Analysis of the angucycline biosynthetic gene cluster in Streptomyces sp. QL37 and implications for lugdunomycin production

Heul, H.U. van der

Phylogenetic and transcriptional analysis of the biosynthetic gene cluster for angucyclines and lugdunomycin in *Streptomyces* sp. QL37

Helga U. van der Heul
Changsheng Wu
Guillermo Guerrero Egido
Somayah S. Elsayed
Victor J. Carrión Bravo
Gilles P. Van Wezel

P art of this chapter is published as:
Wu, C., van der Heul, H.U., Melnik, A.V., Lubben, J., Dorrestein, P.C., Minnaard, A.J., Choi, Y. H., Wezel, G.P. (2019) Lugdunomycin, an angucycline-derived molecule with unprecedented chemical architecture. *Angew Chem Int Ed Engl* **58**: 2809–2814

ABSTRACT

Angucyclines form the largest group of polycyclic aromatic polyketides. These extensively studied compounds are well known for their anti-tumor and antibacterial activities. Previously we discovered the extensively rearranged angucycline, lugdunomycin, which is produced by *Streptomyces* sp. QL37. Despite high production of the angucyclines, lugdunomycin is produced in low amounts under all studied growth conditions. To obtain more insight into the mode of action of lugdunomycin, obtaining higher titers is important. Here, we study the polyketide type II gene cluster for angucyclines (designated *lug*) in *Streptomyces* sp. QL37, using targeted mutagenesis, RNA sequencing and phylogeny. This revealed that the *lug* gene cluster drives the production of non-rearranged and C-ring rearranged angucyclines and is also required for lugdunomycin biosynthesis. Transcriptomics in combination with bioinformatic comparisons of the *lug* gene cluster with other angucycline BGCs indicate that the *lug* gene cluster likely comprises 28 genes. Two oxygenase genes (*lugOIII* and *lugOIV*) and one gene with an unknown product (*lugX*) were not detected in BGCs directing the production of non-rearranged and A- and B-ring rearranged angucyclines, suggesting possible novel enzymology. Orthologues of *lugF–lugE*, genes required for the production of the angucycline backbone were found in some 25% out of 1020 searched Actinobacterial genomes, suggesting that angucycline production is widely spread. Angucycline BGCs with similar architecture as *lug* were identified in at least 36 genomes; these strains may be a good source for the discovery of similar novel rearranged angucyclines.

INTRODUCTION

Polyketides are one of the largest family of natural products, and many compounds or semisynthetic derivatives thereof are clinically relevant as antibiotics (oxytetracycline, erythromycin, monensin), anticancer agents (doxorubicin, mithramycin), immunosuppressants (FK506), antifungals (amphotericin B), cholesterol lowering agents (lovastatin) and antiparasitic agents (Ivermectin) (Tibrewal & Tang, 2014, Zhang *et al.*, 2006). Though polyketides are synthesised by multiple kingdoms of life, the above mentioned molecules are mostly produced by *Streptomyces*, a genus of Gram-positive bacteria, characterised by their filamentous growth and spore formation (Risdian *et al.*, 2019). Streptomycetes harbour different types of polyketide synthases (PKS I, PKS II and PKS III) (Hertweck *et al.*, 2007) that form distinct core structures, using small organic acids as substrates. The generated framework is further modified by post-PKS tailoring enzymes, resulting in final products with a remarkable chemical diversity (Patrikainen, 2015).

In our search for novel natural products, we previously discovered the extensively rearranged polyketide lugdunomycin produced by *Streptomyces* sp. QL37, which was isolated from the Qinling mountains in China (Wu *et al.*, 2019). Besides lugdunomycin, a variety of angucyclines is produced by this strain (Wu *et al.*, 2019), these include non-rearranged angucyclines (**2**–**6**) and rearranged angucyclines, such as the limamycins (**8**–**9**). Recently the structure of one of the limamycin derivatives with $[M+H]^+ = 334.11$ was revised to the structure of pratensilin A (**7**) (Figure 1)(Mikhaylov *et al.*, 2021, Wu *et al.*, 2019). The postulated biosynthesis pathway of lugdunomycin includes C-ring cleavage of a non-rearranged angucycline by Baeyer–Villiger oxidation, followed by a structural rearrangement, and a proposed Diels-Alder reaction to mediate the reaction between the modified angucycline as a diene and *iso*-maleimycin as a dienophile (Wu *et al.*, 2019, Uiterweerd, 2020). The new structural features of lugdunomycin not only make the molecule interesting in terms of drug discovery, but also imply potentially novel enzymology involved in its biosynthesis (Wu *et al.*, 2019). This may be used for the modification of other molecules. However, lugdunomycin is produced in extremely low amounts, and 7.5 L of agar was required to obtain only 0.5 mg of the compound (Wu *et al.*, 2019). Accordingly, it is vital to increase the production of lugdunomycin, to obtain more knowledge on its biological activity.

Angucyclines exhibit a variety of biological activities, including antitumor, antibacterial and enzyme inhibitory activities (Kharel *et al.*, 2012, Korynevska *et al.*, 2007, Yixizhuoma *et al.*, 2017, Fakhruzzaman *et al.*, 2015). These molecules are polycyclic aromatic polyketides, synthesised by type II polyketide synthases (Kharel *et al.*, 2012). The polyketide carbon chain is produced by the so-called minimal PKS, consisting of an heterodimer of ketosynthases, KSα and KSβ, and an acyl carrier protein (ACP) which catalyse consecutive Claisen condensations of acetyl-CoA and malonyl-CoA into a long polyketide chain, which is further modified by ketoreductases and cyclases, resulting in a framework featuring a four ring system (Risdian *et al.*, 2019). For the production of angucyclines, a distinct cyclase is required that folds the fourth ring in an angular position (Sharif & O'Doherty, 2012). This cyclase differentiates the angucyclines, with a typical polycyclic aromatic benz[a]anthracene structure, from the anthracyclines (e.g. doxorubicin) (Metsa-Ketela *et al.*, 2003). The genes encoding the minimal PKS are situated in a biosynthetic gene cluster (BGC) that further contains modification-, transporter- and regulatory genes, essential for structural diversity, resistance and the control of gene expression in response to environmental and intracellular signals (Yushchuk *et al.*, 2019, Zou *et al.*, 2014).

Biosynthesis pathways of angucyclines have been widely studied *in vivo* as well as *in vitro*; especially the early post-PKS modification steps are very well known (Kharel & Rohr, 2012). Oxygenases are crucial modification enzymes and differentiate angucyclines in non-rearranged (typical) and rearranged (atypical) angucyclines (Fan & Zhang, 2018). Non-rearranged angucyclines are exemplified by urdamycins, landomycins and the angucyclines **2**–**6**, and are specified by the typical benz[*a*] anthracene skeleton. Rearranged (atypical) angucyclines are biosynthesised via oxidative ring opening (cleavage) of one of the angucycline rings, followed by rearrangement of the intermediates (Fan & Zhang, 2018). This group of molecules is exemplified by the jadomycins and kinamycins, biosynthesised via oxidative B-ring opening of dehydrorabelomycin; pratensilins (**7**), limamycins (**8**,**9**), and lugdunomycin (**1**), derived from a postulated C- ring opening and rearrangement (Fan & Zhang, 2018) (Fan *et al.*, 2012a, Pan *et al.*, 2017, Zhang *et al.*, 2017b, Wu *et al.*, 2019, Yixizhuoma *et al.*, 2017). In addition, the A-ring can be extensively modified and cleaved resulting in the compounds BE-7585A, grincamycins, fridamycins and urdamycin L (Lai *et al.*, 2018, Sasaki *et al.*, 2010, Yoon *et al.*, 2019, Maskey *et al.*, 2003, Rix *et al.*, 2003). It is postulated that the A-, B- and C-ring opening

involve a Baeyer–Villiger oxidation, due to the discovery of lactone intermediates (Figure 2)(Rix *et al.*, 2003, Tibrewal *et al.*, 2012, Wu *et al.*, 2019). For discussion on oxidative ring opening and the possible oxygenases involved in this reaction, I refer to Chapter 5.

Besides lugdunomycin many other C-ring cleaved angucyclines have been isolated from various *Streptomyces* species and this group of molecules has grown into a distinct subclass of natural products (Mikhaylov *et al.*, 2021). However, not much is known about their cognate BGC, and the reaction mechanism involved in C-ring cleavage (Mikhaylov *et al.*, 2021, Cao *et al.*, 2021).

In this study we identified the BGC required for the production of angucyclines, limamycins, pratensilin A and lugdunomycin in *Streptomyces* sp. QL37, which we designated *lug*. The role of this BGC was confirmed with targeted gene deletion. The borders of the *lug* gene cluster were predicted using RNA-seq and by comparison with angucycline BGCs from other *Streptomyces* spp. These comparisons also revealed genes specific to the *lug* gene cluster and other potential producers of C-ring rearranged angucyclines. Phylogenetic analysis of various *Streptomyces* strains and *Kitasatospora* strains revealed that the presence of the *lug* gene cluster and angucycline production was not limited to a specific phylogenetic clade but were distributed all over the phylogenetic tree.

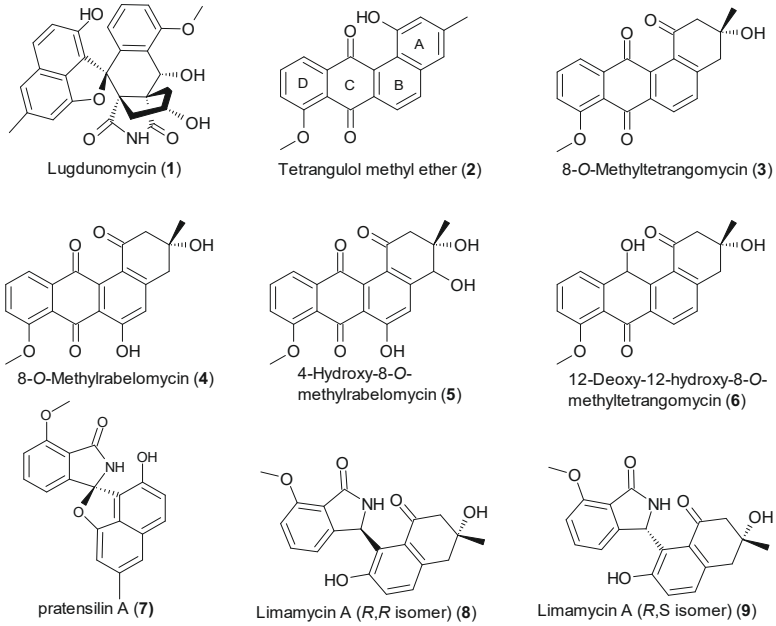


Figure 1 Structures of the compounds discussed in this study.
Lugdunomycin (**1**), non-rearranged angucyclines (**2**, **3**, **4**, **5**, **6**) and limamycins (**8**, **9**) were previously isolated from *Streptomyces* sp. QL37 (Wu *et al.*, 2019). The structure of compound **7** was later revised to be pratensilin A (Mikhaylov *et al.*, 2021).

