The genomic evaluation of estrogen degradation by *Rhodococcus equi* ATCC 13557

Sarah Louisa Flint

A thesis submitted to Newcastle University in partial fulfilment of the requirements for

the Degree of Doctor of Philosophy in the faculty of Science, Agriculture and Engineering.

School of Civil Engineering and Geosciences

Newcastle University

Newcastle upon Tyne

NE1 7RU

February 2019

Abstract

Estrogens estrone (E1), 17β-Estradiol (E2), and estriol (E3) are steroidal hormones produced naturally, and the synthetic estrogen 17α-Ethinylestradiol (EE2) is contained in the contraceptive pill. In 2012, the European Commission proposed international environmental limits in freshwater bodies on EE2 and E2, with Environmental Quality Standard (EQS) values of 0.035 and 0.4 ng/L, respectively. Despite their low concentrations (ng/L) in the environment, several studies demonstrated their harmful effect to aquatic organisms. Some bacteria have been identified to biodegrade estrogen with varied efficiency. Biodegradation is a low cost and sustainable method that could help in the removal of estrogens.

This study aims to provide the framework with which to elucidate, identify and characterise those genes involved in estrogen degradation by a genomic evaluation of steroid degrading bacteria. Firstly, a putative estrogen biodegradation pathway was postulated, and a database of the major genes potentially involved in it was compiled, by combining existing information on the degradation of estrogen, and its better characterised analogue – testosterone, for which there is a complete degradation pathway.

The genome of *Rhodococcus equi* ATCC13557, a known estrogen degrading bacterium, was sequenced, assembled and mapped. Genome analysis revealed a gene cluster potentially coding for enzymes involved in estrogen degradation; 3-carboxyethylcatechol 2,3-dioxygenase, 2-hydroxymuconic semialdehyde hydrolase, 3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase, 3-(3-hydroxy-phenyl)propionate hydroxylase, 3-oxosteroid 1-dehydrogenase and transcriptional regulator IcIR family. Specific primer sets were developed to target these genes, and using quantitative reverse transcription PCR (RT-qPCR), gene expression when exposed to E2 and EE2, was quantified relative to the control, without estrogen, and normalised to the reference genes, malate dehydrogenase and arginine deiminase. Gene expression of putative cytochrome P450 monooxygenase was upregulated initially in the presence of estrogen, therefore may be involved in one of the first steps in estrogen degradation. The RT-qPCR experiment was complemented by biodegradation studies, however, metabolites analysis is required to confirm the postulated estrogen biodegradation pathway.

Author's declaration

I hereby certify that all work presented is my own, except where otherwise acknowledged.

Sarah Louisa Flint

Table of Contents

Table of figures

Table of tables

Table of equations

Abbreviations

3-HSA	3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione
4-9 DHSA	4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oic acid
4-AD	Androst-4-ene-3,17-dione
4-OH-E1	4-hydroxyestrone
4-OH-E2	4-hydroxyestradiol
4-OH-EE2	4-hydroxyethinylestradiol
AcN:MeOH	Acetonitrile:methanol
ADD	Androsta-1,4-diene-3,17-dione
AOB	Ammonia oxidising bacteria
AOP	Advanced oxidation processes
AS	Activated sludge
ATCC	American Type Culture Collection
BLAST	Basic Local Alignment Search Tool
CA	Cellulose acetate
CCL	Contaminant Candidate List
CCT	CGView Comparison Tool
cDNA	complementary DNA
CDs	coding sequences
CFU/Ml	Colony forming units per millilitre
CIC	Combined injectable contraceptives
CNTs	Carbon nanotubes
COCP	Combined oral contraceptive pill
DCMD	Direct contact membrane distillation
ddTTP	2',3'-dideoxythymidine triphosphate
DNA	Deoxyribonucleic acid
dNMP	deoxyribonucleoside monophosphates
dsDNA	double-standed DNA
dT	thymidylic acid
dw	dry weight
E0	Estratetraenol
E1	Estrone

E2	17β-estradiol
E3	Estriol
ECDs	Endocrine disrupting chemicals
EDSP	Endocrine Disruptor Screening Program
EDTA	Endocrine Disruptors Testing and Assessment
EE2	17α-Ethynylestradiol
EQS	Environmental Quality Standard
FDA	US Food and Drug Administration
GAC	Granular activated carbon
GAGE-B	Genome Assembly Gold-Standard Evaluation for Bacteria
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC	guanine-cystosine
GC/MS	Gas chromatography mass spectrometry
gDNA	genomic DNA
HPLC	High-performance liquid-chromatography
HRT	Hormone Replacement Therapy
HRT	Hydraulic Retention Time
LC	Liquid-Chromatography
LC/MS/MS	Liquid-Chromatography tandem mass spectrometry
LOD	Limit of detection
Log K_{ow}	partition coefficient octanol:water
MIP	Molecularly imprinted polymer
MIQE	Minimum Information for Publication of Quantitative Real-Time PCR experiments
MP	Mate-paired
mRNA	messenger RNA
MS/MS	Tandem mass spectrometry
MSM	Minimal Salts medium
ND	Not detected
NGS	Next Generation Sequencing
NTC	No Template Control
OD	Optical Density
OECD	Organisation for Economic Cooperation and Development
PAC	Powdered activated carbon
PAH	Polycyclic Aromatic Hydrocarbon
PDBGs	Paired de Bruijn graphs

PE	Paired-end
PES	Polyethersulfone
PGM	Personal Genome Machine
PNEC	Predicted-no-effect concentration
PTFE	Polytetrafluoroethylene
PVDF	Polyvinylidene fluoride
qPCR	real-time PCR
QUAST	Quality Assessment Tool for Genome Assemblies
RAST	Rapid Annotation using Subsystem Technology
RNA	Ribonucleic acid
-RT	negative reverse transcription
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
SBS	Sequencing by synthesis
SNP	Single nucleotide polymorphism
SPE	Solid phase extraction
SPEED	Strategic Programs on Endocrine Disruptors
SS	Suspended solids
ssDNA	single-stranded DNA
SWCNTs	Single Walled Carbon nanotubes
TCA	Tricarboxylic acid
UCM	Unregulated Contaminant Monitoring
uHPLC	Ultra-high-performance liquid-chromatography
UK	United Kingdom
USA	United States of America
US-EPA	United States- Environmental Protection Agency
VTG	Vitellogin
WFD	Water Framework Directive
WGS	Whole Genome Sequencing
WHI	Womens Health Institute
WHO	World Health Organisation
WwTP	Wastewater Treatment Plant
YES	Yeast Estrogen Screen

Acknowledgements

I would like to thank my supervisors Dr Russell Davenport, Dr Jan Dolfing, Dr Wojciech Mrozik and Dr Paola Meynet for their support, advice and guidance throughout this project.

I would like to thank the Engineering and Physical Sciences Research Council for providing funding to support this research project.

I would like to thank PrimerDesign for awarding the Gold Student Sponsorship to me, as it provided valuable training, support and resources which helped with the gene expression experiments.

I would like to thank Dr Lucy Eland, Dr Martin Sim and Dr Anil Wipat for their collaboration and for providing the genome sequence of *R. equi* ATCC13557.

I would like to thank Dr Ben Allen, Dr Alison MacFadyen and James Skelton for providing bioinformatics support and advice.

Thank you to anyone else who has provided me with support.

Finally, thank you to my partner, daughter and family for their continued support and encouragement.

Chapter 1: General Introduction

1.1 What is estrogen?

Since the 1990's there have been increasing concerns about the concentrations of the endocrine disrupting (ED) steroidal estrogens (Fig. 1- 1), estrone (E1) $C_{18}H_{22}O_2$, 17β-estradiol (E2) $C_{18}H_{24}O_2$, estriol (E3) $C_{18}H_{24}O_3$, and 17α-ethinylestradiol (EE2) $C_{20}H_{24}O_2$, in the aquatic environment and their effects upon aquatic organisms (Sumpter and Jobling, 1995). The estrogens E1, E2 and E3 are naturally occurring steroid hormones derived from cholesterol and released from the ovaries, adrenal cortex, testes, adipose tissue, other non-reproductive tissues and the placenta in humans (Simpson *et al.,* 1994; Cui *et al.,* 2013; Hanukoglu, 1992; Nelson and Bulun, 2001; Simoni *et al.,* 2002; Shutt *et al.,* 1974). E1 is mainly produced in the adipose tissue derived from the androstenedione substrate. E3 is mainly produced in the placenta derived from 16α-hydroxydehydroisoandrosterone sulfate. E2 is mainly produced in the ovaries derived from cholesterol (Simpson *et al.,* 1994; Cui *et al.,* 2013). Although E2 is often considered a female hormone, it may have important reproductive functions in males (Schulster *et al.,* 2016). Furthermore, the biosynthesis of estrogens is also present in animals including mammals (Andaluri *et al.,* 2012; Ray *et al.,* 2013). However, EE2 is a synthetic derivative of E2 used pharmaceutically as birth control (Fig. 1- 1).

The estrogens E1, E2, E3 and EE2 are one of the five major classes of steroid hormones derived from cholesterol, therefore they share a similar structure consisting of a phenolic ring (A ring), two cyclohexane ring (B and C rings) and a cyclo-pentane ring (D ring) (Fig. 1- 1) (Adeel *et al.,* 2017). Whereas, the A ring in progesterone, testosterone, cortisol and aldosterone contain a C-3 ketone, a Δ^4 bond and a C-19 methyl group (Baker, 2013; Berg *et al.,* 2002). The main differences between the different estrogens are within the D ring: where E1 has a carbonyl group on C17; E2 has a hydroxyl group on C17; E3 has two alcohol groups on the C16 and C17; and EE2 has ethinyl groups on the C17α position (Fig. 1-1) (Adeel *et al.,* 2017).

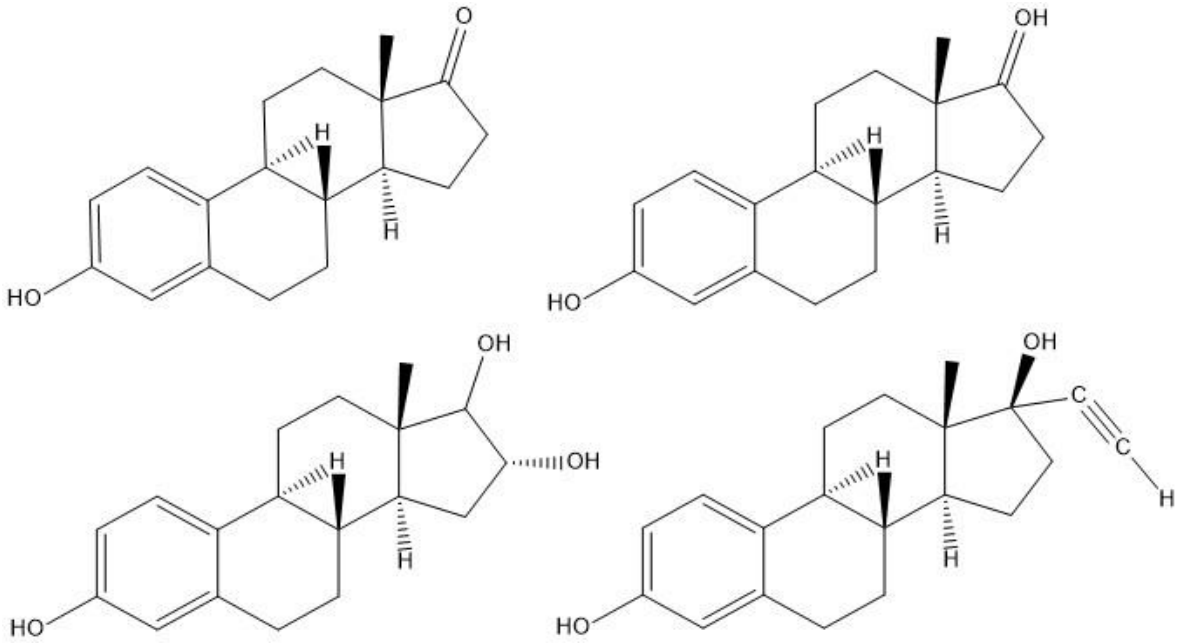

Figure 1-1: The chemical structures of E1 (top left), E2 (top right), E3 (bottom left), and EE2 (bottom right) produced using ChemDrawUltra. The rings beginning from the left of each structure are the aromatic A ring, B ring, C ring and the D ring.

In metabolism of estrogen within the body, the cytochrome P450 enzyme hydroxylates the 'A' ring and leads to its aromatisation (Thomas and Potter, 2013; Xu *et al.,* 2005; Simpson, 1994). Cytochrome P450 catalyses the conversion of androgens into estrogens such as the aromatisation of testosterone to E2 (Thomas and Potter 2013). The 17β-hydroxysteriod dehydrogenase type 1 converts E1 to E2, and type 2 converts E2 to E1 (Thomas and Potter, 2013; Xu *et al.,* 2005). Furthermore E1 is aromatised by cytochrome P450 and converted to 16α-Hydroxyestrone, which is reversibly converted into E3 by 17β-hydroxysteriod dehydrogenase. Thus E3 is a metabolite of E1 (Thomas and Potter, 2013, Xu *et al.,* 2005; Simpson, 1994). Following hydroxylation, the subsequent eliminations of estrogen from the body are through sulfonation, glucuronation or methylation (Thomas and Potter, 2013; Xu *et al.,* 2005; Wang *et al.,* 2004; Helton *et al.,* 1977).

1.2 Pharmaceutical usage of estrogen

EE2 is used in birth control such as the combined oral contraceptive pill (COCP), combined injectable contraceptives (CIC), contraceptive patches and rings. It is also used in hormone replacement therapy (HRT). HRT was first used in the 1950's; however in 2002 it was concluded in the Woman's Health Institute (WHI) that the usage of HRT could lead to complications such as blood clots, stroke and breast cancer (Manson *et al.,* 2013). Therefore, although there is still some usage of HRT, the usage of HRT declined from 2002 to 2010 by 50-77% and usage down to lower than 10% in some European countries (Ameye *et al.,* 2014). In contrast the usage of the COCP has increased globally, especially in Asia and Latin America since its approval by the US Food and Drug Administration (FDA) for usage as a contraceptive in 1960 and in 1961 by the House of Commons in the United Kingdom (UK) (World Health Oragnisation, 2017). According to the World Health Organisation (WHO), the global usage of modern contraceptives, including hormonal contraceptives, was 54% in 1990 and 57.4% in 2015. This included Africa, Asia and America and the Caribbean with usage of 28.5%, 61.8% and 66.7% in 2015 respectively (World Health Organisation, 2017). Furthermore, there has been increase in the uptake of hormonal contraceptives within the UK with an increase of 5.3% between 2002 and 2011 with the COC being the most commonly prescribed (Rashed *et al.,* 2015). In 2007 it was reported by the United Nations that there are over 100 million users of the COC, mainly within Western Europe, the United States of America (USA) and Northern Europe and there are approximately 80% of permanent users in the USA (United Nations, 2007; Brynhildsen, 2014; Guttmacher Institute, 2013). In 2015/16 the usage of COC was the most common method for 45% of women and COC accounted for 5.9% of freely dispensed prescriptions in 2015 (NHS, 2016; Prescribing and Medicines Team Health and Social Care Information Centre, 2016).

1.3 The concentration of estrogen detected in the aquatic environment

Endocrine disrupting steroidal estrogens, E1, E2, E3 and EE2, have been detected in the environment worldwide and at various concentrations within surface water (Table 1- 1) adapted from (Hassani *et al.,* 2016; Wise *et al.,* 2011; Rao *et al.,* 2013; Zhang *et al.,* 2014).

Table 1-1: Concentrations of estrogens detected within surface water.

Country	**Sampling area**	**Concentration (ng/L)**				**Reference**
		E1	E2	E3	EE2	
China	Jiulongjiang river	<LOD-11.23	<LOD-52.71	<LOD-47.85	ND	Yuan *et al.,* 2014
UK	Rivers Arun and Ouse	2.9-4.6	0.6	-	-	Peck *et al.,* 2004
The Netherlands	Dutch surface water	0.3-7.2	0.8-1.0	-	0.3-0.4	Vethaak *et al.,* 2005
Belgium	Scheldt estuary	0.37-10	ND	-	ND	Noppe *et al.,* 2007
Germany	Rhine Agger river	1.3-9.2	-	-	0.3-1.0	Hintemann *et al.,* 2006
China	Dagu river	5.0-55.3	0.9-33.4	2.3-46.4	<LOD-35.6	Lei *et al.,* 2009
China	Rivers in Tianjin	0.64-50.7	<LOD-31.4	<LOD-37.2	<LOD-24.4	Rao *et al.,* 2013
Taiwan	Dan-Shui river	22-66.2	1.4-33.9	12-73.6	7.5-27.4	Zhang *et al.,* 2014
China	Lake Taihu	<LOD-15.8	<LOD-10.8	-	<LOD-16.4	Lu *et al.,* 2011
France	Jalle d'Eysines	17.2-71.4	<0.1-4.4	<0.1-2.9	ND	Labadie and Budzinski, 2005
Austria	>400 samples from Austria	<LOD-4.6	<LOD-1.2	<LOD-1.9	<LOD-0.33	Hohenblum *et al.,* 2004
Japan	Tama River, Tokyo	8.7-19.7	ND	-	0.1-5.1	Kawaguchi *et al.,* 2004
China	The Pearl River Delta	1.5-11.5	-	-	-	Peng *et al.,* 2008

China	The Yangtze River, Nanjing	0.8-1.25	0.8-0.97	5.79	-	Lu *et al.,* 2010
China	Yangtze River	0.9-2.98	0.26-1.8	<LOD-4.37	<LOD-2.67	Jiang *et al.,* 2012
China	Jiangsu	-	5.5-15.5	ND	5.7-30.8	Shen *et al.,* 2001
China	Shandong	14-180	<LOD-134	4-94	7-24	Zhou *et al.,* 2011
China	Beijing	<LOD-2.3	-	<LOD-13.9	5.0-127.9	Sun *et al.,* 2009
China	Beijing	1.1	0.2	ND	ND	Hu *et al.,* 2005
China	The Pearl Rivers	<LOD-75.0	<LOD-7.5	-	-	Zhao *et al.,* 2009
China	Lioaning	<LOD-55.8	<LOD-7.4	-	ND	Wang *et al.,* 2011
China	Taiwan	22.4-66.2	1.4-33.9	12.4-73.6	7.53-27.4	Chen *et al.,* 2007
Iran	Karun river	-	13.66	-	-	Hassani *et al.,* 2016
The Netherlands	Coastal estuarine and freshwater	0.1-3.4	0.3-5.5	-	0.1-4.3	Belfroid *et al.,* 1999
Germany	River Danube	<0.1-4.1	<0.15-3.6	-	<0.1-5.1	Kuch and Ballschmitter, 2000
Germany	3 rivers	0.7-1.6	ND	-	ND	Ternes *et al.,* 1999
Canada	Thames river	16-49	6-14	-	-	Lishman *et al.,* 2006
Portugal	Aveiro district	-	85	-	-	Silva *et al.,* 2013
USA	19 surface waters	0.3	17	-	1.4	Benotti *et al.,* 2009
China	Rivers surrounding Dianchi Lake	2-305.8	1.9-58.5	<LOD-61.5	<LOD-206.5	Huang *et al.,* 2013

ND- Not detected or below the limit of detection (<LOD). (-) Some studies did not measure the concentrations of all the 4 estrogens so the values were not available.

However, the concentrations of estrogens in rivers and surface waters are affected by the sources of estrogen located nearby such as hospitals, population sizes, farming, wastewater treatment plant (WwTP) effluents, and temperature (Hassani *et al.,* 2016).

Furthermore, estrogens have also been detected in the influent and effluent at WwTP, where the effluents are released into rivers (Table 1- 2)

Table 1-2: Example concentrations of estrogens detected in WwTP influent and effluent

Compound	**Influent concentrations from WwTP (Hamid and Eskicioglu, 2012; Huang *et al.,* 2013).**		**Effluent concentrations in WwTP (Hamid and Eskicioglu, 2012; Huang *et al.,* 2013).**	
	Lower (ng/L)	Higher (ng/L)	Lower (ng/L)	Higher (ng/L)
Estrone (E1)	2.4	670.0	<0.3	96.0
17β-estradiol (E2)	2.4	150.0	0.2	30.0
Estriol (E3)	23.0	660.0	<0.5	275
17α-Ethinylestradiol (EE2)	<0.3	70.0	<0.3	26

In addition to detection within surface waters, estrogens have been detected in river sediments. In a study of Rivers Ouse and Arun in the UK, the concentrations of E1 and E2 were higher within the sediment than in the surface water, with concentrations of 24.9-52.5 ng/kg dry weight (dw) and 6.3- 14.6 ng/kg dw, respectively (Peck *et al.,* 2004). Also the estrogenicity of the sediment was 532- to 748- fold higher than in the surface water (Peck *et al.,* 2004). Therefore estrogens adsorb to the suspended and bed sediments within rivers due to the partition coefficient octanol-water (log K_{ow}) values of 3.1-3.4 and 3.8-4.0 for E1 and E2 respectively (Peck *et al.,* 2004; Ternes *et al.,* 1999; Lai *et al.,* 2000). Furthermore higher concentrations of <40-355 ng/kg dw for E1 were detected in rivers Lea and

Nene in the UK, although E2 and EE2 were below the LOD (<100 ng/kg dw) (Williams *et al.,* 2003).

Furthermore estrogens have been detected within soil due to run-off or usage in agriculture with amendment of soil by use of broiler litter, leading to estrogenicity up to 675 ng/kg dw as E2 equivalent and 50-2300 ng/kg dw as E2 equivalent due to surface run-off (Finlay-Moore *et al.,* 2000).

1.4 The sources of estrogen in the environment

The concentrations of estrogen detected in the environment can vary dependent upon the location of the sampling. For example, if it is near a city or rural location and the number of WwTP leading into the rivers. Therefore it is important to consider the different sources of estrogen which may lead to increased levels of estrogen within the environment.

One of the larger sources of estrogens in WwTP has been attributed to the estrogenicity of the natural estrogens being excreted in urine and faeces, with excretion rates of estrogen equivalent for the natural estrogens E1, E2 and E3 accounting for 66-82% (Liu *et al.,* 2009). The excretion rates of estrogen were the highest in pregnant women of 4336.8 µgE2/d, followed by premenopausal women, postmenopausal women and the lowest in men of 9.937, 4.38, and 3.103 µgE2/d, respectively (Liu *at al.,* 2009). Although, according to total population figures of 2001, the excretion of synthetic EE2 as a fraction of total estrogens is likely to be less than 1% in the US. The main contributor of EE2 to the influent at WwTP is excretion, as 60% of ingested EE2 is excreted (Wise *et al.,* 2011; Johnson *et al.,* 2004). Furthermore, 0.21 ng/L of E1 was detected within river water where the source of estrogen was attributed to a few scattered dwellings, although this was lower in concentration than the rivers upstream and downstream of WwTP (Long *et al.,* 2014).

Estrogens have been detected in the effluent and WwTP influent from a hospital in 2007 with 9628 patients. The concentrations of estrogen were measured up to 31.3 ng/L, 4.2 ng/L and 385.5 ng/L for E1, E2 and E3 respectively, but due to the high dilution of hospital effluent the impact upon the WwTP influent was

considered negligible (Avberšek *et al.,* 2011). Another study also found that there was no significant difference between the hospital effluent and the WwTP influent. However E2 was only detected once, whereas E1 and EE2 were detected in all of the samples, with concentrations ranging up to 58.3, 16.9 and 18.3 ng/L in the influent (Pauwels *et al.,* 2008). The average concentrations of E2 detected in the hospital effluent of two hospitals in Iranian cities were much higher, with concentrations of 93.53 and 83.22 ng/L (Hassani *et al.,*2016). Furthermore, for hospital effluent it is not possible to attribute all the estrogen in the effluent to human excretion as there may be other sources in a hospital such as the disposal of pharmaceutical wastes and drugs (Avberšek *et al.,* 2011; Pauwels *et al.,* 2008; Hassani *et al.,* 2016).

Discharge of estrogens in the effluent from pathology laboratories in the pharmaceutical industry were suggested to be the source of increased E2 concentrations in urban wastewater of 31.3-69.4 ng/L (Hassani *et al.,* 2016).

Industrial landfill leachate E2 and EE2 along with bisphenol A were suggested to be the major contributors to *invitro* estrogenicity with concentrations of 2.8 and 17 ng/L (Svenson *et al.,* 2009).

Additionally, agricultural livestock excrete the natural estrogens E1, E2 and E3. The concentrations of estrogen are higher in some cases than for human excretion, with the concentration of E2 being 30-2500ng/L in animals and 0.5-125 ng/L (Wise *et al.,* 2011; Combalbert *et al.,* 2010). The excretion of all natural estrogens by livestock can range from 14-360 µg/d in faeces, 2.5-2000 µg/d in urine and 0.07-990 µg/d in total. The largest amount is excreted by larger animals such as cows, compared to chickens, and is higher when the animals are pregnant (Combalbert *et al.,* 2010). The runoff and effluents from the livestock can enter the environment through leaching into the soil, and livestock have more than twice the rate of human discharge in the US and Europe at 83000 kg/yr (Kjær *et al.,* 2007; Wise *et al.,* 2011). Furthermore, in solid sludge, which can be used in agriculture as a fertiliser and conditioner of soil, there are lower concentrations; LOD - 63, 22, 13, and 18 µg/kg dry weight for E1, E2, E3, and EE2, respectively (Andersen *et al.* 2003; Braga *et al*. 2005; Fernandez *et al*. 2009; Gabet-Giraud *et al.* 2010; Muller *et al.* 2008; Ternes *et al.* 2002; Combalbert and Hernandez-

Raquet, 2010). However, there are also higher concentrations of 50–200 µg/kg dw for E1, 272–406 µg/kg dw for E3 and 0.64–7 µg/kg dw for sulfate-conjugated forms of E2 and E1 (Nieto *et al.,* 2008). Therefore sludge usage in agriculture is another potential source of estrogen in the environment.

Thus the sources of estrogens include excretion and disposal by humans, agriculture and the pharmaceutical industry.

1.5 The environmental fate of estrogens in the environment and their effects

Estrogens E1, E2, E3 and EE2 have been detected at ng/L concentrations within the environment (Table 1- 1). There are a number of studies into the persistence of E1, E2, E3 and EE2 within the environment, in which the half-lives have been found to be 0-1.1 days for E2 (Lee *et al.,* 2003), 2.8-4.9 days for E1 (Lucas and Jones, 2006), 0.7-1.7 days for E3 (Carr *et al.,* 2011) and 4-6 days for EE2 (Yin *et al.,* 2002). However, there are different half-lives reported under different conditions, such as EE2 having a half-life of 36 hours within natural water in the laboratory (Zheng *et al.,* 2011). Furthermore under anaerobic conditions the half-lives are much longer, for example E2 had a half-life of 24 days in soil and little or no degradation of other estrogens was observed (Ying and Kookana, 2005). In rivers the biodegradation rates may depend upon the seasons and the position of the contamination within the river, as the degradation of downstream summer samples was fastest due to factors affecting the microbiological community (Jùrgens *et al.,* 2002). Furthermore aerobic degradation of EE2 was much slower with a half-life of 17 days, compared to E2 with a half-life of 1.2 days (Jùrgens *et al.,* 2002). Generally the half-life of EE2 is the longest and therefore suggests the most persistence of the estrogens, and if compared to the half-life of another steroid hormone testosterone of 0.08-0.8 days (Lee *et al.,* 2003), there is a greater persistence of estrogens. Therefore the risks associated with the presence of estrogens within the environment need to be considered, as some concentrations of estrogens detected in the environment (Table 1- 1) are above the predicted-no-effect concentrations (PNECs) for estrogens for long-term exposure in surface water (>60 days); 6, 2, 60 and 0.1 ng/L for E1, E2, E3 and EE2 respectively

(Caldwell *et al.,* 2012). In some cases the concentrations are above the PNECs for short-term exposure (<60 days); 20, 5, 200 and 0.5 ng/L for E1, E2, E3 and EE2 are recommended (Caldwell *et al.,* 2012).

Vitellogenesis is the process in which vitellogin (VTG) is produced by the liver and then uptaken by the oocytes during maturation to provide nutrition to a growing embryo (Sumpter and Jobling, 1995). Cellular based studies are often used to measure the effects of exposure to estrogens. For example the hepatocytes of rainbow trout have been used to measure the estrogen receptor-mediated VTG synthesis, induced when the cells are exposed to estrogenic substances (Hollert *et al.,* 2005). There are many studies into the effects within living aquatic organisms. For example, it was found that exposure to estrogenic substances within the aquatic environment may lead to the expression of the VTG gene in male fish in a concentration-dependent manner, as fish placed in cages closest to WwTP effluents along a river had the highest concentrations of VTG (Sumpter and Jobling, 1995; Bon *et al.,* 1997; Copeland *et al.,* 1986). A 300-fold increase in the concentration of VTG produced in a male trout placed directly into the effluent of a WwTP was observed after a short exposure of one week (Sumpter and Jobling, 1995). In the UK trouts were placed directly into WwTP effluent in 10 different water authorities. In 13 of the sites the trouts did not survive the 3 weeks and in the other 15 sites the VTG concentrations were increased between 500 to 50000-fold (Sumpter and Jobling, 1995).

Furthermore 5% of roach (*Rutilus rutilus)* in the UK downstream of WwTP have been found to be hermaphrodites (Sumpter and Jobling, 1995). The effect has been found to be concentration dependent and when exposed to various concentrations of effluent, the production of VTG and feminisation of the male reproductive ducts was dose-dependent (Rodgers-Gray *et al.,* 2001). Also the changes to the male reproductive ducts, observed due to early exposure, were permanent. However the concentrations of VTG were reduced when the *Rutilus rutilus* were in water absent of WwTP effluent for 150 days (Rodgers-Gray *et al.,* 2001). Furthermore, the feminisation of male fish by exposure to WwTP effluent contamination, compared to a reference river, led to reduced fertility of *Rutilus rutilus,* as there was a reduction in gamete production leading to a reduction in the number of male fish with moderate or severe feminisation that could produce

sperm and seminal fluid (milt). Furthermore, of those that could, their production of milt was up to 50% less and the motility of sperm was reduced (Jobling *et al.,* 2002). If the fertility of male fish is reduced, there may be a decline in populations of *Rutilis rutilus.*

These studies focused mainly on the effects of exposure to WwTP effluents and did not specifically measure the concentrations of chemicals present or the estrogenicity of the effluent. Therefore these studies demonstrate that there are chemicals within WwTP effluents which have effects on the reproductive function of male fish and in extreme conditions may be fatal and detrimental to population sizes. However these effects are likely attributed to estrogenic activity, as in a laboratory based study there was a response to 0.1-0.5 ng/L of EE2 (Purdom *et al.,* 1994). Also rainbow trout (*Oncorhynchus mykiss*) exposed to 100 ng/L EE2 had increased plasma protein, phosphoprotein and calcium concentrations after two weeks exposure (Verslycke *et al.,* 2002), although this is a concentration higher than those in detected in the aquatic environment and these studies were not within the aquatic environment but within the laboratory.

Declines in fish populations within the Upper Danube River in Germany may be due to estrogen; five of nine sediment extracts had estrogenic activities, of which E1 was the main contributor, contributing to 60% of the sum of E2-equivalents, although E2 and EE2 were below the limits of detection (Grund *et al.,* 2011). However further studies are required to determine if the estrogenic activities are the causative factor for the decline in fish populations (Grund *et al.,* 2011).

Fathead minnow (*Pimephales promelas*) suffered endocrine disruption, genotoxic damage, modulated immune function, and altered metabolism in a sex-specific and dose-dependent manner when exposed to a potent estrogenic and a weakly estrogenic effluent containing E1, E2 and EE2, and in response to EE2 alone (Filby *et al.,* 2007). However fewer of these effects were seen when exposed to a weaker estrogenic effluent and the effects were different when exposed to EE2 in a mixture compared to when EE2 was alone, suggesting that environmental conditions are difficult to replicate within the laboratory (Filby *et al.,* 2007).

In a whole lake study carried out over 7 years in which *Pimephales promelas* were exposed to environmentally relevant concentrations of 5-6 ng/L of EE2, similar

effects were observed to those seen in response to exposure to WwTP effluents and exposure to estrogen within laboratory settings (Kidd *et al.,* 2007). By seven weeks of exposure to EE2 there was increased concentrations of VTG in both male and female fish, but the increase in males was by three orders of magnitude, and the VTG messenger RNA (mRNA) was also increased (Kidd *et al.,* 2007). Additionally reductions in fertility were observed, with delayed spermatogenesis, widespread fibrosis, malformation of tubules and ova-testes in four out of nine male fish after 2-3 years exposure (Kidd *et al.,* 2007). In addition some effects of chronic exposure were observed in females, such as delayed ovarian development and the presence of atretic follicles (Kidd *et al.,* 2007). Overall there were observations of population collapse, which suggests that there is a great risk to populations of short-lived fish species when chronically exposed to concentrations of EE2 detected within the aquatic environment (Kidd *et al.,* 2007).

The main focus of studies is often the effects of exposure to EE2 and E2, but exposure to E1 is rarely investigated. However, exposure to E1 has been attributed to undesirable effects in aquatic organisms. In sediments from the Upper Danube, Germany, the main contributor to estrogenicity was E1, with E2 and EE2 being below the limits of detection. Suggesting that more studies into the effects of E1 are required (Grund *et al.,* 2011; Ankley *et al.,* 2017). Furthermore when *Pimephales promelas* were exposed to E1 for 4 days at environmentally relevant concentrations, this resulted in an increase in the plasma E2 concentration, in a dose-dependent manner at above 15.6 ng/L of E1, to a level comparable to that in females (Ankley *et al.,* 2017). Furthermore VTG mRNA expression and concentrations increased in a dose-dependent manner above 41.9 ng/L and when exposed to C^{13}-E1 lead to C^{13}-E1, β-E2, and α-E2 elevation in the plasma, suggesting that fish have the potential to convert E1 to E2 and therefore the potential for feminisation through exposure to E1 (Ankley *et al.,* 2017).

Therefore low concentrations of estrogens (ng/L) detected in the environment (Table 1- 1) have been found to produce undesired effects upon aquatic organisms, such as vitellogenesis in male fish, endocrine disruption, altered embryo and egg production, altered physiology endpoints and subsequently reduced population sizes (Segner *et al.,* 2003; Porte *et al.,* 2006; Sumpter and

Jobling, 1995; Matozzo *et al.,* 2008; Depledge and Billinghurst, 1999; Jobling *et al*, 2004; Souza *et al.,* 2013; Kidd *et al.,* 2007; Jobling and Owen, 2013).

1.6 Legislation

The effects of endocrine disrupting chemical (ECDs) such as estrogens has led to a number of countries and organisations implementing programs or adding ECDs to their existing chemical safety strategies. These include the European Union, the Joint Working Group on Endocrine Disrupters Testing and Assessment (EDTA) of the Organisation for Economic Cooperation and Development (OECD), the Endocrine Disruptor Screening Program (EDSP) of the United States Environmental Protection Agency (US-EPA), and the Strategic Programs on Endocrine Disruptors (SPEED) of the Japan Environment Agency (Hecker and Hollert, 2011).

However there have been no specific legislations implemented for the regulation of estrogen in the aquatic environment or in drinking water. However, due to increasing concerns, the estrogens, synthetic EE2 and E2 have been recommended for monitoring. In January 2012, the European Commission proposed the estrogens, EE2 and E2 as priority substances, with proposed EQS of 0.035 ng/L and 0.4 ng/L respectively (Gilbert, 2012; Johnson *et al.,* 2013). Furthermore the UK government's Department for Environment, Food and Rural Affairs is proposing an EQS of 0.016 ng/L for EE2 (Gilbert, 2012). Therefore the Water Framework Directive (WFD) passed the 2013/39/EU65 directive in August 2013 in which EE2 and E2 were included in the surveillance list (da Cunha *et al.,* 2016). Despite this, the regulations have been postponed from implementation due opposition by water utilitity companies such as Thames Water, UK which do not accept the inclusion of the additional substances to the Priority substances directive, as they claim that studies and proposals lack evidence of the risks to health, have large economic implications and could lead to increased carbon emissions (Water quality, 2013).

Within the USA, there is only monitoring of EE2 and other endocrine disruptors within drinking water. In 2009 a Contaminant Candidate list (CCL) was produced containing EE2 and endocrine disruptors for monitoring in the Unregulated

Contaminant Monitoring (UCM) program by the US-EPA (da Cunha *et al.,* 2016). However estrogen is not currently monitored in the aquatic environment within the USA.

1.7 Chemical and physical removal methods of estrogen

1.7.1 *Sewage treatment plants*

Removal of E1 and E2 at WwTP has been reported at >98 % (Kanda and Churchley, 2008). Furthermore, some WwTP can remove EE2 to 0.1 ng l^{-1}, concentrations below the PNEC for fish (Lawton *et al.,* 2008). However, removal efficiency is dependent upon the treatment processes and operating parameters (Andersen *et al.,* 2003; He *et al.,* 2013; Ye *et al.,* 2012; Clara *et al.,* 2005; UKWIR, 2009).

Possibilities include advanced oxidation processes (AOP) and physical methods, such as sorption and membrane filtration (Silva *et al.,* 2012) (Table. 1- 2), although these are often energy-intensive and costly.

Table 1-3: Efficiency and disadvantages of estrogen removal methods.

Method	Estrogen removal	Disadvantages	References
Sorption	• 90% E1, E2, E3 and EE2 by granular activated carbon (GAC) and powdered activated carbon (PAC). • 80% EE2 by molecularly imprinted polymer (MIP). • 97% E2 by MIP. • 96.14% to 98.03% E1 by carbon nanotubes (CNTs) with MIP. • 95-98% EE2 by single walled carbon nanotubes (SWCNTs). • 50-75% E1, E2 and EE2 by suspended solids (SS).	• Efficiency affected by the presence of natural organic matter. • pH, temperature and method dependent. • Numerous studies with removal of less than 50%.	Snyder *et al,* 2007; Clara *et al.,* 2004; Neale *et al.,* 2010; Ren *et al.,* 2007; Yamamoto and Lijestrand, 2003; Andersen *et al.,* 2005; Andersen *et al.,* 2003; Muller *et al.,* 2008; Ternes *et al.,* 2004; Meng *et al.,* 2005; Lai *et al.,* 2010; DeMaleki *et al.,* 2010; Gao *et al.,* 2011; Joseph *et al.,* 2011.
Membrane filtration	• 99-100% E1, E2, E3 and EE2 by nanofiltration (NF). • 98 – 100% E1, E2, E3 and EE2 by reverse osmosis or NF. • 99% E1 and E2 by forward osmosis. • ≥99.5% E1 and E2 by direct contact membrane distillation (DCMD).	• Dependent upon equilibrium concentration. • Dependent on physiochemical processes, membrane properties, membrane fouling, pH, temperature and salinity.	Silva *et al.,* 2012; Bellona *et al.,* 2004; Bolong *et al.,* 2009; Liu *et al.,* 2009; Schäfer *et al.,* 2011; Weber *et al.,* 2004; Alturki *et al.,* 2010; Cartinella *et al.,* 2006.
AOP	• 99-100% E2 and EE2 by Chlorination and ozonation.	• pH dependent. • Production of harmful by-products with estrogenic activity	Bila *et al.,* 2007; Kim *et al.,* 2004; Alum *et al.,* 2004.

1.7.2 *Biodegradation of estrogen*

Biodegradation, another removal method, has achieved promising results in some instances, including complete mineralisation of EE2 (Yoshimoto *et al.,* 2004). Estrogen-degrading bacteria have been isolated from WwTP (Table 1- 3); therefore naturally occurring biodegradation processes may provide a sustainable solution not entailing excessive costs or energy. Concentrations of estrogen-degrading bacteria have been measured by real-time PCR (qPCR) (Roh and Chu, 2010; Hashimoto *et al.,* 2009). The concentrations of *Sphingomonas* sp. KC8 were 2-3 orders higher than those of *Novosphingobium tardaugens* sp. ARI-1 in three WwTP; suggesting *Sphingomonas* sp. KC8 may be ubiquitous in WwTP (Roh and Chu, 2010). Additionally, *Novosphingobium* sp. JEM-1 was generally distributed between two WwTP and the cell numbers were correlated with estrogen removal (Hashimoto *et al.,* 2009). Furthermore, *Sphingomonas* sp. AHC-F and *Sphingobium* sp. AX-B immobilised by use of the bacterial cell immobilisation technique degraded 2 mg/L of E2 and its metabolite E1 (Ma *et al.,* 2016). Also when using continuous flow with 850 ng/L of E2 in the influent, the removal efficiency for E2; 94%, and for equivalent estrogenic quantities of E2; 87%, with a hydraulic retention time (HRT) of 12 h (Ma *et al.,* 2016). When the HRT is increased to 72 h the E2 and E1 are undetected suggesting 100% removal efficiency (Ma *et al.,* 2016).

1.8 Known estrogen degrading bacteria

Few estrogen-degrading bacteria have been identified, with fewer EE2-degrading bacteria being identified than other estrogen degrading bacteria (Tables 7 and 8). However, it is unclear whether the removal of estrogen is due to growth-linked metabolism, or co-metabolism catalysed by enzymes or cofactors existing for another function (Yu *et al.,* 2007). Co-metabolism could allow the removal of environmental concentrations of estrogen because it is not necessary to meet the threshold of energy and carbon required for aerobic and anaerobic biodegradation (Hazen, 2010). Co-metabolic estrogen-degrading bacteria have been identified (Pauwels *et al.,* 2008; Shi *et al.,* 2004), including autotrophs. Heterotrophic bacteria are more likely to remove EE2 than ammonia oxidising bacteria (AOB) via co-metabolism (Gaulke *et al*., 2008). Estrogen metabolism by heterotrophs under

aerobic conditions has been identified, but some denitrifying bacteria *Denitratisoma oestradiolicum* gen. nov., sp. nov and *Steroidobacter denitrificans* gen. nov., sp. nov are able to degrade E2 under anoxic conditions (Table 3; Fahrbach *et al.,* 2006; Fahrbach et al., 2008). However, under anaerobic conditions estrogen degradation is absent or has a lower efficiency (Ying *et al.,* 2003; Czajka and Londry, 2006; Sarmah and Northcott, 2008; de Mez *et al.,* 2008; Joss *et al.,* 2004; Jürgens *et al.,* 2002).

In addition to the activated sludge (AS) isolates (Table 1- 3), aquatic bacteria are able to remove E1, E2 and incompletely remove EE2 (Matsuoka *et al.,* 2005; Jürgens *et al*., 2002). However, estrogen-degrading bacteria have not been isolated from river water, and in river sediment only *Escherichia coli* has been found to degrade estrogen conjugates, although poorly (Duong *et al.,* 2011).

Bacteria may preferentially degrade abundant substances, such as glucose, rather than estrogens present at low environmental concentrations (Thayanukul *et al.,* 2010). Most studies (Table 1- 3; Matsuoka *et al.,* 2005; Duong *et al*., 2011) have used higher concentrations of estrogen than those found in the environment (Table 1- 1).

Table 1-4: List of known estrogen degrading bacteria isolated from AS.

Bacterial species	Degradation of estrogen	Starting concentration of estrogen (mg/L)	Estrogen removal (%)	References
***Novosphingobium tardaugens* sp. ARI-1**	Metabolises E1, E2 and E3.	0.3 of E1, E2 and E3	40% E1 in 20 d; 99% E3 in 10 d; 60% E2 in 14 d.	Fujii *et al*., 2002; Fujii *et al*., 2003.
***Aminobacter aminovorans* sp. KC7**	Degrades* E1 and E2.	3 of E2	Transformed E2 to E1 in 7 d; 23-31% E1 removed in 7 d.	Yu *et al.,* 2007.
***Aminobacter* sp. KC6**	Degrades* E1 and E2.	3 of E2	Transformed E2 to E1 in 7 d; 16-26% E1 degraded in 7 d.	Yu *et al*., 2007.
***Sphingomonas* strain KC8**	Metabolises E1 and E2.	3 of E2	Complete mineralisation of E1 and E2 in 7 d.	Roh and Chu, 2010; Yu *et al*., 2007; Yu *et al.,* 2013.
***Brevundimonas diminuta* strain NK 2 (EU35276)**	Transforms E2 to E1	3- 4 of E2	The residual concentration of E2 was 38 ± 1% after 168 h of incubation.	Muller *et al.,* 2010.
***Brevundimonas vesicularis* KC12**	Transforms E2 to E1	3 of E2	Complete transformation within 7 days.	Yu *et al*., 2007.

***Rhodococcus ruber* M2 KC4**	Transforms E2 to E1	3 of E2	Complete transformation within 1 day.	Yu *et al*., 2007.
***Microbacterium testaceum* KC5.**	Transforms E2 to E1	3 of E2	Complete transformation within 1 day.	Yu *et al*., 2007.
***Sphingomonas* sp. KC9-11**	Transforms E2 to E1	3 of E2	Complete transformation within 1 day.	Yu *et al*., 2007.
***Flavobacterium* sp. KC1**	Transforms E2 to E1	3 of E2	Complete transformation within 7 days.	Yu *et al*., 2007.
Bacteroidetes sp. KC2	Transforms E2 to E1	3 of E2	Complete transformation within 7 days.	Yu *et al*., 2007.
***Nocardioides simplex* sp. KC3**	Transforms E2 to E1	3 of E2	Complete transformation within 7 days.	Yu *et al*., 2007.
***Escherichia coli* sp. KC13**	Transforms E2 to E1	3 of E2	Complete transformation within 7 days.	Yu *et al*., 2007.
***Sphingomonas* sp. KC14**	Transforms E2 to E1	3 of E2	Complete transformation within 7 days.	Yu *et al*., 2007.
Achromobacter xylosoxidans	Degrades* E1, E2 and transforms E3.	0.1- 5.5	Complete transformation within 2 d.	Weber *et al*., 2005.
***Ralstonia* sp.**	Degrades* E1, E2 and transforms E3.	0.1 – 5.5	Complete transformation within 2 d.	Weber *et al.,* 2005.
***Pseudomonas aeruginosa* sp. TJ1**	Metabolises E2.	$5 \times 10^{-3} – 1.5 \times 10^{-2}$	Complete mineralisation of E2 in 2 h.	Zeng *et al.,* 2009.

***Bacillus subtilis* sp. E2Y4**	Degrades* E1 and E2.	10	Transformed E2 completely in 4 days; 40% E1 within 9 d.	Jiang *et al.,* 2010.
***Bacillus* sp. E2Y1**	Degrades* E1 and E2.	10	Transformed E2 completely in 6 days; 20% E1 within 9 d.	Jiang *et al*., 2010.
***Bacillus amylonquefaciens* sp. E2Y2**	Transforms E2 to E1.	10	Transformed E2 to E1 within 4 days.	Jiang *et al*., 2010.
***Bacillus* sp. E2Y3**	Transforms E2 to E1.	10	Transformed E2 to E1 within 5 days.	Jiang *et al*., 2010.
***Bacillus cereus* sp. E2Y5**	Transforms E2 to E1.	10	Transformed E2 to E1 within 6 days.	Jiang *et al*., 2010.
***Denitratisoma oestradiolicum* gen. nov., sp. nov**	Metabolises E2 under anoxic conditions.	0.3	Complete mineralisation of E2.	Fahrbach *et al.,* 2006.
***Steroidobacter denitrificans* gen. nov., sp. nov**	Metabolises E2 under anoxic conditions.	0.3	Complete mineralisation of E2.	Fahrbach *et al.,* 2008.
Nitrosomonas europaea	Co-metabolises E1, E2, E3 and EE2 under nitrifying conditions.	1	>95-100% E1, E2 and EE2 in 8 days; >62.5% E3 in 9 d.	Shi *et al.,* 2004.
***Sphingobacterium* sp. JCR5**	Metabolises E1, E2, E3 and EE2.	30	87% EE2 within 10 d.	Haiyan *et al*., 2007.

***Novosphingobium* sp. JEM-1**	Degrades* E1, E2 and EE2.	4.5 $x10^{-5}$ ± 4.8 x 10^{-6} of E1 and 2.0$x10^{-5}$ ± 3.0 $x10^{-6}$ of E2	99-100% E1 and E2 in <1 h.	Hashimoto *et al*., 2009.
***Rhodococcus zopfii* strain Y 50158**	Metabolises E1, E2, E3 and EE2.	100	81% E2 in 2 h; 91% E1 in 3 h; 96% E3 in 4 h; 70% EE2 in 7 h; Complete mineralisation in 24 h.	Yoshimoto *et al*., 2004.
***Rhodococcus equi* strains Y 50155, Y 50156, and Y 50157**	Metabolises E1, E2, E3 and EE2.	100	>80% E2 in 5 h; 99% E2 in 24 h; >80% E1 in 8 h; 99% in 24 h; 72-95% E3 in 24 h; >70% EE2 in 8 h; >80-95% in 24 h.	Yoshimoto *et al*., 2004.
***Pseudomonas citronellolis* SS-2**	Degradation of E1, E2 and EE2, but not E3	4 of EE2 2 of E2	93.6% EE2 removal in 168 hours. 99% E2 in 36 hours.	Fang-Fang et al., 2012
***Pseudomonas citronellolis* SJTE-3**	Metabolises E1, E2, E3 and EE2.	Unpublished data	Unpublished data	Zheng *et al.,* 2016.
***Novosphingobium* sp. E2S**	Degradation of E2	50	63.29% of E2 was removed within 7 days.	Li *et al.,* 2017
***Rhodococcus* sp. DS201**	Metabolises E2 and E1	0.5 - 10	63.4-100% E2 removed in 3 days, highest removal 100% of 1 mg/L E2 in 3 days. Lowest 63.4% removal of 10mg/L E2 in 3 days.	Yu *et al.,* 2016

* Where strains were unable to grow upon estrogen as the sole energy or carbon source, but co-metabolism was not confirmed. All strains are aerobic unless otherwise stated.

Table 1-5: List of known estrogen degrading bacteria from other sources.

Bacterial species	Degradation of estrogen	Source of bacteria	References
Phyllobacterium myrsinacearum	Metabolises E1, E2, E3 and co-metabolises EE2 in the presence of E1, E2 and E3.	Compost	Pauwels et al.(2008)
***Ralstonia picketti* BP2**	Metabolises E1, E2, E3 and co-metabolises EE2 in the presence of E1, E2 and E3.	Compost	Pauwels et al.(2008)
***Pseudomonas aeruginosa* BP3**	Metabolises E1, E2, E3 and co-metabolises EE2 in the presence of E1, E2 and E3.	Compost	Pauwels et al.(2008)
***Pseudomonas* sp. BP7**	Metabolises E1, E2, E3 and co-metabolises EE2 in the presence of E1, E2 and E3.	Compost	Pauwels et al.,(2008)
***Acinetobacter* sp.**	Metabolises E1, E2, E3 and co-metabolises EE2 in the presence of E1, E2 and E3.	Compost	Pauwels et al., (2008)
***Sphingomonas* sp. CYH**	Degrades E1 (aerobic and anoxic) and transforms E2 to E1.	Artificial sandy aquifer	Ke et al., (2007)
***Agromyces* sp. LHJ3**	Degrades E3 (aerobic), E2 (anoxic) and transforms E2 to E1.	Artificial sandy aquifer	Ke et al., (2007)
***Acinetobacter* sp. LHJ1**	Transforms E2 to E1.	Artificial sandy aquifer	Ke et al., (2007)
***Sphingomonas* ED8 and 9.**	Metabolises E1 and E2.	Soil	Kurisu et al., (2010)

***Rhodococcus* ED6, 7 and 10.**	Metabolises E1 and E2.	Soil	Kurisu et al., (2010)
***Pseudomonas putida* strain SJTE-1**	Metabolises E1 and E2	Soil	Liang *et al.,* (2012)
Leptothrix discophora	Manganese Mn^{2+}-dependent degradation of EE2	Belgian coordinated collections of microorganisms.	Sabirova et al., (2008)
***Pseudomonas putida* MnB1, 6 and 29.**	Manganese Mn^{2+}-dependent degradation of EE2	Belgian coordinated collections of microorganisms.	Sabirova et al., (2008)
***Nitrosomonas europaea* ATCC 19718**	Co-metabolism and Nitration of EE2	ATCC	Gaulke et al.(2008); Skotnicka-Pitak et al. (2009)
Buttiauxella	Metabolism of E2 and Testosterone	Baltic Sea	Zhang et al.(2011a)
***Vibrio* sp. H5**	Metabolism of E2 and Testosterone	Baltic Sea	Sang et al.(2012)
***Rhodococcus equi* ATCC 13557**	Partial degradation of EE2 in the presence of a cosubstrate. EE2 61% ± 1% in 300h.	ATCC	O'Grady et al.(2009); Larcher and Yargeau (2013)
***Rhodococcus erythropolis* ATCC 4277**	Partial degradation of EE2 in the presence of a cosubstrate. EE2 46% ± 2% in 300h	ATCC	O'Grady et al.(2009); Larcher and Yargeau (2013)
***Rhodococcus zopfii* ATCC 51349**	Partial degradation of EE2 in the presence of a cosubstrate. EE2 38% ± 1% in 300h	ATCC	O'Grady et al.(2009); Larcher and Yargeau (2013)

***Bacillus subtilis* ATCC 6051**	Partial degradation of EE2 in the presence of a cosubstrate. EE2 27% ± 2% in 300h.	ATCC	Larcher and Yargeau (2013)
Pseudomonas aeruginosa PA01	Partial degradation of EE2 in the presence of a cosubstrate. EE2 34% ± 2% in 300h.	ATCC	Larcher and Yargeau (2013)
Pseudomonas putida ATCC12633	Partial degradation of EE2 in the presence of a cosubstrate. EE2 21% ± 2% in 300h	ATCC	Larcher and Yargeau (2013)
Rhodococcus rhodochrous ATCC13808	Partial degradation of EE2 in the presence of a cosubstrate. No EE2 detected after 48 h (possibly co-metabolic)	ATCC	Larcher and Yargeau (2013)
***Bacillus cereus* Socransky 67**	Conversion of E2 to unknown metabolites	Dental plaque	Ojanotko-Harri et al (1990)
***Streptococcus mutans* Ingbritt**	Conversion of E2 to E1	Dental plaque	Ojanotko-Harri et al (1990)
***Streptococcus mutans* NCTC 10449**	Conversion of E2 to E1	Dental plaque	Ojanotko-Harri et al (1990)
***Streptococcus mutans* NCTC 10904**	Conversion of E2 to E1	Dental plaque	Ojanotko-Harri et al (1990)
Rhodococcus sp. P14	Conversion of E2 to E1	Crude oil	Ye *et al.,* 2017; Song *et al.,* 2011.
***Nocardia sp.* E110**	Degradation of E1	Soil	Coombe *et al.,* 1966

1.9 The genomic potential of bacteria to degrade estrogen

The genomic potential of bacteria to degrade estrogen is not well understood, as the only whole genome sequences are of the E2 and E1 degrading bacteria *Sphingomonas* sp. KC8 (Hu *et al*., 2011), *Pseudomonas putida* strain SJTE-1 (Liang *et al.,* 2012) and *Rhodococcus* sp. P14 which transforms E2 to E1 (Zhang *et al.,* 2012; Ye *et al.,* 2017). Also the genome of an E1, E2, E3 and EE2 degrader *Pseudomonas citronellolis* SJTE-3 (Zheng *et al.,* 2016). Several genes encoding the enzymes suggested to be putatively involved in estrogen degradation, such as hydroxysteroid dehydrogenase, 3-ketosteroid-delta1-dehydrogenase, rieske dioxygenase, catechol 2,3-dioxygenase, were observed in the genomes of *P. putida* strain SJTE-1 and *Sphingomonas* sp. KC8 (Liang *et al.,* 2012; Hu *et al.,* 2011). Furthermore 17β-hydroxysteriod dehydrogenase was upregulated in *Rhodococcus* sp. P14 when exposed to E2, with increases of 5.35 and 1.68-fold at 6 and 12 hours, suggesting that 17β-hydroxysteriod dehydrogenase is responsible for the dehydrogenation of the C-17 within the D ring of E2 to produce E1 (Ye *et al.,* 2017). Additionally the involvement of the dioxygenase enzyme in the oxidation of the C-4 and C5 was proposed in the degradation of E1 by *Nocardia* sp. E110 and in the degradation of EE2 and E1 by *Sphingobacterium* sp. JCR5 (Coombe *et al.,* 1966; Haiyan *et al.,* 2007). Furthermore the enzyme 3β,17β-hydroxysteroid dehydrogenase was suggested to be responsible for the dehydrogenation of C-17 in the D ring leading to the conversion of E2 to E1, and Flavin-dependent monoxygenase is responsible for the conversion of E1 to 4-OH-E1 in *Sphingomonas* sp. KC8 (Chen *et al.,* 2017). The enzyme 4-hydroxyestrone 4,5-dioxygenase is suggested to have a role in the degradation of 4-OH-E1 (Chen *et al.,* 2017). Genes encoding Flavin-dependent monooxygenase and 4-hydroxyestrone 4,5-dioxygenase were found within two separate gene clusters within the genome of *Sphingomonas* sp. KC8, however the gene encoding 3β,17β-hydroxysteroid dehydrogenase was not found clustered with other genes (Chen *et al.,* 2017).

In mammals and vertebrates, cytochrome P450 enzymes, from the monooxygenase family, allow E2 metabolism (Scornaienchi *et al.,* 2010; Nishida *et al.,* 2013). The P450 enzymes are also present in bacteria, P450 monooxygenase is required for bisphenol A, another benzene-containing pollutant, to be degraded by *Sphingomonas bisphenolicum* strain AO1 (Sasaki *et al.,* 2008). Also non-specific

monooxygenase activity may be necessary to degrade E1 (Yu *et al.,* 2007) and ammonia monooxygenase was suggested to be involved in estrogen degradation by *Nitrosomonas europaea* (Shi *et al.,* 2004). Therefore, P450 enzymes may have a role in estrogen degradation, although P450 enzymes responsible for estrogen degradation by bacteria are unidentified.

1.10 Aims

The main focus of this work will be the degradation of E2, EE2 and their metabolite E1, due to the legislative focus being upon E2 and EE2 (Section 1.6). However, E3 will be included in a potential estrogen degradation pathway for its completeness. In a potential estrogen degradation pathway, putative enzymes with the potential for estrogen degradation will be identified. The hypothesis is that strains of bacteria known to degrade estrogens will contain genes encoding at least some of these enzymes. Analysis of their genomes and the use of molecular techniques will confirm the presence and expression of those genes during estrogen degradation. Therefore RT-qPCR experiments will be carried out using specific primer sets to target these genes and measure their expression during a degradation experiment. The expression of ribonucleic acid (RNA) during the degradation of estrogen will suggest the involvement of the gene and allow the pathway of estrogen degradation to be confirmed.

Therefore the aim of the project is to evaluate the potential estrogen degradation pathway through analysis and comparison of the *R. equi* American Type Culture Collection 13557 (ATCC) genome with the genomes of other estrogen degrading bacteria, and through measuring the gene expression and metabolite production during estrogen degradation. *R. equi* ATCC 13557 was selected as a model organism to study the degradation of estrogen because from the limited number of estrogen degrading bacteria (10) available on culture collection websites (Table 1-4), *R. equi* ATCC 13557 had the highest percentage of EE2 degradation reported, 61% ± 1% in 300h, in the presence of a cosubstrate. (Larcher and Yargeau, 2013). Additionally, a large number of other *Rhodococcus* species (14) and other strains of *R. equi* (3) have been reported as estrogen degrading bacteria (Tables 1- 4 and 1-5). Furthermore, *R. equi* previously known as *Nocardia restrictus* is a well-studied

testosterone, another steroid hormone, degrading organism (Kim *et al.,* 2007; Gibson *et al.,* 1966; Sih *et al.,* 1966). To these ends:

- A hypothetical pathway of estrogen degradation will be proposed using existing information from current literature (Chapter 2).
- The genome of *R. equi* ATCC 13557 will be assembled and searched for genes coding for the enzymes that are potentially involved in estrogen degradation (Chapter 3).
- A comparison of the genomes of potential estrogen degrading bacteria will be carried out to localise the genes coding for enzymes that are suggested to be involved in estrogen degradation (Chapter 3).
- An optimised method of estrogen sample preparation and method for the addition of estrogen to the growth media will be developed (Chapter 4).
- Degradation experiments will be carried out to determine estrogen degradation by *R. equi* ATCC13557 under the experimental conditions (Chapter 5).
- New specific sets of primers will be developed to identify and enumerate the identified genes coding for enzymes potentially involved in estrogen degradation using RT-qPCR (Chapter 6).
- Gene expression analysis of *R.equi* ATCC13557, during the degradation of estrogen, will be carried out to determine which genes are actively expressed (Chapter 6).

Chapter 2: Creating a hypothetical estrogen degradation pathway

2.1 Introduction

The physiological process of estrogen metabolism is important in understanding the fate of estrogen and its metabolites within the environment. The identification of genes can help in the use of bacteria as a method of estrogen degradation, biodegradation, by potentially enhancing the degradation through bioengineering (Van der Geize *et al.,* 2007). Furthermore though genome comparison, genes coding for enzymes involved in a metabolic pathway may be identified in the genomes of other bacteria, previously unknown to metabolise the compound.

In 1941, it was proposed that genes control chemical reactions by acting as or producing enzymes (Beadle and Tatum, 1941). Later it was postulated that there was a relationship between the metabolism of organisms and their genes (Horowitz, 1945). The Retrograde Hypothesis suggested that due to the depletion of resources, organisms which mutated to utilise different resources to produce the depleted compound, gained a selective advantage leading to the emergence of metabolic pathways (Horowitz, 1945). Another theory is the Patchwork Hypothesis, in which early cells contained genes coding for many non-specific enzymes to utilise a broad range of compounds. These genes coding for non-specific enzymes went through duplication and evolutionary divergence leading to genes coding for more specific enzymes. (Ycas, 1974; Jensen, 1976). Therefore, the analysis of metabolic pathways can give an understanding of the enzymes involved in each step, thus identification of the genes responsible for a metabolic function. The genes of organisms, such as bacteria, involved in the production of enzymes with specific metabolic functions, may share evolutionary origins, thus there may be a conserved core metabolic pathway for steroid degradation. The degradation pathway for testosterone by *Comamonas testosteroni* is well characterised due to the identification of metabolites and genes. Confirmation of the gene involvement in testosterone degradation was carried out by gene mutation, to stop the activity of specific genes, such as TesB, thought to encode dioxygenase, TesD encoding hydrolase, and TesEFG (Horinouchi *et al.,* 2001; Horinouchi *et al.,* 2003; Horinouchi *et al.,* 2005). In the absence of these genes the metabolism of testosterone was

decreased or absent, leading to the accumulation of metabolites which were identified. Some of these metabolites were compared to the degradation pathway of Androst-4-ene-3,17-dione (4-AD) degradation pathway (Horinouchi *et al.,* 2001; Horinouchi *et al.,* 2003; Horinouchi *et al.,* 2005; Sih *et al.,* 1966; Horinouchi *et al.,* 2012).

Metabolite detection during the degradation of estrogen, by bacteria, can help to elucidate the bacterial degradation pathway of estrogen. In estrogen degradation, there have been some metabolites detected, such as E1, 2-hydroxy-2,4-dienevaleric acid and 2-hydroxy-2,4-diene-1,6-dioic acid during EE2 degradation by *Sphingobacterium* sp. JCR5 (Haiyan *et al.,* 2007). The production of the metabolites E1, pent-4-enoic acid; 5-(4-(2-carboxyethyl)-7a-methyl-1-oxooctahydro-1H-inden-5-yl)pent-4-enoic acid , 2-ethyl-3-hydroxy-6-methylcyclohexane-1-carboxylic acid; 3-(7a-methyl-1,5-dioxooctahydro-1H-inden-4-yl) propanoic acid; and 5-hydroxy-4-(3-hydroxypropyl)-7a-methyloctahydro-1H-inden-1-one during the degradation of E2 by *Rhodococcus sp*. DS201 (Yu *et al.,* 2016). The metabolites 3aα-H-4α-[3'-propanoic acid]-5β-[2-ketopropyll-7aβ-methyl-1-indanone, 3aα(-H-4α-[3'-propanoic acid]-5β-[4'-but-3-enoic acid]-7aβ-methyl-l-indanone, and 2-carboxy-7aβ-methyl-7-keto-9aα-*H*-indano-[5,4f]-5aα, 10, 10aβ, ll-tetrahydroquinoline were produced in the degradation of E1 by *Nocardia* sp. E110 (Coombe *et al.,* 1966). Therefore suggesting that E1 is degraded by cleavage of the A ring followed by the B ring by the metabolic steps; 1) hydroxylation of C-4, 2) Oxidation by dioxygenase between C-4 and C-5, 3) the seco- intermediate reacts non-enzymatically with ammonia to produce pyridine carboxylic acid (Coombe *et al.,* 1966). The degradation of E1 by *Sphingomonas* sp. KC8 also suggests the A ring is hydroxylated to produce 4-OH-E1, followed by cleavage of the A ring by dioxygenase producing a similar metabolite to those identified in the degradation by *Rhodococcus* sp. DS201 and *Nocardia* sp. E110, therefore a common estrogen degradation pathway may occur in both Proteobacteria and Actinobacteria (Chen *et al.,* 2017; Coombe *et al.,* 1966; Yu *et al.,* 2016). Furthermore, the estrogens E1, E2, E3, EE2 and other steroid hormones may lie on interconnecting pathways as bacteria can convert EE2 and E2 to E1 interchangeably, and within soil E2 and E1 were both degraded to E3 interchangeably (Xuan *et al.,* 2008; Adeel *et al.,* 2017). There have been some common metabolites identified within the degradation pathways of estrogen,

testosterone and 9,10-Seco steroids, suggesting that there may be a common pathway for steroid degradation. For example, the metabolite 2-hydroxy-2,4-dienevaleric acid identified in EE2 degradation by *Sphingobacterium* sp. JCR5 is analogous to a metabolite produced in the degradation of testosterone by *Comamonas testosteroni* TA441 and the metabolite 2-hydroxy-2,4-diene-1,6-dioic acid is also similar to a metabolite in testosterone degradation with a different cleavage position (Haiyan *et al.,* 2007; Horinouchi *et al.,* 2003). Also the metabolite 5-(4-(2-carboxyethyl)-7a-methyl-1-oxooctahydro-1H-inden-5-yl)pent-4-enoic acid found in the degradation of E2 by *Rhodococcus* sp*.* DS201 is analogous to 3aα-*H*-4α-[3'-propanoic acid]-5β-[4'-but-3-enoic acid]-7aβ-methyl-1-indanone dimethyl ester found in E1 degradation by *Nocardia* sp. E110 (Yu *et al.,* 2016; Coombe *et al.,* 1966). The metabolite 3-(7a-methyl-1,5-dioxooctahydro-1Hinden-4-yl) propanoic acid in degradation of E2 is analogous to 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid produced in testosterone degradation by *Comamonas testosteroni* TA441 (Yu *et al.,* 2016; Horinouchi *et al.,* 2003). Additionally, an intermediate metabolite proposed, but not identified as a metabolite in the degradation of E2 by *Rhodococcus* sp. DS201, was proposed in the pathway of Androst-4-ene-3,17-dione degradation (Yu *et al.,* 2016; Sih *et al.,* 1966).

However there are suggestions for alternate pathways such as the degradation of EE2 to E1 by oxidisation of C17 of the D ring, followed by cleavage of the B ring of E1 (Haiyan *et al.,* 2007) compared to the degradation of EE2 beginning with electrophilic reactions including hydroxylation and conjugation of the A ring (Yi and Harper, 2007). Furthermore the degradation of E2 may follow three pathways; 1) dehydrogenation to the intermediate estratetraenol (E0), or E1 by C-17 of the D ring followed by hydroxylation of C-4 of the A ring to form 4-hydroxyestrone (4-OH-E1) followed by ring cleavage between C-4 and C-5; 2) reduction of the keto-function of the D ring at C-17 followed by hydroxylation of C-4 of the A ring to produce 4-hydroxyestradiol (4-OH-E2); or 3) hydroxylation of C-4 of the A ring followed by reduction of the keto-function of the D-ring at C-17 (Kurisu *et al.,* 2010; Nakai *et al.,* 2011). Also the metabolites keto-E1 and keto-E2 may be due to hydroxylation of the saturated rings (Kurisu *et al.,* 2010). However, commonly the first step in E2

degradation is conversion to E1 by oxidisation of the C-17 within the D ring to a ketone (Yu *et al.,* 2007; Ke *et al.,* 2007; Lee and Liu, 2002).

Although few studies describe biodegradation pathways, some transient metabolites have been reported that suggest the following possible pathways;

1. EE2 degradation is initiated by either;
   a. Dehydrogenation of C-17 within the saturated D ring to produce E1 (Haiyan *et al.,* 2007).
   b. Electrophilic reactions including conjugation and hydroxylation of the aromatic A ring (Yi and Harper, 2007).
2. E2 degradation is initiated by either;
   a. Dehydrogenation of C-17 within the saturated D ring to produce E1 or the intermediate E0 (Yu *et al.,* 2007; Ke *et al.,* 2007; Lee and Liu, 2002; Nakai *et al.,* 2011). Followed by hydroxylation of the C-4 of the aromatic A ring to produce 4-OH-E1 (Kurisu *et al.,* 2010).
   b. Hydroxylation of C-4 of the aromatic A ring followed by reduction of the keto-function at C-17 of the saturated D ring to produce 4-OH-E2 (Kurisu *et al.,* 2010).
   C. Hydroxylation of the saturated rings B, C or D to produce keto-E1 and keto-E2 (Kurisu *et al.,* 2010).
   d. Hydroxylation of C-16 within the saturated D ring to produce E3 (Xuan *et al.,* 2008).
3. E1 degradation is initiated by either;
   a. Cleavage of the saturated B ring (Haiyan *et al.,* 2007).
   b. Oxidisation of the saturated D ring to produce a lactone (Lee and Liu, 2002).
   c. Hydroxylation of the aromatic A ring to produce 4-OH-E1 followed by cleavage of the saturated B ring (Kurisu *et al.,* 2010; Coombe *et al.,* 1966).
   d. Hydration of the saturated D ring at positions C-16 and C-17 to produce E3 (Xuan *et al.,* 2008).

However, a common metabolic pathway for estrogen degradation is unknown.

2.2 Materials and methods

A literature search was conducted using PubMed to find papers using the search terms “estrogen degradation,” “bacterial estrogen degradation,” “steroid degradation” and other similar search terms. The literature was used to look for a potential estrogen degradation pathway using information on estrogen metabolites from degradation and gene expression studies. Further literature was searched for steroid degradation pathways which were better characterised such as the degradation pathway of *Comamonas testosteroni.* The information was compiled and used to produce a hypothetical estrogen degradation pathway by looking at the reported and suggested metabolites and combining the pathways with analogous metabolites.

The software used to produce the potential pathway and all other figures containing chemical structures was ChemDrawBio Ultra version 14.0.0.117 (1998-2014).

2.3 Results and Discussion

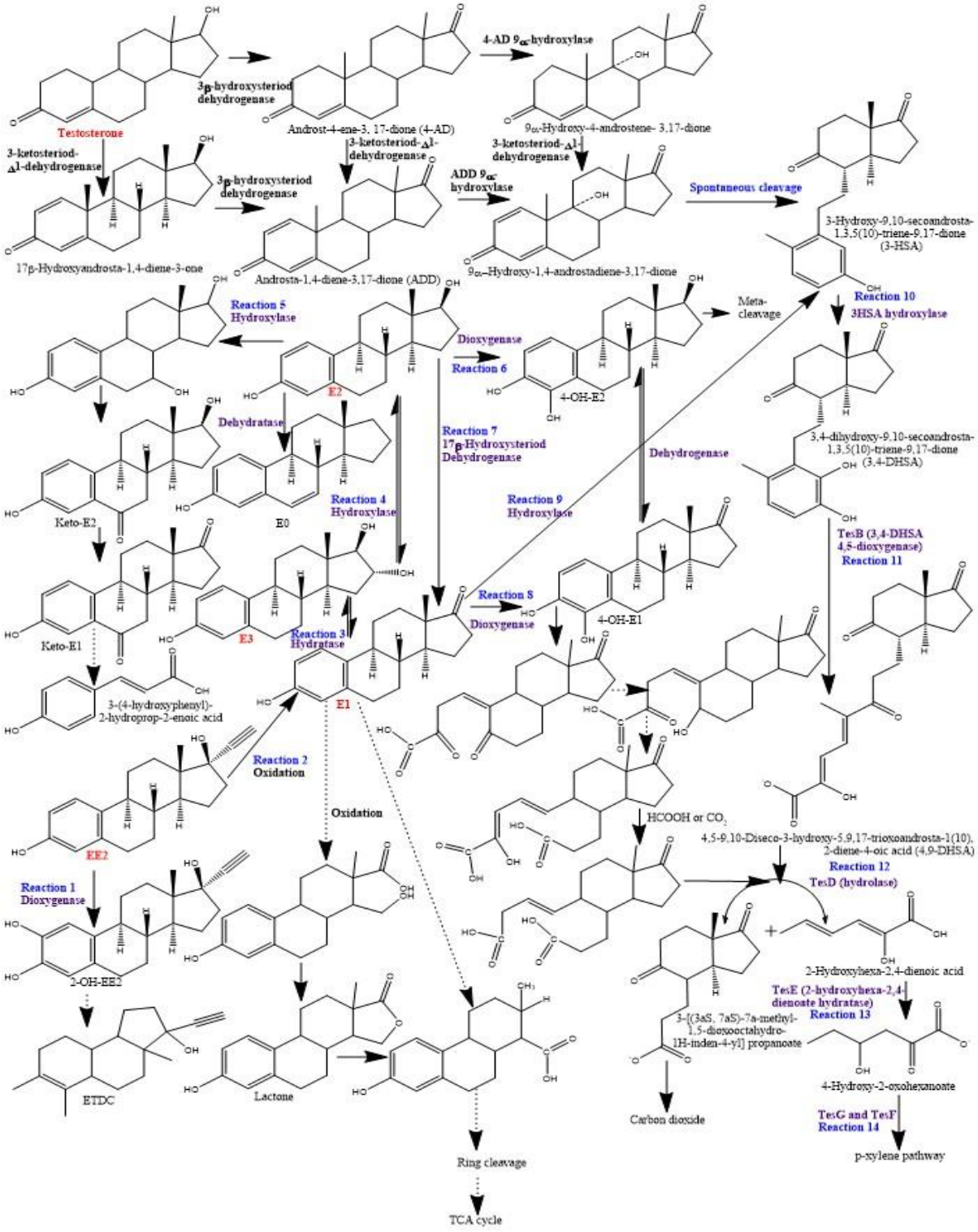

Figure 2-1: The hypothetical estrogen degradation pathway compiled from the information about estrogen degradation (Yu et al., 2007; Lee and Liu, 2002; Ke et al., 2007; Kurisu et al., 2010; Haiyan et al., 2007; Yi and Harper, 2007; Yu et al., 2013; Yu et al., 2016; Adeel et al., 2017; Xuan et al., 2008) and the testosterone degradation pathway by Comamonas testosteroni (Horinouchi et al., 2012). The enzymes or genes coding for enzymes involved in the degradation steps are highlighted in purple, the reactions in blue and the parent compounds E1, E2, E3, EE2 and testosterone in red. The dotted arrows represent unconfirmed pathways. ChemDrawBio Ultra version 14.0.0.117

Combining the existing estrogen degradation pathway information along with the better characterised testosterone degradation pathway, we postulated a potential estrogen degradation pathway (Fig. 2- 1). There are two possible degradation pathways for EE2 (Fig. 2- 2); i) hydroxylation of the aromatic A ring (reaction 1), also a common step in the degradation of polycyclic aromatic hydrocarbons (PAH) (Peng *et al.,* 2008), suggesting there may be involvement of dioxygenase in the degradation of EE2 (Yi and Harper, 2007). ii) Oxidation of the D ring to produce a ketone, converting it into E1 (reaction 2), followed by B ring cleavage and then A ring hydroxylation and ketonisation, followed by further A ring hydroxylation and meta-cleavage (Haiyan *et al.,* 2007).

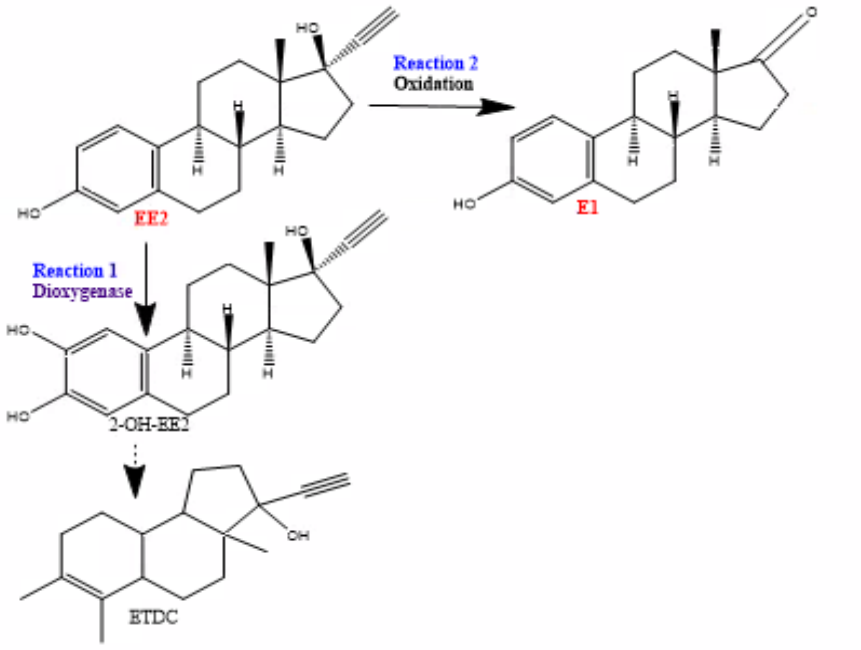

Figure 2- 2: The two possible pathways for EE2 degradation (Yi and Harper, 2007; Haiyan et al., 2007; Yu et al., 2013)

There are five possible pathways in E2 degradation (Fig. 2- 3); i) A ring hydroxylation suggesting the involvement of a dioxygenase enzyme to produce the 4-OH-E2 metabolite (reaction 6). Followed by either spontaneous cleavage and complete metabolism or dehydrogenation of the D ring by a dehydrogenase enzyme to form interchangeably 4-OH-E1. The 4-OH-E1 metabolite then undergoes further cleavage steps resulting in complete metabolism. ii) Ring hydroxylation of the saturated ring B, C or D (reaction 5), producing hydroxy-E2, keto-E2, keto-E1 and 3-(4-hydroxyphenyl)-2-hydroxyprop-2-enoic acid (Kurisu *et al.,* 2010). iii) Dehydration of

D-ring to produce the metabolite E0 (Nakai *et al.,* 2011). iv) Formation of E1 from E2 by dehydrogenation of the D ring (reaction 7) (Yu *et al.,* 2007). v) Hydroxylation of the D ring to produce E3 (reaction 4) (Xuan *et al.,* 2008).

