UNIVERSITA' DEGLI STUDI DELL'INSUBRIA

DOTTORATO DI RICERCA IN SCIENZE DELLA VITA E BIOTECNOLOGIE
XXXIV CICLO

Biosynthesis, self-resistance and regulation of glycopeptide antibiotics in producing actinomycetes

Biosintesi, autoresistenza e regolazione degli antibiotici glicopeptidici in actinomiceti produttori

Docente guida: Prof. Flavia Marinelli
Tutor: Dr. Oleksandr Yushchuk

Tesi di dottorato di:
Andrés Andreo Vidal
Matr. 738935

Dip. Biotecnologie e Scienze della Vita - Università degli Studi dell'Insubria

Anno accademico 2020-2021

Aknowledgements

In the following lines I intend to show my gratitude for the help and support I received throughout these years of my PhD course. When I have finished my PhD in Life Science and Biotechnology at University of Insubria, I look back and realize that the most important thing is the friendships I made. Although this has been a long period, including having to overcome COVID-19 pandemic, I am very happy to have started this experience. I have met extraordinary people from whom I have learned a lot.

First of all, I would like to express my deepest gratitude to my supervisor Prof. Flavia Marinelli for the opportunity to do a PhD in her group. I would like to thank for your guidance and your wise advice in these years. Your laboratory has become the nucleus of my professional life last years. I am extremely grateful to my tutor Dr. Oleksandr Yushchuk for your help in the design of experiments, interpretation of results, personal trainer and your various interesting stories. Although part of this period we have been separated thousands of kilometres due to the pandemic, we have always been in touch by email, video calls or phone calls, being available to solve problems at any time. I really appreciate as well the constant assistance of Dr. Elisa Binda and Dr. Francesca Berini in the laboratory. Thanks for your daily help and your important suggestions with experiments. I would like to extend my sincere thanks to my labmates Elisa, Melissa, Letizia and Elena for the good times we have spent in the University and in our free time.

I also have to mention my beginning as a young scientist in Antonio Sánchez Amat group at the Department of Genetics and Microbiology at the University of Murcia. Prof. Antonio Sánchez Amat, Prof. Patricia Lucas Elío and Dr. Jonatan Cristian Campillo Brocal installed me the curiosity to know new things and work hard every day. There I started my career as a researcher and I made my first contributions to science.

Last but not least, I have to say thanks for their unconditional support to my parents, Andrés and Pepi, my sisters Belén and Marina, and rest of my family. Special thanks to my future wife Laura for the encouragement you have provided me during this time and Italian family for your warm hospitality. With the intention of avoiding giving a long list of names and forgetting someone, I would like to summarize my acknowledgements to my friends from my hometown Alhama, university colleagues from Murcia and friends from Varese.

Index

List of abbreviations

AA – amino acid
***aac(3)IV*p** – promoter of apramycin acetyltransferase gene
aadA – spectinomycin/streptomycin resistance gene
AMR – antibiotic microbial resistance
antiSMASH – antibiotics & secondary metabolites analysis shell
ATCC – American Type Culture Collection
BGC – biosynthetic gene cluster
Bht – β-hydroxytyrosine
BLAST – basic local alignment search tool
Bp – base pair(s)
CDS – coding sequence
CSR – cluster-situated regulator
Da – dalton (unit of mass)
D-Ala-D-Ala – D-alanyl-D-alanine
D-Ala-D-Lac – D-alanyl-D-lactate
D-Ala-D-Ser – D-alanyl-D-serine
Ddl – D-Ala-D-Ala ligase
DMSO – dimethyl sulfoxide
Dpg – 3,5-dihydroxyphenylglycine
DSM – Deutsche Sammlung von Mikroorganismen und Zellkulturen
EICs – extracted ion chromatograms
ermEp – promoter of erythromycin resistance gene
FDA – Food and Drug Administration
G-C – guanine and cytosine content
Glc*N*Ac – *N*-acetyl glucosamine
GPA – glycopeptide antibiotic
HGT – horizontal gene transfer
Hpg – 4-hydroxyphenylglycine
HPLC – high performance liquid chromatography
LB – Lysogeny broth medium
LC-MS – liquid chromatography-mass spectrometry
LGPA – lipoglycopeptide antibiotic
NRPS – nonribosomal peptide synthetase
NRRL – Agricultural Research Service Culture Collection
MDR – multidrug resistance
MGE – mobile genetic elements
MHA – Muller Hinton Agar medium
MRSA – methicillin-resistant *Staphylococcus aureus*
MS/MS – tandem mass spectrometry
MultiGeneBlast – modified BLAST procedure for multiple querries
NMR – nuclear magnetic resonance
OD – optical density
ORF – open reading frame
PCR – polymerase chain reaction
PG – peptidoglycan
PKS – polyketide synthase
Tyr – tyrosine
UV – ultraviolet
VanA – D-Ala-D-Lac ligase
VanR – *van*-operon transcriptional regulator
VanS – *van*-operon sensor histidine kinase
VanY – D,D-carboxypeptidase
VanX – D,D-dipeptidase
vlg – *van*-like gene
VRE – vancomycin-resistant *enterococci*
VRSA – vancomycin-resistant *Staphylococcus aureus*
WAC – Wright Actinomycete Collection
WCB – working cell bank
WHO – World Health Organization
WT – wild type
X-Gal – 5-bromo-4-chloro-3-indolyl-β-D-galactoside
X-Gluc – 5-bromo-4-chloro-3-indolyl-β-D-glucuronide
Δ – deletion (in the description of genotypes)

I. ABSTRACT

There is an urgent and growing need for new antibiotics to treat infections caused by emerging multi-resistant bacterial pathogens and to maintain the advanced medical procedures that we now take for granted. Since the appearance of penicillin in 1928, antibiotic discovery was based on antimicrobial activity-guided screening of thousands of microbial isolates from soil samples (Waksman screening platform). These massive screening campaigns were conducted mainly by pharmaceutical companies and were very successful, resulting in the discovery of most of the major antibiotic classes. The corresponding period is in fact known as the "Golden Age of Antibiotics". Majority of newly discovered antibiotics are secondary metabolites –also named specialized metabolites – produced from genera belonging to the Actinobacteria phylum, which includes high G-C soil-dwelling mycelial organisms with sizable genomes (traditionally named actinomycetes). Unfortunately, after this fruitful period, the chance of finding new antibiotics started to decrease, and it became more and more evident that new methods for accessing and screening still-untapped sources for biologically-active microbial metabolites should be invented. Till then, the large majority of antibiotics were discovered mainly from actinobacteria belonging to *Streptomyces* genus, apparently the most abundant in soil, and the easier-to isolate and -to cultivate among actinomycetes. Thus, unexplored microorganisms (not coming from *Streptomyces* genus) – the so-named "rare" or "uncommon" actinomycetes – were discovered as a rather untapped source of chemically diverse specialized metabolites. These bacteria are defined as the actinomycete strains less frequently isolated than *Streptomyces* spp., even though they may not actually be so rare in the environment. Undoubtfully, we know less on how to handle them and they tend to be more recalcitrant to classical cultivation and manipulation methods. Contrary to better-known *Streptomyces* spp., genetics tools are still much less developed for "rare" actinomycetes today.

Among the most relevant antibiotics produced by "rare" actinomycetes are glycopeptides (GPAs) which are drugs of last resort against severe infections caused by multidrug-resistant Gram-positive pathogens. Clinically important GPAs include first-generation vancomycin and teicoplanin – used in clinic for more than 30 years – and second-generation antibiotics telavancin, oritavancin and dalbavancin, recently approved for medical practice. As for other specialized metabolites, the biosynthetic routes leading to the production of GPAs are encoded within an assemblage of genes which are grouped in biosynthetic gene clusters (BGCs). BGCs contain not only the biosynthetic genes, but also regulatory, resistance or transport genes. All these genes are essential, since resistance genes are necessary to avoid suicide during the production, while regulatory and transport genes are required to control biosynthetic processes and secretion of the molecule at the appropriate lifecycle phase.

In my PhD thesis, I have used recent technologies for rapid bacterial genome sequencing, advanced genetic engineering and bioinformatics prediction tools to identify, characterize and improve the production of GPAs in "rare" actinomycetes. First, I investigated the role of the positive regulatory genes involved in the biosynthesis of clinically relevant GPAs teicoplanin, produced by *Actinoplanes teichomyceticus* and A40926 from *Nonomuraea gerenzanensis*, this last the precursor of the second generation dalbavancin. To perform these studies, I had to develop proper genetic tools to manipulate the "rare" actinomycetes producing these molecules, *i.e.,* selecting the best promoter for heterologous gene expression from a collection of promoter-probe vectors. One final outcome of such work was the improvement of the production yield for teicoplanin and 40926 by means of genetic engineering, contributing to a possible price reduction of these molecules in the future.

In a following part of my work, the sequencing and analysis of genomes of "rare" actinomycetes producing putative GPAs allowed the identification of a new BGC for the synthesis of a novel antibiotic. I was in fact part of the international team which identified a new A40926-like antibiotic produced by *Nonomuraea coxensis* DSM 45129, named A50926.

Finally, I analysed more than 7000 genomes of actinobacteria available in public databases to map GPA resistance genes and GPA BGCs. Our bioinformatic analysis revealed how these resistance genes are widespread within Actinobacteria phylum and pointed to further novel GPA BGCs, awaiting further experimental evaluations.

II. INTRODUCTION

1.- Actinomycetes: soil dwelling bacteria capable to produce secondary metabolites

The Actinobacteria phylum is one of the largest taxonomic groups among *Bacteria* domain.[1] This phylum is further divided into six different classes: *Actinobacteria*, *Acidimicrobiia*, *Coriobacteriia*, *Nitriliruptoria*, *Rubrobacteria* and *Thermoleophilia*. Class *Actinobacteria* is the most diverse and it is divided into many orders, which are: *Actinopolysporales, Actinomycetales, Bifidobacteriales, Catenulisporales, Corynebacteriales, Frankiales, Glycomycetales, Jiangellales, Kineosporiales, Micrococcales, Micromonosporales, Propionibacteriales, Pseudonocardiales, Streptomycetales, Streptosporangiales, etc*. The term "actinomycetes" is a historical way to describe members of the Actinobacteria phylum that exhibit mycelial growth (indeed not all actinobacteria are mycelial in nature, as in case of *Corynebacterium* spp. and others). The term dates back to 1877, when Carl Otto Harz described *Actinomyces bovis*, the microorganism responsible for a serious cattle disease. Harz observed that this microbe grew in thin filaments and the outer layers showed ray-like structures, resembling fungal hyphae. In fact, *Actinomyces* means "ray fungus". This term was derived from the Greek words "aktis" which means ray, and "mykes/mukes" which means fungi.

The first detailed description of an actinomycete was done by Ferdinand Cohn in 1875, when he published a work where he included a microbe called *Streptothrix foersteri* (later found to actually be a *Streptomyces* sp.), which was described as a microorganism with elongated and branching cells reminiscent of those of fungi, but on a minute scale. The next to report on an actinomycete was Robert Koch in 1882, who described the agent causing tuberculosis.[2] However, actinomycetes were not regarded especially important until Selman Waksman discovered numerous clinically-relevant antimicrobial compounds coming from them. During his experiments on nitrogen-fixing bacteria from soil, he realised that some small colonies on agar plates were similar in appearance to those of bacteria but under the microscope observation looked much like fungi. These colonies were conical, compact and frequently pigmented, which are characteristic features of actinomycete colonies. In the past, some people did believe that actinomycetes were fungi and there was much controversy over the classification and origin of these organisms. Indeed, actinomycetes classification was a hard process until Selman Waksman and Arthur Henrici proposed them to be bacteria in 1943.[2]

Thus, modern microbiology views actinomycetes as aerobic spore forming Gram-positive bacteria with high guanine-plus-cytosine (G-C) DNA content in their genomes. Although growing as mycelial hyphae similarly to fungi, cells of actinomycetes lack nucleus, with chromosomes being organized into a nucleoid. Their cell membrane is covered with a peptidoglycan-containing cell wall. In fact, they are susceptible to antibacterial compounds. Actinomycetes are primarily soil inhabitants – especially widespread in alkaline and rich for organic matter soils – but have been also found widely distributed in a diverse range of aquatic ecosystems.[3] Actinomycetes, as the other soil bacteria, are mostly mesophilic, with an optimal growth temperature range between 25-

30 ºC and optimal pH around neutrality (6-9). However, a vast number of actinomycetes have been also isolated from extreme environments as well.[4]

As mentioned above, actinomycetes have complex lifecycle, developing true aerial mycelium and spores from vegetative hyphae. Briefly, lifecycle of a typical actinomycete (exemplified here with *Streptomyces* spp.) starts with the germination of a spore to vegetative hyphae, when all required environmental conditions are met. Vegetative hyphae grow by tip extension and eventually branch. Under adverse conditions, vegetative mycelium differentiates to erected structures called aerial hyphae. During this phase of the lifecycle most antibiotics are produced. This event is due to autolytic degradation of vegetative mycelium required to acquire enough nutrients (amino acids, aminosugars, lipids and nucleotides) needed to erect the aerial hyphae. Therefore, this process inevitably attracts competing microorganisms; antibiotics thus protect this pool of valuable nutrients. Aerial hyphae eventually differentiate to form spores, which are then dispensed and remain dormant until favorable environmental conditions stimulate spore germination (Figure 1). All these processes are strictly regulated by different genetic regulatory mechanisms.[5]

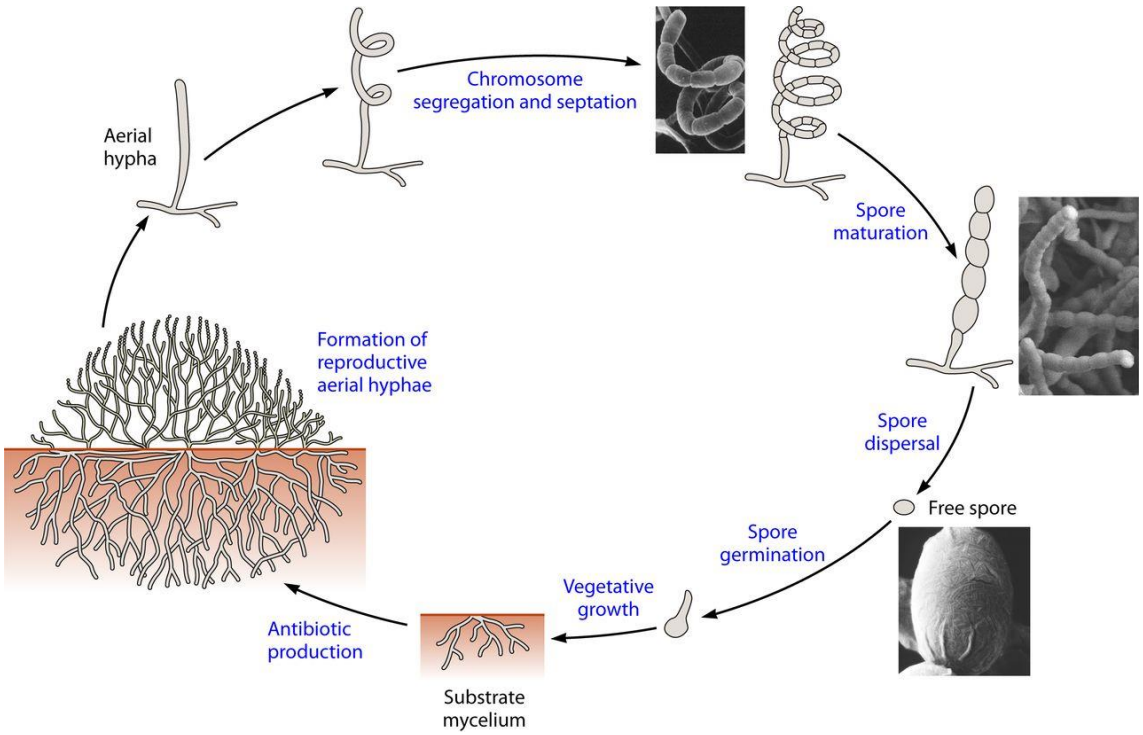


Figure 1. Schematic representation of the life cycle of sporulating actinomycetes (exemplified with *Streptomyces* spp.).[5] Lifecycle goes through vegetative phase, when vegetative mycelium is actively growing and consuming available nutrient sources. After the nutrients are exhausted, vegetative hyphae differentiate into spores, used for reproduction and dispersion. Please see main text for more details.

Because of such complex morphology and morphogenesis, as well as due to being part of multi-component soil ecosystems, actinomycetes have sizable genomes and extensive secondary metabolism, allowing them to produce a large number of natural

products. In fact, they produce approximately two-thirds of all naturally derived antibiotics used in clinics.[5] Secondary metabolites – also called more recently specialized metabolites – are molecules that are not required for survival under laboratory conditions, but which undoubtedly provide some advantage to the producer in its native environment. Since the discovery of actinomycin and streptomycin by the Waksman group, more than half a century ago, an enormous number of natural products coming from actinomycetes has been isolated and characterized. In general, the main roles of these specialized metabolites are defense, regulation and communication within and between different species. Their chemical and physiological functions are very diverse. Secondary metabolites are nowadays being used in different aspects of human life. Not only antibiotics, but also pesticides, antiparasitic drugs, herbicides, anti-inflammatory drugs, cardioactive compounds, antitumor drugs, antiviral drugs, antioxidants, immunoactive modulators or stimulators are produced by actinomycetes. Peculiarly, these bioactivities share a high degree of overlap. Indeed, more than half of all known antibiotics exhibit also some other bioactivities.[6]

The discovery process of secondary metabolites from actinomycetes requires the intervention of several disciplines. This interdisciplinarity was clear since the beginning when the group of Waksman was pioneering the natural product screening approach that was then used by pharma companies in the last century to discover the natural product sourced drugs that are in use today. Thus, microbiologists are interested in the discovery of new species that might produce novel molecules through the screening of environmental samples or analyzing bacterial genomes in databases, biochemists study the characteristics of the molecule *in vitro* and elucidate the chemical structure, geneticists modify BGCs through genetic engineering, bioinformaticians can predict their biosynthesis pathway and design new structures with better properties *in silico*, cell biologists test pharmacological properties *in vivo* of these novel compounds and organic chemists develop semi-synthetic derivatives to improve their characteristics. Today the classical natural product screening has been integrated with new technologies as the Next Generation Sequencing and the advanced bioinformatics tools, which have allowed to determine that actinomycetes possess large genomes encoding multiple – mostly silent – secondary metabolite biosynthetic gene clusters (BGCs), in a number that ranges from 20 to 50 per genome.[7,8,9,10] The bioiformatic search of BGCs in the actinomycete genomes and the prediction of the secondary metabolites that they could produce is named genome mining, which represents a powerful tool for discovering novel molecules. Most likely, a large number of these novel secondary metabolite BGCs is still hiden in the large actinobacterial genomes and requires to be discovered.

The best studied group of actinomycetes are the members of genus *Streptomyces*, which produce more than 70% of all antibiotics coming from actinomycetes. In fact, *Streptomyces* genus also dominates the soil bacterial populations, and accounts to more than 95% of the actinomycete strains isolated from soil.[11] However, more and more representatives of non-*Streptomyces* genera are becoming objects of interest in the last years, being able to produce novel natural compounds. For example, some successful antibiotics have been produced by "rare" actinomycetes such as vancomycin (produced by *Amycolatopsis orientalis*) or erythromycin (produced by *Saccharopolyspora*

erythraea). Thus, the proven capacity of these "rare" or "uncommon" genera to produce novel chemical structures have attracted an increasing attention.[12] "Rare" actinomycetes are defined as the actinomycete strains less frequently isolated than the "common" *Streptomyces* spp. using conventional methods, even though they may not actually be so rare in the environment. Since these microorganisms have not been intensively screened in the past, they undoubtedly represent an important source of novel specialized metabolites. While *Streptomyces* is a genus relativity easy to handle and to genetically manipulate, genetics and physiology of "rare" actinomycetes are still poorly known. Some relevant genera from this group include: *Actinomadura, Actinoplanes, Amycolatopsis, Actinokineospora, Acrocarpospora, Actinosynnema, Catenuloplanes, Cryptosporangium, Dactylosporangium, Kibdelosporangium, Kineosporia, Kutzneria, Microbiospora, Microtetraspora, Nocardia, Nonomuraea, Planomonospora, Planobispora, Pseudonocardia, Saccharomonospora, Saccharopolyspora, Saccharothrix, Streptosporangium, Spirilliplanes, Thermomonospora, Thermobifida*, and *Virgosporangium*.[13]

This thesis is specially focus on the genus *Nonomuraea* (belonging to the *Streptosporangiaceae* family), whose members are considered as "rare" actinomycetes. Nowadays, more than 70 species within *Nonomuraea* genus have been described and they are found widely distributed in soil, freshwater and marine environments. A broad range of valuable bioactive compounds are produced by *Nonomuraea* spp., including antimicrobial and anticancer drugs. Thus, this genus represents a great potential for biotechnological applications. In contrast, only few tools for their genetic manipulations have been developed so far.[14] Probably, the most important bioactive metabolite produced by a *Nonomuraea* species is A40926 which is the precursor of dalbavancin – a clinically relevant glycopeptide antibiotic – produced by *N. gerenzanensis* ATCC 39727.[15] Recently, the genome of this strain was sequenced and bioinformatics tools allowed to identify 32 BGCs of different interesting specialized metabolites.[16] For instance, a putative BGC for an enediyne-like antibiotic was found. Currently, enediyne-like compounds draw big attention, being a very promising molecules for anticancer therapy.

2.- Antibiotics: history and overview

The discovery and clinical use of antibiotics are considered as one of the major scientific achievements and started with the discovery of penicillin and sulfonamide antibiotics at the beginning of 20th century. This gave rise to the so-called "Golden Age of Antibiotics" period (1940s-1960s) during which the most of the antibiotic classes used in the clinic today were discovered (Figure 2).

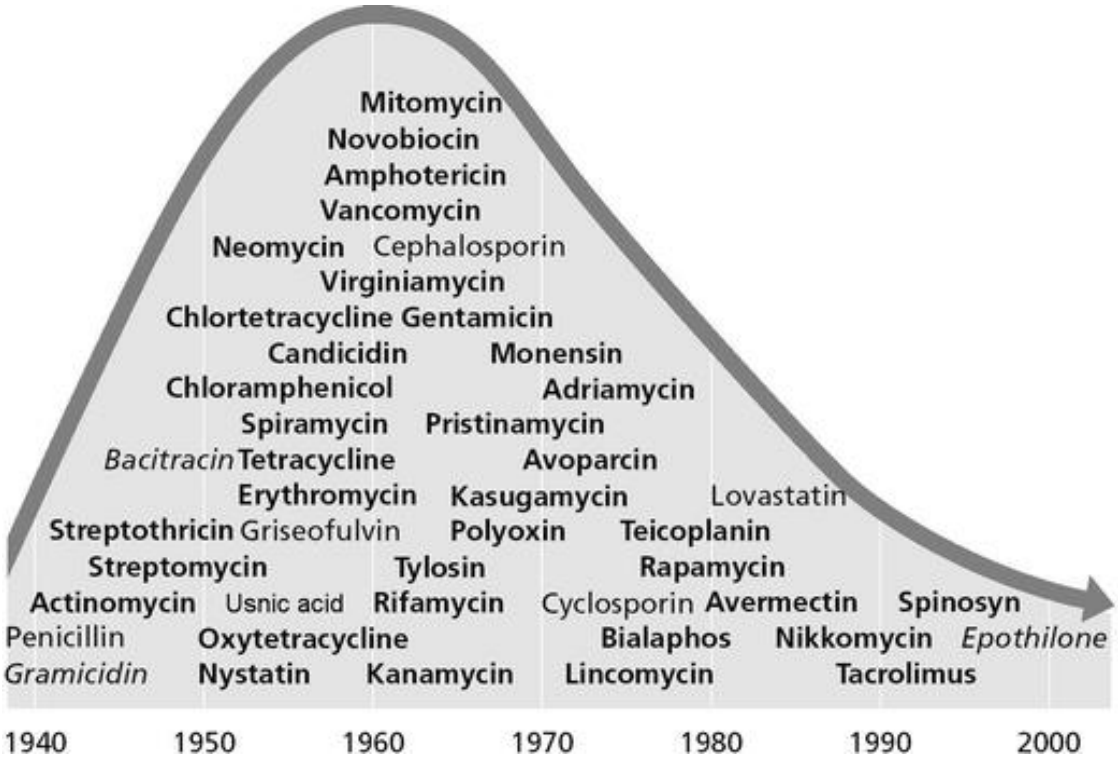


Figure 2. Chronology of the discovery of some new antibiotics and other relevant natural products. Bold type: actinomycete products; italic type: non-actinomycete bacteria products, normal type: fungal products (taken from [2]).

The use of antibiotic-producing microorganisms to prevent some diseases dates back thousands of years. Natural products with antimicrobial activity were probably used in pre-antibiotic era. One study suggests that people from Roman town Herculaneum were protected against some infections due to consumption of dried fruits contaminated by antibiotic producers.[17] In addition, it has been found that human skeletal remains from several archaeological places in Egypt,[18] Sudan[19,20] and Jordan[21] exhibited histological evidence of tetracycline consumption. Others studies suggest that another possibility of exposure to antimicrobial compounds in the pre-antibiotic era could be the utilization of remedies in traditional or alternative medicine in the past. One relevant example is artemisinin which was used in traditional Chinese medicine as an antidote for many illnesses for thousands of years. Nowadays, we know artemisinin is a potent anti-malarial drug.[22]

However, the development of modern antimicrobial drugs started with the synthetic arsenic-based Salvarsan, used for the treatment of *Treponema pallidum* - the causative agent of syphilis. Salvarsan was discovered at the beginning of the 1910s by Paul Ehrlich group.[23] Nevertheless, its toxicity limited the commercialization. Later in 1928, Alexander Fleming discovered penicillin.[24] Then, chemists at Bayer in Germany developed Prontosil in 1931 – a new chemical compound with the ability to treat general bacterial infections in humans – while Fleming was still trying to purify penicillin. Prontosil was the trade name of the first synthetic commercial antibacterial, being cheap

to produce and off-patent. Consequently, it was widely used during World War II saving many human lifes.[25]

Inspired by these events, Selman Waksman started to study bacteria as producers of antimicrobial compounds in the late 1930s. Waksman discovered numerous antibiotics produced by soil-dwelling actinomycetes, such as neomycin and streptomycin, this last really effective agent against *Mycobacterium tuberculosis*, the microbe responsible for tuberculosis.[26] In 1942, Waksman proposed the term "antibiotic" itself, referring to "a compound produced by one microorganism which is capable of killing or inhibiting the growth of the another." Works of Waksman's set a starting point for the Golden Age of antibiotic discovery.

Antibiotics have revolutionized human history and modern medicine in particular. Thanks to antibiotics, leading causes of death changed from communicable diseases to non-communicable diseases such as cancer, cardiovascular disease and stroke: until the beginning of the 20th century, the main causes of mortality worldwide were infectious diseases, caused by bacteria, such as: cholera, diphtheria, pneumonia, typhoid fever, tuberculosis, typhus, syphilis, *etc*. As an example, antibiotics allowed to reduce from 25% mortality in 1900 to less than 1% in 1945 in England.[27] In United Stated, deaths due to bacterial illnesses decreased from 277541 (247.7 death per 100000 population) in 1936 to 93014 (59.7 death per 100000 population) in 1952: life expectancy at birth was 58.5 years in 1936 and became 68.6 years in 1952, showing an impressive difference of 10.1 years in just less than two decades.[28]

Unfortunately, bacteria have developed many mechanisms of resistance towards antibiotics.[29] Dramatic increase of antibiotic resistance could be observed in hospitals, communities and in the environment due to antibiotic overuse. The fight between antibiotics and antibiotic resistance in pathogens has always existed. Penicillin was discovered in 1928, and a bacterial penicillinase was identified shortly after – in 1940, several years before the introduction of penicillin in clinic.[30] In fact, it is currently known that occurrence of the resistance genes predates our use of antibiotics by thousands of years.[31] The current alarm on antibiotic resistance is a complex and relatively old topic which remained rather anecdotal for a long time. However, it was not until the end of the 20th century that consciousness emerged. The massive use by humans has only accelerated this natural process. The increase in antibiotic-resistant pathogens is a consequence of several factors, including – but not limited to – high proportion of medical prescriptions, interruption of therapies and large-scale use as growth promotors in livestock farming.[32] Another important factor is the vertiginous drop in research and development of novel antibiotics in the last years. Presently, only few new molecules are in the clinical trial phase.[33] Indeed, many companies simply stopped this activity since it was not rewarding as before.[34]

According to a recent report, commissioned by an interagency group of the United Nations, antibiotic-resistant infections are going to kill more people in just over three decades. It is estimated that around ten million people are set to lose their lives every year due to drug resistant pathogens by 2050 while eight million people are going to die with cancer (Figure 3).[35] In addition, this problem has been accentuated whith the recent emergence of multidrug resistance (MDR) pathogens – so-called "superbugs" – which

refers to bacteria with enhanced morbidity and mortality due to high levels of resistance to different antibiotics, especially those traditionally used, making the hospital treatment of these infections very difficult.

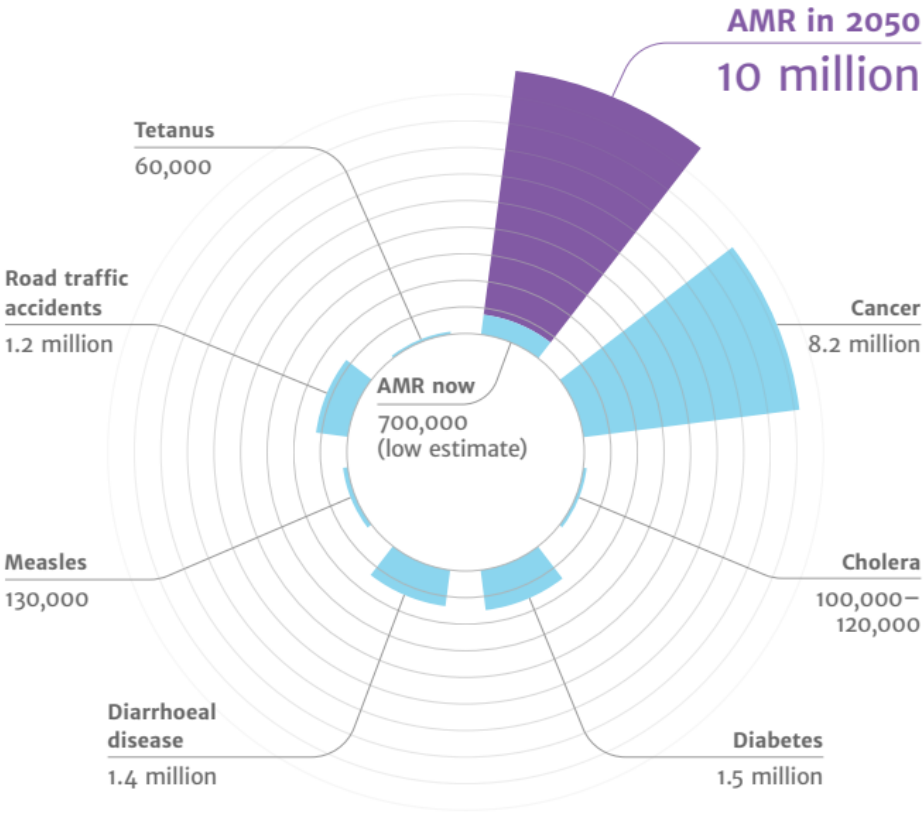


Figure 3. Estimation of the main causes of death by 2050. The greatest cause will be due to antimicrobial resistance (AMR) (adapted from [35]).

3.- Glycopeptide antibiotics

Glycopeptide antibiotics (GPAs) are drugs of last resort for treating severe infections caused by Gram-positive pathogens such as *Clostridioides difficile, Enterococcus* spp. and *Staphylococcus aureus*.[36] GPAs are widely used against methicillin-resistant *S. aureus* (MRSA), which is a major cause of community-acquired infections and results in serious hospital difficulties so far.[37] In addition, *S. aureus* managed to evade not only methicillin, but also a significant number of other classical antibiotics such as tetracycline, penicillin and erythromycin B. Thus, it is one of the most dangerous bacterial pathogens we face today.

Clinically relevant GPAs include first-generation vancomycin and teicoplanin, which are natural products. Second generation GPAs include semi-synthetic molecules recently approved for clinical use (Figure 4).[38] These are dalbavancin, telavancin and oritavancin. In addition to clinically relevant GPAs, others related natural products have been discovered and a vast number of semi-synthetic analogues has been designed and tested.[39]

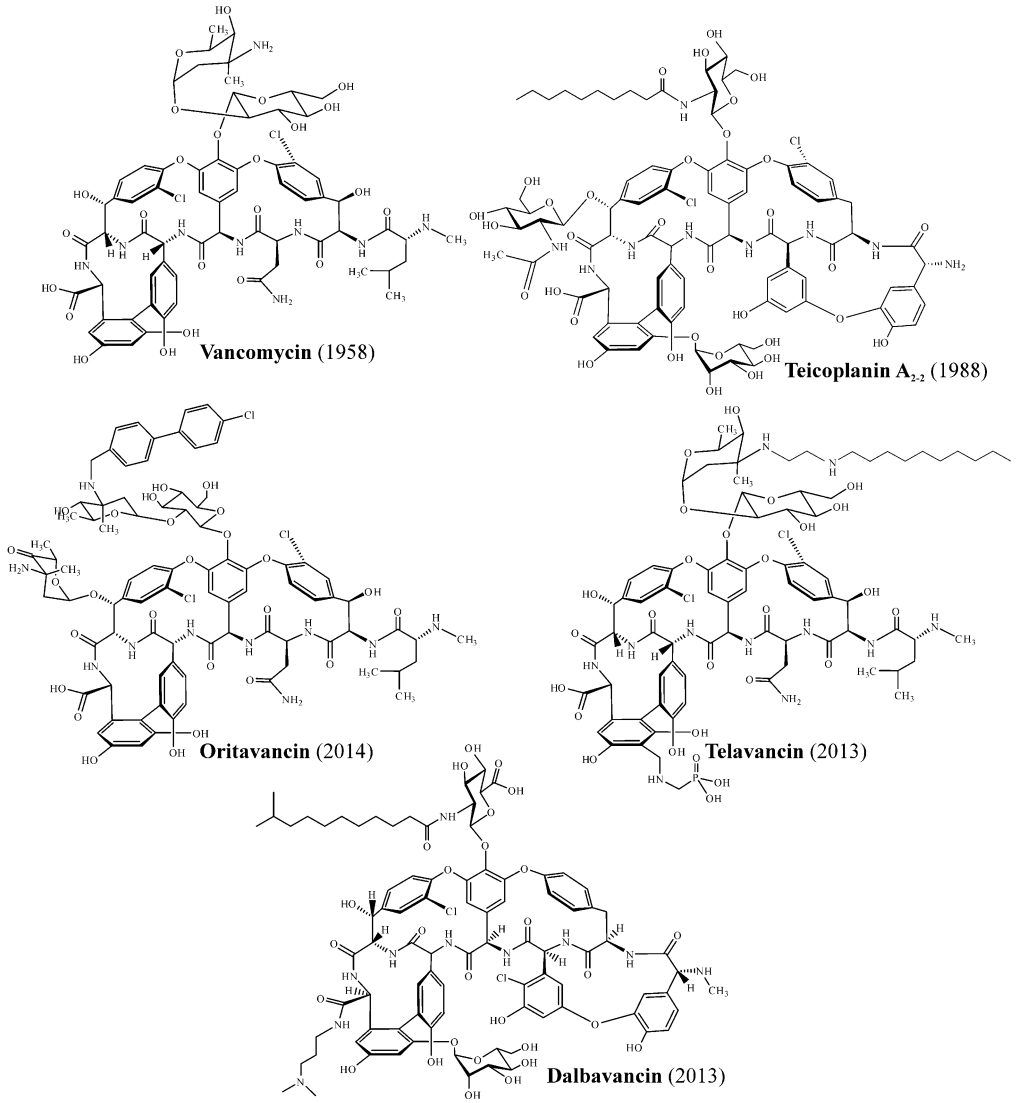


Figure 4. Chemical structures of first- and second-generation clinical GPAs. First generation includes: vancomycin (approved in 1958) and teicoplanin (approved in 1988). Second generation includes: telavancin (trade name Vibativ®, approved in 2013), oritavancin (trade name Orbactiv®, approved in 2014) and dalbavancin (trade names Dalvance® in the United States and Xydalba® in the European Union, approved in 2014).

GPAs inhibit peptidoglycan biosynthesis, thus interrupting the normal formation of bacterial cell walls. GPAs bind with high affinity to the D-alanine-D-alanine (D-Ala-D-Ala) C-terminus of the nascent peptidoglycan chain, blocking the transpeptidation and transglycosylation cross-linking.[40] After several decades of clinical use, the emergence of

GPA-resistant Gram-positive pathogens such as vancomycin-resistant *S. aureus* (VRSA) and vancomycin-resistant enterococci (VRE) has introduced a serious challenge to public health when we are in a situation with only few new drugs are under investigation for commercial use. This has revived the interest in the search for new effective treatments including the development of new GPA derivatives.

3.1.- History of the discovery and use of clinically relevant GPAs

Vancomycin was the first clinically relevant GPA to be discovered in 1953 by Eli Lilly & Co (Indianapolis, US), produced by an *Amycolatopsis orientalis* strain found in a soil sample collected in Borneo.[41] It was approved for the treatment of penicillin-resistant staphylococcal infections by Food and Drug Administration (FDA) in 1958. Ristocetin – also named ristomycin – comes from *Amycolatopsis lurida* NRRL 2430 and was reported by Abbott Laboratories (Chicago, IL, US) in 1956.[42] Ristocetin was introduced in clinics in 1957 but removed from the market very soon because of side effects (such as thrombocytopenia and platelet agglutination).[43] Although ristocetin is no longer used for the treatment of bacterial infections, it is still in use as a diagnostic marker to detect some genetic disorders like von Willebrand disease or Bernard-Soulier syndrome.[44] Teicoplanin was the second natural GPA, produced by *Actinoplanes teichomyceticus* NRRL B-16726, which was developed as a clinical drug. This GPA is still on duty and is successfully used in clinics. Teicoplanin was first reported by Lepetit Research Center (Milan, Italy) in 1978[45] and then introduced to clinical use for treating Gram-positive infections in Europe and in Japan in 1988 and 1998, respectively (Figure 5). Teicoplanin shows antimicrobial activity comparable to vancomycin but possesses superior activity toward enterococci and some other microorganisms.[46]

GPAs used in clinic are not the only ones described. There is an extensive number of other natural compounds but the majority were reported in the second half of the 20th century.[39] For instance, avoparcin and actaplanin were described in 1968[47] and 1976,[48] respectively. Avoparcin was massively used as a growth promoter in animal feeds until VRE emergence in cattle halted the further usage in the United States and many other countries.[49] Other important GPAs, such as balhimycin,[50] chloroeremomycin,[51] A47934[52] or complestatin[53] served as models to investigate GPA biosynthesis, genetics, *etc*.

Since the resistance to vancomycin and teicoplanin significantly increased in pathogens, some companies focused on the creation of novel synthetic GPAs. As a result, semi-synthetic GPAs appeared for use in clinic at the beginning of the 21st century (Figure 5). The most significant modifications in these new drugs were introductions of hydrophobic moieties into the glycopeptide scaffold, which often increased their activity.[54] Hence, telavancin was launched in 2009, followed with dalbavancin and oritavancin being approved in 2014. Telavancin was developed by Theravance Biopharma (California, US). It is a vancomycin derivative, modified with decylaminoethyl lipophilic tail – attached to the vancosamine sugar – and hydrophilic (phosphonomethyl)aminomethyl group attached to the amino acid (AA) 7.[55,56,57]

Oritavancin was synthesized by Eli Lilly & Co and it is a derivative of the naturally occurring GPA chloroeremomycin, produced by *Amycolatopsis orientalis* PA-42867. Addition of a 4'-chlorobiphenylmethyl substituent to the disaccharide moiety, along with the additional 4-epivancosamine moiety attached to the AA6 significantly enhanced the activity against VRE and VRSA.[58,59] Dalbavancin is a derivative of the natural product A40926 which was discovered by Lepetit Research Center in the mid 1980's. It was isolated from the rare actinomycete *N. gerenzanensis* ATCC 39727 – reported previously as *Actinomadura* sp. strain ATCC 39727 before reclassification – isolated from an Indian soil.[60] The difference between A40926 and dalbavancin is the amidation of the peptide carboxyl group with a dimethylaminopropylamide side chain.[61]

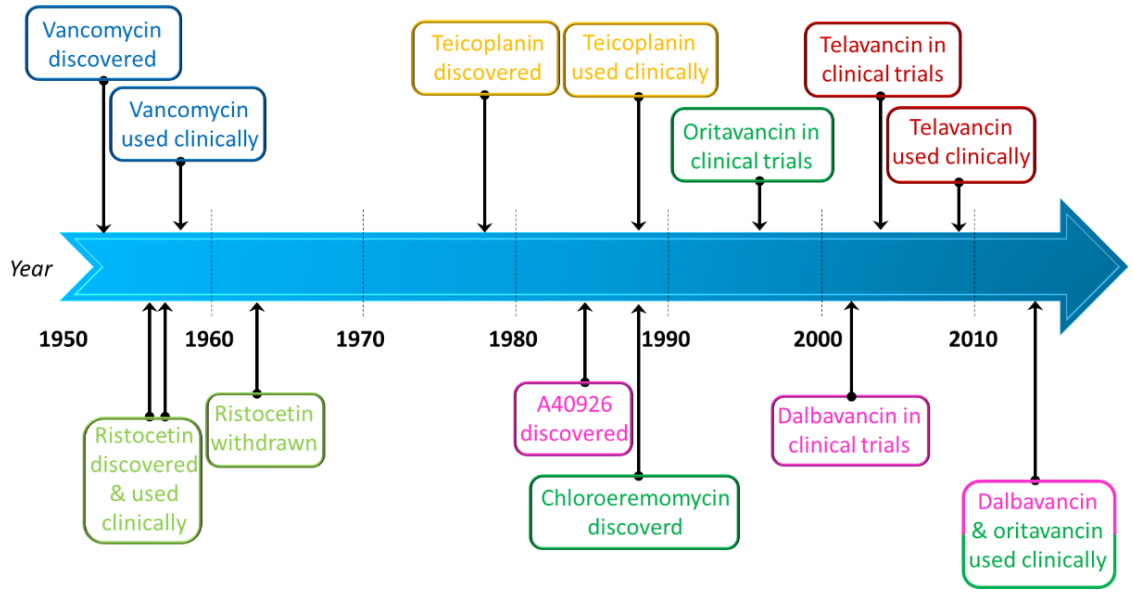


Figure 5. Timeline of discovery and the clinical usage of GPAs. Both dates of discovery, clinical trials and introduction to clinical practice are shown.

3.2.- Classification of glycopeptide antibiotics

Chemically, GPAs consist of a heptapeptide core (aglycone), synthesized by nonribosomal peptide synthetase (NRPS) and then cross-linked by P450-like non-heme oxygenases to yield a rigid peptide scaffold; aglycone undergoes further tailoring steps, which may include attachment of sugar moieties, chlorine atoms, methyl groups, sulfate groups and aliphatic side chains.

The first complete GPA structure to be elucidate was vancomycin in 1982,[62] followed by ristocetin and teicoplanin. Then, crystal structure of vancomycin bound to a *N*-acetate-D-alanyl-D-alanine – which mimics its target – was determined in 1998.[63] Afterwards, a large number of GPA structures were solved. Structural studies of GPAs have clarified the biological mode of action and revealed the structural differences, summarized in the description of five structural types: I to V (Figure 6). The seven amino acids that form aglycone are commonly denoted as AA1 to AA7 from N- to C-terminus of the oligopeptide and the five aromatic rings are named from A to E. Type I GPAs (e.g.,

vancomycin and balhimycin) contain aliphatic amino acids in AA1 and AA3 positions. Type I GPAs have aromatic side rings A and B linked (AB cross-link), as well as C and D (C-O-D cross-link) together with D and E (D-O-E cross-link). By turn, in type II GPAs (*e.g.,* keratinimicin and actinoidin) AA1 and AA3 bear aromatic side rings (F and G, respectively), but all cross-links are the same as in type I compounds. Types III (e.g., ristocetin and actaplanin) and IV (e.g., teicoplanin and A40926) include amino acids with aromatic side chains in AA1 and AA3 positions (like in type II), but here they are cross-linked, forming a F-O-G cross-link. Additionally, type IV contains a long fatty acid-chain attached to a sugar moiety, which type III GPAs lack.[39] Type V (e.g., complestatin and kistamicin) is a quite divergent and heterogeneous group, that maybe will need a reclassification of the molecules actually attributed to in future. Molecules belonging to this type uniquely contain a tryptophan residue at position AA2 and are not glycosylated. Recently, two new GPAs with nine amino acids forming their aglycone were discovered. These molecules, called corbomycin[64] and GP6738[65], were classified into type V. This peculiar group also shows variable oxidative cross-linking patterns. All molecules include D-E and C-O-D cross-links, but in some cases an additional A-O-B (kistamicin) or A-B and F-O-G (corbomycin) cross-linking might be present.

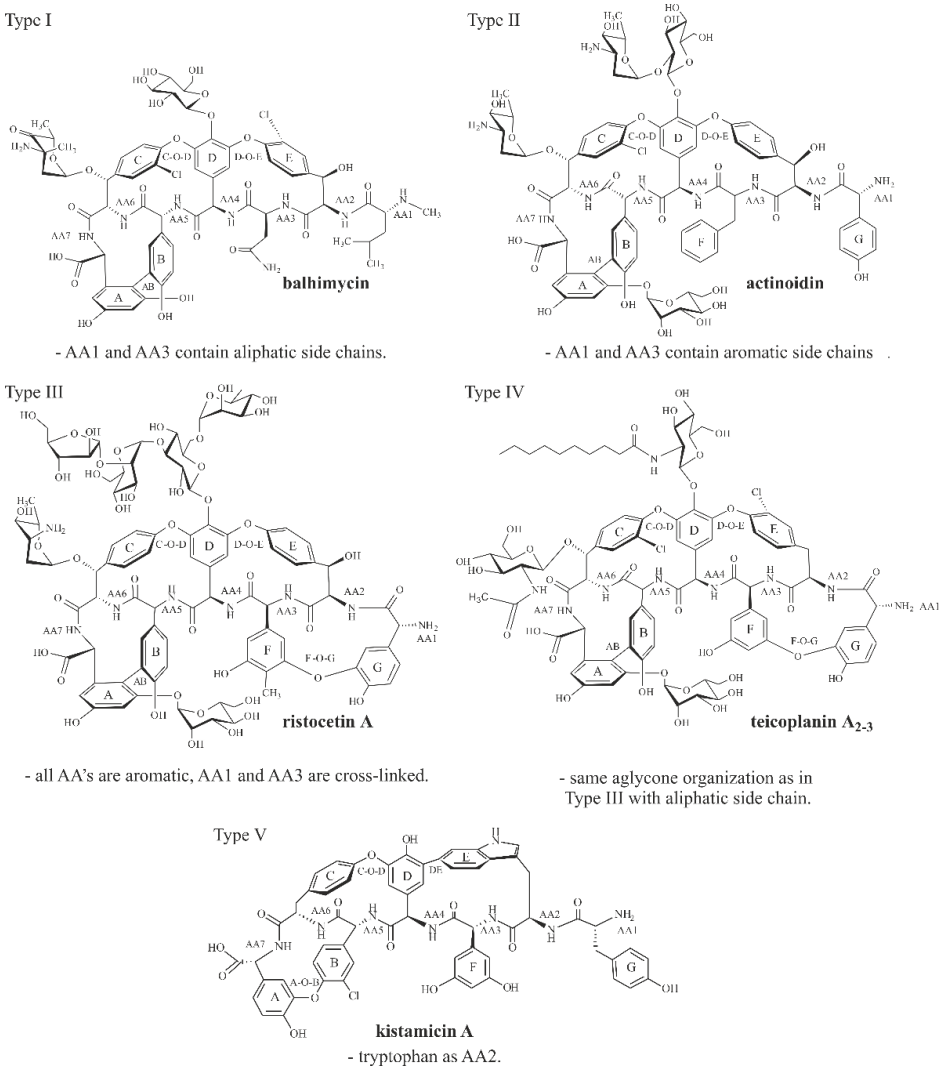


Figure 6: Structural classification of GPAs into five types. Please refer to the main text for more details.

3.3.- GPA biosynthetic pathways: genes and their functions

GPA NRPSs are multimodular giant enzyme machines that assemble seven amino acids in an assembly line manner. They are encoded by three or four genes, oriented in one direction – with a single exception of A40926 NRPS genes, which are bizarrely rearranged due to unknown reasons. Each NRPS module is typically dedicated to the activation and coupling of a single amino acid, and the primary sequence of the peptide product depends on the order of NRPS modules. Moreover, each module includes diverse domains, playing certain roles in oligopeptide assembly. These domains are: condensation (C-domain, responsible for peptide bond formation), adenylation (A-domain, responsible for amino acid recognition), thiolation (T-domain, also known as peptidyl carrier protein – PCP – responsible for the transfer of new amino acids to the growing oligopeptide), epimerization (E-domain, changing the stereochemistry of amino acids) and thioesterase (TE-domain, responsible for the termination of the biosynthesis) domains (Figure 7).[66,67]

As in many other natural products, heptapeptide core of GPAs may contain proteinogenic (Tyr, Leu, Asn, Ala and Glu) as well as non-proteinogenic amino acids, like: 4-hydroxyphenylglycine (Hpg), β-hydroxytyrosine (Bht) and 3,5-dihydroxyphenylglycine (Dpg).[68] Genes necessary for the synthesis of these non-proteinogenic amino acids are always present within GPA BGCs. Hpg synthesis was elucidated in chloroeremomycin producer. Four genes are needed for Hpg production from the common bacterial metabolic intermediate chorismate.[69] HmaS converts 4-hydroxyphenylpyruvate into L-4-hydroxymandelate, then oxidized to L-4-hydroxybenzoylformate by Hmo and finally transformed in the non-proteinogenic amino acids by HpgT.[69, 70] Dpg requires five genes - DpgA, DpgB, DpgC, DpgD and, once again, HpgT – for the biosynthesis from malonyl-CoA, investigated in the examples of chloroeremomycin and balhimycin producers.[71] DpgA, DpgB and DpgD constitute a complex which incorporates malonyl-CoA units. The product – 3,5-dihydroxyphenylacetate – is converted by DpgC into 3,5-dihydroxyphenylglyoxylate and then Dpg synthesis is catalyzed by HpgT.[71, 72] Biosynthetic pathways for Bht are different among the vancomycin-like GPAs and teicoplanin-like ones. Three genes are involved in the in β-hydroxylation of tyrosine in vancomycin-type producers and only one gene in teicoplanin-type.[73] In vancomycin-like balhimycin biosynthetic pathway, *bhp, bpsD* and *oxyD* are involved in the synthesis of Bht in its free form in the cytoplasm, which is then incorporated into the growing heptapeptide core.[74] Indeed, a single β-hydroxylase (Dbv28) acts on tyrosine, when it is already incorporated into the NRPS-bound aglycone, to synthesize Bht in teicoplanin-like A40926.[75]

The last module of GPA NRPSs always contains a unique domain, known as X-domain. It is necessary for the recruitment of cross-linking oxygenases to the NRPS-bound peptide to achieve the final side-chain crosslinking.[76,77] These oxygenases belong to cytochrome P450 superfamily of non-heme monooxygenases and are divided into four functional classes: OxyA, B, C and E. The number of crosslinks and oxygenases depends on the GPA type. Two main cross-linking patterns could be differentiated: vancomycin-type BGCs encode three oxygenases (OxyA, OxyB and OxyC) and corresponding

molecules possess three cross-links; whereas teicoplanin-type BGCs encode an additional oxygenase (OxyE) and the corresponding molecules contain one additional cross-link (Figure 7).[73] The succession of cross-linking reactions was established as OxyB>OxyA>OxyC in vancomycin-type producers[78] and OxyB>OxyE>OxyA>OxyC in teicoplanin-type producers.[79] Cross-linking patterns and involved enzymes for type V are heterogeneous and require further investigation.

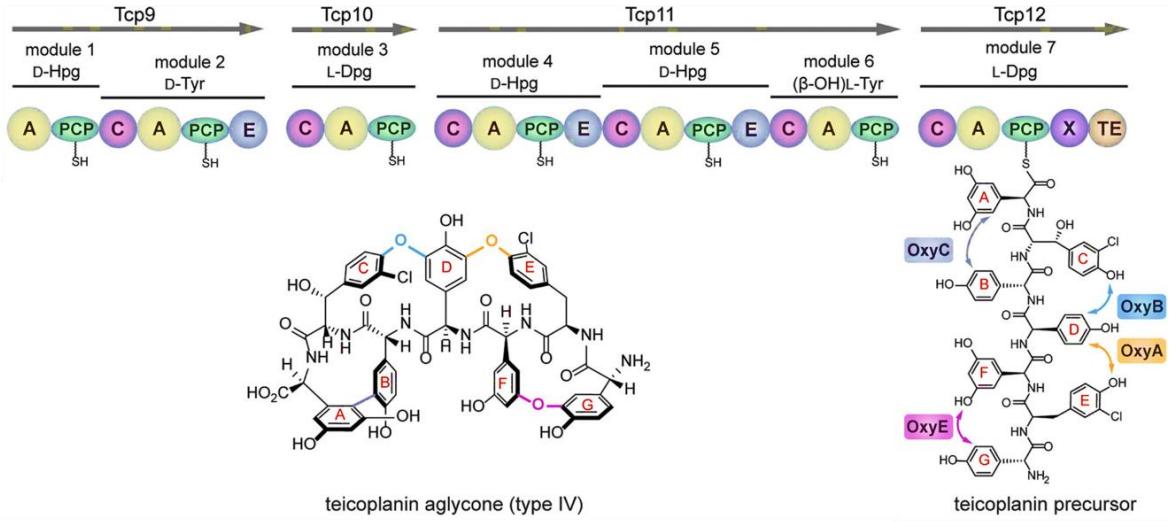


Figure 7: Schematic representation of NRPS biosynthesis of teicoplanin aglycone. NRPS domains are indicated using the following nomenclature: A (adenylation), PCP (peptidyl carrier protein), C (condensation), E (epimerisation), X (domain responsible for P450 cross-linking oxygenases recruitment) and TE (thioesterase). Incorporated amino acids are indicated above the modules: Hpg (4-hydroxyphenylglycine), Dpg (3,5-dihydroxyphenylglycine), Tyr (tyrosine). Oxidative crosslinking patterns are shown. Bonds catalyzed with OxyA are shown in yellow color, with OxyB in sky blue, with OxyC in navy blue and with OxyE in purple (taken from [80]).

As mentioned above, cross-linked aglycone is further decorated with different moieties, such as sugar residues, chlorine atoms, methyl groups or lipid chains (Figure 8). These modifications are carried out by different tailoring enzymes which are coded within BGCs, such as: glycosyltransferases, halogenases, methyltransferases, acyltransferases and sulfotransferases. These enzymes are regiospecific, modifying only certain residues on the aglycone. Tailoring of the GPA scaffolds alter their chemical and biological properties. For example, addition of aliphatic side chain improves pharmacokinetics of the antibiotics,[81] whereas sulfation decreases the induction of resistance mechanisms in actinomycetes - and likely will in pathogens.[82]

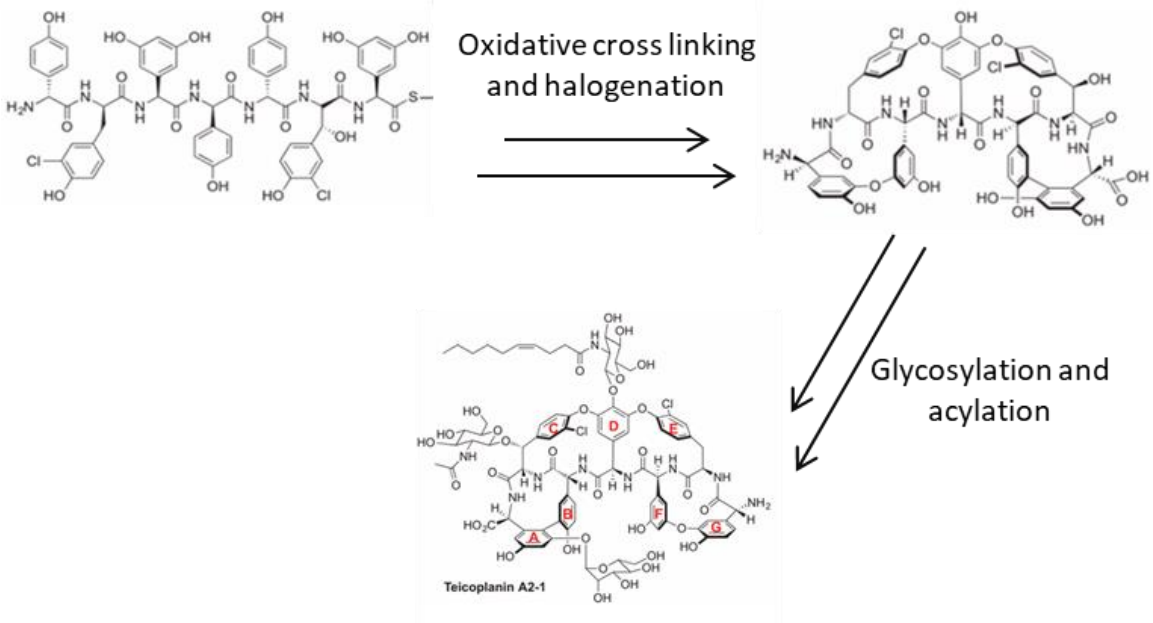


Figure 8. Schematic representation of post-NRPS assembly line tailoring enzymatic modifications to convert teicoplanin aglycone to the final teicoplanin GPA.

Glycosylation. As the name suggests, GPA glycosylation is the most notable decoration observed for these compounds. Glycosylation enhances the solubility of the molecule, playing an important role in the delivery of the antibiotic to its target. Glycosylation reactions modify one specific amino acid, mainly amino acids at position 4 and less frequently at positions 6 and 7. Although glycosylation at position 6 may include a range of different sugar moieties, Hpg4 modifications are restricted to glucose or glucosamine and Dpg7 is modified only with mannose.[83] Different glycosylation degrees could be present in GPAs, ranging from two (A40926), three (teicoplanin A_2) and even up to five (actaplanin A) or six (ristocetin A).[39] One exception is A47934, that does not undergo any glycosylation[52] and its BGC lacks the genes for glycosyltransferases. All modifying sugars could be divided into following categories: the hexo- and 6-deoxyhexopyranosides, such as D-glucose and L-fucose and the aminotrideoxyhexopyranosides, such as L-ristosamine and L-vancosamine. Some other rare carbohydrates have also been described to decorate GPA aglycone, such as L-oxovancosamine in balhimycin or D-glucosamine present in teicoplanin.[39]

Methylation. GPAs are often methylated, although non-methylated compounds also occur. Some GPA BGCs encode proteins that have been shown to have C-, N- or O-methyltransferase activity. They catalyze the addition of a methyl group in the heptapeptide core or sugar moieity.[73] N-methylation occurs before glycosylation.[84]

Halogenation. Chlorination is another common GPA modification. Only few molecules among GPAs lack chlorines attached to the aglycone, for instance ristocetin.[83] The role of chlorination is not understood yet, but the main hypothesis is that it enhances positively the antimicrobial activity.[85] BGCs contain genes for one or two halogenases, responsible for the installment of up to four chlorine atoms. BGCs coding for one halogenase commonly produce GPAs chlorinated at Tyr/Bht2 and Bht6, while BGCs

coding for two halogenases produce compounds additionally halogenated at Hpg1, Dpg3 or Hpg5.[73] Halogenation occurs before the release of the peptide from the NRPS, probably after the first crosslinking by OxyB.[84]

Sulfation. Sulfation is relatively rare in GPAs. Only three natural GPAs have been characterized to contain sulfate groups: A47934, UK-68,597 and pekiskomycin. Sulfate groups were found attached to the residues Hpg1, Dpg3, Hpg4 and Bht6.[86,87,88] Sulfation is likely one of the last modifications but is not still clear if it occurs before or after glycosylation.[89]

Acylation. Acylation is the process of adding an acyl aliphatic side chain to GPA aglycone. Acylated GPAs such as teicoplanin and A40926 are also called lipoglycopeptide antibiotics (LGPAs). This modification adds favorable antibacterial characteristics against some VRE strains due to its anchoring membrane ability.[81] Substrate promiscuity of the acyltransferases involved in acylation results in a variety of GPA products. For instance, five major teicoplanins ($A_{2\text{-}1}$ to $A_{2\text{-}5}$) and additionally sometimes four minor components could be produced having slight differences in the acyl chain length and branching.[67] Clinical teicoplanin complex component is mainly constitute by the 8-methylnonanoic (iso-C10:0) acid, $A_{2\text{-}2}$. Similarly, A40926 is a complex of compounds varying in side fatty acid length and branching (A_0, A_1, B_0, B_1).[90] These complex of fatty acyl variants could be found in the natural producer – depending on the fermentation condition – but control of the feedstock during industrial fermentation yields the commercial antibiotic form.

3.4.- Glycopeptide mode of action and of resistance

Bacterial cell wall is crucial for the maintenance of structural integrity and cell shape. Gram-positive bacteria are surrounded with a thick (10-40 nm) layer of peptidoglycan (PG) which is decorated with lipoteichoic acids and wall-teichoic acids. Gram-negative bacteria, indeed, are covered with a thinner (3-6 nm) PG cell wall, which itself is surrounded by an outer membrane containing lipopolysaccharides. PG – also called murein – has a structure of a layer formed with repeating linear chains of two alternating amino sugars: *N*-acetylglycosamine (Glc*N*Ac or NAG) and *N*-acetylmuramic acid (Mur*N*ac or NAM). These monomers are connected with a β-(1,4)-glycosidic bond. Each *N*-acetylmuramic acid is attached to a short peptide unit consisting commonly of L-Ala-D-Glu-L-Lys-D-Ala-D-Ala oligopeptide (some variations are reported which includes peptide substitutions or D-Glu and meso-diaminopimelic acid amidation).[91]

PG biosynthesis is a target for numerous antibiotics. They usually inhibit the biosynthetic enzymes or directly bind to PG intermediates. GPAs in particular block cell wall synthesis by binding to the D-Ala-D-Ala terminus of pentapeptide stem. This binding is stabilized by five hydrogen bonds and hydrophobic van der Waals force (Figure 9).[54] GPAs are exclusively effective against Gram-positive bacteria because the external membrane in Gram-negative bacteria is impermeable to them, preventing GPAs from reaching PG.

Since the resistance is an essential prerequisite for the antibiotic producer to ensure its survival from the lethal effects of its own product, all producers are resistant to their own compounds.[92] Self-resistance towards GPAs is determined by the so-named *van* genes and it is co-regulated with GPA biosynthesis to avoid suicide during the antibiotic production. In most producers, *vanHAX*-operon is part of corresponding GPA BGCs. This operon, that is similar to one reported in GPA resistant pathogens (enterococci and staphylococci),[38] was found in all the published GPA producers, except *N. gerenzanensis* ATCC 39727, which employs an alternative mechanism of self-protection based on the activity of the only carboxypeptidase VanY.[93] Other particular case is present in *Amycolatopsis balhimycina* where *vanHAX* genes are present in the chromosome, but they are not adjacent to the balhimycin BGC.[94] The similarity of *van* gene sequence and organization between the resistant pathogens and the GPA producing actinomycetes have suggested that the latter might represent the original source of genes involved in the formation of resistant PG precursors in pathogens.[95,96]

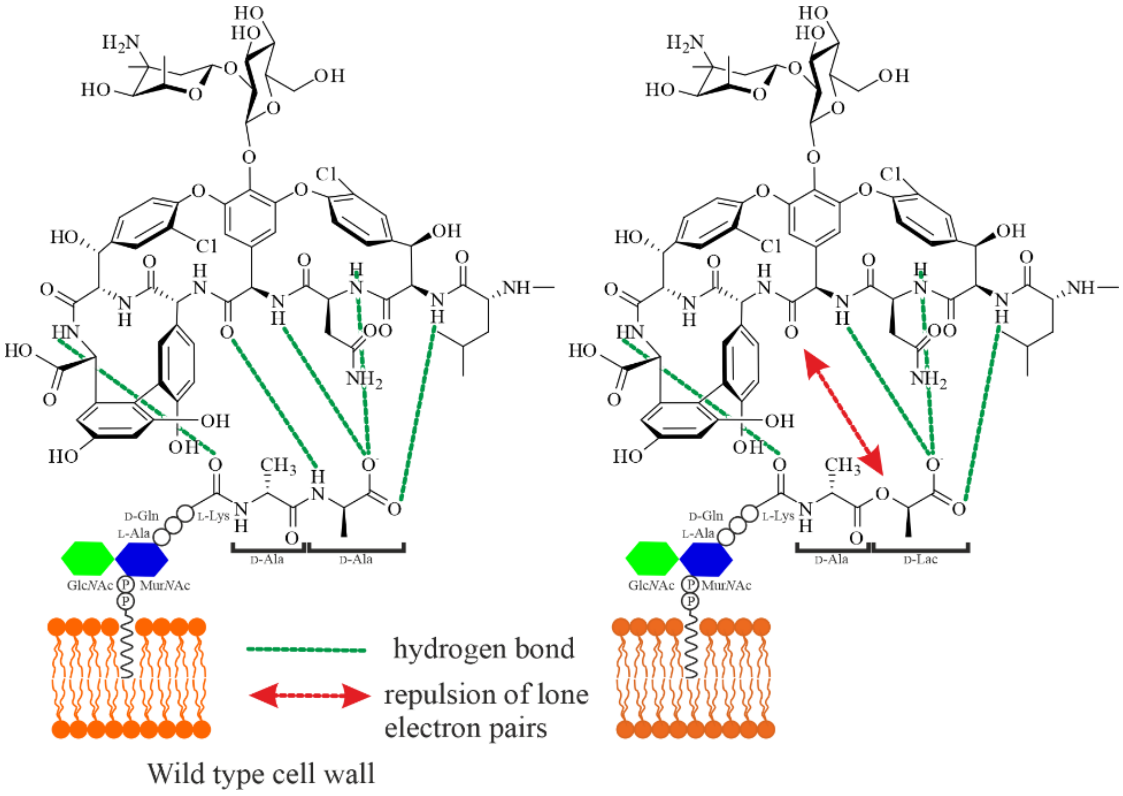


Figure 9. Interaction of vancomycin with *N*-acyl-D-Ala-D-Ala and *N*-acyl-D-Ala-D-Lac termini. Five hydrogen bonds are represented (green lines) between the GPA and the *N*-acyl-D-Ala-D-Ala termini while four hydrogen bonds are present in *N*-acyl-D-Ala-D-Lac termini. Red arrow indicates the repulsion of lone electron pairs.

The *vanHAX* genes encode enzymes for the synthesis of modified PG precursors, in which the usual D-Ala-D-Ala terminus of the growing PG is replaced with D-alanyl-D-lactate (D-Ala-D-Lac). This replacement markedly reduces the GPA affinity to their

molecular target.[97] The most clinically relevant incidence of GPA resistance are reported in VanA enterococci and staphylococci, and VanB in enterococci. The first genotype is highly resistant to vancomycin and teicoplanin, whereas the second genotype is resistant to vancomycin but remains susceptible to teicoplanin. In both of them, *van* genes, often present on Tn1546 transposon or closely related mobile genetic elements, encode for a two-component regulatory system (VanR and VanS), for the synthesis of modified PG precursors ending in D-Lac (VanH and VanA) and finally for the hydrolysis of normal precursors (VanX and VanY). A protein with unknown function (VanZ) is also generally encoded.[98] Briefly, VanS is a membrane-spanning histidine kinase that undergoes auto-phosphorylation in response to GPA presence. VanS subsequently phosphorylates VanR, the cognate response regulator of VanS. Afterwards, the phosphorylated VanR activates the expression of the *vanHAXYZ* genes. VanH is a D-stereospecific lactate dehydrogenase which converts pyruvate to D-Lac.[99] VanA is a D-Ala-D-Ala-ligase that ligates D-Ala and D-Lac to D-Ala-D-Lac-depsipeptide.[97] VanA is a modified form of the traditional D-Ala-D-Ala ligase, with a preference for D-Lac. VanX is a highly selective D,D-carboxypeptidase that eliminates the remaining D-Ala-D-Ala-dipeptide but not hydrolyze D-Ala-D-Lac.[100] The accessory D,D-carboxypeptidase VanY cleaves the terminal D-Ala from the already prepared pentapeptide, leaving a tetrapeptide that is not recognized by GPAs, and thereby increases the resistance level.[93, 101] VanZ – the second possible accessory protein within the *van* cluster – generates low teicoplanin resistance level in absence of the other resistance genes by an unknown mechanism.[102] Therefore, VanY and VanZ are not required resistance proteins – but contribute to high resistance level to vancomycin and teicoplanin, respectively.[103]

The difference of vancomycin and teicoplanin activity towards VanB enterococci is probably due to the difference in their sensor kinase VanS, which is not activated by teicoplanin. Thanks to the lipid chain, teicoplanin-like GPAs, unlike vancomycin-like, anchor the bacterial cell membrane. These LGPAs with an acyl chain do not induce resistance in VanB enterococci. In fact, the addition of a C_{10} acyl chain to vancomycin abolishes its ability to activate the resistance genes in these bacteria.[81] However, the mechanism by which LGPAs avoid resistance is not clear. One hypothesis is that membrane localization of LGPAs makes them inaccessible to the sensor kinase, so the induction necessary to start the resistance process does not occur. Another possibility is that the sensor kinase interacts not with the GPA itself but with PG intermediates or degradation products produced by the metabolic blockade.

GPA resistance in enterococci could also result from the substitution of the terminal D-Ala-D-Ala to D-alanyl-D-serine (D-Ala-D-Ser), but in this case the level of resistance is generally lower than in VanA and VanB phenotype. The best study among the D-Ala-D-Ser operons is the *vanC*. VanC enterococci are resistant to low levels of vancomycin, but sensitive to teicoplanin. VanC-like mechanism was identified in the GPA-resistant *Enterococcus gallinarum*, *E. flavescens and E. casseliflavus*. Some of the proteins that take part in this second resistance route are different from those involved in the previous mechanism which conferred high-level resistance, pointing out a second mechanism by which resistance has emerged. Here, three proteins are involved: a racemase (VanT) that converts L-Ser to D-Ser, a ligase (VanC) that join D-Ala and D-Ser to the final D-Ala-D-

Lac-depsipeptide and a bi-functional D,D-dipeptidase/D,D-carboxypeptidase ($VanXY_C$) that cleaves the D-Ala-D-Ala C-terminal from the mature PG precursors.[104]

Besides the phenotypes describes above, *van*-like gene operons were also found in other Gram-positive pathogens such as streptococci,[105] *Listeria* spp.[106] or *C. difficile*[107] and in some nonpathogenic Gram-positive environmental bacteria such as *Oerskovia turbata*, *Corynebacterium haemolyticum*,[108] *Bacillus circulans*,[109] and *Streptomyces coelicolor*.[110]

4.- A40926: a natural precursor of the clinically valuable second generation semi-synthetic dalbavancin

The World Health Organization (WHO) published in 2017 the list of pathogen bacteria, whose level of resistance to antibiotics is such that they represent an important challenge to medicine and human health. Both *E. faecium* and *S. aureus* are in this list and classified as high priority pathogens for research and development of new drugs due to vancomycin resistance.[111] Novel GPAs are urgently needed for a more efficient action against Gram-positive pathogens, and this medical need led to the discovery and development of the novel semi-synthetic GPAs dalbavancin, oritavancin and telavancin. These second-generation GPAs are now prescribed instead of vancomycin to treat severe infections. Among them, dalbavancin is especially efficient due to its potency and once-a-week dose. As it was mentioned before, dalbavancin is a derivative of naturally occurring GPA A40926.

Structure of A40926. A40926 is a type IV GPA, structurally almost identical to teicoplanin. Unlike teicoplanin, A40926 lacks Glc*N*Ac residue at AA6 of the aglycone. Another major difference is a presence of *N*-acylaminoglucuronic group at AA4, where teicoplanin has Glc*N*-Acyl moiety. Smaller structural differences include methylation of N-terminal end of A40926 aglycone and differences in chlorination pattern. The length and branching pattern of aliphatic side chains at Glc*N*-Acyl moiety also differs between teicoplanin and A40926, but this does not stem to the biosynthetic machinery being rather dependent on the structure of cell membrane (where BGC-encoded acetyltransferases take substrates).[112] These modifications improve antimicrobial activity and pharmacokinetic properties of A40926, compared to teicoplanin, maintaining an excellent safety profile.[61] Its molecular characteristics confer more effective *in vitro* and *in vivo* activity than teicoplanin. As mentioned above, the difference between A40926 and dalbavancin is the amidation of the peptide carboxyl group with a dimethylaminopropylamide side chain (Figure 10).[61] This modification extends half-life of dalbavancin in blood (over 300 h in humans), which concedes a unique once-a-week dose by intravenous injection.[113]

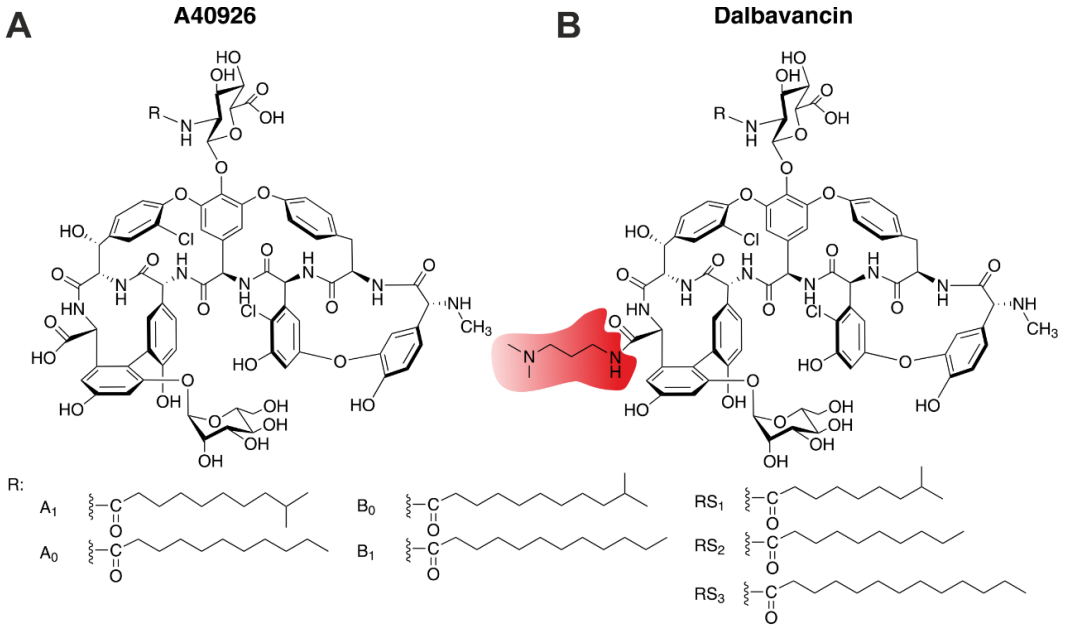


Figure 10. Chemical structures of A40926 complex (A) and dalbavancin (B). The chemical modification present in dalbavancin is highlighted with red colour.

Genes behind A40926 biosynthetic pathway: biosynthesis, regulation and resistance. BGC responsible for the biosynthesis of A40926 was characterized in 2003, almost 13 years before the sequencing of the full *N. gerenzanensis* ATCC 39727 genome and it was called *dbv*.[77] The name *dbv* was derived from ***d***al***b***a***v***ancin. Like other GPA BGCs, *dbv* consisted of structural, regulatory, export and self-resistance genes. Overall, *dbv* contains 37 genes organized in 12 transcriptional units (Figure 1A). 27 out of the 37 *dbv* genes find homologs in at least one other GPA cluster.[77]

Since biosynthesis of all GPAs requires non-proteinogenic amino acids, *dbv* pathway is not an exception. *dbv1*, *dbv2*, *dbv5*, *dbv30-34* and *dbv37* code for enzymes involved in the biosynthesis of non-proteinogenic amino acids Hpg and Dpg. Unlike vancomycin-like BGCs, *dbv* cluster has a single gene (*dbv28*) for Bht synthesis instead of three. With a sufficient pool of non-proteinogenic amino acids, non-ribosomal synthesis of A40926 aglycone begins. *dbv* NRPS is composed of 7 modules and sequentially joins Hpg, Tyr, Dpg, Hpg, Hpg, Tyr and Dpg (Figure 11B). NRPS is coded with four genes, particularly *dbv16-17* and *dbv25-26*. The organization of *dbv* NRPS genes is unprecedented in other GPA BGCs, since they are not collinear to the NRPS assemble line and are coded on the opposite DNA strands. Besides *dbv16-17* and *dbv25-26,* two other *dbv* ORFs are likely to participate in heptapeptide synthesis: *dbv15*, encoding for a short, highly conserved MbtH-like protein, working as a NRPS assembly chaperone; and *dbv36*, coding for a type II thioesterase, which might be involved in hydrolysis of misprimed or misacylated T domains, thus proofreading the NRPS assembly line.[77,114] After the synthesis of the linear heptapeptide, the crosslinking of the aglycone occurs, involving four P450 monooxygenases encoded by *dbv11-14*.[77]

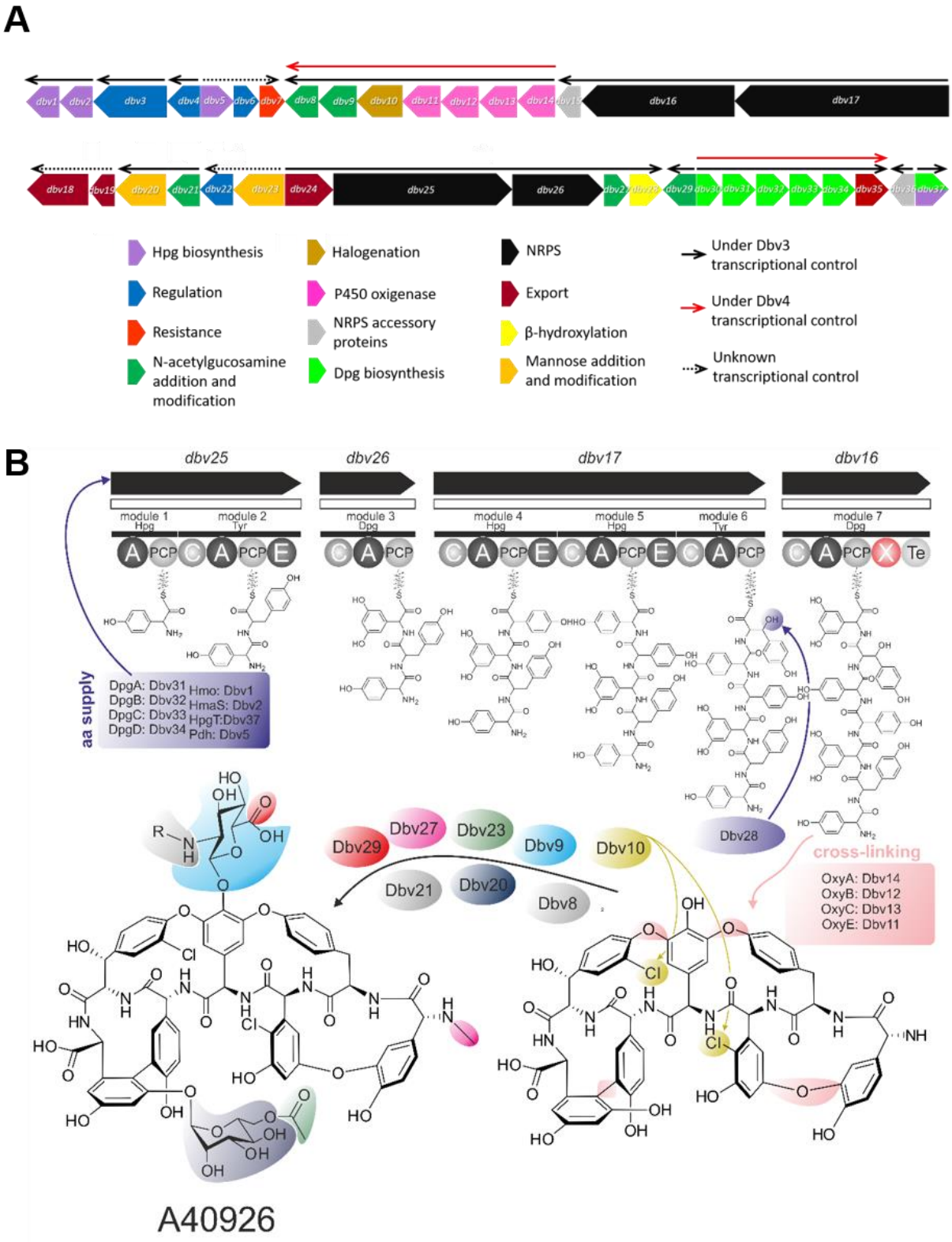


Figure 11. Genetic organization of the *dbv* cluster (A) and A40926 biosynthesis scheme (B). A) Thin arrows indicate experimentally established operons. Black arrows represent operons controlled by Dbv3, red arrows indicate operons controlled by Dbv4 and dashed arrows indicate operons whose control is unknown. For more details, please refer to the main text.

Other *dbv* genes are involved in the further modifications of A40926 aglycone (Figure 11B). As most GPAs, A40926 is decorated with sugar residues. Two glycosyltranferases – Dbv9 and Dbv20 – are involved in this, attaching Glc*N*Ac and mannose, respectively. Then, Glc*N*Ac is oxidized by Dbv29, deacetylated by Dbv21, and acylated by Dbv8. Finally, mannose moiety is acetylated by Dbv23, giving *O*-acetyl-A40926. Thus, products of *dbv8*, *dbv9*, *dbv20*, *dbv21* and *dbv29* are involved in the attachment and modification of sugars.[114,115,116,117] Although A40926 shows two chlorination sites in the structure, *dbv* cluster possesses a single halogenase, encoded by *dbv10*.[114] A single methyltransferase, encoded with *dbv27,* is required for *N*-methylation of the N-terminal Hpg residue.[114]

The *dbv* cluster also contains four genes with export functions. These are *dbv18*, *dbv19*, and *dbv24*, coding for ABC transporters and *dbv35,* coding for a Na^+/H^+-antiporter. Functions of Na^+/H^+-antiporter are so far not clear, although the corresponding gene is well conserved throughout other GPA BGCs. At the same time, ABC transporters are most likely responsible for the export of A40926. Here once again *dbv* stands aside other GPA BGCs, which usually code only for a single ABC-transporters. Thus, it is not clear whether all these ABC-transporters are responsible for the export of A40926, or maybe only some of them. These ABC transporters might also contribute to glycopeptide resistance through active export of the antibiotic from the cell.

GPA BGCs typically possess *vanHAX* resistance genes, which remodel cell wall making it GPA resistant, to avoid the producer suicide during antibiotic production. Instead, *dbv* is an exception and lacks this cassette. The only *dbv* resistant determinant is the *dbv7* gene, which encodes a D,D-carboxypeptidase belonging to the VanY family.[93]

GPA production is usually controlled through the transcriptional regulation of biosynthetic gene expression. This is achieved with cluster-situated regulators (CSRs). Same transcriptional factors are also often called pathway-specific regulators.[118] The *dbv* cluster contains two regulatory genes: *dbv3* (coding for a LuxR-like transcriptional regulator) and *dbv4* (coding for a StrR-like transcriptional regulator). *dbv* also codes for a putative two-component regulatory response system, *dbv6* (coding for a response regulator) and *dbv22* (coding for a sensory histidine kinase).[119,120,121] The roles of these pathway-specific regulators were experimentally investigated.[120] Dbv4 appeared to be crucial for A40926 production, since the knock-out of the corresponding gene led to the complete abolishment of antibiotic biosynthesis.[120] Interestingly, *dbv4* orthologues from *Am. balhimycina* (the producer of balhimycin)[122] and *A. teichomyceticus* (the producer of teicoplanin)[123] – StrR-like regulators Bbr and Tei15*, respectively – also were found to be crucial for the corresponding pathways. However, both Bbr and Tei15* were found to control the expression of the majority, if not the all, balhimycin and teicoplanin biosynthetic genes. Dbv4 is different, since its regulon is rather small and consists of the genes responsible for the non-proteinogenic amino acid biosynthesis – *dbv14-dbv8* and *dbv30-dbv35* operons. The second *dbv* CSR – Dbv3 – belongs to a LuxR-family of transcriptional factors. Dbv3 is reported to governs the expression of almost all *dbv* transcriptional units, including Dbv4 – at least other six operons (*dbv2-dbv1*, *dbv14-dbv8*, *dbv17-dbv15*, *dbv21-dbv20*, *dbv24-dbv28*, and *dbv30- dbv35*) and four single genes (*dbv4*, *dbv29*, *dbv36*, and *dbv37*).[120] Balhimycin BGC lacks the LuxR-like pathway-

specific regulator, while teicoplanin BGC has one – Tei16*. The latter is also a key regulator of teicoplanin biosynthesis, however both Tei16* and Dbv3 are absolutely unrelated from evolutionary point of view. Although *dbv6* and *dbv22* until very recently were believed to be *vanRS* analogues, it was shown that they are involved in the fine tuning of *dbv*-gene expression, possibly interlinking export and resistance. These genes have a negative effect on *dbv* biosynthesis, while having the resistance gene *dbv7* under positive control.[119,120,121]

Ways of improving A40926 production in *N. gerenzanensis.* The use of second-generation GPAs reactivated the debate about therapy cost. Dalbavancin has demonstrated an efficacy and safety comparable to vancomycin, but the actual problem is the high cost compared with traditional antibiotics (Table 1).[124] Therefore, numerous studies have been carried out to develop genetic modification of *N. gerenzanensis* in order to improve the production of A40926.

Drug	1 vial (mg)	Unit price ($US)	Usual adult dose	Total price ($US)
Vancomycin	1000	9.5	15-20 mg/kg (max 2 g) x2 x 7-14 days	133 - 266
Telavancin (Vibativ®)	750	309.5	10 mg/kg x 7-14 days	2167 - 4333
Oritavancin (Orbactiv®)	400	1160	1200 mg once	2900
Dalbavancin (Dalvance®)	500	1490	1000 mg x1, then 500 mg 1 week later	4470

Table 1. GPA acquisition cost associated with usual adult treatment (modified from [36])

Generally, once the antibiotic had been discovered, the producing strains undergo a process of genetic engineering and fermentation conditions optimization to obtain the highest production for industry. GPAs and A40926 in particular, like all other antibiotics, have also undergone this process. However, most of this work has not published in the scientific literature, remaining part of the intellectual property rights of producing pharma companies. The need to improve the fermentation process (and reduce the cost of a multi-step process) is particularly demanding for second-generation GPAs that undergo semi-synthetic modification of natural scaffolds. In addition, fermentation remains the only approach to obtain A40926 (and other GPAs) on the industrial scale. Therefore, it is particularly important and necessary to improve the yield of production. A low yield of the *N. gerenzanensis* wild-type producing strain is a serious limitation for developing and commercializing this drug and their derivatives.

Actually, industrial fermentation media to produce A40926 and other GPAs consists are rich, containing dextrose, malt extract and yeast extract.[125] It was reported that a low initial concentrations of phosphate and ammonium result in increased productivity of A40926[126] and calcium has a repressive effect.[127] Amino acids are important precursors for the production of many secondary metabolites, including GPAs.[128] L-asparagine and L-glutamine resulted in a dramatic stimulation of production as nitrogen sources instead of ammonium salts, although these amino acids are not the direct precursors of the heptapeptide scaffold.[127] Addition of L-valine also increased A40926 production, being in the case the precursor of the fatty acid tail incorporated in the most produced component of the complex.[125] All these improvements were successfully applied for the wild type producer. A recent report describes a UV-generated *N.*

gerenzanensis mutant to produce high levels of A40926 in an optimized production medium. In this strain, A40926 production was markedly promoted by using poorly assimilated carbon source maltodextrin and nitrogen source soybean meal. Furthermore, L-leucine and Cu^{2+} stimulated biosynthesis while Co^{2+} showed an inhibitory effect.[129] Additionally, rational strain improvement was achieved by deleting and overexpressing certain *dbv* genes. Thus, A40926-overproducing recombinant strain was achieved with the deletion of *dbv23*, which codes for the pathway-specific acetyltransferase, modifying mannose residue. Recombinant strain lacking *dbv23* produced approximately twice A40926 than the wild type.[130] Then, overexpression of *dbv3* regulator under the control of the thiostrepton-inducible *tipA* promoter also increased twice the production of A40926 in R3 liquid medium.[120] Overexpression of some tailoring genes, particularly *dbv9*, *dbv20*, *dbv29* and *dbv36* also increased the antibiotic production.[131] However, ways for the rational improvement of A40926 production are not exhausted yet and offer a lot of space for future maneuvers.