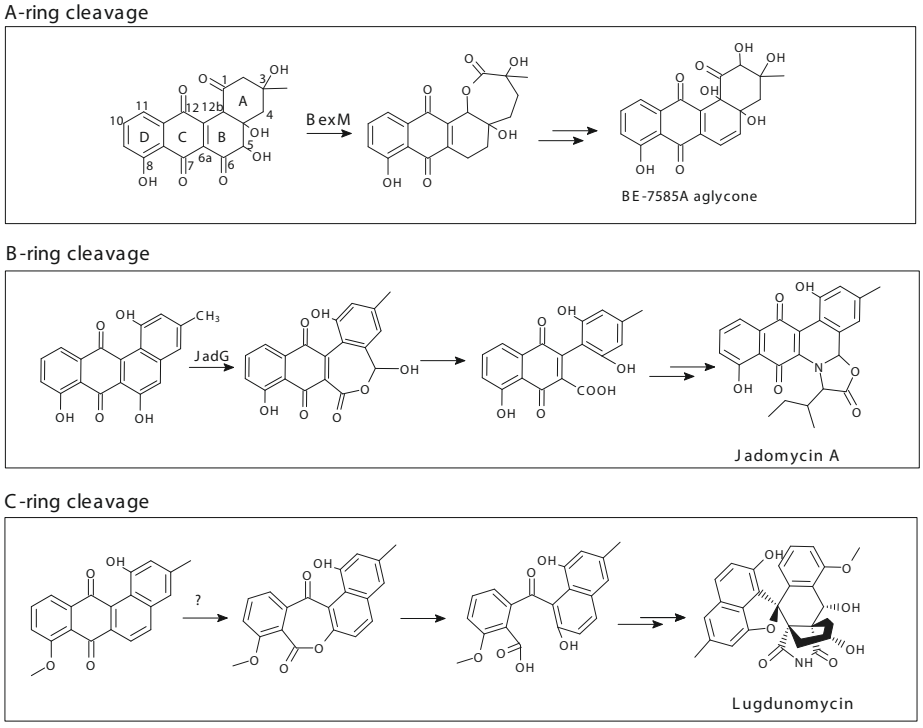


Figure 2 (Proposed) biosynthesis pathways of the BE-7585A aglycone, jadomycin A and lugdunomycin.

Each biosynthesis pathway starts with the non-rearranged or typical angucycline structure (Fan & Zhang, 2018). The ring opening at each ring is postulated to be catalysed by a Baeyer–Villiger mechanism, whereby an oxygen is incorporated in one of the rings, resulting in a lactone intermediate. BexM, an oxygenase-reductase is proposed to introduce an oxygen at position C12b–C1 of the angucycline A-ring, followed by hydrolysis of the lactone ring (Sasaki *et al.*, 2010). JadG introduces an oxygen at position C5–C6 of the angucycline B-ring followed by hydrolysis of the formed lactone and incorporation of *iso*-leucine, forming jadomycin A (Fan *et al.*, 2012a) . The activity of JadG was experimentally confirmed (Fan *et al.*, 2012a). For the production of the C-ring rearranged angucyclines, a similar mechanism is proposed as for the opening of the A-and B-rings, whereby an oxygen is incorporated in the C6a–C7 bond (Wu *et al.*, 2019). However, the question is which enzymes are involved in the opening of the C-ring (for further discussion I refer to below and Chapter 5).

RESULTS

Targeted deletion of the minimal PKS genes in the *lug* cluster

Lugdunomycin is derived from angucyclines (Wu *et al.*, 2019), which are generated by type II polyketide synthases (PKS) (Kharel *et al.*, 2012). Genome analysis of *Streptomyces* sp. QL37 using the bioinformatic tool antiSMASH 5.0 revealed 35 predicted BGCs, including a type II PKS cluster which is designated as *lug* (Figure S1) (Medema *et al.*, 2011). The by antiSMASH proposed region (BGC12) is 69,871 nt long (Figure S2 and Table S1). In this Chapter we propose a region of 29,183 kb as designated in Figure 3A.

To verify whether the *lug* gene cluster is indeed required for the biosynthesis of angucyclines, limamycins and lugdunomycin, genes encoding for minimal PKS enzymes were deleted in the wild-type strain via homologous recombination. LugA–C are the enzymes required for the biosynthesis of the angucycline backbone (Kharel *et al.*, 2012). The *lugA–OII* genes were replaced by double recombination with an apramycin resistance cassette (*aacC4*) using the unstable multi-copy plasmid pWHM3-*oriT* (Wu *et al.*, 2019, Vara *et al.*, 1989). The created mutant is further indicated as *lug-pks* mutant. Primers used for the generation of the knock-out construct are indicated in Table 3. We previously isolated lugdunomycin from cultures grown on minimal medium (MM) supplemented with 0.5% mannitol and 1% glycerol, while pratensilin A (**7**) and the limamycins (**8**, **9**) were isolated from cultures on R5 medium supplemented with 1% mannitol and 0.8% peptone, and angucyclines from both media (Wu *et al.*, 2019). Therefore, the wild-type strain and its *lug-pks* mutant were grown on both MM and R5 agar plates for seven days at 30 °C, and metabolites were extracted using ethyl acetate. The extracts were subsequently dried, re-dissolved in methanol and analysed using liquid chromatography-mass spectrometry (LC-MS). The *lug-pks* deletion mutant failed to produce any of the non-rearranged or rearranged angucyclines, including pratensilin A, the limamycins and lugdunomycin, under all conditions (Figure 3). This confirms that the *lug* gene cluster is indeed required and essential for (non)-rearranged angucyclines and lugdunomycin biosynthesis.

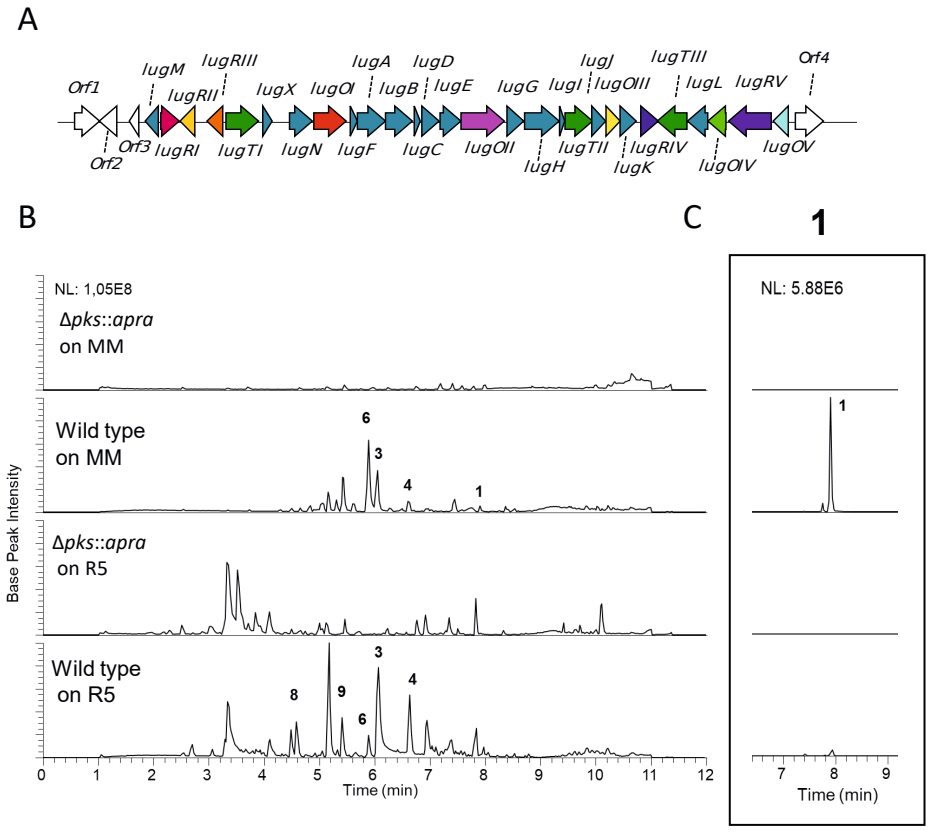


Figure 3 The *lug* gene cluster is required for the biosynthesis of angucyclines, pratensilin A, limamycins and lugdunomycin.
A) Genetic organization of the *lug* gene cluster. More information on the annotation of the genes can be found in Table S2. The marine blue genes indicate genes that are not annotated as either regulatory or oxygenase gene, the white genes are not part of the *lug* gene cluster B) Base peak LC-MS chromatogram of the extracts derived from *Streptomyces* sp. QL37 wild-type and its *lug-pks* mutant grown on MM and R5. Compound numbers are given over their corresponding peaks (Figure 1). The wild-type strain produced angucyclines (**6**,**2**,**3**,**4**) pratensilin A (**7**) and limamycins (**8**,**9**) and lugdunomycin (**1**), whereas the *pks* mutant was unable to produce these compounds (**2** and **7** could only be visualised in an extracted ion chromatogram). C) Extracted ion chromatogram of the water adduct ion of lugdunomycin (**1**, m/z 456.14) in the corresponding extracts shown in B. Lugdunomycin is produced by the wild type grown on MM, whereas the *pks* mutant is incapable of producing the molecule.

Prediction of the extent of the *lug* gene cluster using time-course RNA-seq on the transcriptome of *Streptomyces* sp. QL37

To obtain more insights in the extent of the *lug* gene cluster (described in this Chapter) and whether other BGCs are co-expressed with the cluster (described in Chapter 4), RNA-sequencing (RNA-seq) was performed. For this, *Streptomyces* sp. QL37 was grown on MM agar at 30 °C, and mycelia were harvested after 24 h (T_{24h}; vegetative growth), 36 h (T_{36h}; aerial growth), 48 h (T_{48h}; onset of sporulation) and 60 h (T_{60h}; sporulation). Four biological replicates were collected for all time points. From these samples, RNA was isolated from the mycelium, and secondary metabolites were simultaneously extracted from both the medium and the mycelium. The RNA sequencing (RNA-seq) data were normalised using DeSeq2 (Love *et al.*, 2014).

Principal component analysis (PCA) of the transcriptome data was done to evaluate the different replicates and treatments (Figure S3) This revealed that the replicates of the transcriptome of *Streptomyces* sp. QL37 at the different time points clustered into separate groups. The variation among the replicates of the 24 h samples was small, while that among the 36 h samples was relatively large.

At first the expression of the *lug* gene cluster was analysed (Figure 4) and thereby an indication could be made on the extent of the cluster; a similar gene expression pattern can indicate that these genes are co-regulated and are thus part of one gene cluster. To do so, the expression of all genes in region 12 (68,871 nt) predicted by antiSMASH was studied (Figure S2 and Table S1). Figure 3 depicts the average expression of each gene in region 12 at each of the four time points. In certain regions (QL37_12745–QL37_128760, DNA replication), (*orf1–orf2*, hypothetical protein, GntR-regulator), (*lugRI–lugOV*) the gene expression was relatively high (Figure 4). Comparing the normalised gene expression counts of each gene at T_{36h}, T_{48h} and T_{60h} with the expression counts at T_{24h}, confirmed that the region surrounding the *lugA–C* genes was significantly increased at the time points T_{36h}, T_{48h} and T_{60h}. On the left flank of *lugA–C* this expression trend extended until *lugRI* and on the right border of these genes until *lugOV*, indicating that *lugRI* and *lugOV* may define the borders of the *lug* gene cluster. On the right flank of the gene cluster *orf10* also showed an expression trend similar to that of the *lug* gene cluster, this gene is possibly involved in sugar transport (Table S1).

Most of the *lug* genes showed a similar expression pattern. The lowest expression was seen at T_{24h} and the highest expression was seen at T_{36h}. From T_{36h} onwards, the expression of the genes decreased. The genes *lugRI*, *lugRII*, *lugRIII*, *lugTI* and *lugX* had an expression pattern different from the other *lug* genes. The increase in *lug* gene expression at T_{36h} corresponded well with the appearance of angucyclines at T_{48h} and T_{60h} (Figure S4). These are also the time points were lugdunomycin was observed at T_{60h} and possibly at T_{48h}.

