 3584 are two species of the genus lacking plasmids (Netolitzky, Wu, Jensen, & Roy, 1995), and even the strain S. clavuligerus F613-1 has a different number of plasmids (Cao et al., 2016). These findings suggest that plasmids could provide a vehicle for genetic interchange by horizontal gene transfer or some other mechanism in this species and even in the genus.

Horizontal gene transfer (HGT) corresponds to the transmission of genetic information between related or unrelated organisms (Soucy, Huang, & Gogarten, 2015). This mechanism has been recognized as a driving evolutionary force, and it allows for gene flow between distant evolutionary taxa (Boto, 2010; Syvanen, 2012). Therefore, HGT provides a new combination of sequences that usually confer selective advantages to the host, and it can survive in the “new” genome for long periods of time (Soucy et al., 2015). Furthermore, there are neutral transferred genes that do not confer an immediate benefit but could provide material for variation and posterior innovations (Boto, 2010; Syvanen, 2012).

Horizontal gene transfer has been associated with prokaryotes, so their processes (conjugation, transformation, and transduction) are well understood. However, the advent of sequencing technologies has uncovered a lot of evidence to show that this process also occurs in eukaryotes (Dunning Hotopp et al., 2007; Hirt, Alsmark, & Embley, 2015; Husnik et al., 2013; Sieber, Bromley, & Dunning Hotopp, 2017; Wu et al., 2013). Additionally, there are many instances of HGT events in eukaryotes, for instance the transfer of genes from organelles, such as chloroplasts and mitochondria, to the nucleus (Kleine, Maier, & Leister, 2009) or certain reported cases of gene transmission that occurs in symbiotic relationships between prokaryotes and eukaryotes (Acuña et al., 2012; Boto, 2014; Dunning Hotopp, 2011; Klasson et al., 2014). Examples of HGT involving animals are less detected, presumably because the germline is isolated from somatic cells, meaning that contact with foreign DNA is reduced (Acuña et al., 2012). Despite this observation, several studies have reported animals as acceptors in this process. Genomic sections of the endosymbiotic bacteria Wolbachia pipientis have been found in fruit flies and wasps (Conner et al., 2017; Dunning Hotopp et al., 2007). Anecdotally, HGT in the opposite direction was also observed because genomic regions of the mosquito Aedes aegypti were detected in W. pipientis. Data support the hypothesis that those genes are expressed after an extended evolutionary period because the transfer, with functional significance, produces evolutionary innovation (Woolfit, Iturbe-Ormaetxe, McGraw, & O’Neill, 2009).

Antimicrobial peptides (AMPs), also known as host defense peptides, are molecules from the innate immune system of all living organisms that protect against a host of bacterial, yeast, fungal, and viral infections and modulate immune responses during infections (Zhang & Gallo, 2016). The best-known mode of action of these molecules involves the membrane disruption of microbes by forming cavities and producing cell death (Mahlapuu, Håkansson, Ringstad, & Björn, 2016). However, mechanisms affecting intracellular processes such as cell wall formation, DNA, RNA and protein synthesis, and protein folding have also been described (Cudic & Otvos, 2002; Ho, Shah, Chen, & Chen, 2016; Le, Fang, & Sekaran, 2017). Moreover, certain AMPs have been reported as products of an HGT event. These cases include insect drosomycins and nematode cremycins, which are probably acquired from plants (Zhu & Gao, 2014). The Alo-3 sequence, which is an antifungal peptide found in many plants

| Antibiotic | Reference |
|------------|-----------|
| β-lactam | Cephamycin C | Nabais and da Fonseca (1995) |
| β-lactam | Penicillin N | Nabais and da Fonseca (1995) |
| β-lactam | Clavulanic acid | Reading and Cole (1977) |
| β-lactam | Cephalosporin | Aharonowitz and Demain (1978) |
| β-lactam | Deacetoxycephalosporin C | Nabais and da Fonseca (1995) |
| β-lactam | Ro 22-5417 | Pruess and Kellett (1983) |
| β-lactam | O-carbamoyl-deacetylcephalosporin | Nabais and da Fonseca (1995) |
| Non β-lactam | Holomycin | Li and Walsh (2010) |
| Non β-lactam | MM 19290 (related to tunicamycin) | Kenig and Reading (1979) |
and in the animal kingdom only in the beetle Acrocinus longimanus (Barbault et al., 2003), is the product of an HGT event proposed to explain the presence of this gene in the beetle (Husnik et al., 2013). Finally, a similar case was suggested for thaumatins from nematodes, ticks, and insects, which probably appeared due to HGT from plants (Petre, Major, Rouhier, & Duplessis, 2011).