5.-References

(1) Ludwig, W.; Euzéby, J.; Schumann, P.; Busse, H.-J.; Trujillo, M. E.; Kämpfer, P.; Whitman, W. B. Road map of the phylum Actinobacteria. In *Bergey's manual® of systematic bacteriology*, Springer, **2012**; pp 1-28.

(2) Hopwood, D. A. *Streptomyces in nature and medicine: the antibiotic makers*; Oxford University Press, **2007**.

(3) Bhatti, A. A.; Haq, S.; Bhat, R. A. Actinomycetes benefaction role in soil and plant health. *Microb Pathog* **2017**, *111*, 458-467. DOI: 10.1016/j.micpath.2017.09.036.

(4) Sayed, A. M.; Hassan, M. H. A.; Alhadrami, H. A.; Hassan, H. M.; Goodfellow, M.; Rateb, M. E. Extreme environments: microbiology leading to specialized metabolites. *J Appl Microbiol* **2020**, *128* (3), 630-657. DOI: 10.1111/jam.14386.

(5) Barka, E. A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Meier-Kolthoff, J. P.; Klenk, H. P.; Clément, C.; Ouhdouch, Y.; van Wezel, G. P. Taxonomy, physiology, and natural products of actinobacteria. *Microbiol Mol Biol Rev* **2016**, *80* (1), 1-43. DOI: 10.1128/MMBR.00019-15.

(6) Bérdy, J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J Antibiot (Tokyo)* **2012**, *65* (8), 385-395. DOI: 10.1038/ja.2012.27.

(7) Aigle, B.; Lautru, S.; Spiteller, D.; Dickschat, J. S.; Challis, G. L.; Leblond, P.; Pernodet, J. L. Genome mining of *Streptomyces ambofaciens*. *J Ind Microbiol Biotechnol* **2014**, *41* (2), 251-263. DOI: 10.1007/s10295-013-1379-y.

(8) Bachmann, B. O.; Van Lanen, S. G.; Baltz, R. H. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J Ind Microbiol Biotechnol* **2014**, *41* (2), 175-184. DOI: 10.1007/s10295-013-1389-9.

(9) Challis, G. L. Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *J Ind Microbiol Biotechnol* **2014**, *41* (2), 219-232. DOI: 10.1007/s10295-013-1383-2.

(10) Ikeda, H.; Kazuo, S. Y.; Omura, S. Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of

biosynthetic gene clusters. *J Ind Microbiol Biotechnol* **2014**, *41* (2), 233-250. DOI: 10.1007/s10295-013-1327-x.
(11) Williams, S. Ecology of actinomycetes. *The biology of the actinomycetes* **1983**, 481-528.
(12) Ding, T.; Yang, L.-J.; Zhang, W.-D.; Shen, Y.-H. The secondary metabolites of rare actinomycetes: chemistry and bioactivity. *RSC advances* **2019**, *9* (38), 21964-21988.
(13) Lazzarini, A.; Cavaletti, L.; Toppo, G.; Marinelli, F. Rare genera of actinomycetes as potential producers of new antibiotics. *Antonie Van Leeuwenhoek* **2001**, *79* (3-4), 399-405.
(14) Sungthong, R.; Nakaew, N. The genus *Nonomuraea*: A review of a rare actinomycete taxon for novel metabolites. *J Basic Microbiol* **2015**, *55* (5), 554-565. DOI: 10.1002/jobm.201300691.
(15) Dalmastri, C.; Gastaldo, L.; Marcone, G. L.; Binda, E.; Congiu, T.; Marinelli, F. Classification of *Nonomuraea* sp. ATCC 39727, an actinomycete that produces the glycopeptide antibiotic A40926, as *Nonomuraea gerenzanensis* sp. nov. *Int J Syst Evol Microbiol* **2016**, *66* (2), 912-921. DOI: 10.1099/ijsem.0.000810.
(16) D'Argenio, V.; Petrillo, M.; Pasanisi, D.; Pagliarulo, C.; Colicchio, R.; Talà, A.; de Biase, M. S.; Zanfardino, M.; Scolamiero, E.; Pagliuca, C.; et al. The complete 12 Mb genome and transcriptome of *Nonomuraea gerenzanensis* with new insights into its duplicated "magic" RNA polymerase. *Sci Rep* **2016**, *6* (1), 18. DOI: 10.1038/s41598-016-0025-0.
(17) Capasso, L. Infectious diseases and eating habits at Herculaneum (1st century AD, southern Italy). *International Journal of Osteoarchaeology* **2007**, *17* (4), 350-357.
(18) Cook, M.; Molto, E.; Anderson, C. Fluorochrome labelling in Roman period skeletons from Dakhleh Oasis, Egypt. *Am J Phys Anthropol* **1989**, *80* (2), 137-143. DOI: 10.1002/ajpa.1330800202.
(19) Bassett, E. J.; Keith, M. S.; Armelagos, G. J.; Martin, D. L.; Villanueva, A. R. Tetracycline-labeled human bone from ancient Sudanese Nubia (A.D. 350). *Science* **1980**, *209* (4464), 1532-1534. DOI: 10.1126/science.7001623.
(20) Nelson, M. L.; Dinardo, A.; Hochberg, J.; Armelagos, G. J. Brief communication: mass spectroscopic characterization of tetracycline in the skeletal remains of an ancient population from Sudanese Nubia 350-550 CE. *Am J Phys Anthropol* **2010**, *143* (1), 151-154. DOI: 10.1002/ajpa.21340.
(21) Waterhouse, S. D. *Hesban 10: The necropolis of hesban: A typology of tombs*; Andrews University Press, **1998**.
(22) Cui, L.; Su, X. Z. Discovery, mechanisms of action and combination therapy of artemisinin. *Expert Rev Anti Infect Ther* **2009**, *7* (8), 999-1013. DOI: 10.1586/eri.09.68.
(23) Gelpi, A.; Gilbertson, A.; Tucker, J. D. Magic bullet: Paul Ehrlich, Salvarsan and the birth of venereology. *Sexually transmitted infections* **2015**, *91* (1), 68-69.
(24) Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *British journal of experimental pathology* **1929**, *10* (3), 226.
(25) Durand, G. A.; Raoult, D.; Dubourg, G. Antibiotic discovery: history, methods and perspectives. *Int J Antimicrob Agents* **2019**, *53* (4), 371-382. DOI: 10.1016/j.ijantimicag.2018.11.010.
(26) Waksman, S. A.; Schatz, A.; Reynolds, D. M. Production of antibiotic substances by actinomycetes. *Annals of the New York Academy of Sciences* **2010**, *1213* (1), 112-124.
(27) Smith, P. W.; Watkins, K.; Hewlett, A. Infection control through the ages. *Am J Infect Control* **2012**, *40* (1), 35-42. DOI: 10.1016/j.ajic.2011.02.019.

(28) Gottfried, J. *History Repeating? Avoiding a Return to the Pre-Antibiotic Age*. **2005.** https://dash.harvard.edu/bitstream/handle/1/8889467/Gottfried05.html?sequence=2
(29) Holmes, A. H.; Moore, L. S.; Sundsfjord, A.; Steinbakk, M.; Regmi, S.; Karkey, A.; Guerin, P. J.; Piddock, L. J. Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* **2016**, *387* (10014), 176-187. DOI: 10.1016/S0140-6736(15)00473-0.
(30) Abraham, E. P.; Chain, E. An enzyme from bacteria able to destroy penicillin. 1940. *Rev Infect Dis* **1988**, *10* (4), 677-678.
(31) D'Costa, V. M.; King, C. E.; Kalan, L.; Morar, M.; Sung, W. W.; Schwarz, C.; Froese, D.; Zazula, G.; Calmels, F.; Debruyne, R.; et al. Antibiotic resistance is ancient. *Nature* **2011**, *477* (7365), 457-461. DOI: 10.1038/nature10388.
(32) Rather, I. A.; Kim, B. C.; Bajpai, V. K.; Park, Y. H. Self-medication and antibiotic resistance: Crisis, current challenges, and prevention. *Saudi J Biol Sci* **2017**, *24* (4), 808-812. DOI: 10.1016/j.sjbs.2017.01.004.
(33) Ribeiro da Cunha, B.; Fonseca, L. P.; Calado, C. R. C. Antibiotic discovery: Where have we come from, where do we go? *Antibiotics (Basel)* **2019**, *8* (2). DOI: 10.3390/antibiotics8020045.
(34) Plackett, B. Why big pharma has abandoned antibiotics. **2020**; 586, S50-S52.
(35) O'Neil, J. Tackling a crisis for the health and wealth of nations. **2014**.
(36) Van Bambeke, F. Lipoglycopeptide antibacterial agents in Gram-positive infections: A comparative review. *Drugs* **2015**, *75* (18), 2073-2095. DOI: 10.1007/s40265-015-0505-8.
(37) Rossolini, G. M.; Arena, F.; Pollini, S. Novel infectious diseases and emerging Gram-positive multi-resistant pathogens in hospital and community acquired infections. In *Antimicrobials*, Springer, **2014**; pp 11-28.
(38) Binda, E.; Marinelli, F.; Marcone, G. L. Old and new glycopeptide antibiotics: Action and resistance. *Antibiotics (Basel)* **2014**, *3* (4), 572-594. DOI: 10.3390/antibiotics3040572.
(39) Nicolaou, K. C.; Boddy, C. N.; Bräse, S.; Winssinger, N. Chemistry, Biology, and medicine of the glycopeptide atibiotics. *Angew Chem Int Ed Engl* **1999**, *38* (15), 2096-2152. DOI: 10.1002/(sici)1521-3773(19990802)38:15<2096::aid-anie2096>3.0.co;2-f.
(40) Reynolds, P. E. Structure, biochemistry and mechanism of action of glycopeptide antibiotics. *Eur J Clin Microbiol Infect Dis* **1989**, *8* (11), 943-950. DOI: 10.1007/bf01967563.
(41) Levine, D. P. Vancomycin: a history. *Clinical infectious diseases* **2006**, *42* (Supplement_1), S5-S12.
(42) Philip, J.; Schenck, J.; Hargie, M. Ristocetins A and B, two new antibiotics; isolation and properties. *Antibiotics Annual* **1956**, 699-705.
(43) Jenkins, C.; Meyer, D.; Dreyfus, M.; Larreu, M. J. Willebrand factor and ristocetin I. Mechanism of ristocetin-induced platelet aggregation. *British journal of haematology* **1974**, *28* (4), 561-578.
(44) Sarji, K. E.; Stratton, R. D.; Wagner, R. H.; Brinkhous, K. M. Nature of von Willebrand factor: a new assay and a specific inhibitor. *Proc Natl Acad Sci U S A* **1974**, *71* (8), 2937-2941. DOI: 10.1073/pnas.71.8.2937.
(45) Parenti, F.; Beretta, G.; Berti, M.; Arioli, V. Teichomycins, new antibiotics from *Actinoplanes teichomyceticus* Nov. Sp. I. Description of the producer strain, fermentation studies and biological properties. *J Antibiot (Tokyo)* **1978**, *31* (4), 276-283. DOI: 10.7164/antibiotics.31.276.
(46) Cynamon, M. H.; Granato, P. A. Comparison of the *in vitro* activities of teichomycin A2 and vancomycin against staphylococci and enterococci. *Antimicrob Agents Chemother* **1982**, *21* (3), 504-505. DOI: 10.1128/aac.21.3.504.

(47) Redin, G. LL-AV290, a new antibitic II. Antibacterial efficancy in mice and *in vitro*. *Antimicrob. Agents Chemother.* **1969**, 246-248.
(48) Hamill, R. L.; Stark, W. M.; DeLong, D. C. Novel antibiotic and a process for the production thereof. Google Patents: **1976**.
(49) Casewell, M.; Friis, C.; Marco, E.; McMullin, P.; Phillips, I. The European ban on growth-promoting antibiotics and emerging consequences for human and animal health. *J Antimicrob Chemother* **2003**, *52* (2), 159-161. DOI: 10.1093/jac/dkg313.
(50) Pelzer, S.; Süssmuth, R.; Heckmann, D.; Recktenwald, J.; Huber, P.; Jung, G.; Wohlleben, W. Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908. *Antimicrob Agents Chemother* **1999**, *43* (7), 1565-1573.
(51) van Wageningen, A. M.; Kirkpatrick, P. N.; Williams, D. H.; Harris, B. R.; Kershaw, J. K.; Lennard, N. J.; Jones, M.; Jones, S. J.; Solenberg, P. J. Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem Biol* **1998**, *5* (3), 155-162. DOI: 10.1016/s1074-5521(98)90060-6.
(52) Pootoolal, J.; Thomas, M. G.; Marshall, C. G.; Neu, J. M.; Hubbard, B. K.; Walsh, C. T.; Wright, G. D. Assembling the glycopeptide antibiotic scaffold: The biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL15009. *Proc Natl Acad Sci U S A* **2002**, *99* (13), 8962-8967. DOI: 10.1073/pnas.102285099.
(53) Chiu, H. T.; Hubbard, B. K.; Shah, A. N.; Eide, J.; Fredenburg, R. A.; Walsh, C. T.; Khosla, C. Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc Natl Acad Sci U S A* **2001**, *98* (15), 8548-8553. DOI: 10.1073/pnas.151246498.
(54) James, R. C.; Pierce, J. G.; Okano, A.; Xie, J.; Boger, D. L. Redesign of glycopeptide antibiotics: back to the future. *ACS Chem Biol* **2012**, *7* (5), 797-804. DOI: 10.1021/cb300007j.
(55) Leadbetter, M. R.; Adams, S. M.; Bazzini, B.; Fatheree, P. R.; Karr, D. E.; Krause, K. M.; Lam, B. M.; Linsell, M. S.; Nodwell, M. B.; Pace, J. L.; et al. Hydrophobic vancomycin derivatives with improved ADME properties: discovery of telavancin (TD-6424). *J Antibiot (Tokyo)* **2004**, *57* (5), 326-336. DOI: 10.7164/antibiotics.57.326.
(56) Higgins, D. L.; Chang, R.; Debabov, D. V.; Leung, J.; Wu, T.; Krause, K. M.; Sandvik, E.; Hubbard, J. M.; Kaniga, K.; Schmidt, D. E.; et al. Telavancin, a multifunctional lipoglycopeptide, disrupts both cell wall synthesis and cell membrane integrity in methicillin-resistant Staphylococcus aureus. *Antimicrob Agents Chemother* **2005**, *49* (3), 1127-1134. DOI: 10.1128/AAC.49.3.1127-1134.2005.
(57) Lunde, C. S.; Hartouni, S. R.; Janc, J. W.; Mammen, M.; Humphrey, P. P.; Benton, B. M. Telavancin disrupts the functional integrity of the bacterial membrane through targeted interaction with the cell wall precursor lipid II. *Antimicrob Agents Chemother* **2009**, *53* (8), 3375-3383. DOI: 10.1128/AAC.01710-08.
(58) Cooper, R. D.; Snyder, N. J.; Zweifel, M. J.; Staszak, M. A.; Wilkie, S. C.; Nicas, T. I.; Mullen, D. L.; Butler, T. F.; Rodriguez, M. J.; Huff, B. E.; et al. Reductive alkylation of glycopeptide antibiotics: synthesis and antibacterial activity. *J Antibiot (Tokyo)* **1996**, *49* (6), 575-581. DOI: 10.7164/antibiotics.49.575.
(59) Mitra, S.; Saeed, U.; Havlichek, D. H.; Stein, G. E. Profile of oritavancin and its potential in the treatment of acute bacterial skin structure infections. *Infect Drug Resist* **2015**, *8*, 189-197. DOI: 10.2147/IDR.S69412.
(60) Goldstein, B. P.; Selva, E.; Gastaldo, L.; Berti, M.; Pallanza, R.; Ripamonti, F.; Ferrari, P.; Denaro, M.; Arioli, V.; Cassani, G. A40926, a new glycopeptide antibiotic with anti-Neisseria activity. *Antimicrob Agents Chemother* **1987**, *31* (12), 1961-1966. DOI: 10.1128/aac.31.12.1961.

(61) Malabarba, A.; Goldstein, B. P. Origin, structure, and activity *in vitro* and *in vivo* of dalbavancin. *J Antimicrob Chemother* **2005**, *55 Suppl 2*, ii15-20. DOI: 10.1093/jac/dki005.
(62) Harris, C. M.; Harris, T. M. Structure of the glycopeptide antibiotic vancomycin. Evidence for an asparagine residue in the peptide. *Journal of the American Chemical Society* **1982**, *104* (15), 4293-4295.
(63) Loll, P. J.; Miller, R.; Weeks, C. M.; Axelsen, P. H. A ligand-mediated dimerization mode for vancomycin. *Chemistry & biology* **1998**, *5* (5), 293-298.
(64) Culp, E. J.; Waglechner, N.; Wang, W.; Fiebig-Comyn, A. A.; Hsu, Y. P.; Koteva, K.; Sychantha, D.; Coombes, B. K.; Van Nieuwenhze, M. S.; Brun, Y. V.; et al. Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature* **2020**, *578* (7796), 582-587. DOI: 10.1038/s41586-020-1990-9.
(65) Xu, M.; Wang, W.; Waglechner, N.; Culp, E. J.; Guitor, A. K.; Wright, G. D. GPAHex-A synthetic biology platform for Type IV-V glycopeptide antibiotic production and discovery. *Nat Commun* **2020**, *11* (1), 5232. DOI: 10.1038/s41467-020-19138-5.
(66) Sosio, M.; Kloosterman, H.; Bianchi, A.; de Vreugd, P.; Dijkhuizen, L.; Donadio, S. Organization of the teicoplanin gene cluster in *Actinoplanes teichomyceticus*. *Microbiology* **2004**, *150* (1), 95-102.
(67) Li, T.-L.; Huang, F.; Haydock, S. F.; Mironenko, T.; Leadlay, P. F.; Spencer, J. B. Biosynthetic gene cluster of the glycopeptide antibiotic teicoplanin: characterization of two glycosyltransferases and the key acyltransferase. *Chemistry & biology* **2004**, *11* (1), 107-119.
(68) Marcone, G. L.; Marinelli, F. Glycopeptides: an old but up-to-date successful antibiotic class. In *Antimicrobials*, Springer, 2014; pp 85-107.
(69) Hubbard, B. K.; Thomas, M. G.; Walsh, C. T. Biosynthesis of L-p-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics. *Chem Biol* **2000**, *7* (12), 931-942. DOI: 10.1016/s1074-5521(00)00043-0.
(70) Li, T. L.; Choroba, O. W.; Charles, E. H.; Sandercock, A. M.; Williams, D. H.; Spencer, J. B. Characterisation of a hydroxymandelate oxidase involved in the biosynthesis of two unusual amino acids occurring in the vancomycin group of antibiotics. *Chem Commun (Camb)* **2001**, (18), 1752-1753. DOI: 10.1039/b103548g.
(71) Chen, H.; Tseng, C. C.; Hubbard, B. K.; Walsh, C. T. Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proc Natl Acad Sci U S A* **2001**, *98* (26), 14901-14906. DOI: 10.1073/pnas.221582098.
(72) Pfeifer, V.; Nicholson, G. J.; Ries, J.; Recktenwald, J.; Schefer, A. B.; Shawky, R. M.; Schröder, J.; Wohlleben, W.; Pelzer, S. A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (S)-3, 5-dihydroxyphenylglycine. *Journal of Biological Chemistry* **2001**, *276* (42), 38370-38377.
(73) Yim, G.; Thaker, M. N.; Koteva, K.; Wright, G. Glycopeptide antibiotic biosynthesis. *J Antibiot (Tokyo)* **2014**, *67* (1), 31-41. DOI: 10.1038/ja.2013.117.
(74) Mulyani, S.; Egel, E.; Kittel, C.; Turkanovic, S.; Wohlleben, W.; Süssmuth, R. D.; van Pée, K. H. The thioesterase Bhp is involved in the formation of beta-hydroxytyrosine during balhimycin biosynthesis in *Amycolatopsis balhimycina*. *Chembiochem* **2010**, *11* (2), 266-271. DOI: 10.1002/cbic.200900600.
(75) Stinchi, S.; Carrano, L.; Lazzarini, A.; Feroggio, M.; Grigoletto, A.; Sosio, M.; Donadio, S. A derivative of the glycopeptide A40926 produced by inactivation of the beta-hydroxylase gene in *Nonomuraea* sp. ATCC39727. *FEMS Microbiol Lett* **2006**, *256* (2), 229-235. DOI: 10.1111/j.1574-6968.2006.00120.x.

(76) Haslinger, K.; Peschke, M.; Brieke, C.; Maximowitsch, E.; Cryle, M. J. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature* **2015**, *521* (7550), 105-109. DOI: 10.1038/nature14141.
(77) Sosio, M.; Stinchi, S.; Beltrametti, F.; Lazzarini, A.; Donadio, S. The gene cluster for the biosynthesis of the glycopeptide antibiotic A40926 by nonomuraea species. *Chem Biol* **2003**, *10* (6), 541-549. DOI: 10.1016/s1074-5521(03)00120-0.
(78) Stegmann, E.; Pelzer, S.; Bischoff, D.; Puk, O.; Stockert, S.; Butz, D.; Zerbe, K.; Robinson, J.; Süssmuth, R. D.; Wohlleben, W. Genetic analysis of the balhimycin (vancomycin-type) oxygenase genes. *J Biotechnol* **2006**, *124* (4), 640-653. DOI: 10.1016/j.jbiotec.2006.04.009.
(79) Hadatsch, B.; Butz, D.; Schmiederer, T.; Steudle, J.; Wohlleben, W.; Süssmuth, R.; Stegmann, E. The biosynthesis of teicoplanin-type glycopeptide antibiotics: assignment of p450 mono-oxygenases to side chain cyclizations of glycopeptide a47934. *Chem Biol* **2007**, *14* (9), 1078-1089. DOI: 10.1016/j.chembiol.2007.08.014.
(80) Peschke, M.; Brieke, C.; Cryle, M. J. F-O-G ring formation in glycopeptide antibiotic biosynthesis is catalysed by OxyE. *Sci Rep* **2016**, *6*, 35584. DOI: 10.1038/srep35584.
(81) Kahne, D.; Leimkuhler, C.; Lu, W.; Walsh, C. Glycopeptide and lipoglycopeptide antibiotics. *Chem Rev* **2005**, *105* (2), 425-448. DOI: 10.1021/cr030103a.
(82) Kalan, L.; Perry, J.; Koteva, K.; Thaker, M.; Wright, G. Glycopeptide sulfation evades resistance. *J Bacteriol* **2013**, *195* (1), 167-171. DOI: 10.1128/JB.01617-12.
(83) Al Toma, R. S.; Brieke, C.; Cryle, M. J.; Süssmuth, R. D. Structural aspects of phenylglycines, their biosynthesis and occurrence in peptide natural products. *Nat Prod Rep* **2015**, *32* (8), 1207-1235. DOI: 10.1039/c5np00025d.
(84) Schmartz, P. C.; Wölfel, K.; Zerbe, K.; Gad, E.; El Tamany, e. S.; Ibrahim, H. K.; Abou-Hadeed, K.; Robinson, J. A. Substituent effects on the phenol coupling reaction catalyzed by the vancomycin biosynthetic P450 enzyme OxyB. *Angew Chem Int Ed Engl* **2012**, *51* (46), 11468-11472. DOI: 10.1002/anie.201204458.
(85) Williams, D. H.; Bardsley, B. The Vancomycin group of antibiotics and the fight against resistant bacteria. *Angew Chem Int Ed Engl* **1999**, *38* (9), 1172-1193. DOI: 10.1002/(SICI)1521-3773(19990503)38:9<1172::AID-ANIE1172>3.0.CO;2-C.
(86) Lamb, S. S.; Patel, T.; Koteva, K. P.; Wright, G. D. Biosynthesis of sulfated glycopeptide antibiotics by using the sulfotransferase StaL. *Chem Biol* **2006**, *13* (2), 171-181. DOI: 10.1016/j.chembiol.2005.12.003.
(87) Yim, G.; Kalan, L.; Koteva, K.; Thaker, M. N.; Waglechner, N.; Tang, I.; Wright, G. D. Harnessing the synthetic capabilities of glycopeptide antibiotic tailoring enzymes: characterization of the UK-68,597 biosynthetic cluster. *Chembiochem* **2014**, *15* (17), 2613-2623. DOI: 10.1002/cbic.201402179.
(88) Bick, M. J.; Banik, J. J.; Darst, S. A.; Brady, S. F. The 2.7 Å resolution structure of the glycopeptide sulfotransferase Teg14. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 12), 1278-1286. DOI: 10.1107/S0907444910036681.
(89) Bischoff, D.; Bister, B.; Bertazzo, M.; Pfeifer, V.; Stegmann, E.; Nicholson, G. J.; Keller, S.; Pelzer, S.; Wohlleben, W.; Süssmuth, R. D. The biosynthesis of vancomycin-type glycopeptide antibiotics--a model for oxidative side-chain cross-linking by oxygenases coupled to the action of peptide synthetases. *Chembiochem* **2005**, *6* (2), 267-272. DOI: 10.1002/cbic.200400328.
(90) Marcone, G. L.; Binda, E.; Berini, F.; Marinelli, F. Old and new glycopeptide antibiotics: From product to gene and back in the post-genomic era. *Biotechnol Adv* **2018**, *36* (2), 534-554. DOI: 10.1016/j.biotechadv.2018.02.009.
(91) Rohde, M. The Gram-positive bacterial cell wall. *Microbiol Spectr* **2019**, *7* (3). DOI: 10.1128/microbiolspec.GPP3-0044-2018.

(92) Cundliffe, E.; Demain, A. L. Avoidance of suicide in antibiotic-producing microbes. *J Ind Microbiol Biotechnol* **2010**, *37* (7), 643-672. DOI: 10.1007/s10295-010-0721-x.
(93) Marcone, G. L.; Beltrametti, F.; Binda, E.; Carrano, L.; Foulston, L.; Hesketh, A.; Bibb, M.; Marinelli, F. Novel mechanism of glycopeptide resistance in the A40926 producer *Nonomuraea* sp. ATCC 39727. *Antimicrob Agents Chemother* **2010**, *54* (6), 2465-2472. DOI: 10.1128/AAC.00106-10.
(94) Frasch, H. J.; Kalan, L.; Kilian, R.; Martin, T.; Wright, G. D.; Stegmann, E. Alternative pathway to a glycopeptide-resistant cell wall in the balhimycin Pproducer *Amycolatopsis balhimycina*. *ACS Infect Dis* **2015**, *1* (6), 243-252. DOI: 10.1021/acsinfecdis.5b00011.
(95) Marshall, C. G.; Broadhead, G.; Leskiw, B. K.; Wright, G. D. D-Ala-D-Ala ligases from glycopeptide antibiotic-producing organisms are highly homologous to the enterococcal vancomycin-resistance ligases VanA and VanB. *Proc Natl Acad Sci U S A* **1997**, *94* (12), 6480-6483. DOI: 10.1073/pnas.94.12.6480.
(96) Marshall, C. G.; Lessard, I. A.; Park, I.; Wright, G. D. Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. *Antimicrob Agents Chemother* **1998**, *42* (9), 2215-2220.
(97) Bugg, T. D.; Wright, G. D.; Dutka-Malen, S.; Arthur, M.; Courvalin, P.; Walsh, C. T. Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. *Biochemistry* **1991**, *30* (43), 10408-10415. DOI: 10.1021/bi00107a007.
(98) Périchon, B.; Courvalin, P. VanA-type vancomycin-resistant *Staphylococcus aureus.* *Antimicrob Agents Chemother* **2009**, *53* (11), 4580-4587. DOI: 10.1128/AAC.00346-09.
(99) Arthur, M.; Molinas, C.; Bugg, T. D.; Wright, G. D.; Walsh, C. T.; Courvalin, P. Evidence for *in vivo* incorporation of D-lactate into peptidoglycan precursors of vancomycin-resistant enterococci. *Antimicrob Agents Chemother* **1992**, *36* (4), 867-869. DOI: 10.1128/aac.36.4.867.
(100) Wu, Z.; Wright, G. D.; Walsh, C. T. Overexpression, purification, and characterization of VanX, a D-, D-dipeptidase which is essential for vancomycin resistance in *Enterococcus faecium* BM4147. *Biochemistry* **1995**, *34* (8), 2455-2463. DOI: 10.1021/bi00008a008.
(101) Wright, G. D.; Molinas, C.; Arthur, M.; Courvalin, P.; Walsh, C. T. Characterization of vanY, a DD-carboxypeptidase from vancomycin-resistant *Enterococcus faecium* BM4147. *Antimicrob Agents Chemother* **1992**, *36* (7), 1514-1518. DOI: 10.1128/aac.36.7.1514.
(102) Arthur, M.; Depardieu, F.; Molinas, C.; Reynolds, P.; Courvalin, P. The vanZ gene of Tn1546 from *Enterococcus faecium* BM4147 confers resistance to teicoplanin. *Gene* **1995**, *154* (1), 87-92. DOI: 10.1016/0378-1119(94)00851-i.
(103) Arthur, M.; Reynolds, P. E.; Depardieu, F.; Evers, S.; Dutka-Malen, S.; Quintiliani, R.; Courvalin, P. Mechanisms of glycopeptide resistance in enterococci. *J Infect* **1996**, *32* (1), 11-16. DOI: 10.1016/s0163-4453(96)80003-x.
(104) Reynolds, P. E.; Courvalin, P. Vancomycin resistance in enterococci due to synthesis of precursors terminating in D-alanyl-D-serine. *Antimicrob Agents Chemother* **2005**, *49* (1), 21-25. DOI: 10.1128/AAC.49.1.21-25.2005.
(105) Poyart, C.; Pierre, C.; Quesne, G.; Pron, B.; Berche, P.; Trieu-Cuot, P. Emergence of vancomycin resistance in the genus *Streptococcus*: characterization of a vanB transferable determinant in *Streptococcus bovis*. *Antimicrob Agents Chemother* **1997**, *41* (1), 24-29. DOI: 10.1128/AAC.41.1.24.

(106) Biavasco, F.; Giovanetti, E.; Miele, A.; Vignaroli, C.; Facinelli, B.; Varaldo, P. E. *In vitro* conjugative transfer of VanA vancomycin resistance between *Enterococci* and *Listeriae* of different species. *Eur J Clin Microbiol Infect Dis* **1996**, *15* (1), 50-59. DOI: 10.1007/BF01586185.
(107) Peltier, J.; Courtin, P.; El Meouche, I.; Catel-Ferreira, M.; Chapot-Chartier, M. P.; Lemée, L.; Pons, J. L. Genomic and expression analysis of the vanG-like gene cluster of *Clostridium difficile*. *Microbiology (Reading)* **2013**, *159* (Pt 7), 1510-1520. DOI: 10.1099/mic.0.065060-0.
(108) Power, E. G.; Abdulla, Y. H.; Talsania, H. G.; Spice, W.; Aathithan, S.; French, G. L. VanA genes in vancomycin-resistant clinical isolates of *Oerskovia turbata* and *Arcanobacterium* (*Corynebacterium*) *haemolyticum*. *J Antimicrob Chemother* **1995**, *36* (4), 595-606. DOI: 10.1093/jac/36.4.595.
(109) Fontana, R.; Ligozzi, M.; Pedrotti, C.; Padovani, E. M.; Cornaglia, G. Vancomycin-resistant *Bacillus circulans* carrying the vanA gene responsible for vancomycin resistance in enterococci. *Eur J Clin Microbiol Infect Dis* **1997**, *16* (6), 473-474. DOI: 10.1007/BF02471915.
(110) Hong, H. J.; Hutchings, M. I.; Neu, J. M.; Wright, G. D.; Paget, M. S.; Buttner, M. J. Characterization of an inducible vancomycin resistance system in *Streptomyces coelicolor* reveals a novel gene (vanK) required for drug resistance. *Mol Microbiol* **2004**, *52* (4), 1107-1121. DOI: 10.1111/j.1365-2958.2004.04032.x.
(111) *WHO publishes list of bacteria for which new antibiotics are urgently needed*. **2017**. https://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed
(112) Malabarba, A.; Ciabatti, R. Glycopeptide derivatives. *Curr Med Chem* **2001**, *8* (14), 1759-1773. DOI: 10.2174/0929867013371716.
(113) Smith, J. R.; Roberts, K. D.; Rybak, M. J. Dalbavancin: a novel lipoglycopeptide antibiotic with extended activity against Gram-positive infections. *Infectious diseases and therapy* **2015**, *4* (3), 245-258.
(114) Alduina, R.; Sosio, M.; Donadio, S. Complex regulatory networks governing production of the glycopeptide A40926. *Antibiotics (Basel)* **2018**, *7* (2). DOI: 10.3390/antibiotics7020030.
(115) Zou, Y.; Brunzelle, J. S.; Nair, S. K. Crystal structures of lipoglycopeptide antibiotic deacetylases: implications for the biosynthesis of A40926 and teicoplanin. *Chem Biol* **2008**, *15* (6), 533-545. DOI: 10.1016/j.chembiol.2008.05.009.
(116) Kruger, R. G.; Lu, W.; Oberthür, M.; Tao, J.; Kahne, D.; Walsh, C. T. Tailoring of glycopeptide scaffolds by the acyltransferases from the teicoplanin and A-40,926 biosynthetic operons. *Chem Biol* **2005**, *12* (1), 131-140. DOI: 10.1016/j.chembiol.2004.12.005.
(117) Li, Y. S.; Ho, J. Y.; Huang, C. C.; Lyu, S. Y.; Lee, C. Y.; Huang, Y. T.; Wu, C. J.; Chan, H. C.; Huang, C. J.; Hsu, N. S.; et al. A unique flavin mononucleotide-linked primary alcohol oxidase for glycopeptide A40926 maturation. *J Am Chem Soc* **2007**, *129* (44), 13384-13385. DOI: 10.1021/ja075748x.
(118) Bibb, M. J. Understanding and manipulating antibiotic production in actinomycetes. *Biochem Soc Trans* **2013**, *41* (6), 1355-1364. DOI: 10.1042/BST20130214.
(119) Alduina, R.; Lo Piccolo, L.; D'Alia, D.; Ferraro, C.; Gunnarsson, N.; Donadio, S.; Puglia, A. M. Phosphate-controlled regulator for the biosynthesis of the dalbavancin precursor A40926. *J Bacteriol* **2007**, *189* (22), 8120-8129. DOI: 10.1128/JB.01247-07.

(120) Lo Grasso, L.; Maffioli, S.; Sosio, M.; Bibb, M.; Puglia, A. M.; Alduina, R. Two master switch regulators trigger A40926 biosynthesis in *Nonomuraea* sp. strain ATCC 39727. *J Bacteriol* **2015**, *197* (15), 2536-2544. DOI: 10.1128/JB.00262-15.
(121) Alduina, R.; Tocchetti, A.; Costa, S.; Ferraro, C.; Cancemi, P.; Sosio, M.; Donadio, S. A two-component regulatory system with opposite effects on glycopeptide antibiotic biosynthesis and resistance. *Sci Rep* **2020**, *10* (1), 6200. DOI: 10.1038/s41598-020-63257-4.
(122) Shawky, R. M.; Puk, O.; Wietzorrek, A.; Pelzer, S.; Takano, E.; Wohlleben, W.; Stegmann, E. The border sequence of the balhimycin biosynthesis gene cluster from *Amycolatopsis balhimycina* contains bbr, encoding a StrR-like pathway-specific regulator. *J Mol Microbiol Biotechnol* **2007**, *13* (1-3), 76-88. DOI: 10.1159/000103599.
(123) Horbal, L.; Kobylyanskyy, A.; Truman, A. W.; Zaburranyi, N.; Ostash, B.; Luzhetskyy, A.; Marinelli, F.; Fedorenko, V. The pathway-specific regulatory genes, *tei15** and *tei16**, are the master switches of teicoplanin production in *Actinoplanes teichomyceticus*. *Appl Microbiol Biotechnol* **2014**, *98* (22), 9295-9309. DOI: 10.1007/s00253-014-5969-z.
(124) Agarwal, R.; Bartsch, S. M.; Kelly, B. J.; Prewitt, M.; Liu, Y.; Chen, Y.; Umscheid, C. A. Newer glycopeptide antibiotics for treatment of complicated skin and soft tissue infections: systematic review, network meta-analysis and cost analysis. *Clin Microbiol Infect* **2018**, *24* (4), 361-368. DOI: 10.1016/j.cmi.2017.08.028.
(125) Beltrametti, F.; Jovetic, S.; Feroggio, M.; Gastaldo, L.; Selva, E.; Marinelli, F. Valine influences production and complex composition of glycopeptide antibiotic A40926 in fermentations of *Nonomuraea* sp. ATCC 39727. *J Antibiot (Tokyo)* **2004**, *57* (1), 37-44. DOI: 10.7164/antibiotics.57.37.
(126) Gunnarsson, N.; Bruheim, P.; Nielsen, J. Glucose metabolism in the antibiotic producing actinomycete *Nonomuraea* sp. ATCC 39727. *Biotechnol Bioeng* **2004**, *88* (5), 652-663. DOI: 10.1002/bit.20279.
(127) Technikova-Dobrova, Z.; Damiano, F.; Tredici, S. M.; Vigliotta, G.; di Summa, R.; Palese, L.; Abbrescia, A.; Labonia, N.; Gnoni, G. V.; Alifano, P. Design of mineral medium for growth of *Actinomadura* sp. ATCC 39727, producer of the glycopeptide A40926: effects of calcium ions and nitrogen sources. *Appl Microbiol Biotechnol* **2004**, *65* (6), 671-677. DOI: 10.1007/s00253-004-1626-2.
(128) Li, Z. L.; Wang, Y. H.; Chu, J.; Zhuang, Y. P.; Zhang, S. L. Effect of branched-chain amino acids, valine, isoleucine and leucine on the biosythesis of bitespiramycin 4"-O-acylspiramycins. *Braz J Microbiol* **2009**, *40* (4), 734-746. DOI: 10.1590/S1517-83822009000400003.
(129) Chen, M.; Xu, T.; Zhang, G.; Zhao, J.; Gao, Z.; Zhang, C. High-yield production of lipoglycopeptide antibiotic A40926 using a mutant strain *Nonomuraea* sp. DP-13 in optimized medium. *Prep Biochem Biotechnol* **2016**, *46* (2), 171-175. DOI: 10.1080/10826068.2015.1015561.
(130) Sosio, M.; Canavesi, A.; Stinchi, S.; Donadio, S. Improved production of A40926 by *Nonomuraea* sp. through deletion of a pathway-specific acetyltransferase. *Appl Microbiol Biotechnol* **2010**, *87* (5), 1633-1638. DOI: 10.1007/s00253-010-2579-2.
(131) Dong, H.; Yue, X.; Yan, B.; Gao, W.; Wang, S.; Li, Y. Improved A40926 production from *Nonomuraea gerenzanensis* using the promoter engineering and the co-expression of crucial genes. *J Biotechnol* **2020**, *324*, 28-33. DOI: 10.1016/j.jbiotec.2020.09.017.

III. AIMS and OUTLINES

The aim of this thesis was to perform further studies about relevant industrial "rare" actinomycetes such us *N. gerenzanensis* and *A. teichomyceticus*, which produce important GPAs used in clinic. For this purpose, we used new technologies such as genetic engineering tools to manipulate GPA BGCs, and genome sequencing and genome mining for the identification of new GPAs. In particular, we aimed at:

1. Improvement of A40926 production in *N. gerenzanensis.* By generating and testing a collection of the promoter probe vectors, we aimed to estimate the strongest promoters for gene expression in *N. gerenzanensis.* The best promoter was then chosen to overexpress in *N. gerenzanensis* CSRs of A40926 BGC – *dbv3* and *dbv4*, as well as *dbv3* orthologue from a putative novel GPA producer *N. coxensis.*
2. Studying GPA production in *N. coxensis*. This strain itself was a very interesting object for us, since we wanted to investigate whether this strain is able to produce a novel GPA. Thus, we planned to assembly the complete genome of *N. coxensis,* annotate the putative GPA BGC, screen different production conditions to find the proper ones for GPA production, and isolate the new GPA by D-Ala-D-Ala affinity chromatography for proceeding in its structure elucidation.
3. Heterologous expression of StrR-like (*dbv4)* and LuxR-like (*dbv3*) CSRs from *dbv* in *A. teichomyceticus* mutants lacking genes for teicoplanin (*tei*) BGC-situated CSRs, i.e., *tei15** (StrR, *dbv4* orthologue) and *tei16** (LuxR, not related to *dbv3*) and *vice versa.* Such investigation was aimed to identify wether CSRs from distant GPA BGCs would be able to "cross-talk" between the pathways.
4. Comparative genomic analysis of *van*-genes distribution in Actinobacteria phylum to test the hypothesis that *van*-genes originate from GPA producers, investigate their distribution and discover new putative GPA BGCs.

These studies alltogether might contribute to a better understanding of the regulatory networks controlling GPAs biosynthesis, opening new ways to improve GPA production and thus reducing their final costs. In addition, since "rare" actinomycetes represent an untapped source of new compounds, we could discover new GPAs by their genome mining, highlight their potential to be used in counteracting antimicrobial resistance. Finally, since the producing actinomycetes are considered the evolutionary source of GPA resistance genes, it was important to examine this hypothesis by analyzing the abundance of genes either in the genomes of GPA-producers and non-producing actinobacteria, available in public databases at the moment of this work.

Thus, briefly, the outline of the chapters that compose this thesis are:

- **Chapter 1**: Here we describe novel approaches for A40926 production improvement. A collection of 11 promoters (heterologous and native) was tested in *N. gerenzanensis* and in *N. coxensis* using the *gusA*-reporter system to select the strongest one for the expression of GPA regulatory genes in these "rare" actinomycetes. Thus, the strongest constitutive promoter among those analyzed – *aac(3)IVp* – was successfully applied to increase A40926 production in *N. gerenzanensis* by overexpressing the CSR from *dbv* and from novel *N. coxensis* GPA BGC.

- **Chapter 2**: In this chapter we report the discovery of a novel GPA – named A50926 – from *N. coxensis* DSM 45129 by genome mining. Corresponding BGC – *noc*

– is similar to *dbv*. The only difference is the absence of the *dbv29*-like gene in *noc*, which codes for an enzyme responsible for the oxidation of the *N*-acylglucosamine moiety.

- **Chapter 3:** Here we report results of large comparative genomic study, revealing that GPA resistance genes – so called *van*-genes – are widespread within the Actinobacteria phylum, either in GPA-producers or non-producing bacteria. For this purpose, we have analysed more than 7000 genomes of actinobacteria available in public databases to map *van*-genes and GPA BGCs. We have also investigated the phylogeny of main *van*-gene-encoded proteins and discovered new putative GPA BGCs.

- **Chapter 4**: Since StrR and LuxR cluster-situated regulators are essential for the production of GPAs in some relevant producers like *N. gerenzanenesis* or *A. teichomyceticus,* in this chapter we investigated if they are able to function in heterologous hosts. We have expressed *tei15** and *tei16** in *dbv4* and *dbv3* knocked out mutants of *N. gerenzanensis* and *vice versa* to demonstrate that it is possible to complement GPA production using non-native CSRs.

IV. RESULTS

CHAPTER 1:

New molecular tools for regulation and improvement of A40926 glycopeptide antibiotic production in *Nonomuraea gerenzanensis* ATCC 39727

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 January 2020
doi: 10.3389/fmicb.2020.00008

New Molecular Tools for Regulation and Improvement of A40926 Glycopeptide Antibiotic Production in *Nonomuraea gerenzanensis* ATCC 39727

Oleksandr Yushchuk[1], Andres Andreo-Vidal[1], Giorgia Letizia Marcone[1], Mervyn Bibb[2], Flavia Marinelli[1] and Elisa Binda[1]*

[1]*Department of Biotechnology and Life Sciences, University of Insubria, Varese, Italy,* [2]*Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich, United Kingdom*

OPEN ACCESS

Edited by:
Yvonne Mast, German Collection of Microorganisms and Cell Cultures GmbH (DSMZ), Germany

Reviewed by:
Max Julian Cryle, Monash University, Australia
Marta Vaz Mendes, University of Porto, Portugal

***Correspondence:**
Flavia Marinelli
flavia.marinelli@uninsubria.it

Specialty section:
This article was submitted to Antimicrobials, Resistance and Chemotherapy, a section of the journal Frontiers in Microbiology

Received: *11 November 2019*
Accepted: *06 January 2020*
Published: *21 January 2020*

Citation:
Yushchuk O, Andreo-Vidal A, Marcone GL, Bibb M, Marinelli F and Binda E (2020) New Molecular Tools for Regulation and Improvement of A40926 Glycopeptide Antibiotic Production in Nonomuraea gerenzanensis ATCC 39727. Front. Microbiol. 11:8. doi: 10.3389/fmicb.2020.00008

Genome sequencing has revealed that *Nonomuraea* spp. represent a still largely unexplored source of specialized metabolites. *Nonomuraea gerenzanensis* ATCC 39727 is the most studied representative species since it produces the glycopeptide antibiotic (GPA) A40926 – the precursor of the clinically relevant antibiotic dalbavancin, approved by the FDA in 2014 for the treatment of acute skin infections caused by multi-drug resistant Gram-positive pathogens. The clinical relevance of dalbavancin has prompted increased attention on A40926 biosynthesis and its regulation. In this paper, we investigated how to enhance the genetic toolkit for members of the *Nonomuraea* genus, which have proved quite recalcitrant to genetic manipulation. By constructing promoter-probe vectors, we tested the activity of 11 promoters (heterologous and native) using the GusA reporter system in *N. gerenzanensis* and in *Nonomuraea coxensis*; this latter species is phylogenetically distant from *N. gerenzanesis* and also possesses the genetic potential to produce A40926 or a very similar GPA. Finally, the strongest constitutive promoter analyzed in this study, *aac(3)IVp*, was used to overexpress the cluster-situated regulatory genes controlling A40926 biosynthesis (*dbv3* and *dbv4* from *N. gerenzanensis* and *nocRI* from *N. coxensis*) in *N. gerenzanensis*, and the growth and productivity of the best performing strains were assessed at bioreactor scale using an industrial production medium. Overexpression of positive pathway-specific regulatory genes resulted in a significant increase in the level of A40926 production in *N. gerenzanensis*, providing a new knowledge-based approach to strain improvement for this valuable glycopeptide antibiotic.

Keywords: A40926, *Nonomuraea*, glycopeptide antibiotics, pathway-specific regulators, strain improvement

INTRODUCTION

Research on glycopeptide antibiotics (GPAs) – drugs of "last resort" for treating severe infections caused by multi-drug resistant Gram-positive pathogens – has experienced a "renaissance" over the last decade (Marcone et al., 2018). Clinically important GPAs include two natural products (vancomycin and teicoplanin) and three second generation antibiotics (telavancin,

dalbavancin, and oritavancin), which are semisynthetic derivatives of natural products endowed with an increased antimicrobial potency and superior pharmacokinetic properties. The urgent need for new potent antibiotics has driven much recent interest in GPAs. For example, chemical synthesis recently resulted in the generation of a plethora of vancomycin derivatives with novel modifications that show superior antimicrobial activities (Okano et al., 2017; Wu and Boger, 2019), while teicoplanin has been conjugated with nanoparticles resulting in increased activity against biofilm-forming pathogens (Armenia et al., 2018).

In contrast, genetic manipulation of GPA producers to yield novel potent derivatives is in its infancy. Recent work (Haslinger et al., 2015; Peschke et al., 2017; Schoppet et al., 2019) has revealed new details of the specificity and timing of non-ribosomal peptide synthesis, including chlorination and cross-linking steps, suggesting that the use of combinatorial biosynthesis to generate GPAs with completely novel oligopeptide scaffolds should be possible. In parallel, heterologous expression of enzymes involved in later stages of GPA biosynthesis (glycosylation, sulfation, acylation etc.) in known producers or *in vitro* has already generated novel GPA derivatives that could not be prepared easily by chemical synthesis (Banik and Brady, 2008; Banik et al., 2010; Yim et al., 2014).

In the meantime, genome sequencing has revealed the organization of GPA biosynthetic gene clusters (BGCs) in industrially valuable actinobacteria (D'Argenio et al., 2016), including long known GPA-producers (Vongsangnak et al., 2012; Truman et al., 2014; Kusserow and Gulder, 2017; Nazari et al., 2017; Adamek et al., 2018) as well as in novel producing strains (Thaker et al., 2013; Stegmann et al., 2014). Although the global regulation of GPA biosynthesis is still largely unexplored, the pathway-specific regulation controlling the expression of BGCs is being elucidated in model systems (Bibb, 2013). The roles of cluster-situated regulatory genes have been investigated in *Amycolatopsis balhimycina* (Shawky et al., 2007), *Nonomuraea gerenzanensis* (Lo Grasso et al., 2015; Alduina et al., 2018), and *Actinoplanes teichomyceticus* (Horbal et al., 2014b; Yushchuk et al., 2019), producing balhimycin, A40926 (the natural precursor of dalbavancin), and teicoplanin, respectively. Overexpression of the teicoplanin cluster-situated regulatory genes (*tei15**, coding for a StrR-like transcriptional regulator, and *tei16**, coding for a LuxR-type regulator) in *A. teichomyceticus* markedly increased teicoplanin production in the wild type strain, representing one of the most successful examples of using molecular tools for improving antibiotic production (Horbal et al., 2012, 2014b). In this work, we investigated the potential of molecular tools to improve the production of A40926, the dalbavancin precursor.

The BGC for A40926, named *dbv*, contains two genes encoding transcriptional regulators: *dbv3* and *dbv4* (Lo Grasso et al., 2015). *dbv3* encodes a LuxR-type regulator, which, however, is non-orthologous to the *tei* cluster encoded LuxR-regulator – Tei16* (Yushchuk et al., 2019). *dbv4* codes for a StrR-like transcriptional regulator with close homologues in every GPA BGC (Yushchuk et al., 2019). The A40926-producing strain, recently re-classified as *N. gerenzanensis* (Dalmastri et al., 2016), belongs to a still poorly investigated genus of actinobacteria that was only recently identified as an untapped source of novel antibiotics and other bioactive metabolites (Sungthong and Nakaew, 2015). More recently, fully sequenced genomes of *N. gerenzanensis* (D'Argenio et al., 2016) and of the kistamicin producer *Nonomuraea* sp. ATCC 55076, previously classified as *Actinomadura parvosata* subsp. *kistnae* S382–8 (Kusserow and Gulder, 2017; Nazari et al., 2017), have confirmed the hidden potential of these uncommon actinomycetes as prolific producers of specialized metabolites. In this paper, we report that another member of this genus, *Nonomuraea coxensis* DSM 45129, which was isolated in Bangladesh in 2007 (Ara et al., 2007), has the genetic potential to produce A40926 or a very similar GPA; its BGC contains two regulatory genes, *nocRI* and *nocRII*, which are close homologs of *dbv3* and *dbv4*, respectively. Thus, we first developed the molecular tools to manipulate both *N. gerenzanensis* and *N. coxensis*, we then selected the strongest heterologous promoter to drive gene expression in *Nonomuraea* spp., and finally we overexpressed both native and heterologous cluster-specific regulatory genes in *N. gerenzanensis*, assessing the best performers at flask and bioreactor scale in industrial media. The overexpression of the positive pathway-specific regulators significantly increased the level of A40926 production in *N. gerenzanensis*, paving the way for knowledge-based strain improvement for the production of this valuable GPA.

MATERIALS AND METHODS

Plasmids, Bacterial Strains, Antibiotics, and Culture Conditions

Plasmids and bacterial strains used in this work are summarized in **Table 1**. Compositions of media are given in ESM. Unless otherwise stated, all media components and antibiotics were supplied by Sigma-Aldrich, St. Louis, MO, United States. For routine maintenance, actinobacterial strains were cultivated on ISP3 or VM0.1 agar media supplemented with 50 μg/ml apramycin-sulfate when appropriate. For genomic DNA isolation, *N. gerenzanensis* ATCC 39727 and *N. coxensis* DSM 45129 were grown in 250 ml Erlenmeyer flasks containing 10 glass beads (ø5 mm) with 50 ml of liquid VSP medium on an orbital shaker at 220 rpm and at 30°C. Working cell banks (WCB) of *Nonomuraea* spp. were prepared as described previously (Marcone et al., 2014). *Escherichia coli* DH5α was used as a routine cloning host and *E. coli* ET12567 pUZ8002 as a donor for intergeneric conjugations. *E. coli* strains were grown at 37°C in LB liquid or agar media supplemented with 100 μg/ml of apramycin-sulfate, 50 μg/ml of kanamycin-sulfate and 25 μg/ml of chloramphenicol when appropriate.

Generation of Recombinant Plasmids

Construction of Promoter-Probe Vectors

To test the activity of different native *N. gerenzanensis* promoters, pSAGA (Koshla et al., 2019), where *gusA* (Myronovskyi et al., 2011) is expressed from *aac(3)IVp*, was chosen as a chassis. Genomic DNA, extracted from *N. gerenzanensis* according to the Kirby procedure (Kieser et al., 2000), was used as a template

to amplify the putative promoter regions of: *BN4615_P641* (coding for the DNA gyrase B subunit – $GyrB_{ng}$) 342 bp; *BN4615_P8899* (coding for the RNA polymerase sigma factor RpoD – $HrdB_{ng}$), 524 bp; *BN4615_P604* (coding for the single-stranded DNA-binding protein – Ssb_{ng}), 207 bp; *BN4615_P1543* (coding for the SSU ribosomal protein S12p – $RpsL_{ng}$), 339 bp; *BN4615_P7269* (coding for the rifamycin-resistant RNA polymerase subunit β – $RpoB_{Rng}$), 558 bp; and *BN4615_P1539* (coding for the rifamycin-sensitive RNA polymerase subunit β – $RpoB_{ng}$), 493 bp. Amplicons were generated using Q5 High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, United States) according to the supplier's protocol and the oligonucleotide primers listed in **Table 2**. All of the amplicons were digested with *Bam*HI and *Eco*RV and cloned in pSAGA cleaved with the same endonucleases, thus replacing *aac(3)IVp* in front of *gusA* with each of the amplified promoter regions. The resulting recombinant plasmids were named pGBP (carrying $gyrB_{ng}p$), pHBP ($hrdB_{ng}p$), pSBP ($ssb_{ng}p$), pRLP ($rpsL_{ng}p$), pRBP1 ($rpoB_{ng}p$), and pRBP2 ($rpoB_{Rng}p$).

Construction of the *dbv3*, *dbv4*, and *nocRI* Overexpression Plasmids

The coding sequences of *dbv3* (2,635 bp) and *dbv4* (1,006 bp) were amplified from the A40Y cosmid (**Table 1**; Marcone et al., 2010a) using Q5 High-Fidelity DNA Polymerase and the dbv3_F/R or dbv4_F/R primer pairs (**Table 2**). The obtained amplicons were digested with *Eco*RI and *Eco*RV and cloned into pSET152A cleaved with the same enzymes. The resulting plasmids were named pSAD3 and pSAD4. To generate a vector for the overexpression of both *dbv3* and *dbv4*, the regions containing the coding sequences of both genes were amplified using the dbv4_F/dbv3_R primer pair and cloned into pSET152A in a similar fashion, generating pSAD3–4.

To construct the vector for overexpression of *nocRI* (the *dbv3* ortholog from *N. coxensis*), the coding sequence of *A3G7_RS0138355* was amplified from the genomic DNA of *N. coxensis* isolated using the Kirby procedure (Kieser et al., 2000) using Q5 High-Fidelity DNA Polymerase and the orfR1_F/R primer pair (**Table 2**). The obtained amplicon (2,661 bp) was digested with *Eco*RI and *Eco*RV and cloned into pSET152A cut with the same enzymes to generate pSAR1. All of the generated recombinant plasmids were verified by restriction endonuclease mapping and sequencing at BMR Genomics (University of Padua, Italy).

Conjugative Transfer of Plasmids Into *Nonomuraea* spp. and Verification of the Recombinant Strains

Conjugative transfer of plasmids into *N. gerenzanensis* was performed essentially as described previously (Marcone et al., 2010c). All recombinant plasmids were transferred individually into the non-methylating *E. coli* ET12567 pUZ8002 and the resulting derivatives used as donor strains for intergeneric conjugation. To prepare fresh vegetative mycelium of *N. gerenzanensis* prior to conjugal transfer, one vial of WCB was inoculated into 50 ml of VSP medium (250 ml Erlenmeyer

TABLE 1 | Bacterial strains and plasmids used in this work.

Name	Description	Source of reference
N. gerenzanensis	Wild type, A40926 producer	ATCC 39727
N. coxensis	Wild type	DSM 45129
N. gerenzanensis pSET152A$^+$	Wild type derivative carrying pSET152A	This work
N. gerenzanensis pSAD3$^+$	Wild type derivative carrying pSAD3	This work
N. gerenzanensis pSAD4$^+$	Wild type derivative carrying pSAD4	This work
N. gerenzanensis pSAD3–4$^+$	Wild type derivative carrying pSAD3–4	This work
N. gerenzanensis pSAR1$^+$	Wild type derivative carrying pSAR1	This work
N. gerenzanensis pSAGA$^+$	Wild type derivative carrying pSAGA	This work
N. gerenzanensis pTEGA$^+$	Wild type derivative carrying pTEGA	This work
N. gerenzanensis pGUSmoeE5script$^+$	Wild type derivative carrying pGUSmoeE5script	This work
N. gerenzanensis pGCymRP21$^+$	Wild type derivative carrying pGCymRP21	This work
N. gerenzanensis pGT2p$^+$	Wild type derivative carrying pGT2p	This work
N. gerenzanensis pGBP$^+$	Wild type derivative carrying pGBP	This work
N. gerenzanensis pHBP$^+$	Wild type derivative carrying pHBP	This work
N. gerenzanensis pSBP$^+$	Wild type derivative carrying pSBP	This work
N. gerenzanensis pRLP$^+$	Wild type derivative carrying pRLP	This work
N. gerenzanensis pRBP1$^+$	Wild type derivative carrying pRBP1	This work
N. gerenzanensis pRBP2$^+$	Wild type derivative carrying pRBP2	This work
N. coxensis pSAGA$^+$	Wild type derivative carrying pSAGA	This work
N. coxensis pTEGA$^+$	Wild type derivative carrying pTEGA	This work
N. coxensis pGUSmoeE5script$^+$	Wild type derivative carrying pGUSmoeE5script	This work
N. coxensis pGCymRP21$^+$	Wild type derivative carrying pGCymRP21	This work
N. coxensis pGT2p$^+$	Wild type derivative carrying pGT2p	This work
N. coxensis pGBP$^+$	Wild type derivative carrying pGBP	This work
N. coxensis pHBP$^+$	Wild type derivative carrying pHBP	This work
N. coxensis pSBP$^+$	Wild type derivative carrying pSBP	This work
N. coxensis pRLP$^+$	Wild type derivative carrying pRLP	This work
N. coxensis pRBP1$^+$	Wild type derivative carrying pRBP1	This work
N. coxensis pRBP2$^+$	Wild type derivative carrying pRBP2	This work
E. coli DH5α	General cloning host	MBI Fermentas, USA
E. coli ET12567 pUZ8002	(*dam-13*::Tn9 *dcm-6*), pUZ8002$^+$ (Δ*oriT*), used for conjugative transfer of DNA	Kieser et al. (2000)

(Continued)

TABLE 1 | Continued

Name	Description	Source of reference
A40Y	SuperCos1 derivative, including 22 kb of *dbv* cluster (*dbv1-dbv17*)	Marcone et al. (2010a)
pGUS	pSET152 derivative, containing promoterless *gusA*	Myronovskyi et al. (2011)
pSET152A	pSET152 derivative, containing *aac(3)IVp* from pIJ773	Horbal et al. (2013)
pSAD3	pSET152A derivative, containing *dbv3* under the control of *aac(3)IVp*	This work
pSAD4	pSET152A derivative, containing *dbv4* under the control of *aac(3)IVp*	This work
pSAD3–4	pSET152A derivative, containing *dbv4* together with *dbv3*	This work
pSAR1	pSET152A derivative, containing *nocRI* under the control of *aac(3)IVp*	This work
pSAGA	pSET152A derivative, containing *gusA* under the control of *aac(3)IVp*	Koshla et al. (2019)
pTEGA	pTES derivative, containing *gusA* under the control of *ermEp*	Yushchuk et al. (2020a)
pGUSmoeE5script	pGUS derivative, containing *gusA* under the control of *moeE5p*	Makitrynskyy et al. (2013)
pGCymRP21	pGUS derivative, containing CymR operator, *P21* promoter and *cymR* gene	Horbal et al. (2014a)
pGT2p	pGUS derivative, containing *gusA* under the control of *tei2p*	Yushchuk et al. (2020b)
pGBP	pSAGA derivative, containing *gusA* under the control of $gyrB_{Rng}p$	This work
pHBP	pSAGA derivative, containing *gusA* under the control of $hrdB_{Rng}p$	This work
pSBP	pSAGA derivative, containing *gusA* under the control of $ssb_{Rng}p$	This work
pRLP	pSAGA derivative, containing *gusA* under the control of $rpsL_{Rng}p$	This work
pRBP1	pSAGA derivative, containing *gusA* under the control of $rpoB_{Rng}p$	This work
pRBP2	pSAGA derivative, containing *gusA* under the control of $rpoB_{RngP}$	This work

flask with 10 ø5 mm glass beads) and incubated for 48 h on the orbital shaker at 220 rpm and at 30°C. The mycelium was collected by centrifugation (10 min, 3,220 × *g*), washed twice with sterile 20% v/v glycerol, resuspended in the same solution to a final volume of 20 ml, and stored at −80°C. 1 ml of mycelial suspension was mixed with approximately 10^9 of donor *E. coli* cells and the mixtures were plated on well dried VM0.1 agar plates supplemented with 20 mM of $MgCl_2$. After 12–16 h of incubation at 30°C, each plate was overlaid with 1 ml of sterile deionized water containing 1.25 mg of apramycin-sulfate and 750 μg of nalidixic acid sodium salt. Transconjugants were selected as resistant to 50 μg/ml of apramycin-sulfate.

Spore suspensions of *N. coxensis* were prepared from lawns grown on ISP3 agar for 7 days. Spores from one plate were collected in deionized water and filtered through one layer of Miracloth (Merck KGaA, Darmstadt, Germany) to remove vegetative mycelial fragments. Then, spores were pelleted from a 50 ml suspension by centrifugation (15 min, 3,220 × *g*), resuspended in 1 ml of 15% v/v glycerol, and stored at −80°C. For conjugation, approx. 10^6 spores were mixed with 10^7 *E. coli* donor cells and plated on VM0.1 agar plates supplemented with 20 mM of $MgCl_2$. The overlay for the selection of transconjugants was performed as described previously for *N. gerenzanensis*.

To verify the integration of promoter-probe vectors, an ~1 kbp internal fragment of *gusA* was amplified from the genomic DNA of recombinant *N. gerenzanensis* or *N. coxensis* strains using the gusA_ver_F/R primer pair (**Table 2**). To verify the integration of pSET152A, *aac(3)IV* was amplified using the aac(3)IV_F/R primer pair (**Table 2**). To verify the integration of pSAD4, an ~1 kbp fragment of pSAD4 was amplified with the PAM_seq_F/dbv4_R primer pair (**Table 2**), in which PAM_seq_F anneals within the *aac(3)IVp* sequence. Verification of pSAD3 and pSAD3–4 integration was made by amplification of an ~2 kbp fragment (for pSAD3) or an ~3 kbp fragment (for pSAD3–4), using the PAM_seq_F/dbv3_seq_R primer pair (**Table 2**), in which dbv3_seq_R anneals in the middle of *dbv3*. Finally, to verify the integration of pSAR1 an ~2 kbp fragment was amplified using the PAM_seq_F/orfR1mid_EcoRV_R (**Table 2**) primer pair, in which orfR1mid_EcoRV_R anneals in the middle of *nocRI*.

β-Glucuronidase Activity Assay

β-Glucuronidase (GusA) activity in *Nonomuraea* strains growing on VM0.1 agar medium was assessed by adding, after 6 days of cultivation at 30°C, 10 μl drops of 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-Gluc, Thermo Fisher Scientific, Waltham, MA, United States) 50 mg/ml in DMSO to the surfaces of the lawns. Chromogenic conversion of X-Gluc into the blue-colored 5,5′-dibromo-4,4′-dichloro-indigo was monitored after 1 h of incubation. For the quantitative measurements of GusA activity, *Nonomuraea* strains were grown in liquid media. One WCB vial of each of the strains was inoculated into a baffled 500 ml Erlenmeyer flask containing 100 ml of E26 (*N. gerenzanensis* strains) or of VSP (*N. coxensis* strains). After 72 h of cultivation, 10% v/v of this preculture was transferred into a baffled 500 ml Erlenmeyer flask containing 100 ml of FM2 (*N. gerenzanensis* strains) or ISP2 (*N. coxensis* strains). To induce *P21-cmt*-driven *gusA*-expression, cumate was added at the final concentration of 50 μM to cultures carrying pGCymRP21 24 h after inoculation. After 120 h of cultivation, mycelial lysates were prepared as previously reported by Horbal et al., 2013. Glucuronidase activity was measured as previously described (Myronovskyi et al., 2011; Horbal et al., 2013) using

TABLE 2 | Oligonucleotide primers used in this work.

Primer	Nucleotide sequence (5′-3′)*	Purpose
Dbv3_F	TTTGATATCGGAGGGCAAGAGTGCTGTTCGGGC	Cloning of *dbv3* into pSET152A
Dbv3_R	TTTGAATTCCTCTACAGCCGCACTGCCT	
dbv4_F	TTTGATATCGGAGGGGCTAGGTGGACCCGACGG	Cloning of *dbv4* into pSET152A
dbv4_R	TTTGAATTCTCCACTCGTGCTCATCCAG	
PAM_seq_F	GATGTCATCAGCGGTGGAG	Verification of recombinant strains
dbv3_seq_R	CCAGCGCTGGACCGCCTGC	
orfR1nid_EcoRV_rev	TTTGATATCGCAAGGGGCCTCCCCGCCG	
orfR1_F	TTTGATATCGGAGGACTGCGTTGACGAACCGCT	Cloning of *nocRI* into pSET152A
orfR1_R	TTTGAATTCGCGTCATGGGACCACCGCC	
hrdBp_F	TTTGGATCCACCGAAGCGCCGCCTGAGG	Cloning of $hrdB_{ng}p$
hrdBp_R	TTTGATATCGAAGGCCTGACGGACATCC	
rpoB1p_F	TTTGGATCCTCTCGCTGGCTGGTGGCCG	Cloning of $rpoB_{ng}p$
rpoB1p_R	TTTGATATCTCGCGGCGGACTGACTACA	
rpoB2p_F	TTTGGATCCTGTCGTACTGCTC	Cloning of $rpoB_{Rng}p$
rpoB2p_R	TTTGATATCATACGAAGGCGAGGGAGGG	
rpsLp_F	TTTGGATCCATGGACGGCGGAGCTGTAG	Cloning of $rpsL_{ng}p$
rpsLp_R	TTTGATATCTTGGCCGGTGTTACGTCA	
ssbp_F	TTTGGATCCAAGTCCGAAGGCATCTACG	Cloning of $ssb_{ng}p$
ssbp_R	TTTGATATCTGCACGCCTTCGCTTGGGT	
gyrBAp_F	TTTGGATCCAGGCTTCGCACAGTAACGG	Cloning of $gyrB_{ng}p$
gyrBAp_R	TTTGATATCGCGGACACGCGGCGGGGGA	
aac(3)IV_F	ATCGACTGATGTCATCAGCG	Amplification of *aac(3)IV*
aac(3)IV_R	CGAGCTGAAGAAAGACAAT	
gusA_ver_F	GGCGGCTACACGCCCTTCGA	Amplification of *gusA* internal fragment
gusA_ver_R	TGATGGGCCGGGTGGGGTC	

*Restriction sites are underlined in primer sequence.

a spectrophotometric assay following the conversion of the colorless *p*-nitrophenyl-β-D-glucuronide (Thermo Fisher Scientific, Waltham, MA, United States) into the colored *p*-nitrophenol at 415 nm using an Infinite 200 PRO microplate reader (Tecan, Switzerland). Glucuronidase activity was normalized to dry biomass weight as previously reported (Marcone et al., 2014). One unit of activity is defined as the amount of enzyme that is able to convert 1 μM of substrate in 1 min.

A40926 Production

One WCB vial was inoculated into 300 ml baffled flasks containing 50 ml of vegetative medium E26 with 10 glass beads (ø5 mm). Flask cultures were incubated for 72 h on a rotary shaker at 220 rpm and 30°C and then used to inoculate (10% v/v) 500 ml baffled Erlenmeyer flasks containing 100 ml of FM2 medium or a 3-l P-100 Applikon glass reactor (height 250 mm, ø130 mm) equipped with a AD1030 Biocontroller and AD1032 motor, and containing 2 l of the same production medium. Cultivations in FM2 in shake-flasks were conducted at 30°C and 220 rpm. Bioreactor fermentations were conducted at 30°C, with stirring at 450 rpm (corresponding to 1.17 m/s of tip speed) and 2 l/min aeration rate. Dissolved oxygen (measured as % pO2) was monitored using an Ingold polarographic oxygen electrode. The pH values of culture broths were monitored using a pH meter. Foam production was controlled by adding Hodag antifoam (Hodag Chemical Corporation, Chicago, IL, United States) through an antifoam sensor. Samples were collected at regular cultivation time intervals and analyzed to estimate biomass (dry weight), glucose consumption (Diastix sticks, Bayer AG, Leverkusen, Germany), and A40926 production.

HPLC Analysis of Culture Extracts

A40926 was extracted from *Nonomuraea* spp. cultures as previously reported (Marcone et al., 2014). Chromatography was performed with a VWR Hitachi diode array L-2455 HPLC system with detection at 254 nm. The A40926 titers in the batch cultivations were estimated by injecting 50 μl of sample onto a 5 μm-particle-size Ultrasphere ODS (Beckman) HPLC column (4.6 by 250 mm) and eluting at a flow rate of 1 ml/min with a 30 min linear gradient from 15 to 64% of phase B. Phase A was 32 mM $HCOONH_4$ (pH 7) – CH_3CN [90:10 (vol/vol)], and phase B was 32 mM $HCOONH_4$ (pH 7) – CH_3CN [30:70 (vol/vol)]. A volume of 50 μl of a pure sample of 200 μg/ml A40926 (Sigma-Aldrich, St. Louis, MO, United States) was used as an internal standard.

Tools for the Bioinformatics Analysis

Blastp was used to search for homologs (Altschul, 1990); protein sequence alignments were performed with Clustal Omega (EMBL-EBI, Sievers et al., 2011).

RESULTS

Genetic Manipulation of *Nonomuraea coxensis*, a Novel Putative Producer of a A40926-Like Molecule

According to the 16S rRNA gene-based reconstruction of *Nonomuraea* phylogeny (Dalmastri et al., 2016), *N. coxensis* occupies a relatively distant position from *N. gerenzanensis*. Conversely, mining the partially sequenced genome of *N. coxensis*

(ASM37988v1), we found a close homolog of the *N. gerenzanensis* regulatory gene *dbv3* (locus *A3G7_RS0138355* in the *N. coxensis* genome, coding for WP_020547054.1 with 86.6% predicted amino acid sequence identity with Dbv3, ESM **Figure 1**). Dbv3 is a unique LuxR-regulator controlling the expression of the A40926 BGC and it is not closely related to the better-characterized family of Tei16*-like regulators controlling teicoplanin biosynthesis, sharing only 32% of amino acid sequence identity with Tei16* (Yushchuk et al., 2019). This *dbv3*-like gene (we named it *nocRI*) was found on a short *N. coxensis* contig (NZ_KB904006) flanking a gene coding for a putative StrR-like transcriptional regulator apparently orthologous to *dbv4* (locus *A3G7*_RS0138355, coding for WP_026215141.1 with 94.39% of amino acid sequence identity with Dbv4); *dbv4* is the other known cluster-situated regulatory gene in the *dbv* BGC and we named the *N. coxensis* homolog *nocRII* (ESM **Figure 2**). AntiSMASH (Blin et al., 2019) analysis of *N. coxensis* genome revealed the presence of three short contigs (NZ_KB904006, NZ_KB903995, NZ_KB903969, ESM **Figure 3**) covering the majority of a *dbv*-like BGC, suggesting that *N. coxensis* might produce A40926 or a very similar GPA. Thus, we considered *N. coxensis* an interesting candidate for developing *Nonomuraea*-targeted genetic tools.

Initially, we tried to transfer *φC31*-based integrative plasmids (**Table 1**) into *N. coxensis* by using the protocol of intergeneric conjugation optimized for conjugal transfer from a DNA-non-methylating *E. coli* donor strain to *N. gerenzanensis* vegetative mycelium (Marcone et al., 2010c). Transconjugants were obtained at a very low frequency (*ca.* 1×10^{-7}) and only for the relatively small pSAGA-based promoter-probe vectors (approx. 6 kbp). Increasing the amount of donor and recipient cells, changing the time of overlay, and adjusting medium composition (including increasing or decreasing $MgCl_2$ concentration) did not allow the transfer of the larger pGUS-based promoter-probe vectors (such as pGUSmoeE5script or pGCymRP21, both more than 9 kbp). Since, unlike *N. gerenzanensis*, *N. coxensis* sporulates abundantly when grown on ISP3 agar medium (ESM **Figure 4**), we tried to use spores for conjugal transfer. This resulted in transfer rates of approximately 1×10^{-3} when 10^6 spores were mixed with 10^7 *E. coli* donor cells, regardless of plasmid size.

Using the GusA-Reporter System to Assess the Activity of Native and Heterologous Promoters in *Nonomuraea* Species

The set of *φC31*-based integrative plasmids transferred to *N. gerenzanensis* and *N. coxensis* were promoter-probe vectors utilizing the GusA reporter system and carrying a selection of native and heterologous promoters, the latter having been used previously to drive gene expression in streptomycetes or actinoplanetes (Horbal et al., 2013, 2014a; Makitrynskyy et al., 2013). Glucuronidase activity of the recombinant *Nonomuraea* strains was assessed qualitatively (on agar plates) and quantitatively (in cell lysates obtained from mycelium grown in liquid medium). Both of the *N. coxensis* and *N. gerenzanensis* wild type strains did not display any glucuronidase activity (**Figure 1**). The heterologous promoters tested were: *aac(3)IVp* (in pSAGA) – the apramycin acetyltransferase gene promoter, derived from pSET152A (Horbal et al., 2013); *ermEp* (in pTEGA, Yushchuk et al., 2020a) – the erythromycin resistance gene promoter from pTES (Herrmann et al., 2012); *moeE5p* (in pGUSmoeE5script) – the *S. ghanaensis* moenomycin biosynthesis gene *moeE5* promoter (Makitrynskyy et al., 2013); *tei2p* (pGT2p, Yushchuk et al., 2020b) – the *A. teichomyceticus* teicoplanin resistance gene *tei2* promoter; and *P21*, a synthetic promoter fused with the cumate inducible *Pseudomonas putida* F1 *cmt* operon operator (in pGCymRP21, Horbal et al., 2014a). In parallel, we tested the activity of six native promoters derived from *N. gerenzanensis* house-keeping genes: $gyrB_{ng}p$ (in pGBP – the promoter of the DNA gyrase B subunit gene); $hrdB_{ng}p$

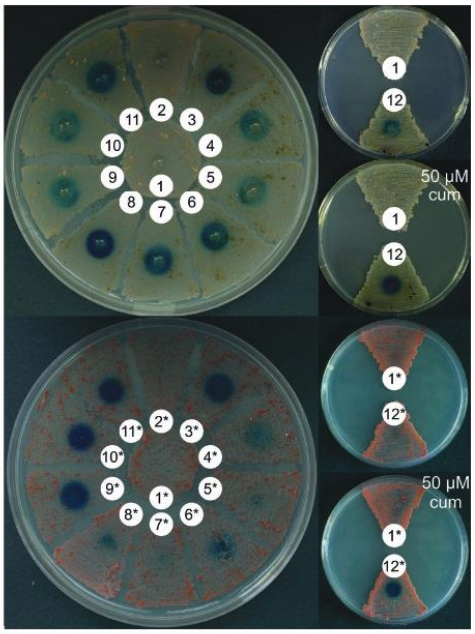


FIGURE 1 | Comparison of GusA activity in recombinant strains of *N. gerenzanensis* **(A)** and *N. coxensis* **(B)**, carrying the following promoter-probe vectors: 2, 2* – pTEGA (*gusA* under the control of *ermEp*); 3, 3* – pSAGA (*gusA* under the control of *aac(3)IVp*); 4, 4* – pHBP (*gusA* under the control of $hrdB_{ng}p$); 5, 5* – pRBP2 (*gusA* under the control of $rpoB_{Rng}p$); 6, 6* – pRBP1 (*gusA* under the control of $rpoB_{ng}p$); 7, 7* – pSBP (*gusA* under the control of $ssb_{ng}p$); 8, 8* – pRLP (*gusA* under the control of $rpsL_{ng}p$); 9, 9* – pGUSmoeE5script (*gusA* under the control of *moeE5p*); 10, 10* – pGT2p (*gusA* under the control of *tei2p*); 11, 11* – pGBP (*gusA* under the control of $gyrB_{ng}p$); 12, 12* – pGCymRP21 (*gusA* under the control of *P21-cmt*). The control parental strains (1, *N. gerenzanensis* and 1*, *N. coxensis*) do not display chromogenic conversion of X-Gluc. Strains were cultivated for 6 days on VM0.1 medium.

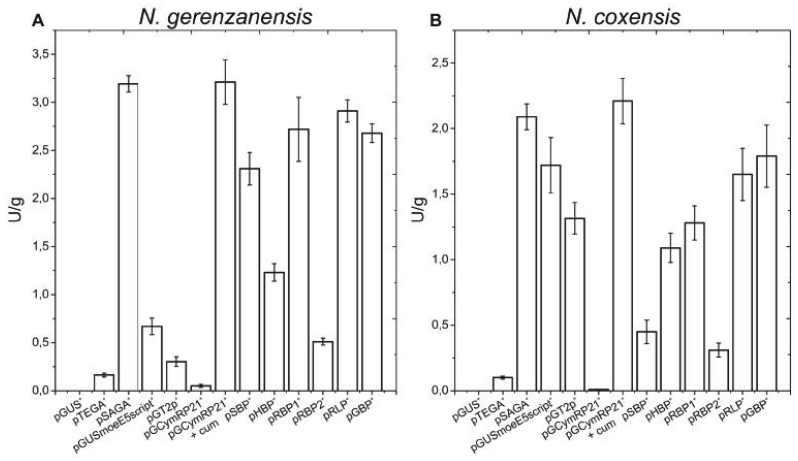


FIGURE 2 | Quantitative measurement of GusA activity (U/gram of dry biomass) in *N. gerenzanensis* **(A)** and *N. coxensis* **(B)** recombinant strains grown in liquid media for 120 h (see growth curves in ESM **Figure 5**), carrying different promoter-probe plasmids. Control strains carried the promoterless pGUS vector and exhibited no GusA activity. Activities are the mean values of three independent experiments. Error bars represent standard deviations.

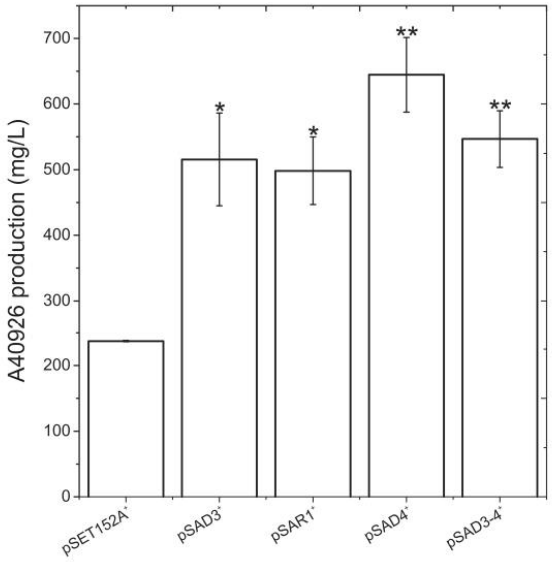


FIGURE 3 | A40926 production in *N. gerenzanensis* carrying the empty vector pSET152A^{-} and in its recombinant strains overexpressing *dbv3* (pSAD3), *nocRI* (pSAR1), *dbv4* (pSAD4), and co-expressing *dbv3* and *dbv4* genes (pSAD3–4) cultivated in FM2 industrial medium at 500 ml Erlenmeyer flasks-scale. A40926 production was measured after 120 h of cultivation. The results given represent three independent fermentations, error bars represent standard deviations. Statistical significance of the differences in A40926 production between the control and the recombinant strains was estimated using Welch's *t*-test: $^{*}p < 0.05$; $^{**}p < 0.01$.

(in pHBP – the promoter of the RNA polymerase σ-factor RpoD gene); $ssb_{ng}p$ (in pSBP – the promoter of the single-stranded DNA-binding protein); $rpsL_{ng}p$ (in pRLP – the promoter of the SSU ribosomal protein S12p gene); $rpoB_{Rng}p$ (in pRBP2 – the promoter of the rifamycin-resistant RNA polymerase subunit β gene); and $rpoB_{ng}p$ (in pRBP1 – the

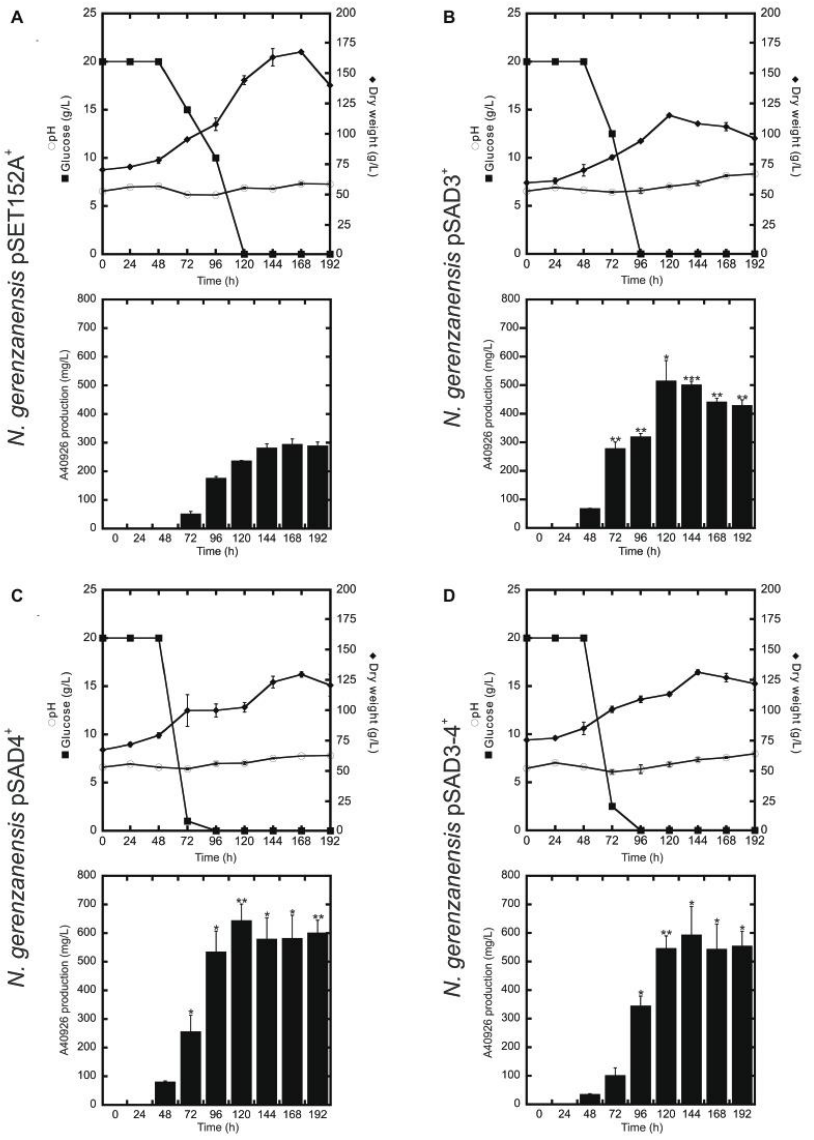


FIGURE 4 | Time courses of *N. gerenzanensis* pSET152A+ **(A)** and recombinant strains overexpressing *dbv3* **(B)**, *dbv4* **(C)** and co-expressing both genes **(D)** cultivated in FM2 industrial medium in 500 ml Erlenmeyer flasks. Glucose consumption (filled rectangles), biomass accumulation (filled rhombi), pH (circles), and A40926 production were monitored every 24 h. Results given are mean values of three independent experiments. Error bars represent standard deviations. Statistical significance of the differences in A40926 production between the control and the recombinant strains was estimated using Welch's *t*-test: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

promoter of the rifamycin-sensitive RNA polymerase subunit β gene). In contrast to the wild type strains, all of the recombinant derivatives grown on VM0.1 agar plates converted X-Gluc to its colored derivative 5,5′-dibromo-4,4′-dichloro-indigo, albeit to different extents (**Figure 1**). *N. gerenzanensis* pSAGA+ and *N. coxensis* pSAGA+, carrying *aac(3)IVp*, produced

the most intensive color (**Figures 1A,B**), while the chromogenic conversion of X-Gluc in *N. gerenzanensis* pTEGA$^+$ and *N. coxensis* pTEGA$^+$, carrying the *ermEp*, was only slightly visible. In strains carrying the inducible pGCymRP21, GusA activity was induced by the addition of 50 μM of cumate, proving inducible gene expression (**Figures 1A,B**). However, in *N. gerenzanensis* pGCymRP21$^+$, a basal level of GusA activity was detected even in the absence of cumate (**Figure 1A**). Contrary to this, no basal level of expression from *P21-cmt* was detected in *N. coxensis* pGCymRP21$^+$ (**Figure 1B**).

When the glucuronidase activity present in cell lysates obtained from liquid cultures grown for 120 h (late exponential/early stationary growth phase, see ESM **Figure 5**) was measured using a spectrophotometric assay, overall these quantitative results (normalized for the dry weight of the differently growing strains) correlated with those observed on agar plates: *aac(3)IVp* behaved as a strong promoter in both *N. gerenzanensis* and *N. coxensis*, surpassed only by *cmt-P21p* when induced by the addition of 50 μM cumate (**Figures 2A,B**). Also $rpsL_{Rng}p$ and $hrdB_{Rng}p$ proved to be strong promoters in both of the *Nonomuraea* spp. whereas the weakest was the *ermEp* (**Figures 2A,B**). Interestingly, the activity of $rpoB_{Rng}P$ was higher than that of $rpoB_{Rng}p$ (**Figures 1**, **2**), consistent with previously reported data about differences in transcription levels of the two alleles (Vigliotta et al., 2004).

Knowledge-Based Generation of A40926 Overproducing Strains

Since *aac(3)IVp* was identified as the strongest constitutive promoter for *Nonomuraea* spp., we used it to overexpress *dbv3* and *dbv4* in *N. gerenzanensis*. Both genes were cloned into the integrative pSET152A vector (Horbal et al., 2013) yielding the recombinant vectors pSAD3 and pSAD4, respectively (**Table 1**). Benefiting from the neighboring positions of *dbv3* and *dbv4*, we also cloned them together, generating pSAD3–4, where *dbv4* was directly under the control of *aac(3)IVp*, but *dbv3* remained under the control of its native promoter (**Table 1**). Additionally, the *dbv3*-like *nocRI* from *N. coxensis* was cloned into pSET152A, generating pSAR1, and transferred to *N. gerenzanensis* to determine if it could improve A40926 production, and to assess possible cross-talk between the regulators from the two different *Nonomuraea* spp. All the *N. gerenzanensis* recombinant strains were grown for 120 h in parallel with the control strain carrying the empty vector and A40926 production was measured (**Figure 3**). The production of A40926 in *N. gerenzanensis* carrying pSET152A reached almost 250 mg/l after 120 h of cultivation (**Figure 3**). The four recombinant strains overexpressing the cluster-situated regulatory genes produced more antibiotic than the control strain (**Figure 3**). A40926 production in *N. gerenzanensis* pSAD3$^+$ and pSAR1$^+$ was comparable (*ca.* 500 mg/l), proving that *nocRI* had a similar impact in *N. gerenzanensis* as *dbv3*. The recombinant strain carrying pSAD3–4 vector produced slightly more, around 550 mg/l, whereas the best producer in these conditions was the strain overexpressing *dbv4*, which produced more than 650 mg/l.