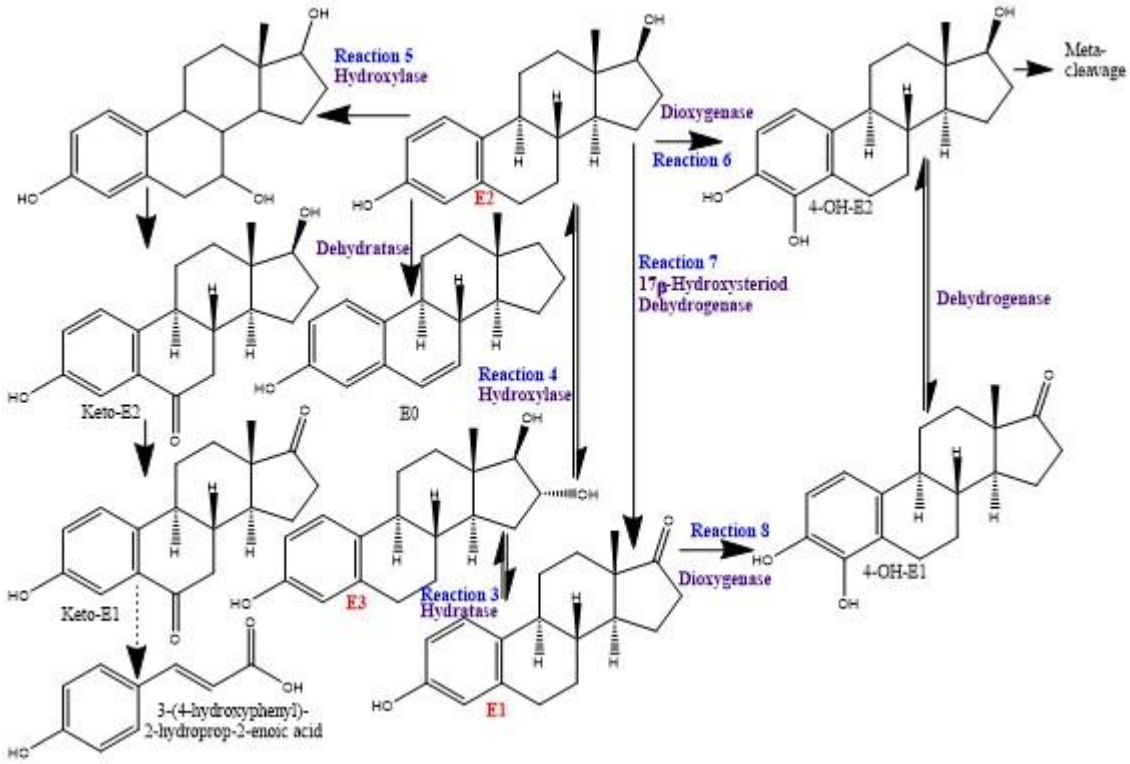

Figure 2-2: The five possible pathways of E2 degradation (Yu et al., 2007; Xuan et al., 2008; Kurisu et al., 2010; Nakai et al., 2010; Yu et al., 2013; Adeel et al., 2017)

There are four possible pathways for E1 degradation (Fig. 2- 4); i) Dehydrogenation of the D ring to form the X1 metabolite, with a lactone at ring D, which is then degraded by the tricarboxylic acid (TCA) cycle (Lee and Liu, 2002). ii) Hydroxylation at C-4 of the A ring to produce 4-OH-E1 (reaction 8) which then undergoes cleavage between C-4 and C-5 of the A ring leading to oxidation, the C-5 is then oxidised to a carbonyl group releasing carbon dioxide, then formic acid or carbon dioxide is released, then C-9 is hydroxylated to form a keto group and the B ring is cleaved (Kurisu *et al*., 2010; Yu *et al.,* 2016). At this point the pathway is combined with that of testosterone degradation by *Comamonas testosteroni* due to the analogous metabolite 4,9-DSHA (Horinouchi *et al.,* 2003; Yu *et al.,* 2016). iii) Hydration of the D ring lead to the production of E3 (reaction 3) (Xuan *et al.,* 2008). iv) E1 undergoes B ring cleavage by hydroxylation to produce 3-HSA (reaction 9) and at this point the pathway was combined with the information about the well characterised

degradation pathway of testosterone by *Comamonas testosteroni*, another steroid hormone due to the analogous metabolite found in the degradation of EE2 by *Sphingobacterium* sp. JCR5 (Horinouchi *et al*., 2005; Haiyan *et al.,* 2007). In the hypothetical estrogen degradation pathway it is suggested that the degradation of the E1 would begin at a later point in the degradation pathway than testosterone due to the saturated A ring being more similar to the later point where the B ring is cleaved in testosterone. Following this A ring hydroxylation is carried out by C4 hydroxylase, and metacleavage of the A ring by a dioxygenase enzyme (Coombe *et al.,* 1966). Followed by the action of hydrolase and then hydratase for complete metabolism (Horinouchi *et al*., 2005).

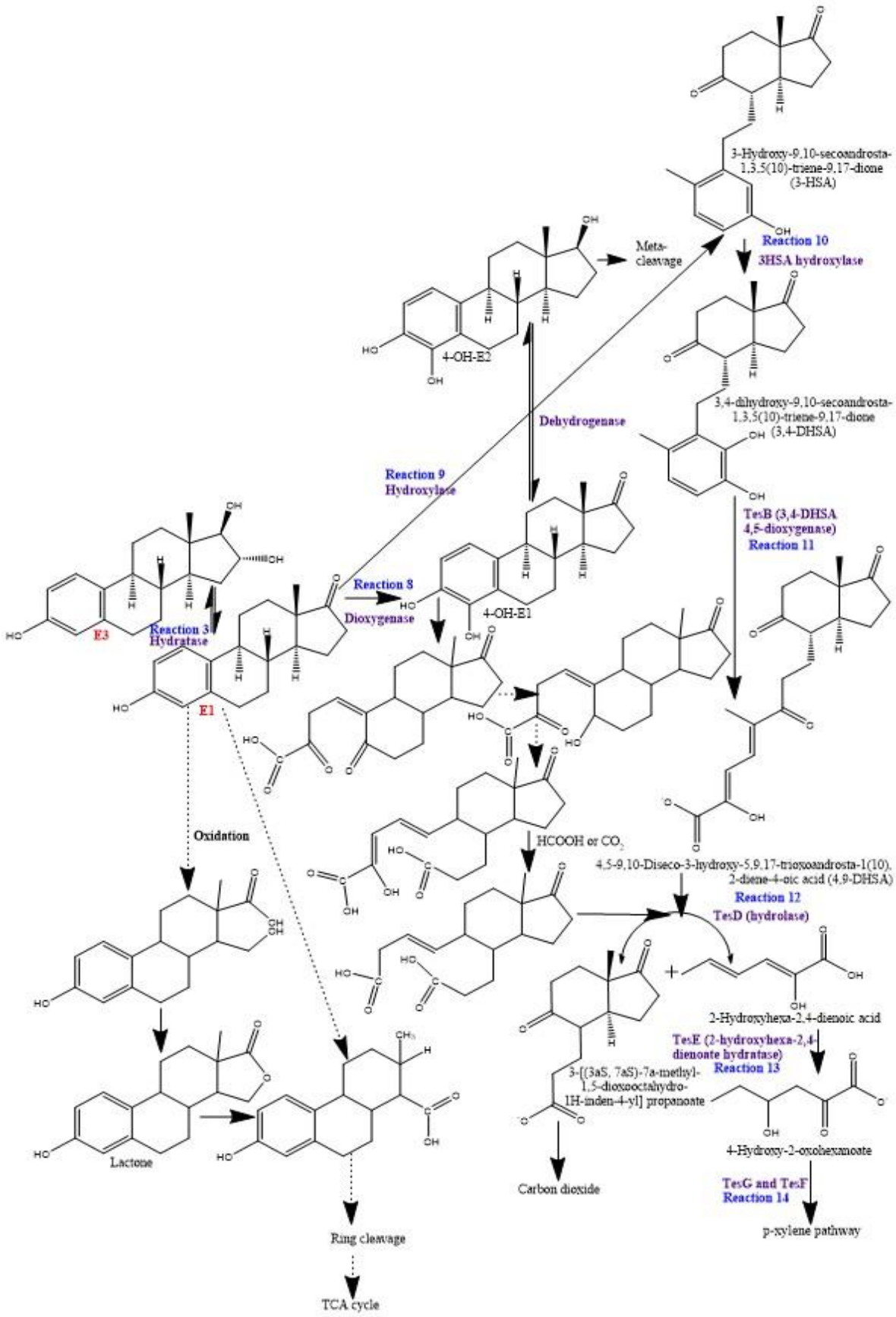

Figure 2-3: The four possible pathways of E1 degradation (Lee and Liu, 2002; Kurisu et al., 2010; Yu et al., 2016; Horinouchi et al., 2003; Xuan et al., 2008; Horinouchi et al., 2005; Haiyan et al., 2007; Coombe et al., 1966; Yu et al., 2013)

Therefore, using the potential pathway of estrogen degradation (Fig. 2- 1), the major enzymes suggested to be involved in the degradation of estrogen are; dioxygenase, hydroxylase, dehydrogenase, hydrolase and hydratase (Table 2- 1). Genes encoding some or all of these enzymes are likely to be present in the genomes of estrogen degrading bacteria, thus the genome of *R. equi* ATCC13557 will be evaluated for the presence of these genes (Table 2- 1). The proposed hypothetical estrogen pathway provides a more comprehensive understanding of the potential metabolites and genes coding for the enzymes responsible for each step. There are still multiple pathways suggested which may depend on the bacteria and its growth conditions. However given that degradation of E1 by *Nocardia* sp. E110 shares commonalities with the degradation of cholesterol by *Nocardia restrictus* (ATCC 14887) and additionally the degradation of E1 by *Rhodococcus* sp. DS201 and *Sphingomonas* sp. KC8 suggests that the most common pathway for E1 degradation begins with the hydroxylation of the A ring by a dioxygenase enzyme to produce 4-OH-E1 (reaction 10) (Coombe *et al.,* 1966; Sih *et al.,* 1965; Chen *et al.,* 2017; Yu *et al.,* 2016). This is further supported by the EAWAG Biocatalysis/Biodegradation database pathway prediction for E1 suggests that the hydroxylation of the A ring (rule bt004) is a likely first step (Hou *et al.,* 2011). The conversion of EE2 to E1 proposed, is unlikely due to the ethinyl group, which is resistant to microbial degradation (García *et al.*, 2011). Additionally, bacteria able to degrade EE2 to E1 by the *seco* pathway, would not be able to further degrade E1 (García *et al.*, 2011; Coombe *et al.,* 1966; Wang and Sih, 1963). Therefore, using the hypothetical estrogen degradation pathway (Fig. 2- 1), metabolites detected during estrogen degradation may help to determine the pathway of estrogen degradation by bacteria. Additionally, searching for the genes coding for enzymes (Table 2- 1), identified in steps within the hypothetical estrogen pathway (Fig. 2- 1), within bacterial genomes may also help to identify estrogen degrading bacteria and identify which pathway steps are involved in biodegradation of estrogen.

Table 2-1: The major enzymes potentially involved in estrogen degradation

Reaction number	Metabolic step	Suggested enzyme	EC number	References
1	EE2 to 2-OH-EE2	Dioxygenase	1.13.11.-	Yi and Harper, 2007
2	EE2 to E1	*Oxygenase	1.13.- or 1.14.-	Haiyan *et al.,* 2007
3	E1 to E3	Hydratase	4.2.1.-	Xuan *et al.,* 2008; Adeel *et al.,* 2017
4	E2 to E3	Hydroxylase	1.14.16.-	Xuan *et al.,* 2008; Adeel *et al.,* 2017
5	E2 to keto E2	Hydroxylase	1.14.16.-	Kurisu *et al.,* 2010
6	E2 to 4-OH-E2	Dioxygenase	1.13.11.-	Kurisu *et al.,* 2010
7	E2 to E1	17β-hydroxysteriod dehydrogenase	1.1.1.51	Yu *et al.,* 2007; Ke *et al.,* 2007; Lee and Liu, 2002
8	E1 to 4-hydroxy-E1	Dioxygenase	1.13.11.-	Chen *et al.,* 2017; Coombe *et al.,* 1966; Yu *et al.,* 2016
9	E1 to 3-HSA	Hydroxylase	1.14.16.-	Haiyan *et al.,* 2007
10	3-HSA to 3,4-DHSA	Hydroxylase	1.14.16.-	Horinouchi *et al.,* 2004
11	3,4-DHSA to 4,9-DHSA	Dioxygenase	1.13.11.-	Horinouchi *et al.,* 2001
12	4,9-DHSA to 2-hydroxyhexa-2,4-dienoic acid and 3-[(3aS, 7aS)-7a-methyl-1,5-dioxooctahydro-1H-inden-4-yl] propanoate	Hydrolase	3.7.1.-	Horinouchi *et al.,* 2003
13	2-hydroxyhexa-2,4-dienoic acid to 4-hydroxy-2-oxohexanoate	Hydratase	4.2.1.-	Horinouchi *et al.,* 2005
14	4-hydroxy-2-oxohexanoate to p-xylene pathway	BphI and BphJ (Aldolase and dehydrogenase complex)	4.1.3.43	Horinouchi *et al.,* 2005; Carere *et al.,* 2011

* Suggested oxygenation of the ethinyl group, however a specific enzyme is unknown.

Chapter 3: Genome assembly and evaluation

3.1 Introduction

Whole genome sequencing (WGS) is an important tool in understanding the genomic potential of bacteria through the identification of genes coding for enzymes with specific functions and through comparative genomics where genomes of bacteria with similar functions can be compared to identify common genes. Although the cost, speed, average read length and preparation times need to be considered when choosing a sequencing platform (Table 3- 1).

The method of sequencing deoxyribonucleic acid (DNA), termed Sanger sequencing, was first developed in 1977 by use of the incorporation of 2',3'-dideoxythymidine triphosphate (ddTTP) within DNA polymerase I, leading to chain-termination at the positions where thymidylic acid (dT) is replaced. Therefore, when fractionated by electrophoresis the band patterns can be observed (Sanger *et al.,* 1977). Although widely used, Sanger sequencing, which produced long sequence reads (~1 kb), has since been replaced with higher-throughput next-generation sequencing (NGS) methods. The first of which was 454 sequencing (Ekblom and Wolf, 2014; Rothberg and Leamon, 2008). 454 sequencing miniaturised the pyrosequencing reaction which generated short read length and emitted light upon the incorporation of nucleotides. This meant that smaller volumes could be used and allowed sequencing to be carried out in parallel (Rothberg and Leamon, 2008).

The use of 454 sequencing was replaced with the use of Illumina HiSeq (Solexa) and SOLiD, which produce shorter read lengths and a lower cost-per-base (Ekblom and Wolf, 2014; Shendure and Ji, 2008). Illumina uses sequencing by synthesis (SBS) bridge-PCR in which forward and reverse primers are tethered to a solid substrate by a linker, leading to the clustering of amplicons (Shendure and Ji, 2008). SOLiD uses emulsion PCR with use of DNA ligase to ligate octomers, removing the fluorescent label and a two-base sequencing method (Shendure and Ji, 2008; Liu *et al.,* 2012).

Furthermore in 2010 and 2011 the compact benchtop sequencers, the Ion Personal genome machine (PGM) from Ion Torrent, and MiSeq from Illumina were released respectively (Liu *et al.,* 2012). The Ion PGM uses semiconductor sequencing

technology, releasing a proton when polymerase incorporates a nucleotide, thus does not require fluorescence (Liu *et al.,* 2012). Therefore the Ion PGM is faster, cheaper and smaller in size compared to other NGS, with a higher sequence quality and higher mapping rate than the Illumina HiSeq (Liu *et al.,* 2012). The Miseq uses SBS but produces higher integrity data and has a broader range of applications including amplicon sequencing, clone checking, ChIP-Seq and small genome sequencing (Liu *et al.,* 2012). Furthermore the read length is longer and therefore contig assembly is better than the HiSeq (Liu *et al.,* 2012). Another benchtop sequencer is the Polonator G.007 by Complete Genomics which is a ligation based system but is recently introduced and not yet widely used (Liu *et al.,* 2012; Zhang *et al.,* 2011b).

Third-generation sequencers have been newly released and are not yet widely used. These include Helicos HeliScope using Single-molecule real-time technology from PacBio in which an enzyme incorporates nucleotides into the complementary strand and cleaves the fluorescent dye which is captured by a camera in real-time (Liu *et al.,* 2012; Zhang *et al.,* 2011b). Its advantages are faster sample preparation, reduced error and bias from PCR based methods, fast run time, longer average read length (Liu *et al.,* 2012; Zhang *et al.,* 2011b). However, the throughput is lower compared to NGS (Liu *et al.,* 2012; Zhang *et al.,* 2011b).

Another third-generation sequencing technology is Nanopore in which a single-strand of DNA is threaded into a biopore, which self-assembles to produce a heptameric transmembrane channel to which a continuous ionic flow is applied (Liu *et al.,* 2012; Zhang *et al.,* 2011b). The current disruption is detected using electrophysiological technique based on the size of the deoxyribonucleoside monophosphates (dNMP) (Liu *et al.,* 2012; Zhang *et al.,* 2011b). The Nanopore produces long read lengths quickly and is less sensitive to temperature than enzyme dependent technologies (Liu *et al.,* 2012; Zhang *et al.,* 2011b).

Table 3-1: Comparison of genome sequencing platforms (adapted from Shendure and Ji, 2008).

Platform	Read length (bp)	per-base accuracy (%)	Cost per Mb ($)	Cost per instrument ($)	Time taken (h)	SBS	Features	References
Sanger	400-1000	99.999	500	95000	0.33-3	Polymerase	Chain termination	Shendure and Ji, 2008; Foundation for Genomics, and Population Health, 2011; Liu *et al.,* 2012.
454	250	99.9	60	500000	10	Polymerase Pyrosequencing	Emulsion PCR	Margulies *et al.,* 2005; Dressman *et al.,* 2003; Liu *et al.,* 2012).
SOLiD	35	99.99	2	591000	168	Ligase	Emulsion PCR	Shendure *et al.,* 2005; McKernan *et al.,* 2006; Liu *et al.,* 2012
HiSeq	36	98	2	430000	192	Polymerase	Bridge PCR	Bentley, 2006; Fedurco *et al.,* 2006; Liu *et al.,* 2012
MiSeq	36-300	99.8	0.5	99000	5.25- 29.25 (incl. prep.)	Polymerase	Bridge PCR	Liu *et al.,* 2012; Loman *et al.,* 2012; Illumina, 2011a
Ion PGM	200	99	0.63	80490	8 (incl. prep.)	Polymerase	Semiconductor	Liu *et al.,* 2012; Loman *et al.,* 2012
Polonator G.007	13	99.999	1	155000	80	Ligase	Emulsion PCR	Shendure *et al.,* 2005; Dressman *et al.,* 2003; Li, 2015.
Helicos HeliScope	30	99.999	1	1350000	4-6	Polymerase	Single molecule	Harris *et al.,* 2008; Braslavsky *et al.,* 2003; Shendure and Ji, 2008; Liu *et al.,* 2012
Nanopore	>5000	N/A	<1.5	1000	48	Electrophysiological	Biopore	Liu *et al.,* 2012; Timp *et al.,* 2010; Goodwin *et al.,* 2016

N/A: Information not available at present.

Following genome sequencing the raw sequencing data requires assembly into contigs. Genome assembly can be carried out using numerous different assembly platforms, which produce genome assemblies of different sizes and qualities, for example a guanine-cytosine (GC) rich bacterial genome sequenced using Ilumina Miseq SPAdes preformed better compared to other assemblers (Scott and Ely, 2015). Due to the differences in assembly methods, the performance of an assembler can depend on factors such as; GC content, genome size, genome sequencing method, and genome target (Ekbolm and Wolf, 2014). Therefore, it is important to consider the advantages and disadvantages of different assemblers with respect to the requirements of the specific genome assembly project and the available resources (Ekbolm and Wolf, 2014). It is difficult to compare between different genome assemblers as the validation of each assembler often uses a different genome, however Genome Assembly Gold-Standard Evaluation for Bacteria (GAGE-B) evaluated eight different bacterial genomes using different assemblers to ensure reproducibility (Magoc *et al.,* 2013). In the comparison of some genome assemblers; SPAdes, CLC, MaSuRCA and ALLPATHS-LG (Table 3- 2), it was considered that in this project the genome of *R. equi* ATCC13557, likely a GC-rich bacterial genome (Letek *et al.,* 2010), was sequenced using IIlumina MiSeq (Table 3- 2). Genome assembly by SPAdes uses a multisized de Bruijn graph in which bubbles and chimeric reads are removed, the software estimates the distance between k-mers, constructs a paired assembly graph based upon Paired de Bruijn graphs (PDBGs), and constructs the contigs and maps the reads back to the contigs using graph simplification (Bankevich *et al.,* 2012). MaSuRCA combines de Bruijn graph with overlap-based assembly methods (Zimin *et al.,* 2013). CLC and ALLPATHS-LG use the de Bruijn graph approach, but ALLPATHS-LG requires both a mate-paired (MP) and paired-ends (PE) libraries and functions to reduce the number of errors such as mismatches within the assembly (Butler *et al.,* 2008; Gnerre *et al.,* 2011; CLC Genomics Workbench 9.5.3). SPAdes and CLC were chosen to assemble the genome of *R. equi* ATCC13557 due to ease of use and availability of CLC and the quality of assembly for PE Illumina data (Table 3- 2; Utturkar *et al.,* 2014).

Both SPAdes and CLC used a de Bruijn graph based assembly method, in which sub-sequences of a certain word size found in the reads are tabulated (CLC

Genomics Workbench 9.5.3). The table also contains neighboring forwards and backwards words, where the starting word shares its first bases with the backwards and its last bases with the forward (CLC Genomics Workbench 9.5.3). A graph is then produced where the nodes represent the words and neighbouring nodes are connected (CLC Genomics Workbench 9.5.3). Branching of the nodes leads to separate nodes which overlap, if there is a sequencing error or single nucleotide polymorphism (SNP) present a bubble is formed, but bubbles are corrected by the assembler choosing the path of highest coverage (CLC Genomics Workbench 9.5.3). Most of these parameters are within a simple user-interface on CLC where the parameters may be set or they can be automatically calculated, additionally trimming, quality assessment and genome finishing can all be carried out within the assembly software. However, there is a need to use additional programmes to carry out additional assembly steps when using SPAdes, for example before assembly the quality of the raw sequencing data needs to be assessed using, for example FastQC (Andrews, 2010), and then trimming can be carried out using trimming software such as Trimmomatic (Bolger *et al.,* 2014). Trimmomatic is a command-line based software which removes adapter sequences and carries out quality filtering to remove poor quality reads from the sequencing data before assembly (Bolger *et al.,* 2014). Following SPAdes assembly, unlike CLC which produces the assembly statistics, a quality assessment software such as, Quality Assessment Tool for Genome Assemblies (QUAST) is required to assess the size of the assembled genome, misassemblies and the N50 score, which is the average length of the longest 50% of contigs assembled (Gurevich *et al.,* 2013). In order to carry out genome finishing in CLC the genome finishing module is required, but once purchased can be used within the CLC workbench (CLC Genomics Workbench 9.5.3). In order to carry out genome finishing for SPAdes assemblies the reads can be aligned to the assembled contigs using Mauve which uses algorithmic methods to scale the sequence lengths being aligned (Darling*,* 2004). Following mapping the contigs can be reordered using Bowtie2 software which aligns sequence reads to a reference genome (Langmead and Salzberg, 2012). Genome analysis can then be carried out using Artemis to browse the genome and identify any SNPs, insertions or deletions (Rutherford *et al.,* 2000).

Table 3-2: Comparing the advantages and disadvantages of different genome assemblers.

Assembler	Advantages	Disadvantages
CLC	- User friendly and fast assembly - Can be used to design primers and probes using graphs and algorithms (CLC Genomics Workbench 9.5.3). - Low memory 4GB RAM recommended 2GB required (CLC Genomics Workbench 9.5.3). - Tracks to display and compare data. - Can be used to find ORFs in a sequence.	- Cost - Difficult to alter algorithms (Ekblom and Wolf, 2014). - Poorer assembly statistics compared to SPAdes, MaSuRCA and ALLPATHS-LG (Utturkar *et al.,* 2014).
SPAdes	-Cost - High genome coverage and assembly accuracy compared to MaSuRCA and ALLPATHS-LG (Utturkar *et al.,* 2014). -Identified as one of the best algorithms for bacterial genome assemblies using Illumina data (Utturkar *et al.,* 2014) - The best summary statistics using Illumina PE reads (Utturkar *et al.,* 2014)	-Sometimes generates small unaligned contigs indicating that they should be discarded (Magoc *et al*., 2013). - Can handle only one library at a time. - Requires preliminary read error correction (e.g. with BayesHammer or IonHammer) or the assembly will require more time and memory (Bankevich *et al.,* 2012). - Other software, such as QUAST, are needed to generate summary statistics and for gene finding.

MaSuRCA	-Cost - For *B. cereus* produced the highest N50 and N50 corrected contigs compared to SPAdes; N50 was twice any other assembler (Magoc *et al.,* 2013). - For *M. abscessus* generated larger contigs than SPAdes (Magoc *et al.,* 2013). - Produced largest contigs for a 4.75 Mb genome measured by either N50 or corrected N50 values, for 10 of the 12 experiments (Magoc *et al.,* 2013).	-Can create good contigs that it labels 'degenerate' based on internal coverage statistics, which can cause it to omit some parts of the genome from the assembly (Magoc *et al.,* 2013). - Assembly can be genome and data dependent (Utturkar *et al.,* 2014).
ALLPATHS-LG	-Cost - Outperformed the SPAdes assemblies in terms of contig numbers and generated superior hybrid assemblies (Nagarajan and Pop, 2013).	- Requires a lot of memory, 32GB recommended (Ekblom and Wolf, 2014). - Lower contig numbers than SPAdes - Requires high sequence coverage in order to compensate for the shortness of the reads. The precise coverage required depends on the length and quality of the paired reads, but typically is of the order 100x or above (Ekblom and Wolf, 2014). - Requires a short and jumping libraray (Magoc *et al.,* 2013).

Exploring the genome may help in understanding the evolution of the estrogen degradation pathway and through genome comparison may help with the identification of unknown estrogen degrading bacteria and degradation gene clusters. For example, a study searched genome databases for gene clusters coding enzymes involved in steroid degradation, identifying 8000 taxa containing genes coding enzymes involved in steroid degradation, and predicted steroid degradation in taxa (Bergstrand *et al.,* 2016). Using genomics the cholate degradation pathway was found to be evolutionarily conserved, and that genes encoding enzymes within cholate and cholesterol degradation pathways were located on a single gene cluster, within the genome of *R. equi* (Bergstrand *et al.,* 2016). Furthermore, analysis of the *Rhodococcus* sp. RHA1 genome revealed its diverse catabolic potential, including the identification of genes coding for enzymes involved in steroid degradation (McLeod *et al.,* 2006).

As mentioned previously in the general introduction (Section 1.9), the only whole genome sequences for estrogen degrading bacteria, are of the E2 and E1 degrading bacteria *Sphingomonas* sp. KC8 (Hu *et al*., 2011) and *Pseudomonas putida* strain SJTE-1 (Liang *et al.,* 2012), *Rhodococcus* sp. P14 which transforms E2 to E1 (Zhang *et al.,* 2012; Ye *et al.,* 2017), and *Pseudomonas citronellolis* SJTE-3 an E1, E2, E3 and EE2 degrader (Zheng *et al.,* 2016). Several genes encoding the enzymes suggested to be putatively involved in estrogen degradation, such as hydroxysteroid dehydrogenase, 3-ketosteroid-delta1-dehydrogenase, rieske dioxygenase, catechol 2,3-dioxygenase, were observed in the genomes of *P. putida* strain SJTE-1 and *Sphingomonas* sp. KC8 (Liang *et al.,* 2012; Hu *et al.,* 2011). Furthermore, the gene coding for 17β-hydroxysteriod dehydrogenase was upregulated in *Rhodococcus* sp. P14 when exposed to E2, suggesting that 17β-hydroxysteriod dehydrogenase is responsible for the dehydrogenation of the C-17 within the D ring of E2 to produce E1 (Ye *et al.,* 2017).

There are a limited number of estrogen degrading bacteria (10) available on culture collection websites (Table 1- 4). *Rhodococcus equi* ATCC® 13557™ was identified in the literature as an EE2 degrading bacterium (Larcher and Yargeau, 2013; O'Grady *et al.,* 2009), and from the bacteria available by culture collection it had the highest removal of EE2 (Table 1- 4), therefore it was selected as a model organism for the biodegradation studies.

The genome of *R. equi* ATCC13557 will be assembled and evaluated for the genes identified as coding for enzymes potentially involved in estrogen degradation from the hypothetical estrogen degradation pathway (Fig. 2- 1). Furthermore the genome of *R. equi* ATCC13557 will be compared with the genomes of *Sphinomonas* sp. KC8 (Hu et al., 2011) and *Pseudomonas putida* strain SJTE-1 (Liang et al., 2012).Therefore the hypotheses are that genes coding for enzymes from the hypothetical estrogen degradation pathway (Chapter 2) will be present in the genome of *R. equi* ATCC13557 and there will be genes coding for enzymes potentially involved in estrogen degradation shared between the genomes of *R. equi* ATCC13557, *P. putida* SJTE-1 (Liang et al., 2012) and *Sphingomonas* sp. KC8 (Hu et al., 2011).

3.2 Materials and methods

3.2.1 *Genome sequencing*

The genome of *R. equi* ATCC13557 was sequenced and prepared by Dr Lucy Eland, School of Computing.

3.2.1.1 Bacterial cultivation and genomic DNA extraction

The following method was provided from Dr Lucy Eland, School of Computing:

“Qiagen DNeasy DNA extraction protocol for bacterial cultures

Adapted from Qiagen DNeasy handbook, July 2006

A culture of *R. equi* ATCC13557 was prepared by seeding 10 mL of nutrient broth with a single colony taken from a nutrient agar plate. The culture was incubated over-night at 26°C. To pellet the bacteria, 1.75 mL of the culture was added to a labelled 2 mL spin tube and centrifuged at 20, 000 x g for 3 minutes, the supernatant was then removed and discarded. To lyse the bacterial cells, 180 µL of enzymatic lysis buffer, prepared using; 20mM Tris-Cl (pH), 2mM Sodium EDTA and 1.2% Triton (x100), was added to the spin tube and vortexed for 10-20 seconds. The tube was then incubated for 30 minutes at 37°C. Following incubation, 25 µL of proteinase K and 200 µL of buffer AL were added to the tube, and vortexed to mix. The tube was

then incubated for 30 minutes at 56°C. It was allowed to cool for 2 minutes, before adding 200 µL of 100% ethanol to the tube, incubating at room temperature for 2 minutes, and then vortexing to mix. The contents of the tube (~600 µL) were transferred to a labelled 2 mL collection tube and centrifuged at 10,000 x g for 1 minute. The column was then placed into a new collection tube, and the supernatant was discarded. Then 500 µL of buffer AW1 was added to the column, incubated at room temperature for 2 minutes, and centrifuged at 10,000 x g for 1 minute. Again the column was placed into a new collection tube and the supernatant was discarded. Next, 500 µL of buffer AW2 was added to the column and incubated at room temperature for 2 minutes, then the tube was centrifuged at 20,000 x g for 3 minutes. The tube was carefully removed from the centrifuge to ensure the flow-through did not come into contact with the column. The column was transferred to a 1.5 mL tube, to which 100 µL of sterile molecular grade water (Sigma Aldrich) was added to the column and incubated at room temperature for 5 minutes. The tube was then centrifuged at 10,000 x g for 1 minute. The column was transferred to a new 1.5 mL tube and 100 µL of sterile molecular grade water (Sigma Aldrich) was added to the column and incubated at room temperature for 5 minutes. The column was centrifuged at 10,000 x g for 1 minute, then the column was discarded. Following assessment of the quantity and quality of the two elutions using nanodrop, in which the 260/280 = 1.8 and 260/230 = 2.0, the better elution was chosen and the DNA was stored at -20°C.

3.2.1.2 Library preparation and Illumina MiSeq sequencing

The following method was provided by Dr Lucy Eland, School of computing:

The genomic DNA elution was diluted using molecular-grade water (Sigma Aldrich) to a concentration of 1 ng. Sequencing was carried out using the Illumina Nextera XT library prep kit using amplicons >300 bp. During the Nextera XT library preparation the DNA was quantified using the Qubit (high sensitivity dsDNA kit) and the fragment sizes were measured using the Bioanalyser (high sensitivity DNA kit) as quality control steps. The Bioanalyser trace from the Illumina Miseq raw sequence data, which was carried out after PCR and the first clean-up stage, including the fragment and adapter sequence lengths, was provided by Dr Lucy Eland (School of

Computing) (Appendix, Fig. 0- 16) (Appendix, Tables 0- 1, 2, 3). The Nextera XT library was sequenced using an Illumina MiSeq 300 bp paired-end read run on a V3 cartridge, on the 'FASTQ only' setting to remove the adapter sequences.

3.2.2 *Genome assembly*

Various genome assembly methods were considered (Table 3- 2), however CLC and SPAdes were selected due to their low computational requirements, availability and ease of use.

3.2.2.1 CLC

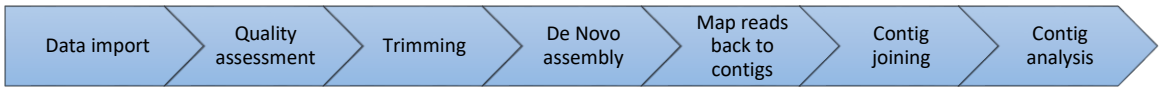

Figure 3-1: Workflow for the genome assembly of R. equi ATCC13557 using CLC.

Following the workflow (Fig. 3- 1), the genome of *R. equi* ATCC13557 was assembled using CLC. The raw sequence data was imported into CLC workbench 7.5 as Illumina fastq paired-end data with a minimum distance of 397 and maximum distance of 3,520, which were determined by looking at the bioanalyser trace information, and with the failed reads removed and the quality assessment was created using quality scores using NCBI/Sanger or Illumina 1.8 and later.

Trimming was carried out using a limit of 0.05 and a maximum number of ambiguities of 2, reads below 36 bases were discarded.

De Novo assembly was carried out with the following parameters; word size of 23, bubble size of 228, minimum contig length of 200, perform scaffolding and auto detect paired distances. For the read mapping options, "map reads back to contigs (slow)" was selected. The following parameters were the default; mismatch cost of 2, insertion cost of 3, deletion cost of 3, Length fraction of 0.5 and similarity fraction of

0.8 with a list of unmapped reads created. Word size and bubble size were automatically calculated by CLC.

Using the genome finishing module, the contigs were joined using *Rhodococcus equi* 103S (Letek *et al.,* 2010) as a reference genome. The reference genome was selected by submitting the largest contig sequence, contig 5, from the *R. equi* ATCC13557 assembled genome with a size of 598658 bp to Genbank and selecting the genome with the highest similarity, which was 99% identity and a 0.0 E value, and then the contigs were analysed.

3.2.2.2 SPAdes

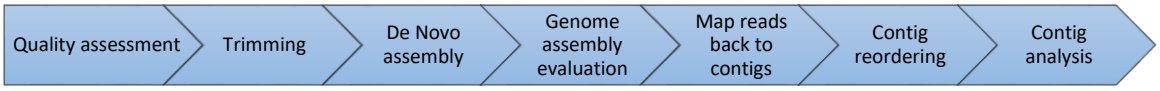

Figure 3-2: Workflow for the genome assembly of R. equi ATCC13557 using SPAdes

Following the workflow (Fig. 3- 2) the genome of *R. equi* ATCC13557 was assembled using SPAdes (Bankevich *et al.,* 2012). The quality of the raw sequence data was assessed during the CLC assembly so was not repeated, then the data was trimmed using Trimmomatic based on the average quality of 4 bases and to remove any reads below the length of 36 bases. SPAdes was then used to assemble the paired-end Illumina raw data using the default parameters except “--cov-cutoff auto” to remove contigs with lower coverage and "--careful" was used to reduce the number of mismatches and short indels within the genome. QUAST (Gurevich, *et al.,* 2013) was used to evaluate the SPAdes genome assembly. Mauve (Darling, 2004) was used to align the original reads back to the assembled genome. Bowtie 2 on the usegalaxy.org (Afgan, *et al.,* 2016; Langmead and Salzberg, 2012) was used to reorder the contigs. Artemis was used to browse the genome and look for SNPs and insertions or deletions (Rutherford, *et al.,* 2000).

3.2.3 *Annotation*

Both the CLC and SPAdes genome assemblies were submitted to RAST for annotation, and the SEED viewer was used to search for the genes identified in the hypothetical estrogen degradation pathway (Chapter 2) (Aziz *et al.,* 2008; Overbeek, *et al.,* 2014; Brettin, *et al.,* 2015).

3.2.4 *Creating a graphical map of the genome*

The CGView Comparison Tool (CCT) (Grant *et al.,* 2012) was used to create a graphical representation of the genome. As outlined in the CCT website (Grant *et al.,* 2012), the command "cgview_comparison_tool.pl –p CCT" to create a project, the assembled genome the *R. equi* 103S (Letek *et al.,* 2010) were added to the project folders for comparison along with the genes to label and a file containing information about the genes. The project settings folder was edited to produce a large map using nucleotide sequences. The map was then customised to resize the diagram and text using the command "cgview_comparison_tool.pl -p CCT --cct --custom useInnerLabels=false featureThickness=150 _cct_blast_thickness=200 titleFontSize=150 rulerFontSize=70 rulerPadding=100 legendFontSize=70 backboneRadius=2200 arrowheadLength=20 labelLineLength=80 featureOpacity=1 labelFontSize=80 maxLabelLength=500 maxLegendLength=400 gene_labels=t height=6000".

3.2.5 *Genomic analysis*

A database of genes containing enzymes potentially involved in estrogen degradation was compiled using the hypothetical estrogen degradation pathway (Fig.2- 1) (Table 2- 1). In addition to these, 3-ketosteriod Δ1-dehydrogenase, also named 3-oxosteriod 1-dehydrogenase, responsible for the conversion of 4-AD to Androsta-1,4-diene-3,17-dione (ADD) within the testosterone degradation pathway, was included in a database of potential estogen degradation genes due to upregulation when exposed to E2 (Sang *et al.,* 2011). Genes identified in the genome of *Comamonas testosteroni* (Horinouchi *et al*., 2003; Horinouchi *et al*., 2010; Horinouchi *et al.,* 2004; Horinouchi *et al.,* 2005) and the two genome sequences of the estrogen degrading bacteria *Pseudomonas putida* SJTE-1 (Liang *et al.,* 2012) and *Sphingomonas* KC8 (Hu *et al.,* 2011), along with the PAH dioxygenase encoding genes (Meynet *et al*., 2015) and dehydrogenase encoding genes (Kisiela *et al.,* 2012) have been stored in Fasta format. The compiled list was then used to search for other genes with at least 85% identity and any Basic Local Alignment Search Tool (BLAST) results were added to the list (Appendix Tables 0- 4 and 5), these included monoxygenase and isomerase due to their similarities to dioxygenase and dehydrogenase, respectively. A RefSeq ID list is provided for some of the genes coding for dehydrogenase (Appendix Table 0- 6) (Kisiela *et al.,* 2012).

Analysis of the *R. equi* ATCC13557 was carried out using BLAST in BioLinux to search the genome for the presence of the genes encoding the enymes within the compiled list of gene sequences coding for enzymes which are potentially involved estrogen degradation, this was also carried out for the only other two genomes of known estrogen degrading bacteria *Sphingomonas* sp. KC8 and *Pseudomonas putida* SJTE-1 (Hu *et al.,* 2011; Liang *et al.,* 2012). A database was created using the command “formatdb -p F Sequence_files/genome.fasta” then the database was searched using tblastn, using the command “tblastn -query Protein_files/*.fasta -db Sequence_files/genome.fasta -outfmt 7 -out Outputfiles/*.txt”

Using Rapid Annotation using Subsystem Technology (RAST) (Aziz *et al.,* 2008; Overbeek *et al.,* 2014; Brettin *et al.,* 2015), genomes with similarity to *R. equi* ATCC13557 were identified. Furthermore, the genome was compared to the genomes of *Sphingomonas* sp. KC8 and *Pseudomonas putida* SJTE-1 (Hu *et al.,*

2011; Liang *et al.,* 2012) by using The SEED Viewer sequence based comparison tool within RAST (Aziz *et al.,* 2008; Overbeek *et al.,* 2014; Brettin *et al.,* 2015).

3.3 Results and Discussion

3.3.1 *Quality assessment*

The quality distribution of the raw sequencing data produced by CLC workbench 7.5 (Fig. 3- 3) PHRED-scores of 30 and above are considered high quality, in the raw sequencing data over 60% of sequences being observed above 30, however there are some being observed below which could lead to false-positive variant calls (Illumina. 2011b), thus smaller sequence lengths were discarded in the later assembly.

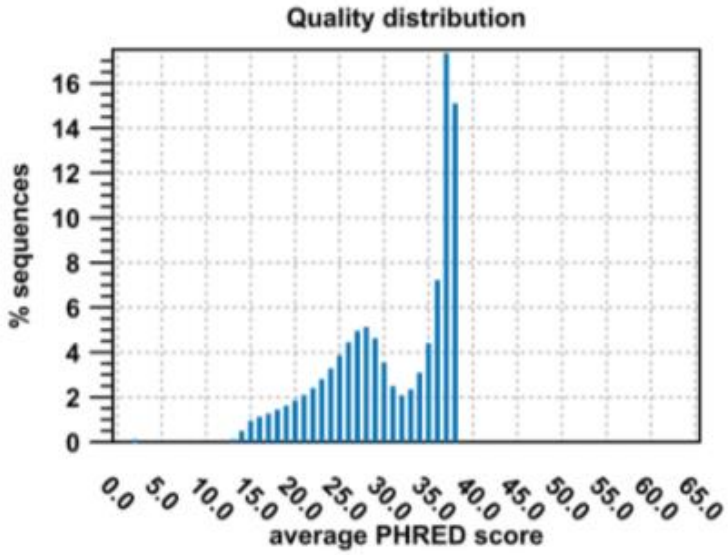

Figure 3-3: The quality of raw sequence data assessed by CLC workbench 7.5. The distribution of average sequence quality scores. The quality of a sequence is calculated as the average of its base qualities. x: PHRED-score and y: number of sequences observed at that quality score normalised to the total number of sequences.

3.3.2 *CLC assembly*

The CLC assembly of *Rhodococcus equi* ATCC13557 was a total size of 3.93 Mb and an N50, 50% of the total, contig length of 180.7 Kb and a total of 663 contigs assembled (Fig 3- 4; Tables 3- 3). There is a large range of contig lengths, the maximum contig length is 598.7 Kb and the minimum is 200 bases, but the average length is 8.2 Kb (Table 3- 3). In the mapping there are 159109 unmapped reads, with a length of 121.4 bases and 3,018,040 broken paired reads, with a length of 203.49 bases. When compared to the *R. equi* 103S reference genome there were a large number of BLAST hits above 98%, however the genome size of the CLC assembly of 3.93 Mb is much smaller than the 5.04 Mb genome of *R. equi* 103S (Fig. 3- 4) (Letek *et al.,* 2010).

Table 3-3: The assembly statistics for the R. equi genome assembled using CLC.

	Length	**Count**	**Total bases**
Reads	228.4	8,884,954	2,029,296,984
Matched	230.35	8,725,845	2,009,981,356
Not matched	121.4	159,109	19,315,628
Contigs	8,183	663	
Reads in pairs	456.56	5,707,774	
Broken paired reads	203.49	3,018,040	
N75	108,697		
N50	180,906		
N25	392,386		
Minimum	200		
Maximum	598,670		
Average	8,184		

According to the RAST annotation, the CLC assembled *R. equi* ATCC 13557 genome (Fig. 3- 4), contains 3748 predicted coding sequences (CDSs) and 354 subsystems. Furthermore 37% of genes were categorised within a subsystem, whilst 63% were not categorised within a subsystem. The largest four subsystems are amino acids and derivatives (347), carbohydrates (294), cofactors, vitamins, prosthetic groups and pigments (248), and fatty acids, lipids, and isoprenoids (200) (Fig. 3- 5). A number of CDSs, related to virulence, disease and defense (59), stress response (77) and secondary metabolism (77), suggesting there are diverse catabolic genes which assist in the adaptation of the bacteria to the environment. Furthermore, there are 62 aromatic pathways for the metabolism of aromatic compounds of which 13 are peripheral pathways and 47 are involved in the metabolism of central aromatic intermediates, which could assist in the degradation of steroids.

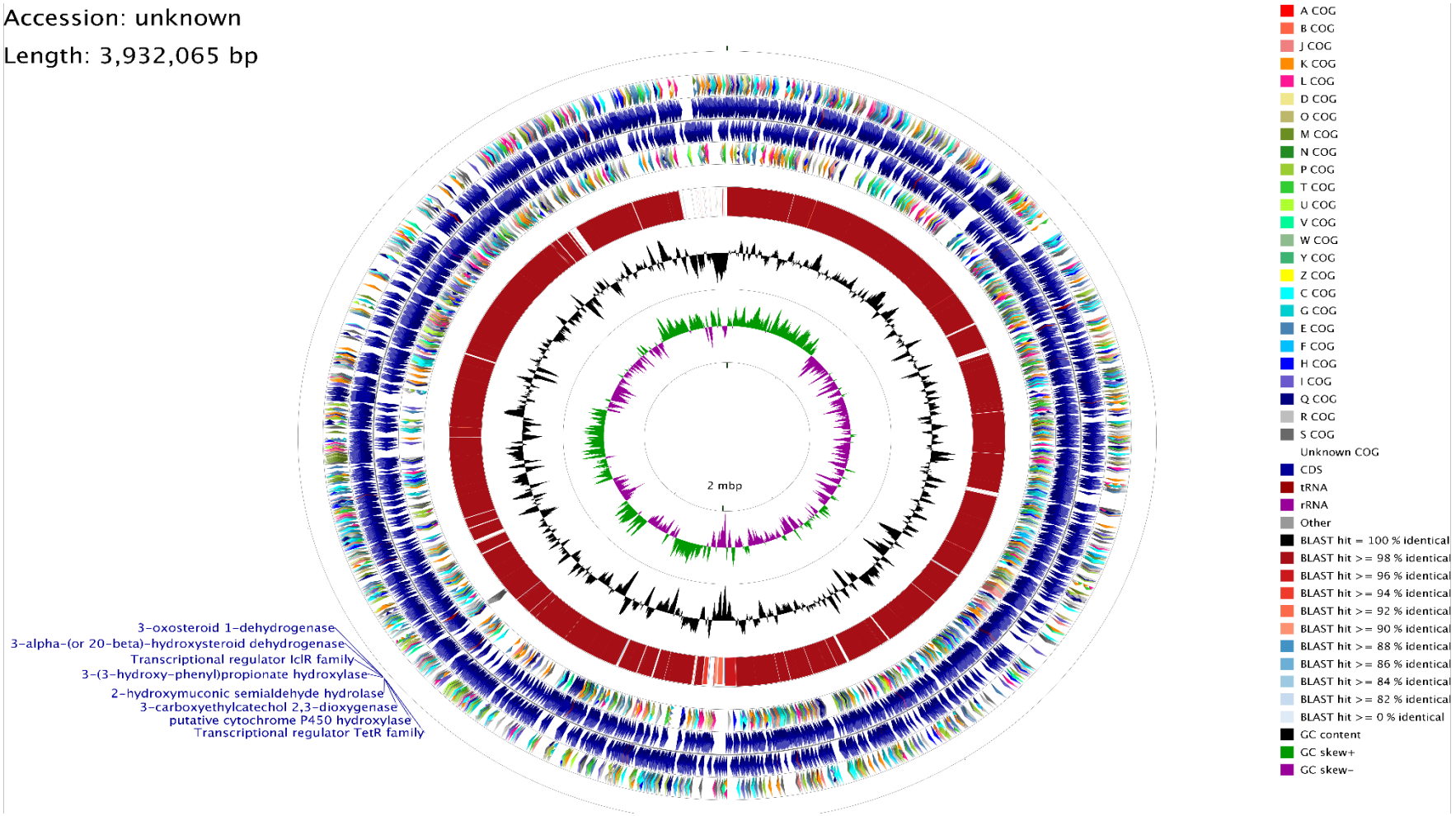

Figure 3-4: The genome map comparing the CLC assembled R. equi ATCC 13557 to the R. equi 103S reference genome. Starting from the outermost ring the feature rings depict: 1. Forward strand sequence features of R.equi ATTC 13557; 2. Reverse strand sequence features of R.equi ATCC 13557; 3. The sequence similarity detected by BLAST comparisons conducted between nucleotide sequences from R.equi ATCC 13557 and R.equi 103S; 4. The COG features of the reverse strand sequence; 5. The sequence similarity detected by BLAST comparisons conducted between nucleotide sequences from R.equi ATCC 13557 and R.equi 103S; The final rings display the GC content and the GC skew. A gene cluster of potential estrogen degradation genes is labelled on the outermost ring. Diagram produced using CCT (Grant et al., 2012).

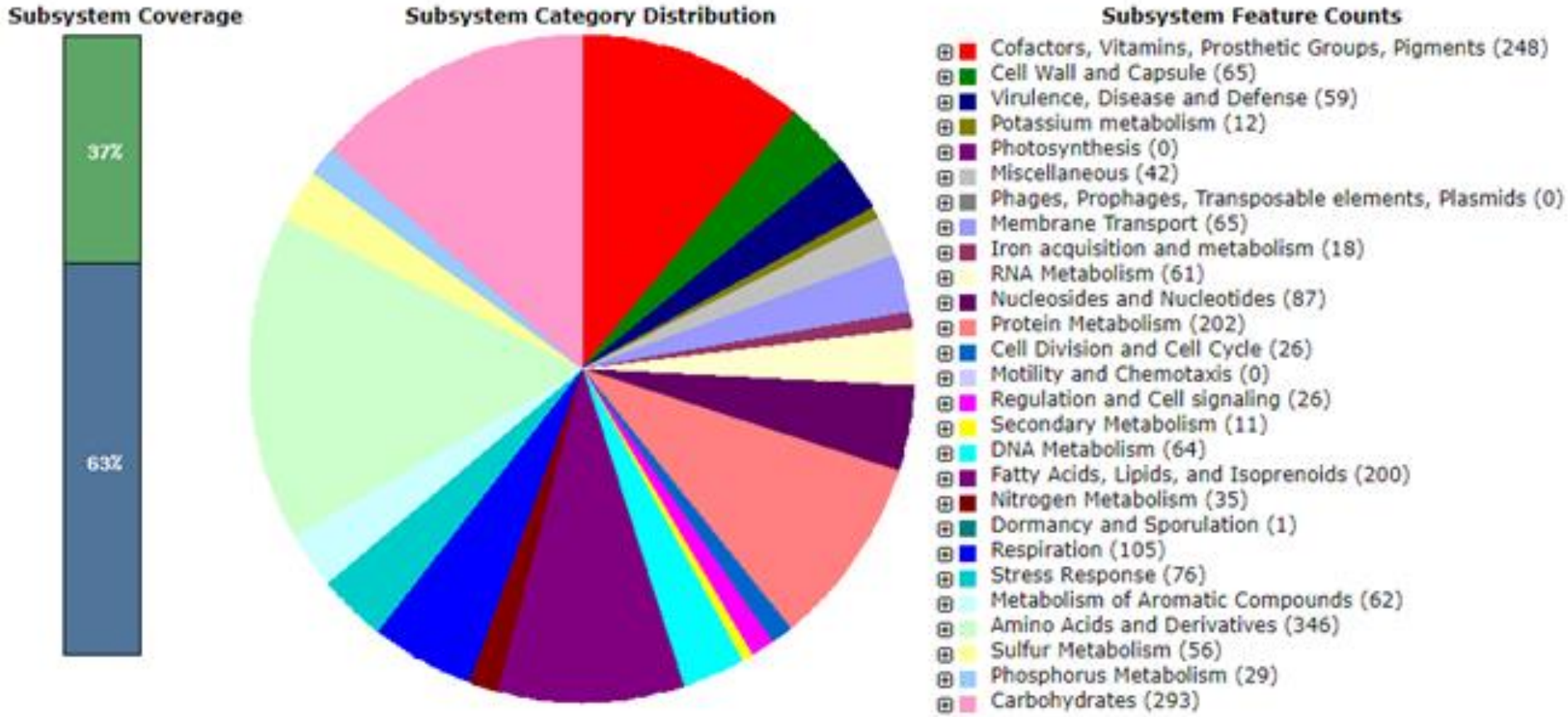

Figure 3-5: The subsystem distribution, coverage and counts within the CLC genome assembly of R. equi ATCC13557 as annotated by Rapid Annotation System Technology (RAST) server (Aziz, RK et al., 2008; Overbeek, R et al., 2014; Brettin, T et al., 2015).

3.3.3 *SPAdes assembly*

3.3.3.1 QUAST assembly statistics

The SPAdes assembled *R. equi* ATCC13557 genome had a size of 5.24 Mb with an N50 contig length of 196.4 Kb, and a total of 53 contigs (Fig. 3- 5; Table 3- 4). Furthermore the length of the longest contig is 715.1 Kb, with the 45 of the contigs being over 1000 bp in length (Table 3- 4). Therefore, the size of the SPAdes assembly, 5.24 Mb, was of a similar size to the 5.04 Mb genome of *R. equi* 103S (Letek *et al.,* 2010).

Table 3-4: Assembly statistics for the SPAdes assembled R. equi ATCC13557 genome.

Statistics	***Contigs***
# contigs	53
# contigs (>= 0 bp)	55
# contigs (>=1000 bp)	45
Largest contig	715095
Total length	5243649
Total length (>= 0 bp)	5243905
Total length (>= 1000 bp)	5238676
N50	196389
N75	119737
L50	7
L75	15
GC (%)	68.73
Mismatches	
# N's	0
# N's per 100 kbp	0
Predicted genes	
# predicted genes (unique)	4870
# predicted genes (>= 0 bp)	4873
# predicted genes (>= 300 bp)	4436
# predicted genes (>= 1500 bp)	659
# predicted genes (>= 3000 bp)	69

According to the RAST annotation the genome of *R. equi* ATCC 13557 (Fig. 3- 6), assembled using SPAdes, contains 4924 predicted coding sequences (CDSs) and 414 subsystems. Furthermore 39% of genes were categorised within a subsystem, whilst 61% were not categorised within a subsystem. The largest four subsystems are Amino Acids and Derivatives (494), Carbohydrates (415), Cofactors, Vitamins, Prosthetic Groups, Pigments (355), and Fatty Acids, Lipids, and Isoprenoids (274) (Fig. 3- 7). A large number of CDSs, similar in number to those in the genomes of *Sphingomonas* sp. KC8 (Hu *et al.,* 2011) and *Pseudomonas putida* SJTE-1 (Liang *et al*., 2010) are found to be related to virulence, disease and defense (81), stress response (129) and secondary metabolism (11) which suggest there are diverse catabolic genes which assist in the adaptation of the bacteria to the environment. Furthermore, there are 75 pathways for the metabolism of aromatic compounds, of which 14 are peripheral pathways and 59 are involved in the metabolism of central aromatic intermediates, which could assist in the degradation of steroids.

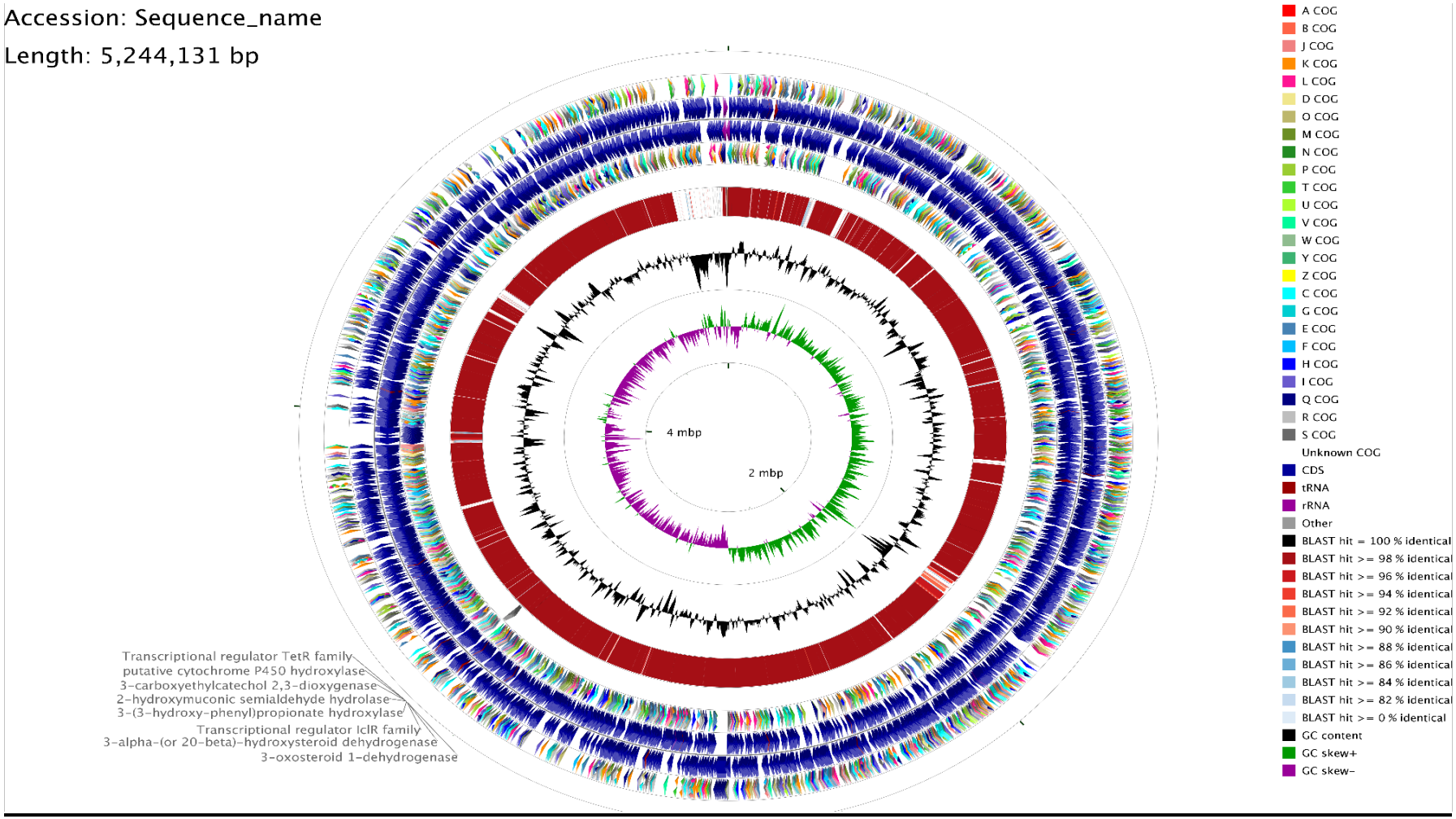

Figure 3-6: The genome map comparing the SPAdes assembled R. equi ATCC 13557 to the R. equi 103S reference genome. Starting from the outermost ring the feature rings depict: 1. COG features of the forward strand sequence; 2. Forward strand sequence features of R.equi ATTC 13557; 3. Reverse strand sequence features of R.equi ATCC 13557; 4. The COG features of the reverse strand sequence; 5. The sequence similarity detected by BLAST comparisons conducted between nucleotide sequences from R.equi ATCC 13557 and R.equi 103S; The final rings display the GC content and the GC skew. A gene cluster encoding enzymes potentially involved in estrogen degradation is labelled on the outermost ring. Diagram produced using CCT (Grant et al., 2012)

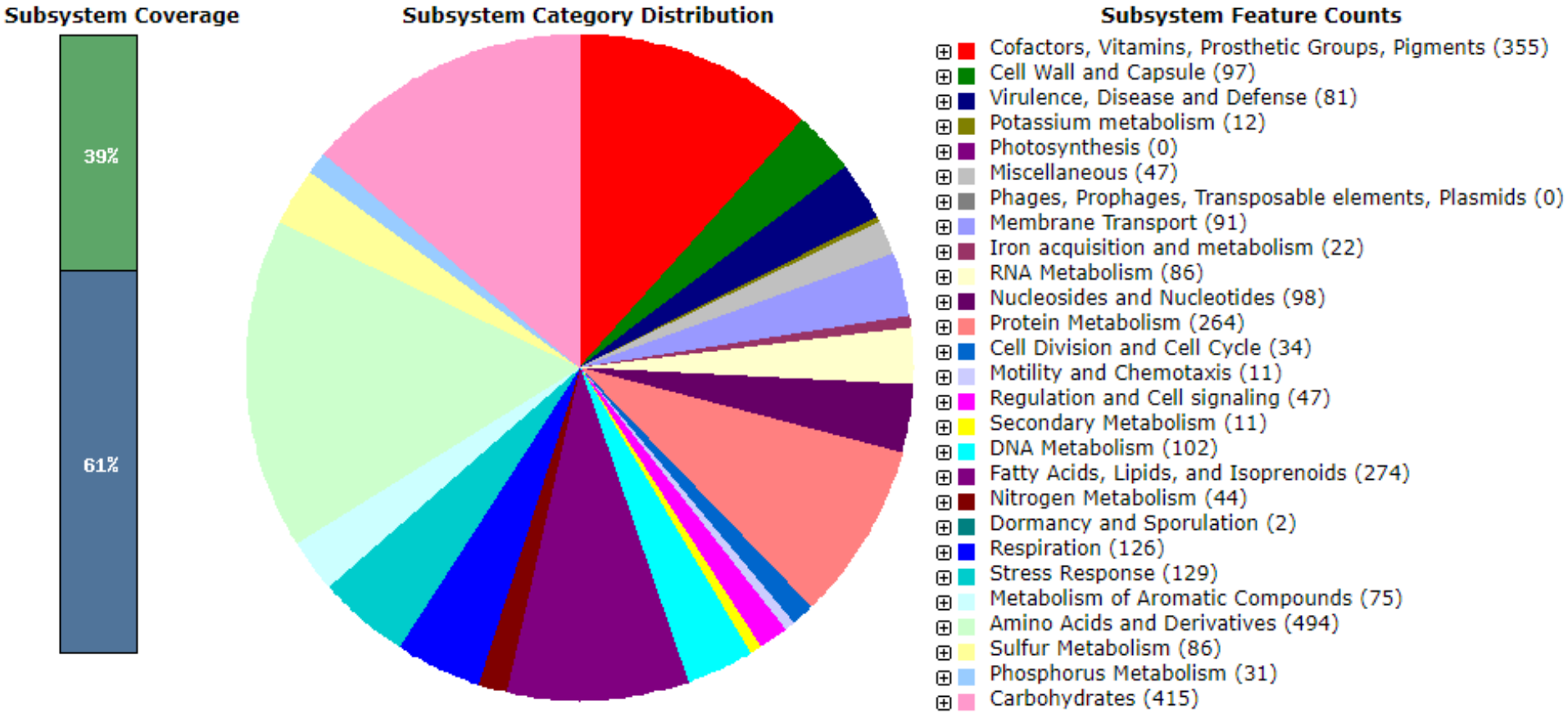

Figure 3-7: The subsystem distribution, coverage and counts within the SPAdes assembly of R. equi ATCC13557 as annotated by Rapid Annotation System Technology (RAST) server (Aziz, RK et al., 2008; Overbeek, R et al., 2014; Brettin, T et al., 2015).

3.3.4 *Comparison of the SPAdes and CLC assemblies of R. equi ATCC13557*

The genome assembly using SPAdes had a larger genome size of 5.24Mb compared to the CLC assembly of 3.93Mb. Furthermore, the N50 score of the SPAdes assembly was higher at 196.4Kb than the CLC assembly of 180.7Kb. Also there were 53 contigs assembled by the SPAdes compared to 663 assembled by CLC. Meaning the SPAdes assembly produced a smaller number of higher quality contigs, due to their longer average length, compared to the contigs assembled using CLC. Furthermore, in the RAST annotation fewer CDSs were found in the CLC assembly (3748), compared to the SPAdes assembly (4924), the numbers of subsystems were lower and the percentage of genes classified in a subsystem was slightly lower for the CLC assembly (37%) compared to SPAdes assembly (39%). The lower number of CDSs and subsystems within the CLC genome assembly suggests that the assembly was not as complete compared to the SPAdes assembly. The SPAdes assembly, is according to the assembly statistics, a better assembly, so the *R.equi* ATCC13557 genome assembled using SPAdes was used in any further analysis and comparisons.

Table 3-5: Comparison of the assembly statistics of R. equi ATCC13557 assembled using CLC and SPAdes

Assembly statistic	Genome assembler	
	CLC	SPAdes
Genome size (Mb)	3.93	5.24
N50 (Kb)	196.4	180.7
Number of Contigs	663	53

3.3.5 *Genome comparison*

According to the functional comparison in the RAST server the closest neighbors of *R.equi* ATCC13557 are *Rhodococcus jostii* RHA1 (score 504), followed by *Rhodococcus opacus* B4 (score 494), another genome of *Rhodococcus jostii* RHA1 (score 489) and *Rhodococcus equi* 103S (score 449) (Table 3 –6). The neighbor score assigned by RAST is the number of top BLAST hits between genes within the genome of *R.equi* ATCC13557 and the neighbor genomes and therefore indicates the functional similarities. All of the 30 closest neighbors to *R.equi* ATCC13557; *Rhodococcus, Nocardia, Mycobacterium* and *Gordonia,* were all from the phylogenetically-related sub-order *Corynebacterineae*. The closest neighbor, *Rhodococcus jostii* RHA1, is a soil bacterium with a diverse catabolic potential, which can degrade aromatic compounds such as biphenyl, chloate and cholesterol (McLeod *et al.,* 2006; Bergstrand *et al.,* 2016). *Rhodococcus erythropolis* (score 437), although a different strain, is suggested to have the ability to partially degrade EE2 in the presence of a cosubstrate (O'Grady et al., 2009; Larcher and Yargeau, 2013).

Table 3-6: The closest neighbours of R. equi ATCC13557 according to RAST

Genome ID	Score	Genome name
101510.15	504	*Rhodococcus jostii* RHA1
632772.3	494	*Rhodococcus opacus* B4
101510.16	489	*Rhodococcus jostii* RHA1
685727.3	449	*Rhodococcus equi* 103S
234621.6	437	*Rhodococcus erythropolis* PR4
596309.3	429	*Rhodococcus erythropolis* SK121
247156.1	369	*Nocardia farcinica* IFM 10152
247156.8	340	*Nocardia farcinica* IFM 10152
745408.3	228	*Rhodococcus* sp. JVH1
186196.5	218	*Rhodococcus* sp. DK17
525370.3	199	*Rhodococcus equi* ATCC 33707
1206729.4	180	*Nocardia exalbida* NBRC 100660
246196.19	154	*Mycobacterium smegmatis* str. MC2 155
1206721.4	147	*Nocardia asiatica* NBRC 100129
1206735.3	143	*Nocardia takedensis* NBRC 100417
1165867.3	137	*Rhodococcus imtechensis* RKJ300
543736.3	133	*Rhodococcus opacus* PD630
1078016.3	130	*Rhodococcus erythropolis* XP
1206737.4	129	*Nocardia pneumoniae* NBRC 100136
1206730.4	120	*Nocardia higoensis* NBRC 100133
1077976.3	119	*Gordonia polyisoprenivorans* NBRC 16320
1089454.3	116	*Gordonia terrae* NBRC 100016
1206722.4	115	*Nocardia araoensis* NBRC 100135
246196.1	115	*Mycobacterium smegmatis* str. MC2 155
1214102.3	112	*Mycobacterium fortuitum* subsp. fortuitum DSM 46621
350058.5	108	*Mycobacterium vanbaalenii* PYR-1
278137.3	106	*Mycobacterium* sp. Spyr1
164756.15	105	*Mycobacterium* sp. MCS
350058.8	105	*Mycobacterium vanbaalenii* PYR-1

Score: The number of top blast hits for genes shared between *R.equi* ATCC13557 and the neighbor genome.

3.3.6 *Genome evaluation for genes coding for enzymes potentially involved in estrogen degradation*

RAST annotation suggested there may be a gene cluster, containing some of genes coding for enzymes potentially involved in estrogen degradation, which was present in both the CLC and SPAdes genome assemblies (Fig. 3- 8). The presence of the genes within a cluster suggests that they are evolutionarily conserved, like the cholate and cholesterol degradation pathways which are located on a single gene cluster, within the genome of *R. equi* (Bergstrand *et al.,* 2016). The gene cluster contains genes coding for the enzymes identified in the hypothetical estrogen degradation pathway (Fig. 2-1; Table 2- 1); 3-carboxyethylcatechol 2,3-dioxygenase (EC 1.13.11.16) which catalyses an oxidation reaction by the addition of two oxygen molecules, which is likely to be responsible for reactions 1, 6, 8, and 11 (Fig 2- 1; Table 2-1); 2-hydroxymuconic semialdehyde hydrolase (EC 3.7.1.9) catalyses the hydrolysis of carbon-carbon bonds in ketonic substances, responsible for reaction 12 (Fig 2- 1; Table 2-1); 3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase is involved in steroid degradation by acting on the CH-OH groups with NAD+ or NADP+ as an acceptor, responsible for reaction 7 (Fig 2- 1; Table 2-1); 3-(3-hydroxy-phenyl)propionate hydroxylase (EC 1.14.13.127) is involved in the degradation of aromatic compounds by incorporation or reduction of oxygen, responsible for reactions 4, 5, 9 and 10 (Fig 2- 1; Table 2-1); 3-oxosteroid 1-dehydrogenase (EC 1.3.99.4), named 3-ketosteriod Δ1-dehydrogenase within the testosterone degradation pathway (Horinouchi *et al.,* 2012) (Fig. 2- 1), by acting upon CH-CH groups (Kanehisa *et al.,* 2000; Kanehisa *et al.,* 2016; Kanehisa *et al.,* 2017); and transcriptional regulator IcIR family regulates the genes involved in the degradation of aromatic compounds (Molina-Henares *et al.,* 2006). Within the hypothetical estrogen degradation pathway the genes encoding the enzymes were previously stated as the class which contain many enzymes which carry out a similar function; dioxygenase, hydroxylase, 17β-hydroxysteriod dehydrogenase, and hydrolase (Fig 2- 1; Table 2-1). As this study is focused upon the model organism *R.equi* ATCC13557 the enzymes will be henceforth referred to more specifically as their annotated name given by RAST of 3-carboxyethylcatechol 2,3-dioxygenase, 3-(3-hydroxy-phenyl)propionate hydroxylase, 3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase, and 2-hydroxymuconic semialdehyde hydrolase respectively.

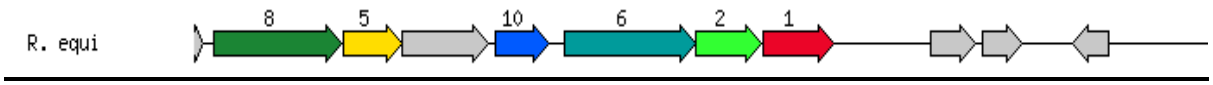

Figure 3-8: A RAST diagram of a chromosomal region around the focus gene coding for 3-carboxyethylcatechol 2,3-dioxygenase (red, 1). The other genes present code for 2-hydroxymuconic semialdehyde hydrolase (light green, 2), 3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase (yellow, 5), 3-(3-hydroxy-phenyl)propionate hydroxylase (turquoise, 6), 3-oxosteroid 1-dehydrogenase (dark green, 8) and transcriptional regulator IcIR family (blue, 10). The grey arrows are genes with the relative position conserved found in at least four other species.

Protein BLAST (BLASTP) comparisons of the *R. equi* ATCC 13557, *Sphingomonas* KC8 and *P. putida* SJTE-1 genomes with a database of gene sequences coding for enzymes potentially involved in estrogen degradation were carried out in Bio-Linux 8.0 software (Field et al., 2006). The BLASTP hits above 70% identity are summarised and the genes with significant similarity, having a low E-value and large coverage are highlighted yellow (Tables 3- 7, 8, 9).

For *Rhodococcus equi* there are 11 sequences with identity >70% to 3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase), which within the degradation pathway of testosterone catalyses the conversion of 4-AD to ADD (Fig. 2- 1), within other species of *Rhodococcus* and also in *Mycobacterium ulcerans* Agy99, all of which are from the phylogenetically-related mycolic-acid containing sub-order *Corynebacterineae* (Table 3- 7). Although other genes showed similarity to genes coding for enzymes involved in steroid degradation, such as hydratase (TesD), which catalyses the conversion of 4,9-DHSA to 2-hydroxyhexa-2,4-dienoic acid and 3-[(3aS, 7aS)-7a-methyl-1,5-dioxooctahydro-1H-inden-4-yl] propanoate in reaction 12 (Fig. 2- 1 and Table 2- 1), the sequence alignment length is very small (14 amino acids) suggesting only a small length of the gene had similarity (Horinouchi *et al.,* 2003). However, this may be important when considering the design of primers to target genes coding for enzymes involved in steroid degradation.

In *Sphingomonas* KC8, the genes with significant identity, greater than 70% and 0.0 e^{-10}, to those in the database of genes coding for enzymes potentially involved in estrogen degradation were different to those found in *Rhodococcus equi;* 3-ketosteriod 5-dehydrogenase (71.14%) was identified in another *Sphingomonas* species, *Sphingomas wittichii* RW1, and a dioxygenase (84.81%) present in *Pseudomonas* sp. SKA58 (Table 3- 8). All of these genes with significant identity

were from *Proteobacteria*, which is a group phylogenetically distinct and unrelated to the sub-order *Corynebacterineae,* to which *Rhodococcus equi* belongs. However, the gene coding for 3-ketosteriod 1-dehydrogenase of *R. equi* ATCC13557 shared 71.14% identity with *Sphingomonas* sp. KC8 but had an E value of 5.9 over a small sequence length of 17 amino acids.

In *Pseudomonas putida* SJTE-1, the genes with significant identity, greater than 70% and 0.0 e^{-10}, to those in the database of genes encoding enzymes potentially involved in estrogen degradation were two genes coding for dioxygenase (100% and 96.91%) and a genes coding for catechol dioxygenase (98.08%) present in other species of *Pseudomonas* (Table 3- 9).

Table 3-7: BLASTP comparisons of the R. equi ATCC 13557 genome with the database of genes coding for enzymes potentially involved in estrogen degradation.

Enzyme	BLASTP hit					
	Bacteria	Accession/ Protein Seq ID	Percentage identity (%)	Alignment length	Mismatches	E value (e^{-10})
3-ketosteriod 5-dehydrogenase	*Comamonas testosteroni* CNB-2	YP_003277400.1	71.43	14	4	1.6
3-ketosteriod 5-dehydrogenase	*Comamonas testosteroni* TA441	AB076368	71.43	14	4	1.7
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Rhodococcus erythropolis* PR4	NC_012490	87.39	555	66	0
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Rhodococcus jostii* RHA1	NC_00826	80.04	551	110	0
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Rhodococcus jostii* RHA1	YP_704476	78.30	553	115	0
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Mycobacterium smegmatis* str. MC2 155 chromosome	NC_008596.1	70.05	551	163	0
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Rhodococcus jostii* RHA1	YP_705733	80.04	551	110	0
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Rhodococcus erythropolis* PR4	YP_002764184	87.21	555	67	0

Enzyme	BLASTP hit					
	Bacteria	Accession/ Protein Seq ID	Percentage identity (%)	Alignment length	Mismatches	E value (e^{-10})
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Rhodococcus opacus* B4	YP_002783051	80.59	546	106	0
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Rhodococcus erythropolis* SK121	ZP_04388196	87.21	555	67	0
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Rhodococcus erythropolis* PR4	NC_012490.1	87.39	555	66	0
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Rhodococcus jostii* RHA1	NC_008268.1	80.04	551	110	0
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Mycobacterium ulcerans* Agy99	NC_008611.1	67.03	558	178	0
3-betahydroxysteriod Delta-5 dehydrogenase	*Mycobacterium abscessus*	NC_01039	80.30	264	52	3.00E-132
Dioxygenase	*Rhodococcus* sp. R04	JF502261	80.00	300	59	4.00E-147
Dioxygenase	*Rhodococcus* sp. YK2	AB070456	80.00	15	2	4.8
Catechol dioxygenase	*Burkholderia xenovorans* LB400	NC_007953	88.24	17	2	0.023
Hydratase (TesD)	*Comamonas testosterone*	LC_010134	78.57	14	3	6.3
Hydroxylase (3-ketosteriod 9-alpha monooxygenase)	*Rhodococcus* sp. RD6	CVQP01000005.1	70.20	349	104	1.00E-44

Table 3-8: BLASTP comparisons of the Sphingomonas sp. KC8 genome with the database of genes coding for enzymes potentially involved in estrogen degradation.

Enzyme	BLASTP hit					
	Bacteria	Accession/ Protein Seq ID	Percentage identity (%)	Alignment length	Mismatches	E value (e^{-10})
3-ketosteriod 5-dehydrogenase	*Sphingomonas wittichii* RW1	YP_001263819.1	71.14	551	159	0
3-ketosteriod 1-dehydrogenase (3-oxosteriod 1-dehydrogenase)	*Rhodococcus equi* ATCC 13557		70.59	17	5	5.9
3beta Delta5-steroid dehydrogenase	*Mycobacterium abscessus*	NC_010397.1	75.00	12	3	6.3
Dioxygenase (homogentisate 1,2-dioxygenase)	*Pseudomonas putida* SJTE-1	NZ_AKCL01000092.1	72.73	11	3	0.71
Dioxygenase (taurine dioxygenase)	*Pseudomonas putida* SJTE-1	NZ_AKCL01000007.1	70.59	17	5	1.2
Dioxygenase	*Pseudomonas* sp. SKA58	AAQG01000001	84.81	441	67	0
Catechol dioxygenase	*Bradyrhizobium elkanii* USDA 76	NZ_KB900701.1	73.33	15	4	4
Hydratase	*Comamonas testosteroni* TK102	CP006704	90.91	11	1	8.5

Table 3-9: BLASTP comparisons of the P. putida SJTE-1 genome with the database of genes coding for enzymes potentially involved in estrogen degradation.