Antimicrobial peptides exhibit certain properties that give them extraordinary antimicrobial activity and, in some cases, even antitumoral activity. Their shortness in length (usually <100 residues) enables the easy penetration of membranes. In the case of antibacterial peptides, a net positive charge ensures the electrostatic interaction with negatively charged membranes as well as selectivity action against prokaryotes given the neutrality of eukaryote membranes. Another important AMP characteristic is the amphiphilicity (hydrophobic and hydrophilic regions) that facilitates peptide entrance and membrane lysis (Kang, Kim, Seo, & Park, 2017; Mahlapuu et al., 2016; Sun, Xia, Li, Du, & Liang, 2014). Thus, due to the effectiveness of AMPs against microbes, tumoral cells, and their immunomodulatory capacity, these molecules have emerged as a new approach for the treatment of multidrug resistant bacteria, cancer, and other diseases (Mahlapuu et al., 2016).

Antimicrobial peptides have been classified based on their biological activity (antibacterial, antiviral, antifungal, anticancer, and so on), synthesis machinery (gene encoded and nongene encoded), and different properties, such as peptide charge, length, and hydrophobic content, chemical modifications, and others (Wang, 2017). The most extended classification is based on the secondary structure of peptides, which consists of four families (α-helices, β-sheets, αβ structures, and non-αβ structures) determined by the presence or absence of α and β secondary structures in the three-dimensional structures (Wang, 2017). However, most AMPs lack a 3D structure; thus, a classification has been proposed considering the connection mode of polypeptide chains (Wang, 2015).

Insects are one of the largest producers of AMPs in nature. This is explained by the lack of an adaptive immune system in these animals, meaning that they need to produce a large amount of AMPs from different types as a defense mechanism against pathogens (Bulet & Stöcklin, 2005; Mylonakis, Podziadlowski, Muhammed, & Vilcinskas, 2016; Yi, Chowdhury, Huang, & Yu, 2014). They release AMPs from the fat body (analogous to a vertebrate’s liver) into the hemolymph; thus, they can be distributed throughout the insect body, although other epithelial cells and hemocytes in certain species can also secrete AMPs (Bulet & Stöcklin, 2005). Other important sources of AMPs are insect venom (e.g., melittins and apamins from bees and mastoparans from wasps), which serves as a defense against pathogens or for infecting prey (Moreno & Giralt, 2015), and saliva (e.g., drosomycin from flies and termicin from termites) (Bulet & Stöcklin, 2005; Mylonakis et al., 2016), which protects eggs from infections (Józefiak & Engberg, 2017).

In this study, we report and characterize a putative AMP of the S. clavuligerus strain ATCC 27064 transferred from Hymenoptera insects. We used several programs to predict the AMP annotation. Additionally, public RNAseq data were used to corroborate the expression of the transferred gene, and we predicted the potential cleavage sites of the protein in congruence with a post-translational scenario. To our knowledge, this is the first time that such an event has been reported and provides hints of the complexity of S. clavuligerus and its capacity to produce antimicrobial compounds.

### 2 | METHODS

#### 2.1 | Interdomain HGT candidate selection

A BLASTp search was performed (default values, e-value threshold of 1e-5) using the S. clavuligerus strain ATCC 27064 proteome (Song et al., 2010) as the query and UniProt (http://www.uniprot.org) as the database. Next, we retrieved the taxonomy of each blast hit using custom python scripts and a MySQL database. The S. clavuligerus proteins with at least 80% of BLAST hits with a taxonomic classification different to bacteria were selected as HGT candidates. The 80% BLAST hit threshold was established through the combination of HGT candidates reported in Richards, Leonard, Soanes, and Talbot (2011) and Schmitt and Lumbsch (2009), allowing for the creation of guidelines to validate the HGT candidates, as proposed in Armijos-Jaramillo, Sukno, and Thon (2015).

#### 2.2 | Alignments and phylogenetic reconstruction

We performed global alignment with the BLAST hit candidates obtained in the previous step, using MAFFT 7 (Katoh, Misawa, Kuma, & Miyata, 2002) option auto (offset value 0.123, gap open penalty 1.53, and a BLOSUM62 scoring matrix). Next, we reconstructed a phylogenetic tree using PhyML 3.0 (Guindon et al., 2010) with an LG amino acid replacement matrix, 100 nonparametric bootstrap repetitions, and all other options as default. We manually analyzed the tree topology to establish an HGT pattern in the candidate’s trees. DNA alignments were also performed with MAFFT 7 using the same parameters described above.

#### 2.3 | Contamination elimination and HGT region determination

To avoid obtaining HGT candidates due to contamination in the S. clavuligerus genome project, we verified the existence of similar proteins in our HGT candidates through BLASTp searches in the PATRIC database (https://www.patricbrc.org/). The first searches were performed using PATRIC’s representative proteome database (e-value threshold 10). Next, we used the Streptomyces database and finally the S. clavuligerus database, which has available the complete proteome of the two strains (ATCC 27064 from three different projects and F613-1) plus several plasmids.

To verify the expression of HGT candidates, we used the RNAseq data available in GEO (GSE104738, bioproject PRJNA413703). These data contain deep sequencing of S. clavuligerus (strains F613-1 and ATCC27064) mRNA. We selected the genetic region that encodes one of the HGT protein candidates, and then used this as a template
to map readings from RNAseq experiments. To perform the mapping process, we used Geneious mapper (iterations of up to 5 times) from Geneious 10 (http://www.geneious.com; Kearse et al., 2012).