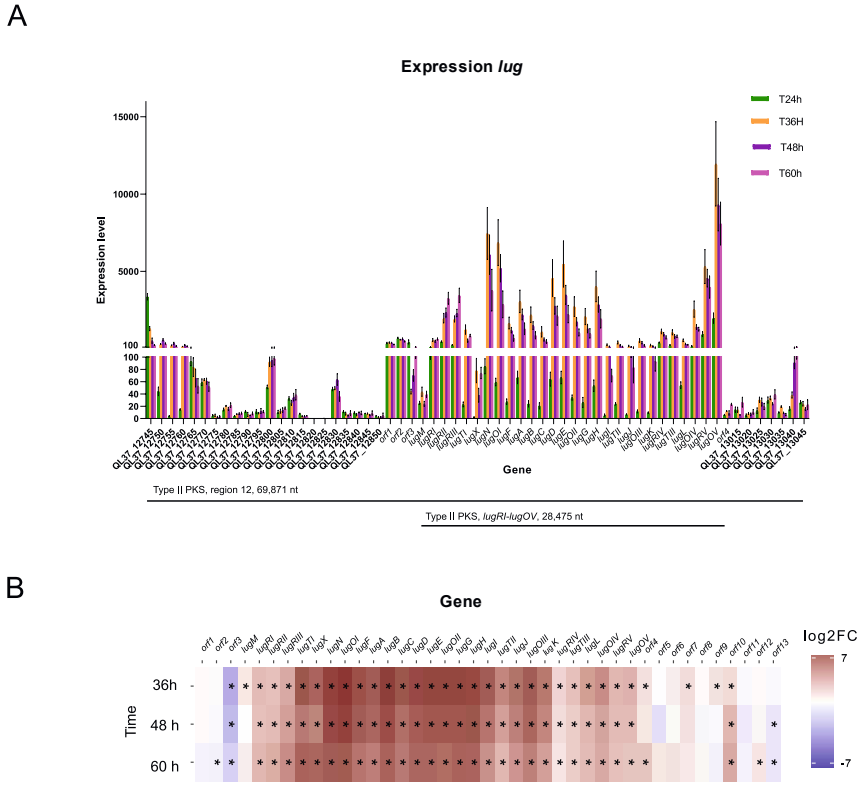


Figure 4 Expression level of the *lug* genes at T_{24h}, T_{36h},T_{48h} and T_{60h}
A) Barplot indicates the absolute normalised expression level of each gene in the predicted *lug* gene cluster. Error bars indicate the standard error of the mean (SEM). B) The heatmap indicates the log2FC of the expression of each gene in or around the *lug* cluster at $T_{36h,}$ T_{48h} and T_{60h} compared to the expression of each gene at T_{24h}. Data with *p*-value <0.05 are indicated with (*) . The expression level of each gene is increased from T_{36h}, which is in line with the production of angucyclines at T_{48h} (Figure S4). The heatmap suggests that the *lug* gene cluster includes the genes from *lugRI–lugOV*.

Prediction of the functional *lug* gene cluster using bioinformatic comparison

The RNA-seq data suggested that the *lug* gene cluster runs from *lugRI* to *lugOV* (Figure 4). We then analysed which of these genes are conserved in other *Streptomyces* spp. This should facilitate finding other *Streptomyces* strains that may produce C-ring rearranged angucyclines and perhaps lugdunomycin. Despite its differential regulation as shown by the RNAseq, we also included *lugM* (encoding an NADPH-dependent FMN reductase) in the bioinformatic comparison.

For the gene comparisons a database was generated of 1020 fully sequenced genomes of *Streptomyces* and *Kitasatospora* spp., obtained from the NCBI database and from our MBT strain collection (Zhu *et al.*, 2014b). As a cut-off for the presence of a gene, we used a similarity threshold of 40% aa identity between the predicted gene products. The core feature of the *lug* gene cluster is that it is an angucycline gene cluster and thus the *lug* type gene clusters should contain the genes *lugA–C* typical of type II PKS BGCs, encoding the ketosynthase α (KSα) and β (KSβ) subunits and an acyl carrier protein (ACP), respectively, which form the minimal PKS (Kharel *et al.*, 2012). These genes must be flanked by genes encoding cyclases (*lugF* and *lugE*) and a ketoreductase (*lugD*). The cyclase *lugF* is required for the cyclisation of the fourth angular ring (ring A), which leads to the typical angucycline framework (Kharel *et al.*, 2012, Kulowski *et al.*, 1999). Thus, first of all we filtered the collection for strains containing these core genes. In the searched genomes 262 strains contained genes similar to *lugF–lugE*, suggesting that these strains may all produce angucyclines. In other words, some 25% of all *Streptomyces* and *Kitasatospora* strains produce angucyclines.

Of the 262 strains containing the angucycline core genes, 36 strains contained at least 18 genes with significant similarity to those in the *lug* cluster. The 36 putative *lug*-like gene clusters were analysed using antiSMASH and compared using the software tool clinker using the complete region 12 from the genome of *Streptomyces* sp. QL37 (Figure 5 and Figure S5) (Gilchrist & Chooi, 2021, Blin *et al.*, 2019). This showed that *orf3* and the left flanking region of this gene are absent in all of the 36 strains and it showed that the gene *orf4* and the right flanking region of this gene are not found in the neighbourhood of the *lug*-type gene clusters of the 36 strains, using a 40% similarity cut-off. Hierarchical clustering using Ward's method mapped to the generated table showed a distribution of BGCs in different groups (or clusters) based on the presence or absence of a gene in the *lug*-type BGCs.

The genes from *lugM* to *lugOV* are frequently shared with the genes in the *lug*-like clusters, which indicates that *lugM* and *lugOV* most likely define the boundaries of the *lug* gene cluster. The region *lugOI –lugOII* was conserved between all BGCs. Some BGCs contained the full-length *lugOII*, while in other strains *lugOII* was split so as to encode a separate oxygenase and reductase (Xiao *et al.*, 2020). Only *lugX* and *lugK* were not found near any of the 36 *lug*-like clusters. Homologues of *lugK* were identified elsewhere in other regions of the chromosome in 28 of the 36 strains. LugK a phosphopantetheinyl transferase (PPTase), which can be used in multiple pathways (Bunet *et al.*, 2014). LugX is a hypothetical protein, and it is not only absent from BGCs, but in fact no homologues of this protein are found in the NCBI database. The role of this putative protein in lugdunomycin biosynthesis (if any) still has to be established. The gene *lugRI* was shared with only one *lug*-type gene cluster; that of *Streptomyces* sp. 94 (Hulcr *et al.*, 2011). However, there was a distance of 23 genes between the *lugRI* homologue and the rest of the *lug*-type gene cluster of *Streptomyces* sp. 94, including many mobile genetic elements.

The BGCs with highest similarity to the *lug* gene cluster were those from *S. scopuliridis* RB72 strain NRRL B-24574, *Streptomyces* spp. CB00072, b84, w007 (Figure 6). Remarkably, the *lug*-like BGCs in the strains *S. scopuliridis*, *Streptomyces* spp. CB00072, b84 and w007 were very similar to each other and to the *lug* cluster in terms of gene order and orientation. Additionally, the four clusters lacked the genes which are homologous to *lugM*, *lugRI*, *lugX*, *lugJ*, *lugK*, and *lugTIII*. *Streptomyces* sp. w007 and *Streptomyces* sp. CB00072 are known angucycline producers (Zhang *et al.*, 2012, Zhang *et al.*, 2015a, Cao *et al.*, 2021). They also produce C-fragmented angucyclines, suggesting that the genes that are absent in these clusters (*lugM*, *lugRI*, *lugX*, *lugJ*, *lugK*, and *lugTIII*) are not required for the production or transport of the C-ring rearranged angucycline that is required for lugdunomycin biosynthesis.

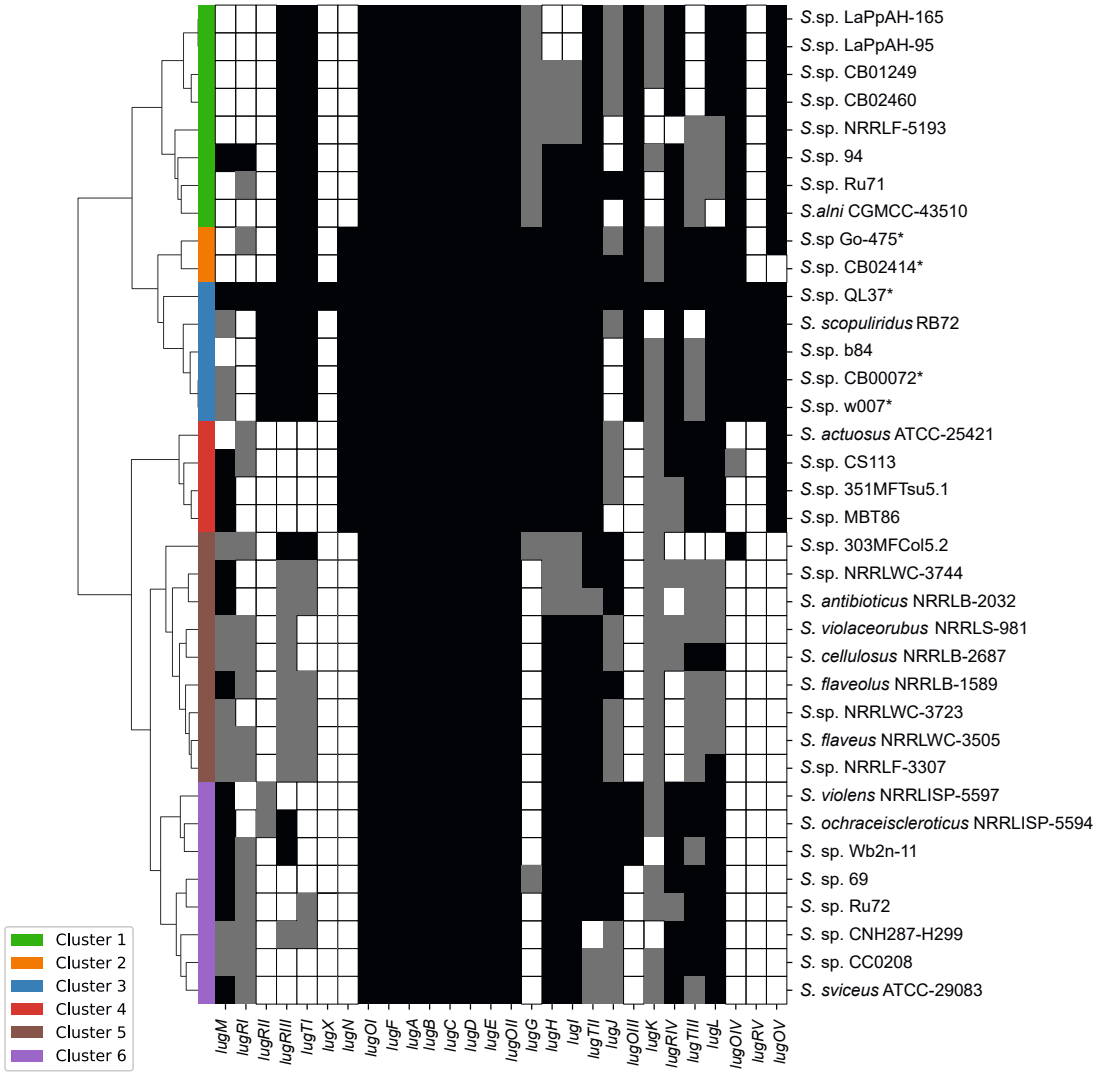


Figure 5 Absence or presence of the *lug* genes in other Streptomyces strains. Hierarchical clustering analysis using Ward's method mapped against a table indicating if a gene is absent (white) present in a polyketide type II BGC (black) or present outside a polyketide type II BGC (grey) in the genome of *Streptomyces* strains that contain at least 18 genes of the *lug* gene cluster. More information about the annotation of each gene can be found in the Table S2.

