
Appendices

List of Figures

| | | |
|-----|---|----|
| 2.1 | A snapshot of the diversity of complex bioactive polyketides. PubChem CIDs were accessed on June 21, 2022. | 5 |
| 2.2 | Biosynthesis of erythromycin A by 6-deoxyerythronolide B synthase (DEBS), overlaid with the structure of its corresponding biosynthetic gene cluster (MIBiG: BGC0000054) from <i>Aeromicobium erythreum</i> . <i>EryAI</i> ⇒ <i>DEBS1</i> , <i>EryAII</i> ⇒ <i>DEBS2</i> and <i>EryAIII</i> ⇒ <i>DEBS3</i> . The grey circle illustrates the lack of KR domain. Figure reprinted from Kwan et al., 2011 [13]. | 8 |
| 2.3 | Representations of catalytic domains with their abbreviations. | 8 |
| 2.4 | Monomers incorporated during the chain initiation and elongation catalyzed by polyketide synthases typically derive from acyl-CoA thioesters. | 10 |
| 2.5 | Monomers incorporated during the chain initiation and elongation catalyzed by Non-Ribosomal Peptide Synthases typically derive from amino acids and aryl acids. | 10 |
| 2.6 | Condensations are catalyzed by the combined action of ketosynthase, acyltransferase and thiolation domains during polyketide biosynthesis. Figure reprinted from Fischbach et al., 2006 [14]. | 11 |
| 2.7 | Acylation and ketosynthase domains activities catalyzed by PKSs. (A) AT domains catalyze the selection and loading of carboxylic extender units on the assembly line, then their transfer to the downstream T_n domain. (B) KS domains catalyze the formation of C-C bonds, that is a Claisen-like condensation reaction, as a result of (methyl)-malonyl-S- T_n decarboxylation. Figure reprinted from Fischbach et al., 2006 [14]. | 12 |
| 2.8 | β -carbon reduction cascade in PKSs catalyzed by ketoreductases, dehydratases and enoylreductases. The KR domain catalyzes the reduction of the β -ketoacyl-S- T_n substrate into its β -hydroxyacyl-S-T form. The DH domains then may serve as a catalyst to dehydrate the earlier reduction product, yielding a α,β -enoxy-S-T substrate. The ER domain's activity may result in a further reduced substrate in the form of a completely saturated acyl-S-T. Reduced variants will be created dependent on the extent of the reduction. Figure reprinted from Fischbach et al., 2008 [14]. | 14 |

