

Contents

Introduction	1
Specialised metabolites and microbial metabolism	1
Why is it important to find new specialised metabolites?	2
<i>Streptomyces</i> as a source of specialised metabolites	3
<i>Streptomyces leeuwenhoekii</i>	4
Polyketides	4
Non-ribosomal peptides	6
Hybrid non-ribosomal peptide/ <i>trans</i> -AT polyketide	6
Whole genome sequencing	8
From the genome to new compounds	9
To study the metabolism: Genome scale models and Flux Balance Analysis	9
Description of the thesis	11
Objectives	13
1 Identification and characterization of gene clusters of <i>Streptomyces leeuwenhoekii</i> C34	14
1.1 Abstract	14
1.2 Introduction	15
1.3 Methodology	16
1.3.1 Bacterial strains and general procedures	16
1.3.2 Sequencing of <i>S. leeuwenhoekii</i> C34 genome	16
1.3.3 Identification of putative BGCs for specialised metabolites biosynthesis	16
1.3.4 Generation of mutant strains to study BGCs	17
1.3.5 Metabolite profile comparison analysis	17
1.3.6 Implementation and software usage	17
1.4 Results	17
1.4.1 BGC in the genome of <i>S. leeuwenhoekii</i> C34	17
1.4.2 Lasso-peptides	20
1.4.3 Hybrid <i>trans</i> -AT PKS/NRPS BGC	23
1.4.4 Prediction of the possible structure of the hybrid PK-NP	26
1.4.5 Generation of mutants to study the production of the hybrid PK-NP	26
1.4.6 Metabolic profile comparison	29
1.4.7 Bioassays against <i>B. subtilis</i>	33
1.4.8 Effect of increasing salt concentration in the production of specialised metabolites	35
1.5 Discussion	37

1.5.1	The sequencing of <i>S. leeuwenhoekii</i> C34 genome allowed the identification of several biosynthetic gene clusters	37
1.5.2	The <i>trans</i> -AT PKS/NRPS BGC compared to the leinamycin BGC	37
1.5.3	Bioinformatic analysis of the <i>trans</i> -AT PKS/NRPS allows to predict the product of the BGC	38
1.5.4	Could the hybrid PK-NP be a halogenated compound?	38
1.5.5	Why has it been so difficult to detect/identify the specialised metabolite produced by the hybrid <i>trans</i> -AT PKS/NRPS BGC?	39
1.5.6	Utility of the antibiotic activity bioassay	40
1.5.7	The heterologous expression of the hybrid <i>trans</i> -AT PKS/NRPS BGC did not work	41
1.5.8	Possible candidate for the product of the hybrid <i>trans</i> -AT PKS/NRPS BGC	41
1.5.9	Metabolic profiling as a tool to find new specialised metabolites	42
1.5.10	Future work	42
1.6	Conclusions	43
2	Analysis of metabolic networks of <i>Streptomyces leeuwenhoekii</i> C34 by means of a genome scale model: prediction of modifications that enhance the production of specialised metabolites.	44
2.1	Abstract	44
2.2	Introduction	45
2.3	Methodology	46
2.3.1	Bacterial strains	46
2.3.2	Sole carbon sources assay	46
2.3.3	Reconstruction of the genome scale model	46
2.3.4	Biomass composition	47
2.3.5	Incorporation of specialised metabolites pathways	48
2.3.6	Curation of the model	48
2.3.7	Simulations of the metabolism	48
2.3.8	Validation of the model	49
2.3.9	Gene knockout analysis and experimental studies	50
2.3.10	Identification of gene's overexpression targets	50
2.3.11	Implementation and software usage	50
2.4	Results	51
2.4.1	Interface used to reconstruct the GSM of <i>S. leeuwenhoekii</i> , and perform simulations	51
2.4.2	Curation of the model	52
2.4.3	Validation of the model	53
2.4.4	The model	57
2.4.4.1	Fatty acid biosynthesis	58
2.4.4.2	Specialised metabolites pathways	59
2.4.5	Essential gene analysis	62
2.4.6	Gene knockout analysis	62
2.4.7	Identification of targets for overexpression	66
2.5	Discussion	69
2.5.1	Tools used for the reconstruction	69
2.5.2	Discrepancies between model predictions and the experimental data	70