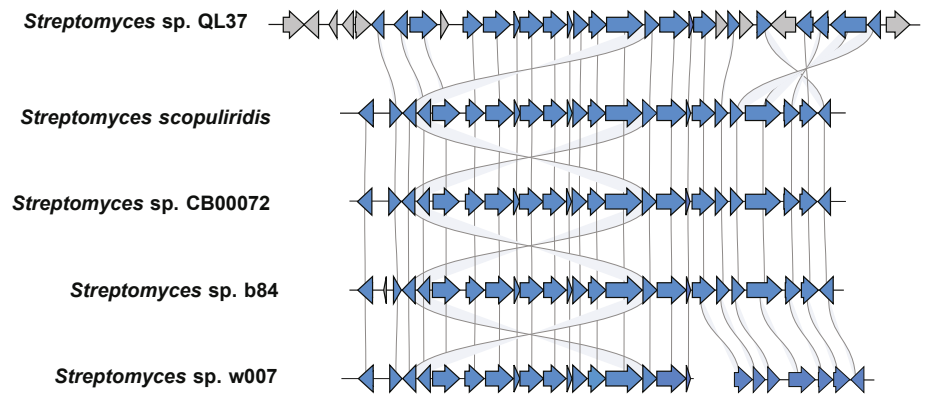


Figure 6 Gene by gene comparison of the *lug* gene cluster with the highest similar ones in other *Streptomyces* strains.
The gene clusters were compared using Clinker (Gilchrist & Chooi, 2021, Berdy, 2005). The strokes between the BGCs link genes that share at least 40% identity. Grey coloured genes, like ORF1, ORF2, ORF3, *lugM*, *lugRI*, *lugX*, *lugK* and ORF4 have no homologues in any of the compared gene clusters.

Prediction of genes that are specific to *lug*-type gene clusters

To obtain a further indication on which genes are specific to the *lug*-type gene clusters, the *lug* gene cluster was compared with 27 BGCs directing non-rearranged angucyclines and A- and B-ring rearranged angucyclines using the software tool Clinker (Table 1)**.** (Gilchrist & Chooi, 2021). Most of these BGCs were derived from the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database, as the organisms containing these BGCs are not fully sequenced, or are detected in species other than *Streptomyces* or *Kitasatospora* (Kautsar *et al.*, 2020). Furthermore, the region of *lugOI–lugE* was highly conserved in all angucycline BGCs. The oxidoreductase gene *lugOII* was not found in all BGCs, but those lacking *lugOII* contained a second paralogue of *lugOI*, a gene similar to *jadF* from the jadomycin BGC (Xiao *et al.*, 2020, Kallio *et al.*, 2008a, Kharel & Rohr, 2012, Fan & Zhang, 2018, Patrikainen *et al.*, 2012). The *lugM* gene was frequently found in BGCs known to produce angucyclines, further supporting the idea that this gene is a *bonafide* member of the *lug* gene cluster. Interestingly, the entire region spanning from *lugRI* to *lugX*, as well as *lugOIII*, *lugOIV* and *lugRV* were absent in all 27 BGCs. This suggests that all of these genes may be specific for biosynthetic pathways involved in the production of C-ring cleaved angucyclines. Again, *lugK* was rarely found; only the simocyclinone BGC of *S. antibioticus* Tü 6040 has an orthologue of the gene, suggesting it plays a minor role in angucycline biosynthesis. Finally,

lugN and *lugOV* were only detected in the hatomarubigin BGC of *Streptomyces* sp. 2238 SVT-4 (Izawa *et al*., 2014)

Table 1 Overview of *lug* genes detected in BGC specifying the production of known non-rearranged angucyclines and A- and B-ring rearranged angucyclines*.

Strain	BGC	*lugM*	*lugRI*	*lugRII*	*lugRIII*	*lugTI*	*lugX*	*lugN*	*lugOI*	*lugF*	*lugA*	*lugB*	*lugC*	*lugD*	*lugE*	*lugOII*	*lugG*	*lugH*	*lugI*	*lugTII*	*lugJ*	*lugOIII*	*lugK*	*lugRIV*	*lugTIII*	*lugL*	*lugOIV*	*lugRV*	*lugOV*		
Streptomyces sp. 120454	Mayamycin																												Non rearragend (typical)	Angucycline backbone
S. antibioticus ATCC 11891	Oviedomycin																													
S. globisporus 1912	Landomycin																													
S. cyanogenus S136	Landomycin																													
S. griseus NTK97	Frigocyclinone																													
Streptomyces sp. 2238-SVT4	Hatomarubigin																													
Streptomyces sp. PGA64	Gaudimycin																													
Streptomyces sp. CS057	Warkmycin																													
Streptomyces sp. TK08046	Saprolmycin																													
S. antibioticus Tü 6040	Simocyclinone																													
Streptomyces sp. NRRL B24484	Simocyclinone																													
Streptomyces sp. KY 40-1.	Saquayamycin																													
Streptomyces sp. MBT86	Saquayamycin																													
Streptomyces sp. SCC2136	Sch-47554 and Sch-47555																													
Kibdelosporangium sp. MJ126-NF4	Azicemycin																													
Streptomyces sp. CNZ748	Grincamycin																												A-ring cleaved	
Amycolatopsis orientalis subsp. vinearia BA-07585	BE-7585A																													
S. griseoflavus Gö 3592	Gilvocarcin																												B-ring cleaved	
S. albaduncus C38291	Chrysomycin																													
S. ravidus NRRL 11300	Ravidomycin																													
S. venezuelae ISP5230	Jadomycin																													
S. murayamaensis sp. nov. Hata et Ohtani	Kinamycin																													
S. ambofaciens ATCC 23877	Kinamycin																													
S. albus DSM 41398	Fluostatin																													
Salinispora pacifica DPJ0016	Lomaiviticin																													
Micromonospora echinospora SCSIO 0408	Nenestatin																													
S. aureofaciens CCM3239/*S. lavenduleae* subsp. *lavendulae* CCM 3239	auricin																												Pyranonaphthoquinone (**)	

*As cut-off for the presence of a gene we used a threshold of 40% aa identity between the predicted gene product. Gene products with less than 40% aa identity are indicated in white. Gene products with an identity of 40≤aa<60% identity are indicated in light grey; gene products with an identity of 60≤aa<80% are indicated in dark grey and gene products with an identity of 80≤aa≤100% are indicated in black.

** Though auricin has been characterised as part of the angucycline family, due to the typical angucycline cyclase gene in its BGC, recent developments in the characterisation of the auricin structure and extensive research on its BGC, revealed it possibly belongs to the pyranonaphtoquinones instead (Matulova *et al.*, 2019).

Phylogenetic analysis of *Streptomyces* sp. QL37 and its relatives

A phylogenetic tree of the strains analysed in this Chapter was generated using the PhyloPhlAn algorithm (Figure 7A) (Asnicar *et al.*, 2020). The phylogenetic analysis revealed that *Streptomyces* sp. QL37 is closely related to *Streptomyces* sp. LamerLS-316 (Book *et al.*, 2016). It is important to note that *Streptomyces* sp. LamerLS-316 lacks an angucycline BGC (Figure 7B). In fact, strains that contain *lug*-like gene clusters are scattered over the phylogenetic tree, indicating that the presence of the *lug* cluster is not correlated to overall phylogeny of the strains. Still, some of the strains that contain a *lug*-like gene cluster are closely related to each other, for example *Streptomyces* spp. w007, CB00072 and b84.

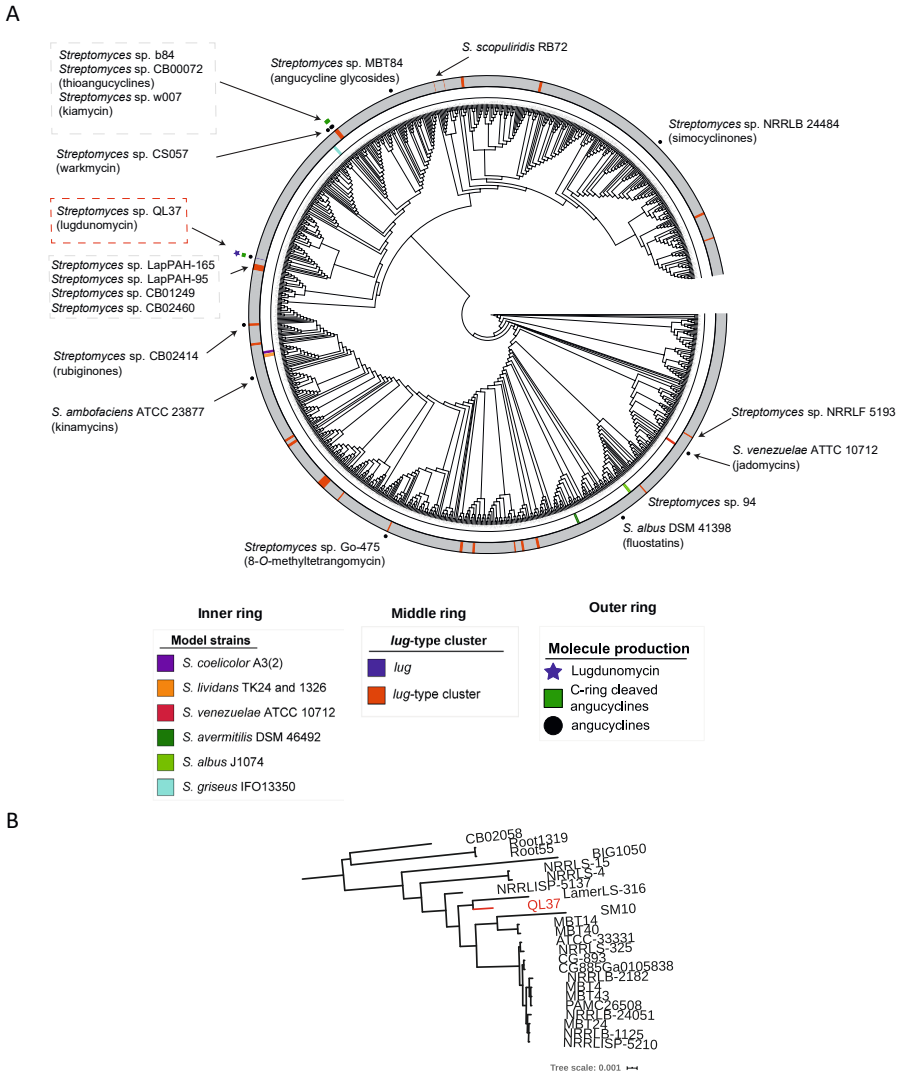


Figure 7 Phylogenetic tree of 1020 *Streptomyces* and *Kitasatospora* strains from the NCBI database and the Leiden MBT collection.
A) The phylogenetic tree was generated using the PhyloPhlAn algorithm (Asnicar *et al.*, 2020) based on four housekeeping genes (*atpD*, *recA*, *trpB* and *gyrB*). The colours of the inner ring represent different well-known *Streptomyces* spp. In the second ring, the colours indicate strains containing different *lug*-type gene clusters. The outer ring shows the *Streptomyces* spp. that are known to be producers of angucyclines, C-ring rearranged angucyclines or lugdunomycin. B) Clade of the phylogenetic tree, showing *Streptomyces* sp. QL37 and its close relatives

DISCUSSION

Streptomyces sp. QL37 produces the highly complex angucycline derivative lugdunomycin (Wu *et al.*, 2019). A biosynthesis pathway was proposed wherein the C-ring of the angucycline backbone is cleaved and subsequently rearranged. The rearranged angucycline was proposed to subsequently react with *iso*-maleimycin through a Diels-Alder [4+2] cycloaddition to form lugdunomycin (Uiterweerd, 2020, Wu *et al.*, 2019). This new study provides clues as to which genes might be involved in lugdunomycin biosynthesis. In this study an angucycline BGC (*lug*) was characterised and analysed using deletion experiments, RNA-seq data analysis and bioinformatic comparisons.