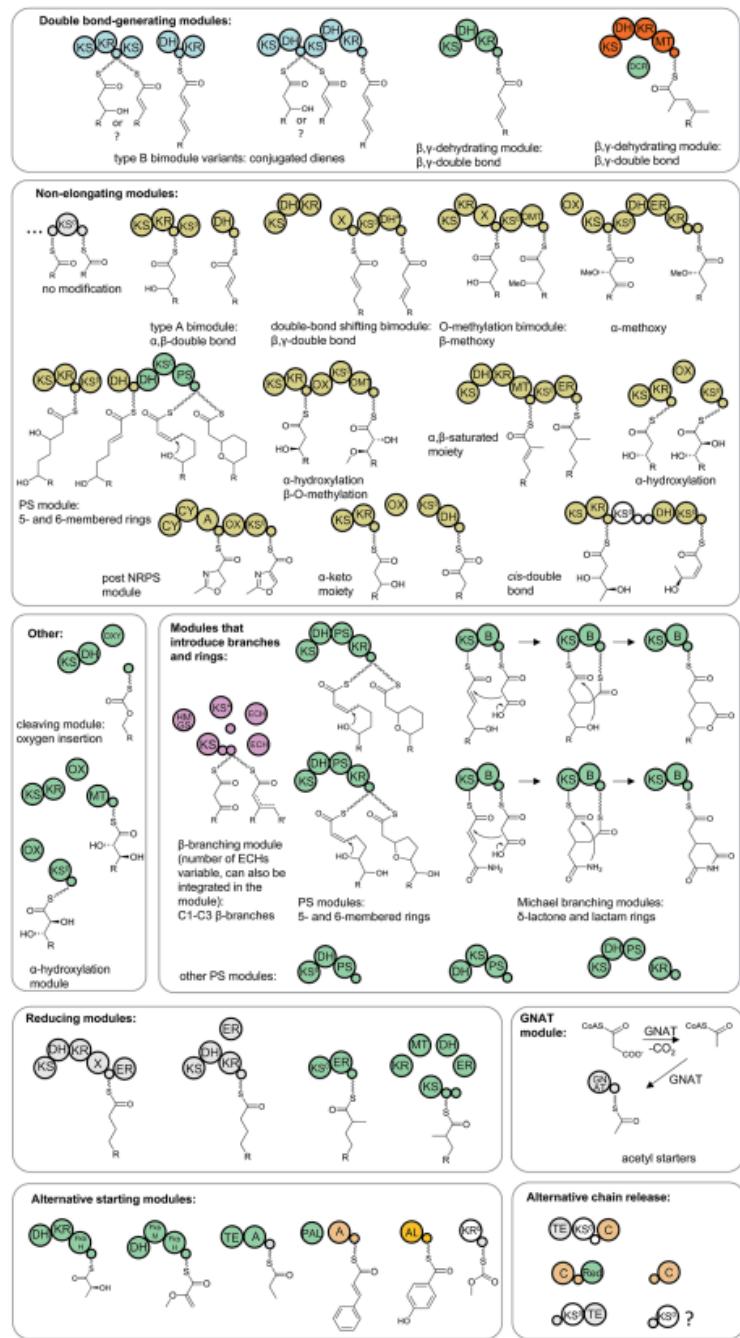
| | | |
|------|---|----|
| 2.9 | Formation of erythromycin A from a 6-dEB polyketide intermediate results from post-PKS tailoring processes. 6-deoxyerythronolide B (6-dEB) is a biologically inactive intermediate that must be tailored post-PKS to become active. In the context of erythromycin, two cyclohexyl functional groups are added to produce erythromycin A, the active form. Figure reprinted from Jurgen <i>et al.</i> , 2002 [28]. | 15 |
| 2.10 | Comparison of the molecular structure of Erythromycin A and C. The methylation post-assembly-line tailoring process that 6-dEB undergoes to produce erythromycin A is highlighted by the black boxes. A lack of methylation post-PKS tailoring yields erythromycin C. Both molecular structures were accessed on May 17, 2022 under the PubChem CID 12560 and 441095, respectively. | 16 |
| 2.11 | Core domains involved in NRP assembly lines. The joint activity of condensation (C), adenylation (A), and thiolation (T) domains is requisite for NRPS backbone elongation. The condensation domain is tasked with the creation of amide bonds. The adenylation domain catalyzes amino or aryl acid recruitment onto the NRPS assembly line as well as the acyl transfer to the neighbouring thiolation domain. Thiolation domains, like in PKS, are in charge for covalently tethering substrates to the biosynthetic assembly line. Figure reprinted from Fischbach <i>et al.</i> , 2006 [14]. | 17 |
| 2.12 | Adenylation and condensation domains activities catalyzed by NRPSs. (A) Adenylation recruits amino acids, which are then acylated and thus attached to the subsequent thiolation domain. (B) The C domain catalyzes the creation of C-N bonds. Figure reprinted from Fischbach <i>et al.</i> , 2006 [14]. | 18 |
| 2.13 | The NRPS ACV synthetase catalyzes the assembly of the ACV tripeptide intermediate. The release of ACV occurs in a linear manner. Isopenicillin N is a crucial intermediate in the synthesis of penicillin G. Figure reprinted from Felnagle <i>et al.</i> , 2008 [40]. | 18 |
| 2.14 | Amino acid epimerization strategies in NRP assembly lines. (A) The preponderant strategy corresponds to the epimerization of amino acids catalyzed by the E domain from an $E-D_{CL}$ didomain. (B) In other NRPSs, single bi-functional D_{CL}/E domains catalyze the epimerization of the L-intermediate to its D-form, which is subsequently selected for condensation. C/E stands for condensation/epimerization. D_{CL} stands for a condensation domain taking D-intermediate as donor and L-intermediate as output. Figure reprinted from Fischbach <i>et al.</i> , 2006 [14]. | 20 |
| 2.15 | Cy-catalyzed formation of heterocycles in NRP backbones. The nucleophilic attack of the side chain of cysteine (X = S), serine (X = O), or threonine (X = O) on the previously incorporated carbonyl is catalyzed by Cy domains, followed by the dehydration of the resulting product to form oxazoline or thiazoline functional groups. Figure reprinted from Fischbach <i>et al.</i> , 2006 [14]. | 20 |
| 2.16 | Comparison of the biosynthetic architecture between <i>cis</i>-AT PKSs and <i>trans</i>-AT PKSs for a fictional polyketide. ACP domains are shown by tiny unmarked circles. At the top are functional annotations of the resulting intermediate. Figure reprinted from Helfrich <i>et al.</i> , 2016 [47]. | 22 |