Time Courses of A40926 Production in *Nonomuraea gerenzanensis* Recombinant Strains at Flask and at Bioreactor Scale

Although further investigations will be devoted to the expression of the heterologous *nocRI* and *nocRII in N. gerenzanensis*, in our strain improvement work we then focused on *N. gerenzanensis* strains overexpressing native regulators. Consequently, *N. gerenzanensis* strains containing pSET152A, pSAD3, pSAD4, or pSAD3–4 were grown for 192 h in parallel with the control strain carrying the empty vector using the previously optimized industrial medium FM2 (Marcone et al., 2010a, 2014) at flask scale. Samples for the analysis of dry weight, pH, glucose consumption, and A40926 production were collected at regular 24 h intervals. All three recombinant strains expressing the regulatory genes from *aac(3)IVp* accumulated detectable amounts of A40926 earlier than the empty vector control (**Figure 4**). A40926 production reached its peak after 120 h of growth for *N. gerenzanensis* pSAD3$^+$ (*ca.* 500 mg/l) (**Figure 4B**) and pSAD4$^+$ (nearly 650 mg/l) (**Figure 4C**), whereas in pSAD3–4$^+$ the maximum productivity (*ca.* 600 mg/l) was delayed to 144 h (**Figure 4D**). The control strain produced *ca.* 300 mg/l after 144–168 h from inoculation (**Figure 4A**). Although glucose consumption was faster in the recombinant strains containing the cloned regulatory genes in comparison to the control, they accumulated less biomass than *N. gerenzanensis* pSET152A$^+$ (**Figure 4**). Maximum biomass production was around 125 g/l (dry weight) in the overexpression strains versus the 175 g/l produced by the empty vector control strain. These data suggest that part of the consumed glucose was used by the strains carrying the regulatory genes under the control of the strong constitutive promoter *aac(3)IVp* to support antibiotic production at the detriment of biomass formation.

When the cultivation of recombinant strains was scaled up in 3 l vessel-bioreactors containing 2 l working volume of FM2, all of the strains produced significantly more A40926 and grew better than in flask culture with the exception of *N. gerenzanensis* pSAD3$^+$ (**Figure 5**). The control strain with the empty vector grew and produced the antibiotic faster than at the flask level (**Figure 5A**). Maximum biomass (300 g/l dry weight, more than the double of that achieved in flasks) and A40926 production (nearly 400 mg/l) were reached after 120 h from inoculation; glucose was completely consumed within 96 h versus the 120 h needed at flask level. The recombinant strain pSAD4$^+$ grew more (maximum biomass production of 220 g/l after 144 h of growth) than in flask culture, although less than the control strain in the bioreactor. Glucose was consumed faster than in the control strain and glucose concentration tended to zero at 48 h of fermentation (**Figure 5C**), and antibiotic production reached a peak of 700 mg/l after 168 h. The best performance in terms of A40926 productivity in the bioreactor was shown by the pSAD3–4$^+$ strain, which grew better (nearly 250 g/l dry weight biomass) than in flask culture and produced the maximum concentration of the antibiotic (800 mg/l) after 168 h of cultivation (**Figure 5D**). Conversely, the pSAD3$^+$ strain showed a reduced biomass production in comparison to all of the other strains, with consumption of glucose markedly delayed and A40926 production

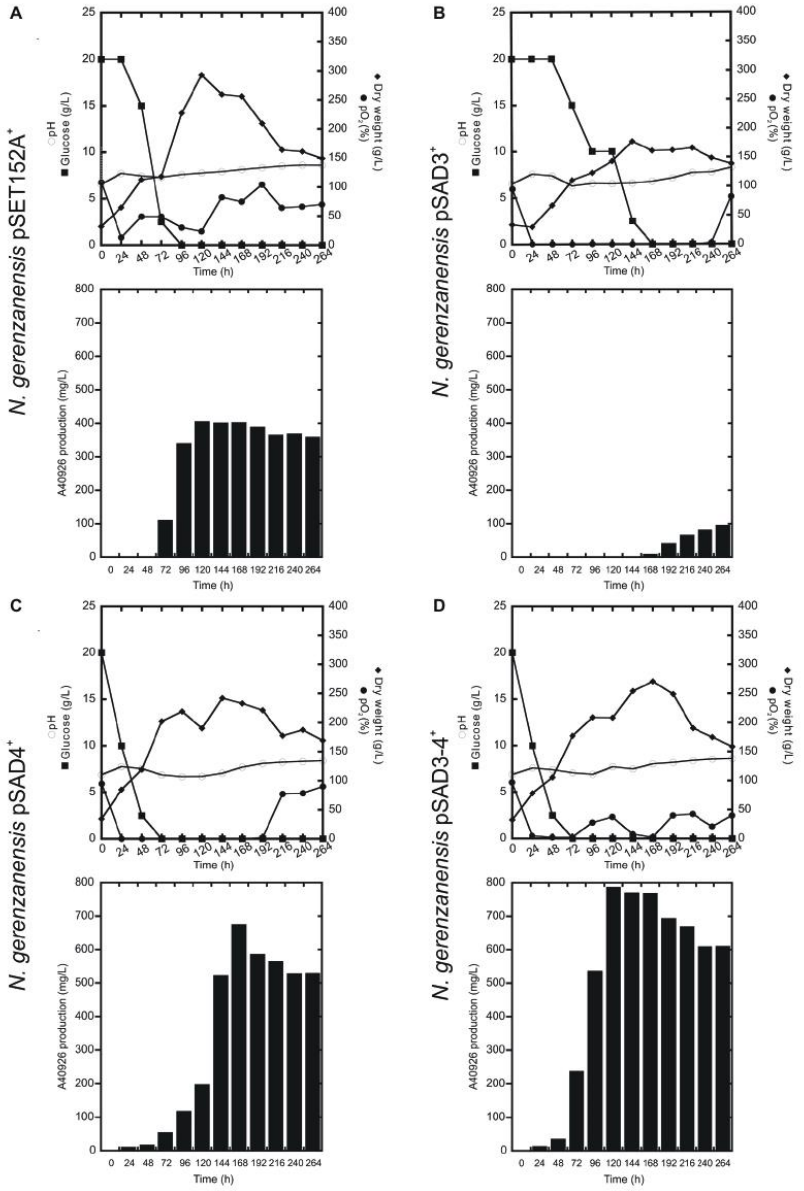


FIGURE 5 | Time courses of *N. gerenzanensis* pSET152A+ **(A)** and recombinant strains, overexpressing *dbv3* **(B)**, *dbv4* **(C)** and co-expressing both genes **(D)** cultivated in 2 l FM2 industrial medium in a 3 l fermenter. Glucose consumption (filled rectangles), biomass accumulation (filled rhombi), pH (circles), O_2 level (filled circles), and A40926 production were monitored every 24 h.

starting very late (after 168 h of cultivation) and never exceeding 100 mg/l (**Figure 5B**). Microscopical observation of pSAD3+ strain showed that the mycelium was highly fragmented (data not shown), suggesting some kind of physiological stress resulting from *dbv3* overexpression. Some fragmentation of the pSAD3+ strain (as well as of the pSAR1 strain carrying the *dbv3*-like

nocRI from *N. coxensis*) was also observed in flask culture, in contrast to the other strains which produced dense mycelial pellets; the fragmentation was less pronounced, indicating that scaling up in the bioreactor dramatically enhanced this effect. Interestingly, in the best performer at the bioreactor scale – the pSAD3–4^{+} strain – *dbv4* was expressed from the strong constitutive *aac(3)IVp* but *dbv3* was left under its endogenous promoter, suggesting that the balance between the expression level of the two cluster-situated regulatory genes is important for optimal improvement of A40926 production.

DISCUSSION

Apart from some reports on cultivating and manipulating the industrially valuable A40926 producer *N. gerenzanensis* (Stinchi et al., 2003, 2006; Marcone et al., 2010a,b,c; Alt et al., 2019) and the kistamicin producer *Nonomuraea* sp. ATCC 55076 (Greule et al., 2019), we are not aware of any other attempt to develop genetic tools for manipulating *Nonomuraea* spp. However, some other glycopeptide producers, like *A. teichomyceticus*, already possess well-developed toolkits for genetic manipulation which has greatly simplified investigations in these strains (Horbal et al., 2013; Yushchuk et al., 2016). Therefore, our first goal in this work was to develop a set of genetic tools for manipulating diverse species of *Nonomuraea*. To do this, we decided to work in parallel with the better-known A40926 producer, recently re-classified as *N. gerenzanensis* (Dalmastri et al., 2016), and with the little investigated *N. coxensis*, which was isolated in Bangladesh in 2007 (Ara et al., 2007). Although the available *N. coxensis* genome sequence is still incomplete, we could identify three contigs covering most of a *dbv*-like gene cluster including the *dbv*-like cluster-situated regulatory genes, which we named *nocRI* (*dbv3* homolog) and *nocRII* (*dbv4* homolog). The next step of our work will be additional sequencing to yield a properly annotated *N. coxensis* genome. Interestingly, during the course of our investigations, Waglechner et al. (2019), systematically screening the available sequences in genomic databases, also reported the presence of a BGC encoding for a A40926-like GPA in the genome of *N. coxensis*. In addition, the same authors reported that a newly isolated *Nonomuraea* sp. WAC01424 possess a BGC which could produce another A40926-related compound (Waglechner et al., 2019). Our preliminary analysis of this BGC suggests that it could be a sulfated A40926-like GPA, lacking the aliphatic side chains.

In the meantime, we tested both in *N. gerenzanensis* and *N. coxensis*, a set of heterologous and native promoters (the latter derived from a set of house-keeping genes in *N. gerenzanensis*) with the final goal of using them for driving gene expression in these strains. Besides the practical outcome of this screening (all of the generated promoter-probe vectors could be easily used as expression vectors simply by exchanging *gusA* for a gene of interest, offering a set of variable tools for gene expression), it is interesting to observe that the studied promoters had similar strengths in the two phylogenetically distant *Nonomuraea* species and that the native promoters from *N. gerenzanensis* worked similarly in *N. coxensis*. Consistently, we detected a comparable difference in the two strains between the strength of promoters driving the expression of rifamycin-sensitive and rifamycin-resistant *rpoB* alleles of *N. gerenzanensis* ($rpoB_{ng}p$ and $rpoB_{Rng}p$). Since we found the two *rpoB* alleles in *N. coxensis* draft genome (on a short genomic scaffolds KB904038 and KB904038), we might suppose that their manipulation could improve antibiotic production in this strain as already reported for *N. gerenzanensis* (Vigliotta et al., 2004). More generally, these vectors should be useful for the genetic manipulation of other members of the genus *Nonomuraea*, which have the potential to produce novel valuable specialized metabolites (Sungthong and Nakaew, 2015; Nazari et al., 2017).

As in the case of *Actinoplanes* spp. (Horbal et al., 2013), the strongest heterologous promoter was *aac(3)IVp*, although some *N. gerenzanensis* native promoters like $hrdB_{ng}p$ and $rpsL_{ng}p$ appeared to have comparable strength and merit further investigations. When we used *aac(3)IVp* to overexpress the cluster-situated regulatory genes *dbv3* and *dbv4* from the *N. gerenzanensis dbv* gene cluster and *nocRI* (*dbv3*-like) from the A40926-like BGC of *N. coxensis*, the recombinant *N. gerenzanensis* strains produced significantly more A40926 than the parental strain. The evidence that the heterologous expression of *nocRI* increased A40926 production in *N. gerenzanensis* confirmed its role in regulating the expression of a A40926-like BGC in *N. coxensis*. Additionally, it represents another case of cross-talk between regulators controlling GPA BGCs in producing actinomycetes (Spohn et al., 2014). Our next goal will be to investigate if and how (in which cultivation conditions) *N. coxensis* produces A40926 or a A40926-like molecule.

Previous work (Lo Grasso et al., 2015) reported that overexpression of *dbv3* (under the control of the thiostrepton-inducible *tipA** promoter in the integrative plasmid pIJ8600) in *N. gerenzanensis* increased A40926 production from 13 to 27 mg/l using the laboratory medium R3 (Lo Grasso et al., 2015). In this paper, we tested the real industrial potential of overexpressing not only *dbv3*, but also *dbv4*, and *dbv3* and *dbv4* together, cloning them under the strong constitutive *aac(3)IVp* promoter, scaling up their cultivation at bioreactor scale and using a previously optimized industrial medium where A40926 is produced in hundreds of milligrams per liter (Marcone et al., 2010a, 2014). At the bioreactor level, where strains could have a different performance from the flask-cultivation due to different mixing and mass transfer rates of nutrients and oxygen, the best performer was the strain carrying both *dbv4* under *aac(3)IVp* and *dbv3* under its own endogenous promoter. This strain produced nearly 800 mg/l of A40926, which is twice that of the parental strain grown under the same conditions. Conversely, the strain carrying only *dbv3* expressed from *aac(3)IVp* showed reduced production capacity and an altered phenotype particularly after scaling up from flask to bioreactor; in contrast, the *dbv4* overexpressing recombinant grew similarly under both conditions. It is widely recognized that any potentially higher producing mutant or derivative needs to be validated in a bioreactor-scale fermentation since unpredictable discrepancies

in strain performance can occur during scaling up from flask culture (Lee and Kim, 2015). In the case of the *dbv3* overexpressing recombinant, we believe that the fragmented mycelial phenotype, which was much more apparent in the bioreactor, could be a specific consequence of the overexpression of the *dbv7* gene encoding a D,D-carboxypeptidase known as VanYn (Binda et al., 2012). The level of VanYn activity in cell extracts from the *dbv3*-carrying recombinant cultivated at bioreactor scale was found to be much higher than in the parental strain (unpublished data). Consistent with this, overexpression of VanYn altered the mycelial morphology in *N. gerenzanensis* as well as in heterologous hosts such as streptomycetes strains (Binda et al., 2013).

In conclusion, only a few GPAs are used in clinical practice and those produced by semi-synthesis from natural products, such as dalbavancin derived from A40926, are still quite expensive. Dalbavancin is the first antibiotic designated as a Qualified Infection Diseases Product by the FDA because of its potency, extended dosing interval, and unique dose regimen, but its cost largely exceeds that of first-generation GPAs and consequently its use in hospitals is still limited (Chiasson and White, 2016; Agarwal et al., 2018). Improving A40926-producing strains might lead to a decrease in the cost of dalbavancin. As demonstrated in this paper, A40926 production could be significantly enhanced by manipulating the expression of *dbv* cluster-situated regulators. An important and often-neglected aspect is testing the genetic stability and productivity of the selected recombinant strains in a fully developed industrial process at bioreactor level, which mimics the conditions of antibiotic large scale production. We were able to demonstrate here that the improvements we made to A40926 production levels in shake flasks were also achieved in the bioreactor, indicating the relevance of this approach to industrial-scale strain improvement.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

OY, EB, MB, and FM conceived and designed the experiments and wrote the paper. OY, AA-V, GM, and EB performed the experiments. OY, GM, and EB analyzed the data.

FUNDING

This work was supported by public grants "Fondo di Ateneo per la Ricerca" 2018 and 2019 to FM and GM and by Federation of European Microbiological Societies (FEMS) Research Fellowship 2015 to EB.

ACKNOWLEDGMENTS

Consorzio Interuniversitario per le Biotecnologie is acknowledged for its support to traveling and attending congresses for EB and GM.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00008/full#supplementary-material

REFERENCES

Adamek, M., Alanjary, M., Sales-Ortells, H., Goodfellow, M., Bull, A. T., Winkler, A., et al. (2018). Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in *Amycolatopsis* species. *BMC Genomics* 19, 1–15. doi: 10.1186/s12864-018-4809-4

Agarwal, R., Bartsch, S., Kelly, B., Prewitt, M., Liu, Y., Chen, Y., et al. (2018). Newer glycopeptide antibiotics for treatment of complicated skin and soft tissue infections: systematic review, network meta-analysis and cost analysis. *Clin. Microbiol. Infect.* 24, 361–368. doi: 10.1016/j.cmi.2017.08.028

Alduina, R., Sosio, M., and Donadio, S. (2018). Complex regulatory networks governing production of the glycopeptide A40926. *Antibiotics* 7, pii: E30. doi: 10.3390/antibiotics7020030

Alt, S., Bernasconi, A., Sosio, M., Brunati, C., Donadio, S., and Maffioli, S. I. (2019). Toward single-peak dalbavancin analogs through biology and chemistry. *ACS Chem. Biol.* 14, 356–360. doi: 10.1021/acschembio.9b00050

Altschul, S. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Ara, I., Kudo, T., Matsumoto, A., Takahashi, Y., and Omura, S. (2007). *Nonomuraea bangladeshensis* sp. nov. and *Nonomuraea coxensis* sp. nov. *Int. J. Syst. Evol. Microbiol.* 57, 1504–1509. doi: 10.1099/ijs.0.65054-0

Armenia, I., Marcone, G. L., Berini, F., Orlandi, V. T., Pirrone, C., Martegani, E., et al. (2018). Magnetic nanoconjugated teicoplanin: a novel tool for bacterial infection site targeting. *Front. Microbiol.* 9, 1–17. doi: 10.3389/fmicb.2018.02270

Banik, J. J., and Brady, S. F. (2008). Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. *Proc. Natl. Acad. Sci.* 105, 17273–17277. doi: 10.1073/pnas.0807564105

Banik, J. J., Craig, J. W., Calle, P. Y., and Brady, S. F. (2010). Tailoring enzyme-rich environmental DNA clones: a source of enzymes for generating libraries of unnatural natural products. *J. Am. Chem. Soc.* 132, 15661–15670. doi: 10.1021/ja105825a

Bibb, M. J. (2013). Understanding and manipulating antibiotic production in actinomycetes. *Biochem. Soc. Trans.* 41, 1355–1364. doi: 10.1042/BST20130214

Binda, E., Marcone, G. L., Berini, F., Pollegioni, L., and Marinelli, F. (2013). *Streptomyces* spp. as efficient expression system for a D,D-peptidase/D,D-carboxypeptidase involved in glycopeptide antibiotic resistance. *BMC Biotechnol.* 13, 1–14. doi: 10.1186/1472-6750-13-24

Binda, E., Marcone, G. L., Pollegioni, L., and Marinelli, F. (2012). Characterization of VanYn, a novel D,D-peptidase/D,D-carboxypeptidase involved in glycopeptide antibiotic resistance in *Nonomuraea* sp. ATCC 39727. *FEBS J.* 279, 3203–3213. doi: 10.1111/j.1742-4658.2012.08706.x

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019). AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81–W87. doi: 10.1093/nar/gkz310

Chiasson, J., and White, R. (2016). Comparison of dalbavancin, oritavancin and vancomycin pharmacodynamics using clinical targets against ABSSSI pathogens. *Open Forum Infect. Dis.* 3:S515. doi: 10.1093/ofid/ofw172.1506

Dalmastri, C., Gastaldo, L., Marcone, G. L., Binda, E., Congiu, T., and Marinelli, F. (2016). Classification of *Nonomuraea* sp. ATCC 39727, an actinomycete that produces the glycopeptide antibiotic A40926, as *Nonomuraea gerenzanensis* sp. nov. *Int. J. Syst. Evol. Microbiol.* 66, 912–921. doi: 10.1099/ijsem.0.000810

D'Argenio, V., Petrillo, M., Pasanisi, D., Pagliarulo, C., Colicchio, R., Talà, A., et al. (2016). The complete 12 Mb genome and transcriptome of *Nonomuraea gerenzanensis* with new insights into its duplicated "magic" RNA polymerase. *Sci. Rep.* 6, 1–13. doi: 10.1038/s41598-016-0025-0

Greule, A., Izoré, T., Iftime, D., Tailhades, J., Schoppet, M., Zhao, Y., et al. (2019). Kistamicin biosynthesis reveals the biosynthetic requirements for production of highly crosslinked glycopeptide antibiotics. *Nat. Commun.* 10, 1–15. doi: 10.1038/s41467-019-10384-w

Haslinger, K., Peschke, M., Brieke, C., Maximowitsch, E., and Cryle, M. J. (2015). X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature* 521, 105–109. doi: 10.1038/nature14141

Herrmann, S., Siegl, T., Luzhetska, M., Petzke, L., Jilg, C., Welle, E., et al. (2012). Site-specific recombination strategies for engineering actinomycete genomes. *Appl. Environ. Microbiol.* 78, 1804–1812. doi: 10.1128/AEM.06054-11

Horbal, L., Fedorenko, V., and Luzhetskyy, A. (2014a). Novel and tightly regulated resorcinol and cumate-inducible expression systems for *Streptomyces* and other actinobacteria. *Appl. Microbiol. Biotechnol.* 98, 8641–8655. doi: 10.1007/s00253-014-5918-x

Horbal, L., Kobyłyanskyy, A., Truman, A. W., Zaburranyi, N., Ostash, B., Luzhetskyy, A., et al. (2014b). The pathway-specific regulatory genes, tei15* and tei16*, are the master switches of teicoplanin production in *Actinoplanes teichomyceticus*. *Appl. Microbiol. Biotechnol.* 98, 9295–9309. doi: 10.1007/s00253-014-5969-z

Horbal, L., Kobyłyanskyy, A., Yushchuk, O., Zaburannyi, N., Luzhetskyy, A., Ostash, B., et al. (2013). Evaluation of heterologous promoters for genetic analysis of *Actinoplanes teichomyceticus* – producer of teicoplanin, drug of last defense. *J. Biotechnol.* 168, 367–372. doi: 10.1016/j.jbiotec.2013.10.018

Horbal, L., Zaburannyy, N., Ostash, B., Shulga, S., and Fedorenko, V. (2012). Manipulating the regulatory genes for teicoplanin production in *Actinoplanes teichomyceticus*. *World J. Microbiol. Biotechnol.* 28, 2095–2100. doi: 10.1007/s11274-012-1013-6

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000). *Practical streptomyces genetics*. Norwich: John Innes Foundation.

Koshla, O., Yushchuk, O., Ostash, I., Dacyuk, Y., Myronovskyi, M., Jäger, G., et al. (2019). Gene *miaA* for post-transcriptional modification of tRNA XXA is important for morphological and metabolic differentiation in *Streptomyces*. *Mol. Microbiol.* 112, 249–265. doi: 10.1111/mmi.14266

Kusserow, K., and Gulder, T. A. M. (2017). Complete genome sequence of *Actinomadura parvosata* subsp. *kistnae*, a rich source of novel natural product (bio-)chemistry. *J. Genomics* 5, 75–76. doi: 10.7150/jgen.19673

Lee, S. Y., and Kim, H. U. (2015). Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* 33, 1061–1072. doi: 10.1038/nbt.3365

Lo Grasso, L. L., Maffioli, S., Sosio, M., Bibb, M., Puglia, A. M., and Alduina, R. (2015). Two master switch regulators trigger A40926 biosynthesis in *Nonomuraea* sp. strain ATCC 39727. *J. Bacteriol.* 197, 2536–2544. doi: 10.1128/JB.00262-15

Makitrynskyy, R., Ostash, B., Tsypik, O., Rebets, Y., Doud, E., Meredith, T., et al. (2013). Pleiotropic regulatory genes *bldA*, *adpA* and *absB* are implicated in production of phosphoglycolipid antibiotic moenomycin. *Open Biol.* 3:130121. doi: 10.1098/rsob.130121

Marcone, G. L., Beltrametti, F., Binda, E., Carrano, L., Foulston, L., Hesketh, A., et al. (2010a). Novel mechanism of glycopeptide resistance in the A40926 producer *Nonomuraea* sp. ATCC 39727. *Antimicrob. Agents Chemother.* 54, 2465–2472. doi: 10.1128/AAC.00106-10

Marcone, G. L., Binda, E., Berini, F., and Marinelli, F. (2018). Old and new glycopeptide antibiotics: from product to gene and back in the post-genomic era. *Biotechnol. Adv.* 36, 534–554. doi: 10.1016/j.biotechadv.2018.02.009

Marcone, G. L., Binda, E., Carrano, L., Bibb, M., and Marinelli, F. (2014). Relationship between glycopeptide production and resistance in the actinomycete *Nonomuraea* sp. ATCC 39727. *Antimicrob. Agents Chemother.* 58, 5191–5201. doi: 10.1128/AAC.02626-14

Marcone, G. L., Carrano, L., Marinelli, F., and Beltrametti, F. (2010b). Protoplast formationa and reversion to normal growth in antibiotic-producing uncommon actinomycetes. *J. Antibiot.* 63, 83–88. doi: 10.1038/ja.2009.127

Marcone, G. L., Foulston, L., Binda, E., Marinelli, F., Bibb, M., and Beltrametti, F. (2010c). Methods for the genetic manipulation of *Nonomuraea* sp. ATCC 39727. *J. Ind. Microbiol. Biotechnol.* 37, 1097–1103. doi: 10.1007/s10295-010-0807-5

Myronovskyi, M., Welle, E., Fedorenko, V., and Luzhetskyy, A. (2011). β-Glucuronidase as a sensitive and versatile reporter in actinomycetes. *Appl. Environ. Microbiol.* 77, 5370–5383. doi: 10.1128/AEM.00434-11

Nazari, B., Forneris, C. C., Gibson, M. I., Moon, K., Schramma, K. R., and Seyedsayamdost, M. R. (2017). *Nonomuraea* sp. ATCC 55076 harbours the largest actinomycete chromosome to date and the kistamicin biosynthetic gene cluster. *MedChemComm* 8, 780–788. doi: 10.1039/C6MD00637J

Okano, A., Isley, N. A., and Boger, D. L. (2017). Peripheral modifications of [Ψ[CH_2NH]Tpg4]vancomycin with added synergistic mechanisms of action provide durable and potent antibiotics. *Proc. Natl. Acad. Sci.* 114, E5052–E5061. doi: 10.1073/pnas.1704125114

Peschke, M., Brieke, C., Goode, R. J. A., Schittenhelm, R. B., and Cryle, M. J. (2017). Chlorinated glycopeptide antibiotic peptide precursors improve cytochrome P450-catalyzed cyclization cascade efficiency. *Biochemistry* 56, 1239–1247. doi: 10.1021/acs.biochem.6b01102

Schoppet, M., Peschke, M., Kirchberg, A., Wiebach, V., Süssmuth, R. D., Stegmann, E., et al. (2019). The biosynthetic implications of late-stage condensation domain selectivity during glycopeptide antibiotic biosynthesis. *Chem. Sci.* 10, 118–133. doi: 10.1039/C8SC03530J

Shawky, R. M., Puk, O., Wietzorrek, A., Pelzer, S., Takano, E., Wohlleben, W., et al. (2007). The border sequence of the balhimycin biosynthesis gene cluster from *Amycolatopsis balhimycina* contains *bbr*, encoding a StrR-like pathway-specific regulator. *J. Mol. Microbiol. Biotechnol.* 13, 76–88. doi: 10.1159/000103599

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol. Syst. Biol.* 7, 1–6. doi: 10.1038/msb.2011.75

Spohn, M., Kirchner, N., Kulik, A., Jochim, A., Wolf, F., Muenzer, P., et al. (2014). Overproduction of ristomycin a by activation of a silent gene cluster in *Amycolatopsis japonicum* MG417-CF17. *Antimicrob. Agents Chemother.* 58, 6185–6196. doi: 10.1128/AAC.03512-14

Stegmann, E., Albersmeier, A., Spohn, M., Gert, H., Weber, T., Wohlleben, W., et al. (2014). Complete genome sequence of the actinobacterium *Amycolatopsis japonica* MG417-CF17T (=DSM 44213T) producing (S,S)-N,N′-ethylenediaminedisuccinic acid. *J. Biotechnol.* 189, 46–47. doi: 10.1016/j.jbiotec.2014.08.034

Stinchi, S., Azimonti, S., Donadio, S., and Sosio, M. (2003). A gene transfer system for the glycopeptide producer *Nonomuraea* sp. ATCC39727. *FEMS Microbiol. Lett.* 225, 53–57. doi: 10.1016/S0378-1097(03)00490-7

Stinchi, S., Carrano, L., Lazzarini, A., Feroggio, M., Grigoletto, A., Sosio, M., et al. (2006). A derivative of the glycopeptide A40926 produced by inactivation of the β-hydroxylase gene in *Nonomuraea* sp. ATCC39727. *FEMS Microbiol. Lett.* 256, 229–235. doi: 10.1111/j.1574-6968.2006.00120.x

Sungthong, R., and Nakaew, N. (2015). The genus *Nonomuraea*: a review of a rare actinomycete taxon for novel metabolites. *J. Basic Microbiol.* 55, 554–565. doi: 10.1002/jobm.201300691

Thaker, M. N., Wang, W., Spanogiannopoulos, P., Waglechner, N., King, A. M., Medina, R., et al. (2013). Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat. Biotechnol.* 31, 922–927. doi: 10.1038/nbt.2685

Truman, A. W., Kwun, M. J., Cheng, J., Yang, S. H., Suh, J.-W., and Hong, H.-J. (2014). Antibiotic resistance mechanisms inform discovery: identification and characterization of a novel *Amycolatopsis* strain producing ristocetin. *Antimicrob. Agents Chemother.* 58, 5687–5695. doi: 10.1128/AAC.03349-14

Vigliotta, G., Tredici, S. M., Damiano, F., Montinaro, M. R., Pulimeno, R., Summa, R. D., et al. (2004). Natural merodiploidy involving duplicated *rpoB* alleles affects secondary metabolism in a producer actinomycete. *Mol. Microbiol.* 55, 396–412. doi: 10.1111/j.1365-2958.2004.04406.x

Vongsangnak, W., Figueiredo, L. F., Förster, J., Weber, T., Thykaer, J., Stegmann, E., et al. (2012). Genome-scale metabolic representation of *Amycolatopsis balhimycina*. *Biotechnol. Bioeng.* 109, 1798–1807. doi: 10.1002/bit.24436

Waglechner, N., Mcarthur, A. G., and Wright, G. D. (2019). Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nat. Microbiol.* 4, 1862–1871. doi: 10.1038/s41564-019-0531-5

Wu, Z.-C., and Boger, D. L. (2019). Exploration of the site-specific nature and generalizability of a trimethylammonium salt modification on vancomycin: A-ring derivatives. *Tetrahedron* 75, 3160–3165. doi: 10.1016/j.tet.2019.02.008

Yim, G., Kalan, L., Koteva, K., Thaker, M. N., Waglechner, N., Tang, I., et al. (2014). Harnessing the synthetic capabilities of glycopeptide antibiotic tailoring enzymes: characterization of the UK-68,597 biosynthetic cluster. *ChemBioChem* 15, 2613–2623. doi: 10.1002/cbic.201402179

Yushchuk, O., Homoniuk, V., Datsiuk, Y., Ostash, B., Marinelli, F., and Fedorenko, V. (2020a). Development of a gene expression system for the uncommon actinomycete *Actinoplanes rectilineatus* NRRL B-16090. *J. Appl. Genet.* doi: 10.1007/s13353-019-00534-7 [Epub ahead of print].

Yushchuk, O., Homoniuk, V., Ostash, B., Marinelli, F., and Fedorenko, V. (2020b). Genetic insights into the mechanism of teicoplanin self-resistance in *Actinoplanes teichomyceticus*. *J. Antibiot.* (in press). doi: 10.1038/s41429-019-0274-9

Yushchuk, O., Horbal, L., Ostash, B., Marinelli, F., Wohlleben, W., Stegmann, E., et al. (2019). Regulation of teicoplanin biosynthesis: refining the roles of *tei* cluster-situated regulatory genes. *Appl. Microbiol. Biotechnol.* 103, 4089–4102. doi: 10.1007/s00253-019-09789-w

Yushchuk, O., Ostash, B., Pham, T. H., Luzhetskyy, A., Fedorenko, V., Truman, A. W., et al. (2016). Characterization of the post-assembly line tailoring processes in teicoplanin biosynthesis. *ACS Chem. Biol.* 11, 2254–2264. doi: 10.1021/acschembio.6b00018

Conflict of Interest: The 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.

frontiers

Supplementary Material

1 Supplementary Data

Compositions of media used in the work:

Unless otherwise stated, components are from Sigma-Aldrich, St. Louis, MO, United States.

ISP2 (g/l of distilled water):
Yeast extract – 4;
Malt extract – 10;
Dextrose – 4;
Agar – 20;
pH 7.5.
ISP3 (g/l of tap water):
Fine ground whole oats (Kozub, Poltava, Ukraine) – 34;
Agar – 20;
pH 7.5.
VSP (g/l of distilled water):
Soluble starch (Difco, Franklin Lakes, NJ, United States) – 24;
Dextrose – 1;
Meat extract – 3;
Yeast extract – 5;
Tryptose – 5;
L-proline – 0.5;
Sucrose – 50;
pH 7.5.
VM0.1 (g/l of distilled water):
Soluble starch (Difco) – 2.4;
Dextrose – 0.1;
Meat extract – 0.3;
Yeast extract – 0.5;
Tryptose – 0.5;
Agar – 20;
pH 7.2.
E26 (g/l of distilled water):
Dextrose – 25;
Soy flour – 20;
Yeast extract – 4;
NaCl – 1.25;
$CaCO_3$ – 5;

pH 7.5.
FM2 (g/l of distilled water):
Dextrose – 30;
Soy flour – 30;
Yeast extract – 8;
Malt extract – 15;
$CaCO_3$ – 5;
L-valine – 1;
pH 7.5.
SFM (g/l of distilled water):
Mannitol – 20;
Soy flour – 20;
Agar – 20;
pH 7.5.

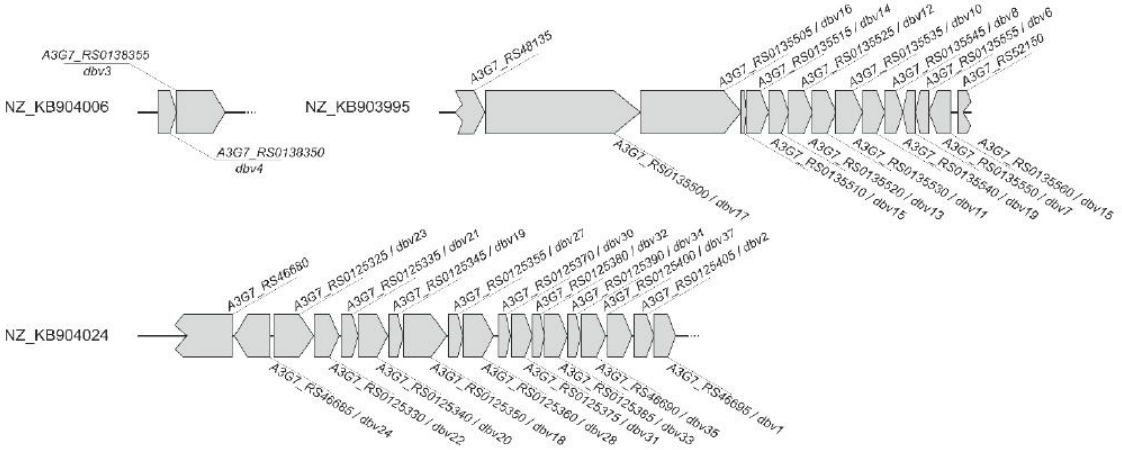
2 Supplementary Figures and Tables

2.1 Supplementary Figures

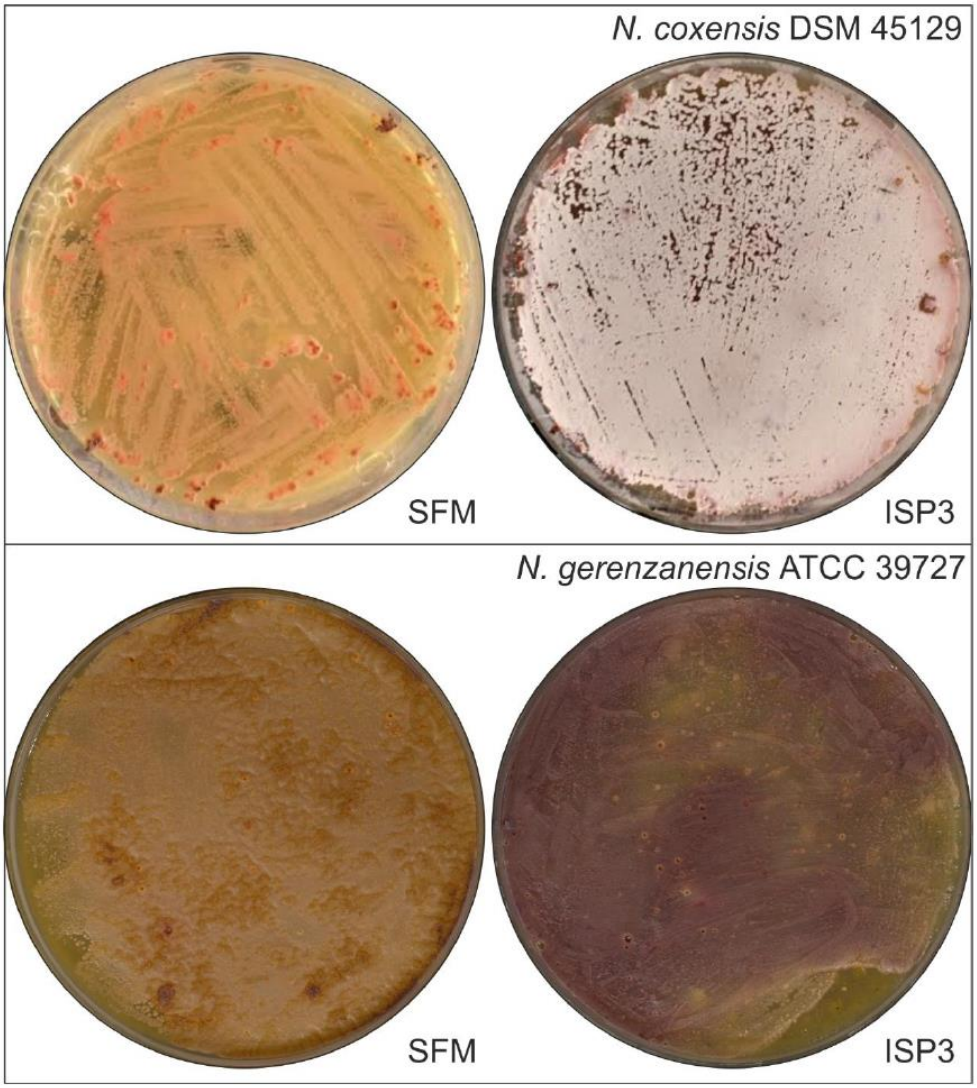
Supplementary Figure 1. Alignment of the amino acid sequences of Dbv3 and its orthologue from *Nonomuraea coxensis* – NocRI. Alignment was generated using ClustalOmega (EMBL-EBI).

```
NocRII      VDPTGVDIVALPVVEIELSRLSSVYSPRTSGEDPEHVETLLSAQGELPPILVHRPTMRVI        60
Dbv4        VDPTGVDIATLPVVEIELSRLSSVYSPRTSGEDPEHVETLLSAQGELPPILVHRPTMRVI        60
            ********.:**************************************************

NocRII      DGLHRLRVARVRGETKIAVRLIDGTESDAFVLAVEANVRHGLPLSLADRKRAAVQIIGTH        120
Dbv4        DGLHRLKVARVRGETTISVRLIDGTESDAFVLAVEANVRHGLPLSLADRKRAAVRIIGTH        120
            ******:********.*:***********************************:*****

NocRII      PQWSDRRVASATGISAGTVADLRKRRGQDGDEARIGRDGRIRPVDSSEGRRLAAEIIRSH        180
Dbv4        PQWSDRRVASATGISAGTVADLRRRRGQGGDEARIGRDGRIRPVDSSEGRRLAAELIRSH        180
            **********************:****.*************************:****

NocRII      PDLSLRQVAKQVGISPETVRDVRGRLEQGESPIPDGSRRLRAKPESLRRPEQDFGHAGGR        240
Dbv4        PDLSLRQVAKQVGISPETVRDVRGRLEHGESPIPDGSRRLRTKPELLRRAEQDFGHVDGR        240
            **************************:************:*** *** ******..**

NocRII      DRQAVLERLKADPALRLTETGRILLRMLSLHSIDGQEWERILRGVPPHWDAVVARCARDH        300
Dbv4        DRQAVLERLKADPALRLTETGRILLRMLSLHSIDGQEWERILRGVPPHWGTVVARCARDH        300
            ***********************************************.:*********

NocRII      AQIWAAFADRLEGRATDLAAG       321
Dbv4        AQIWAAFADRLEGRATDLAAG       321
            *********************
```

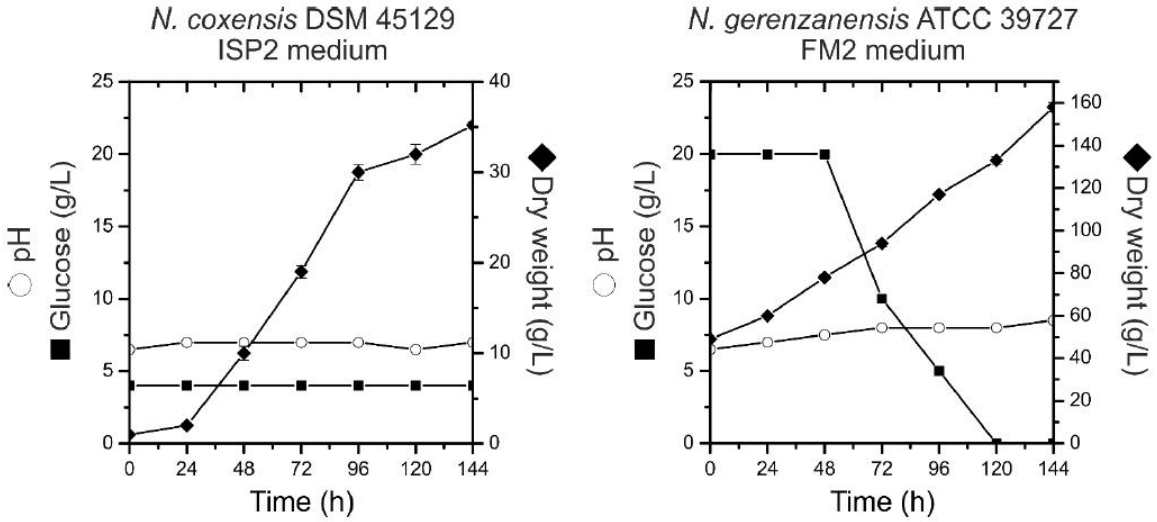
Supplementary Figure 2. Alignment of the amino acid sequences of Dbv4 and its orthologue from *N. coxensis* – NocRII. Alignment was generated using ClustalOmega (EMBL-EBI).



Supplementary Figure 3. Scheme of the parts of putative GPA BGC from *N. coxensis* that could be found on three contigs available in GenBank. Locus tags for individual genes are given together with the names of putative orthologues from *dbv* cluster.



Supplementary Figure 4. 7-day-growth of *N. coxensis* DSM 45129 and *Nonomuraea gerenzanensis* ATCC 39727 on SFM and ISP3 media. Both strains did not differentiate spores on SFM, which indeed promotes sporulation of *Streptomyces* spp. and of other uncommon actinobacteria like *Actinoplanes* spp. (Kieser et al., 2000; Gren et al., 2016). Only *N. coxensis* produces abundant spores on ISP3 in these conditions. *N. gerenzanensis* was reported to produce very few spores only after 14-21 days of cultivation on ISP3 (Dalmastri et al., 2016).



Supplementary Figure 5. Growth curves of *N. coxensis* DSM 45129 and *N. gerenzanensis* ATCC 39727 in ISP2 and FM2 liquid media, respectively. Glucose consumption (rectangles), biomass accumulation (rhombi) and pH (circles) were monitored every 24 h. Results given are mean values of three independent experiments. Error bars represent standard deviations.

3 Supplementary references

Dalmastri, C., Gastaldo, L., Marcone, G. L., Binda, E., Congiu, T., and Marinelli, F. (2016). Classification of *Nonomuraea* sp. ATCC 39727, an actinomycete that produces the glycopeptide antibiotic A40926, as *Nonomuraea gerenzanensis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 66, 912–921. doi:10.1099/ijsem.0.000810.

Gren T., Ortseifen. V., Wibberg, D., Schneiker-Bekel, S., Bednarz, H., Niehaus, K., Zemke, T., Persicke, M., Pühler, A., Kalinowski, J., (2016). Genetic engineering in *Actinoplanes* sp. SE50/110 – development of an intergeneric conjugation system for the introduction of actinophage-based integrative vectors. *Journal of Biotechnology* 232, 79-88. doi: 10.1016/j.jbiotec.2016.05.012.

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000). *Practical Streptomyces Genetics.* Norwich: John Innes Foundation.

CHAPTER 2:

Genomic-led discovery of a novel glycopeptide antibiotic by *Nonomuraea coxensis* DSM 45129

ACS chemical biology

pubs.acs.org/acschemicalbiology

Articles

Genomic-Led Discovery of a Novel Glycopeptide Antibiotic by *Nonomuraea coxensis* DSM 45129

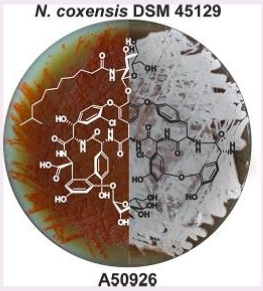
Oleksandr Yushchuk, Natalia M. Vior, Andres Andreo-Vidal, Francesca Berini, Christian Rückert, Tobias Busche, Elisa Binda, Jörn Kalinowski, Andrew W. Truman,* and Flavia Marinelli*

Cite This: *ACS Chem. Biol.* 2021, 16, 915–928

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Glycopeptide antibiotics (GPAs) are last defense line drugs against multidrug-resistant Gram-positive pathogens. Natural GPAs teicoplanin and vancomycin, as well as semisynthetic oritavancin, telavancin, and dalbavancin, are currently approved for clinical use. Although these antibiotics remain efficient, emergence of novel GPA-resistant pathogens is a question of time. Therefore, it is important to investigate the natural variety of GPAs coming from so-called "rare" actinobacteria. Herein we describe a novel GPA producer—*Nonomuraea coxensis* DSM 45129. Its *de novo* sequenced and completely assembled genome harbors a biosynthetic gene cluster (BGC) similar to the *dbv* BGC of A40926, the natural precursor to dalbavancin. The strain produces a novel GPA, which we propose is an A40926 analogue lacking the carboxyl group on the *N*-acylglucosamine moiety. This structural difference correlates with the absence of *dbv29*—coding for an enzyme responsible for the oxidation of the *N*-acylglucosamine moiety. Introduction of *dbv29* into *N. coxensis* led to A40926 production in this strain. Finally, we successfully applied *dbv3* and *dbv4* heterologous transcriptional regulators to trigger and improve A50926 production in *N. coxensis*, making them prospective tools for screening other *Nonomuraea* spp. for GPA production. Our work highlights genus *Nonomuraea* as a still untapped source of novel GPAs.



1. INTRODUCTION

Nonomuraea is a genus of so-called "rare" actinomycetes whose potential to produce specialized (secondary) metabolites is still rather poorly explored.[1,2] Recently sequenced genomes of *Nonomuraea* species appear to be generally larger than the reference *Streptomyces* ones. The mean genome size of *Nonomuraea* (based on the three available complete assemblies[2,3]) is around 12 Mbp, whereas the mean genome size of *Streptomyces* (calculated on 251 fully assembled genomes available in GenBank) equals 8.6 Mbp. The larger genomes of *Nonomuraea* spp. encode dozens of putative biosynthetic gene clusters (BGCs).[2−4] *Nonomuraea* spp. were initially found to be recalcitrant to commonly used genetic engineering manipulations, but new tools are now being developed for this genus.[5−7] This paves the way for unravelling the huge hidden biosynthetic potential of these organisms.

Probably the most important bioactive metabolite produced by a *Nonomuraea* species is the type IV[8] glycopeptide antibiotic (GPA) A40926[9] (Figure 1) produced by *Nonomuraea gerenzanensis* ATCC 39727. Like other GPAs, A40926 acts as a selective and potent inhibitor of cell-wall biosynthesis in Gram-positive bacteria. A40926 is structurally related to the clinically relevant GPA teicoplanin (Figure 1), produced by *Actinoplanes teichomyceticus* ATCC 31121[10,11] and to ristocetin (Figure 1), previously isolated from numerous *Amycolatopsis* spp. (i.e., *A. lurida* NRRL 2430, *A. japonicum* MG417-CF17, and *Amycolatopsis* sp. MJM2582).[12−14] Like teicoplanin, A40926 is produced as a mixture of related compounds (major components are A40926 B and A40926 A factors), which differ in the length and branching of an aliphatic side chain (Figure 1). It was recently clarified that *N. gerenzanensis* produces the GPA in the form of *O*-acetyl-A40926 (with an *O*-acetylated mannose residue), but the acetyl group is lost during the alkaline extraction of the antibiotic.[15,16] Since it was this deacetylated GPA that was initially named A40926, we will refer to it as A40926 hereafter.

A40926 is the precursor of the second-generation semisynthetic GPA dalbavancin (Figure 1), which is currently applied in clinics to treat severe infections caused by multidrug-resistant Gram-positive pathogens.[17] Dalbavancin (marketed in Europe and USA under the trade names xydalba and dalvance, respectively) is the first antibiotic designated as a qualified infectious disease product by FDA because of its potency, extended dosing interval, and unique dose regimen (once-a-week), but its cost still largely exceeds that of first-generation GPAs.[10] Therefore, improvement of A40926 production by recombinant engineering of *N. gerenzanensis*

Received: March 8, 2021
Accepted: April 14, 2021
Published: April 29, 2021

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

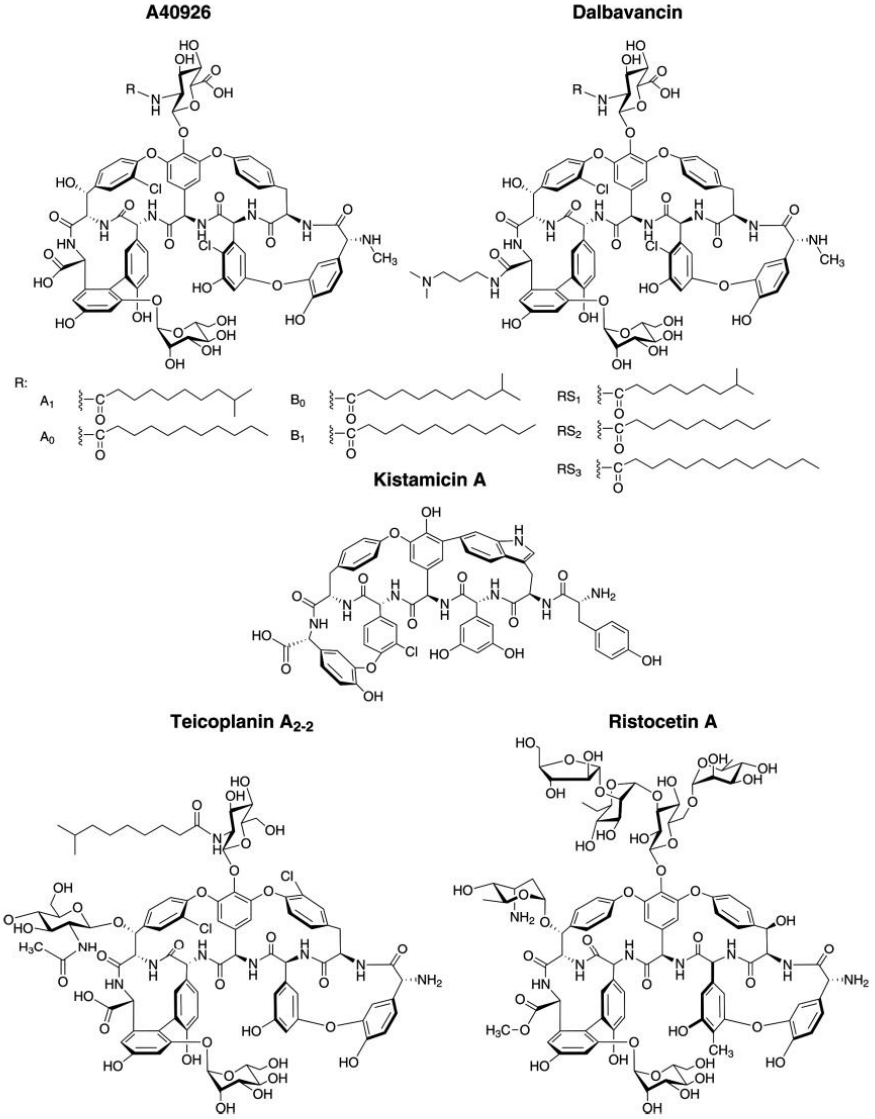


Figure 1. Structures of the GPAs found in genus *Nonomuraea*: type IV A40926 and type V kistamicin. Clinically used dalbavancin is obtained from A40926 by conversion of the C-terminal carboxyl group into a (3-dimethylamino)-1-propylamide. Type IV teicoplanin and type III ristocetin are shown due to their structural similarities with A40926. For teicoplanin, the main factor (TA_{2-2}) of the complex produced by *A. teichomyceticus* is shown, the other factors are differing by the length and branching of the lipid chain. Ristocetin is produced by numerous *Amycolatopsis* species.

has become increasingly relevant.[6,16] Following the sequencing of the A40926 BGC (*dbv*) almost two decades ago,[18] multiple aspects of A40926 biosynthesis were investigated, including nonribosomal aglycone assembly and tailoring steps,[15,19,20] self-resistance,[21,22] and pathway-specific regulation of its production.[6,23,24] *N. gerenzanensis* was also engineered to produce A40926 derivatives that are better suited for downstream chemical modification to dalbavancin.[16] Another GPA produced by a *Nonomuraea* species is the type V GPA kistamicin (Figure 1) from *Nonomuraea* sp. ATCC 55076, which was reported to exhibit potent antiviral activity as well as mild antibiosis against Gram-positive bacteria.[2,25] Its structure contains an unusual indole−phenol cross-link which makes this GPA unique among those already known.[5,26]

Genome mining has recently shown that other species from the genus *Nonomuraea* also possess BGCs for GPAs,[27] as in the cases of *Nonomuraea* sp. WAC 01424 and *Nonomuraea coxensis* DSM 45129. Notwithstanding the low quality of the available

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

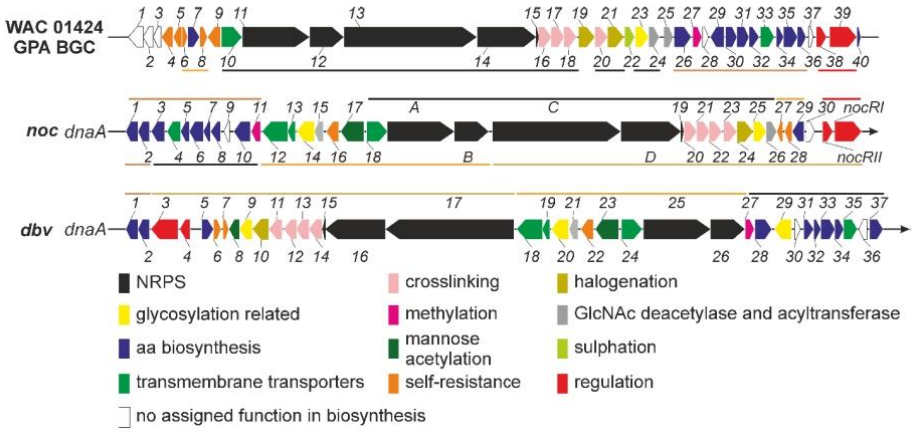


Figure 2. Comparison of BGCs from *N. gerenzanensis* (*dbv*), *N. coxensis* (*noc*), and *Nonomuraea* sp. WAC 01424. Colored lines indicate the homology segments among the BGCs. For *dbv* and *noc*, the orientation of the BGC genes is in relation to the orientation of the *dnaA* gene of the chromosome. This orientation was not possible for the WAC 01424 GPA BGC, since the corresponding genome is fragmented across multiple contigs. Details on gene function and homology are given in Table 1 and in the main text.

draft genomic data, we recently showed that *N. coxensis* DSM 45129 carries a BGC remarkably similar to *dbv*.[6] We found that this BGC contains a putative regulatory gene orthologous to *dbv3*, which encodes the pathway-specific regulator of LuxR-type in *N. gerenzanensis*.[6] The heterologous expression of this gene from *N. coxensis* (named *nocRI*) led to A40926 overproduction in *N. gerenzanensis*, indicating that it might be functional in *N. coxensis* as well. Thus, in this paper we present the fully assembled genome of *N. coxensis*, which has allowed us to properly describe the putative GPA BGC (called *noc*). Additionally, we report that *N. coxensis* produces a novel GPA complex, which we named A50926. Structural characterization of A50926 by liquid chromatography–mass spectrometry (LC-MS) and tandem MS (MS/MS) showed it has high similarity to A40926, although A50926 lacks the carboxyl group on the *N*-acylglucosamine (Glc*N*-Acyl) moiety. Consistently, the *noc* BGC lacks an orthologue of *dbv29*, which in *N. gerenzanensis* encodes the enzyme oxidizing the Glc*N*-Acyl moiety to an *N*-acylaminoglucuronic group.[19] Introduction of *dbv29* into *N. coxensis* changed the GPA production profile of this strain to A40926. Finally, we have introduced *dbv3* and *dbv4* pathway-specific regulatory genes in *N. coxensis* to trigger and overproduce A50926 by regulatory gene cross-talking. In conclusion, our results describe the biosynthesis of a novel GPA, which may have superior properties to A40926[28] and thus may contribute to developing a platform for the combinatorial biosynthesis of third generation lipo-GPAs.

2. RESULTS AND DISCUSSION

2.1. Complete Assembly of *N. coxensis* Genome Reveals the Presence of a Novel GPA BGC. The presence of a novel GPA BGC in the genome of *N. coxensis* was recently anticipated.[6,27] However, due to the poor quality of the available draft, fragments of the BGC were found on different contigs and did not cover the full expected sequence of the BGC. Therefore, we sequenced and fully assembled the genome of *N. coxensis* DSM 45129 using a combination of HiSeq Illumina and GridION ONT technologies. The circular chromosome of *N. coxensis* was found to have a smaller size in comparison to the other two previously published *Nonomuraea* genomes—only 9.07 Mbp compared to 11.85 Mbp in *N. gerenzanensis*[3] and 13.05 Mbp in *Nonomuraea* sp. ATCC 55076.[2] The average GC-content was 71.8%. Annotation of the *N. coxensis* genome revealed 8398 predicted protein coding sequences, five operons for 16S-23S-5S rRNA, and 73 tRNA genes. Genome analysis by antiSMASH 5.0,[29] a specialized metabolite BGC identification tool, led to the discovery of 27 putative BGCs when used in the "relaxed" search mode. However, only a few BGCs showed more than 20% similarity to known BGCs (Table S1).

We thus focused our attention on the GPA-like BGC, which we denoted as *noc* (from ***N****onomuraeacoxensis*). The *noc* BGC is the fourth GPA BGC described from *Nonomuraea* genus, following the *dbv* BGC from *N. gerenzanensis*,[18] a putative GPA BGC from *Nonomuraea* sp. WAC 01424[27] and the type V GPA kistamicin (*kis*) BGC from *Nonomuraea* sp. ATCC 55076.[2] Overall, *noc* contains 36 open reading frames (ORFs) with 35 among them homologous to *dbv* genes (the nonhomologous *noc* gene encoding for a putative transposase) and 32 being homologous to genes in the *Nonomuraea* sp. WAC 01424 GPA BGC (Figure 2, Table 1). The *kis* BGC differed from *noc* most significantly (data not shown).

2.2. Comparative Genomics of *Nonomuraea* GPA Producers. At the time of writing, genomic information for 34 *Nonomuraea* species was available in GenBank, although there are only three complete assemblies (Table S2). Along with the four reported *Nonomuraea* GPA BGCs, we found a *kis*-like BGC in the draft genome of *Nonomuraea* sp. NN258 (Figure S1). We have then reconstructed the multilocus phylogeny (MLP) of all *Nonomuraea* species with available genomic data using conserved house-keeping proteins (Table S3). It revealed *N. coxensis* to be most closely related to *N. wenchangensis* CGMCC 4.5598, *N. polychroma* DSM 43925, and *N. turkmeniaca* DSM 43926 (Figure S2). None of these species have GPA BGCs in their genomes. *N. gerenzanensis* is most closely related to *Nonomuraea* sp. FMUSA5–5 and to the kistamicin producer *Nonomuraea* sp. ATCC 55076, whereas

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915–928

Table 1. Characterization of *noc* BGC genes and their comparison to the *dbv* and WAC 01424 GPA BGCs from *Nonomuraea* spp.

noc BGC genes	homologues from *dbv* BGC (aa identity of protein product with *noc* homologue, %)	homologues from WAC 01424 GPA BGC (numbered as in Figure 2) (aa identity of protein product with *noc* homologue, %)	encoded protein
noc1	*dbv1* (90.6%)	*DMB42_RS42735* (31) (60%)	hydroxymandelate oxidase (Hmo)
noc2	*dbv2* (89.3%)	*DMB42_RS42740* (30) (62%)	hydroxymandelate synthase (HmaS)
noc3	*dbv37* (90.9%)	*DMB42_RS42745* (29) (83%)	hydroxyphenylglycine aminotransferase (HpgT)
noc4	*dbv35* (90.9%)	*DMB42_RS42730* (32) (63%)	Na^+−H^+ antiporter
noc5	*dbv34* (93.9%)	*DMB42_RS42710* (36) (87%)	enoyl-CoA hydratase (DpgD)
noc6	*dbv33* (89.2%)	*DMB42_RS42715* (35) (84%)	dihydroxyphenylacetyl-CoA dioxygenase (DpgC)
noc7	*dbv32* (85.1%)	*DMB42_RS42720* (34) (74%)	enoyl-CoA hydratase (DpgB)
noc8	*dbv31* (94.3%)	*DMB42_RS42725* (33) (91%)	type III polyketide synthase (DpgA)
noc9	*dbv30* (83.5%)	*DMB42_RS42750* (28) (69%)	4HB-CoA thioesterase
noc10	*dbv28* (92.4%)	*DMB42_RS42760* (26) (86%)	β-hydroxylase
noc11	*dbv27* (91.8%)	*DMB42_RS42755* (27) (58%)	methyltransferase
noc12	*dbv18* (87.3%)	[a]	ABC transporter
noc13	*dbv19* (92.2%)	[a]	ABC transporter
noc14	*dbv20* (89.7%)	[a]	mannosyltransferase
noc15	*dbv21* (86.6%)	*DMB42_RS42765* (25) (64%)	deacetylase
noc16	*dbv22* (92.3%)	*DMB42_RS42850* (9) (77%)	sensory histidine kinase
noc17	*dbv23* (88.1%)	[a]	acetyltransferase
noc18	*dbv24* (92.4%)	*DMB42_RS42845* (10) (81%)	ABC transporter
nocA	*dbv25* (88.7%)	*DMB42_RS42840* (11) (76%)	NRPS modules 1−2
nocB	*dbv26* (91%)	*DMB42_RS42835* (12) (78%)	NRPS module 3
nocC	*dbv17* (89.6%)	*DMB42_RS42830* (13) (77%)	NRPS modules 4−5−6
nocD	*dbv16* (91.7%)	*DMB42_RS42825* (14) (79%)	NRPS module 7
noc19	*dbv15* (94.2%)	*DMB42_RS42820* (15) (93%)	MbtH-like protein
noc20	*dbv14* (91.8%)	*DMB42_RS42815* (16) (78%)	cross-linking oxygenase (OxyA)
noc21	*dbv13* (89.8%)	*DMB42_RS42810* (17) (77%)	cross-linking oxygenase (OxyC)
noc22	*dbv12* (93.5%)	*DMB42_RS42805* (18) (77%)	cross-linking oxygenase (OxyB)
noc23	*dbv11* (91.9%)	*DMB42_RS42795* (20) (78%)	cross-linking oxygenase (OxyE)
noc24	*dbv10* (94.1%)	*DMB42_RS42790* (21) (87%)	halogenase
noc25	*dbv9* (90.4%)	*DMB42_RS42780* (23) (74%)	glycosyltransferase (GtfB)
noc26	*dbv8* (87.5%)	*DMB42_RS42775* (24) (77%)	acyltransferase
noc27	*dbv7* (87.3%)	*DMB42_RS42865* (6) (78%)	VanY-carboxypeptidase
noc28	*dbv6* (95.9%)	*DMB42_RS42855* (8) (92%)	response regulator
noc29	*dbv5* (92.8%)	*DMB42_RS42860* (7) (85%)	prephenate dehydrogenase (Pdh)
noc30	[a]	[a]	putative transposase
nocRII	*dbv4* (94.4%)	*DMB42_RS42700* (38) (85%)	StrR-like transcriptional regulator
nocRI	*dbv3* (86.3%)	*DMB42_RS42695* (39) (70%)	LuxR-like transcriptional regulator

[a]Homologue is absent.

Nonomuraea sp. WAC 01424 is distantly related to both *N. coxensis* and *N. gerenzanensis* (Figure S2). Thus, GPA-producing *Nonomuraea* species do not form a single phylogenetic group, which is different from what occurs in the majority of *Amycolatopsis* spp. producing GPAs.[30]

Since *N. gerenzanensis* and *Nonomuraea* sp. ATCC 55076 are closely related and their genomes had been completely assembled, we compared their sequences using the MAUVE genome alignment tool.[31] We found that the two genomes are very similar, having few rearranged homologous segments (Figure S3A). Interestingly, the regions flanking the *dbv* BGC in *N. gerenzanensis* show synteny in *Nonomuraea* sp. ATCC 55076, but in this genome, they flank a miscellaneous assemblage of GPA-unrelated genes instead of the *dbv* genes. No *dbv*-like BGC is present in *Nonomuraea* sp. ATCC 55076. Similarly, no *kis*-like BGC is in the *N. gerenzanensis* genome, but the regions flanking the *kis* BGC in *Nonomuraea* sp. ATCC 55076 have their homologous counterparts in the *N.*

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

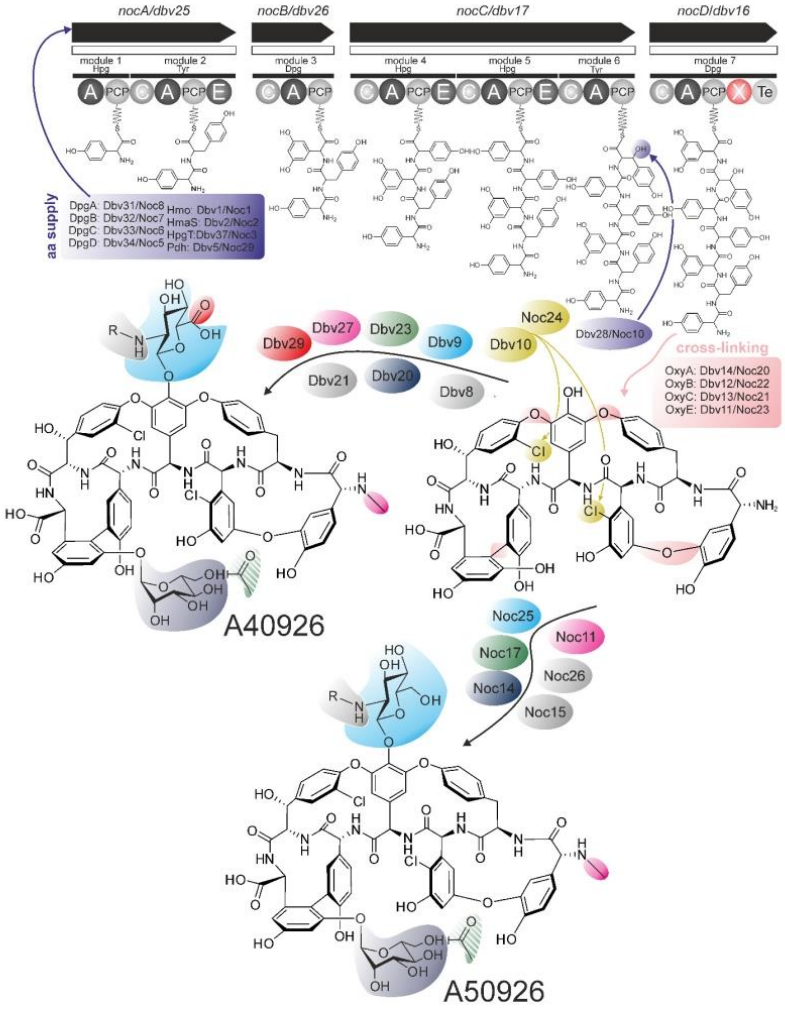


Figure 3. Conceptual scheme of the biosynthesis of A40926 and of the GPA (named A50926) from *N. coxensis*. Please note the dashed acyl group at the mannose residue, which is installed by Dbv23/Noc17 but consequentially lost during antibiotic extraction. For more details and encoded protein names, please refer to the main text and Table 1.

gerenzanensis genome (Figure S3A). Dot plots of *N. gerenzanensis* and *Nonomuraea* sp. ATCC 55076 confirm the high homology between the two strains (Figure S3B). A possible explanation is that *Nonomuraea* sp. ATCC 55076 and *N. gerenzanensis* genomes might have acquired different GPA BGCs independently through horizontal gene transfer (HGT) events from other *Nonomuraea* (or not) species.

Dot plots of *N. coxensis* and *Nonomuraea* sp. ATCC 55076 genomes (Figure S3C) as well as of *N. coxensis* and *N. gerenzanensis* (Figure S3D) indicate that *N. coxensis* is more distantly related to the other GPA producing species. Unfortunately, it was impossible to compare the genome of *N. coxensis* with its closest relatives *N. wenchangensis* CGMCC 4.5598, *N. polychroma* DSM 43925, and *N. turkmeniaca* DSM 43926 (Figure S2), due to the incompleteness of their genome assemblies. Overall, it seems that the position of GPA BGCs is not conserved within *Nonomuraea* genomes, which contrasts to what was observed in most *Amycolatopsis* spp.[30]

2.3. Comparing *noc* and *dbv* Biosynthetic Pathways: From Genes to Products. The biosynthesis of A40926 is well understood (Figure 3). The heptapeptide core of this antibiotic is synthesized by a nonribosomal peptide synthetase (NRPS) assembly line involving Dbv25, Dbv26, Dbv17, and Dbv16 proteins. The linear peptide is cross-linked by four monooxygenases (Dbv14, Dbv12, Dbv13, and Dbv11) and halogenated by Dbv10, giving the core aglycone. This aglycone is further modified with the glycosyltransferases Dbv9 and Dbv20, which attach *N*-acetyl glucosamine (GlcNAc) and mannose, respectively.[32] Then, GlcNAc is oxidized by Dbv29, deacetylated by Dbv21, and acylated by Dbv8. Finally, the

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

mannose moiety is acetylated by Dbv23, giving *O*-acetyl-A40926.

Considering the A40926 pathway, it was possible to predict the biosynthetic pathway of the putative GPA from *N. coxensis* (Figure 3). Sets of genes required for the biosynthesis of the nonproteinogenic precursor amino acids 4-hydroxyphenylglycine (Hpg), 3,5-dihydroxyphenylglycine (Dpg), and β-hydroxytyrosine (further used as substrates for NRPS) are the same in *noc* and *dbv* BGCs (Table 1, Figures 2 and 3). Next, the NRPS, encoded within *noc* BGC, was found to have the same organization and A-domain specificities as the *dbv* NRPS (Figure S4, Table S4). All other genes, responsible for the cross-linking and tailoring steps, were identical in both the *noc* and *dbv* pathways (Table 1, Figures 2 and 3). However, one notable difference between *dbv* and *noc* was the absence of a *dbv29* orthologue in the latter. As mentioned above, Dbv29 is a hexose oxidase responsible for the oxidation of the Glc*N*-Acyl moiety of A40926.[19] On this basis, we predicted that the *noc* pathway might produce an A40926 analogue lacking the carboxylic group on the Glc*N*-Acyl residue and therefore resembling teicoplanin in this moiety (Figures 1 and 3).

Beyond the biosynthetic genes, *noc* and *dbv* feature homologous regulatory genes. Two master regulators of A40926 biosynthesis—LuxR-like Dbv3 and StrR-like Dbv4—have orthologues coded within *noc*—NocRI (94% aa sequence identity) and NocRII (86% aa sequence identity), respectively.[6] In *N. gerenzanensis*, both Dbv3 and Dbv4 are crucial for biosynthesis activation.[23] Dbv4 was shown to bind the promoter regions of operons *dbv30-35* (mainly coding for Dpg biosynthesis enzymes) and *dbv14-8* (including the genes coding for cross-linking monooxygenases), and its binding sites were identified.[33] Our *in silico* analysis indicates that identical binding sites are present in the promoter regions of *noc20* and *noc8*, orthologues of *dbv14* and *dbv30*, respectively (Figure S5). DNA-binding sites of Dbv3 remain uncharacterized, but its regulon was defined from gene expression analysis and includes other biosynthetic genes and Dbv4.[23] Given all these similarities, we presume that NocRI/NocRII have functions identical to Dbv3/Dbv4 and both regulatory pairs might cross-talk between these species. Our previous results,[6] where heterologous expression of *nocRI* in *N. gerenzanensis* improved A40926 production, support this assumption. The single GPA resistance determinant encoded within *noc* is Noc27, a close (87%) orthologue of Dbv7 ($VanY_n$), which is a D,D-carboxypeptidase involved in A40926 self-resistance.[21,34]

Although the biosynthetic, regulatory, and resistance genes are apparently shared by the *dbv* and *noc* BGCs, their genetic organization is different. So far, almost all GPA BGCs have NRPS genes located on one strand in an order that is colinear to the order of the modules in the NRPS assembly line. The only exception is the *dbv* BGC, where the NRPS genes are coded on different strands and are separated by other biosynthetic genes.[18] The *noc* BGC, although sharing a remarkable similarity with *dbv*, features an organization of NRPS genes that is typical of all the other GPAs. Interestingly, only two chromosomal inversion events are needed to rearrange *noc* into *dbv* (Figure S6), indicating how a *dbv*-like gene arrangement might have derived from a *noc*-like BGC in a common ancestor of *N. coxensis* and *N. gerenzanensis* (or in an ancestralprotocluster).

The putative GPA BGC in *Nonomuraea* sp. WAC 01424 (Figure 2) differs more substantially from both *noc* and *dbv*. It lacks a *noc14/dbv20* homologue encoding for a mannosyltransferase, as well as a *noc17/dbv23* homologue encoding for a mannose-*O*-acetyltransferase (Table 1). Instead, WAC 01424 GPA BGC contains a close homologue of *staL* (Figure S7), which encodes for a sulfotransferase involved in the biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL 15009.[35] Additionally, the WAC 01424 BGC-encoded halogenases seem more related to the ones from the A47934 BGC than to Noc24 and Dbv8 (Figure S7). Thus, we suggest that WAC 01424 GPA is a nonmannosylated, but sulfated, A40926 analogue, putatively with a halogenation pattern different from A40926 (Figure S8).

2.4. Optimization of GPA-Producing Conditions for *N. coxensis*. *N. coxensis* was first described in 2007,[36] but as far as we know, it was never tested for the production of antimicrobials. Considering the predicted similarity between the putative GPA produced by *N. coxensis* with A40926, we first applied to *N. coxensis* the cultivation and A40926 production conditions that we had previously optimized for *N. gerenzanensis*.[22,37] In these conditions (namely a vegetative preculture in E26 medium and a GPA production step in FM2 medium using baffled flasks), *N. coxensis* tended to grow poorly, and no antimicrobial activity was detectable throughout the 168h cultivation from inoculum. Thus, we further screened different media and fermentation conditions previously used for growing other GPA producing strains, such as TM1 used for teicoplanin production by *A. teichomyceticus*[38] and R5 adopted for balhimycin production in *Amycolatopsis balhimycina*,[39] as well as VM0.1 and ISP2l previously employed for the vegetative cultivation of *N. coxensis*[6] (media composition detailed in the Supporting Information). The production of antimicrobial activity toward *Bacillus subtilis* ATCC 6633 was observed only in TM1 and ISP2l media when glass beads were added to favor dispersed growth (Figure S9). Indeed, adding glass beads to E26 medium cultures allowed us to use it for a successful vegetative preculture step (Figure S10A). Interestingly, routine analysis of glucose consumption in all media described above indicated that *N. coxensis* did not visibly consume glucose during growth (data not shown). We thus tested the glucose-lacking E26 (named E27), TM1 (TM1m), and ISP2l (ISP2lm) media variants for *N. coxensis* growth and putative GPA production. We found that biomass accumulation was similar in E26 and E27 (Figure S10A) and that biomass and antimicrobial production were equivalent in TM1 and ISP2l as well as in their glucose lacking variants TM1m and ISP2lm (Figure S10B and C). Currently, it is impossible to say why *N. coxensis* fails to use glucose throughout cultivation given that all necessary genes are present within its genome (Figure S10D). Thus, for all the following work with *N. coxensis*, E27, TM1m, and ISP2lm were used.

2.5. Expression of VanY-like Activity in *N. coxensis*. As already mentioned, the *noc* BGC encodes a Dbv7 orthologue—Noc27. We therefore tested whether D,D-carboxypeptidase activity could be detected in GPA-producing cultures of *N. coxensis*. This was measured in membrane extracts as previously reported for *N. gerenzanensis* and its mutant strains.[22] D,D-carboxypeptidase activity was measurable in *N. coxensis* extracts, although at an inferior level than in *N. gerenzanensis* (Figure S11). This indicated that Noc27 is functional and its expression correlates with the antimicrobial producing conditions. These results corroborate the hypothesis that *noc* genes are expressed and a novel GPA active versus *B. subtilis* is produced by *N. coxensis*. As in *dbv*[40] and WAC 01424

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

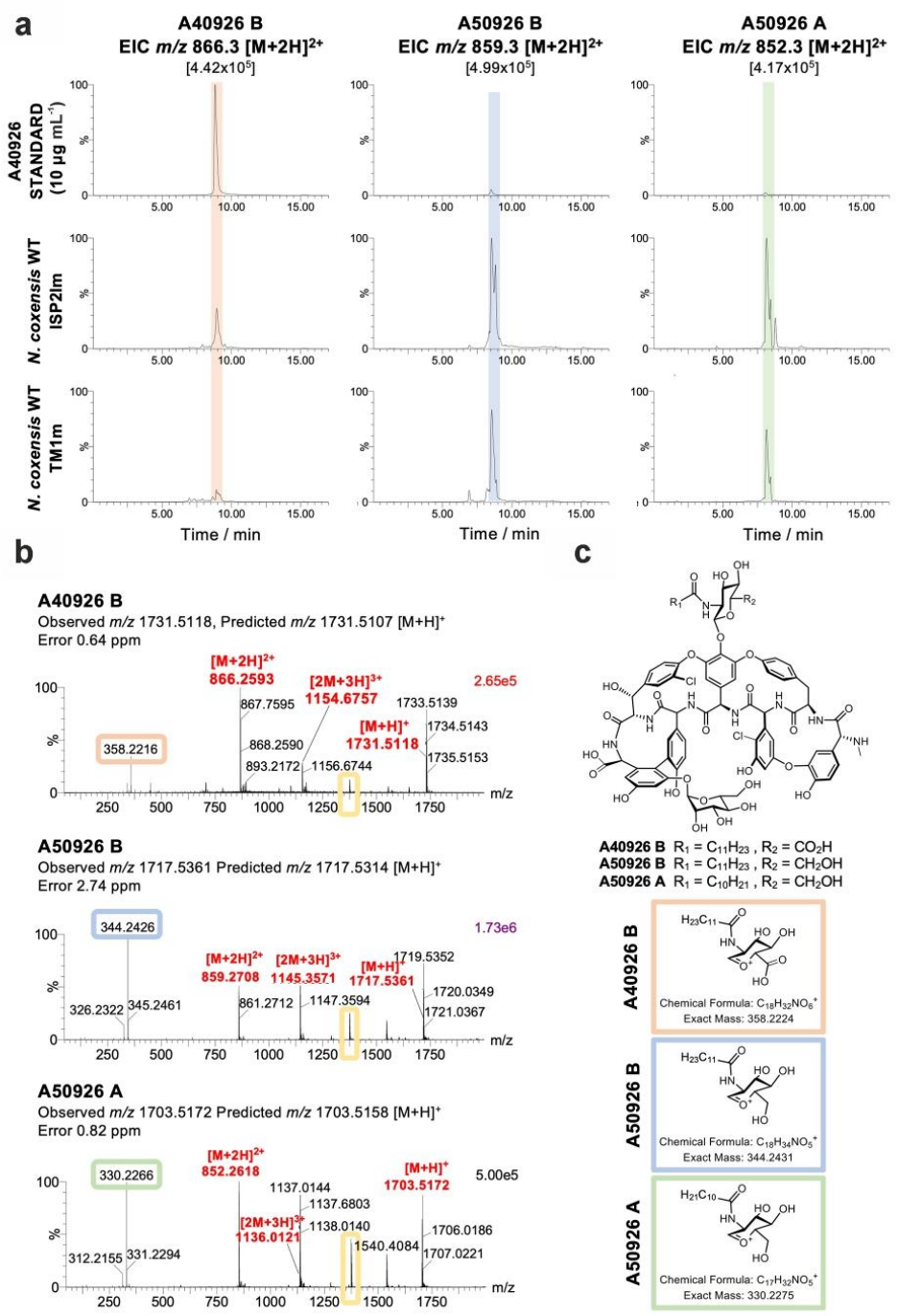


Figure 4. MS characterization of novel GPA complex produced by wild type *N. coxensis* grown in ISP2lm and TM1m media for 7 days. (a) Extracted ion chromatograms (EICs) of masses corresponding to A40926 B (left column) and the major components of the A50926 complex produced by *N. coxensis* WT, A50926 B (*m*/*z* 859.3, second column), and A50926 A (*m*/*z* 852.3, third column). The top row corresponds to a

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

Figure 4. continued

commercial standard of A40926 and the middle and bottom rows to culture extracts from ISP2lm and TM1m, respectively. For each mass, peak heights are normalized relative to the intensity of the largest peak in the sample set, shown in brackets at the top of each column. (b) MS spectra for A40926 B, A50926 B, and A50926 A. Peak heights are normalized to the intensity of the top peak in each spectrum, shown on the top right corner of each plot. Signature in-source fragments for each of the analyzed molecules are circled in pink, blue, and green, respectively, whereas the fragment corresponding to the mannosylated aglycone common to all of them is highlighted in yellow. (c) Proposed structure for the A50926 molecules. The top schematic represents a generic proposed structure common to A40926 and A50926 while the insets below represent the differential fragments for each of the analyzed molecules, as inferred from MS and MS/MS data.

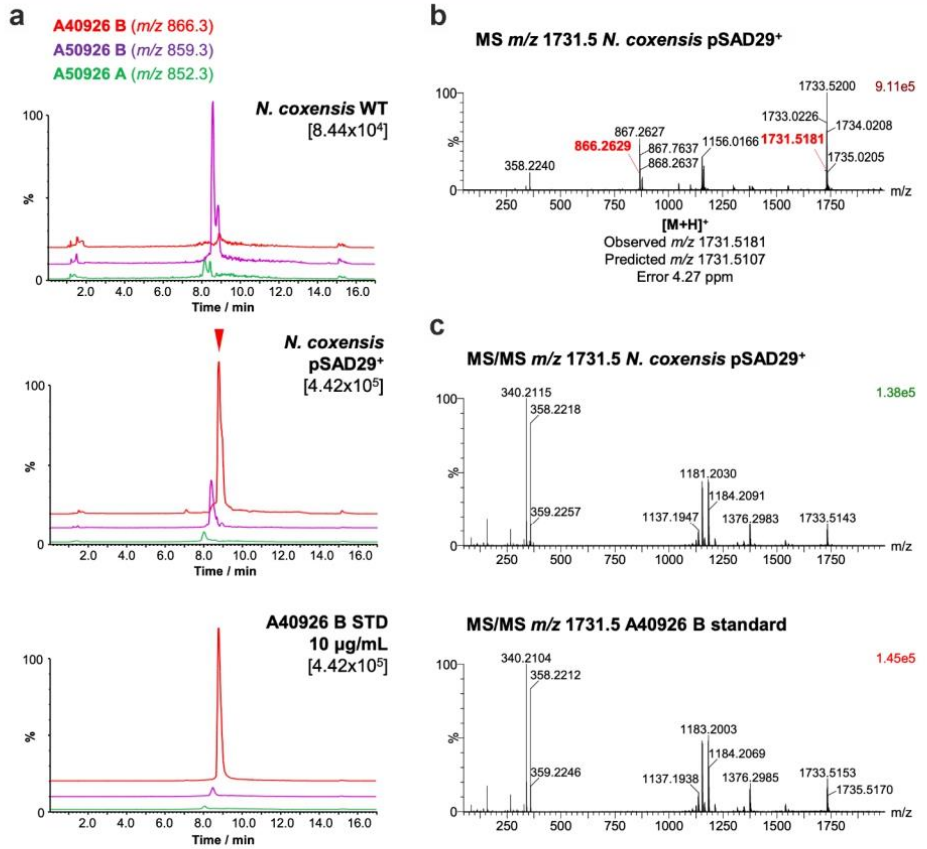


Figure 5. Production of A40926 in *N. coxensis* pSAD29⁺ grown in ISP2lm for 7 days. (a) EICs for masses corresponding to A40926 B (red trace), A50926 B (purple), and A50926 A (green) in purified extracts of *N. coxensis* pSAD29⁺ (top chromatogram) and *N. coxensis* WT (middle) in comparison to an A40926 commercial standard. The intensity for the top peak in each chromatogram is shown in brackets under the sample name. (b) MS spectrum of A40926 B from *N. coxensis* pSAD29⁺ cultures. Monoisotopic masses corresponding to $[M + 2H]^{2+}$ and $[M + H]^{+}$ adducts are highlighted in red, and the deviation between the observed accurate mass and the predicted mass for A40926 is represented in parts per million. (c) MS/MS spectra of A40926 B produced by *N. coxensis* pSAD29⁺ and an A40926 B commercial standard.

BGCs, a *vanY* gene seems to be the only cluster-situated determinant of self-resistance in *N. coxensis*.

2.6. Purification and Identification of the Novel Glycopeptide Complex Produced by *N. coxensis*. D-Alanine-D-Alanine (D-Ala-D-Ala) affinity resin chromatography was used to capture the putative GPA from cultures of *N. coxensis* grown in ISP2lm and TM1m media. ISP2lm appeared to be the most suitable medium for GPA purification, since the rich composition and high viscosity of TM1m interfered with affinity chromatography. Analyzed by HPLC, the affinity resin eluates contained two major peaks with the characteristic UV spectra of the commercially available A40926 standard, but with a different retention time (Figure S12). LC-MS analysis of these peaks revealed they corresponded to ions with *m/z* 852.3 and 859.3 ($[M + 2H]^{2+}$), 28 and 14 Da smaller respectively than an A40926 standard ($[M + 2H]^{2+}$ = 866.3, corresponding to A40926 B). We therefore tentatively named this new GPA complex A50926 (Figure 4a and c).