The *lug* gene cluster is a type II PKS gene cluster that shows high similarity to angucycline BGCs. Deletion of *lugA–OII*, of which *lugA–C* encode the minimal PKS enzymes KSα, KSβ and ACP, yielded a strain that did not produce any angucyclines, limamycins or lugdunomycin. Indeed, deletion of the minimal PKS of BGC12 abolished production of all angucycline-related metabolites, including lugdunomycin. AntiSMASH predicted a 70 kb genomic region which includes a complete angucycline BGC and flanking regions, and we performed a more in depth analysis to determine the likely borders of the BGC. Bioinformatic comparison of 1020 *Streptomyces* spp. identified that some 25% of all strains contained angucycline BGCs, a huge number that again underlines how abundant angucyclines are in nature. Of these, 36 strains had a BGC that shared at least 18 genes with the *lug* gene cluster, which were therefore called *lug*-like gene clusters. These include the known angucycline producers *Streptomyces* sp. w007, *Streptomyces* sp. CB00072, *Streptomyces* sp. CB02414 and *Streptomyces* sp. Go-475 (Zhang *et al.*, 2012, Zhang *et al.*, 2015a, Cao *et al.*, 2021, Kibret *et al.*, 2018). We did not find a clear correlation between strain phylogeny and the presence or absence of angucycline BGCs, except for a few strains sharing a highly similar BGC. Based on our comparative analysis, the *lug* gene cluster spans from *lugM* to *lugOV*, considering that the region left of *lugM* and right of *lugOV* were not conserved in any of the 36 *lug*-type gene clusters, nor in the 27 angucycline BGCs directing the production of non-rearranged and A- and B-ring rearranged angucyclines. Time course RNA-seq of the transcriptome of *Streptomyces* sp. QL37 under conditions that lugdunomycin is produced (MM agar) was performed, which showed that the genes from *lugRI–lugOV* had a significantly higher transcription than the

flanking genes. Although, the expression of *lugM* was relatively low compared to the expression of *lugRI–lugOV*, the gene is part of the *lug* gene cluster according to the bioinformatic comparisons. Possibly the expression of *lugM* is regulated differently compared to the expression of *lugRI–lugOV*. Thus, the RNA-seq data combined with the bioinformatic predictions support that the *lug* gene cluster spans from *lugM* to *lugOV*, with in total 28 genes, comprising 29,183 nt of DNA.

The *lug* gene cluster includes among others five oxygenase genes (Chapter 5), five regulatory genes (Chapter 6), and one gene of unknown function (*lugX*) that was exclusively found in *Streptomyces* sp. QL37. Although all 36 strains with *lug*-like gene clusters shared the core set of minimal PKS genes, the post-PKS tailoring genes varied between strains. Accordingly, they could be a good source for the discovery of novel angucycline derivatives, and thus to help understand the evolutionary relationships in the biosynthesis of angucyclines. Of these 36 strains, only *Streptomyces* sp. 94 contains an orthologue of *lugRI* (Hulcr *et al.*, 2011), while *lugK* was not found in any of the angucycline BGCs, but instead was found elsewhere in the genome for 28 out of the 36 strains; *lugK* encodes a phosphopantetheinyl transferase (PPTase), which can be used in multiple pathways (Bunet *et al.*, 2014), and it remains to be seen whether the orthologues play a role in angucycline biosynthesis. Further comparison of the *lug* gene cluster with the 27 previously studied angucycline BGCs that specify the biosynthesis of non-rearranged and A- and B-ring rearranged angucyclines showed that besides *lugX*, seven other *lug* genes were absent in any of these clusters, suggesting they are specific to *lug*-type gene clusters. This included the regulatory genes *lugRI–lugRIII* and *lugRV*, a transporter gene (*lugTI*), and the oxygenase genes *lugOIII* and *lugOIV*. Orthologues of *lugRIV*, the only regulatory genes found in most of the BGCs, typically activate the angucycline BGC they belong to, including *jadR1* and *aur1P* from the jadomycin and auricin BGC in *S. venezuelae* ISP5230 and *S. aureofaciens* CCM3239, respectively (Wang *et al.*, 2009, Novakova *et al.*, 2005). Of the 27 angucycline BGCs, only the hatomarubigin BGC harboured copies of *lugN* and *lugOV*. it is important to note that the *lug* gene cluster and the hatomarubigin BGC both direct the production of 8-*O*-methylated angucyclines, which may require the methyltransferase encoded by *lugN* (Izawa *et al.*, 2014). HrbF shares homology with LugOV (48 % amino acid identity, 84% coverage) and is an oxygenase that regulates the regiospecificity of oxygenase enzymes during hatomarubigin biosynthesis in *Streptomyces* sp. 2238 SVT-4 (Izawa *et al.*, 2014). No orthologues

of *lugOIII* and *lugOIV* could be found in any of the 27 previously studied angucycline BGCs; however, they were conserved in BGCs that likely direct the biosynthesis of C-ring cleaved angucyclines, suggesting a direct correlation to opening of the C-ring. The Lug oxygenases are described in detail in Chapter 5.

In summary, the *lug* gene cluster drives the production of C-ring-rearranged angucyclines and is required for the complex angucycline derivative lugdunomycin. Comparative analysis could clearly separate BGCs for C-ring rearranged angucycline from those for B-ring cleaved angucyclines. In fact, the entire region spanning from *lugRI* to *lugX*, as well as *lugOIII*, *lugOIV* and *lugRV* were absent in all 27 previously studied BGCs for non-rearranged and A- and B-ring rearranged angucyclines. This provides important leads for the understanding of these distinct biosynthetic pathways of angucycline biosynthesis. Further studies should reveal which genes drive the production of lugdunomycin itself.

Acknowledgments
We thank Du Chao for his assistance in creating Figure 5. The work was supported by NACTAR grant 16439 from the Netherlands Organization for Scientific research (NWO).

MATERIALS AND METHODS

Bacterial strains and growth conditions

Streptomyces sp. QL37 was obtained from the MBT strain collection (Zhu *et al.*, 2014b). The strain was isolated from soil in the Qinling mountains (P. R. China) as described previously and was deposited to the collection of the Centraal Bureau voor Schimmelcultures (CBS) in Utrecht, the Netherlands, under deposit number 138593 (Wu *et al.*, 2019). *Streptomyces* sp. QL37 was grown on soya flour mannitol (SFM) for seven days at 30 °C, then spores were collected and stored in 20% glycerol at -20 °C (Kieser *et al.*, 2000). For conjugation *Streptomyces* sp. QL37 was grown on SFM containing 60 mM $MgCl_2$ and 60 mM $CaCl_2$ (Wang & Jin, 2014). Ex-conjugants were selected using thiostrepton (20 µg/ml) and apramycin (50 µg/ml) (Wu *et al.*, 2019). For routine cloning, *Escherichia coli* JM109 was used and grown on Luria Broth (LB), where needed supplemented with antibiotics as selection markers. *E. coli* ET12567/pUZ8002 was used for conjugation of plasmids towards *Streptomyces sp.* QL37. Strains containing plasmids were selected on

LB containing ampicillin (100 µg/ml), apramycin (50 µg/ml), chloramphenicol (25 µg/ml) and kanamycin (50 µg/ml). All strains and plasmids used in this study are listed in Table 2.

Construction of knock-out mutants

Targeted deletion of *lugA–OII* was executed as previously described (Wu *et al.*, 2019, Vara *et al.*, 1989). To obtain a construct for gene disruption, approximately 1.5 kb regions up- and downstream of *lugA–OII* were amplified from the chromosome of *Streptomyces* sp. QL37 using primer pairs 1+2 and 3+4, respectively (Table 3) and cloned as EcoRI-XbaI and XbaI-BamHI fragments into a derivative of the unstable multi-copy plasmid pWHM3 [14] that harbours *oriT* in the PstI site to allow its conjugative transfer. The apramycin resistance cassette *aac(3)IV* flanked by *loxP* sites was then inserted in-between using an engineered XbaI site. The correct knock-out construct was transformed to the methylase-deficient strain *E. coli* ET12567/pUZ8002 [15] and subsequently further introduced into *Streptomyces* sp. QL37 by conjugation, following the protocol as described previously [8]. The correct mutant was selected by resistance to apramycin (50 µg/ml) and sensitivity to thiostrepton (10 µg/ml) [16]. The introduction of the apramycin cassette was verified by PCR with the primers 5-8 (Table 3) and Sanger sequencing of the PCR products (Baseclear).

Table 2 Bacterial strains used in this study

Strains		
Name	Description	Reference
***Streptomyces sp.* QL37**	Wild type, CBS 138593	(Wu *et al.*, 2019), (Zhu *et al.*, 2014b)
Streptomyces sp.* QL37Δ *lug-pks	Deletion mutant of the *lugA–OII* genes (the minimal PKS), Apra^r.	(Wu *et al.*, 2019)
***Escherichia coli* ET12567/ pUZ8002**	*E. coli* ET12567 containing pUZ8002. The strain was used for conjugal transfer. pUZ8002 is a derivative of RK2 with a mutation in *oriT* and containing the *tra* gene; Cm^{r}Km^r.	(MacNeil *et al.*, 1992), (Paget *et al.*, 1999)
***Escherichia coli ET12567/*pUWL-Cre**	*E. coli* ET12567 harbouring pUWL-Cre, a pUWL-*oriT* derivative containing the *cre* gene under the control of the *ermE* promoter. Cre recombinase recognises the *loxP* site, used for removal of *aac (3)IV* (Apra^R); Cm^r, Thio^r.	(Fedoryshyn *et al.*, 2008)
***E. coli* JM109**	*end*A1, *rec*A1, *gyr*A96, *thi, hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F´ *tra*D36, *pro*AB, *laq*I^{q}ZΔM15 Used for general cloning	(Yanisch-Perron *et al.*, 1985)
Plasmids		
Name	Description	Reference
pWHM3-*oriT*	High copy number *E. coli-Streptomyces* shuttle vector. The *oriT* from pSET152 was inserted in the PstI site of pWHM3.	(Wu *et al.*, 2019, Garg & Parry, 2010, Vara *et al.*, 1989)

Table 3 PCR primers used in this study

No.	Name	Sequence 5' --> 3'^
1	MinPKS_LF_Fw	ctagGAATTCCCGCCACCACCGAGCTCTTC
2	MinPKS_LF_RV	GAAGTTATCCATCACCTCTAGAGATACCGGTGATGACGACCC
3	MinPKS_RF_Fw	GAAGTTATCGCGCATCTCTAGAGCCGAGCAGCTCGACCGTTAC
4	MinPKS_RF_Rv	ctagGGATCCCTGCCCTTGTCGAGAAGCAGTG
5	MinPKS_Apra_ Fw	GGATCAGAGATGATCTGCTCTGCCT
6	MinPKS_Apra_Rv	GTCGCCCGTGTCCATGAACTC
7	MinPKS_LF_ Check_Fw	GGTAGCCGCACCCACCGGAC
8	MinPKS_RF_ Check_Rv	TCCTTCGGTCGGGAACCGGG

^ restriction sites underlined. GAATTC, EcoRI; GGATCC, BamHI; TCTAGA, XbaI

Metabolite extraction

For angucycline and lugdunomycin production *Streptomyces* sp. QL37 was grown on minimal media agar plates (MM) supplemented with 0.5% mannitol and 1% glycerol (w/v) as the carbon source. The strains were grown on R5 agar medium plates supplemented with 0.8% peptone and 1% mannitol (w/v) for angucycline and limamycin production (Wu *et al.*, 2019, Kieser *et al.*, 2000). For MM we used Iberian agar (TM Duche & Sons Ltd, batch from 2017) and for R5 Bacto agar (Brunswig Chemie). After seven days of growth at 30 °C the agar plates were cut into small pieces and soaked in 25 ml of ethyl acetate for 12 hours. Subsequently the ethyl acetate was decanted and evaporated at room temperature. This process was repeated two times. The dried extract was re-dissolved in methanol (MeOH) and centrifugated to remove any undissolved matters. Subsequently the MeOH solutions were transferred to new pre-weighed glass vials, where it was dried under nitrogen. The crude extracts were weighed and dissolved in methanol to a fixed concentration for LC-MS analysis. The prepared solutions were centrifuged again for 20 min at 4 °C to remove any suspended matters.