| | | |
|------|---|----|
| 2.17 | Workflow of the antiSMASH comprehensive analysis tool for secondary metabolites. The antiSMASH comprehensive tool: https://antismash.secondarymetabolites.org/ and the antiSMASH database: https://antismash-db.secondarymetabolites.org/ . Figure reprinted and adapted from Medema <i>et al.</i> , 2011 [52]. | 25 |
| 2.18 | Workflow of the transPACT software. (A) This step corresponds to the phylogenetic placement of KS sequences to a reference phylogeny in order to predict their substrate activities. (B) This step corresponds to phylogenetic comparison of the PKSs sharing module or domain blocks using network and dendrogram representations. Figure reprinted and adapted from Helfrich <i>et al.</i> , 2021 [50]. | 26 |
| 3.1 | Workflow of the genetic engineering investigation of the assembly lines under study from <i>Caballeronia udeis</i> and <i>Massilia flava</i>. On the left side is illustrated the bacterial cultivation on five different poor media. On the right side is depicted the disruption of targeted genes in order to disable the operation of the assembly lines. Created with BioRender.com. | 30 |
| 3.2 | Experimental workflow for the construction of recombinant pGPI plasmids, inserted with bacterial DNA inserts. The gene under disruption is GO485_00250 for <i>M. flava</i> and AWB69_2900 for <i>C. udeis</i> . Created with BioRender.com. | 31 |
| 3.3 | Experimental workflow for the transformation of chemically competent <i>E. coli</i> cells (i.e. SY327 strain) using the ligation mixture that contains native and recombinant pGPI plasmids. The red gene corresponds to the gDNA insert from <i>C. udeis</i> or <i>M. flava</i> . The antibiotic is Trimethoprim. Created with BioRender.com. | 32 |
| 3.4 | Experimental workflow of the pre-screening, sequencing and selection step. Red boxes and circles correspond to the selected wells, that is the ones containing recombinant plasmids. Created with BioRender.com. | 33 |
| 3.5 | Experimental workflow of the conjugation step. The conjugation step is two-folded, corresponding to a tri-parental mating design. Created with BioRender.com. | 34 |
| 3.6 | Predictions of the biosynthetic pathways encoded by <i>trans</i>-AT PKSs in <i>C. udeis</i> and <i>M. flava</i>. HMGS: 3-hydroxy-3-methylglutaryl synthase (branching enzyme). The rest of the domain acronyms can be translated using Figure 2.3. Figure created by Angus Weir. | 36 |
| 3.7 | Files continuum of the transPACT software. (A) Alignment of the KS sequences and extraction of the KS numbers. (B) Annotation of KS sequences with their substrate specificity. (C) Creation of all vs. all diamond table. (D) Processing of the annotation file. (E) Generation of dendrogram visualization. | 38 |
| 3.8 | Growth of both <i>M. flava</i> and <i>C. udeis</i> on BSM plates completed with four different carbon sources. From top left to bottom right, glucose, ribose, fructose, glycerol and control. The control consists of no addition of carbon sources, that is BSM medium. | 40 |
| 3.9 | Stacked LCMS spectra of both <i>M. flava</i> and <i>C. udeis</i> on BSM plates completed with four different carbon sources. | 40 |