To detect putative HGT regions in the plasmid pSCL4 of *S. clavuligerus* ATCC 27064, we used the software Alien Hunter (Vernikos & Parkhill, 2006), optimizing predicted boundaries with a change-point detection 2 state 2nd order HMM.

### 2.4 | AMP candidate annotation and post-translational modification prediction

Annotation of our candidates as AMPs was made through several web servers available for this purpose. To estimate the accuracy of the AMP prediction methods, we selected a list of verified AMPs as positive controls and housekeeping genes as negative controls (see Supporting Information Table S1). The list of positive controls was obtained from the Antimicrobial Peptide Database (APD) (http://aps.unmc.edu/AP) using four arthropod peptides of each AMP family with structural information (α-helices, β-sheets, αβ structures, and non-α structures) and four without an available structure in the database. We used these data to calculate the sensitivity and specificity of each AMP predictor web server. In total, we evaluated seven AMP predictors: CAMP3 (Waghu, Barai, Gurung, & Idicula-Thomas, 2016), AMPA (Torrent et al., 2012), ClassAMP (Joseph, Karnik, Nilawe, Jayaraman, & Idicula-Thomas, 2012), AntiBP (Lata, Sharma, & Raghava, 2007), MLAMP (Lin & Xu, 2016), AMP Scanner Vr.2 (Veltri, Kamath, & Shehu, 2018), and ADP3 (Wang, Li, & Wang, 2016). We used the servers with the highest sensitivity and specificity to predict our candidates as AMPs. The signal peptide prediction was performed with SignalP4.1 (Petersen, Brunak, Heijne, & Nielsen, 2011).

We used the web server PROSPER (https://prosper.erc.monash.edu.au/) to predict the post-translational modifications for our candidates to generate peptides from proteins. This tool predicts the cleavage sites for different proteases detectable in the query sequence. Additionally, to explore the serine protease machinery of *S. clavuligerus*, we predicted and classified these sequences with BLASTp searches against the MERPOPS database (Rawlings, Barrett, & Bateman, 2012).

3 | RESULTS AND DISCUSSION

### 3.1 | Interdomain HGT candidates in *Streptomyces clavuligerus* ATCC 27064

We applied a pipeline adapted from Armijos-Jaramillo et al. (2015) and Armijos-Jaramillo, Santander-Gordón, Soria, Pazmiño-Betancourth, and Echeverría (2017) to detect interdomain HGT events in the proteome of the *S. clavuligerus* strain ATCC 27064. Because of this search, we found a hypothetical protein (GenBank: EFG03676) highly similar to antimicrobial peptides (AMPs) of arthropods (60% identical to the most similar protein), particularly from the order Hymenoptera and mostly from the suborder Apocrita. We found this protein only in the *S. clavuligerus* strain ATCC 27064, and no other homolog sequences were detected in other *S. clavuligerus* strains or bacterial species.

We found three paralogs to EFG03676 in the *S. clavuligerus* ATCC 27064 proteome: EDY50506, EFG03588, and EDY49959. These sequences are 97.9% identical at the amino acid level, and all were predicted with a signal peptide. From the NCBI databases, we found that the codifying genes of EDY50506 and EDY49959 are located in the bacterial chromosome, whereas EFG03676 and EFG03588 are located in the pSCL4 megaplasmid. Nevertheless, the similarity between sequences led us to analyze the genomic regions that encode the candidate proteins. Thus, we used the DS570654 contig for the EDY50506 codifying region and DS570647 contig for the EDY49959 codifying region, both part of the BioProject PRJNA28551 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA28551). At the same time, we used the pSCL4 plasmid contig CM000914 to detect codifying regions of EFG03676 and EFG03588 proteins. The sequence alignment demonstrated that the genes of DS570654 and DS570647 originally annotated as part of the *S. clavuligerus* chromosome are identical to two different pSCL4 plasmid regions (contig CM000914, BioProject PRJNA42475, Medema et al., 2010). Thus, we determined that there are only two copies of Hymenoptera AMP-like genes in the *S. clavuligerus* ATCC 27064 genome, and both are located in the megaplasmid pSCL4. There is no evidence of the presence of these genes in the chromosome contigs of the sequencing projects PRJNA19249 and PRJNA42475. Additionally, the alignment between pSCL4 plasmid contigs CM000914 (BioProject PRJNA42475) and CM001019 (BioProject PRJNA19249, Song et al., 2010) showed that the regions containing Hymenoptera AMP-like genes are identical. The current evidence showed that pSCL4 was originated from the excision of a *S. clavuligerus* chromosome section (Álvarez-Álvarez, Martínez-Burgo, Rodríguez-García, & Liras, 2017). Thus, it could be that the original copy of the gene is in the chromosome and is then excised to the plasmid. An alternative scenario is the appearance of one copy in the pSCL4 megaplasmid without trespessing to the chromosome. In any case, the similarity between these two paralogs suggests a recent duplication event.