Enzyme	BLASTP hit					
	Bacteria	Accession/Protein Seq ID	Percentage identity (%)	Alignment length	Mismatches	E value (e^{-10})
3 or 17-beta hydroxysteriod dehydrogenase	*Pseudomonas protegens* Pf-5	NC_004129.6	79.69	325	66	7.00E-170
7-alpha hydroxysteriod dehydrogenase	*Acinetobacter radioresistens* DSM 6976	NZ_KB849749.1	70.59	17	5	6.5
12alpha hydroxysteriod dehydrogenase	*Bacteriodes pectinophilus* ATCC 43243	ZP_03463615.1	73.33	15	4	3
3-ketosteriod delta 4(5alpha) – dehydrogenase (3-oxosteroid 1-dehydrogenase)	*Comamonas testosteroni* ATCC 17410	L23428.1	83.33	12	2	7.1
Dioxygenase	*Pseudomonas putida* KT2440	AE015451	100	453	0	0
Dioxygenase	*Pseudomonas monteilii* SB3101	CP006979	84.33	453	71	2.00E-115
Dioxygenase	*Pseudomonas putida* DOT-T1E	CP003734	98.02	202	4	4.00E-89
Dioxygenase	*Pseudomonas monteilii* SB3078	CP006978	84.33	453	71	2.00E-115
Dioxygenase	*Pseudomonas putida* F1	CP000712	96.91	453	14	0
Dioxygenase	*Sphingomonas* sp. LB126	EU024110	71.42	14	4	5.1
Catechol dioxygenase	*Pseudomonas putida* KT220	NC_002947.3	98.08	312	6	0
Hydroxylase (3-ketosteroid-9-alpha-monooxygenase oxygenase)	*Mycobacterium chubuense* strain DSM 44219	JYNX01000038.1	78.57	14	3	5.7

When the genomes of *R. equi* ATCC13557, *Pseudomonas putida* SJTE-1 and *Sphingomonas* KC8 were compared to each other, the genes with greater than 70% identity coded for the enzymes; malate synthase G, which catalyses the condensation of acetyl coenzyme A in the glyoxylate pathway (Roucourt *et al.,* 2009), with 70.64% identity between *R. equi* ATCC13557 and *P. putida* SJTE-1; translation elongation factor Tu, which accelerates the binding of aminoacyl-tRNA to the ribosome within the elongation phase of protein synthesis (Doerfel and Rodnina, 2013), with 71.75% identity between *Sphingomonas* KC8 and *P. putida* SJTE-1; SSU ribosomal protein S12p (S23e), involved in protein synthesis (Restreppo *et al.,* 2016), with over 70% between all three genomes; alkyl hydroperoxide reductase protein F (EC 1.6.4.-), which has a role in oxidative stress response (Rocha and Smith, 1999), shared 77.3% identity between *R. equi* ATCC13557 and *P. putida* SJTE-1; and aliphatic amidase AmiE (EC 3.5.1.4), which has a role in virulence and biofilm formation (Clamens *et al.,* 2017), shared 80.47% identity between *R. equi* ATCC13557 and *P. putida* SJTE-1 (Appendix Table 6).

There was no presence of genes coding for the following enzymes putatively involved in estrogen degradation; hydroxylase, responsible for the conversion of E2 to E3 in reaction 4, E2 to keto-E2 in reaction 5, E1 to 3-HSA in reaction 9 and 3-HSA to 3,4-DHSA in reaction 10 (Fig. 2- 1 and Table 2- 1); monoxygenase, included due to similarity to dioxyenase sequences; or cytochrome P450, included as it may have a role in the conversion of E1 to 16α-Hydroxyestrone (Thomas and Potter, 2013, Xu *et al.,* 2005; Simpson, 1994). However, the presence of genes coding for the following enzymes suggested to be involved in estrogen degradation were discovered; dioxygenase, catalyses the conversion of EE2 to 2-OH-EE2 in reaction 1, E2 to 4-OH-E2 in reaction 6, E1 to 4-OH-E1 in reaction 8 (Fig. 2- 1 and Table 2- 1); dehydrogenase, catalyses the conversion of E2 to E1 in reaction 7 (Fig. 2- 1 and Table 2- 1); hydratase, catalyses the conversion of E1 to E3 in reaction 3; hydrolase, catalyses the conversion of 4,9-DHSA to 2-hydroxyhexa-2,4-dienoic acid and 3-[(3aS, 7aS)-7a-methyl-1,5-dioxooctahydro-1H-inden-4-yl] propanoate in reaction 12 (Fig. 2- 1 and Table 2- 1); and isomerase, included due to similarity to dehydrogenase. There are five genes coding for hydratase present in all three genomes and seven coding for isomerase present in all three, with one gene coding for isomerase being present in *R. equi* ATCC13557 and *P. putida* SJTE-1. However,

some genes coding for these enzymes are present in two or all three of the genomes (Fig. 3- 9, 10, 11).

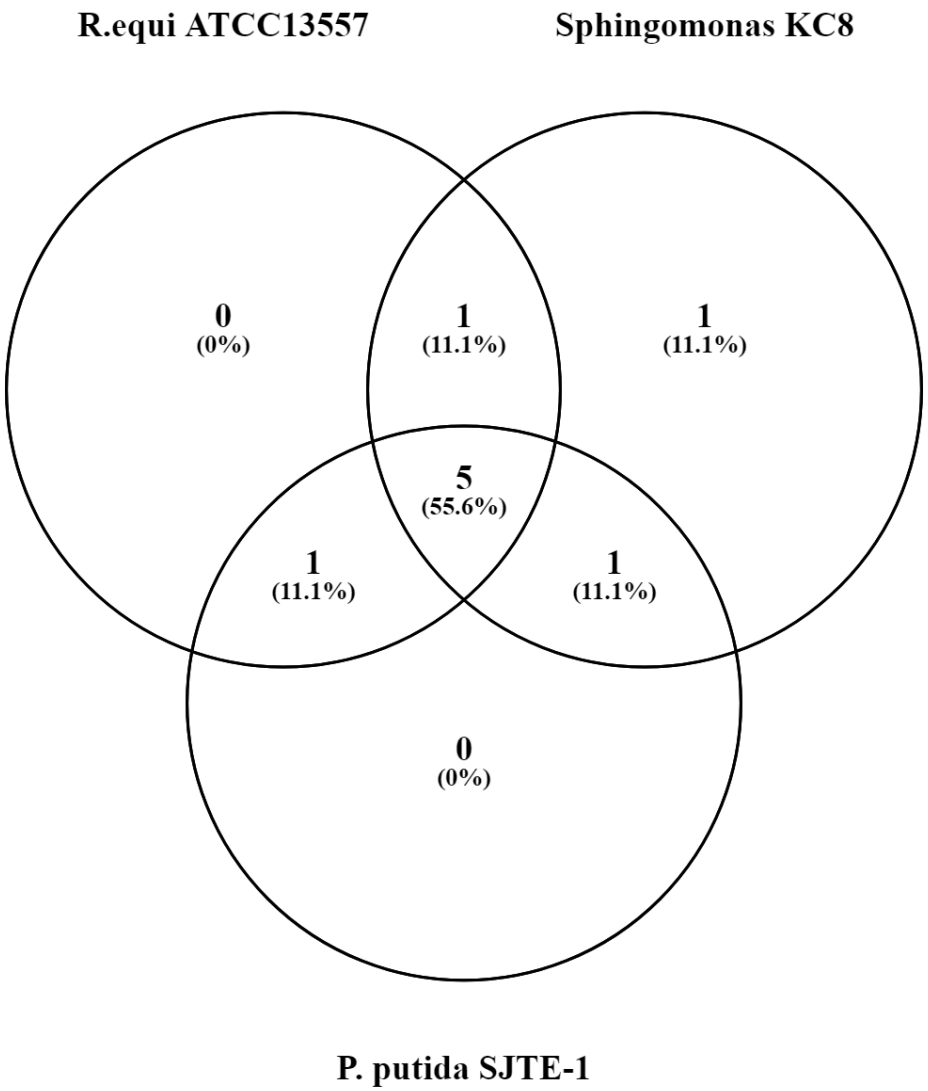

Figure 3-9: Venn diagram produced using Venny 2.1.0 (Oliveros, 2007-15) showing the presence of genes coding for dioxygenase in the genomes of R. equi ATCC13557, P. putida SJTE-1 and Sphingomonas sp. KC8.

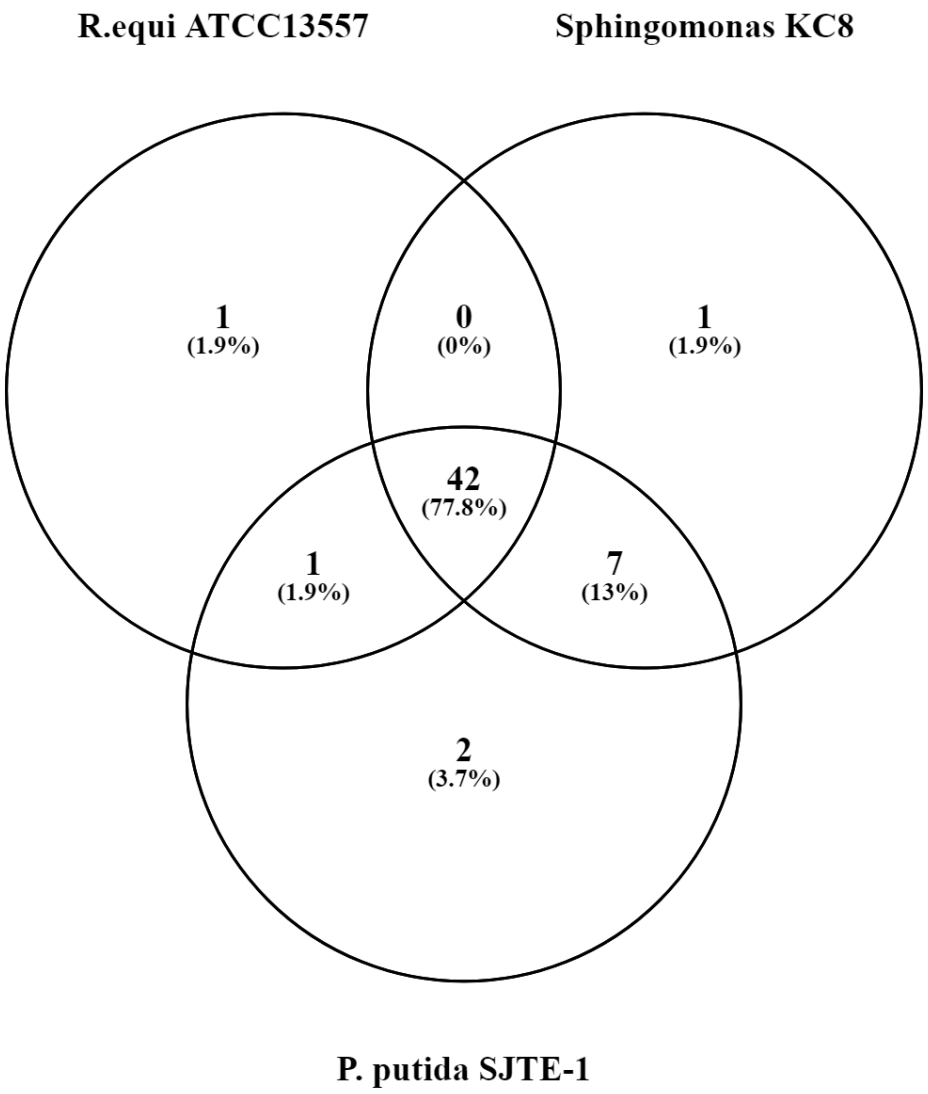

Figure 3-10: Venn diagram produced using Venny 2.1.0 (Oliveros, 2007-15) showing the presence of genes coding for dehydrogenase in the genomes of R. equi ATCC13557, P. putida SJTE-1 and Sphingomonas sp. KC8.

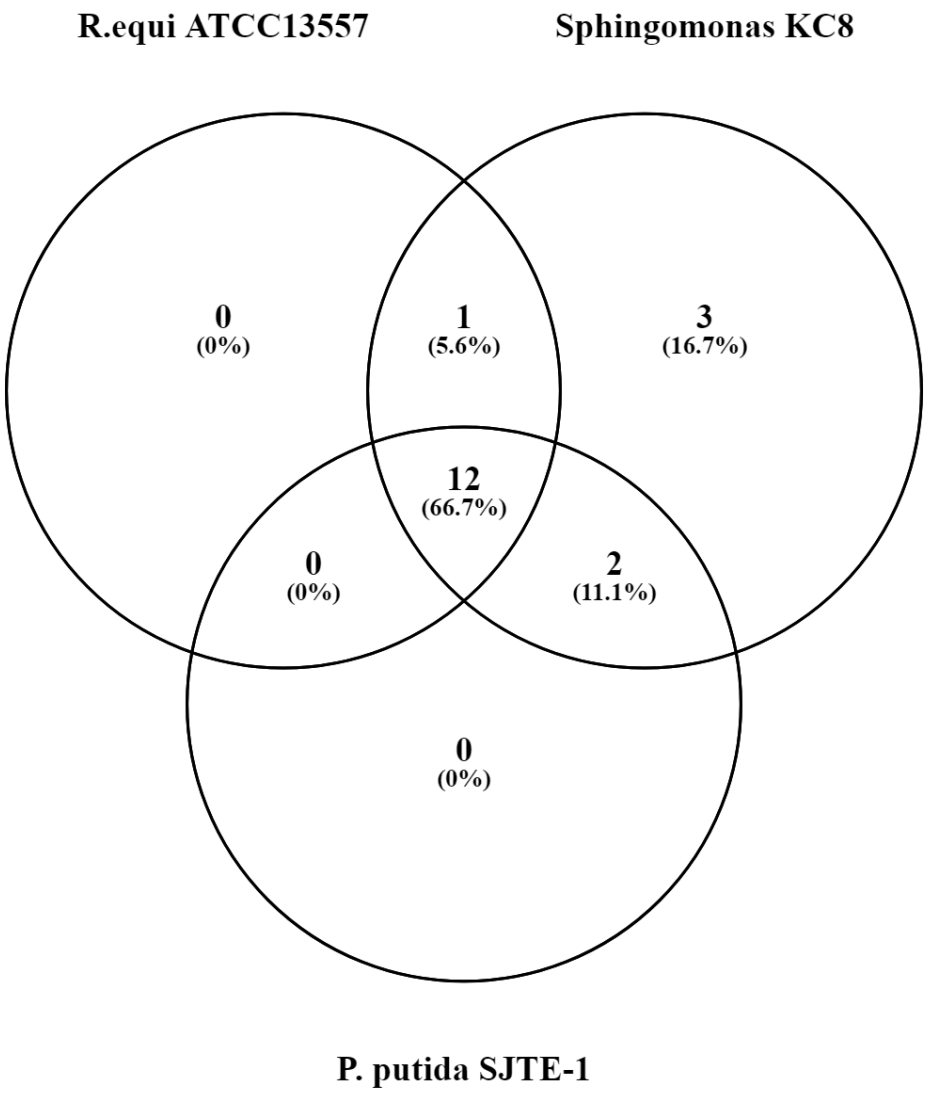

Figure 3-11: Venn diagram produced using Venny 2.1.0 (Oliveros, 2007-15) showing the presence of genes encoding hydrolase in the genomes of R. equi ATCC13557, P. putida SJTE-1 and Sphingomonas sp. KC8.

Additionally, when searching for the genes coding for enzymes, in the BLAST results, there was a cluster of genes coding for enzymes potentially involved in estrogen degradation identified, highlighted yellow (Appendix, Table 0- 7), within *R. equi* ATCC13557 and also present in *Sphingomonas* KC8. In the cluster, a number of the genes coding for enzymes suggested to be involved in estrogen degradation were on the same contig. The cluster contained the genes coding for Enoyl-CoA hydratase (EC 4.2.1.17), 3-hydroxyisobutyrate dehydrogenase (EC 1.1.1.31), 3-hydroxyisobutyryl-CoA hydrolase (EC 3.1.2.4), Butyryl-CoA dehydrogenase (EC

1.3.99.2), and Methylmalonate-semialdehyde dehydrogenase (EC 1.2.1.27), however the function of these enzymes, is suggested to be valine degradation, according to RAST (Aziz *et al.,* 2008; Overbeek *et al.,* 2014; Brettin *et al.,* 2015). Furthermore, the percentage identity between the gene sequences in the two genomes was low (42.18% - 64.26%), suggesting that the sequences don't match very well.

Whilst searching the genomes individually for the genes coding for enzymes potentially involved in estrogen degradation using the RAST SEED viewer (Aziz *et al.,* 2008; Overbeek *et al.,* 2014; Brettin *et al.,* 2015), a gene cluster, was identified. The presence of a gene cluster suggests that the genes are evolutionarily conserved, containing genes similar to those identified as coding for enzymes potentially involved in estrogen degradation (Fig. 3 – 13). The genome of *Novosphingobium aromaticivorans* (Balkwill *et al.,* 1997) contained a similar cluster (Fig. 3 – 13). The gene cluster contained genes coding for 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase, which catalyses the conversion of 4,9-DHSA to 2-hydroxyhexa-2,4-dienoic acid and 3-[(3aS, 7aS)-7a-methyl-1,5-dioxooctahydro-1H-inden-4-yl] propanoate in reaction 12 (Fig. 2- 1 and Table 2- 1). Also a gene coding for dioxygenase was present, which catalyses; EE2 to 2-hydroxy-EE2 in reaction 1, E2 to 4-hydroxy-E2 in reaction 6, E1 to 4-hydroxy-E1 in reaction 8 of the potential estrogen degradation pathway (Fig 2- 1 and Table 2- 1).A gene coding acyl-CoA dehydrogenase within the cluster may have a role in the degradation of estrogen as in the degradation of E1 by the lactone pathway, it is degraded via the TCA cycle (Fig. 2- 1). The presence of these genes encoding for enzymes within the hypothetical estrogen degradation pathway (Fig. 2- 1) suggest that *Sphingomonas* sp. KC8 may be able to metabolise E1, E2 and EE2, furthermore the metabolism of E1, E2 and testosterone has been previously confirmed (Roh and Chu, 2010). However, the metabolism of EE2 has not been reported in *Sphingomonas* sp. KC8 but in a different species *Sphingomonas* sp. JCR5 (Haiyan *et al.,* 2007) and if testosterone and estrogen share a conserved core metabolic pathway then EE2 degradation may be possible. The metabolism of estrogens by the PAH degrader *Novosphingobium aromaticivorans* (Bell and Wong, 2007) is unreported however, *Novosphingobium tardaugens* sp. AR1 metabolises E1, E2 and E3 (Fujii *et al.,* 2002;

Fujii *et al.,* 2003), *Novosphingobium* sp. JEM-1 degrades E1, E2 and EE2 (Hashimoto *et al.,* 2009), and *Novosphingobium* sp. E2S degrades E2 (Li *et al.,* 2017). Additionally, the presence of dioxygenase within the gene cluster within *Sphingomonas* sp. KC8 and *N. aromaticovorans* may suggest a conserved mechanism of A ring cleavage leading to the complete metabolism of EE2.

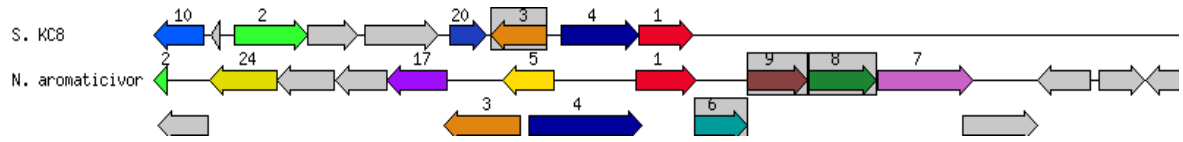

Figure 3-12: RAST SEED viewer diagram of Sphingomonas KC8 (Hu et al., 2011) gene cluster showing the genes coding for 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase, annotated as Tes D in N. aromaticivorans (EC 3.7.1.-) (red, 1), Acyl-CoA dehydrogenase, short-chain specific (EC 1.3.99.2) (Green, 2), 2,3-dihydroxybiphenyl 1,2-dioxygenase (orange, 3), Acyl-CoA dehydrogenase; probable dibenzothiophene desulfurization enzyme (blue, 4), Rieske (2Fe-2S) domain-containing protein (blue, 10), and transcriptional regulator, TetR family (blue, 20). The grey arrows between genes labelled 2 and 20 are genes where the relative position is conserved in at least four other species and are functionally coupled, these genes are coding for the enzymes Enoyl-[acyl-carrier-protein] reductase [NADPH] (EC 1.3.1.10) and Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases; sulfonate monooxygenase. In N. aromaticivorans the additional genes code for Probable short-chain type dehydrogenase/reductase (EC 1.-.-.-) (yellow, 5), 2-keto-4-pentenoate hydratase (EC 4.2.1.80) (turquoise, 6), fumarate reductase/succinate dehydrogenase flavoprotein, N-terminal:FAD dependent oxidoreductase (pink, 7), 4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.39) (green, 8), Acetaldehyde dehydrogenase, acetylating, (EC 1.2.1.10) in gene cluster for enzymes involved in the degradation of phenols, cresols, catechol (brown, 9), putative CoA-transferase subunit alpha Rv3551/MT3655 (EC 2.8.3.-) (purple, 17), and 2-nitropropane dioxygenase (EC 1.13.11.32) (yellow, 24). Additionally the grey arrows represent in order left to right, genes coding for probable enoyl-CoA hydratase EchA20, Enoyl-CoA hydratase/carnithine racemase, putative CoA-transferase subunit beta Rv3552/MT3656, putative oxidoreductase, Sugar phosphate isomerases/epimerases, hypothetical protein and Cytochrome P450.

In summary, the genome of *R. equi* ATCC13557 a known estrogen degrader was sequenced and assembled for the first time using both CLC and SPAdes, SPAdes produced a better genome assembly using the PE Illumina MiSeq data for the GC rich bacterial genome. The genome of *R.equi* ATCC13557 is slightly larger than that of *R. equi* 103S (5.04 Mb) but smaller than the genome of *R. equi* sp. P14 (5.68 Mb) (Letek *et al.,* 2010; Zhang *et al.,* 2012). Furthermore, the *R.equi* ATCC13557 genome contained a greater number of CDs involved in the central aromatic pathway (59), compared to the diverse catabolic genome of *R.* sp. RHA1 (8) (McLeod *et al.,* 2006). Although less peripheral aromatic pathways (14), compared to *R.* sp. RHA1 (26) (McLeod *et al.,* 2006), the percentage relative to the total number of CDSs is the same in both genomes (0.28%). The assembled *R. equi* ATCC13557 genome was compared with the genomes of *Sphingomonas* sp. KC8 and *Pseudomonas putida* SJTE-1 (Hu *et al.,* 2011; Liang *et al*., 2010), however only a small number of genes,

mostly coding for enzymes essential for survival, and none of the genes with above 70% identity between the genomes were involved in estrogen degradation. When the genomes *R. equi* ATCC13557, *Sphingomonas* sp. KC8 and *Pseudomonas putida* SJTE-1 (Hu *et al.,* 2011; Liang *et al*., 2010) were searched using a database of gene sequences coding for enzymes potentially involved in estrogen degradation, again there were few similarities between the genomes. However, a gene cluster was identified within the genome of *R. equi* ATCC13557 coding for enzymes that may catabolise reactions within the potential estrogen degradation pathway. To confirm the genes are coding for enzymes involved estrogen degradation, gene expression analysis experiments will be required to determine if the genes are active during estrogen degradation. If active, the presence of the gene encoding 3-carboxyethylcatechol 2,3-dioxygenase, suggests that *R. equi* ATCC13557 may be able to degrade; EE2 in reaction 1, E2 in reaction 6, E1 and its metabolites in reactions 8 and 11 (Fig 2-1; Table 2- 1). If active, the gene encoding 2-hydroxymuconic semialdehyde hydrolase, suggests that *R. equi* ATCC13557 can degrade E1 metabolites in reaction 12 (Fig 2-1; Table 2- 1). If active, the presence of the gene coding for 3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase, suggests that *R. equi* ATCC13557 can convert E2 to E1 in reaction 7 (Fig 2-1; Table 2- 1). If active, the gene encoding 3-(3-hydroxy-phenyl)propionate hydroxylase, suggests that *R. equi* ATCC13557 can convert E2 to E3 or keto E2 in reactions 4 and 5 and convert E1 to 3-HSA and further degrade 3-HSA to 3,4-DHSA in reactions 9 and 10 (Fig 2-1; Table 2- 1). However, the genes coding for hydratase and the aldolase-dehydrogenase complex were not significantly identified within the genome of *R. equi* ATCC13557 (Fig 2-1; Table 2- 1), therefore the conversion of E1 to E3 in reaction 3 is not possible (Fig 2-1; Table 2- 1) and the further degradation of the E1 metabolites in reactions 13 and 14 are not possible (Fig 2-1; Table 2- 1). Thus it is not likely that *R. equi* ATCC13557 will be able to convert E1 to E3 or completely metabolise E1.

Chapter 4: Developing a sample preparation method for the quantification of estrogens in degradation experiments

4.1 Introduction

Estrogens have a limited aqueous solubility ranging from 0.8-19.1 mg/L and are moderately hydrophobic with $LogK_{ow}$'s of 3.1-4.15 (Table 4- 1). Furthermore the aqueous solubilities of estrogen vary dependent upon the study conditions and EE2 has been reported as the estrogen with the lowest solubility in some studies and with the highest solubility in others (Table 4- 1). The variations in solubility may be due to the pH or ionic strength of the solution, as estrogens have greater solubility in alkaline conditions (pH10) and are less soluble within solutions with higher ionic strength (Shareef *et al.,* 2006).

Table 4-1: The physiochemical properties of estrogens E1, E2 and EE2 (adapted from Shareef et al., 2006)

E1		**E2**		**EE2**		**References**
Solubility in water (mg/L)	LogKow (octanol-water partition coefficient)	Solubility in water (mg/L)	LogKow (octanol-water partition coefficient)	Solubility in water (mg/L)	$LogK_{ow}$ (octanol-water partition coefficient)	
0.8-12.4	3.1-3.4	3.9-13.3	3.1-4.0	-	-	Hanselman *et al.,* 2003
1.30±0.08	-	1.51±0.04	-	9.20±0.09	-	Shareef *et al.,* 2006
0.8	-	3.9	-	9.7	-	Hurwitz and Liu, 1977
12.4	-	12.96	-	4.83	-	Tabak *et al.,* 1981
1.53	-	3.85	-	19.1	-	Yalkowsky, 1999
2.1	-	3.1	-	3.1	-	Yu *et al.,* 2004
13	3.43	13	3.94	4.8	4.15	Lai *et al.,* 2000
-	-	5	-	10	-	Kabasakalian *et al.,* 1966

In order to measure the relatively high estrogen concentrations necessary for biodegradation studies it is therefore necessary to optimise the sample preparation method in order to; i) ensure that the estrogens are added to, and dissolved fully, into the aqueous growth media, ii) halt the degradation at a specific time point for sample storage prior to sample measurement and iii) reduce and/or account for losses to ensure measured estrogen concentrations are reliable. There are different methods to remove the bacterial biomass from samples (Table 4- 2), but filtration is used most commonly within pure culture degradation experiments. Following filtration to remove the bacterial biomass it is assumed that the concentration in the filtered sample would remain stable and undegraded until measurement. However, the filter material has also shown to retain estrogen and therefore would lead to a lower measurement of estrogen in the sample (Han, *et al.,* 2012). Furthermore, there are numerous methods of adding estrogen to the growth media by dissolving first in an organic solvent and then addition to the aqueous media, which is necessary due to the low solubility of estrogen in aqueous solution (Shareef, *et al.,* 2006) (Table 4-2). Due to the concerns regarding concentrations of estrogen within the environment there are numerous developed methods for the chemical analysis of estrogen within environmental samples such as river water for example the sample preparation by use of solid phase extraction (SPE) using an Oasis MCX and LiChrolut EN, elution with methanol, ethyl acetate followed by analysis using the Luna C8 (Positive-ion mode: Solvent A: formic acid 0.1% in Milli-Q water, pH 2 Solvent B: acetonitrile) (Negative-ion mode: Solvent A: triethylamine, pH 8, 0.05% in water Solvent B: acetonitrile) using high-performance liquid chromatography (HPLC) (Castiglioni *et al.,* 2005). However there are fewer methods for detection of estrogens within pure culture and the methods differ from each other in both how the samples are prepared and how the estrogens are analysed, such as different filter material and size, different stock solutions, and different analytical detectors (Table 4- 2). Furthermore there is often an absence of the filtration material or sizing.

Table 4-2: Methods for sample preparation and estrogen analysis.

Bacterial removal method	Filter size (µm)	Stock solution.	Chemical analysis method	References
Millex LG13 filter	0.2	Powder dissolved directly into media	GC/MS and NMR	Fujii et al., 2002; Fujii et al., 2003
Ethyl ester extraction	N/A	Above 3 mg/L in acetone	GC/MS and YES assay	Yu et al., 2007
Schleicher and Schuell, Germany filter	0.2	Acetone	HPLC-UV/DAD-FLD	Weber et al., 2005.
Centrifugation and methylene chloride extraction	N/A	Methanol	GC/MS and YES assay	Jiang et al., 2010.
Ethyl acetate extraction	N/A	Acetone	HPLC and MS/MS	Haiyan et al., 2007.
Microporous membrane filter	0.22	Methanol	HPLC	Yu *et al.,* 2016
Solid phase extraction, Sep-Pack plus Florisil and NH2 cartiridges	N/A	N/A	LC/MS/MS	Hashimoto et al., 2009
PVDF filter	0.22	100% ethanol	HPLC equipped with a Diode Array Detector.	Larcher and Yargeau, 2013
Whatman GF/B filter paper	1	Methanol	GC/MS	Yoshimoto et al., 2004

N/A – information was not provided. Abbreviations: Gas Chromatography mass spectrometry (GC/MS); Yeast estrogen screen (YES); Tandem mass spectrometry (MS/MS); Liquid chromatography tandem mass spectrometry (LC/MS/MS).

With few studies focusing upon degradation in pure culture and the absence of a 'gold' standard method for chemical analysis, method development will be required to overcome issues encountered with estrogen detection within different growth media.

The hypotheses are that different filter materials will lead to different percentage retentions of estrogens and that the starting estrogen concentrations will vary between the different organic solvents and evaporation times, as with longer evaporation times the temperature of the growth media will increase and therefore the solubility may increase (Rashid *et al.,* 2014).

4.2 Materials and methods

4.2.1 *Comparison of different methods to separate bacterial biomass from the aqueous phase*

Powdered E1, E2, and EE2 was dissolved using a magnetic stirrer bar into aqueous MSM to achieve the known starting concentration of 5 mg/L in 500mL. An aliquot of 1 mL was taken from this stock solution using a 1.5 mL syringe and was filtered through different filters (Table 4- 4), Polytetrafluoroethylene (PTFE) was pre-treated by adding 1 mL ethanol followed by 1 mL distilled water to the filter using a syring to create a hydrophilic membrane, into amber glass vials in triplicate. Furthermore a set of triplicate samples was centrifuged using the Mikro 200R Hettich zentrifugen at 13,000 x g for 3 minutes and was unfiltered to account for potential losses of estrogens during this step. The samples were then measured using HPLC Shimadzu VP using a Phenomenex Gemini-NX 5u C18 110A column with the settings of flow=1, B = 20%, max. pressure = 4000 psi, low pressure= 200 psi, wavelength= 230nm, Acquisition = 5, runtime = 13 min, injection volume = 10 µL, and the concentration gradients (Table 4- 3). The estrogen samples were quantified relative to the standard curve produced for the duplicate standards of known concentrations (0.5, 1, 2, 3, 4, 6 and 8 mg/L) of E1, E2 and EE2, which were produced in 50:50 AcN: MeOH (Appendix Fig. 0- 1, 2 and 3).

Table 4-3: HPLC concentration gradient used for the quantification of estrogens

Concentration of acetonitrile (%)	Time (min)
40	1
95	7
95	9
40	10
40	13

Table 4-4: The different filtration materials assessed for sample preparation for the quantification of estrogens.

Filter	Material description	Pore size (µm)	Diameter (mm)
Whatman GD/X Cellulose Acetate (CA)	Hydrophillic, suitable for aqueous and solvent media. For biological solutions, high-loading capacity, and low-protein-binding.[a]	0.2	25
Millex® Durapore® Polyvinylidene fluoride (PVDF)	Hydrophillic, the most suitable membrane for HPLC.[a]	0.22	13
Nalgene™ Polyethersulfone (PES)	Hydophillic, offers higher flow rates, lower protein binding and lower extractables than competitive cellulosic membranes. For serum, plasma, and tissue culture solutions.[a]	0.2	25
Sigma Nylon membrane (Nylon)	Hydrophillic, for aqueous and aqueous organics. High non-specific binding, they are not recommended for protein solutions.[a]	0.2	25
VWR Polytetrafluoroethylene filters (PTFE treated)	Hydrophillic due to pretreatment. Designed for filtration of corrosive gases and organic solvents. Excellent chemical resistance. Ultimate compatibility for filtering harsh chemicals.[b]	0.2	25
VWR Polytetrafluoroethylene hydrophobic (PTFE hydrophobic)	Hydrophobic. Designed for filtration of corrosive gases and organic solvents. Excellent chemical resistance. Ultimate compatibility for filtering harsh chemicals.[b]	0.2	13

Material descriptions taken directly from (a) Sigma Aldrich website, (b) VWR website.

4.2.2 *Comparison of methods for dissolving estrogen into aqueous growth media*

The aim of this experiment was to improve solubility by; i) experimenting with different evaporation times, ii) dissolving estrogen into a solvent either acetonitrile: methanol (AcN:MeOH) or acetone, and iii) experimenting with the composition of the growth medium to which the estrogen is dissolved due to the possibility that addition to a more ionic solution may lead to reduced solubility (Shareef *et al.,* 2006).

Stock solutions of 1 g/L of mixed estrogen (E1, E2, and EE2) were prepared in acetone and 50:50 AcN:MeOH. 50 mL glass flasks were prepared containing 5 mL of *R. equi* ATCC13557 bacterial culture in 40 mL of MSM containing 0.6 g/L of yeast extract.

The time taken to fully evaporate 0.25 mL of Acetone and 0.25 mL of AcN:MeOH, within a total volume of 1 mL MSM at 80°C using a LABCONCO RapidVap vertex evaportator within a fume hood were measured as 50 and 80 minutes respectively.

Then the methods of adding the estrogen solutions to the MSM to achieve a concentration of 5 mg/L and different evaporation times were varied (Table 4- 5). Triplicate samples were taken from each condition and centrifuged using the Mikro 200R Hettich zentrifugen at 13,000 x g for 3 minutes and filtered through a 0.2 µm PVDF filter into amber glass vials using a 1mL syringe and analysed by uHPLC (ThermoScientific Dionex Ultimate 3000 RS autosampler, pump, column compartment and electrochemical detector) using a ThermoScientific Accucore C18 100 x 2.1 mm, 2.6µm particle size column with the settings of; flow rate = 0.4 mL/min, temperature = 40°C, injection volume = 50 – 90 µL, electrochemical potential = 1800 mV. The run was isocratic using 62% of mobile phase A and 38% of mobile phase B for 4- 6.5 minutes. Mobile phase A contained 95% water, 5% acetonitrile and 0.1% formic acid. Mobile phase B contained 95% acetonitrile and 5% water with 0.1% formic acid. The detection limit of estrogen was 100ng/L. The estrogen samples were quantified relative to the standard curve produced for the duplicate standards of known concentrations (1, 2.5, 5, 7.5 and 10 mg/L) of E1, E2 and EE2, which were produced in 50:50 AcN:MeOH (Appendix Fig. 0- 4, 5 and 6).

Table 4-5: The different conditions of the 50 mL flasks.

Evaporation	Stock solvent	Composition of media
1. No evaporation	a. Acetone	MSM
	b. AcN:MeOH	MSM
2. Half (25 min)	c. Acetone	MSM
Half (40 min)	d. AcN:MeOH	MSM
3. Full (50 min)	e. Acetone	MSM
Full (80 min)	f. AcN:MeOH	MSM
4. Full (80 min)	a. AcN:MeOH	20 mL water
	b. AcN:MeOH	Salts solution

The method of adding the estrogen to the flask was further tested to determine if the method was valid for use in the 500 mL flasks for the degradation experiments. A 1 g/L 50:50 AcN:Me 25 mL stock solution was prepared for mixed estrogen (E1, E2, and EE2) and for EE2 only. Then 2 mL of the stock solution was evaporated to 0.25 mL using the LABCONCO RapidVap vertex evaporator at 60°C and then added to the 400ml of MSM to produce a concentration of 5 mg/L. Followed by further evaporation on a hot plate at 80°c for 80 minutes. Triplicate samples of 1 mL were taken from each flask and centrifuged at 13,000 x g for 3 minutes and filtered through a 0.2 µm PVDF filter into amber glass vials using a syringe. The concentrations of estrogen were analysed using uHPLC (ThermoScientific Dionex Ultimate 3000 RS autosampler, pump, column compartment and electrochemical detector) using a ThermoScientific Accucore C18 100 x 2.1 mm, 2.6µm particle size column with the settings of; flow rate = 0.4 mL/min, temperature = 40°C, injection volume = 50 – 90 µL, electrochemical potential = 1800 mV. The run was isocratic using 62% of mobile phase A and 38% of mobile phase B for 4- 6.5 minutes. Mobile phase A contained 95% water, 5% acetonitrile and 0.1% formic acid. Mobile phase B contained 95% acetonitrile and 5% water with 0.1% formic acid. The detection limit of estrogen was 100ng/L. The estrogen samples were quantified relative to the standard curve produced for the duplicate standards of known concentrations (1, 2.5, 5, 7.5 and 10 mg/L) of E1, E2 and EE2, which were produced in 50:50 AcN:MeOH (Appendix Fig. 0- 7, 8 and 9).

4.2.3 *Statistical analysis and graphing software*

Statistical analysis was carried out using Minitab 17:

The data was determined to be not normal using the Anderson Darling normality test.

Kruskal-Wallis test was used to determine the significance of the losses of estrogen due to the different bacterial removal methods, followed by the use of the Games-Howell method to identify which methods were significantly different to the unfiltered control samples.

Kruskal-Wallis test was used to determine if there was a significant difference between the losses of estrogen at different evaporation times.

Mann-Whitney test was used to determine the significance of the differences in estrogen losses due to the different stock solutions prepared with different organic solvents.

Kruskal-Wallis test was used to determine the significance in the losses of estrogens due to the addition of estrogen into the different components of MSM.

All graphs were produced using SigmaPlot version 12.5.

4.3 Results

4.3.1 *Comparison of different methods to separate bacterial biomass from the aqueous phase.*

The concentration of EE2 was not quantifiable and was not detected in any samples. Estrogen retention varied widely between replicates (n=3) for centrifugation and PTFE with average coeffiecients of variation of 35.8% and 35.6% for E1 respectively, and 36.6% and 19.5% for E2 respectively. The average retentions of E1 and E2 were the lowest for the filter material PVDF of 9.09% and 21.44% respectively. Furthermore there was less variation between replicates (n=3) for the method using PVDF filters with average coefficients of variation of 0.57% and 15.9% for both E1 and E2 respectively. Losses of E1 and E2 using untreated PTFE were 36.36% and

55.56% respectively. There were significant differences for E2 or E1 loss between the different biomass removal methods (P=0.038). There was a significant loss of estrogen using PTFE (hydrophobic) (P=0.021), but there was no significant loss of estrogen, compared to the unfiltered samples, when using PVDF (P=0.159), or centrifugation (P=0.284). Estrogens were undetected after filtering with the following materials; CA, PES, Nylon and PTFE (treated), indicating that they were removed completely by these filters. There were additional losses of estrogen due to centrifugation of the samples prior to filtration of 27.27% and 3.33% for E1 and E2 respectively (Fig. 4- 1).

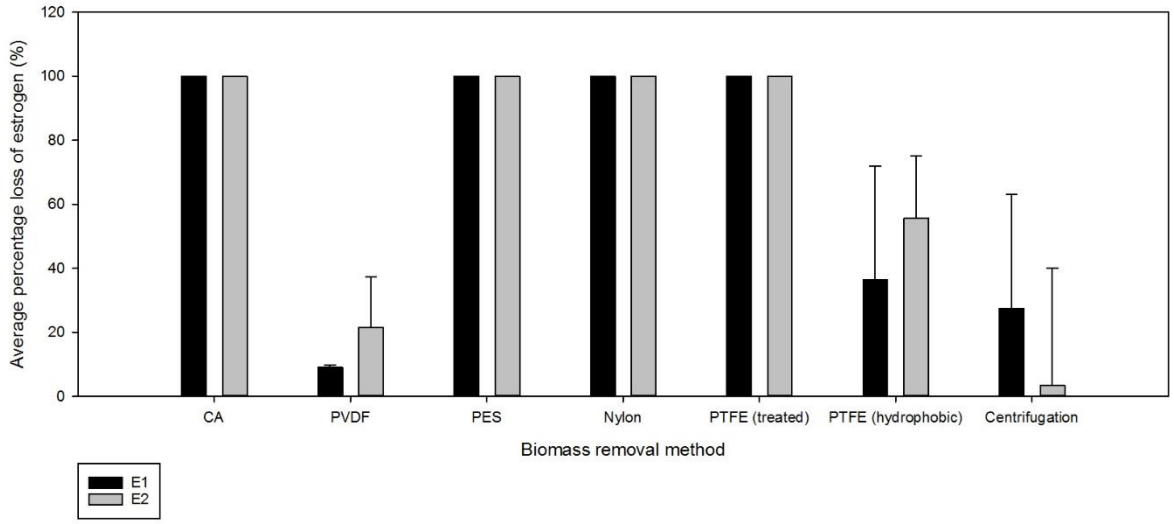

Figure 4-1: Comparison of the average percentage loss of estrogen (%) resulting from different bacteria removal methods from aqueous samples. The methods included the different filter materials; Cellulose Acetate filtration (CA), Polyvinylidene fluoride filtration (PVDF), Polyethersulfone filtration (PES), Nylon membrane filtration (Nylon), Polytetrafluoroethylene filters treated with ethanol and distilled water to create a hydrophilic membrane (PTFE treated), Polytetrafluoroethylene hydrophobic filtration (PTFE hydrophobic), and centrifugation of the aqueous samples (Centrifugation). The error bars represent standard deviation of the mean.

4.3.2 *Comparison of methods for dissolving estrogen into aqueous growth media*

Firstly, considering the effect of evaporation time upon the solubility of estrogens of the samples prepared by the addition of the AcN:MeOH mixed estrogen stock solution. The average percentage loss of E2 was significantly less when the evaporation time was 80 and 50 minutes; 5.13±0.42% and 12.10±0.24% respectively (P=0.006) (n=9, Fig. 4- 2 and 3). The average percentage loss of E1 was significantly lower when the evaporation time was 80 and 25 minutes; 6.46±0.43% and 14.37±0.41% respectively (P=0.002) (n=9, Fig. 4- 2 and 3). Although the

average percentage loss of EE2 was least at the higher evaporation times, there was no significant difference between the different evaporation times (P=0.079) (n=9, Fig. 4- 2 and 3).

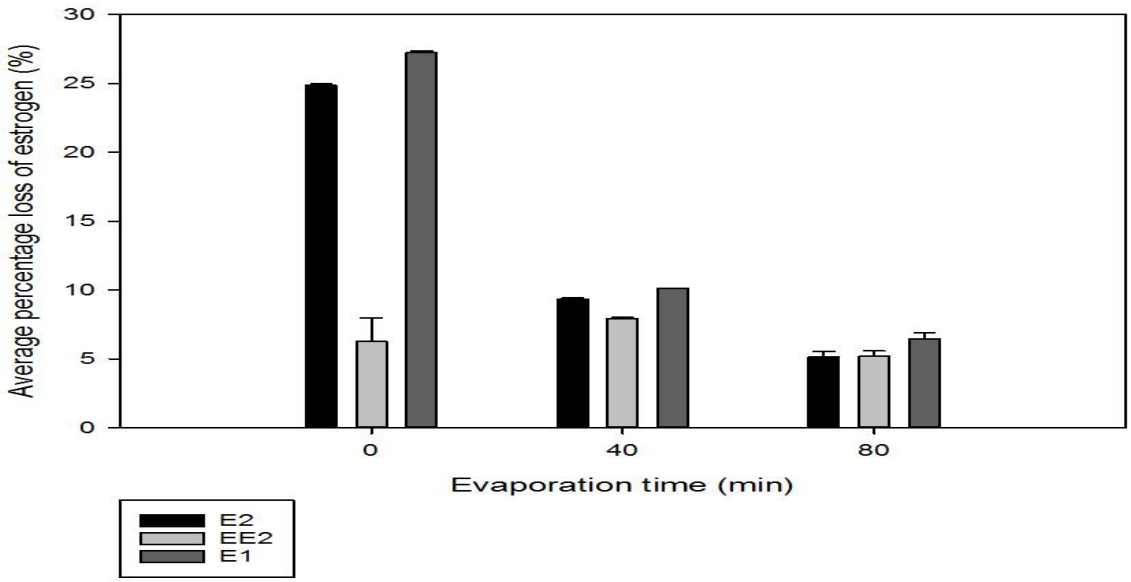

Figure 4-2: Comparison of the average percentage loss of estrogen at different evaporation times; 0, 40 and 80 minutes, following the addition of the estrogens from a 1g/L AcN:MeOH stock solution. The error bars represent standard deviation of the mean.

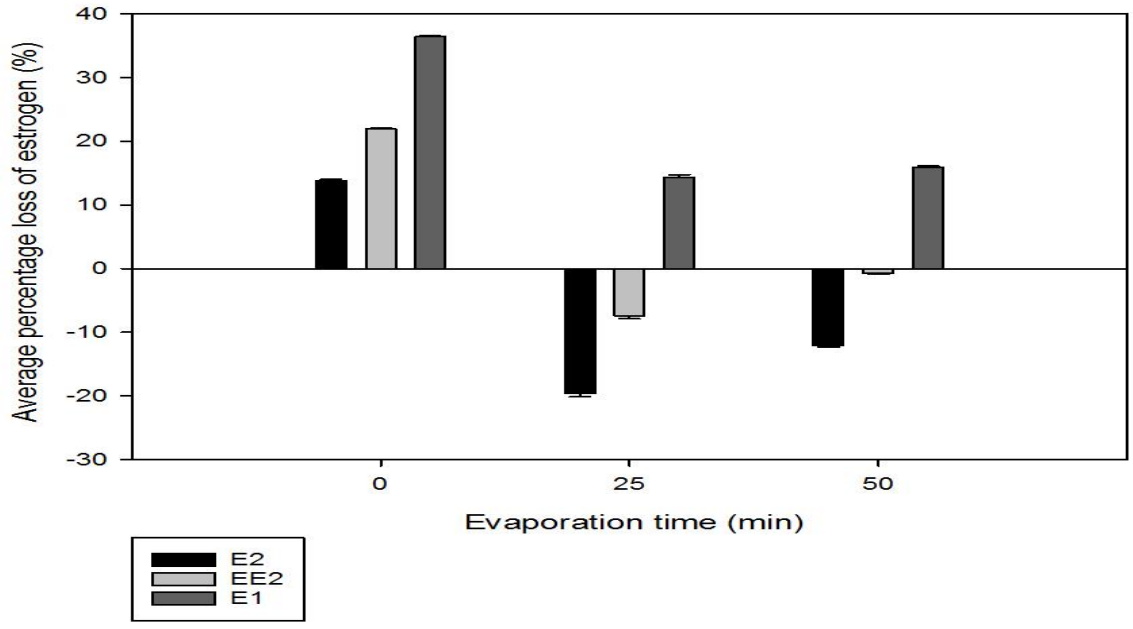

Figure 4-3: Comparison of the average percentage loss of estrogen at different evaporation times; 0, 25 and 50 minutes, following the addition of the estrogens from a 1g/L Acetone stock solution. The error bars represent standard deviation of the mean.

In the comparison of the solubility of estrogens when added from stock solutions containing different organic solvents, the average percentage loss of E2 was significantly lower for added in the AcN:MeOH stock solution with 40 and 80 minutes of evaporation; 4.53±0.07% and 5.13±0.42% respectively (P=0.0341) (n=9, Fig. 4- 5 and 6). The average percentage loss of E1 was significantly lower when the estrogen was added in the AcN:MeOH stock solution for 40 and 80 minutes of evaporation; 10.10±0.03% and 6.46±0.43% (P=0.0453) (n=6, Fig. 4- 5 and 6). The average percentage loss of EE2 was lower for methods when added in AcN:MeOH stock without evaporation, but was significantly lower when added in the acetone stock solution for 25 and 50 minutes of evaporation; -7.40±0.50% and -0.68±0.22% (P=0.0306) (n=6, Fig. 4- 4, 5, 6).

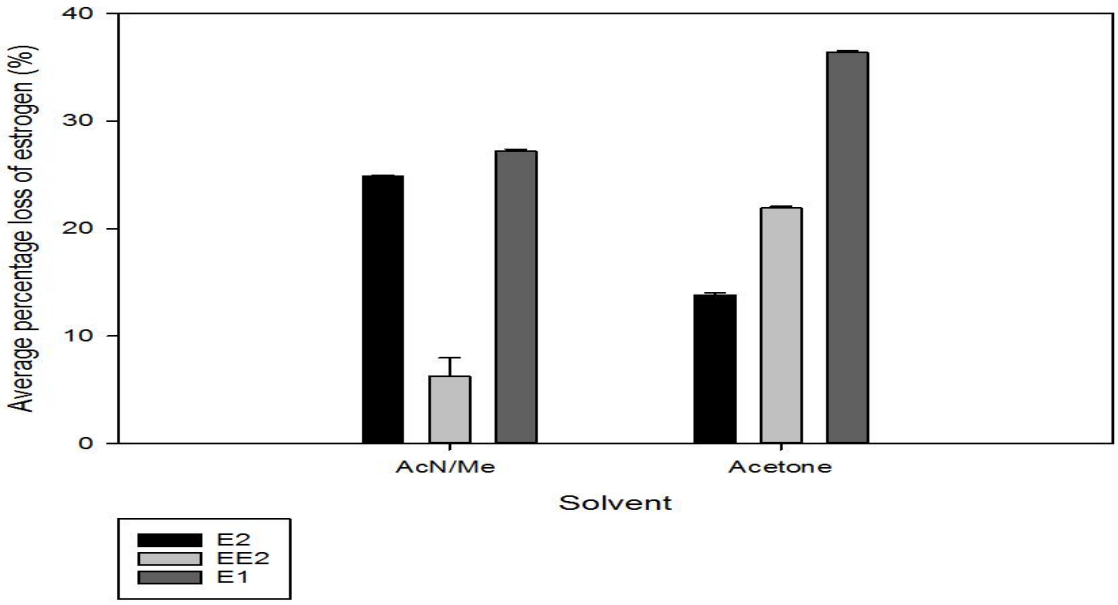

Figure 4-4: Comparison of the average percentage loss of estrogen without evaporation, when the estrogen is added to the MSM growth media using the different solvents Acetone and AcN:MeOH. The error bars represent standard deviation of the mean.

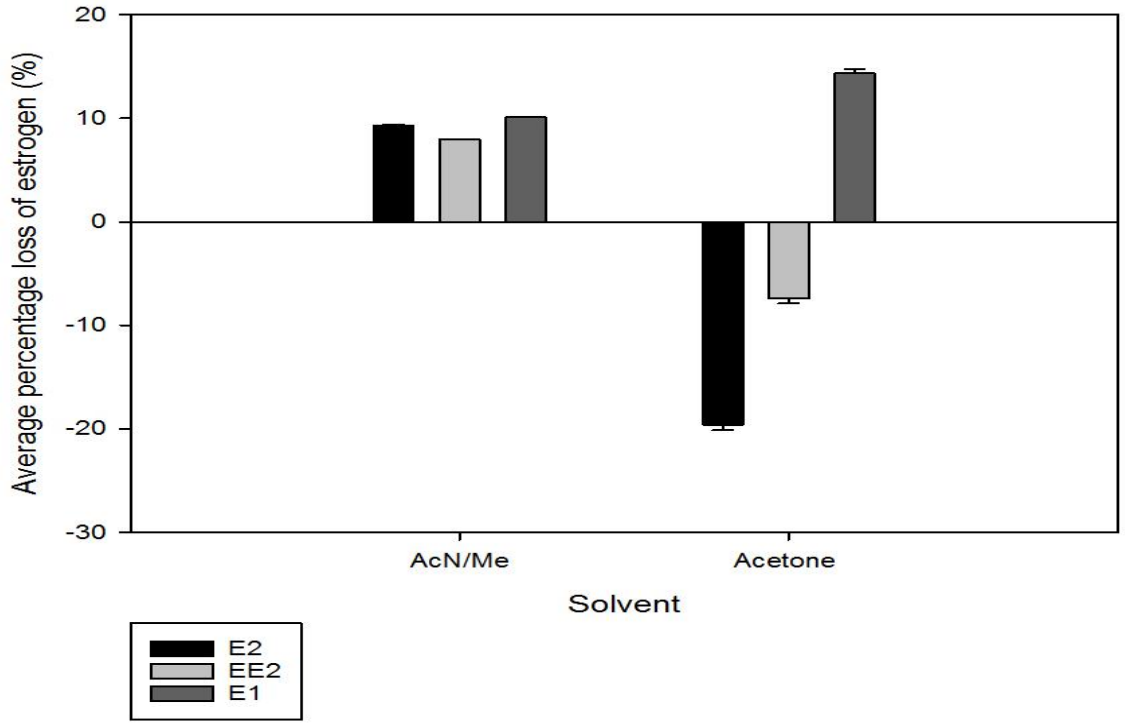

Figure 4-5: Comparison of the average percentage loss of estrogen with half evaporation, when the estrogen is added to the MSM growth media using the different solvents Acetone and AcN:MeOH. The error bars represent standard deviation of the mean.

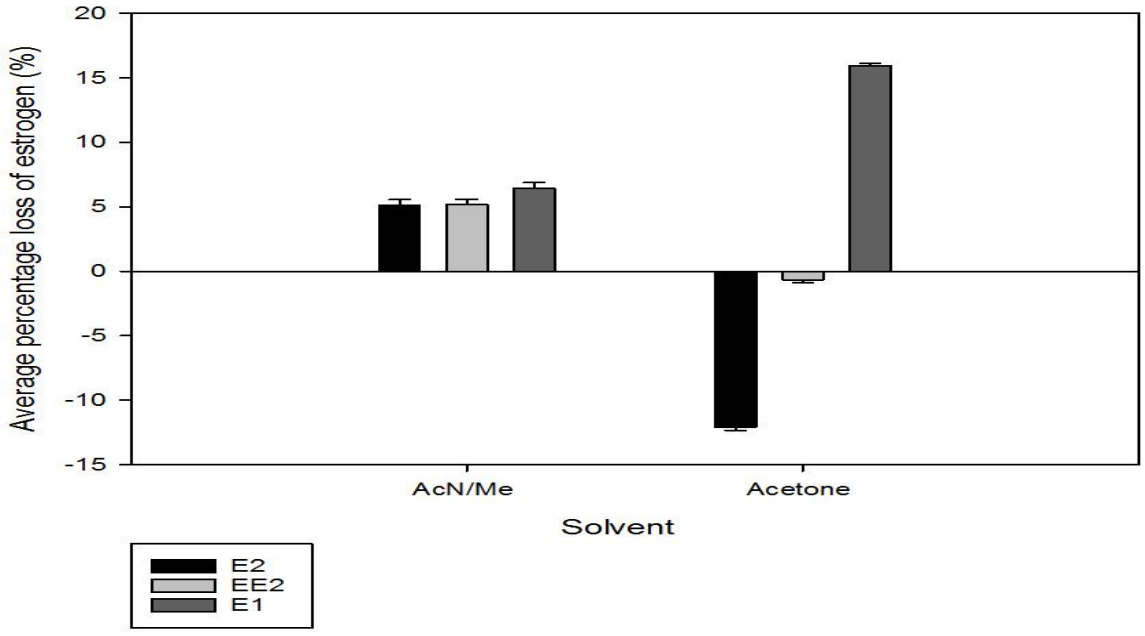

Figure 4-6: Comparison of the average percentage loss of estrogen with full evaporation, when the estrogen is added to the MSM growth media using the different solvents Acetone and AcN:MeOH. The error bars represent standard deviation of the mean.

Then comparing the addition of estrogen to the different components of the MSM to determine if the ionic strength of the salts solution, MSM or water only would affect the solubility of estrogens. The percentage loss of E1, E2 and EE2 was significantly less for addition of the AcN:MeOH mixed estrogen stock into the MSM growth media with 80 min evaporation of 6.46±0.43%, 5.13±0.42% and 5.18±0.4% respectively when comparing the addition of estrogen to the different components of the growth media (P=0.027) (n=3, Fig. 4- 7).

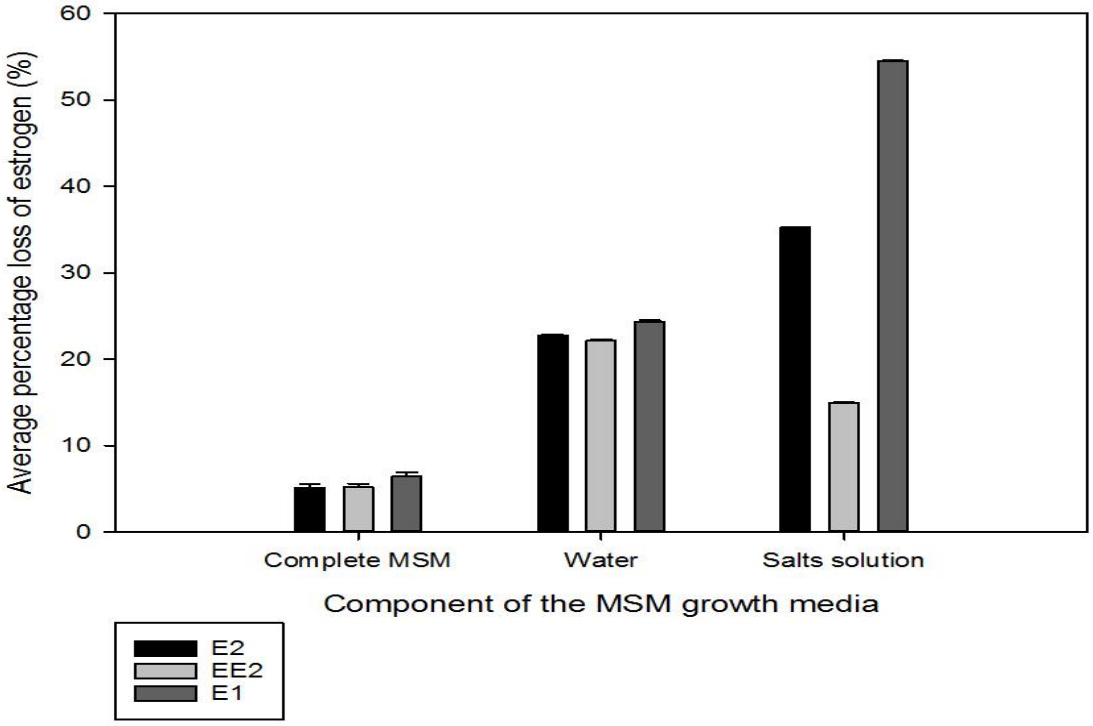

Figure 4-7: Comparison of the average percentage loss of estrogen when adding a 1g/L mixed estrogen AcN:MeOH stock solution into the different components, water, salts and complete MSM growth medium. The error bars represent standard deviation of the mean.

Following the method development there was still a difference between the quantified estrogen concentrations and the expected concentrations of 5 mg/L. The concentration of EE2 was detected at 3.32±0.07 mg/L higher than expected in the EE2 only flask. In the mixed estrogen flask there was a higher than expected concentration of both E2 and EE2 of 0.17±0.38 mg/L and 0.39±0.47 mg/L respectively, and a lower than expected concentration for E1 of 0.28±0.61 mg/L (Fig. 4- 8).

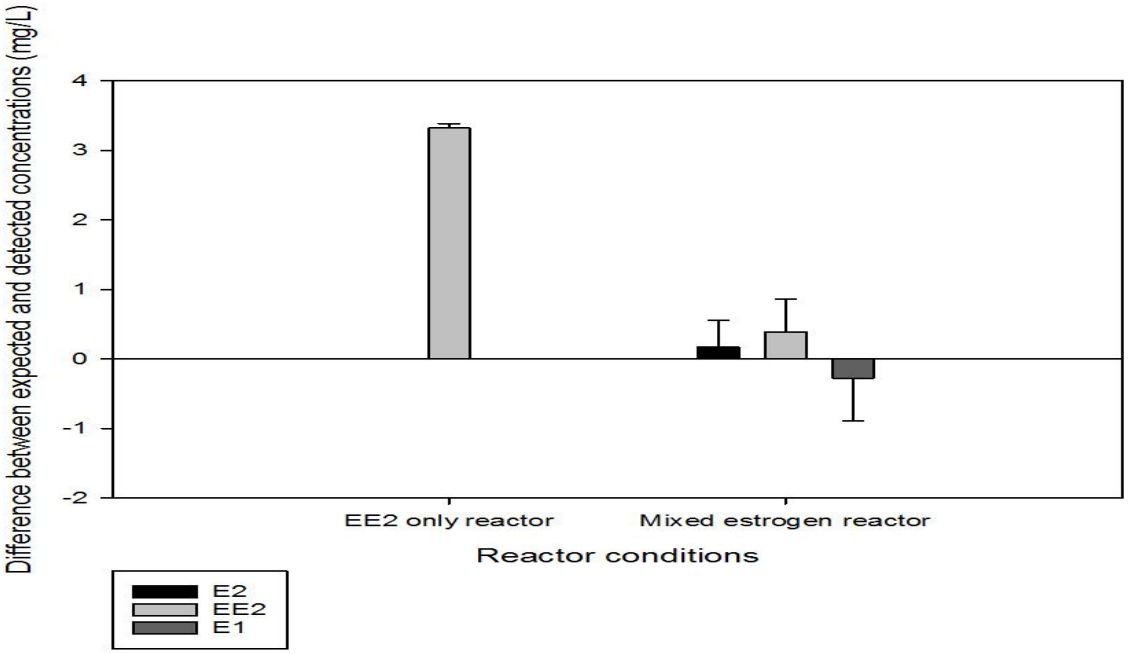

Figure 4-8: Comparison of the difference between the initial concentration detected and the actual concentration of 5 mg/L when the 50:50 AcN:MeOH estrogen stock solution is added to the MSM, and the residual organic solvent evaporated off for 80 min. The error bars represent standard deviation of the mean.

4.4 Discussion

The filter material with the least retention of estrogen was PVDF so it was chosen as the best filter for these experimental conditions. The filter material PVDF has also been used previously (Table 4- 2) (Larcher and Yargeau, 2013). The non-detection and significant loss of estrogen from the other filters is likely due to adsorption. It has been previously reported that estrogens adsorb to various filter materials including cellulose containing filters such as CA, paper, and glass fibre at varying amounts (Schicksnus and Müller-Goymann, 2000; Walker and Watson, 2010). During this experiment it was found that the filter would clog when there was a greater amount of bacteria present in samples, therefore it was necessary to first centrifuge the samples and filter the supernatant. Centrifugation led to the loss of some estrogen, in addition to the loss of estrogen due to filtration. Therefore, these losses of estrogen will be considered in the measured estrogen concentrations during the degradation experiments. The loss in centrifugation is likely due to sorption of the estrogen to the polypropylene centrifuge tubes or to the plastic syringe used during filtration. Although, as in these experiments, the losses due to the syringe or plastics

were determined not significant compared to the losses due to filtration (Walker and Watson, 2010). To account for any potential losses of estrogen, however low, the standards will be produced by using the same method of preparation as the samples, by centrifugation and filtration through PVDF to control for losses of estrogen not due to biodegradation. In previous studies, where filtration is the method of sample preparation, the use of a control to account for losses of estrogen due to filtration is unreported (Table 4- 2). Furthermore, it was shown in these experiments that EE2 could not be detected when added in a powder form, like the other estrogens, therefore the solubility of EE2 (<5 mg/L) in aqueous media is much less than that of the other estrogens under these experimental conditions which has been previously reported (Table 4- 1) (Tabak *et al.,* 1981; Lai *et al.,* 2000).

The method of addition of the estrogen to the growth media (MSM) showed that there was a significant difference between the concentrations of estrogen detected following addition of estrogen in various different methods, evaporation time, the solvent used to produce the stock solution of estrogen and to which component of the growth media the estrogen was added. The longest evaporation times of 80 minutes and 50 minutes led to the significantly lower percentage loss of E2. Furthermore the average percentage loss of E1 was significantly lower at 25 minutes and 80 minutes evaporation. Solubility of E1 in acetone at 40°C has been reported as 20.244 mg/L previously and its solubility in alcohols may be lower due to hydrogen-bonding (Ruchelman, 1967). Therefore this may explain the lower percentage loss of E1 when using acetone at 25 min evaporation. The difference due to the evaporation time was not significant for EE2. The AcN:MeOH stock solution led to the lowest percentage loss of E1 and E2 when the evaporation time was 80 minutes and 50 minutes. However, the acetone stock solution with at 25 and 50 minutes evaporation time led to the significantly lower average percentage loss of EE2, this was unsurprising as EE2 is soluble in acetone (Veronez *et al.,* 2015). The average percentage loss of estrogen was significantly lower when the estrogen stock solution was added to the complete MSM, rather than the salts or water components. This is likely due to the low solubility of estrogens in water and the increased ionic strength of the salts solution alone could have led to reduced solubility (Shareef *et al.,* 2006). A single method to add the mixed estrogen to the flask for the degradation study was required, therefore it was decided that the method with the overall lower

percentage loss of estrogen would be to use the 50:50 AcN:MeOH stock solution with an 80 minute evaporation at 80°C into the complete media MSM. Although this was the better method, the difference between the expected and quantified concentrations; EE2 of 0.39±0.47 mg/L when in mixed stock solution and 3.32±0.07 mg/L, E2 of 0.17±0.38 mg/L and E1 of -0.28±0.61 mg/L need to be considered in the degradation experiments. The losses in estrogen are likely to have been in transferring the solution from the test tube into the growth media, because there was visible precipitation of the estrogen within the test tube. This may have been due to salting out due to the salts solution within the MSM having an ionic effect, hydrogen-bonding within the water component of the MSM, or due to the pH of the MSM (Carter and Sluss, 2013; Shareef *et al.,* 2006; Ruchelman, 1967). Although evaporation is carried out it is possible that some residual solvent could remain and the addition of organic solvents to the growth media may add an additional carbon source during degradation studies, although AcN:MeOH is not known to have an effect in enhancing the degradation of estrogens, unlike ethanol (Larcher and Yargeu, 2013). Furthermore, dissolving estrogens into solvents and then cystalising them out before adding to aqueous media, may effect their physicochemical properties, such that the solubility may be altered (Park *et al.,* 2008). For example, the solubility of E2 was reduced when crystalised from acetone (Park *et al.,* 2008). Therefore, due to these findings it would be recommended that in estrogen degradation studies; 1) the filter material should be carefully chosen to reduce the retention of estrogen; 2) the loss of estrogen due to filtration and other sample preparation should be determined and if these losses are significant a control such as processing the standards in the same way or ensuring these losses are accounted for in the degradation concentrations; 3) choice of solvent is carefully considered to reduce solubility issues if present; 4) the effect upon degradation is considered if the solvent remains unremoved.

Chapter 5: Degradation experiments

5.1 Introduction

Chromatography techniques are based on the separation of molecules between the mobile and stationary phases (Forgács and Cserháti, 2003). Liquid chromatography (LC) used a liquid, usually organic solvent mobile phase and a solid stationary phase upon the column (Forgács and Cserháti, 2003). The equipment for LC, including HPLC which uses pressure to force the mobile phase through the system, and ultra-high-performance liquid-chromatography (uHPLC) which uses even higher pressure, by pumping the mobile phase, which is usually an organic solvent such as methanol or acetonitrile, through the system at concentrations specified within the program (Pitt, 2009). By using LC, concentrations of chemical compounds, such as estrogens, can be quantified by injecting samples into the mobile phase which carries the compounds onto the solid phase column where the compounds bind, the compound is then eluted from the column as the concentration gradient of the mobile phase is increased and the compound passes through the detector producing a peak with a peak area relative to the concentration of the compound within the sample (Waters, 2017). Monitoring of estrogen concentrations within the environment is of increasing importance due to the increasing concerns about the effects of estrogen upon aquatic organism and the proposal by the EC of EE2 and E2 as priority substances, with EQS of 0.035 ng/L and 0.4 ng/L respectively (Gilbert, 2012; Johnson *et al.,* 2013).

Furthermore, chemical analysis of estrogens by LC has been used in biodegradation studies to quantify the concentrations of estrogen in the presence of bacteria to study their potential to metabolise estrogens. There have been a number of examples in which *Rhodococci* have been reported to degrade estrogen; *Rhodococcus* sp. P14 was able to convert E2 into E1, *Rhodococcus* sp. DS201 was able to completely metabolise E2 and *R. equi* ATCC13557 was able to metabolise EE2 (Ye *et al.,* 2017; Yu *et al.,* 2016; Larcher and Yargeau, 2013). In order to confirm that *R. equi* ATCC13557 is able to degrade estrogen within my experimental conditions it was necessary to carry out degradation experiments in which the concentration of estrogen is measured at regular intervals using HPLC for the duration of the bacterial growth phase. The hypothesis is that the concentrations of

E2 and EE2 would decrease over time and would possibly be completely metabolised.

5.2 Materials and methods

5.2.1 *Bacterial culture*

R. equi ATCC13557 was acclimated to estrogens for 24 hours, by being cultured in 500 mL Erlenmeyer flasks with vented bungs, containing MSM (See 5.2.1.1) with 0.6 g/L of yeast extract. The conditions in each of the acclimation flasks were without estrogen (control), 5 mg/L of EE2, 5 mg/L of E2, and 5 mg/L mixed estrogen of E1, E2 and EE2. *R. equi* ATCC13557 was previously subcultured numerous times in MSM (see 5.2.1.1) containing 5 mg/L of mixed estrogen of E1, E2 and EE2, 5 mg/L of EE2, 5 mg/L of E2 and without estrogen (control).

5.2.1.1 Preparing the growth media

M9, Minimal Salts, 5X MSM (Sigma Aldrich) containing 33.9g/L Na_2HPO_4, 15g/L KH_2PO_4, 5g/L NH_4Cl and 2.5g/L NaCl was prepared by weighing 56.6g into 1L of distilled water. Separating the 1L salts solution into 5x 200mL Duran flasks and autoclaving at 121°C. The 200 mL salts solutions were the stored at -4°C until required.

To produce the final growth media the 200 mL salts solution was added to 800 mL of distilled autoclaved water containing 0.6 g/L of yeast extract.

5.2.2 *Preparation of the experimental flasks*

To add the estrogen to the MSM (containing 0.6 g/L of yeast extract), 2 mL of a stock solution of 1g/L of estrogen stock solution (see 5.2.2.3) was evaporated by nitrogen gas down to a volume of 0.25 mL. The 0.25 mL estrogen stock was then added to the MSM containing 0.6 g/L of yeast extract and evaporated for a further 80 minutes on a hot plate to remove the remaining acetonitrile and methanol. The solution was

removed from the heat and allowed to cool down, for the 311 hour experiment it was left overnight. Next, 40 mL of the acclimated *R. equi* culture was inoculated into 360 ml of MSM with 0.6 g/L of yeast extract, containing the estrogen at a concentration of 5 mg/L. To prepare the abiotic control the method of adding the estrogen into the MSM containing 0.6 g/L was the same as above however a 40 mL of a 1% formalin solution was added to give a final concentration of 0.1% formalin. The other control was prepared without any estrogen solution.

5.2.2.1 Experimental flask conditions for the 45.5 hour experiment

The conditions were; containing E2; “no estrogen control”; the abiotic control containing E1, E2 and EE2; and a mixed estrogen flask containing E1, E2 and EE2.

5.2.2.2 Experimental flask conditions for the 311 hour experiment

The flasks were set up in triplicate for each condition, the conditions were; containing EE2; containing E2; containing mixed E2 and EE2; “no estrogen control;” and the abiotic control which contained E2 and EE2.

5.2.2.3 Preparing the estrogen stock solution

To prepare the 1 g/L mixed estrogen stock solution, 0.025 g of each estrogen powder (E1, E2, and EE2) were weighed into an amber glass 25 cm^3 volumetric flask. A 50:50 solution of acetonitrile: methanol (AcN:MeOH) was prepared and mixed in a 500 cm^3 volumetric flask, and dispensed into a beaker. The AcN:MeOH solution was transferred into the 25 cm^3 volumetric using a 25 cm^3 Pasteur pipette. The solution was then mixed by inversion and stored at 4°C.

For individual estrogen stock solutions (1g/L) the same protocol was used, however only one estrogen powder was added to the solvent.

5.2.3 *Monitoring for the degradation experiments*

At each time point the following samples were taken (triplicates of each);

- 1 mL - to measure bacterial growth
- 2 mL - for HPLC analysis

To assess bacterial growth optical density (OD), measured in a Spectrophotomer (ATI Unicam 8625 UV/Vis).

Prior to HPLC analysis, samples were processed through sample preparation protocol and stored for estrogen quantification.

5.2.3.1 Monitoring the 45.5 hour degradation experiment

The flasks were incubated at 25°C for 45.5 hours with 155 rpm rotation. There were 10 sampling points; an initial sample was taken at 0 hours, then samples were taken every 1.5 hours between hours 16 to 24.5, followed by a final sample point at 45.5 hours. In addition, 100uL samples were taken from each flask to provide colony counts, which was done by serial dilutions in Ringer's solution and plating in triplicates onto nutrient agar.

5.2.3.2 Monitoring the 311 hour degradation experiment

The flasks were incubated at 25°C for 311 hours with 155 rpm rotation. There were 8 sampling points, which were at every 3 hours for 18 hours and a final sample at 311 hours.

5.2.3.3 Estrogen sample preparation

In order to prepare the samples for uHPLC analysis, the samples were immediately centrifuged at 13000 x g for 3 min and 1 mL of the supernatant was filtered through 0.2 µm pore size 13mm EMD Millipore Durapore^{TM} PVDF membrane syringe filters

into a 2 mL wide screw neck 12x32 mm Chromacol™ amber glass vial with Chromacol™ Blue screw cap, open top with 6 mm hole and a PTFE septum for HPLC analysis. Furthermore standards of known concentrations (1, 2.5, 5, 7.5 and 10 mg/L) were produced in MSM and were processed in the same way as the samples to calculate the concentration of estrogen in the samples (Appendix Fig. 0-10-15).

5.2.3.4 uHPLC analysis of estrogen concentration

Quantification of estrogen concentration was carried out using uHPLC (ThermoScientific Dionex Ultimate 3000 RS autosampler, pump, column compartment and electrochemical detector) using a ThermoScientific Accucore C18 100 x 2.1 mm, 2.6µm particle size column with the settings of; flow rate = 0.4 mL/min, temperature = 40°C, injection volume = 50 – 90 µL, electrochemical potential = 1800 mV. The run was isocratic using 62% of mobile phase A and 38% of mobile phase B for 4- 6.5 minutes. Mobile phase A contained 95% water, 5% acetonitrile and 0.1% formic acid. Mobile phase B contained 95% acetonitrile and 5% water with 0.1% formic acid. The detection limit of estrogen was 100ng/L.

5.2.4 Statistical testing and graphing software

Statistical analysis was carried out using Minitab 17:

The data was determined to be not normal using the Anderson Darling normality test.

A Kruskal-Wallis test was carried out to determine if the difference in optical density and CFU/mL were significantly different compared to each other and to the control.

A Pearson's correlation was carried out to determine if there is a significant correlation between time and the concentration of estrogen. Moreover, correlation between the concentration of E1 and E2, as the conversion of E2 to E1 was identified as a potential pathway step within estrogen degradation (Chapter 2).

All graphs were produced using SigmaPlot version 12.5.

5.3 Results and discussion

5.3.1 *The 45.5 hour degradation experiment*

5.3.1.1 Bacterial growth

The average OD at 16 hours for the control flasks is slightly higher at 0.791±0.006 than for the mixed estrogen flask and the E2 containing flask of 0.061±0.012 and 0.520±0.015 respectively (Fig. 5-1). The average OD of the control flasks remains higher until 45.5 hours where the average OD of the E2 containing flasks is higher at 0.528±0.007 compared to the control and mixed estrogen flasks of 0.445±0.006 and 0.424±0.004 respectively. The number of colony forming units per mL (CFU/mL) of the control growth was also greater at $6.5x10^7 \pm 4.5x10^5$ CFU/mL than for the mixed estrogen flask and the E2 containing flask of $5.1x10^7 \pm 9.6x10^5$ CFU/mL and $4.4x10^7 \pm 1.2x10^5$ CFU/mL respectively (Fig. 5- 2). Furthermore between the hours 16 to 23.5 the growth is in the stationary phase and after 23.5 hours the bacteria are entering the death phase, therefore the bacterial exponential growth phase was identified during the first 16 hours. During the whole experiment, the average OD of *R. equi* ATCC13557 in the control flasks is significantly different to the average OD and CFU/mL of the mixed flasks (P=0.0173) and the E2 flasks (P=0.0173), however there is no significant difference (P=0.0639) in average OD or CFU/ml within the mixed estrogen flasks compared to the average OD and CFU/ml of the E2 flasks. Furthermore, there was no bacterial growth within the abiotic control.

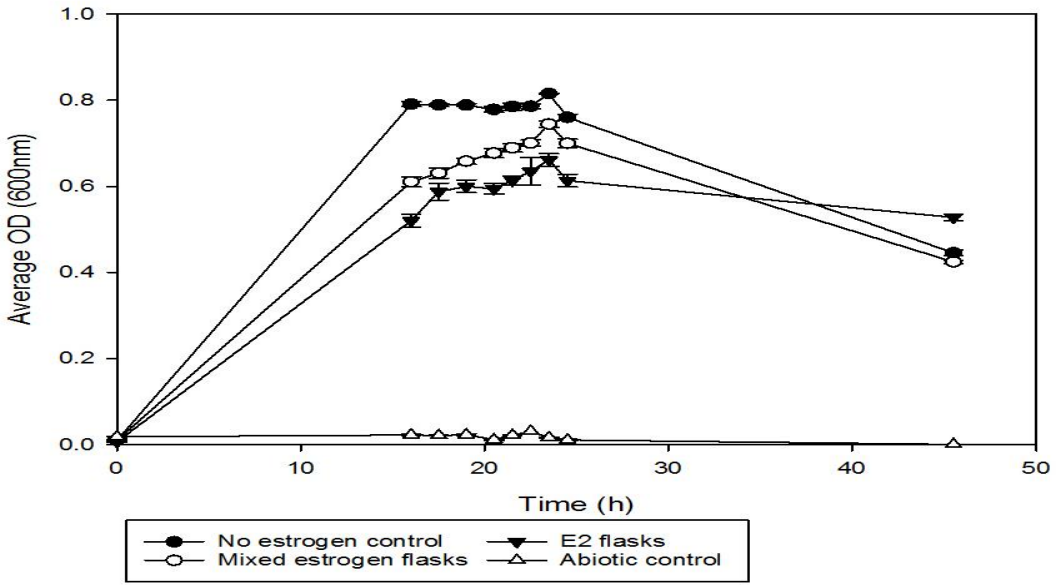

Figure 5-1: The average OD of R. equi ATCC13557 whilst being grown in the different conditions, without estrogen (control), exposed to mixed estrogens E1, E2 and EE2 (Mix), exposed to E2 (E2) and 0.1% formalin with E1, E2 and EE2 (abiotic). The error bars represent standard deviation.