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

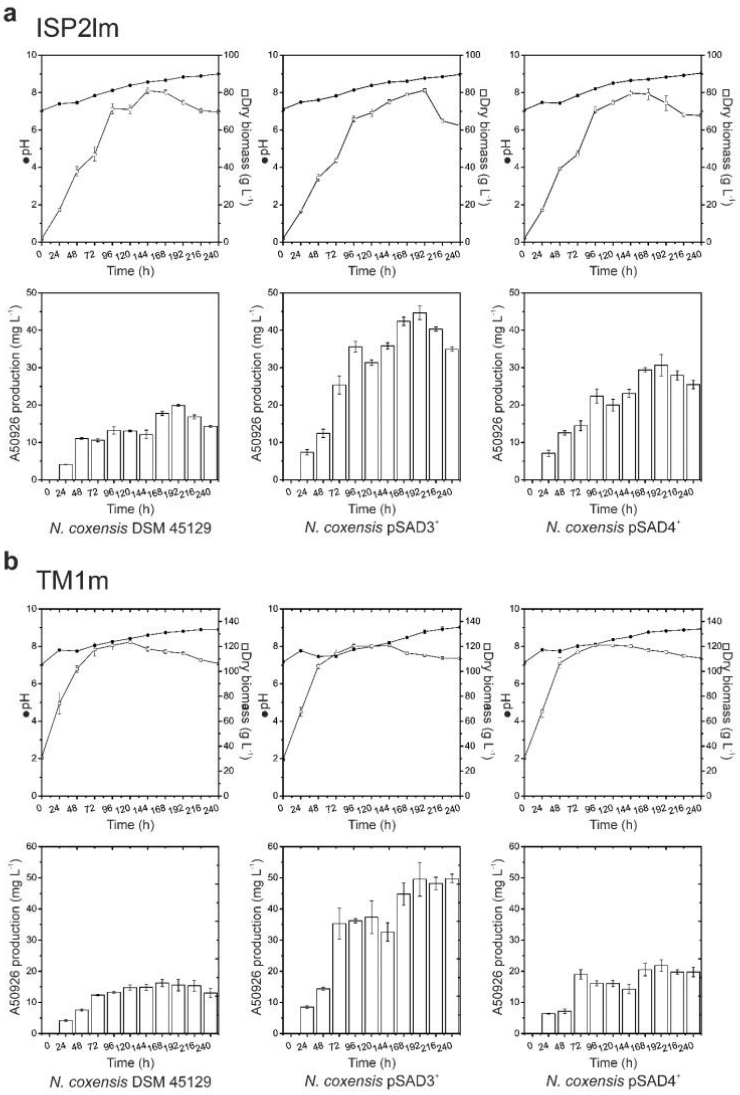


Figure 6. Time courses of *N. coxensis* wild type and of the recombinant strains overexpressing *dbv3* (pSAD3$^+$) and *dbv4* (pSAD4$^+$) cultivated in ISP2lm (a) or TM1m (b) in 500 mL Erlenmeyer flasks. pH (filled circles), biomass accumulation (empty squares), and A50926 production were monitored every 24 h. Results given are mean values of three independent experiments, and error bars represent standard deviations.

All three molecules showed similar MS spectra with single, double, and triple charge proton adducts as well as in-source fragments corresponding to the aglycone carrying the mannose moiety and the Glc*N*-Acyl moiety (Figure 4b,c). The mannosylated aglycone fragment (m/z 1374.3) was common to all three peaks (Figures 4b and S13), indicating that they share the same aglycone structure and mannose decoration. In contrast, the in-source fragment corresponding to the acylated sugar carried the signature mass difference for each molecule (Figures 4b, S14, and S15): the main A50926 peak ($[M + H]^+$ = 1717.5361) had a fragment with m/z 344.2, whereas the A40926 standard had a fragment with m/z 358.22 (Figures 4b, S14, and S15). Further MS and MS/MS analyses of these fragments (Figures S14 and S16) allowed us to assign this 14 Da mass difference to the glucosamine moiety. The masses are consistent with this sugar featuring a regular 6-hydroxyl group in A50926 versus being carboxylated in A40926 (Figures 1, 3, 4c, S14, S15, and S16). This correlates with the lack of a homologue of *dbv29* in the *noc* BGC, as it encodes the enzyme responsible for the oxidation of the C-6 hydroxyl group of Glc*N*-Acyl into a carboxylic acid in A40926. The second A50926 peak ($[M + H]^+$ = 1703.5172) had a further 14 Da

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

mass difference in the GlcN-Acyl moiety (Figures 4b and S14), but in this case MS/MS showed this difference to be in the acyl chain (Figure S16), which is consistent with an A50926 congener with a C11 acyl chain instead of a C12 acyl chain. This is equivalent to the A and B series of congeners in the A40926 complex.[41] Based on this analysis and accurate mass data (Figure 4b), we named the compound with $[M + H]^+$ = 1717.54 A50926 B (Figure 4c) and the compound with $[M + H]^+$ = 1703.52 A50926 A (Figure 4c).

2.7. Single Gene Expression Leads to A40926 Production in *N. coxensis*. To support our MS-based characterization of A50926, we hypothesized that we could convert *N. coxensis* into an A40926 producer by overexpression of the *dbv29* gene from *N. gerenzanensis*, which encodes the hexose oxidase required for oxidation of the C-6 hydroxyl group of GlcN-Acyl into the corresponding carboxylic acid. To achieve this, we used the pSET152A expression platform, which has proven to be very effective for gene overexpression in both *N. coxensis* and *N. gerenzanensis*.[6] *dbv29* was cloned into pSET152A to generate pSAD29, which was then introduced into *N. coxensis* by conjugation from *Escherichia coli*. *N. coxensis* pSAD29$^+$ was grown in ISP2lm medium for 168 h, and the resulting GPA complex was purified using D-Ala-D-Ala affinity resin. LC-MS analysis determined that *N. coxensis* pSAD29$^+$ was able to produce a molecule with an identical retention time and MS spectrum to that of A40926 (observed m/z 1731.5181, calculated A40926 $[M + H]^+$ 1731.5107, 4.27 ppm difference) (Figure 5a and b).

MS/MS analysis of the molecule showed it also had an identical fragmentation pattern to the A40926 standard, including the in-source fragment with m/z 358.22 characteristic of the carboxylated GlcN-Acyl moiety (Figures 5c and S17). Traces of A50926 could also be detected in the extract of the complemented strain, indicating that while complementation was very efficient, conversion from A50926 to A40926 was not complete (Figure 5a). Alongside the BGC homology (Figure 2), this provides strong evidence that A50926 is chemically identical to A40926 with the exception of the carboxylated GlcN-Acyl. However, we cannot completely rule out small differences, such as acyl chain branching.

2.8. Heterologous Expression of Transcriptional Regulators *dbv3* and *dbv4* to Enhance the Production of A50926 in *N. coxensis*. In previous work, we overexpressed the two *dbv* BGC situated master regulators in *N. gerenzanensis* (*dbv4* and *dbv3*) to successfully improve A40926 production.[6] Therefore, hereby we used the previously constructed expression vectors pSAD4 and pSAD3 carrying *dbv4* and *dbv3*, respectively, in *N. coxensis* to trigger and improve A50926 production. First, we observed that *N. coxensis* pSAD3$^+$ and pSAD4$^+$ recombinant strains grown in the E27 and VSP vegetative media produced an antimicrobial activity against *B. subtilis* (Figure S18A), whereas their parental wild type strain did not exhibit any antimicrobial activity in these media. Overexpression of *dbv3* also triggered antimicrobial activity on VM0.1 and ISP2 solid media, whereas the wild type was not active (Figure S18B). Consistently, in both ISP2lm and TM1m production media *N. coxensis* pSAD3$^+$ and pSAD4$^+$ produced more antibiotic than the wild type (Figures S18C and 6a and b). In ISP2lm (Figure 6a), at 192 h *N. coxensis* pSAD3$^+$ reached the maximum production of approximately 45 μg mL^{-1}, exceeding both wild type (approximately 20 μg mL^{-1}) and *N. coxensis* pSAD4$^+$ (approximately 30 μg mL^{-1}) productivities. In TM1m medium (Figure 6b), *N. coxensis* pSAD3$^+$ produced approximately 50 μg mL^{-1} after 192 h of cultivation. At the same time point in TM1m the wild type and *N. coxensis* pSAD4$^+$ produced approximately 16 and 22 μg mL^{-1} of antibiotic, respectively. The control strain carrying the "empty" pSET152A vector performed exactly as the wild type (data not shown). No significant differences between biomass accumulation or pH were observed among the recombinant strains, or in comparison with the parental *N. coxensis* wild type strain. Thus, overexpression of *dbv3* and *dbv4* regulatory genes triggered or improved the production of A50926 in *N. coxensis* under different cultivation conditions.

3. CONCLUSIONS

A novel GPA, A50926, was identified from *N. coxensis* DSM 45129. Detailed MS and MS/MS analysis indicates that A50926 differs from the previously characterized A40926 GPA by lacking the carboxyl group on the GlcN-Acyl moiety attached to Hpg4 of the GPA aglycone, resembling teicoplanin in this part of the molecule. A compound with the same chemical structure was described 25 years ago as a chemically prepared derivative of A40926 (named RA[28]). Extensive study of antibacterial activities of RA *in vitro*[28] indicated that RA has slightly better antimicrobial activity than A40926: minimal inhibitory concentrations (MICs) of RA were 2−4 times lower against different staphylococcal and enterococcal strains when compared to A40926. The difference of chemical structure between the newly described A50926 and A40926 correlates with the absence of *dbv29* orthologue in the A50926 BGC (*noc*). Consistently, when *dbv29* was introduced into *N. coxensis*, we obtained A40926 production in the recombinant strain. Otherwise, both *noc* and *dbv* BGCs share all biosynthetic genes, which are closely related. Heterologous expression of A40926 regulatory genes *dbv3* and *dbv4* in *N. coxensis* improved A50926 production.

Although the majority of *noc* and *dbv* genes are orthologous, the *dbv* BGC is significantly rearranged in comparison to the *noc* BGC, as well as all other characterized GPA BGCs. We have proposed a series of genetic inversions that could have occurred in a common *Nonomuraea* ancestor to explain these different genetic architectures. Both BGCs are quite similar to the putative GPA BGC from *Nonomuraea* sp. WAC 01424. The latter lacks genes required for the addition of mannose, but possesses a gene encoding a sulfotransferase and an additional gene encoding a halogenase. Thus, the putative nonmannosylated GPA from *Nonomuraea* sp. WAC 01424 might be sulfated and have a different chlorination pattern than A40926/A50926. Consequently, *Nonomuraea* sp. WAC 01424 GPA BGC seems an attractive source for new tailoring genes to obtain A40926 derivatives with altered pharmacological properties. Notwithstanding the GPA BGC similarity, multilocus phylogeny of *Nonomuraea* spp. shows that GPA producers are not clustered together: GPA producers are found in distinct clades within the genus. Our analysis indicates that type IV and V GPA BGCs are common in *Nonomuraea* spp., which is in contrast to how rare these BGCs were believed to be. This is comparable to studies that show that BGCs for types I−III−IV GPAs are common in *Amycolatopsis*, and type V GPAs in *Streptomyces*.[27,42,43] This highlights how rare actinomycete genera, such as *Nonomuraea*, may represent a rich untapped source of novel GPAs, as well as GPA tailoring enzymes for the diversification of existing GPA scaffolds.

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

4. METHODS

4.1. Bacterial Strains and Cultivation Conditions. Bacterial strains and plasmids used in this work are summarized in Table S6. Compositions of all the media used for cultivation and GPA production are also given in the Supporting Information. All media components and antibiotics were supplied by Sigma-Aldrich, unless otherwise stated. For routine maintenance, *N. gerenzanensis* and *N. coxensis* strains were cultivated on ISP3 agar medium supplemented with 50 μg mL^{-1} apramycin-sulfate when appropriate. For genomic DNA isolation, *N. gerenzanensis* and *N. coxensis* strains were cultivated in liquid VSP medium on an orbital shaker at 220 rpm and at 30 °C. The working cell banks (WCBs) for *N. gerenzanensis* and *N. coxensis* strains were prepared as described previously.[22,37] *E. coli* DH5α was used as a routine cloning host, and *E. coli* ET12567 pUZ8002 was used as a donor for intergeneric conjugations. *E. coli* strains were cultivated at 37 °C in LB liquid or agar media supplemented with 100 μg mL^{-1} of apramycin-sulfate, 50 μg mL^{-1} of kanamycin-sulfate, and 25 μg mL^{-1} of chloramphenicol when appropriate.

4.2. Plasmid Construction and Generation of Recombinant *N. coxensis* Strains. To construct the pSAD29 expression vector, the coding sequence of *dbv29* (1601 bp) was amplified from the genomic DNA of *N. gerenzanensis* using dbv29_F/R primer pair (Table S7) and Q5 high-fidelity DNA polymerase (New England Biolabs). The resulting amplicon was digested with *Eco*RI and *Eco*RV restriction endonucleases and cloned into pSET152A[44,45] cleaved at the same binding sites. The resulting plasmid was verified by endonuclease restriction mapping and sequencing at BMR Genomics.

pSAD29, as well as pSAD3,[6] pSAD4,[6] and pSET152A,[44] were transferred to *N. coxensis* conjugatively, as described previously.[6] Transconjugants were selected as resistant to 50 μg mL^{-1} of apramycin-sulfate. Obtained strains were verified by PCR. To verify the integration of pSAD29, a ∼1.1 kbp fragment of pSAD29 was amplified using the dbv29_seq_int/PAM_seq_R (Table S7) primer pair, in which dbv29_seq_int anneals within *dbv29* and PAM_seq_R anneals upstream the *Eco*RV cleavage site of pSET152A. To verify the integrations of pSAD4 and pSAD3, ∼1 kbp and ∼2 kbp fragments were amplified respectively using PAM_seq_F/dbv4_R and PAM_seq_F/dbv3_seq_R primer pairs (Table S7). Finally, the integration of pSET152A was verified by amplifying *aac(3)IV* with the aac(3)IV_F/R primer pair (Table S7). In all cases, genomic DNA was isolated using the Kirby procedure.[46]

4.3. *N. coxensis* Cultivation for A50926 Production. To initiate the cultivation of *N. coxensis*, one WCB vial was inoculated into a 250 mL Erlenmeyer flask with 50 mL of VSP reactivation medium containing 6 glass beads (ø5 mm). After 72 h of incubation on a rotary shaker at 220 rpm, 30 °C the culture was used to inoculate (10% v/v) 500 mL Erlenmeyer flasks containing 100 mL of E27 vegetative medium and 12 glass beads (ø 5 mm). Following 72 h of incubation on a rotary shaker at 220 rpm, 30 °C this culture was used to inoculate (10% v/v) 500 mL Erlenmeyer flasks with 100 mL of ISP2lm or TM1m production media containing 12 glass beads (ø 5 mm). A50926 production cultures were then incubated up to 240 h on a rotary shaker at 220 rpm, 30 °C. Samples were collected at regular time points to estimate biomass accumulation (dry weight), pH, and A50926 production.

4.4. VanY-Related Activity Measurement. D,D-carboxypeptidase activity in *Nonomuraea* spp. was measured in FM2 production medium (*N. gerenzanensis*) and ISP2lm (*N. coxensis*) at 24, 48, 72, 96, 120, and 144 h time points. Mycelial lysates were prepared as described previously.[34] The enzyme activity releasing D-Ala from the tripeptide *N*-Acetyl-L-Lys-D-Ala-D-Ala (10 mM) was followed spectrophotometrically by a D-amino acid oxidase/peroxidase coupled reaction that oxidizes the colorimetric substrate 4-aminoantipyrine to chinonemine. D,D-carboxypeptidase activity was normalized to dry biomass weight, as previously reported.[22] One unit is defined as the amount of enzyme that is able to convert 1 μmol of substrate in 1 min.

4.5. HPLC and LC-MS Analysis of GPAs. For quantitative measurement, A40926 and A50926 were extracted from *N. coxensis* cultures with equal volumes of borate buffer composed of 100 mM H_3BO_3 (Sigma-Aldrich) and 100 mM NaOH (Sigma-Aldrich), pH 12. During this extraction the *O*-acetylated forms were converted in the corresponding deacetylated GPAs A40926 and A50926. A40926 and A50926 were analyzed using HPLC as previously reported.[6,22,37] In all cases the injection volumes of studied samples and standards were the same (50 μL). Concentration of A50926 was estimated as follows:

$$\text{A50926 concentration}\left(\frac{\text{mg}}{\text{L}}\right) = \frac{C(\text{A40926 std}) \times A(\text{A50926})}{A(\text{A40926 std})} \times 2$$

Where, C(A40926 std) is the concentration of the commercial A40926 sample; A(A50926) is the area sum of the peaks corresponding to A50926 B; A(A40926 std) is the area of the peak corresponding to the standard A40926 factor B_0; and 2 is the dilution factor.

High resolution liquid chromatography−mass spectrometry (LC-MS) and fragmentation (MS/MS) analysis of A40926 and A50926 was carried out on a SYNAPT G2-Si mass spectrometer equipped with an Acquity UPLC (Waters). Samples were injected onto a Waters Acquity UPLC BEH 1.7 μm, 1 × 100 mm C18 column, and eluted with a gradient of (B) acetonitrile/0.1% formic acid in (A) water/0.1% formic acid with a flow rate of 0.08 mL min^{-1} at 45 °C. The concentration of B was kept at 1% for 2 min followed by a gradient up to 40% B over 9 min, ramping to 99% B in 1 min, kept at 99% B for 2 min and re-equilibrated at 1% B for 4 min. MS data were collected in positive mode with the following parameters: resolution mode, scan time 0.5 s, mass range m/z 50−2000 calibrated with sodium iodide, capillary voltage = 2.5 kV; cone voltage = 40 V; source temperature = 120 °C; desolvation temperature = 350 °C. Leu-enkephalin peptide was used to generate a lock-mass calibration with m/z 556.2766 for positive mode, measured every 90 s during the run. For MS/MS fragmentation, a data directed analysis (DDA) method was used with the following parameters: precursor selected from the 4 most intense ions; MS/MS threshold 5000; scan time 2 s; no dynamic exclusion. Collision energy (CE) was ramped between 8 and 35 at low mass (m/z 50) and 10−70 at high mass (m/z 1200).

4.6. Purification of GPAs Using D-Ala-D-Ala Based Affinity Resin. GPAs were purified by affinity chromatography with a D-Alanine-D-Alanine (D-Ala-D-Ala) based resin. Activation of 5 mL HiTrap NHS-activated HP affinity columns (GE Healthcare) and ligand binding was conducted as described before[47] with modifications. Briefly, the resin was activated with 30 mL of 1 mM HCl, followed by injection of 200 mM D-Ala-D-Ala dipeptide, dissolved into 5 mL of coupling buffer (0.2 M $NaHCO_3$, pH 7.0). After 30 min incubation, the resin was washed with three cycles of 0.5 M ethanolamine hydrochloride, 0.5 M NaCl (pH 4.0, 30 mL), followed by 0.1 M sodium acetate, 0.5 mM NaCl (pH 4.0, 30 mL), alternately. Finally, the resin was washed with 50 mL coupling buffer and left to equilibrate for at least 1 h before use.

N. coxensis cultures were extracted in borate buffer as reported above, the pH in the obtained extracts was adjusted to 7.5 with HCl, and they were applied to the affinity chromatography system. Thus, extracts in borate buffer, coming from *N. coxensis* strains cultivated in TM1m or ISP2lm media, were filtered with 0.45 μm cutoff and loaded onto a D-Ala-D-Ala column at a flow rate of 0.5 mL min^{-1}. After extensive washing with coupling buffer, the bound GPA was eluted with 0.1 M NaOH and the eluate was lyophilized.

4.7. Bioassays for the Detection of A50926. Agar plug or Whatman paper disc (GE Healthcare) antibiotic diffusion assays were used to determine antimicrobial activities. An overnight *B. subtilis* ATCC 6633 culture in Mueller-Hinton broth II (cation adjusted, Sigma-Aldrich) was used to inoculate (1% v/v) a fresh culture, which was grown to OD_{600} = 0.6. A 200 μL portion of this culture was then added to 25 mL of 0.7% (w/v) Mueller-Hinton agar (Condalab) and plated. After solidification of the media, agar plugs cut from the plates with *N. coxensis* lawns, or Whatman paper discs containing GPAs,

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

were placed on the agar surface. Bioassay plates were incubated for 16 h at 37 °C before examination.

4.8. Sequencing and Annotation of the *N. coxensis* Genome. The genome of *N. coxensis* was sequenced using a combination of HiSeq Illumina and GridION ONT technologies. The Illumina data was obtained from SRA (PRJNA165411), while for the ONT data, a sequencing library (SQK-LSK109) was prepared using the Ligation Sequencing Kit (Oxford Nanopore Technologies) according to the manufacturer's instructions and run on a GridION sequencer in an R9.4.1 flowcell (both Oxford Nanopore Technologies). Base-calling of the raw data was performed with GUPPY-FOR-GRIDION v3.0.6. The assembly and polishing were performed as described previously,[48] using canu v.1.8 instead of v.1.6. The ONT data was assembled into 5 contigs, while the Illumina data were assembled into 87 scaffolds containing 310 contigs using NEWBLER v2.8. After manual curation using CONSED,[49] the complete genome of *N. coxensis* DSM 45129, consisting of one circular chromosome of 9,073,954 bp (72.12% G + C) was obtained. Annotation was performed using PROKKA v1.11[50] resulting in the prediction of 8,398 coding sequences (CDS), 5 rRNA operons, 73 tRNAs, and 5 noncoding RNA elements. The annotated genome and ONT raw data were deposited at DDBJ/ENA/GenBank under the BioProject accession number PRJNA693185.

4.9. *In Silico* Analysis Tools and Approaches. Routine analysis of nucleotide and amino acid sequences was performed in GENEIOUS v4.8.5.[51] Multiple sequence alignments, selection of the best models for the phylogenetic reconstruction and phylogenetic reconstruction itself were done with the MEGA X package.[52] To reconstruct the multilocus phylogeny of *Nonomuraea*, orthologues of 30 *S. coelicolor* house-keeping proteins (Table S3,[53]) were identified within the genomes of 34 *Nonomuraea* spp. (Table S2) using reciprocal best hit (RBH) BLAST. Sequences of these proteins from each *Nonomuraea* spp. were concatenated, and these concatenates were used for the upstream phylogenetic reconstruction.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.1c00170.

Tables S1−S7 and Figures S1−S18 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Andrew W. Truman − *Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, United Kingdom;* orcid.org/0000-0001-5453-7485; Email: andrew.truman@jic.ac.uk

Flavia Marinelli − *Department of Biotechnology and Life Sciences, University of Insubria, 21100 Varese, Italy;* orcid.org/0000-0001-9195-6777; Email: flavia.marinelli@uninsubria.it

Authors

Oleksandr Yushchuk − *Department of Biotechnology and Life Sciences, University of Insubria, 21100 Varese, Italy;* orcid.org/0000-0003-3296-8297

Natalia M. Vior − *Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, United Kingdom*

Andres Andreo-Vidal − *Department of Biotechnology and Life Sciences, University of Insubria, 21100 Varese, Italy*

Francesca Berini − *Department of Biotechnology and Life Sciences, University of Insubria, 21100 Varese, Italy*

Christian Rückert − *Technology Platform Genomics, CeBiTec, Bielefeld University, 33615 Bielefeld, Germany*

Tobias Busche − *Technology Platform Genomics, CeBiTec, Bielefeld University, 33615 Bielefeld, Germany*

Elisa Binda − *Department of Biotechnology and Life Sciences, University of Insubria, 21100 Varese, Italy*

Jörn Kalinowski − *Technology Platform Genomics, CeBiTec, Bielefeld University, 33615 Bielefeld, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.1c00170

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the public grant "Fondo di Ateneo per la Ricerca" 2019 to F.M., by the Biotechnology and Biological Sciences Research Council MfN Institute Strategic Programme grant (BBS/E/J/000PR9790) for the John Innes Centre (JIC) and by a Royal Society University Research Fellowship for A.W.T. A.A.-V. is a Ph.D. student of the "Life Science and Biotechnology" course at Università degli Studi dell'Insubria. We thank graduate student E. Salvadè of University of Insubria for her contribution to this research and Consorzio Interuniversitario per le Biotecnologie for supporting E.B. and F.B. We thank C. de Oliveira Martins (JIC) for technical support with mass spectrometry.

■ REFERENCES

(1) Sungthong, R., and Nakaew, N. (2015) The genus *Nonomuraea*: a review of a rare actinomycete taxon for novel metabolites. *J. Basic Microbiol. 55*, 554−565.

(2) Nazari, B., Forneris, C. C., Gibson, M. I., Moon, K., Schramma, K. R., and Seyedsayamdost, M. R. (2017) *Nonomuraea* sp. ATCC 55076 harbours the largest actinomycete chromosome to date and the kistamicin biosynthetic gene cluster. *MedChemComm 8*, 780−788.

(3) D'Argenio, V., Petrillo, M., Pasanisi, D., Pagliarulo, C., Colicchio, R., Talà, A., De Biase, M. S., Zanfardino, M., Scolamiero, E., Pagliuca, C., Gaballo, A., Cicatiello, A. G., Cantiello, P., Postiglione, I., Naso, B., Boccia, A., Durante, M., Cozzuto, L., Salvatore, P., Paolella, G., Salvatore, F., and Alifano, P. (2016) The complete 12 Mb genome and transcriptome of *Nonomuraea gerenzanensis* with new insights into its duplicated "magic" RNA polymerase. *Sci. Rep. 6*, 18.

(4) Narsing Rao, M. P., Dong, Z. Y., Jiao, J. Y., Zhou, Y., Zhao, J., Xiao, M., and Li, W. J. (2020) Genome sequence and comparative analysis of DRQ-2, the type strain of *Nonomuraea indica*. *Genomics 112*, 2842−2844.

(5) Greule, A., Izoré, T., Iftime, D., Tailhades, J., Schoppet, M., Zhao, Y., Peschke, M., Ahmed, I., Kulik, A., Adamek, M., Goode, R. J. A., Schittenhelm, R. B., Kaczmarski, J. A., Jackson, C. J., Ziemert, N., Krenske, E. H., De Voss, J. J., Stegmann, E., and Cryle, M. J. (2019) Kistamicin biosynthesis reveals the biosynthetic requirements for production of highly crosslinked glycopeptide antibiotics. *Nat. Commun. 10*, 2613.

(6) Yushchuk, O., Andreo-Vidal, A., Marcone, G. L., Bibb, M., Marinelli, F., and Binda, E. (2020) New molecular tools for regulation and improvement of A40926 glycopeptide antibiotic production in *Nonomuraea gerenzanensis* ATCC 39727. *Front. Microbiol. 11*, 8.

(7) Yue, X., Xia, T., Wang, S., Dong, H., and Li, Y. (2020) Highly efficient genome editing in *N. gerenzanensis* using an inducible CRISPR/Cas9−RecA system. *Biotechnol. Lett. 42*, 1699−1706.

(8) Nicolaou, K. C., Boddy, C. N. C., Bräse, S., and Winssinger, N. (1999) Chemistry, biology, and medicine of the glycopeptide antibiotics. *Angew. Chem., Int. Ed. 38*, 2096−2152.

(9) Goldstein, B. P., Selva, E., Gastaldo, L., Berti, M., Pallanza, R., Ripamonti, F., Ferrari, P., Denaro, M., Arioli, V., and Cassani, G. (1987) A40926, a new glycopeptide antibiotic with anti-*Neisseria* activity. *Antimicrob. Agents Chemother. 31*, 1961−1966.

(10) Marcone, G. L., Binda, E., Berini, F., and Marinelli, F. (2018) Old and new glycopeptide antibiotics: from product to gene and back in the post-genomic era. *Biotechnol. Adv. 36*, 534−554.
(11) Yushchuk, O., Ostash, B., Truman, A. W., Marinelli, F., and Fedorenko, V. (2020) Teicoplanin biosynthesis: unraveling the interplay of structural, regulatory, and resistance genes. *Appl. Microbiol. Biotechnol. 104*, 3279−3291.
(12) Grundy, W. E., Sinclair, A. C., Theriault, R. J., Goldstein, A. W., Rickher, C. J., Warren, H. B., Oliver, T. J., and Sylvester, J. C. (1956) Ristocetin, microbiologic properties. *Antibiot. Annu.*, 687−692.
(13) Truman, A. W., Kwun, M. J., Cheng, J., Yang, S. H., Suh, J. W., and Hong, H. J. (2014) Antibiotic resistance mechanisms inform discovery: identification and characterization of a novel *Amycolatopsis* strain producing ristocetin. *Antimicrob. Agents Chemother. 58*, 5687−5695.
(14) Spohn, M., Kirchner, N., Kulik, A., Jochim, A., Wolf, F., Muenzer, P., Borst, O., Gross, H., Wohlleben, W., and Stegmann, E. (2014) Overproduction of ristomycin a by activation of a silent gene cluster in *Amycolatopsis japonicum* MG417-CF17. *Antimicrob. Agents Chemother. 58*, 6185−6196.
(15) Sosio, M., Canavesi, A., Stinchi, S., and Donadio, S. (2010) Improved production of A40926 by *Nonomuraea* sp. through deletion of a pathway-specific acetyltransferase. *Appl. Microbiol. Biotechnol. 87*, 1633−1638.
(16) Alt, S., Bernasconi, A., Sosio, M., Brunati, C., Donadio, S., and Maffioli, S. I. (2019) Toward single-peak dalbavancin analogs through biology and chemistry. *ACS Chem. Biol. 14*, 356−360.
(17) Soriano, A., Rossolini, G. M., and Pea, F. (2020) The role of dalbavancin in the treatment of acute bacterial skin and skin structure infections (ABSSSIs). *Expert Rev. Anti-Infect. Ther. 18*, 415−422.
(18) Sosio, M., Stinchi, S., Beltrametti, F., Lazzarini, A., and Donadio, S. (2003) The gene cluster for the biosynthesis of the glycopeptide antibiotic A40926 by *Nonomuraea* species. *Chem. Biol. 10*, 541−549.
(19) Li, Y. S., Ho, J. Y., Huang, C. C., Lyu, S. Y., Lee, C. Y., Huang, Y. T., Wu, C. J., Chan, H. C., Huang, C. J., Hsu, N. S., Tsai, M. D., and Li, T. L. (2007) A unique flavin mononucleotide-linked primary alcohol oxidase for glycopeptide A40926 maturation. *J. Am. Chem. Soc. 129*, 13384−13385.
(20) Kaniusaite, M., Tailhades, J., Kittilä, T., Fage, C. D., Goode, R. J. A., Schittenhelm, R. B., and Cryle, M. J. (2021) Understanding the early stages of peptide formation during the biosynthesis of teicoplanin and related glycopeptide antibiotics. *FEBS J. 288*, 507−529.
(21) Binda, E., Marcone, G. L., Pollegioni, L., and Marinelli, F. (2012) Characterization of VanYn, a novel D,D-peptidase/D,D-carboxypeptidase involved in glycopeptide antibiotic resistance in *Nonomuraea* sp. ATCC 39727. *FEBS J. 279*, 3203−3213.
(22) Marcone, G. L., Binda, E., Carrano, L., Bibb, M., and Marinelli, F. (2014) Relationship between glycopeptide production and resistance in the actinomycete *Nonomuraea* sp. ATCC 39727. *Antimicrob. Agents Chemother. 58*, 5191−5201.
(23) Lo Grasso, L., Maffioli, S., Sosio, M., Bibb, M., Puglia, A. M., and Alduina, R. (2015) Two master switch regulators trigger A40926 biosynthesis in *Nonomuraea* sp. strain ATCC 39727. *J. Bacteriol. 197*, 2536−2544.
(24) Alduina, R., Tocchetti, A., Costa, S., Ferraro, C., Cancemi, P., Sosio, M., and Donadio, S. (2020) A two-component regulatory system with opposite effects on glycopeptide antibiotic biosynthesis and resistance. *Sci. Rep. 10*, 6200.
(25) Naruse, N., Tenmyo, O., Kobaru, S., Hatori, M., Tomita, K., Hamagishi, Y., and Oki, T. (1993) New antiviral antibiotics, kistamicins A and B I. Taxonomy, production, isolation, physico-chemical properties and biological activities. *J. Antibiot. 46*, 1804−1811.
(26) Naruse, N., Oka, M., Konishi, M., and Oki, T. (1993) New Antiviral Antibiotics, Kistamicins A and B II. Structure determination. *J. Antibiot. 46*, 1812−1818.
(27) Waglechner, N., McArthur, A. G., and Wright, G. D. (2019) Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nat. Microbiol. 4*, 1862−1871.
(28) Malabarba, A., Ciabatti, R., Scotti, R., Goldstein, B. P., Ferrari, P., Kurz, M., Andreini, B. P., and Denaro, M. (1995) New semisynthetic glycopeptides MDL 63,246 and MDL 63,042, and other amide derivatives of antibiotic A-40,926 active against highly glycopeptide-resistant VanA enterococci. *J. Antibiot. 48*, 869−883.
(29) Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., and Weber, T. (2019) AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res. 47*, W81−W87.
(30) Adamek, M., Alanjary, M., Sales-Ortells, H., Goodfellow, M., Bull, A. T., Winkler, A., Wibberg, D., Kalinowski, J., and Ziemert, N. (2018) Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in *Amycolatopsis* species. *BMC Genomics 19*, 426.
(31) Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004) MAUVE: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res. 14*, 1394−1403.
(32) Alduina, R., Sosio, M., and Donadio, S. (2018) Complex regulatory networks governing production of the glycopeptide A40926. *Antibiotics (Basel, Switz.) 7*, 30.
(33) Alduina, R., Piccolo, L. L., D'Alia, D., Ferraro, C., Gunnarsson, N., Donadio, S., and Puglia, A. M. (2007) Phosphate-controlled regulator for the biosynthesis of the dalbavancin precursor A40926. *J. Bacteriol. 189*, 8120−8129.
(34) Binda, E., Marcone, G. L., Berini, F., Pollegioni, L., and Marinelli, F. (2013) *Streptomyces* spp. as efficient expression system for a D,D-peptidase/D,D-carboxypeptidase involved in glycopeptide antibiotic resistance. *BMC Biotechnol. 13*, 24.
(35) Pootoolal, J., Thomas, M. G., Marshall, C. G., Neu, J. M., Hubbard, B. K., Walsh, C. T., and Wright, G. D. (2002) Assembling the glycopeptide antibiotic scaffold: the biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL15009. *Proc. Natl. Acad. Sci. U. S. A. 99*, 8962−8967.
(36) Ara, I., Kudo, T., Matsumoto, A., Takahashi, Y., and Omura, S. (2007) *Nonomuraea bangladeshensis* sp. nov. and *Nonomuraea coxensis* sp. nov. *Int. J. Syst. Evol. Microbiol. 57*, 1504−1509.
(37) Marcone, G. L., Beltrametti, F., Binda, E., Carrano, L., Foulston, L., Hesketh, A., Bibb, M., and Marinelli, F. (2010) Novel mechanism of glycopeptide resistance in the A40926 producer *Nonomuraea* sp. ATCC 39727. *Antimicrob. Agents Chemother. 54*, 2465−2472.
(38) Taurino, C., Frattini, L., Marcone, G. L., Gastaldo, L., and Marinelli, F. (2011) *Actinoplanes teichomyceticus* ATCC 31121 as a cell factory for producing teicoplanin. *Microb. Cell Fact. 10*, 82.
(39) Puk, O., Bischoff, D., Kittel, C., Pelzer, S., Weist, S., Stegmann, E., Süssmuth, R. D., and Wohlleben, W. (2004) Biosynthesis of chloro-β-hydroxytyrosine, a nonproteinogenic amino acid of the peptidic backbone of glycopeptide antibiotics. *J. Bacteriol. 186*, 6093−6100.
(40) Yushchuk, O., Binda, E., and Marinelli, F. (2020) Glycopeptide antibiotic resistance genes: distribution and function in the producer actinomycetes. *Front. Microbiol. 11*, 1.
(41) Zerilli, L. F., Edwards, D. M. F., Borghi, A., Gallo, G. G., Selva, E., Denaro, M., and Lancini, G. C. (1992) Determination of the acyl moieties of the antibiotic complex A40926 and their relation with the membrane lipids of the producer strain. *Rapid Commun. Mass Spectrom. 6*, 109−114.
(42) Xu, M., Wang, W., Waglechner, N., Culp, E. J., Guitor, A. K., and Wright, G. D. (2020) GPAHex - a synthetic biology platform for type IV−V glycopeptide antibiotic production and discovery. *Nat. Commun. 11*, 5232.
(43) Culp, E. J., Waglechner, N., Wang, W., Fiebig-Comyn, A. A., Hsu, Y. P., Koteva, K., Sychantha, D., Coombes, B. K., Van Nieuwenhze, M. S., Brun, Y. V., and Wright, G. D. (2020) Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature 578*, 582−587.

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

(44) Horbal, L., Kobylyanskyy, A., Yushchuk, O., Zaburannyi, N., Luzhetskyy, A., Ostash, B., Marinelli, F., and Fedorenko, V. (2013) Evaluation of heterologous promoters for genetic analysis of *Actinoplanes teichomyceticus* - producer of teicoplanin, drug of last defense. *J. Biotechnol. 168*, 367−372.
(45) Horbal, L., Kobylyanskyy, A., Truman, A. W., Zaburranyi, N., Ostash, B., Luzhetskyy, A., Marinelli, F., and Fedorenko, V. (2014) The pathway-specific regulatory genes, *tei15** and *tei16**, are the master switches of teicoplanin production in *Actinoplanes teichomyceticus*. *Appl. Microbiol. Biotechnol. 98*, 9295−9309.
(46) Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000) *Practical Streptomyces genetics*; John Innes Foundation, Norwich, England.
(47) Holding, A. N., and Spencer, J. B. (2008) Investigation into the mechanism of phenolic couplings during the biosynthesis of glycopeptide antibiotics. *ChemBioChem 9*, 2209−2214.
(48) Bekiesch, P., Zehl, M., Domingo-Contreras, E., Martín, J., Pérez-Victoria, I., Reyes, F., Kaplan, A., Rückert, C., Busche, T., Kalinowski, J., and Zotchev, S. B. (2020) Viennamycins: lipopeptides produced by a *Streptomyces* sp. *J. Nat. Prod. 83*, 2381−2389.
(49) Gordon, D., and Green, P. (2013) Consed: A Graphical editor for next-generation sequencing. *Bioinformatics 29*, 2936−2937.
(50) Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics 30*, 2068−2069.
(51) Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics 28*, 1647−1649.
(52) Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol. 35*, 1547−1549.
(53) Gao, B., and Gupta, R. S. (2012) Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. *Microbiol. Mol. Biol. Rev. 76*, 66−112.

https://doi.org/10.1021/acschembio.1c00170
ACS Chem. Biol. 2021, 16, 915−928

Supporting Information

Genomic-led discovery of a novel glycopeptide antibiotic by *Nonomuraea coxensis* DSM 45129

Oleksandr Yushchuk[§], Natalia M.Vior[†], Andres Andreo-Vidal[§], Francesca Berini[§], Kristian Rückert[‡], Tobias Busche[‡], Elisa Binda[§], Jörn Kalinowski[‡], Andrew W. Truman[†*] and Flavia Marinelli[§*]

[§]Department of Biotechnology and Life Sciences, University of Insubria, via J. H. Dunant 3, 21100 Varese, Italy.

[†]Department of Molecular Microbiology, John Innes Centre, Norwich, NR4 7UH, United Kingdom

[‡]Technology Platform Genomics, CeBiTec, Bielefeld University, Sequenz 1, 33615 Bielefeld, Germany

* Corresponding author for chemistry Andrew W. Truman (andrew.truman@jic.ac.uk) and for microbiology Flavia Marinelli (flavia.marinelli@uninsubria.it).

Inventory of Supporting Information

Composition of media used in the study:

Unless otherwise stated, components are from Sigma-Aldrich, St. Louis, MO, United States.

ISP2 (g/L of distilled water):
Yeast extract – 4;
Malt extract – 10;
Dextrose – 4;
Agar – 20;
pH 7.5.

ISP2l (g/L of distilled water):
Yeast extract – 4;
Malt extract – 10;
Dextrose – 4;
pH 7.5.

ISP2lm (g/L of distilled water):
Yeast extract – 4;
Malt extract – 10;
pH 7.5.

ISP3 (g/L of tap water):
Fine ground whole oats (Kozub, Poltava, Ukraine) – 34;
Agar – 20;
pH 7.5.

VSP (g/L of distilled water):
Soluble starch (Difco, Franklin Lakes, NJ, United States) – 24;
Dextrose – 1;
Meat extract – 3;
Yeast extract – 5;
Tryptose – 5;
L-proline – 0.5;
Sucrose – 50;
pH 7.5.

VM0.1 (g/L of distilled water):
Soluble starch (Difco) – 2.4;
Dextrose – 0.1;
Meat extract – 0.3;
Yeast extract – 0.5;
Tryptose – 0.5;
Agar – 20;
pH 7.2.

E25 (g/L of distilled water):
Dextrose – 25;

Meat extract – 4;
Yeast extract – 1;
Soybean flour – 10;
Bacto peptone – 4;
NaCl – 2.5;
pH 7.5

E26 (g/L of distilled water):
Dextrose – 25;
Soy flour – 20;
Yeast extract – 4;
NaCl – 1.25;
$CaCO_3$ – 5;
pH 7.5.

E27 (g/L of distilled water):
Soy flour – 20;
Yeast extract – 4;
NaCl – 1.25;
$CaCO_3$ – 5;
pH 7.5.

R5 (g/L of distilled water):
Sucrose – 103;
Glucose – 10;
K_2SO_4 – 0.25;
$MgCl_2$ x 6 H_2O 10.12;
Casamino acids – 0.1;
Yeast extract – 5;
TES buffer – 5.73;
Trace elements solution[1] – 2 mL
pH 7.2
ddH_2O add up to 900 ml.
To be added at the time of use:
$CaCl_2$ 3.68% (w/v) 80 mL
KH_2PO_4 0.54% (w/v) 10 mL
L-proline 20% (w/v) 15 mL

FM2 (g/L of distilled water):
Dextrose – 30;
Soy flour – 30;
Yeast extract – 8;
Malt extract – 15;
$CaCO_3$ – 5;
L-valine – 1;
pH 7.5.

TM1 (g/L of distilled water):

Dextrose – 10;
Malt extract – 30;
Yeast extract – 2.5;
Soybean flour – 15;
$CaCO_3$ – 4;
pH 7.5.

TM1m (g/L of distilled water):
Malt extract – 30;
Yeast extract – 2.5;
Soybean flour – 15;
$CaCO_3$ – 4;
pH 7.5.

5

Table S1. Summary of the putative BGCs predicted within *N. coxensis* genome using antiSMASH

BGC-like region	Location, bp	Most similar BGC in MIBIG database/ similarity (%)	Comments
1	219,542-229,750	desferrioxamine E BGC from *Streptomyces* sp. ID38640/100	Identical to desferrioxamine E BGC
2	309,384-330,034	blasticidin S biosynthetic gene cluster from *Streptomyces griseochromogenes*/7	Only two genes show similarity to blasticidin BGC
3	1,918,176-1,932,871	no	Putative HIV-1 Rev response element containing BGC
4	2,580,016-2,596,430	isorenieratene BGC from *Streptomyces griseus* subsp. *griseus* NBRC 13350/28	Only two genes show similarity to isorenieratene BGC
5	2,605,255-2,664,589	natamycin BGC from *Streptomyces gilvosporeus*/9	Encodes for an NRPS-PKS hybrid, two genes for ABC transporters share similarity to transporter genes from natamycin BGC
6	2,667,388-2,731,534	lysolipin I biosynthetic gene cluster from *Streptomyces tendae*/43	Contains type I PKS and NRPS genes, putative tailoring genes share similarity to lysolipin I BGC tailoring genes
7	2,813,563-2,896,967	A40926 BGC from *N. gerenzanensis*/61	*noc* BGC
8	2,931,058-2,989,541	acarviostatin I03 BGC from *Streptomyces coelicoflavus* ZG0656/22	Only six genes show similarity to acarviostatin I03 BGC
9	4,126,909-4,149,389	no	Putative class IV lanthipeptide BGC
10	3,075,747-3,094,693	chlortetracycline BGC from *Kitasatospora aureofaciens*/5	Only two genes show similarity to chlortetracycline BGC
11	4,241,436-4,337,989	kirromycin BGC from *Streptomyces collinus* Tü 365/10	Four genes show similarity to kirromycin BGC
12	5,036,130-5,090,294	meridamycin BGC from *Streptomyces* sp. NRRL 30748/5	Transporter gene shows similarity to meridamycin BGC transporter
13	5,369,623-5,479,734	butyrolactol A biosynthetic gene cluster from *Streptomyces* sp. NBRC 110030/53	Superficial similarity to butyrolactol A BGC
14	5,619,580-5,685,245	calicheamicin BGC from *Micromonospora echinospora*/2	Only two genes show similarity to calicheamicin BGC

15	6,052,873-6,094,253	mirubactin BGC from *Actinosynnema mirum* DSM 43827/21	Three transport-related genes show similarity to mirubactin BGC
16	6,544,775-6,564,867	no	Putative terpene BGC
17	6,756,057-6,797,112	lagunapyrone A BGC from *Streptomyces* sp. MP131-18/22	Only two genes show similarity to lagunapyrone BGC
18	7,117,439-7,169,843	no	Putative NRPS/PKS BGC
19	7,766,570-7,773,806	no	Putative siderophore BGC
20	7,947,755-7,958,612	no	Putative RiPP BGC
21	7,975,769-7,998,173	no	Putative lassopeptide BGC
22	8,119,609-8,148,927	no	Putative betalctone BGC
23	8,163,803-8,209,012	chlorothricin BGC from *Streptomyces antibioticus*/6	Only two genes show similarity to chlorothricin BGC
24	8,274,638-8,320,463	murayaquinone BGC from *Streptomyces griseoruber*/3	One gene show similarity to murayaquinone BGC
25	8,672,394-8,685,664	no	Putative siderophore BGC
26	8,703,672-8,724,947	geosmin BGC from *Nostoc punctiforme* PCC 73102/100	Identical to geosmin BGC
27	8,874,658-8,903,559	hopene BGC from *Streptomyces coelicolor* A3(2)	Six genes show similarity tp hopene BGC

Table S2. *Nonomuraea* genomes available in GenBank.

Strain	Assembly accession number	Assembly state
Nonomuraea sp. ATCC 39727 (=*N. gerenzanensis* ATCC 39727)	LT559118.1 (GenBank accession)	complete genome
Nonomuraea polychroma DSM 43925	ASM401150v1	single contig
Nonomuraea sp. FMUSA5-5	ASM1203427v1	contigs
Nonomuraea sp. 160415	ASM589312v1	contigs
Nonomuraea sp. SBT364	*Nonomuraea* sp. SBT364_SPAdes_SSPACE	scaffolds
Nonomuraea sp. KC712	ASM434901v1	contigs
Nonomuraea sp. 6K102	ASM435280v1	contigs
Nonomuraea sp. K271	ASM990821v1	contigs
Nonomuraea sp. KC310	ASM434868v1	contigs
Nonomuraea sp. KC401	ASM577475v1	contigs
Nonomuraea sp. CH32	ASM434899v1	contigs
Nonomuraea sp. KC201	ASM434834v1	contigs
Nonomuraea sp. KC333	ASM323639v1	contigs
Nonomuraea sp. NEAU-YG30	ASM331339v1	contigs
Nonomuraea kuesteri NRRL B-24325	ASM71613v1	contigs
Nonomuraea sp. PA05	ASM808604v1	contigs
Nonomuraea sp. p1410	ASM976092v1	scaffolds
Nonomuraea sp. WAC 01424	ASM394731v1	contigs
Nonomuraea sp. NBRC 110462	ASM141775v1	few scaffolds
Nonomuraea sp. C10	ASM801751v1	few contigs
Nonomuraea sp. WYY166	ASM949707v1	complete genome
Nonomuraea sp. ATCC 55076	ASM205745v1	complete genome
Nonomuraea zeae DSM 100528	ASM588972v1	contigs
Nonomuraea turkmeniaca DSM 43926	ASM588973v1	contigs
Nonomuraea indica DRQ-2	ASM285074v1	scaffolds
Nonomuraea maritima CGMCC 4.5681	IMG-taxon 2675903066 annotated assembly	scaffolds
Nonomuraea coxensis DSM 45129	ASM37988v1	scaffolds
Nonomuraea candida NRRL B-24552	Doro.v1.0	scaffolds
Nonomuraea jiangxiensis CGMCC 4.6533	IMG-taxon 2675903065 annotated assembly	scaffolds
Nonomuraea pusilla DSM 43357	IMG-taxon 2599185218 annotated assembly	scaffolds
Nonomuraea wenchangensis CGMCC 4.5598	IMG-taxon 2675903060 annotated assembly	scaffolds
Nonomuraea fuscirosea CGMCC 4.7104	ASM300193v1	contigs

Nonomuraea solani CGMCC 4.7037	IMG-taxon 2675903140 annotated assembly	contigs
Nonomuraea phyllanthi PA1-10	ASM633498v2	contigs

Table S3. Set of house-keeping proteins used for reconstruction of multi locus phylogeny (MLP) of *Nonomuraea* species and locus tags of their orthologues in *S. coelicolor*.

Protein	**Protein in *S. coelicolor***
30S ribosomal protein S5 201	SCO4719
Dimethyladenosine transferase (KsgA) 286	SCO3149
Ribosomal protein S13 126	SCO4727
30S ribosomal protein S7 156	SCO4660
50S ribosomal protein L13P 147	SCO4734
Elongation factor Tu 397	SCO4662
30S ribosomal protein S9 170	SCO4735
50S ribosomal protein L11 144	SCO4648
30S ribosomal protein S15 95	SCO5736
50S ribosomal protein L1 241	SCO4649
Ribosomal protein S19 93	SCO4706
Gyrase B 686 RNA polymerase β' subunit (RpoC) 1299	SCO4655
Gyrase A 857 50S	SCO3873
50S ribosomal protein L10/L16 176	SCO4652
Ribosomal protein L4 219	SCO1505
Preprotein translocase subunit SecY 437	SCO4722
Ribosomal protein L2 278	SCO5624
RNA polymerase subunit alpha 340	SCO4792
50S ribosomal protein L22 125	SCO4707
Threonyl-tRNA-synthetase 658	SCO3778
30S ribosomal protein S3 277	SCO4708
Molecular chaperone DnaK (Hsp70) 618	SCO3671
50S ribosomal protein L14 122	SCO4712
Chaperonin GroEL (Hsp60) 541	SCO4762
Ribosomal protein L5 185	SCO4714
O-sialoglycoprotein endopeptidase	SCO4752
374 30S ribosomal protein S8 132	SCO4716
Phosphatidate cytidylyltransferase 391	SCO5628
50S ribosomal protein L6 179	SCO4717
CDP-diglyceride synthase (CdsA) 391	SCO5628
Ribosomal protein L3 214	SCO4702

Table S4. Prediction of NRPS A-domain specificities in *noc*-encoded and related GPA NRPSs; Hpg – 4-hydroxyphenylglycine; Tyr – tyrosine; Dpg – 3,5-dihydroxyphenylglycine; Bht – β-hydroxytyrosine. A-domain specificities were predicted using NRPSpredictor2[2].

BGC	NRPS	Position of the A-domain	Extracted Stachelhaus code	A-domain specificity prediction, based on the Stachelhaus code
dbv	Dbv25	A1	DAFHLGLLCK	Hpg
		A2	DASTVAAVCK	Tyr
	Dbv26	A3	DAYNAGTLCK	Dpg
	Dbv17	A4	DIFHLGLLCK	Hpg
		A5	DALHLGLLCK	Hpg
		A6	DASTVAAVCK	Tyr
	Dbv16	A7	DPYHGGTLCK	Dpg
noc	NocA	A1	DAFHLGLLCK	Hpg
		A2	DASTVAAVCK	Tyr
	NocB	A3	DAYNAGTLCK	Dpg
	NocC	A4	DIFHLGLLCK	Hpg
		A5	DALHLGLLCK	Hpg
		A6	DASTVAAVCK	Tyr
	NocD	A7	DPYHGGTLCK	Dpg
WAC01424 GPA BGC	WP_125645195	A1	DAYHLGLLCK	Hpg
		A2	DASTVAAVCK	Tyr
	WP_125645193	A3	DAYNAGTLCK	Dpg
	WP_125645191	A4	DIFHLGLLCK	Hpg
		A5	DALHLGLLCK	Hpg
		A6	DASTVAAVCK	Tyr
	WP_125645189	A7	DPYHGGTLCK	Dpg
tei	TeiA	A1	DAFHLGLLCK	Hpg
		A2	DASTVAAVCK	Tyr
	TeiB	A3	DAYNLGTLCK	Dpg
	TeiC	A4	DIFHLGLLCK	Hpg
		A5	DALHLGLLCK	Hpg
		A6	DASTIAGVCK	Tyr
	TeiD	A7	DPYHGGTLCK	Dpg
NRRL 2430 ristocetin BGC	AIE77057.1	A1	DACHLGLLCK	Hpg
		A2	DTSKTAAICK	Bht
	AIE77058.1	A3	DPYNQGTFCK	Dpg
	AIE77059.1	A4	DIFHLGLLCK	Hpg
		A5	DAVHLGLLCK	Hpg
		A6	DASTLGAICK	Bht
	AIE77060.1	A7	DPYHGGTLCK	Dpg

Table S5. Sources and accession numbers of GPA BGC-encoded halogenases and sulphotransferases used for the phylogenetic reconstruction in Figure S7.

Sulphotransferase	Protein accession number	Source
Teg12	ACJ60995	Uncultured soil bacterium clone D30 TEG GPA biosynthetic gene cluster (BGC)
Teg13	ACJ60996	Uncultured soil bacterium clone D30 TEG GPA BGC
Teg14	ACJ60997	Uncultured soil bacterium clone D30 TEG GPA BGC
n/a	AGO98990	*Streptomyces* sp. WAC4229 pekiskomycin BGC
Pek25	AGF91760	*Streptomyces* sp. WAC1420 pekiskomycin BGC (*pek*)
n/a	AGS49779	Uncultured bacterium esnapd15 GPA BGC
StaL	AAM80529	*Streptomyces toyocaensis* NRRL 15009 A47934 BGC (*sta*)
n/a	WP_125645175	*Nonomuraea* sp. WAC01424 GPA BGC
Auk20	AGS77324	*Actinoplanes* sp. ATCC 53533 UK-68,597 BGC (*auk*)
n/a	WP_030455526	*Herbidospora cretaceae,* sulphotransferase with unknown function
Halogenase	**Protein accession number**	**Source**
Veg13	ACJ60955	Uncultured soil bacterium clone B128 VEG GPA BGC
VhaA	CCD33142	*Amycolatopsis orientalis* ATCC19795 vancomycin BGC
BhaA	CAA76550	*Amycolatopsis balhimycina* DSM 5908 balhimycin BGC
CepH	n/a	*Amycolatopsis orientalis* chloroeremomycin BGC (*cep*)
n/a	AGO98992	*Streptomyces* sp. WAC4229 pekiskomycin BGC
Pek27	AGF91762	*Streptomyces* sp. WAC1420 pekiskomycin BGC (*pek*)
	WP_125645180	*Nonomuraea* sp. WAC01424 GPA BGC
Teg16	ACJ60999	Uncultured soil bacterium clone D30 TEG GPA biosynthetic gene cluster (BGC)
Dbv10	CAD91205	*Nonomuraea gerenzanensis* ATCC 39727 A40926 BGC (*dbv*)
Noc24	n/a	*Nonomuraea coxensis* DSM 45129 A50926 BGC (*noc*)
ComH	AAK81830	*Streptomyces lavendulae* complestatin BGC (*com*)
KisU	AQZ69237	*Nonomuraea* sp. ATCC 55076 kistamicin BGC (*kis*)
n/a	WP_173522348	*Nonomuraea* sp. NN258 *kis*-like GPA BGC
Auk21	AGS77325.1	*Actinoplanes* sp. ATCC 53533 UK-68,597 BGC (*auk*)
n/a	AGS49782	Uncultured bacterium esnapd15 GPA BGC
Tei8	CAG15020.1	*Actinoplanes teichomyceticus* ATCC 31121 teicoplanin BGC (*tei*)
CA915-35	ADU56156	Uncultured organism CA915 GPA BGC
Auk23	AGS77327.1	*Actinoplanes* sp. ATCC 53533 UK-68,597 BGC (*auk*)
CA37-38	ADU56061	Uncultured organism CA37 GPA BGC
n/a	WP_125645176	*Nonomuraea* sp. WAC01424 GPA BGC
n/a	AGS49780	Uncultured bacterium esnapd15 GPA BGC
StaK	AAM80530.1	*Streptomyces toyocaensis* NRRL 15009 A47934 BGC (*sta*)

SCO1275	2SCG18	*Streptomyces coelicolor* A3(2), halogenase with unknown function

n/a – not available.

13

Table S6. Bacterial strains and plasmids used in this work.

Name	Description	Source or reference
N. gerenzanensis	Wild type, A40926 producer	ATCC 39727
N. coxensis	Wild type, A50926 producer	DSM 45129
B. subtilis	GPA test-culture	ATCC 6633
N. coxensis pSET152A$^+$	Wild type derivative carrying pSET152A	This work
N. coxensis pSAD3$^+$	Wild type derivative carrying pSAD3	This work
N. coxensis pSAD4$^+$	Wild type derivative carrying pSAD4	This work
N. coxensis pSAD29$^+$	Wild type derivative carrying pSAD29	This work
E. coli DH5α	General cloning host	MBI Fermentas
E. coli ET12567 (pUZ8002$^+$)	(*dam-13*::Tn9 *dcm-6*), pUZ8002$^+$ (Δ*oriT*), used for conjugative transfer of DNA into *Nonomuraea* spp.	1
pSET152A	pSET152 derivative, containing *aac(3)IVp* from pIJ773	3, 4
pSAD3	pSET152A derivative, containing *dbv3* under the control of *aac(3)IVp*	5
pSAD4	pSET152A derivative, containing *dbv4* under the control of *aac(3)IVp*	5
pSAD29	pSET152A derivative, containing *dbv29* under the control of *aac(3)IVp*	This work

Table S7. Oligonucleotide primers used in this work.

Primer	Nucleotide sequence (5'-3')*	Purpose
dbv29_F	TTTGATATCGGAGGGCGGTGGTGACCGGCGGCAC	Cloning of *dbv29* into pSET152A
dbv29_R	TTTGAATTCTCAGGGCCGGATCGACAACGCG	
PAM_seq_F	GATGTCATCAGCGGTGGAG	Verification of recombinant strains
PAM_seq_R	TGAGCGGATAACAATTTCA	
dbv29_seq_int	TGTCACGGCAGTTCGGCTC	
dbv3_seq_R	CCAGCGCTGGACCGCCTGC	
dbv4_R	TTTGAATTCTCCACTCGTGCTCATCCAG	
aac(3)IV_F	ATCGACTGATGTCATCAGCG	Amplification of *aac(3)IV*
aac(3)IV_R	CGAGCTGAAGAAAGACAAT	

* recognition sites of restriction endonucleases are underlined.

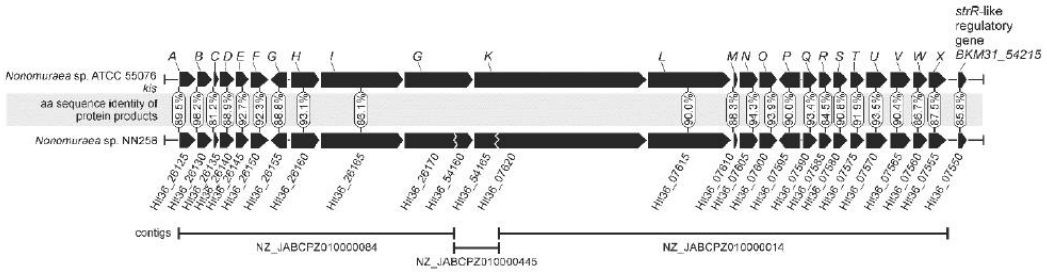


Figure S1. Comparison of kistamicin – *kis*[6] – BGC from *Nonomuraea* sp. ATCC 55076 with the putative type V GPA BGC from *Nonomuraea* sp. NN258. The sequence of the latter was reconstructed from three contigs of the *Nonomuraea* sp. NN258 genomic draft. Both clusters share an identical genetic organization and the same numbers of genes; the corresponding protein products possess high levels of amino acid sequence identity.

Figure S2. MLP tree of *Nonomuraea* spp. whose genomic records are available (Table S2). Well-supported clades containing GPA producers are highlighted in orange. Strains known to possess GPA BGCs in their genomes are in blue. The tree was built using concatenated protein sequences of 30 housekeeping proteins, orthologous to *S. coelicolor* proteins given in Table 3S. Phylogeny was inferred using Mega X[7] with Maximum Likelihood method and JTT matrix-based model with Gamma distribution and invariable sites; final topology and bootstrap support values (indicated at the base of the branching points) are based on 500 bootstraps.

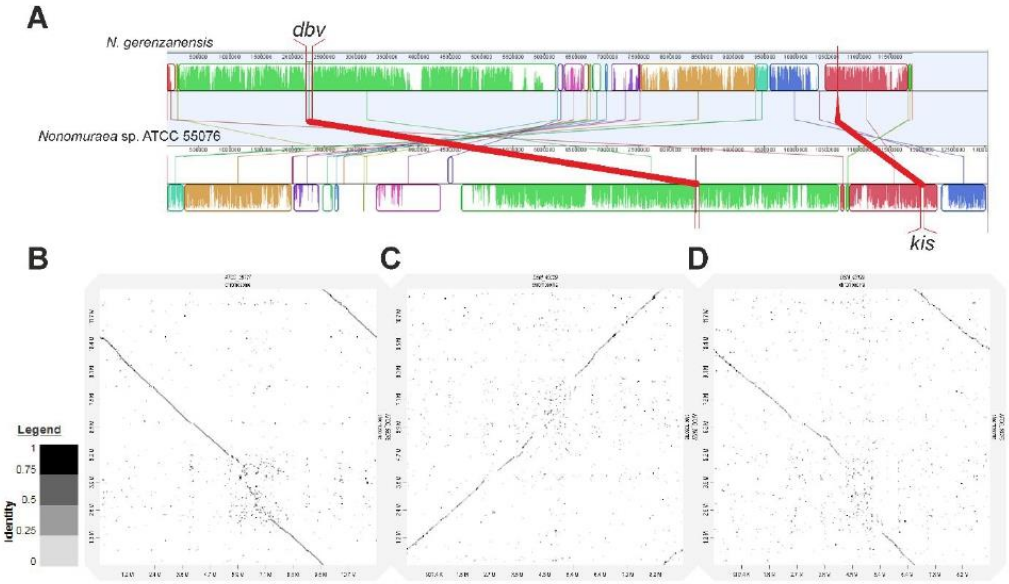


Figure S3. (A) MAUVE[8] alignment of *N. gerenzanensis* and *Nonomuraea* sp. ATCC 55076 genomes; significant similarity of both genomes could be observed. Location of the insertion loci for the *dbv* and *kis* BGCs is indicated in red. (B) Dot plots of the *N. gerenzanensis* and *Nonomuraea* sp. ATCC 55076 chromosomes showing similarity of both genomes. (C) Dot plots of *N. coxensis* and *Nonomuraea* sp. ATCC 55076 and (D) of *N. coxensis* and *N. gerenzanensis,* showing that the genome of *N. coxensis* is less similar to the other two. Dot plots were constructed using D-genies[9].

Figure S4. (A) Organization of the NRPSs encoded in *noc, dbv* and WAC 01424 GPA BGCs. Epimerization (E-) domain in module 3 of WAC 01424 GPA NRPS (orange) might be non-functional. (B) Sequence alignments of the seven A-domains of the three NRPSs, which are highly similar and share the same amino acid specificity (see also Table S4). Alignments were built in Clustal Omega[10]. Hpg – 4-hydroxyphenylglycine; Tyr – tyrosine; Dpg – 3,5-dihydroxyphenylglycine.

Figure S5. Pairwise alignments of the putative Dbv4 binding regions[11] found in the promoters of *dbv30* and *noc9* as well as in *dbv14* and *noc20.* Inverted repeats are in boxes with arrows, asterisks indicate identical nucleotides. Alignment was built with Clustal Omega[10].

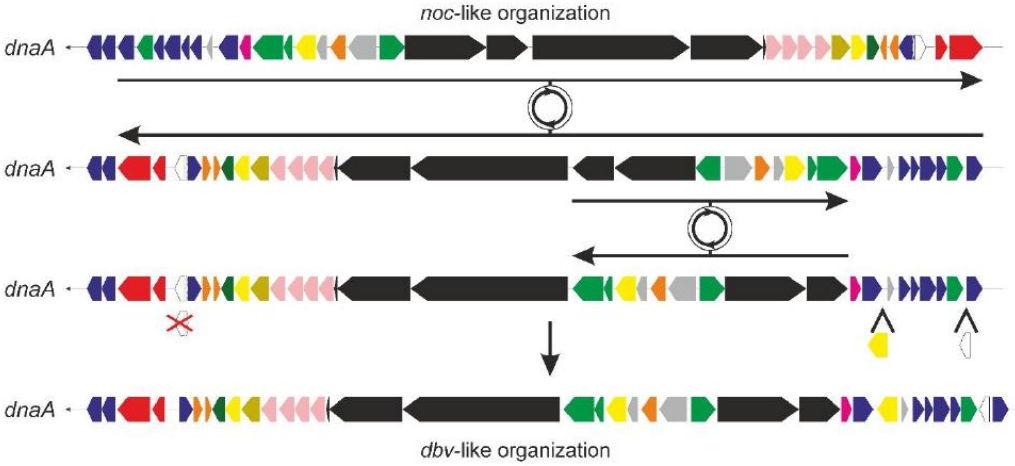


Figure S6. A scheme for the proposed recombination events involving two chromosomal inversions (as well as gene loss/gain events) that may have led to the uncommon organization of *dbv* BGC, starting from the *noc* BGC organization which is similar to those of other know GPA BGCs. In the *dbv* BGC, NRPS genes are encoded on different strands and are separated by other biosynthetic genes, which contrasts with the *noc* and other known GPA BGCs.

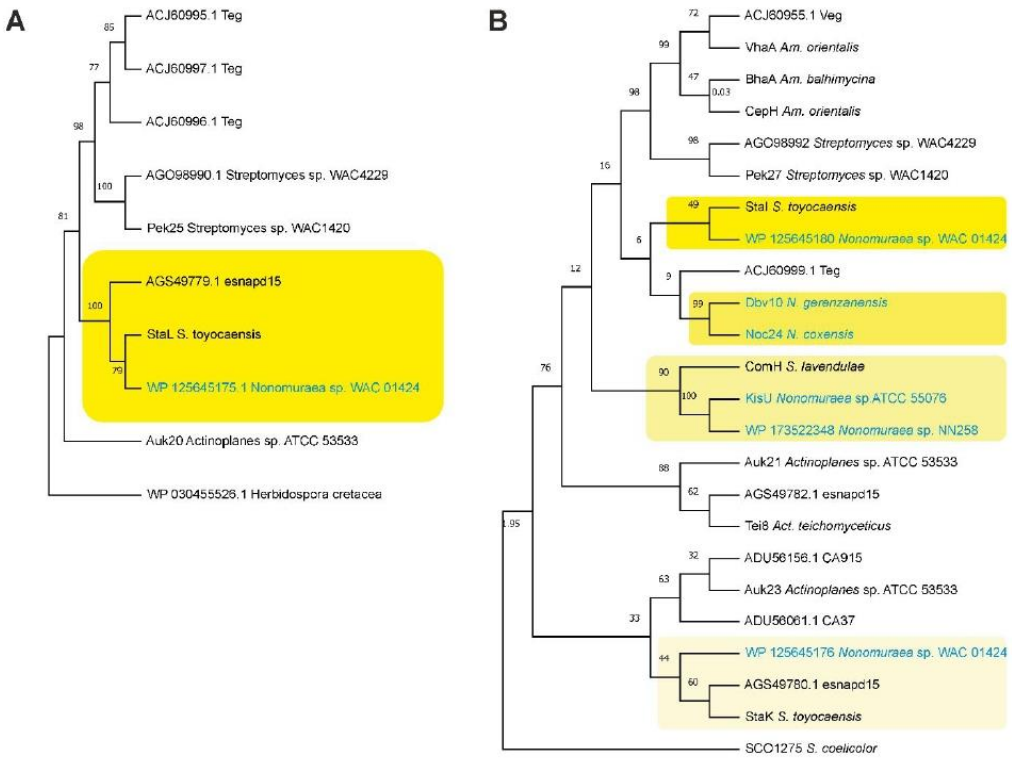
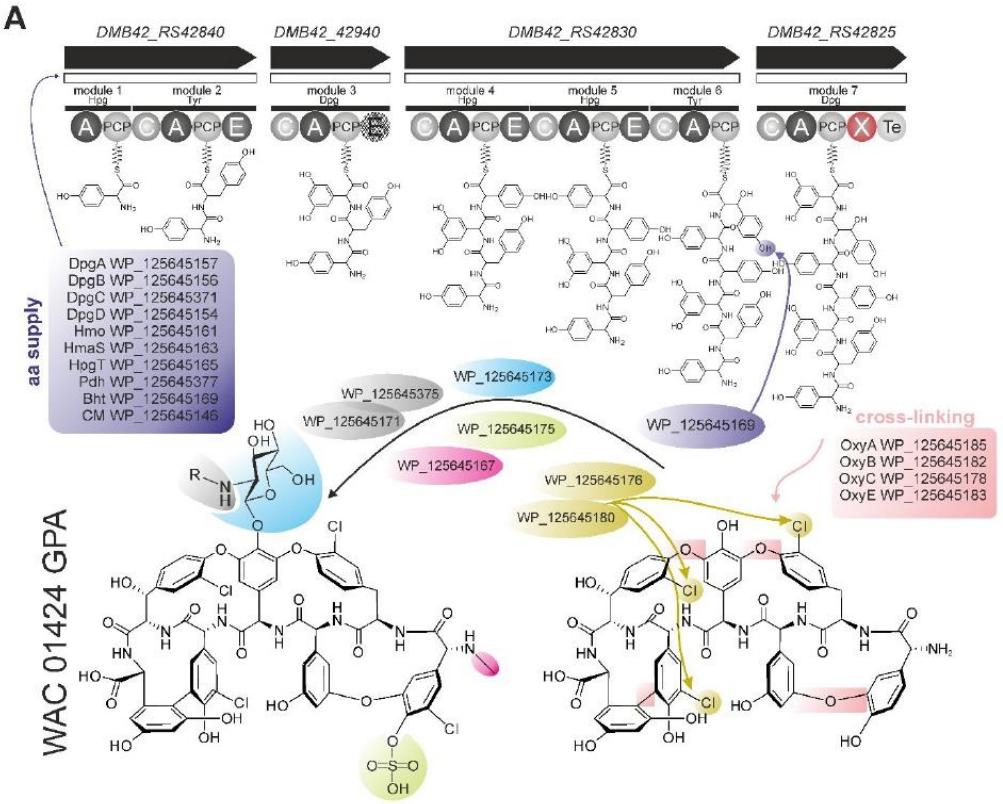


Figure S7. (A) Cladogram of the sulphotransferases, encoded within GPA BGCs. The sulphotransferase encoded within WAC 01424 putative GPA BGC (blue) is closely related to the one encoded within A47934 BGC (StaL) from *Streptomyces toyocaensis*. (B) Cladogram of the halogenases encoded within GPA BGCs. Halogenases from *Nonomuraea* spp. are in blue. Clades grouping different halogenases from *Nonomuraea* spp. are highlighted in different shades of yellow. The two halogenases encoded within *Nonomuraea* sp. WAC 01424 GPA BGC are sharing clades with halogenases coded within A47934 BGC – StaI and StaK, being distantly related to Dbv10 and Noc24. Phylogeny was inferred using Mega X[7] with Maximum Likelihood method and JTT matrix-based model; final topologies and bootstrap support values (indicated at the base of the branching points) are based on 500 bootstraps. Detailed information for the amino acid sequences used in this reconstruction is given in Table S5.



B

```
WAC         GELPWTPVMRLWGDGVASAGFAQWVVVGAPSDLTEKVLVAGVAALVDTHDMLRLRVQ---      57
A47934      GDVARTPMTKVWGPG-----AARWVTVGAPGDLSVDTLAAGLSVVIDAHDMLRSRTVGAG      55
            *::  **: ::** *      *:**.****.**: ..*.**::.::*:***** *.

WAC         SGRLVVGERGSVDAARLVRRVDG--GSLQEVV----ENAVRRLNPAAGVVAQVVWMDAGP      111
A47934      TPRMVVSGRGAVDAAHLIERVEADTGDLDDVAERCAREAAGRLDPAAGVMVRAVWVDAGP      115
            : *:**. **:****:*:.**:.  *.*::*.    .:*. **:*****:.:.**:****
                        "His-motif"
                        HHxxxDxxSW
WAC         DRVGRLVVVAHHLAVDEASWRVLVPDLRAACEGE-------PEPVGMSFRRWAGLLEEWA      164
A47934      GRVGRLVLVAHPAVVDARSWEILLRDLRAACEAVAAGREPSVDPVDVSFRRWAAALAEQA      175
            .******:***  .**  **.:*: *******.         :**.:******. * * *

WAC         VSAERAKELAAWKAIVGEDDQPIGQGLSRGKVRRRSWVVSGAEASVLAGRAPAVFFCDVQ      224
A47934      TSAERVAELTGWTALLDGEEPRLGEPGTGSPAGRRSWTLPKDRAGALAEQVTSAFHCGVH      235
            .****. **:.*.*::. ::  :*:  : . . ****.:   .*..** :. :.*.*.*:

WAC         DVLLAGLAGSVARWRGADVVLVDLEGHGRVP----DVSRTVGWFTSVHPVRLDM      274
A47934      EVLLTGLAGAVARWRGGHVVLVDVEDHGRHPVGGMDPSRTVGRFTSVHPVRLDV      289
            :***:****:******..*****:*.*** *    * ***** **********:
```

Figure S8. (A) Putative biosynthetic route encoded by WAC 01424 GPA BGC and proposed structure of the putative resulting GPA. (B) Amino acid sequence alignment of module 3 E-domain (crosshatched) from A47934 biosynthesis NRPS and WAC 01424 GPA biosynthesis NRPS. We cannot rule out that module 3 E-domain might be non-functional, like in the case of NPRS module 3 E-domain involved in A47934 biosynthesis from *Streptomyces toyocaensis* NRRL 15009[12]; however, the "His-motif" (HHxxxDxxSW, involved in the racemase activity[13]) in WAC 01424 GPA module 3 E-domain seems to be intact, differently from the A47934 NRPS module 3 E-domain.

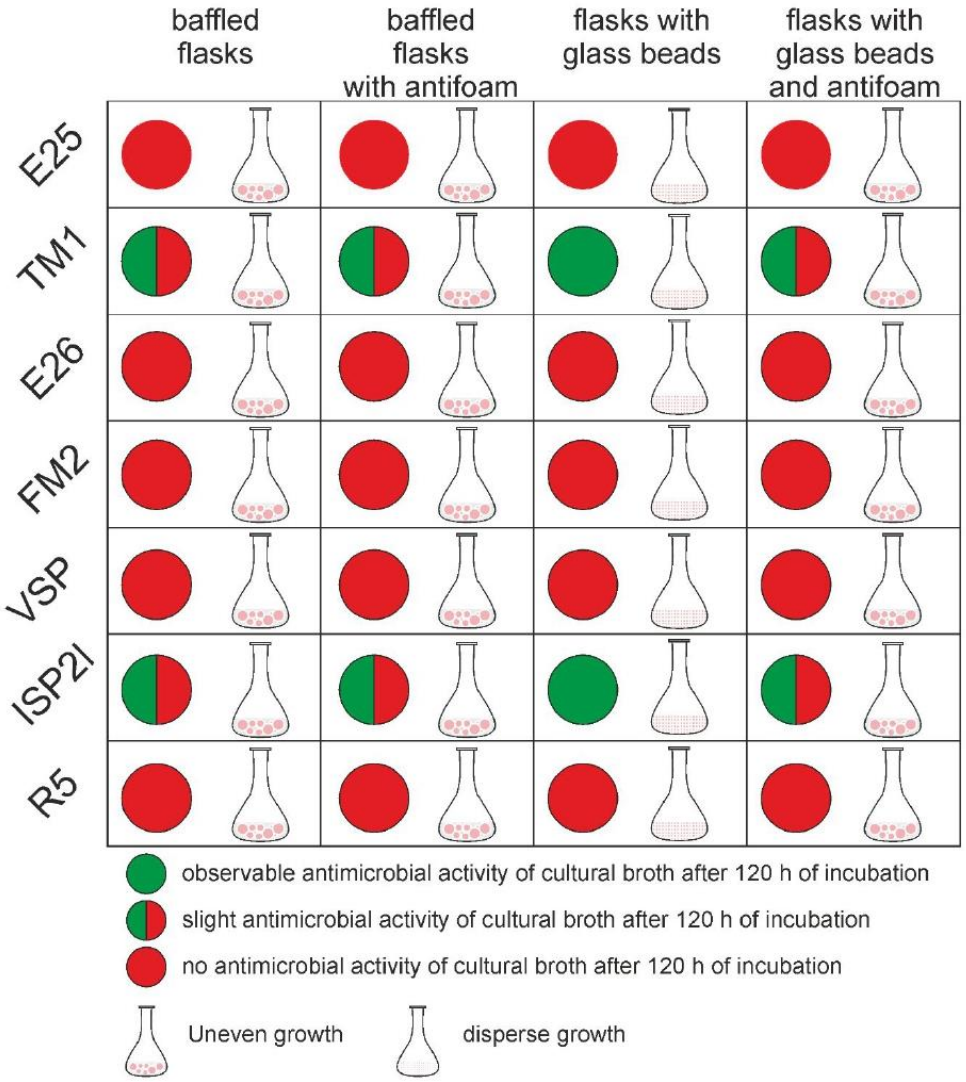


Figure S9. Scheme representing the growth pattern and the cell-free broth antimicrobial activity of *N. coxensis* cultivated in different media. Antimicrobial activity production was observed in two media (TM1 and ISP2l) out of the seven tested. In these media, antimicrobial activity seemed correlated with the mycelium dispersed growth that was observed only under cultivation in Erlenmeyer flasks with the addition of 5 mm glass beads. When baffled flasks were used and antifoam was added, culture tended to form irregular mycelial pellets and antimicrobial activity was negatively affected. Antimicrobial activity was tested against *B. subtilis* ATCC 6633 in Whatman paper disc antibiotic diffusion assays.

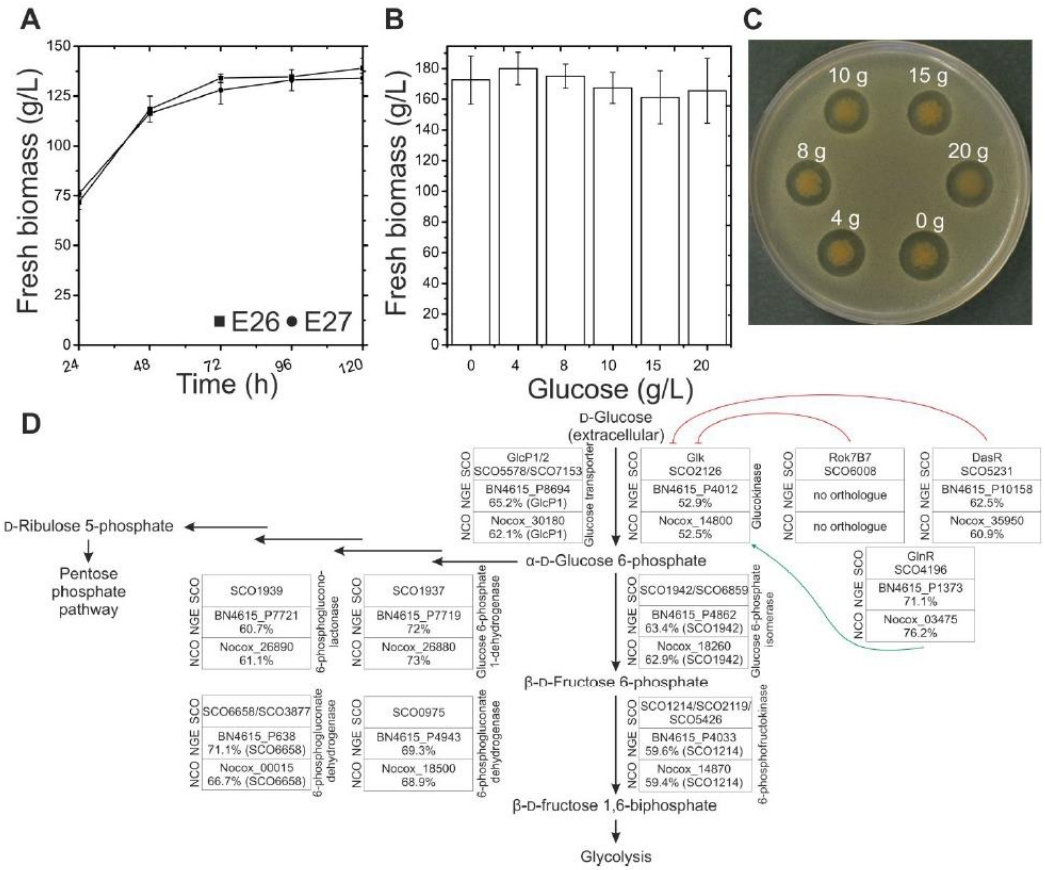


Figure S10. (A) Biomass accumulation of *N. coxensis* in E26 culture medium and its glucose-lacking variant E27. Growth curves were similar in the two vegetative media and biomass produced after 72 h of cultivation was used to inoculate production media. (B) Biomass accumulation in *N. coxensis* cultures grown for 120 h in ISP2l medium supplemented with increasing concentrations of glucose. The variant lacking glucose was named ISP2lm. (C) Antimicrobial activities of culture broths of *N. coxensis* grown in ISP2l media with different glucose concentrations (g L^{-1}) against *B. subtilis* ATCC 6633; in this bioassay 50 µL of production culture were placed on the surface of 0.7% Mueller-Hinton agar plate containing 10^6 *B. subtilis* cells. In both (B) and (C), the addition of different concentrations of glucose (from 0 to 20 g L^{-1}) to ISP2l did not exert any significant impact on either biomass accumulation or antimicrobial activity. Similar results were obtained also in TM1 and TM1m lacking-glucose (data not shown). (D) Comparative genomics of early steps of glucose metabolism in *N. coxensis* (NCO), compared to *N. gerenzanensis* (NGE) and *S. coelicolor* (SCO). All key enzymes of glucose catabolism (as well as some regulators) from *S. coelicolor* have orthologues in both *Nonomuraea* spp.; orthologues were discovered using Reciprocal Best Hit (RBH) BLAST approach, percent of amino acid sequence identity of *Nonomuraea* proteins to *S. coelicolor* proteins is shown under locus tag identifiers. Inability of *N. coxensis* to consume glucose might stem from the lack of expression of some of these genes and requires further experimental evaluation.

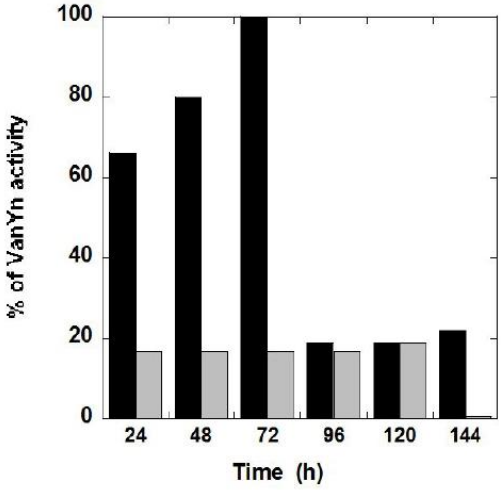


Figure S11. D,D-Carboxypeptidase activities in GPA-producing cultures of *N. gerenzanensis* in FM2 medium (black bars) and *N. coxensis* in ISP2lm medium (grey bars). Activity values were expressed as relative % of the maximum obtained at 72 h for *N. gerenzanensis* and were determined by measuring the amount of D-Ala released by hydrolysis of the *N*-acetyl-L-Lys-D-Ala-D-Ala tripeptide using a D-amino acid oxidase coupled to a peroxidase[14,15].

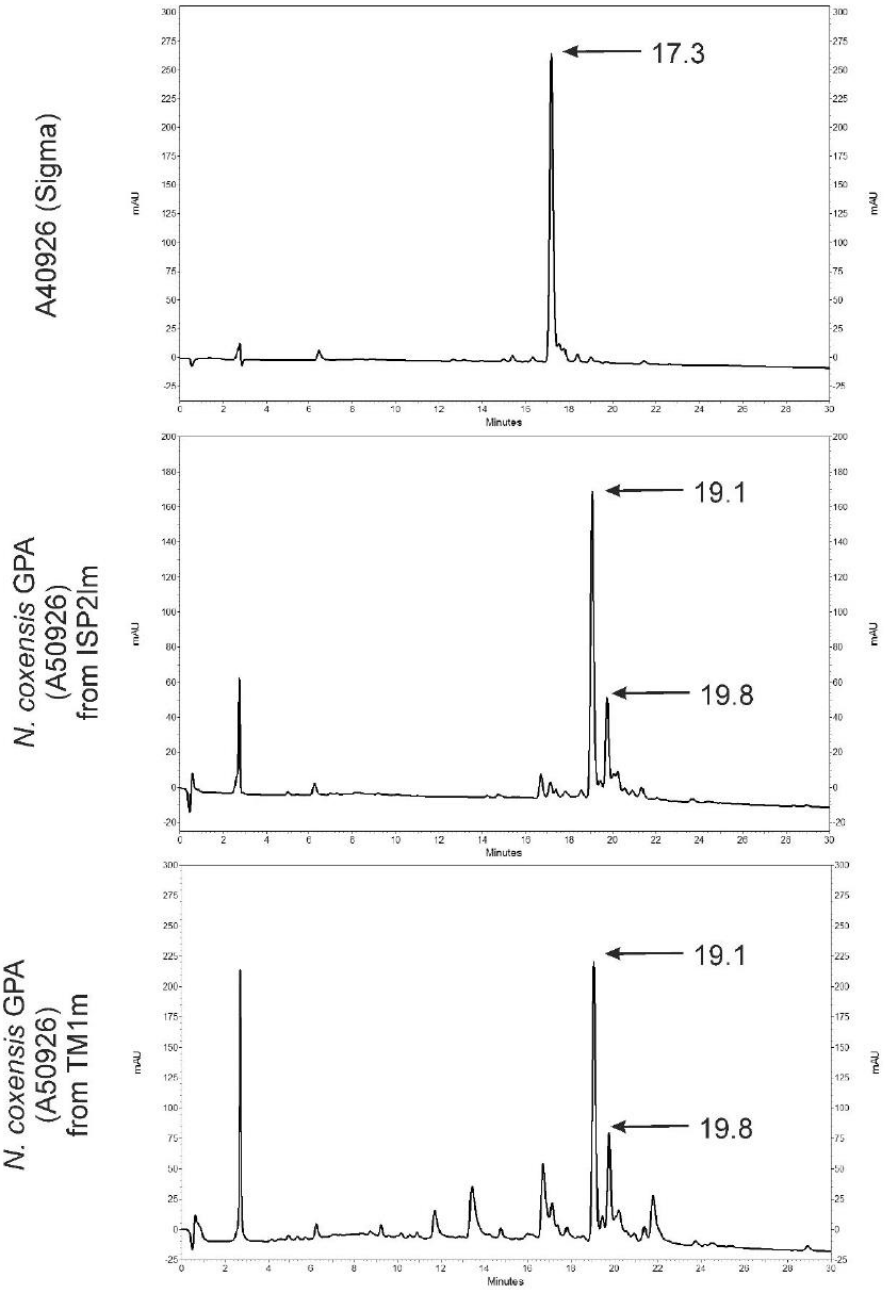


Figure S12. HPLC chromatograms with detection at 236 nm wavelength showing the partially purified GPA from *N. coxensis* (that we named A50926) cultivated for 168 h in either ISP2lm or TM1m. In both cases, the antibiotic complex contained two major peaks with retention times of 19.1 and 19.8 min, respectively. In the same HPLC conditions, the retention time of commercial A40926 standard (A40926 B) was 17.3 min.