| | |
|--|----|
| 3.10 PCR outputs confirming the correct construction of recombinant plasmids. <i>C. udeis</i> 's AWB69_2900 gene and <i>M. flava</i> 's GO485_00250 gene were inserted in the original genome of pGPI plasmids. The first column outlines a ladder of known length. The second column reflects the size of recombinant plasmid's genomes while the third column represents the size of original plasmid's genome. | 41 |
| 3.11 Colony PCR outputs determining the chemically competent <i>E. coli</i> cells that have taken up recombinant plasmids The first column outlines a ladder of known length. | 41 |
| 3.12 antiSMASH identification of PKS/NRPS domains within predicted gene clusters responsible for <i>trans-AT</i> PKSs in the <i>Massilia flava</i> and <i>Caballeronia udeis</i> genomes. ECH stands for enoyl-CoA hydratase. Blue circles represent <i>trans-AT</i> docking locations. Blue ellipses represent PKS PP domains, that is the phosphopantetheine attachment site of ACPs. | 42 |
| 3.13 Rectangular dendrogram visualization of conserved PKS motifs across the <i>trans-AT</i> PKSs under study placed on the the PKSs of reference. Generated using the transPACT software as described on Figure 3.7. A fully detailed colors code is given in Appendix C. | 46 |
| 3.14 Module blocks similarities between <i>C. udeis</i> and <i>M. flava</i> -like <i>trans-AT</i> PKSs. A. Horizontal black lines reflects KS substrate specificities conservation. The PKSs from <i>C. udeis</i> and <i>M. flava</i> are highly similar. The KS10-11-12 block is highly conserved across the six PKSs, ruling the formation of two double bonds and a hydroxyl group. Also, the KS5-6 block is particularly well conserved, ruling the formation of a double bond followed by an exomethylene group. B. Core molecular structures were predicted using the transATor software while the dendrogram was generated using the transPACT software thus specificity annotations slightly differ. | 49 |
| 3.15 TransATor-predicted KS substrate specificity annotations for <i>C.udeis</i> and <i>M.flava</i> PKSs. A color legend is given on the figure. β and α refer to the position of the group on the carbon backbone. | 50 |
| 3.16 Circular dendrogram representation of conserved PKS motifs across 10 to 16 KS domains long <i>trans-AT</i> PKSs from the antiSMASH ClusterBlast database, overlaid with PKSs under study. The figure was generated using the transPACT software as described on Figure 3.7. A fully detailed colors code is given in Appendix C. Circular and rectangular dendrogram visualizations can be accessed on https://itol.embl.de/export/85201185207360831658172247 | 54 |
| 3.17 Module blocks conservation between <i>C. udeis</i> , <i>M. flava</i> and related <i>trans-AT</i> PKSs, identified following the genome mining approach. An additional PKS from <i>Chromobacterium vaccinii</i> XC014 was identified to be highly related to <i>C. udeis</i> and <i>M. flava</i> PKSs, that can be accessed under NZ_CP017707 on NCBI RefSeq database. | 54 |

List of Tables

| | |
|---|----|
| 3.1 Annotation of the KS substrates specificity using transPACT software. PBC = <i>Pseudomonas baetica</i> , GO485 = <i>Massilia flava</i> , AWB69 = <i>Caballeronia udeis</i> , DK441 = <i>Aquimarina</i> sp. AU58, XB90 = <i>Burkholderia pseudomallei</i> MSHR4299, DO63 = <i>Burkholderia pseudomallei</i> strain BDE, BLU84 = <i>Lysobacter enzymogenes</i> , ASC93 = <i>Massilia</i> sp. Root335, CR152 = <i>Massilia violaccinigra</i> B2 and K350 = <i>Sporocytophaga myxococcoides</i> A clade not conserved corresponds to a KS specificity not catalogued in the reference phylogeny. | 43 |
|---|----|



Appendix A: Non-canonical components of trans-AT PKSs and their proposed or demonstrated functions. The biosynthetic products are shown attached to the ACP domains. Connected circles belong to the same proteins. X can be additional domains. For X $\frac{1}{4}$ MT, one or two additional α -methyl groups can be present in the product. KS0, non-elongating KS, DH* dehydratase-like domain involved in double bond migration; OXY, oxygenase; PS pyran synthase; PAL, phenylalanine ammonia lyase; AL, acyl-ligase, FkbH, hydroxylase; FkbM, methyltransferase; Red, reductase; ECH, enoyl-CoA hydratase (syn.crotonase); B, branching domain; OMT, O-methyl-transferase. The following color coding of enzymatic modules is used in this and all following figures: grey, canonical PKS module architecture; orange, NRPS module (not shown here); white, putatively inactive domains lacking active site residues; pink, β -branching module; gold, bimodule involving a non-elongating KS0 ; blue, type B bimodule; red, iterative

module; green, other noncanonical module.