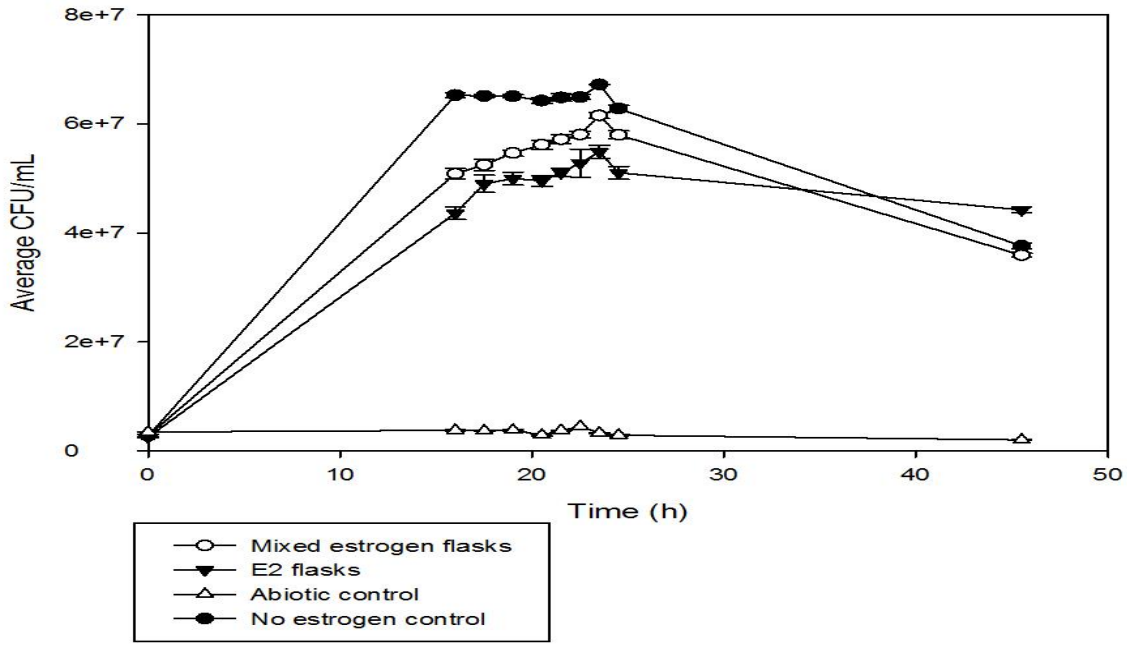

Figure 5-2: The average CFU/mL of R. equi ATCC13557 whilst being grown in the different conditions, without estrogen (control), exposed to mixed estrogens E1, E2 and EE2 (Mix), exposed to E2 (E2) and 0.1% formalin with E1, E2 and EE2 (abiotic). The error bars represent standard deviation.

5.3.1.2 Estrogen concentrations

In the mixed estrogen flask there was no change detected in the average concentration of EE2 which remained relatively stable. Still, due to slight fluctuation between 4.15±0.02 and 4.58±0.04 mg/L, there is a significant positive correlation between concentration and time (P=0.005) (Fig. 5- 3). However there was a significant decrease (P= 0.001) in the average concentration of E2 from the initial 4.10±0.05 mg/L, to an average concentration of 0.88±0.01 mg/L at 24.5 hours, and not detected by 45.5 hours (Fig. 5- 3). Contrastingly there was a significant increase (P=0.01) in the average concentration of E1 from the initial starting concentration of 4.33±0.11 mg/L to 8.16±0.11 mg/L at 24.5 hours, but it decreased slightly at 45.5 hours to 7.97±0.10 mg/L (Fig. 5- 3). Furthermore the average decrease in E2 concentration correlated significantly (P=0.000) with the average increase in E1 concentration.

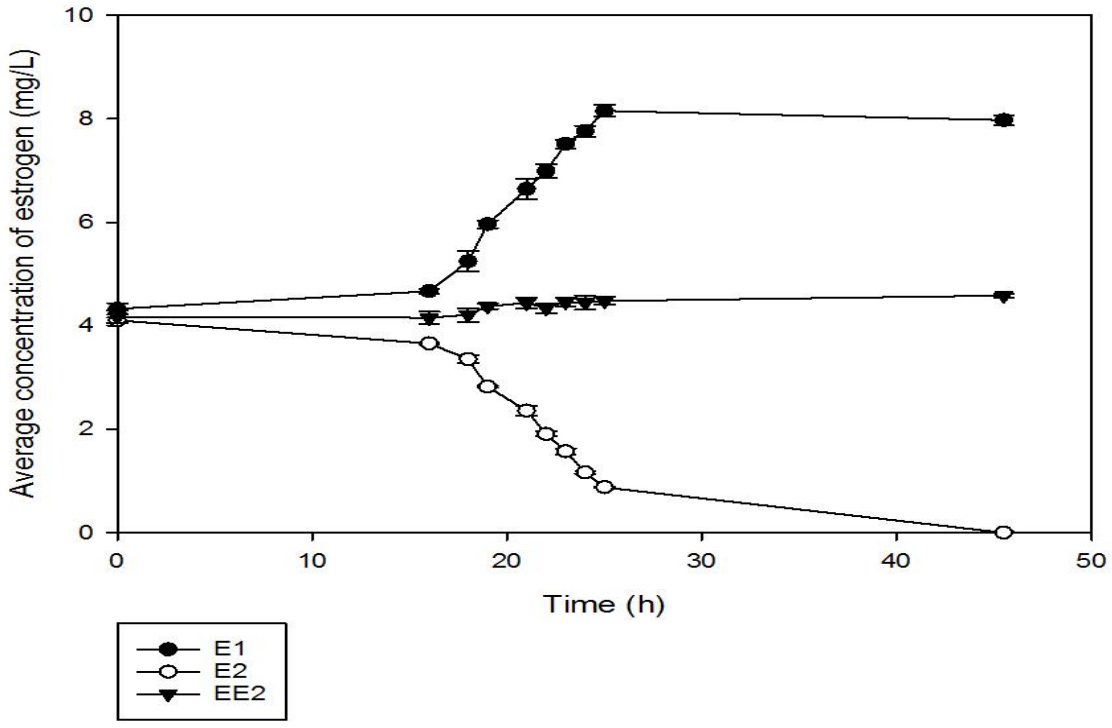

Figure 5-3: Measuring the average concentrations of mixed estrogens E1, E2 and EE2 over 45.5 hours to measure the degradation of estrogen by R. equi ATCC13557. The error bars represent standard deviation.

In the E2 containing flask the average concentration of E2 significantly decreased (P=0.012) from an initial concentration of 6.57±0.09 mg/L to 0.94±0.04 mg/L at 23.5 hours and then remained relatively unchanged with a slight decrease in the average concentration of 0.86±0.06 mg/L at 45.5 hours (Fig. 5- 4). Furthermore there was a significant increase (P=0.003) in the average concentration of E1 with none detectable initially to 3.10±0.09 mg/L by 45.5 hours (Fig. 5- 4). Again the decrease in the average concentration of E2 significantly correlated (P= 0.000) with an increase in the average concentration of E1.

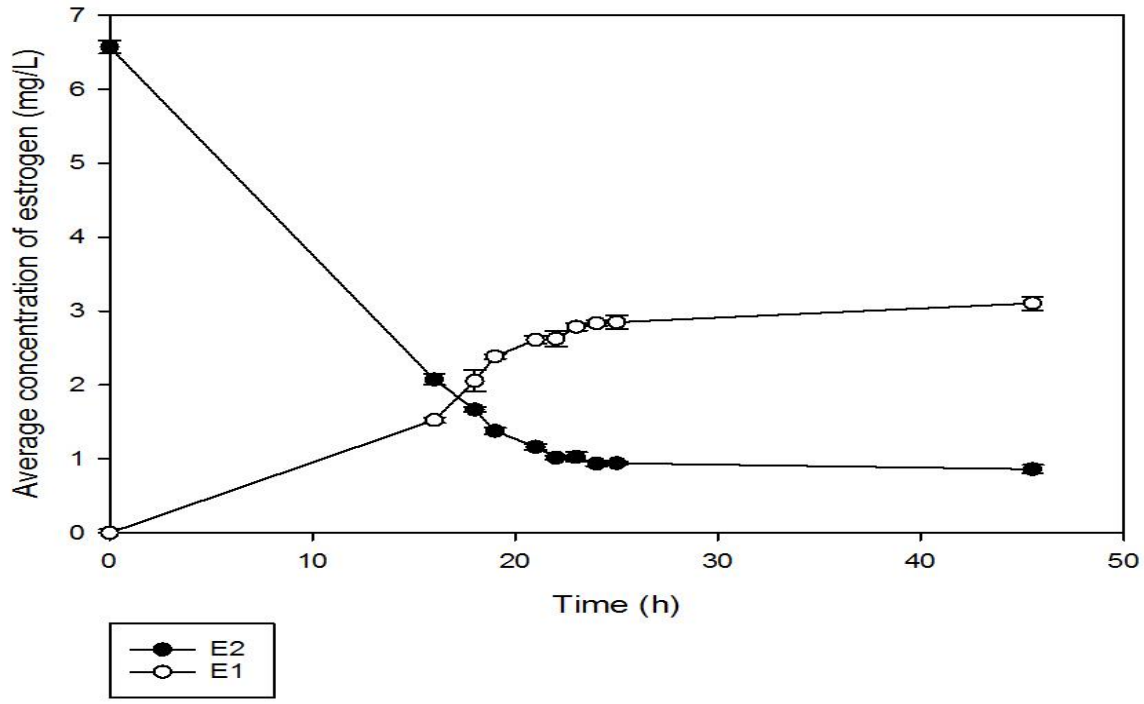

Figure 5-4: Measuring the average concentration of E2 and E1 over 45.5 hours to measure the degradation of estrogen by R. equi ATCC13557. The error bars represent standard deviation.

In the abiotic control containing E1, E2 and EE2 there was no significant variation (P=0.287, 0.307, 0.207) in the average concentrations, with E1 being 5.19±0.09 mg/L at the lowest and 5.99±0.09 mg/L at the highest, an increase of 0.80 mg/L (Fig. 5- 5). The lowest average concentration of E2 was 4.93±0.04 mg/L and 5.56 ±0.12 mg/L at the highest, an increase of 0.63 mg/L (Fig. 5- 5). Furthermore the lowest average concentration of EE2 was 4.99±0.08 mg/L and 5.56±0.06 mg/L at the highest, an increase of 0.57 mg/L (Fig. 5- 5).

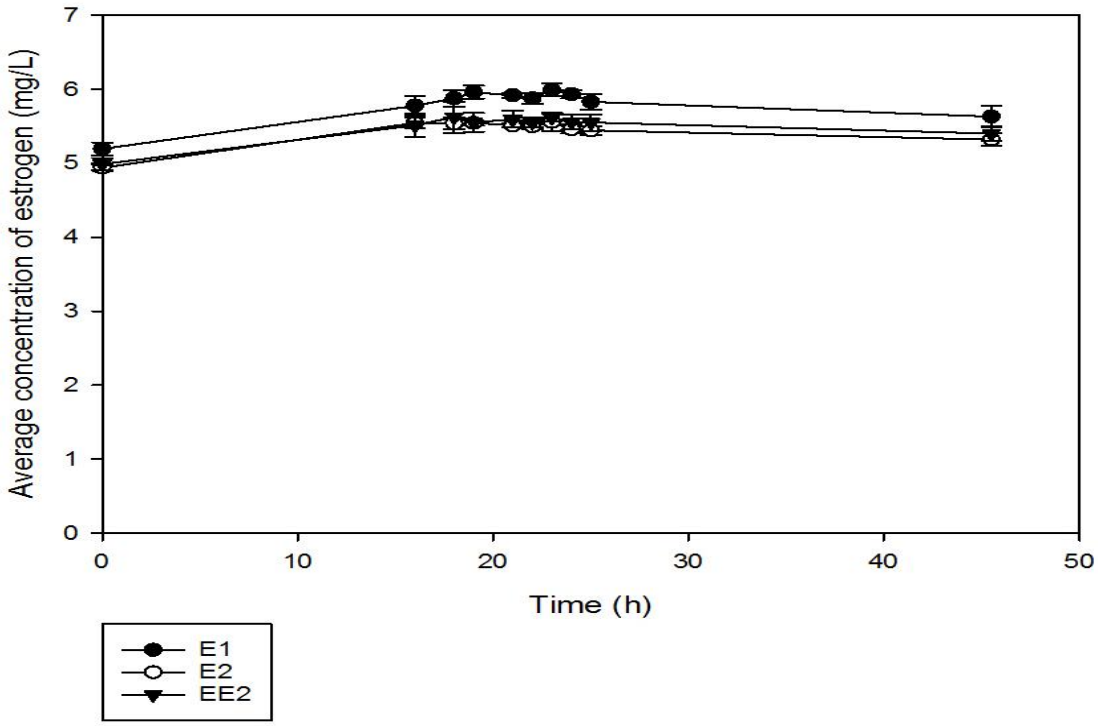

Figure 5-5: Measuring the average concentration of EE2, E2 and E1 over 45.5 hours of the 0.1% formalin containing abiotic control. The error bars represent standard deviation.

5.3.1.3 Metabolite analysis

In the E2 flasks an unknown metabolite peak was present at 16 hours at retention time of 1.425 min and a peak area of 50.196nA*min (Fig. 5- 6), which appeared to decrease in area at each sampling point for the duration of the 45.5 hours (Fig. 5- 7).

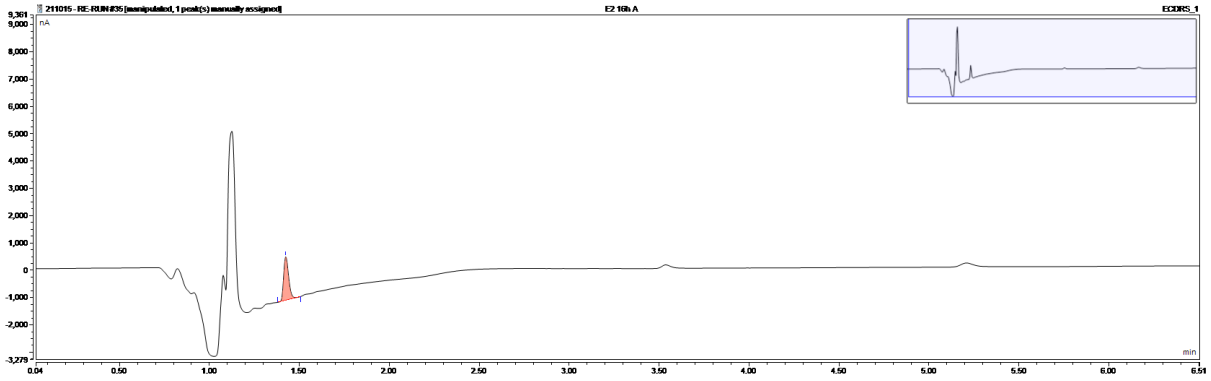

*Figure 5-6: Unknown metabolite peak (red) present at 16 hours at retention time 1.425 min and with a peak area of 50.196nA*min.*

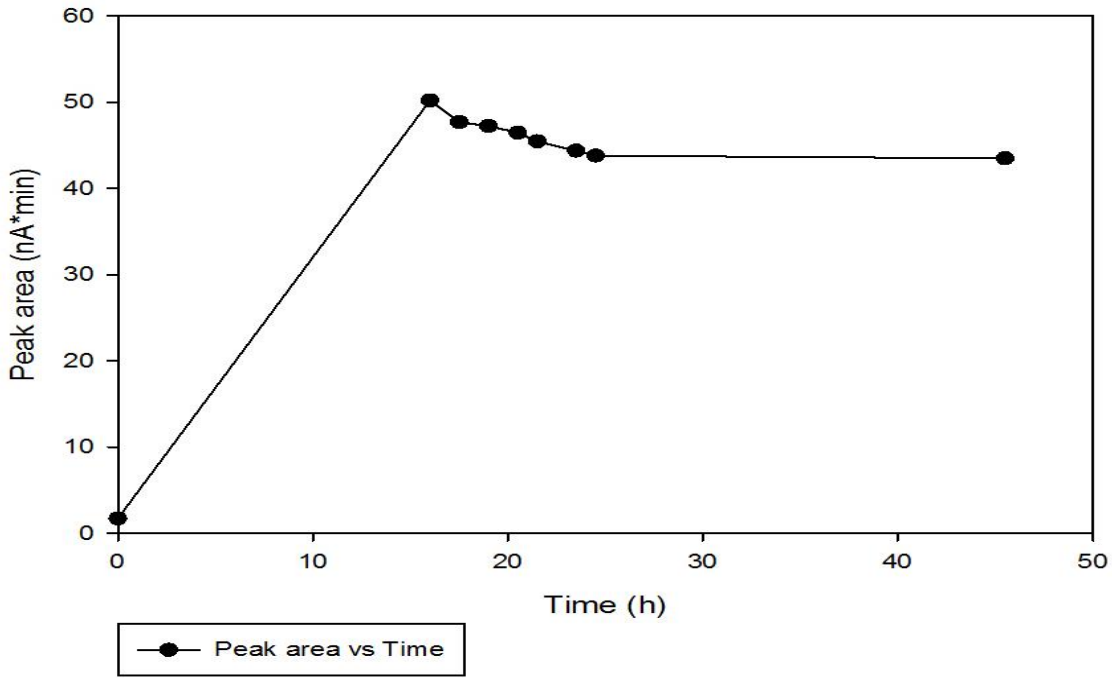

Figure 5-7: Peak area of an unknown metabolite at retention time 1.425 min in the E2 only flasks over 45.5 hours.

5.3.2 *The 311 hour degradation experiment*

5.3.2.1 Bacterial growth

The average OD between all the experimental conditions remained similar until 12 hours. At that point, the average OD of the flasks containing E2 was significantly higher (P=0.000) than the other conditions at 0.054±0.004 and remained significantly higher (P=0.000) at 15 hours at 0.085±0.004 (Fig. 5- 8). At 18 hours the OD of the control was significantly higher (P=0.000) at 0.108±0.008, but at 311 hours the average OD of the E2 containing flasks was significantly higher (P=0.000) at 0.156±0.016 (Fig. 5- 8). Furthermore there is no growth of the abiotic control (Fig. 5- 8).

When comparing the triplicate flasks for each condition, the average OD is significantly different at 12 hours (P=0.016). Flasks A and B at 12 hours, 0.03±0.001, have a higher average OD than flask C of 0.02±0.001, but the average OD of the control is the highest at 0.043±0.002 in the mixed estrogen flasks (Fig. 5- 9). The average OD of flask C increases between 15 and 18 hours, therefore the flasks A, B and C have similar OD and there is no significant difference (P=0.060) at 18 hours;

0.064 ± 0.002, 0.062 ± 0.001 and 0.047 ± 0.02, respectively. Flask C increases in OD between 18 and 311 hours, the average OD; 0.022, is significantly higher (P=0.023), than flasks A and B (Fig. 5- 9).

The average OD of flask C in the EE2 condition is significantly higher (P=0.024) than flasks A and B at 12-311 hours (Fig. 5- 10). However, the average OD of the control is significantly higher (P=0.020, 0.016) than all the flasks containing EE2 at 15 and 18 hours, respectively. Again, at 311 hours the EE2 containing flask C has a significantly higher (P=0.032) average OD, by 0.02 (Fig. 5- 10).

The average OD of the flasks containing E2 remain similar until between 18-311 hours where the average OD of flasks A, B and C; 0.155 ± 0.001, 0.138 ±0.003, and 0.175 ± 0.005, are significantly higher (P=0.016) than average OD of the control, 0.129 ±0.006 (Fig. 5- 11).

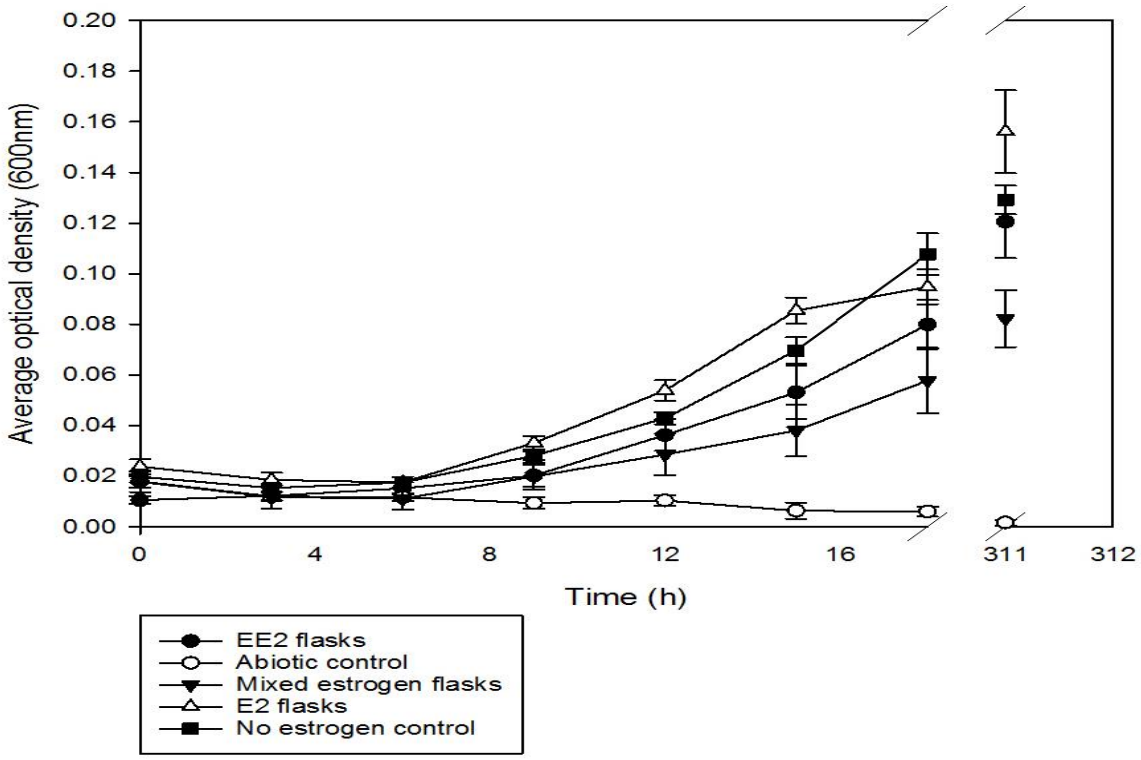

Figure 5-8: The average OD of R. equi ATCC13557 whilst being grown in the different conditions, without estrogen (control), exposed to mixed estrogens E2 and EE2 (Mix), exposed to E2 (E2), exposed to EE2 (EE2) and 0.1% formalin with E2 and EE2 (abiotic). The error bars represent standard deviation.

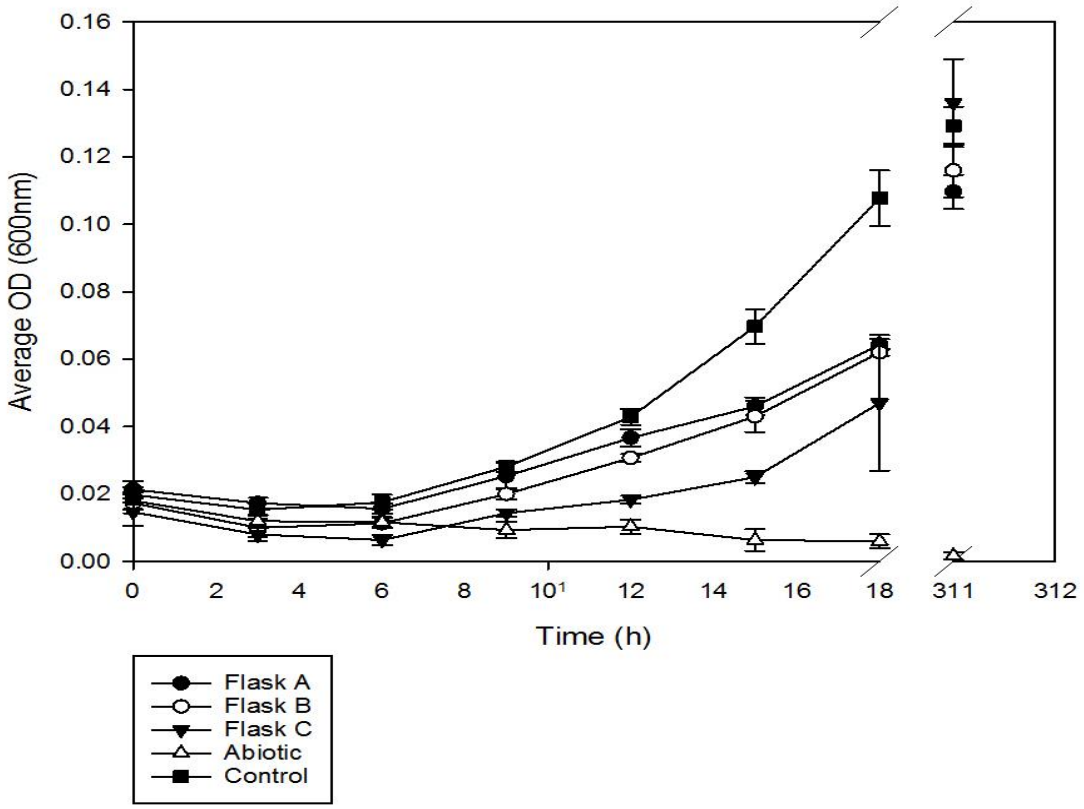

Figure 5-9: The average OD of R.equi ATCC13557 whilst being grown in mixed estrogen conditions (E2 and EE2), compared to the without estrogen (control), and 0.1% formalin with E2 and EE2 (abiotic)

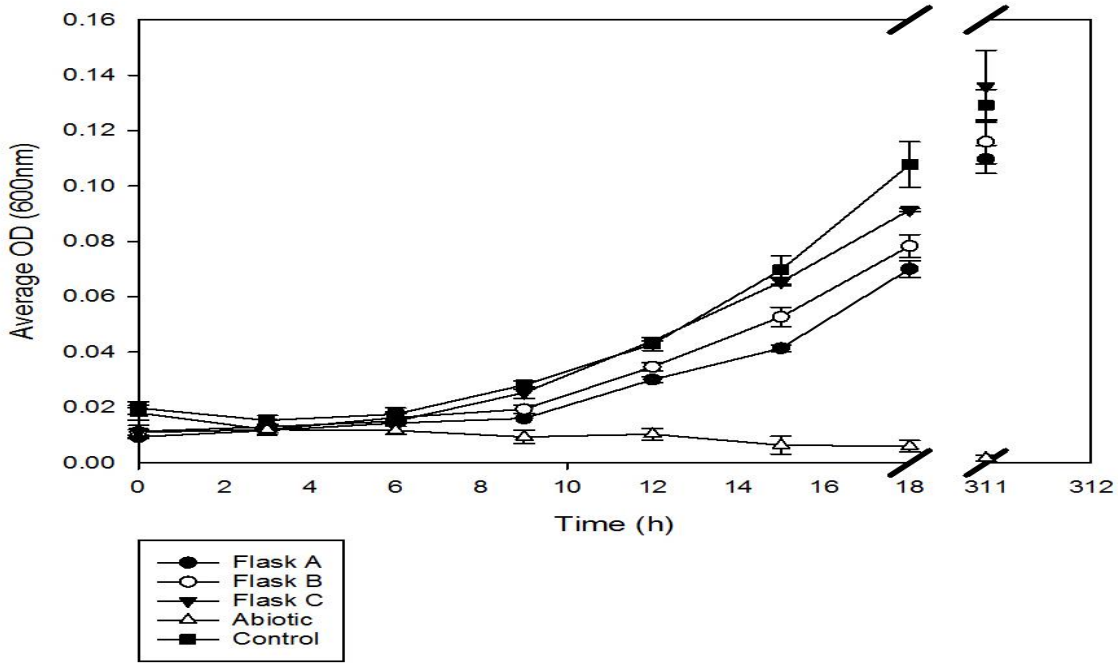

Figure 5-10: The average OD of R.equi ATCC13557 whilst being grown in EE2, compared to the without estrogen (control), and 0.1% formalin with E2 and EE2 (abiotic)

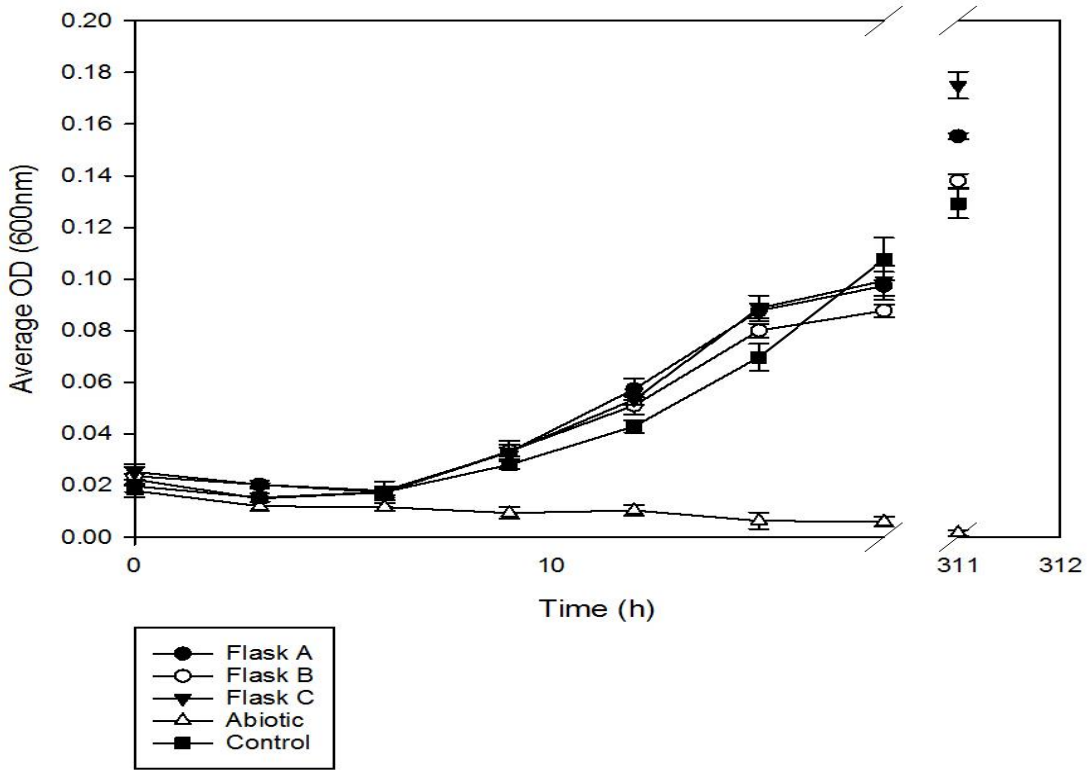

Figure 5-11: The average OD of R.equi ATCC13557 whilst being grown in E2, compared to the without estrogen (control), and 0.1% formalin with E2 and EE2 (abiotic)

5.3.2.2 Estrogen concentration

In all of the three mixed (E2 and EE2) estrogen flasks, there was a significant decrease (P=0.028, 0.046, 0.001) in the average concentration of E2 over time and with none detectable by 18 hours in flask C (Fig. 5- 12, 13, 14). At 18 hours the average concentration of E2 had decreased to 1.6±0.02 mg/L and to 1.5±0.03 mg/L in flask A and B respectively (Fig. 5- 12, 13). However a further set of samples was taken at 311 hours in which and there was no detectable E2 in flasks A and B. Therefore, there was an overall decrease of 5.4±0.02 mg/L in E2 in flask A, 6.6±0.16 mg/L in flask B and 5.7±0.08 mg/L in flask C. However the average concentration of EE2 remained relatively stable in all of the flasks with no significant correlation (P=0.384, 0.828, 0.558) over time. Furthermore there was a significant positive correlation (P=0.015, 0.039, 0.001) in the average concentration of E1 with time. E1 was detectable in all of the flasks at 6 hours and the increase significantly correlated (P=0.000) with the average concentration of E2 decreasing in all biological replicates.

The concentration of E1 increased to 4.8±0.02 mg/L, 5.7±0.06, and 6.81±0.05 mg/L in flasks A, B and C respectively by 18 hours. Furthermore the average concentration of E1 continued to increase in flasks A and B to 7.7±0.15 mg/L at 311 hours, but the average concentration of E1 in flask C had decreased from 18 hours to 6.2±0.02 mg/L at 311 hours.

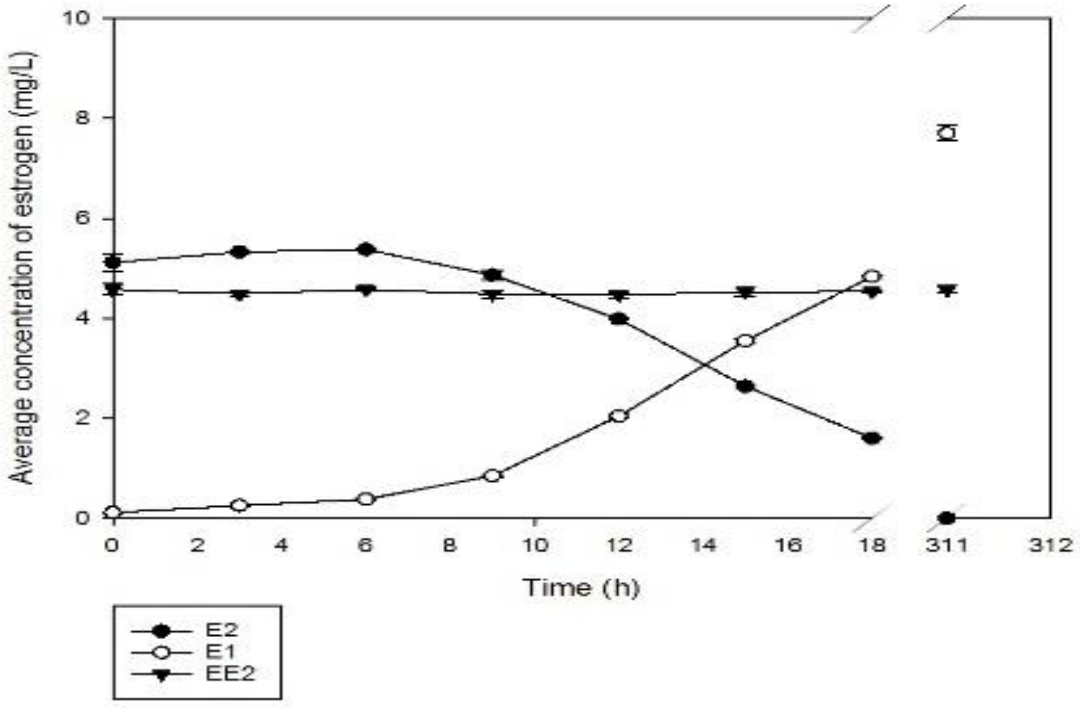

Figure 5-12: The average concentration of estrogen measured over time in mixed estrogen flask A containing E2 and EE2 initially.

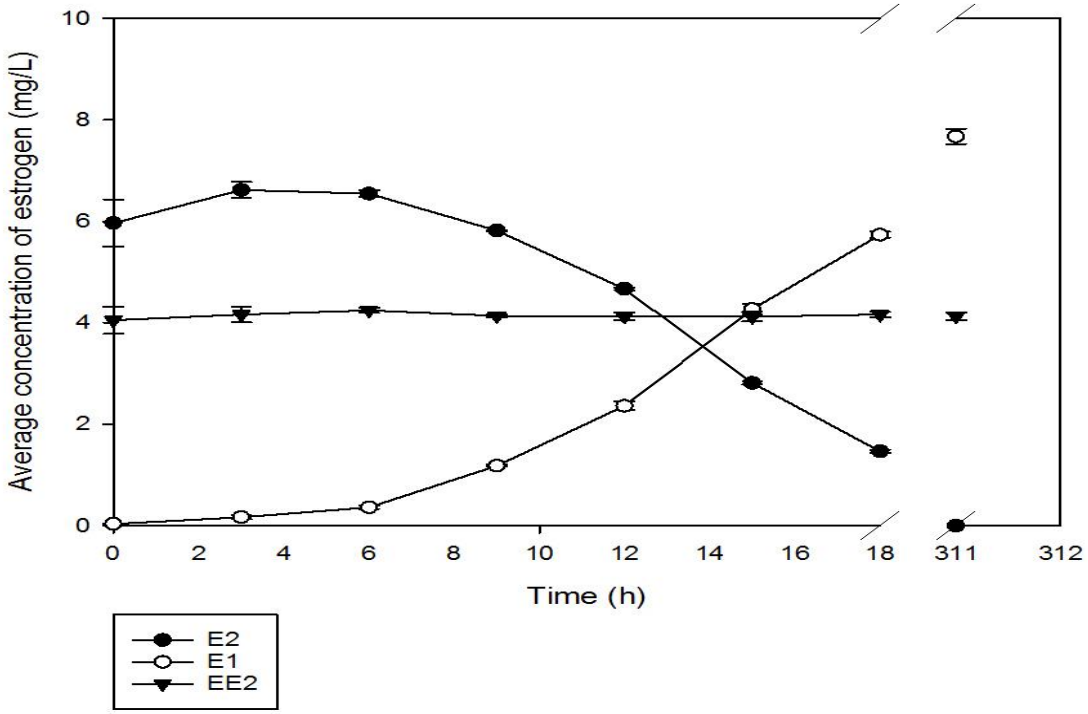

Figure 5-13: The average concentration of estrogen measured over time in mixed estrogen flask B containing E2 and EE2 initially.

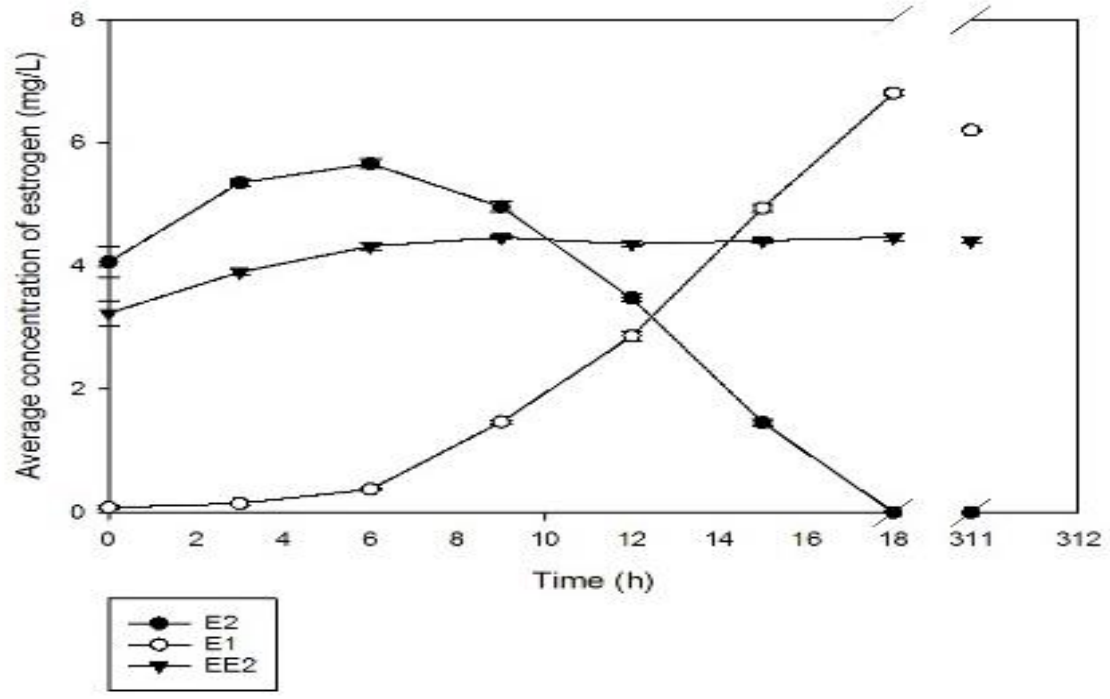

Figure 5-14: The average concentration of estrogen measured over time in mixed estrogen flask C containing E2 and EE2 initially.

In the three flasks containing only E2, during the first 18 hours there was no decrease in the average concentration of E2 or any detectable E1 (Fig. 5- 15). However at 311 hours there was no detectable E2 in flasks A, B or C and there was a significant increase (0.000) in the average concentration of E1 of 3.2±0.09 mg/L, 2.2±0.04 mg/L and 3.1±0.02 mg/L in flasks A, B and C respectively. There was a significant decrease (P=0.000, 0.000, 0.002) in E2 of 4.2±0.01 mg/L, 6.2±0.02 and 1.6±0.01 mg/L in flasks A, B and C respectively by 311 hours. The increase in the average concentration of E1 significantly correlated (P=0.000, 0.000, 0.001) with the decrease in the average concentration of E2 in flasks A, B and C.

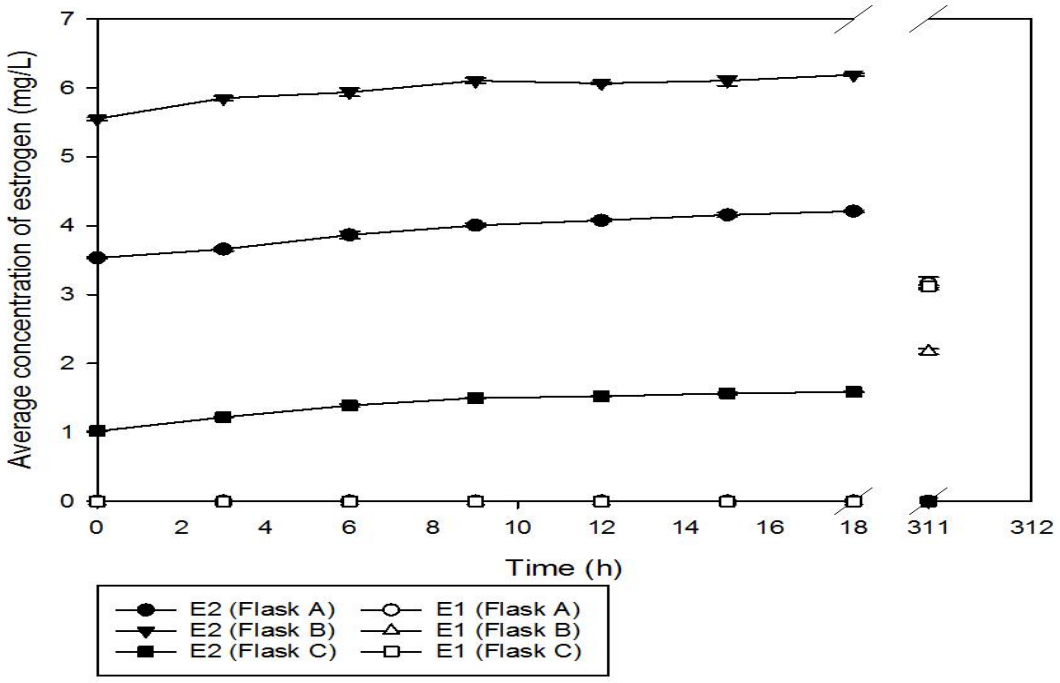

Figure 5-15: The average concentration of estrogen measured over time in flasks A B and C containing only E2 initially.

The average concentration of EE2 in flask A showed a gradual increase in concentration over time which reached 4.5±0.09 mg/L by 18 hours (Fig. 5- 16). In flask B the average concentration of EE2 increased at a faster rate than in flask A, reaching 4.5±0.03 mg/L by 15 hours and remaining the same at 18 hours (Fig. 5-16). However the average concentration of EE2 in flask C remained relatively unchanged for the 18 hours and remained at around 2.3 mg/L between hours 9 until

18 (Fig. 5- 16). At 311 hours the average concentration in flask A decreased slightly to 4.4±0.03 mg/L, in flask B decreased to 2.7±0.03 mg/L, and in flask C decreased slightly to 2.0±0.005 mg/L. The decrease in average concentrations of EE2 in flasks B and C correlated significantly (P=0.001, 0.002) with time between 9 and 311 hours, however there was no significant positive or negative correlation (P=0.444) with time for flask A.

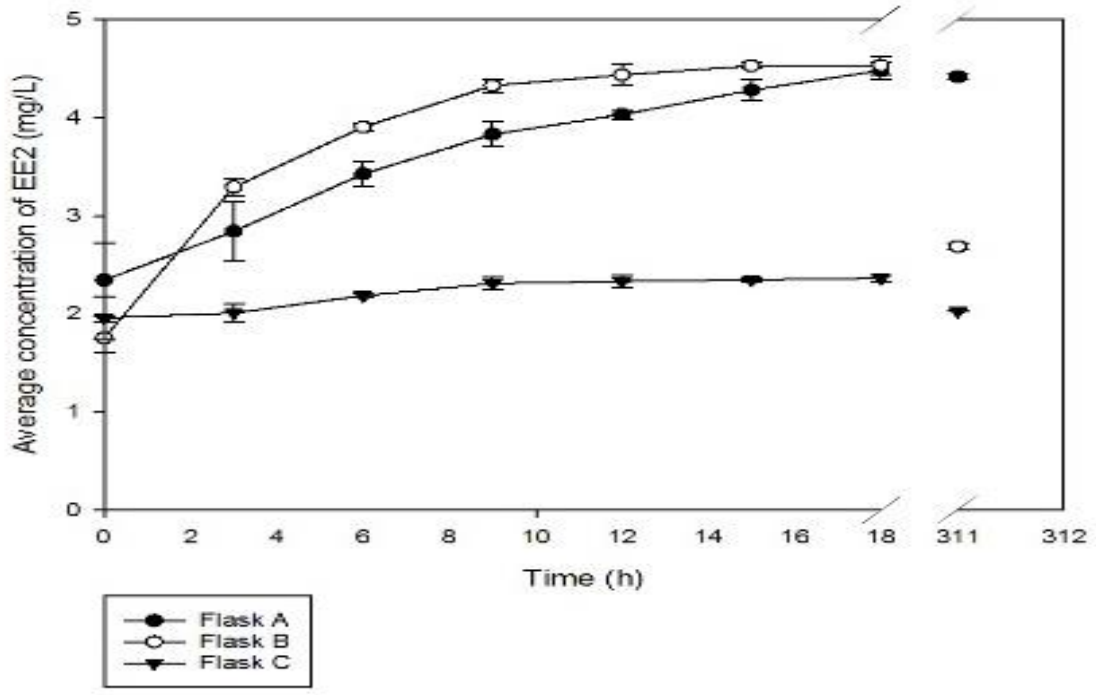

Figure 5-16: The average concentration of estrogen measured over time in flasks A B and C containing only EE2 initially.

In the abiotic control the average concentrations for E2 and EE2 increased between 0 and 3 hours by 1.7 mg/L and 0.4 mg/L respectively, but remained unchanged until 18 hours (Fig. 5- 17). However at 311 hours there was a decrease in the average concentration of E2 and EE2 of 1.3 mg/L and 0.7 mg/L respectively. The decrease in average concentration of EE2 significantly correlated (P=0.027) with time, however the decrease in E2 did not significantly correlate (P=0.270) with time. The increase in the concentrations of E2 and EE2 initially are due to solubility issues, caused by the estrogens precipitating out of the solution after being prepared the previous day.

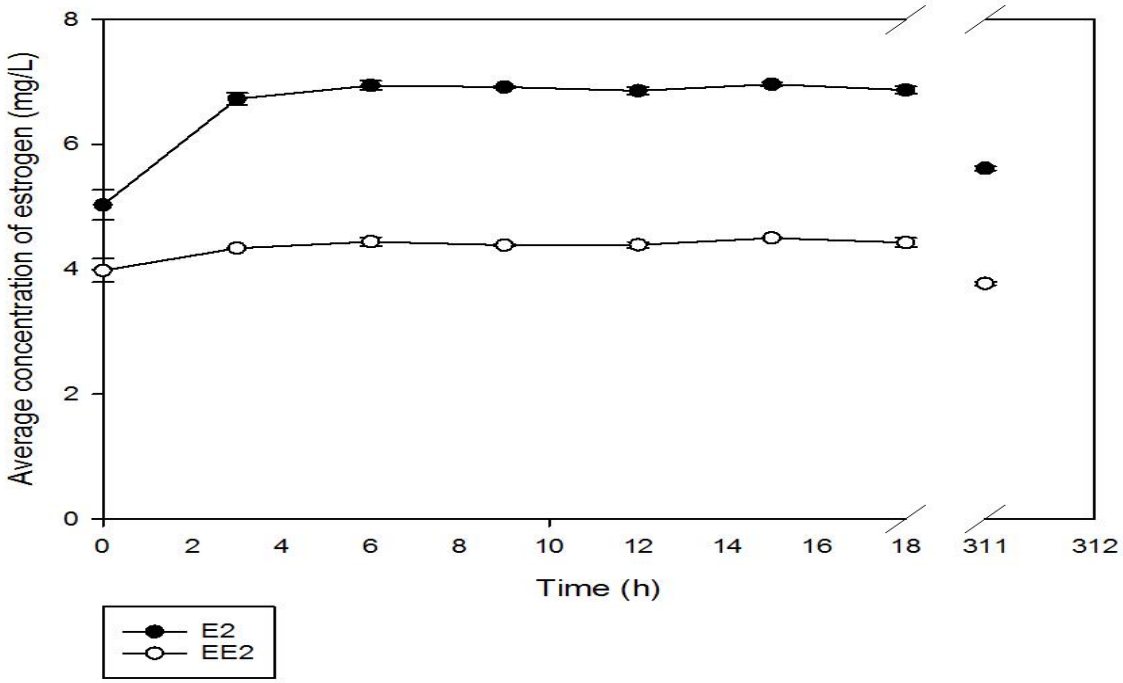

Figure 5-17: The average concentration of estrogen measured over time in the abiotic control treated with 0.1% Formalin solution and containing E2 and EE2 initially.

In conclusion, in the 45.5 hour experiment there was no significant change in the average concentrations of estrogen in the abiotic control for the duration of the 45.5 hours so therefore the changes in the average concentrations in the E2 only and mixed estrogen flasks can be attributed to biotic factors. However in the 311 hour experiment, there is a decrease in E2 and EE2 at 311 hours, therefore any changes in the concentrations of estrogen between 0 and 18 hours are likely due to biotic factors but any change in the EE2 concentration after 18 hours may have been due to abiotic factors.

The time taken for E2 to be completely metabolised was shorter in the presence of other estrogens, all E2 was degraded between 18-311 h in the presence of other estrogens, however, in the E2 only flasks there was still E2 detectable at 45.5 h but were not detected by 311 h. Furthermore in mixed estrogen the conversion from E2 to E1 was complete, however in the E2 only flasks the amount of E1 detected was less than the amount of E2 degraded, therefore there is some complete degradation when E2 is the sole estrogen. There was a significant difference in the growth of *R. equi* ATCC13557 in the presence of E2 compared to the control, suggesting that the

metabolism of E2 and the degradation of E2 corresponds to an increase in growth, therefore the degradation of E2 is growth-linked when E2 is the sole estrogen. The degradation of E2 in the mixed estrogen flasks of the 45.5h experiment began after the exponential growth phase and during the stationary phase and therefore the degradation of E2 within mixed estrogen is not growth-linked. The growth in the 311 hour experiment of the control is, at the majority of time points, higher than that of the experimental flasks, except for the increase in average OD within flask C containing mixed estrogen which corresponded with the degradation of E2 which is completely metabolised by 18 hours. The faster complete degradation of E2 in the presence of other estrogens and non growth-linked degradation suggest that E2 may be co-metabolised in the presence of other estrogens (Dalton and Stirling, 1982). Co-metabolism has been reported previously in numerous studies (Pauwels et al., 2008; Shi et al., 2004). This is possibly beneficial as with co-metabolism it is not necessary to meet the threshold of energy and carbon required for aerobic and anaerobic biodegradation, which may reduce the costs associated with estrogen removal (Hazen, 2010). However, the decrease in E2 correlated significantly with the increase in E1, meaning E2 was converted into E1. Increased concentration of E1 within the environment may lead to undesired effects in aquatic organisms (Ankley *et al.,* 2017).

There was no evidence of EE2 degradation by *Rhodococcus equi* ATCC13557, and although the average concentration of EE2 in the EE2 only flasks A and B appeared to increase over time, this is likely due to the solubility issues encountered during the method development in which the addition of EE2 in an AcN:MeOH stock solution appeared to be 3.32±0.07 mg/L lower than the expected concentration of 5 mg/L (Fig. 5- 8). There was a decreased in the average concentration of EE2 in flasks A, B and C of 0.1 mg/L, 1.8 mg/L and 0.3 mg/L respectively at between 18 and 311 hours but the decrease was only significant for flasks B and C. However due to a significant reduction of E2 and EE2 in the abiotic control this decrease may have been due to abiotic factors.

In some of the flasks the average increase in E1 is greater than the average decrease in E2. The average concentrations of E1 in flasks A, B and C at 311 hours, and flask C at 18 hours, were outside the range of the standards and therefore are unreliable (Appendix Fig. 0- 13). Thus, it is possible that all the E2 is being converted

to E1, but also possible that there may be a lower concentration of E1 than being measured as the concentration at 18 hours is not accurate, therefore there is a possibility of other metabolites. The elevated concentration of E1 may be due to the decrease in flask volume over time due to the number of samples being taken, or an issue of separating the peaks during the analysis of the HPLC trace where E2 and E1 are overlapping. However, there is no change in the average concentration of EE2 in any of the flasks.

The hypothesis was correct in that the concentration of E2 decreased over time within all of the conditions. However, the conversion of E2 to E1 and absence of EE2 degradation in *R. equi* ATCC13557 is different than what was reported previously, but the presence of 0.5% (v/v) ethanol in the other study may have increased estrogen degradation via co-metabolism (Larcher and Yargeu, 2013). Other reasons for the difference in degradation may be due to pH, inoculum concentration, and temperature which were reported to have effects upon the degradation of E2 by *Rhodococcus* sp. DS201 (Yu *et al.,* 2016). However these findings were similar to the conversion of E2 to E1 by *Rhodococcus* sp. P14 (Ye *et al.,* 2017). Furthermore the additional peak with a retention time similar to that of E3, has been reported previously however this was under anaerobic conditions using lake sediment and sludge (Czajka and Londry, 2006). The conversion of E1 and E2 to E3 has been reported in Silt Loam Soil (Xuan *et al.,* 2008).

Furthermore, the degradation studies combined with the hypothetical degradation pathway (Fig. 2-1; Table 2-1) and the gene cluster within the genome of *R. equi* ATCC13557 (Fig. 3- 6), can be used to hypothesize which genes will be expressed in *R.equi* ATCC13557 when exposed to estrogen. The conversion of E2 to E1 is present in all of the flasks, with complete conversion occurring in the mixed estrogen flasks, which suggests that the gene coding for 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase will be expressed in the presence of estrogen as shown in reaction 7 of the hypothetical estrogen degradation pathway (Fig. 2- 1; Table 2- 1; Fig. 3- 6). However, in the degradation studies with E2 as the sole estrogen, not all the E2 is converted to E1, there is evidence of an unknown metabolite (Fig. 5- 6; Fig. 5- 7), and there is no decrease in the concentration of E1, suggesting that the genes coding for the enzymes catechol 2,3 dioxygenase or putative cytochrome P450 monooxygenase in reaction 6 and 3-(3-hydroxy-phenyl)proprionate hydroxylase in

reactions 4 and 5 (Fig. 2- 1; Table 2- 1; Fig. 3- 6) may be possible. Additionally, the decrease in the average peak area of the unknown metabolite (Fig. 5- 6; Fig. 5- 7) may suggest a decrease in its concentration and therefore that it is further degraded. If the unknown metabolite is 4-OH-E2 and it is converted to 4-OH-E1, then there may be expression of the gene coding for 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase after 16 hours, in addition to the initial expression for the conversion of E2 to E1. As neither E1 nor EE2 were degraded by *R.equi* ATCC13557 the genes coding for 2-hydroxymuconate-semialdehyde hydrolase shown in reactions 12 of the hypothetical estrogen degradation pathway (Fig. 2- 1; Table 2- 1; Fig. 3- 6), is unlikely to be expressed in the presence of estrogen. However, if the unknown metabolite was 4-OH-E2 and it was converted to 4-OH-E1 then the expression of the gene coding for 2-hydroxymuconate-semialdehyde hydrolase in reaction 12 of the hypothetical estrogen degradation pathway may be possible (Fig. 2- 1; Table 2- 1; Fig. 3- 6). The identity of the unknown metabolite cannot be confirmed without further analysis and therefore the concentration cannot be determined from the peak area to determine if there is a significant decrease in concentration over time.

As the conversion of E2 to E1 is observed early within the mixed estrogen flasks it is expected that the expression the gene encoding 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase will be early and will decrease as the concentration of E2 decreases. However, in the E2 only flasks the expression of 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase is likely to be much later, after 18 hours, due to the delayed degradation. Furthermore, in the E2 flasks where the conversion of E2 to E1 was incomplete, the genes encoding catechol 2,3 dioxygenase, putative cytochrome P450 monooxygenase and 3-(3-hydroxy-phenyl)proprionate hydroxylase are expected to be expressed early. If expressed the gene coding for 2-hydroxymuconate-semialdehyde hydrolase is expected to be expressed much later in the degradation.

Chapter 6: Gene expression analysis

6.1 Introduction

There have been few studies into the genes coding for enzymes involved in estrogen degradation with the focus being mainly upon the metabolites, in a previous study using a naphthalene oxidase assay, non-specific monooxygenase activity was suggested to be involved in E2 degradation by *Sphingomonas* sp. KC8 (Yu *et al.,* 2007). Using RT-qPCR, the 17β-hydroxysteriod dehydrogenase enzyme was suggested in the degradation of E2 by *Rhodococcus* sp. P14 (Ye *et al.,* 2017). In order to determine if the genes identified in the genome of *R. equi* ATCC13557 (Chapter 3), coding for 2-hydroxymuconate-semialdehyde hydrolase, putative cytochrome P450 monoxygenase, 3-oxosteriod 1-dehydrogenase, IcIR family transcriptional regulator, catechol 2,3 dioxygenase, 3-(3-hydroxy-phenyl)proprionate hydroxylase and 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase, are involved in estrogen degradation, it is necessary to measure their expression over time using RT-qPCR when exposed to estrogen and compare this to expression in the absence of estrogen. The hypothesis is that there will be an increased expression of the genes, coding for enzymes involved in estrogen degradation, identified in *R. equi* ATCC13557 when exposed to estrogen, and that their expression will vary over the duration of the degradation study. Following the degradation study, gene expression of the genes coding for the enzymes identified in the genome of *R. equi* ATCC13557 can be hypothesized for mixed estrogen conditions (Table 6- 1).

Table 6-1: The hypothesized expression of the genes coding for the enzymes identified within the genome of R. equi ATCC13557

Enzyme	Reaction number(s) (Fig. 2-1 & Table 2-1)	Hypothesis
2-hydroxymuconate-semialdehyde hydrolase	12	Not expected to be expressed as there was no evidence of E1 degradation.
Putative cytochrome P450 monooxygenase	2	May be expressed in addition to or in place of dioxygenase to form 4-OH-E2 or 4-OH-E1, which were undetectable in the degradation studies. Non specific monooxygenase activity has been previously reported (Yu *et al.,* 2007). However, there is no degradation of EE2.
3-oxosteriod 1-dehydrogenase	N/A	Although the function within the pathway is uncertain there has been a previous study that has suggested expression in the presence of E2 (Sang *et al.,* 2011). Within the context of the hypothesised pathway, not expected to be expressed.
IcIR family transcriptional regulation	N/A	Transcriptional regulator so expected to be expressed where the other genes are expressed.
Catechol 2,3 dioxygenase	6	Expression unlikely as all E2 was converted to E1, but may have formed undetectable metabolites such as 4-OH-E2 or 4-OH-E1.
3-(3-hydroxy-phenyl)proprionate hydroxylase	4 & 5	Expression unlikely as all E2 was converted to E1, however may be possible if forming transient metabolite.
3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase	7	Expression expected early due to conversion of E2 to E1 beginning at 0-3 hours.

RT-qPCR is the most effective molecular technique in which the gene expression (RNA) of specific genes is quantified relative to a control and to reference gene(s) due to its detection efficiency and precision (Heid *et al.,* 1996; Rocha *et al.,* 2016). During RT-qPCR a fluorescent dye such as SYBR green is incorporated into the double-stranded DNA (dsDNA), producing a fluorescent signal directly proportional to the number of PCR products within the sample during amplification (Rocha *et al.,* 2016). RT-qPCR requires the production of complementary DNA (cDNA) by reverse transcription of good quality RNA. The stages that follow in RT-qPCR are denaturation, where the double-stranded DNA (dsDNA) are separated to single stranded DNA (ssDNA), the temperature is reduced to allow binding of the primers, and extension where polymerase carries out replication and extension. These steps are then repeated to produce many copies of the PCR product and the fluorescence is measured in real-time (Rocha *et al.,* 2016). In order for qPCR to be successful, the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines suggest that there are a number of considerations in the production of relevant, accurate, reliable and correct interpretation of RT-QPCR data (Bustin *et al.,* 2009). These suggestions include ensuring that; 1) the RNA is of sufficient quantity and quality. The quality can be ensured by the 260/280 and 260/230 ratios between 1.8-2.1 and greater than 1.5 respectively, in addition to gel electrophoresis where the ribosomal 23S RNA band which is twice the intensity of the 16S RNA band and the absence of smearing, or by use of a bioanalyser (ThermoFisher Scientific, 2018); 2) the primers are efficient and specific; 3) there are negative reverse transcription (-RT) and no template controls (NTC) on each plate or batch of samples; 4) use of standard error bars; 5) appropriate normalisation to stable reference gene(s) (Bustin *et al.,* 2009). Furthermore the use of multiple reference genes may improve the accuracy of RT-qPCR data. Often studies use a single reference gene, without validation to ensure that under their experimental conditions that the gene is stably expressed, however reference gene expression can vary considerably (Vandesompele *et al.,* 2002). For example, Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is the most commonly used reference gene, however under different experimental conditions its expression may vary, and incorrect normalisation may lead to false interpretations of genes being upregulated or downregulated (Kozera and Rapacz, 2013). Therefore the use of geNorm to determine the most stable reference genes, genes with the least variation in gene

expression, and the minimum number of reference genes required to calculate a reliable normalisation factor are recommended (Vandesompele *et al.,* 2002). The geNorm assay consists of a number (n=6-12) of commonly used reference genes, which are from different functional classes, to reduce the likelihood that they are co-regulated (Vandesompele *et al.,* 2002). These reference genes are plated along with cDNA from all of the experimental and control conditions within the planned experiment, RT-qPCR is carried out to measure the gene expression of the reference genes. The gene-stability is then determined by the principle that the expression ratio of two ideal reference genes is the same in all samples, a lower expression ratio suggests the genes are more stable (Vandesompele *et al.,* 2002). Therefore, geNorm M is used as a stability measure in which the average pairwise variation of a particular reference gene with all the other reference genes as standard deviation of the fold change in expression (Vandesompele *et al.,* 2002). In order to select the number of reference genes to allow accurate normalisation, and to prevent the need for excessive numbers of reference genes, a normalisation factor geNorm V is calculated (Vandesompele *et al.,* 2002). The geNormV calculation compares the geometric mean of gene expression for the reference genes identified as the most stable, with the lowest geNorm M values, by an inclusion of reference genes which significantly effect the pair-wise variation (Vandesompele *et al.,* 2002).

6.2 Materials and methods

6.2.1 *Bacterial culture*

R. equi ATCC13557 was acclimated to estrogens for 24 hours, by being cultured in 500 mL Erlenmeyer flasks with vented bungs, containing MSM (See 6.2.1.1) with 0.6 g/L of yeast extract. The conditions in each of the acclimation flasks were without estrogen (control), 5 mg/L of EE2, 5 mg/L of E2, and 5 mg/L mixed estrogen of E2 and EE2. *R. equi* ATCC13557 was previously subcultured numerous times in MSM (see 6.2.1.1) containing 5 mg/L of mixed estrogen of E1, E2 and EE2, 5 mg/L of EE2, 5 mg/L of E2, and without estrogen (control).

6.2.1.1 Preparing the growth media

M9, Minimal Salts, 5X MSM (Sigma Aldrich) containing 33.9g/L Na_2HPO_4, 15g/L KH_2PO_4, 5g/L NH_4Cl and 2.5g/L NaCl was prepared by weighing 56.6g into 1L of distilled water. Separating the 1L salts solution into 5 200mL Duran flasks and autoclaving at 121°C. The 200 mL salts solutions were the stored at -4°C until required.

To produce the final growth media the 200 mL salts solution was added to 800 mL of distilled autoclaved water containing 0.6 g/L of yeast extract.

6.2.2 *Preparation of the experimental flasks*

To add the estrogen to the MSM (containing 0.6 g/L of yeast extract, 2 mL of a stock solution of 1g/L of estrogen in 50:50 acetonitrile:methanol (AcN:MeOH) stock solution (see 6.2.2.2) was evaporated using nitrogen gas to a volume of 0.25 mL. The 0.25 mL estrogen stock was then added to the MSM containing 0.6 g/L of yeast extract and evaporated for a further 80 minutes on a hot plate to remove the remaining acetonitrile and methanol. The solution was removed from the heat and allowed to cool and then 40 mL of the acclimated *R. equi* culture was inoculated into 360 ml of MSM with 0.6 g/L of yeast extract containing the estrogen at a concentration of 5 mg/L. The control was prepared without any estrogen solution.

6.2.2.1 Experimental flask conditions

The flasks were set up in triplicate for each condition, the conditions were containing mixed E2 and EE2 and the no estrogen control.

6.2.2.2 Preparing the estrogen stock solution

To prepare the 1 g/L mixed estrogen stock solution, 0.025 g of each estrogen powder (E2 and EE2) were measured into an amber glass 25 cm^3 volumetric flask. A

50:50 solution of AcN:MeOH was prepared and mixed in a 500 cm^3 volumetric flask, and dispensed into a beaker. The AcN:MeOH solution was transferred into the 25 cm^3 volumetric using a 25 cm^3 Pasteur pipette. The solution was then mixed by inversion and stored at 4°C.

6.2.3 *Monitoring for the gene expression experiments*

The flasks were incubated at 25°C for 311 hours with 155 rpm rotation and samples were taken every 3 hours for 18 hours and a final sample at 311 hours. However, the gene expression analysis will only include samples taken between 0 and 18 hours because the bacteria were in stationary/death phase at 311 hours, the growth and degradation at 311 hours (Chapter 5, Section 5.3.1).

At each time point the following samples were taken (triplicates of each);

- 1 mL - to measure bacterial growth
- 2 mL - for HPLC analysis
- 2 mL – for gene expression analysis

To assess bacterial growth optical density (OD), measured in a Spectrophotomer (ATI Unicam 8625 UV/Vis) (Fig. 5- 9).

Prior to HPLC analysis, samples were processed through sample preparation protocol and stored for estrogen quantification (see 6.2.3.1).

Prior to RNA extraction the samples for gene expression analysis were snap frozen at -80°C.

6.2.3.1 Estrogen sample preparation

In order to prepare the samples for uHPLC analysis, the samples were immediately centrifuged at 13000 x g for 3 min and 1 mL of the supernatant was filtered through 0.2 µm pore size 13mm EMD Millipore Durapore™ PVDF membrane syringe filters into a 2 mL wide screw neck 12x32 mm Chromacol™ amber glass vial with Chromacol™ Blue screw cap, open top with 6 mm hole and a PTFE septum for

HPLC analysis. Furthermore standards of known concentrations (1, 2.5, 5, 7.5 and 10 mg/L) were produced in MSM and were processed in the same way as the samples to calculate the concentration of estrogen in the samples (Appendix Fig. 0-10 -15).

6.2.3.2 uHPLC analysis of estrogen concentration

Quantification of estrogen concentration was carried out using uHPLC (ThermoScientific Dionex Ultimate 3000 RS autosampler, pump, column compartment and electrochemical detector) using a ThermoScientific Accucore C18 100 x 2.1 mm, 2.6µm particle size column with the settings of; flow rate = 0.4 mL/min, temperature = 40°C, injection volume = 50 – 90 µL, electrochemical potential = 1800 mV. The run was isocratic using 62% of mobile phase A and 38% of mobile phase B for 4- 6.5 minutes. Mobile phase A contained 95% water, 5% acetonitrile and 0.1% formic acid. Mobile phase B contained 95% acetonitrile and 5% water with 0.1% formic acid. The detection limit of estrogen was 100ng/L.

6.2.4 *Extraction of RNA*

The RNA was extracted from the samples for gene expression analysis in a Grade II Microbiological cabinet in which all equipment had been cleaned using 70% ethanol and RNA erase ® (MP Biomedicals) solution. The extraction method was a Phenol:Chloroform extraction with alcohol precipitation modified from (White, 2004).

1. The samples were thawed on ice and then 1 mL of the sample was added to a lysing matrix B tubes (MP Biomedicals) with 1 mL of Phenol:Chloroform, pH 4.5 (Invitrogen), and ribolysed (MPBio FastPrep®-24) for 30s to disrupt the cell walls.
2. The samples were then incubated on ice for 15 minutes.
3. Then centrifuged at 8000 x g for 15 min at 4°C.
4. The aqueous phase was carefully removed and transferred into a fresh 2 mL tube and the volume was 500 µL.
5. 50 µL of 3 M Sodium acetate was added and mixed by vortexing. Sodium acetate was added to assist in the ethanol precipitation of the RNA.

6. 500 µL of Phenol:Chloroform, pH 4.5 (Invitrogen), was added and mixed by vortexing for 1 min and then incubated on ice for 15 min.
7. Then the homogenate was centrifuged at 8000 x g for 15 min at 4°C.
8. The 500 µL aqueous phase was transferred to a fresh 2 mL tube.
9. 500 µL of cold 100% Isopropanol was added and mixed by inversion and then incubated at -20°C for 1 hour.
10. Centrifugation at 8000 x g for 20 min at 4°C until a pellet was formed.
11. The supernatant was removed and discarded and the pellet washed in cold 70% Ethanol by vortexing.
12. Following centrifugation at 4000 x g for 5 min at 4°C the residual ethanol was removed and the pellet air dried for 1 min.
13. The pellet was then resuspended in 100 µL of RNase free water heated to 55°C and then stored at -80°C.

6.2.4.1 Quantification of RNA

Nanodrop was used to assess the quality by the use of the 260/280 and 260/230 ratios which were expected to be 1.8-2.1 and above 1.5, respectively. Additionally, the quality was checked by electrophoresis on 1% agarose gel.

Furthermore the Qubit® 2.0 Fluorometer (Invitrogen by LifeTechnologies) was used along with the Qubit® RNA HS Assay kit to gain a more accurate RNA concentration measurement and all the RNA samples were diluted to a concentration of 1 ng/µL using Molecular grade water (Sigma Aldrich).

6.2.4.2 Reverse transcription of the RNA

To generate cDNA from the RNA the QIAGEN® QuantiTect® Reverse Transcription kit was used in which the RNA was thawed on ice. Firstly the genomic DNA (gDNA) elimination reactions were carried out on a 42°C heat block for 2 min and placed on ice until the reverse-transcription reactions were set up. To carry out the reverse-transcription reactions a PCR machine (Alpha PCR Max) was used with an incubation step at 42°C for 30 min, followed by 95°C for 3 min. The reverse-transcription reactions were then further diluted to 1:200 uL using Molecular grade

water (Sigma Aldrich) and stored at -20°C for usage in the geNorm™ and RT-qPCR experiments.

6.2.5 *geNorm™*

A geNorm™ Reference gene selection kit specifically for *R. equi* was developed by Primerdesign using synthetic oligonucleotides for the optimisation. The reference gene selection kit was used to compare the expression stability of 12 reference genes (Table 6- 2) across the estrogen treated and control samples over time in order to select the most stable, the lowest variation in gene expression, gene combination over the whole growth period. The 12 reference genes selected, were genes identified within gene expression analysis, used commonly within the literature, for *Rhodococcus*.

Table 6-2: The 12 reference genes within the geNormTM Reference gene selection kit for R. equi ATCC13557

REQ number	Uniprot gene name	Protein
00090	GyrB	DNA gyrase subunit B
21700	Ihf	Integration host factor
37720	proC	Pyrridine-5-carboxylase reductase
17550	atpB	ATP synthase subunit A
11880	arcA	Arginine deiminase
46170	G3P	Glyceraldehyde-3-phosphate dehydrogenase
36840	rplA	50S ribosomal protein L1
35570	rpoA	DNA-directed RNApolymerase subunit
14600	MDH	Malate dehydrogenase
17970	Rim	Ribosomal maturation factor
35650	Adk	Adenylate kinase
00010	dnaA	Chromosomal replication initiator protein

Using the reverse-transcription products produced, the samples were plated in triplicate and so that all samples for a given reference gene were within the same plate (Hellemans *et al.,* 2007). The BrightWhite Real-Time PCR plates (Primerdesign) were set up to have wells containing a total volume of 10 µL, of which 2.5 µL was template and 7.5 µL of Primerdesign 2x PrecisionPLUS™ Mastermix which contained 0.5 µL of the reference gene primer mix, 2 µL of RNase/DNase free water, and 5 µL of Primerdesign 2x PrecisionPLUS™ mastermix containing SYBR green, per reaction. The plates also contained a no template control (NTC) and no reverse transcriptase control (-RT) to ensure there was an absence of contamination and no formation of primer dimers. RT-qPCR was carried out using the CFX 96™

Real-Time System C1000^{TM} Thermal cycler BioRad in the CFX Manager ^{TM} 3.0 software with the amplification conditions for 40 cycles (Table 6- 3).

Table 6-3: geNorm^{TM} amplification conditions (PrimerDesign, 2016a)

Step	Time (s)	Temperature (°C)
Enzyme activation	120	95
Denaturation	10	95
Data collection	60	60
Melt curve	10	60
	20	95

The data produced was analysed using Biogazelle's qbasePLUS 3.1 software and the optimal number of reference genes was calculated where geNorm V < 0.15 when comparing the normalisation factor based on the geometric mean of the reference targets. The genes with the highest reference target stability across all the experimental conditions was determined where geNorm M ≤ 0.5.

6.2.6 *Selection of gene targets and creation of primers*

The gene sequences of the genes coding for enzymes potentially involved in estrogen degradation, which were identified within the genome of *R. equi* ATCC 13557, 3-carboxyethylcatechol 2,3-dioxygenase, putative cytochrome P450 monooxygenase, 2-hydroxymuconic semialdehyde hydrolase, 3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase, 3-(3-hydroxy-phenyl)propionate hydroxylase, 3-oxosteroid 1-dehydrogenase and transcriptional regulator IcIR family, were used for the design of the primer sets (Table 6- 4). The specific primer sets were developed by PrimerDesign as a custom designed real-time PCR assay for use with SYBR green.

Table 6-4: Primer sets provided by PrimerDesign

Enzyme	Function of enzyme within the hypothetical estrogen degradation pathway (Table 2- 1 and Fig. 2- 1)	Primer	Sequence	Tm	GC%	Dimer
2-hydroxymuconate-semialdehyde hydrolase	Reaction 12	Sense	CTACACGACCGACGCCTTC	57.3	63.2	0.0
		Antisense	CCGTGCAGGAACAGGATCG	57.6	63.2	-1.2
Putative cytochrome P450 monooxygenase	Reactions 6, 8, and/or 11	Sense	GCAGTGGTCGAGCTGATGA	56.5	57.9	-0.9
		Antisense	CGAGGTGCGAGATGGTGTC	57.3	63.2	0.0
3-oxosteriod 1-dehydrogenase	Function uncertain within estrogen degradation pathway	Sense	CCTGTTCATCACATCCAAGGAC	56.5	50	0.0
		Antisense	CGACTTCACGGCGTTGAAC	56.6	57.9	-0.2
IclR family transcriptional regulator	Regulation of the genes	Sense	CATCAGCTTCGGCGAGGTC	57.6	63.2	-0.9
		Antisense	CGATGAGTTCCTTCGATGCG	56.2	55	-0.8
Catechol 2,3 dioxygenase	Reactions 6, 8, and/or 11	Sense	GGACCACTTCAACGGGTTCT	56.8	55	0.0
		Antisense	AGGTCTTCGGCGATCTCCT	56.7	57.9	-0.1
3-(3-hydroxy-phenyl)proprionate hydroxylase	Reactions 5 and/or 9	Sense	TCTGTTGCCGCAGGAATCG	57.5	57.9	-0.3
		Antisense	CGTCGTCGTCCATGTACTTCT	57.2	52.4	0.0
3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase	Reaction 7	Sense	GGCAGGTCCTCGACATCAA	56.1	57.9	-0.8
		Antisense	AGCGGACGAGATGTTCACG	56.9	57.9	0.0

6.2.7 *RT-qPCR*

Using the reverse-transcription products produced, the samples were plated in with two biological samples per flask (three flasks experimental, three control) and each with triplicate technical replicates. The plate was set up to have wells containing a total volume of 10 µL, of which 2.5 µL was template and 7.5 µL of Primerdesign 2x PrecisionPLUS™ Mastermix which contained 0.5 µL of the reference gene primer mix, 2 µL of RNase/DNase free water, and 5 µL of Primerdesign 2x PrecisionPLUS™ mastermix containing SYBR green, per reaction. The plates also contained a no template control (NTC) and no reverse transcriptase control (-RT) to ensure there was an absence of contamination and no formation of primer dimers. RT-qPCR was carried out using the CFX 96™ Real-Time System C1000™ Thermal cycler BioRad in the CFX Manager ™ 3.0 software with the amplification conditions for 40 cycles (Table 6- 5).

Table 6-5: RT-qPCR amplification conditions (PrimerDesign, 2016b)

Step	Time (s)	Temperature (°C)
Enzyme activation	120	95
Denaturation	10	95
Data collection	60	60
Melt curve	10	60
	20	95

To analyse the results the delta delta CT method was used (Equation 6- 1) and then fold change was calculated using (Equation 6- 2) (Livak *et al.,* 2001). The arithmetic mean of the reference genes and samples were calculated for the experimental and control samples, however the geometric mean was taken to combine the average CT for the reference genes.

Equation 6-1: Calculation of delta delta CT normalised to the reference genes.

$$\Delta\Delta\mathrm{CT} = \big((\mathrm{CT}(\mathrm{target}, \mathrm{control}) - \mathrm{CT}(\mathrm{reference}, \mathrm{control})\big) - ((\mathrm{CT}(\mathrm{target}, \mathrm{treated}) - \mathrm{CT}(\mathrm{reference}, \mathrm{treated}))$$

Equation 6-2: Calculation of the fold change relative to the control.

$$\text{fold change} = 2^{\Delta\Delta CT}$$

To produce the standard errors associated with the fold change the variance within all six values for CT needed to be included in the final error, therefore the Taylor series was used (Equation 6- 3) (Vandesompele *et al.,* 2002).

Equation 6-3: Calculation of the standard errors associated with the delta delta CT values.

$$\text{Error } (g(a,b) = \sqrt{(a.\text{error } (a)^2) + (b.\text{error } (b^2)}$$

The upper and lower standard error was then calculated for the delta delta CT values and the standard errors of the upper and lower value were then given to the power of 2 to calculate the standard error within the fold change.