The presence of Hymenoptera AMP-like proteins in three different sequencing projects of *S. clavuligerus* ATCC 27064 (BioProjects PRJNA28551, PRJNA42475 and PRJNA19249) demonstrates that those sequences are not artifacts. The lack of EFG03588 homologs in bacteria and their similarity to Hymenoptera AMPs strongly suggests horizontal gene transfer (HGT) as the origin of these molecules in *S. clavuligerus*. The alternative hypothesis (vertical transfer) implies the loss of these Hymenoptera AMP-like genes in all bacterial species except *S. clavuligerus* ATCC 27064, but this explanation is less likely. Additionally, regarding Hymenoptera AMPs, we found a sequence of the common carp *Cyprinus carpio* (XP_018951382) within the homologs of EFG03676-EFG03588. We tracked XP_018951382 in the scaffold where their codifying gene is located and found no evidence of HGT in the adjacent regions (data not shown). The origin of the arthropod AMP-like protein of *Cyprinus carpio* is not within the scope of this study, but we suggest
that it could be the result of contamination during the sequencing process or a very recent case of HGT. This suggestion is based on the discovery of a 99.2% identity between *C. carpio* and the red harvester ant *Pogonomyrmex barbatus* proteins, and there is no evidence of homology with Chordata species. The phylogenetic relationship of EFG03676 with its paralogs and animal xenologs can be observed in Figure 1. The topology observed in this tree strongly suggests the direction of the transference from Hymenoptera to *S. clavuligerus* or at least this is the last bacterial descendent (discovered so far) that maintains a copy of that gene in its lineage.

Despite the fact that the origin of HGT events cannot be determined with certainty, several reports conclude that close ecological interactions increase the chances of lateral gene transmissions (Degnan, 2014; Venner et al., 2017; Wang & Liu, 2016). In that sense, the interaction between the Streptomyces species and Hymenoptera has been constantly reported. That is the case for the fungus-growing ants, which use antibiotics produced by Streptomyces bacteria to protect their cultivations (Currie, Scott, Summerbell, & Malloch, 1999). From there, other protective uses of Streptomyces antibiotics have been reported in different Hymenoptera species (de Souza et al., 2013; Engl et al., 2018; Kaltenpoth, Yildirim, Gürbüz, Herzer, & Strohm, 2012; Kroiss et al., 2010; Van Arnem, Currie, & Clardy, 2018). Particularly, *S. clavuligerus* has not been observed in one of these interactions; however, several unidentified Streptomyces species were reported by Haeder, Wirth, Herz, and Spitteler (2009) as being associated with leaf-cutting ants. Some of them were identified as similar to *S. griseus*, a species related to *S. clavuligerus*, which is

**FIGURE 1** Phylogenetic tree of *Streptomyces clavuligerus* ATCC 27064 Hymenoptera antimicrobial peptide-like proteins and their homologs. Proteins from bacteria are represented in red, the sequence from Hymenoptera insects are colored in black, and blue highlights the *Cyprinus carpio* (Chordata) sequence. IDs of sequence and their taxonomy can be observed in terminal nodes. The values of internal nodes are the nonparametric bootstrap percentage.
in agreement with Sembiring (2009). All these arguments lead us to think that a close symbiotic relationship between S. clavuligerus and Hymenoptera could lead the transference of AMP-like gene.

3.2 | Expression evidence of the HGT candidates

We mapped RNAseq data from S. clavuligerus ATCC 27064 to the genomic scaffold DS570654 (BioProject PRJNA28551), from 42,968 to 43,734. This region is the same in the three sequencing projects (PRJNA28551, PRJNA42475, and PRJNA19249) and encodes one of the two copies of Hymenoptera AMP-like proteins (EFG03588). We observed a clear expression signal along this region (Figure 2). Using RNAseq data from S. clavuligerus F613-1, we did not map readings in the region between 42,968 and 43,734. This result is congruent with the absence of EFG03588 orthologs in bacteria, including other S. clavuligerus strains. Additionally, the RNAseq data demonstrate that the codifying gene of EFG03588 is expressed under in vitro culture conditions.

3.3 | Inference of HGT recentness

Lawrence and Ochman (1997) proposed the idea that foreign genes undergo amelioration in order to adapt their sequence to the host genome. This proposition suggests that laterally transferred genes could be differentiated from indigenous ones by several specific features of the donor and receptor genomes. However, by the same effect of amelioration, only "recently" transferred genes could be identified by this approximation because transferred genes over sufficient time ("old" transferred genes) evolve to imitate the receptor genomic features. Considering this argument, we predicted unusual regions in the pSCL4 plasmid genome (contig CM001019).

We performed this task through interpolated variable order motifs, a technique implemented in the software Alien Hunter (Vernikos & Parkhill, 2006). This program located six potentially foreign regions in pSCL4: 1–7,500, 19,165–26,714, 46,125–62,763, 73,381–80,082, 102,991–113,397, and 134,187–144,263 (these positions are relative to CM001019). The codifying gene of EFG03588 is located in the region 108,662–109,105 and that of EFG03676 is located in the region 196,642–197,085. Thus, only the codifying gene of EFG03676 is located inside an Alien Hunter HGT predicted region. EFG03676 and EFG03588 are 98% identical; thus, it is improbable that the difference in the predicted regions was produced by these sequences.