Figure and caption reprinted from Helfrich *et al.*, 2016 [47].

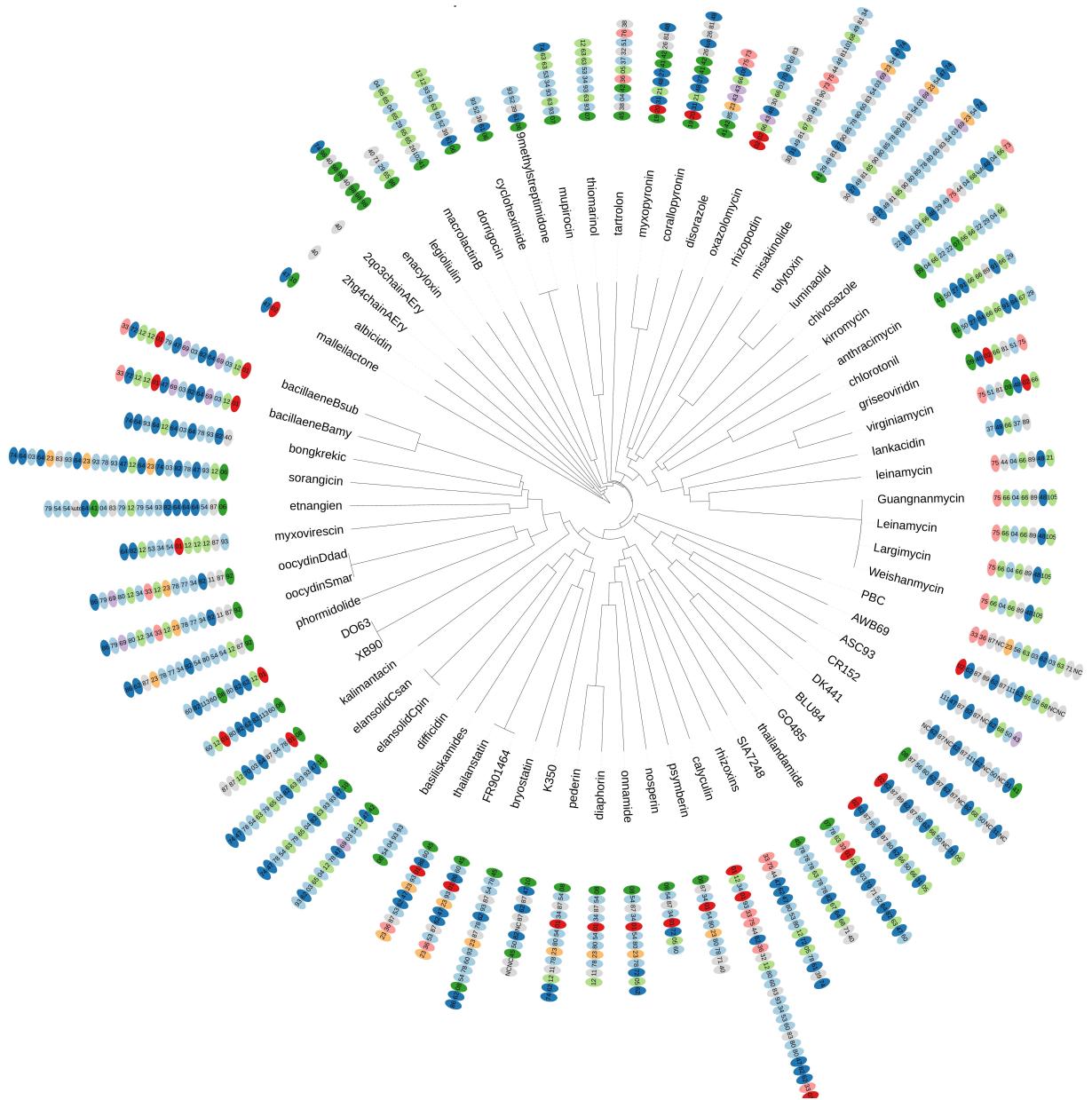
| Group | Color | Clade number | Clade description | Group | Color | Clade number | Clade description |
|--------------|-------|--------------|------------------------------------|----------------|-------|--------------|----------------------------|
| Amino acids | | 1 | amino_acids | eDB | | 70 | eDB |
| Amino acids | | 2 | amino_acids | eDB | | 72 | non_elongating_aMeeDB |
| Amino acids | | 121 | amino_acids | eDB | | 74 | non_elongating_DB |
| bOH | | 3 | bimod_bOH | eDB | | 82 | b_MeDB |
| bOH | | 4 | bimod_bOH | eDB | | 84 | aMe_z_shifted_DB |
| bOH | | 22 | aMe_bketo/bOH | eDB | | 86 | non_elongating_DB |
| bOH | | 29 | b_OH/keto | eDB | | 90 | red_a_Me/shifted_DB_a_Me |
| bOH | | 34 | non_elongating_hemiacetal/b_OH | eDB | | 97 | bMeDB |
| bOH | | 36 | non_elongating_reduced/bOH | eDB | | 107 | eDB |
| bOH | | 49 | non_elongating_b_OH | eDB | | 110 | non-elongating_zDB |
| bOH | | 50 | b_OH | Non-elongating | | 33 | non_elongating |
| bOH | | 51 | b_OH | Non-elongating | | 73 | non_elongating |
| bOH | | 52 | b_OH/eDB | Non-elongating | | 75 | non_elongating_oxazole |
| bOH | | 53 | mostly_a_OH_b_OH | Non-elongating | | 108 | non_elongating_pyran/furan |
| bOH | | 56 | b_L_OH | Starter | | 6 | unusual_Starter_AMT |
| bOH | | 58 | a_LMe_b_DOH | Starter | | 7 | Starter |
| bOH | | 59 | b_DOH | Starter | | 8 | GNAT_Starter |
| bOH | | 60 | non_elongating_b_OH | Starter | | 10 | aromatic_Starter |
| bOH | | 77 | aOH_bOH | Starter | | 19 | methoxycarbonyl_Starter |
| bOH | | 78 | b_L_OH | Starter | | 41 | acetyl_Starter |
| bOH | | 79 | non_elongating_b_L_OH | Starter | | 42 | DB/De_Starter |
| bOH | | 80 | aMe_b_L_OH | Starter | | 45 | lactate_Starter |