Method LC-MS runs on extracts from the *lug-pks* mutant

LC-DAD-HRESIMS spectra were obtained using a Waters Acquity UPLC system, equipped with Waters Acquity PDA, and coupled to a Thermo Instruments MS

system (LTQ XL/LTQ Orbitrap XL). The UPLC system was run using Acquity UPLC HSS T3 C_{18} column (1.8 µm, 100 Å, 2.1 × 100 mm). Solvent A was 0.1% formic acid, 95% H_2O and 5% ACN. Solvent B was 0.1% formic acid, 95% ACN and 5% H_2O. The gradient used was 2% B for 0.5 min, 2-40% for 5.5 min, 40-100% for 2 min, and 100% for 3 min. The flow rate used was 0.5 ml/min. The MS conditions used were: capillary voltage 5 V, capillary temperature 300 °C, auxiliary gas flow rate 5 arbitrary units, sheath gas flow rate 50 arbitrary units, spray voltage 3.5 kV, mass range 100-2000 amu, FT resolution 30000. Spectra were analysed using Thermo Scientific Xcalibur.

RNA isolation

To collect biomass for RNA isolation, 6×10^6 spores of *Streptomyces* sp. QL37 were confluently spread over an MM agar plate supplemented with 0.5 % mannitol and 1 % glycerol, and samples were taken at 24, 48, 36 and 60 hours at 30 °C. For each time point four replica plates were incubated. At each time point 12 small (1.5 × 0.5 cm) agar pieces were cut from the agar from each of the four replicas. The excess agar below the mycelium was sliced off. It was ensured that the vegetative mycelium was still present in the agar. Two pieces of agar were transferred to a 2 ml tube containing 1 × 4 mm and 2 × 3 mm metal beads and were immediately frozen in liquid nitrogen. This led to 6 samples per plate. In total 96 samples were required for the experiment. The samples were stored at -80 °C until RNA isolation. The following protocol was followed-up in four separate rounds for each replica plate of each time point. Each round of isolation included 6 tubes (6 × 2 pieces of agar) per time point. In the final steps of the protocol, the RNA derived from the 6 samples was pooled. After removal of the samples from -80 °C, the samples were homogenized using a TissueLyserII (Qiagen, Hilden Germany). In order to do so, the tubes were transferred to a pre-cooled TissueLyser block. The samples were shaken at 30 Hz for 15 s and this cycle was repeated four times. Between each cycle the machine was stopped for 5 s. When no large pieces were detected in the samples, the tubes were transferred to liquid nitrogen.

To each sample 0.9 ml of pre-warmed (40 °C) lysis buffer (5% (v/v) Triton X-100 and 10 mM EDTA pH 8.0) was added. The samples were vortexed for 30 s. Subsequently 0.9 ml phenol:chloroform (50:50, pH 4.9, VWR) was added. The mixture was vortexed for 1 min, followed by a centrifugation step (15 min, 12,000 × g). The aqueous phase was transferred to a new tube. The phenol:chloroform RNA

extraction was repeated for another two cycles, in order to make sure the sample was clean. After the final phenol:chloroform cleaning step, the aqueous phase was divided over two tubes. The nucleic acids were precipitated by addition of 62.5 µL 1 M Tris-HCL pH 8.0, 25 µL 5 M NaCl and 0.75 ml cold absolute ethanol. After overnight incubation at -20 °C, the samples were centrifuged (5 min at 7,500 × g) and the supernatant was removed. The pellet was washed with 1 ml 80% ethanol. The ethanol was removed, and the pellet was dried at room temperature for 10 min. The pellets derived from the samples that were earlier divided were pooled by dissolving in 87.5 µL nuclease-free water (Ambion). To that, 10 µL DNase buffer and 5 units of DNase I (New England Biolabs) were added, and samples were incubated for 45 min at 37 °C. This was followed by another cleaning step using phenol: chloroform. To do so, nuclease free water was added up to a volume of 200 µL. Then 200 µL phenol:chloroform were added to each sample and vortexed for 1 min. The samples were centrifuged (10 min, 7500 × g) and one other round of phenol:chloroform extraction was executed. The aqueous phase was transferred to a new tube and precipitated by addition of 25 µL 1M Tris-HCl pH 8.0, 10 µL 5M NaCl and 300 µL cold ethanol absolute. This was incubated for at least 10 min at -20 °C. Again, the samples were centrifuged (5 min, 7,500 × g) and the supernatant was removed. The pellet was washed with 200 µL 80% ethanol, dried and subsequently dissolved in 50 µL nuclease free water. The six samples of each different time point were pooled, which led to a final volume of 300 µL RNA. The purity of the RNA was tested using the A260/A280 and A260/A230 ratio using the Nanodrop ND-1000 Spectrophotometer (PEQLAB) and the integrity of the RNA was assessed on a 1.2% 1x TAE gel before sending to Baseclear. The RNA integrity number was estimated using the Bioanalyzer 2100 (Agilent).

Library preparation and RNA-sequencing

rRNA depletion and RNA sequencing were performed by BaseClear. RNA samples were subjected to the rRNA depletion kit using the siTOOLS riboPOOLs Actinobacteria kit. Single-end or paired-end sequence reads were generated using the Illumina NovaSeq 6000 or MiSeq system at a depth of 6–9 million reads each. The sequences generated with the MiSeq system were performed under accreditation according to the scope of BaseClearB.V. (L457; NEN-EN-ISO/IEC 17025).

Quality assessment and analysis of the RNA-seq data

Illumina sequences were trimmed and filtered with FastQC using a threshold of 25 (quality value [Q] > 25). Reads were mapped to the *Streptomyces* sp. QL37 reference genes using the software Bowtie 2 v.2.1.0 (Langmead *et al.*, 2009). The Bioconductor package DESeq2 (Love *et al.*, 2014) was used for normalization and differential expression analyses. The *p*-value was obtained from the differential gene expression test, and False discovery rate (FDR) was used to correct for multiple testing. Accordingly, differentially expressed genes (DEGs) were regarded significant using an FDR adjusted *p*-value of <0.05.

Bioinformatics

For the prediction of the BGCs in the genome sequence of *Streptomyces* sp. QL37 (NZ_PTJS00000000.1) and other *Streptomyces* strains, the bioinformatic tool antiSMASH 5.0 was used (Blin *et al.*, 2019). The bioinformatic tool Clinker was applied to compare the *lug*-like and angucycline and betalactone gene clusters (Gilchrist & Chooi, 2021). As cut-off for the presence of a gene we used a threshold of 40% aa identity between the predicted gene products.

Phylogenetic analysis

A set of 1020 genomes, including 1004 *Streptomyces* and 16 *Kitasatospora* genomes, was downloaded from the NCBI database by querying the fasta files in combination with the taxonomic identifier. 116 unpublished draft genome sequences of an in-house collection of actinomycetes (Zhu *et al.*, 2014b) were added to the set. A maximum-composite likelihood phylogeny was constructed using PhyloPhlAn (Asnicar *et al.*, 2020). Ortholog identification and alignment was performed in Phylophlan using the "-u" command. iTOL (70) was used for the visualization of the phylogenetic tree.

SUPPLEMENTARY TABLES

Table S1 Annotation of the genes in region 12 in the genome of *Streptomyces* sp. QL37 as predicted by antiSMASH (see also Figure S2) (Blin *et al.*, 2019)

Gene	Gene name	Prokka annotation	Gene	gene name	Prokka annotation (Seemann, 2014)
QL37_12745	-	hypothetical protein	QL37_12900	*lugN*	3-hydroxy-5-methyl-1-naphthoate 3-O-methyltransferase
QL37_12750	-	Single-stranded DNA-binding protein 1	QL37_12905	*lugOI*	Anhydrotetracycline monooxygenase
QL37_12755	-	GTPase Era	QL37_12910	*lugF*	Tetracenomycin F2 cyclase
QL37_12760	-	GTPase Era	QL37_12915	*lugA*	Actinorhodin polyketide putative beta-ketoacyl synthase 1
QL37_12765	-	NA	QL37_12920	*lugB*	Actinorhodin polyketide putative beta-ketoacyl synthase 2
QL37_12770	-	Tyrosine recombinase XerC	QL37_12925	*lugC*	Oxytetracycline polyketide synthase acyl carrier protein
QL37_12775	-	hypothetical protein	QL37_12930	*lugD*	Putative ketoacyl reductase
QL37_12780	-	hypothetical protein	QL37_12935	*lugE*	Putative polyketide cyclase
QL37_12785	-	hypothetical protein	QL37_12940	*lugOII*	Anhydrotetracycline monooxygenase
QL37_12790	-	hypothetical protein	QL37_12945	*lugG*	3-oxoacyl-[acyl-carrier-protein] reductase FabG
QL37_12795	-	hypothetical protein	QL37_12950	*lugH*	Methylmalonyl-CoA carboxyltransferase 12S subunit
QL37_12800	-	hypothetical protein	QL37_12955	*lugI*	hypothetical protein
QL37_12805	-	hypothetical protein	QL37_12960	*lugTII*	putative MFS-type transporter EfpA

Table S1 Annotation of the genes in region 12 in the genome of *Streptomyces* sp. QL37 as predicted by antiSMASH (see also Figure S2) (Blin *et al.*, 2019) *(continued)*

Gene	Gene name	Prokka annotation	Gene	gene name	Prokka annotation (Seemann, 2014)
QL37_12810	-	hypothetical protein	QL37_12965	*lugJ*	Chromate reductase
QL37_12815	-	hypothetical protein	QL37_12970	*lugOIII*	Tetracenomycin-F1 monooxygenase
QL37_12820	-	hypothetical protein	QL37_12975	*lugK*	4'-phosphopantetheinyl transferase Npt
QL37_12825	-	hypothetical protein	QL37_12980	*lugRIV*	Transcriptional regulatory protein BaeR
QL37_12830	-	hypothetical protein	QL37_12985	*lugTIII*	Enterobactin exporter EntS
QL37_12835	-	hypothetical protein	QL37_12990	*lugL*	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase
QL37_12840	-	Agglutinin receptor	QL37_12995	*lugOIV*	putative oxidoreductase YciK
QL37_12845	-	hypothetical protein	QL37_13000	*lugRV*	Regulatory protein AfsR
QL37_12850	-	hypothetical protein	QL37_13005	*lugOV*	Tetracenomycin-F1 monooxygenase
QL37_12855	-	hypothetical protein	QL37_13010	-	Putative phosphoribosyl transferase
QL37_12860	-	Methanol dehydrogenase activator	QL37_13015	-	hypothetical protein
QL37_12865	-	hypothetical protein	QL37_13020	-	hypothetical protein
QL37_12870	*lugM*	Chromate reductase	QL37_13025	-	hypothetical protein
QL37_12875	*lugRI*	HTH-type transcriptional regulator DdrOP3	QL37_13030		hypothetical protein