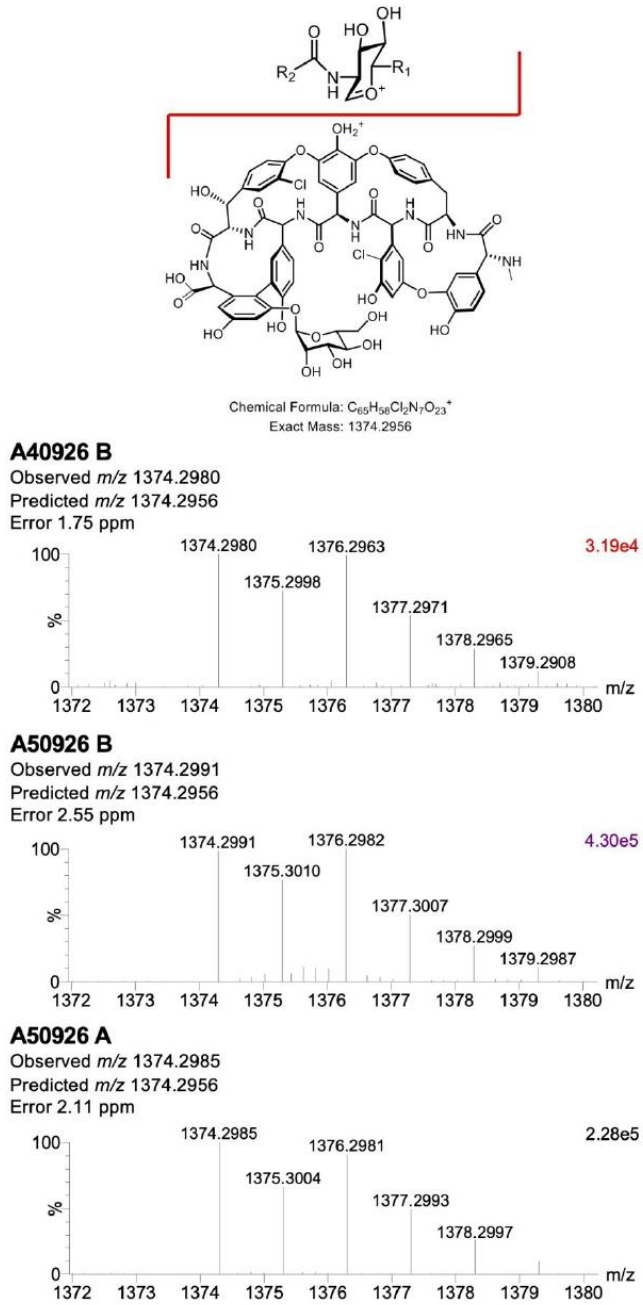


Figure S13. MS analysis of the in-source fragment corresponding to the mannosylated aglycone of the A40926 and A50926 GPAs. The schematic on top depicts the fragmentation leading to this peak, whereas the spectra below show masses and isotopic patterns for the fragment in all three molecules analyzed. Errors in ppm between the accurate observed masses of these fragments and the predicted mass are presented above each chromatogram. The intensity of the top peak in each spectrum is shown on the top right corner of each plot.

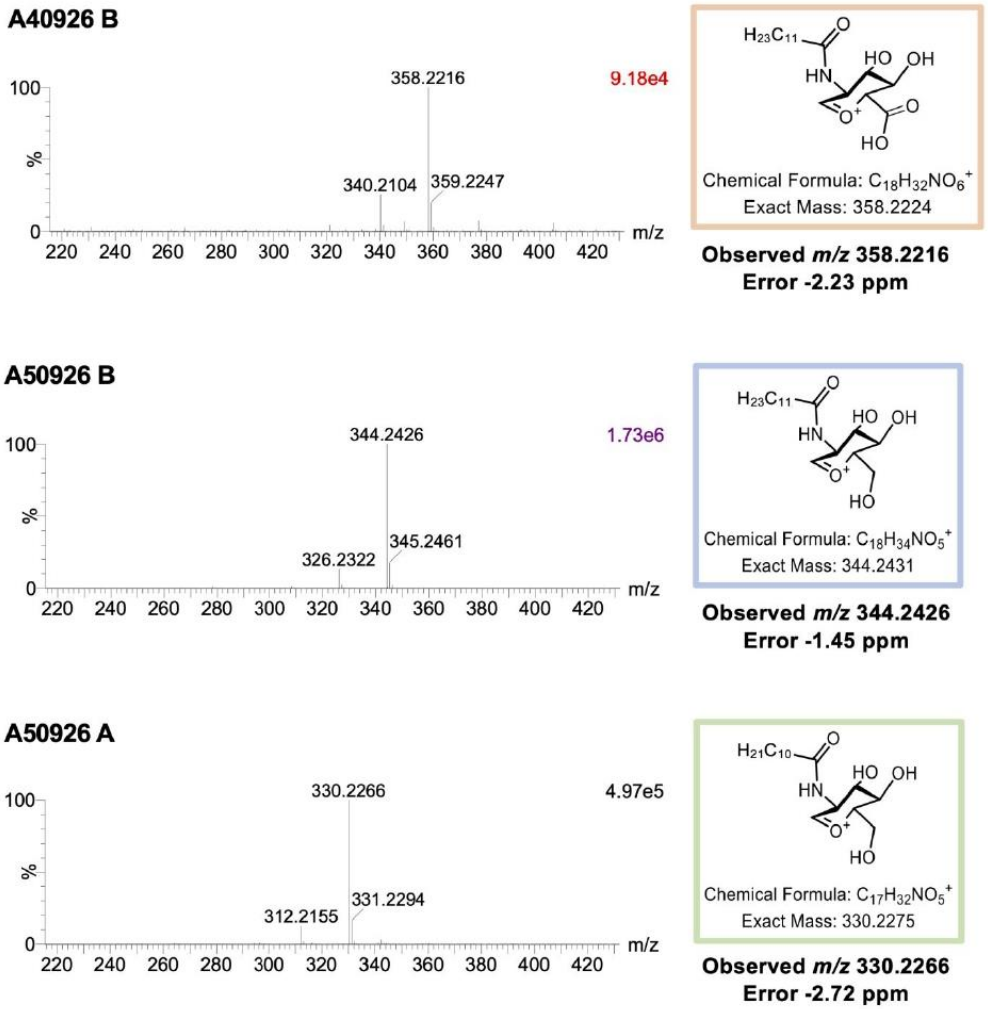


Figure S14. MS analysis of the in-source fragment corresponding to the GlcN-Acyl moieties of the A40926 and A50926 GPAs. The proposed structure for this moiety in each of the molecules analyzed is depicted next to its corresponding MS spectrum. Deviation in ppm between the accurate observed masses of these fragments and the predicted mass are presented below each proposed structure. The intensity of the top peak in each spectrum is shown on the top right corner of each plot.

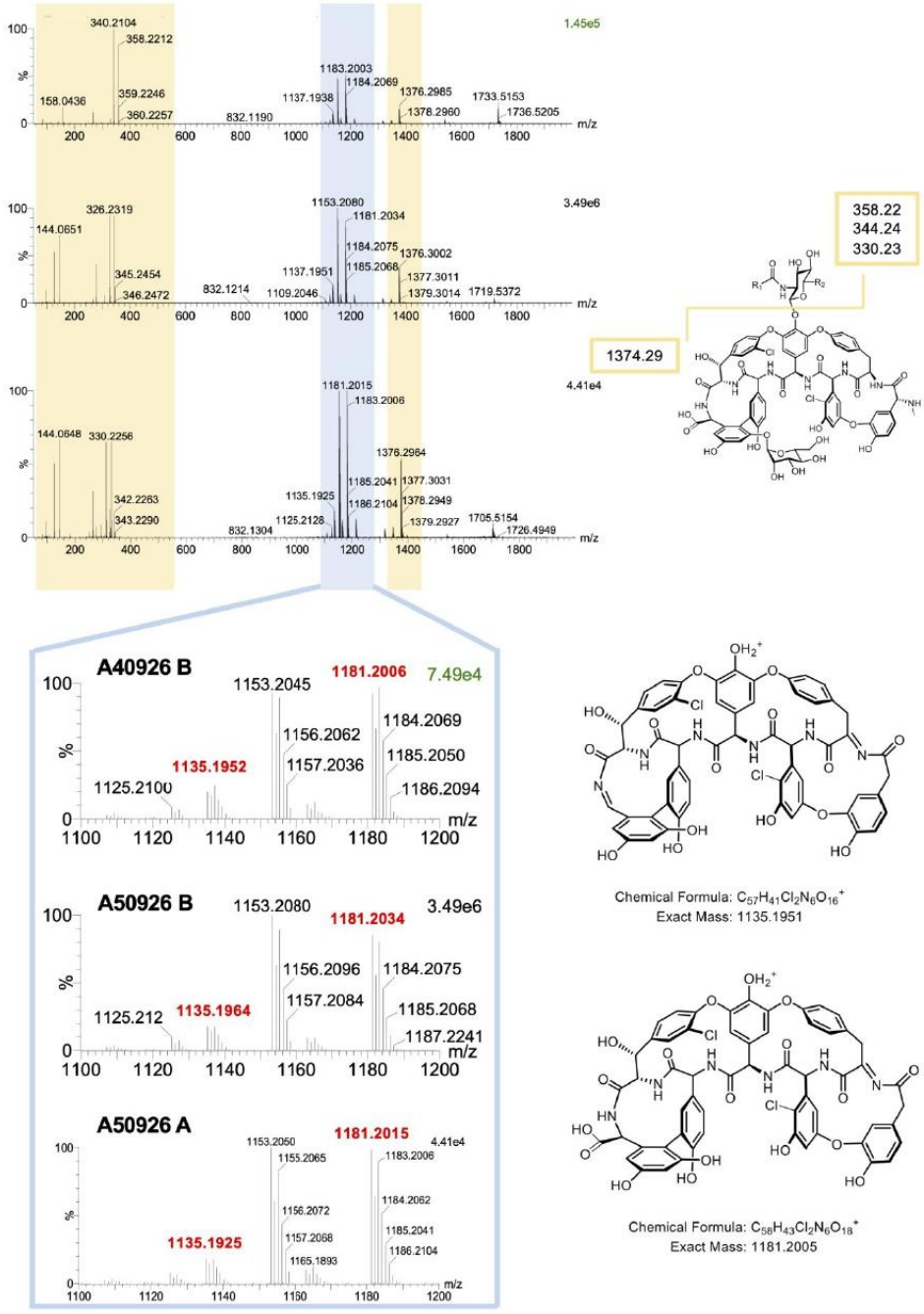


Figure S15. MS/MS analysis of A40926 B, A50926 B and A50926 A GPAs. Regions of the spectra corresponding to in-source fragments analyzed previously are highlighted in yellow, whereas a detailed view of aglycone fragmentation (highlighted in blue) is shown below. Monoisotopic masses of fragments for which a structure is proposed are highlighted in red in the spectra. Peak intensities are shown in the top right corner of each spectrum.

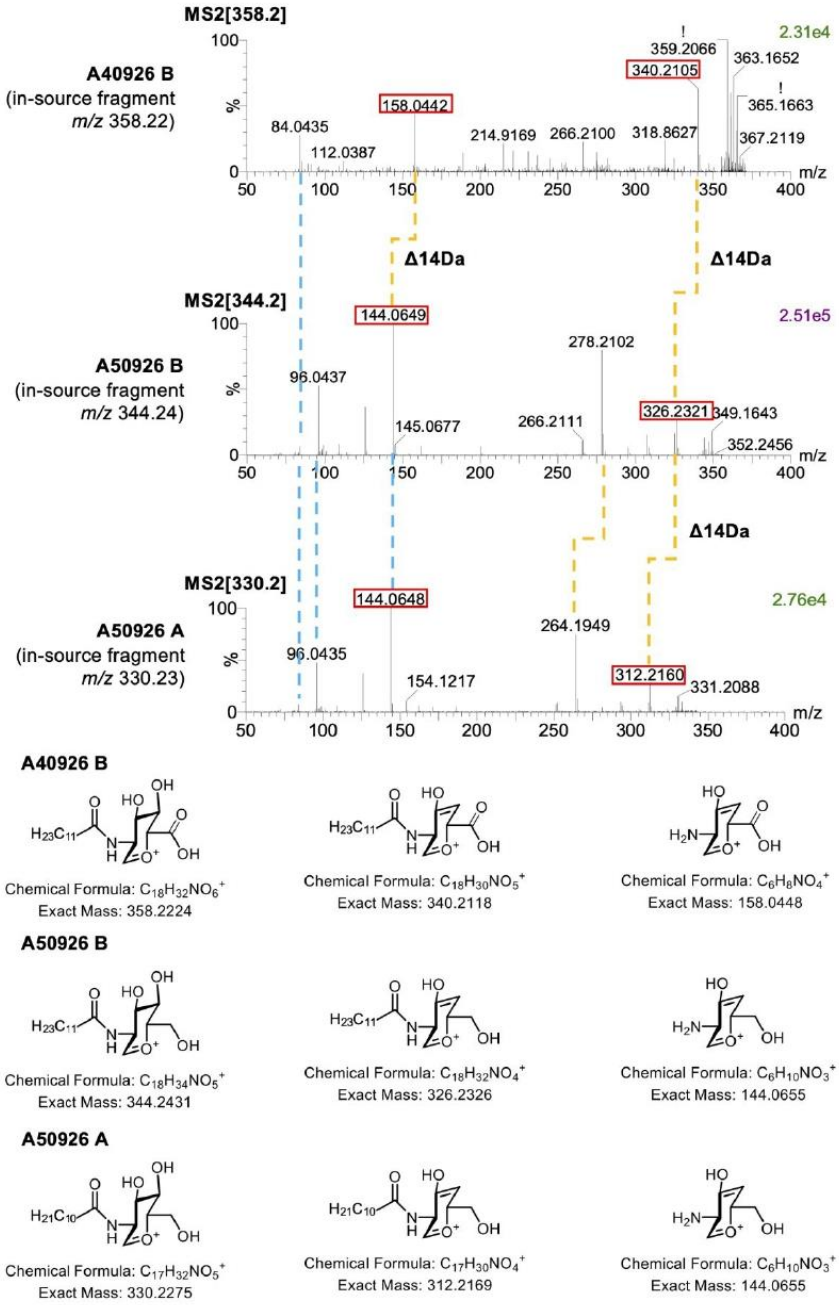


Figure S16. MS/MS analysis of the in-source signature fragment corresponding to the GlcN-Acyl moieties of the A40926 and A50926 GPAs. Identical fragments across several molecules are indicated with blue dashed lines, whereas fragments with distinctive mass losses are indicated with yellow dashed lines. Fragments for which proposed structures are presented in the bottom of the figure are circled in red in the MS/MS spectra. The intensity of the top peak in each spectrum is shown on the top right corner of each plot.

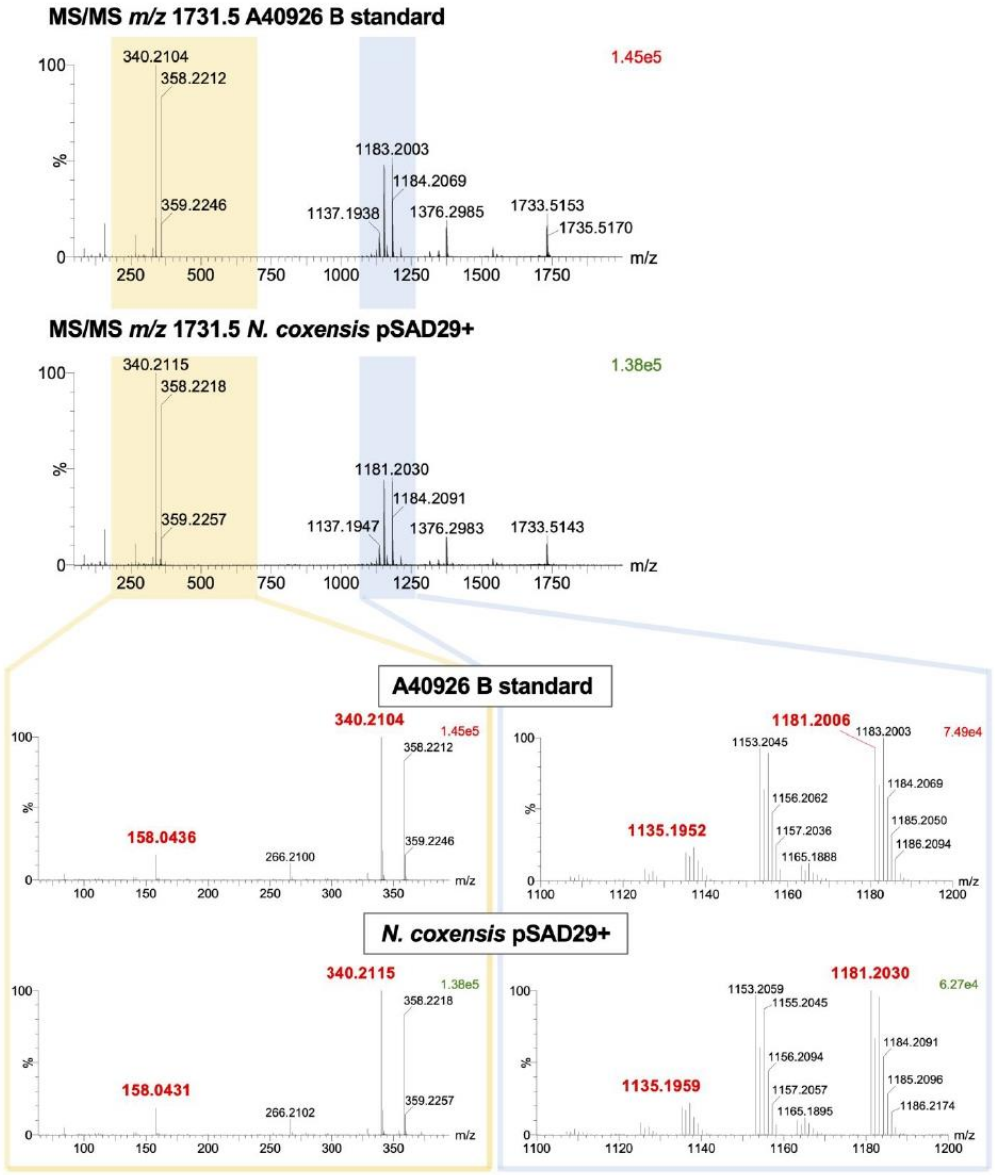


Figure S17. Comparison of the MS/MS spectra of a commercial standard of A40926 and A40926 produced by *N. coxensis* pSAD29^{+}. The areas corresponding to the signature GlcN-Acyl moiety and aglycone fragments of the spectra are zoomed out at the bottom of the figure for clarity. Monoisotopic masses of fragments previously identified are highlighted in red and top peak intensities for each spectrum are shown in their top right corner.

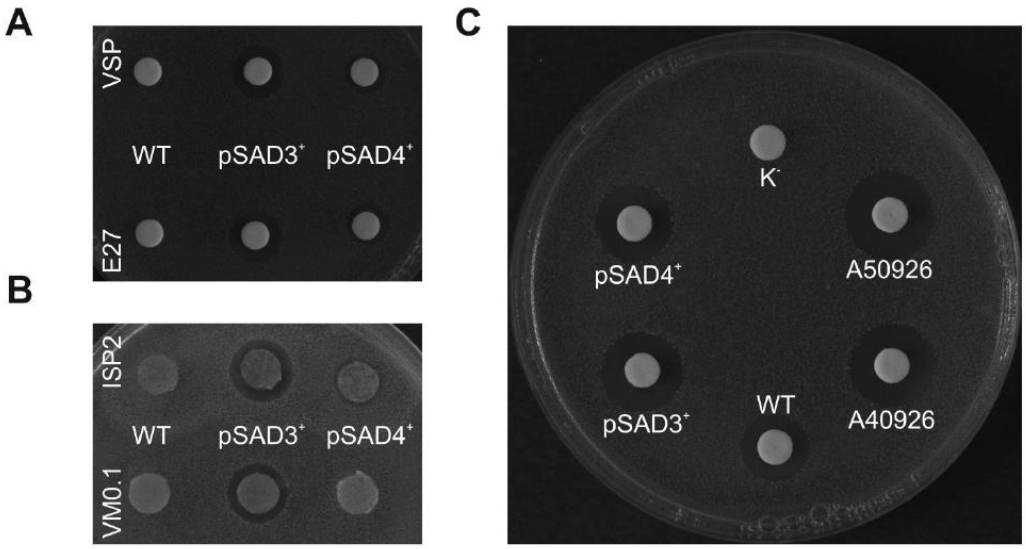


Figure S18. Antimicrobial activity assays towards *B. subtilis* ATCC 6633 showing: (A) activation of antimicrobial activity in recombinant strains overexpressing *dbv3* (pSAD3$^+$) and *dbv4* (pSAD4$^+$) cultivated in liquid VSP and E27 media, where the wild type does not produce any antimicrobial activity; (B) activation of the antimicrobial activity in solid media, where agar plugs cut from the lawns of *N. coxensis* pSAD3$^+$ grown on ISP2 and VM0.1 gave antimicrobial activity against *B. subtilis,* while the wild type and *N. coxensis* pSAD4$^+$ did not; (C) antimicrobial activity of borate buffer extracts from 168 h cultures of *N. coxensis* wild type and of recombinant strains pSAD3$^+$ and pSAD4$^+$ in ISP2lm compared with 1 µg of A40926 and 1 µg of partially purified A50926; K$^-$ was 50 µL of borate buffer as a control.

Supplementary references

(1) Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., Hopwood, D. A. Practical *Streptomyces* genetics. *The John Innes Foundation, Norwich, England*, 2000.

(2) Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., Kohlbacher, O. NRPSpredictor2 – a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.*, 2011, *39*. https://doi.org/10.1093/nar/gkr323.

(3) Horbal, L., Kobylyanskyy, A., Yushchuk, O., Zaburannyi, N., Luzhetskyy, A., Ostash, B., Marinelli, F., Fedorenko, V. Evaluation of heterologous promoters for genetic analysis of *Actinoplanes teichomyceticus* - producer of teicoplanin, drug of last defense. *J. Biotechnol.*, 2013, *168*, 367–372. https://doi.org/10.1016/j.jbiotec.2013.10.018.

(4) Horbal, L., Kobylyanskyy, A., Truman, A. W., Zaburranyi, N., Ostash, B., Luzhetskyy, A., Marinelli, F., Fedorenko, V. The pathway-specific regulatory genes, *tei15** and *tei16**, are the master switches of teicoplanin production in *Actinoplanes teichomyceticus*. *Appl. Microbiol. Biotechnol.*, 2014, *98*, 9295–9309. https://doi.org/10.1007/s00253-014-5969-z.

(5) Yushchuk, O., Andreo-Vidal, A., Marcone, G. L., Bibb, M., Marinelli, F., Binda, E. New molecular tools for regulation and improvement of A40926 glycopeptide antibiotic production in *Nonomuraea gerenzanensis* ATCC 39727. *Front. Microbiol.*, 2020, *11*. https://doi.org/10.3389/fmicb.2020.00008.

(6) Nazari, B., Forneris, C. C., Gibson, M. I., Moon, K., Schramma, K. R., Seyedsayamdost, M. R. *Nonomuraea* sp. ATCC 55076 harbours the largest actinomycete chromosome to date and the kistamicin biosynthetic gene cluster. *MedChemComm*, 2017, *8*, 780–788. https://doi.org/10.1039/c6md00637j.

(7) Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, 2018, *35*, 1547–1549. https://doi.org/10.1093/molbev/msy096.

(8) Darling, A. C. E., Mau, B., Blattner, F. R., Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.*, 2004, *14*, 1394–1403. https://doi.org/10.1101/gr.2289704.

(9) Cabanettes, F., Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, 2018, *6,* e4958. https://doi.org/10.7717/peerj.4958.

(10) Sievers, F., Higgins, D. G. Clustal Omega. *Curr. Protoc. Bioinformatics*, 2014, *48*, 3.13.1-3.13.16. https://doi.org/10.1002/0471250953.bi0313s48.

(11) Alduina, R., Lo Piccolo, L., D'Alia, D., Ferraro, C., Gunnarsson, N., Donadio, S., Puglia, A. M. Phosphate-controlled regulator for the biosynthesis of the dalbavancin precursor A40926. *J. Bacteriol.*, 2007, *189*, 8120–8129. https://doi.org/10.1128/JB.01247-07.

(12) Pootoolal, J., Thomas, M. G., Marshall, C. G., Neu, J. M., Hubbard, B. K., Walsh, C. T., Wright, G. D. Assembling the glycopeptide antibiotic scaffold: the biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL15009. *PNAS*, 2002, *99*, 8962–8967. https://doi.org/10.1073/pnas.102285099.

(13) Stachelhaus, T., Walsh, C. T. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry*, 2000, *39*, 5775–5787. https://doi.org/10.1021/bi9929002.

(14) Binda, E., Marcone, G. L., Pollegioni, L., Marinelli, F. Characterization of VanYn, a novel D,D-peptidase/D,D-carboxypeptidase involved in glycopeptide antibiotic resistance in

Nonomuraea sp. ATCC 39727. *FEBS Journal*, 2012, *279*, 3203–3213. https://doi.org/10.1111/j.1742-4658.2012.08706.x.

(15) Binda, E., Marcone, G. L., Berini, F., Pollegioni, L., Marinelli, F. *Streptomyces* spp. as efficient expression system for a D,D-peptidase/D,D-carboxypeptidase involved in glycopeptide antibiotic resistance. *BMC Biotechnol.*, 2013, *13*. https://doi.org/10.1186/1472-6750-13-24.

CHAPTER 3:

Genomic insights into the distribution and phylogeny of glycopeptide resistance determinants within the Actinobacteria phylum

antibiotics

MDPI

Article

Genomic insights into the distribution and phylogeny of glycopeptide resistance determinants within the Actinobacteria phylum

Andrés Andreo-Vidal [1], Elisa Binda [1], Victor Fedorenko [2], Flavia Marinelli [1*] and Oleksandr Yushchuk [1,2]

[1] Department of Biotechnology and Life Sciences, University of Insubria, Varese, Italy; aandreovidal@uninsubria.it (A.A-V.); elisa.binda@uninsubria.it (E.B.); flavia.marinelli@uninsubria.it (F.M.); oleksandr.yushchuk@uninsubria.it (O.Y.)

[2] Department of Genetics and Biotechnology, Ivan Franko National University of Lviv, Lviv, Ukraine; viktor.fedorenko@lnu.edu.ua (V.F.)

* Correspondence: flavia.marinelli@uninsubria.it; Tel.: +39-0332-42-1546

Abstract: The spread of antimicrobial resistance (AMR) creates a challenge for the global health security, rendering many previously successful classes of antibiotics useless. Glycopeptide antibiotics (GPAs) are so-called last resort drugs, which represent the ultimate line of defense against multidrug-resistant (MDR) Gram-positive pathogens. Unfortunately, we are assisting to the spread of resistance towards the first-generation GPAs vancomycin and teicoplanin, risking to limit the clinical use of this antibiotic class. It is widely recognized that GPA resistance determinants – *van*-genes – might been originated from GPA-producers, soil-dwelling Gram-positive actinobacteria, that use them for self-protection. In the current work, we present a comprehensive bioinformatic study on the distribution and phylogeny of GPA resistance determinants within the Actinobacteria phylum. Interestingly, *van*-like genes (*vlgs*) were found distributed in different arrangements not only among GPA producing actinobacteria, but in the non-producers as well: more than 10% of the screened actinobacterial genomes contained one or multiple *vlgs*, while only less than 1% encoded for a biosynthetic gene cluster (BGC). By phylogenetic reconstructions, our results highlight the co-evolution of the different *vlgs*, indicating that the most diffused are the ones coding for putative VanY carboxylases, which can be found alone in the genomes or associated with a *vanS/R* regulatory pair.

Keywords: glycopeptide antibiotics, multidrug-resistance, antibiotic resistance, *van*-genes.

Citation: Andreo-Vidal, A.; Binda, E.; Fedorenko, V.; Marinelli F.; Yushchuk O. Genomic insights into the distribution and phylogeny of glycopeptide resistance determinants within the Actinobacteria phylum. *Antibiotics* **2021**, *10*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

1. Introduction

Starting with the discovery of penicillin [1], humanity has been involved in a never-ending arms race between life-threatening bacterial pathogens and antibiotics, either natural, semisynthetic or completely synthetic. An estimation made by the United Nations Interagency Coordination Group on Antimicrobial Resistance in 2014 predicted that the spread of antimicrobial resistance (AMR) will cause up to dramatic 10 million deaths per year by the 2050 [2]. However, recent events might make this number even more grim: the worldwide health crisis caused by SARS-CoV-2 has led to an increase in antibiotic use and misuse, which, in turn, is likely to further accelerate AMR diffusion [3]. AMR has rendered many previously successful groups of antibiotics non-functional, leaving us hiding behind the "thin red line" of last resort drugs, capable to combat multidrug-resistant (MDR) pathogens. Glycopeptide antibiotics (GPAs) are a class of non-ribosomally synthesized, highly cross-linked, halogenated and glycosylated natural products, which are considered frontline drugs against Gram-positive MDR pathogens, such as *Staphylococcus aureus*, *Enterococcus* spp., *Clostridioides difficile*, etc. [4].

GPAs are produced by high G-C content soil-dwelling actinobacteria and could be divided into five types, according to their chemical structures and molecular targets [5]. Types I-IV group compounds having a cross-linked heptapeptide core highly glycosylated and/or lipidated. Their molecular target in pathogens is lipid II [6], a highly conserved macromolecule among bacteria which is essential for the cell-wall biosynthesis [7]. Types I-IV GPAs form five hydrogen bonds with D-alanyl-D-alanine (D-Ala-D-Ala) terminus of the lipid II pentapeptide stem (Figure 1a) [8–10]. Such binding terminates upstream transpeptidation and transglycosylation reactions, preventing the formation of mature

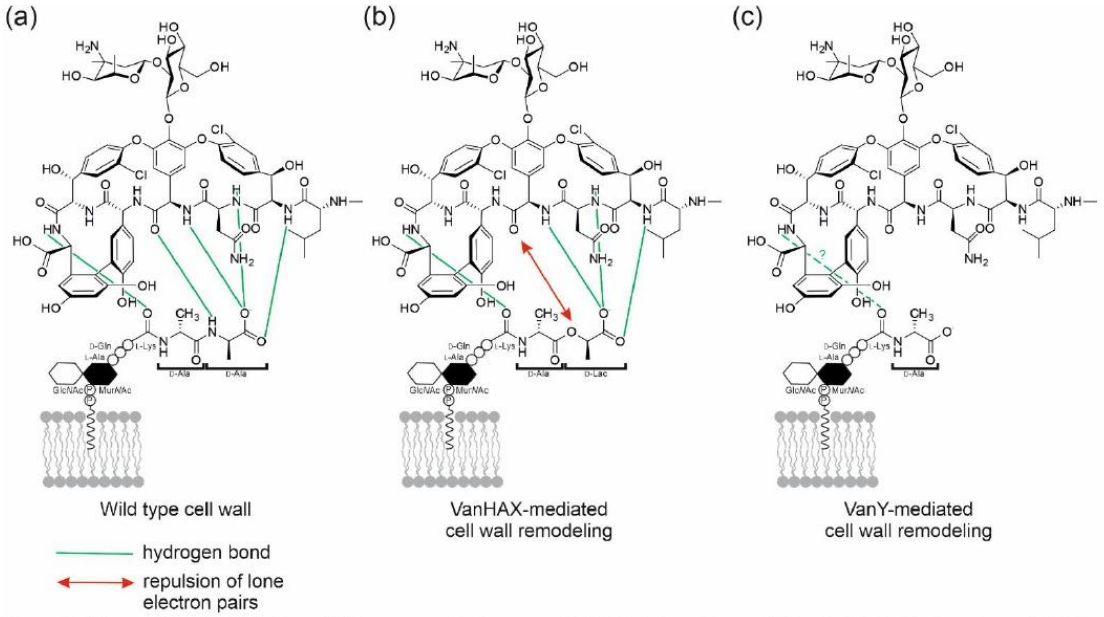


Figure 1. Schematic representation of (**a**) vancomycin (as a model GPA), interacting with the D-Ala-D-Ala terminus of lipid II pentapeptide stem of a Gram-positive cell wall, forming five hydrogen bonds (after [10]); (**b**) in the cell wall remodeled by the action of VanHAX, D-Ala-D-Lac termini of lipid II pentapeptide stem interacts with a significantly lower (1000 fold lower) affinity with vancomycin, due to the formation of only four hydrogen bonds and the repulsion of lone electron pairs between oxygen atoms [18]; (**c**) in the cell wall remodeled by the action of VanY D,D-carboxypeptidase, lipid II pentapeptide stem is truncated by the excision of the terminal D-Ala, and vancomycin affinity for this target appears significantly reduced, although at which extent has not been investigated yet.

peptidoglycan (PG). An old term describes type I-IV GPAs very precisely: dalbaheptides [12], meaning D-Ala-D-Ala-binding antibiotics with heptapeptide structures. Lipid II is obviously involved in the cell wall biosynthesis of dalbaheptide-producers as well. Thus, to avoid suicide during GPA production, producing strains need self-resistance mechanisms. This topic was recently revised in detail [13,14]. In brief, two main mechanisms of cell-wall remodeling exist in dalbaheptide-producers as well as in the resistant pathogens [15,16]. The first involves three genes – *vanHAX* – coding for: an α-ketoacid dehydrogenase VanH, that stereospecifically reduces pyruvate to D-lactate (D-Lac); a D-Ala-D-Lac ligase VanA; and a D,D-dipeptidase VanX. VanX removes the D-Ala-D-Ala termini of lipid II pentapeptide stems, while VanA prepares a pool of D-Ala-D-Lac dipeptides, which MurF, a UDP-*N*-acetylmuramoyl-tripeptide ligase, installs instead of the terminal D-Ala-D-Ala. As a result, GPAs form four, instead of five, hydrogen bonds with such D-Ala-D-Lac termini, and the repulsion of lone electron pairs between oxygen atoms contributes to make these GPA-lipid II complexes unstable (Figure 1b). The second resistance mechanism requires the expression of a D,D-carboxypeptidase VanY, which cleaves the terminal D-Ala residues of lipid II pentapeptide stems, hampering GPA complex formation with such

truncated lipid II derivatives (Figure 1c). So formed tetrapeptide-carrying lipid II molecules are still able to enter transpeptidation and transcglycosylation steps yielding a 3-3 cross-linked mature peptidoglycan [17]. Expression of either *vanHAX* or *vanY* (known overall as *van*-genes) is often inducible and regulated by a VanRS two-component regulatory pair. There, a sensor histidine kinase VanS (de)phosphorylates transcriptional regulator VanR in response to the extracellular presence of GPAs, and phosphorylated VanR activates the expression of *van*-genes involved in cell-wall remodeling. In dalbaheptide-producers, *van*-genes are always localized inside the borders of GPA biosynthetic gene clusters (BGCs), which group genes devoted to antibiotic biosynthesis, transport and regulation, guaranteeing co-regulation of self-resistance with antibiotic production. The only exception is the chloroeremomycin BGC in *Kibdelosporangium aridum* A82846, which apparently does not include *van*-genes [18].

Type V GPAs are not *glyco*peptides *sensu stricto*, since there are not glycosylated; moreover, they are not *dalbahept*ides, since they include also nonapeptide antibiotics (like corbomycin and GP6738 [19,20]), and they target autolysins (named also murein hydrolases) instead of lipid II. Autolysins are enzymes breaking the bonds within the peptidoglycan to allow bacterial growth and cell division [21]; thus, type V GPAs block cell-wall remodeling arresting cell division. Consistently, type V GPA BGCs lack cluster-situated *van*-genes needed to remodel lipid II termini [19,22]. *van*-genes are also missed in the related BGC of the uncrosslinked nonglycosylated peptide antibiotic known as feglymycin [23], which inhibits peptidoglycan synthesis targeting MurA (the UDP-*N*-acetylglucosamine l-carboxyvinyltransferase catalyzing phosphoenolpyruvate transfer to UDP-*N*-acetylglucosamine) and MurB (the UDP-*N*-acetylenolpyruvoylglucosamine reductase catalyzing the last step of the formation of UDP-*N*-acetylmuramic acid), both acting at the cytoplasmic step of cell-wall biosynthesis. However, type V GPA and feglymycin BGCs carry genes for a two-component regulatory pair, consisting of a sensor histidine kinase and a response regulator, which remind VanS and VanR of dalbaheptide BGCs, but whose function remains unclear.

Intriguingly, functional *van*-genes are also found in non-producing actinobacteria like in *Streptomyces coelicolor* [24] as well as in various other environmental non-infectious low G-C content bacteria (like *Paenibacillus popilliae* [25] or *Bacillus circulans* [26]). As a consequence, different hypotheses have been proposed to explain *van*-gene distribution and evolution in bacteria. The most accredited one is that pathogens such as enterococci (and the GPA non-producing environmental bacteria) might have acquired *van*-genes from dalbaheptide-producers through a series of horizontal gene transfer events, likely promoted by the selective pressure exerted by antibiotic environmental contamination [27]. An alternative hypothesis suggests that pathogens received *van*-genes from soil low G-C Gram-positives (Firmicutes phylum), as the mentioned above *Pnb. popilliae* and *Bac. circulans* [28], implying that in these bacteria *van*-genes evolved independently.

To get an insight into the phylogeny of *van*-genes, in this work we analyze their distribution and arrangement within the different orders belonging to Actinobacteria phylum, using the genomic data available in public databases. We have investigated more than 7000 actinobacterial genomes and found *van*-like genes (defined hereafter *vlgs*) in more than one tenth of them. Thus, we can confirm that *vlgs* presence is not limited to the dalbaheptide-producers, but these genes are also widely distributed among genomes of type V GPA producers and in non-producing actinobacterial taxa, which do not need them for self-resistance. In addition, phylogenetic reconstructions made for VanY-like proteins as well as for the VanHAX triads and for the two-component VanS/VanR regulatory system, highlight the evolutionary independent stories of the corresponding gene acquisitions. Thanks to this comparative genomic analysis, novel transposon-like mobile element carrying *vlgs* are here for the first time described, originated from poorly investigated orders as *Eggerthellales* and *Coriobacteriales*. Finally, as a control, the same bioinformatic analysis applied to more than 2000 *Bacillales* complete genome assemblies, yields

only a few *vlgs*, often adjacent to transposase-like open reading frames (ORFs). Taken altogether, these data reveal that phylum Actinobacteria is an incredibly vast source of variable GPA resistance determinants, which might potentially continue to move towards pathogens, contributing to the alarming diffusion of AMR. Their study might help surveillance of AMR spread in compliance with the One-Health approach [29].

2. Results

2.1. Organization of vlgs in GPA producers and beyond

Until now, *vlgs* in actinobacteria were reported as BGC-situated in more than 20 "classical" dalbaheptide-producers [13]. The case of the GPA non-producer actinobacterium *S. coelicolor* A3(2) – having a whole set of *vlgs* [30,31] – was rather interpreted as an exception. Thanks to the today abundance of genomic data on actinobacteria, our aim is to prove or disprove the assumption that *vlgs* are peculiar of GPA-producers, and to clarify how these genes are eventually distributed and organized among different orders belonging to Actinobacteria phylum. Thus, we screened all the genome assemblies available for actinobacteria in GenBank at the moment of this work preparation (April, 2020, Supplementary Excel File 1). This search covered 28 orders of the Actinobacteria phylum (Table 1), including two "candidate" ones (namely *candidatus Actinomarinales* and *Nanopelagicales*). We searched for *vanY*- and *vanHAX*- sequences, co-localized with *vanRS*-like two component regulatory pairs, and then we analyzed the genetic context of these genes.

Table 1. List of orders belonging to Actinobacteria phylum for whom genome assemblies were available in GenBank at the moment of this work preparation (April, 2020, Supplementary Excel File 1) and summary of the *vlgs* found in them.

Order	Number of genome assemblies analyzed	Number of genome assembling containing at least one *vlg*	Occurrence (%)	*vanY*-like	*vanR*-like	*vanS*-like	*vanH*	*vanA*	*vanX*
Acidimicrobiales	216	0	0	-	-	-	-	-	-
Actinomycetales	223	1	0,45	1	1	1	-	-	-
Actinopolysporales	10	0	0	-	-	-	-	-	-
Bifidobacteriales	1028	0	0	-	-	-	-	-	-
Candidatus *Actinomarinales*	214	0	0	-	-	-	-	-	-
Candidatus *Nanopelagicales*	26	1	3,85	1	-	-	-	-	-
Catenulisporales	3	1	34	1	1	1	1	1	1
Coriobacteriales	217	2	0,9	1	2	2	2	2	2
Corynebacteriales	707	110	15,6	122	23	25	14	14	14
Cryptosporangiales	3	2	67	-	1	1	1	1	1
Eggerthellales	106	1	0,9	-	1	1	1	1	1
Egibacterales	3	0	0	-	-	-	-	-	-
Frankiales	46	7	15	-	6	7	8	8	8
Gaiellales	3	0	0	-	-	-	-	-	-
Geodermatophilales	60	11	18	12	10	9	-	-	-
Glycomycetales	12	11	92	10	10	10	8	8	8
Jiangellales	11	10	91	12	11	10	7	7	7
Kineosporiales	12	1	8,4	1	1	1	-	-	-
Micrococcales	1741	15	0,86	2	13	13	15	15	14
Micromonosporales	200	83	42	14	86	88	5	57	56

Nakamurellales	6	6	100	6	1	1	-	-	-
Nitriliruptorales	6	1	17	1	-	-	-	-	-
Propionibacteriales	593	16	2,7	13	16	16	16	16	16
Pseudonocardiales	243	135	56	141	129	139	105	105	105
Rubrobacterales	8	1	12,5	2	1	1	-	-	-
Solirubrobacterales	45	4	8,9	6	6	6	2	2	2
Streptomycetales	1138	418	37	126	414	429	93	92	94
Streptosporangiales	228	63	28	52	67	63	7	7	6

At least one *vlg*-sequence was found in the majority of orders (22, Table 1); only the orders *Acidimicrobiales, Actinomarinales, Actinopolysporales, Bifidobacteriales, Egibacteriales* and *Gaiellales* lacked any *vlgs*. *vlgs* were most abundant (found in more than 10% of the genomic records per each order) in orders *Catenulisporales, Corynebacteriales, Cryptosporangiales, Frankiales, Geodermatophilales, Glycomycetales, Jiangellales, Micromonosporales, Nakamurellales, Nitriliruptorales, Pseudonocardiales, Rubrobacterales, Streptomycetales* and *Streptosporangiales* (Table 1, Supplementary Excel File 2). Since known GPA-producers belong to *Pseudonocardiales, Streptosporangiales, Micromonosporales*, *Streptomycetales* [5], we started to investigate *vlgs* organization in these orders in detail, to move then in the analysis of the still unexplored taxa.

2.1.1. Order *Pseudonocardiales*

Order *Pseudonocardiales* is the most abundant source of types I-IV GPAs [5,32]. So far, GPA BGCs were described only in *Amycolatopsis* and *Kibdelosporangium* genera. In our screening of 242 genomes from *Pseudonocardiales* spp., we found at least one *vlgs*-sequence in 150 assemblies (Table 1, Supplementary Excel Files 1 and 2). Only 30 assemblies contained GPA BGCs (Supplementary Excel File 2). Besides the known GPA producing genera *Amycolatopsis* and *Kibdelosporangium* [18], a GPA BGC was, for the first time, found in species belonging to the genus *Actinokineospora*. Overall, none correlation between quality and quantity of *vlgs* in GPA producers and non-producers was observed. The repertoire of *vlgs* was quite different in each strain. The following combinations were found: *vanYRS, vanHAXRS, vanHAX, vanYHAX, vanYHAXRS*. A significant portion of *vanY*-like genes was found to be "orphan" (meaning not co-localized with any other *vlgs*). Sequences similar to *vanJ* and *vanZ* genes, that were previously sporadically reported as involved in GPA resistance, but apparently without an essential role [14], were rare and always co-localized with *vanRS*-like pair (except cases in *Pseudonocardia* sp. CNS-139 – *vanHAXZ* arrangement and in *Pseudonocardia cypriaca* DSM 45511 – *vanHAXJ* arrangement).

To study the arrangement of *vlgs* in detail, we focused on the 30 genomes of the known GPA producers and, as a control, on 25 genomes of never previously investigated non-producing *Pseudonocardiales* spp. (Figure 2). Thus, in the majority of the GPA producing *Amycolatopsis*, *vanHAX* operon was found just upstream the *bbr*-orthologue (coding for a StrR-like cluster-situated-pathway-specific regulator of GPA biosynthesis [33]), in rare cases having *vanY*-like gene in between (Figure 2). At the same time, genomes of these *Amycolatopsis* GPA-producers contained two copies of *vanY*-like genes located outside the GPA BGCs (Figure 2). One of these copies was always co-localized with *vanRS*-like two-component regulatory genes. *Amycolatopsis balhimycina* DSM 5908 (balhimycin producer) and *Amycolatopsis* sp. H5 were notable exceptions: belonging to a different clade than other *Amycolatopsis* spp. GPA producers, they had a *vanSRY*-genes cluster-situated and a *vanHAX* operon outside the BGC (Figure 2). In *Amycolatopsis bartoniae* DSM 45807, only a *vanY*-like gene was found upstream the *bbr*-orthologue, while a *vanHAX* operon coupled with *vanRS*-like two-component regulatory genes was placed somewhere else on the chromosome (Figure 2). However, according to 16S rRNA gene phylogeny (see phylogenetic framework on Figure 2), *Am. bartoniae* appeared to outgroup all other *Amycolatopsis* spp. together with *Prauserella muralis* DSM 45305, *Tamaricihabitans halophyticus* DSM 45765 and

Amycolatopsis sp. KNN50.9b. Thus, it is likely that *Am. bartoniae* (as well as *Amycolatopsis* sp. KNN50.9b) might not belong to the *Amycolatopsis* genus at all. GPA producing *Kibdelosporangium* spp. had no cluster-situated *vlgs*, but *vanHAX* operons and multiple

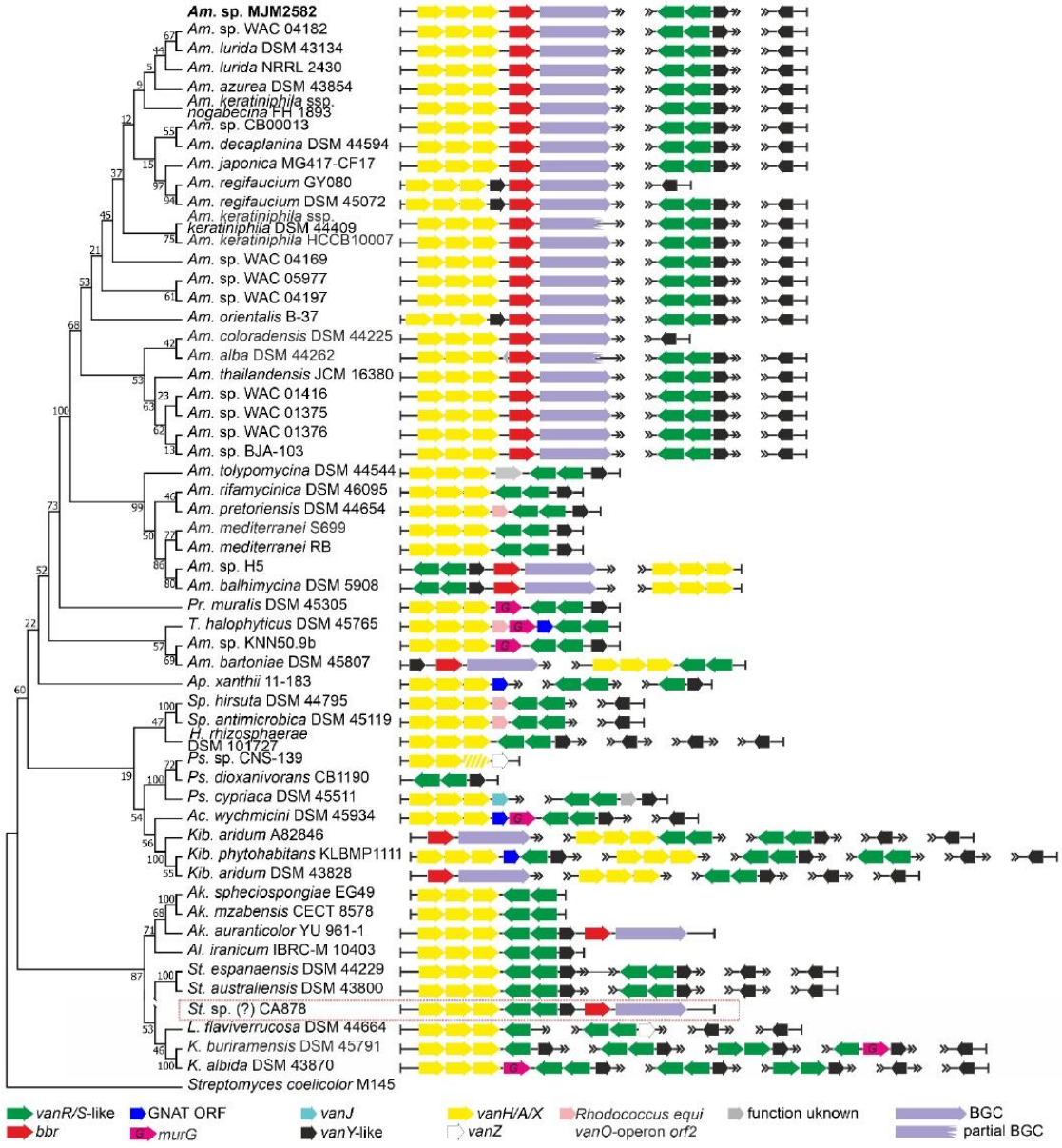


Figure 2. Organization of *vlgs* found within the set of 55 chosen genomes from *Pseudonocardiales* spp. Maximum Likelihood phylogenetic tree of 16S rRNA genes of corresponding species (see Methods section for details) served as a phylogenetic framework for the scheme. *Amycolatopsis* sp. MJM2582 (in bold at the top of the figure) is outside the framework due to the lack of full 16S rRNA gene in the corresponding genome assembly. *vlgs* from the metagenomic CA878 BGC (highlighted with red frame) were arbitrary introduced in this scheme since the retrieved sequences likely belong to *Saccharothrix* spp. Names of *Pseudonocardiales* genera were abbreviated according to ESM Table 1. Legend below the figure explains the color-coding of the scheme. Please refer to the text for the role of the single genes. Pseudogenes are shaded.

copies of *vanY*-like genes were found somewhere else on the chromosome (Figure 2). Finally, in *Actinokineospora auranticolor* YU 961-1 *vlgs* were only GPA cluster-situated – a set of *vanHAXRSY* genes was found upstream the *bbr*-orthologue (Figure 2).

Notably, in some cases (*e.g. Saccharopolyspora hirsuta* DSM 44795*), vanHAX* operon was followed by a homologue of *orf2* (a gene coding for a protein with unknown function) previously identified in *Rhodococcus equi* S7B *vanO*-operon [34]. As it does occur in *vanO*-operon, *orf2* homologue was in some species followed by a *murG*-like gene (*e.g. T. halophyticus* DSM 45765), coding for an essential peptidoglycan glycosyltransferase [35]. It is so far unknown how exactly MurG contributes to GPA-resistance, but it might be assisting *van*-mediated cell-wall remodeling. In few species as *Pr. muralis*_DSM 45305, *murG* was following *vanHAX* directly. Genes coding for a GCN5-related *N*-acetyltransferases (GNATs) were also found often co-localized with *Pseudonocardiales* spp. *vlgs*.

Finally, we introduced in our phylogenetic framework the only metagenomics-derived GPA BGCs – CA878 [36] – which was recently shown to be rather closely related to BGCs from *Amycolatopsis* spp. [18]. We speculated that this BGC might also belong to some unknown species of *Pseudonocardiales* – the organization of cluster-situated *vlgs* here seemed identical to the one from *Ak. auranticolor* YU 961-1. Fortunately, at the moment of its discovery, CA878 BGC was sequenced together with unannotated DNA flanks (*ca.* 26 and 3 kbp). This allowed to annotate 9 ORFs upstream and 6 ORFs downstream the borders of CA878 BGC (ESM Figure 1). It appeared, that the vast majority of these ORFs coded for proteins with orthologues in *Saccharothrix* spp. (ESM Figure 1). In our opinion, it is possible that CA878 comes from an unknown species belonging to *Saccharothrix* genus, further expanding the list of GPA-producing *Pseudonocardiales* genera.

2.1.2. Order *Streptosporangiales*

Order *Streptosporangiales* is the source of valuable lipidated dalbaheptides like the A40926 from *Nonomuraea gerenzanensis* ATCC 39727, which is the precursor of second-generation dalbavancin [4,37]. In addition, *N. gerenzanensis* was the first model for studying the role of VanY-like carboxypeptidases in self-resistance [38,39]. Other known GPA producers are *Nonomuraea coxensis* DSM 45129, recently reported to produce the lipopeptide A50496 [40] and *Nonomuraea* sp. ATCC 5507, which produces the type V kistamicin [41]. Here, we analyzed the genomic records from 228 *Streptosporangiales* spp. (Supplementary Excel File 1). *vlgs* were found in 63 out of them (Table 1, Supplementary Excel File 2). *vanY*-like genes were the most abundant, in almost all cases being co-localized with *vanRS*-like regulatory pairs (although few "orphan" ones were also found). *vanHAX*-operons were not found so often, coming exclusively from *Actinomadura* spp.; anyhow, *vanHAX*-operon was never found associated with *vanY*-like genes. Accessory *vlgs*, such as *vanJ*, were found rarely and tended to be co-localized with *vanRS*-like regulatory pairs. Only in one case – in *Actinomadura* sp. H3C3 – a *vanZ* pseudogene was discovered, co-localized with *vanRSY*.

Going into more detail, we analyzed the genomes from the three known GPA producers, those from five strains carrying putative GPA-like BGCs, including *Nonomuraea* sp. WAC 01424 [18], together with 17 genomes of other *Streptosporangiales* spp. lacking any GPA BGCs (Figure 3). We found that the two type IV GPA producers – *N. gerenzanensis* ATCC 39727 and *N. coxensis* DSM 45129 – carried one BGC-situated *vanY*-like gene and an additional *vanY*-like gene, co-localized with *vanRS*-like regulatory pair, distantly from the BGCs. The similarity of both distant- and cluster-encoded VanY-carboxypeptidases (amino acid sequence identity of 82.4% in *N. coxensis* and of 80.7% in *N. gerenzanensis*) was remarkable. The *vanRSY*-triad was found also in other *Nonomuraea* spp., lacking any GPA BGCs, as in *Nonomuraea fuscirosea* CGMCC 4.7104 (Figure 3). As well, it was present on the chromosome of the kistamicin producer *Nonomuraea* sp. ATCC 55076 away from its BGCs, although this type V GPA probably targets autolysins (like corbomycin and complestatin [19]), thus not requiring *van*-genes for self-resistance. Peculiarly, in the putative type IV GPA producer *Nonomuraea* sp. WAC 01424 [18], GPA BGC seemed to be localized

just downstream the *vanRSY*-triad (which indeed was not cluster-situated in other three GPA-producing *Nonomuraea* spp., Figure 3). WAC 01424 BGC carried another two-component regulatory pair, but without an additional copy of *vanY*.

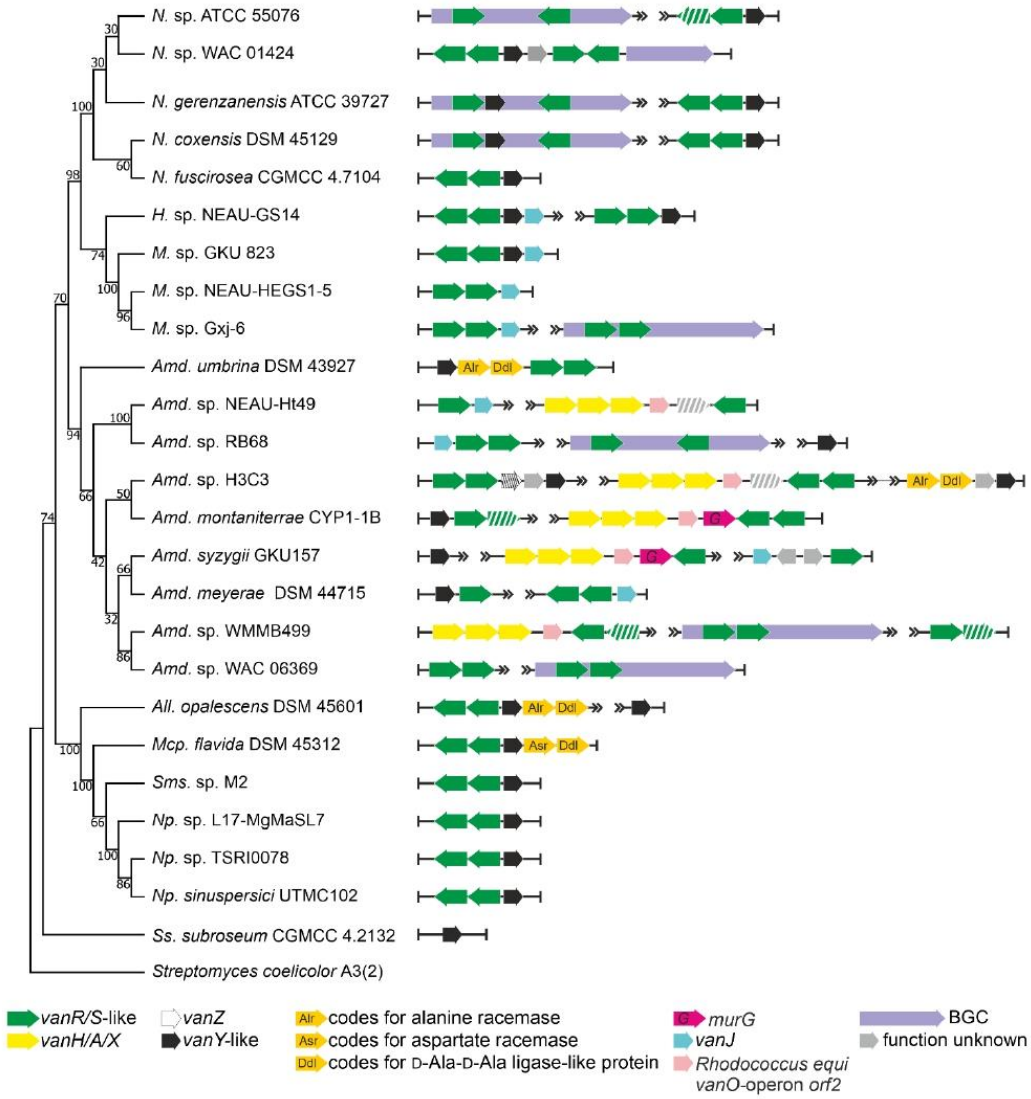


Figure 3. Organization of *vlgs* found within the set of 25 chosen genomes from *Streptosporangiales* spp. Maximum Likelihood phylogenetic tree of 16S rRNA genes of corresponding species (see Methods section for details) served as a phylogenetic framework for the scheme. Names of *Streptosporangiales* genera were abbreviated according to ESM Table 1. Legend below the figure explains the color-coding of the scheme. Please refer to the text for the role of the single genes. Pseudogenes are shaded.

Another two findings are worth of mention. Similar to what was observed for *Pseudonocardiales*, we found that all *vanHAX*-operons (present exclusively in *Actinomadura* spp. among *Streptosporangiales*) had a homologue of *vanO*-operon *orf2* (and sometimes a *murG*-like gene, too) downstream *vanX* (Figure 3). Finally, *vanRSY*-triad was in rare cases co-localized with genes coding for alanine/aspartate racemase- and D-Ala-D-Ala-ligase (Ddl)-

like proteins (*e.g. Allonocardiopsis opalecscens* DSM 45601 and *Murinocardiopsis flavida* DSM 45312, Figure 3), whose role would merit further investigation, indicating a possible alternative mechanism of cell wall remodeling such the one reported in few enterococci based on the incorporation of D-Ala-D-Ser termini in the resistant peptidoglycan precursors [14].

2.1.3. Order *Micromonosporales*

Order *Micromonosporales* is rich for GPA producers coming from genus *Actinoplanes* [42]. These are: *i)* the clinically relevant lipo-GPA teicoplanin, coming from *Actinoplanes teichomyceticus* ATCC 31121 [43,44], *ii)* the sulfated GPA UK-68,597 [45], coming from *Actinoplanes* sp. ATCC 53533, *iii)* the hyperglycosylated GPA actaplanin from *Actinoplanes missouriensis* ATCC 23342 [46]. We screened 200 genome assemblies (Supplementary Excel File 1) and we found *vlgs* in 83 of them (Table 1, Supplementary Excel File 2). Complete sets of *vanHAXRS* genes were found only in GPA producers *Apl. teichomyceticus*.

Figure 4. Organization of *vlgs* found within the set of 28 chosen genomes of *Micromonosporales* spp. Maximum Likelihood phylogenetic tree of 16S rRNA genes of corresponding species (see Methods section for details) served as a phylogenetic framework for the scheme. Names of *Micromonosporales* genera were abbreviated according to ESM Table 1. *vlgs* from the metagenomic CA915 and CA37 BGCs (highlighted with red frame) were arbitrarily introduced within the scheme since the retrieved sequences likely belong to some *Actinoplanes* spp. Legend below the figure explains the color-coding of the scheme. Please refer to the text for the role of the single genes. Pseudogenes are shaded.

and *Actinoplanes* sp. ATCC 53533 as well as in GPA non-producer *Actinoplanes derwentensis* DSM 43941 (Figure 4).

We chose a total of 26 genomes (including only two genomes of GPA producers, since the genome of *Apl. missouriensis* ATCC 23342 is not yet available) for more detailed examination. *vanY*-like genes were found in *Actinoplanes* spp., although they were not co-localized with *vanRS*-like regulatory pairs. Instead, *vanRS*-like regulatory pairs were often found co-localized with *vanZ* genes (like in *Actinoplanes missouriensis* 431, Figure 4) or found without any other close *vlg* (like in *Actinoplanes italicus* DSM 43146). However, one peculiarity specifically attracted our attention. In the course of our screenings, a particular gene arrangement was found to occur very often in the genomes of different *Micromonosporales* spp., especially in those belonging to the *Micromonospora* genus (Figure 4, Supplementary Excel File 2). This arrangement included a triad of genes coding for a PALP (pyridoxal-phosphate dependent)-like serine-threonine dehydratase, a Ddl-like protein and a VanX-like dipeptidase (further referred to as *pdx*-operon). Such triad was accompanied with a VanRS-like regulatory pair. Overall, such arrangement strikingly resembled typical *vanHAX-vanRS* operons, although the gene for lactate dehydrogenase was replaced by a serine-threonine dehydratase.

Finally, two metagenome-derived GPA-BGCs were described as related to the *Actinoplanes*-derived ones [18]. These were CA915 and CA37 [47]. Both of them were submitted to GenBank with rather long un-annotated flanking regions. As in the case of CA878 (see above), we annotated the genes present on these flanks. Majority of the BGCs flanking genes seemed to be homologous to *Actinoplanes* spp. genes (ESM Figure 2). Thus, CA915 and CA37 most likely belong to some unknown *Actinoplanes* spp.

2.1.4. Order *Streptomycetales*

Although multiple genomes of *Streptomyces* spp. were sequenced (definitively more than in the other orders belonging to Actinobacteria phylum), only few type I-IV GPA BGCs are known for this genus. These are A47934 BGCs from *Streptomyces toyocaensis* NRRL 15009 [48] and pekiskomycin BGCs from *Streptomyces* spp. WAC 04229 (WAC4229) and WAC1420 [49]. Our current analysis involved 1138 genome assemblies of *Streptomycetales* spp. (Supplementary Excel File 1), but none other novel BGC for types I-IV GPA was found (Supplementary Excel File 2). On a contrary, BGCs for type V GPAs and *feg*-like BGCs were found in 44 assemblies – all *Streptomyces* spp. except two *Kitasatospora* spp. Some of these BGCs were already reported [18], but we identified new ones (see Supplementary Excel File 2). *vlgs* were found exceptionally widespread in *Streptomycetales* spp.: more than one third of the analyzed genomes (418) contained *vlgs* (Table 1, Supplementary Excel File 2). Once again, we observed no correlation between the distribution of *vlgs* and of the GPA-like BGCs. More detailed analysis of 76 *Streptomycetales* spp. genomes (including 46 genomes carrying GPA- and *feg*-like BGCs) showed several different combinations, where certain strains carried a putative type V GPA BGC and no *vlgs* (e.g. *Streptomyces fradiae* NKZ-259, Figure 5) along with strains carrying type V GPA BGCs and a full set of *vlgs* (e.g. *Streptomyces* sp. NRRL WC-3897, Figure 5). Additionally, different combinations of *vlgs* were found in strains carrying no GPA-like BGCs. One peculiar trait of *Streptomycetales* spp. carrying the canonical *vanHAX-vanRS* operons is the presence of *vanK*, coding for an enzyme, belonging to the Fem family, which adds the branch amino acid(s) to the stem pentapeptide of peptidoglycan precursors carrying the D-Ala-D-Lac termini [50].

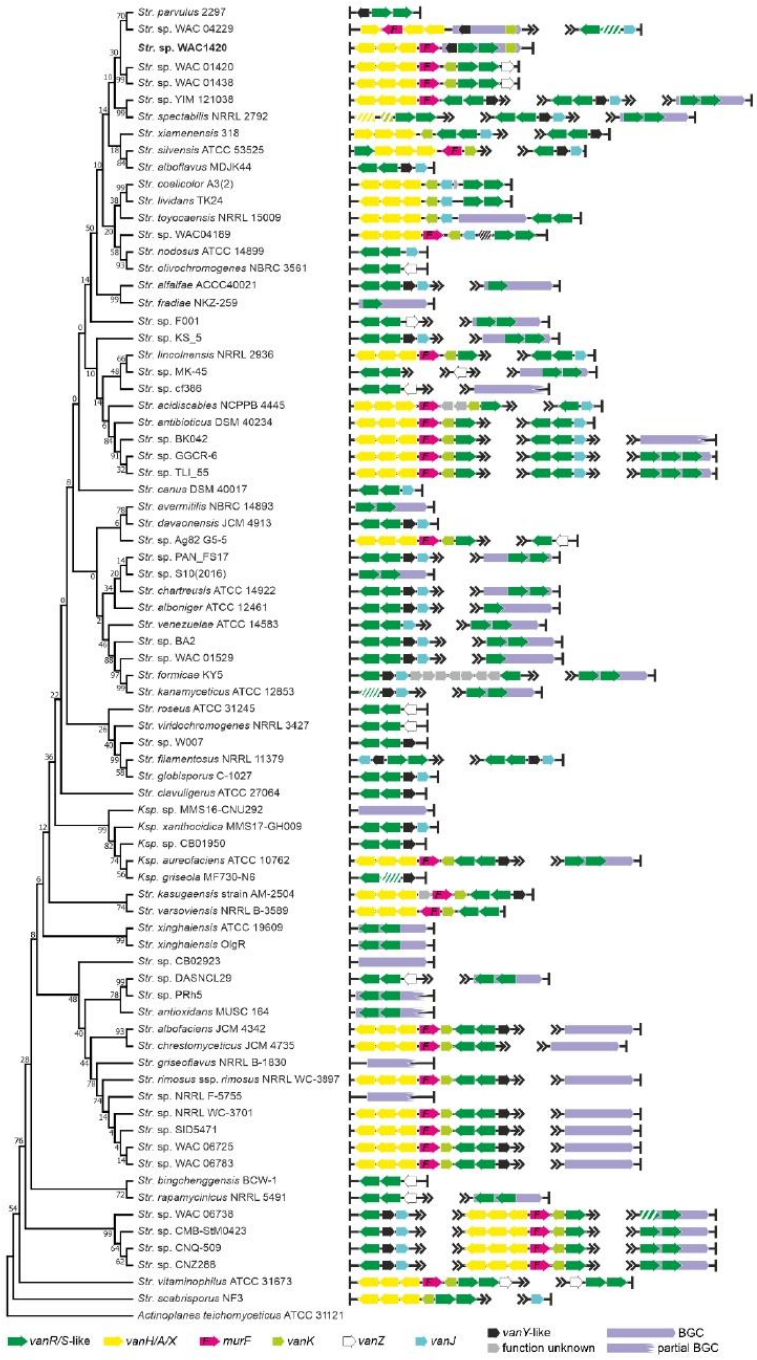


Figure 5. Organization of *vlgs* found within the set of 74 chosen genomes of *Streptomycetales* spp. Maximum Likelihood phylogenetic tree of 16S rRNA genes of corresponding species (see Methods section for details) served as a phylogenetic framework for the scheme. Names of *Streptomycetales* genera were abbreviated according to ESM Table 1. We were unable to detect the BGC for pekiskomycin within the published genome assembly of *Streptomyces* sp. WAC 01420, although the assembly contained *vlgs*; thus, *vlgs* in pekiksomycin BGC (as published in [49]) are given outside the phylogenetic framework (highlighted in bold). Please refer to the text for the role of the single genes. Legend below the figure explains the color-coding of the scheme. Pseudogenes are shaded.

2.1.5. Occurrence of *vlg*s in GPA non-producing groups

Order Actinomycetales. Although being known for various opportunistic animal and human pathogens, *Actinomycetales* spp. did not carry multiple *vlgs*. The only taxon (out of the 200 genome assembles screened, Supplementary Excel Files 1 and 2) carrying *vanYRS*-genes was *Actinomycetales* bacterium JB111 (Table 1, Figure 6a).

Order Catenulisporales. Only three genomes of *Catenulisporales* spp. were available, an in one of them – *Catenulispora acidiphila* DSM 44928 – a complete set of *vanHAXRSY* was found (Table 1, Figure 6b, Supplementary Excel Files 1 and 2).

Order Coriobacteriales. Only two genomes, out of the 217 available for *Coriobacteriales* spp., contained a complete set of *vanHAXRSY* (Table 1, Supplementary Excel Files 1 and 2 see next paragraph for more detailed description).

Order Corynebacteriales. *vlgs* appeared to be quite common in *Corynebacteriales* spp. (707 genomes available for screening) (Table 1, Supplementary Excel Files 1 and 2). Remarkably, a full set of *vlgs* (*vanHAXRS*) was discovered within the genome of *Williamsia marianensis* DSM 44944 – a species isolated from Mariana trench (10.898 m below the sea level) in 1998 [51]. Other *vlgs* combinations included: *vanY*-like genes paired with *vanRS*-like regulatory pair; "orphan" *vanY*-like genes; *vanHAXRSY*-genes, often accompanied with *murG* genes and homologues of *vanO*-operon *orf2* (see typical examples on Figure 6c). Peculiarly, during this analysis, a putative unknown GPA BGC was found in the genome of *Nocardia terpenica* NC_YFY_NT001, which is a clinical isolate derived from a human cerebrospinal fluid (see CP023778 genome assembly information).

Order Cryptosporangiales. *vlgs* were found in two out of the three available genome assemblies of *Cryptosporangiales* spp. and were arranged as *vanHAXRS*/*HAXRSZ* (Table 1, Supplementary Excel Files 1 and 2); indeed, multiple copies of "orphan" *vanZ*-like genes were also present in the genome of *Cryptosporangium.* sp. A-T5661 (Figure 6d). In the genome of the latter, a gene coding for a GNAT was co-localized with *vanHAXRS*-genes (Figure 6d).

Order Eggerthellales. Only in one genome out of the 106 available for *Eggerthellales*, a *vlg* was found (Table 1, Supplementary Excel Files 1 and 2, see next paragraph).

Order Frankiales. Approximately 15% of the analyzed *Frankiales* spp. genomes (46 on total, Supplementary Excel Files 1 and 2) possessed *vlgs*, arranged most often as *vanHAXRS* (and sometimes co-localized with *murF*- and *murG*-like genes) (Table 1, Figure 6e).

Order Geodermathophilales. A large portion of the analyzed *Geodermathophilales* spp. genomes (60, Supplementary Excel Files 1 and 2) carried *vlgs*, namely *vanY*-like genes co-localized with *vanRS*-like regulatory pairs (Table 1). Rarely, *vanYRS*-like genes were found together with genes coding for a Ddl and for an alanine-racemase (as it was observed in *Alr. opalescens* DSM 45601, Figure 6f). Other unique feature (discovered only in *Geodermathophilales* spp.) was the presence of genes coding for putative VanY-VanZ fusion proteins (*e.g.* in *Modestobacter* sp. I12A-02628, Figure 6f).

Order Glycomycetales. *vlgs* were ubiquitously found in *Glycomycetales* spp. genomes (12 available in total), organized as *vanYRS*, *vanHAXY* or *vanHAXRSY* (Table 1, Supplementary Excel Files 1 and 2, Figure 6g). Latter arrangements were coupled with the genes coding for MurF and GNAT. Multiple copies of "orphan" *vanZ*-like genes were also found (Figure 6g).

Order Jiangellales. *vlgs* were found in more than 90% of the eleven analyzed *Jiangellales* spp. genomes (Table 1, Supplementary Excel Files 1 and 2). The most frequent arrangement was *vanHAXRSY,* although in *Jiangella anatolica* GTF31 *vanHAXS*-genes were co-localized with *vanK* and with genes coding for MurF-, MurG- and GNAT-like proteins (Figure 6h).

Order Kineosporiales. A single set of *vlgs* in this order – *vanRSY* – was found in *Pseudokineococcus lusitanus* CECT 7306 (Figure 6i) among the twelve genomes analyzed (Table 1, Supplementary Excel Files 1 and 2).

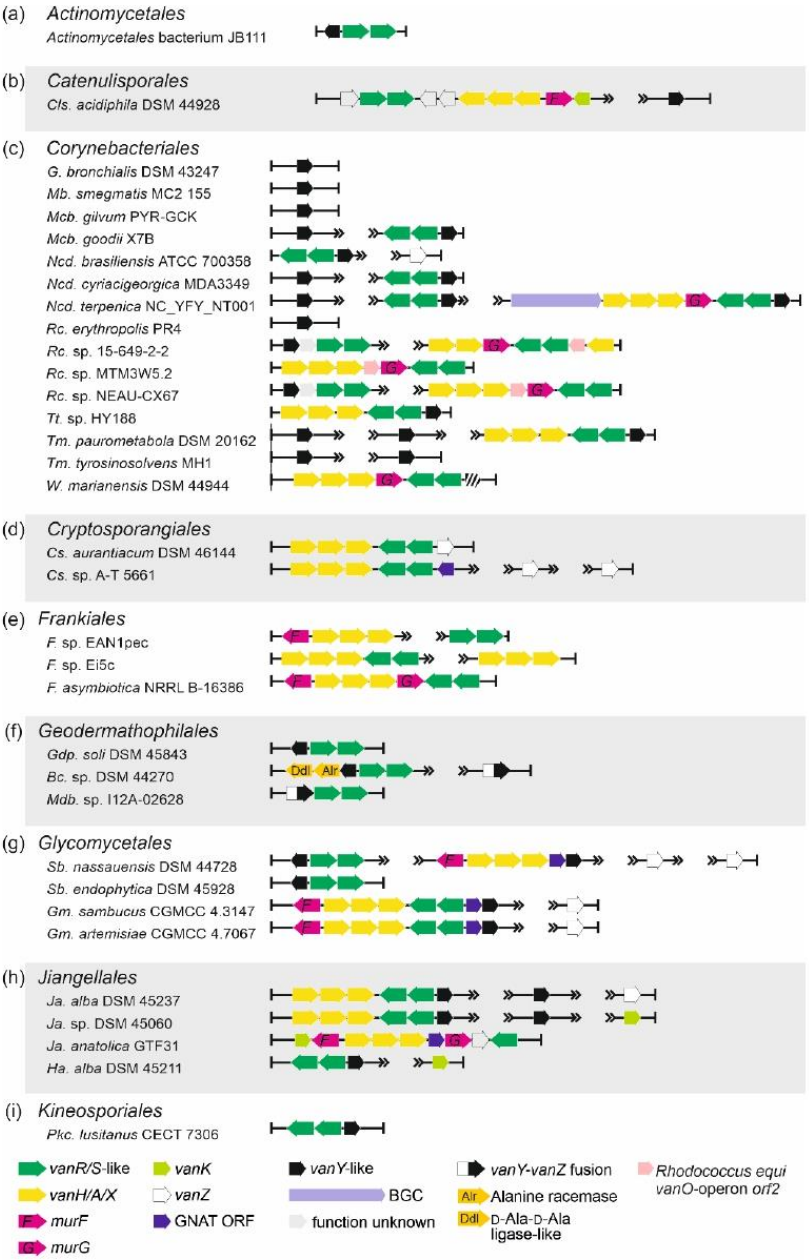


Figure 6. Arrangements of *vlgs* discovered in representative genomes from orders (**a**) *Actinomycetales*, (**b**) *Catenulisporales*, (**c**) *Corynebacteriales*, (**d**) *Cryptosporangiales*, (**e**) *Frankiales*, (**f**) *Geodermathophilales*, (**g**) *Glycomycetales*, (**h**) *Jiangellales* and (**g**) *Kineosporiales*. Please refer to the main text for more details; genus names were abbreviated according to the ESM Table 1. Legend below the figure explains the color-coding of the scheme. Pseudogenes are shaded.

Order Micrococcales. *vlg*s were found in less than 1% of analyzed *Micrococcales* spp. genomes (a total of 1741, Supplementary Excel Files 1 and 2) and were represented mainly as *vanHAX, vanHAXRS* or *vanHAXRSY* arrangements (Table 1, Figure 7a). *vanZ*-like genes, coupled with *vanRS*-like regulatory pairs were also observed as well as "orphan" *vanY*-like genes co-localized with *vanZ* genes (Figure 7a).

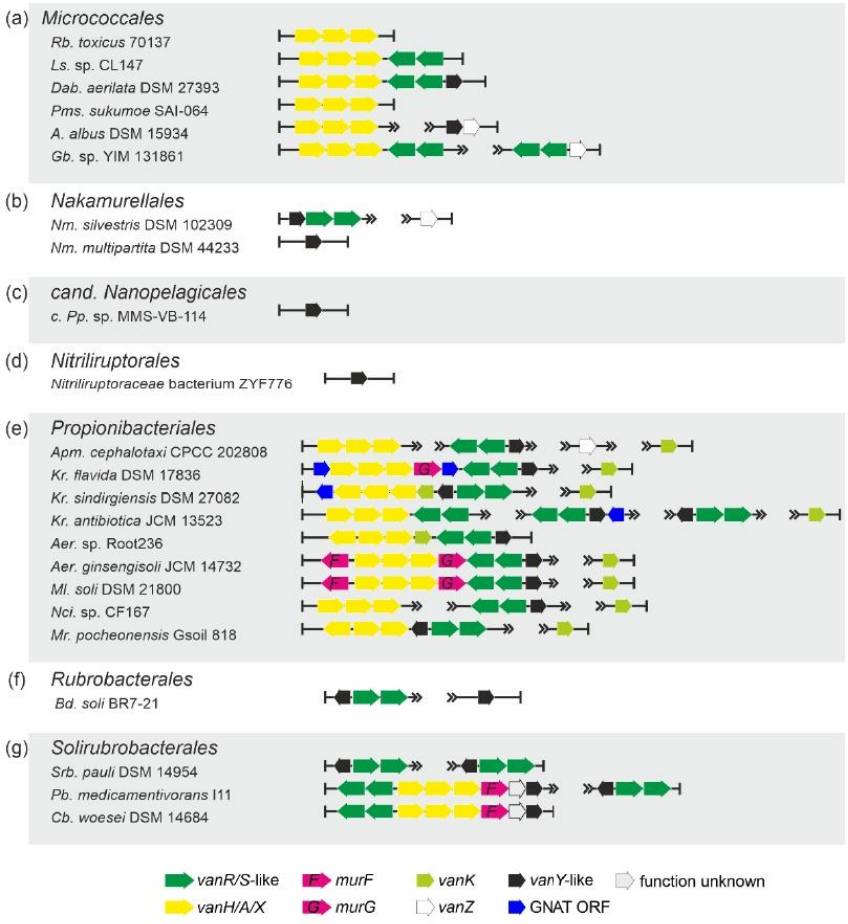


Figure 7. Arrangements of *vlgs* discovered in representative genomes from orders (**a**) *Micrococcales,* (**b**) *Nakamurellales*, (**c**) *Nanopelagicales*, (**d**) *Nitriliruptorales*, (**e**) *Propionibacteriales*, (**f**) *Rubrobacterales* and (**g**) *Solirubrobacterales*. Please refer to the main text for more details; genus names were abbreviated according to the ESM Table 1. Legend below the figure explains the color-coding of the scheme.

Order Nakamurellales. *vlgs* were found within all the 6 genome assemblies of *Nakamurella* spp., either as *vanYRS* or as "orphan" *vanY*-like and *vanZ* genes (Table 1, Figure 7b, Supplementary Excel Files 1 and 2).

Order cand. Nanopelagicales and order Nitriliruptorales. In the few genome assemblies belonging to the species of both orders, only "orphan" *vanY*-like genes were rarely found (Table 1, Supplementary Excel Files 1 and 2, Figure 7c and d).

Order Propionibacteriales. *vlgs* were found in the genome assemblies belonging to few genera of *Propionibacteriales* (Table 1, Supplementary Excel File 2), although it was

possible to analyze 593 genomes (Supplementary Excel File 1). There, *vlgs* exhibited different arrangements, summarized on Figure 7e. Genes coding for MurF, MurG and GNAT proteins were often co-localized with *vlgs*.

Orders Rubrobacterales and Solirubrobacterales. Only a small portion of species, belonging to both orders, carried *vlgs* within their genomes (Table 1, Supplementary Excel Files 1 and 2). *vlgs* mainly were arranged as either *vanYRS* or *vanRSHAXY* (sometimes co-localized with a *murF* gene, Figure 7f and g).

2.1.6. Putatively novel transposable elements carrying *vlgs* in *Eggerthellales* and *Coriobacteriales* spp.

Analyzing the genomes of actinobacteria belonging to orders *Eggerthellales* and *Coriobacteriales,* we found *vlgs* in *Enterorhabdus mucosicola* NM66_B29, *Parvibacter caecicola* DSM 22242 and *Atopobium minutum* 10063974. When we examined the genetic neighborhood of these genes, it emerged that they might belong to the family of transposon-like mobile genetic elements (MGEs), involving multiple genes deputed to DNA transfer (Figure 8). Moreover, it resulted that these *vlgs*-carrying putative MGEs were almost identical in *Er. mucosicola* NM66_B29 (order *Eggerthellales*) and *Pb. caecicola* DSM 22242 (order *Coriobacteriales*), while the MGE from *Atp. minutum* 10063974 (order *Coriobacteriales*) significantly differed from both (Figure 8). Transposons and other MGEs are believed to be one

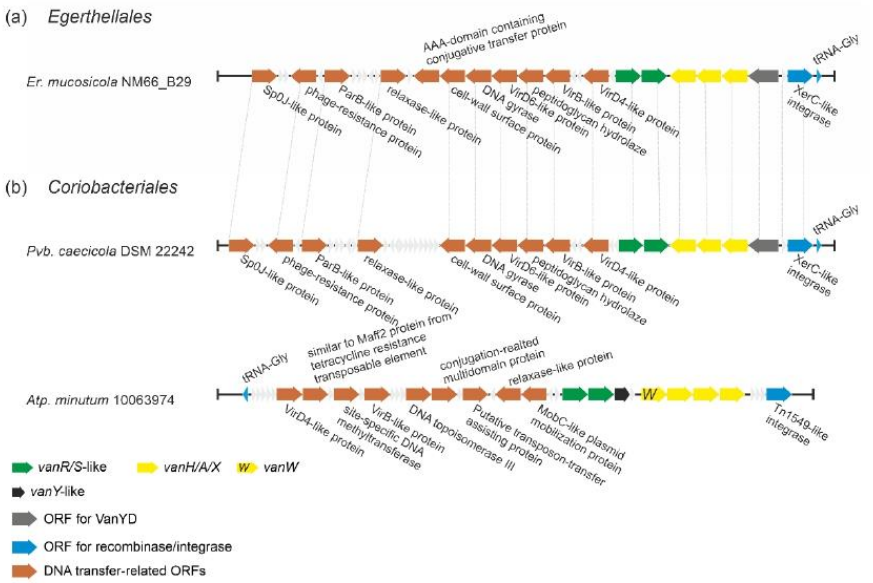


Figure 8. A scheme of putative MGEs found in (**a**) *Er. mucosicola* (*Eggerthellales*) and (**b**) *Coriobacteriales* spp. *Pvb. caecicola* and *Atp. minutum.* MGEs from *Er. mucosicola* and *Pvb. caecicola* seem to be almost identical (identical genes are joined with gray dashed lines), while MGE from *Atp. minutum* differs from both. Legend below the figure explains the color-coding of the scheme.

of the main sources of *van*-genes dissemination throughout pathogens [52]. Thus, we decided to check whether MGEs from the three abovementioned species corresponded to those already known. For this, we compared putative integrases/recombinases from *Er. mucosicola* NM66_B29, *Pb. caecicola* DSM 22242 and *Atp. minutum* 10063974 with integrases found in known MGEs carrying *van*-genes (ESM Table 2). It came out that the putative integrase from *Atp. minutum* 10063974 (EMZ42128) was identical to the integrase

from *Enterococcus faecalis* transposon Tn1549 [53]. Further comparison of Tn1549 genes with the genes from *Atp. minutum* 10063974 MGE confirmed that they were identical. However, both integrases from *Er. mucosicola* NM66_B29 and *Pb. caecicola* DSM 22242 (EMZ42128 and MVX60893, respectively) were only slightly related to the transposases/integrases from Tn1549 and *Enterococcus faecium* insertion sequence IS1216V [54], sharing only 18.6% and 14.6% of amino acid sequence identity respectively. Thus, MGE found in *Er. mucosicola* NM66_B29 and *Pb. caecicola* DSM 22242 might represent a novel MGE, carrying *van*-genes. Notably, this last putative MGE seemed to code for a VanYD protein [55], which is a D-Ala-D-Ala carboxypeptidase belonging to the penicillin binding protein family, structurally unrelated with VanY M15 peptidases (see below).

2.1.7. Occurrence of *vlgs* in *Bacillales* spp. (Firmicutes phylum)

Soil low G-C Gram-positives and in particular bacilli were considered as a one of possible sources of *vlgs* for pathogens, as it was considered that ancestral *vanHAX*-cluster might had evolved in one of such species (e.g. those belonging to the *Paenibacillus* genus) and then disseminated to pathogens in transposon-mediated fashion [28]. We decided to test such hypothesis by screening 2379 full genome assemblies of *Bacillales* spp. (Supplementary Excel File 3), available in GenBank, searching for *vlgs*. The results indicated that *vlgs* were quite rare in *Bacillales* spp. genomes: *vanHAXRS* were found in the genomes assemblies of *Brevibacillus laterosporus* E7593-50, *Thermoactinomyces vulgaris* and *Paenibacillus sonchi* LMG 24727 (Figure 9, Supplementary Excel File 3). *vanAX*-pseudogenes were also found in *Paenibacillus yonginensis* DCY84, while *vanHA*-genes (degraded to different extents) were found in 7 strains of *Paenibacillus larvae* (Figure 9, Supplementary Excel File 3). Only one *vanY*-like gene was found in *Bbac. laterosporus* E7593-50. Most of *vlgs* were found adjacent to transposase-related genes, except the cases of *Pnb. yonginensis* DCY84, *Pnb. sonchi* LMG 24727 and *Tam. vulgaris*, by the way this last strain being known to be naturally competent for exogenous DNA [56]. Overall, such findings indicated that the occurrence of *vlgs* in bacilli is not comparable to their distribution in most of the orders belonging to actinobacteria, making it highly unlikely for *van*-genes to arise independently in bacilli.

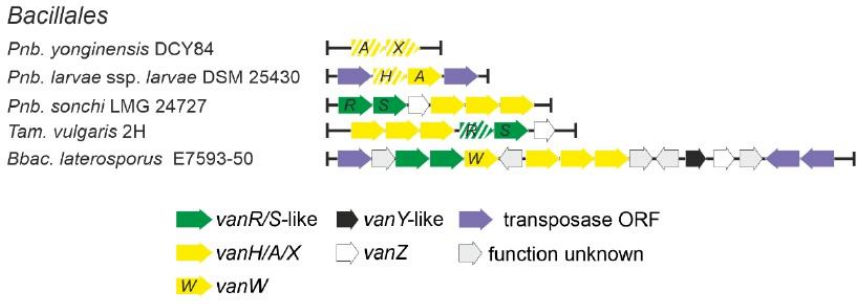


Figure 9. Arrangements of *vlgs* discovered in species belonging to *Bacillales* order. Please refer to the main text for more details; genus-names were abbreviated according to ESM Table 1. Legend below the figure explains the color-coding of the scheme. Pseudogenes are shaded.