| bOH | | 85 | aMe_bOH | Starter | | 92 | unusual_Starter |
| bOH | | 93 | aMe_red/bOH/bketo | Starter | | 104 | acetyl/aromatic_Starter |
| bOH | | 98 | b_OH/eDB(enacyloxins) | Starter | | 125 | aromatic_starter |
| bOH | | 99 | aMe_bOH | zDB | | 43 | zDB |
| bOH | | 100 | aMe_b_L_OH | zDB | | 69 | zDB |
| bOH | | 111 | b_DOH | zDB | | 119 | zDB |
| bOH | | 112 | b_DOH | Other | | 11 | oxygen_insertion |
| bOH | | 113 | a_Me_b_OH | Other | | 23 | pyran/furan |
| bOH | | 115 | b_DOH | Other | | 26 | reduced |
| bOH | | 116 | b_LOH | Other | | 30 | glycine |
| bOH | | 117 | b_OH | Other | | 38 | mainly_reduced |
| bOH | | 122 | b_DOH | Other | | 39 | vinylogous_chain_branching |
| bOH | | 124 | b_DOH | Other | | 40 | cis-AT_PKS-like |
| bOH | | 128 | b_OH | Other | | 44 | oxazole |
| Double bonds | | 5 | shifted_double_bonds | Other | | 67 | various |
| Double bonds | | 12 | reduced/shifted_double_bonds | Other | | 71 | b_keto |
| Double bonds | | 65 | double_bonds | Other | | 81 | b_OMe |
| Double bonds | | 66 | double_bonds | Other | | 83 | exometh/red_bMe |
| Double bonds | | 68 | double_bonds | Other | | 87 | b_keto |
| Double bonds | | 105 | reduced/shifted_double_bonds | Other | | 89 | b_Me |
| Double bonds | | 106 | shifted_double_bonds | Other | | 101 | b_Me |
| eDB | | 31 | eDB | Other | | 102 | b_Me |
| eDB | | 47 | aMe_eDB | Other | | 103 | aMe_reduced |
| eDB | | 48 | aMe_eDB | Other | | 109 | non-elongating_aMe_reduced |
| eDB | | 55 | non_elongating_DB | Other | | 114 | unknown |
| eDB | | 57 | eDB | Other | | 120 | various |
| eDB | | 61 | non_elongating_DB_before_branching | Other | | 123 | non-canonical_pyran |
| eDB | | 62 | eDB | Other | | 126 | a_oxygination |
| eDB | | 63 | mainly_eDB | Other | | 127 | beet |
| eDB | | 64 | eDB | Other | | NC | not_conserved |