Following the idea of amelioration of foreign genes, the Hymenoptera AMP-like sequences should have sufficient time to adapt to the pSCL4 genome, but there are several other explanations for this observation. In contrast to the amelioration hypothesis, Medrano-Soto, Moreno-Hagelsieb, Vinuesa, Christen, and Collado-Vides (2004) demonstrated that certain transferred genes require a level of codon usage compatibility between the transferred gene and receptor genome to survive over time. Additionally, certain indigenous regions in the genomes possess distinctive features (GC content or codon bias) that can be confused with HGT regions, such as genes with translational robustness (Drummond, Bloom, Adami, Wilke, & Arnold, 2005). Considering these arguments, the lack of HGT signatures in the codifying regions of AMP-like proteins (particularly EFG03588) might be produced by an amelioration process, codon usage compatibility between donor and receptor or the existence of genomic regions in pSCL4 with certain features that cover up the real HGT regions.

The phylogenetic distribution of the S. clavuligerus AMP-like homologs (Figure 1) suggests that the transfer should occur

**FIGURE 2** Mapping of the scaffold DS570654 (region from 42,968 to 43,734) of Streptomyces clavuligerus ATCC 27064 using RNAseq data (GSE104738). In the upper panel, a graphical representation of the coverage is observed. In the lower panel, the relative position of the mapping readings to the scaffold region is observed.
before the appearance of Apocrita (Hymenoptera suborder). In agreement with the TimeTree database (Kumar, Stecher, Suleski, & Hedges, 2017), this event occurred between 227 and 254 million years ago. The correlation of this time range with HGT occurrence should be considered with precaution because it is based upon the impossibility of associating S. clavuligerus proteins with a particular Apocrita clade. When more Apocrita sequences become available, a more precise determination of the time of lateral transfer can be achieved. Additionally, it is possible to identify the specific Apocrita species that underwent transfer. In this scenario, the estimated time of the transfer could be ostensibly lower.

### 3.4 Size estimation of the HGT region

We performed alignment within the codifying sections of Hymenoptera AMP-like paralogs (EFG03676 and EFG03588) and their neighbor genes in the pSCL4 genome (Figure 3a). In this alignment, we identified a region of 736 bp with 97.8% identity that contains the coding genes of EFG03676 and EFG03588, the last section of the upstream coding genes of the hypothetical proteins EFG03677 and EFG03589 and part of the intergenic section with the downstream genes that codify proteins EFG03675 and EFG03587. Thus, we propose that the entire section of 736 bp was duplicated inside the pSCL4 plasmid.

To explore the size of the HGT event in the current S. clavuligerus genome, we performed a BLASTn search using the 736-bp section annotated as duplicated as a query. We found that only a region of 293 nucleotides has sufficient similarity to retrieve results from Hymenoptera genome sequences (this region excludes the signal peptide) (Figure 3b). In addition, we performed BLASTp searches of the contiguous proteins of AMP-like sequences (data not shown). The results for EFG03676 and EFG03588 (left contiguous sequences) showed these proteins had a vertical transmission phylogenetic pattern that was highly similar to that of other Streptomyces proteins. By contrast, the BLAST results of EFG03676 and EFG03588 proteins (right contiguous sequences) showed similarity to their own or partial sequences of S. clavuligerus. This result shows a misannotation or the possibility that these proteins can be unique to S. clavuligerus. In any case, the contiguous sequences to Hymenoptera AMP-like proteins showed no HGT signal, at least in the current annotation state.
3.5 Antimicrobial peptide prediction

To explore whether the HGT candidates maintain the same annotation of their xenologs, we used several AMP prediction software platforms (see Section 2). First, we established the sensitivity and specificity of each predictor (Supporting Information Table S1) because we observed heterogeneity in the results obtained from preliminary analysis. Next, we submitted the candidate sequences to the predictors with the highest sensitivity and specificity values (ADP3, AMPA, MLAMP, and AMP Scanner Vr.2).

ADP3 predicted the entire EFG03588 and EFG03676 sequences as AMPs. Its output provides estimations of physicochemical parameters and sequence structure. In the case of HGT candidates, the formation of alpha helices and hydrophobic surfaces was predicted, so they may interact with membranes and function as AMPs. AMPA predicted the peptide TIRFKSQRGHI (from 126 to 137) of the sequence EFG03676 as an AMP; meanwhile, the sequence EFG03588 was not predicted as an AMP. Both sequences differ in position 134 with G instead of S. Using the regions between 126 and 137 of the sequences EFG03588 and EFG03676, the MLAMP program predicted these peptides as AMPs with a probability of 0.8 and 0.9, respectively, and as antibacterial AMPs with probabilities higher than 0.8 in both cases. AMP Scanner Vr.2 predicted that neither the entire sequence nor the fragments are AMPs.