2.2. Phylogeny of VanY-like carboxypeptidases

Surprisingly, *vanY*-like genes were the most common *vlgs* found in actinobacteria. We decided to reconstruct the phylogeny of VanY-like proteins to comprehend such variety and its relation with similar proteins described in low G-C Gram-positives, including pathogens such as GPA resistant enterococci To this purpose, we selected 251 proteins (see Methods section) coming either from actinobacteria or from low G-C Gram-positives, which, according to MEROPS peptidase database [57,58], belonged to M15B subfamily of M15-family of metallopeptidases (mostly carboxypeptidases and dipeptidases). Other

subfamilies are M15A, composed by a specific group of the so-called *Streptomyces*-type zinc-D-Ala-D-Ala carboxypeptidases [59], and M15D, to which VanX D-Ala-D-Ala dipeptidases belong (see the following section) [60,61]. To check that our selected proteins were VanY-like and no other carboxypeptidases, we controlled their sequence by CD-Search [58] and excluded those sharing the putative peptidoglycan binding domain on *N*-terminal region, which is typical of M15A proteins.

Reconstruction of Maximum-likelihood phylogeny of VanY-like proteins from our dataset yielded a tree, where 5 distinct clusters might be differentiated (Figure 10, ESM Figure 3). Cluster Y1 (Figure 10, ESM Figures 3 and 4) did outgroup the tree and contained

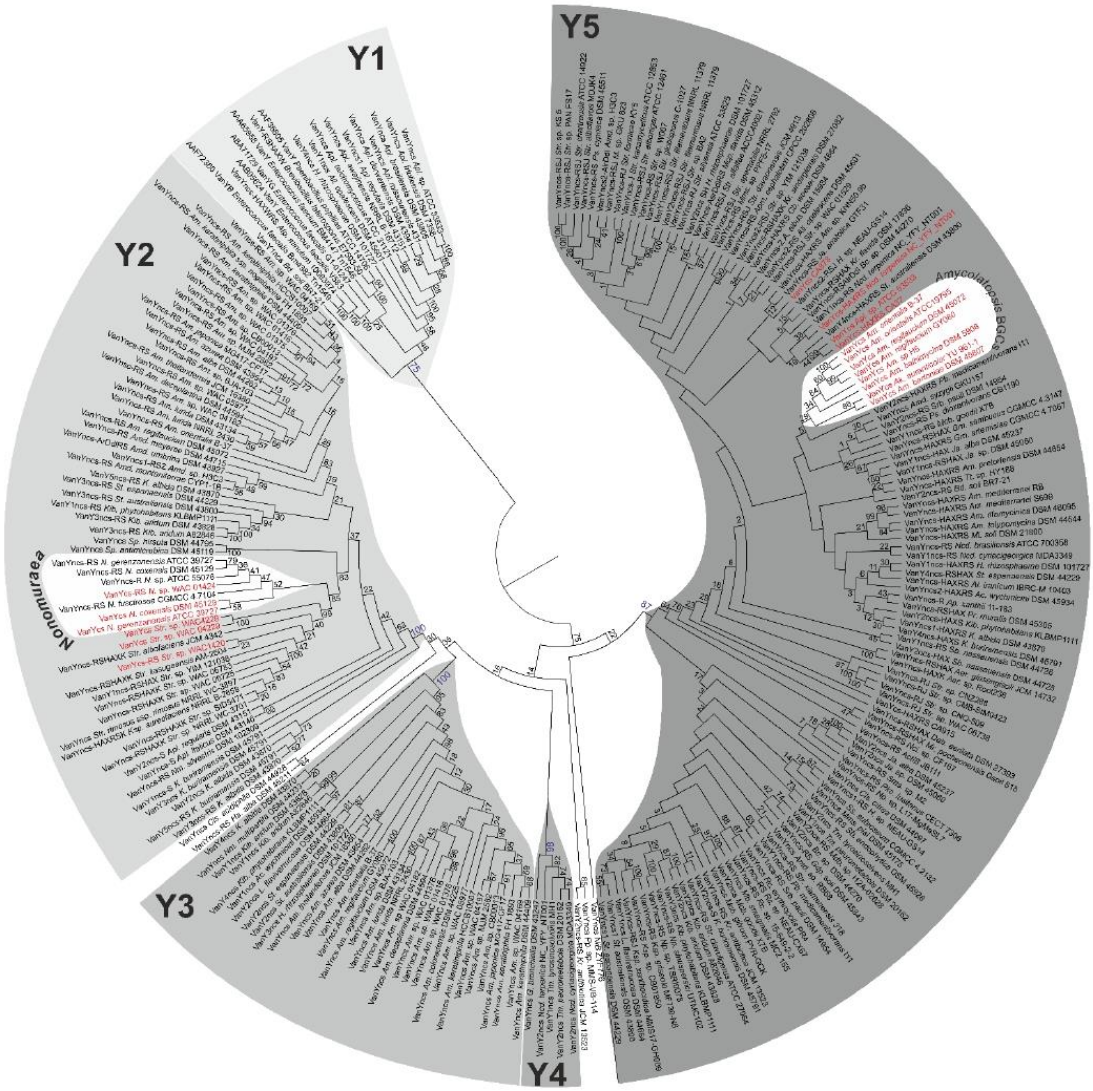


Figure 10. Maximum-likelihood phylogenetic tree of 251 VanY-like M15B carboxypeptidases. To show the topology of the tree better, branch lengths were ignored; the same tree drawn to scale is given in ESM Figures 3-7. Five well supported clusters – Y1 to Y5 – were distinguished on the tree. "cs/ncs" abbreviations in the label at the tip of each branch mean "cluster-situated/non-cluster-situated". BGC-encoded proteins are given in red. Importantly, the mostly studied VanY$_n$ (Dbv7) from the A40926 producing *N. gerenzanensis* ATCC 397272 [62] belongs to Y2, while VanY$_{Ab}$ (coming from the balhimycin producer *Am. balhimycina* DSM 5908 [63]) to Y5 cluster.

VanY-like proteins originating from enterococci and other low G-C Gram-positives, consistently with previous reports [37]. Differently from what described a decade before [38], this cluster included also VanY-like proteins coming from different *Actinoplanes* spp., from *Bd. soli* BR7-21, *H. rhizosphaerae* DSM 101727 and *All. opalescens* DSM 45601 (Figure 10, ESM Figure 4). This discrepancy is due to the increasing number of genomes that have become accessible meanwhile. Notably, all the genes corresponding to Y1 proteins in actinobacteria were 'orphan' (i.e. not co-localized with any other *vlgs*, see Figures 3, 4 and 7, and Supplementary Excel File 2).

Contrary, VanY-like proteins from large Y2 cluster were almost exclusively encoded by genes adjacent to *vanRS*-like pairs and/or other *vlgs* (Figure 10, ESM Figure 5, Supplementary Excel File 2). They were found in the genomes of multiple orders of actinobacteria (Figure 10, ESM Figure 5, Supplementary Excel File 2). Importantly, Y2 included VanY-like proteins from *Nonomuraea* GPA BGCs (including the mostly studied $VanY_n$ from *N. gerenzanensis* ATCC 39727 [62]) and pekiskomycin BGCs (*Strtreptomyces* spp. WAC1420, WAC4229, WAC 04229 [49]).

Next clusters – Y3 and Y4 (Figure 10, ESM Figure 6) – were composed of VanY-like proteins coming from *Pseunocardiales* and *Corynebacteriales*, respectively, with corresponding genes being 'orphan', again.

Finally, Y5 was the last cluster defined within the tree of VanY-like proteins. It was the biggest and well separated from the others, although the internal branching pattern was not completely clear, often lacking a trustable bootstrap support (Figure 10, ESM Figure 6). Nevertheless, genes, corresponding to Y5-proteins, were found either co-localized with different other *vlgs*, or 'orphan'. VanY-like proteins coded by *Pseudonocardiales* spp. GPA BGCs, including the mostly studied $VanY_{Ab}$ from the balhimycin producer *Am. balhimycina* [63], formed a distinct subclade within Y5. VanY-like proteins encoded within CA878, CA37, CA915, *auk* and *Ncd. terpenica* NC_YFY_NT001 BGCs were also found in Y5.

2.3. Phylogeny of VanHAX

VanH lactate dehydrogenases. For the phylogenetic reconstruction of VanH, we used a dataset of 156 VanH proteins from actinobacteria and low G-C Gram-positives, with SCO2118 lactate dehydrogenase serving as an outgroup (see Methods section). It appeared that VanH proteins were quite conserved, and a branching pattern with completely reliable bootstrap support was difficult to obtain (ESM Figure 8). Nevertheless, few features could be presumed with certainty. First, VanH-proteins coded within the putative MGEs of *Er. mucosicola* NM66_B29 and *Pvb. caecicola* DSM 22242 reliably outgrouped all other proteins (ESM Figure 8). Then, a well-defined cluster (named VH1, ESM Figure 8) was composed of VanH proteins from pathogens and from soil low G-C Gram-positives. VH1 was not that close to the base of the tree, with multiple actinobacterial VanH proteins out-grouping it (ESM Figure 8). Another well-supported cluster – VH2 (ESM Figure 8) – was composed of VanH proteins coming from different *Streptomyces* spp., including those from A47934 BGC (*Str. toyocaensis* NRRL 15009) and pekiskomycin BGC (*Streptomyces* sp. WAC 04229). A third big cluster (VH3) was formed by VanH proteins coming from different GPA non-producing *Streptomyces* spp. Finally, VanH-proteins coded within *tei*, *auk*, CA37 and CA915 also grouped together on the tree (ESM Figure 8).

VanA D-Ala-D-Lac-ligases. Overall, from previous works, it seems plausible that VanA D-Ala-D-Lac-ligases are a specialized evolutionary branch of D-Ala-D-Ala ligases (Ddl) involved in the "primary metabolism" of cell wall in actinobacteria [64]. As reported above, Ddl-like proteins were also found encoded in metagenomic-sourced CA37, CA915 and in the *auk* BGCs; additionally, in *Micromonosporales* we found a peculiar putative *pdx*-operon (see section *2.1.3*) composed with genes for a PALP threonine dehydratase, a Ddl-like ligase and a VanX-like dipeptidase, adjacent to a *vanRS*-like regulatory pair. Thus, here we decided to test the phylogeny of VanA ligases together with the above-mentioned

Ddl-like proteins on a background of 'house-keeping' Ddl-ligases from main actinobacterial orders. The protein set (see Methods section) used for this phylogenetic reconstruction contained 153 VanA ligases from actinobacteria and low G-C Gram-positives, 12 Ddl-like ligases from *Micromonosporales* spp., the 3 Ddl-like ligases from CA37, CA915 and *auk* BGCs, as well as 81 'house-keeping' actinobacterial Ddl-ligases.

In the resulting phylogenetic tree (ESM Figure 9), 'house-keeping' Ddl-ligases formed a number of distinct clades which corresponded to the orders of origin (ESM Figure 9). There, Ddl-like ligases coded in CA915, CA37 and *auk* BGCs grouped together with "house-keeping" Ddl-ligases from order *Eggerthellales,* while Ddl-like ligases from *Micromonosporales* (coded in putative *pdx*-operons) formed a distinct clade among the other clades of 'house-keeping' Ddl-ligases (ESM Figures 9 and 10). Most strikingly, clades for VanA ligases from low G-C Gram-positives (ESM Figure 11) and actinobacteria (ESM Figure 12) (both out-grouped by VanA ligases coded within the putative MGEs of *Er. mucosicola* NM66_B29 and *Pvb. caecicola* DSM 22242) clustered together with 'house-keeping' Ddl-ligases from the order *Coriobacteriales* (ESM Figure 9).

Composition of the main actinobacterial VanA-clade (ESM Figures 9 and 12) required further comments. The resolution of this clade was high-enough to distinguish four well-supported clusters – VA1-4. VA1 was formed by the VanA ligases from GPA non-producing streptomycetes together with the one coded in A47934 GPA BGC from *Str. toyocaensis* NRRL 15009. Other VanA ligases from GPA non-producing streptomycetes composed VA2, while VA3 contained proteins from different groups of actinobacteria. Finally, VA4 was formed by VanA ligases coded within GPA BGCs of *Amycolatopsis* spp. VanA-ligases from CA915, CA37 and *tei* BGCs grouped together on the tree. Notably, VanA-ligases from pekiskomycin BGCs (from *Str.* sp. WAC 04229 and *Str.* sp. WAC1420) grouped together with VanA from *Str. varsoviensis* NRRL B-3589, which does not carry any GPA BGC.

VanX M15D dipeptidases. According to MEROPS database, VanX-dipeptidases belong to the same M15 family as VanY-carboxypeptidases, but group in the M15D subfamily. The dataset used for the phylogenetic reconstruction contained 155 VanX dipeptidases coded within different *vanHAX*-operons from actinobacteria and low G-C Gram-positives and 12 VanX-like dipeptidases coded in the putative *pdx*-operon found in *Micromonosporales.* So-obtained phylogenetic tree revealed that VanX-like dipeptidases coded in *pdx*-operon and VanX proteins from low G-C Gram-positives formed well-supported clusters (VX1 and VX2, respectively, ESM Figure 13). All other actinobacterial VanX-like dipeptidases were out-grouped by VanX-like dipeptidases coded within the putative MGEs of *Er. mucosicola* NM66_B29 and *Pvb. caecicola* DSM 22242 (ESM Figure 13). Unfortunately, internal topology of the latter clade had non-optimal bootstrap support, since VanX-proteins appeared well conserved (ESM Figure 14). However, it was possible to distinguish two clusters – VX3 and VX4. Interestingly, composition of VX3 and VX4 corresponded to the clusters VA4 and VA3, respectively, of VanA-like ligases tree (ESM Figure 12). Finally, VanX dipeptidases coded in pekiskomycin BGCs (from *Str.* sp. WAC 04229 and *Str.* sp. WAC1420) again grouped together with VanX from *Str. varsoviensis* NRRL B-3589, which is not a GPA producer.

van-genes from pathogenic enterococci often code a peculiar group of bifunctional D,D-peptidases/D,D-carboxypeptidases known as VanXY [65]. In 2014, a comparative structural study of VanXY, VanX and VanY [57] assumed VanXY-peptidases to be a specialized evolutionary branch of VanY-carboxypeptidases in pathogens. This assumption was supported with a phylogenetic reconstruction of M15 family peptidases [57]. However, rather few protein sequences were available at that time. Thus, we decided to check the phylogeny of VanXY-peptidases in relation to our VanX- and VanY-datasets simultaneously. We used a dataset of 425 proteins, including 7 VanXY-peptidases. The topology of the resulting tree correlated with the topologies of the trees received for the VanX- and VanY-datasets separately. At the same time, VanXY-peptidases appeared to root deeply

within VanY-clade (ESM Figure 15) from low G-C Gram-positives (corresponding to Y1 cluster of VanY-carboxypeptidases, Figure 10). Therefore, our large-scale reconstruction was in line with the previously made assumption about VanXY origin [57].

2.4. Phylogeny of VanRS-like two-component regulatory pairs

In the course of our screen for *vlgs*, many of them were found co-localized with *vanRS*-like regulatory pairs. *vanRS*-like regulatory pairs were also found adjacent to *i)* putative *pdx*-operon in *Micromonosporales, ii)* putative operons formed with genes coding for alanine/aspartate racemase, Ddl and VanY-like carboxypeptidase (like in *All. opalescens* DSM 45601 and *Mcp. flavida* DSM 45312, respectively, Figure 3), or *iii)* genes for β-lactamases (as in *Xa. phaseoli* DSM 45730, Figure 4). Additionally, multiple BGCs for type V GPAs and *feg*-like BGCs from *Streptomyces, Microbispora* and *Actinomadura* spp., as well as for type

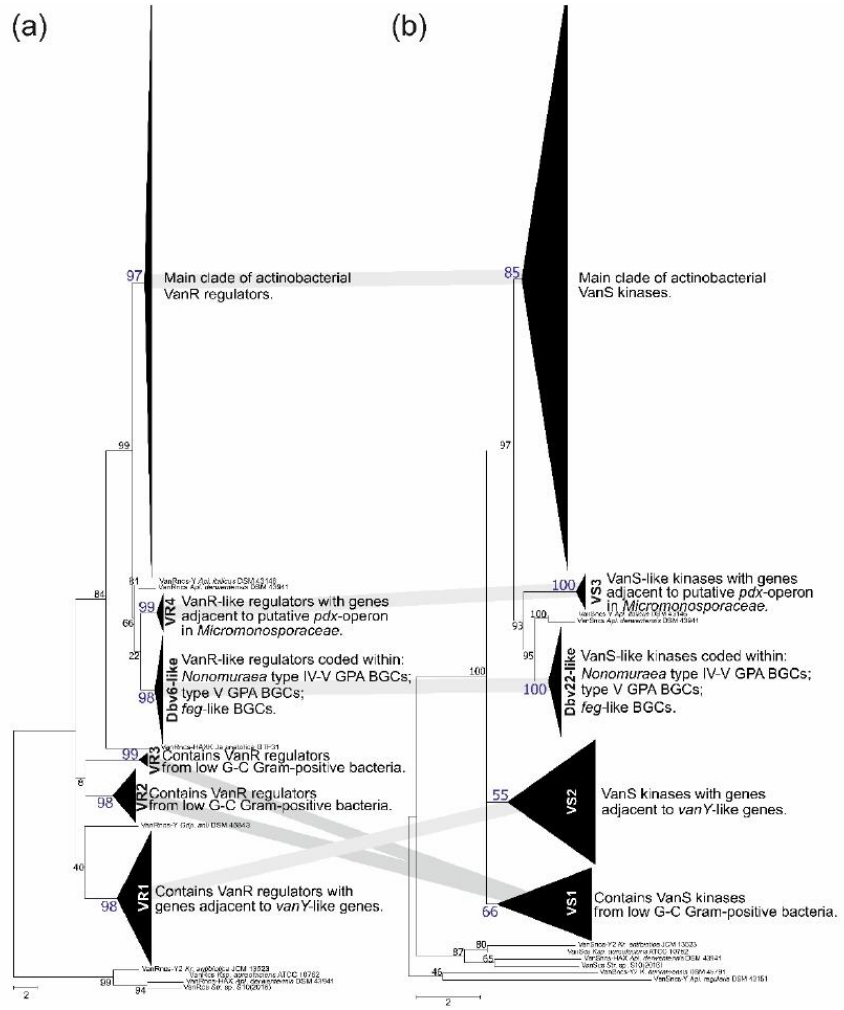


Figure 11. Phylogenetic trees of actinobacterial VanR-like response regulators (**a**) and VanS-sensor histidine kinases (**b**). Defined clusters on each tree were collapsed; for the expanded versions please refer to ESM Figures 16-31. Coherent clusters are joined with thick grey lines. Please see main text for more details. Scale bar represents number of substitutions per site.

IV GPAs from *Nonomuraea* spp. were found to carry *vanRS*-like regulatory pairs. To clarify if, how, and to what extent all these VanR-like response regulators and VanS-like sensor histidine kinases are related to each other and with the VanRS-proteins from low G-C Gram-positives, we reconstructed separately phylogenies of two datasets, one for VanR- and the other for VanS-like proteins. First dataset contained 295 proteins, while the second was composed of 313 proteins. Such discrepancy in numbers derives from the fact that *vanRS*-like pairs often lacked one of the genes, or had it impaired (as pseudogene). Overall, final topology of both trees was quite coherent, implying the co-evolution of VanR- and VanS-like proteins coded within one gene pair (Figure 11); this allowed us to define a set of well-supported clusters (named VS in VanS- and VR in VanR-trees, respectively). Both trees showed a set of clusters formed as basal clades as well as defined crown groups. Basal clades formed VR1-4 clusters plus a Dbv6-like cluster in VanR-phylogenetic tree and VS1-3 clusters plus a Dbv22-like cluster in VanS-phylogenetic tree. These data were consistent with what previously reported on Dbv6/Dbv22, which is a cluster-situated two-component regulatory pair in A40926 BGC from *N. gerenzanensis,* not grouping with "classic" VanRS-like proteins from other GPA BGCs [66].

Proteins from coherent clusters VR1 (VanR-tree, Figure 11a, ESM Figure 16) and VS2 (VanS-tree, Figure 11b, ESM Figure 17) had genes co-localized with *vanY*-like genes, coding for Y2-cluster VanY-like proteins (Figure 10); overall topology of Y2 was found similar to both VR1 and VS2 (ESM Figures 4, 16 and 17). Then, VR2+3 and VS1 clusters were also coherent and contained VanS- and VanR-like proteins found in low G-C Gram-positives, as well as the ones coded within the MGEs from *Coriobacteriales* and *Eggerthellales* spp. (Figure 11, ESM Figures 18 and 19). Next pair of coherent clusters contained VanS- and VanR-like proteins coded within *feg*-like BGCs, type V GPA BGCs and type IV GPA BGCs from *Nonomuraea* spp. (Figure 11). Basically, both clades were formed with orthologues of either Dbv6 or Dbv22 from *dbv* BGCs, thus they received corresponding naming (ESM Figures 20 and 21). Finally, last pair of coherent basal clusters were VR4 and VS3 (Figure 11), formed by VanR- and VanS-like proteins coded adjacent to putative *pdx*-operon from *Micromonosporales* (ESM Figures 22 and 23).

Similar to what was previously observed for VanX-, VanA- and VanH-phylogenies, the resolution of VanR and VanS-trees crown groups was not perfect (ESM Figures 24 and 25). Nevertheless, we defined three additional clusters in the VanR-phylogenetic tree – VR5-7 (ESM Figures 24 and 26). VanR-regulators from clusters VR5-6 were coded adjacent to *vanY*-like genes in different actinomycetes, while VR7 was composed with VanR-regulators coded adjacent to *vanHAXYK* in *Streptomyces* spp. VR7 contained VanR coded within pekiskomycin BGC from *Str.* sp. WAC1420. The resolution of VanS-tree crown group was better, allowing to distinguish there five additional clusters – VS4-8 (ESM Figure 25). Here, VS4 and VS5 clusters were formed by VanS-kinases coming from different actinobacterial orders and coded adjacent to *vanY*-like genes and *vanZ*-like genes, respectively (ESM Figures 27 and 28). Then, proteins forming VS6 cluster were coded adjacent to *vanHAX*-genes in different actinobacteria (ESM Figure 29), while VS7 was formed by VanS-kinases coded adjacent to *vanHAXYK*-genes from *Streptomyces* spp., including the one from pekiskomycin BGC (*Str.* sp. WAC1420, ESM Figure 30). Finally, proteins from VS8 were coded adjacent to *vanHAXK*-genes in streptomycetes; VS8 also included VanS from model *Str. coelicolor* (ESM Figure 31).

Some other notable features in VanS- and VanR-crown groups phylogenies require comments. First, both reconstructions placed VanS- and VanR-proteins encoded within *tei* and A47934 BGCs together (ESM Figures 25 and 26). Second, both trees showed evidence of a possible evolution of VanRS-like regulatory pair, expanding its regulon control from *van* genes to some other genes: although VanRS-pair from *All. opalescens* DSM 45601 and *Mcp. flavida* DSM 45312 were related to VanRS-proteins coded adjacent to *vanHAX*-genes, corresponding *vanRS*-like gene pairs actually were co-localized with genes for alanine or aspartate racemases, together with *ddl* and *vanY*-like genes.

3. Discussion

In the current work we aimed to address certain unclear issues about *van*-genes, their distribution and phylogeny. Although we are aware that our results might risk to generate more questions than answers, we tried in the following section to summarize what are, in our opinion, the most relevant findings.

Actinobacteria are the most likely primary sources of vlgs. First of all, the results of our screens, covering more than 7000 actinobacterial genomes and 2000 *Bacillales* genomes, revealed that *vlgs* are abundant within Actinobacteria phylum (with an incidence of *ca.* 13%), while vanishingly rare in bacilli and related spp. belonging to Firmicutes phylum. This disproves the idea of *van*-like genes and operons emerging independently in soil-dwelling actinobacteria and bacilli [28]. Abundance and context variability of actinobacterial *vlgs* point to Actinobacteria phylum as to the original source of *van*-genes. At the same time, the assumption that ubiquitous low G-C soil Gram-positives served as a bridge for *vlgs* to arrive in pathogens [28] seems likely. Such transfer was probably achieved via MGEs, which often were proved to carry *vlgs* in low G-C soil Gram-positives and pathogens [52]. In fact, we found the classical Tn1549 transposon in the actinobacterium *Atp. minutum* 10063974, although, until now, Tn1549-like transposons were described only in enterococci [53]. It is hard to say whether *Atp. minutum* 10063974 might represent the original actinobacterial source of Tn1549, or if this is an example of reverse HGT event: G-C content (estimated from *vanRSYWHAX*-genes) of *Atp. minutum* 10063974 Tn1549 is the same as in *Enterococcus faecalis* BM4382 Tn1549 – 47%. However, this is comparable to the overall genome G-C content of *Atp. minutum* 10063974 which is *ca.* 48%. At the same time, two novel, putative MGEs (very similar to each other) were also found in the genomes of the actinobacteria *Er. mucosicola* NM66_B29 and *Pvb. caecicola* DSM 22242. These MGEs coded unusual transposases, carrying *vlgs* coding proteins that in our phylogenetic reconstructions did not cluster with Van-proteins derived from low G-C Gram-positives. Once again, it is not clear if these MGEs might represent the 'original' actinobacterial *vlg*-carrying elements. G-C content of MGEs from *Er. mucosicola* NM66_B29 and *Pvb. caecicola* DSM 22242 (as estimated from the G-C content of their *vanRSHAXY-D*-genes) is *ca.* 55%. This is higher than the usual G-C content of MGEs from low G-C Gram-positives, but lower than the overall genome G-C content of *Er. mucosicola* NM66_B29 and *Pvb. caecicola* DSM 22242 (64.6% and 62.4% respectively). Thus, the study of *vlgs* in soil mobilome requires more detailed and focused research, which could in future contribute to a better understanding on how *vlgs* were disseminated from actinobacteria.

vlgs are distributed in actinobacteria without any evident strict correlation to GPA BGCs distribution, although a complex co-evolution with BGCs likely occurred. Our comparative genomic analysis of the different orders belonging to Actinobacteria phylum showed that *vlgs* are not necessarily co-localized with type I-IV GPA BGCs; moreover, in the majority of the cases, *vlgs* were found in GPA non-producers as well as in type V GPA producers. Consequently, it is reasonable to assume that non-cluster-situated *vlgs* existed independently from GPA BGCs and might actually be considered a preadaptation feature, which then facilitated the spread of GPA BGCs within Actinobacteria phylum. Next, GPA cluster-situated *vlgs* are not monophyletic, meaning that different GPA BGCs likely acquired these genes from different sources of the vast actinobacterial pool. All our phylogenetic reconstructions in fact showed that cluster-situated *vlgs* emerged randomly on the trees, surrounded by non-cluster-situated *vlgs*. One of the most prominent evidence came from VanY-phylogeny, where *Nonomuraea* spp. and *Amycolatopsis* spp. GPA cluster-encoded VanY-carboxypeptidases belonged to distant clusters (Y2 and Y5, Figure 9), separated by multiple other non-cluster-encoded VanY-proteins. Such evidence cautions against overt reliance on out-of-context phylogenetic reconstructions of cluster-situated *vlgs* [18], which might distort the real situation.

Moreover, *vlgs* and GPA BGCs seem to have shared complex co-evolution patterns. We were able to reconstruct one of the most obvious, observed for the self-resistance phenotype in the A40926 producer *N. gerenzanensis*, which relied on the expression of the

cluster-situated *vanY*-like gene – *dbv7*. In the corresponding *dbv7* knockout mutant, GPA resistance was significantly reduced, but not completely abolished [62]. Screening the full genome of *N. gerenzanensis* [67], the reason became evident: an additional *vanY* allele was found far away the *dbv* BGC. Interestingly, this allele was co-localized with a *vanRS*-like regulatory pair, whereas *dbv7* was not. Same situation was observed for *N. coxensis* DSM 45129 – a recently described producer of A40926-like type IV GPA [68]: a cluster-situated *dbv7* homologue (almost identical) and a non-cluster-situated *vanRSY*-triad were identified in its genome. Other two screened genomes of *Nonomuraea* spp. – the GPA non-producing *N. fuscirosea* CGMCC 4.7104 and the kistamicin producer *Nonomuraea* sp. ATCC 55076 – contained close homologues of the *vanRSY*-triad. In addition, the genome of the other putative GPA producer *Nonomuraea* sp. WAC 01424 presented the GPA BGC just upstream the *vanRSY*-triad, while no *vanY*-allele was found in the BGC itself. Finally, in our reconstructed phylogeny of VanY-like carboxypeptidases, *dbv* and *noc* cluster-encoded VanY-proteins out-grouped (Y1 cluster) all other ones (Figure 10). We believe that such a picture is a result of a series of HGT events. In our hypothesis, summarized in Figure 11, first, *noc* and *dbv* protoclusters recruited *vanY*-gene from the *vanRSY*-triad present in the common ancestor of *Nonomuraea* spp., whereas WAC 01424 and kistamycin ancestral BGCs did not acquire it. Then, *noc* and *dbv*, through another HGT event, were introduced in *Nonomuraea* ancestors already carrying the *vanRSY*-triad. Another HGT event delivered WAC 01424 BGC to *Nonomuraea* sp. WAC 01424 and kistamycin BGC to *Nonomuraea* sp. ATCC 55076, explaining why *vanRSY*-triad is the only resistance determinant in these strains. Finally, none of such HGT event occurred in *N. fuscirosea* CGMCC 4.7104, leaving this strain without any GPA BGCs, but anyhow carrying the *vanRSY*-triad.

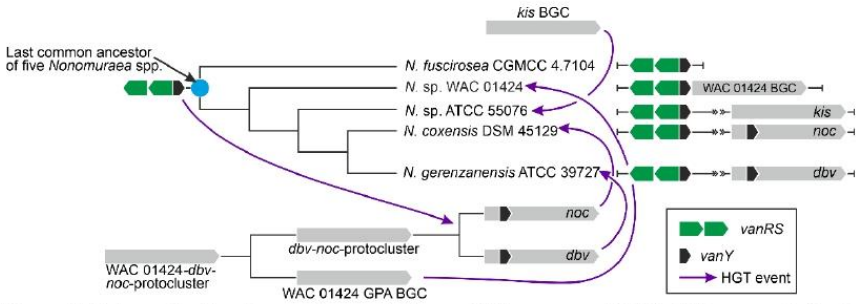


Figure 12. Scheme for the discussed scenario for *vanY*-like genes and GPA BGCs co-evolution in *Nonomuraea* spp. Please refer to the main text for more details.

Is the GPA resistance the original function of vlgs? As reported above, actinobacteria are an extremely abundant source of *vlgs*. It is quite difficult to consider that such variety of *vlgs* in GPA non-producing actinobacteria is needed to protect them versus the GPA producers, which definitively seemed quite rare. Alternatively, *vlgs* might have natural functions in cell wall remodeling in response to environmental and/or developmental triggers. For instance, VanY-like M15B carboxypeptidases are extremely abundant in actinobacteria, forming diverged evolutionary lineages; it is tempting to suspect their role in, for instance, cell wall remodeling during the life cycle (a function indeed shown for some other D-Ala-D-Ala carboxypeptidase [69]). Our finding of co-called *pdx*-operon, common for *Micromonosporales*, contributed to such idea. This putative operon is composed of a PALP-like serine-threonine dehydratase, a Ddl ligase and a VanX-like dipeptidase genes, co-localized with a *vanRS*-like regulatory pair. This organization suggests an alternative mechanism of peptidoglycan precursor remodeling, which reminds the introduction of D-Ala-D-Ser termini in the cell wall of some enterococci, showing a low level of GPA resistance [14]. It is unknown whether the *pdx*-operon is functional and if its expression

leads to some cell wall remodeling, but such case indeed merits further experimental evaluation.

To conclude, we must say that current work only scratched the surface of actinobacterial *vlgs*, leaving many issues without a complete final evaluation and understanding. However, we hope that this should provoke other *in silico, in vitro* and *in vivo* studies, which will shed light on such an important question as GPA resistance. Our findings portray actinobacteria as a Pandora's box, hosting myriads of putative GPA resistance genes which might be transferred sooner or later to pathogens, significantly contributing to AMR. A thorough understanding of GPA resistance in actinobacteria may prove useful in the future surveillance of emerging mechanisms of resistance to clinically used GPAs. Although further experiments are necessary to show that the discovered *in silico* putative *vlgs* have a real function *in vivo* conferring GPA resistance, their study may reveal new insights of their biological functions in actinobacteria, augmenting our comprehension of this remarkable phylum.

4. Materials and Methods

4.1. Routine analysis of nucleic and amino acid sequences

All routine analytic work with nucleic acid and amino acid sequences was performed using Geneious 4.8.5 [70], Mega X [71]; routine amino acid sequence alignments were performed with Clustal Omega (EMBL-EBI) [72].

4.2. vlgs search pipeline

To perform the search for *van*-like genes (named *vlgs*), all genome assemblies of Actinobacteria – available at the time of work preparation (April 2020) as either full sequences or uncompleted ones – were retrieved from GenBank database (in few exceptions, genome assemblies were retrieved from RefSeq database), for a total of 7108 assemblies from species belonging to 26 established and 2 tentative orders within the Actinobacteria phylum, represented by 653111 corresponding nucleic acid sequences. Full list of the genome assemblies and corresponding nucleic acid sequences is given in Supplementary Excel Table 1. MultiGeneBlast tool [73] was utilized to screen chosen set of genome assemblies for *vlgs*. To do that, chosen set of genome assemblies belonging to each order was downloaded from GenBank (or RefSeq) in a genomic GenBank format (*.gbff). These files were then used to create offline MultiGeneBlast custom databases for each order belonging to the Actinobacteria phylum. MultiGeneBlast was run in a "homology" mode with the default settings, which included 25% minimal sequence coverage of the BLAST hits and 30% minimal amino acid sequence identity of BLAST hits. The maximum distance between genes in locus was increased to 40 kb (considering that *vlgs* inside a BGC might be separated by some other genes – like in A47934 BGC [48]). Two input files were used as queries for MultiGeneBlast search: one included *van*-genes from *Str. coelicolor* – *SCO3589-3590-3594-3595-3596* (*vanSRHAX*); and the other included balhimycin BGC-situated *van*-genes from *Am. balhimycina* – *DMA12_00360, DMA12_00365* and *DMA12_00370* (*vanS, vanR* and *vanY* respectively). First input files allowed to detect *vanHAX* orthologues co-localized with *vanRS*-like regulatory pairs, while the second helped to detect cases when *vanY*-like genes were co-localized with *vanRS* but lacked other *van*-genes in their genetic neighborhood. MultiGeneBlast outputs were then carefully examined, and amino acid sequences of the proteins coded with the so-identified *vlgs* were extracted. The information about these *vlgs*, including corresponding protein accession numbers, nucleic acid accession numbers and taxa are summarized in Supplementary Excel Table 2. To refine this initial screening, chosen sets of genomes for each order (highlighted in red in Supplementary Excel Table 2) were manually reexamined for *vlgs* using BLASTP [74] with SCO3589 (VanS), SCO3590 (VanR), SCO3592 (VanJ), SCO3593 (VanK), SCO3594 (VanH), SCO3595 (VanA), SCO3596 (VanX), CAG25753 (VanY), ELS50663 (VanZ) as queries. These last sets of genomes were chosen to cover all the genera within a certain order and to include all the genomes carrying known and putative type I-V GPA BGCs as well as *feg*-like BGCs. Selected hits were tested for orthology with queries using Reciprocal Best Hit BLAST approach. Information received here was used to build Figures 2-9.

4.3. Search for putative GPA-like BGCs

MultiGeneBlast tool [73] was used to screen all the genome assemblies for GPA-like BGCs. We utilized the same offline custom databases created for *vlgs*-screening. MultiGeneBlast was also run in a "homology" mode with the default settings, however the maximum distance between genes in one locus was increased to 60 kb. An input file was composed of teicoplanin BGC [75] genes: *tei4**, *tei8**, *tei15**, *tei17**, *tei23**, *tei24**, *tei28** and *tei29**, coding for: an ABC transporter, teicoplanin halogenase, StrR-like transcriptional regulator, prephenate dehydrogenase, l-4-hydroxyphenylglycine biosynthesis enzyme HpgT, type III polyketide synthase DpgA (involved in the biosynthesis of both l-4-hydroxyphenylglycine and l-3,5-dihydroxyphenylglycine), HmaS and Hmo (l-3,5-dihydroxyphenylglycine biosynthesis enzymes), respectively. All these genes have their orthologues in BGCs for type I-V GPAs and in *feg*. MultiGeneBlast outputs were manually examined and the nucleic acid sequences containing MultiGeneBlast hits were applied for upstream antiSMASH [76] analysis. A list of obtained putative GPA- and *feg*-like BGCs is given in Supplementary Excel Table 2.

4.4. Phylogenetic reconstruction

Since screening revealed hundreds of *vlgs*, it was difficult to use all the sequences for a comprehensive phylogenetic analysis. Therefore, phylogenies were reconstructed for sets of proteins coded with *vlgs* from chosen genomes for each order (highlighted in red in Supplementary Excel Table 2). Final protein datasets for the reconstruction of phylogenies of VanY-like carboxypeptidases; VanH; VanA and Ddl; VanX and VanX-like dipeptidases; VanY-like carboxypeptidases, VanX and VanX-like dipeptidases; VanR-like regulators; VanS-like sensor histidine kinases; are given in Supplementary FASTA Files 1-7, respectively. Some additional information was coded in the name of each protein sequence to indicate *i)* whether this protein is BGC-encoded or not; *ii)* whether this protein is coded with an 'orphan' gene or corresponding gene is co-localized with other *vlgs*. For instance, "VanYncs-HAXRS_Tt_sp_HY188" indicates that this is VanY-like non-BGC-encoded peptidase, with the corresponding gene co-localized with *vanHAXRS*, coming from *Tomitella* sp. HY188.

Mega X [71] package was used to perform phylogenetic reconstructions. On a road to representative phylogenetic trees we always followed the next algorithm. First, multiple amino acid sequence alignments for each dataset were generated using Muscle; obtained alignments were manually curated to preserve as much meaningful data as possible. Then curated multiple sequence alignments were analyzed using Mega X model finder to discover the most appropriate evolutionary models and the best scoring models were applied to generate Maximum likelihood phylogenies for each protein dataset. Similar approach was used to generate 16S rRNA gene phylogenetic trees, which could be found on Figures 2-5. Final topologies of either protein or gene trees were derived from 500 bootstraps.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Supplementary materials file; Supplementary Excel Files 1-3; Supplementary FASTA Fils 1-7.

Author Contributions: O.Y. conceived the research with the conceptual input from V.F. and F.M., A.A-V. and E.B. collected the data, E.B. and O.Y. analyzed the data, O.Y. and F.M. wrote the manuscript, all authors reviewed and approved the manuscript.

Funding: This work was supported by the public grant "Fondo di Ateneo per la Ricerca" 2019 and 2020 to F.M. and by the BG-09F grant of Ministry of Education and Science of Ukraine to V.F.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
Data Availability Statement: All the data is available from the corresponding author upon reasonable request.

Acknowledgments: A.A.-V. is a Ph.D. student of the "Life Science and Biotechnology" course at Università degli Studi dell'Insubria.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. 1929. *Bull. World Health Organ.* **2001**, *79*, 780–790.
2. O' Neil, J. Review on Antibiotic resistance. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. *Heal. Wealth Nations* **2014**, 1–16.
3. Ghosh, S.; Bornman, C.; Zafer, M.M. Antimicrobial resistance threats in the emerging COVID-19 pandemic: where do we stand? *J. Infect. Public Health* **2021**, *14*, 555–560, doi:10.1016/j.jiph.2021.02.011.
4. Marcone, G.L.; Binda, E.; Berini, F.; Marinelli, F. Old and new glycopeptide antibiotics: from product to gene and back in the post-genomic era. *Biotechnol. Adv.* **2018**, *36*, 534–554, doi:10.1016/j.biotechadv.2018.02.009.
5. Nicolaou, K.C.; Boddy, C.N.C.; Bräse, S.; Winssinger, N. Chemistry, biology, and medicine of the glycopeptide antibiotics. *Angew. Chemie - Int. Ed.* **1999**, *38*, 2096–2152, doi:10.1002/(sici)1521-3773(19990802)38:15<2096::aid-anie2096>3.0.co;2-f.
6. Sarkar, P.; Yarlagadda, V.; Ghosh, C.; Haldar, J. A review on cell wall synthesis inhibitors with an emphasis on glycopeptide antibiotics. *Medchemcomm* 2017, *8*, 516–533, doi:10.1039/c6md00585c.
7. Vollmer, W.; Blanot, D.; De Pedro, M.A. Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* **2008**, *32*, 149–167, doi:10.1111/j.1574-6976.2007.00094.x.
8. Perkins, H.R. Specificity of combination between mucopeptide precursors and vancomycin or ristocetin. *Biochem. J.* **1969**, *111*, 195–205, doi:10.1042/bj1110195.
9. Nitanai, Y.; Kikuchi, T.; Kakoi, K.; Hanamaki, S.; Fujisawa, I.; Aoki, K. Crystal structures of the complexes between vancomycin and cell-wall precursor analogs. *J. Mol. Biol.* **2009**, *385*, 1422–1432, doi:10.1016/j.jmb.2008.10.026.
10. Williams, D.H. The glycopeptide story - how to kill the deadly "Superbugs." *Nat. Prod. Rep.* **1996**, *13*, 469–477, doi:10.1039/NP9961300469.
11. Marschall, E.; Cryle, M.J.; Tailhades, J. Biological, chemical, and biochemical strategies for modifying glycopeptide antibiotics. *J. Biol. Chem.* **2019**, *294*, 18769–18783, doi:10.1074/JBC.REV119.006349.
12. Parenti, F.; Cavalleri, B. Proposal to name the vancomycin-ristocetin like glycopeptides as dalbaheptides. *J. Antibiot. (Tokyo).* **1989**, *42*, 1882–1883, doi:10.7164/antibiotics.42.1882.
13. Yushchuk, O.; Binda, E.; Marinelli, F. Glycopeptide antibiotic resistance genes: distribution and function in the producer actinomycetes. *Front. Microbiol.* **2020**, *11*, doi:10.3389/fmicb.2020.01173.
14. Binda, E.; Marinelli, F.; Marcone, G.L. Old and new glycopeptide antibiotics: Action and resistance. *Antibiotics* **2014**, *3*, 572–594, doi:10.3390/antibiotics3040572.
15. Arthur, M.; Quintiliani R., J. Regulation of VanA- and VanB-type glycopeptide resistance in enterococci. *Antimicrob. Agents Chemother.* **2001**, *45*, 375–381, doi:10.1128/AAC.45.2.375-381.2001.
16. Arthur, M.; Molinas, C.; Bugg, T.D.H.; Wright, G.D.; Walsh, C.T.; Courvalin, P. Evidence for *in vivo* incorporation of D-lactate into peptidoglycan precursors of vancomycin-resistant enterococci. *Antimicrob. Agents Chemother.* **1992**, *36*, 867–869, doi:10.1128/AAC.36.4.867.
17. Stegmann, E.; Frasch, H.J.; Kilian, R.; Pozzi, R. Self-resistance mechanisms of actinomycetes producing lipid II-targeting antibiotics. *Int. J. Med. Microbiol.* **2015**, *305*, 190–195, doi:10.1016/J.IJMM.2014.12.015.
18. Waglechner, N.; McArthur, A.G.; Wright, G.D. Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nat. Microbiol.* **2019**, *4*, 1862–1871, doi:10.1038/s41564-019-0531-5.
19. Culp, E.J.; Waglechner, N.; Wang, W.; Fiebig-Comyn, A.A.; Hsu, Y.P.; Koteva, K.; Sychantha, D.; Coombes, B.K.; Van Nieuwenhze, M.S.; Brun, Y. V.; et al. Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature* **2020**, *578*, 582–587, doi:10.1038/s41586-020-1990-9.
20. Xu, M.; Wang, W.; Waglechner, N.; Culp, E.J.; Guitor, A.K.; Wright, G.D. GPAHex-A synthetic biology platform for Type IV–V glycopeptide antibiotic production and discovery. *Nat. Commun.* **2020**, *11*, doi:10.1038/s41467-020-19138-5.
21. Mitchell, S.J.; Verma, D.; Griswold, K.E.; Bailey-Kellogg, C. Building blocks and blueprints for bacterial autolysins. *PLOS Comput. Biol.* **2021**, *17*, e1008889, doi:10.1371/JOURNAL.PCBI.1008889.
22. Cheng, A. V.; Wuest, W.M. Phylogeny-guided approach yields glycopeptides with unique action. *Trends Pharmacol. Sci.* **2020**, *41*, 297–299, doi:10.1016/j.tips.2020.03.002.
23. Gonsior, M.; Mühlenweg, A.; Tietzmann, M.; Rausch, S.; Poch, A.; Süssmuth, R.D. Biosynthesis of the peptide antibiotic feglymycin by a linear nonribosomal peptide synthetase mechanism. *ChemBioChem* **2015**, *16*, 2610–2614, doi:10.1002/cbic.201500432.
24. Hong, H.J.; Paget, M.S.B.; Buttner, M.J. A signal transduction system in *Streptomyces coelicolor* that activates expression of a putative cell wall glycan operon in response to vancomycin and other cell wall-specific antibiotics. *Mol. Microbiol.* **2002**, *44*, 1199-1211, doi:10.1046/j.1365-2958.2002.02960.x.
25. Patel, R.; Piper, K.; Cockerill, F.R.; Steckelberg, J.M.; Yousten, A.A. The biopesticide *Paenibacillus popilliae* has a vancomycin resistance gene cluster homologous to the enterococcal VanA vancomycin resistance gene cluster. *Antimicrob. Agents Chemother.* **2000**, *44*, 705–709, doi:10.1128/AAC.44.3.705-709.2000.

26. Fontana, R.; Ligozzi, M.; Pedrotti, C.; Padovani, E.M.; Cornaglia, G. Vancomycin-resistant *Bacillus circulans* carrying the *vanA* gene responsible for vancomycin resistance in enterococci. *Eur. J. Clin. Microbiol. Infect. Dis.* **1997**, *16*, 473–474, doi:10.1007/BF02471915.
27. Marshall, C.G.; Broadhead, G.; Leskiw, B.K.; Wright, G.D. D-Ala-D-Ala ligases from glycopeptide antibiotic-producing organisms are highly homologous to the enterococcal vancomycin-resistance ligases VanA and VanB. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 6480–6483, doi:10.1073/pnas.94.12.6480.
28. Aminov, R.I.; Mackie, R.I. Evolution and ecology of antibiotic resistance genes. *FEMS Microbiol. Lett.* **2007**, *271*, 147–161, doi:10.1111/j.1574-6968.2007.00757.x.
29. One Health Available online: https://www.who.int/news-room/q-a-detail/one-health (accessed on Oct 21, 2021).
30. Hutchings, M.I.; Hong, H.J.; Buttner, M.J. The vancomycin resistance VanRS two-component signal transduction system of *Streptomyces coelicolor*. *Mol. Microbiol.* **2006**, *59*, 923–935, doi:10.1111/j.1365-2958.2005.04953.x.
31. Hong, H.J.; Hutchings, M.I.; Neu, J.M.; Wright, G.D.; Paget, M.S.B.; Buttner, M.J. Characterization of an inducible vancomycin resistance system in *Streptomyces coelicolor* reveals a novel gene (*vanK*) required for drug resistance. *Mol. Microbiol.* **2004**, *52*, 1107–1121, doi:10.1111/j.1365-2958.2004.04032.x.
32. Adamek, M.; Alanjary, M.; Sales-Ortells, H.; Goodfellow, M.; Bull, A.T.; Winkler, A.; Wibberg, D.; Kalinowski, J.; Ziemert, N. Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in *Amycolatopsis* species. *BMC Genomics* **2018**, *19*, doi:10.1186/s12864-018-4809-4.
33. Shawky, R.M.; Puk, O.; Wietzorrek, A.; Pelzer, S.; Takano, E.; Wohlleben, W.; Stegmann, E. The border sequence of the balhimycin biosynthesis gene cluster from *Amycolatopsis balhimycina* contains *bbr*, encoding a StrR-like pathway-specific regulator. *J. Mol. Microbiol. Biotechnol.* **2007**, *13*, 76–88, doi:10.1159/000103599.
34. Gudeta, D.D.; Moodley, A.; Bortolaia, V.; Guardabassi, L. VanO, a new glycopeptide resistance operon in environmental *Rhodococcus equi* isolates. *Antimicrob. Agents Chemother.* **2014**, *58*, 1768–1770, doi:10.1128/AAC.01880-13.
35. Mengin-Lecreulx, D.; Texier, L.; Rousseau, M.; Van Heijenoort, J. The *murG* gene of *Escherichia coli* codes for the UDP-*N*-acetylglucosamine:*N*-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol *N*-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis. *J. Bacteriol.* **1991**, *173*, 4625–4636, doi:10.1128/JB.173.15.4625-4636.1991.
36. Banik, J.J.; Craig, J.W.; Calle, P.Y.; Brady, S.F. Tailoring enzyme-rich environmental DNA clones: a source of enzymes for generating libraries of unnatural natural products. *J. Am. Chem. Soc.* **2010**, *132*, 15661–15670, doi:10.1021/ja105825a.
37. Goldstein, B.P.; Selva, E.; Gastaldo, L.; Berti, M.; Pallanza, R.; Ripamonti, F.; Ferrari, P.; Denaro, M.; Arioli, V.; Cassani, G. A40926, a new glycopeptide antibiotic with anti-*Neisseria* activity. *Antimicrob. Agents Chemother.* **1987**, *31*, 1961–1966, doi:10.1128/AAC.31.12.1961.
38. Binda, E.; Marcone, G.L.; Pollegioni, L.; Marinelli, F. Characterization of VanYn, a novel D, D- peptidase/ D, D -carboxypeptidase involved in glycopeptide antibiotic resistance in *Nonomuraea* sp. ATCC 39727. *FEBS J.* **2012**, *279*, 3203–3213, doi:10.1111/j.1742-4658.2012.08706.x.
39. Binda, E.; Marcone, G.L.; Berini, F.; Pollegioni, L.; Marinelli, F. *Streptomyces* spp. as efficient expression system for a D,D-peptidase/D,D-carboxypeptidase involved in glycopeptide antibiotic resistance. *BMC Biotechnol.* **2013**, *13*, doi:10.1186/1472-6750-13-24.
40. Yushchuk, O.; Vior, N.M.; Andreo-Vidal, A.; Berini, F.; Rückert, C.; Busche, T.; Binda, E.; Kalinowski, J.; Truman, A.W.; Marinelli, F. Genomic-led discovery of a novel glycopeptide antibiotic by *Nonomuraea coxensis* DSM 45129. *ACS Chem. Biol.* **2021**, *16*, 915–928, doi:10.1021/acschembio.1c00170.
41. Nazari, B.; Forneris, C.C.; Gibson, M.I.; Moon, K.; Schramma, K.R.; Seyedsayamdost, M.R. *Nonomuraea* sp. ATCC 55076 harbours the largest actinomycete chromosome to date and the kistamicin biosynthetic gene cluster. *Medchemcomm* **2017**, *8*, 780–788, doi:10.1039/c6md00637j.
42. Yushchuk, O.; Ostash, B.; Truman, A.W.; Marinelli, F.; Fedorenko, V. Teicoplanin biosynthesis: unraveling the interplay of structural, regulatory, and resistance genes. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 3279–3291, doi:10.1007/s00253-020-10436-y.
43. Bardone, M.R.; Paternoster, M.; Coronelli, C. Teichomycins, new antibiotics from *Actinoplanes teichomyceticus* nov. sp. II. Extraction and chemical characterization. *J. Antibiot. (Tokyo).* **1978**, *31*, 170–177, doi:10.7164/antibiotics.31.170.
44. Borghi, A.; Coronelli, C.; Faniuolo, L.; Allievi, G.; Pallanza, R.; Gallo, G.G. Teichomycins, new antibiotics from *Actinoplanes teichomyceticus* nov. sp.: IV. Separation and characterization of the components of teichomycin (teicoplanin). *J. Antibiot. (Tokyo).* **1984**, *37*, 615–620, doi:10.7164/antibiotics.37.615.
45. Yim, G.; Kalan, L.; Koteva, K.; Thaker, M.N.; Waglechner, N.; Tang, I.; Wright, G.D. Harnessing the synthetic capabilities of glycopeptide antibiotic tailoring enzymes: characterization of the UK-68, 597 biosynthetic cluster. *ChemBioChem* **2014**, *15*, 2613–2623, doi:10.1002/cbic.201402179.
46. Debono, M.; Merkel, K.E.; Molloy, R.M.; Barnhart, M.; Prestí, E.; Hunt, A.H.; Hamill, R.L. Actaplanin, new glycopeptide antibiotics produced by *Actinoplanes missouriensis* the isolation and preliminary chemical characterization of actaplanin. *J. Antibiot. (Tokyo).* **1984**, *37*, 85–95, doi:10.7164/antibiotics.37.85.
47. Owen, J.G.; Reddy, B.V.B.; Ternei, M.A.; Charlop-Powers, Z.; Calle, P.Y.; Kim, J.H.; Brady, S.F. Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11797–11802, doi:10.1073/pnas.1222159110.

48. Pootoolal, J.; Thomas, M.G.; Marshall, C.G.; Neu, J.M.; Hubbard, B.K.; Walsh, C.T.; Wright, G.D. Assembling the glycopeptide antibiotic scaffold: the biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL15009. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 8962–8967, doi:10.1073/pnas.102285099.
49. Thaker, M.N.; Wang, W.; Spanogiannopoulos, P.; Waglechner, N.; King, A.M.; Medina, R.; Wright, G.D. Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat. Biotechnol.* **2013**, *31*, 922–927, doi:10.1038/nbt.2685.
50. Hong, H.J.; Hutchings, M.I.; Hill, L.M.; Buttner, M.J. The role of the novel fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*. *J. Biol. Chem.* **2005**, *280*, 13055–13061, doi:10.1074/JBC.M413801200.
51. Pathom-Aree, W.; Nogi, Y.; Sutcliffe, I.C.; Ward, A.C.; Horikoshi, K.; Bull, A.T.; Goodfellow, M. *Williamsia marianensis* sp. nov., a novel actinomycete isolated from the Mariana Trench. *Int. J. Syst. Evol. Microbiol.* **2006**, *56*, 1123–1126, doi:10.1099/ijs.0.64132-0.
52. Hegstad, K.; Mikalsen, T.; Coque, T.M.; Werner, G.; Sundsfjord, A. Mobile genetic elements and their contribution to the emergence of antimicrobial resistant *Enterococcus faecalis* and *Enterococcus faecium*. *Clin. Microbiol. Infect.* **2010**, *16*, 541–554, doi:10.1111/j.1469-0691.2010.03226.x.
53. Garnier, F.; Taourit, S.; Glaser, P.; Courvalin, P.; Galimand, M. Characterization of transposon Tn1549, conferring VanB-type resistance in *Enterococcus* spp. *Microbiology* **2000**, *146*, 1481–1489, doi:10.1099/00221287-146-6-1481.
54. Handwerger, S.; Skoble, J. Identification of chromosomal mobile element conferring high-level vancomycin resistance in *Enterococcus faecium*. *Antimicrob. Agents Chemother.* **1995**, *39*, 2446–2453, doi:10.1128/AAC.39.11.2446.
55. Reynolds, P.E.; Ambur, O.H.; Casadewall, B.; Courvalin, P. The VanYD DD-carboxypeptidase of *Enterococcus faecium* BM4339 is a penicillin-binding protein. *Microbiology* **2001**, *147*, 2571–2578, doi:10.1099/00221287-147-9-2571.
56. Hopwood, D.A.; Wright, H.M. Transformation in *Thermoactinomyces vulgaris*. *J. Gen. Microbiol.* **1972**, *71*, 383–398, doi:10.1099/00221287-71-2-383.
57. Meziane-Cherif, D.; Stogios, P.J.; Evdokimova, E.; Savchenko, A.; Courvalin, P. Structural basis for the evolution of vancomycin resistance D,D-peptidases. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 5872–5877, doi:10.1073/pnas.1402259111.
58. Rawlings, N.D.; Waller, M.; Barrett, A.J.; Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **2014**, *42*, D503–D509, doi:10.1093/nar/gkt953.
59. Charlier, P.; Wery, J.-P.; Dideberg, O.; Frère, J.-M. *Streptomyces albus* G D-Ala-D-Ala carboxypeptidase. *Handb. Met.* **2006**, doi:10.1002/0470028637.MET032.
60. Marshall, C.G.; Lessard, I.A.D.; Park, I.S.; Wright, G.D. Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. *Antimicrob. Agents Chemother.* **1998**, *42*, 2215–2220, doi:10.1128/aac.42.9.2215.
61. Bussiere, D.E.; Pratt, S.D.; Katz, L.; Severin, J.M.; Holzman, T.; Park, C.H. The structure of VanX reveals a novel amino-dipeptidase involved in mediating transposon-based vancomycin resistance. *Mol. Cell* **1998**, *2*, 75–84, doi:10.1016/S1097-2765(00)80115-X.
62. Marcone, G.L.; Beltrametti, F.; Binda, E.; Carrano, L.; Foulston, L.; Hesketh, A.; Bibb, M.; Marinelli, F. Novel mechanism of glycopeptide resistance in the A40926 producer *Nonomuraea* sp. ATCC 39727. *Antimicrob. Agents Chemother.* **2010**, *54*, 2465–2472, doi:10.1128/AAC.00106-10.
63. Kilian, R.; Frasch, H.J.; Kulik, A.; Wohlleben, W.; Stegmann, E. The VanRS homologous two-component system VnlRS$_{Ab}$ of the glycopeptide producer *Amycolatopsis balhimycina* activates transcription of the *vanHAX*$_{Sc}$ genes in *Streptomyces coelicolor*, but not in *A. balhimycina*. *Microb. Drug Resist.* **2016**, *22*, 499–509, doi:10.1089/mdr.2016.0128.
64. Mainardi, J.-L.; Villet, R.; Bugg, T.D.; Mayer, C.; Arthur, M. Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in Gram-positive bacteria. *FEMS Microbiol. Rev.* **2008**, *32*, 386–408, doi:10.1111/J.1574-6976.2007.00097.X.
65. Podmore, A.H.B.; Reynolds, P.E. Purification and characterization of VanXYc, a D,D-dipeptidase/D,D-carboxypeptidase in vancomycin-resistant *Enterococcus gallinarum* BM4174. *Eur. J. Biochem.* **2002**, *269*, 2740–2746, doi:10.1046/j.1432-1033.2002.02946.x.
66. Alduina, R.; Tocchetti, A.; Costa, S.; Ferraro, C.; Cancemi, P.; Sosio, M.; Donadio, S. A two-component regulatory system with opposite effects on glycopeptide antibiotic biosynthesis and resistance. *Sci. Rep.* **2020**, *10*, doi:10.1038/s41598-020-63257-4.
67. D'Argenio, V.; Petrillo, M.; Pasanisi, D.; Pagliarulo, C.; Colicchio, R.; Talà, A.; De Biase, M.S.; Zanfardino, M.; Scolamiero, E.; Pagliuca, C.; et al. The complete 12 Mb genome and transcriptome of *Nonomuraea gerenzanensis* with new insights into its duplicated "magic" RNA polymerase. *Sci. Rep.* **2016**, *6*, doi:10.1038/s41598-016-0025-0.
68. Yushchuk, O.; Andreo-Vidal, A.; Marcone, G.L.; Bibb, M.; Marinelli, F.; Binda, E. New molecular tools for regulation and improvement of A40926 glycopeptide antibiotic production in *Nonomuraea gerenzanensis* ATCC 39727. *Front. Microbiol.* **2020**, *11*, doi:10.3389/fmicb.2020.00008.
69. Rioseras, B.; Yagüe, P.; López-Garciá, M.T.; Gonzalez-Quiñõnez, N.; Binda, E.; Marinelli, F.; Manteca, A. Characterization of SCO4439, a D-alanyl-D-alanine carboxypeptidase involved in spore cell wall maturation, resistance, and germination in *Streptomyces coelicolor*. *Sci. Rep.* **2016,** *6*, doi:10.1038/srep21659.
70. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649, doi:10.1093/bioinformatics/bts199.
71. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549, doi:10.1093/molbev/msy096.
72. Sievers, F.; Higgins, D.G. Clustal Omega. *Curr. Protoc. Bioinforma.* **2014**, *2014*, 3.13.1-3.13.16, doi:10.1002/0471250953.bi0313s48.

73. Medema, M.H.; Takano, E.; Breitling, R. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* **2013**, *30*, 1218–1223, doi:10.1093/molbev/mst025.
74. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410, doi:10.1016/S0022-2836(05)80360-2.
75. Li, T.L.; Huang, F.; Haydock, S.F.; Mironenko, T.; Leadlay, P.F.; Spencer, J.B. Biosynthetic gene cluster of the glycopeptide antibiotic teicoplanin: characterization of two glycosyltransferases and the key acyltransferase. *Chem. Biol.* **2004**, *11*, 107–119, doi:10.1016/S1074-5521(04)00002-X.
76. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S.Y.; Medema, M.H.; Weber, T. AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019**, *47*, W81–W87, doi:10.1093/nar/gkz310.

Electronic supplementary materials

Genomic insights into the distribution and phylogeny of glycopeptide resistance determinants within the Actinobacteria phylum

Andres Andreo-Vidal [1], Elisa Binda [1], Victor Fedorenko [2], Flavia Marinelli [1]* and Oleksandr Yushchuk [1,2]

[1] Department of Biotechnology and Life Sciences, University of Insubria, Varese, Italy.

[2] Department of Genetics and Biotechnology, Ivan Franko National University of Lviv, Lviv, Ukraine.

*Correspondence: Flavia Marinelli, flavia.marinelli@uninsubria.it

Inventory of Electronic supplementary materials

ESM Table 1. Abbreviations of Actinobacteria and *Bacillales* genera names, used throughout the work.

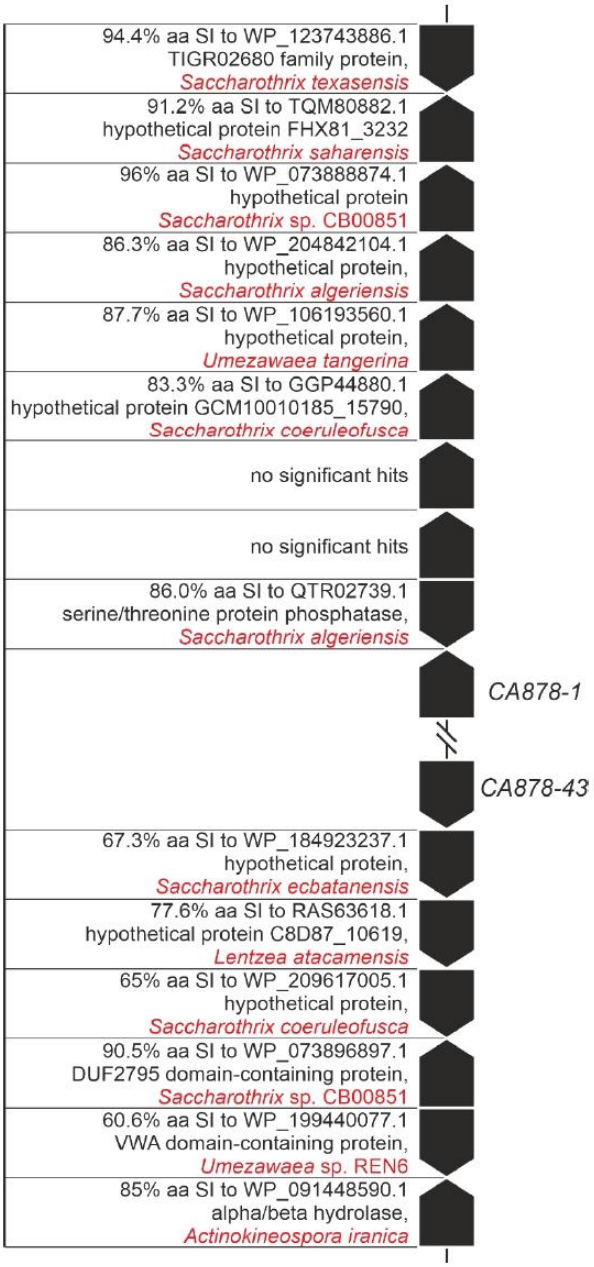
Full genus name:	Abbreviation used in the work:
Agromyces.	*A.*
Actinocrispum	*Ac.*
Aeromicrobium	*Aer.*
Actinokineospora	*Ak.*
Alloactinosynnema	*Al.*
Allonocardiopsis	*All.*
Amycolatopsis	*Am.*
Actinomadura	*Amd.*
Actinophytocola	*Ap.*
Actinoplanes	*Apl.*
Actinopolymorpha	*Apm.*
Atopobium.	*Atp.*
Bacillus	*Bac.*
Blastococcus	*Bc.*
Brevibacillus	*Bbac.*
Baekduia	*Bd.*
cand. Planktophila	*c. Pp.*
Catellatospora	*Cls.*
Diaminobutiricimonas	*Dab.*
Enterorhabdus	*Er.*
Frankia	*F.*
Gordonia	*G.*
Glacibacter	*Gb.*
Geodermathophilus	*Gdp.*
Glycomyces	*Gm.*
Herbihabitans	*H.*
Haloactinospora	*Ha.*
Herbidospora	*Hs.*
Jiangella	*Ja.*
Jishengella	*Js.*
Kutzneria	*K.*
Kibdelosporangium	*Kib.*
Kribbella	*Kr.*
Kitasatospora	*Ksp.*
Lentzea	*L.*
Leifsonia	*Ls.*
Microbispora	*M.*
Mycobacterium	*Mb.*
Mycolicibacterium	*Mcb.*
Murinocardiopsis	*Mcp.*
Modestobacter	*Mdb.*
Micromonospora.	*Mms.*
Nonomuraea	*N.*
Nocardia	*Ncd.*
Nitriliruptoraceae bacterium	*NiB*
Nakamurella	*Nm.*
Nocardiopsis	*Np.*

4

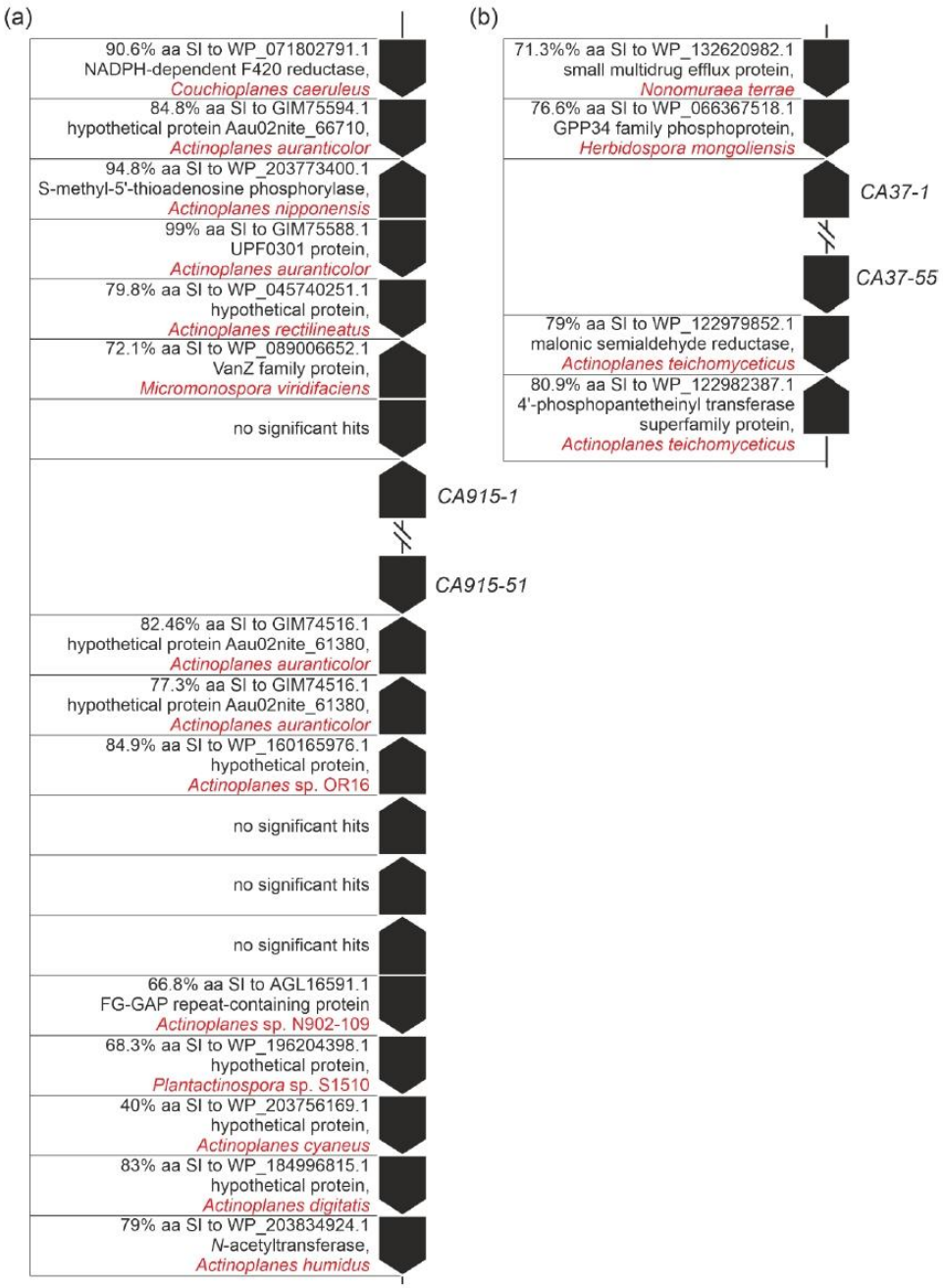
Full genus name:	Abbreviation used in the work:
Paenibacillus	*Pnb.*
Pseudokineococcus	*Pkc.*
Promicromonospora	*Pms.*
Prauserella	*Pr.*
Pseudonocardia	*Ps.*
Parvibacter	*Pbv.*
Rathayibacter	*Rb.*
Rhodococcus	*Rc.*
Stackebrandtia	*Sb.*
Streptomonospora	*Sms.*
Saccharopolyspora	*Sp.*
Streptosporangium	*Ss.*
Saccharothtix	*St.*
Streptomyces	*Str.*
Tamaricihabitans	*T.*
Thermoactinomyces	*Tam.*
Tsukamurella	*Tm.*
Tomitella	*Tt.*
Verrucosispora	*Vcs.*
Williamsia	*W.*
Xiangella	*Xa.*
Cryptosporangium	*Cs.*

ESM Table 2. List of transposases from known MGEs carrying *vlgs* used for comparing with the transposases encoded within *Atopobium minutum* 10063974, *Enterorhabdus mucosicola* NM66_B29 and *Parvibacter caecicola* DSM 22242 MGEs.

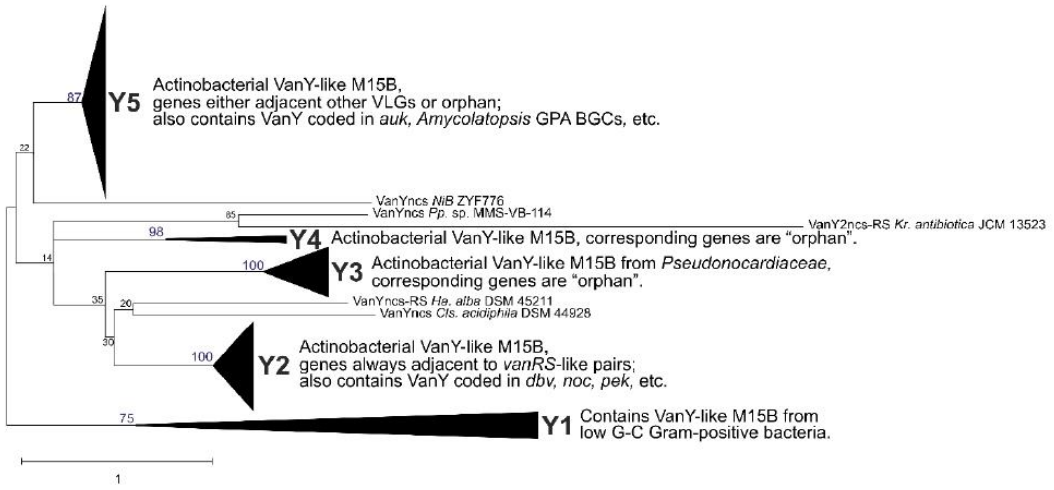
Protein ID:	Name:	Source:	Reference:
AAQ17155	Tn1546 transposase	*Staphylococcus aureus* plasmid pLW043	[1]
AAQ17125	IS431mec transposase	*Staphylococcus aureus* plasmid pLW043	[1]
AAQ17171	Tn552 transposase	*Staphylococcus aureus* plasmid pLW043	[1]
AAB42161	Tn5506 IS1252 transposase	*Enterococcus faecium* plasmid pHKK701	[2]
RNL10277	site-specific integrase	*Parvibacter caecicola* DSM 22242	-
AAB00677	Tn5482 IS3-like transposase	*Enterococcus faecium* insertion sequence IS1216V	[3]
AAB00676	Tn5482 ISS1 homolog transposase	*Enterococcus faecium* insertion sequence IS1216V	[3]
AAO83056	IS256 transposase	*Enterococcus faecalis* V583 plasmid pTEF1	[4]
AAO83057	IS1216 transposase	*Enterococcus faecalis* V583 plasmid pTEF1	[4]
AAF72368	Tn1549 integrase	*Enterococcus faecalis* transposon Tn1549	[5]
AAL27448	Tn6202 Int410 integrase	*Enterococcus faecalis* N00-0410 transposon Tn6202	[6]
AAC44460	Tn1547 transposase	*Enterococcus faecalis* insertion sequence IS16, Tn1547	[7]
EMZ42128	site-specific recombinase/integrase	*Atopobium minutum* 10063974	-
MVX60893	tyrosine-type recombinase/integrase	*Enterorhabdus mucosicola* NM66_B29	-



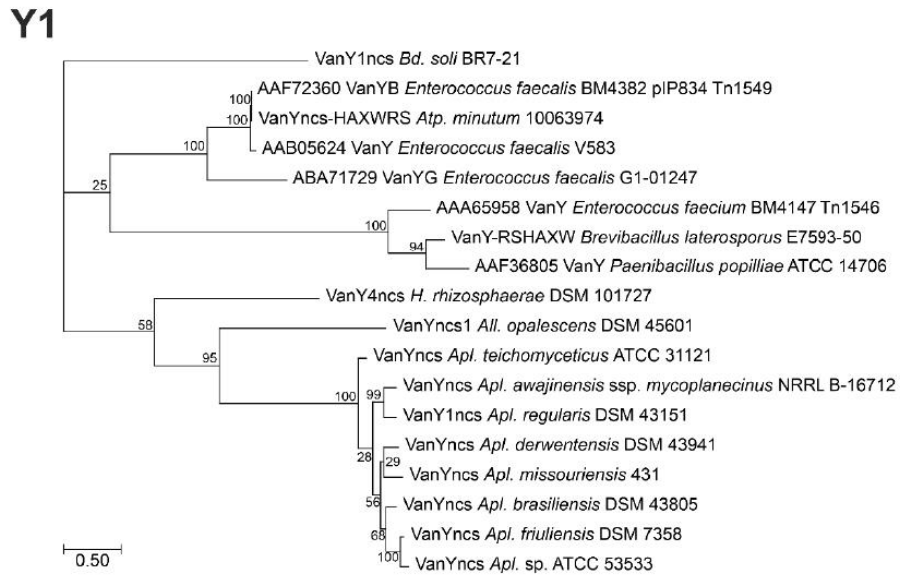
ESM Figure 1. Defining the genes annotated within previously unannotated regions flanking the CA878 GPA BGC (HM486075, genes *CA878-1* to *CA878-43*), coming from metagenomic sample. Majority of newly annotated genes encode proteins which have homologues with highest amino acid sequence identity percentages (aa SI) in *Saccharothrix* spp.



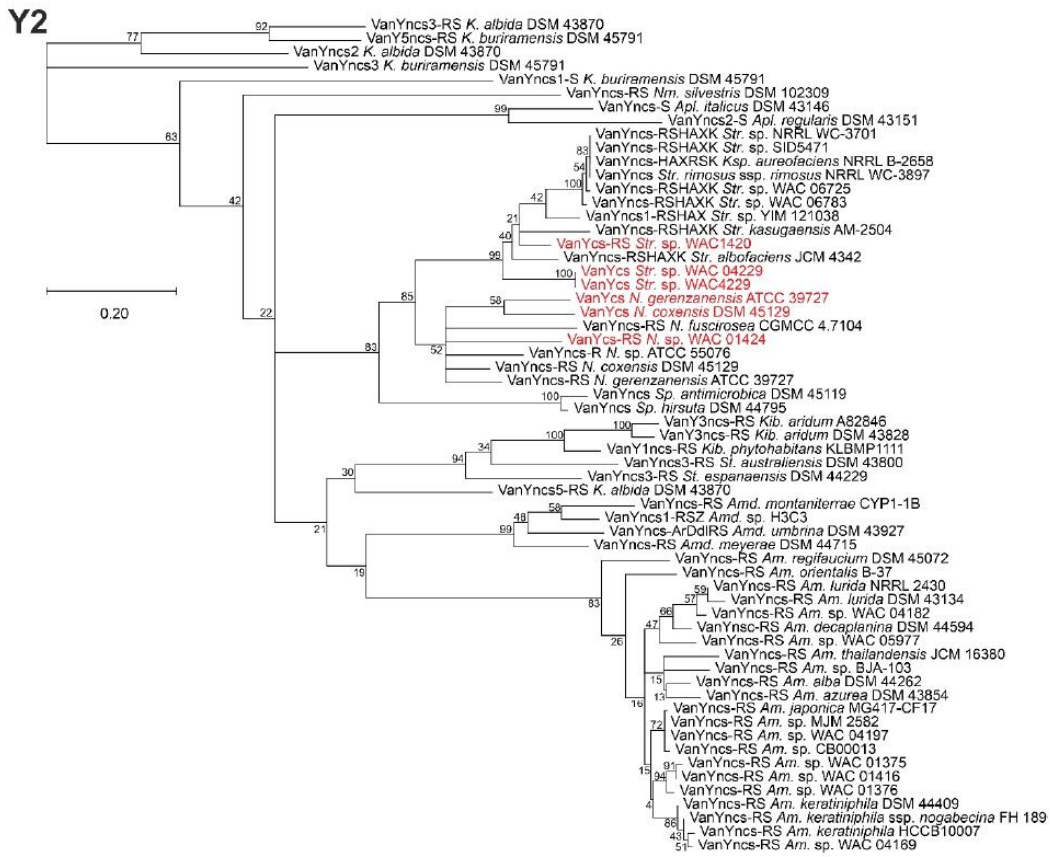
ESM Figure 2. Defining the genes annotated within previously unannotated flanks of (**a**) CA915 (HM486076, genes from *CA915-1* to *CA915-51*) and (**b**) CA37 (HM486074, genes from *CA37-1* to *CA37-55*) GPA BGCs, coming from metagenomic samples. Majority of newly annotated genes encode proteins which have homologues with highest amino acid sequence identity percentages (aa SI) in *Actinoplanes* spp.



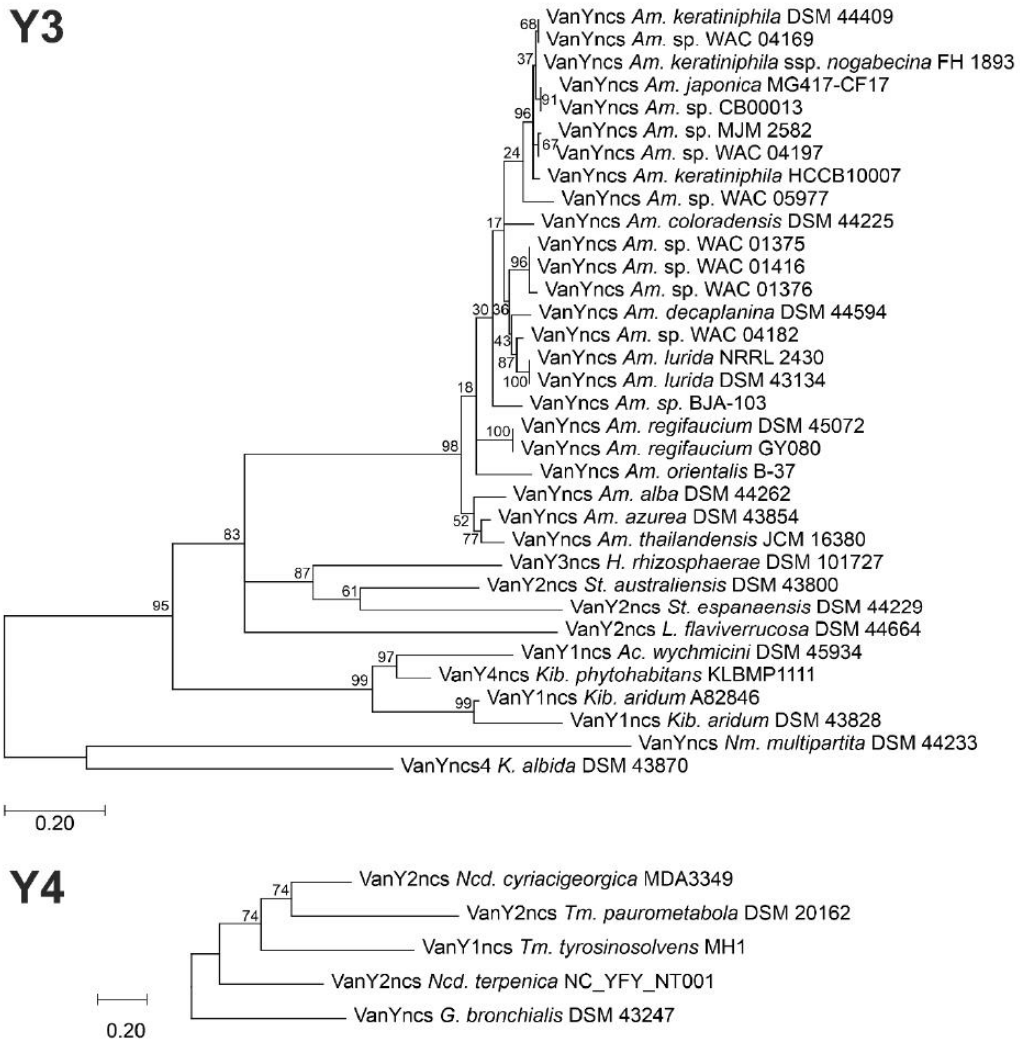
ESM Figure 3. Phylogenetic tree showing the overall phylogeny of VanY-like M15B carboxypeptidase dataset. Five main clusters (Y1-Y5) were collapsed. Expanded clusters are given separately: Y1 in ESM Figure 4; Y2 in ESM Figure 5; Y3 and Y4 in ESM Figure 6; Y5 in ESM Figure 7. Phylogenetic tree was constructed as described in Methods section. "cs/ncs" abbreviations in the label at the tip of each branch here and further means "cluster-situated/non-cluster-situated". Scale bar represents number of substitutions per site.



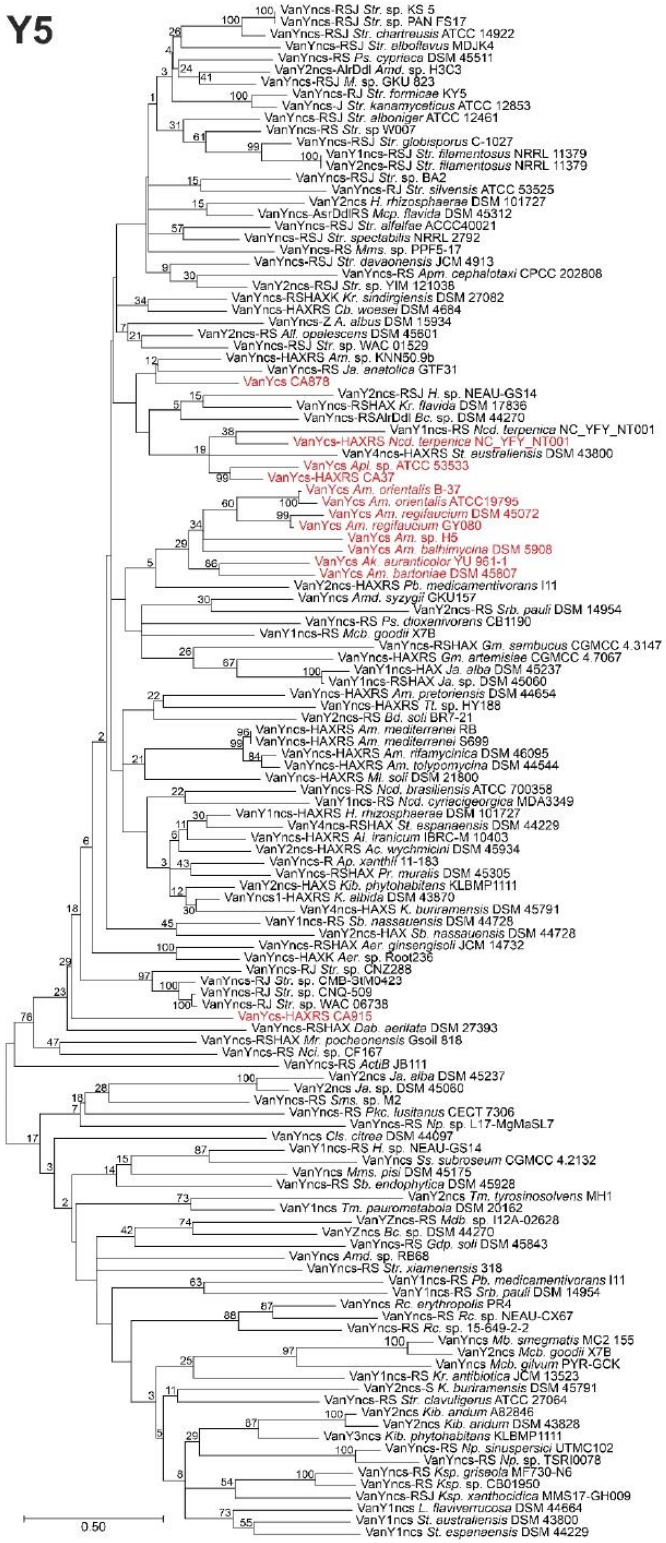
ESM Figure 4. Expanded clade corresponding to the Y1 cluster from ESM Figure 3. Scale bar represents number of substitutions per site.



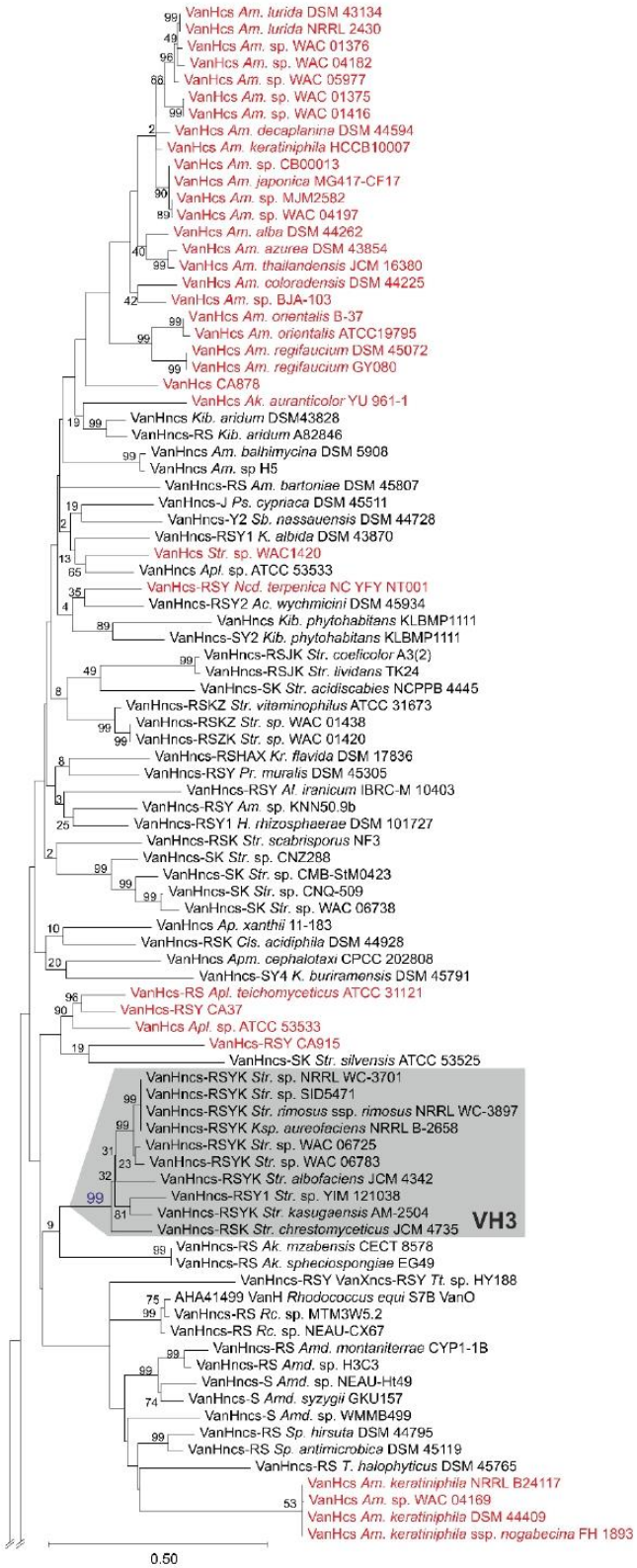
ESM Figure 5. Expanded clade corresponding to the Y2 cluster from ESM Figure 3. BGC-encoded proteins are given in red. Scale bar represents number of substitutions per site.