6.2.8 *Statistical testing and graphing software*

For the geNorm the values of geNorm M ≤ 0.5 and geNorm V < 0.15 are considered statistically significant. The graphs were produced using qbase+ software version 3.0 (Biogazelle, Zwijnaarde, Belgium - www.qbaseplus.com).

For the RT-qPCR the fold change in expression of a gene must be greater than 2 or less than 0.5 to suggest a significant difference in the relative gene expression of a gene, because the control fold change is equal to 1 (IRIC., 2017). Graphs were produced using SigmaPlot 12.5.

Additonally to determine the statistical significance of the difference in gene expression over the 18 hour period, the delta CT values of the experimental samples were compared to the delta CT values of the control samples, using a two-tailed Mann-Whitney test (Minitab 17, 2010). The data was determined to be not normal using the Anderson Darling normality test.

6.3 Results

6.3.1 *GeNorm™*

The geNorm V calculation compares the geometric mean of gene expression for the reference genes identified as the most stable, by an inclusion of reference genes which significantly effect the pair-wise variation (Vandesompele *et al.,* 2002). The optimal number of reference gene targets where geNorm V < 0.15 was determined to be the two most stably expressed reference gene targets (Fig. 6- 1). All the possible numbers of reference genes combinations were < 0.15, however the minimal but stable number of two reference genes was selected.

The geNorm M is used as a stability measure in which the average pairwise variation of a particular reference gene with all the other reference genes as standard deviation of the fold change in expression (Vandesompele *et al.,* 2002). Although there were three genes with geNorm M ≤ 0.5, the optimum number of reference gene targets was two, therefore the reference gene targets selected were 11880 Arginine deiminase (arcA) and 14600 Malate dehydrogenase (MDH) which respectively had the lowest average expression stability geNorm M values for the whole growth period (Fig. 6- 2).

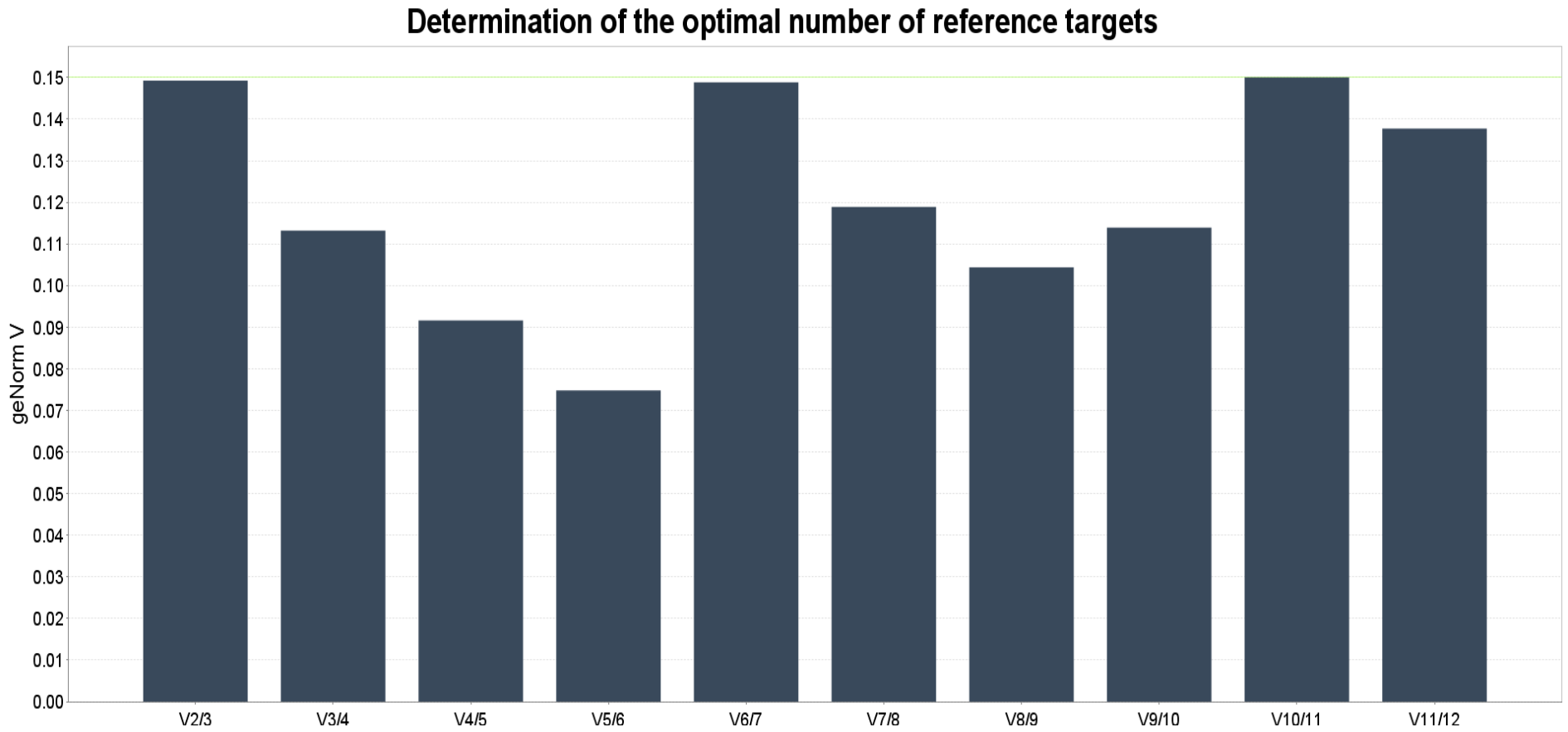

Figure 6-1: geNorm V <0.15 was used to determine the optimal number of reference gene targets.

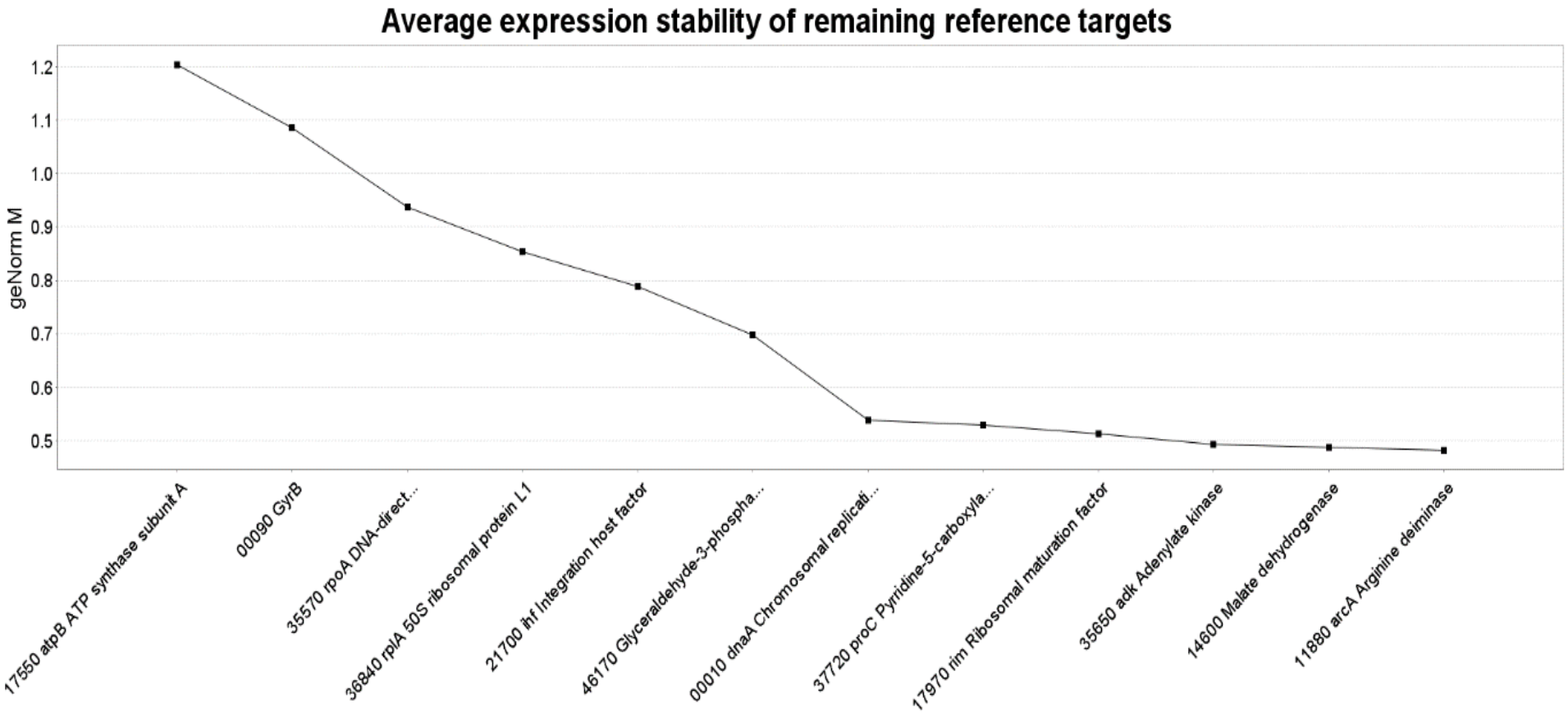

Figure 6-2: geNorm M ≤ 0.5 was used to determine the reference genes with the highest average expression stability. The genes with the highest average stability have the lowest geNorm M values (right) and the genes with the lowest average stability have the highest geNorm M values (left).

6.3.2 *RT-qPCR*

The average fold change of the gene coding for 3-(3-hydroxy-phenyl)proprionate hydroxylase was downregulated in the presence of estrogen at 12 hours and 15 hours by 3.4 times and 9.6 times respectively (Fig. 6- 3). Overall the gene expression of the gene coding for 3-(3-hydroxy-phenyl)proprionate hydroxylase over the 18 hours was not significantly different compared to the control (P=0.0967).

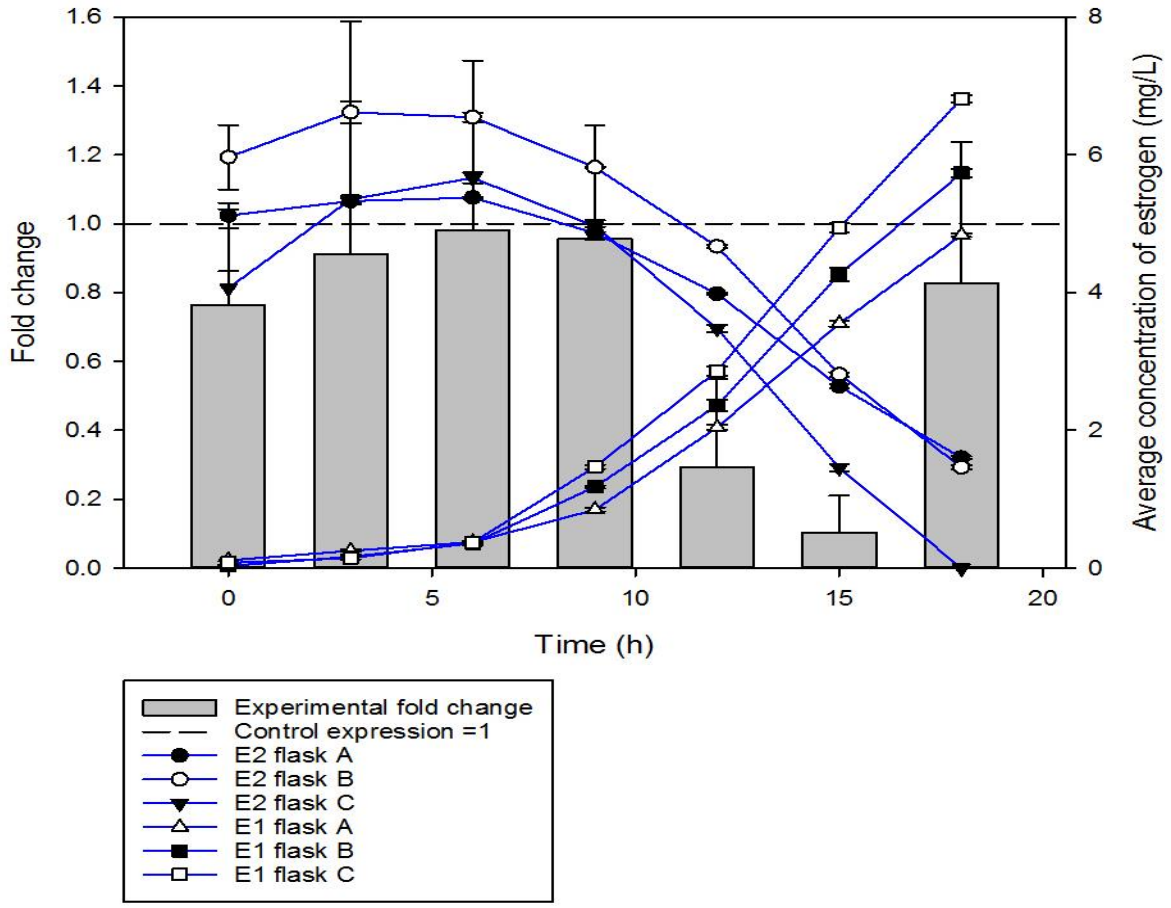

Figure 6-3: Fold change in expression of the gene coding for 3-(3-hydroxy-phenyl)proprionate hydroxylase when exposed to mixed estrogen relative to the control without estrogen. Normalised to the reference genes Malate dehydrogenase and Arginine deiminase. The error bars represent standard error. The dashed line at y=1 represents the expresson of the control

The average fold change of the gene coding for 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase when exposed to estrogen was downregulation between 6 and 18 hours by 5.3, 17.2, 6.1, 4.6, and 9.3 times respectively, with the expression being the lowest compared to the control by 17.2 times at 9 hours (Fig. 6- 4). Furthermore, over the 18 hour duration the gene expression compared to the control was statistically significant (P=0.0106).

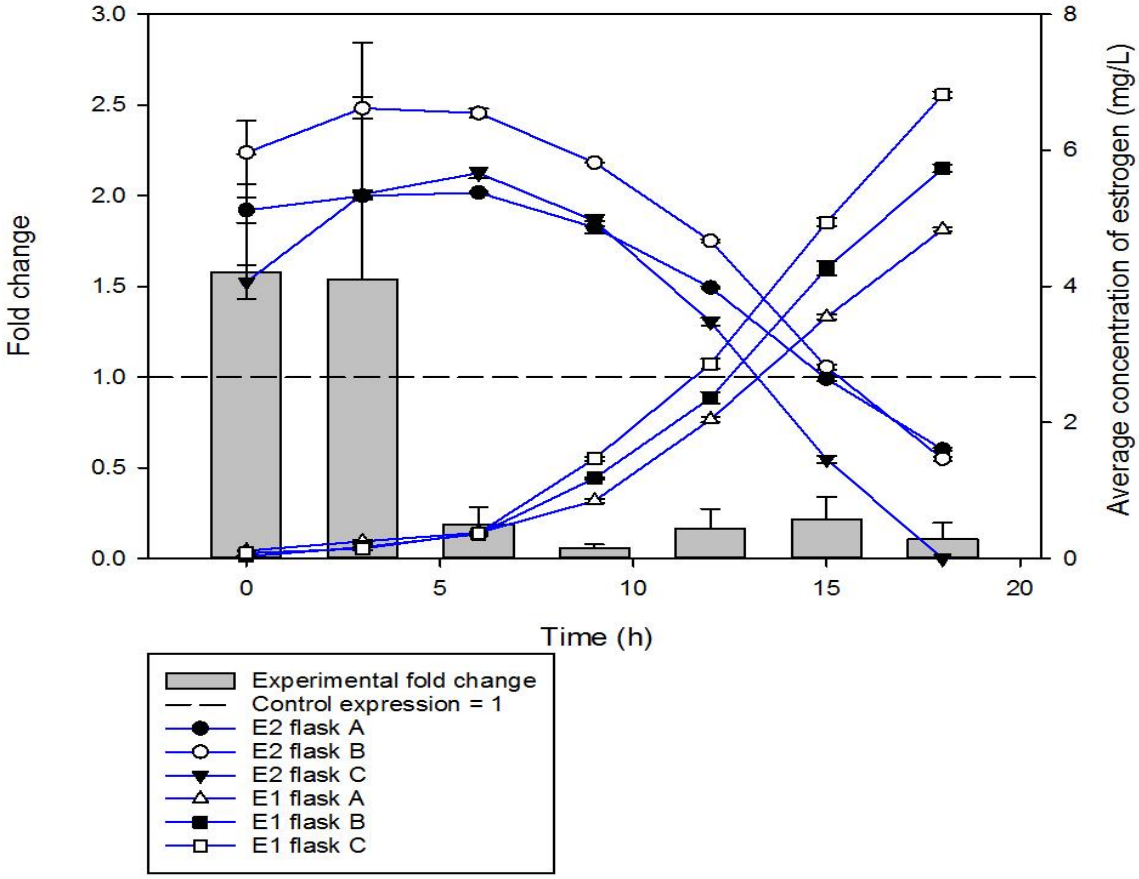

Figure 6-4: Fold change in expression of the gene coding for 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase when exposed to mixed estrogen relative to the control without estrogen. Normalised to the reference genes Malate dehydrogenase and Arginine deiminase. The error bars represent standard error. The dashed line at y=1 represents the expresson of the control.

The gene coding for catechol 2,3-dioxygenase gene was downregulated in the presence of estrogen at 12, 15 and 18 hours by 3.9, 2.9 and 2.4 times relative to the control (Fig. 6- 5). The expression was therefore lowest compared to the control at 12 hours by 3.9 times (Fig. 6- 5). However, over the 18 hour duration the gene expression was not significantly different when compared to the control (P= 0.0967).

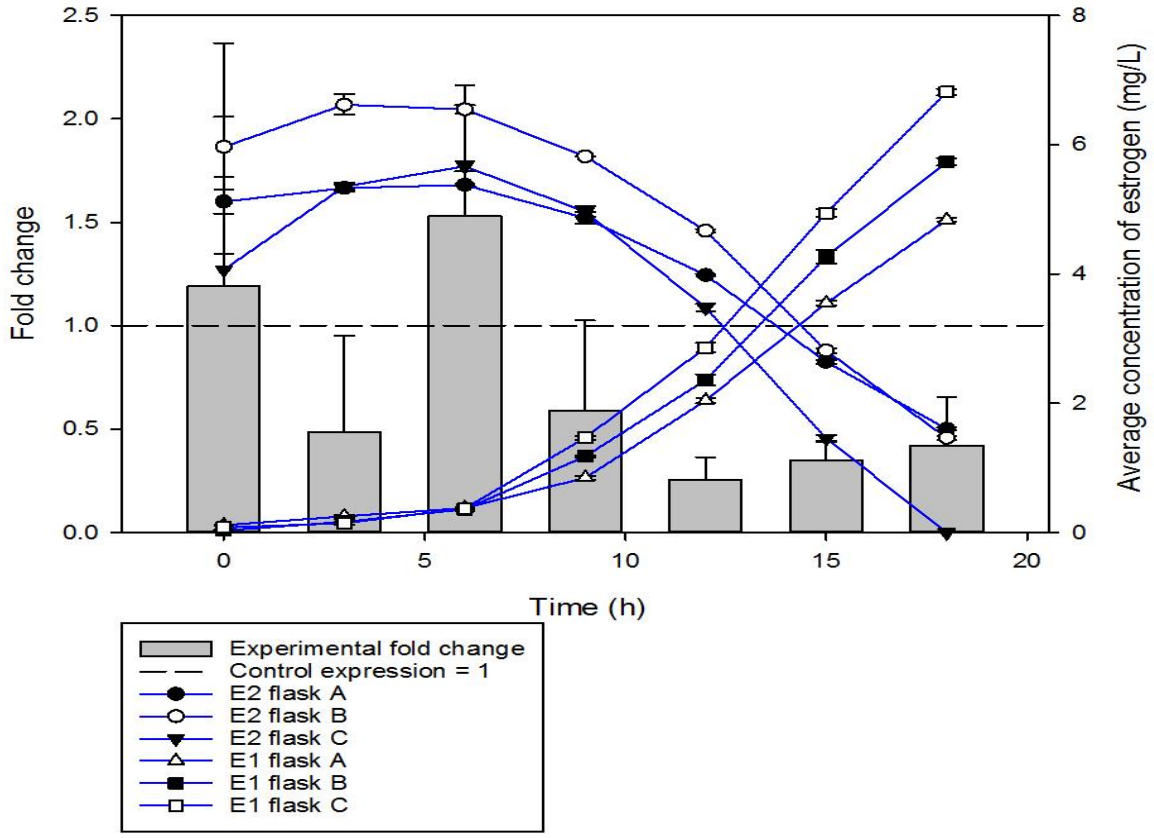

Figure 6-5: Fold change in expression of the gene coding for catechol 2,3-dioxygenase when exposed to mixed estrogen relative to the control without estrogen. Normalised to the reference genes Malate dehydrogenase and Arginine deiminase. The error bars represent standard error. The dashed line at y=1 represents the expresson of the control.

The average fold change of the gene coding for the IcIR family transcriptional regulator was downregulated at 9, 15 and 18 hours by 3.2, 2.0 and 2.4 times respectively (Fig. 6- 6). The lowest expression compared to the control was 3.2 times at 9 hours. Furthermore the gene expression over the 18 hours compared to the control was significantly different (P= 0.0106).

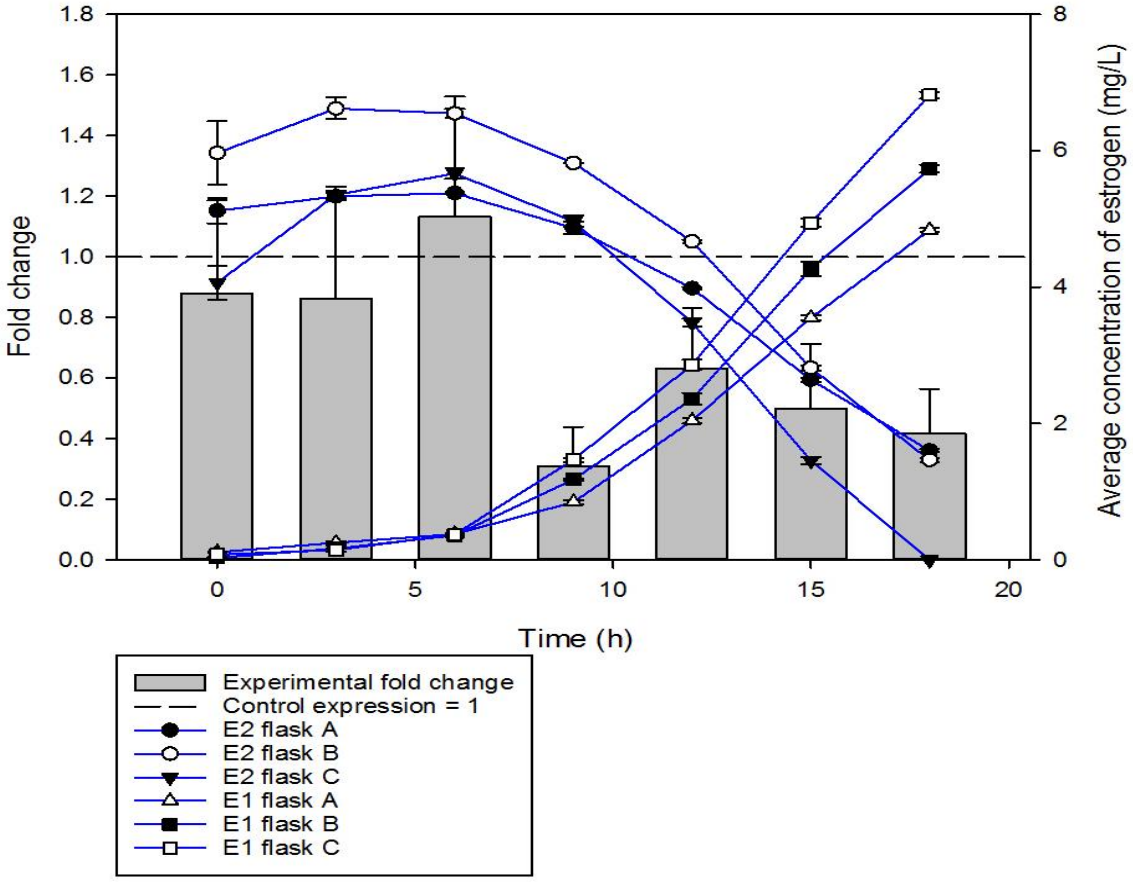

Figure 6-6: Fold change in expression of the gene coding for the IcIR family transcriptional regulator when exposed to mixed estrogen relative to the control without estrogen. Normalised to the reference genes Malate dehydrogenase and Arginine deiminase. The error bars represent standard error. The dashed line at y=1 represents the expresson of the control.

The gene coding for 3-oxosteriod 1-dehydrogenase was neither upregulated nor downregulated in the presence of estrogen compared to the control (Fig. 6- 7). Furthermore the gene expression over the 18 hour duration was not significantly different compared to the control (P= 0.3067).

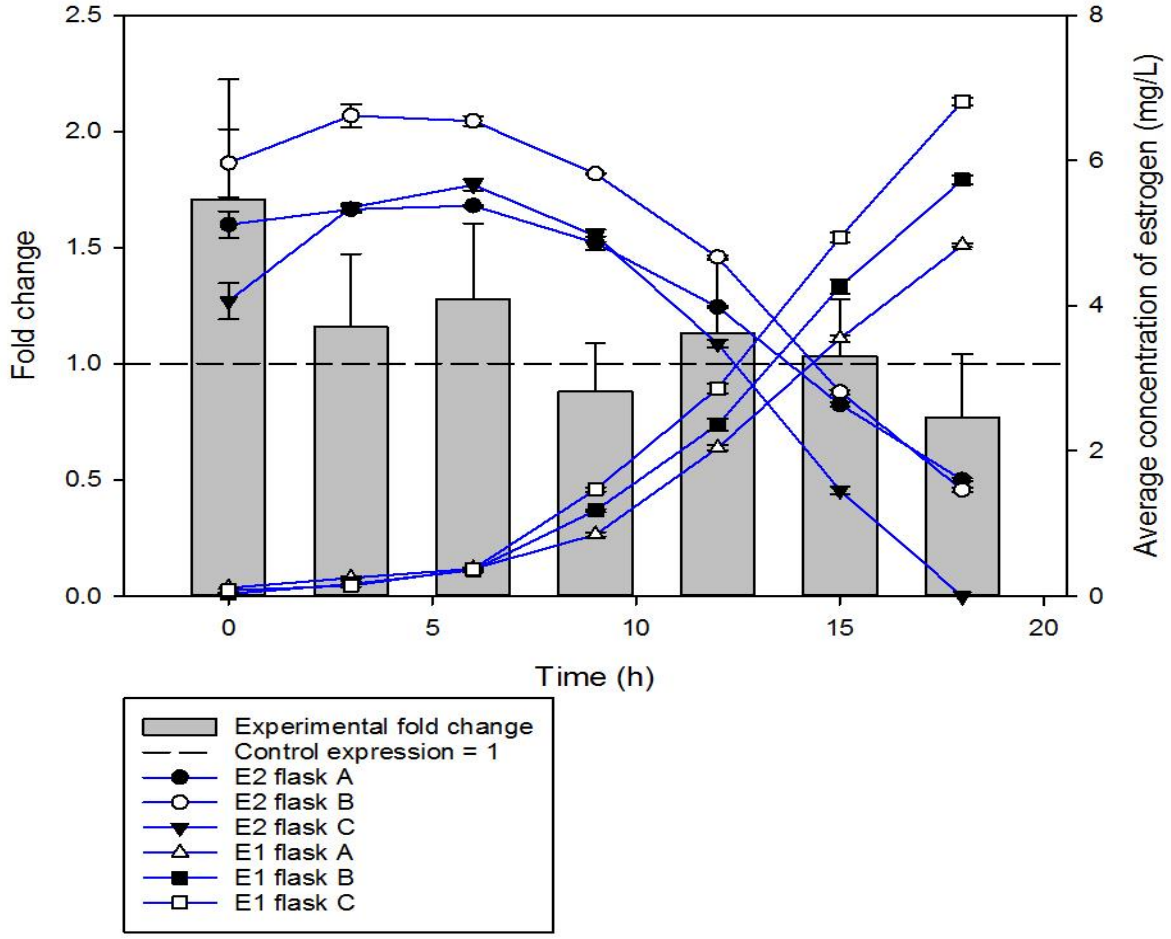

Figure 6-7: Fold change in expression of the gene coding for 3-oxosteriod 1-dehydrogenase when exposed to mixed estrogen relative to the control without estrogen. Normalised to the reference genes Malate dehydrogenase and Arginine deiminase. The error bars represent standard error. The dashed line at y=1 represents the expresson of the control.

The gene coding for putative cytochrome P450 monooxygenase was upregulated at 0 hours by 4.4 times more than the control, however at 18 hours it was downregulated by 3.1 times less than the control (Fig. 6- 8). However, the gene expression over the 18 hours was not statistically different compared to the control (P= 0.6093).

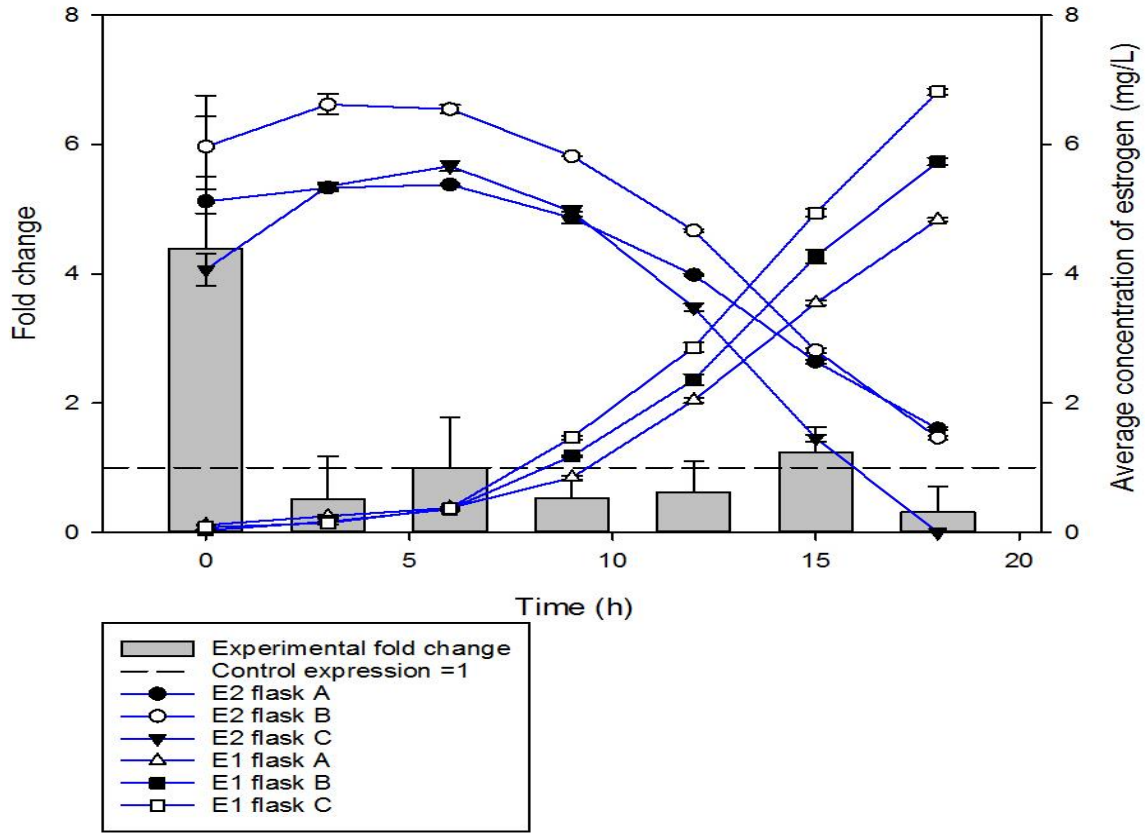

Figure 6-8: Fold change in expression of the gene coding for putative cytochrome P450 monooxygenase when exposed to mixed estrogen relative to the control without estrogen. Normalised to the reference genes Malate dehydrogenase and Arginine deiminase. The error bars represent standard error. The dashed line at y=1 represents the expresson of the control.

The gene expression of the gene coding for 2-hydroxymuconate-semialdehyde hydrolase was downregulated at 15 hours by 2.8 times less than the control (Fig. 6-9). Furthermore the gene expression over the 18 hour duration was statistically different compared to the control (P= 0.0049).

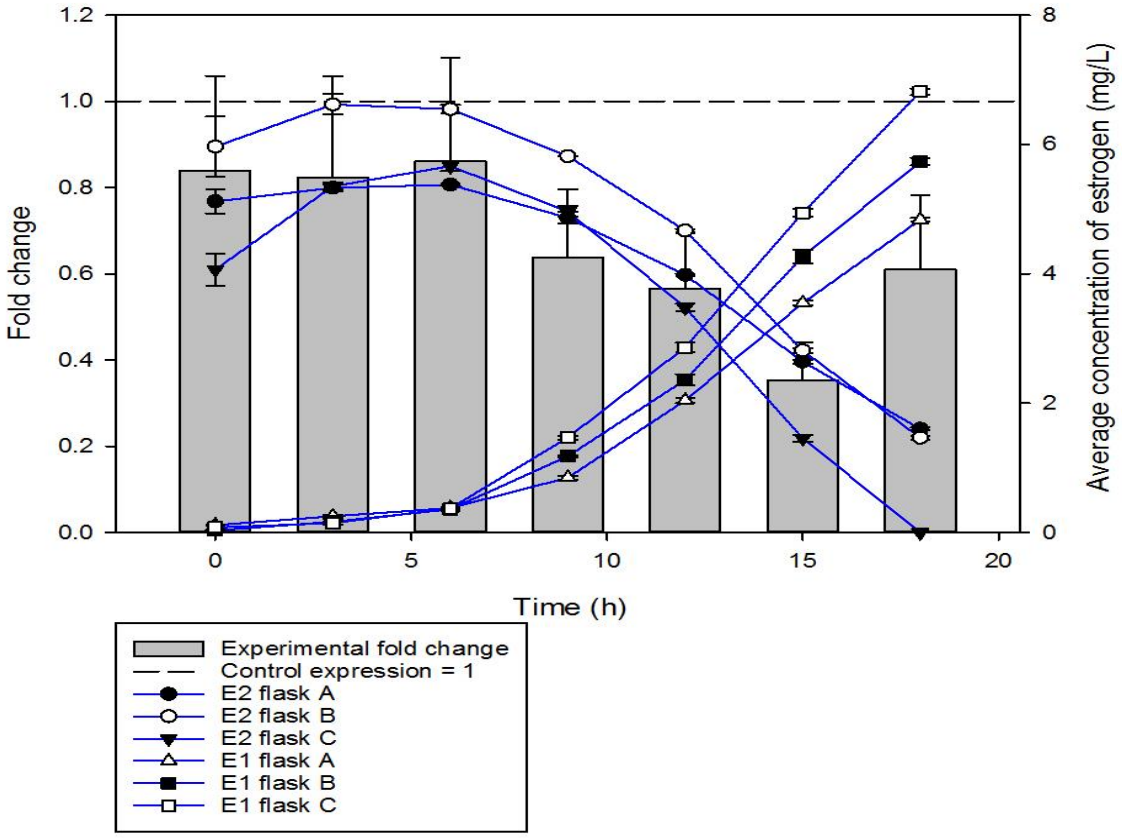

Figure 6-9: Fold change in expression of the gene coding for 2-hydroxymuconate-semialdehyde hydrolase when exposed to mixed estrogen relative to the control without estrogen. Normalised to the reference genes Malate dehydrogenase and Arginine deiminase. The error bars represent standard error. The dashed line at y=1 represents the expresson of the control.

6.4 Discussion

The gene expression of the genes coding for enzymes potentially involved in estrogen degradation seemed to vary between the experimental flasks and all of the genes were expressed lower than the control, with the exception of the genes coding for 3-oxosteriod 1-dehydrogenase which was expressed the same as the control for the 18 hour duration, and the putative cytochrome P450 monooxygenase gene which was expressed higher by 4.4 times more than the control initially. Furthermore the average expression of the genes seemed to be decrease relative to the control at at the later time points, mostly between 12 and 18 hours. The downregulation of these genes corresponds with the time point in the degradation experiment in which the growth of *R.equi* ATCC13557 began to enter the exponential phase and where the degradation of E2 began. Therefore, either the increase in E1 or other undetectable metabolites, the decrease in concentration of E2, or a combination of both led to a reduction in expression of the genes coding for 3-(3-hydroxy-phenyl)proprionate hydroxylase, 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase, catechol 2,3 dioxygenase, IcIR family transcriptional regulator, putative cytochrome P450 monoxygenase, and 2-hydroxymuconate-semialdehyde hydrolase, with the exception of 3-oxosteriod 1-dehydrogenase which remained unchanged.

When comparing the gene expression over the full 18 hours and not considering the expression at each time point individually, expression of the genes coding for 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase, IcIR family transcriptional regulator, and 2-hydroxymuconate-semialdehyde hydrolase are different compared to the control. However, the only gene upregulated in response to exposure to estrogen was the gene coding for putative cytochrome P450 monoxygenase which was significantly different than the control at 0 hours. The expression at 0 hours is likely due to a delay in taking samples from the flasks due to the large number of samples required, and it is likely that due to enrichment, sufficient time had passed for expression of genes to begin. The early expression of the gene encoding putative cytochrome P450 monoxygenase supports the hypothesis that the first step in estrogen degradation by *R. equi* ATCC13557 is the hydroxylation of the A ring in reaction 6 to produce 4-OH-E2 (Table 2- 1; Fig. 2- 1). The hydroxylation of the A ring is suggested to be the most common first step in estrogen degradation (Coombe *et*

al., 1966; Sih *et al.,* 1965; Chen *et al.,* 2017; Yu *et al.,* 2016) and supported by the EAWAG Biocatalysis/Biodegradation database pathway prediction (Hou *et al.,* 2011). However, in all of the experimental flasks the E2 was converted to E1 completely with the average concentration of E1 being equal or greater than the average decrease in E2, although being greater or equal may be due to issues with the standard curve for E1 (Chapter 5). Furthermore, the only metabolite detectable was E1, suggesting that the first step in estrogen metabolism involves 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase, from reaction 7 of the hypothetical estrogen degradation pathway (Table 2- 1; Fig. 2- 1), which has the function of converting E2 to E1 (Ye *et al.,* 2017). However, there was no upregulation of the gene encoding 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase in response to estrogen, so it is possible that another gene encoding 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase outside of this gene cluster is responsible for the conversion of E2 to E1, as the primers used in this experiment were only specific to the gene identified within the gene cluster (Fig. 3- 8). Furthermore, expression of the gene coding for 3-oxosteriod 1-dehydrogenase was unchanged, despite being identified as one of the first enzymes involved in steroid degradation (Sang *et al.,* 2012). However, the function within the hypothetical estrogen degradation pathway (Fig. 2- 1) was unknown, so the unchanged expression of 3-oxosteriod 1-dehydrogenase in the presence of estrogen is unsurprising.

One of the first possible steps in the degradation of E2 is the conversion of E1 by 17β-hydroxysteroid dehydrogenase in reaction 7 (Table 2- 1; Fig. 2- 1) therefore early expression at 0-3 hours of 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase was expected due to the conversion of E2 to E1 in *R. equi* ATCC13557. However, there was no upregulation of this gene, but another gene within the genome of *R. equi* ATCC13557 may be expressed to produce this enzyme. The other two pathways for E2 degradation are the conversion of E2 to keto-E2 by the action of hydroxylase (reaction 5) or conversion to 4-OH-E2 by dioxygenase (reaction 6) (Table 2- 1; Fig. 2- 1), therefore the genes coding for 2-hydroxymuconate-semialdehyde hydrolase, catechol 2,3 dioxygenase or putative cytochrome P450 monooxygenase, were expected to be expressed early at timepoints 0 to 6 hours. Initially the expression of the gene coding for catechol 2,3 dioxygenase was unchanged relative to the control and then downregulated at 12 to 18 hours as the

concentration of E1 increased. Therefore the gene encoding catechol 2,3 dioxygenase within the gene cluster doesn't have a role in the degradation of E2 by *R. equi* ATCC13557. However, expression of putative P450 monooxygenase, which can function in a similar way to dioxygenase, at 0 hours may suggest conversion to 4-OH-E2 in reaction 6 or E1 to 4-OH-E1 in reaction 8 (Table 2- 1; Fig. 2- 1), however, without metabolite analysis this cannot be confirmed. The gene coding for 3-(3-hydroxy-phenyl) propionate hydroxylase is suggested to be responsible for conversion to keto-E2 by hydroxylase in reaction 5 or conversion of E1 to 3-HSA in reaction 9 (Table 2- 1; Fig. 2- 1). However, the gene coding for 3-(3-hydroxy-phenyl) proprionate hydroxylase was not changed in response to estrogen exposure until at 12 and 15 hours where it was downregulated. Therefore 3-(3-hydroxy-phenyl) proprionate hydroxylase is unlikely to have a role in the degradation of estrogen by *R. equi* ATCC13557, this was expected as there was no significant reduction in E1 concentration during the degradation experiments. Furthermore, expression of the gene coding for 2-hydroxymuconate-semialdehyde hydrolase enzyme would be expected at a later timepoint because within the degradation pathway of E1, reaction 12 (Table 2- 1; Fig. 2- 1) happens at a later point, and again as there was no upregulation of this gene in the presence of estrogen or any significant reduction in the concentration of E1 during the degradation studies the gene coding for 2-hydroxymuconate-semialdehyde hydrolase is not involved in estrogen degradation by *R. equi* ATCC13557. Additonally the function of 3-oxosteriod 1-dehydrogenase is unclear in the degradation of E2 as in the degradation of testosterone it functions to assist in the cleavage of the A ring by the dehydrogenation leading to a double bond, however the double bond already exists within the aromatic A ring of both E1 and E2 (Horinouchi *et al.,* 2012; Guevara *et al.,* 2017), therefore as expected there was no change in expression when exposed to estrogen the gene coding for 3-oxosteriod 1-dehydrogenase has no role in estrogen degradation by *R. equi* ATCC13557. The IcIR family transcriptional regulator has the function of regulating the genes within the cluster involved in specific functions such as carbon metabolism (Fillet *et al.,* 2011), the IcIR family transcriptional regulator was not expressed differently from the control until downregulation at 9, 15 and 18 hours which correspond with the timepoints during which the majority of the other genes were downregulated.

Although the triplicate flask were set up as biological replicates, the degradation in each flask and gene expression were different from each other, for example there was complete metabolism of E2 by 18 hours in flask C but E2 was still detected in flasks A and B (Fig. 5- 12; 5- 13; 5- 14). The differences in degradation, may be due to factors such as differences in the starting concentration in each flask, pH, inoculum concentration (Yu *et al.,* 2016). The gene expression experiment used samples from the mixed estrogen containing flasks stored from the degradation experiments, it is likely that the degradation within the mixed flasks was co-metabolic as it was not growth-linked metabolism of estrogen. Therefore, in the future gene expression of these genes should be measured within the E2 containing flasks which was growth-linked metabolism (Fig. 5- 15).

If the gene expression experiment is interpreted in the contect of the mixed estrogen flask of the 45.5 hour experiment in which the quantification of E1 was much more reliable due to a good standard curve, there was 0.46 mg/L of E2 degraded and a decrease in the concentration of E1 at 45.5 hours (Fig. 5- 3). There was E2 degraded which was not converted to E1 and/or the E1 was degraded further, therefore the expression of the gene coding for putative cytochrome P450 monooxygenase within the gene cluster identified within *R.equi* ATCC13557 is involved in estrogen degradation, but these pathway reactions would be more strongly confirmed by the use of metabolite detection in addition to the RT-qPCR experiment. Further gene expression experiments need to be carried out to confirm the involvement of a gene coding for 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase within the genome of *R. equi* ATCC13557 in the conversion of E2 to E1. Additionally, gene expression of these genes should be measured during the degradation of E2 as the sole estrogen where the degradation is growth-linked.

Chapter 7: Final discussion

An updated and comprehensive potential estrogen degradation pathway for bacteria has been proposed, which combined the estrogen degradation pathway with that of testosterone degradation by *Comamonas testosteroni* (Horinouchi *et al.,* 2003). Using the pathway, a number of enzymes likely necessary in the metabolism of estrogen by bacteria were identified including; 17β-hydroxysteriod dehydrogenase, dioxygenase, hydroxylase, oxosteroid 1-dehydrogenase, hydratase and hydrolase. Some of these enzymes had also been previously implied in the degradation of estrogen, such as 17β-hydroxysteriod dehydrogenase, oxosteroid 1-dehydrogenase, rieske dioxygenase, catechol 2,3-dioxygenase. Furthermore genes encoding these enzymes were identified in the genomes of *Sphingomonas* sp. KC8 and *Pseudomonas putida* SJTE-1 (Liang *et al.,* 2012; Hu *et al.,* 2011). A gene cluster containing genes encoding these enzymes in addition to hydrolase identified within the testosterone degradation pathway, which may be a common steroid degradation pathway (Horinouchi *et al.,* 2003). Previously, only the gene responsible for producing the 17β-hydroxysteriod dehydrogenase enzyme, had been found to be upregulated in the presence of estrogen by use of RT-qPCR experiments (Ye *et al.,* 2017). However, in this work, evidence of upregulation of putative cytochrome P450 monooxygenase initially in the present of estrogen was presented, therefore the enzyme may have a role in estrogen degradation. However, the genes coding for 3-(3-hydroxy-phenyl)proprionate hydroxylase, catechol 2,3 dioxygenase, IcIR family transcriptional regulator, 3-oxosteriod 1-dehydrogenase, 2-hydroxymuconate-semialdehyde hydrolase, and surprisingly 3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase were not upregulated in the presence of estrogen but at later timepoints downregulated as the concentration of E2 decreased. However, another gene coding for 3-alpha-(or 20-beta)-hydroxysteriod dehydrogenase is likely responsible for the conversion of E2 to E1 in reaction 5 of the proposed estrogen degradation pathway as E1 was the only detectable metabolite (Fig. 2- 1). Therefore, metabolite studies are required in addition to the RT-qPCR experiments to fully evaluate the potential estrogen degradation pathway.

Additionally, the stability of reference genes in measuring the gene expression in the presence of estrogens has not been reported previously. In this work, the stability of twelve reference genes was compared and the use of multiple reference genes were

evaluated (Vandesompele *et al.,* 2002). It was found that the use of reference genes coding for malate dehydrogenase and arginine deiminase were the most stably expressed reference genes under these experimental conditions.

Method development of the sample preparation method for the quantification of estrogen demonstrated that, as shown previously, choice of filtration is highly important in the separation of biomass from estrogen containing samples due to their sorption to various filtration materials (Walker and Watson, 2010). Furthermore it is important to ensure that estrogen is soluble, and remains soluble, within the conditions of an estrogen degradation experiment due to the effects of ionic interactions, hydrogen-bonding, and pH (Carter and Sluss, 2013; Shareef *et al.,* 2006; Ruchelman, 1967). It appears that although dissolved initially in the estogen degradation experiments that the estrogen had precipitated out of the MSM and re-dissolved over time, resulting in the increase in initial estrogen concentrations of E2 and EE2. Therefore additional measures such as ensuring the pH is constant by use of a buffer and beginning the experiment immediately after dissolving the estrogen are required. In the degradation experiment *R. equi* ATCC13557 conversion of E2 into E1. Although an unknown metabolite was present within the E2 only degradation of the 45.5 hour experiments, there was no significant decrease in the concentration of E1 suggesting that there is not complete metabolism. Therefore further experiments are required to identify the unknown metabolite. The degradation of estrogen may not be linked to growth so may be due to co-metabolism in the presence of other estrogens, yeast extract or residual solvent within the MSM. The co-metabolic conversion of E2 to E1 by *R. equi* ATCC13557 has not been reported previously. There was no degradation of EE2 by *R. equi* ATCC13557 but this may have been due to differences in the pH, inoculum concentration, temperature fluctuations or absence of ethanol (Yu *et al.,* 2016; Larcher and Yargeau, 2013). Therefore experiments to investigate these effects upon the degradation of *R. equi* ATCC13557 are required.

Therefore, the findings of this study have successfully explored the study aims (Table 7- 1). The findings suggest that there may be more than one estrogen degradation pathway present in *R. equi* ATCC13557 as the conversion of E2 to E1 suggests that the first step in estrogen degradation is the dehydrogenation of the D ring and the upregulation of P450 monooxygenase suggests the first step is

hydroxylation of the A ring. Non specific monooxygenase activity has been reported previously in the degradation of E1 (Yu *et al.,* 2007) and ammonia monooxygenase was suggested to be involved in estrogen degradation by *Nitrosomonas europaea* (Shi *et al.,* 2004), and P450 monoxygenase is involved in E2 degradation within mammals (Scornaienchi *et al.,* 2010; Nishida *et al.,* 2013). Therefore, the gene coding for putative cytochrome P450 enzyme may have a role in estrogen degradation. Therefore, both genes coding for 17β-hydroxysteriod dehydrogenase and putative cytochrome P450 monooxygenase are important in estrogen degradation within *R. equi* ATCC13557. However, a bacterium with a greater ability to degrade estrogen or an improvement in the degradation experimental conditions to enhance the degradation by *R. equi* ATCC13557 is required to fully evaluate the potential estrogen degradation pathways. The current data could be improved if metabolite analysis was carried out in addition to RT-qPCR, allowing both the identification of the products and the enzymes which catalyse the reactions. Additionally, gene expression should be measured in the degradation experiments containing E2 as the sole estrogen as the degradation appeared to be growth-linked. Although costly, a transciptomic, proteomic and metabolomic approach, may provide a more comprehensive understanding of the genomic potential of estrogen degrading bacteria. Further primer sets should be developed to target the genes coding for the enzymes identified in the hypothetical estrogen degradation pathway (Fig. 2- 1; Table 2- 1) as the primers within this study were specific only to the gene cluster identified in *R. equi* ATCC13557 (Fig. 3- 8).

This research has shown that it is of importance that genome assembly is chosen carefully to suit the sequencing data by considering factors such as sequencing platform, sequence length and quality. Additionally, in the design of estrogen degradation studies there are a number of considerations to ensure the estrogen concentrations detected truly represent the losses of estrogen such as; 1) the filter material should be carefully chosen to reduce the retention of estrogen; 2) the loss of estrogen due to filtration and other sample preparation should be determined and if these losses are significant a control such as processing the standards in the same way or ensuring these losses are accounted for in the degradation concentrations; 3) choice of solvent is carefully considered to reduce solubility issues if present; 4) the effect upon degradation is considered if the solvent remains unremoved. Further

recommendations are that by searching the genomes of bacteria for genes coding for the enzymes necessary for estrogen degradation, using the hypothetical estrogen degradation pathway, may allow identification of previously unknown estrogen degrading bacteria. Furthemore, in addition to isolating and confirming estrogen degradation by biodegradation studies, a greater availability of genome sequences of estrogen degrading bacteria, particularly EE2 degraders, within public databases would assist in the understanding of estrogen biodegradation through a genomic approach.

Table 7-1: A table showing how the aims of the project were met.

Chapter	Aims	Findings
2	To compile a hypothetical pathway of estrogen degradation using existing information from current literature on estrogen and testosterone degradation.	A hypothetical estrogen degradation pathway was compiled (Fig. 2- 1). The key steps (reactions) and enzymes involved were summarised into a table (Table 2- 1).
3	To assemble the genome of *R. equi* ATCC 13557 and search it for genes coding for the enzymes that are potentially involved in estrogen degradation. To compare the genomes of potential estrogen degrading bacteria to localise the genes coding for enzymes that are suggested to be involved in estrogen degradation	The genome of *R. equi* ATCC13557 was assembled using both CLC and SPAdes. However, the SPAdes assembly was used in further analysis due to a better quality assembly and larger genome size. A gene cluster was identified within the genome of *R. equi* ATCC13557 containing genes from the hypothetical estrogen degradation pathway. The genome of *R. equi* ATCC13557 was compared with the genomes of *Sphingomonas* sp. KC8 and *Pseudomonas putida* SJTE-1, but the only genes with significant similarity were those encoding enzymes essential for survival. The genomes were searched for genes coding for the enzymes identified in the hypothetical estrogen degradation pathway and 3-oxosteriod 1-dehydrogenase was identified within *R. equi* ATCC13557. In *Sphingomonas* sp. KC8 3-ketosteriod 5-dehydrogenase, and dioxygenase were identified. In *P. putida* SJTE-1 only dioxygenase was identified.
4	To optimise the method of estrogen sample preparation and addition of estrogen to the growth media.	The better sample preparation to remove biomass was determined to be centrifugation followed by PVDF filtration.

		The method to add estrogen to the growth media was determined to be producing a stock solution in AcN/Me and adding to the MSM, followed by evaporation at 80℃ for 80 minutes.
5	Carry out degradation experiments to determine estrogen degradation under the experimental conditions.	*R. equi* ATCC13557 converted E2 to E1 and the degradation was cometabolic in the presence of other estrogens. When E2 was the sole estrogen there was production of an unidentified metabolite and some E2 may have been converted to this metabolite or possibly degraded as not all E2 was converted to E1. There was no degradation of EE2.
6	Develop new specific primer sets to identify and enumerate the identified genes coding for enzymes potentially involved in estrogen degradation using RT-qPCR. To quantify gene expression of *R. equi* ATCC13557 during degradation to determine the active genes.	Primers were developed to specifically target the genes within the gene cluster of potential estrogen degradation genes found in the genome of *R. equi* ATCC13557. RT-qPCR experiments using samples taken from the degradation experiment containing mixed estrogen in 5.3.2 (Fig. 5- 9, 5- 12, 5- 13 and 5- 14). The gene coding for putative cytochrome P450 monooxygenase was upregulated initially in response to estrogen, however the other genes, with the exception of 3-oxosteriod 1-dehydrogenase which remained unchanged, were downregulated at later timepoints as the concentration of E2 decreased and E1 increased. Thus, suggesting the first step in estrogen degradation by *R. equi* ATCC13557 is the hydroxylation of the A ring.

Reference list

ADEEL, M., SONG, X., WANG, Y., FRANCIS, D. & YANG, Y. 2017. Environmental impact of estrogens on human, animal and plant life: A critical review. Environ Int, 99, 107-119.

AFGAN, E., BAKER, D., VAN DEN BEEK, M., BLANKENBERG, D., BOUVIER, D., CECH, M., CHILTON, J., CLEMENTS, D., CORAOR, N., EBERHARD, C., GRUNING, B., GUERLER, A., HILLMAN-JACKSON, J., VON KUSTER, G., RASCHE, E., SORANZO, N., TURAGA, N., TAYLOR, J., NEKRUTENKO, A. & GOECKS, J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res, 44, W3-w10.

ALTURKI, A. A., TADKAEW, N., MCDONALD, J. A., KHAN, S. J., PRICE, W. E. & NGHIEM, L. D. 2010. Combining MBR and NF/RO membrane filtration for the removal of trace organics in indirect potable water reuse applications. Journal of Membrane Science, 365, 206-215.

ALUM, A., YOON, Y., WESTERHOFF, P. & ABBASZADEGAN, M. 2004. Oxidation of bisphenol A, 17beta-estradiol, and 17alpha-ethynyl estradiol and byproduct estrogenicity. Environ Toxicol, 19, 257-64.

AMEYE, L., ANTOINE, C., PAESMANS, M., DE AZAMBUJA, E. & ROZENBERG, S. 2014. Menopausal hormone therapy use in 17 European countries during the last decade. Maturitas, 79, 287-91.

ANDALURI, G., SURI, R. P. & KUMAR, K. 2012. Occurrence of estrogen hormones in biosolids, animal manure and mushroom compost. Environ Monit Assess, 184, 1197-205.

ANDERSEN, H., SIEGRIST, H., HALLING-SORENSEN, B. & TERNES, T. A. 2003. Fate of estrogens in a municipal sewage treatment plant. Environ Sci Technol, 37, 4021-6.

ANKLEY, G. T., FEIFAREK, D., BLACKWELL, B., CAVALLIN, J. E., JENSEN, K. M., KAHL, M. D., POOLE, S., RANDOLPH, E., SAARI, T. & VILLENEUVE, D. L. 2017. Re-evaluating the Significance of Estrone as an Environmental Estrogen. Environ Sci Technol, 51, 4705-4713.

AVBERSEK, M., SOMEN, J. & HEATH, E. 2011. Dynamics of steroid estrogen daily concentrations in hospital effluent and connected waste water treatment plant. J Environ Monit, 13, 2221-6.

AZIZ, R. K., BARTELS, D., BEST, A. A., DEJONGH, M., DISZ, T., EDWARDS, R. A., FORMSMA, K., GERDES, S., GLASS, E. M., KUBAL, M., MEYER, F., OLSEN, G. J., OLSON, R., OSTERMAN, A. L., OVERBEEK, R. A., MCNEIL, L. K., PAARMANN, D., PACZIAN, T., PARRELLO, B., PUSCH, G. D., REICH, C., STEVENS, R., VASSIEVA, O., VONSTEIN, V., WILKE, A. & ZAGNITKO, O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics, 9, 75.

BAKER, M. E. 2013. What are the physiological estrogens? Steroids, 78, 337-40.

BALKWILL, D. L., DRAKE, G. R., REEVES, R. H., FREDRICKSON, J. K., WHITE, D. C., RINGELBERG, D. B., CHANDLER, D. P., ROMINE, M. F., KENNEDY, D. W. & SPADONI, C. M. 1997. Taxonomic study of aromatic-degrading bacteria from deep-terrestrial-subsurface sediments and description of Sphingomonas aromaticivorans sp. nov., Sphingomonas subterranea sp. nov., and Sphingomonas stygia sp. nov. Int J Syst Bacteriol, 47, 191-201.

BANKEVICH, A., NURK, S., ANTIPOV, D., GUREVICH, A. A., DVORKIN, M., KULIKOV, A. S., LESIN, V. M., NIKOLENKO, S. I., PHAM, S., PRJIBELSKI, A. D., PYSHKIN, A. V., SIROTKIN, A. V., VYAHHI, N., TESLER, G., ALEKSEYEV, M. A. & PEVZNER, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol, 19, 455-77.

BEADLE, G. W. & TATUM, E. L. 1941. Genetic Control of Biochemical Reactions in Neurospora. Proc Natl Acad Sci U S A, 27, 499-506.

BELFROID, A. C., VAN DER HORST, A., VETHAAK, A. D., SCHÄFER, A. J., RIJS, G. B. J., WEGENER, J. & COFINO, W. P. 1999. Analysis and occurrence of estrogenic hormones and their glucuronides in surface water and waste water in The Netherlands. Science of The Total Environment, 225, 101-108.

BELL, S. G., WONG, L-L. 2007. P450 enzymes from the bacterium Novosphingobium aromaticivorans. Biochem Biophys Res Commun, 360, 666-672.

BELLONA, C., DREWES, J. E., XU, P. & AMY, G. 2004. Factors affecting the rejection of organic solutes during NF/RO treatment--a literature review. Water Res, 38, 2795-2809.

BENOTTI, M. J., TRENHOLM, R. A., VANDERFORD, B. J., HOLADY, J. C., STANFORD, B. D. & SNYDER, S. A. 2009. Pharmaceuticals and endocrine disrupting compounds in U.S. drinking water. Environ Sci Technol, 43, 597-603.

BENTLEY, D. R. 2006. Whole-genome re-sequencing. Curr Opin Genet Dev, 16, 545-552.

BERG, J. M., TYMOCZKO, J. L., STRYER, L (2002). Biochemistry. 5th ed. New York: Freeman, WH. Section 26.4, Important Derivatives of Cholesterol Include Bile Salts and Steroid Hormones. Available from: https://www.ncbi.nlm.nih.gov/books/NBK22339/

BERGSTRAND, L. H., CARDENAS, E., HOLERT, J., VAN HAMME, J. D. & MOHN, W. W. 2016. Delineation of Steroid-Degrading Microorganisms through Comparative Genomic Analysis. MBio, 7, e00166.

BILA, D., MONTALVAO, A. F., AZEVEDO DDE, A. & DEZOTTI, M. 2007. Estrogenic activity removal of 17beta-estradiol by ozonation and identification of by-products. Chemosphere, 69, 736-46.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30, 2114-20.

BOLONG, N., ISMAIL, A. F., SALIM, M. R. & MATSUURA, T. 2009. A review of the effects of emerging contaminants in wastewater and options for their removal. Desalination, 239, 229-246.

BON, E., BARBE, U., NUNEZ RODRIGUEZ, J., CUISSET, B., PELISSERO, C., SUMPTER, J. P. & LE MENN, F. 1997. Plasma vitellogenin levels during the annual reproductive cycle of the female rainbow trout (Oncorhynchus mykiss): establishment and validation of an ELISA. Comp Biochem Physiol B Biochem Mol Biol, 117, 75-84.

BRAGA, O., SMYTHE, G. A., SCHÄFER, A. I. & FEITZ, A. J. 2005. Fate of Steroid Estrogens in Australian Inland and Coastal Wastewater Treatment Plants. Environmental Science & Technology, 39, 3351-3358.

BRASLAVSKY, I., HEBERT, B., KARTALOV, E. & QUAKE, S. R. 2003. Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci U S A, 100, 3960-4.

BRENDA. 2017. GeneXplain GmbH. The Comprehensive Enzyme Information System. Release 2017.2. https://www.brenda-enzymes.org/index.php [Accessed 31/08/2017].

BRETTIN, T., DAVIS, J. J., DISZ, T., EDWARDS, R. A., GERDES, S., OLSEN, G. J., OLSON, R., OVERBEEK, R., PARRELLO, B., PUSCH, G. D., SHUKLA, M., THOMASON III, J. A., STEVENS, R., VONSTEIN, V., WATTAM, A. R. & XIA, F. 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. 5, 8365.

BRYNHILDSEN, J. 2014. Combined hormonal contraceptives: prescribing patterns, compliance, and benefits versus risks. Ther Adv Drug Saf, 5, 201-13.

BUSTIN, S. A., BENES, V., GARSON, J. A., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W., SHIPLEY, G. L., VANDESOMPELE, J. & WITTWER, C. T. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem, 55, 611-22.

BUTLER, J., MACCALLUM, I., KLEBER, M., SHLYAKHTER, I. A., BELMONTE, M. K., LANDER, E. S., NUSBAUM, C. & JAFFE, D. B. 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. Genome Res, 18, 810-20.

CALDWELL, D. J., MASTROCCO, F., ANDERSON, P. D., LANGE, R. & SUMPTER, J. P. 2012. Predicted-no-effect concentrations for the steroid estrogens estrone, 17beta-estradiol, estriol, and 17alpha-ethinylestradiol. Environ Toxicol Chem, 31, 1396-406.

CARERE, J., BAKER, P. & SEAH, S. Y. 2011. Investigating the molecular determinants for substrate channeling in BphI-BphJ, an aldolase-dehydrogenase

complex from the polychlorinated biphenyls degradation pathway. Biochemistry, 50, 8407-16.

CARR, D. L., MORSE, A. N., ZAK, J. C. & ANDERSON, T. A. 2011. Microbially Mediated Degradation of Common Pharmaceuticals and Personal Care Products in Soil Under Aerobic and Reduced Oxygen Conditions. Water, Air, & Soil Pollution, 216, 633-642.

CARTER, J. E. A. & SLUSS, P. M. 2013. Estradiol Solubility in Aqueous Systems: Effect of Ionic Concentrations, pH, and Organic Solvents. Journal of Hormones, 2013, 4.

CARTINELLA, J. L., CATH, T. Y., FLYNN, M. T., MILLER, G. C., HUNTER, K. W., JR. & CHILDRESS, A. E. 2006. Removal of natural steroid hormones from wastewater using membrane contactor processes. Environ Sci Technol, 40, 7381-6.

CASTIGLIONI, S., BAGNATI, R., CALAMARI, D., FANELLI, R. & ZUCCATO, E. 2005. A multiresidue analytical method using solid-phase extraction and high-pressure liquid chromatography tandem mass spectrometry to measure pharmaceuticals of different therapeutic classes in urban wastewaters. Journal of Chromatography A, 1092, 206-215.

CHEMDRAWBIO ULTRA. 1998-2014. Version 14.0.0.117 [computer software] PerkinElmer, CambridgeSoft Corporation.

CHEN, C.-Y., WEN, T.-Y., WANG, G.-S., CHENG, H.-W., LIN, Y.-H. & LIEN, G.-W. 2007. Determining estrogenic steroids in Taipei waters and removal in drinking water treatment using high-flow solid-phase extraction and liquid chromatography/tandem mass spectrometry. Science of The Total Environment, 378, 352-365.

CHEN, Y. L., YU, C. P., LEE, T. H., GOH, K. S., CHU, K. H., WANG, P. H., ISMAIL, W., SHIH, C. J. & CHIANG, Y. R. 2017. Biochemical Mechanisms and Catabolic Enzymes Involved in Bacterial Estrogen Degradation Pathways. Cell Chem Biol, 24, 712-724.e7.

CLAMENS, T., ROSAY, T., CREPIN, A., GRANDJEAN, T., KENTACHE, T., HARDOUIN, J., BORTOLOTTI, P., NEIDIG, A., MOOIJ, M., HILLION, M., VIEILLARD, J., COSETTE, P., OVERHAGE, J., O'GARA, F., BOUFFARTIGUES, E.,

DUFOUR, A., CHEVALIER, S., GUERY, B., CORNELIS, P., FEUILLOLEY, M. G. & LESOUHAITIER, O. 2017. The aliphatic amidase AmiE is involved in regulation of Pseudomonas aeruginosa virulence. Sci Rep, 7, 41178.

CLARA, M., KREUZINGER, N., STRENN, B., GANS, O. & KROISS, H. 2005. The solids retention time-a suitable design parameter to evaluate the capacity of wastewater treatment plants to remove micropollutants. Water Res, 39, 97-106.

CLARA, M., STRENN, B., SARACEVIC, E. & KREUZINGER, N. 2004. Adsorption of bisphenol-A, 17 beta-estradiole and 17 alpha-ethinylestradiole to sewage sludge. Chemosphere, 56, 843-51.

CLC GENOMICS WORKBENCH 9.5.3 Available at: http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/754/index.php?manual=Introduction_CLC_Genomics_Workbench.html. [Accessed 15 Sep. 2017].

COMBALBERT, S. & HERNANDEZ-RAQUET, G. 2010. Occurrence, fate, and biodegradation of estrogens in sewage and manure. Appl Microbiol Biotechnol, 86, 1671-92.

COOMBE, R. G., TSONG, Y. Y., HAMILTON, P. B. & SIH, C. J. 1966. Mechanisms of Steroid Oxidation by Microorganisms: X. OXIDATIVE CLEAVAGE OF ESTRONE. Journal of Biological Chemistry, 241, 1587-1595.

COPELAND, P. A., SUMPTER, J. P., WALKER, T. K. & CROFT, M. 1986. Vitellogenin levels in male and female rainbow trout (Salmo gairdneri Richardson) at various stages of the reproductive cycle. Comp Biochem Physiol B, 83, 487-93.

CUI, J., SHEN, Y. & LI, R. 2013. Estrogen synthesis and signaling pathways during ageing: from periphery to brain. Trends in molecular medicine, 19, 197-209.

CZAJKA, C. P. & LONDRY, K. L. 2006. Anaerobic biotransformation of estrogens. Science of the total environment, 367, 932-941.

DA CUNHA, D. L., DA SILVA, S. M., BILA, D. M., OLIVEIRA, J. L., SARCINELLI PDE, N. & LARENTIS, A. L. 2016. Regulation of the synthetic estrogen 17alpha-ethinylestradiol in water bodies in Europe, the United States, and Brazil. Cad Saude Publica, 32, e00056715.

DALTON, H., STIRLING, D. I. & QUAYLE, J. R. 1982. Co-Metabolism [and Discussion]. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 297, 481-496.

DARLING, A. C., MAU, B., BLATTNER, F. R. & PERNA, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res, 14, 1394-403.

Darling, A. C. 2004. [Computer Program]. Mauve, (the Darling lab | computational (meta) genomics, University of Technology Sydney.

DE MES, T. Z., KUJAWA-ROELEVELD, K., ZEEMAN, G. & LETTINGA, G. 2008. Anaerobic biodegradation of estrogens--hard to digest. Water Sci Technol, 57, 1177-82.

DEMALEKI, Z., LAI, E. P. & DABEK-ZLOTORZYNSKA, E. 2010. Capillary electrophoresis characterization of molecularly imprinted polymer particles in fast binding with 17beta-estradiol. J Sep Sci, 33, 2796-803.

DEPLEDGE, M. H. & BILLINGHURST, Z. 1999. Ecological Significance of Endocrine Disruption in Marine Invertebrates. Marine Pollution Bulletin, 39, 32-38.

DOERFEL, L. K. & RODNINA, M. V. 2013. Elongation factor P: Function and effects on bacterial fitness. Biopolymers, 99, 837-45.

DRESSMAN, D., YAN, H., TRAVERSO, G., KINZLER, K. W. & VOGELSTEIN, B. 2003. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc Natl Acad Sci U S A, 100, 8817-22.

DUONG, C. N., LEE, J. H., LIM, B. J. & KIM, S. D. 2011. Biodegradation of estrogen conjugates by bacteria isolated from river sediments. Water Sci Technol, 64, 1750-8.

EKBLOM, R. & WOLF, J. B. 2014. A field guide to whole-genome sequencing, assembly and annotation. Evol Appl, 7, 1026-42.

FAHRBACH, M., KUEVER, J., MEINKE, R., KAMPFER, P. & HOLLENDER, J. 2006. Denitratisoma oestradiolicum gen. nov., sp. nov., a 17beta-oestradiol-degrading, denitrifying betaproteobacterium. Int J Syst Evol Microbiol, 56, 1547-52.

FAHRBACH, M., KUEVER, J., REMESCH, M., HUBER, B. E., KAMPFER, P., DOTT, W. & HOLLENDER, J. 2008. Steroidobacter denitrificans gen. nov., sp. nov., a steroidal hormone-degrading gammaproteobacterium. Int J Syst Evol Microbiol, 58, 2215-23.

FANG-FANG, Z., DAN, Q., LIANG-MIN, G. & CHANG-PING, Y. 2012. Microbial degradation of estrogens in the environment. Microbiol. China, 39, 711-721.

FEDURCO, M., ROMIEU, A., WILLIAMS, S., LAWRENCE, I. & TURCATTI, G. 2006. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res, 34, e22.

FERNANDEZ, M. P., NOGUEROL, T.-N., LACORTE, S., BUCHANAN, I. & PIÑA, B. 2009. Toxicity identification fractionation of environmental estrogens in waste water and sludge using gas and liquid chromatography coupled to mass spectrometry and recombinant yeast assay. Analytical and Bioanalytical Chemistry, 393, 957-968.

FIELD, D., TIWARI, B., BOOTH, T., HOUTEN, S., SWAN, D., BERTRAND, N. & THURSTON, M. 2006. Open software for biologists: from famine to feast. Nat Biotechnol, 24, 801-3.

FILBY, A. L., NEUPARTH, T., THORPE, K. L., OWEN, R., GALLOWAY, T. S. & TYLER, C. R. 2007. Health impacts of estrogens in the environment, considering complex mixture effects. Environ Health Perspect, 115, 1704-10.

FILLET, S., KRELL, T., MOREL, B., LU, D., ZHANG, X. & RAMOS, J. L. 2011. Intramolecular signal transmission in a tetrameric repressor of the IcIR family. Proc Natl Acad Sci U S A, 108, 15372-15377.

FINLAY-MOORE, O., HARTEL, P. G. & CABRERA, M. L. 2000. 17β-Estradiol and Testosterone in Soil and Runoff from Grasslands Amended with Broiler Litter. Journal of Environmental Quality, 29, 1604-1611.

FORGÁCS, E. & CSERHÁTI, T. 2003. Chromatography: Principles. In: Encyclopedia of Food Sciences and Nutrition (Second Edition). Page 1259.

FOUNDATION FOR GENOMICS AND POPULATION HEALTH (PHG). 2011. Next steps in the sequence. The implications of whole genome sequencing for health in

the UK [Internet]. Cambridge (UK): PHG Foundation. Available from: http://www.phgfoundation.org/file/10363/ [Accessed 17 Aug. 2017].

FUJII, K., KIKUCHI, S., SATOMI, M., USHIO-SATA, N. & MORITA, N. 2002. Degradation of 17beta-estradiol by a gram-negative bacterium isolated from activated sludge in a sewage treatment plant in Tokyo, Japan. Appl Environ Microbiol, 68, 2057-60.

FUJII, K., SATOMI, M., MORITA, N., MOTOMURA, T., TANAKA, T. & KIKUCHI, S. 2003. Novosphingobium tardaugens sp. nov., an oestradiol-degrading bacterium isolated from activated sludge of a sewage treatment plant in Tokyo. Int J Syst Evol Microbiol, 53, 47-52.

GABET-GIRAUD, V., MIEGE, C., HERBRETEAU, B., HERNANDEZ-RAQUET, G. & COQUERY, M. 2010. Development and validation of an analytical method by LC-MS/MS for the quantification of estrogens in sewage sludge. Analytical and Bioanalytical Chemistry, 396, 1841-1851.

GAO, R., SU, X., HE, X., CHEN, L. & ZHANG, Y. 2011. Preparation and characterisation of core-shell CNTs@MIPs nanocomposites and selective removal of estrone from water samples. Talanta, 83, 757-64.

GARCÍA, J. J., UHÍA, I., GARCÍA, E., GALÁN, B. 2011. Degradation of Steriods In: Microbial Bioremediation of Non-metals. Koukkou, AI(ed). Caister Academic Press, Norfolk UK.

GAULKE, L. S., STRAND, S. E., KALHORN, T. F. & STENSEL, H. D. 2008. 17alpha-ethinylestradiol transformation via abiotic nitration in the presence of ammonia oxidizing bacteria. Environ Sci Technol, 42, 7622-7.

GIBSON, D. T., WANG, K. C., SIH, C. J. & WHITLOCK JR, H. 1966. Mechanisms of Steroid Oxidation by Microorganisms IX. On the mechanism of ring A cleavage in the degradation of 9,10-seco steroids by microorganisms. J Biol Chem, 241, 551-9.

GILBERT, N. 2012. Drug-pollution law all washed up. Nature, 491, 503-4.

GNERRE, S., MACCALLUM, I., PRZYBYLSKI, D., RIBEIRO, F. J., BURTON, J. N., WALKER, B. J., SHARPE, T., HALL, G., SHEA, T. P., SYKES, S., BERLIN, A. M., AIRD, D., COSTELLO, M., DAZA, R., WILLIAMS, L., NICOL, R., GNIRKE, A.,

NUSBAUM, C., LANDER, E. S. & JAFFE, D. B. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A, 108, 1513-8.

GOODWIN, S., MCPHERSON, J. D. & MCCOMBIE, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet, 17, 333-51.

GRANT, J. R., ARANTES, A. S. & STOTHARD, P. 2012. Comparing thousands of circular genomes using the CGView Comparison Tool. BMC Genomics, 13, 202-202.

GRUND, S., HIGLEY, E., SCHONENBERGER, R., SUTER, M. J., GIESY, J. P., BRAUNBECK, T., HECKER, M. & HOLLERT, H. 2011. The endocrine disrupting potential of sediments from the Upper Danube River (Germany) as revealed by in vitro bioassays and chemical analysis. Environ Sci Pollut Res Int, 18, 446-60.

GUEVARA, G., FERNANDEZ DE LAS HERAS, L., PERERA, J. & NAVARRO LLORENS, J. M. 2017. Functional differentiation of 3-ketosteroid Delta1-dehydrogenase isozymes in Rhodococcus ruber strain Chol-4. Microb Cell Fact, 16, 42.

GUREVICH, A., SAVELIEV, V., VYAHHI, N. & TESLER, G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics, 29, 1072-5.

GUTTMACHER INSTITUTE (2013) Contraception in the United States. http://www.guttmacher.org/pubs/fb_contr_use.html [Accessed 14 Sep. 2017].

HAIYAN, R., SHULAN, J., UD DIN AHMAD, N., DAO, W. & CHENGWU, C. 2007. Degradation characteristics and metabolic pathway of 17alpha-ethynylestradiol by Sphingobacterium sp. JCR5. Chemosphere, 66, 340-6.

HAMID, H. & ESKICIOGLU, C. 2012. Fate of estrogenic hormones in wastewater and sludge treatment: A review of properties and analytical detection techniques in sludge matrix. Water Res, 46, 5813-33.

HAN, J., QIU, W., HU, J. & GAO, W. 2012. Chemisorption of estrone in nylon microfiltration membranes: Adsorption mechanism and potential use for estrone removal from water. Water Res, 46, 873-81.

HANSELMAN, T. A., GRAETZ, D. A. & WILKIE, A. C. 2003. Manure-Borne Estrogens as Potential Environmental Contaminants: A Review. Environmental Science & Technology, 37, 5471-5478.

HANUKOGLU, I. 1992. Steroidogenic enzymes: Structure, function, and role in regulation of steroid hormone biosynthesis. The Journal of Steroid Biochemistry and Molecular Biology, 43, 779-804.

HARRIS, T. D., BUZBY, P. R., BABCOCK, H., BEER, E., BOWERS, J., BRASLAVSKY, I., CAUSEY, M., COLONELL, J., DIMEO, J., EFCAVITCH, J. W., GILADI, E., GILL, J., HEALY, J., JAROSZ, M., LAPEN, D., MOULTON, K., QUAKE, S. R., STEINMANN, K., THAYER, E., TYURINA, A., WARD, R., WEISS, H. & XIE, Z. 2008. Single-molecule DNA sequencing of a viral genome. Science, 320, 106-9.

HASSANI, G., BABAEI, A. A., TAKDASTAN, A., SHIRMARDI, M., YOUSEFIAN, F., MOHAMMADI, M. J. 2016. OCCURRENCE AND FATE OF 17β-ESTRADIOL IN WATER RESOURCES AND WASTEWATER IN AHVAZ, IRAN. Global. NEST. J, 18, 855-66.

HAZEN, T. C. 2010. Cometabolic Bioremediation. In: TIMMIS, K. N. (ed.) Handbook of Hydrocarbon and Lipid Microbiology. Berlin, Heidelberg: Springer Berlin Heidelberg.

HE, Y. J., CHEN, W., ZHENG, X. Y., WANG, X. N. & HUANG, X. 2013. Fate and removal of typical pharmaceuticals and personal care products by three different treatment processes. Sci Total Environ, 447, 248-54.

HECKER, M. & HOLLERT, H. 2011. Endocrine disruptor screening: Regulatory perspectives and needs.

HEID, C. A., STEVENS, J., LIVAK, K. J. & WILLIAMS, P. M. 1996. Genome Methods: Real Time Quantitative PCR. Genome. Res, 6: 986-994.

HELLEMANS, J., MORTIER, G., DE PAEPE, A., SPELEMAN, F. & VANDESOMPELE, J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol, 8, R19.