Despite the increasing number of AMP predictors (Liu, Fan, Sun, Lao, & Zheng, 2017; Porto, Pires, & Franco, 2017) and various approximations used for this task, the diversity of AMPs makes it difficult to use them as an accurate prediction of these types of molecules. Considering this limitation, the prediction of our candidates using 3 of the 4 most sensible programs evaluated in this study makes us confident of the veracity of the annotation.

3.6 Post-translational modification of AMP candidates

We performed the prediction of the cleavage sites of our candidate proteins using the web server PROSPER, and we found that they could be cut in different positions by several proteases. Remarkably, the server predicted cutting sites (by serine proteases) near to the AMP prediction of the AMPA and MLAMP programs. Thus, the PROSPER results suggest that an elastase-2 makes cuts in the sequence EFG03588 at positions 25, 96, 99, 127, 140, and 141 and in EFG03676 at positions 25, 33, 96, 99, 127, 140, and 141. The cleavage of the residues 127 and 141 includes the peptide predicted with AMPA and MLAMP. The resultant peptide of the cleavage in these two positions (RFKSQRG/SHGIDFV) was also predicted as AMP by the AMPA and MLAMP programs.

Elastase-2 belongs to the serine proteases family, which are widely extended in all kingdoms of life (Tripathi & Sowdhamini, 2008), including S. clavuligerus, which has thirteen of these proteins annotated in its proteome. We performed BLASTp searches of S. clavuligerus serine proteases in the MEROPS database (https://www.ebi.ac.uk/merops/index.shtml) and found eight serine proteases of the S1 family, four of the S8 family and one of the M6 family. Elastase-2 belongs to the S1A subfamily, and we found three of these sequences in the S. clavuligerus proteome. The presence of these enzymes suggests that S. clavuligerus possesses the machinery to cleave EFG03676 and EFG03588 proteins. What is more, the signal peptide of these proteins could be used to translocate the sequence to some transporter (like an ABC transporter) to process the protein and then secrete the peptide outside the cell. A similar mechanism has been observed in class Ila bacteriocins (Drider, Finland, Héchard, McMullen, & Prévost, 2006; Perez, Zendo, & Sonomoto, 2014). In this scenario, Hymenoptera AMP-like sequences could be translocated, guided by the signal peptide, and then be submitted to post-translational proteolytic cleavage using serine proteases (and likely many other enzymes and complexes) before being secreted outside the cell (Figure 4).

With this scenario, we highlight the machinery available in bacteria to complete the processing of AMPs even in sequences that arise from distant organisms. In fact, several approaches have been reported to process and secrete polypeptides from longer proteins (Dores, Lecaudé, Bauer, & Danielson, 2002; Fukudome & Yoshikawa, 1992; Ivanov, Karelín, Philippova, Nazimov, & Pletnev, 1997; Lewis & Stern, 1983; Meisel & Bockelmann, 1999; Vanhoye, Bruston, Nicolas, & Amiche, 2003; Zhao, Garreau, Sannier, & Piot, 1997), and only experimental procedures can unveil the real processing of our candidates.

Although some other peptides (especially with cyclic conformation) with antimicrobial and/or antitumoral effects have been reported in the Streptomyces genus (Shetty, Buddana, Tatipamula, Naga, & Ahmad, 2014; Um et al., 2013; Zhou et al., 2014), this is the first study to detect an AMP that is laterally transferred in this group of organisms. More than 2,500 natural or encrypted AMPs have been described previously (Zhang & Gallo, 2016). However, some of those have low antimicrobial activity or an unknown function. Therefore, the stable HGT genes were demonstrated to play...
relevant roles in receptor organisms (Belbahri, Calmin, Mauch, & Andersson, 2008; Friesen et al., 2006; Slot & Rokas, 2010; Tiburcio et al., 2010; Wenzl, Wong, Kwang-won, & Jefferson, 2005). The rationale behind this observation suggests that the increase in the fitness generated by the HGT gene in the receptor organism compensates or overtakes the negative selection that a gene of this type would undergo, because a foreign gene (especially in the interdomain HGT) should overcome the incompatibility of promoters and an alternative genetic code, without mentioning the cost generated by the synthesis and duplication of new genetic material, especially in prokaryotic genomes that tend to be highly efficient (Kuo & Ochman, 2009; Mira, Ochman, & Moran, 2001). Thus, the presence of Hymenoptera AMP-like genes in \textit{S. clavuligerus} ATCC 27064 should play an important role in this strain. This assertion is supported by the estimated time in which the Hymenoptera transferred genes are retained in the \textit{S. clavuligerus} genome. Additionally, the presence of two copies in the genome suggests an expansion of the family, perhaps to produce dosage duplication. These lines of evidence lead us to propose that Hymenoptera AMP-like proteins could be useful antimicrobial molecules for \textit{S. clavuligerus} ATCC 27064 with several biotechnological applications that can be explored.