2.5.3	Essential genes of <i>S. leeuwenhoekii</i> C34	70
2.5.4	Identification of metabolic engineering targets for increasing chaxamycins, chaxalactins and hybrid PK-NP production	70
2.5.4.1	Prediction of gene knockout targets	70
2.5.4.2	Prediction of overexpression gene targets	71
2.5.5	Addressing functionality and genetic redundancy	74
2.5.6	The genome scale model of <i>S. leeuwenhoekii</i> C34 would allow experimen- tal design	74
2.6	Future work	74
2.7	Conclusions	75
	Wrapping up: general conclusions and perspectives	76
	Bibliography	79
	Appendix A Strains and plasmids	91
	Appendix B List of primers and maps of vectors used in this work	95
B.1	List of primers	95
	Appendix C Buffers, solutions and culture media	97
C.1	Buffer solutions	97
C.1.1	Buffer SET	97
C.1.2	Buffer STET	97
C.1.3	Buffer TE	98
C.2	Microbiology solutions	98
C.3	Culture medium	99
C.3.1	Agar media	99
C.3.2	Liquid media	100
	Appendix D Protocols	103
D.1	Boiling plasmid DNA extraction	103
D.2	Genomic DNA extraction	103
D.3	Chemical transformation of <i>E. coli</i>	104
D.4	Preparation of <i>E. coli</i> electro-competent cells	104
D.5	Transformation of <i>E. coli</i> by electroporation	105
D.6	Conjugation between <i>E. coli</i> and <i>Streptomyces</i>	105
D.7	Triparental mating using <i>S. coelicolor</i> as recipient	106
D.8	Growth conditions for production of specialised metabolites	107
D.9	Bioassays: testing the antibiotic activity	107
D.10	Chromatographic conditions	108
D.11	Metabolic profile comparison: analysis of the samples	108
	Appendix E Supplementary information for Chapter One	110
E.1	Draws of the BGC	110
E.1.1	Code used to generate the draws of the BGCs	110
E.1.2	Draw of the BGCs of <i>S. leeuwenhoekii</i>	112
E.2	Alignment of the lasso-peptides cyclization proteins	113

E.3	Similarities of the hybrid <i>trans</i> -AT PKS/NRPS proteins	114
E.4	Alignment of the hybrid <i>trans</i> -AT PKS/NRPS domains and regulator genes	117
E.5	Supplementary chromatograms	118
Appendix F Supplementary information for Chapter Two		122
F.1	Formulation of the biomass equation and their components	122
F.1.1	Biomass	122
F.1.2	Protein	122
F.1.3	RNA	124
F.1.4	DNA	124
F.1.5	Phospholipid and TAG composition	126
F.1.6	Small molecules pool	127
F.1.7	Peptidoglycan synthesis	127
F.1.8	Carbohydrate biosynthesis	128
F.1.9	Teichoic acid biosynthesis	128
F.1.10	Ions pool	128
F.1.11	Essential gene list	129
F.1.12	Growth in complex media	131
F.2	<i>S. leeuwenhoekii</i> map of reactions	132
F.3	Script used to generate and work with the GSM	132
F.3.1	Get the reaction information from KEGG database	132
F.3.2	Perform Blast and sort the output file	134
F.3.3	Get compound information	138
F.3.4	Abbreviate compounds name	138
F.3.5	Write reactions in COBRAPy	142
F.3.6	To create a python module that would contain the information to be loaded when creating the GSM	144
F.3.7	Create the model	145
F.3.8	Check mass balance	145
F.3.9	Balance protons or water	147
F.3.10	Simulate growth in complex media	148
F.3.11	Apply FSEOF	148
F.4	List of abbreviations of compounds	151
Abbreviations		164
Nomenclature		166