HELTON, E. D., WILLIAMS, M. C. & GOLDZIECHER, J. W. 1977. Oxidative metabolism and de-ethynylation of 17alpha-ethynylestradiol by baboon liver microsomes. Steroids, 30, 71-83.

HINTEMAN, T., SCHNEIDER, C., SCHOLER, H. F. & SCHNEIDER, R. J. 2006. Field study using two immunoassays for the determination of estradiol and ethinylestradiol in the aquatic environment. Water Res, 40, 2287-94.

HOHENBLUM, P., GANS, O., MOCHE, W., SCHARF, S. & LORBEER, G. 2004. Monitoring of selected estrogenic hormones and industrial chemicals in groundwaters and surface waters in Austria. Sci Total Environ, 333, 185-93.

HOLLERT, H., DURR, M., HOLTEY-WEBER, R., ISLINGER, M., BRACK, W., FARBER, H., ERDINGER, L. & BRAUNBECK, T. 2005. Endocrine disruption of water and sediment extracts in a non-radioactive dot blot/RNAse protection-assay using isolated hepatocytes of rainbow trout. Environ Sci Pollut Res Int, 12, 347-60.

HORINOUCHI, M., HAYASHI, T., KOSHINO, H., KURITA, T. & KUDO, T. 2005. Identification of 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid, 4-hydroxy-2-oxohexanoic acid, and 2-hydroxyhexa-2,4-dienoic acid and related enzymes involved in testosterone degradation in Comamonas testosteroni TA441. Appl Environ Microbiol, 71, 5275-81.

HORINOUCHI, M., HAYASHI, T., KOSHINO, H., YAMAMOTO, T. & KUDO, T. 2003. Gene encoding the hydrolase for the product of the meta-cleavage reaction in testosterone degradation by Comamonas testosteroni. Appl Environ Microbiol, 69, 2139-52.

HORINOUCHI, M., HAYASHI, T. & KUDO, T. 2012. Steroid degradation in Comamonas testosteroni. J Steroid Biochem Mol Biol, 129, 4-14.

HORINOUCHI, M., KURITA, T., HAYASHI, T. & KUDO, T. 2010. Steroid degradation genes in Comamonas testosteroni TA441: Isolation of genes encoding a Delta4(5)-isomerase and 3alpha- and 3beta-dehydrogenases and evidence for a 100 kb steroid degradation gene hot spot. J Steroid Biochem Mol Biol, 122, 253-63.

HORINOUCHI, M., KURITA, T., YAMAMOTO, T., HATORI, E., HAYASHI, T. & KUDO, T. 2004. Steroid degradation gene cluster of Comamonas testosteroni

consisting of 18 putative genes from meta-cleavage enzyme gene tesB to regulator gene tesR. Biochem Biophys Res Commun, 324, 597-604.

HORINOUCHI, M., YAMAMOTO, T., TAGUCHI, K., ARAI, H. & KUDO, T. 2001. Meta-cleavage enzyme gene tesB is necessary for testosterone degradation in Comamonas testosteroni TA441. Microbiology, 147, 3367-75.

HOROWITZ, N. H. 1945. On the Evolution of Biochemical Syntheses. Proc Natl Acad Sci U S A, 31, 153-7.

HOU, B. K., TURNBULL, M., YU, J. 2011. The EAWAG-BBD biotransformation rule, ruleID# bt0014. EAWAG-BBD 2017. doi: http://eawag-bbd.ethz.ch/servlets/rule.jsp?rule=bt0014 [Accessed 4 Sep. 2017]

HU, A., HE, J., CHU, K. H. & YU, C. P. 2011. Genome sequence of the 17beta-estradiol-utilizing bacterium Sphingomonas strain KC8. J Bacteriol, 193, 4266-7.

HU, A., HE, J., CHU, K.-H. & YU, C.-P. 2011. Genome Sequence of the 17β-Estradiol-Utilizing Bacterium Sphingomonas Strain KC8. Journal of Bacteriology, 193, 4266-4267.

HU, J., ZHANG, H. & CHANG, H. 2005. Improved method for analyzing estrogens in water by liquid chromatography–electrospray mass spectrometry. Journal of Chromatography A, 1070, 221-224.

HUANG, B., WANG, B., REN, D., JIN, W., LIU, J., PENG, J. & PAN, X. 2013. Occurrence, removal and bioaccumulation of steroid estrogens in Dianchi Lake catchment, China. Environment International, 59, 262-273.

HURWITZ, A. R. & LIU, S. T. 1977. Determination of aqueous solubility and pKa values of estrogens. J Pharm Sci, 66, 624-7.

ILLUMINA. 2011a. MiSeqTM Personal Sequencing System. Available at: https://www.illumina.com/Documents/seminars/presentations/2011_09_anderson.pdf [Accessed 17 Aug. 2017].

ILLUMINA. 2011b. Quality Scores for Next-Generation Sequencing. Technical Note: Sequencing. Available at https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf. [Accessed 26 July 2017]

IRIC. 2017. Understanding qPCR results. Genomic Platform Institute of research in immunology and cancer, University of Montreal. Available from: https://genomique.iric.ca/resources/files/Understanding_qPCR_results.pdf [Accessed 4 Aug. 2017]

JENSEN, R. A. 1976. Enzyme recruitment in evolution of new function. Annu Rev Microbiol, 30, 409-25.

JIANG, L., YANG, J. & CHEN, J. 2010. Isolation and characteristics of 17beta-estradiol-degrading Bacillus spp. strains from activated sludge. Biodegradation, 21, 729-36.

JIANG, W., YAN, Y., MA, M., WANG, D., LUO, Q., WANG, Z. & SATYANARAYANAN, S. K. 2012. Assessment of source water contamination by estrogenic disrupting compounds in China. J Environ Sci (China), 24, 320-8.

JOBLING, S., CASEY, D., ROGERS-GRAY, T., OEHLMANN, J., SCHULTE-OEHLMANN, U., PAWLOWSKI, S., BAUNBECK, T., TURNER, A. P. & TYLER, C. R. 2004. Comparative responses of molluscs and fish to environmental estrogens and an estrogenic effluent. Aquat Toxicol, 66, 207-22.

JOBLING, S., COEY, S., WHITMORE, J. G., KIME, D. E., VAN LOOK, K. J., MCALLISTER, B. G., BERESFORD, N., HENSHAW, A. C., BRIGHTY, G., TYLER, C. R. & SUMPTER, J. P. 2002. Wild intersex roach (Rutilus rutilus) have reduced fertility. Biol Reprod, 67, 515-24.

JOBLING, S. & OWEN, R. 2013. In Late lessons from early warnings: science, precaution, innovation. Late lessons II: Chapter 13 – Ethinyl oestradiol in the aquatic environment. Copenhagen: European Environment Agency. https://www.eea.europa.eu/publications/late-lessons-2 [Accessed 14 Sep. 2017].

JOHNSON, A. C., DUMONT, E., WILLIAMS, R. J., OLDENKAMP, R., CISOWSKA, I. & SUMPTER, J. P. 2013. Do concentrations of ethinylestradiol, estradiol, and diclofenac in European rivers exceed proposed EU environmental quality standards? Environ Sci Technol, 47, 12297-304.

JOHNSON, A. C. & WILLIAMS, R. J. 2004. A Model To Estimate Influent and Effluent Concentrations of Estradiol, Estrone, and Ethinylestradiol at Sewage Treatment Works. Environmental Science & Technology, 38, 3649-3658.

JOSEPH, L., BOATENG, L., R. V. FLORA, J., PARK, Y.-G., SON, A., BADAWY, M. & YOON, Y. 2013. Removal of bisphenol A and 17α-ethinyl estradiol by combined coagulation and adsorption using carbon nanomaterials and powdered activated carbon.

JOSS, A., ANDERSEN, H., TERNES, T., RICHLE, P. R. & SIEGRIST, H. 2004. Removal of estrogens in municipal wastewater treatment under aerobic and anaerobic conditions: consequences for plant optimization. Environ Sci Technol, 38, 3047-55.

JURGENS, M. D., HOLTHAUS, K. I., JOHNSON, A. C., SMITH, J. L., HETHERIDGE, M. & WILLIAMS, R. J. 2002. The potential for estradiol and ethinylestradiol degradation in English rivers. Environ Toxicol Chem, 21, 480-8.

KABASAKALIAN, P., BRITT, E. & YUDIS, M. D. 1966. Solubility or some steroids in water. Journal of Pharmaceutical Sciences, 55, 642-642.

KANDA, R. & CHURCHLEY, J. 2008. Removal of endocrine disrupting compounds during conventional wastewater treatment. Environ Technol, 29, 315-23.

KANEHISA, M., FURUMICHI, M., TANABE, M., SATO, Y. & MORISHIMA, K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res, 45, D353-d361.

KANEHISA, M. & GOTO, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 28, 27-30.

KANEHISA, M., SATO, Y., KAWASHIMA, M., FURUMICHI, M. & TANABE, M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res, 44, D457-62.

KAWAGUCHI, M., ISHII, Y., SAKUI, N., OKANOUCHI, N., ITO, R., INOUE, K., SAITO, K. & NAKAZAWA, H. 2004. Stir bar sorptive extraction with in situ derivatization and thermal desorption–gas chromatography–mass spectrometry in

the multi-shot mode for determination of estrogens in river water samples. Journal of Chromatography A, 1049, 1-8.

KE, J., ZHUANG, W., GIN, K. Y., REINHARD, M., HOON, L. T. & TAY, J. H. 2007. Characterization of estrogen-degrading bacteria isolated from an artificial sandy aquifer with ultrafiltered secondary effluent as the medium. Appl Microbiol Biotechnol, 75, 1163-71.

KIDD, K. A., BLANCHFIELD, P. J., MILLS, K. H., PALACE, V. P., EVANS, R. E., LAZORCHAK, J. M. & FLICK, R. W. 2007. Collapse of a fish population after exposure to a synthetic estrogen. Proc Natl Acad Sci U S A, 104, 8897-901.

KIM, S.-E., YAMADA, H. & TSUNO, H. 2004. Evaluation of Estrogenicity for 17 β-Estradiol Decomposition During Ozonation. Ozone: Science & Engineering, 26, 563-571.

KIM, Y. U., HAN, J., LEE, S. S., SHIMIZU, K., TSUTSUMI, Y. & KONDO, R. 2007. Steroid 9alpha-hydroxylation during testosterone degradation by resting rhodococcus equi cells. Arch Pharm (Weinheim), 340, 209-14.

KISIELA, M., SKARKA, A., EBERT, B. & MASER, E. 2012. Hydroxysteroid dehydrogenases (HSDs) in bacteria: a bioinformatic perspective. J Steroid Biochem Mol Biol, 129, 31-46.

KJÆR, J., OLSEN, P., BACH, K., BARLEBO, H. C., INGERSLEV, F., HANSEN, M. & SØRENSEN, B. H. 2007. Leaching of Estrogenic Hormones from Manure-Treated Structured Soils. Environmental Science & Technology, 41, 3911-3917.

KOZERA, B. & RAPACZ, M. 2013. Reference genes in real-time PCR. J Appl Genet, 54, 391-406.

KUCH, H. M. & BALLSCHMITER, K. 2000. Determination of endogenous and exogenous estrogens in effluents from sewage treatment plants at the ng/L-level. Fresenius J Anal Chem, 366, 392-5.

KURISU, F., OGURA, M., SAITOH, S., YAMAZOE, A. & YAGI, O. 2010. Degradation of natural estrogen and identification of the metabolites produced by soil isolates of Rhodococcus sp. and Sphingomonas sp. J Biosci Bioeng, 109, 576-82.

LABADIE, P. & BUDZINSKI, H. 2005. Determination of steroidal hormone profiles along the Jalle d'Eysines River (near Bordeaux, France). Environ Sci Technol, 39, 5113-20.

LAI, E. P. C., MALEKI, Z. D. & WU, S. 2010. Characterization of molecularly imprinted and nonimprinted polymer submicron particles specifically tailored for removal of trace 17β-estradiol in water treatment. Journal of Applied Polymer Science, 116, 1499-1508.

LAI, K. M., JOHNSON, K. L., SCRIMSHAW, M. D. & LESTER, J. N. 2000. Binding of Waterborne Steroid Estrogens to Solid Phases in River and Estuarine Systems. Environmental Science & Technology, 34, 3890-3894.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods, 9, 357-9.

LARCHER, S. & YARGEAU, V. 2013. Biodegradation of 17alpha-ethinylestradiol by heterotrophic bacteria. Environ Pollut, 173, 17-22.

LAWTON, E. J., BUTWELL, A. J., GARDNER, M., JOHNSON, I. & ROCKETT, L. 2008. Removal of endocrine disrupting chemicals in sewage treatment – The National Demonstration. 20th SETAC conference. Seville, Spain.

LEE, H. B. & LIU, D. 2002. Degradation of 17β-Estradiol and its Metabolites by Sewage Bacteria. Water, Air, and Soil Pollution, 134, 351-366.

LEE, L. S., STROCK, T. J., SARMAH, A. K. & RAO, P. S. 2003. Sorption and dissipation of testosterone, estrogens, and their primary transformation products in soils and sediment. Environ Sci Technol, 37, 4098-105.

LEI, B., HUANG, S., ZHOU, Y., WANG, D. & WANG, Z. 2009. Levels of six estrogens in water and sediment from three rivers in Tianjin area, China. Chemosphere, 76, 36-42.

LETEK, M., GONZALEZ, P., MACARTHUR, I., RODRIGUEZ, H., FREEMAN, T. C., VALERO-RELLO, A., BLANCO, M., BUCKLEY, T., CHEREVACH, I., FAHEY, R., HAPESHI, A., HOLDSTOCK, J., LEADON, D., NAVAS, J., OCAMPO, A., QUAIL, M. A., SANDERS, M., SCORTTI, M. M., PRESCOTT, J. F., FOGARTY, U., MEIJER, W. G., PARKHILL, J., BENTLEY, S. D. & VAZQUEZ-BOLAND, J. A. 2010. The genome

of a pathogenic rhodococcus: cooptive virulence underpinned by key gene acquisitions. PLoS Genet, 6, e1001145.

LI, L. 2015. Chapter 7: Molecular Microecological Techniques. In: Infectious Microecology: Theory and Applications. Page 178.

LI, S., LIU, J., SUN, M., LING, W. & ZHU, X. 2017. Isolation, Characterization, and Degradation Performance of the 17beta-Estradiol-Degrading Bacterium Novosphingobium sp. E2S. Int J Environ Res Public Health, 14.

LIANG, R., LIU, H., TAO, F., LIU, Y., MA, C., LIU, X. & LIU, J. 2012. Genome sequence of Pseudomonas putida strain SJTE-1, a bacterium capable of degrading estrogens and persistent organic pollutants. J Bacteriol, 194, 4781-2.

LISHMAN, L., SMYTH, S. A., SARAFIN, K., KLEYWEGT, S., TOITO, J., PEART, T., LEE, B., SERVOS, M., BELAND, M. & SETO, P. 2006. Occurrence and reductions of pharmaceuticals and personal care products and estrogens by municipal wastewater treatment plants in Ontario, Canada. Sci Total Environ, 367, 544-58.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012. Comparison of Next-Generation Sequencing Systems. Journal of Biomedicine and Biotechnology, 2012, 11.

LIU, Z. H., KANJO, Y. & MIZUTANI, S. 2009. Removal mechanisms for endocrine disrupting compounds (EDCs) in wastewater treatment - physical means, biodegradation, and chemical advanced oxidation: a review. Sci Total Environ, 407, 731-48.

LIU, Z.-H., KANJO, Y. & MIZUTANI, S. 2009. Urinary excretion rates of natural estrogens and androgens from humans, and their occurrence and fate in the environment: A review. Science of The Total Environment, 407, 4975-4985.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods, 25, 402-8.

LOMAN, N. J., CONSTANTINIDOU, C., CHAN, J. Z., HALACHEV, M., SERGEANT, M., PENN, C. W., ROBINSON, E. R. & PALLEN, M. J. 2012. High-throughput

bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol, 10, 599-606.

LONG, M., STRAND, J., LASSEN, P., KRUGER, T., DAHLLOF, I., BOSSI, R., LARSEN, M. M., WIBERG-LARSEN, P. & BONEFELD-JORGENSEN, E. C. 2014. Endocrine-disrupting effects of compounds in Danish streams. Arch Environ Contam Toxicol, 66, 1-18.

LU, G., YAN, Z., WANG, Y. & CHEN, W. 2011. Assessment of estrogenic contamination and biological effects in Lake Taihu. Ecotoxicology, 20, 974-81.

LU, G. H., SONG, W. T., WANG, C. & YAN, Z. H. 2010. Assessment of in vivo estrogenic response and the identification of environmental estrogens in the Yangtze River (Nanjing section). Chemosphere, 80, 982-90.

LUCAS, S. D. & JONES, D. L. 2006. Biodegradation of estrone and 17 β-estradiol in grassland soils amended with animal wastes. Soil Biology and Biochemistry, 38, 2803-2815.

MA, C., QIN, D., SUN, Q., ZHANG, F., LIU, H. & YU, C. P. 2016. Removal of environmental estrogens by bacterial cell immobilization technique. Chemosphere, 144, 607-14.

MAGOC, T., PABINGER, S., CANZAR, S., LIU, X., SU, Q., PUIU, D., TALLON, L. J. & SALZBERG, S. L. 2013. GAGE-B: an evaluation of genome assemblers for bacterial organisms. Bioinformatics, 29, 1718-25.

MANSON, J. E., CHLEBOWSKI, R. T., STEFANICK, M. L., ARAGAKI, A. K., ROSSOUW, J. E., PRENTICE, R. L., ANDERSON, G., HOWARD, B. V., THOMSON, C. A., LACROIX, A. Z., WACTAWSKI-WENDE, J., JACKSON, R. D., LIMACHER, M., MARGOLIS, K. L., WASSERTHEIL-SMOLLER, S., BERESFORD, S. A., CAULEY, J. A., EATON, C. B., GASS, M., HSIA, J., JOHNSON, K. C., KOOPERBERG, C., KULLER, L. H., LEWIS, C. E., LIU, S., MARTIN, L. W., OCKENE, J. K., O'SULLIVAN, M. J., POWELL, L. H., SIMON, M. S., VAN HORN, L., VITOLINS, M. Z. & WALLACE, R. B. 2013. Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. JAMA, 310, 1353-68.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437, 376-80.

MATOZZO, V., GAGNE, F., MARIN, M. G., RICCIARDI, F. & BLAISE, C. 2008. Vitellogenin as a biomarker of exposure to estrogenic compounds in aquatic invertebrates: a review. Environ Int, 34, 531-45.

MATSUOKA, S., KIKUCHI, M., KIMURA, S., KUROKAWA, Y. & KAWAI, S. I. 2005. Determination of Estrogenic Substances in the Water of Muko River Using <I>in Vitro</I> Assays, and the Degradation of Natural Estrogens by Aquatic Bacteria. Journal of Health Science, 51, 178-184.

MCKERNAN, K., BLANCHARD, A., KOTLER, L. & COSTA, G. 2006. Reagents, methods, and libraries for bead-based sequencing. US patent application 20080003571

MCLEOD, M. P., WARREN, R. L., HSIAO, W. W., ARAKI, N., MYHRE, M., FERNANDES, C., MIYAZAWA, D., WONG, W., LILLQUIST, A. L., WANG, D., DOSANJH, M., HARA, H., PETRESCU, A., MORIN, R. D., YANG, G., STOTT, J. M., SCHEIN, J. E., SHIN, H., SMAILUS, D., SIDDIQUI, A. S., MARRA, M. A., JONES, S. J., HOLT, R., BRINKMAN, F. S., MIYAUCHI, K., FUKUDA, M., DAVIES, J. E., MOHN, W. W. & ELTIS, L. D. 2006. The complete genome of Rhodococcus sp. RHA1 provides insights into a catabolic powerhouse. Proc Natl Acad Sci U S A, 103, 15582-7.

MENG, Z., CHEN, W. & MULCHANDANI, A. 2005. Removal of estrogenic pollutants from contaminated water using molecularly imprinted polymers. Environ Sci Technol, 39, 8958-62.

MEYNET, P., HEAD, I. M., WERNER, D. & DAVENPORT, R. J. 2015. Re-evaluation of dioxygenase gene phylogeny for the development and validation of a quantitative assay for environmental aromatic hydrocarbon degraders. FEMS Microbiol Ecol, 91.

MINITAB 17 STATISTICAL SOFTWARE. 2010. [Computer software]. State College, PA: Minitab, Inc. (www.minitab.com)

MOLINA-HENARES, A. J., KRELL, T., EUGENIA GUAZZARONI, M., SEGURA, A. & RAMOS, J. L. 2006. Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. FEMS Microbiol Rev, 30, 157-86.

MULLER, M., PATUREAU, D., GODON, J. J., DELGENES, J. P. & HERNANDEZ-RAQUET, G. 2010. Molecular and kinetic characterization of mixed cultures degrading natural and synthetic estrogens. Appl Microbiol Biotechnol, 85, 691-701.

MULLER, M., RABENOELINA, F., BALAGUER, P., PATUREAU, D., LEMENACH, K., BUDZINSKI, H., BARCELO, D., DE ALDA, M. L., KUSTER, M., DELGENES, J. P. & HERNANDEZ-RAQUET, G. 2008. Chemical and biological analysis of endocrine-disrupting hormones and estrogenic activity in an advanced sewage treatment plant. Environ Toxicol Chem, 27, 1649-58.

NAGARAJAN, N. & POP, M. 2013. Sequence assembly demystified. Nat Rev Genet, 14, 157-167.

NAKAI, S., YAMAMURA, A., TANAKA, S., SHI, J., NISHIKAWA, M., NAKASHIMADA, Y. & HOSOMI, M. 2011. Pathway of 17β-estradiol degradation by Nitrosomonas europaea and reduction in 17β-estradiol-derived estrogenic activity. Environmental Chemistry Letters, 9, 1-6.

NEALE, P. A., MASTRUP, M., BORGMANN, T. & SCHAFER, A. I. 2010. Sorption of micropollutant estrone to a water treatment ion exchange resin. J Environ Monit, 12, 311-7.

NELSON, L. R. & BULUN, S. E. 2001. Estrogen production and action. J Am Acad Dermatol, 45, S116-24.

NHS (2016) Statistics on Sexual and Reproductive Health Services 2015/2016. http://www.content.digital.nhs.uk/catalogue/PUB21969/srh-serv-eng-15-16-rep.pdf [Accessed 8 Aug. 2017].

NIETO, A., BORRULL, F., POCURULL, E. & MARCE, R. M. 2008. Determination of natural and synthetic estrogens and their conjugates in sewage sludge by pressurized liquid extraction and liquid chromatography-tandem mass spectrometry. J Chromatogr A, 1213, 224-30.

NISHIDA, C. R., EVERETT, S. & ORTIZ DE MONTELLANO, P. R. 2013. Specificity determinants of CYP1B1 estradiol hydroxylation. Mol Pharmacol, 84, 451-8.

NOPPE, H., VERSLYCKE, T., DE WULF, E., VERHEYDEN, K., MONTEYNE, E., VAN CAETER, P., JANSSEN, C. R. & DE BRABANDER, H. F. 2007. Occurrence of estrogens in the Scheldt estuary: a 2-year survey. Ecotoxicol Environ Saf, 66, 1-8.

O'GRADY, D., EVANGELISTA, S. & YARGEAU, V. 2009. Removal of Aqueous 17α-Ethinylestradiol by Rhodococcus Species. Environmental Engineering Science, 26, 1393-1400.

OJANOTKO-HARRI, A., NIKKARI, T., HARRI, M. P. & PAUNIO, K. U. 1990. Metabolism of progesterone and testosterone by Bacillus cereus strain Socransky 67 and Streptococcus mutans strain Ingbritt. Oral Microbiol Immunol, 5, 237-9.

OLIVEROS, J. C. 2007-2015. Venny. An interactive tool for comparing lists with Venny's diagrams. Available at: http://bioinfogp.cnb.csic.es/tools/venny/index.html [Accessed 27 Aug. 2017].

OVERBEEK, R., BEGLEY, T., BUTLER, R. M., CHOUDHURI, J. V., CHUANG, H. Y., COHOON, M., DE CRECY-LAGARD, V., DIAZ, N., DISZ, T., EDWARDS, R., FONSTEIN, M., FRANK, E. D., GERDES, S., GLASS, E. M., GOESMANN, A., HANSON, A., IWATA-REUYL, D., JENSEN, R., JAMSHIDI, N., KRAUSE, L., KUBAL, M., LARSEN, N., LINKE, B., MCHARDY, A. C., MEYER, F., NEUWEGER, H., OLSEN, G., OLSON, R., OSTERMAN, A., PORTNOY, V., PUSCH, G. D., RODIONOV, D. A., RUCKERT, C., STEINER, J., STEVENS, R., THIELE, I., VASSIEVA, O., YE, Y., ZAGNITKO, O. & VONSTEIN, V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res, 33, 5691-702.

PARK, J.-S., PARK, Y.-J., KANG, H. W. & KIM, C.-K. 2008. Solvent effects on physicochemical behavior of estradiols recrystallized for transdermal delivery. Archives of Pharmacal Research, 31, 111-116.

PAUWELS, B., NOPPE, H., DE BRABANDER, H. & VERSTRAETE, W. 2008. Comparison of Steroid Hormone Concentrations in Domestic and Hospital Wastewater Treatment Plants. Journal of Environmental Engineering, 134, 933-936.

PAUWELS, B., WILLE, K., NOPPE, H., DE BRABANDER, H., VAN DE WIELE, T., VERSTRAETE, W. & BOON, N. 2008. 17alpha-ethinylestradiol cometabolism by bacteria degrading estrone, 17beta-estradiol and estriol. Biodegradation, 19, 683-93.

PECK, M., GIBSON, R. W., KORTENKAMP, A. & HILL, E. M. 2004. Sediments are major sinks of steroidal estrogens in two United Kingdom rivers. Environ Toxicol Chem, 23, 945-52.

PENG, R. H., XIONG, A. S., XUE, Y., FU, X. Y., GAO, F., ZHAO, W., TIAN, Y. S. & YAO, Q. H. 2008. Microbial biodegradation of polyaromatic hydrocarbons. FEMS Microbiol Rev, 32, 927-55.

PENG, R.-H., XIONG, A.-S., XUE, Y., FU, X.-Y., GAO, F., ZHAO, W., TIAN, Y.-S. & YAO, Q.-H. 2008. Microbial biodegradation of polyaromatic hydrocarbons. FEMS Microbiology Reviews, 32, 927-955.

PITT, J. J. 2009. Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. The Clinical Biochemist Reviews, 30, 19-34.

PORTE, C., JANER, G., LORUSSO, L. C., ORTIZ-ZARRAGOITIA, M., CAJARAVILLE, M. P., FOSSI, M. C. & CANESI, L. 2006. Endocrine disruptors in marine organisms: approaches and perspectives. Comp Biochem Physiol C Toxicol Pharmacol, 143, 303-15.

PRESCRIBING AND MEDICINES TEAM HEALTH AND SOCIAL CARE INFORMATION CENTRE (2016). Prescriptions Dispensed in the Community England 205-2015. http://content.digital.nhs.uk/catalogue/PUB20664/pres-disp-com-eng-2005-15-rep.pdf [Accessed 8 Aug. 2017].

PRIMERDESIGN. 2016a. geNormTM and geNormPLUSTM Reference Gene Selection Kit: geNorm primer only kit Handbook. Available from:

http://www.primerdesign.co.uk/assets/files/genorm_sybrgreen_handbook.pdf?timestamp=1503952745 [Accessed 28 Aug 2017].

PRIMERDESIGN. 2016b. Custom designed real-time PCR assay: Handbook HB08.03.03. Available from: http://www.primerdesign.co.uk/assets/files/primer_only_handbook.pdf?timestamp=1503952717 [Accessed 28 Aug 2017].

PURDOM, C. E., HARDIMAN, P. A., BYE, V. V. J., ENO, N. C., TYLER, C. R. & SUMPTER, J. P. 1994. Estrogenic Effects of Effluents from Sewage Treatment Works. Chemistry and Ecology, 8, 275-285.

RAO, K., LEI, B., LI, N., MA, M. & WANG, Z. 2013. Determination of estrogens and estrogenic activities in water from three rivers in Tianjin, China. J Environ Sci (China), 25, 1164-71.

RASHED, A. N., HSIA, Y., WILTON, L., ZILLER, M., KOSTEV, K. & TOMLIN, S. 2015. Trends and patterns of hormonal contraceptive prescribing for adolescents in primary care in the U.K. J Fam Plann Reprod Health Care, 41, 216-22.

RASHID, A., WHITE, E. T., HOWES, T., LITSTER, J. D. & MARZIANO, I. 2014. Effect of Solvent Composition and Temperature on the Solubility of Ibuprofen in Aqueous Ethanol. Journal of Chemical & Engineering Data, 59, 2699-2703.

RAY, S., JOHNSTON, R., CAMPBELL, D. C., NUGENT, S., MCDADE, S. S., WAUGH, D. & PANOV, K. I. 2013. Androgens and estrogens stimulate ribosome biogenesis in prostate and breast cancer cells in receptor dependent manner. Gene, 526, 46-53.

REN, Y. X., NAKANO, K., NOMURA, M., CHIBA, N. & NISHIMURA, O. 2007. A thermodynamic analysis on adsorption of estrogens in activated sludge process. Water Res, 41, 2341-8.

RESTREPO, L., BAYOT, B., BETANCOURT, I. & PINZON, A. 2016. Draft genome sequence of pathogenic bacteria Vibrio parahaemolyticus strain Ba94C2, associated with acute hepatopancreatic necrosis disease isolate from South America. Genom Data, 9, 143-4.

ROCHA, A. J., MIRANDA, R. S., SOUSA, A. J. S. & DA SILVA, A. L. C. 2016. Guidelines for Successful Quantitative Gene Expression in Real- Time qPCR Assays, Polymerase Chain Reaction for Biomedical Applications, (Ed.), InTech, DOI: 10.5772/65850. Available from: https://www.intechopen.com/books/polymerase-chain-reaction-for-biomedical-applications/guidelines-for-successful-quantitative-gene-expression-in-real-time-qpcr-assays [Accessed 15 Sep. 2017]

ROCHA, E. R. & SMITH, C. J. 1999. Role of the alkyl hydroperoxide reductase (ahpCF) gene in oxidative stress defense of the obligate Anaerobe bacteroides fragilis. J Bacteriol, 181, 5701-10.

RODGERS-GRAY, T. P., JOBLING, S., KELLY, C., MORRIS, S., BRIGHTY, G., WALDOCK, M. J., SUMPTER, J. P. & TYLER, C. R. 2001. Exposure of juvenile roach (Rutilus rutilus) to treated sewage effluent induces dose-dependent and persistent disruption in gonadal duct development. Environ Sci Technol, 35, 462-70.

ROH, H. & CHU, K. H. 2010. A 17beta-estradiol-utilizing bacterium, Sphingomonas strain KC8: part I - characterization and abundance in wastewater treatment plants. Environ Sci Technol, 44, 4943-50.

ROTHBERG, J. M. & LEAMON, J. H. 2008. The development and impact of 454 sequencing. Nat Biotechnol, 26, 1117-24.

ROUCOURT, B., MINNEBO, N., AUGUSTIJNS, P., HERTVELDT, K., VOLCKAERT, G. & LAVIGNE, R. 2009. Biochemical characterization of malate synthase G of P. aeruginosa. BMC Biochem, 10, 20.

RUCHELMAN, M. W. 1967. Solubility studies of estrone in organic solvents using gas-liquid chromatography. Anal Biochem, 19, 98-108.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M.-A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. Bioinformatics, 16, 944-945.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and

annotation. [Computer Software]. Bioinformatics (Oxford, England) 16;10;944-5 PUBMED: 11120685

SABIROVA, J. S., CLOETENS, L. F., VANHAECKE, L., FORREZ, I., VERSTRAETE, W. & BOON, N. 2008. Manganese-oxidizing bacteria mediate the degradation of 17alpha-ethinylestradiol. Microb Biotechnol, 1, 507-12.

SANG, Y., XIONG, G. & MASER, E. 2011. Steroid degradation and two steroid-inducible enzymes in the marine bacterium H5. Chem Biol Interact, 191, 89-94.

SANG, Y., XIONG, G. & MASER, E. 2012. Identification of a new steroid degrading bacterial strain H5 from the Baltic Sea and isolation of two estradiol inducible genes. J Steroid Biochem Mol Biol, 129, 22-30.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A, 74, 5463-7.

SARMAH, A. K. & NORTHCOTT, G. L. 2008. Laboratory degradation studies of four endocrine disruptors in two environmental media. Environ Toxicol Chem, 27, 819-27.

SASAKI, M., TSUCHIDO, T. & MATSUMURA, Y. 2008. Molecular cloning and characterization of cytochrome P450 and ferredoxin genes involved in bisphenol A degradation in Sphingomonas bisphenolicum strain AO1. J Appl Microbiol, 105, 1158-69.

SCHAFER, A. I., AKANYETI, I. & SEMIAO, A. J. 2011. Micropollutant sorption to membrane polymers: a review of mechanisms for estrogens. Adv Colloid Interface Sci, 164, 100-17.

SCHICKSNUS, T. & MÜLLER-GOYMANN, C. 2000. 7β-Estradiol Solubility In Aqueous Systems – Influence Of Ionic Strength, pH and Adsorption To Packaging Material.

SCHULSTER, M., BERNIE, A. M. & RAMASAMY, R. 2016. The role of estradiol in male reproductive function. Asian J Androl, 18, 435-40.

SCORNAIENCHI, M. L., THORNTON, C., WILLETT, K. L. & WILSON, J. Y. 2010. Cytochrome P450-mediated 17beta-estradiol metabolism in zebrafish (Danio rerio). J Endocrinol, 206, 317-25.

SCOTT, D. & ELY, B. 2015. Comparison of genome sequencing technology and assembly methods for the analysis of a GC-rich bacterial genome. Curr Microbiol, 70, 338-44.

SEGNER, H., NAVAS, J. M., SCHAFERS, C. & WENZEL, A. 2003. Potencies of estrogenic compounds in in vitro screening assays and in life cycle tests with zebrafish in vivo. Ecotoxicol Environ Saf, 54, 315-22.

SHAREEF, A., ANGOVE, M. J., WELLS, J. D. & JOHNSON, B. B. 2006. Aqueous Solubilities of Estrone, 17β-Estradiol, 17α-Ethynylestradiol, and Bisphenol A. Journal of Chemical & Engineering Data, 51, 879-881.

SHEN, J. H., GUTENDORF, B., VAHL, H. H., SHEN, L. & WESTENDORF, J. 2001. Toxicological profile of pollutants in surface water from an area in Taihu Lake, Yangtze Delta. Toxicology, 166, 71-8.

SHENDURE, J. & JI, H. 2008. Next-generation DNA sequencing. Nat Biotech, 26, 1135-1145.

SHENDURE, J., PORRECA, G. J., REPPAS, N. B., LIN, X., MCCUTCHEON, J. P., ROSENBAUM, A. M., WANG, M. D., ZHANG, K., MITRA, R. D. & CHURCH, G. M. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. Science, 309, 1728-32.

SHI, J., FUJISAWA, S., NAKAI, S. & HOSOMI, M. 2004. Biodegradation of natural and synthetic estrogens by nitrifying activated sludge and ammonia-oxidizing bacterium Nitrosomonas europaea. Water Res, 38, 2322-9.

SHUTT, D. A., SMITH, I. D. & SHEARMAN, R. P. 1974. OESTRONE, OESTRADIOL-17β AND OESTRIOL LEVELS IN HUMAN FOETAL PLASMA DURING GESTATION AND AT TERM. Journal of Endocrinology, 60, 333-341.

SIH, C. J., LEE, S. S., TSONG, Y. Y. & WANG, K. C. 1966. Mechanisms of steroid oxidation by microorganisms. 8. 3,4-Dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione, an intermediate in the microbiological degradation of ring A of androst-4-ene-3,17-dione. J Biol Chem, 241, 540-50.

SIH, C. J., WANG, K. C., GIBSON, D. T. & WHITLOCK, H. W., JR. 1965. ON THE MECHANISM OF RING A CLEAVAGE IN THE DEGRADATION OF 9,10-SECO STEROIDS BY MICROORGANISMS. J Am Chem Soc, 87, 1386-7.

SILVA, C. P., LIMA, D. L. D., SCHNEIDER, R. J., OTERO, M. & ESTEVES, V. I. 2013. Development of ELISA methodologies for the direct determination of 17β-estradiol and 17α-ethinylestradiol in complex aqueous matrices. Journal of Environmental Management, 124, 121-127.

SILVA, C. P., OTERO, M. & ESTEVES, V. 2012. Processes for the elimination of estrogenic steroid hormones from water: a review. Environ Pollut, 165, 38-58.

SIMONI, R. D., HILL, R. L. & VAUGHAN, M. 2002. The Discovery of Estrone, Estriol, and Estradiol and the Biochemical Study of Reproduction. The Work of Edward Adelbert Doisy. Journal of Biological Chemistry, 277, e17.

SIMPSON, E. R., MAHENDROO, M. S., MEANS, G. D., KILGORE, M. W., HINSHELWOOD, M. M., GRAHAM-LORENCE, S., AMARNEH, B., ITO, Y., FISHER, C. R., MICHAEL, M. D., MENDELSON, C. R. & BULUN, S. E. 1994. Aromatase Cytochrome P450, The Enzyme Responsible for Estrogen Biosynthesis*. Endocrine Reviews, 15, 342-355.

SKOTNICKA-PITAK, J., KHUNJAR, W. O., LOVE, N. G. & AGA, D. S. 2009. Characterization of Metabolites Formed During the Biotransformation of 17α-Ethinylestradiol by Nitrosomonas europaea in Batch and Continuous Flow Bioreactors. Environmental Science & Technology, 43, 3549-3555.

SNYDER, S. A., ADHAM, S., REDDING, A. M., CANNON, F. S., DECAROLIS, J., OPPENHEIMER, J., WERT, E. C. & YOON, Y. 2007. Role of membranes and activated carbon in the removal of endocrine disruptors and pharmaceuticals. Desalination, 202, 156-181.

SONG, X., XU, Y., LI, G., ZHANG, Y., HUANG, T. & HU, Z. 2011. Isolation, characterization of Rhodococcus sp. P14 capable of degrading high-molecular-weight polycyclic aromatic hydrocarbons and aliphatic hydrocarbons. Marine Pollution Bulletin, 62, 2122-2128.

SOUZA, M. S., HALLGREN, P., BALSEIRO, E. & HANSSON, L. A. 2013. Low concentrations, potential ecological consequences: synthetic estrogens alter life-history and demographic structures of aquatic invertebrates. Environ Pollut, 178, 237-43.

SUMPTER, J. P. & JOBLING, S. 1995. Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. Environ Health Perspect, 103 Suppl 7, 173-8.

SUN, L., YONG, W., CHU, X. & LIN, J. M. 2009. Simultaneous determination of 15 steroidal oral contraceptives in water using solid-phase disk extraction followed by high performance liquid chromatography-tandem mass spectrometry. J Chromatogr A, 1216, 5416-23.

SVENSON, A., SJOHOLM, S., ALLARD, A. S. & KAJ, L. 2011. Antiestrogenicity and estrogenicity in leachates from solid waste deposits. Environ Toxicol, 26, 233-9.

TABAK, H. H., BLOOMHUFF, R. N. & BUNCH, R. L. 1981. Steroid hormones as water pollutants II. Studies on the persistence and stability of natural urinary and synthetic ovulation-inhibiting hormones in untreated and treated wastewaters. Dev. Ind. Microbiol, 22, 497−519

TERNES, T. A., ANDERSEN, H., GILBERG, D. & BONERZ, M. 2002. Determination of estrogens in sludge and sediments by liquid extraction and GC/MS/MS. Anal Chem, 74, 3498-504.

TERNES, T. A., HERRMANN, N., BONERZ, M., KNACKER, T., SIEGRIST, H. & JOSS, A. 2004. A rapid method to measure the solid-water distribution coefficient (Kd) for pharmaceuticals and musk fragrances in sewage sludge. Water Res, 38, 4075-84.

TERNES, T. A., STUMPF, M., MUELLER, J., HABERER, K., WILKEN, R. D. & SERVOS, M. 1999. Behavior and occurrence of estrogens in municipal sewage treatment plants — I. Investigations in Germany, Canada and Brazil. Science of The Total Environment, 225, 81-90.

THAYANUKUL, P., ZANG, K., JANHOM, T., KURISU, F., KASUGA, I. & FURUMAI, H. 2010. Concentration-dependent response of estrone-degrading bacterial

community in activated sludge analyzed by microautoradiography-fluorescence in situ hybridization. Water Res, 44, 4878-87.

THERMOFISHER SCIENTIFIC (2018). Is Your RNA Intact? Methods to Check RNA Integrity. https://www.thermofisher.com/uk/en/home/references/ambion-tech-support/rna-isolation/tech-notes/is-your-rna-intact.html. [Accessed 13 May 2018].

THOMAS, M. P. & POTTER, B. V. 2013. The structural biology of oestrogen metabolism. J Steroid Biochem Mol Biol, 137, 27-49.

TIMP, W., MIRSAIDOV, U. M., WANG, D., COMER, J., AKSIMENTIEV, A. & TIMP, G. 2010. Nanopore Sequencing: Electrical Measurements of the Code of Life. IEEE transactions on nanotechnology, 9, 281-294.

TOSHIKAZU, H., KENSUKE, O., TOMOYUKI, M., BEGUM SHAILA, L., KEITAROU, T., AKIKO, M. & TAKAO, M. 2010. Contribution of the Estrogen-Degrading Bacterium Novosphingobium sp. Strain JEM-1 to Estrogen Removal in Wastewater Treatment. Journal of Environmental Engineering, 136, 890-896.

UTTURKAR, S. M., KLINGEMAN, D. M., LAND, M. L., SCHADT, C. W., DOKTYCZ, M. J., PELLETIER, D. A. & BROWN, S. D. 2014. Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences. Bioinformatics, 30, 2709-16.

UNITED KINGDOM WATER INDUSTRY RESEARCH (UKWIR). 2009. Endocrine Disrupting Chemicals: Assessment of the Performance of Wastewater Treatment Works (WwTW) in Removing Oestrogenic Substances. www.ukwir.org.uk [Accessed 28 Aug. 2017].

UNITED NATIONS (2007) World contraceptive use. http://www.un.org/esa/population/publications/contraceptive2007/contraceptive2007.htm [Accessed 14 Sept. 2017]

VAN DER GEIZE, R., YAM, K., HEUSER, T., WILBRINK, M. H., HARA, H., ANDERTON, M. C., SIM, E., DIJKHUIZEN, L., DAVIES, J. E., MOHN, W. W. & ELTIS, L. D. 2007. A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into Mycobacterium tuberculosis survival in macrophages. Proc Natl Acad Sci U S A, 104, 1947-52.

VANDESOMPELE, J., DE PRETER, K., PATTYN, F., POPPE, B., VAN ROY, N., DE PAEPE, A. & SPELEMAN, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol, 3, Research0034.

VERONEZ, I. P., DANIEL, J. S. P., JÚNIOR, C. E. C., GARCIA, J. S. & TREVISAN, M. G. 2015. Development, characterization, and stability studies of ethinyl estradiol solid dispersion. Journal of Thermal Analysis and Calorimetry, 120, 573-581.

VERSLYCKE, T., VANDENBERGH, G. F., VERSONNEN, B., ARIJS, K. & JANSSEN, C. R. 2002. Induction of vitellogenesis in 17alpha-ethinylestradiol-exposed rainbow trout (Oncorhynchus mykiss): a method comparison. Comp Biochem Physiol C Toxicol Pharmacol, 132, 483-92.

VETHAAK, A. D., LAHR, J., SCHRAP, S. M., BELFROID, A. C., RIJS, G. B., GERRITSEN, A., DE BOER, J., BULDER, A. S., GRINWIS, G. C., KUIPER, R. V., LEGLER, J., MURK, T. A., PEIJNENBURG, W., VERHAAR, H. J. & DE VOOGT, P. 2005. An integrated assessment of estrogenic contamination and biological effects in the aquatic environment of The Netherlands. Chemosphere, 59, 511-24.

WALKER, C. W. & WATSON, J. E. 2010. Adsorption of estrogens on laboratory materials and filters during sample preparation. J Environ Qual, 39, 744-8.

WANG, B., SANCHEZ, R. I., FRANKLIN, R. B., EVANS, D. C. & HUSKEY, S. E. 2004. The involvement of CYP3A4 and CYP2C9 in the metabolism of 17 alpha-ethinylestradiol. Drug Metab Dispos, 32, 1209-12.

WANG, K. C. & SIH, C. J. 1963. MECHANISMS OF STEROID OXIDATION BY MICROORGANISMS. IV. SECO INTERMEDIATES. Biochemistry, 2, 1238-43.

WANG, W., REN, S., ZHANG, H., YU, J., AN, W., HU, J. & YANG, M. 2011. Occurrence of nine nitrosamines and secondary amines in source water and drinking water: Potential of secondary amines as nitrosamine precursors. Water Research, 45, 4930-4938.

WATER QUALITY. 2013. Written evidence submitted by Thames Water (WQ07) https://publications.parliament.uk/pa/cm201213/cmselect/cmsctech/writev/932/wq07.htm [Accessed 6 Aug. 2017]

WATERS. 2017. How Does High Performance Liquid Chromatography Work? Available from: http://www.waters.com/waters/en_GB/How-Does-High-Performance-Liquid-Chromatography-Work%3F/nav.htm?cid=10049055&locale=en_GB [Accessed 7 Sep. 2017]

WEBER, S., GALLENKEMPER, M., MELIN, T., DOTT, W. & HOLLENDER, J. 2004. Efficiency of nanofiltration for the elimination of steroids from water. Water Sci Technol, 50, 9-14.

WEBER, S., LEUSCHNER, P., KAMPFER, P., DOTT, W. & HOLLENDER, J. 2005. Degradation of estradiol and ethinyl estradiol by activated sludge and by a defined mixed culture. Appl Microbiol Biotechnol, 67, 106-12.

WHITE, P 2004. RNA extraction from Mammalian Tissue- Traditional Method. Functional Genomics Core, University of Pennsylvania. Available from: http://www.cbil.upenn.edu/downloads_orig/EPConDB/download/Protocols/RNA%20Extraction%20from%20mammalian%20tissue%20-%20Traditional%20Method.pdf [Accessed 15 Sep. 2017].

WILLIAMS, R. J., JOHNSON, A. C., SMITH, J. J. L. & KANDA, R. 2003. Steroid Estrogens Profiles along River Stretches Arising from Sewage Treatment Works Discharges. Environmental Science & Technology, 37, 1744-1750.

WISE, A., O'BRIEN, K. & WOODRUFF, T. 2011. Are oral contraceptives a significant contributor to the estrogenicity of drinking water? Environ Sci Technol, 45, 51-60.

WORLD HEALTH ORGANISATION. (2017). Family planning/Contraception. [online] Available at: http://www.who.int/mediacentre/factsheets/fs351/en/ [Accessed 14 Sep. 2017].

XU, X., VEENSTRA, T. D., FOX, S. D., ROMAN, J. M., ISSAQ, H. J., FALK, R., SAAVEDRA, J. E., KEEFER, L. K. & ZIEGLER, R. G. 2005. Measuring fifteen endogenous estrogens simultaneously in human urine by high-performance liquid chromatography-mass spectrometry. Anal Chem, 77, 6646-54.

XUAN, R., BLASSENGALE, A. A. & WANG, Q. 2008. Degradation of estrogenic hormones in a silt loam soil. J Agric Food Chem, 56, 9152-8.

YALKOSKY, S. H. 1999. Solubility and Solubilization in Aqueous Media; Oxford University Press: New York.

YAMAMOTO, H. & LILJESTRAND, H. M. 2003. The fate of estrogenic compounds in the aquatic environment: sorption onto organic colloids. Water Sci Technol, 47, 77-84.

YCAS, M. 1974. On earlier states of the biochemical system. J Theor Biol, 44, 145-60.

YE, X., GUO, X., CUI, X., ZHANG, X., ZHANG, H., WANG, M. K., QIU, L. & CHEN, S. 2012. Occurrence and removal of endocrine-disrupting chemicals in wastewater treatment plants in the Three Gorges Reservoir area, Chongqing, China. J Environ Monit, 14, 2204-11.

YE, X., WANG, H., KAN, J., LI, J., HUANG, T., XIONG, G. & HU, Z. 2017. A novel 17β-hydroxysteroid dehydrogenase in Rhodococcus sp. P14 for transforming 17β-estradiol to estrone. Chemico-Biological Interactions.

YI, T. & HARPER, W. F., JR. 2007. The link between nitrification and biotransformation of 17alpha-ethinylestradiol. Environ Sci Technol, 41, 4311-6.

YIN, G. G., KOOKANA, R. S. & RU, Y. J. 2002. Occurrence and fate of hormone steroids in the environment. Environ Int, 28, 545-51.

YING, G. G. & KOOKANA, R. S. 2005. Sorption and degradation of estrogen-like-endocrine disrupting chemicals in soil. Environ Toxicol Chem, 24, 2640-5.

YING, G. G., KOOKANA, R. S. & DILLON, P. 2003. Sorption and degradation of selected five endocrine disrupting chemicals in aquifer material. Water Res, 37, 3785-91.

YOSHIMOTO, T., NAGAI, F., FUJIMOTO, J., WATANABE, K., MIZUKOSHI, H., MAKINO, T., KIMURA, K., SAINO, H., SAWADA, H. & OMURA, H. 2004. Degradation of estrogens by Rhodococcus zopfii and Rhodococcus equi isolates from activated sludge in wastewater treatment plants. Appl Environ Microbiol, 70, 5283-9.

YU, C. P., DEEB, R. A. & CHU, K. H. 2013. Microbial degradation of steroidal estrogens. Chemosphere, 91, 1225-35.

YU, C. P., ROH, H. & CHU, K. H. 2007. 17beta-estradiol-degrading bacteria isolated from activated sludge. Environ Sci Technol, 41, 486-92.

YU, Q., WANG, P., LIU, D., GAO, R., SHAO, H., ZHAO, H., MA, Z., WANG, D. & HUO, H. 2016. Degradation characteristics and metabolic pathway of 17β-estradiol (E2) by Rhodococcus sp. DS201. Biotechnology and Bioprocess Engineering, 21, 804-813.

YU, Z., XIAO, B., HUANG, W. & PENG, P. 2004. Sorption of steroid estrogens to soils and sediments. Environ Toxicol Chem, 23, 531-9.

YUAN, X., LI, T., ZHOU, L. & ZHAO, X. 2014. Characteristics and Risk Assessment of Estrogenic Compounds in Rivers of Southern Jiangsu Province, China. IERI Procedia, 9, 176-184.

ZENG, Q., LI, Y., GU, G., ZHAO, J., ZHANG, C. & LUAN, J. 2009. Sorption and Biodegradation of 17β-Estradiol by Acclimated Aerobic Activated Sludge and Isolation of the Bacterial Strain. Environmental Engineering Science, 26, 783-790.

ZHANG, A., LI, Y. & CHEN, L. 2014. Distribution and seasonal variation of estrogenic endocrine disrupting compounds, N-nitrosodimethylamine, and N-nitrosodimethylamine formation potential in the Huangpu River, China. J Environ Sci (China), 26, 1023-33.

ZHANG, J., CHIODINI, R., BADR, A. & ZHANG, G. 2011. The impact of next-generation sequencing on genomics. Journal of genetics and genomics = Yi chuan xue bao, 38, 95-109.

ZHANG, T., XIONG, G. & MASER, E. 2011. Characterization of the steroid degrading bacterium S19-1 from the Baltic Sea at Kiel, Germany. Chem Biol Interact, 191, 83-8.

ZHANG, Y., QIN, F., QIAO, J., LI, G., SHEN, C., HUANG, T. & HU, Z. 2012. Draft genome sequence of Rhodococcus sp. strain P14, a biodegrader of high-molecular-weight polycyclic aromatic hydrocarbons. J Bacteriol, 194, 3546.

ZHAO, J.-L., YING, G.-G., WANG, L., YANG, J.-F., YANG, X.-B., YANG, L.-H. & LI, X. 2009. Determination of phenolic endocrine disrupting chemicals and acidic pharmaceuticals in surface water of the Pearl Rivers in South China by gas

chromatography–negative chemical ionization–mass spectrometry. Science of The Total Environment, 407, 962-974.

ZHENG, D., WANG, X., WANG, P., PENG, W., JI, N. & LIANG, R. 2016. Genome Sequence of Pseudomonas citronellolis SJTE-3, an Estrogen- and Polycyclic Aromatic Hydrocarbon-Degrading Bacterium. Genome Announc, 4.

ZHENG, M., WANG, L., BI, Y. & LIU, F. 2011. Improved method for analyzing the degradation of estrogens in water by solid-phase extraction coupled with ultra performance liquid chromatography-ultraviolet detection. J Environ Sci (China), 23, 693-8.

ZIMIN, A. V., MARCAIS, G., PUIU, D., ROBERTS, M., SALZBERG, S. L. & YORKE, J. A. 2013. The MaSuRCA genome assembler. Bioinformatics, 29, 2669-77.

Appendix

Figures

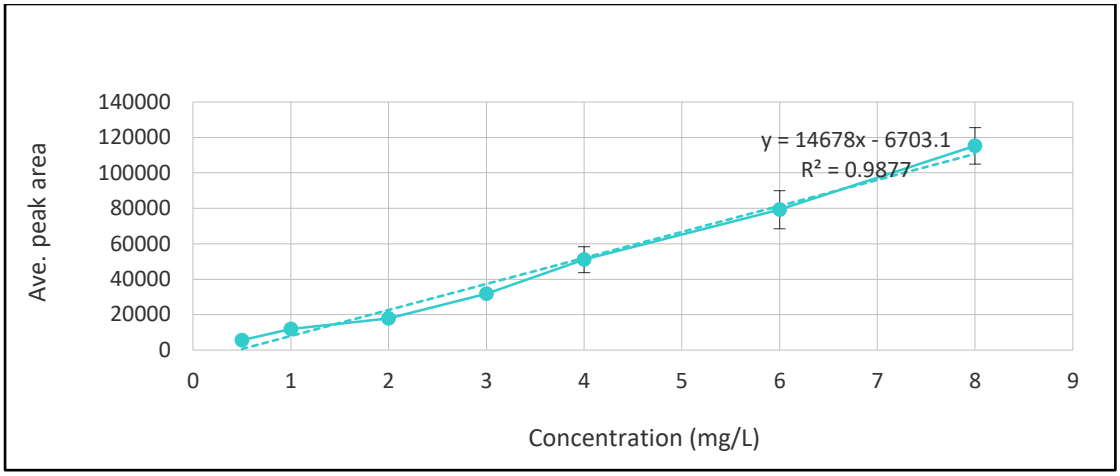

Figure Appendix 0-1: The standard curve for E1 used to determine the concentration of the E1 samples of the biomass removal experiment. Error bars represent the standard deviation.

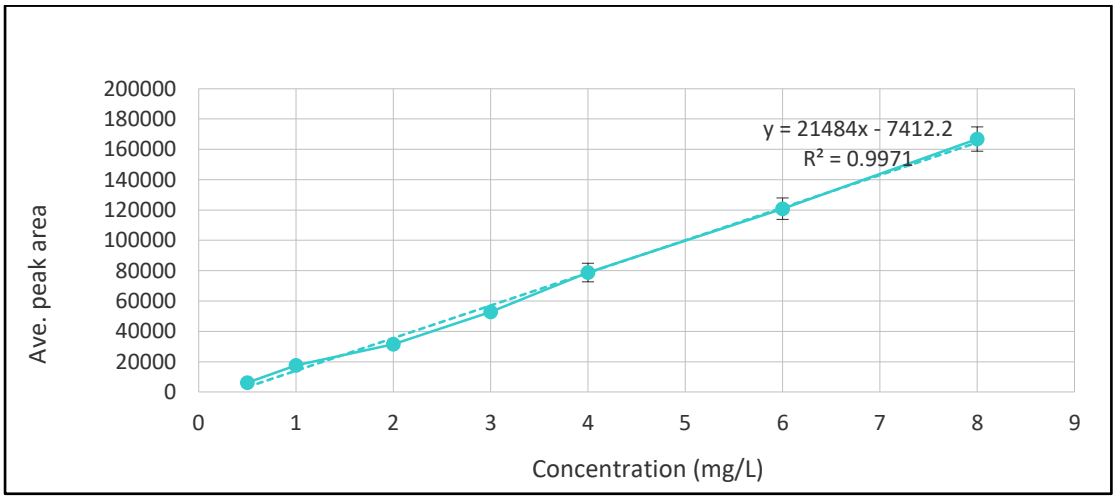

Figure Appendix 0-2: The standard curve for E2 used to determine the concentration of the E2 samples of the biomass removal experiment. Error bars represent the standard deviation.

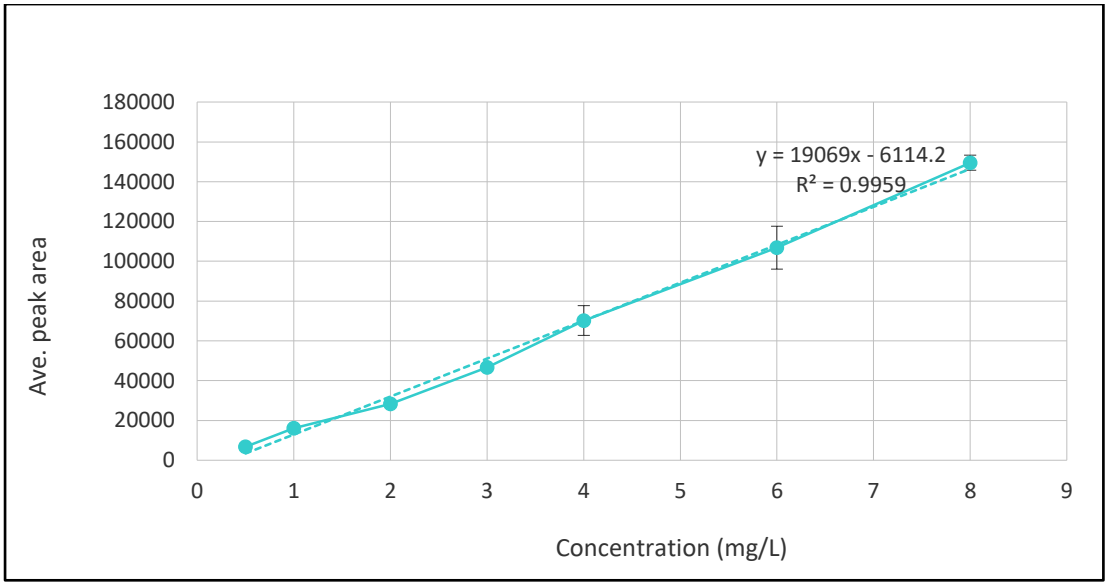

Figure Appendix 0-3: The standard curve for EE2 used to determine the concentration of the EE2 samples of the biomass removal experiment. Error bars represent the standard deviation.

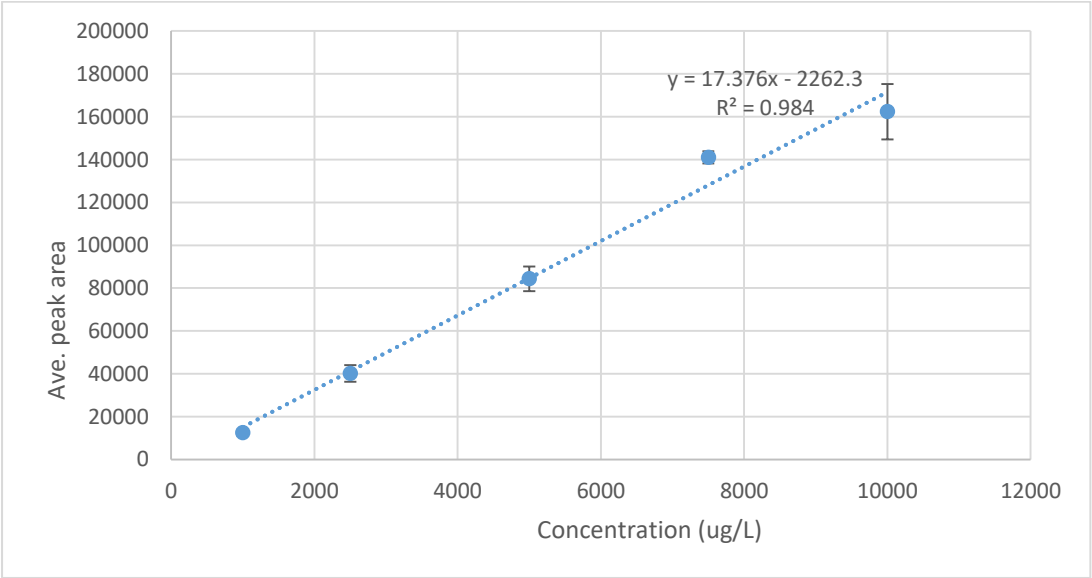

Figure Appendix 0-4: The standard curve for E1 used to determine the concentration of the E1 samples of the methods of dissolving estrogen into MSM. Error bars represent the standard deviation.

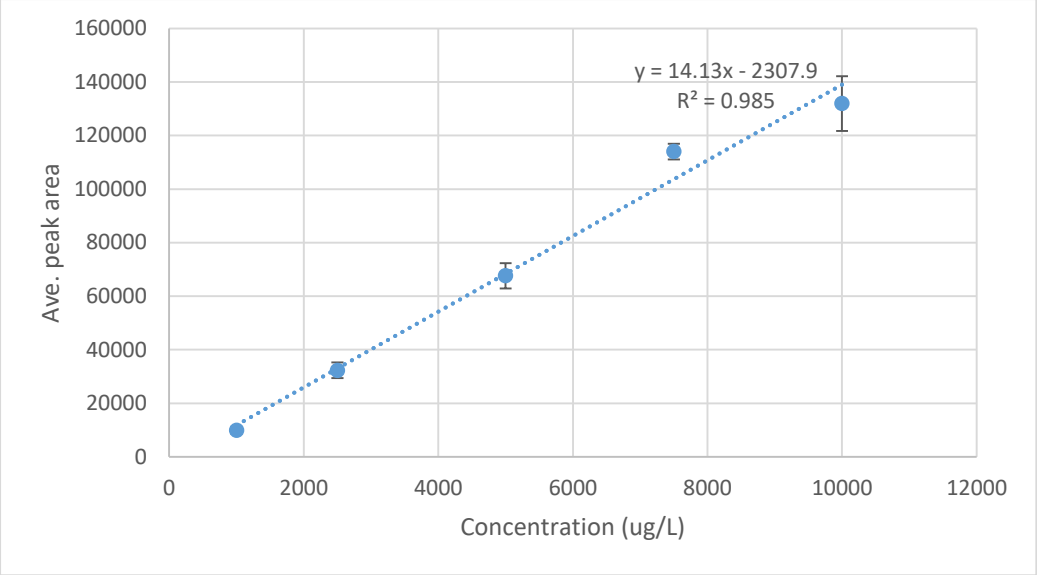

Figure Appendix 0-5: The standard curve for E2 used to determine the concentration of the E2 samples of the methods of dissolving estrogen into MSM. Error bars represent the standard deviation.

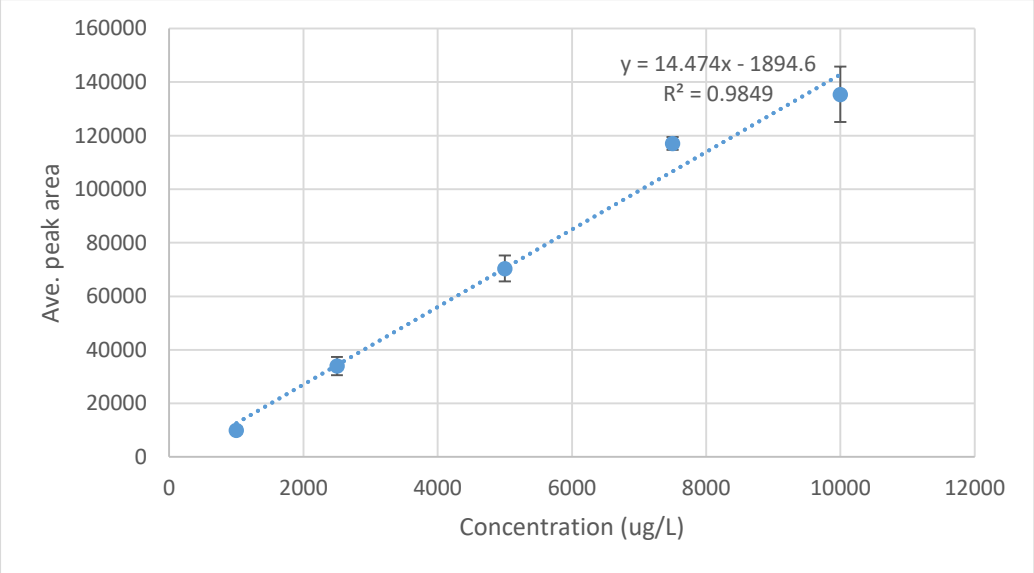

Figure Appendix 0-6: The standard curve for EE2 used to determine the concentration of the EE2 samples of the methods of dissolving estrogen into MSM. Error bars represent the standard deviation.

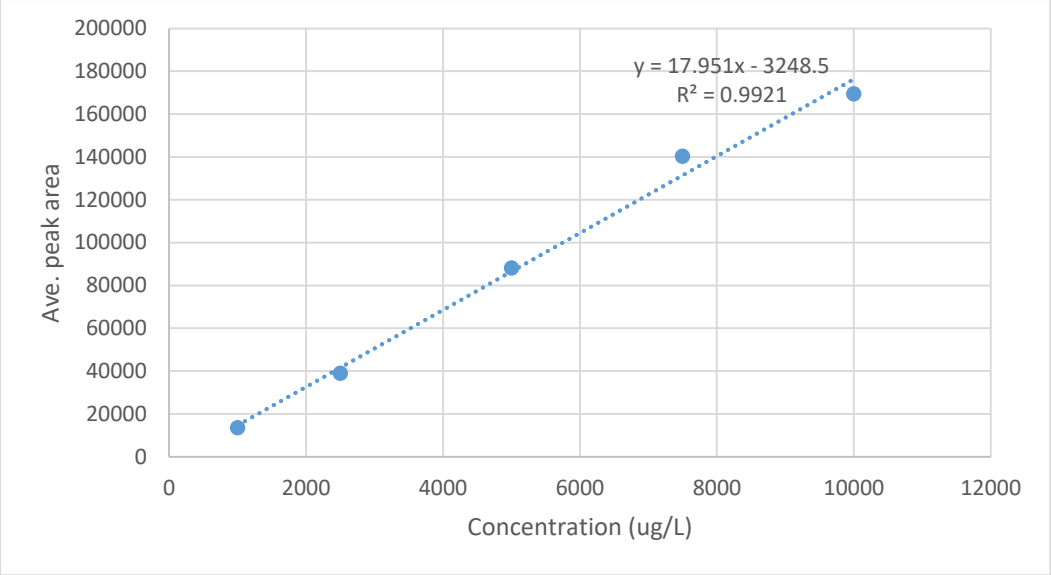

Figure Appendix 0-7: The standard curve for E1 used to determine the concentration of the E1 samples of the initial concentrations experiment. Error bars represent the standard deviation.

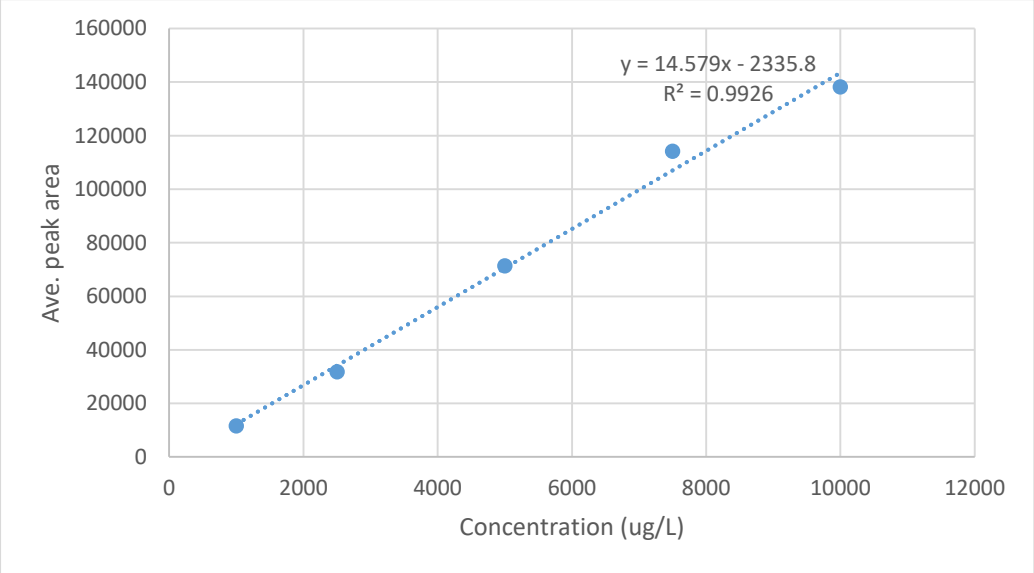

Figure Appendix 0-8: The standard curve for E2 used to determine the concentration of the E2 samples of the initial concentrations experiment. Error bars represent the standard deviation.

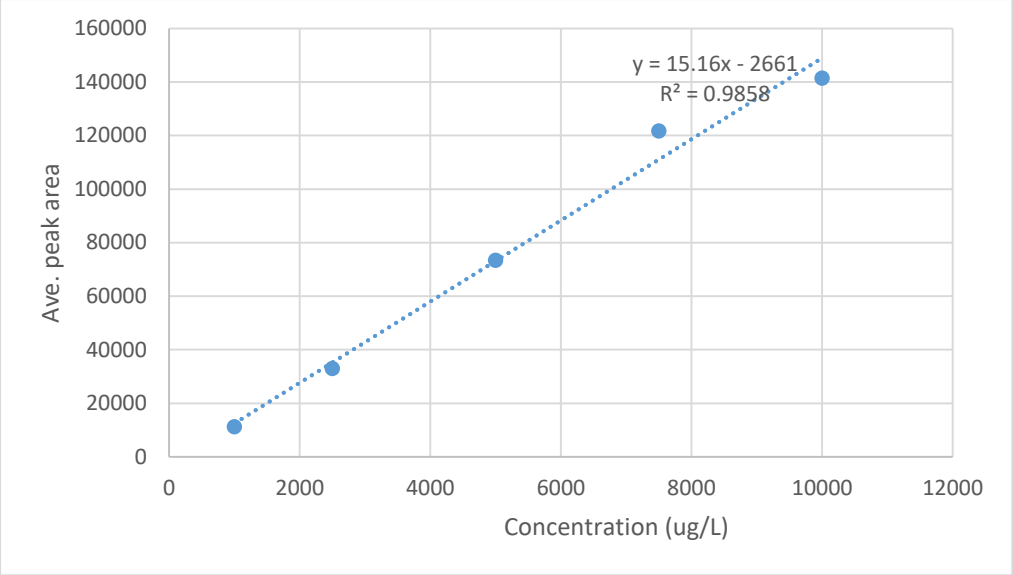

Figure Appendix 0-9: The standard curve for EE2 used to determine the concentration of the EE2 samples of the initial concentrations experiment. Error bars represent the standard deviation.

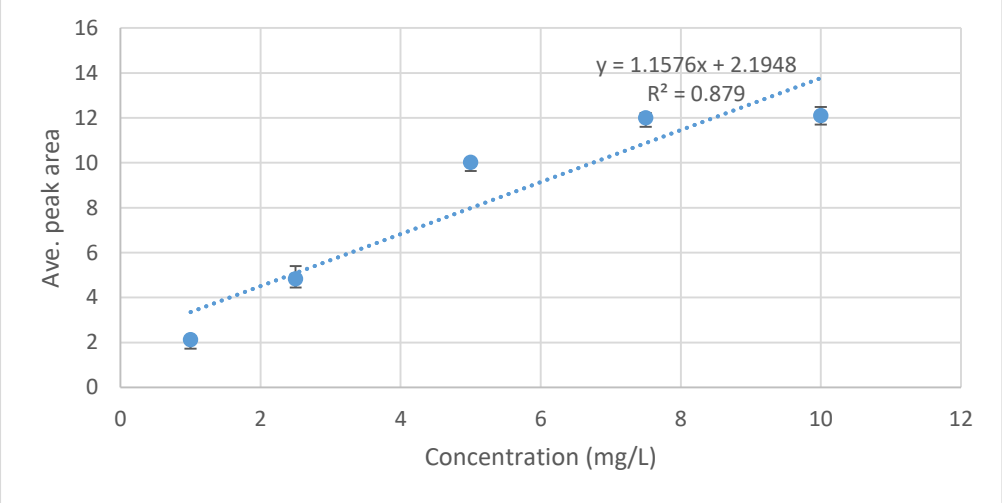

Figure Appendix 0-10: The standard curve for E1 used to determine the concentration of the E1 samples of the 45.5h degradation experiment. Error bars represent the standard deviation.

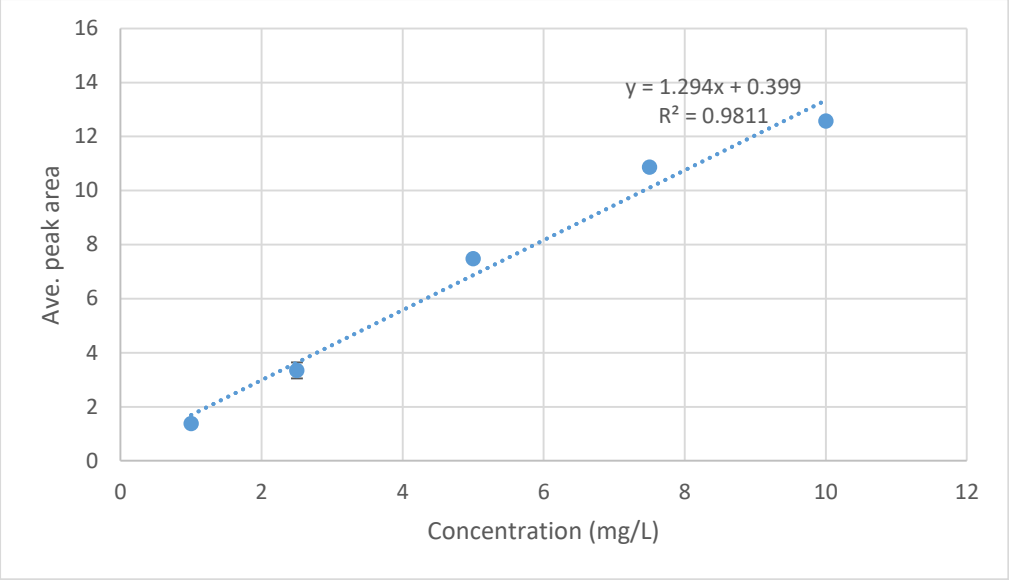

Figure Appendix 0-11: The standard curve for E2 used to determine the concentration of the E2 samples of the 45.5 hour degradation experiment. Error bars represent the standard deviation.

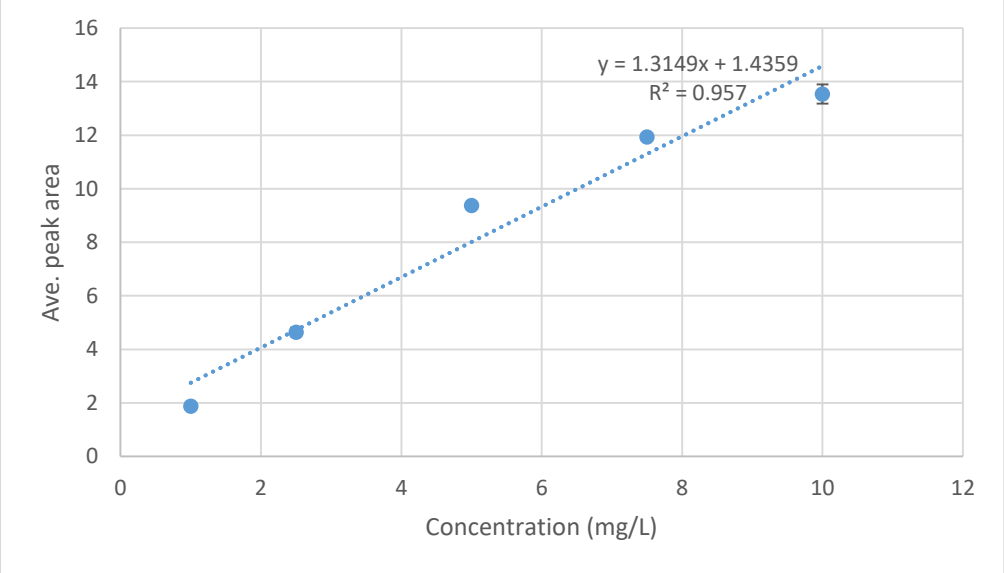

Figure Appendix 0-12: The standard curve for EE2 used to determine the concentration of the EE2 samples of the 45.5 hour degradation experiment. Error bars represent the standard deviation.

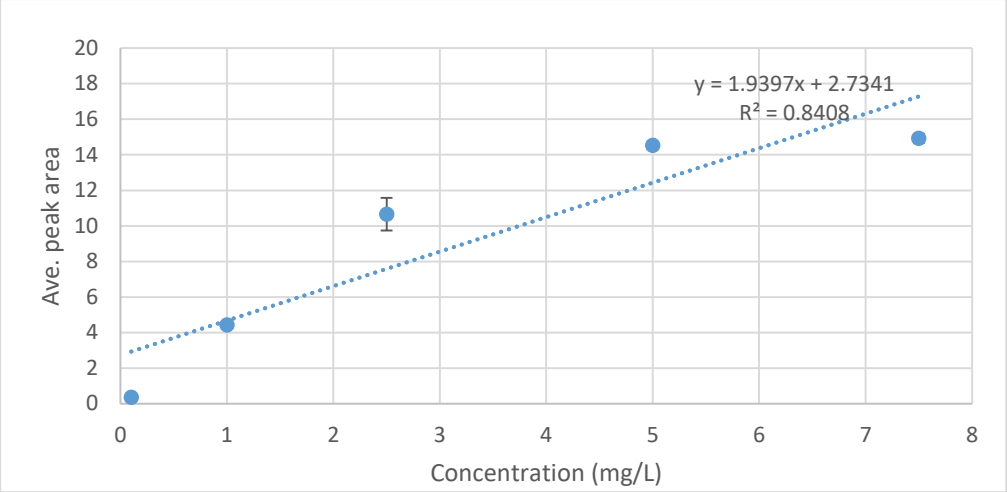

Figure Appendix 0-13: The standard curve for E1 used to determine the concentration of the E1 samples of the 311 hour degradation experiment. Error bars represent the standard deviation.