Appendix B: List of the color code used to annotate the dendrogram representation in transPACT.

Figure reprinted from Helfrich *et al.*, 2021 [50].



Appendix C: Assembly lines of relevant PKSs. The two first assembly lines are the ones under study. The remaining ones are similar to the PKSs expressed in *Caballeronia udeis* and *Massilia flava*. They all contain an Domain of Unknown Function (DUF) in a conserved module.



Appendix D: Circular dendrogram representation of module blocks conservation using the *transPACT* comparison tool in a query targeted set-up.

| R2A medium | Yeast Maltose Medium Tryptic Soy Medium LB Medium |
|--|--|
| Proteose peptone: 0.5 Soluble starch: 0.5 Yeast extract: 0.5 D-glucose: 0.5 Casamino acids: 0.5 Dipotassium phosphate: 0.3 Sodium pyruvate: 0.3 Magnesium sulfate: 0.03 (Agar: 15) | D-glucose: 10 Peptone A: 5 Yeast extract: 3 Malt extract: 3 (Agar: 0.5) Tryptone: 15 Soya digest: 5 NaCl: 5 (Agar: 15) Premix: 20 (Agar: 15) |

| Basal Salts Medium (BSM) | Final concentration in g/L |
|---|----------------------------|
| Phosphate Salts 20x Stock | |
| $K_2HPO_4 \cdot 3H_2O$ | 4.25 |
| $NaH_2PO_4 \cdot H_2O$ | 1.0 |
| Ammonium Chrolide 20x Stock | 2.0 |
| Nitrilotriacetic Acid 100x Stock | 0.1 |
| Metal Salts 100x Stock | |
| $MgSO_4 \cdot 7H_2O$ | 0.2 |
| $FeSO_4 \cdot 7H_2O$ | 0.012 |
| $MnSO_4 \cdot H_2O$ | 0.003 |
| $ZnSO_4 \cdot 7H_2O$ | 0.003 |
| $CoSO_4 \cdot 7H_2O$ | 0.001 |
| Casamino Acids 5% Stock | 0.5 |
| Yeast Extract 5% Stock | 0.5 |

Appendix F: Culture media experimental for bacterial genome engineering methods.
 Quantities are give in g/L. Original reference for BSM: Hareland *et al.*, 1975 [65].

| | <i>Caballeronia udeis</i> | | <i>Massilia flava</i> | |
|---------------------------------------|---------------------------|--|-----------------------|--|
| 5X Q5 Reaction Buffer | 5 | | 5 | |
| 10 μM dNTPs | 0.5 | | 0.5 | |
| 20 μM F Primer | 0.625 | | 0.625 | |
| 20 μM R Primer | 0.625 | | 0.625 | |
| Template DNA (10 nG) | 0.27 | | 0.11 | |
| Q5 H-FD DNAP | 0.25 | | 0.25 | |
| 5X Q5 High GC | 5 | | 5 | |
| Water | 12.73 | | 12.89 | |
| Primer TM (melting temperature in °C) | 69 | | 64 | |

Appendix G: **Q5 PCR experimental.** Quantities are given in μL .