List of Tables

1.1	Characteristics of the <i>S. leeuwenhoekii</i> C34 genome. Adapted from (Gomez-Escribano et al., 2015).	18
1.2	Gene clusters of <i>S. leeuwenhoekii</i> C34.	19
1.3	Description of the proteins found in the lasso-peptide 1 BGC and closest NCBI database homologous.	20
1.4	Description of the proteins found in the lasso-peptide 2 BGC and closest NCBI database homologous.	21
1.5	Description of the proteins found in the lasso-peptide 3 BGC and closest NCBI database homologous.	23
2.1	Statistics of <i>iVR1007</i>	58
2.2	Essential gene analysis.	62
2.3	Predicted genes knockout targets for increasing chaxamycin A, chaxalactin A, and hybrid PK-NP production.	65
2.4	No. of overexpression targets found for each specialised metabolites.	66
2.5	Overexpression targets not directly related to the production of the compound of interest ¹	67
2.6	Overexpression targets directly related to the production of the compound of interest ¹	68
A.1	List of strains used in this work.	91
A.2	List of plasmids used in this work.	93
B.1	List of primers used in this work.	95
C.1	Microbiology solutions	98
D.1	Volumes of antibiotics used for overlay the conjugation plates.	106
E.1	Description of the proteins found in the hybrid <i>trans</i> -AT PKS/NRPS, with closest NCBI database and leinamycin BGC homologous.	114
F.1	Components of the biomass equation at a dilution rate of 0.109 h ⁻¹	123
F.2	Aminoacid composition used for <i>S. leeuwenhoekii</i> based on aminoacid composition of <i>S. tsukubaensis</i> (Huang et al., 2013). considering the energy required for polymerization of <i>E. coli</i> (Ingraham et al., 1983).	123
F.3	RNA composition of <i>S. leeuwenhoekii</i>	124
F.4	DNA composition of <i>S. leeuwenhoekii</i>	124

F.5	Fatty acid composition of <i>S. leeuwenhoekii</i> (Busarakam, 2014) and proportion present in phospholipids and in TAGs.	126
F.6	Small molecules pool.	127
F.7	Peptidoglycan composition.	127
F.8	Carbohydrates composition of cell wall.	128
F.9	Teichoic acid biosynthesis composition.	128
F.10	Ions pool composition.	129
F.11	Essential genes for <i>S. leeuwenhoekii</i> C34 for growth in complex media.	129
F.12	Uptake rates of compound used to simulate growth in complex media (ISP2).	131
F.13	Compound abbreviations.	151

List of Figures

1	Time-line of discovery of new antibiotic classes	2
2	Percentage of resistance to antibiotics of Chilean clinical isolates from 1991 until 2015	3
3	Reactions catalyzed by PKS domains	5
4	Reactions catalysed by most common NRPS domains	7
5	Representation of the general mechanism of β -branching	8
6	Leinamycin BGC as an example of hybrid <i>trans</i> -AT PKS/NRPS	8
7	Methodology used to identify specialised metabolites products of silent BGC	10
1.1	Percentages of the different BGC types found in <i>S. leeuwenhoekii</i> C34 genome.	18
1.2	Lasso-peptide 1 and lasso-peptide 2 BGC	21
1.3	Alignment of the amino-acids of the proteins Sle29720, Sle05791 and Sle2_133, and their homologous protein of lariatins BGC.	22
1.4	Lasso-peptide 3 BGC	22
1.5	Hybrid <i>trans</i> -AT PKS/NRPS BGC	23
1.6	Comparison of the hybrid <i>trans</i> -AT PKS/NRPS and leinamycin BGC.	24
1.7	Alignment of the amino-acids of the putative chlorinating enzyme (Sle09470) detected in the hybrid <i>trans</i> -AT PKS/NRPS biosynthetic gene cluster, with similar known chlorinating enzymes.	25
1.8	Prediction of the structure of the hybrid <i>trans</i> -AT PKS/NRPS	27
1.9	Metabolite profile comparison between <i>S. leeuwenhoekii</i> C34, M1600 and M1601	29
1.10	Comparison of fragmentation pattern of the ion m/z 717.87 $[M + H]^+$ RT 40.6 min	30
1.11	Metabolite profile comparison between <i>S. leeuwenhoekii</i> C34, M1614 and M1619 in negative ionization	31
1.12	Fragmentation pattern of the ion of m/z 609.77 $[M - H]^-$, and m/z 609.51 $[M - H]^-$ and m/z 611.53 $[M + H]^+$, found in the sample of mycelium extract of 2 or 5 days old liquid culture in mISP2 of <i>S. leeuwenhoekii</i> M1619, respectively	32
1.13	Mass spectrum at RT 5.2 min. of mix samples of supernatant and mycelium extract of <i>S. leeuwenhoekii</i> M1614 and M1619	33
1.14	Mass spectrum at RT 5.2 min. of mix samples of supernatant and mycelium extract of <i>S. leeuwenhoekii</i> M1614 and M1619	34
1.15	Comparison of fragmentation pattern of the ion m/z 715.16 $[M - H]^-$ RT 5.21 from sample of <i>S. leeuwenhoekii</i> M1614 and <i>S. leeuwenhoekii</i> M1619, with ion m/z 715.22 $[M - H]^-$ RT 40.09 from sample of <i>S. leeuwenhoekii</i> C34 and with ion m/z 715.84 $[M - H]^-$ RT 40.82 from sample of <i>S. leeuwenhoekii</i> M1601	35