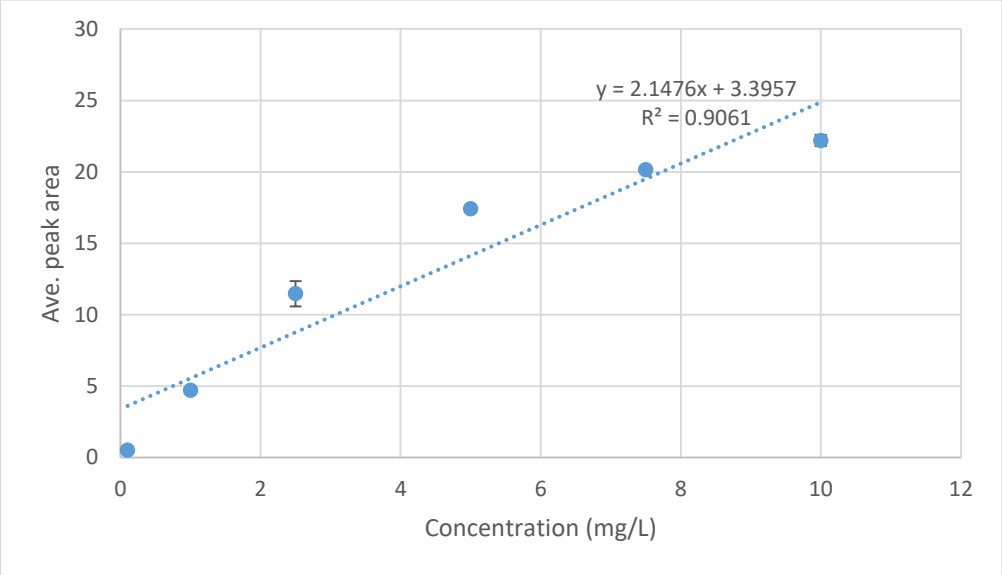

Figure Appendix 0-14: The standard curve for E2 used to determine the concentration of the E2 samples of the 311 hour degradation experiment. Error bars represent the standard deviation.

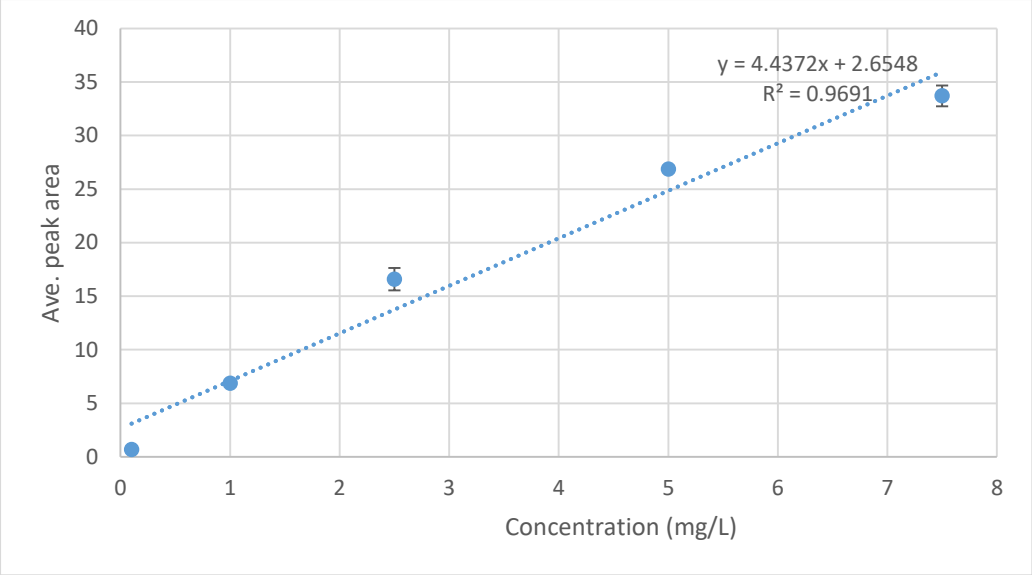

Figure Appendix 0-15:The standard curve for EE2 used to determine the concentration of the EE2 samples of the 311 hour degradation experiment. Error bars represent the standard deviation.

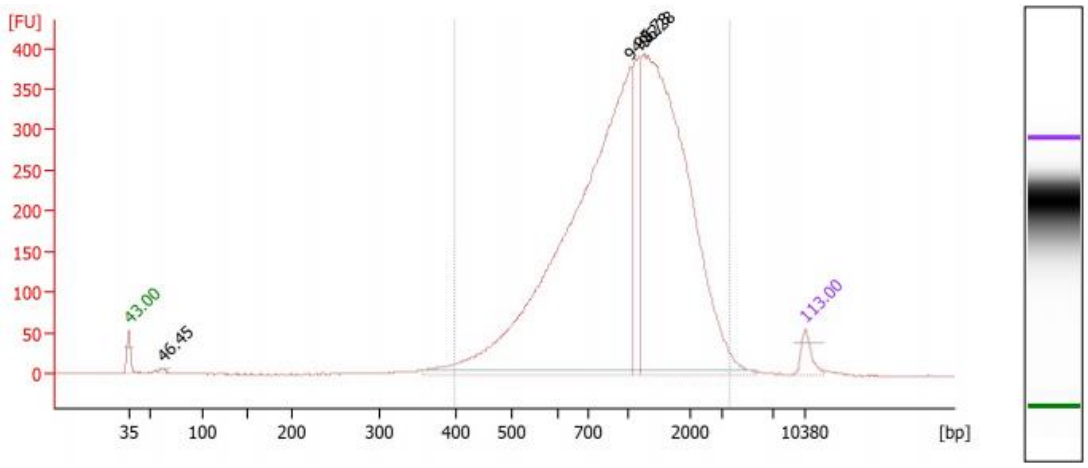

Figure Appendix 0-16: Bioanalyser trace for R.equi ATCC13557.

Tables:

Table Appendix 0-1: Bioanalyser data overall results

Number of peaks found	4
Noise	0.4
Corr. Area 1	5,629.6

Table Appendix 0-2: Bioanalyser peak data

Peak	Size (bp)	Conc. (pg/µL)	Molarity (pmol/µL)	Observations
1	35	125.00	5,411.3	Lower marker
2	61	19.26	477.2	
3	1,036	4,034.09	5,900.8	
4	1,182	531.78	681.7	
5	1,259	3,137.26	3,775.9	
6	10,380	75.00	10.9	Upper marker

Table Appendix 0-3: Bioanalyser region table

From (bp)	To (bp)	Corr. Area	% of Total	Average size (bp)	Size distribution in CV (%)	Conc. (pg/µL)	Molarity (pmol/L)
397	3,520	5,629.6	97	1,191	47.9	8,063.36	13,404.1

Table Appendix 0-4: Database of genes coding for enzymes potentially involved in estrogen degradation

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
ORF11	AB063482 a	Comamonas testosteroni ORF11, tesA, tesD, tesE, tesF, tesG genes, complete cds	Horinouchi,M Appl. Environ. Microbiol. 69 (4), 2139-2152 (2003)
TesA/ORF12 (hydroxylase)	AB063482 b	Comamonas testosteroni ORF11, tesA, tesD, tesE, tesF, tesG genes, complete cds	Horinouchi, M Appl. Environ. Microbiol. 69 (4), 2139-2152 (2003)
TesA2 (hydroxylase of 3-hydroxy-9, 10-secoandrosta -1, 3, 5(10)-triene-9, 17-dione, flavin-dependent monooxygenase oxygenase subunit)	LC010134	Comamonas testosteroni DNA, putative mega-cluster of steroid degradation genes, strain: T441.	Horinouchi Microbiology (Reading, Engl.) 147 (PT 12), 3367-3375 (2001)
flavin-dependent monooxygenase oxygenase subunit	CP006704	Comamonas testosteroni TK102, complete genome.	Fukuda Genome Announc 2 (5), e00865-14 (2014)
ORF17 (3-ketosteroid-9-alpha-hydroxylase reductase) 1a (EC(1.14.13.142)	AB076368	Comamonas testosteroni tesH, tesI, ORF17, ORF18 genes for 3-ketosteriod-delta1-dehydrogenase, 3-ketosteriod-delta4(5alpha)-dehydrogenase, hypothetical protein, complete cds.	Horinouchi, M. Appl. Environ. Microbiol. 69 (8), 4421-4430 (2003)
3-ketosteroid-9-alpha-hydroxylase reductase subunit	CP006704	Comamonas testosteroni TK102, complete genome.	Fukuda Genome Announc 2 (5), e00865-14 (2014)
3-(3-hydroxy-phenyl)proprionate hydroxylase	R.equi ATCC13557	R.equi ATCC13557 rast annotation of spades assembly	This thesis
Chromosome I	LN879547	Comamonas testosteroni P19 genome assembly Comamonas testosteroni P19, chromosome : I	Park,Soo.-Je. Submitted (15-SEP-2015) JEJU NATIONAL UNIVERSITY, Lab of Microbial Microbial Ecology and Genomics, Dept. of Biology, Jeju National University, 102 Jejudaehak-ro, Jeju 690-756, Jeju, 690-756, South Korea

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	BCMM01000037.1	Streptomyces scabiei DNA, scaffold: NODE_037_cov_57.5993, strain: S58, whole genome shotgun sequence	Tomihama,T., Ikenaga,M., Sakai,M., Okubo,T. and Ikeda,S. Ecology and Genomics, Dept. of Biology, Jeju National University, 102 Jejudaehak-ro, Jeju 690-756, Jeju, 690-756, South Korea
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	LLWJ01000056.1	Pseudomonas sp. TAA207 contig000056, whole genome shotgun sequence	Submitted (20-OCT-2015) University of Florence, Via Madonna del Piano 6, Sesto Fiorentino, Italy
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	LLWI01000067.1	Pseudomonas sp. TAD18 contig000067, whole genome shotgun sequence	Submitted (20-OCT-2015) Dept. of Biology, University of Florence, Via Madonna del Piano 6, Sesto Fiorentino, Italy
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	BCMK01000014.1	Streptomyces acidiscabies DNA, scaffold: NODE_014_cov_52.5735, strain: a10, whole genome shotgun sequence	Tomihama et al., Genome Announc 4 (2), e00062-16 (2016)
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	BCMK01000200.1	Streptomyces acidiscabies DNA, scaffold: NODE_200_cov_56.898, strain: a10, whole genome shotgun sequence	Tomihama et al., Genome Announc 4 (2), e00062-16 (2016)
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	BCMN01000002.1	Streptomyces turgidiscabies DNA, scaffold: NODE_002_cov_60.061, strain: T45, whole genome shotgun sequence	Tomihama et al., Genome Announc 4 (2), e00062-16 (2016)
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	CP012042.1	Burkholderia pseudomallei strain Bp1651 chromosome 2, complete sequence	Bugrysheva et al., Genome Announc. 2015 Dec 3;3(6). pii: e01427-15. doi: 10.1128/genomeA.01427-15.
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	CVQP01000002.1	Rhodococcus sp. RD6.2 genome assembly RD6.2, scaffold RHCRD62_Contig_10, whole genome shotgun sequence	De Marco Submitted (12-MAY-2015) IINFACTS, R. Central de Gandra, 1317 Gandra PRD, Portugal
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	CVQP01000005.1	Rhodococcus sp. RD6.2 genome assembly RD6.2, scaffold RHCRD62_Contig_2, whole genome shotgun sequence	De Marco Submitted (12-MAY-2015) IINFACTS, R. Central de Gandra, 1317 Gandra PRD, Portugal
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	CVQP01000007.1	Rhodococcus sp. RD6.2 genome assembly RD6.2, scaffold RHCRD62_Contig_4, whole genome shotgun sequence	De Marco Submitted (12-MAY-2015) IINFACTS, R. Central de Gandra, 1317 Gandra PRD, Portugal

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Rieske (2Fe-2S) domain-containing protein	CP011072.1	Azoarcus sp. CIB, complete genome	Martin-Moldes et al Submitted (04-MAR-2015) Environmental Biology, Centro de Investigaciones Biologicas (CSIC), Ramiro de Maeztu 9, Madrid, Madrid 28040, Spain
3-ketosteroid 9alpha-monooxygenase (1.14.13.142)	JYNL01000001.1	Mycobacterium chlorophenolicum strain DSM 43826 MCHLDSM_contig000001, whole genome shotgun sequence	Das et al Genome Biol Evol. 2015 Jun 16;7(7):1871-86. doi: 10.1093/gbe/evv111.
3-ketosteroid-9-alpha-monooxygenase oxygenase subunit	JYNL01000008.1	Mycobacterium chlorophenolicum strain DSM 43826 MCHLDSM_contig000008, whole genome shotgun sequence	Das et al., Genome Biol Evol. 2015 Jun 16;7(7):1871-86. doi: 10.1093/gbe/evv111.
3-ketosteroid-9-alpha-monooxygenase oxygenase	JYNU01000001.1	Mycobacterium obuense strain DSM 44075 MOBUDSM44075_contig000001, whole genome shotgun sequence	Das et al Genome Biol Evol. 2015 Jun 16;7(7):1871-86. doi: 10.1093/gbe/evv111.
3-ketosteroid-9-alpha-monooxygenase oxygenase	JYNU01000002.1	Mycobacterium obuense strain DSM 44075 MOBUDSM44075_contig000002, whole genome shotgun sequence	Das et al Genome Biol Evol. 2015 Jun 16;7(7):1871-86. doi: 10.1093/gbe/evv111.
3-ketosteroid-9-alpha-monooxygenase oxygenase	JYNX01000035.1	Mycobacterium chubuense strain DSM 44219 MCHUDSM44219_contig000035, whole genome shotgun sequence	Das et al Genome Biol Evol. 2015 Jun 16;7(7):1871-86. doi: 10.1093/gbe/evv111.
3-ketosteroid-9-alpha-monooxygenase oxygenase	JYNX01000038.1	Mycobacterium chubuense strain DSM 44219 MCHUDSM44219_contig000038, whole genome shotgun sequence.	Das et al Genome Biol Evol. 2015 Jun 16;7(7):1871-86. doi: 10.1093/gbe/evv111.
3-ketosteroid-9-alpha-monooxygenase oxygenase	JYNL01000069.1	Mycobacterium chlorophenolicum strain DSM 43826 MCHLDSM_contig000070, whole genome shotgun sequence	Das et al Genome Biol Evol. 2015 Jun 16;7(7):1871-86. doi: 10.1093/gbe/evv111.
3-ketosteroid-9-alpha-monooxygenase oxygenase	CDHG01000028.1	Mycobacterium caprae genome assembly Genome assembly of M caprae MB2, contig 28, whole genome shotgun sequence	Manrique Submitted (12-NOV-2014) ERA7, Oh no sequences! group, Plaza Campoverde 3 atico, 18001, Spain
3-ketosteroid-9-alpha-monooxygenase oxygenase	CDHE01000065.1	Mycobacterium bovis genome assembly Assembly of Mycobacterium bovis MB4 genome, contig 65, whole genome shotgun sequence	Manrique Submitted (12-NOV-2014) ERA7, Oh no sequences! group, Plaza Campoverde 3 atico, 18001, Spain

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
3-ketosteroid-9-alpha-monooxygenase oxygenase	CP006850.1	Nocardia nova SH22a, complete genome	Luo et al., Appl Environ Microbiol. 2014 Jul;80(13):3895-907. doi: 10.1128/AEM.00473-14. Epub 2014 Apr 18.
3-ketosteroid-9-alpha-monooxygenase oxygenase	CDHH01000076.1	Mycobacterium bovis genome assembly Assembly of the genome MB3, contig 76, whole genome shotgun sequence	Manrique Submitted (12-NOV-2014) ERA7, Oh no sequences! group, Plaza Campoverde 3 atico, 18001, Spain
3-ketosteroid-9-alpha-monooxygenase oxygenase	CDHF01000008.1	Mycobacterium bovis genome assembly Genome assembly of Mycobacterium bovis MB1, contig 8, whole genome shotgun sequence	Manrique Submitted (12-NOV-2014) ERA7, Oh no sequences! group, Plaza Campoverde 3 atico, 18001, Spain
3-ketosteroid-9-alpha-monooxygenase oxygenase	HG322950.1	Pseudomonas knackmussii B13 complete genome	Miyazaki et al., Submitted (08-MAR-2013) Bielefeld University, Center For Biotechnology, Universitaetsstrasse 25, D-33501 Bielefeld, GERMANY
3-ketosteroid-9-alpha-monooxygenase oxygenase	CCSD01000032.1	Rhodococcus ruber genome assembly BDK_PRJEB6917_v1, contig BDK_RHRU231_Contig_23, whole genome shotgun sequence	Ivshina et al., Genome Announc. 2014 Dec 11;2(6). pii: e01297-14. doi: 10.1128/genomeA.01297-14
3-ketosteroid-9-alpha-monooxygenase oxygenase	CCSD01000043.1	Rhodococcus ruber genome assembly BDK_PRJEB6917_v1, contig BDK_RHRU231_Contig_33, whole genome shotgun sequence	Ivshina et al., Genome Announc. 2014 Dec 11;2(6). pii: e01297-14. doi: 10.1128/genomeA.01297-14
3-ketosteroid-9-alpha-monooxygenase oxygenase	CCSD01000089.1	Rhodococcus ruber genome assembly BDK_PRJEB6917_v1, contig BDK_RHRU231_Contig_75, whole genome shotgun sequence	Ivshina et al., Genome Announc. 2014 Dec 11;2(6). pii: e01297-14. doi: 10.1128/genomeA.01297-14
3-ketosteroid-9-alpha-monooxygenase oxygenase	LN482603.1	Rhodococcus ruber genome assembly BDK_PRJEB6917_v1, scaffold BDK_RHRU231_scaffold13, whole genome shotgun sequence	Ivshina et al., Genome Announc. 2014 Dec 11;2(6). pii: e01297-14. doi: 10.1128/genomeA.01297-14
3-ketosteroid-9-alpha-monooxygenase oxygenase	LN482615.1	Rhodococcus ruber genome assembly BDK_PRJEB6917_v1, scaffold	Ivshina et al., Genome Announc. 2014 Dec 11;2(6). pii: e01297-14. doi: 10.1128/genomeA.01297-14

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
		BDK_RHRU231_scaffold25, whole genome shotgun sequence	
flavodoxin reductase family protein	JTJI01000003.1a	Prauserella sp. Am3 HQ32_scaffold_2.3, whole genome shotgun sequence	Sarkar et al., Submitted (25-NOV-2014) Cellular, Molecular and Biomedical Science, University of New Hampshire, 46 College Road, Durham, NH 03824, USA
Rieske (2Fe-2S) domain-containing	JTJI01000003.1b	Prauserella sp. Am3 HQ32_scaffold_2.3, whole genome shotgun sequence	Sarkar et al., Submitted (25-NOV-2014) Cellular, Molecular and Biomedical Science, University of New Hampshire, 46 College Road, Durham, NH 03824, USA
3-ketosteroid-9-alpha-monooxygenase oxygenase	CP008953.1	Amycolatopsis japonica strain MG417-CF17, complete genome	Stegmann et al., J Biotechnol. 2014 Nov 10;189:46-7. doi: 10.1016/j.jbiotec.2014.08.034. Epub 2014 Sep 3.
3-ketosteroid-9-alpha-monooxygenase oxygenase	JNNW01000040.1	Mycobacterium tuberculosis strain A4 Contig_136_len_305152, whole genome shotgun sequence	Liao et al., Genome Announc. 2014 Sep 4;2(5). pii: e00672-14. doi: 10.1128/genomeA.00672-14.
3-ketosteroid-9-alpha-monooxygenase oxygenase	JNGF01000142.1	Mycobacterium tuberculosis strain A2 Contig_56, whole genome shotgun sequence	Liao et al., Genome Announc. 2014 Sep 4;2(5). pii: e00672-14. doi: 10.1128/genomeA.00672-14.
3-ketosteroid-9-alpha-monooxygenase oxygenase	AZIN01000001.1	Gammaproteobacteria bacterium MOLA455 MOLA455_Contig1, whole genome shotgun sequence	Courties et al., Genome Announc. 2014 Jan 30;2(1). pii: e01203-13. doi: 10.1128/genomeA.01203-13.
putative oxidoreductase	FN554889.1a	Streptomyces scabiei 87.22 complete genome	Bignell et al., Mol. Plant Microbe Interact. 23 (2), 161-175 (2010)
putative oxidoreductase	FN554889.1b	Streptomyces scabiei 87.22 complete genome	Bignell et al., Mol. Plant Microbe Interact. 23 (2), 161-175 (2010)
rieske [2Fe-2S] domain protein	CP009160.1	Burkholderia pseudomallei TSV 48 chromosome 2, complete sequence	Johnson et al., Genome Announc 3 (1) (2015)
Rieske (2Fe-2S) protein	CP012577.1	Burkholderia pseudomallei strain 982 chromosome 2, complete sequence	Hong Submitted (31-AUG-2015) Institute of Biological Sciences University of Malaya, Microbiome Lab, High Impact Research, Building, Kuala Lumpur, Kuala Lumpur 50603, Malaysia

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
rieske [2Fe-2S] domain protein	CP010066.1	Burkholderia mallei strain 2002721276 chromosome 2, complete sequence.	Johnson et al., Genome Announc 3 (2) (2015)
rieske [2Fe-2S] domain protein	CP009943.1	Burkholderia mallei strain KC_1092 chromosome 2 sequence.	Johnson et al., Genome Announc 3 (2) (2015)
rieske [2Fe-2S] domain protein	CP009898.1	Burkholderia pseudomallei Pasteur 52237 chromosome 2, complete sequence.	Johnson et al., Genome Announc 3 (2) (2015)
rieske [2Fe-2S] domain protein	CP009588.1	Burkholderia mallei strain 11 chromosome II, complete sequence.	Johnson et al., Genome Announc 3 (2) (2015)
rieske [2Fe-2S] domain protein	CP009537.1	Burkholderia pseudomallei K96243 chromosome II, complete sequence.	Johnson et al., Genome Announc 3 (2) (2015)
rieske [2Fe-2S] domain protein	CP009484.1	Burkholderia pseudomallei MSHR491 chromosome II, complete sequence.	Johnson et al., Genome Announc 3 (2) (2015)
rieske [2Fe-2S] domain protein	CP009338.1	Burkholderia mallei strain 2002734299 chromosome 2, complete	Johnson et al., Genome Announc 3 (2) (2015)
rieske [2Fe-2S] domain protein	CP010974.1	Burkholderia pseudomallei strain vgh07 chromosome 2, complete sequence.	Chen et al., Genome Announc. 2015 Apr 30;3(2). pii: e00345-15. doi: 10.1128/genomeA.00345-15.
rieske [2Fe-2S] domain protein	CP008731.2	Burkholderia mallei strain 2000031063 chromosome 2, complete	Daligault et al., Genome Announc 2 (6) (2014)
rieske [2Fe-2S] domain protein	CP009156.1	Burkholderia sp. TSV202 chromosome 2, complete sequence.	Johnson et al., Genome Announc 3 (1) (2015)
rieske [2Fe-2S] domain protein	CP009150.1	Burkholderia pseudomallei B03 chromosome 2, complete sequence.	Johnson et al., Genome Announc 3 (1) (2015)
rieske [2Fe-2S] domain protein	CP009164.1	Burkholderia pseudomallei A79A chromosome 2, complete sequence.	Johnson et al., Genome Announc 3 (1) (2015)
rieske [2Fe-2S] domain protein	CP009152.1	Burkholderia pseudomallei MSHR3965 chromosome 2, complete sequence.	Johnson et al., Genome Announc 3 (1) (2015)
rieske [2Fe-2S] domain protein	CP004378.1	Complete genome sequences for 59 burkholderia isolates, both pathogenic and near neighbour	Johnson et al., Genome Announc 3 (2) (2015)
rieske [2Fe-2S] domain protein	CP008917.1	Burkholderia sp. BGK chromosome 2 sequence.	Daligault et al Genome Announc 2 (6) (2014)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
rieske [2Fe-2S] domain protein	CP007801.1	Burkholderia mallei NCTC 10247 chromosome 2, complete sequence.	Daligault et al Genome Announc 2 (6) (2014)
rieske [2Fe-2S] domain protein	CP009147.1	Burkholderia mallei strain FMH 23344 chromosome 2, complete	Daligault et al Genome Announc 2 (6) (2014)
rieske [2Fe-2S] domain protein	CP008912.1	Burkholderia pseudomallei HBPUB10134a chromosome 2, complete	Daligault et al Genome Announc 2 (6) (2014)
rieske [2Fe-2S] domain protein	CP008910.1	Burkholderia pseudomallei MSHR5848 chromosome 2, complete sequence.	Daligault et al Genome Announc 2 (6) (2014)
rieske [2Fe-2S] domain protein	CP008783.1	Burkholderia pseudomallei MSHR5855 chromosome 2, complete sequence.	Daligault et al Genome Announc 2 (6) (2014)
rieske [2Fe-2S] domain protein	CP008754.1	Burkholderia pseudomallei strain 9 chromosome 2, complete sequence.	Daligault et al Genome Announc 2 (6) (2014)
rieske [2Fe-2S] domain protein	CP008722.1	Burkholderia mallei strain BMQ chromosome 2, complete sequence.	Daligault et al Genome Announc 2 (6) (2014)
rieske [2Fe-2S] domain protein	CP008710.1	Burkholderia mallei strain 6 chromosome 2, complete sequence.	Daligault et al Genome Announc 2 (6) (2014)
rieske [2Fe-2S] domain protein	CP008705.1	Burkholderia mallei strain 23344 chromosome 2, complete sequence.	Daligault et al Genome Announc 2 (6) (2014)
Rieske iron-sulphur domain protein	LK936443.1	Burkholderia pseudomallei genome assembly BP_3921g, chromosome : 2.	Aslett and De Silva Submitted (03-JUL-2014) SC, Pathogen Informatics, Wellcome Trust Genome Campus, Hinxton, Cambridge, Cambridgeshire., CB10 1SA, United Kingdom
iron-sulfur cluster-binding protein, Rieske family	CP000547.1	Burkholderia mallei NCTC 10247 chromosome II, complete sequence.	DeShazer et al., Submitted (21-OCT-2009) The J. Craig Venter Institute, 9704 Medical Center Dr, Rockville, MD 20850, USA
iron-sulfur cluster-binding protein, Rieske	CP000545.1	Burkholderia mallei NCTC 10229 chromosome II, complete sequence.	DeShazer et al., Submitted (21-OCT-2009) The J. Craig Venter Institute, 9704 Medical
iron-sulfur cluster-binding protein, rieske	CP000525.1	Burkholderia mallei SAVP1 chromosome II, complete sequence.	DeShazer et al., Submitted (18-DEC-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
iron-sulfur cluster-binding protein, rieske	CP000125.1	Burkholderia pseudomallei 1710b chromosome II, complete sequence.	Woods and Nierman Submitted (28-SEP-2005) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
iron-sulfur cluster-binding protein, rieske	CP000011.2	Burkholderia mallei ATCC 23344 chromosome 2, complete sequence.	Nierman et al., Proc. Natl. Acad. Sci. U.S.A. 101 (39), 14246-14251 (2004)
Rieske iron-sulphur domain protein	BX571966.1	Burkholderia pseudomallei strain K96243, chromosome 2, complete	Holden et al., Proc Natl Acad Sci U S A. 2004 Sep 28;101(39):14240-5. Epub 2004 Sep 17.
rieske [2Fe-2S] domain protein	CP009546.1	Burkholderia pseudomallei strain MSHR668 chromosome II, complete	Johnson et al., Genome Announc. 2015 Apr 30;3(2). pii: e00159-15. doi: 10.1128/genomeA.00159-15.
rieske [2Fe-2S] domain protein	CP009536.1	Burkholderia pseudomallei 7894 chromosome II, complete sequence.	Johnson et al., Genome Announc. 2015 Apr 30;3(2). pii: e00159-15. doi: 10.1128/genomeA.00159-15.
rieske [2Fe-2S] domain protein	CP009477.1	Burkholderia pseudomallei MSHR2543 chromosome II, complete	Johnson et al., Genome Announc. 2015 Apr 30;3(2). pii: e00159-15. doi: 10.1128/genomeA.00159-15.
rieske [2Fe-2S] domain protein	CP009473.1	Burkholderia pseudomallei MSHR840 chromosome II, complete sequence.	Johnson et al., Genome Announc. 2015 Apr 30;3(2). pii: e00159-15. doi: 10.1128/genomeA.00159-15.
rieske [2Fe-2S] domain protein	CP009234.1	Burkholderia pseudomallei MSHR62 chromosome 2, complete sequence.	Johnson et al., Genome Announc. 2015 Feb 12;3(1). pii: e01282-14. doi: 10.1128/genomeA.01282-14.
rieske [2Fe-2S] domain protein	CP009269.1	Burkholderia pseudomallei MSHR2243 chromosome 2 sequence.	Bunnell et al., Submitted (26-AUG-2014) Bioscience, Los Alamos National Laboratory, P.O. Box 1663, MS M888, Los Alamos, NM 87545, US
rieske [2Fe-2S] domain protein	CP009272.1	Burkholderia pseudomallei MSHR1153 chromosome 2 sequence	Bunnell et al., Submitted (26-AUG-2014) Bioscience, Los Alamos National Laboratory,
rieske [2Fe-2S] domain protein	CP009210.1	Burkholderia pseudomallei strain BDP chromosome 2, complete	Daligault et al., Genome Announc. 2014 Nov 20;2(6). pii: e01106-14. doi: 10.1128/genomeA.01106-14.
rieske [2Fe-2S] domain protein	CP008891.1	Burkholderia pseudomallei MSHR5858 chromosome 2, complete sequence.	Daligault et al., Genome Announc. 2014 Nov 20;2(6). pii: e01106-14. doi: 10.1128/genomeA.01106-14.
rieske [2Fe-2S] domain protein	CP008779.1	Burkholderia pseudomallei strain MSHR1655 chromosome 2, complete	Daligault et al., Genome Announc. 2014 Nov 20;2(6). pii: e01106-14. doi: 10.1128/genomeA.01106-14.
rieske [2Fe-2S] domain protein	CP004369.1	Burkholderia pseudomallei MSHR520 chromosome 2, complete sequence	Johnson et al., Genome Announc. 2015 Feb 12;3(1). pii: e01282-14. doi: 10.1128/genomeA.01282-14.

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
rieske [2Fe-2S] domain protein	CP003977.1	Burkholderia pseudomallei NCTC 13179 chromosome 2, complete	Johnson et al., Genome Announc. 2015 Feb 12;3(1). pii: e01282-14. doi: 10.1128/genomeA.01282-14.
rieske [2Fe-2S] domain protein	CP006469.1	Burkholderia pseudomallei MSHR305 chromosome 2, complete sequence.	Stone et al., Genome Announc. 2013 Aug 22;1(4). pii: e00656-13. doi: 10.1128/genomeA.00656-13.
iron-sulfur cluster-binding protein, Rieske	CP000571.1	Burkholderia pseudomallei 668 chromosome II, complete sequence.	DeShazer et al., Submitted (12-JAN-2010) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
Rieske (2Fe-2S) protein	CP012518.1	Burkholderia pseudomallei strain vgh16W chromosome 2, complete	Hsueh et al., Genome Announc. 2015 Oct 15;3(5). pii: e01194-15. doi: 10.1128/genomeA.01194-15.
Rieske (2Fe-2S) protein	CP012516.1	Burkholderia pseudomallei strain vgh16R chromosome 2, complete	Hsueh et al., Genome Announc. 2015 Oct 15;3(5). pii: e01194-15. doi: 10.1128/genomeA.01194-15.
Rieske (2Fe-2S) protein	CP012093.1	Burkholderia pseudomallei strain 350105 chromosome 2 sequence.	Song et al., Genome Announc. 2015 Oct 15;3(5). pii: e01162-15. doi: 10.1128/genomeA.01162-15.
Rieske (2Fe-2S) protein	CP009550.1	Burkholderia pseudomallei PB08298010 chromosome II, complete	Johnson et al., Genome Announc. 2015 Feb 12;3(1). pii: e01282-14. doi: 10.1128/genomeA.01282-14.
Rieske (2Fe-2S) protein	CP009297.1	Burkholderia pseudomallei 406e chromosome 2, complete sequence.	Johnson et al., Genome Announc. 2015 Feb 12;3(1). pii: e01282-14. doi: 10.1128/genomeA.01282-14.
Rieske (2Fe-2S) protein	CP004380.1	Burkholderia pseudomallei 1026b chromosome 2, complete sequence.	Johnson et al., Genome Announc. 2015 Feb 12;3(1). pii: e01282-14. doi: 10.1128/genomeA.01282-14.
Rieske (2Fe-2S) protein	CP009163.1	Burkholderia pseudomallei K42 chromosome 2, complete sequence.	Johnson et al., Genome Announc. 2015 Feb 12;3(1). pii: e01282-14. doi: 10.1128/genomeA.01282-14.
Rieske (2Fe-2S) protein	CP009127.1	Burkholderia pseudomallei strain BSR chromosome 2, complete	Daligault et al., Genome Announc. 2014 Nov 20;2(6). pii: e01106-14. doi: 10.1128/genomeA.01106-14.
Rieske (2Fe-2S) protein	CP008835.1	Burkholderia pseudomallei strain BGR chromosome 2, complete	Daligault et al., Genome Announc. 2014 Nov 20;2(6). pii: e01106-14. doi: 10.1128/genomeA.01106-14.
Rieske (2Fe-2S) protein	CP008782.1	Burkholderia pseudomallei strain Mahidol-1106a chromosome 2,	Daligault et al., Genome Announc. 2014 Nov 20;2(6). pii: e01106-14. doi: 10.1128/genomeA.01106-14.
Rieske (2Fe-2S) protein	CP008778.1	Burkholderia pseudomallei 576 chromosome 2, complete sequence.	Daligault et al., Genome Announc. 2014 Nov 20;2(6). pii: e01106-14. doi: 10.1128/genomeA.01106-14.
Rieske (2Fe-2S) protein	CP008759.1	Burkholderia pseudomallei strain 1106a chromosome 2, complete	Daligault et al., Genome Announc. 2014 Nov 20;2(6). pii: e01106-14. doi: 10.1128/genomeA.01106-14.

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Rieske (2Fe-2S) protein	CP004043.1	Burkholderia pseudomallei MSHR146 chromosome 2, complete sequence.	Tuanyok et al., Submitted (14-JAN-2013) Genome Science, Los Alamos National
Rieske (2Fe-2S) protein	CP004024.1	Burkholderia pseudomallei MSHR511 chromosome 2, complete sequence.	Tuanyok et al., Submitted (14-JAN-2013) Genome Science, Los Alamos National
Rieske (2Fe-2S) protein	CP004002.1	Burkholderia pseudomallei NCTC 13178 chromosome 2, complete	Tuanyok et al., Submitted (18-DEC-2012) Genome Science, Los Alamos National
Rieske (2Fe-2S) protein	CP004004.1	Burkholderia pseudomallei NAU20B-16 chromosome 2, complete	Tuanyok et al., Submitted (18-DEC-2012) Genome Science, Los Alamos National
Rieske family iron-sulfur cluster-binding	CP003782.1	Burkholderia pseudomallei BPC006 chromosome II, complete sequence.	Fang et al J. Bacteriol. 194 (23), 6604-6605 (2012)
Rieske (2Fe-2S) protein	CP002834.1	Burkholderia pseudomallei 1026b chromosome 2, complete sequence.	Hayden et al PLoS ONE 7 (5), E36507 (2012)
Rieske (2Fe-2S) protein	CP000573.1	Burkholderia pseudomallei 1106a chromosome II, complete sequence.	Harkins et al., Submitted (13-FEB-2007) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
Rieske (2Fe-2S) protein	CP009586.1	Burkholderia pseudomallei strain PHLS 112 chromosome II, complete	Johnson et al., Genome Announc. 2015 Feb 12;3(1). pii: e01282-14. doi: 10.1128/genomeA.01282-14.
Rieske (2Fe-2S) protein	CP008893.1	Burkholderia pseudomallei HBPUB10303a chromosome 2, complete	Daligault et al., Genome Announc. 2014 Nov 20;2(6). pii: e01106-14. doi: 10.1128/genomeA.01106-14.
Rieske (2Fe-2S) protein	CP008763.1	Burkholderia pseudomallei strain MSHR346 chromosome 2, complete	Daligault et al., Genome Announc. 2014 Nov 20;2(6). pii: e01106-14. doi: 10.1128/genomeA.01106-14.
Rieske (2Fe-2S) protein	CP002600.1	Burkholderia gladioli BSR3 chromosome 2, complete sequence.	Seo et al., J. Bacteriol. 193 (12), 3149 (2011)
putative FAD-dependent monooxygenase	FN563149.1a	Rhodococcus equi 103S chromosome.	Letek et al., PLoS Genet. 6 (9), E1001145 (2010)
cytochrome P450 monooxygenase	FN563149.1b	Rhodococcus equi 103S chromosome.	Letek et al., PLoS Genet. 6 (9), E1001145 (2010)
iron-sulfur binding oxidoreductase	KJ598877.1	Rhodococcus equi strain U-S-A-18 clone S3A2 iron-sulfur binding oxidoreductase gene, complete cds.	Yeh et al., Submitted (19-MAR-2014) Graduate Institute of Biotechnology, National Chung Hsing University, 250 Kuo-Kuang Road, Taichung

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
iron-sulfur binding oxidoreductase	FN563149.1c	Rhodococcus equi 103S chromosome.	Letek et al., PLoS Genet. 6 (9), E1001145 (2010)
3-ketosteroid-9-alpha-hydroxylase oxygenase subunit	CP003949.1a	Rhodococcus opacus PD630, complete genome.	Chen et al Submitted (26-NOV-2012) National Laboratory of Macromolecules, Institute of Biophysics, Institute of Biophysics, Chinese Academy of Sciences, Beijing, 15 Datun Road, Chaoyang District, Beijing, Beijing 100101, China
3-ketosteroid 9alpha-hydroxylase component KshA	AP011115.1a	Rhodococcus opacus B4 DNA, complete genome.	Takarada et al., Submitted (30-MAR-2009) Contact:Director-General Department of Biotechnology National Institute of Technology and Evaluation, NITE Genome Analysis Center (NGAC), Department of Biotechnology; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan URL :http://www.bio.nite.go.jp/
3-ketosteroid-9-alpha-hydroxylase	CP011341.1a	Rhodococcus aetherivorans strain IcdP1, complete genome.	Qu et al., Genome Announc 3 (4), e00711-15 (2015)
ketosteroid-9-alpha-hydroxylase, oxygenase	CP000431.1a	Rhodococcus jostii RHA1, complete genome.	McLeod et al., Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
3-ketosteroid-9-alpha-hydroxylase	CP008947.1a	Rhodococcus opacus strain R7 sequence.	Di Gennaro et al., Genome Announc 2 (4) (2014)
KshA-like protein	CP011269.1a	Mycobacterium fortuitum strain CT6, complete genome.	Costa et al., MBio 6 (6), e01520-15 (2015)
3-ketosteroid-9-alpha-hydroxylase	CP014258.1a	Mycobacterium fortuitum subsp. fortuitum DSM 46621 = ATCC 6841	Singh et al., Submitted (30-NOV-2015) Microbiology, CSIR-CDRI, Sitapur Road, Lucknow, Uttar Pradesh 226021, India
3-ketosteroid-9-alpha-hydroxylase oxygenase subunit	CP009914.1a	Mycobacterium sp. VKM Ac-1817D, complete genome.	Bragin et al., J. Steroid Biochem. Mol. Biol. 138, 41-53 (2013
hypothetical protein	FN563149.1	Rhodococcus equi 103S chromosome.	Letek et al., PLoS Genet. 2010 Sep 30;6(9):e1001145. doi: 10.1371/journal.pgen.1001145.

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
3-ketosteroid-9-alpha-hydroxylase	CP011341.1b	Rhodococcus aetherivorans strain IcdP1, complete genome.	Qu et al., Genome Announc 3 (4), e00711-15 (2015)
3-ketosteroid 9alpha-hydroxylase oxygenase	HQ425873.1	Rhodococcus rhodochrous strain DSM 43269 hypothetical protein gene, partial cds; and 3-ketosteroid 9alpha-hydroxylase oxygenase(kshA1), putative ferredoxin (fd1), hypothetical protein, and 3-ketosteroid 9alpha-hydroxylase reductase (kshB) genes, complete cds.	Petrusma et al., Submitted (22-OCT-2010) Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), Kerklaan 30, Haren 9751 NN, The Netherlands
3-ketosteroid-9-alpha-hydroxylase oxygenase subunit	CP006996.1	Rhodococcus pyridinivorans SB3094, complete genome.	Dueholm et al., Genome Announc 2 (3), e00525-14 (2014)
3-ketosteroid-9-alpha-hydroxylase	CP011295.1	Rhodococcus erythropolis strain BG43, complete genome.	Ruckert et al., Submitted (17-APR-2015) CeBiTec, Bielefeld University, Sequenz 1, Bielefeld, NRW 33615, Germany
3-ketosteroid 9alpha-hydroxylase component KshA	CP003761.1	Rhodococcus erythropolis CCM2595, complete genome.	Strnad et al., Genome Announc. 2014 Mar 20;2(2). pii: e00208-14. doi: 10.1128/genomeA.00208-14.
probable 3-ketosteroid 9alpha-hydroxylase component KshA	AP008957.1	Rhodococcus erythropolis PR4 DNA, complete genome.	Sekine et al Environ. Microbiol. 8 (2), 334-346 (2006
probable dioxygenase Rieske iron-sulfur component	CP000431.1b	Rhodococcus jostii RHA1, complete genome.	McLeod et al., Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
3-ketosteroid 9alpha-hydroxylase component KshA2	AP011115.1b	Rhodococcus opacus B4 DNA, complete genome.	Takarada et al., Submitted (30-MAR-2009) Contact:Director-General Department of Biotechnology National Institute of Technology and Evaluation, NITE Genome Analysis Center (NGAC), Department of Biotechnology; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan URL :http://www.bio.nite.go.jp/

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
3-ketosteroid-9-alpha-hydroxylase	CP007255.1	Rhodococcus erythropolis R138, complete genome.	Kwasiborski et al Genome Announc 2 (2) (2014)
3-ketosteroid-9-alpha-hydroxylase oxygenase subunit	CP003949.1b	Rhodococcus opacus PD630, complete genome.	Chen et al., Submitted (26-NOV-2012) National Laboratory of Macromolecules, Institute of Biophysics, Institute of Biophysics, Chinese Academy of Sciences, Beijing, 15 Datun Road, Chaoyang District, Beijing, Beijing 100101, China
3-ketosteroid-9-alpha-hydroxylase	CP008947.1b	Rhodococcus opacus strain R7 sequence.	Di Gennaro et al., Genome Announc. 2014 Aug 21;2(4). pii: e00827-14. doi: 10.1128/genomeA.00827-14.
putative iron-sulfur binding oxidoreductase	FN563149.1e	Rhodococcus equi 103S chromosome.	Letek et al., PLoS Genet. 2010 Sep 30;6(9):e1001145. doi: 10.1371/journal.pgen.1001145.
3-ketosteroid 9alpha-hydroxylase oxygenase	HQ425877.1	Rhodococcus rhodochrous strain DSM 43269 hypothetical proteins and 3-ketosteroid 9alpha-hydroxylase oxygenase (kshA5) genes, complete cds; and hypothetical protein gene, partial cds.	Petrusma et al., Submitted (22-OCT-2010) Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), Kerklaan 30, Haren 9751 NN, The Netherlands
3-ketosteroid-9-alpha-hydroxylase oxygenase subunit	LN868938.1	Nocardia farcinica genome assembly NCTC11134, chromosome : 1.	Submitted (09-MAR-2015) SC, Wellcome Trust Sanger Institute, CB10 1SA, United Kingdom
putative terminal oxygenase	AP006618.1	Nocardia farcinica IFM 10152 DNA, complete genome.	Ishikawa et al., Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14925-14930 (2004)
3-ketosteroid 9alpha-hydroxylase component KshA	AP011115.1c	Rhodococcus opacus B4 DNA, complete genome.	Takarada et al., Submitted (30-MAR-2009) Contact:Director-General Department of Biotechnology National Institute of Technology and Evaluation, NITE Genome Analysis Center (NGAC), Department of Biotechnology; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan URL :http://www.bio.nite.go.jp/
3-ketosteroid-9-alpha-hydroxylase	CP008947.1c	Rhodococcus opacus strain R7 sequence.	Di Gennaro et al., Genome Announc 2 (4) (2014)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Ketosteroid-9-alpha-hydroxylase, oxygenase	FO082843.1	Nocardia cyriacigeorgica GUH-2 chromosome complete genome.	Zoropogui et al., J. Bacteriol. 194 (8), 2098-2099 (2012)
probable dioxygenase Rieske iron-sulfur	CP000431.1c	Rhodococcus jostii RHA1, complete genome.	McLeod et al., Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
3-ketosteroid-9-alpha-hydroxylase oxygenase	CP003949.1c	Rhodococcus opacus PD630, complete genome.	Chen et al., Submitted (26-NOV-2012) National Laboratory of Macromolecules, Institute of Biophysics, Institute of Biophysics, Chinese Academy of Sciences, Beijing, 15 Datun Road, Chaoyang District, Beijing, Beijing 100101, China
3-ketosteroid-9-alpha-hydroxylase	CP011853.1	Gordonia sp. QH-11, complete genome.	Jin et al., Submitted (16-JUN-2015) Key Laboratory of Environmental Biotechnology, Research Center for Eco-Environmental Sciences, Shuangqing Road 18, Beijing 100085, China
2Fe-2S)-binding protein	CP014646.1	Thauera humireducens strain SgZ-1, complete genome.	Ma et al., Submitted (28-JUN-2015) Guangdong Institute of Eco-Environmental and Soil Science, Tianyuan road, Guangzhou 510650, China
Rieske (2Fe-2S) domain protein	CP000511.1a	Mycobacterium vanbaalenii PYR-1, complete genome.	Copeland et al., Submitted (13-DEC-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Rieske (2Fe-2S	CP002385.1a	Mycobacterium gilvum Spyr1, complete genome.	Kallimanis et al., Stand Genomic Sci 5 (1), 144-153 (2011)
Rieske (2Fe-2S)	CP003053.1	Mycobacterium chubuense NBB4, complete genome.	Lucas et al., Submitted (05-JUN-2012) US DOE Joint Genome Institute, 2800 Mitchell Drive B310, Walnut Creek, CA 94598-1698, USA
ring-hydroxylating dioxygenase, large terminal subunit	CP003078.1a	Mycobacterium sp. JS623, complete genome.	Lucas et al., Submitted (07-OCT-2011) US DOE Joint Genome Institute, 2800 Mitchell Drive B310, Walnut Creek, CA 94598-1698, USA

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
2Fe-2S ferredoxin	CP014258.1b	Mycobacterium fortuitum subsp. fortuitum DSM 46621 = ATCC 6841	Singh et al., Submitted (30-NOV-2015) Microbiology, CSIR-CDRI, Sitapur Road, Lucknow, Uttar Pradesh 226021, India
Rieske (2Fe-2S) domain-containing protein	CP009914.1b	Mycobacterium sp. VKM Ac-1817D, complete genome.	Bragin et al., J Steroid Biochem Mol Biol. 2013 Nov;138:41-53. doi: 10.1016/j.jsbmb.2013.02.016. Epub 2013 Mar 6.
2Fe-2S ferredoxin	CP011022.1	Mycobacterium sp. NRRL B-3805, complete genome.	Rodriguez-Garcia et al., Submitted (06-MAR-2015) Proteomics, INBIOTEC (Instituto de Biotecnologia de Leon), Parque Cientifico de Leon. Avda. Real, 1, Leon, Leon 24006, Spain
2Fe-2S ferredoxin	CP006936.2	Mycobacterium neoaurum VKM Ac-1815D, complete genome.	Shtratnikova et al., Genome Announc 2 (1) (2014)
2Fe-2S ferredoxin	CP009496.1	Mycobacterium smegmatis strain INHR2, complete genome.	Mohan et al., Genome Announc 3 (1) (2015)
2Fe-2S ferredoxin	CP009495.1	Mycobacterium smegmatis strain INHR1, complete genome.	Mohan et al., Genome Announc 3 (1) (2015)
2Fe-2S ferredoxin	CP009494.1	Mycobacterium smegmatis str. MC2 155, complete genome.	Mohan et al., Genome Announc 3 (1) (2015)
2Fe-2S ferredoxin	CP001663.1	Mycobacterium smegmatis str. MC2 155, complete genome.	Perrodou et al., Nucleic Acids Res. 34 (DATABASE ISSUE), D338-D343 (2006)
2Fe-2S ferredoxin	CP000480.1	Mycobacterium smegmatis str. MC2 155, complete genome.	Fleischmann et al., Submitted (19-OCT-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
2Fe-2S ferredoxin	LN831039.1	Mycobacterium smegmatis genome assembly NCTC8159, chromosome : 1.	Submitted (09-MAR-2015) SC, Wellcome Trust Sanger Institute, CB10 1SA, United Kingdom
2Fe-2S ferredoxin	CP010114.1a	Mycobacterium avium subsp. paratuberculosis strain E93, complete	Amin et al., Submitted (28-NOV-2014) Pathobiological Sciences, University of Wisconsin-Madison, 1656 Linden Dr, Madison, WI 53706, USA
2Fe-2S ferredoxin	CP010113.1a	Mycobacterium avium subsp. paratuberculosis strain E1, complete	Amin et al., Submitted (28-NOV-2014) Pathobiological Sciences, University of Wisconsin-Madison, 1656 Linden Dr., Madison, WI 53706, USA

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
putative oxidoreductase, rieske (2Fe-2S	CP005928.1a	Mycobacterium avium subsp. paratuberculosis MAP4, complete genome.	Bannantine et al., Genome Announc 2 (1) (2014)
hypothetical protein	AE016958.1a	Mycobacterium avium subsp. paratuberculosis str. k10, complete	Li et al., Proc. Natl. Acad. Sci. U.S.A. 102 (35), 12344-12349 (2005)
2Fe-2S ferredoxin	CP009614.1a	Mycobacterium avium subsp. avium strain DJO-44271, complete genome.	Hazbon et al., Submitted (14-OCT-2014) Bacterial Collections, BEI Resources/American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110, USA
2Fe-2S ferredoxin	CP009482.1a	Mycobacterium avium subsp. avium 2285 (S), complete genome.	Hazbon et al., Submitted (22-SEP-2014) Bacterial Collections, BEI Resources/American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110, USA
2Fe-2S ferredoxin	CP009493.1a	Mycobacterium avium subsp. avium 2285 (R), complete genome.	Hazbon et al., Submitted (22-SEP-2014) Bacterial Collections, BEI Resources/American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110, USA
2Fe-2S ferredoxin	CP000479.1a	Mycobacterium avium 104, complete genome.	Fleischmann et al., Submitted (19-OCT-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
hypothetical protein	CP011269.1b	Mycobacterium fortuitum strain CT6, complete genome.	Costa et al., MBio 6 (6), e01520-15 (2015)
2Fe-2S ferredoxin	AP012555.1a	Mycobacterium avium subsp. hominissuis TH135 chromosomal DNA,	Uchiya et al., PLoS ONE 8 (8), E71831 (2013)
2Fe-2S ferredoxin	CP002275.1	Mycobacterium indicus pranii MTCC 9506, complete genome.	Ahmed et al., PLoS ONE 2 (10), E968 (2007)
rieske (2Fe-2S)	CP003324.1	Mycobacterium intracellulare MOTT-64, complete genome.	Kim et al., J. Bacteriol. 194 (12), 3268 (2012)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
3-ketosteroid-9-alpha-monooxygenase oxygenase subunit	CP014475.1	Mycobacterium phlei strain CCUG 21000, complete genome.	Das et al., Genome Biol Evol (2016) In press
rieske (2Fe-2S)	CP003347.1a	Mycobacterium sp. 05-1390, complete genome.	Kim et al., Genome Announc 1 (4), e00604-13 (2013)
rieske (2Fe-2S)	CP003323.1	Mycobacterium intracellulare MOTT-02, complete genome.	Kim et al., J. Bacteriol. 194 (10), 2771 (2012)
rieske (2Fe-2S)	CP003322.1	Mycobacterium intracellulare ATCC 13950, complete genome.	Kim et al., J. Bacteriol. 194 (10), 2750 (2012)
rieske (2Fe-2S)	CP012150.1	Mycobacterium goodii strain X7B, complete genome.	Yu et al., Submitted (24-JUL-2015) CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences, No.1 Beichen West Road, Chaoyang District, Beijing 100101, China
rieske (2Fe-2S)	CP003491.1a	Mycobacterium sp. MOTT36Y, complete genome.	Kim et al., J. Bacteriol. 194 (15), 4141-4142 (2012)
oxidoreductase	CP002329.1	Mycobacterium sinense strain JDM601, complete genome.	Zhang et al., J. Bacteriol. 193 (16), 4300-4301 (2011)
rieske (2Fe-2S)	CP009499.1a	Mycobacterium intracellulare 1956, complete genome.	Hazbon et al., Submitted (09-SEP-2014) Bacterial Collections, BEI Resources/American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110, USA
Rieske (2Fe-2S) domain-containing protein	CP003053.1	Mycobacterium chubuense NBB4, complete genome.	Lucas et al., Submitted (05-JUN-2012) US DOE Joint Genome Institute, 2800 Mitchell Drive B310, Walnut Creek, CA 94598-1698, USA
3-ketosteroid-9-alpha-hydroxylase	CP014258.1c	Mycobacterium fortuitum subsp. fortuitum DSM 46621 = ATCC 6841	Singh et al., Submitted (02-FEB-2016) Microbiology, CSIR-CDRI, Sitapur Road, Lucknow, Uttar Pradesh 226021, India
Rieske (2Fe-2S) domain-containing protein	CP009914.1c	Mycobacterium sp. VKM Ac-1817D, complete genome.	Bragin et al., J. Steroid Biochem. Mol. Biol. 138, 41-53 (2013)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
KshA-like protein	CP011269.1c	Mycobacterium fortuitum strain CT6, complete genome.	Costa et al., MBio 6 (6), e01520-15 (2015)
3-ketosteroid-9-alpha-hydroxylase	CP011773.1	Mycobacterium sp. EPa45, complete genome.	Kato et al., Genome Announc 3 (4), e00782-15 (2015)
oxygenase KshA	CP003347.1b	Mycobacterium sp. 05-1390, complete genome.	Kato et al., Genome Announc 1 (4), e00604-13 (2013)
ring-hydroxylating dioxygenase, large terminal	CP002385.1b	Mycobacterium gilvum Spyr1, complete genome.	Kallimanis et al., Stand Genomic Sci 5 (1), 144-153 (2011)
3-ketosteroid-9-alpha-hydroxylase	CP009614.1b	Mycobacterium avium subsp. avium strain DJO-44271, complete genome.	Hazbon et al., Submitted (14-OCT-2014) Bacterial Collections, BEI Resources/American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110, USA
3-ketosteroid-9-alpha-hydroxylase	CP009482.1b	Mycobacterium avium subsp. avium 2285 (S), complete genome.	Hazbon et al., Submitted (22-SEP-2014) Bacterial Collections, BEI Resources/American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110, USA
3-ketosteroid-9-alpha-hydroxylase	CP009493.1b	Mycobacterium avium subsp. avium 2285 (R), complete genome.	Hazbon et al., Submitted (22-SEP-2014) Bacterial Collections, BEI Resources/American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110, USA
oxygenase KshA	CP000479.1b	Mycobacterium avium 104, complete genome.	Fleischmann et al., Submitted (19-OCT-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
Rieske (2Fe-2S) domain protein	CP000656.1	Mycobacterium gilvum PYR-GCK, complete genome.	Badejo et al., PLoS ONE 8 (2), E58066 (2013)
oxygenase KshA	CP003491.1b	Mycobacterium sp. MOTT36Y, complete genome.	Kim et al., J. Bacteriol. 194 (15), 4141-4142 (2012)
3-ketosteroid-9-alpha-hydroxylase	CP010114.1b	Mycobacterium avium subsp. paratuberculosis strain E93, complete	Amin et al., Microbiology (Reading, Engl.) 161 (PT 4), 807-818 (2015)
3-ketosteroid-9-alpha-hydroxylase	CP010113.1b	Mycobacterium avium subsp. paratuberculosis strain E1, complete	Amin et al., Microbiology (Reading, Engl.) 161 (PT 4), 807-818 (2015)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
oxidoreductase	CP005928.1b	Mycobacterium avium subsp. paratuberculosis MAP4, complete genome.	Bannantine et al., Genome Announc. 2014 Feb 6;2(1). pii: e01252-13. doi: 10.1128/genomeA.01252-13.
hypothetical protein	AE016958.1b	Mycobacterium avium subsp. paratuberculosis str. k10, complete	Li et al., Proc Natl Acad Sci U S A. 2005 Aug 30;102(35):12344-9. Epub 2005 Aug 22.
3-ketosteroid-9-alpha-hydroxylase	CP009499.1b	Mycobacterium intracellulare 1956, complete genome.	Hazbon et al., Submitted (09-SEP-2014) Bacterial Collections, BEI Resources/American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110, USA
oxygenase KshA	AP012555.1b	Mycobacterium avium subsp. hominissuis TH135 chromosomal DNA,	Uchiya et al., PLoS One. 2013 Aug 21;8(8):e71831. doi: 10.1371/journal.pone.0071831. eCollection 2013.
Rieske (2Fe-2S) domain protein	CP000518.1	Mycobacterium sp. KMS, complete genome.	Copeland et al., Submitted (18-DEC-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Rieske (2Fe-2S) region	CP000384.1	Mycobacterium sp. MCS, complete genome.	Copeland et al., Submitted (05-JUN-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Rieske (2Fe-2S) domain protein	CP000580.1	Mycobacterium sp. JLS, complete genome.	Copeland et al., Submitted (20-FEB-2007) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Rieske (2Fe-2S) domain-containing protein	CP003078.1b	Mycobacterium sp. JS623, complete genome.	Lucas et al., Submitted (07-OCT-2011) US DOE Joint Genome Institute, 2800 Mitchell Drive B310, Walnut Creek, CA 94598-1698, USA
Rieske (2Fe-2S) domain protein	CP000511.1b	Mycobacterium vanbaalenii PYR-1, complete genome.	Copeland et al., Submitted (13-DEC-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Rieske (2Fe-2S) region	CP000656.1b	Mycobacterium gilvum PYR-GCK, complete genome.	Badejo et al., PLoS ONE 8 (2), E58066 (2013)
Rieske (2Fe-2S) domain-containing protein	CP002385.1c	Mycobacterium gilvum Spyr1, complete genome.	Kallimanis et al., Stand Genomic Sci. 2011 Oct 15;5(1):144-53. doi: 10.4056/sigs.2265047. Epub 2011 Oct 1
hydroxylase of 3-hydroxy-9, 10-	LC010134.1b	Comamonas testosteroni DNA, putative mega-cluster of steroid degradation genes, strain: T441.	Horinouchi et al., Microbiology. 2001 Dec;147(Pt 12):3367-75.

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
secoandrosta -1, 3, 5(10)-triene-9, 17-dione, flavin-dependent monooxygenase oxygenase subunit			
Acyl-CoA dehydrogenase, type 2-like protein	CP001220.2	Comamonas testosteroni CNB-2, complete genome.	Ma et al., Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
flavin-dependent monooxygenase oxygenase subunit	CP006704.1	Comamonas testosteroni TK102, complete genome.	Fukuda et al., Genome Announc 2 (5), e00865-14 (2014)
hydroxylase of 3-hydroxy-9, 10-secoandrosta -1, 3, 5(10)-triene-9, 17-dione	LC010134.1c	Comamonas testosteroni DNA, putative mega-cluster of steroid t441	Horinouchi et al., Microbiology. 2001 Dec;147(Pt 12):3367-75.
chromosome I	LN879547.1b	Comamonas testosteroni P19 genome assembly Comamonas testosteroni P19, chromosome : I	Park Submitted (15-SEP-2015) JEJU NATIONAL UNIVERSITY, Lab of Microbial Ecology and Genomics, Dept. of Biology, Jeju National University, 102 Jejudaehak-ro, Jeju 690-756, Jeju, 690-756, South Korea
flavin reductase-like, FMN-binding protein	CP001220.2b	Comamonas testosteroni CNB-2, complete genome	Fukuda et al., Genome Announc 2 (5), e00865-14 (2014)
flavin reductase	CP006704.1b	Comamonas testosteroni TK102, complete genome.	Fukuda et al., Genome Announc. 2014 Sep 11;2(5). pii: e00865-14. doi: 10.1128/genomeA.00865-14.
ORF11	AB063482.1c	Comamonas testosteroni ORF11, tesA, tesD, tesE, tesF, tesG genes, complete cds.	Horinouchi et al., Appl. Environ. Microbiol. 69 (4), 2139-2152 (2003)
3-ketosteroid-9-alpha-hydroxylase reductase subunit	LC010134.1d	Comamonas testosteroni DNA, putative mega-cluster of steroid degradation genes, strain: T441.	Horinouchi et al., Microbiology (Reading, Engl.) 147 (PT 12), 3367-3375 (2001)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
chromosome I	LN879547.1c	Comamonas testosteroni P19 genome assembly Comamonas testosteroni P19, chromosome : I.	Park Submitted (15-SEP-2015) JEJU NATIONAL UNIVERSITY, Lab of Microbial Ecology and Genomics, Dept. of Biology, Jeju National University, 102 Jejudaehak-ro, Jeju 690-756, Jeju, 690-756, South Korea
FAD-binding oxidoreductase	CP001220.2c	Comamonas testosteroni CNB-2, complete genome	Ma et al., Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
3-ketosteroid-9-alpha-hydroxylase reductase subunit	CP006704.1d	Comamonas testosteroni TK102, complete genome.	Fukuda et al., Genome Announc 2 (5), e00865-14 (2014)
FAD-binding oxidoreductase	CP001220.2d	Comamonas testosteroni CNB-2, complete genome	Ma et al., Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
chromosome i	LN879547.1d	Comamonas testosteroni P19 genome assembly Comamonas testosteroni P19, chromosome : I.	Park Submitted (15-SEP-2015) JEJU NATIONAL UNIVERSITY, Lab of Microbial Ecology and Genomics, Dept. of Biology, Jeju National University, 102 Jejudaehak-ro, Jeju 690-756, Jeju, 690-756, South Korea
Acyl-CoA dehydrogenase, type 2-like protein	CP001220.2e	Comamonas testosteroni CNB-2, complete genome	Ma et al., Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
flavin-dependent monooxygenase oxygenase subunit	CP006704.1e	Comamonas testosteroni TK102, complete genome.	Fukuda et al., Genome Announc 2 (5), e00865-14 (2014)
3-ketosteriod isomerase	AB489116	Comamonas testosteroni genes for 3-ketosteroiod isomerase, 3-alpha steroid dehydrogenase, complete cds, strain: TA441.	Horinouchi,M. J. Steroid Biochem. Mol. Biol. 122 (4), 253-263 (2010)
ketosteroid isomerase	NZ_AKCL01000160	Pseudomonas putida SJTE-1 contig000160, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
ketosteroid isomerase	NZ_AKCL01000166	Pseudomonas putida SJTE-1 contig000166, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
steroid delta-isomerase	NZ_AFMP01000031	Sphingomonas sp. KC8 contig31, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
steroid delta-isomerase	NZ_AFMP01000031	Sphingomonas sp. KC8 contig31, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
putative 3-ketosteroid-5-isomerase	NZ_AFMP01000031	Sphingomonas sp. KC8 contig31, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
steroid delta-isomerase	NZ_AFMP01000018	Sphingomonas sp. KC8 contig18, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
steroid delta-isomerase	LC010134	Comamonas testosteroni DNA, putative mega-cluster of steroid degradation genes, strain: T441.	Horinouchi Microbiology (Reading, Engl.) 147 (PT 12), 3367-3375 (2001)
conserved hypothetical protein	CP001220	Comamonas testosteroni CNB-2, complete genome.	Ma Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
ketosteroid isomerase	M22749	Pseudomonas testosteroni ketosteroid isomoerase (ksi) gene, complete cds.	Choi Gene 69 (1), 121-129 (1988)
delta-5-3-ketosteroid isomerase	J03568	P.testosteroni delta-5-3-ketosteroid isomerase gene, complete cds.	Kuliopulos Proc. Natl. Acad. Sci. U.S.A. 84 (24), 8893-8897 (1987)
steroid delta-isomerase	CP006704	Comamonas testosteroni TK102, complete genome.	Fukuda Genome Announc 2 (5), e00865-14 (2014)
TesF	AB063482	Comamonas testosteroni ORF11, tesA, tesD, tesE, tesF, tesG genes, complete cds.	Horinouchi et al., Appl. Environ. Microbiol. 69 (4), 2139-2152 (2003)
TesF (acetaldehyde dehydrogenase)	LC010134	Comamonas testosteroni DNA, putative mega-cluster of steroid degradation genes, strain: T441.	Horincouchi Microbiology (Reading, Engl.) 147 (PT 12), 3367-3375 (2001)
Acetaldehyde dehydrogenase	CP001220	Comamonas testosteroni CNB-2, complete genome.	Ma Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
Acetaldehyde dehydrogenase	CP001220	Comamonas testosteroni CNB-2, complete genome.	Ma Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
acetaldehyde dehydrogenase	CP006704	Comamonas testosteroni TK102, complete genome.	Fukuda Genome Announc 2 (5), e00865-14 (2014)
TesG (possibly 4-hydroxy-2oxovalerate aldolase)	AB063482	Comamonas testosteroni ORF11, tesA, tesD, tesE, tesF, tesG genes, complete cds.	Horinouchi,M Appl. Environ. Microbiol. 69 (4), 2139-2152 (2003)
4-hyroxy-2-oxovalerate aldolase	CP006704	Comamonas testosteroni TK102, complete genome.	Fukuda Genome Announc 2 (5), e00865-14 (2014)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
4-hyroxy-2-oxovalerate aldolase	CP006704	Comamonas testosteroni TK102, complete genome.	Fukuda Genome Announc 2 (5), e00865-14 (2014)
4-hyroxy-2-oxovalerate aldolase	LC010134	Comamonas testosteroni DNA, putative mega-cluster of steroid degradation genes, strain: T441	Horinouchi Microbiology (Reading, Engl.) 147 (PT 12), 3367-3375 (2001)
aphG 4-hydroxy-2-oxovalerate aldolase	AB029044	Comamonas testosteroni orfX, orfY, aphT, aphC, aphE, aphF, aphG, aphH, orfJ, aphI gene cluster for meta-pathway enzymes required for degradation of phenol, complete cds	Arai Microbiology (Reading, Engl.) 146 (PT 7), 1707-1715 (2000)
4-hydroxy-2-oxovalerate aldolase (tdnJ)	NG_035478	Pseudomonas putida UCC22 plasmid pTDN1 genes for conversion of aniline to catechol, aniline degradation lower-pathway gene cluster, regulatory protein and transposase, partial and complete cds	Fukumori Unpublished
TadJ 4-hydroxy-2-oxovalerate aldolase	AY940090	Delftia tsuruhatensis strain AD9 transposase (tnpA-L1) gene, partial cds; tad gene cluster, complete sequence; transcriptional regulator (orfX), hydrolase/acyltransferase (orfY), and muconate cycloisomerase (orfZ) genes, complete cds; and transposase (tnpA-L2) gene, partial cds	Liang Microbiology (Reading, Engl.) 151 (PT 10), 3435-3446 (2005)
4-hydroxy-2-oxovalerate aldolase	CP000539	Acidovorax sp. JS42, complete genome.	Copeland unpublished
4-hydroxy-2-oxovalerate aldolase	AF190463	Comamonas sp. JS765 hypothetical CdoX1 (cdoX1), hypothetical CmpX-like protein (cdoX2), hypothetical CdoFa (cdoFa), hypothetical CdoFb (cdoFb), hypothetical CdoR2 (cdoR2), hypothetical CdoX3(cdoX3), 2-hydroxymuconic semialdehyde dehydrogenase (cdoG), 2-oxo-4-pentenoate hydratase (cdoH), acetaldehyde dehydrogenase(cdoI), 4-hydroxy-2-oxovalerate aldolase (cdoJ), and4-oxalocrotonate decarboxylase	He J. Ind. Microbiol. Biotechnol. 34 (2), 99-104 (2007)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
		(cdoK) genes, complete cds; and CdoL (cdoL) gene, partial cds.	
4-hydroxy-2-oxovalerate aldolase (pcaJ)	FJ601374	Diaphorobacter sp. PCA039 pca gene cluster, complete sequence.	Zhang World J. Microbiol. Biotechnol. 26 (4), 665-673 (2010)
4-hydroxy-2-oxovalerate aldolase	FR687359	Burkholderia rhizoxinica HKI 454, complete genome.	Lackner J. Bacteriol. 193 (3), 783-784 (2011)
TesE	AB063482	Comamonas testosteroni ORF11, tesA, tesD, tesE, tesF, tesG genes, complete cds.	Horinouchi,M Appl. Environ. Microbiol. 69 (4), 2139-2152 (2003)
2-keto-4-pentenoate hydratase (hydratase/decarboxylase)	CP001220	Comamonas testosteroni CNB-2, complete genome.	Ma Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
2-keto-4-pentenoate hydratase	CP006704	Comamonas testosteroni TK102, complete genome.	Fukuda Genome Announc 2 (5), e00865-14 (2014)
2-hydroxyhexa-2,4-dienoate hydratase	LC010134	Comamonas testosteroni DNA, putative mega-cluster of steroid degradation genes, strain: T441.	Horinouchi Microbiology (Reading, Engl.) 147 (PT 12), 3367-3375 (2001)
TesD/ORF13	AB063482	Comamonas testosteroni ORF11, tesA, tesD, tesE, tesF, tesG genes, complete cds.	Horinouchi,M Appl. Environ. Microbiol. 69 (4), 2139-2152 (2003)
TesD (hydrolase for 4, 5-9, 10-Diseco-3-hydroxy-5, 9, 17-trioxoandrosta-1(10), 2-dien-4-oic acid)	LC010134	Comamonas testosteroni DNA, putative mega-cluster of steroid degradation genes, strain: T441.	Horinouchi Microbiology (Reading, Engl.) 147 (PT 12), 3367-3375 (2001)
alpha/beta hydrolase fold protein	CP001220	Comamonas testosteroni CNB-2, complete genome.	Ma Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
3-oxoacyl-ACP reductase	CP006704	Comamonas testosteroni TK102, complete genome.	Fukuda Genome Announc 2 (5), e00865-14 (2014)
Chromosome I	LN879547.1	Comamonas testosteroni P19 genome assembly Comamonas testosteroni P19, chromosome : I	Park,Soo.-Je.
2-hydroxymuconic semialdehyde hydrolase		Rhodococcus equi ATCC13557	This thesis
putative alpha/beta hydrolase	FN563149.1	Rhodococcus equi 103S chromosome	Letek et al., PLoS Genet. 2010 Sep 30;6(9):e1001145. doi: 10.1371/journal.pgen.1001145.
hypothetical protein	AB117721.1	Rhodococcus rhodochrous orf31, orf32, bphC3 genes for hypothetical proteins and 2, 3-dihydroxybiphenyl 1, 2-dioxygenase, partial and complete cds.	Taguchi et al., Biosci Biotechnol Biochem. 2004 Apr;68(4):787-95.
Oxosteriod 1-dehydrogenase	LGTW01000000	Streptomyces scabiei DNA, scaffold: NODE_015_cov_67.063, strain: S58, whole genome shotgun sequence	de Man,T.J., Perry,K.A., Coulliette,A.D., Jensen,B., Toney,N.C., Limbago,B.M. and Noble-Wang,J. Genome Announc 4 (2) (2016)
3- Oxosteriod 1-dehydrogenase	BCMM01000015	Streptomyces scabiei DNA, scaffold: NODE_015_cov_67.063, strain: S58, whole genome shotgun sequence	Submitted (17-NOV-2015) Contact:Tsuyoshi Tomihama Kagoshima Prefectural Institute for Agricultural Development, Plant Pathology and Entomology; Oono2200, Minamisatsuma, Kagoshima 899-3401, Japan
Oxosteriod 1-dehydrogenase	BCMM01000026	Streptomyces scabiei DNA, scaffold: NODE_026_cov_60.1768, strain: S58, whole genome shotgun sequence	Submitted (17-NOV-2015) Contact:Tsuyoshi Tomihama Kagoshima Prefectural Institute for Agricultural Development, Plant Pathology and Entomology; Oono2200, Minamisatsuma, Kagoshima 899-3401, Japan
3- Oxosteriod 1-dehydrogenase	LMXT01000039.1	Comamonas testosteroni strain WDL7 contig_68, whole genome shotgun sequence	Submitted (31-OCT-2015) Lab of Soil and Water Management, KU Leuven, Kasteelpark Arenberg, Heverlee 3001, Belgium