**ACKNOWLEDGMENTS**

This research was supported by internal funds of the Universidad de Las Américas-Ecuador under the project BIO.VA.17.04. We would also like to thank Jennifer Gallardo for her valuable work on this study.

**CONFLICT OF INTEREST**

None declared.

**AUTHOR CONTRIBUTIONS**

SA-R performed experiments and wrote the first draft of the manuscript; DS-G performed experiments and edited the manuscript; ET and YP-C helped in experimental design and in manuscript editing; and VA-J designed and performed the experiments and wrote and edited the manuscript.

**DATA ACCESSIBILITY**

Table S1 is available in the supporting information section.
REFERENCES

Acuña, R., Padilla, B. E., Flórez-Ramos, C. P., Rubio, J. D., Herrera, J. C., Benavides, P., ... Rose, J. K. C. (2012). Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. Proceedings of the National Academy of Sciences USA, 109, 4197–4202. https://doi.org/10.1073/pnas.1121190109

Aharonowitz, Y., & Demain, A. L. (1978). Carbon catabolite regulation of cephalosporin production in Streptomyces clavuligerus. Antimicrobial Agents and Chemistry, 14, 159–164. https://doi.org/10.1128/AAC.14.2.159

Álvarez-Alvarez, R., Martínez-Burgo, Y., Rodríguez-García, A., & Liras, P. (2017). Discovering the potential of S. clavuligerus for bioactive compound production: Cross-talk between the chromosome and the pSCL4 megaplasmid. BMC Genomics, 18, 907. https://doi.org/10.1186/s12864-017-4289-y

Armijos-Jaramillo, V., Santander-Gordón, D., Soria, R., Pazmiño-Betancourt, M., & Echeverría, M. C. (2017). A whole genome analysis reveals the presence of a plant PR1 sequence in the potato pathogen Streptomyces scabies and other Streptomyces species. Molecular Phylogenetics and Evolution, 114, 346–352. https://doi.org/10.1016/j.ympev.2016.08.006

Armijos-Jaramillo, V. D., Sukno, S. A., & Thon, M. R. (2015). Identification of horizontally transferred genes in the genus Colletotrichum reveals a steady tempo of bacterial to fungal gene transfer. BMC Genomics, 16, 2. https://doi.org/10.1186/1471-2164-16-2

Barbault, F., Landon, C., Guenneugues, M., Meyer, J.-P., Schott, V., Dimarco, J.-L., & Vovelle, F. (2003). Solution structure of Alo-3: A new knottin-type antifungal peptide from the insect Acrocinus longimanus. Biochemistry, 42, 14434–14442. https://doi.org/10.1021/bi035400o

Belbahri, L., Calmin, G., Mauch, F., & Andersson, J. O. (2008). Evolution of the cutinase gene family: Evidence for lateral gene transfer of a candidate Phytophthora virulence factor gene. Gene, 408, 1–8. https://doi.org/10.1016/j.gene.2007.10.019

Bobek, J., Šmídová, K., & Čihák, M. (2017). A waking review: Old and novel insights into the spore germination in streptomyces. Frontiers in Microbiology, 8, 2205. https://doi.org/10.3389/fmicb.2017.02205

Boto, L. (2010). Horizontal gene transfer in evolution: Facts and challenges. Proceedings of the Royal Society B: Biological Sciences, 277, 819–827. https://doi.org/10.1098/rspb.2009.1679

Boto, L. (2014). Horizontal gene transfer in the acquisition of novel traits by metazoans. Proceedings of the Royal Society B: Biological Sciences, 281, 20123450. https://doi.org/10.1098/rspb.2013.2450

Brand, G. D., Magalhães, M. T. Q., Tinoco, M. L. P., Aragão, F. J. L., Nicoli, J., Kelly, S. M., ... Bloch, C. (2012). Probing protein sequences as sources for encrypted antimicrobial peptides. PLoS ONE, 7, e45848. https://doi.org/10.1371/journal.pone.0045848

Bulet, P., & Stöcklin, R. (2005). Insect antimicrobial peptides: Structures, properties and gene regulation. Protein and Peptide Letters, 12, 3–11

Cao, G., Zhong, C., Zong, G., Fu, J., Liu, Z., Zhang, G., & Qin, R. (2016). Complete genome sequence of Streptomyces clavuligerus F613-1, an industrial producer of clavulanic acid. Genome Announcements, 4, e01020-16. https://doi.org/10.1128/genomeA.01020-16

Conner, W. R., Blaxter, M. L., Anfora, G., Ometto, L., Rota-Stabelli, O., & Turelli, M. (2017). Genome comparisons indicate recent transfer of wR-like Wolbachia between sister species Drosophila suzukii and D. subpudchrella. Ecology and Evolution, 7, 9391–9404. https://doi.org/10.1002/ece3.3449

Cudic, M., & Otvos, L. (2002). Intracellular targets of antibacterial peptides. Current Drug Targets, 3, 101–106.