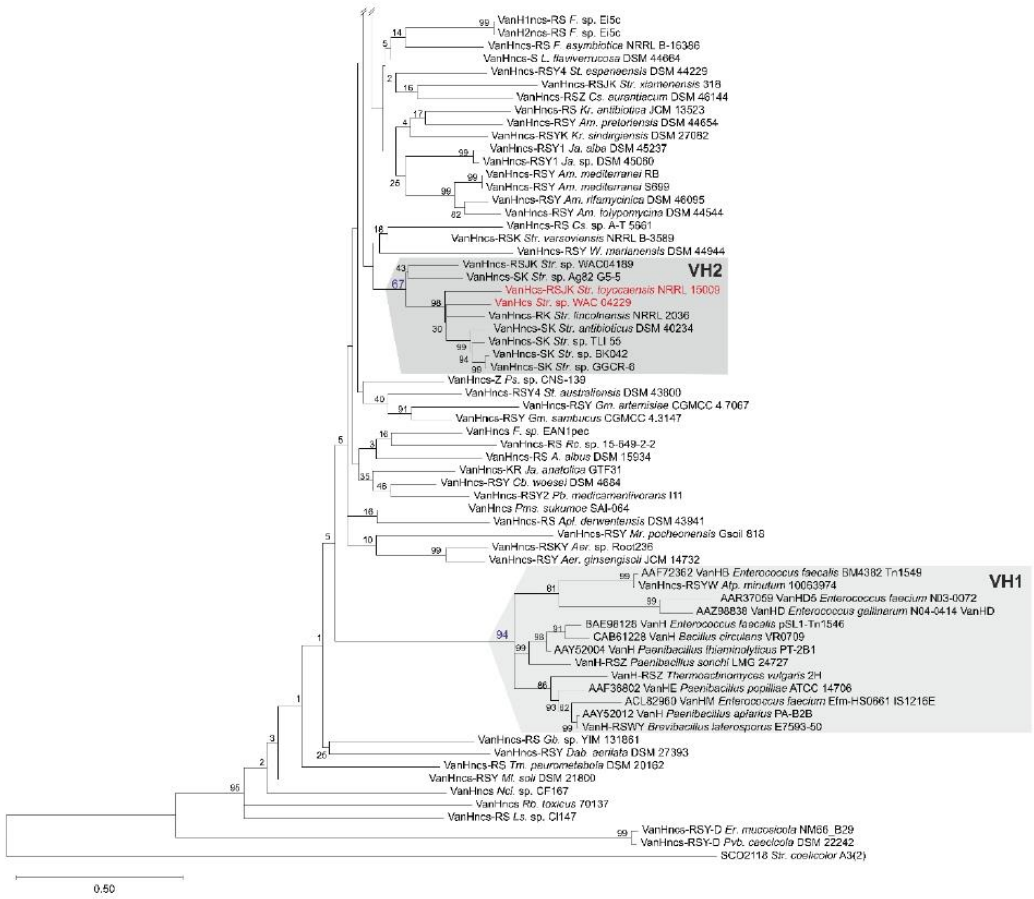
ESM Figure 6. Expanded clades corresponding to the Y3 and Y4 clusters from ESM Figure 3. Scale bar represents number of substitutions per site.



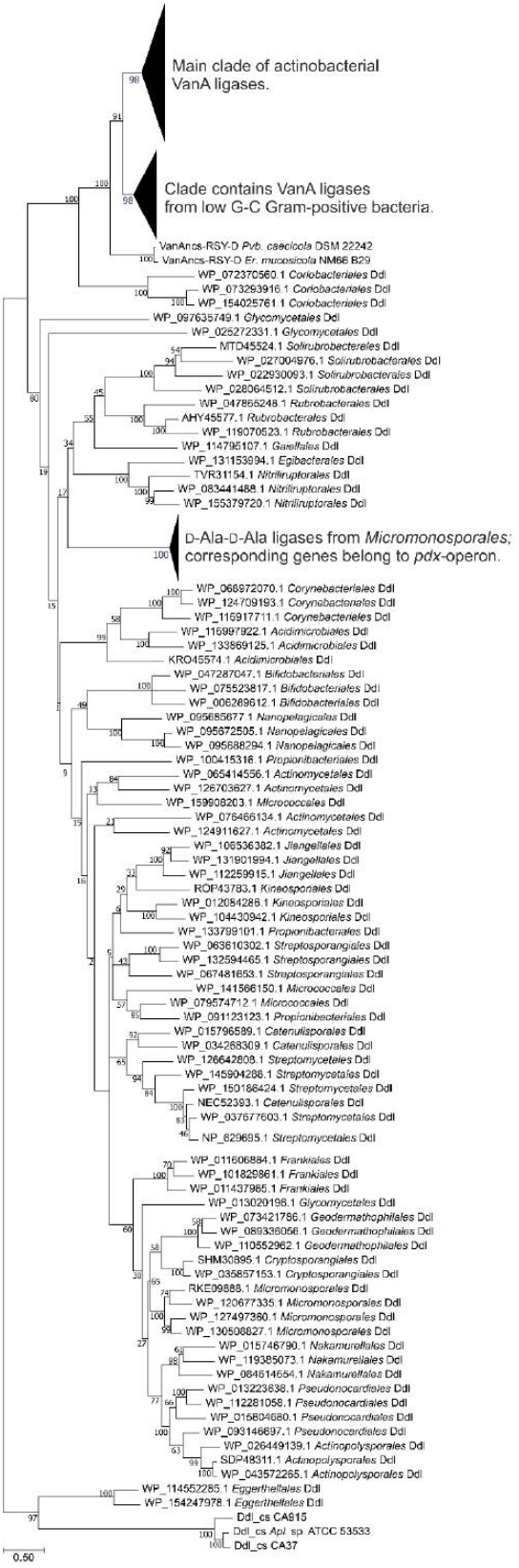
ESM Figure 7. Expanded clade corresponding to the Y5 cluster from ESM Figure 3. BGC-encoded proteins are given in red. Scale bar represents number of substitutions per site.



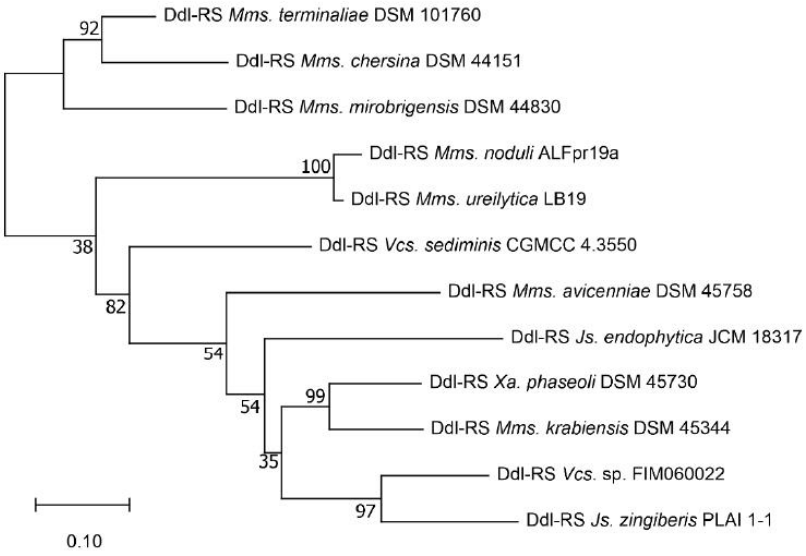
ESM Figure 8. Continued on the next page.



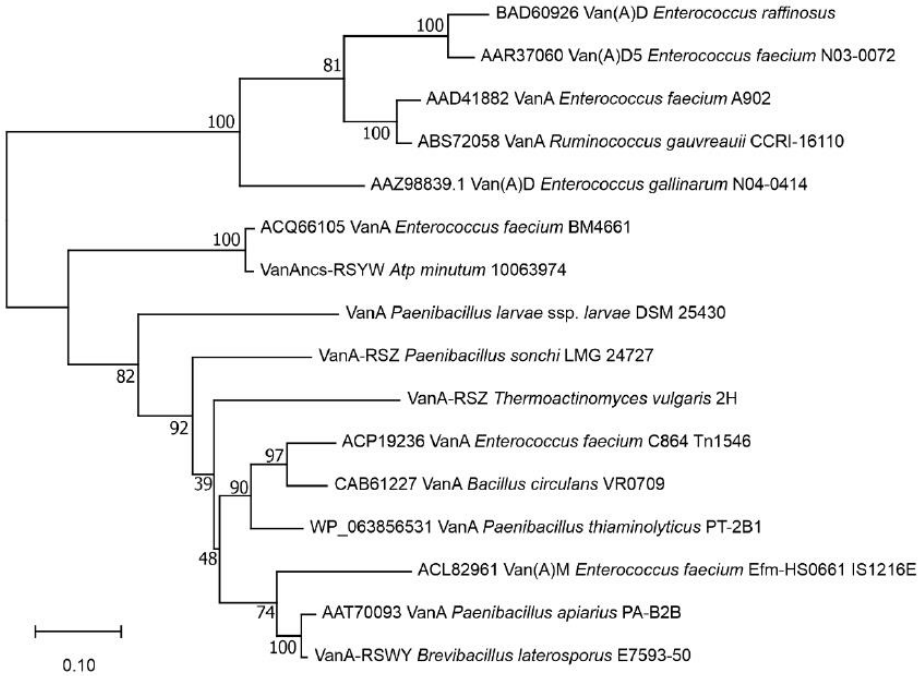
ESM Figure 8. Phylogenetic tree showing the overall phylogeny of VanH-dataset. Three well-supported clusters (VH1-3) are highlighted. Phylogenetic tree was constructed as described in Methods section, SCO2118 – a putative D-lactate dehydrogenase – was used as an outgroup. BGC-encoded proteins are given in red. Scale bar represents number of substitutions per site.



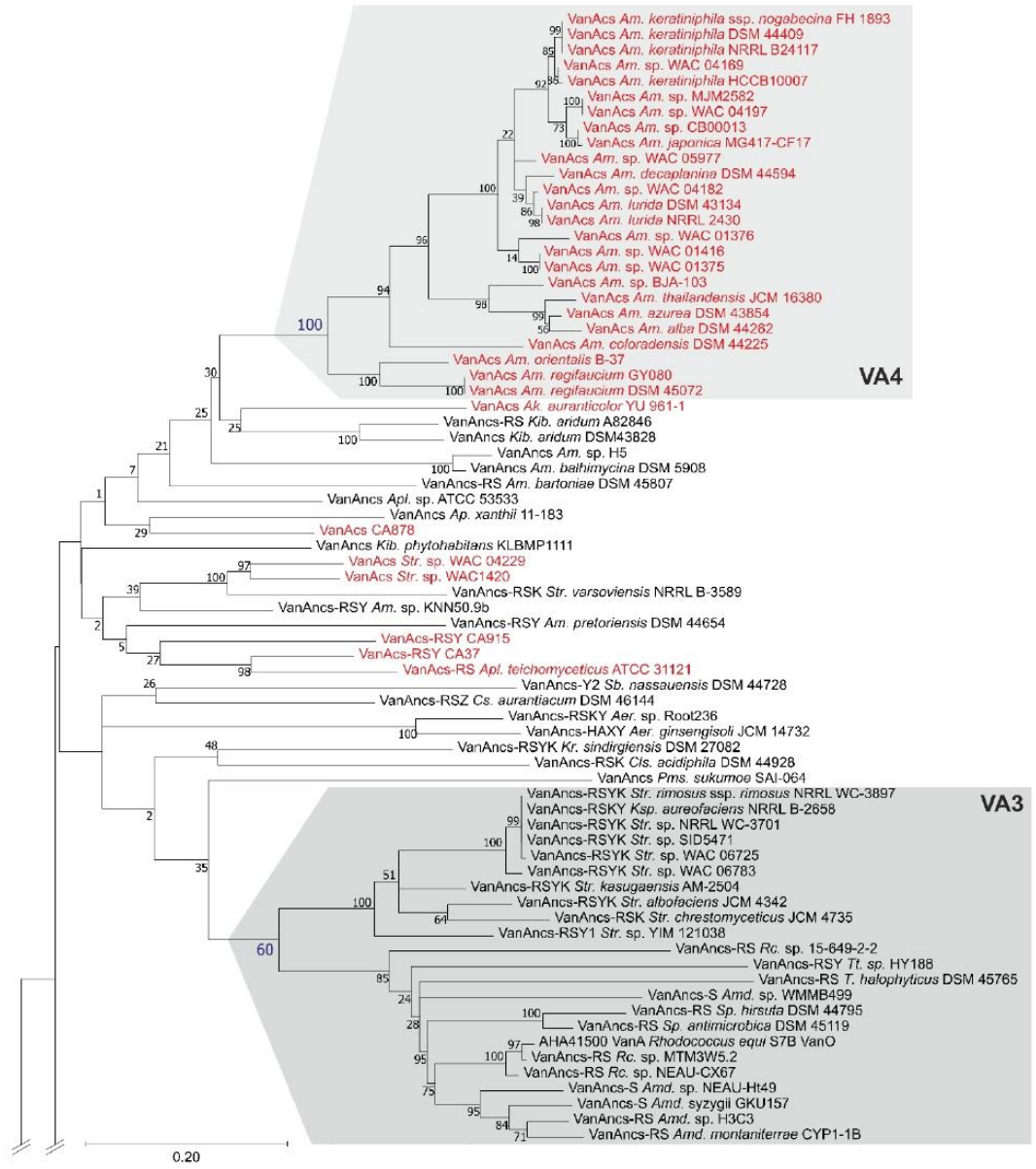
ESM Figure 9. Phylogenetic tree showing the overall phylogeny of the dataset composed with VanA proteins, Ddl-ligases coded in *Micromonosporales* putative *pdx*-operons, *auk,* CA915 and CA37 BGCs as well as actinobacterial "house-keeping" Ddl-ligases. Clades containing Ddl-ligases coded in putative *pdx*-operons and VanA-ligases from low G-C Gram-positive bacteria are collapsed; for these clades expansion please refer to ESM Figures 10 and 11. Expanded crown group of the tree (main clade of actinobacterial VanA-ligases) is given in ESM Figure 12. Phylogenetic tree was constructed as described in Methods section. Scale bar represents number of substitutions per site.



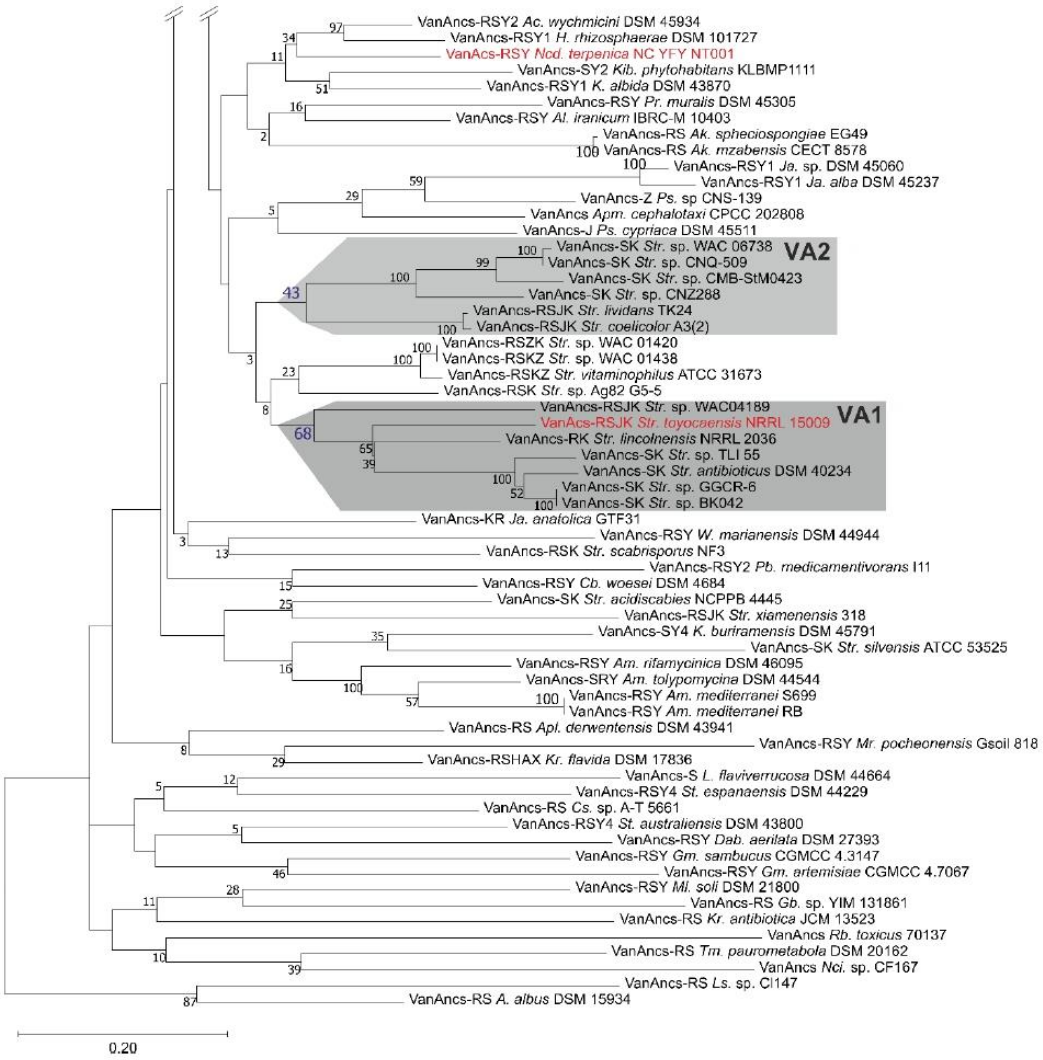
ESM Figure 10. Expanded clade of Ddl-ligases coded in putative *pdx*-operons, which was collapsed on the ESM Figure 9. Scale bar represents number of substitutions per site.



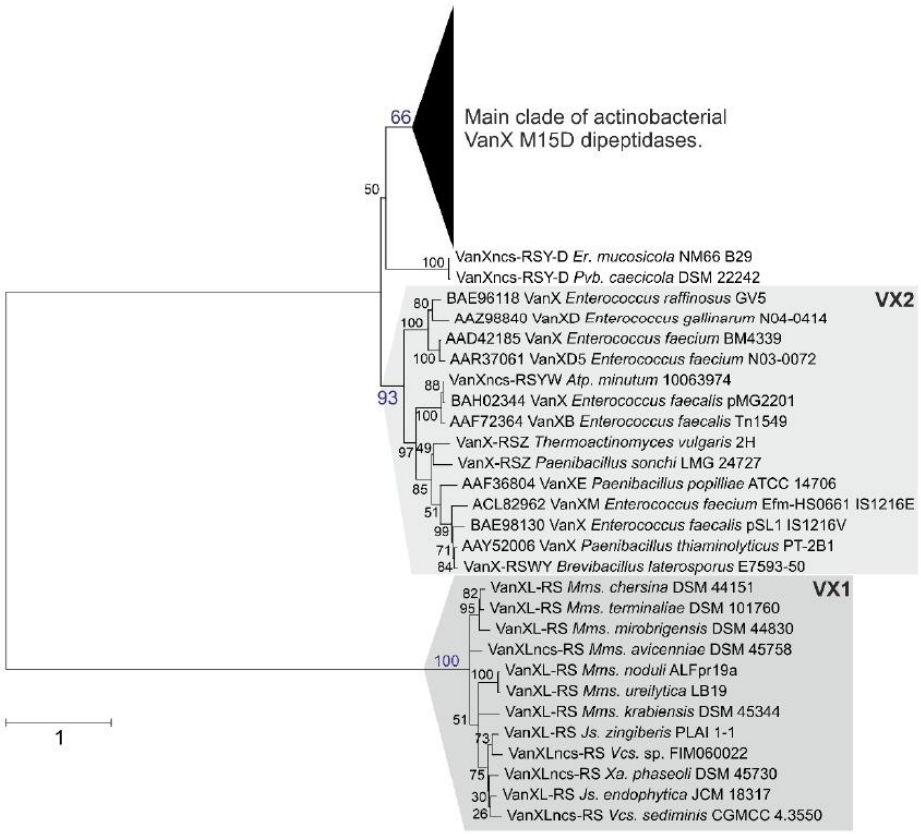
ESM Figure 11. Expanded clade containing VanA-ligases from low G-C Gram-positive bacteria, which was collapsed on the ESM Figure 9. Scale bar represents number of substitutions per site.



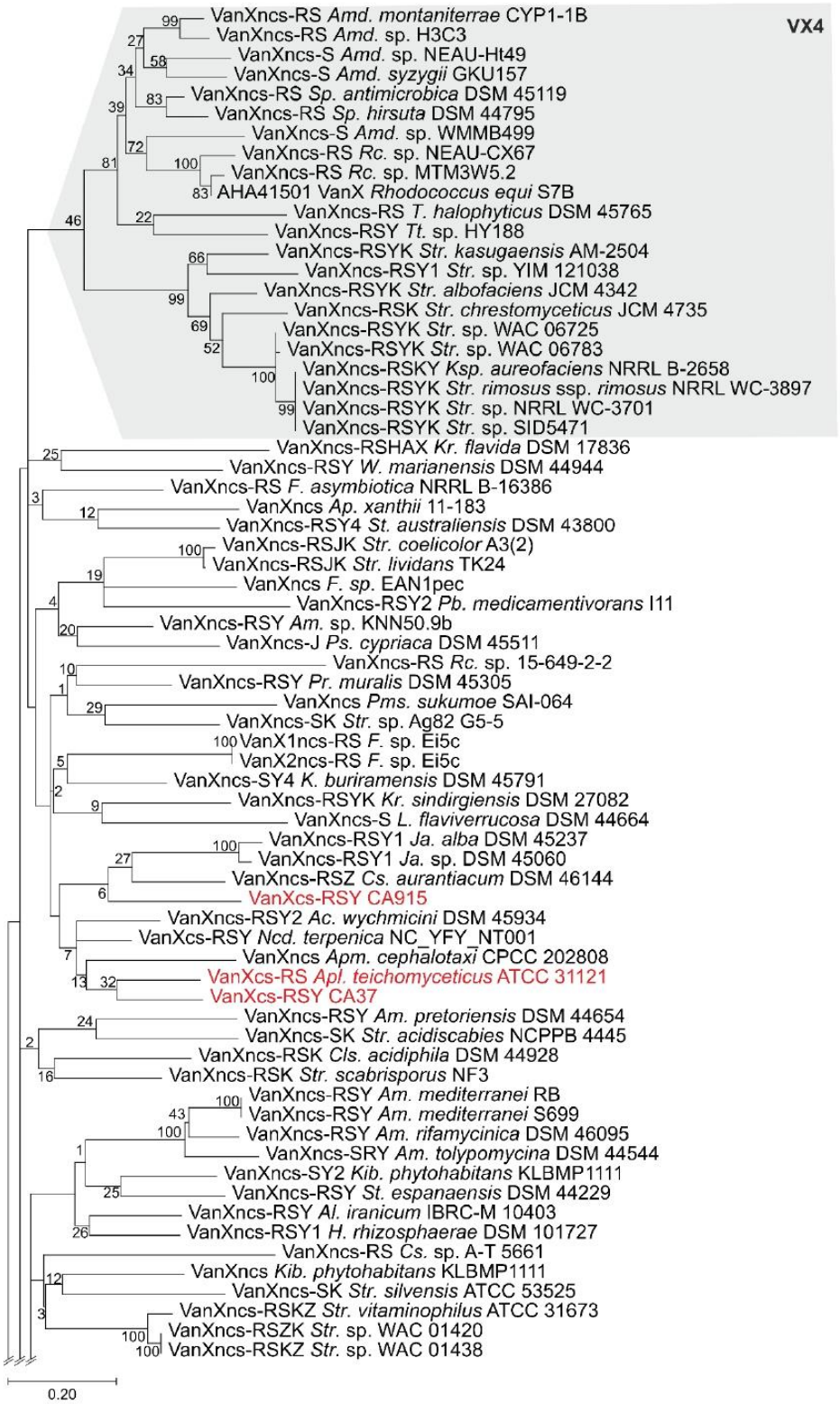
ESM Figure 12. Continued on the next page.



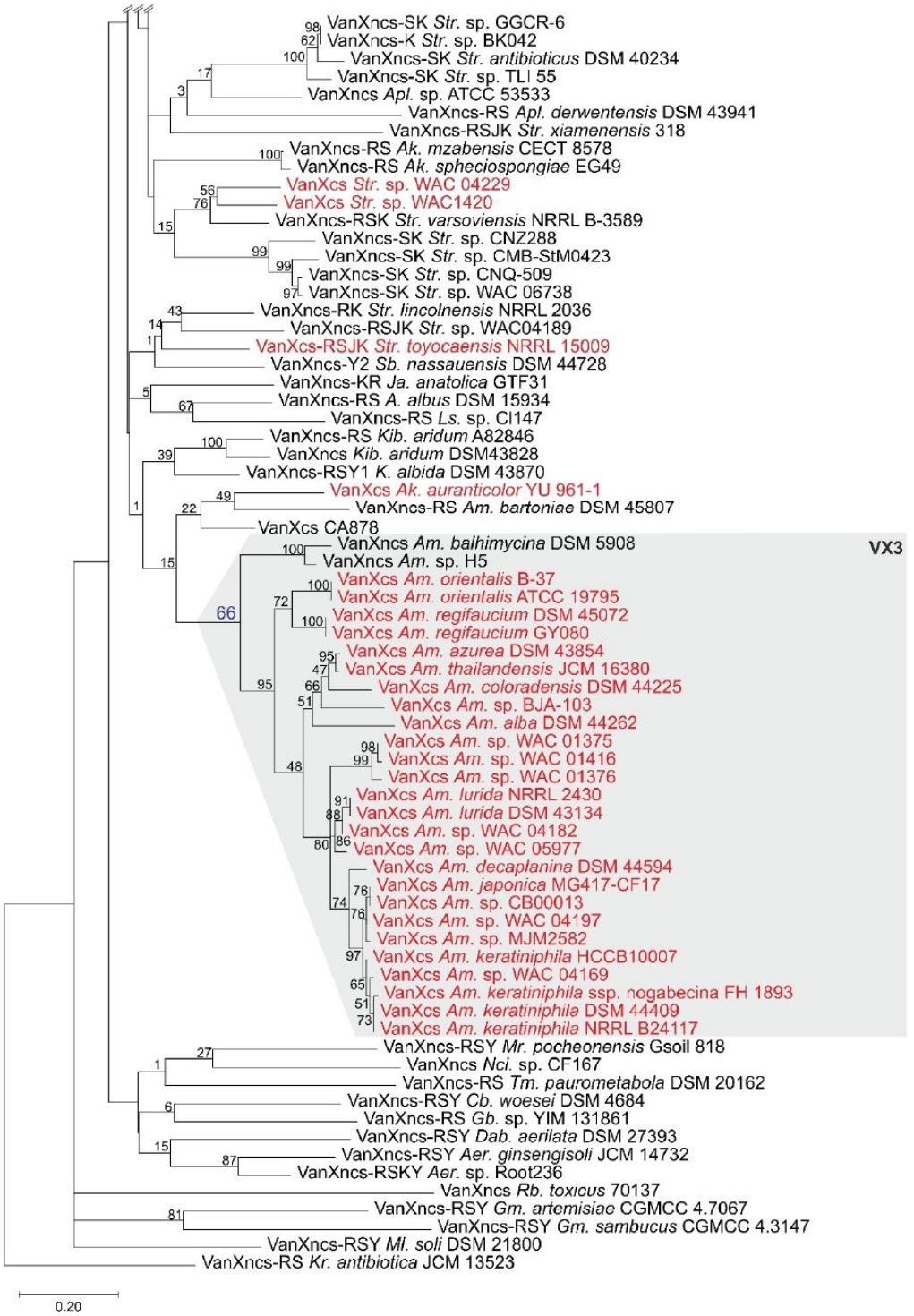
ESM Figure 12. Expanded crown group of VanA-phylogenetic tree, which was collapsed on the ESM Figure 9. Four clusters with the best bootstrap support were distinguished – VA1-4. BGC-encoded proteins are given in red. Please see main text for more details. Scale bar represents number of substitutions per site.



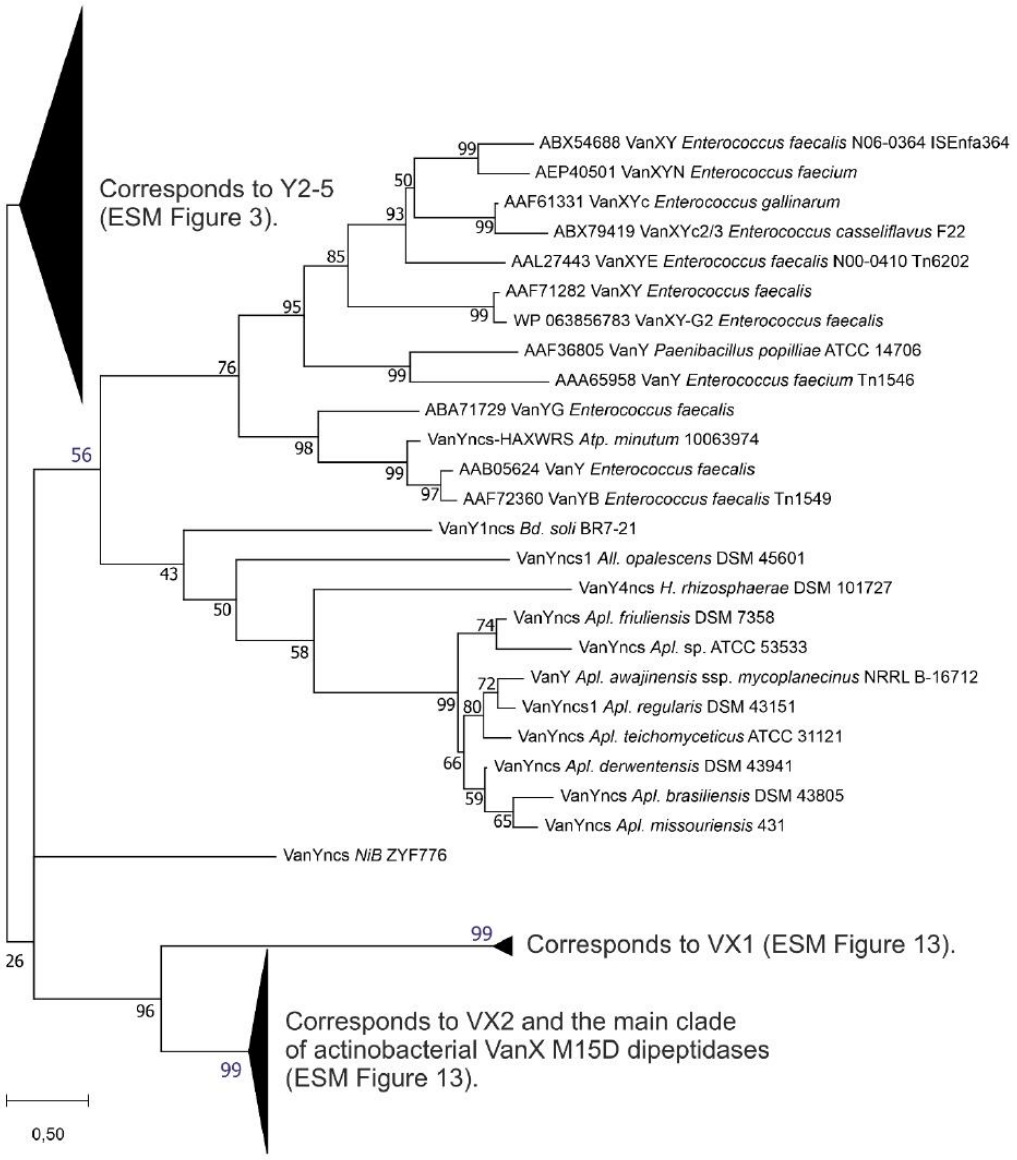
ESM Figure 13. Phylogenetic tree showing the overall phylogeny of the dataset composed with VanX-like proteins, including the ones coded in *Micromonosporales* putative *pdx*-operons. Tree contains well-separated clusters containing VanX-like proteins coded in putative *pdx*-operons (VX1) and VanX proteins from low G-C Gram-positive bacteria (VX2). The crown group of the tree (main clade of actinobacterial VanX M15D dipeptidases) is collapsed; for the expanded version please see ESM Figure 14. Phylogenetic tree was constructed as described in Methods section. Scale bar represents number of substitutions per site.



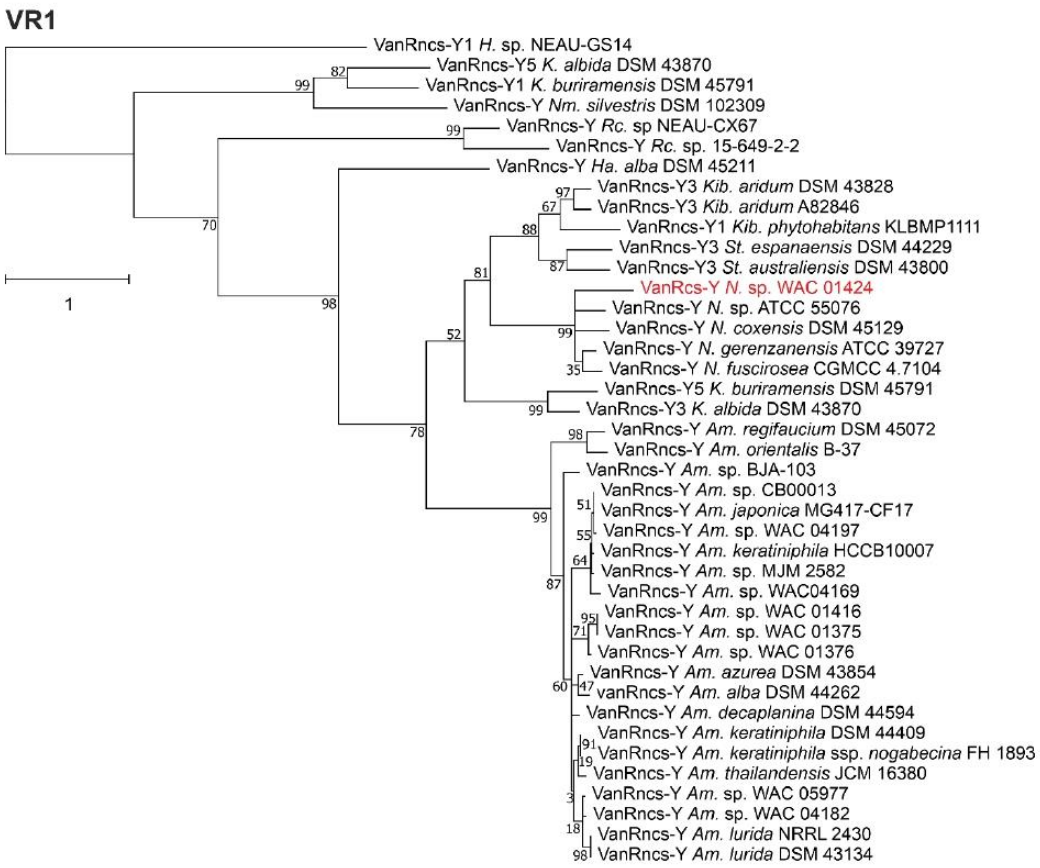
ESM Figure 14. Continued on the next page.



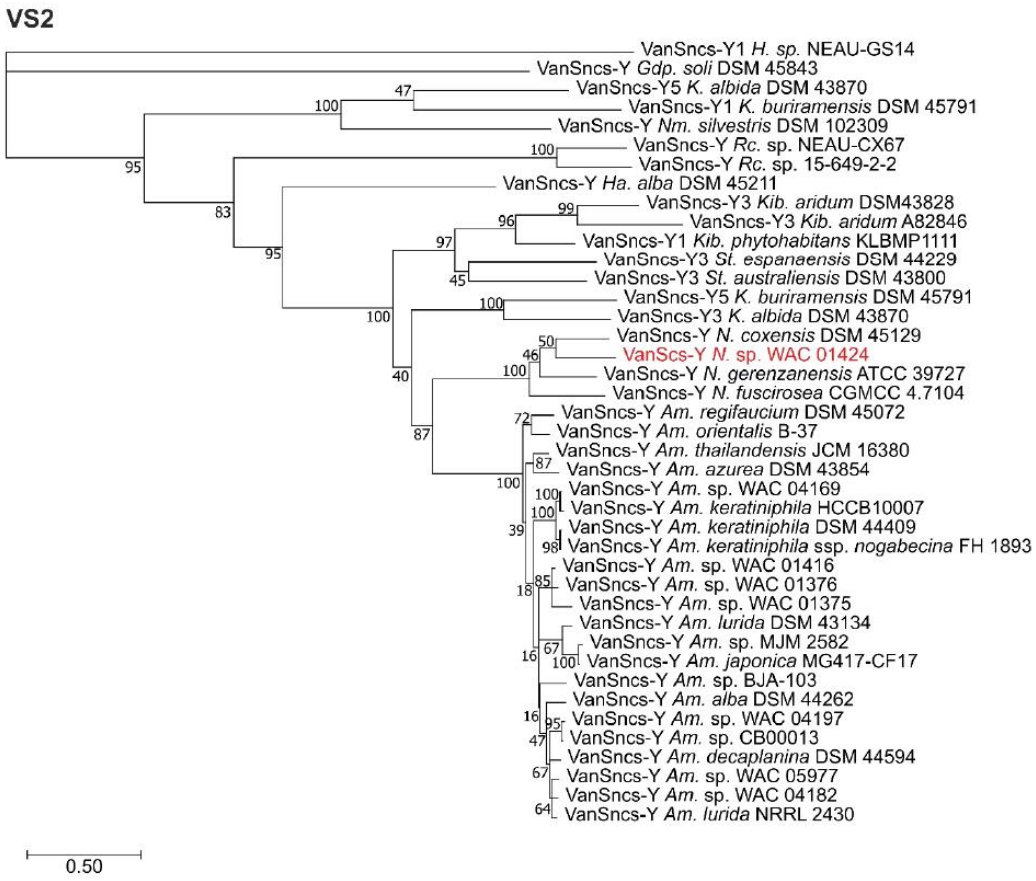
ESM Figure 14. Expanded version of the crown group of actinobacterial VanX M15D dipeptidases from ESM Figure 13. Please refer to the main text for more details. BGC-encoded proteins are given in red. Scale bar represents number of substitutions per site.



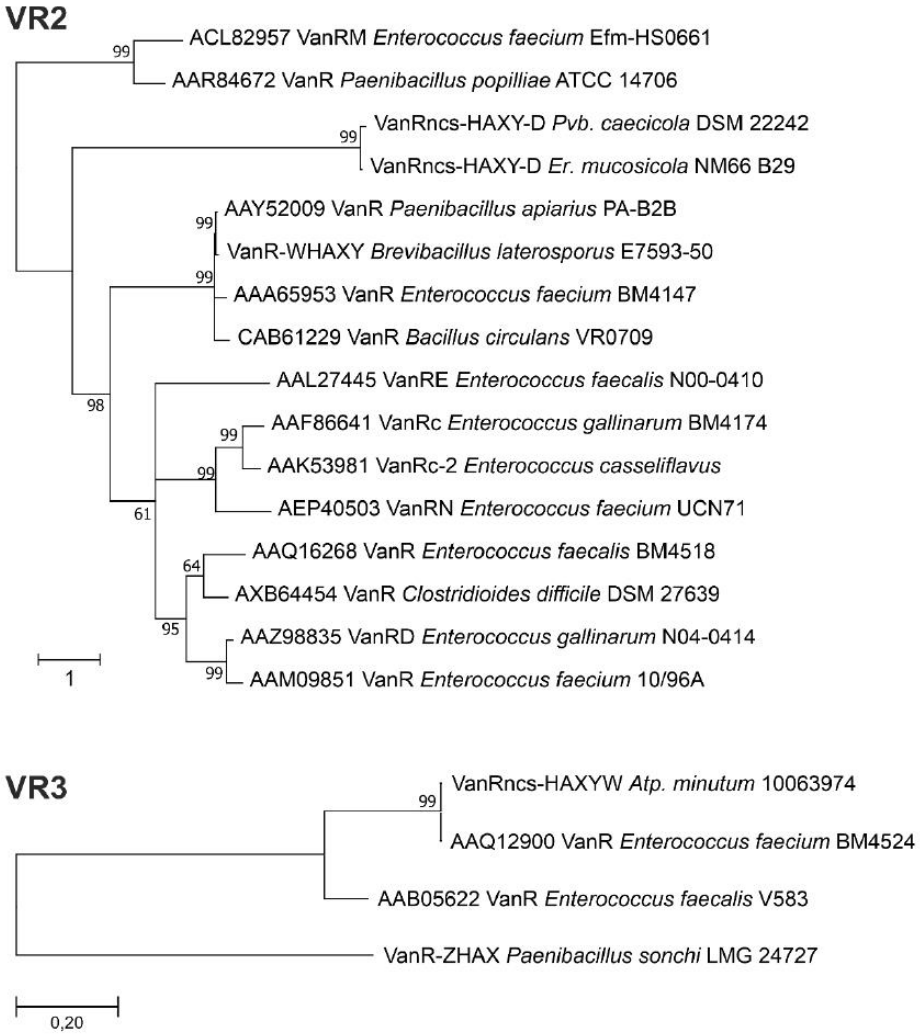
ESM Figure 15. Phylogeny of combined datasets for VanY-like M15B carboxypeptidases and VanX M15D dipeptidases, including also VanXY-proteins. Phylogenetic tree was constructed as described in Methods section. Scale bar represents number of substitutions per site.



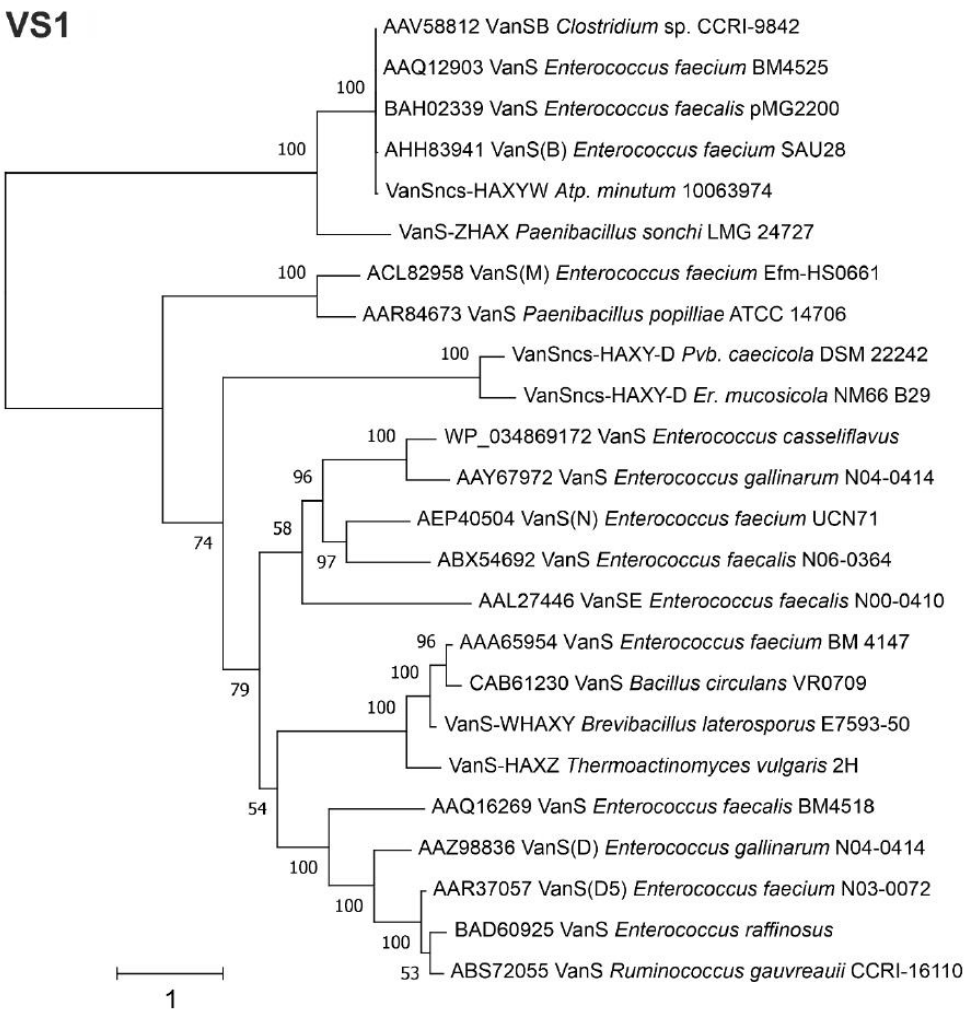
ESM Figure 16. Expanded version of the VR1 cluster from the phylogenetic tree of actinobacterial VanR-like response regulators (Figure 11a). Please refer to the main text for more details. BGC-encoded proteins are given in red. Scale bar represents number of substitutions per site.



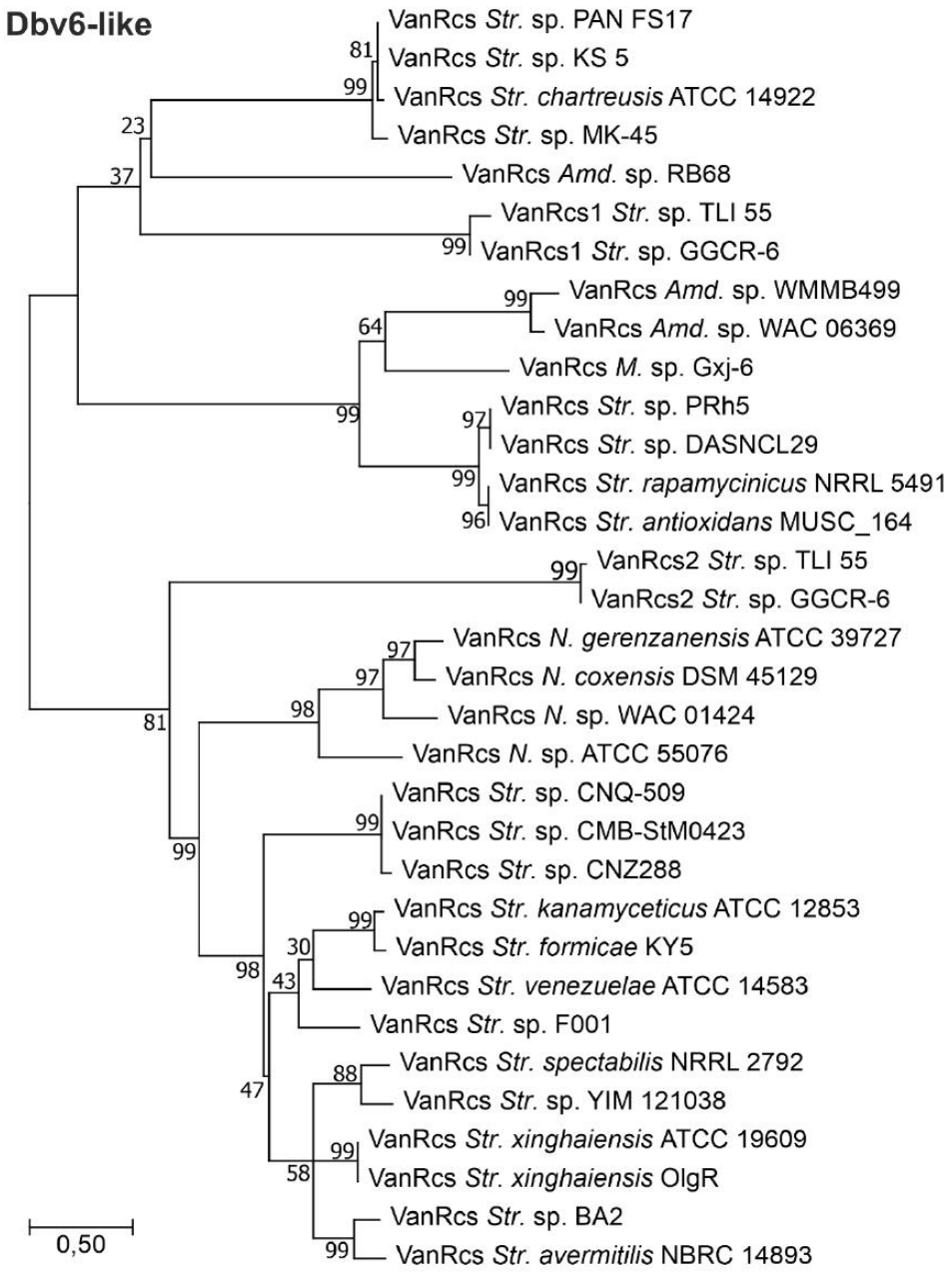
ESM Figure 17. Expanded version of the VS2 cluster from the phylogenetic tree of actinobacterial VanS-like sensor histidine kinases (Figure 11b). Please refer to the main text for more details. BGC-encoded proteins are given in red. Scale bar represents number of substitutions per site.



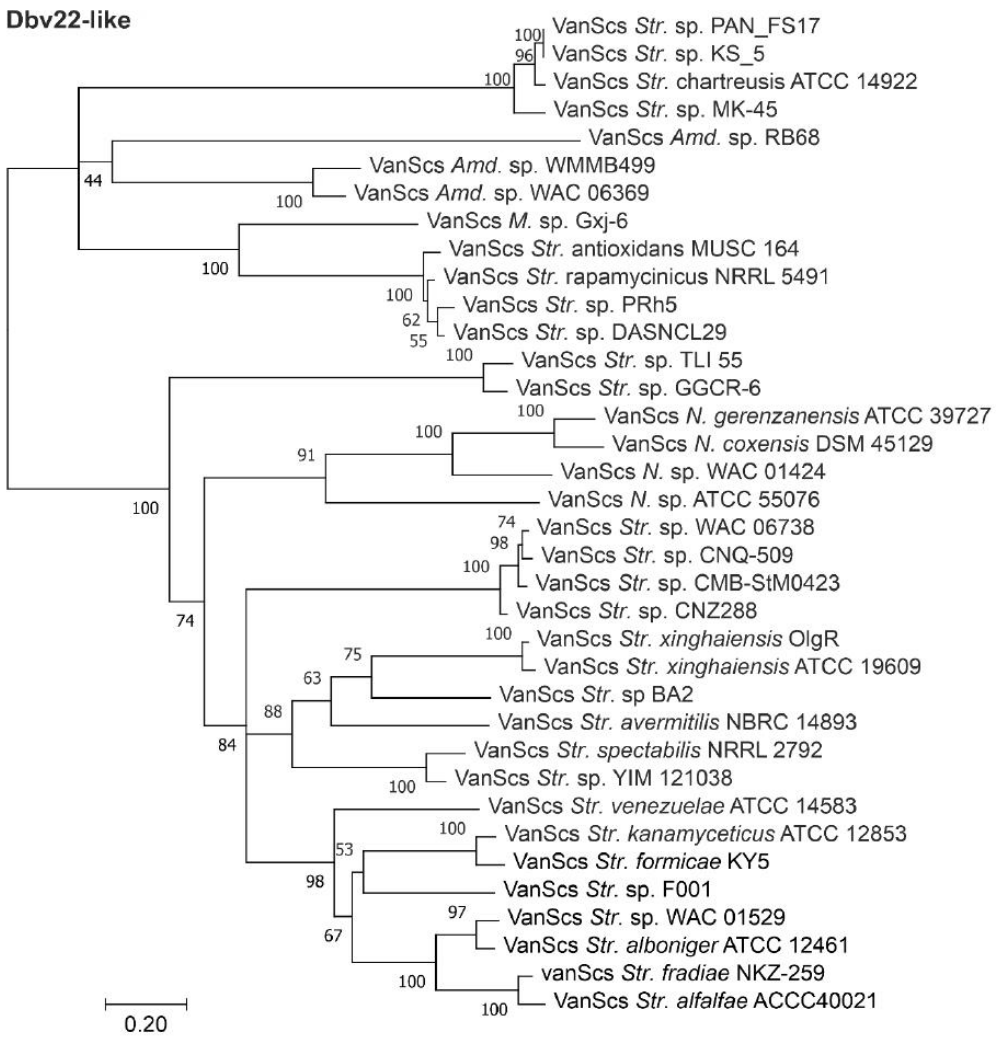
ESM Figure 18. Expanded versions of the VR2 and VR3 clusters from the phylogenetic tree of actinobacterial VanR-like response regulators (Figure 11a). Please refer to the main text for more details. Scale bar represents number of substitutions per site.



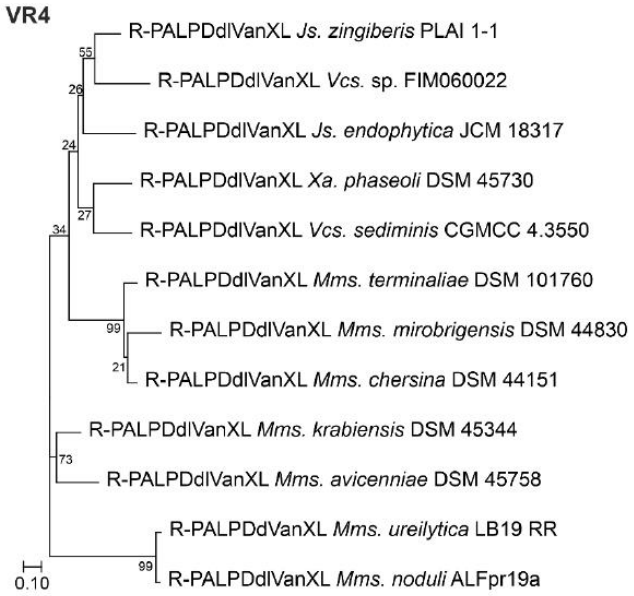
ESM Figure 19. Expanded version of the VS1 cluster from the phylogenetic tree of actinobacterial VanS-like sensor histidine kinases (Figure 11b). Please refer to the main text for more details. Scale bar represents number of substitutions per site.



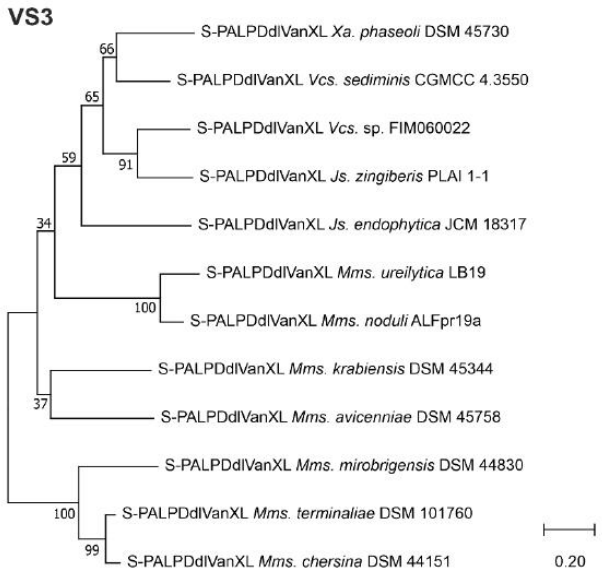
ESM Figure 20. Expanded version of the Dbv6-like cluster from the phylogenetic tree of actinobacterial VanR-like response regulators, coded within type V GPA BGCs, *Nonomuraea* spp. type IV GPA BGCs, and *feg*-like BGCs (Figure 11a). Please refer to the main text for more details. Scale bar represents number of substitutions per site.



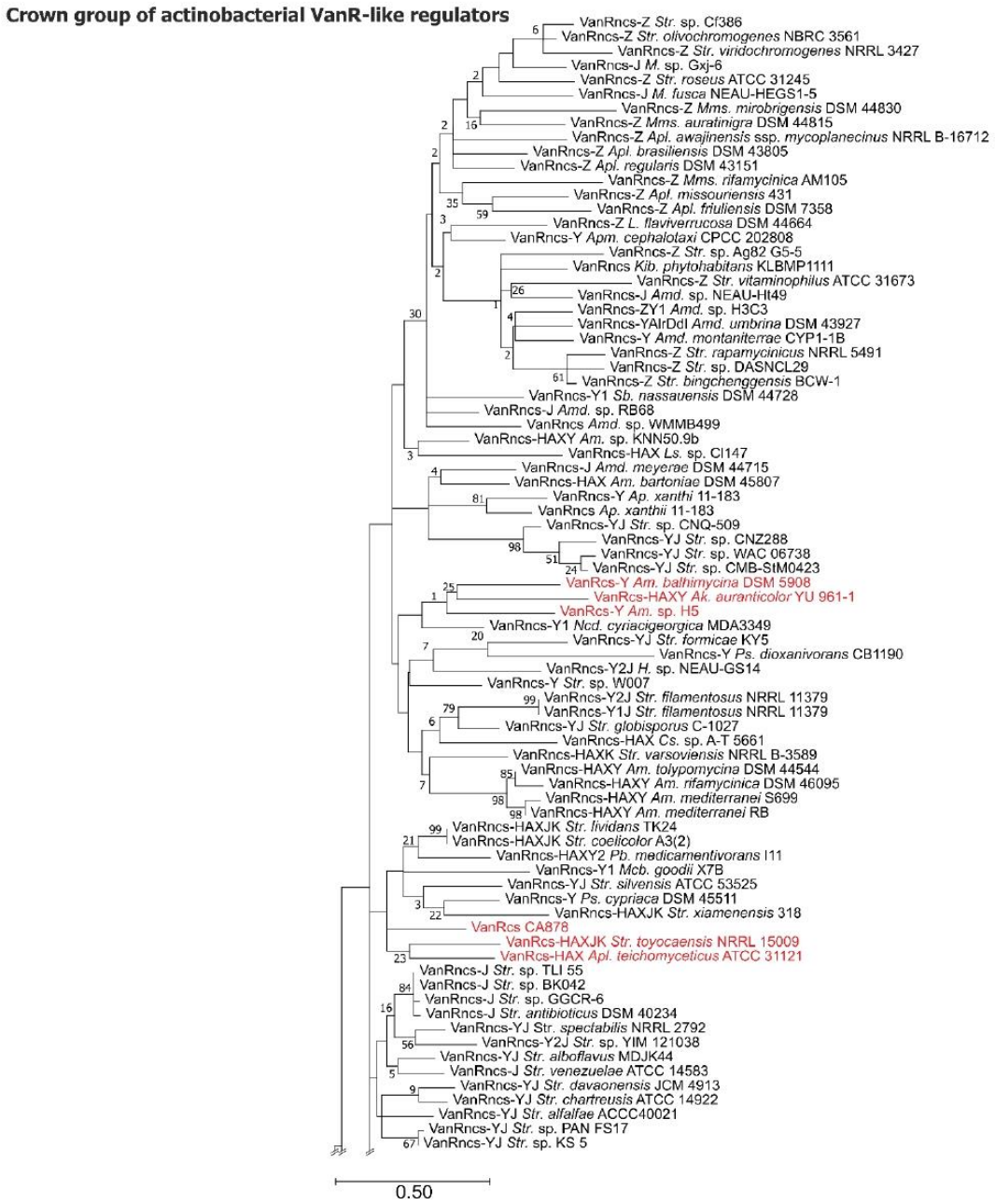
ESM Figure 21. Expanded version of the Dbv22-like cluster from the phylogenetic tree of actinobacterial VanS-like sensor histidine kinases, coded within type V GPA BGCs, *Nonomuraea* spp. type IV GPA BGCs, and *feg*-like BGCs (Figure 11a). (Figure 11b). Please refer to the main text for more details. Scale bar represents number of substitutions per site.



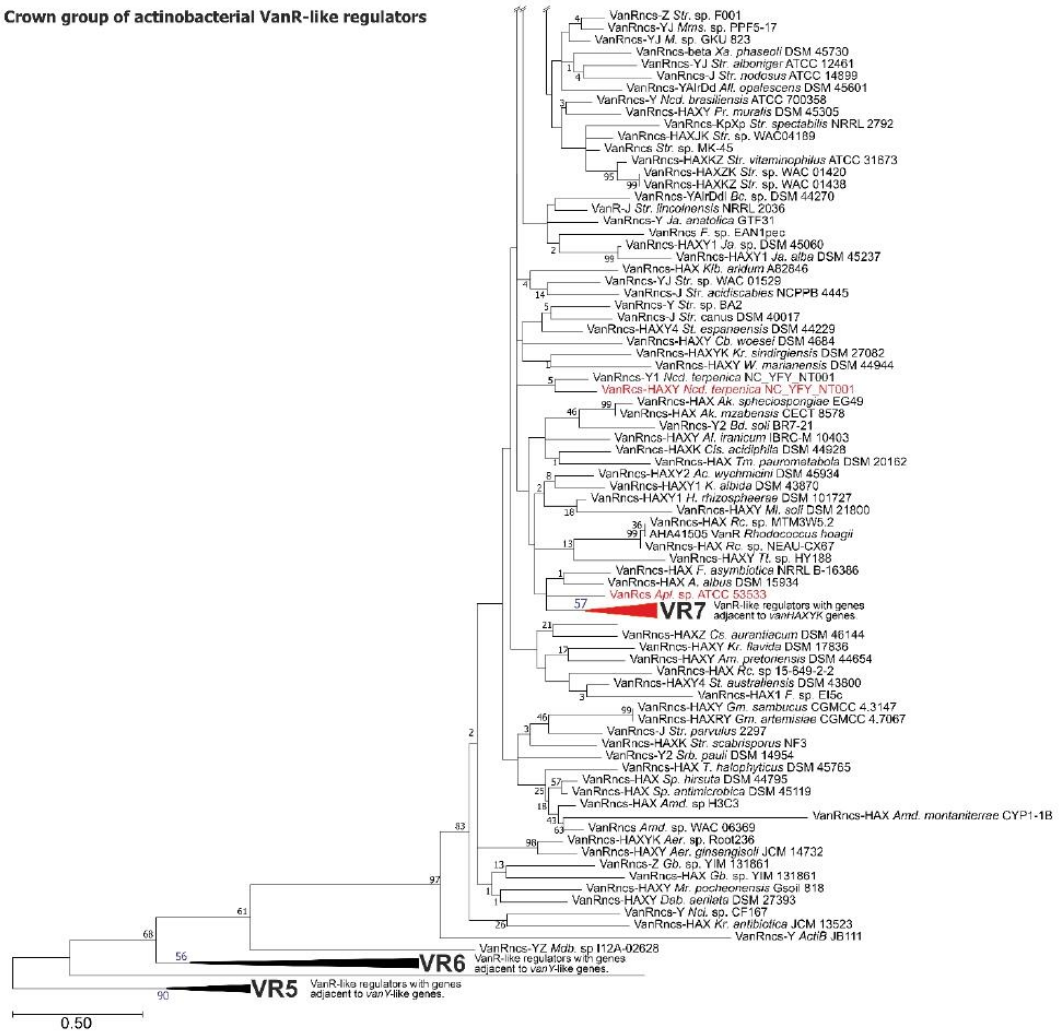
ESM Figure 22. Expanded version of the VR4 cluster from the phylogenetic tree of actinobacterial VanR-like response regulators, coded within putative *pdx*-operons from *Micromonosporales* (Figure 11a). Please refer to the main text for more details. Scale bar represents number of substitutions per site.



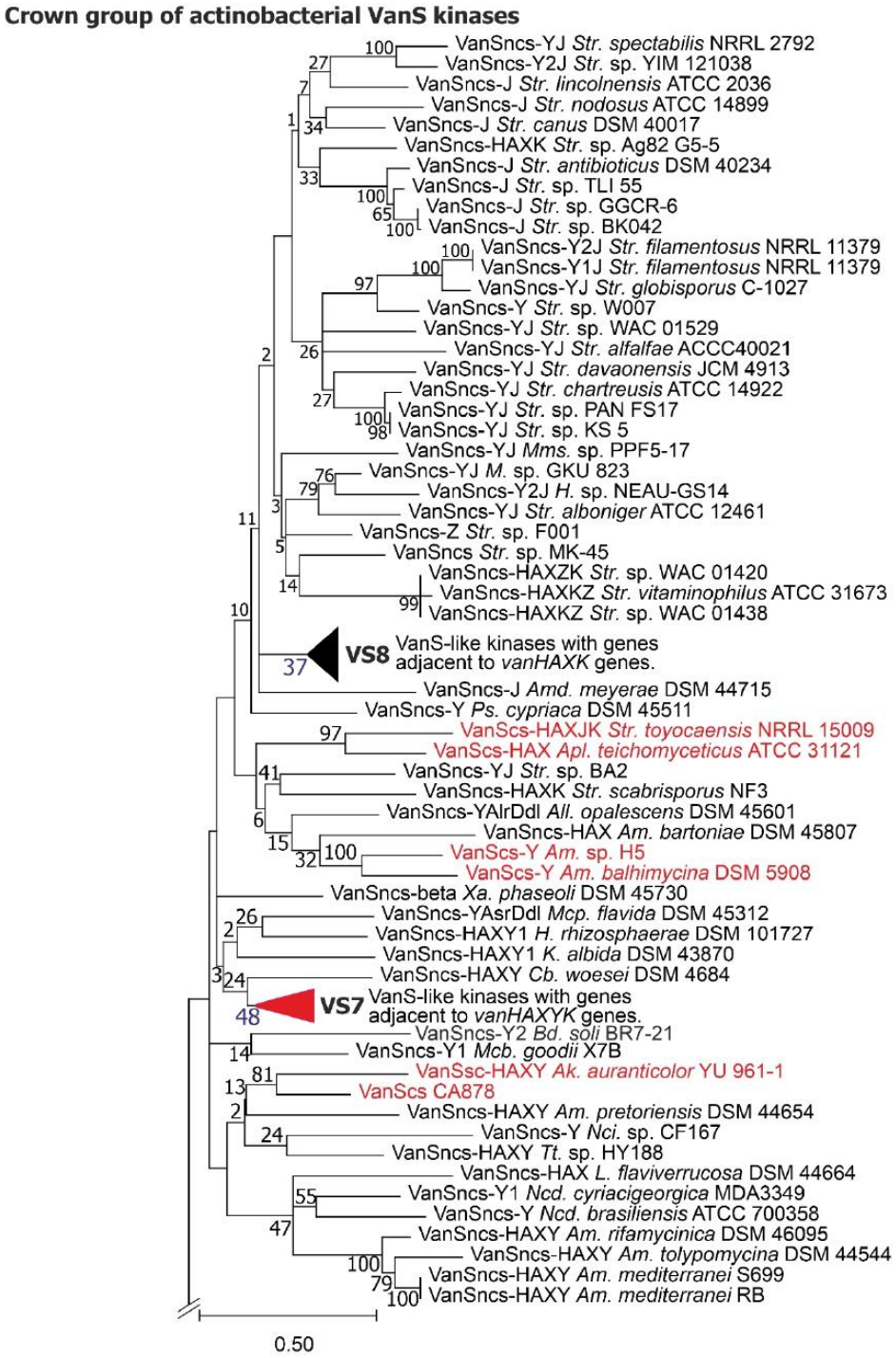
ESM Figure 23. Expanded version of the VS3 cluster from the phylogenetic tree of actinobacterial VanS-like sensor histidine kinases, coded within putative *pdx*-operons from *Micromonosporales* (Figure 11b). Please refer to the main text for more details. Scale bar represents number of substitutions per site.



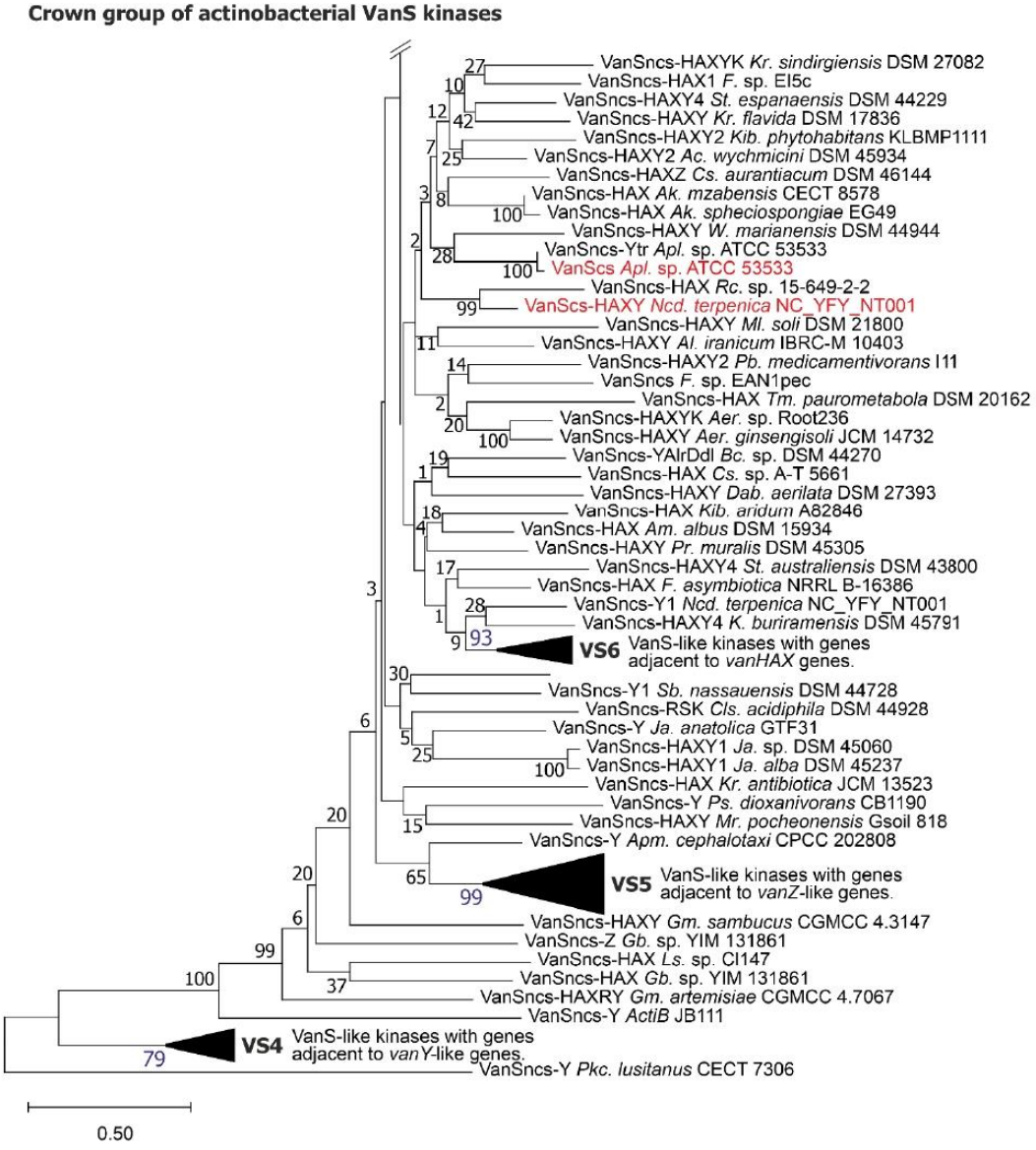
ESM Figure 24. Continued on the next page.



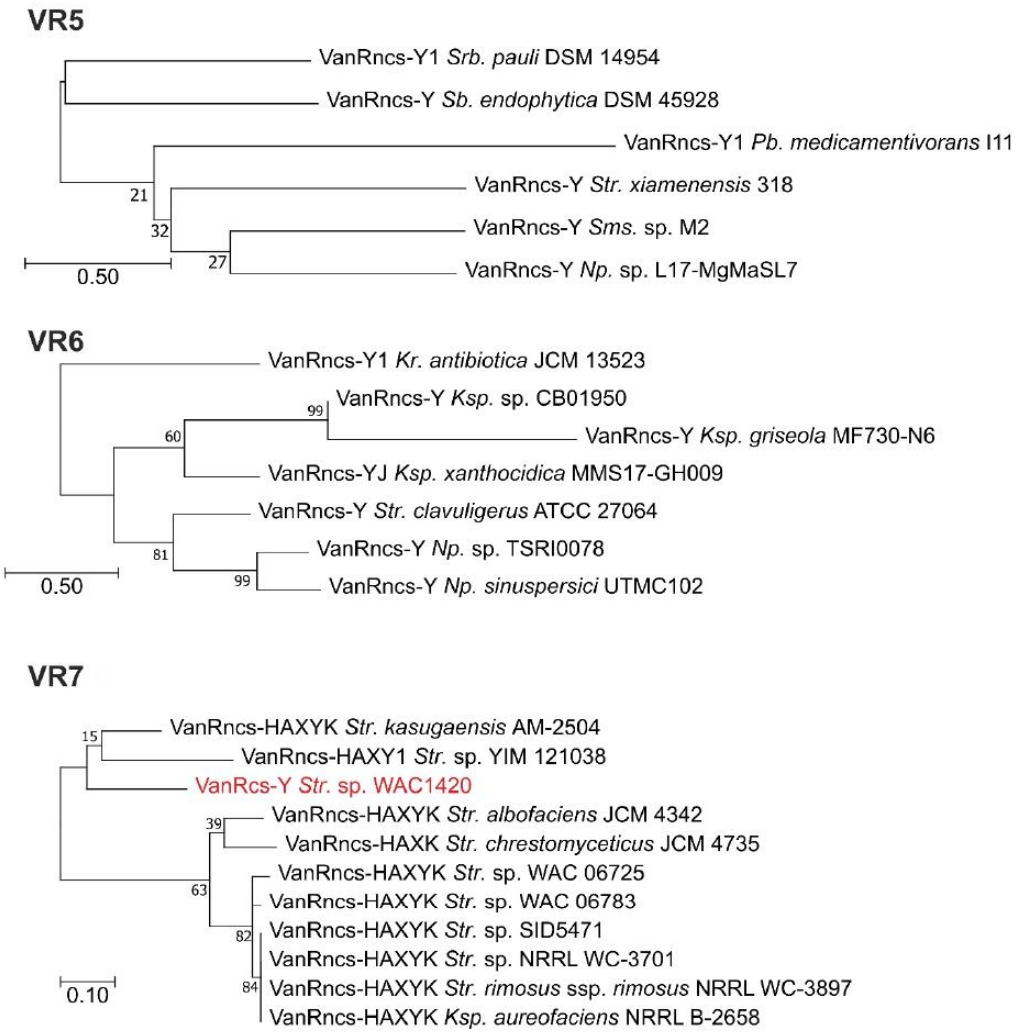
ESM Figure 24. Expanded version of the clade corresponding to the crown group of bacterial VanR response regulators from tree on Figure 11a. Well-defined clusters VR5-7 were collapsed; their expanded versions are given on ESM Figure 26. Please refer to the main text for more details. BGC-encoded proteins are given in red. Scale bar represents number of substitutions per site.



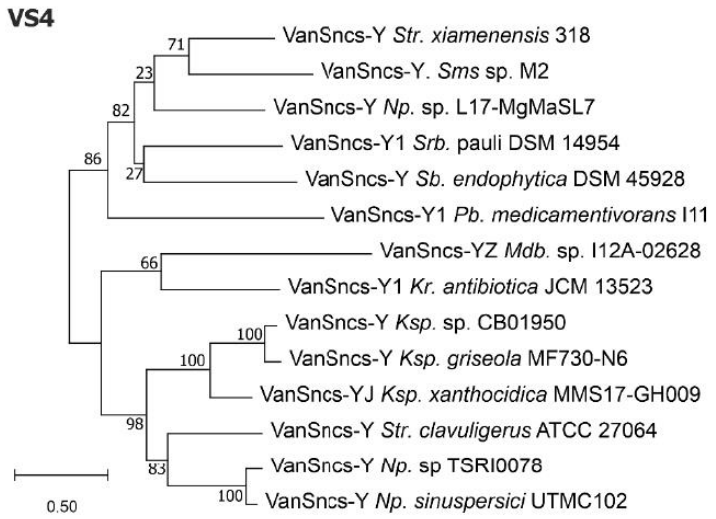
ESM Figure 25. Continued on the next page.



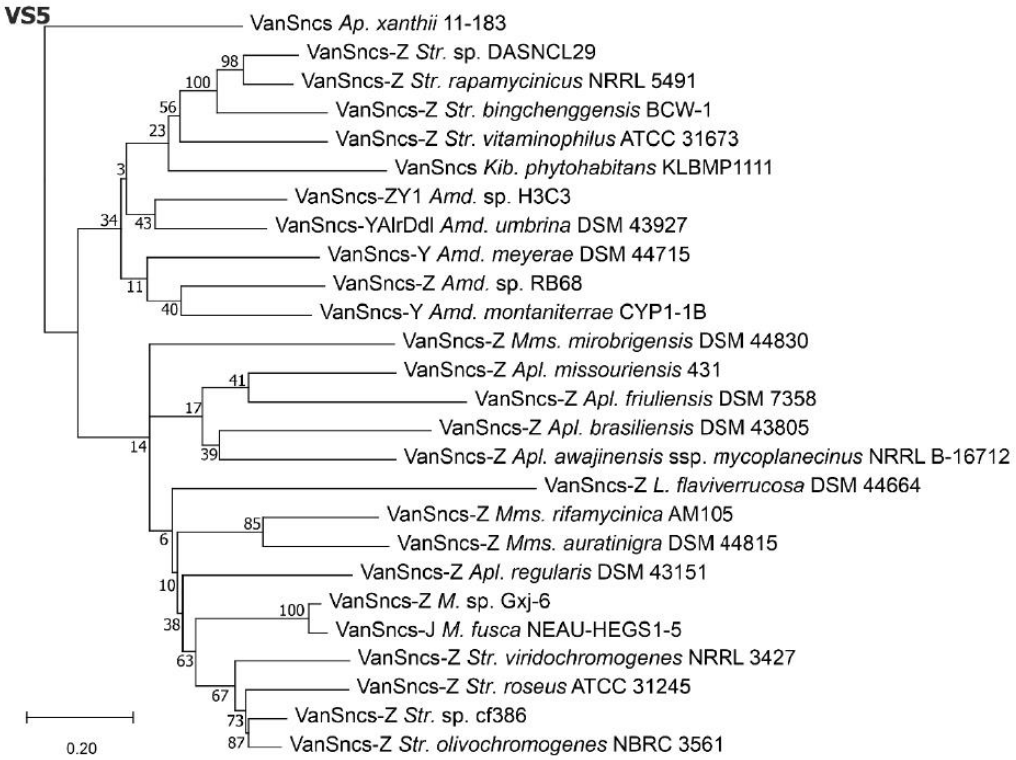
ESM Figure 25. Expanded version of the crown group of bacterial VanS-like sensor histidine kinases from tree in Figure 11b. Well-defined clusters VS4-8 were collapsed; their expanded versions are given in ESM Figures 27-31. Please refer to the main text for more details. BGC-encoded proteins are given in red. Scale bar represents number of substitutions per site.



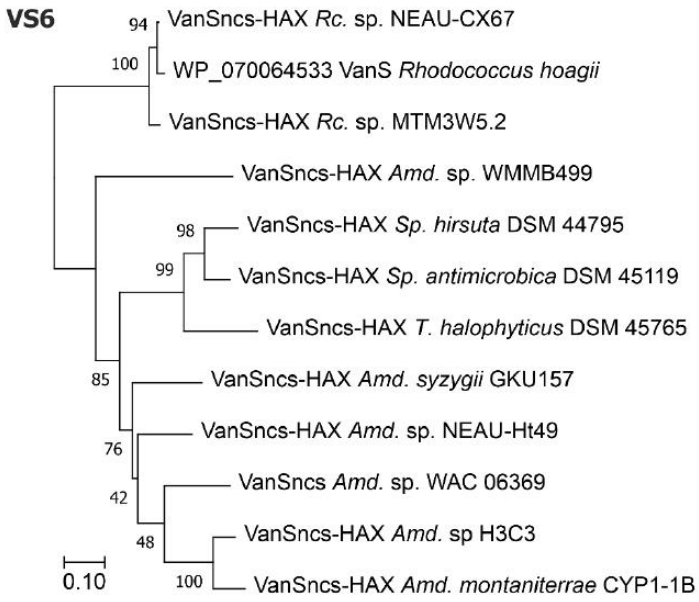
ESM Figure 26. Expanded versions of the VR5-7 clusters of the phylogenetic tree from ESM Figure 24. Please refer to the main text for more details. BGC-encoded proteins are given in red. Scale bar represents number of substitutions per site.



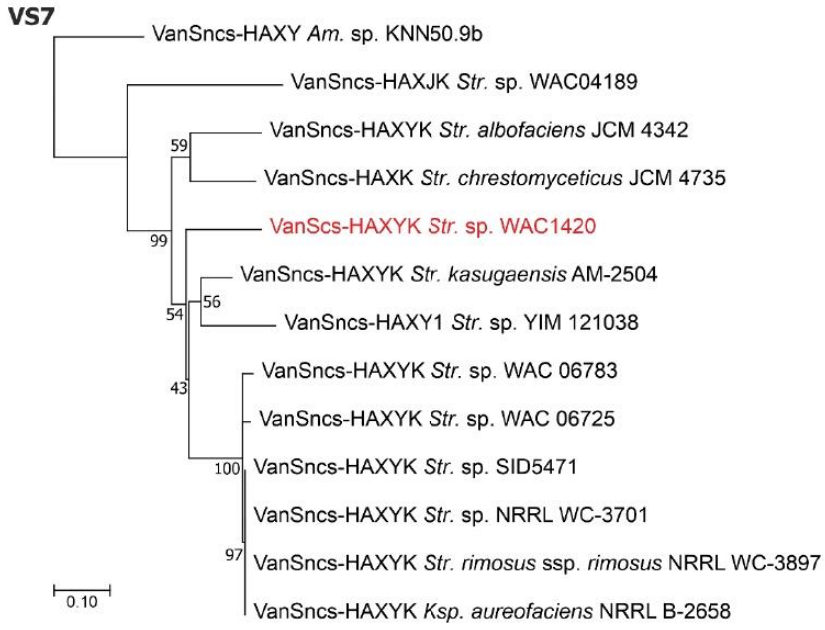
ESM Figure 27. Expanded version of the VS4 cluster of the phylogenetic tree from ESM Figure 25. Please refer to the main text for more details. Scale bar represents number of substitutions per site.



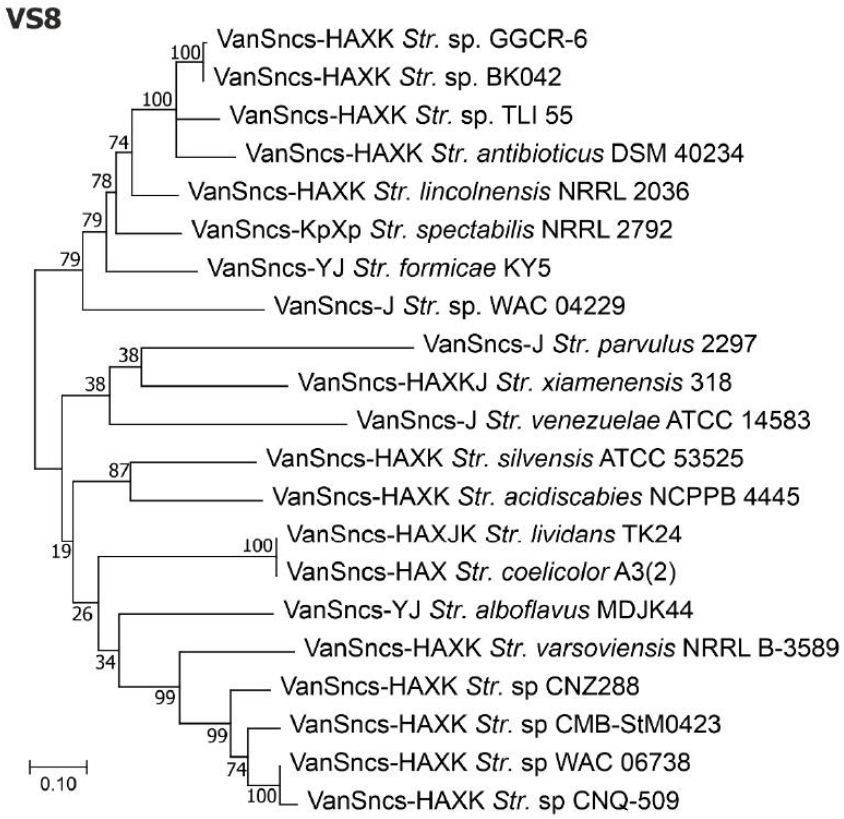
ESM Figure 28. Expanded version of the VS5 cluster of the phylogenetic tree from ESM Figure 25. Please refer to the main text for more details. Scale bar represents number of substitutions per site.



ESM Figure 29. Expanded version of the VS6 cluster of the phylogenetic tree from ESM Figure 25. Please refer to the main text for more details. Scale bar represents number of substitutions per site.



ESM Figure 30. Expanded version of the VS7 cluster of the phylogenetic tree from ESM Figure 25. Please refer to the main text for more details. BGC-encoded proteins are given in red. Scale bar represents number of substitutions per site.



ESM Figure 31. Expanded version of the VS8 cluster of the phylogenetic tree from ESM Figure 25. Please refer to the main text for more details. Scale bar represents number of substitutions per site.

Supplementary references

1. Weigel, L.M.; Clewell, D.B.; Gill, S.R.; Clark, N.C.; McDougal, L.K.; Flannagan, S.E.; Kolonay, J.F.; Shetty, J.; Killgore, G.E.; Tenover, F.C. Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. *Science.* **2003**, *302*, 1569–1571, doi:10.1126/science.1090956.
2. Heaton, M.P.; Discotto, L.F.; Pucci, M.J.; Handwerger, S. Mobilization of vancomycin resistance by transposon-mediated fusion of a VanA plasmid with an *Enterococcus faecium* sex pheromone-response plasmid. *Gene* **1996**, *171*, 9–17, doi:10.1016/0378-1119(96)00022-4.
3. Handwerger, S.; Skoble, J. Identification of chromosomal mobile element conferring high-level vancomycin resistance in *Enterococcus faecium*. *Antimicrob. Agents Chemother.* **1995**, *39*, 2446–2453, doi:10.1128/AAC.39.11.2446.
4. Paulsen, I.T.; Banerjei, L.; Hyers, G.S.A.; Nelson, K.E.; Seshadri, R.; Read, T.D.; Fouts, D.E.; Eisen, J.A.; Gill, S.R.; Heidelberg, J.F.; et al. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science.* **2003**, *299*, 2071–2074, doi:10.1126/science.1080613.
5. Garnier, F.; Taourit, S.; Glaser, P.; Courvalin, P.; Galimand, M. Characterization of transposon Tn1549, conferring VanB-type resistance in *Enterococcus* spp. *Microbiology* **2000**, *146*, 1481–1489, doi:10.1099/00221287-146-6-1481.
6. Boyd, D.A.; Cabral, T.; Van Caeseele, P.; Wylie, J.; Mulvey, M.R. Molecular characterization of the *vanE* gene cluster in vancomycin-resistant *Enterococcus faecalis* N00-410 isolated in Canada. *Antimicrob. Agents Chemother.* **2002**, *46*, 1977–1979, doi:10.1128/AAC.46.6.1977-1979.2002.
7. Quintiliani, R.; Courvalin, P. Characterization of Tn1547, a composite transposon flanked by the IS16 and IS256-like elements, that confers vancomycin resistance in *Enterococcus faecalis* BM4281. *Gene* **1996**, *172*, 1–8, doi:10.1016/0378-1119(96)00110-2.