Table S1 Annotation of the genes in region 12 in the genome of *Streptomyces* sp. QL37 as predicted by antiSMASH (see also Figure S2) (Blin *et al.*, 2019) *(continued)*

Gene	Gene name	Prokka annotation	Gene	gene name	Prokka annotation (Seemann, 2014)
QL37_12880	*lugRII*	Transcriptional regulatory protein LiaR	QL37_13035	-	hypothetical protein
QL37_12885	*lugRIII*	Fatty acid metabolism regulator protein	QL37_13040	-	Ribose import ATP-binding protein RbsA
QL37_12890	*lugTII*	putative transport protein HsrA	QL37_13045	-	Ribose transport system permease protein RbsC
QL37_12895	*lugX*	hypothetical protein		-	

Table S2 Functional annotation of the gene products of the *lug* gene cluster in *Streptomyces* sp. QL37

Protein	Gene Annotation	Aa	Putative function	Nearest homologue	Identity (%)	Coverage (%)	Accession
Orf1	WP_187355741.1	378	Hypothetical protein	*Streptomyces mediolani*	97,4	100	WP_030801858.1
Orf2	WP_104791672.1	264	GntR-like regulator, Nudix hydrolase	*Streptomyces pratensis*	98,5	100	MBD2831978.1
Orf3	WP_203186292.1	143	Hypothetical protein	*Streptomyces pratensis*	84,6	100	MBD2831977.1
LugM	WP_104785779.1	199	NADPH-dependent FMN reductase	*Streptomyces* sp. M3	91.4	100	WP_129263039.1
LugRI	WP_104785781.1	280	XRE family transcriptional regulator	*Streptomyces* sp. M3	88.2	100	WP_129263041.1
LugRII	WP_104785782.1	223	LuxR family transcriptional regulator	*Streptomyces* sp. W007	52.1	97	EHM27498.1
LugRIII	WP_104785784.1	239	TetR family transcriptional regulator	*Streptomyces* sp. L-9-10	63.1	96	RYJ29277.1
LugTI	WP_104785785.1	493	Transporter	*Streptomyces* sp. W007	76	98	WP_032792571.1
LugX	WP_104785787.1	144	Hypothetical protein	---	---	---	---
LugN	WP_104791674.1	346	O-methyltransferase	*Streptomyces* sp. CB00072	71.2	99	WP_073867019.1
LugOI	WP_104785789.1	490	FAD-dependent Monooxygenase	*Streptomyces* scopuliridis	81.4	100	WP_030349271.1
LugF	WP_104785790.1	109	Cyclase	*Streptomyces scopuliridis*	83.	98	WP_030349270.1
LugA	WP_104785792.1	427	Polyketide-α-ketoacyl synthase II	*Streptomyces scopuliridis*	87.3	100	WP_030349269.1
LugB	WP_104785793.1	407	Polyketide-β-ketoacyl synthase	*Streptomyces Scopuliridis*	80.3	99	WP_030349268.1

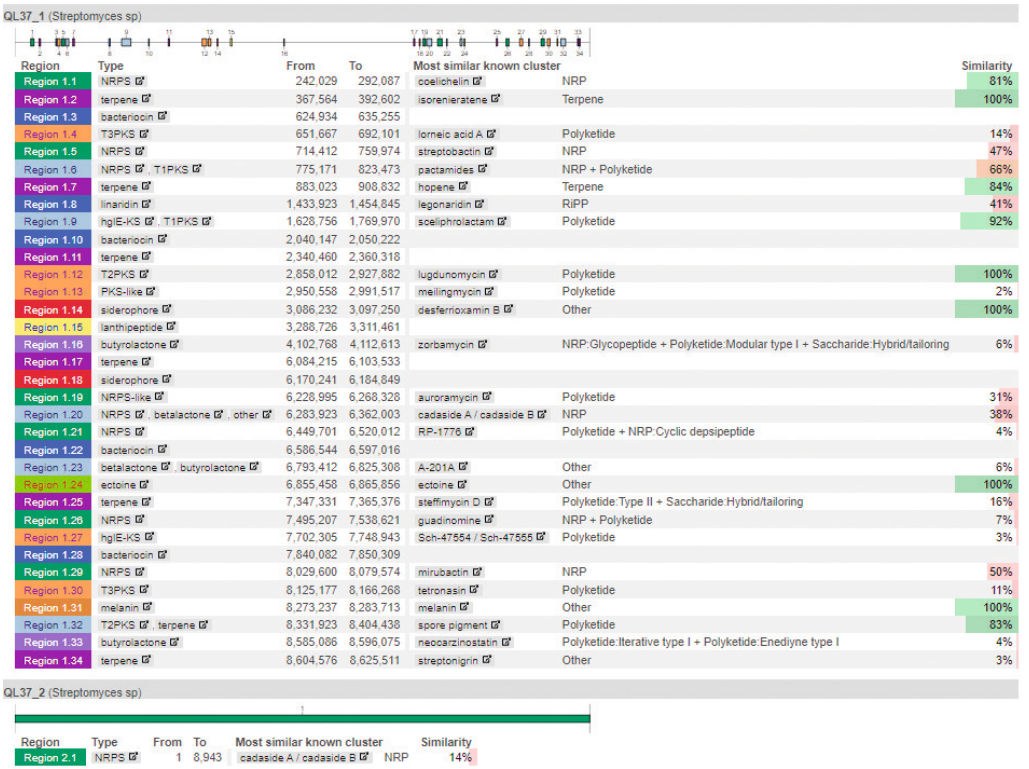
Table S2 Functional annotation of the gene products of the *lug* gene cluster in *Streptomyces* sp. QL37 *(continued)*

Protein	Gene Annotation	Aa	Putative function	Nearest homologue	Identity (%)	Coverage (%)	Accession
LugC	WP_104785795.1	91	Acyl carrier protein	*Streptomyces scopuliridis*	70.8	98	WP_030349267.1
LugD	WP_104785796.1	262	Ketoacyl reductase	*Streptomyces scopuliridis*	85.1	100	WP_030349266.1
LugE	WP_104785798.1	527	Cylclase	*Streptomyces scopuliridis*	81.3	100	WP_030349265.1
LugOII	WP_104785800.1	656	FAD dependent Monooxygenase-SDR family reductase	*Streptomyces* sp. CB00072	76.4	100	WP_073867012.1
LugG	WP_107528480.1	262	Short chain dehydrogenase (SDR)/reductase	*Streptomyces* sp. b84	77.	98	WP_097910307.1
LugH	WP_104791675.1	527	Acyl-coA carboxylase subunit beta	*Streptomyces* sp. w007	87.1	97	WP_050987785.1
LugI	WP_104785801.1	79	Acyl-CoA carboxylase subunit epsilon	*Streptomyces* sp. NBRC 110028	53.8	94	WP_055547361.1
LugTII	WP_104785803.1	417	Transporter	*Streptomyces scopuliridis*	67.7	96	WP_051745235.1
LugJ	WP_104785804.1	211	NAD(P)H dependent FMN reductase	*Streptomyces* sp. Ru71	63.9	90	WP_103781763.1
LugOIII	WP_104785806.1	214	Antibiotic biosynthesis monooxygenase	*Streptomyces* sp. CB00072	62.3	99	WP_073867008.1
LugK	WP_104785807.1	245	4'-phosphopantetheinyl transferase superfamily protein	*Goodfellowiella* sp. AN110305	51.1	94	WP_104785807.1
LugRIV	WP_104785809.1	267	Response regulator transcription factor	*Streptomyces violens*	62	91	WP_078601061.1

Table S2 Functional annotation of the gene products of the *lug* gene cluster in *Streptomyces* sp. QL37 *(continued)*

Protein	Gene Annotation	Aa	Putative function	Nearest homologue	Identity (%)	Coverage (%)	Accession
LugTIII	WP_104785811.1	458	Transporter	*Streptomyces violens*	75.5	84	WP_078601068.1
LugL	WP_104785812.1	307	Alpha/beta hydrolase	*Streptomyces scopuliridis*	69	94	WP_030349255.1
LugOIV	WP_104785814.1	275	SDR family NAD(P)-dependent oxidoreductase	*Streptomyces* sp. TSRI0281	75.6	100	WP_107468115.1
LugRV	WP_104785815.1	646	AfsR/SARP family transcriptional regulator	*Streptomyces scopuliridis*	60.8	93	WP_078490261.1
LugOV	WP_146111254.1	229	Hypothetical protein	*Streptomyces zhaozhouensis*	57.5	90	WP_141514657.1
Orf 4	WP_104785818.1	425	Phosphoribosyltransferase	*Streptomyces* sp. ADI93-02]	93	100	WP_124276433.1
Orf 5	WP_104791676.1	989	Alpha_L-rhamnosidase	*Streptomyces* sp. wa22	93	96	WP_147960798.1
Orf 6	WP_104785820.1	868	glycoside hydrolase family 127	*Streptomyces* sp. TRM S81-3]	96	76.1	WP_188185628.1
Orf 7	WP_104785821.1	1133	glycoside hydrolase family 106 (alpha L-rhamnosidase)	*Streptomyces* sp. TRM S81-3	100	74.6	WP_188185460.1
Orf 8	WP_104785823.1	1092	glycoside hydrolase family 78	*Streptomyces* sp. TRM S81-3	99	77.7	WP_188185461.1
Orf 9	WP_104785824.1	184	GNAT-family acetyltransferase	*Streptomyces* sp. For3	98	81.8	WP_202079774.1
Orf 10	WP_104785825.1	507	Sugar ABC-transport system	*Streptomyces* sp. wa22	100	97.8	WP_147960792.1

SUPPLEMENTARY FIGURES



QL37_1 (Streptomyces sp)

Region	Type	From	To	Most similar known cluster		Similarity
Region 1.1	NRPS	242,029	292,087	coelichelin	NRP	81%
Region 1.2	terpene	367,564	392,602	isorenieratene	Terpene	100%
Region 1.3	bacteriocin	624,934	635,255			
Region 1.4	T3PKS	651,667	692,101	lorneic acid A	Polyketide	14%
Region 1.5	NRPS	714,412	759,974	streptobactin	NRP	47%
Region 1.6	NRPS, T1PKS	775,171	823,473	pactamides	NRP + Polyketide	66%
Region 1.7	terpene	883,023	908,832	hopene	Terpene	84%
Region 1.8	linaridin	1,433,923	1,454,845	legonaridin	RiPP	41%
Region 1.9	hglE-KS, T1PKS	1,628,756	1,769,970	scelphrolactam	Polyketide	92%
Region 1.10	bacteriocin	2,040,147	2,050,222			
Region 1.11	terpene	2,340,460	2,360,318			
Region 1.12	T2PKS	2,858,012	2,927,882	lugdunomycin	Polyketide	100%
Region 1.13	PKS-like	2,950,558	2,991,517	mellingmycin	Polyketide	2%
Region 1.14	siderophore	3,086,232	3,097,250	desferrioxamin B	Other	100%
Region 1.15	lanthipeptide	3,288,726	3,311,461			
Region 1.16	butyrolactone	4,102,768	4,112,613	zorbamycin	NRP:Glycopeptide + Polyketide:Modular type I + Saccharide:Hybrid/tailoring	6%
Region 1.17	terpene	6,084,215	6,103,533			
Region 1.18	siderophore	6,170,241	6,184,849			
Region 1.19	NRPS-like	6,228,995	6,268,328	auroramycin	Polyketide	31%
Region 1.20	NRPS, betalactone, other	6,283,923	6,362,003	cadaside A / cadaside B	NRP	38%
Region 1.21	NRPS	6,449,701	6,520,012	RP-1776	Polyketide + NRP:Cyclic depsipeptide	4%
Region 1.22	bacteriocin	6,586,544	6,597,016			
Region 1.23	betalactone, butyrolactone	6,793,412	6,825,308	A-201A	Other	6%
Region 1.24	ectoine	6,855,458	6,865,856	ectoine	Other	100%
Region 1.25	terpene	7,347,331	7,365,376	steffimycin D	Polyketide:Type II + Saccharide:Hybrid/tailoring	16%
Region 1.26	NRPS	7,495,207	7,538,621	guadinomine	NRP + Polyketide	7%
Region 1.27	hglE-KS	7,702,305	7,748,943	Sch-47554 / Sch-47555	Polyketide	3%
Region 1.28	bacteriocin	7,840,082	7,850,309			
Region 1.29	NRPS	8,029,600	8,079,574	mirubactin	NRP	50%
Region 1.30	T3PKS	8,125,177	8,166,268	tetronasin	Polyketide	11%
Region 1.31	melanin	8,273,237	8,283,713	melanin	Other	100%
Region 1.32	T2PKS, terpene	8,331,923	8,404,438	spore pigment	Polyketide	83%
Region 1.33	butyrolactone	8,585,086	8,596,075	neocarzinostatin	Polyketide:Iterative type I + Polyketide:Enediyne type I	4%
Region 1.34	terpene	8,604,576	8,625,511	streptonigrin	Other	3%