Currie, C. R., Scott, J. A., Sommerbell, R. C., & Malloch, D. (1999). Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. Nature, 398, 701–704. https://doi.org/10.1038/19519

de Souza, D. J., Lenoir, A., Kasuya, M. C. M., Ribeiro, M. M. R., Devers, S., Couceiro, J. C., ... Della Lucia, T. M. C. (2013). Ectosymbionts and immunity in the leaf-cutting ant Acromyrmex subterraneus subterraneus. Brain, Behavior, and Immunity, 28, 182–187. https://doi.org/10.1016/j.bbi.2012.11.014

Degnan, S. M. (2014). Think laterally: Horizontal gene transfer from symbiotic microbes may extend the phenotype of marine sessile hosts. Frontiers in Microbiology, 5, 638. https://doi.org/10.3389/fmicb.2014.00638

Dores, R. M., Lecadé, S., Bauer, D., & Danielson, P. B. (2002). Analyzing the evolution of the opioid/orphanin gene family. Mass Spectrometry Reviews, 21, 220–243. https://doi.org/10.1002/mas.10029

Drider, D., Finland, G., Héchard, Y., McMullen, L. M., & Prévost, H. (2006). The continuing story of class Ila bacteriocins. Microbiology and Molecular Biology Reviews, 70, 564–582. https://doi.org/10.1128/MMBR.00016-05

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., & Arnold, F. H. (2005). Why highly expressed proteins evolve slowly. Proceedings of the National Academy of Sciences USA, 102, 14338–14343. https://doi.org/10.1073/pnas.0504070102

Dunning Hotopp, J. C. (2011). Horizontal gene transfer between bacteria and animals. Trends in Genetics, 27, 157–163. https://doi.org/10.1016/j.tig.2011.01.005

Dunning Hotopp, J. C., Clark, M. E., Oliveira, D. C. S. G., Foster, J. M., Fischer, P., Muñoz Torres, M. C., ... Werren, J. H. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science, 317, 1753–1756. https://doi.org/10.1126/science.1142490

Engl, T., Kroiss, J., Kai, M., Netchitaylo, T. Y., Svatôš, A., & Kaltenpohl, M. (2018). Evolutionary stability of antibiotic protection in a defensive symbiosis. Proceedings of the National Academy of Sciences USA, 115, E2020–E2029. https://doi.org/10.1073/pnas.1719797115

Friesen, T. L., Stukenberg, E. H., Liu, Z., Meinhardt, S., Ling, H., Faris, J. D., ... Oliver, R. P. (2006). Emergence of a new disease as a result of interspecific virulence gene transfer. Nature Genetics, 38, 953–956. https://doi.org/10.1038/ng1839

Fukudome, S., & Yoshikawa, M. (1992). Opioid peptides derived from wheat gluten: Their isolation and characterization. FEBS Letters, 296, 107–111.

Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology, 59, 307–321. https://doi.org/10.1093/sysbio/syq010

Haeder, S., Wirth, R., Herz, H., & Spitterell, D. (2009). Candidicin-producing Streptomyces support leaf-cutting ants to protect their fungus garden against the pathogenic fungus Escovopsis. Proceedings of the National Academy of Sciences USA, 106, 4742–4746. https://doi.org/10.1073/pnas.0812082106

Higgins, C. E., & Kastner, R. E. (1971). Streptomyces clavuligerus sp. nov., a β-lactam antibiotic producer. International Journal of
Yi, H.-Y., Chowdhury, M., Huang, Y.-D., & Yu, X.-Q. (2014). Insect antimicrobial peptides and their applications. Applied Microbiology and Biotechnology, 98, 5807–5822. https://doi.org/10.1007/s00253-014-5792-6
Zhang, L.-J., & Gallo, R. L. (2016). Antimicrobial peptides. Current Biology, 26, R14–19. https://doi.org/10.1016/j.cub.2015.11.017
Zhao, Q., Garreau, I., Sannier, F., & Piot, J. M. (1997). Opioid peptides derived from hemoglobin: Hemorphins. Biopolymers, 43, 75–98. https://doi.org/10.1002/1097-0282(1997)43:2<75:AID-BIP2>3.0.CO;2-X
Zhou, H., Yang, Y., Yang, X., Li, W., Xiong, Z., Zhao, L., ... Ding, Z. (2014). A new cyclic tetrapeptide from an endophytic Streptomyces sp. YIM67005. Natural Product Research, 28, 318–323. https://doi.org/10.1080/14786419.2013.863198
Zhu, S., & Gao, B. (2014). Nematode-derived drosomycin-type antifungal peptides provide evidence for plant-to-ecdysozoan horizontal transfer of a disease resistance gene. Nature Communications, 5, 3154. https://doi.org/10.1038/ncomms4154

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ayala-Ruano S, Santander-Gordón D, Tejera E, Perez-Castillo Y, Armijos-Jaramillo V. A putative antimicrobial peptide from Hymenoptera in the megaplasmid pSCL4 of Streptomyces clavuligerus ATCC 27064 reveals a singular case of horizontal gene transfer with potential applications. Ecol Evol. 2019;9:2602–2614. https://doi.org/10.1002/ece3.4924