1.16	Comparison of bioactivity of samples of 5 days production culture of <i>S. leeuwenhoekii</i> C34 from mDM and mLPM against <i>B. subtilis</i>	36
1.17	Comparison of bioactivity of methanol extraction of samples of 2 days seed culture in mISP2 against <i>B. subtilis</i>	36
1.18	Metabolite profile of <i>S. leeuwenhoekii</i> C34 grew in presence of different salt concentration	37
2.1	Methodology of genome scale reconstruction of <i>Streptomyces leeuwenhoekii</i> C34	47
2.2	GeMRA interface developed to help in the reconstruction of GSMs and to perform simulations	51
2.3	Examples of the result obtained when applied the domain comparison of proteins	53
2.4	Comparison of experimental information of growth in different carbon sources with model predictions	55
2.5	Comparison of experimental information of growth in different nitrogen and phosphorous sources with model predictions	56
2.6	Distribution of reactions of <i>iVR1007</i> in each metabolism category	58
2.7	Reactions and gaps distribution of <i>iVR1007</i>	59
2.8	Chaxamycin A biosynthesis pathway	60
2.9	Chaxalactin A biosynthesis pathway	60
2.10	Desferrioxamine B, E and G biosynthesis pathway	61
2.11	Ectoine biosynthesis pathway	61
2.12	Hybrid transAT PKS/NRPS biosynthesis pathway	62
2.13	Gene knockout search for increased chaxalactin A production	63
2.14	Gene knockout search for increased hybrid PK-NP production	64
2.15	Gene knockout search for increasing hybrid PK-NP production	64
2.16	Gene knockout search for increasing chaxamycin A production	65
2.17	Number of overexpression gene targets non-related to the biosynthesis pathways of specialised metabolites that are shared or unique	67
D.1	Generation of double recombinants of <i>Streptomyces leeuwenhoekii</i>	106
E.1	Draw of the identified BGC of <i>S. leeuwenhoekii</i>	112
E.2	Alignment of the amino-acids of the proteins Sle29740, Sle05810 and Sle2_131, and their homologous protein of lariatine BGC.	113
E.3	Alignment of the KS domains of the hybrid <i>trans</i> -AT PKS/NRPS BGC.	117
E.4	Alignment of the KR domains of the hybrid <i>trans</i> -AT PKS/NRPS BGC.	118
E.5	Alignment of the ACP domains of the hybrid <i>trans</i> -AT PKS/NRPS BGC.	118
E.6	Alignment of the PCP domains of the hybrid <i>trans</i> -AT PKS/NRPS BGC.	118
E.7	Alignment of the amino-acids of the putative regulator enzyme (Sle09280) detected in the hybrid <i>trans</i> -AT PKS/NRPS BGC, with similar proteins.	119
E.8	Alignment of the amino-acids of the putative regulator enzyme (Sle09560) detected in the hybrid <i>trans</i> -AT PKS/NRPS BGC, with TetR transcriptional regulator protein.	119
E.9	Metabolite profile comparison between <i>S. coelicolor</i> M1152 and <i>S. coelicolor</i> M1607120	120
E.10	Chromatograms of samples extracted with acetone-methanol and with methanol from <i>S. leeuwenhoekii</i> M1614 and <i>S. leeuwenhoekii</i> M1619	120
E.11	Fragmentation pattern of the ions m/z 715.16, 746.14 and 791.14 observed in the samples of supernatant and mycelium extract of <i>S. leeuwenhoekii</i> M1614 and M1619121	121

F.1 Map of reactions of *iVR1007* of *Streptomyces leeuwenhoekii* 132