QL37_2 (Streptomyces sp)

Region	Type	From	To	Most similar known cluster		Similarity
Region 2.1	NRPS	1	8,943	cadaside A / cadaside B	NRP	14%

Figure S1 Overview of antiSMASH 5.0 secondary metabolites predicted from the genome of *Streptomyces* sp. QL37. Region 12 contains 100% of the genes of the lugdunomycin BGC, since this cluster is already deposited in the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database(Kautsar *et al.*, 2020).

Figure S2 Region 12 in the genome of *Streptomyces* sp. QL37 contains a type II polyketide gene cluster as predicted by antiSMASH
For further details on this region, see Table S1.

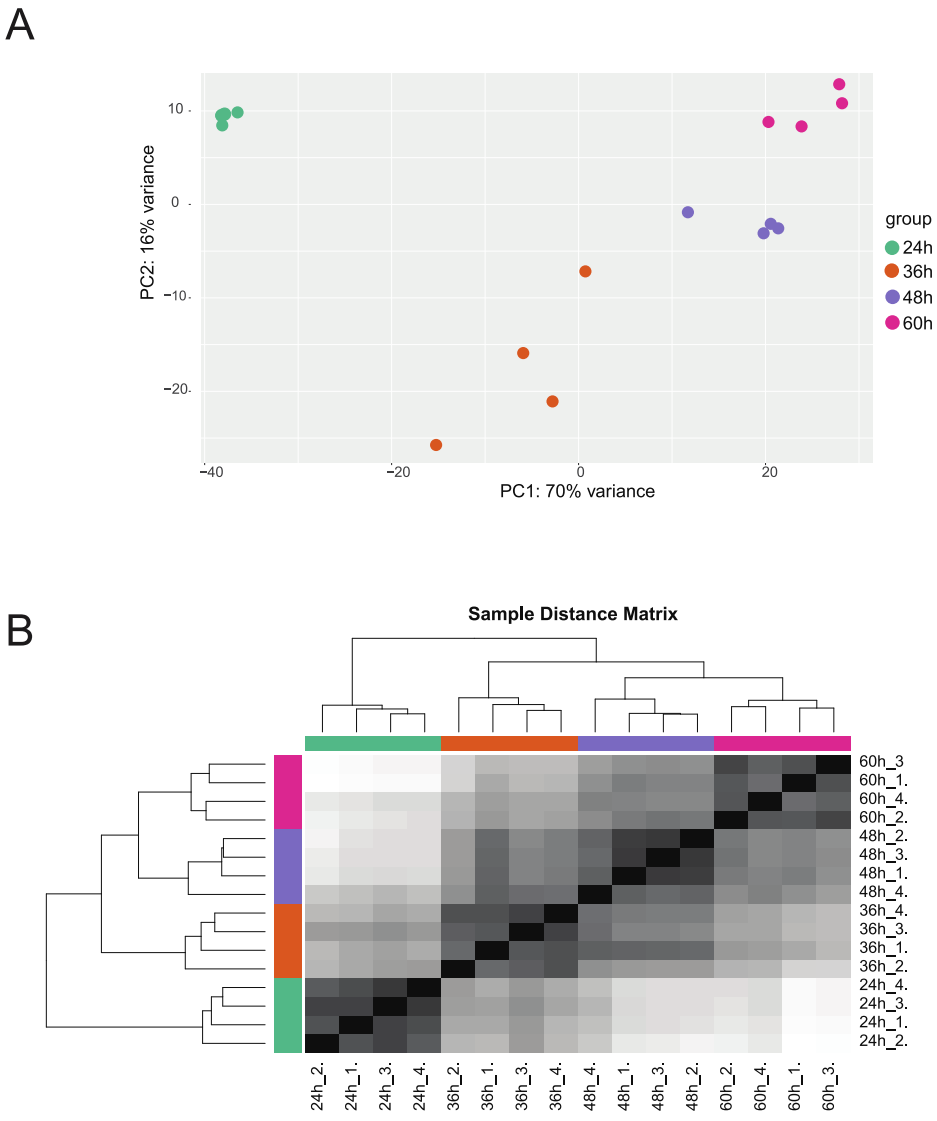


Figure S3 Global analysis of the RNA-seq data of the transcriptome of *Streptomyces* sp. QL37.
RNA was isolated from *Streptomyces* sp. QL37 after 24 hours, 36 hours, 48 hours and 60 hours of growth on MM at 30 °C. A) Principal component analysis (PCA) and B) sample distance matrix of the normalised expression counts of the transcriptome of *Streptomyces* sp. QL37. The samples distance matrix shows the levels of similarity between samples. The square at the intersection of a certain row and a certain column is shaded according to the similarity between the samples. Black indicates a high degree of similarity between two samples, and lighter shades of gray indicate lower levels of similarity (white, no relevant similarity).

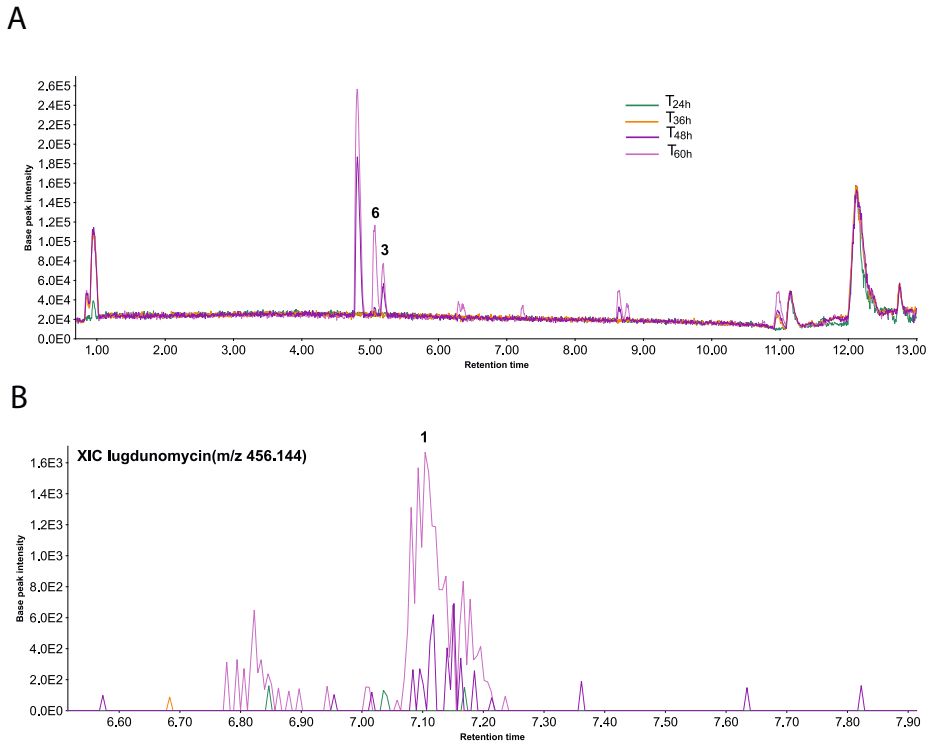


Figure S4 LC-MS chromatogram of the extracts derived from *Streptomyces* sp. QL37 grown on MM for four different time points.
The strain was grown for 24 h, 36 h, 48 h or 60 h on MM agar supplemented with 0.5% mannitol and 1% glycerol. Angucyclines were produced from 48 h (A) and a minute amount of lugdunomycin was identified after 48 and 60 h (B).

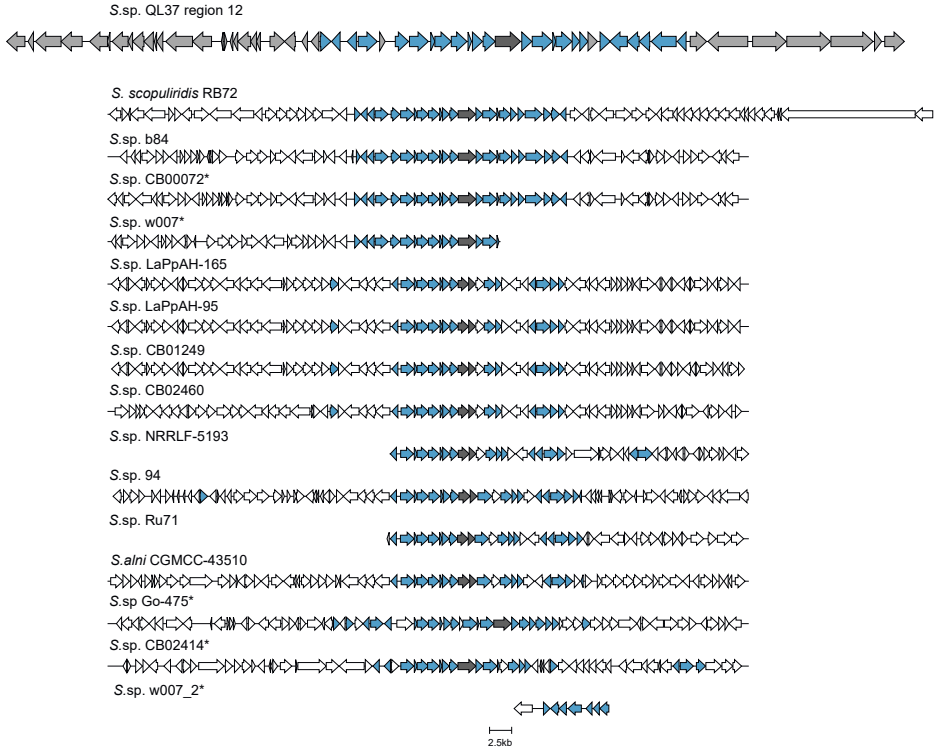


Figure S5 Comparison of region 12 from *Streptomyces* sp. QL37 with *lug*-type gene clusters in other *Streptomyces* strains.
Blue genes are detected in other BGCs, besides the *lug* gene cluster; light grey genes are only detected in the *lug* gene cluster. Dark grey genes indicates either the full length *lugOII*, or the separate oxygenase and reductase encoding genes of which the products show similarity with the oxygenase and reductase domain of LugOII. White genes are genes unrelated tot he *lug* gene cluster. The other 22 *lug*-type gene clusters, indicated in Figure 5 also did not contain the light grey genes (not shown) flanking *lugM* and *lugOV*, indicating these genes mark the borders of the *lug* gene cluster. * indicates known angucycline producers.