CHAPTER 4:

Exchange of StrR and LuxR pathway-specific regulators activates the production of glycopeptide antibiotics in *N. gerenzanenesis* and *A. teichomyceticus*

Title:

Exchange of StrR and LuxR pathway-specific regulators activates the production of glycopeptide antibiotics in *N. gerenzanenesis* and *A. teichomyceticus*

Abstract:

StrR and LuxR cluster-situated regulators are essential for the production of glycopeptide antibiotics in some relevant producers like *Nonomuraea gerenzanenesis* and *Actinoplanes teichomyceticus*. In fact, knocking out *dbv3* and *dbv4* in *N. gerenzanensis* and *tei15** and *tei16** in *A. teichomyceticus,* completely abolishes A40926 and teicoplanin production, respectively. Knowledge of these regulators is crucial to understand how these biosynthetic gene clusters are expressed and how to improve antibiotic production. In this work, we demonstrate that a cross-talk between LuxR regulators from *N. gerenzanenesis* and *A. teichomyceticus* is possible. Thus, heterologous *dbv3* and *tei16** LuxR genes are able to complement GPA production when the genes for native LuxR-like regulators are knocked out. Although a complementation of antibiotic production with *tei15**-StrR gene was also achieved in *N. gerenzanensis* Δ*dbv4*, restored production in *A. teichomyceticus* Δ*tei15** was not obtained using the *dbv4*-StrR gene. Indeed, a significant increase of teicoplanin production came out following the overexpression of *dbv3* gene into *A. teichomyceticus* under the control of apramycin promoter. Our work highlights that although both *dbv4/tei15** StrR regulators share high sequence similarity, their cross talking was quite ineffective, while unrelated *dbv3/tei16** LuxR regulators gave better and promising results.

Introduction:

Actinobacteria produce more than two-thirds of antibiotics used in medicine and agriculture so far, as well as other numerous specialized metabolites.[1] One example of these valuable compounds are glycopeptide antibiotics (GPAs) which are last resort drugs against multidrug-resistant Gram-positive pathogens such as staphylococci, enterococci, and *Clostridioides difficile*.[2] They are produced by filamentous actinobacteria mainly from genera *Amycolatopsis, Actinoplanes, Nonomuraea* and *Streptomyces.* GPAs are divided into five types accordingly to their structure, and among them types I-IV are clinically relevant. Clinically important GPAs include first-generation vancomycin and teicoplanin, which are natural products, coming from *Amycolatopsis orientalis* and *Actinoplanes teichomyceticus* NRRL B-16726, respectively. Vancomycin was introduced in clinics first, in 1958, followed by the onset of teicoplanin application in 1988 in Europe and in 1998 for Japan. Second generation GPAs are semisynthetic molecules recently approved for clinical use such as telavancin, oritavancin and dalbavancin (in 2009 in case of telavancin and 2014 for telavancin and oritavancin).[3]

As majority of specialized metabolites, GPAs biosynthesis is strictly regulated by cluster-situated regulators (CSRs). CSRs directly address antibiotic biosynthesis activated by pleiotropic[4] or global regulators,[5,6] and consequently control the expression of the whole biosynthetic gene cluster (BGC). All described GPA BGCs contain at least one CSR, but an additional one could also be present. The omnipresent GPA BGC CSR belongs to the family of StrR-like transcriptional regulators; additional CSRs come from LuxR-family of transcriptional factors. So far three different models for pathway-specific regulation of GPA biosynthesis exist; one, occurring mainly in *Amycolatopsis*-derived BGCs, involves a single StrR-like regulator. Contrary, in *Actinoplanes* and *Nonomuraea*-derived BGCs an additional LuxR-like regulator takes part in the regulation. Notably,

LuxR-like regulators from different GPA BGCs are not homologs and they might be involved in different regulatory mechanisms. The roles of these regulators were further investigated in the relevant strain *Amycolatopsis balhimycina* DSM5908,[7] *A. teichomyceticus* NRRL B-16726[8,9] and *Nonomuraea gerenzanensis* ATCC 39727,[10,11] producing balhimycin, teicoplanin and A40926 (the natural precursor of dalbavancin), respectively. *Am. balhimycina* BGC (called *bal*) is regulated by only one regulator – Bbr (StrR-like) while *A. teichomyceticus* BGC (called *tei*) encodes two CSRs: Tei15* (StrR-like) and Tei16* (LuxR-like) as well as *N. gerenzanensis* BGC (called *dbv*) which contains Dbv3 (LuxR-like) and Dbv4 (StrR-like). Although StrR-like genes are quite similar between *bal*, *tei* and *dbv* clusters, *dbv3* and *tei16** LuxR-like are non-orthologous genes.[9] The experimental evidence obtained from mutants of the producer strains indicated that they positively regulate biosynthesis. Since knockout of one of these regulators in *A. teichomyceticus* or *N. gerenzanensis* completely abolishes GPA production, the expression of these genes are considered crucial for antibiotic production.[8,10] In fact, overexpression of *tei15**, *tei16**, *dbv3* and *dbv4* improves significantly antibiotic biosynthesis in the homologous producers.[8,12]

StrR-like proteins are present in all GPA BGCs known so far.[13] Moreover, DNA-binding sites of StrR-like CSR in GPA BGCs are identified with some certainty. The target binding region of Bbr, Tei15* and Dbv4 contains the highly conserved sequence [GTCCAa$(N)_{17}$TtGGAC].[7,8,14] Bbr binds *in vitro* to five regions within *bal* cluster: the promoters of *bbr* (own gene), *tba* (ABC-transporter gene) *orf7* (Na^+/K^+ antiporter gene), *dvaA* (dehydrovancosamine biosynthesis gene) and *oxyA* (cross-linking oxygenase gene).[7] Dbv4 positively regulates the expression of two operons: *dbv14-8* (oxidative cross-linking and tailoring enzymes genes) and *dbv30-35* (Dpg biosynthesis genes).[14] Tei15* directly controls the transcription of at least 17 genes and binding sites were found upstream *tei2-3* (resistance genes), *teiA* (NRPS gene), *tei2** (tailoring enzyme gene), *tei5** (oxidative cross-linking gene), *tei16** (LuxR-like CSR gene), *tei17** (DpgA biosynthesis gene), *tei23** (HpgT biosynthesis gene), *tei26** (tailoring enzyme gene), *tei28** (Hpg biosynthesis gene), *tei30** (tailoring enzyme gene) and *tei31** (AfsR-like regulatory gene).

LuxR-like proteins are the second pathway-specific regulators and they belong to the large ATP-binding regulators of the LuxR protein family. They have been studied in *A. teichomyceticus* (*tei16**) and *N. gerenzanensis* (*dbv3*). Nevertheless, these regulators are less understood than StrR-like and their binding sites were not yet identified. Tei16* DNA-binding analysis *in vitro* did not reveal putative targets within the cluster.[8] Dbv3 positively controls the expression of four genes within *dbv* BGC: *dbv4* (gene of StrR-like regulator), *dbv29* (tailoring enzyme gene), *dbv36* (type II thioesterase gene) and *dbv37* (HpgT biosynthetic gene); and six operons: *dbv2-dbv1* (Hpg biosynthetic genes), *dbv14-dbv8* (oxidative cross-linking and tailoring genes), *dbv17-dbv15* (NRPS genes), *dbv21-dbv20* (tailoring enzymes genes), *dbv24-dbv28* (NRPS, export and tailoring enzymes genes), and *dbv30-dbv35* (Dpg biosynthesis and export genes).

Recently, we showed that the overexpression of a close *dbv3* homolog, called *nocRI* from *Nonomuraea coxensis* DSM 45129, was able to increase A40926 production in *N. gerenzanensis*. This proves that a succesfull cross-talk between regulators from two different *Nonomuraea* spp. is possible.[12] In the current work, we decided to study cross-talk between the well-studied pathway-specific regulators StrR-like and between the LuxR-like present in *A. teichomyceticus* and *N. gerenzanensis,* which are much less related to each other than *dbv3* with *nocRI.* In fact, *dbv4* and *tei15** are distantly related on the overall phylogeny of StrR-like GPA biosynthesis regulators, while *dbv3* and *tei16** are not related at all.[9]

Materials and Methods

Bacterial strains and growth conditions

N. gerenzanensis ATCC 39727 and *A. teichomyceticus* NRRL-B16726 (ATCC 31121) were cultivated on ISP3 agar medium. *N. gerenzanensis* was grown in 50 mL E26 as vegetative medium for 72 h and then 10% (v/v) of the preculture was inoculated in 100 mL FM2 production medium for A40926 production.[15] *A. teichomyceticus* was grown in 50 mL of E25 vegetative medium for 72 h and then 10% (v/v) of the preculture was inoculated in 100 mL TM1 production medium for teicoplanin production.[16] Both strains were grown on an orbital shaker in baffled Erlenmeyer flasks at 30ºC, 220 rpm. *Escherichia coli* DH5α was used for routine DNA cloning. *E. coli* strains carrying recombinant plasmids were grown in Lysogeny broth (LB) agar at 37 ºC supplemented with 100 μg/mL apramycin-sulfate, 50 μg/mL hygromycin B, 50 μg/mL of kanamycin-sulfate and 25 μg/mL of chloramphenicol when necessary. All strains and plasmids used in this work are listed in Table 1.

Extraction of the genomic DNA from *N. gerenzanensis* and *A. teichomyceticus* (as well as recombinant derivatives) was done with Kirby procedure.[17] Prior to DNA isolation, *N. gerenzanensis* and *A. teichomyceticus* strains were grown in 250 mL baffled Erlenmeyer flasks with 50 mL of E25 or E26 on an orbital shaker at 220 rpm and at 30 °C for 96 h.

Table 1. Strains and plasmids used in this work

Strain or plasmid	Description	Source or reference
pSAD3	pSET152A derivative containing *dbv3* regulatory gene under the control of *aac(3)IVp* promoter	[12]
pSAD4	pSET152A derivative containing *dbv4* regulatory gene under the control of *aac(3)IVp* promoter	[12]
pIJ10700	Template vector for the amplification of hygromycin resistance cassette for pSHAD3 and pSHAD4 construction	[18]
pSHAD3	pSAD3 derivative containing hygromycin resistance gene instead of apramycin	This work
pSHAD4	pSAD4 derivative containing hygromycin resistance gene instead of apramycin	This work
pIJ12551	ϕC31-actinophage based integrative expression vector with *ermE** promoter	[19]
pIJ12551dbv4	pIJ12551 derivative containing *dbv4* regulatory gene under the control of *ermE** promoter	Dr. Elisa Binda (unpublished data)
pSET152Atei15*	pSET152A derivative containing *tei15** regulatory gene under the control of *aac(3)IVp* promoter	[8]
pSET152Atei16*	pSET152A derivative containing *tei16** regulatory gene under the control of *aac(3)IVp* promoter	[8]

pIJ12551tei15*	pIJ12551 derivative containing *tei15** regulatory gene under the control of *ermE** promoter	Dr. Elisa Binda (unpublished data)
pKC1132	Suicide plasmid lacking actinobacterial oriR	[20]
pKCKOD3	pKC1132 derivative containing *dbv3* flanking regions and streptomycin resistance cassette	This work
pKCKOD4	pKC1132 derivative containing *dbv4* flanking regions and streptomycin resistance cassette	This work
pKCKOD3-4	pKC1132 derivative containing *dbv3* and *dbv4* flanking regions and streptomycin resistance cassette	This work
pIJ778	Template vector for the amplification of streptomycin resistance cassette for PCR-targeted mutagenesis	[21]
A40Y	SuperCos1 derivative, including 22 kb fragment of *dbv* cluster (*dbv1-dbv17*)	[22]
A40dbv3::aadA	A40Y derivative with *dbv3* replaced by *aadA-oriT* cassette derived from plasmid pIJ778	This work
A40dbv4::aadA	A40Y derivative with *dbv4* replaced with *aadA-oriT* cassette derived from plasmid pIJ778	This work
N. gerenzanensis	Wild type, A40926 producer	ATCC 39727
N. gerenzanensis Δ*dbv3*	Wild type with *dbv3* replaced with *aadA-oriT* cassette	This work
N. gerenzanensis Δ*dbv4*	Wild type with *dbv4* replaced with *aadA-oriT* cassette	This work
N. gerenzanensis Δ*dbv3-4*	Wild type with *dbv3* and *dbv4* replaced with *aadA-oriT* cassette	This work
N. gerenzanensis Δ*dbv3* pSAD3	*N. gerenzanensis* Δ*dbv3* derivative carrying pSAD3 plasmid	This work
N. gerenzanensis Δ*dbv4* pSAD4	*N. gerenzanensis* Δ*dbv4* derivative carrying pSAD4 plasmid	This work
N. gerenzanensis Δ*dbv4* pIJ12551dbv4	*N. gerenzanensis* Δ*dbv4* derivative carrying pIJ12551dbv4 plasmid	This work
N. gerenzanensis pSET152Atei15*	Wild type derivative carrying pSET152Atei15* plasmid	This work
N. gerenzanensis pSET152Atei16*	Wild type derivative carrying pSET152Atei16* plasmid	This work
N. gerenzanensis pIJ12551tei15*	Wild type derivative carrying pIJ12551tei15* plasmid	This work
N. gerenzanensis Δ*dbv4* pSET152Atei15*	*N. gerenzanensis* Δ*dbv4* derivative carrying pSET152Atei15* plasmid	This work
N. gerenzanensis Δ*dbv3* pSET152Atei16*	*N. gerenzanensis* Δ*dbv3* derivative carrying pSET152Atei16* plasmid	This work

N. gerenzanensis Δ*dbv4* pIJ12551dbv4	*N. gerenzanensis* Δ*dbv4* derivative carrying pIJ12551dbv4 plasmid	This work
A. teichomyceticus	Wild type, teicoplanin producer	NRRL-B16726
A. teichomyceticus pSAD3	Wild type derivative carrying pSAD3 plasmid	This work
A. teichomyceticus pSAD4	Wild type derivative carrying pSAD4 plasmid	This work
A. teichomyceticus pIJ12551dbv4	Wild type derivative carrying pIJ12551dbv4 plasmid	This work
A. teichomyceticus Δ*tei15**	Wild type derivative with *tei15** gene replaced with *aac(3)IV-oriT* cassette	[8]
A. teichomyceticus Δ*tei16**	Wild type derivative with *tei16** gene replaced with *aac(3)IV-oriT* cassette	[8]
A. teichomyceticus Δ*tei15** pSHAD4	*A. teichomyceticus* Δ*tei15** derivative carrying pSHAD4 plasmid	This work
A. teichomyceticus Δ*tei16** pSHAD3	*A. teichomyceticus* Δ*tei16** derivative carrying pSHAD3 plasmid	This work
A. teichomyceticus Δ*tei15** pIJ12551dbv4	*A. teichomyceticus* Δ*tei15** derivative carrying pIJ12551dbv4 plasmid	Dr. Elisa Binda (unpublished data)
E. coli DH5α	General cloning host	MBI Fermentas, US
E. coli ET12567 pUZ8002	(*dam*-13::Tn9 *dcm*-6), pUZ8002^{+} (Δ*oriT*), used for conjugative transfer of DNA	[17]
Bacillus subtilis HB0933	CU1065 *liaR*::*kan*	[23]
Bacillus subtilis HB0950	CU1065 SPβ2Δ2::Tn*917*::Φ($P_{liaI\text{-}74}$-*cat-lacZ*)	[24]

dbv3* and *dbv4* inactivation in *N. gerenzanensis

First of all, *dbv3* and *dbv4* genes were replaced on A40Y cosmid,[22] which contains the majority of *dbv* BGC genes, with a spectinomycin resistance cassette *aadA-oriT* derived from plasmid pIJ778 using the λ-Red recombination process,[21] yielding A40dbv4::aadA and A40dbv3::aadA. Then, using the previous recombinant cosmids, we amplified *aadA–oriT* with *ca.* 2 Kbp flanking regions of *dbv3* and *dbv4* (which previously were replaced by streptomycin cassette) and cloned the amplicons into pKC1132 suicide vector, carrying apramycin resistance cassette. Thus, *dbv3* and *dbv4* flanking regions were intercalated with streptomycin resistance gene. For inactivation of both genes together same approach was applied. In this way pKCKOD3, pKCKOD4 and pKCKOD3-4 knockout suicide plasmids were generated. Transconjugants carrying spectinomycin (Sm^R) and apramycin resistance cassette (Am^R) were grown several times on ISP3 solid medium without antibiotic selection until we obtained Sm^R Am^S strains, due to a second crossover event, resulting in the loss of Am^R cassette, leaving only Sm^R cassette, replacing the target gene. Mutant strains were verified by PCR, using 3/4KOVER_F/R primer pair. Steps in details are discussed in Results.

Cloning of *tei15 and *dbv4* into pIJ12551**

The coding sequences of *tei15** and *dbv4* were amplified from the genomic DNA of *A. teichomyceticus* and *N. gerenzanensis*, respectively and cloned into pIJ12551 plasmid. Amplicons were generated using Q5 High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, United States) and oligonucleotide primers listed in Table 2. Then, amplicons were digested with *Nde*I and *Not*I and cloned via the same recognition sites into pIJ12551, generating pIJ12551tei15* and pIJ12551dbv4.

Table 2. Primers used in this work

Primer	Nucleotide sequence (5′-3′)[#]	Purpose
dbv3KO_R	TTTACTAGTCCCCGATGAGCCTCGGTCC	Amplification of *dbv3* with 2 Kbp flanks
dbv3KO_F	TTTAAGCTTGCCGATCAGACCGGTGCCG	
dbv4KO_R	CCTCGCGAGCTCGTAGCCG	Amplification of *dbv4* with 2 Kbp flanks
dbv4KO_F	CGTGGAAGCGCAGTGCCTC	
3/4KOVER_R	GATATCCTGCCCGAGGCCG	Verification of deletion for *dbv3* and *dbv4* genes
3/4KOVER_F	CCAGATGCTGCAGGCGCGA	
dbv3_P1	GCAAACCAAGTCGACGAACCGCTTGGGGGA CGAGCAAGAATTCCGGGGATCCGTCGACC	Replacement of *dbv3* with the *aadA-oriT* cassette within the cosmid A40Y
dbv3_P2	GCCCGAGGCCGGCGAATTCGGCTTGTCGAA CTCTTCGCTTGTAGGCTGGAGCTGCTTC	
dbv4_P1	CCCCGGCTCCGATATGACGCTAATCGAATC GGAGGCTAGATTCCGGGGATCCGTCGACC	Replacement of *dbv4* with the *aadA-oriT* cassette within the cosmid A40Y
dbv4_P2	GTCGCTCTACATACGGCCGCCCGGCTCATC CACTCGTGCTGTAGGCTGGAGCTGCTTC	
dbv3_FWpIJ	AAAACATATGCTGTTCGGGCGAGATCGT	Amplification of *dbv3*
dbv3_RVpIJ	AAAAGCGGCCGCCTACAGCCGCACTGCCTC	
dbv4_FWpIJ	AAAAAACATATGGACCCGACGGGAGTTGAC ATA	Cloning of *dbv4* into pIJ12551
dbv4_RVpIJ	TTTATTAGCGGCCGCTCATCCAGCGGCCAGA TCGGTCG	
tei15_FWpIJ	GGGCATATGACACCTGACGAAGAG	Cloning of *tei15** into pIJ12551
tei15_RVpIJ	AAAAGCGGCCGCTCAGCTCGCCATC	
pSET_ver_F	GCATCGGCCGCGCTCCCGA	Verification of *tei15** and *tei16** cloned into pSET152Atei15* and pSET152Atei16*, respectively

tei15*_ver_R	CAGCTCAGCGCCGCTGAGCA	Verification of *tei15** cloned into pSET152Atei15*
tei16*_ver_R	CTCGCACACGCCCGGGCC	Verification of *tei16** cloned into pSET152Atei16*
aac(3)IV_Fw	GTTTTCCCAGTCACGACGTT	Verification of apramycin resistance cassette
aac(3)IV_Rv	TATCCGCTCACAATTCCACA	

[#] Binding sites of restriction endonucleases are underlined.

Conjugative transfer of plasmids into *N. gerenzanensis* and *A. teichomyceticus* and verification of the recombinant strains

Conjugative transfer of plasmids into *N. gerenzanensis* and *A. teichomyceticus* was performed essentially as described previously.[17,25] All recombinant plasmids were transferred individually into the non-methylating *E. coli* ET12567 pUZ8002 and the resulting derivatives were used as donor strains for intergeneric conjugation. To verify the integration of plasmids, target genes and apramycin resistance gene were amplified by PCR from the genomic DNA isolated from the recombinant strains.

To prepare fresh vegetative mycelium of *N. gerenzanensis* prior to conjugal transfer, one vial of WCB was inoculated into 50 mL of VSP medium (250 mL Erlenmeyer flask with 10 ø5 mm glass beads) and incubated for 48 h on the orbital shaker at 30 °C, 220 rpm. The mycelium was collected by centrifugation (10 min, 3,220 × *g*), washed twice with sterile 20% v/v glycerol, resuspended in the same solution to a final volume of 20 mL, and stored at −80°C. 1 mL of mycelial suspension was mixed with approximately 10^9 of donor *E. coli* cells and the mixtures were plated on well dried VM0.1 agar plates supplemented with 20 mM of $MgCl_2$. After 12–16 h of incubation at 30 °C, each plate was overlaid with 1 mL of sterile deionized water containing 1.25 mg of apramycin-sulfate and 750 μg of nalidixic acid sodium salt. Transconjugants were selected as resistant to 50 μg/mL of apramycin-sulfate.

Spore suspensions of *A. teichomyceticus* were prepared from lawns grown on ISP3 agar for 7 days. Sporangia from one plate were collected in deionized water and filtered through one layer of Miracloth (Merck KGaA, Darmstadt, Germany) to remove vegetative mycelial fragments. Then sporangia were incubated in an orbital shaker at 30 ºC until spores were released from sporangia. Then, spores were centrifuged (15 min, 3,220 × *g*) and resuspended in 1 mL of 15% v/v glycerol, and stored at −80°C. For conjugation, approx. 10^6 spores were mixed with 10^7 *E. coli* donor cells and plated on SFM agar plates supplemented with 20 mM of $MgCl_2$. The overlay for the selection of transconjugants was performed as described previously for *N. gerenzanensis*.

Analytical bioassay

To prepare the plates for analytical bioassays, *B. subtilis* HB0933[23] and HB0950[24] were grown in LB liquid medium at 37 ºC in a shaker (150 rpm) for 15-16 h. Then, 10% (v/v) of the overnight culture was inoculated in a fresh LB medium and left to grow under the same incubation conditions up to 0.6 OD. Subsequently, 100 μl of so-obtained *B. subtilis* culture were inoculated to 30 mL of Muller Hinton Agar (MHA). In case of *B. subtilis* HB0950, additional 50 μg/mL of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) were added for GPA-induced chromogenic conversion. 6 mm paper discs socked

with antibiotic extracts or agar plugs were placed on the surface of these plates. Plates were incubated overnight at 37 ºC, growth inhibition halos were onserved and documented.

HPLC analysis of teicoplanin and A40926

GPAs were generally extracted by mixing 1 volume of broth with 1 volume of borate buffer [100 mM H_3BO_3 (Sigma-Aldrich), 100 mM NaOH (Sigma-Aldrich), pH 12]. Extraction of teicoplanin was perfomed by shaking samples on a rotary shaker at 200 rpm and 37 °C, then samples were centrifuged (16,000 × *g* for 15 min) to obtain debris-free supernatants. Extracts containing A40926 were centrifuged (16,000 × *g* for 15 min) and supernatants were incubated at 50 ºC, 1 h.

In some cases, lyophilization was required in order to concentrate samples containing amounts of GPAs hard to trace in non-concentrated samples. Thus, culture extracts were first collected at the maximum production timepoint and adjusted to pH 12 with 10 M NaOH to extract GPAs from mycelium. Then, samples were centrifuged (16,000 × *g* for 10 min), supernatants were lyophilized in a VirTis Sentry vacuum chamber for 24 h and subsequently reconstituted in MilliQ, concentrating the sample 10 times.

HPLC was performed with a VWR Hitachi diode array L-2455 HPLC system with detection at 254 nm. Samples were estimated by injecting 50 µL of sample onto a 5 µm-particle-size Ultrasphere ODS (Beckman) HPLC column (4.6 by 250 mm). A40926 and teicoplanin samples eluting at a flow rate of 1 mL/min with a 30 min linear gradient from 15 to 64% of phase B. Phase A was 32 mM $HCOONH_4$ (pH 7) – CH_3CN [90:10 (vol/vol)], and phase B was 32 mM $HCOONH_4$ (pH 7) – CH_3CN [30:70 (vol/vol)]. A volume of 50 µL of a pure sample of 150 µg/mL A40926 (Sigma-Aldrich, St. Louis, MO, United States) and 100 µg/mL teicoplanin (Sigma-Aldrich, St. Louis, MO, United States) were used as standards.

Results

Generation of *dbv3* and *dbv4* knockout mutants in *N. gerenzanensis*

Previously it was reported that *tei15** and *tei16** knockouts completely abolished the production of teicoplanin. Expectedly, reintroduction of the wild-type alleles restored teicoplanin production.[8] The same was also shown for *dbv3* and *dbv4*, pathway-specific regulators of A40926 production in *N. gerenzanensis.*[10] Since ultimate aim of our work was to cross-complement knockouts of *dbv3*/*dbv4* with *tei15**/*tei16** and *vice versa*, we required apramycin-sensitive mutants of *N. gerenzanensis.* Unfortunately, *dbv3* and *dbv4* knockout mutants, described in the literature[10] were indeed apramycin-resistant, since *aac(3)IV* apramycin resistance gene was used to knock out either *dbv3* or *dbv4*. Thus, we decided to create our own *dbv3* and *dbv4* knockout mutants, using *aadA* – spectinomycin/streptomycin resistance gene – instead. The majority of *dbv* BGC (including *dbv3* and *dbv4*) could be found on A40Y cosmid. Therefore, we constructed two derivatives of A40Y, carrying λ-RED induced single gene substitutions. In both cases, the *dbv* genes were replaced with a spectinomycin resistance cassette *aadA-oriT* derived from plasmid pIJ778.[21]

The *dbv3* gene was replaced with the *aadA-oriT* cassette within the cosmid A40Y (Table 1) using the λ-Red recombination process.[21] The primers dbv3_P1 and dbv3_P2 used for replacement are listed in Table 2. The cosmid A40dbv3::aadA was generated. Then, *dbv4* was replaced in similar fashion, with the *aadA-oriT* cassette within the cosmid

A40Y (Table 1) using the λ-Red recombination process. The primers dbv4_P1 and dbv4_P2 used for replacement are listed in Table 2. Thus, the cosmid A40dbv4::aadA was obtained. Following the standard protocol, we then attempted to transfer A40dbv4::aadA and A40dbv3::aadA into *N. gerenzanensis* by means of intergeneric conjugation with *E. coli* ET12567 pUZ8002 using spectinomycin for the selection of transconjugants (which were expected to be a result second-crossover event, and, thus, should already be desired mutants). Unfortunately, all of the obtained spectinomycin-resistant "transconjugants" carried wild type alleles of *dbv3* and *dbv4* when tested by PCR-genotyping. We speculate that these "transconjugants" were actually not the result of conjugation, but rather spontaneous spectinomycin-resistant mutants.

Therefore, we decided to change the strategy. A40dbv4::aadA and A40dbv3::aadA recombinant cosmids were thus used to amplify *aadA–oriT* with *ca.* 2 Kbp flanking regions (which actually were either *dbv3* or *dbv4* flanks). Primer pairs dbv3KO_F/R and dbv4KO_F/R (Table 2) were used to generate 5,497 and 5,404 bp amplicons, respectively. Both amplicons were digested with *Hind*III and *Spe*I restriction endonucleases and cloned into the *Hind*III and *Xba*I recognition sites of pKC1132 suicide vector. Obtained plasmids were named pKCKOD3 and pKCKOD4 and transferred into *N. gerenzanensis* via intergeneric conjugation with *E. coli* ET12567 pUZ8002. Transconjugants were selected as resistant to 50 μg/mL of apramycin-sulfate, having pKCKOD3 and pKCKOD4 integrated into the chromosome via first crossover event. One transconjugant was then randomly picked for *N. gerenzanensis* pKCKOD3$^+$ and pKCKOD4$^+$ each and passaged four times on ISP3 agar without any selective factors to induce second crossover event. After fourth passage, *N. gerenzanensis* pKCKOD3$^+$ and pKCKOD4$^+$ were streaked out to obtain single colonies which then were screened for apramycin-sensitive ones. Overall, we obtained 4 apramycin-sensitive colonies out of *ca.* 2000 pKCKOD4$^+$ screened colonies and 2 apramycin-sensitive colonies out of *ca.* 4000 pKCKOD3$^+$ screened colonies pKCKOD3$^+$. Such differences in the second-crossover efficiency were likely due to larger size of *dbv3,* which is 2,682 bp. Genomic DNA was isolated from all these strains using Kirby procedure[17] and substitution of either *dbv3* or *dbv4* was verified by PCR-genotyping using 3/4KOVER_F/R primer pair (Table 2). A summarizing scheme of the whole knockout approach is given in Figure 1 whereas the results of the PCR verification are given in Figure 1C. For the following experiments, one mutant strain among those knocked out in *dbv3* or in *dbv4* was randomly chosen; mutants were named *N. gerenzanensis* Δ*dbv3* and Δ*dbv4.*

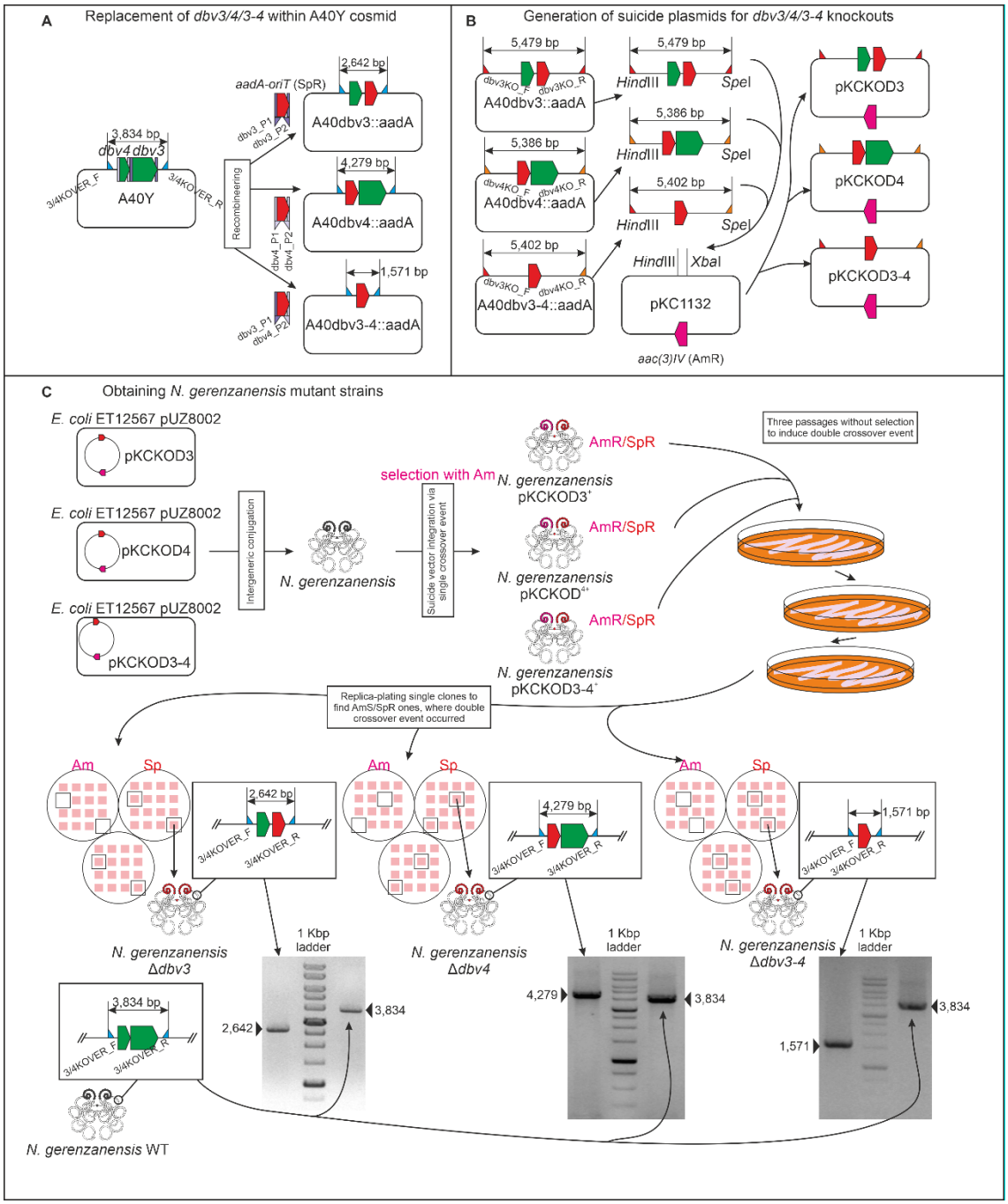


Figure 1. Generation of *dbv3* and *dbv4* knockout strains in *N. gerenzanensis.* A) Replacement of *dbv3*, *dbv4* and both genes together within A40Y cosmid, yielding A40dbv3::aadA, A40dbv4::aadA and A40dbv3-4::aadA, respectively. *dbv3* and *dbv4* are represented with green arrows and *aadA-oriT* cassette with red arrow. Primers to replace *dbv3* are shown in purple triangles while primers to replace *dbv4* are shown as grey triangles. 3/4KOVER primers, which are illustrated as blue triangles, were used to verify the integration of the resistance cassette. B) *aadA–oriT* with *ca.* 2 Kbp flanking regions were amplified using dbv3KO_F/R (red triangles) or dbv4KO_F/R (orange triangles) and introduced in pKC1132 suicide vector. Resulting plasmids were named pkCKOD3, pkCKOD4 and pkCKOD3-4, carrying apramycin resistance cassette (pink arrow). C) Scheme of process followed to obtain and verify *N. gerenzanensis* Δ*dbv3,* Δ*dbv4* and Δ*dbv3-4* strains. Please see main text for more details.

A40926 production in Δ*dbv3* and Δ*dbv4* and complementation of these strains

To verify that both knockouts abolish A40926 production, we cultivated *N. gerenzanensis* Δ*dbv3* and Δ*dbv4* in FM2 production medium according to the previously described protocol.[15] Bioassays (antimicrobial activity on *B. subtilis*) and HPLC analysis indeed showed that neither *N. gerenzanensis* Δ*dbv3* nor Δ*dbv4* were able to produce A40926 (Figure 2).

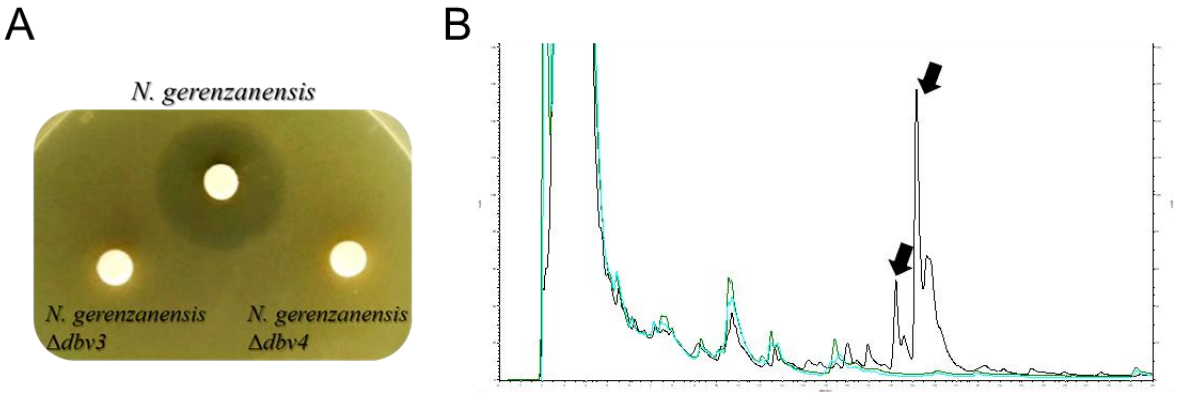


Figure 2. In *N. gerenzanensis* Δ*dbv3* and Δ*dbv4* A40926 production is abolished. A) Bioassays against *B. subtilis* HB0933 in MHA medium. *N. gerenzanensis dbv3* and *dbv4* knockout strains did not show any antimicrobial activity. Samples were collected at 144 h in FM2 medium and A40926 extraction was done by raising pH up to 12. 50 µL sample was loaded in a 6 mm paper disk. B) HPLC analysis confirmed that these recombinants did not show any A40926 production, which indeed is present in the chromatogram from the wild type: *N. gerenzanensis* chromatogram is showed in black colour, *N. gerenzanensis* Δ*dbv3* in green and *N. gerenzanensis* Δ*dbv4* in blue. A40926 complex peaks are indicated with black arrows.

Then, to complement *N. gerenzanensis* Δ*dbv3* and Δ*dbv4,* two different plasmids carrying *dbv3* and *dbv4* were used: pSET152A and pIJ12551 (Table 1). In the first case, pSAD3 and pSAD4 plasmids were previously used in our work for the overexpression of *dbv3* and *dbv4* into *N. gerenzanensis.*[12] *dbv4* was also cloned into pIJ12551, yielding pIJ12551dbv4. pSET152A-derived plasmids contain strong apramycin resistance gene promoter (*aac3(IV)p*) which was previously used to markedly increase the production of GPAs in *Nonomuraea* spp.[12,26] At the same time, pIJ12551 carries the erythromycin resistance gene promotor widely used for gene overexpression in *Streptomyces* (*ermEp*), which was shown previously to be rather weak in *Nonomuraea* sp.[12] Both plasmids are φC31-based integrative vectors and they provide stable gene expression. The three plasmids were transferred into respective deletion mutants to generate the complemented strains *N. gerenzanensis* Δ*dbv3* pSAD3$^+$, *N. gerenzanensis* Δ*dbv4* pSAD4$^+$ and *N. gerenzanensis* Δ*dbv4* pIJ12551dbv4$^+$ via intergeneric mating with *E. coli* ET12567 pUZ8002; obtained recombinants were verified by PCR (Figure 3A). All recombinant strains showed restored antimicrobial activity in bioassays (Figure 3B). In addition, HPLC analysis revealed that *N. gerenzanensis* Δ*dbv3* pSAD3$^+$ and Δ*dbv4* pSAD4$^+$ restored 30% biosynthesis while pIJ12551 restored 100% production in *N. gerenzanensis* Δ*dbv4* pIJ12551dbv4$^+$. We believe that this peculiar result is due to the strength of *aac(3)IVp*; probably strong, "time-ectopic" expression of pathway-specific regulators

leads to A40926 production from the very beginning of the cultivation and, thus, to additional stress in the production culture.

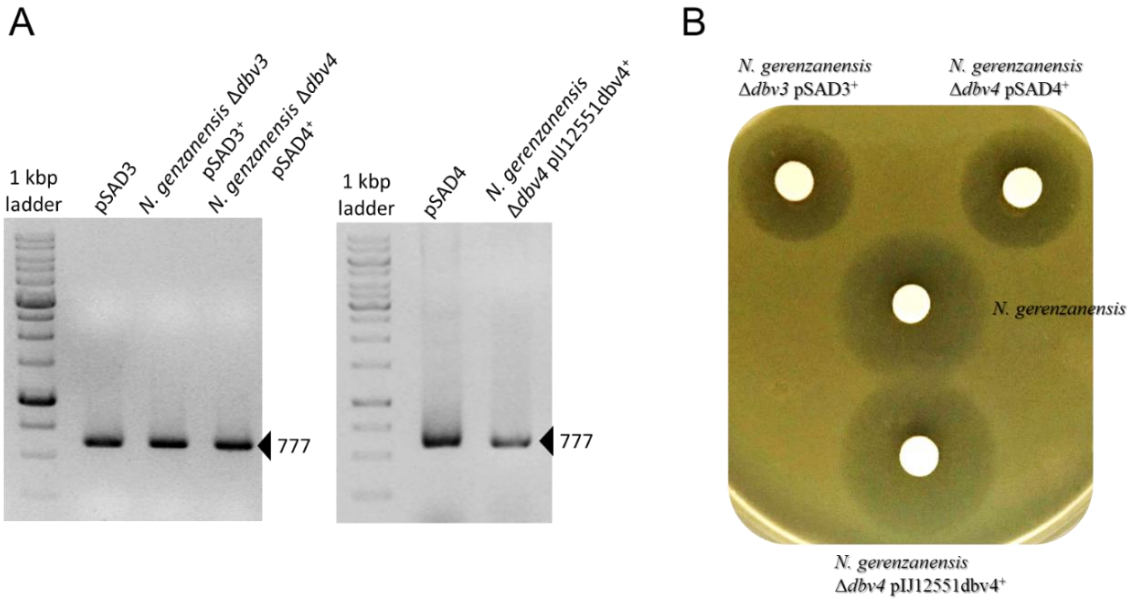


Figure 3. *N. gerenzanensis* Δ*dbv3* and Δ*dbv4* complemented with native regulatory genes have A40926 production restored. A) Verification of recombinant strains using aac(3)IV_Fw and aac(3)IV_Rv primers for the amplification of apramycin cassette (777 bp). B) Bioassays against *B. subtilis* HB0933 in MHA medium. Complemented strains showed clear inhibition halos. Samples were collected at 144 h in FM2 medium and A40926 extraction was done by raising pH up to 12. 50 µL sample were loaded in a 6 mm paper disk.

Generation of recombinant strains for the pathway specific regulatory cross-talking between *N. gerenzanensis* and *A. teichomyceticus*

StrR-like proteins are orthologues in the *bal* (balhimycin), *tei* (teicoplanin) and *dbv* (A40926) BGCs. Bbr and Dbv4 share 80% of amino acid sequence identity, while Tei15* and Dbv4 share only 53% of amino acid sequence identity. Finally, Tei15* and Bbr share 50% identity of amino acid sequence identity. Instead, Dbv3 and Tei16* LuxR-regulator are not closely related sharing only 30% of amino acid sequence identity. To investigate the eventual cross-talking among the regulators controlling teicoplanin and A40926 production, we started exchanging StrR-like and LuxR-like genes from *N. gerenzanensis* and *A. teichomyceticus*. We used pSAD3 and pSAD4 plasmids previously constructed for the overproduction of A40926 in *N. gerenzanensis*[12] as well as pSET152Atei15* and pSET152Atei16* for the overproduction of teicoplanin in *A. teichomyceticus*.[8] All these plasmids were derivatives of pSET152A, where desired genes are under control of strong *aac(3)IVp*. Since *A. teichomyceticus* Δ*tei15** and *A. teichomyceticus* Δ*tei16** mutants are resistant to apramycin (having *tei15** and *tei16** replaced with *aac(3)IV-oriT* cassette)[8], we had to initially disrupt *aac(3)IV* gene in pSAD3 and pSAD4 introducing a hygromycin resistance cassette (*hygR*), derived from pIJ10700, and yielding the recombinant vectors pSHAD3 and pSHAD4 (Table 1). pSHAD4 and pSHAD3 were then transferred into *A. teichomyceticus* Δ*tei15** and *A. teichomyceticus* Δ*tei16**, respectively, by means of intergeneric conjugation with *E. coli* ET12567 pUZ8002. Resulting strains were called *A. teichomyceticus* Δ*tei15** pSHAD4$^+$ and *A. teichomyceticus* Δ*tei16** pSHAD3$^+$. The same

approach was performed for *N. gerenzanensis* Δ*dbv4* and Δ*dbv3,* where pSET152Atei15* and pSET152Atei16* (respectively) were transferred. Obtained strains were named *N. gerenzanensis* Δ*dbv3* pSET152Atei16*$^{+}$ and *N. gerenzanensis* Δ*dbv4* pSET152Atei15*$^{+}$. All recombinant strains were verified by PCR (Figure 4).

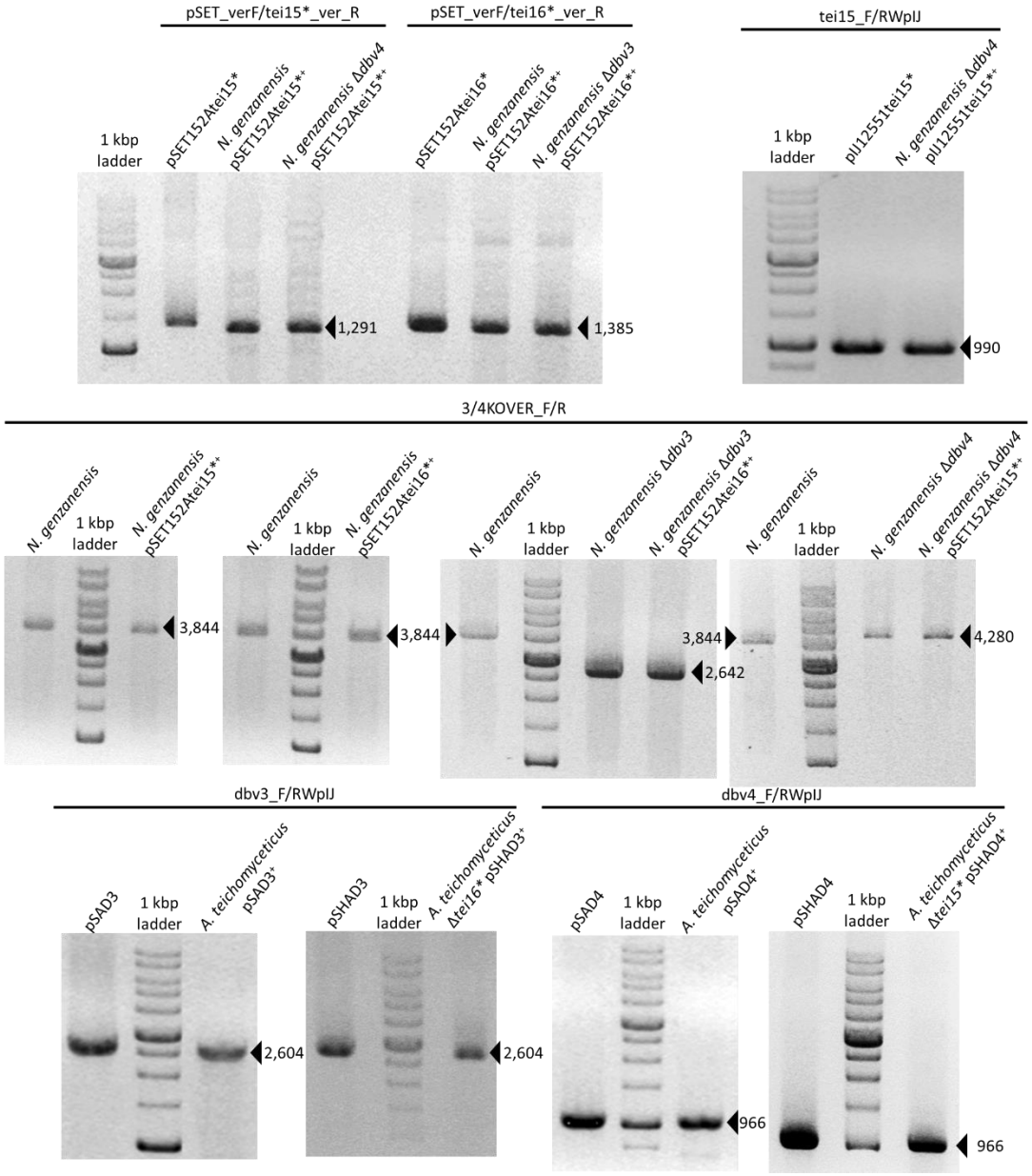


Figure 4. Verification of the recombinant strains. pSET_ver_F/tei15*_ver_R, pSET_ver_F/tei16*_ver_R, dbv3_F/RWpIJ, dbv4_F/RWpIJ, tei15_F/RWpIJ primer pairs were used to amplify an internal region of pSET152A and *tei15** (1,291 bp), internal region of pSET152A and *tei16** (1,385 bp), *dbv3* (2,604 bp), *dbv4* (966 bp) and *tei15** (990 bp), respectively. 3/4KOVER_F/R were used to check knockout (3,834 bp for *N. gerenzanensis*, 2,642 bp for *N. gerenzanensis* Δ*dbv3* and 4,840 bp for *N. gerenzanensis* Δ*dbv4*).

Pathway specific regulatory cross-talking for A40926 production in *N. gerenzanensis*

Once generated the mutants described above, we first studied A40926 production in *N. gerenzanensis Δdbv4* and *Δdbv3*, complemented with *tei15** and *tei16**. As a result, bioassays of culture broth extracts from the recombinants fermented in FM2 production medium showed small inhibition halos towards *B. subtilis* in case of *N. gerenzanensis* Δ*dbv3* pSET152Atei16*$^{+}$ and no antimicrobial activity for *N. gerenzanensis* Δ*dbv4* pSET152Atei15*$^{+}$. HPLC analysis showed no A40926 production in both *N. gerenzanensis* Δ*dbv3* pSET152Atei16*$^{+}$ and *N. gerenzanensis* Δ*dbv4* pSET152Atei15*$^{+}$ (Figure 5A). We also tried to complement *N. gerenzanensis* Δ*dbv4* with *tei15** expressed under the control of *ermEp* (pIJ12551tei15* plasmid, see previous section). *N. gerenzanensis* Δ*dbv4* pIJ12551tei15*$^{+}$ showed a small inhibition halo towards *B. subtilis* (Figure 5A), but, as previously, A40926 presence was not detectable by HPLC analyses.

To further investigate if the poor antimicrobial activity observed *N. gerenzanensis* Δ*dbv3* pSET152Atei16*$^{+}$ and in *N. gerenzanensis* Δ*dbv4* pIJ12551tei15*$^{+}$ was due to A40926 production under the limit of HPLC detection or instead to the unspecific activation of other silent BGCs, we constructed a new *N. gerenzanensis* Δ*dbv3-4* double knock-out mutant (Table 1 and Figure 1) following the same approach above described for the generation of *N. gerenzanensis* Δ*dbv3 and* Δ*dbv4*. We could not in fact rule out that, although Str-like and LuxR-like transcriptional factors are pathway-specific regulators, they could activate pleiotropically the expression of other still-unexplored BGCs present in *N. gerenzanensis* genome, whose products might have antimicrobial activity.[27] In the double knockout mutant, none antimicrobial activity on bioassays nor GPA detection in HPLC were noticed, like in *N. gerenzanensis* Δ*dbv3* and Δ*dbv4.* The introduction of *tei15** and *tei16** in the double knocked-out mutant (strains called *N. gerenzanensis* Δ*dbv3-4* pIJ12551tei15*$^{+}$ and *N. gerenzanensis* Δ*dbv3-4* pSET152Atei16*$^{+}$) did not trigger any antimicrobial activities, demonstrating that the low antimicrobial activity observed in both *N. gerenzanensis* Δ*dbv3* pSET152Atei16*$^{+}$ and in *N. gerenzanensis* Δ*dbv4* pIJ12551tei15*$^{+}$ was likely due to A40926 production, albeit in trace level. Thus, either *tei15** or *tei16** seemed able to induce A40926 production in *N. gerenzanensis* Δ*dbv4* and Δ*dbv3,* although to a negligibly low level.

To confirm the production of A40926 was triggered by the introduction of the heterologous *tei15** and *tei16**, we also tested the extracts of both *N. gerenzanensis* Δ*dbv3* pSET152Atei16*$^{+}$ and in *N. gerenzanensis* Δ*dbv4* pIJ12551tei15*$^{+}$ against a special recombinant strain of *B. subtilis* named HB0950 (Figure 5B). This strain in fact contains a *lacZ* gene (coding for β-galactosidase) fused to the *liaI* promoter (PliaI), which activates *lacZ* in response to the cell wall stress caused with lipid II binders.[24] GPA production can then be easily identified when *B. subtilis* HB0950 is grown on MHA medium containing X-Gal. GPAs, being a lipid II binders, induce the expression of β-galactosidase which hydrolyses X-Gal giving a green edge of the growth inhibition halo. Green halos were observed in the wild type *N. gerenzanensis* as well as in the *N. gerenzanensis* Δ*dbv3* pSET152Atei16*$^{+}$ and in *N. gerenzanensis* Δ*dbv4* pIJ12551tei15*$^{+}$ recombinants, indicating that they produce A40926 (Figure 5B).

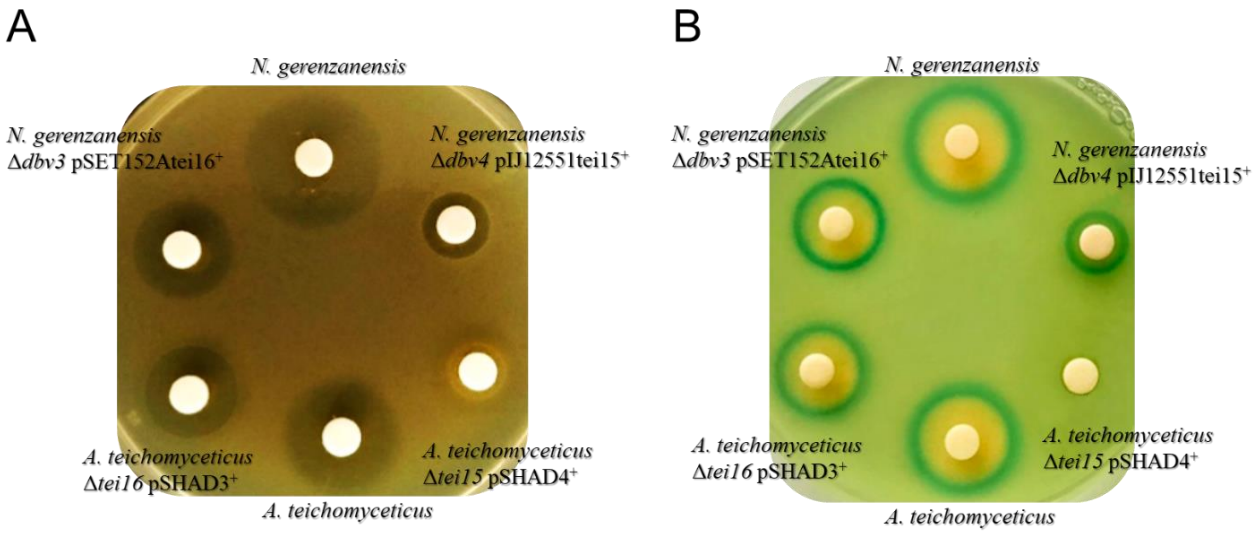


Figure 5. Complemented mutants of *N. gerenzanensis* and *A. teichomyceticus,* lacking CSR-genes, showed restored GPA production. Bioassays against *B. subtilis* HB0933 (A) and HB0950 (B) in MHA medium. Samples were collected at 144 h in FM2 medium in case of *N. gerenzanensis* and TM1 at 96 h in case of *A. teichomyceticus*. *A. teichomyceticus* Δ*tei16** pSHAD3^{+} sample was collected at 96 h in E25 + 1g/L L-valine. GPA extraction was done by raising pH up to 12. 50 µL sample were loaded in a 6 mm paper disk. Green halo in *B. subtilis* HB0950 bioassay indicated the presence of a GPA.

Finally, to quantify A40926 production in *N. gerenzanensis* Δ*dbv3* pSET152Atei16*$^{+}$ and in *N. gerenzanensis* Δ*dbv4* pIJ12551tei15*$^{+}$, extracts from their liquid cultures at the time of maximum production were concentrated by lyophilization and then analyzed by HPLC. It was thus possible to estimate that *N. gerenzanensis* Δ*dbv3* pSET152Atei16*$^{+}$ produced ca. 20 mg/L A40926, whereas *N. gerenzanensis* Δ*dbv4* pIJ12551tei15*$^{+}$only 2 mg/L (Figure 6), confirming that heterologous expression of *tei15** and *tei16**could somehow restore A40926 production in *N. gerenzanensis.*

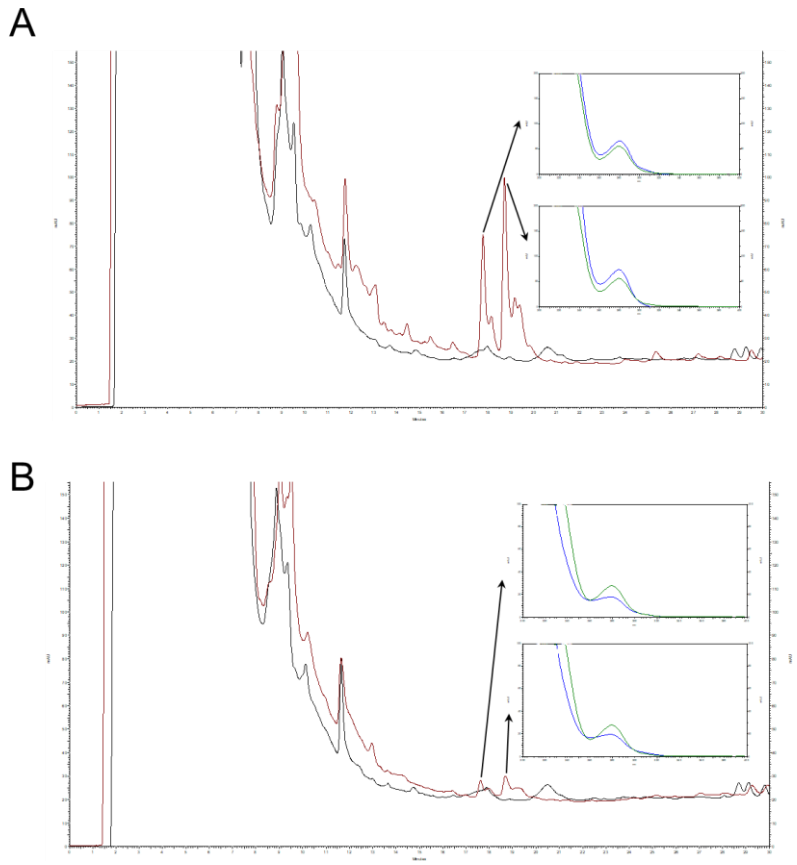


Figure 6. Chromatograms of concentrated samples. *N. gerenzanensis* Δ*dbv3* pSET152Atei16*$^{+}$ (A) and *N. gerenzanensis* Δ*dbv4* pIJ12551tei15*$^{+}$ (B) showed A40926 peaks with its typical UV spectrum. Recombinant strains are showed in red and corresponding knockout in black. Absorption spectrum of A40926 standard is represented in green and absorption spectrum of the peaks identified in the recombinant strains in blue.

Pathway specific regulatory cross-talking for teicoplanin production in *A. teichomyceticus*

A complementary approach was used in *A. teichomyceticus* mutants knocked out in *tei16** and *tei15**, which were complemented with *dbv3* or *dbv4,* respectively. The recombinant strains were grown in the traditional TM1 production medium, previously optimized for teicoplanin production.[16] In bioassays, none antimicrobial activity was detectable versus *B.subtilis* in the case of *A. teichomyceticus* Δ*tei15** pSHAD4^{+} (Figure 5A), while *A. teichomyceticus* Δ*tei16** pSHAD3^{+} did not even grow in TM1 production medium. As in the case of *N. gerenzanensis*, we also tried to complement *A. teichomyceticus Δtei15** with pIJ12551dbv4 where *dbv4* was cloned under the control of *ermEp* (unpublished data), but none antimicrobial activity was observed, too. To overcome these aspects, different combinations of vegetative and production media – previously used for growing other GPA producing strains – were tested such as modified TM1, ISP2, SFM, VSP or the vegetative media E25 with 1 g/L L-valine, which is the precursor amino acid contributing to the increased production of teicoplanin in production media.[26] In none of these conditions, mutants complemntedwith the heterologous *dbv4* produced any antimicrobial activity. Surprisingly, bioassays of *A. teichomyceticus*

Δ*tei16** pSHAD3^{+} in vegetative medium revealed the production of a significant antimicrobial activity. When the strain was grown in E25 vegetative medium with additional 1 g/L of L-valine, HPLC analysis showed that it produced approximately 150 mg/L teicoplanin, despite the relatively poor growth (Figure 7). Although further investigations are needed, we speculated that a strong and not timely controlled heterologous expression of *dbv3* might cause an overproduction of teicoplanin in the vegetative step, which then likely hampers the further growth of the mutant in the production medium. Confirming the data above reported, when the recombinant extracts were tested against *B. subtilis* HB0950, green halos were observed for *A. teichomyceticus* wild type and *A. teichomyceticus Δtei16** pSHAD3^{+} indicating GPA production, but not in Δ*tei15** pSHAD4^{+} which did not show any antimicrobial activity (Figure 5B). The extract of this last strain was concentrated by lyophilization and tested again for the microbiological activity and in HPLC analysis, but the negative results confirmed that *dbv4,* differently from *dbv3,* was unable to restore teicoplanin production in *A. teichomyceticus.*

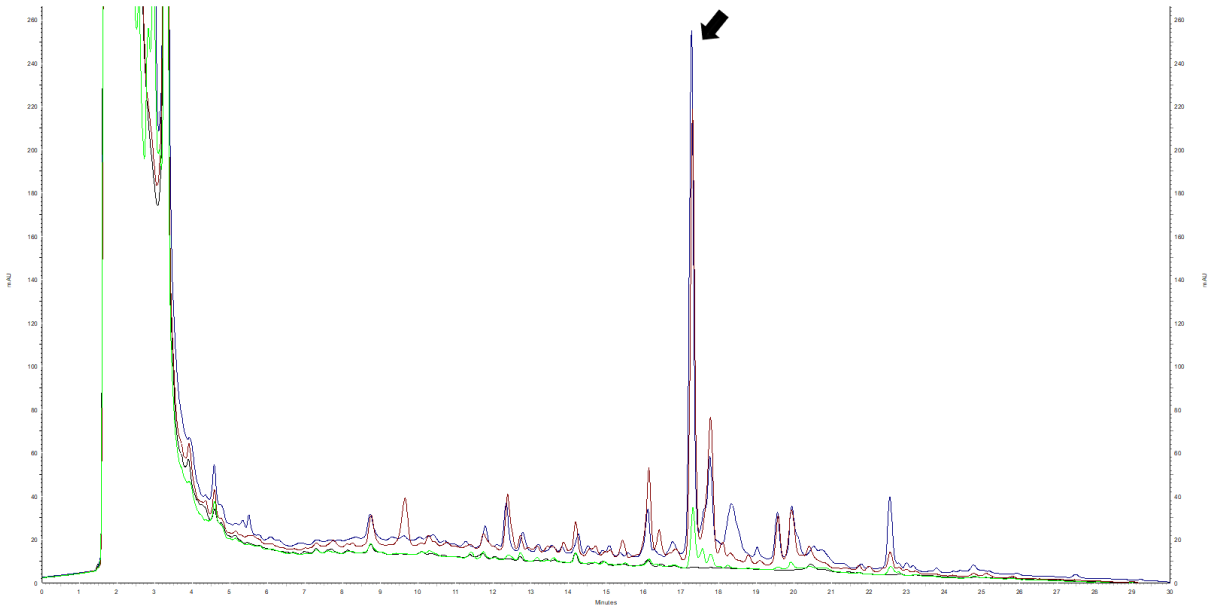


Figure 7. HPLC revealed that *A. teichomyceticus* Δ*tei16** pSHAD3^{+} restored teicoplanin production. Blue chromatogram shows the production profile of *A. teichomyceticus* in TM1 at 96 h, in red *A. teichomyceticus* Δ*tei16** pSHAD3^{+} in E25 + 1g/L L-valine at 96 h, in green *A. teichomyceticus* in E25 + 1g/L L-valine at 96 h and in black *A. teichomyceticus* Δ*tei16** in TM1 at 96h. Teicoplanin main peak is indicated with a black arrow.

Cross-overexpression of *dbv* and *tei* pathway-specific regulators in wild type strains of *A. teichomyceticus* and *N. gerenzanensis*

Since previous data indicated that *tei15** and *tei16** restored A40926 production in *N. gerenzanensis* Δ*dbv3* and Δ*dbv4,* respectively, we tested if their overexpression in *N. gerenzanensis* wild type would affect antibiotic production. The result was that *N. gerenzanensis* pSET152tei15*$^{+}$, *N. gerenzanensis* pIJ12551tei15*$^{+}$ and *N. gerenzanensis* pSET152tei16*$^{+}$antibiotic productivity in FM2 production medium was the same as in the wild type (Figure 8A). When the same approach was used for *A. teichomyceticus* pSAD4^{+} and *A. teichomyceticus* pSAD3^{+}, we found that the first carrying *dbv4* produced as the wild type whereas the second with the *dbv3* approximately 3 times more (Figure 8 B), although its growth was reduced generating half biomass than the wild type (data not shown).

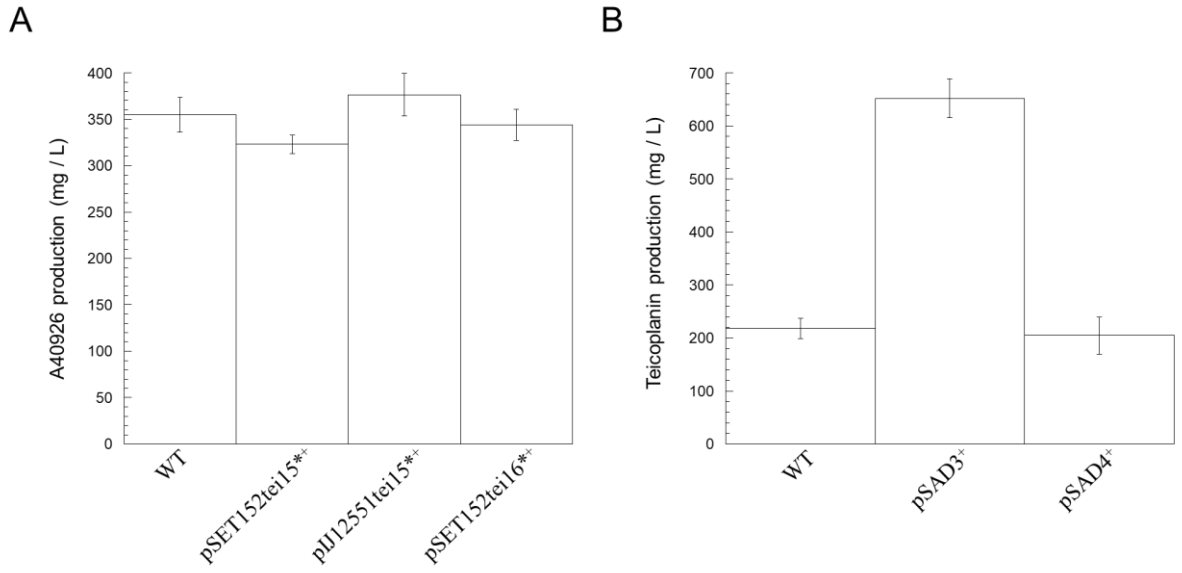


Figure 8. GPA production in cross-overexpression recombinants. A) Maximum A40926 production in FM2 by *N. gerenzanensis*, *N. gerenzanensis* pSET152tei15*$^{+}$, *N. gerenzanensis* Δ*dbv4* pIJ12551tei15*$^{+}$ and *N. gerenzanensis* pSET152tei16*$^{+}$ B) Maximum teicoplanin production in TM1 by *A. teichomyceticus*, *A. teichomyceticus* pSAD3^{+} and *A. teichomyceticus* pSAD4^{+} overexpressing heterologous *dbv3* and *dbv4,* respectively.

Discussion

Previous works reported that StrR-like CSR from different GPA BGCs were able to "cross-talk" between different pathways. StrR-like CSRs Bbr and Ajr (the first from the balhimycin BGC in *Am. balhimycina* and the second from ristocetin producer *Am. japonicum*) could replace each other in governing the reciprocal GPA synthesis.[28, 29] They shared more than 80% of amino acid sequence identity, likely because they originate from species belonging to the same genus. Similarly, *nocRI* from *N. coxensis* was able to increase A40926 production in *N. gerenzanensis*, being 94% similar to *dbv3*.[12] In our study we were going wider and wanted to test whether *dbv4* and *tei15** (sharing only 53% amino acid sequence identity) would be able to "cross-talk" between pathways coming

from species belonging to different families of GPA producers – *Micromonosporaceae* in case of teicoplanin and *Streptosporangiaceae* in case of A40926. Our results did not show a reciprocal similar behavior of the two StrR-like CSRs, since *tei15** was able to restore A40926 production in *N. gerenzanensis* Δ*dbv4*, albeit at very low extent giving less than 1% of A40926 production than in the wild type, whereas at the same time, *dbv4* was not able to restore teicoplanin production in *A. teichomyceticus* Δ*tei15**.

To this purpose, it is important to note that StrR-binding regions [GTCCAa(N)$_{17}$TtGGAC] are not exclusively found in GPA BGCs. These consensus sequences were also described in other clusters present in non-producing GPA strains.[30,31] In fact, StrR-like regulators control the synthesis of a large variety of specialized metabolites in actinobacteria such as different sorts of antibiotics, enedyines or angucyclines, accordingly with MIBiG database.[32] On the other hand, it is also true that they were found in less than 50 actinobacterial BGCs. Thus, we cannot exclude that *dbv4* and *tei15** may not be orthologous proteins, less functionally related than believed before. While being both StrR-like regulators, they share basic architecture of structural domains, as well as high degree of amino acid sequence similarity, at the same time not having DNA-binding activity shaped to achieve sufficient "cross-talk". Anyhow, further investigations are needed on their mode of action at transcriptional level. A high-resolution phylogenetic analysis of all StrR-like proteins coming from Actinobacteria phylum might be also useful to understand better their evolution and reciprocal relations.

Emerging picture with LuxR-like CSR regulators appeared even to be more puzzling. Dbv3 and Tei16* are clearly not related to each other, sharing only 30% of amino acid sequence identity and possessing dramatically different regulons. Surprisingly, our results revealed that complemented strain *A. teichomyceticus* Δ*tei16** pSHAD3^{+} had teicoplanin production fully restored, while *N. gerenzanensis* Δ*dbv3* pSET152tei16*$^{+}$ was able to produce only a couple of mg/L of A40926. We also achieved a significant increase in teicoplanin production when *dbv3* was overexpressed in the wild type *A. teichomyceticus*.

LuxR-like transcriptional factors positively regulate different antibiotic BGCs. They control the synthesis of different compounds in Gram-negative bacteria, being also involved in quorum-sensing.[33] Moreover, LuxR-like regulators are necessary to modulate the expression of virulence factors, immune response, intracellular signaling or biofilms formation in Gram-negative bacteria.[34] However, they are still poorly characterized for Gram-positive bacteria. In addition, LuxR might supposedly function as single transcription factors or as part of two-component systems in actinobacteria.[34] Since LuxR are abundant in nature and they have an important role in a wide variety of BGCs in actinobacteria, it is reasonable to assume that they could "cross-talk" in similar BGCs. Therefore, they might be less specific than StrR-like transcriptional factors explaining why *dbv3* has significant positive effect in teicoplanin biosynthesis.

Another interesting point is that overexpression of *dbv3* generated significant stress preventing *A. teichomyceticus* from a normal growth. *A. teichomyceticus* Δ*tei16** pSHAD3^{+} did not grow in TM1 industrial medium after inoculation from vegetative medium, where it started to produce the antibiotic. This early activation of antibiotic production can be the cause of the premature death of the culture. The maximum production observed in E25 with the addition of L-valine was *ca.*150 mg/mL of teicoplanin at 96 h of cultivation, which is comparable with the maximum production of *A. teichomyceticus* in the TM1 industrial medium. Finally, overexpression of *dbv3* in *A. teichomycetius* wild type allowed to reach almost 700 mg/L of teicoplanin, three time more than wild type. However, this strain had a poor growth, generating half the biomass compared to wild type. These results could be somehow correlated with our previous

results about the overexpression of *dbv3* into *N. gerenzanensis*, which produced two times more A40926 than wild type in flask fermentation, but when we grew it in a 3-L fermenter, it had a poor growth and hardly produced A40926.[12]

To conclude, if *tei15** and *dbv4* are only partially able (*tei15**) or unable (*dbv4*) to "cross-talk" between the pathways, *dbv3* and *tei16** appear more active in influencing heterologous pathways. In particular *dbv3* exerts a significant effect on teicoplanin production in *A. teichomyceticus* and a still–to-be-clarified impact on its growth. Also, we discussed some speculative explanations of this phenomenon, but molecular background of these events is yet to be elucidated.

References:

(1) van der Meij, A.; Worsley, S. F.; Hutchings, M. I.; van Wezel, G. P. Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol Rev* **2017**, *41* (3), 392-416. DOI: 10.1093/femsre/fux005.

(2) Marcone, G. L.; Binda, E.; Berini, F.; Marinelli, F. Old and new glycopeptide antibiotics: From product to gene and back in the post-genomic era. *Biotechnol Adv* **2018**, *36* (2), 534-554. DOI: 10.1016/j.biotechadv.2018.02.009.

(3) Binda, E.; Marinelli, F.; Marcone, G. L. Old and new glycopeptide antibiotics: Action and resistance. *Antibiotics (Basel)* **2014**, *3* (4), 572-594. DOI: 10.3390/antibiotics3040572.

(4) Ohnishi, Y.; Yamazaki, H.; Kato, J. Y.; Tomono, A.; Horinouchi, S. AdpA, a central transcriptional regulator in the A-factor regulatory cascade that leads to morphological development and secondary metabolism in *Streptomyces griseus*. *Biosci Biotechnol Biochem* **2005**, *69* (3), 431-439. DOI: 10.1271/bbb.69.431.

(5) Uguru, G. C.; Stephens, K. E.; Stead, J. A.; Towle, J. E.; Baumberg, S.; McDowall, K. J. Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. *Mol Microbiol* **2005**, *58* (1), 131-150. DOI: 10.1111/j.1365-2958.2005.04817.x.

(6) Higo, A.; Horinouchi, S.; Ohnishi, Y. Strict regulation of morphological differentiation and secondary metabolism by a positive feedback loop between two global regulators AdpA and BldA in *Streptomyces griseus*. *Mol Microbiol* **2011**, *81* (6), 1607-1622. DOI: 10.1111/j.1365-2958.2011.07795.x.

(7) Shawky, R. M.; Puk, O.; Wietzorrek, A.; Pelzer, S.; Takano, E.; Wohlleben, W.; Stegmann, E. The border sequence of the balhimycin biosynthesis gene cluster from *Amycolatopsis balhimycina* contains bbr, encoding a StrR-like pathway-specific regulator. *J Mol Microbiol Biotechnol* **2007**, *13* (1-3), 76-88. DOI: 10.1159/000103599.

(8) Horbal, L.; Kobylyanskyy, A.; Truman, A. W.; Zaburranyi, N.; Ostash, B.; Luzhetskyy, A.; Marinelli, F.; Fedorenko, V. The pathway-specific regulatory genes, *tei15** and *tei16**, are the master switches of teicoplanin production in *Actinoplanes teichomyceticus*. *Appl Microbiol Biotechnol* **2014**, *98* (22), 9295-9309. DOI: 10.1007/s00253-014-5969-z.

(9) Yushchuk, O.; Horbal, L.; Ostash, B.; Marinelli, F.; Wohlleben, W.; Stegmann, E.; Fedorenko, V. Regulation of teicoplanin biosynthesis: refining the roles of tei cluster-situated regulatory genes. *Appl Microbiol Biotechnol* **2019**, *103* (10), 4089-4102. DOI: 10.1007/s00253-019-09789-w.

(10) Lo Grasso, L.; Maffioli, S.; Sosio, M.; Bibb, M.; Puglia, A. M.; Alduina, R. Two master switch regulators trigger A40926 biosynthesis in *Nonomuraea* sp. strain ATCC 39727. *J Bacteriol* **2015**, *197* (15), 2536-2544. DOI: 10.1128/JB.00262-15.
(11) Alduina, R.; Sosio, M.; Donadio, S. Complex regulatory networks governing production of the glycopeptide A40926. *Antibiotics (Basel)* **2018**, *7* (2). DOI: 10.3390/antibiotics7020030.
(12) Yushchuk, O.; Andreo-Vidal, A.; Marcone, G. L.; Bibb, M.; Marinelli, F.; Binda, E. New molecular tools for regulation and improvement of A40926 glycopeptide antibiotic production in *Nonomuraea gerenzanensis* ATCC 39727. *Front Microbiol* **2020**, *11*, 8. DOI: 10.3389/fmicb.2020.00008.
(13) Waglechner, N.; McArthur, A. G.; Wright, G. D. Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nat Microbiol* **2019**, *4* (11), 1862-1871. DOI: 10.1038/s41564-019-0531-5.
(14) Alduina, R.; Lo Piccolo, L.; D'Alia, D.; Ferraro, C.; Gunnarsson, N.; Donadio, S.; Puglia, A. M. Phosphate-controlled regulator for the biosynthesis of the dalbavancin precursor A40926. *J Bacteriol* **2007**, *189* (22), 8120-8129. DOI: 10.1128/JB.01247-07.
(15) Marcone, G. L.; Binda, E.; Carrano, L.; Bibb, M.; Marinelli, F. Relationship between glycopeptide production and resistance in the actinomycete *Nonomuraea* sp. ATCC 39727. *Antimicrob Agents Chemother* **2014**, *58* (9), 5191-5201. DOI: 10.1128/AAC.02626-14.
(16) Taurino, C.; Frattini, L.; Marcone, G. L.; Gastaldo, L.; Marinelli, F. *Actinoplanes teichomyceticus* ATCC 31121 as a cell factory for producing teicoplanin. *Microb Cell Fact* **2011**, *10*, 82. DOI: 10.1186/1475-2859-10-82.
(17) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. *Practical streptomyces genetics*; John Innes Foundation Norwich, **2000**.
(18) Gust, B.; Chandra, G.; Jakimowicz, D.; Yuqing, T.; Bruton, C. J.; Chater, K. F. Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Adv Appl Microbiol* **2004**, *54*, 107-128. DOI: 10.1016/S0065-2164(04)54004-2.
(19) Sherwood, E. J.; Hesketh, A. R.; Bibb, M. J. Cloning and analysis of the planosporicin lantibiotic biosynthetic gene cluster of *Planomonospora alba*. *J Bacteriol* **2013**, *195* (10), 2309-2321. DOI: 10.1128/JB.02291-12.
(20) Bierman, M.; Logan, R.; O'Brien, K.; Seno, E. T.; Rao, R. N.; Schoner, B. E. Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **1992**, *116* (1), 43-49. DOI: 10.1016/0378-1119(92)90627-2.
(21) Gust, B.; Kieser, T.; Chater, K. PCR targeting system in *Streptomyces coelicolor* A3 (2). *John Innes Centre* **2002**, *3* (2), 1-39.
(22) Marcone, G. L.; Beltrametti, F.; Binda, E.; Carrano, L.; Foulston, L.; Hesketh, A.; Bibb, M.; Marinelli, F. Novel mechanism of glycopeptide resistance in the A40926 producer *Nonomuraea* sp. ATCC 39727. *Antimicrob Agents Chemother* **2010**, *54* (6), 2465-2472. DOI: 10.1128/AAC.00106-10.
(23) Mascher, T.; Margulis, N. G.; Wang, T.; Ye, R. W.; Helmann, J. D. Cell wall stress responses in *Bacillus subtilis*: the regulatory network of the bacitracin stimulon. *Mol Microbiol* **2003**, *50* (5), 1591-1604. DOI: 10.1046/j.1365-2958.2003.03786.x.
(24) Mascher, T.; Zimmer, S. L.; Smith, T. A.; Helmann, J. D. Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of *Bacillus subtilis*. *Antimicrob Agents Chemother* **2004**, *48* (8), 2888-2896. DOI: 10.1128/AAC.48.8.2888-2896.2004.
(25) Marcone, G. L.; Foulston, L.; Binda, E.; Marinelli, F.; Bibb, M.; Beltrametti, F. Methods for the genetic manipulation of *Nonomuraea* sp. ATCC 39727. *J Ind Microbiol Biotechnol* **2010**, *37* (10), 1097-1103. DOI: 10.1007/s10295-010-0807-5.

(26) Yushchuk, O.; Vior, N. M.; Andreo-Vidal, A.; Berini, F.; Rückert, C.; Busche, T.; Binda, E.; Kalinowski, J.; Truman, A. W.; Marinelli, F. Genomic-led discovery of a novel glycopeptide antibiotic by *Nonomuraea coxensis* DSM 45129. *ACS Chem Biol* **2021**, *16* (5), 915-928. DOI: 10.1021/acschembio.1c00170.
(27) D'Argenio, V.; Petrillo, M.; Pasanisi, D.; Pagliarulo, C.; Colicchio, R.; Talà, A.; de Biase, M. S.; Zanfardino, M.; Scolamiero, E.; Pagliuca, C.; et al. The complete 12 Mb genome and transcriptome of *Nonomuraea gerenzanensis* with new insights into its duplicated "magic" RNA polymerase. *Sci Rep* **2016**, *6* (1), 18. DOI: 10.1038/s41598-016-0025-0.
(28) Spohn, M.; Kirchner, N.; Kulik, A.; Jochim, A.; Wolf, F.; Muenzer, P.; Borst, O.; Gross, H.; Wohlleben, W.; Stegmann, E. Overproduction of ristomycin A by activation of a silent gene cluster in *Amycolatopsis japonicum* MG417-CF17. *Antimicrob Agents Chemother* **2014**, *58* (10), 6185-6196. DOI: 10.1128/AAC.03512-14.
(29) Liu, K.; Hu, X. R.; Zhao, L. X.; Wang, Y.; Deng, Z.; Tao, M. Enhancing ristomycin A production by overexpression of ParB-Like StrR family regulators controlling the biosynthesis genes. *Appl Environ Microbiol* **2021**, *87* (19), e0106621. DOI: 10.1128/AEM.01066-21.
(30) Retzlaff, L.; Distler, J. The regulator of streptomycin gene expression, StrR, of *Streptomyces griseus* is a DNA binding activator protein with multiple recognition sites. *Mol Microbiol* **1995**, *18* (1), 151-162. DOI: 10.1111/j.1365-2958.1995.mmi_18010151.x.
(31) van der Heul, H. U.; Bilyk, B. L.; McDowall, K. J.; Seipke, R. F.; van Wezel, G. P. Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. *Nat Prod Rep* **2018**, *35* (6), 575-604. DOI: 10.1039/c8np00012c.
(32) Kautsar, S. A.; Blin, K.; Shaw, S.; Navarro-Muñoz, J. C.; Terlouw, B. R.; van der Hooft, J. J. J.; van Santen, J. A.; Tracanna, V.; Suarez Duran, H. G.; Pascal Andreu, V.; et al. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res* **2020**, *48* (D1), D454-D458. DOI: 10.1093/nar/gkz882.
(33) Nealson, K. H.; Platt, T.; Hastings, J. W. Cellular control of the synthesis and activity of the bacterial luminescent system. *J Bacteriol* **1970**, *104* (1), 313-322. DOI: 10.1128/jb.104.1.313-322.1970.
(34) Santos, C. L.; Correia-Neves, M.; Moradas-Ferreira, P.; Mendes, M. V. A walk into the LuxR regulators of Actinobacteria: phylogenomic distribution and functional diversity. *PLoS One* **2012**, *7* (10), e46758. DOI: 10.1371/journal.pone.0046758.

V. GENERAL CONCLUSIONS

Nowadays, bacterial infections are rather unconcerned diseases although they were the primary causes of death around the world at the beginning of 20th century. This important shift of paradigm was due to important advances in science and medicine such as the discovery and use of antibiotics. However, over- and misuse contributed to the spread of antibiotic resistance in bacteria, most importantly pathogens. In fact, according to recent predictions, we will return to the 19th century-like scenario in only few decades, where nosocomial diseases might strike painfully once again.

So far, one of the biggest challenges in modern science and medicine is the development of new and more effective drugs. Thus, my PhD course was focused on the study of GPAs, which are used clinically as the last line drugs towards Gram-positive bacteria, in case first line antibiotics are not effective anymore. Currently, there are only five GPAs used in clinic. The first-discovered natural products are vancomycin and teicoplanin, used for more than 40 years. Unfortunately, effectiveness of these compounds has been reduced in the last years due to the emergence ofresistant enterococci and staphylococciamong pathogens. Additionally, there are semi-synthetic GPA molecules telavancin, dalbavancin and oritavancin, approved in the last decade, with improved antimicrobial spectrum and potency.

One of the main problems about the discovery of new antibiotics is the difficulty to find new ones, due to significant rediscovery rates. Moreover, the success during the second half of the 20th century decreased the danger coming from infectious diseases, causing companies to dramatically reduce investment in the research of new antibiotics. Another great difficulty was always the economically feasible largescale industrial production of antibiotics, demanded by the market.

Consequently, this PhD thesis contributes to all these critic issues.

First, we have developed a set of novel genetic tools for gene expression in *Nonomuraea* spp. By constructing a collection of promoter-probe vectors we have found the most optimal promoters for the heterologous gene expression in *Nonomuraea* spp. Belonging to so-called "rare" actinobacteria, there were only few tools available for the genetic-engineering of the *Nonomuraea* spp. Thus, the strongest constitutive promoter – *aac(3)IVp* – was successfully used to overexpress A40926 BGC (*dbv*) regulatory genes. As a result, a remarkable increase of GPA production was obtained when we overexpressed *dbv3* and *dbv4* in *N. gerenzanensis*. In course of this work, we have also studied whether there are orthologues of *dbv3* and *dbv4*, which might also be used as tools for the improvement of A40926 production. Such were found in *Nonomuraea coxensis* DSM 45129, and they were within a novel GPA BGCs, similar to *dbv.* Overexpression of *dbv3* orthologue from *N. coxensis* improved A40926 production.

Pursuing the idea that a novel GPA might be produced in *N. coxensis,* we have studied this strain *in vivo* and *in silico.* A novel GPA was thus discovered, named A50926. It appeared to be similar to A40926 – the precursor of commercial antibiotic dalbavancin - although lacking the carboxyl group on the *N*-acylglucosamine moiety. This structural difference correlates with the absence of *dbv29*-orthologue within the A50926 BGC, which codes for an enzyme responsible for the oxidation of the *N*-acylglucosamine moiety. In fact, introduction of *dbv29* from *N. gerenzanensis* into *N. coxensis* led to A40926 production in the latter, creating a novel A40926 production platform. Peculiarly, absence of the carboxyl group on the *N*-acylglucosamine moiety might be making A50926 more effective than A40926, according to previous results published on the chemical derivative.

Importantly, we have also studied the relations of cluster-situated transcriptional regulators coming from different GPA pathways. We were able to demonstrate that in the case of unrelated *tei16**/*dbv3* LuxR regulators, they were able to induce GPA synthesis

in *N. gerenzanensis* and *A. teichomyceticus* mutants where the native LuxR was inactivated. Expression of *dbv3* in *A. tecihomyceticus* lead to significant increase in teicoplanin production. In contrast, complementation with orthologous StrR regulators revealed that *tei15** was able to restore A40926 production in *N. gerenzanensis* strains lacking *dbv4*, while *dbv4* was not able to activate teicoplanin production in *A. teichomyceticus* where *tei15** was inactivated. This shows that the interactions of these regulators are complex and further studies are required to understand their role completely.

Finally, keeping in mind an alarming spread of antibiotic microbial resistance (AMR) genes in pathogens, we performed an in-depth bioinformatics analysis searching for the genes involved in GPA resistance, coming from Actinobacteria phylum. These so-called *van* genes were found widely spread within the Actinobacteria phylum, either in GPA-producers or non-producing bacteria – occurring much more often in GPA non-producers. Phylogenetic reconstructions made in this work highlighted that GPA BGCs and *van*-genes have undergone complicated co-evolution history. This makes actinobacteria the most likely primary source of *van*-genes, eventually able to migrate to pathogens, contributing to the problem of AMR. Our results might be useful in the future surveillance of emerging mechanisms of resistance to clinically used GPAs. Another important outcome of this screening was the discovery of multiple novel GPA BGCs, that will merit further efforts aimed at discovery novel GPAs.

To conclude, this thesis significantly contributes to our knowledge of a clinically-important antibiotic class – GPAs. A novel GPA was described in course of the study, while multiple other novel GPA BGCs discovered here await further experimental evaluation. Novel tools were set up for the rational engineering of *N. gerenzanensis* producing A40926 and of *A. teichomyceticus* producing teicoplanin. Improved strains were generated which could be used for developing more productive processes. Elucidating the details of pathway-specific regulation of GPA production might contribute to develop novel platforms for their combinatorial biosynthesis as well as for their more convenient production. Finally, and in-depth analysis of GPA resistant determinants allowed to clarify the evolution and spread of GPA resistance within Actinobacteria phylum. Results summarised in this thesis pave new venues for novel GPA discovery, combinatorial biosynthesis of rationally designed GPAs and effective GPA overproducing platforms.