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
3-oxosteroid 1-dehydrogenase	LLWJ01000045.1	Pseudomonas sp. TAA207 contig000045, whole genome shotgun sequence	Submitted (20-OCT-2015) University of Florence, Via Madonna del Piano 6, Sesto Fiorentino, Italy
3-oxosteroid 1-dehydrogenase	LLWI01000055.1	Pseudomonas sp. TAD18 contig000055, whole genome shotgun sequence	Submitted (20-OCT-2015) Dept. of Biology, University of Florence, Via Madonna del Piano 6, Sesto Fiorentino, Italy
3-oxosteroid 1-dehydrogenase	BCMK01000015.1	Streptomyces acidiscabies DNA, scaffold: NODE_015_cov_60.072, strain: a10, whole genome shotgun sequence	Tomihama,T., Ikenaga,M., Sakai,M., Okubo,T. and Ikeda,S. Genome Announc 4 (2), e00062-16 (2016)
3-oxosteroid 1-dehydrogenase	BCMN01000006.1	Streptomyces turgidiscabies DNA, scaffold: NODE_006_cov_57.7002, strain: T45, whole genome shotgun sequence	Tomihama,T., Ikenaga,M., Sakai,M., Okubo,T. and Ikeda,S Genome Announc 4 (2), e00062-16 (2016)
3-oxosteroid 1-dehydrogenase	BCMN01000025.1	Streptomyces turgidiscabies DNA, scaffold: NODE_025_cov_64.0769, strain: T45, whole genome shotgun sequence	Tomihama,T., Ikenaga,M., Sakai,M., Okubo,T. and Ikeda,S Genome Announc 4 (2), e00062-16 (2016)
3-oxosteroid 1-dehydrogenase	BCMN01000034.1	Streptomyces turgidiscabies DNA, scaffold: NODE_034_cov_67.2766, strain: T45, whole genome shotgun sequence	Tomihama,T., Ikenaga,M., Sakai,M., Okubo,T. and Ikeda,S Genome Announc 4 (2), e00062-16 (2016)
3-oxosteroid 1-dehydrogenase	CP012901.1	Pseudomonas aeruginosa strain N15-01092, complete sequence	Mataseje et al., Submitted (13-OCT-2015) Antimicrobial Resistance and Nosocomial Infections, Public Health Agency of Canada, 1015 Arlington Street, Winnipeg, MB, Canada
3-oxosteroid 1-dehydrogenase	LOEI01000003.1	Klebsiella pneumoniae strain OC511 AOT23_contig000003, whole genome shotgun sequence	Lopez,L.L., Rusconi,B., Mclaughlin,M., Qi,C., Scheetz,M.H., Seshu,J. and Eppinger,M. Genome Announc 4 (2) (2016)
3-oxosteroid 1-dehydrogenase	LOEJ01000018.1	Klebsiella pneumoniae strain K1 AOT20_contig000018, whole genome shotgun sequence	Lopez,L.L., Rusconi,B., Mclaughlin,M., Qi,C., Scheetz,M.H., Seshu,J. and Eppinger,M. Genome Announc 4 (2) (2016)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
3-oxosteroid 1-dehydrogenase	LOEF01000003.1	Klebsiella pneumoniae strain OC217 AOT21_contig000003, whole genome shotgun sequence	Lopez,L.L., Rusconi,B., Mclaughlin,M., Qi,C., Scheetz,M.H., Seshu,J. and Eppinger,M. Genome Announc 4 (2) (2016)
3-oxosteroid 1-dehydrogenase	LOEH01000004.1	Klebsiella pneumoniae strain OC648 AOT24_contig000004, whole genome shotgun sequence	Lopez,L.L., Rusconi,B., Mclaughlin,M., Qi,C., Scheetz,M.H., Genome Announc 4 (2) (2016)
3-oxosteroid 1-dehydrogenase	LOEG01000006.1	Klebsiella pneumoniae strain Z3209 AOT25_contig000006, whole genome shotgun sequence	Lopez,L.L., Rusconi,B., Mclaughlin,M., Qi,C., Scheetz,M.H., Genome Announc 4 (2) (2016)
3-oxosteroid 1-dehydrogenase	CP007255.1	Rhodococcus erythropolis R138, complete genome	Kwasiborski,A., Mondy,S., Beury-Cirou,A. and Faure,D. Genome Announc 2 (2) (2014)
3-ketosteroid-delta-1-dehydrogenase	AZXY01000001.1	Rhodococcus pyridinivorans KG-16 contig00001, whole genome shotgun sequence	Aggarwal,R.K. and Dawar,C. Genome Announc 4 (1), e01704-15 (2016)
3-ketosteroid-delta-1-dehydrogenase	LNQH01000008.1	Rhodococcus enclensis strain NIO-1009 contig008, whole genome shotgun sequence	Dastager et al., Int J Syst Evol Microbiol. 2014 Aug;64(Pt 8):2693-9. doi: 10.1099/ijs.0.061390-0. Epub 2014 May 22.
catechol 1,2-dioxygenase (1a)	NZ_AKCL01000002	Pseudomonas putida SJTE-1 contig000002, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
catechol 1,2-dioxygenase	NZ_AKCL01000193	Pseudomonas putida SJTE-1 contig000193, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
catechol 2,3-dioxygenase	NC_008275	Pseudomonas putida MT53 plasmid pWW53, complete sequence	Miyakoshi, M et al Environ. Microbiol. 14 (11), 2946-2959 (2012) Yano, H et al J. Mol. Biol. 369 (1), 11-26 (2007) Tsuda J. Bacteriol. 183 (21), 6215-6224 (2001) Sentchilo Appl. Environ. Microbiol. 66 (7), 2842-2852 (2000) Gallegos J. Bacteriol. 179 (16), 5024-5029 (1997) Assinder J. Gen. Microbiol. 139 (3), 557-568 (1993) Assinder Nucleic Acids Res. 20 (20), 5476 (1992) Osbourne J. Gen. Microbiol. 134 (11), 2965-2975 (1988) Keil J. Gen. Microbiol. 133 (5), 1149-1158 (1987) Keil J. Bacteriol. 169 (2), 764-770 (1987) Keil J. Bacteriol. 164 (2), 887-895 (1985)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Catechol 2,3 dioxygenase	NC_003350	Pseudomonas putida plasmid pWW0, complete sequence.	Greated, A Environ. Microbiol. 4 (12), 856-871 (2002)
Catechol 2,3 dioxygenase	NC_007926	Pseudomonas putida plasmid NAH7, complete sequence.	Sota J. Bacteriol. 188 (11), 4057-4067 (2006) Tsuda Mol. Gen. Genet. 223 (1), 33-39 (1990)
Catechol 2,3 dioxygenase	NC_004999	Pseudomonas putida NCIB 9816-4 plasmid pDTG1, complete sequence.	Dennis J. Mol. Biol. 341 (3), 753-768 (2004)
Catechol 2,3 dioxygenase	NC_008275	Pseudomonas putida MT53 plasmid pWW53, complete sequence.	Miyakoshi, M Environ. Microbiol. 14 (11), 2946-2959 (2012) Yano, H et al J. Mol. Biol. 369 (1), 11-26 (2007) Tsuda J. Bacteriol. 183 (21), 6215-6224 (2001) Sentchilo Appl. Environ. Microbiol. 66 (7), 2842-2852 (2000) Gallegos J. Bacteriol. 179 (16), 5024-5029 (1997) Assinder J. Gen. Microbiol. 139 (3), 557-568 (1993) Assinder Nucleic Acids Res. 20 (20), 5476 (1992) Osbourne J. Gen. Microbiol. 134 (11), 2965-2975 (1988) Keil J. Gen. Microbiol. 133 (5), 1149-1158 (1987) Keil J. Bacteriol. 169 (2), 764-770 (1987) Keil J. Bacteriol. 164 (2), 887-895 (1985)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Catechol 2,3 dioxygenase	NC_014124	Pseudomonas putida plasmid pDK1, complete sequence.	Mikakoshi Environ. Microbiol. 14 (11), 2946-2959 (2012) Yano, H et al J. Bacteriol. 192 (17), 4337-4347 (2010) Assinder J. Gen. Microbiol. 139 (3), 557-568 (1993) Assinder Nucleic Acids Res. 20 (20), 5476 (1992) Benjamin J. Bacteriol. 173 (8), 2724-2728 (1991) Voss SAAS Bull. Biochem. Biotechnol. 3, 54-57 (1990) Shaw J. Gen. Microbiol. 134 (9), 2463-2474 (1988) Kunz J. Bacteriol. 148 (1), 72-82 (1981) Kunz J. Bacteriol. 146 (3), 952-964 (1981)
Catechol 2,3 dioxygenase	NC_010678	Ralstonia pickettii 12J chromosome 2, complete sequence.	Lucas, S Submitted (01-MAY-2008) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Catechol 2,3 dioxygenase	NC_007494	Rhodobacter sphaeroides 2.4.1 chromosome 2, complete sequence	Copeland A, Unpublished Kontur Submitted (21-OCT-2013) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Catechol 2,3 dioxygenase	NC_010175	Chloroflexus aurantiacus J-10-fl chromosome, complete genome.	Copeland A, Unpublished/Submitted (05-DEC-2007) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Catechol 2,3 dioxygenase	NC_012674	Pseudomonas fluorescens strain PC20 plasmid pNAH20, complete sequence	Merimaa Arch. Microbiol. 186 (4), 287-296 (2006) FEMS Microbiol. Ecol. 31 (3), 195-205 (2000)
Catechol 2,3 dioxygenase	NC_002033	Novosphingobium aromaticivorans plasmid pNL1, complete sequence.	Romine J. Bacteriol. 181 (5), 1585-1602 (1999)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Catechol 2,3 dioxygenase (a)	NC_002754	Sulfolobus solfataricus P2 chromosome, complete genome.	She Proc. Natl. Acad. Sci. U.S.A. 98 (14), 7835-7840 (2001) Charlebois Genome 43 (1), 116-136 (2000)
Catechol 2,3 dioxygenase	NC_008269	Rhodococcus jostii RHA1 plasmid pRHL1, complete sequence.	McLeod Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
Catechol 2,3 dioxygenase	NC_016644	Pseudomonas sp. MC1 plasmid KOPRI126573, complete sequence.	Ahn Unpublished/ Submitted (13-JUL-2011) Department of Systems Biology, Yonsei University, 50 Yonsei-Ro, Seodaemun-Gu, Seoul 120-749, Republic of Korea
Catechol 2,3 dioxygenase	NC_021250	Pseudomonas migulae strain D2RT plasmid pD2RT, complete sequence.	Jutkina Submitted (10-JUL-2012) Department of Genetics, Institute of Molecular and Cell Biology, University of Tartu, 23 Riia Street, Tartu 51010, Estonia
Catechol 2,3 dioxygenase	NC_013446	Comamonas testosteroni CNB-2, complete genome.	Ma Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
Catechol 2,3 dioxygenase	NC_018028	Pseudomonas stutzeri CCUG 29243, complete genome.	Brunet-Galmes J. Bacteriol. 194 (23), 6642-6643 (2012)
Catechol 2,3 dioxygenase	NC_000964	Bacillus subtilis subsp. subtilis str. 168 chromosome, complete genome.	Barbe, V Microbiology (Reading, Engl.) 155 (PT 6), 1758-1775 (2009) Eichenberger PLoS Biol. 2 (10), E328 (2004) Kunst Nature 390 (6657), 249-256 (1997) Mizuno Microbiology (Reading, Engl.) 142 (PT 11), 3103-3111 (1996)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Catechol 2,3 dioxygenase (a)	NC_008268	Rhodococcus jostii RHA1, complete genome.	McLeod Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
Catechol 2,3 dioxygenase	NC_012589	Sulfolobus islandicus L.S.2.15 chromosome, complete genome.	Reno Proc. Natl. Acad. Sci. U.S.A. 106 (21), 8605-8610 (2009)
Catechol 2,3 dioxygenase	NZ_KB900701	Bradyrhizobium elkanii USDA 76 BraelDRAFT_scaffold1.1, whole genome shotgun sequence	Kyrpides Submitted (16-APR-2013) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598-1698, USA
Catechol 2,3 dioxygenase	NC_006462	Thermus thermophilus HB8 plasmid pTT27, complete sequence.	Masui Submitted (11-NOV-2004) Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5, Takayamacho, Ikoma, Nara 630-0192, Japan
Catechol 2,3 dioxygenase	NC_007973	Cupriavidus metallidurans CH34, complete genome.	Copeland A Submitted (18-APR-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Catechol 2,3 dioxygenase	NC_012726	Sulfolobus islandicus M.16.4 chromosome, complete genome.	Reno Proc. Natl. Acad. Sci. U.S.A. 106 (21), 8605-8610 (2009)
Catechol 2,3 dioxygenase	NC_012623	Sulfolobus islandicus Y.N.15.51 chromosome, complete genome.	Reno Proc. Natl. Acad. Sci. U.S.A. 106 (21), 8605-8610 (2009)
Catechol 2,3 dioxygenase	NC_013769	Sulfolobus islandicus L.D.8.5 chromosome, complete genome.	Reno Unpublished Copeland Submitted (03-SEP-2009) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Catechol 2,3 dioxygenase	NC_012586	Sinorhizobium fredii NGR234 plasmid pNGR234b, complete sequence.	Schmeisser Appl. Environ. Microbiol. 75 (12), 4035-4045 (2009) Streit J. Bacteriol. 186 (2), 535-542 (2004) Liesegang Submitted (29-OCT-2007) Institute of Microbiology and Genetics, Goettingen Genomics Laboratory, Grisebachstrasse 8, Goettingen, Lower Saxony D-37077, Germany
Catechol 2,3 dioxygenase	NC_020995	Enterococcus casseliflavus EC20, complete genome.	Russ unpublished Russ Submitted (27-MAR-2013) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
Catechol 2,3 dioxygenase (b)	NC_002754	Sulfolobus solfataricus P2 chromosome, complete genome.	She Proc. Natl. Acad. Sci. U.S.A. 98 (14), 7835-7840 (2001) Charlebois Genome 43 (1), 116-136 (2000)
Catechol 2,3 dioxygenase	NC_017276	Sulfolobus islandicus REY15A chromosome, complete genome.	Guo J. Bacteriol. (2011) In press
Catechol 2,3 dioxygenase	NC_017275	Sulfolobus islandicus HVE10/4 chromosome, complete genome.	Guo J. Bacteriol. (2011) In press
Catechol 2,3 dioxygenase	NC_014318	Amycolatopsis mediterranei U32 chromosome, complete genome.	Zhao Cell Res. 20 (10), 1096-1108 (2010)
Catechol 2,3 dioxygenase	NC_007181	Sulfolobus acidocaldarius DSM 639 chromosome, complete genome.	Chen J. Bacteriol. 187 (14), 4992-4999 (2005)
Catechol 2,3 dioxygenase	NC_016978	Comamonas testosteroni plasmid pI2, complete sequence.	Krol Appl. Environ. Microbiol. (2011) In press Boon Appl. Environ. Microbiol. 67 (3), 1107-1115 (2001)
Catechol 2,3 dioxygenase	NC_020247	Sulfolobus acidocaldarius Ron12/I, complete genome.	Mao ISME J 6 (8), 1613-1616 (2012)
Catechol 2,3 dioxygenase	NC_020246	Sulfolobus acidocaldarius N8, complete genome.	Mao ISME J 6 (8), 1613-1616 (2012)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Catechol 2,3 dioxygenase	NC_006270	Bacillus licheniformis ATCC 14580, complete genome.	Rey Genome Biol. 5 (10), R77 (2004)
Catechol 2,3 dioxygenase	NC_003997	Bacillus anthracis str. Ames chromosome, complete genome.	Read Nature 423 (6935), 81-86 (2003)
Catechol 2,3 dioxygenase	NC_014551	Bacillus amyloliquefaciens DSM7 complete genome.	Borriss Int. J. Syst. Evol. Microbiol. 61 (PT 8), 1786-1801 (2011)
Catechol 1,2 dioxygenase	NC_007953	Burkholderia xenovorans LB400 chromosome 3, complete sequence.	Chain Proc. Natl. Acad. Sci. U.S.A. 103 (42), 15280-15287 (2006) Copeland A Submitted (13-MAR-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Catechol 2,3 dioxygenase	NC_016047	Bacillus subtilis subsp. spizizenii TU-B-10, complete genome.	Earl J. Bacteriol. 194 (9), 2378-2379 (2012)
Catechol 2,3 dioxygenase	NC_014622	Paenibacillus polymyxa SC2, complete genome.	Ma J. Bacteriol. 193 (1), 311-312 (2011)
Catechol 2,3 dioxygenase	NC_005945	Bacillus anthracis str. Sterne chromosome, complete genome.	Brettin Submitted (12-JAN-2004) Joint Genome Institute, Department of Energy, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Catechol 2,3 dioxygenase	NZ_CP007640	Bacillus atrophaeus subsp. globigii strain BSS.	Daligault Genome Announc 2 (5) (2014) Bishop-Lilly Submitted (05-MAY-2014) Bioscience, Los Alamos National Laboratory, P.O. Box 1663, MS M888, Los Alamos, NM 87545, USA
Catechol 2,3 dioxygenase (short removed)	NC_005042	Prochlorococcus marinus subsp. marinus str. CCMP1375 complete genome	Dufresne Proc. Natl. Acad. Sci. U.S.A. 100 (17), 10020-10025 (2003)
Catechol 1,2 dioxygenase	NC_006351	Burkholderia pseudomallei K96243 chromosome 2, complete sequence.	Holden Proc. Natl. Acad. Sci. U.S.A. 101 (39), 14240-14245 (2004)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Catechol 1,2 dioxygenase (a)	NC_002947	Pseudomonas putida KT2440 chromosome, complete genome.	Nelson Environ. Microbiol. 4 (12), 799-808 (2002)
Catechol 1,2 dioxygenase	NC_011002	Burkholderia cenocepacia J2315 chromosome 3, complete genome.	Holden J. Bacteriol. 191 (1), 261-277 (2009)
Catechol 1,2 dioxygenase (b)	NC_002947	Pseudomonas putida KT2440 chromosome, complete genome.	Nelson Environ. Microbiol. 4 (12), 799-808 (2002)
Catechol 1,2 dioxygenase (a)	NZ_HG938353	Neorhizobium galegae, complete genome.	Osterman BMC Genomics 15, 500 (2014)
Catechol 1,2 dioxygenase (a)	NC_010682	Ralstonia pickettii 12J chromosome 1, complete sequence.	Lucas Submitted (01-MAY-2008) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
Catechol 1,2 dioxygenase	NZ_KI530699	Acinetobacter gyllenbergii NIPH 230 adfcq-supercont1.1, whole genome shotgun sequence	Cerqueira Submitted (04-OCT-2013) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
catechol 1,2-dioxygenase	NC_003078	Sinorhizobium meliloti 1021 plasmid pSymB, complete sequence.	Finlan Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9889-9894 (2001)
Catechol 1,2 dioxygenase (a)	NC_003450	Corynebacterium glutamicum ATCC 13032 chromosome, complete genome.	Nagwa (in) PROCEEDINGS OF THE 9TH INTERNATIONAL SYMPOSIUM ON THE GENETICS OF INDUSTRIAL MICROORGANISMS: 21; Silberbach Appl. Environ. Microbiol. 71 (5), 2391-2402 (2005) Ikeda Appl. Microbiol. Biotechnol. 62 (2-3), 99-109 (2003)
catechol 1,2-dioxygenase	NC_004463	Bradyrhizobium japonicum USDA 110 chromosome, complete genome.	Kaneto DNA Res. 9 (6), 225-256 (2002) Kaneto DNA Res. 9 (6), 189-197 (2002) Gottfert J. Bacteriol. 183 (4), 1405-1412 (2001)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
catechol 1,2-dioxygenase	NZ_KB849655	Acinetobacter junii CIP 64.5 acLZZ-supercont1.3, whole genome	Cerqueira Submitted (19-FEB-2013) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA Feldgarden Submitted (19-FEB-2013) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
Catechol 1,2 dioxygenase	NC_015663	Enterobacter aerogenes KCTC 2190 chromosome, complete genome.	Shin J. Bacteriol. 194 (9), 2373-2374 (2012)
Catechol 1,2 dioxygenase	NZ_KB851227	Acinetobacter lwoffii NCTC 5866 = CIP 64.10 acLsp-supercont1.12, whole shotgun sequence.	Cerqueira Submitted (19-FEB-2013) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA Felgarden Submitted (19-FEB-2013) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
catechol 1,2 dioxygenase	NZ_KB849749	Acinetobacter radioresistens DSM 6976 = NBRC 102413 = CIP 103788 acLrZ-supercont1.9, whole genome shotgun sequence.	Cerqueira Submitted (19-FEB-2013) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA Felgarden Submitted (19-FEB-2013) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
Catechol 1,2 dioxygenase (a)	NZ_KI421499	Bradyrhizobium genosp. SA-4 str. CB756 BrageDRAFT_scaffold1.1, whole genome shotgun sequence.	Kyrpides Submitted (23-AUG-2013) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598-1698, USA
Catechol 1,2 dioxygenase	NC_025133	Sphingobium wenxiniae strain JZ-1 plasmid pPBA, complete sequence.	Wang Appl. Environ. Microbiol. 80 (13), 3811-3818 (2014)
Catechol 2,3 dioxygenase (b)	NZ_KI421499	Bradyrhizobium genosp. SA-4 str. CB756 BrageDRAFT_scaffold1.1, whole genome shotgun sequence.	Kyrpides Submitted (23-AUG-2013) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598-1698, USA
Catechol 1,2 dioxygenase	NC_002516	Pseudomonas aeruginosa PAO1 chromosome, complete genome.	Winsor Nucleic acids Res. 37 (DATABASE ISSUE), D483-D488 (2009) Cirz J. Bacteriol. 188 (20), 7101-7110 (2006) Palmer J. Bacteriol. 187 (15), 5267-5277 (2005) Salunkhe J. Bacteriol. 187 (14), 4908-4920 (2005) Filiatrault Infect. Immun. 73 (6), 3764-3772 (2005) Rassmussen J. Bacteriol. 187 (5), 1799-1814 (2005) Winsor Nucleic acids Res. 33 (DATABASE ISSUE), D338-D343 (2005) Wagner J. Bacteriol. 185 (7), 2080-2095 (2003) Schuster J. Bacteriol. 185 (7), 2066-2079 (2003) Stover Nature 406 (6799), 959-964 (2000)
extradiol catechol dioxygenase (a)	NZ_AKCL01000023	Pseudomonas putida SJTE-1 contig000023, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
putative aromatic ring-opening dioxygenase	FN563149.1	Rhodococcus equi 103S chromosome.	Letek, M et al., PLoS Genet. 2010 Sep 30;6(9):e1001145. doi: 10.1371/journal.pgen.1001145.
2, 3-dihydroxybiphenyl 1, 2-dioxygenase	AB117721.1	Rhodococcus rhodochrous orf31, orf32, bphC3 genes for hypothetical proteins and 2, 3-dihydroxybiphenyl 1, 2-dioxygenase, partial and complete cds.	Taguchi *et al., Biosci Biotechnol Biochem.* 2004 Apr;68(4):787-95.
protocatechuate 4,5-dioxygenase	CP003588.1	Pseudomonas putida ND6, complete genome.	Li *et al*., J Bacteriol. 2012 Sep;194(18):5154-5. doi: 10.1128/JB.01190-12.
protocatechuate 4,5-dioxygenase beta subunit / protocatechuate 4,5-dioxygenase alpha subunit	CP000712.1	Pseudomonas putida F1, complete genome.	Copeland *et al*., Submitted (25-MAY-2007) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
protocatechuate 4,5-dioxygenase	CP003734.1	Pseudomonas putida DOT-T1E, complete genome.	Udaondo et al., Submitted (04-JUL-2012) Environmental Protection, EEZ-CSIC, Profesor Albareda 1, Granada, Granada 18008, Spain
gallate dioxygenase	AE015451.2	Pseudomonas putida KT2440 complete genome	Nelson et al., Environ Microbiol. 2002 Dec;4(12):799-808.
Protocatechuate 4,5-dioxygenase	CP002290.1	Pseudomonas putida BIRD-1, complete genome.	Matilla et al., J Bacteriol. 2011 Mar;193(5):1290. doi: 10.1128/JB.01281-10. Epub 2010 Dec 23.
protocatechuate 4,5-dioxygenase subunit alpha	CP005976.1	Pseudomonas putida H8234, complete genome.	Molina et al Genome Announc. 2013 Jul 18;1(4). pii: e00496-13. doi: 10.1128/genomeA.00496-13.
Protocatechuate 4,5-dioxygenase	CP000949.1	Pseudomonas putida W619, complete genome.	Copeland et al Submitted (21-FEB-2008) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
hydroxyquinol 1,2-dioxygenase	CP005959.1	Corynebacterium glutamicum MB001, complete genome.	Baumgart et al., Appl Environ Microbiol. 2013 Oct;79(19):6006-15. doi: 10.1128/AEM.01634-13. Epub 2013 Jul 26.

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
putative hydroxyquinol/catechol 1,2-dioxygenase	HE802067.1	Corynebacterium glutamicum K051 complete genome, strain ATCC 13032, sub-strain K051.	Binder et al., Genome Biol. 2012 May 28;13(5):R40. doi: 10.1186/gb-2012-13-5-r40.
Protocatechuate 3,4-dioxygenase beta subunit	BA000036.3	Corynebacterium glutamicum ATCC 13032 DNA, complete genome.	Submitted (24-MAY-2002) Contact:Satoshi Nakagawa Kyowa Hakko Kogyo Co. Ltd., Tokyo Research Laboratories; 3-6-6, Asahi-machi, Machida, Tokyo 194-8533, Japan
CATECHOL 1,2-DIOXYGENASE	BX927157.1	Corynebacterium glutamicum ATCC 13032, IS fingerprint type 4-5, complete genome; segment 10/10.	Kalinowski et al., J Biotechnol. 2003 Sep 4;104(1-3):5-25.
hydroxyquinol 1,2-dioxygenase	CP012194.1	Corynebacterium glutamicum strain CP, complete genome.	Gui,Y., Ma,Y., Xie,X. and Chen,N. Submitted (30-JUL-2015) Metabolic Engineering Laboratory, College of Biotechnology, Tianjin University of Science and Technology, 13 Main Street, Tianjin Economic and Technological Development, Tianjin, Tianjin 300457, China
hydroxyquinol 1,2-dioxygenase	CP011309.1	[Brevibacterium] flavum strain ATCC 15168, complete genome.	Ahn,J., Park,G., Jeon,W., Jang,Y., Jang,M., Lee,H. and Lee,H. Submitted (22-APR-2015) Biotechnology Process Engineering, KRIBB, 30 Yenongudanjo-ro, Ochang-eup, Cheongwon-gu, Cheong-joo 363-883, Republic of Korea
hydroxyquinol 1,2-dioxygenase	CP010451.1	Corynebacterium glutamicum strain B253, complete genome.	Wu et al., J Biotechnol. 2015 Aug 10;207:10-1. doi: 10.1016/j.jbiotec.2015.04.018. Epub 2015 May 4.
hypothetical protein	AP009044.1	Corynebacterium glutamicum R DNA, complete genome.	Yukawa et al., Microbiology. 2007 Apr;153(Pt 4):1042-58.
hydroxyquinol 1,2-dioxygenase	CP013991.1	Corynebacterium glutamicum strain USDA-ARS-USMARC-56828, complete	Harhay et al Submitted (19-JAN-2016) Genetics, Breeding, and Animal Health Research Unit, USDA-ARS-USMARC, PO BOX 166, Clay Center, NE 68933,

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
			USA
COG3485 Protocatechuate 3,4-dioxygenase beta subunit	CP004048.1	Corynebacterium glutamicum SCgG2, complete genome.	Pei,G., Sun,J. and Zheng,P. Submitted (16-JAN-2013) Laboratory of System&Synthetic Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, N0.32,West 7 Road, Tianjin Airport Economic Park, TianJin, TianJin 300308, China
Protocatechuate 3,4-dioxygenase beta subunit	CP004047.1	Corynebacterium glutamicum SCgG1, complete genome.	Pei,G., Sun,J. and Zheng,P. Submitted (16-JAN-2013) Laboratory of System&Synthetic Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, N0.32,West 7 Road, Tianjin Airport Economic Park, TianJin, TianJin 300308, China
hydroxyquinol 1,2-dioxygenase	CP007724.1	Corynebacterium glutamicum strain AR1, complete genome.	Park et al Nat Commun. 2014 Aug 5;5:4618. doi: 10.1038/ncomms5618.
hydroxyquinol 1,2-dioxygenase	CP007722.1	Corynebacterium glutamicum strain ATCC 21831, complete genome.	Park et al., Submitted (13-MAY-2014) BioInformatics Research Center, KAIST, Daejeon 305-701, Republic of Korea
catechol 1,2-dioxygenase	CP007569.1	Bradyrhizobium japonicum SEMIA 5079 genome.	Siqueira et al BMC Genomics. 2014 Jun 3;15:420. doi: 10.1186/1471-2164-15-420.
catechol 1,2-dioxygenase	AP014685.1	Bradyrhizobium diazoefficiens DNA, complete genome, strain: NK6.	Lida et al., Appl Environ Microbiol. 2015 Jun 15;81(12):4143-54. doi: 10.1128/AEM.00741-15. Epub 2015 Apr 10.
catechol 1,2-dioxygenase	BA000040.2a	Bradyrhizobium japonicum USDA 110 DNA, complete genome.	Kaneko et al., DNA Res. 2002 Dec 31;9(6):189-97.
catechol 1,2-dioxygenase	AP012206.1a	Bradyrhizobium japonicum USDA 6 DNA, complete genome.	Kaneko et al., Genes (Basel). 2011 Oct 28;2(4):763-87. doi: 10.3390/genes2040763.
catechol 1,2-dioxygenase	CP010313.1a	Bradyrhizobium japonicum strain E109, complete genome.	Torres et al., Genome Announc. 2015 Feb 19;3(1). pii: e01566-14. doi: 10.1128/genomeA.01566-14.

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
catechol 1,2-dioxygenase	CP013949.1	Bradyrhizobium sp. CCGE-LA001, complete genome.	Servín-Garcidueñas et al., Mol Phylogenet Evol. 2014 Oct;79:1-11. doi: 10.1016/j.ympev.2014.06.006. Epub 2014 Jun 18.
catechol 1,2-dioxygenase	CU234118.1	Bradyrhizobium sp. ORS278,complete sequence.	Giraud et al., Science. 2007 Jun 1;316(5829):1307-12.
catechol 1,2-dioxygenase	BA000040.2b	Bradyrhizobium japonicum USDA 110 DNA, complete genome.	Kaneko et al., DNA Res. 2002 Dec 31;9(6):189-97.
catechol 1,2-dioxygenase	AP012206.1b	Bradyrhizobium japonicum USDA 6 DNA, complete genome.	Kaneko et al., Genes (Basel). 2011 Oct 28;2(4):763-87. doi: 10.3390/genes2040763.
catechol 1,2-dioxygenase	CP010313.1b	Bradyrhizobium japonicum strain E109, complete genome.	Torres et al., Genome Announc. 2015 Feb 19;3(1). pii: e01566-14. doi: 10.1128/genomeA.01566-14.
catechol 1,2-dioxygenase	CP000494.1	Bradyrhizobium sp. BTAi1, complete genome.	Giraud et al., Science. 2007 Jun 1;316(5829):1307-12.
catechol 1,2-dioxygenase	AP012603.1	Bradyrhizobium oligotrophicum S58 DNA, complete genome.	Okubo et al., Appl Environ Microbiol. 2013 Apr;79(8):2542-51. doi: 10.1128/AEM.00009-13. Epub 2013 Feb 8.
catechol 1,2-dioxygenase	AP014704.1	Methylobacterium aquaticum DNA, complete genome, strain: MA-22A.	Tani et al., Genome Announc 3 (2), e00266-15 (2015)
TesB (meta-cleavage enzyme, 2,3-Dihydroxybiphenyl-1,2-dioxygenase)	AB040808	Comamonas testosteroni tesB, ORF1, ORF2, ORF3 genes, complete cds.	Horinouchi,M Microbiology (Reading, Engl.) 147 (PT 12), 3367-3375 (2001)
homogentisate 1,2-dioxygenase (1)	NZ_AKCL01000092	Pseudomonas putida SJTE-1 contig000092, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
protocatechuate 3,4-dioxygenase alpha chain (2)	NZ_AKCL01000092	Pseudomonas putida SJTE-1 contig000092, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
protocatechuate 3,4-dioxygenase subunit beta (3)	NZ_AKCL01000092	Pseudomonas putida SJTE-1 contig000092, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
benzoate 1,2-dioxygenase subunit beta	NZ_AKCL01000002	Pseudomonas putida SJTE-1 contig000002, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
benzoate 1,2-dioxygenase subunit alpha	NZ_AKCL01000002	Pseudomonas putida SJTE-1 contig000002, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
taurine dioxygenase	NZ_AKCL01000007	Pseudomonas putida SJTE-1 contig000007, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
4-hydroxyphenylpyruvate dioxygenase	NZ_AKCL01000201	Pseudomonas putida SJTE-1 contig000201, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
quercetin 2,3-dioxygenase	NZ_AKCL01000201	Pseudomonas putida SJTE-1 contig000201, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
alpha-ketoglutarate-dependent dioxygenase	NZ_AKCL01000201	Pseudomonas putida SJTE-1 contig000201, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
protocatechuate 4,5-dioxygenase subunit alpha	NZ_AKCL01000023	Pseudomonas putida SJTE-1 contig000023, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
Extradiol ring-cleavage dioxygenase III subunit B.	NZ_AKCL01000055	Pseudomonas putida SJTE-1 contig000055, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
4-hydroxyphenylpyruvate dioxygenase	NZ_AKCL01000022	Pseudomonas putida SJTE-1 contig000022, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
2-nitropropane dioxygenase	NZ_AKCL01000190	Pseudomonas putida SJTE-1 contig000190, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
2-nitropropane dioxygenase	NZ_AKCL01000135	Pseudomonas putida SJTE-1 contig000135, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
taurine dioxygenase	NZ_AKCL01000081	Pseudomonas putida SJTE-1 contig000081, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
DOPA 4,5-dioxygenase	NZ_AKCL01000071	Pseudomonas putida SJTE-1 contig000071, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
taurine dioxygenase	NZ_AKCL01000071	Pseudomonas putida SJTE-1 contig000071, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
ring-cleaving dioxygenase	NZ_AKCL01000205	Pseudomonas putida SJTE-1 contig000205, whole genome shotgun sequence.	Liang,R J. Bacteriol. 194 (17), 4781-4782 (2012)
phytanoyl-CoA dioxygenase	NZ_AFMP01000051	Sphingomonas sp. KC8 contig50, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
dioxygenase	NZ_AFMP01000051	Sphingomonas sp. KC8 contig50, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
dioxygenase	NZ_AFMP01000044	Sphingomonas sp. KC8 contig43, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
dioxygenase	NZ_AFMP01000039	Sphingomonas sp. KC8 contig38, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
4-hydroxyphenylpyruvate dioxygenase	NZ_AFMP01000011	Sphingomonas sp. KC8 contig11, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
homogentisate 1,2-dioxygenase	NZ_AFMP01000011	Sphingomonas sp. KC8 contig11, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
2-nitropropane dioxygenase	NZ_AFMP01000011	Sphingomonas sp. KC8 contig11, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
phytanoyl-CoA dioxygenase	NZ_AFMP01000016	Sphingomonas sp. KC8 contig16, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
phytanoyl-CoA dioxygenase	NZ_AFMP01000016	Sphingomonas sp. KC8 contig16, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
taurine dioxygenase	NZ_AFMP01000016	Sphingomonas sp. KC8 contig16, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
2-nitropropane dioxygenase	NZ_AFMP01000016	Sphingomonas sp. KC8 contig16, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
2-nitropropane dioxygenase	NZ_AFMP01000031	Sphingomonas sp. KC8 contig31, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
2-nitropropane dioxygenase	NZ_AFMP01000031	Sphingomonas sp. KC8 contig31, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
glyoxalase/bleomycin resistance/protein/dioxygenase	NZ_AFMP01000031	Sphingomonas sp. KC8 contig31, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
biphenyl 2,3-dioxygenase	NZ_AFMP01000036	Sphingomonas sp. KC8 contig35, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
Rieske-type ring hydroxylating dioxygenase beta subunit	NZ_AFMP01000026	Sphingomonas sp. KC8 contig26, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
aromatic ring-cleaving dioxygenase	NZ_AFMP01000004	Sphingomonas sp. KC8 contig04, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
2-nitropropane dioxygenase	NZ_AFMP01000059	Sphingomonas sp. KC8 contig58, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
glyoxalase/bleomycin resistance/protein/dioxygenase	NZ_AFMP01000009	Sphingomonas sp. KC8 contig09, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
tryptophan 2 3-dioxygenase	NZ_AFMP01000035	Sphingomonas sp. KC8 contig70, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
alpha-ketoglutarate-dependent dioxygenase	NZ_AFMP01000032	Sphingomonas sp. KC8 contig32, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
dioxygenase beta subunit	NZ_AFMP01000019	Sphingomonas sp. KC8 contig19, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
glyoxalase/bleomycin resistance/protein/dioxygenase	NZ_AFMP01000040	Sphingomonas sp. KC8 contig39, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
3-hydroxyanthranilate 3,4-dioxygenase	NZ_AFMP01000040	Sphingomonas sp. KC8 contig39, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
glyoxalase/bleomycin resistance/protein/dioxygenase	NZ_AFMP01000013	Sphingomonas sp. KC8 contig13, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
protocatechuate 3,4-dioxygenase	NC_008314	Ralstonia eutropha H16 chromosome 2.	Pohlmann, A Nat. Biotechnol. 24 (10), 1257-1262 (2006)
protocatechuate 3,4-dioxygenase	NC_008313	Ralstonia eutropha H16 chromosome 1.	Pohlmann, A Nat. Biotechnol. 24 (10), 1257-1262 (2006)
hydroxyquinol 1,2-dioxygenase	NZ_KI421499	Bradyrhizobium genosp. SA-4 str. CB756 BrageDRAFT_scaffold1.1,	Kyrpides Submitted (23-AUG-2013) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598-1698, USA
protocatechuate dioxygenase	NC_008268	Rhodococcus jostii RHA1, complete genome.	McLeod Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
TesB	LC010134	Comamonas testosteroni DNA, putative mega-cluster of steroid degradation genes, strain: T441	Hourinochi Microbiology (Reading, Engl.) 147 (PT 12), 3367-3375 (2001)
biphenyl-2,3-diol-1,2-dioxygenase	CP001220	Comamonas testosteroni CNB-2, complete genome.	Ma Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
biphenyl-2,3-diol-1,2-dioxygenase	AF493052	Comamonas testosteroni biphenyl-2,3-diol-1,2-dioxygenase gene, complete cds.	Skowasch Biochem. Biophys. Res. Commun. 294 (3), 560-566 (2002)
biphenyl 2,3-dioxygenase	CP006704	Comamonas testosteroni TK102, complete genome.	Fukuda Genome Announc 2 (5), e00865-14 (2014)
Dioxygenase	JF502261	Rhodococcus sp. R04 dioxygenase gene, complete cds.	Yang Submitted (27-FEB-2011) Department of Biotechnology, Institution of Enzyme, Wucheng Road, Taiyuan, Shanxi 030006, China
protocatechuate 3,4-dioxygenase	NC_013446	Comamonas testosteroni CNB-2, complete genome.	Ma Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
protocatechuate 3,4-dioxygenase subunit beta	NC_013446	Comamonas testosteroni CNB-2, complete genome.	Ma Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
protocatechuate 4,5-dioxygenase subunit alpha	NC_010002	Delftia acidovorans SPH-1, complete genome.	Schleheck Appl. Environ. Microbiol. 70 (7), 4053-4063 (2004)
protocatechuate 3,4-dioxygenase subunit beta	NC_010002	Delftia acidovorans SPH-1, complete genome.	Schleheck Appl. Environ. Microbiol. 70 (7), 4053-4063 (2004)
protocatechuate 3,4-dioxygenase subunit beta	NC_003450	Corynebacterium glutamicum ATCC 13032 chromosome, complete genome.	Silberbach Appl. Environ. Microbiol. 71 (5), 2391-2402 (2005) Ikeda Appl. Microbiol. Biotechnol. 62 (2-3), 99-109 (2003)
protocatechuate 3,4-dioxygenase subunit beta	NC_010002	Delftia acidovorans SPH-1, complete genome.	Schleheck Appl. Environ. Microbiol. 70 (7), 4053-4063 (2004)
protocatechuate 3,4-dioxygenase subunit beta	NC_012586	Sinorhizobium fredii NGR234 plasmid pNGR234b, complete sequence.	Schmeisser Appl. Environ. Microbiol. 75 (12), 4035-4045 (2009) Streit J. Bacteriol. 186 (2), 535-542 (2004)
protocatechuate 3,4-dioxygenase	NC_008314	Ralstonia eutropha H16 chromosome 2.	Pohlmann, A Nat. Biotechnol. 24 (10), 1257-1262 (2006)
protocatechuate 3,4-dioxygenase subunit alpha	NC_020800	Xanthomonas axonopodis Xac29-1, complete genome.	Chen Submitted (15-MAR-2013) School of Agriculture and Biology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, Shanghai 200240, China
protocatechuate 3,4-dioxygenase subunit beta	NC_020800	Xanthomonas axonopodis Xac29-1, complete genome.	Chen Submitted (15-MAR-2013) School of Agriculture and Biology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, Shanghai 200240, China
protocatechuate 3,4-dioxygenase subunit beta	NC_006834	Xanthomonas oryzae pv. oryzae KACC 10331, complete genome.	Park Submitted (23-SEP-2004) National Institute of Agricultural Biotechnology, Rural Development

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
			Administration, 225 Seodundong, Suwon 441-707, Republic of Korea
protocatechuate 3,4-dioxygenase	NZ_HG938353	Neorhizobium galegae, complete genome.	Osterman BMC Genomics 15, 500 (2014)
protocatechuate 3,4-dioxygenase	NC_013446	Comamonas testosteroni CNB-2, complete genome.	Ma Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
protocatechuate 4,5-dioxygenase subunit alpha	NC_010002	Delftia acidovorans SPH-1, complete genome.	Schleheck Appl. Environ. Microbiol. 70 (7), 4053-4063 (2004)
protocatechuate 3,4-dioxygenase subunit beta	NC_010002	Delftia acidovorans SPH-1, complete genome.	Schleheck Appl. Environ. Microbiol. 70 (7), 4053-4063 (2004)
protocatechuate 3,4-dioxygenase subunit beta	NC_003450	Corynebacterium glutamicum ATCC 13032 chromosome, complete genome.	Silberbach Appl. Environ. Microbiol. 71 (5), 2391-2402 (2005)
protocatechuate 3,4-dioxygenase subunit beta	NC_010002	Delftia acidovorans SPH-1, complete genome.	Schleheck Appl. Environ. Microbiol. 70 (7), 4053-4063 (2004)
protocatechuate 3,4-dioxygenase subunit beta	NC_012586	Sinorhizobium fredii NGR234 plasmid pNGR234b, complete sequence.	Schmeisser Appl. Environ. Microbiol. 75 (12), 4035-4045 (2009)
protocatechuate 3,4-dioxygenase	NC_008314	Ralstonia eutropha H16 chromosome 2.	Pohlmann, A Nat. Biotechnol. 24 (10), 1257-1262 (2006)
protocatechuate 3,4-dioxygenase subunit alpha	NC_020800	Xanthomonas axonopodis Xac29-1, complete genome.	Chen Submitted (15-MAR-2013) School of Agriculture and Biology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, Shanghai 200240, China

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
protocatechuate 3,4-dioxygenase subunit beta	NC_020800	Xanthomonas axonopodis Xac29-1, complete genome.	Chen Submitted (15-MAR-2013) School of Agriculture and Biology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, Shanghai 200240, China
protocatechuate 3,4-dioxygenase subunit beta	NC_006834	Xanthomonas oryzae pv. oryzae KACC 10331, complete genome.	Park Submitted (23-SEP-2004) National Institute of Agricultural Biotechnology, Rural Development Administration, 225 Seodundong, Suwon 441-707, Republic of Korea
protocatechuate 3,4-dioxygenase	NZ_HG938353	Neorhizobium galegae, complete genome.	Osterman BMC Genomics 15, 500 (2014)
reductase	U49504	Pseudomonas sp.	Parales et al., Gene 181 (1-2), 57-61 (1996)
Dioxygenase	AF252550	Comamonas testosteroni	Moser and Stahl Appl. Microbiol. Biotechnol. 55 (5), 609-618 (2001)
dox operon (doxABCDEFGHIJ) genes	M60405	Pseudomonas sp.	Denome et al., J. Bacteriol. 175 (21), 6890-6901 (1993)
naphthalene dioxygenase	U49496	Pseudomonas sp	Parales et al., Gene 181 (1-2), 57-61 (1996)
naphthalene dioxygenase	AF004284	Pseudomonas putida	Hamann Submitted (14-MAY-1997) Biologie, Universitat Bremen, Leobener Strasse, Bremen 28359, Germany
naphthalene dioxygenase (nahAc) gene	AY048759	Pseudomonas fluorescens	Min and Ji Submitted (26-OCT-2001) Biological Science, Sookmyung Women's University, Chungpa-Dong 2-Ka, Yongsan-Ku, Seoul 140-742, Korea
Dioxygenase	AB004059	Pseudomonas putida	Takizawa et al., J. Bacteriol. 176 (8), 2444-2449 (1994)
naphthalene dioxygenase	M83949	Pseudomonas putida	Simon et al., Gene 127 (1), 31-37 (1993)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
naphthalene dioxygenase	AF039533	Pseudomonas stutzeri	Bosch et al., Submitted (19-DEC-1997) Microbiology, Universitat Illes Balears, Carretera Valldemossa km. 7.5, Palma de Mallorca, Balearic Islands 07071, Spain
naphthalene dioxygenase	M23914	Pseudomonas putida	Kurkela et al., Gene 73 (2), 355-362 (1988)
naphthalene dioxygenase	AB255564	uncultured bacterium	Ono et al., Appl. Microbiol. Biotechnol. 74 (2), 501-510 (2007)
naphthalene dioxygenase	AF004283	Pseudomonas fluorescens	Hamann Submitted (14-MAY-1997) Biologie, Universitat Bremen, Leobener Strasse, Bremen 28359, Germany
naphthalene dioxygenase	D84146	Pseudomonas aeruginosa	Takizawa et al., Submitted (21-MAR-1996) Noboru TAKIZAWA, Okayama University of Science, Applied Chemistry; 1-1 Ridai-cho, Okayama, Okayama 700, JAPAN
nitrobenzene dioxygenase	AF379638	Comamonas sp. JS765	Lessner et al., Appl. Environ. Microbiol. 68 (2), 634-641 (2002)
naphthalene dioxygenase	DQ846881	Rhodococcus opacus	Di Gennaro et al., Appl. Microbiol. Biotechnol. 87 (1), 297-308 (2010)
aromatic dioxygenase	AF121905	Rhodococcus sp. I24	Treadway et al., Submitted (21-JAN-1999) Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Ave., Bldg 68-370, Cambridge, MA 02139, USA
naphthalene dioxygenase	AY392424	Rhodococcus sp. P200	Kulakov et al., Appl. Environ. Microbiol. 71 (4), 1754-1764 (2005)
naphthalene dioxygenase	AY392423	Rhodococcus sp. P400	Kulakov et al., Appl. Environ. Microbiol. 71 (4), 1754-1764 (2005)
o-xylene oxygenase gene	AB206671	Rhodococcus opacus TKN14	Maryuma et al., Appl. Environ. Microbiol. 71 (12), 7705-7715 (2005)
NidA (nidA) gene	AY330100	Mycobacterium sp. KMS	Miller et al., Microb. Ecol. 48 (2), 230-238 (2004)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
dioxygenase large alpha subunit (nidA) gene	DQ537942	Mycobacterium gilvum strain czh-101	Hennessee and Li Submitted (10-MAY-2006) MBBE, UH, 1955 East-West Rd #402, Honolulu, HI 96822, USA
dioxygenase large alpha subunit (nidA) gene	AF548343	Mycobacterium flavescens strain PYR-GCK	Brezna et al., Submitted (25-SEP-2002) Microbiology Division, USFDA/NCTR, 3900 NCTR Rd., Jefferson, AR 72079, USA
Iron-sulfur protein large subunit	AB017794	Nocardioides sp.	Saito et al., Chemosphere 38 (6), 1331-1337 (1999)
dioxygenase	EF026099	Mycobacterium sp. SNP11	Pagnaut et al., Res. Microbiol. 158 (2), 175-186 (2007)
dioxygenase	DQ358754	Mycobacterium sp. CH-1	Morgan and Churchill Submitted (10-JAN-2006) Biological Sciences, University of Alabama, Hackberry Lane, Tuscaloosa, AL 35487-0344, USA
extradiol dioxygenase	DQ157862	Mycobacterium sp. CH-2	Morgan and Churchill Submitted (08-AUG-2005) Department of Biology, University of Alabama, Hackberry Lane, Tuscaloosa, AL 35487-0344, USA
biphenyl dioxygenase	AB113649	Bacillus sp. JF8	Mukerjee-Dhar et al., Microbiology (Reading, Engl.) 151 (PT 12), 4139-4151 (2005)
Ring hydroxylating dioxygenase	AJ494743	Mycobacterium sp. 6PY1	Krivobok et al., J. Bacteriol. 185 (13), 3828-3841 (2003)
extradiol dioxygenase	AB031319	Nocardioides sp. KP7	Saito et al., J. Bacteriol. 182 (8), 2134-2141 (2000)
NidB	DQ157863	Mycobacterium sp. CH-2	Morgan and Churchill Submitted (08-AUG-2005) Department of Biology, University of Alabama, Hackberry Lane, Tuscaloosa, AL 35487-0344, USA
NidA (nidA) gene	AF548345	Mycobacterium frederiksbergense strain FAn9T	Brezna et al., Submitted (25-SEP-2002) Microbiology Division, USFDA/NCTR, 3900 NCTR Rd., Jefferson, AR 72079, USA
phthalate dioxygenase	AY365117	Mycobacterium vanbaalenii strain PYR-1	Stingley et al., Biochem. Biophys. Res. Commun. 322 (1), 133-146 (2004)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
naphthalene inducible dioxygenase	AF249301	Mycobacterium sp. PYR-1	Khan et al., Appl. Environ. Microbiol. 67 (8), 3577-3585 (2001)
NidA (nidA) gene	DQ537941	Mycobacterium pallens strain czh-8	Hennessee and Li Int. J. Syst. Evol. Microbiol. 59 (PT 2), 378-387 (2009)
dioxygenase	DQ358753	Mycobacterium sp. CH-1	Morgan and Churchill Submitted (10-JAN-2006) Biological Sciences, University of Alabama, Hackberry Lane, Tuscaloosa, AL 35487-0344, USA
NidA and NidB	AY330098	Mycobacterium sp. JLS	Miller et al., Microb. Ecol. 48 (2), 230-238 (2004)
NidA	AB179737	Mycobacterium sp. MHP-1	Habe et al J. Biosci. Bioeng. 98 (4), 306-308 (2004)
nidA	AF548347	Mycobacterium gilvum strain BB1	Brezna et al., Submitted (25-SEP-2002) Microbiology Division, USFDA/NCTR, 3900 NCTR Rd., Jefferson, AR 72079, USA
pdoA1	AJ494745	Mycobacterium sp. 6PY1	Krivobok et al., J. Bacteriol. 185 (13), 3828-3841 (2003)
NidA	AY330102	Mycobacterium sp. MCS	Miller et al., Microb. Ecol. 48 (2), 230-238 (2004)
dioxygenase	AF546905	Mycobacterium sp. S65	Sho et al., FEMS Microbiol. Ecol. 48 (2), 209-220 (2004)
dioxygenase	AJ536756	Ralstonia oxalatica transposon Tn4371.	Toussaint et al., Appl. Environ. Microbiol. 69 (8), 4837-4845 (2003)
biphenyl dioxygenase	AB086835	Comamonas testosteroni	Hiraoka et al., Appl. Environ. Microbiol. 68 (10), 5104-5112 (2002)
biphenyl dioxygenase	U47637	Pandoraea pnomenusa strain B-356	Sylvestre et al., Gene 174 (2), 195-202 (1996)
biphenyl dioxygenase	M83673	Pseudomonas pseudoalcaligenes KF707	Taira et al., J. Biol. Chem. 267 (7), 4844-4853 (1992)
biphenyl dioxygenase	AY027651	Pseudomonas sp. Cam-1	Master and Mohn Appl. Environ. Microbiol. 67 (6), 2669-2676 (2001)
biphenyl dioxygenase	M86348	Paraburkholderia xenovorans LB400	Erickson and Mondello J. Bacteriol. 174 (9), 2903-2912 (1992)
biphenyl dioxygenase	U95054	Pseudomonas sp. B4	Ducrocq and Truffaut Submitted (25-MAR-1997) Laboratoire de Genetique Microbienne, Universite de Technologie de Compiegne, Centre de Recherche de Royallieu BP529, Compiegne 60205, France

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
biphenyl dioxygenase	AF049345	Pseudomonas pseudoalcaligenes	Taira et al., J. Biol. Chem. 267 (7), 4844-4853 (1992)
biphenyl dioxygenase	AJ010057	Burkholderia sp. JB1	Dignum Submitted (04-AUG-1998) Dignum M., Laboratory of Microbiology, University of Amsterdam, Nieuwe Achtergracht 127, Amsterdam, NL-1018 WS, THE NETHERLANDS
terpA	EF527236	Rhodococcus sp. L4	Suttinun and Luepromchai Submitted (28-MAR-2007) Department of Microbiology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand
dioxygenase	Y18245	Pseudomonas putida	Mosqueda et al., Gene 232 (1), 69-76 (1999)
benzene dioxygenase	M17904	Pseudomonas putida	Irie et al., J. Bacteriol. 169 (11), 5174-5179 (1987)
biphenyl dioxygenase	AJ251217	Pseudomonas sp. B4	Rodarie et al., Submitted (03-DEC-1999) Rodarie D., DBMS / BBSI, CEA Grenoble, 17 avenue des Martyrs, 38054 Grenoble cedex 09, FRANCE
dioxygenase	AY831463	Pseudomonas putida strain GJ31	Kunze et al., Microbiology (Reading, Engl.) 155 (PT 12), 4069-4083 (2009)
toluene dioxygenase	J04996	Pseudomonas putida	Zylstra and Gibson J. Biol. Chem. 264 (25), 14940-14946 (1989)
dioxygenase	AJ006307	Ralstonia sp. JS705	van der Meer Appl. Environ. Microbiol. 64 (11), 4185-4193 (1998)
dioxygenase	U15298	Pseudomonas sp.	Werlen et al., J. Biol. Chem. 271 (8), 4009-4016 (1996)
chlorobenzene dioxygenase	U78099	Burkholderia sp. PS12	Beil et al., Eur. J. Biochem. 247 (1), 190-199 (1997)
cumene dioxygenase	D37828	Pseudomonas fluorescens	Aoki et al., J. Ferment. Bioeng. 81, 187-196 (1996)
ethylbenzene dioxygenase	AF049851	Pseudomonas fluorescens	Corkery and Dobson FEMS Microbiol. Lett. 166 (2), 171-176 (1998)
dioxygenase	U53507	Pseudomonas JR1	Pflugmacher et al., Appl. Environ. Microbiol. 62 (11), 3967-3977 (1996)
dioxygenase	AJ293587	Pseudomonas putida	Chablain et al., Res. Microbiol. 148 (2), 153-161 (1997)
dioxygenase	AF006691	Pseudomonas putida	Eaton and Timmis J. Bacteriol. 168 (1), 123-131 (1986)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
dioxygenase	AB048707	Rhodococcus sp. RHA1	Kitagawa et al., Biosci. Biotechnol. Biochem. 65 (8), 1907-1911 (2001)
dioxygenase	AB120955	Rhodococcus sp. RHA1	Iwasaki et al., Submitted (24-SEP-2003) Takumi Iwasaki, Nagaoka University of Technology, Department of Bioengineering; Kamitomioka, Nagaoka, Niigata 940-2188, Japan
dioxygenase	AB120956	Rhodococcus sp. RHA1	Iwasaki et al., Submitted (24-SEP-2003) Takumi Iwasaki, Nagaoka University of Technology, Department of Bioengineering; Kamitomioka, Nagaoka, Niigata 940-2188, Japan
dioxygenase	AB048708	Rhodococcus sp. RHA1	Kitagawa et al., Biosci. Biotechnol. Biochem. 65 (8), 1907-1911 (2001)
terminal dioxygenase	U27591	Rhodococcus sp. M5	Wang et al., Gene 164 (1), 117-122 (1995)
dioxygenase	X80041	Rhodococcus globerulus	Asturias et al., Gene 156 (1), 11-18 (1995)
dioxygenase	D32142	Rhodococcus sp. RHA1	Masai et al., Appl. Environ. Microbiol. 61 (6), 2079-2085 (1995)
isopropylbenzene 2,3-dioxygenase	U24277	Rhodococcus erythropolis	Kesseler et al., Microbiology (Reading, Engl.) 142 (PT 11), 3241-3251 (1996)
terminal dioxygenase	D88020	Rhodococcus erythropolis	Kosono et al., Appl. Environ. Microbiol. 63 (8), 3282-3285 (1997)
biphenyl dioxygenase	EF152282	Sphingobium yanoikuyae	Chadhain et al., J. Ind. Microbiol. Biotechnol. 34 (9), 605-613 (2007)
oxidoreductase	AF380367	Burkholderia sp. DBT1	Di Gregorio et al., Biodegradation 15 (2), 111-123 (2004)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
dioxygenase	AB024945	Alcaligenes faecalis	Kiyohara et al., Submitted (14-MAR-1999) Noboru Takizawa, Okayama University of Science, Applied Chemistry; 1-1, Ridai-cho, Okayama 700-0005, Japan
dioxygenase	DQ501245	Cycloclasticus sp. P1	Wang et al., Submitted (19-APR-2006) The Third Institute of Oceanography, State of Oceanic Administration (SOA), Room 329, Daxue Road 184#, Xiamen, Fujian 361005, P.R. China
Ring hydroxylating dioxygenase	AJ633551	Sphingomonas sp. CHY-1	Demaneche et al., Appl. Environ. Microbiol. 70 (11), 6714-6725 (2004)
extradiol dioxygenase	AF169302	Burkholderia cepacia	Johnson et al., Arch. Microbiol. 173 (2), 86-90 (2000)
dioxygenase	CP000511	Mycobacterium vanbaalenii PYR-1	Copeland et al., Submitted (13-DEC-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	CP000479	Mycobacterium avium 104	Fleischmann et al., Submitted (19-OCT-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
dioxygenase	CP000509	Nocardioides sp. JS614,	Coleman et al., J. Bacteriol. 193 (13), 3399-3400 (2011)
dioxygenase	CP000656	Mycobacterium gilvum PYR-GCK	Badejo et al., PLoS ONE 8 (2), E58066 (2013)
dioxygenase	AB091693	Sphingomonas sp. P2	Pinyakong et al., Biochem. Biophys. Res. Commun. 301 (2), 350-357 (2003)
dioxygenase	AAXZ01000003	Rhodobacteraceae bacterium HTCC2150	Kang et al., J. Bacteriol. 192 (23), 6315-6316 (2010)
dioxygenase	AB272984	Rhodococcus rhodochrous	Taguchi et al., Biosci. Biotechnol. Biochem. 68 (4), 787-795 (2004)
dioxygenase	CH482384	Pseudomonas aeruginosa 2192	Mathee et al., Proc. Natl. Acad. Sci. U.S.A. 105 (8), 3100-3105 (2008)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
dioxygenase	EU024110	Sphingomonas sp. LB126 fluorene/dibenzofuran degradation operon,	Schuler et al., Appl. Environ. Microbiol. 74 (4), 1050-1057 (2008)
dioxygenase	AF060489	Sphingomonas sp. CB3	Shepherd and Lloyd-Jones Biochem. Biophys. Res. Commun. 247 (1), 129-135 (1998)
dioxygenase	AAQG01000001	Sphingomonas sp. SKA58	Hagstrom et al., Submitted (17-MAR-2006) J Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
dioxygenase	AB201843	Paenibacillus sp. YK5	Iida et al., Arch. Microbiol. 184 (5), 305-315 (2006)
dioxygenase	U51165	Cycloclasticus oligotrophus XYLX	Wang et al., Appl. Environ. Microbiol. 62 (6), 2169-2173 (1996)
dioxygenase	AB161232	Sphingomonas sp. A4	Pinyakong et al., FEMS Microbiol. Lett. 238 (2), 297-305 (2004)
dioxygenase	AB240454	Sphingomonas sp. A4	Kouzuma et al., Microbiology (Reading, Engl.) 152 (PT 8), 2455-2467 (2006)
dioxygenase	AB121977	Xanthobacter polyaromaticivorans	Hirano et al., Appl. Microbiol. Biotechnol. 69 (6), 672-681 (2006)
dioxygenase	AF157565	Sphingopyxis macrogoltabida strain TFA	Andujar et al., J. Bacteriol. 182 (3), 789-795 (2000)
dioxygenase	AB070456	Rhodococcus sp. YK2	Iida et al., Biosci. Biotechnol. Biochem. 66 (7), 1462-1472 (2002)
dioxygenase	AF061751	Burkholderia sp. strain RP007	Laurie and Lloyd-Jones J. Bacteriol. 181 (2), 531-540 (1999)
dioxygenase	CP000089	Dechloromonas aromatica RCB	Salinero et al., Submitted (04-AUG-2005) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	D89064	Pseudomonas sp.	Sato et al., J. Bacteriol. 179 (15), 4841-4849 (1997)
dioxygenase	U18133	Conidiobolus coronatus	Nakatsu et al., Proc. Natl. Acad. Sci. U.S.A. 88 (19), 8312-8316 (1991)
pobA and pobB genes	X78823	P.pseudoalcaligenes (POB310)	Dehmel et al., Arch. Microbiol. 163 (1), 35-41 (1995)
phthalate dioxygenase	AF095748	Burkholderia cepacia	Chang and Zylstra J. Bacteriol. 180 (24), 6529-6537 (1998)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
2-halobenzoate 1,2-dioxygenase	X79076	Burkholderia cepacia	Haak et al., J. Bacteriol. 177 (3), 667-675 (1995)
dioxygenase	AF009224	Acinetobacter sp. ADP1	Neidle et al., J. Bacteriol. 170 (10), 4874-4880 (1988)
dioxygenase	AF071556	Acinetobacter sp. ADP1	Bundy et al., J. Bacteriol. 180 (17), 4466-4474 (1998)
dioxygenase	AB084235	Terrabacter sp. DBF63	Habe et al., Appl. Microbiol. Biotechnol. 61 (1), 44-54 (2003)
dioxygenase	X72850	Sphingomonas sp.	Happe et al., J. Bacteriol. 175 (22), 7313-7320 (1993)
dioxygenase	D17319	Pseudomonas sp. KKS102	Kimbara et al., J. Bacteriol. 171 (5), 2740-2747 (1989)
napthalene dioxygenase	AF082663	Rhodococcus sp. NCIMB12038	Larkin et al., J. Bacteriol. 181 (19), 6200-6204 (1999)
putative cis-naphthalene 1,2-dioxygenase	AJ401612	Rhodococcus sp. 1BN	Andreoni et al., Environ. Microbiol. 2 (5), 572-577 (2000)
benzene dioxygenase	AF148496	Pseudomonas putida	Tan et al., Gene 130 (1), 33-39 (1993)
dioxygenase	EF600714	Pandoraea pnomenusa	Jiang et al., Arch. Microbiol. 191 (6), 485-492 (2009)
dioxygenase	AP011117	Rhodococcus opacus B4	Na et al., J. Biosci. Bioeng. 99 (4), 378-382 (2005)
dioxygenase	AF452376	Rhodococcus aetherivorans strain I24	Priefert et al., Appl. Microbiol. Biotechnol. 65 (2), 168-176 (2004)
dioxygenase	BA000007	Escherichia coli O157:H7 str	Makino et al., Genes Genet. Syst. 74 (5), 227-239 (1999)
dioxygenase	AE005174	Escherichia coli O157:H7 str. EDL933	Perna et al., Nature 409 (6819), 529-533 (2001)
dioxygenase	Z37966	Escherichia coli K-12	Turlin et al., Submitted (22-SEP-1994) Antoine Danchin, Biochimie et genetique moleculaire, Institut, Pasteur, 28 rue du Docteur Roux, PARIS CEDEX 15, F-75724, FRANCE
dioxygenase	U00096	Escherichia coli str. K-12 substr. MG1655	Blattner et al., Science 277 (5331), 1453-1462 (1997)
dioxygenase	AP009048	Escherichia coli str. K-12 substr. W3110	Musso et al., Proc. Natl. Acad. Sci. U.S.A. 74 (1), 106-110 (1977)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
dioxygenase	CP000038	Shigella sonnei Ss046	Yang et al., Nucleic Acids Res. 33 (19), 6445-6458 (2005)
dioxygenase	CP000802	Escherichia coli HS	Rasko et al., J. Bacteriol. 190 (20), 6881-6893 (2008)
dioxygenase	CP000036	Shigella boydii Sb227	Yang et al., Nucleic Acids Res. 33 (19), 6445-6458 (2005)
dioxygenase	CP000266	Shigella flexneri 5 str. 8401	Nie et al., BMC Genomics 7, 173 (2006)
dioxygenase	ACLI01000099	Corynebacterium efficiens YS-314	Qin et al., Submitted (10-APR-2009) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
dioxygenase	BX571866	Photorhabdus luminescens subsp. laumondii TTO1	Duchaud et al., Nat. Biotechnol. 21 (11), 1307-1313 (2003)
dioxygenase	AB237655	Pseudomonas putida plasmid NAH7	Tsuda and Iino Mol. Gen. Genet. 223 (1), 33-39 (1990)
dioxygenase	AY887963	Pseudomonas fluorescens strain PC20 plasmid pNAH20	Heinaru et al., FEMS Microbiol. Ecol. 31 (3), 195-205 (2000)
dioxygenase	AY125981	Pseudomonas fluorescens plasmid pLP6a	McFarlane and Foght Submitted (25-JUN-2002) Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada
dioxygenase	AF491307	Pseudomonas putida NCIB 9816-4 plasmid pDTG1	Simon et al., Gene 127 (1), 31-37 (1993)
dioxygenase	BX640437	Bordetella bronchiseptica strain RB50	Parkhill et al., Nat. Genet. 35 (1), 32-40 (2003)
dioxygenase	AE005674	Shigella flexneri 2a str. 301	Jin et al., Nucleic Acids Res. 30 (20), 4432-4441 (2002)
dioxygenase	AE014073	Shigella flexneri 2a str. 2457T	Wei et al., Infect. Immun. 71 (5), 2775-2786 (2003)
dioxygenase	AY208917	Pseudomonas sp. ND6 plasmid pND6-1	Li et al., Gene 336 (2), 231-240 (2004)
dioxygenase	AF036940	Ralstonia sp. U2	Fuenmayor et al., J. Bacteriol. 180 (9), 2522-2530 (1998)
dioxygenase	DQ167474	Polaromonas naphthalenivorans CJ2	Jeon et al., Appl. Environ. Microbiol. 72 (2), 1086-1095 (2006)
dioxygenase	CP000529	Polaromonas naphthalenivorans CJ2	Copeland et al., Submitted (21-DEC-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
dioxygenase	CP000539	Acidovorax sp. JS42	Copeland et al., Submitted (21-DEC-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	CP000077	Sulfolobus acidocaldarius DSM 639	Chen et al., J. Bacteriol. 187 (14), 4992-4999 (2005)
dioxygenase	CP000520	Mycobacterium sp. KMS plasmid pMKMS02	Copeland et al., Submitted (18-DEC-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	CP000456	Arthrobacter sp. FB24	Copeland et al., Submitted (28-AUG-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	CP000518	Mycobacterium sp. KMS	Copeland et al., Submitted (18-DEC-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	CP000519	Mycobacterium sp. KMS plasmid pMKMS01	Copeland et al., Submitted (18-DEC-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	CP000580	Mycobacterium sp. JLS	Copeland et al., Submitted (20-FEB-2007) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	CP000384	Mycobacterium sp. MCS	Copeland et al., Submitted (05-JUN-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	CP000385	Mycobacterium sp. MCS	Copeland et al., Submitted (05-JUN-2006) US DOE Joint Genome Institute, 2800

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
			Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	AF331043	Arthrobacter keyseri plasmid pRE1	Eaton J. Bacteriol. 183 (12), 3689-3703 (2001)
dioxygenase	AP008980	Terrabacter sp. DBF63 plasmid pDBF1	Kasuga J. Ferment. Bioeng. 84, 387-399 (1997)
dioxygenase	EF494237	Rhodococcus sp. DK17 plasmid pDK2	Choi et al., Biochem. Biophys. Res. Commun. 357 (3), 766-771 (2007)
dioxygenase	AB048709	Rhodococcus sp. RHA1	Kitagawa et al., Biosci. Biotechnol. Biochem. 65 (8), 1907-1911 (2001)
dioxygenase	AB154536	Rhodococcus sp. RHA1	Kitagawa et al., Submitted (17-DEC-2003) Keisuke Miyauchi, Nagaoka University of Technology, Department of Bioengineering; 1603-1 Kamitomioka, Nagaoka, Niigata 940-2188, Japan
dioxygenase	AB154537	Rhodococcus sp. RHA1	Kitagawa et al., Submitted (17-DEC-2003) Keisuke Miyauchi, Nagaoka University of Technology, Department of Bioengineering; 1603-1 Kamitomioka, Nagaoka, Niigata 940-2188, Japan
putative phthalate dioxygenase	AAR90178	Rhodococcus sp. RHA1	Choi et al., Submitted (16-DEC-2003) Department of Biology, Yonsei University, 134 Shinchon-dong, Seoul 120-749, Korea
dioxygenase	DQ007994	Rhodococcus sp. TFB	Tomas-Gallardo et al Proteomics 6 (SUPPL 1), S119-S132 (2006)
dioxygenase	AF546904	Mycobacterium sp. S65	Sho et al., FEMS Microbiol. Ecol. 48 (2), 209-220 (2004)
dioxygenase	DQ028634	Mycobacterium vanbaalenii PYR-1	Kim et al., Appl. Environ. Microbiol. 72 (2), 1045-1054 (2006)
biphenyl dioxygenase	EF151283	Sphingobium yanoikuyae strain B1	Chadhain et al., J. Ind. Microbiol. Biotechnol. 34 (9), 605-613 (2007)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
dioxygenase	CP000676	Novosphingobium aromaticivorans DSM 12444	Copeland et al., Submitted (12-APR-2007) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	AY502075	Rhodococcus sp. DK17	Kim et al., Appl. Environ. Microbiol. 70 (12), 7086-7092 (2004)
dioxygenase	AF079317	Sphingomonas aromaticivorans plasmid pNL1	Romine et al., J. Bacteriol. 181 (5), 1585-1602 (1999)
dioxygenase	AB272985	Rhodococcus erythropolis	Taguchi et al., Biosci. Biotechnol. Biochem. 71 (5), 1136-1144 (2007)
dioxygenase	AB272986	Rhodococcus sp. HA99	Taguchi et al., Biosci. Biotechnol. Biochem. 71 (5), 1136-1144 (2007)
dioxygenase	DQ403247	Rhodococcus sp. R04	Yang et al., J. Appl. Microbiol. 103 (6), 2214-2224 (2007)
dioxygenase	AF119621	Pseudomonas abietaniphila BKME-9	Mohn et al., Syst. Appl. Microbiol. 22 (1), 68-78 (1999)
dioxygenase	CP000272	Paraburkholderia xenovorans LB400	Chain et al., Proc. Natl. Acad. Sci. U.S.A. 103 (42), 15280-15287 (2006)
dioxygenase	CP000699	Sphingomonas wittichii RW1	Miller et al., J. Bacteriol. 192 (22), 6101-6102 (2010)
dioxygenase	AJ223219	Sphingomonas sp.	Armengaud et al., J. Bacteriol. 180 (15), 3954-3966 (1998)
dioxygenase	AJ223220	Sphingomonas sp.	Armengaud et al., J. Bacteriol. 180 (15), 3954-3966 (1998)
dioxygenase	CP000701	Sphingomonas wittichii RW1	Miller et al., J. Bacteriol. 192 (22), 6101-6102 (2010)
dioxygenase	D88021	Rhodococcus erythropolis	Kosono et al., Appl. Environ. Microbiol. 63 (8), 3282-3285 (1997)
dioxygenase	AB270530	Sphingomonas sp. KA1	Habe et al., FEMS Microbiol. Lett. 211 (1), 43-49 (2002)
dioxygenase	AB110633	Rhodococcus opacus	Kimura et al., Appl. Microbiol. Biotechnol. (2006) In press
dioxygenase	AB102786	Cycloclasticus sp. A5	Kasai et al., Appl. Environ. Microbiol. 69 (11), 6688-6697 (2003)
dioxygenase	AB075242	Terrabacter sp. YK3	Tida et al., Appl. Environ. Microbiol. 68 (8), 3716-3723 (2002)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
dioxygenase	AE015451	Pseudomonas putida KT2440	Nelson et al., Environ. Microbiol. 4 (12), 799-808 (2002)
dioxygenase	CP000432	Rhodococcus jostii RHA1 plasmid pRHL1	McLeod Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
dioxygenase	CP000433	Rhodococcus jostii RHA1 plasmid pRHL2	Takeda et al., J. Bacteriol. 192 (18), 4741-4751 (2010)
dioxygenase	AY502076	Rhodococcus sp. DK17 plasmid pDK3	Choi et al., Submitted (16-DEC-2003) Department of Biology, Yonsei University, 134 Shinchon-dong, Seoul 120-749, Korea
dioxygenase	AB257758	Plasmid pFKY4	Ono et al., Appl. Microbiol. Biotechnol. 74 (2), 501-510 (2007)
dioxygenase	AB257757	Plasmid pFKY1	Ono et al., Appl. Microbiol. Biotechnol. 74 (2), 501-510 (2007)
dioxygenase	NG_034878	Pseudomonas putida plasmid NPL1	Boronin et al., Genetika 25 (2), 226-237 (1989)
napthalene dioxygenase	NG_041567	Pseudomonas putida plasmid pAK5	Izmalkova et al., Res. Microbiol. 164 (3), 244-253 (2013)
dioxygenase	HM623873	Pseudomonas chlororaphis strain SY-02	Zhang et al., Submitted (02-JUL-2010) College of Resources and Environmental Engineering, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China
dioxygenase	JN248563	Pseudomonas sp. MC1	Ahn et al., Submitted (13-JUL-2011) Department of Systems Biology, Yonsei University, 50 Yonsei-Ro, Seodaemun-Gu, Seoul 120-749, Republic of Korea
Naphtalene dioxygenase	HM368649	Pseudomonas sp. N1	Bianchi Submitted (28-MAY-2010) Department of Structural and Functional Biology, University of Insubria, Via J.H. Dunant 3,

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
			Varese 21100, Italy
dioxygenase	CP007510	Pseudomonas stutzeri strain 19SMN4	Brunet-Galmes et al., Submitted (24-MAR-2014) Microbiology, Department of Biology, Univesitat de les Illes Balears (UIB), Carretera de Valldemossa km 7.5, Palma de Mallorca, Illes Balears 07122, Spain
dioxygenase	CP003677	Pseudomonas stutzeri CCUG 29243	Brunet-Galmes et al., J. Bacteriol. 194 (23), 6642-6643 (2012)
dioxygenase	HM204990	Pseudomonas stutzeri strain NJ	Chakraborty et al., Submitted (06-MAY-2010) Microbiology, Bose Institute, P-1/12 C.I.T. Scheme VII M, Kolkata, West Bengal 700054, India
dioxygenase	NG_036701	Rhodococcus sp. TFB	Tomas-Gallardo et al Microb Biotechnol 2 (2), 262-273 (2009)
dioxygenase	CP008952	Rhodococcus opacus strain R7	Di Gennaro et al., Genome Announc 2 (4) (2014)
dioxygenase	CP008951	Rhodococcus opacus strain R7	Di Gennaro et al., Genome Announc 2 (4) (2014)
dioxygenase	AB024936	Rhodococcus sp. CIR2	Takizawa et al., Submitted (12-MAR-1999) Noboru Takizawa, Okayama University of Science, Applied Chemistry; 1-1, Ridai-cho, Okayama 700-0005, Japan
dioxygenase	NG035928	Rhodococcus opacus plasmid pWK301	Kimura et al., Appl. Microbiol. Biotechnol. (2006) In press
dioxygenase	GQ848233	Gordonia sp. CC-NAPH129-6	Lin et al., Microbiol. Res. 167 (7), 395-404 (2012)
napthalene dioxygenase	GQ503240	Rhodococcus sp. B13	Anan et al., Antonie Van Leeuwenhoek 100 (2), 309-316 (2011)
napthalene dioxygenase	GQ503241	Rhodococcus sp. B2-1	Anan et al., Antonie Van Leeuwenhoek 100 (2), 309-316 (2011)

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
napthalene dioxygenase	GQ503239	Rhodococcus sp. DB11	Anan et al., Antonie Van Leeuwenhoek 100 (2), 309-316 (2011)
dioxygenase	CP002385	Mycobacterium gilvum Spyr1	Kallimanis et al., Stand Genomic Sci 5 (1), 144-153 (2011)
NidA	HM049723	Mycobacterium sp. py142	DeBruyn et al., Submitted (31-MAR-2010) Center for Environmental Biotechnology, The University of Tennessee, 676 Dabney Hall, Knoxville, TN 37996, USA
NidA	HM049719	Mycobacterium sp. py137	DeBruyn et al., Submitted (31-MAR-2010) Center for Environmental Biotechnology, The University of Tennessee, 676 Dabney Hall, Knoxville, TN 37996, USA
NidA	HM049714	Bacterium py120	DeBruyn et al., Submitted (31-MAR-2010) Center for Environmental Biotechnology, The University of Tennessee, 676 Dabney Hall, Knoxville, TN 37996, USA
NidA	HM049725	Mycobacterium sp. py145	DeBruyn et al., Submitted (31-MAR-2010) Center for Environmental Biotechnology, The University of Tennessee, 676 Dabney Hall, Knoxville, TN 37996, USA
NidA	HM049713	Bacterium py114	DeBruyn et al., Submitted (31-MAR-2010) Center for Environmental Biotechnology, The University of Tennessee, 676 Dabney Hall, Knoxville, TN 37996, USA
NidA	HM049726	Mycobacterium sp. py146	DeBruyn et al., Submitted (31-MAR-2010) Center for Environmental Biotechnology, The University of Tennessee, 676 Dabney Hall, Knoxville, TN 37996, USA
dioxygenase	CP003169	Mycobacterium rhodesiae NBB3	Lucas et al., Submitted (08-DEC-2011) US DOE Joint Genome Institute, 2800 Mitchell Drive B310, Walnut Creek, CA 94598-1698, USA

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
dioxygenase	AB626849	Mycobacterium sp. NJS-P	Zeng et al., Submitted (20-APR-2011) Contact:Jun Zeng Institute of Soil Science, Chinese Academy of Sciences, Soil Biology and Biochemistry; No. 71 Beijing East Road, Nanjing 210008, China
dioxygenase	JQ916945	Pseudomonas sp. Jpyr-1	Ma Submitted (10-APR-2012) School of Life Science and Biotechnology, Dalian Technology of Technology, No. 2, Ling Gong Road, Gan Jing Zi District, Dalian, Liao Ning 116024, China
dioxygenase	CP003872	Acidovorax sp. KKS102	Ohtsubo et al., J. Bacteriol. 194 (24), 6970-6971 (2012)
dioxygenase	AB706355	Comamonas testosteroni genomic DNA	Hara and Takatsuka Submitted (28-MAR-2012) Contact:Tomijiro Hara Yamagata University, Graduate School of Science and Engineering; 4-3-16 Jonan, Yonezawa, Yamagata 992-8510, Japan
dioxygenase	AB546270	Acidovorax sp. KKS102	Ohtsubo et al., J. Biol. Chem. 276 (39), 36146-36154 (2001)
dioxygenase	CP006704	Comamonas testosteroni TK102	Fukuda et al., Genome Announc 2 (5), e00865-14 (2014)
dioxygenase	CP008761	Burkholderia xenovorans LB400	Bishop-Lilly et al., Submitted (19-JUN-2014) Bioscience, Los Alamos National Laboratory, P.O. Box 1663, MS M888, Los Alamos, NM 87545, USA
dioxygenase	JN874407	Cupriavidus sp. SK-4	Ilori et al., J. Basic Microbiol. (2013) In press
dioxygenase	FJ715926	Pseudomonas putida strain B6-2	Li et al., Environ. Sci. Technol. 43 (22), 8635-8642 (2009)
dioxygenase	CP003093	Pseudoxanthomonas spadix BD-a59	Lee et al., J. Bacteriol. 194 (2), 544 (2012)
dioxygenase	CP000530	Polaromonas naphthalenivorans CJ2	Copeland et al., Submitted (21-DEC-2006) US DOE Joint Genome Institute, 2800

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
			Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
biphenyl dioxygenase	JQ015309	Alcanivorax sp. HA03	Hassan Submitted (09-NOV-2011) Environmental Biotechnology Department, GEBRI Menoufiya Uni., P.O. Box 79, Sadat City, Egypt
biphenyl dioxygenase	AB733643	Janibacter sp. TYM3221	Nguyen et al., J. Biosci. Bioeng. 116 (1), 91-100 (2013)
biphenyl dioxygenase	AB609317	Janibacter sp. TYM3221	Nguyen et al., Enzyme Microb. Technol. 49 (6-7), 532-539 (2011)
dioxygenase	EF635855	Pseudomonas nitroreducens strain J5-1	Jiang and Wang Submitted (29-MAY-2007) Department of Environmental Science and Engineering, Tsinghua University, Beijing 100084, China
dioxygenase	EU825676	Bordetella sp. IITR02	Pathak et al., Submitted (30-MAY-2008) Environmental Biotechnology Division, Indian Institute of Toxicology Research, Mahatma Gandhi Marg, P.O. Box-80, Lucknow, Uttar Pradesh 226 001, India
dioxygenase	AM902716	Bordetella petrii strain DSM 12804	Gross et al., BMC Genomics 9, 449 (2008)
dioxygenase	CP006979	Pseudomonas monteilii SB3101	Dueholm et al., Submitted (11-DEC-2013) Biotechnology, Chemistry, and Environmental Engineering, Aalborg University, Sohngaardsholmsvej 49, Aalborg 9000, Denmark
dioxygenase	CP006978	Pseudomonas monteilii SB3078	Dueholm et al., Genome Announc 2 (3), e00524-14 (2014)
dioxygenase	CP003734	Pseudomonas putida DOT-T1E	Udaondo et al., Submitted (04-JUL-2012) Environmental Protection, EEZ-CSIC, Profesor Albareda 1, Granada, Granada 18008, Spain

Gene/enzyme	Accession n.	Description	Reference/submission information from GenBank
dioxygenase	CP000712	Pseudomonas putida F1	Copeland et al., Submitted (25-MAY-2007) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
dioxygenase	AB828709	Pseudomonas putida	Faizal et al., J. Ind. Microbiol. Biotechnol. 32 (11-12), 542-547 (2005)
dioxygenase	NC_010002	Delftia acidovorans SPH-1	Schleheck et al., Appl. Environ. Microbiol. 70 (7), 4053-4063 (2004)
dioxygenase	NC_020800	Xanthomonas axonopodis Xac29-1	Chen et al., Submitted (15-MAR-2013) School of Agriculture and Biology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, Shanghai 200240, China
dioxygenase	NC_006834	Xanthomonas oryzae pv. oryzae KACC 10331	Lee et al., Nucleic Acids Res. 33 (2), 577-586 (2005)

Table Appendix 0-5: Database of gene sequences coding for dehydrogenase

Gene/enzyme	Recommended enzyme names (BRENDA, 2017)	Accession n.	Description	Reference/submission information from GenBank
TesH (3-ketosteroid-delta1-dehydrogenase)	3-oxosteroid 1-dehydrogenase,	AB076368	Comamonas testosteroni TA441 tesH, tesI, ORF17, ORF18 genes for 3-ketosteriod-delta1-dehydrogenase, 3-ketosteriod-delta4(5alpha)-dehydrogenase, hypothetical protein, complete cds.	Horinouchi, M. Appl. Environ. Microbiol. 69 (8), 4421-4430 (2003)
TesI (3-ketosteroid-delta4(5alpha)-dehydrogenase)	3-oxo-5alpha-steroid 4-dehydrogenase	AB076368	Comamonas testosteroni TA441 tesH, tesI, ORF17, ORF18 genes for 3-ketosteriod-delta1-dehydrogenase, 3-ketosteriod-delta4(5alpha)-dehydrogenase, hypothetical protein, complete cds.	Horinouchi, M. Appl. Environ. Microbiol. 69 (8), 4421-4430 (2003)
3-beta-hydroxysteroid dehydrogenase ORF62	3beta-hydroxy-DELTA5-steroid dehydrogenase	AB474240	Comamonas testosteroni TA441 O genes for hypothetical proteins, 3-beta-hydroxysteriod dehydrogenase, hypothetical protein, partial and complete cds, strain:TA441.	Horinouchi,M. J. Steroid Biochem. Mol. Biol. 122 (4), 253-263 (2010)
3-alpha steroid dehydrogenase (acts as 17β-dehydrogenase)	3alpha-hydroxysteroid 3-dehydrogenase	AB489116	Comamonas testosteroni TA441 genes for 3-ketosteroiod isomerase, 3-alpha steroid dehydrogenase, complete cds, strain: TA441.	Horinouchi,M. J. Steroid Biochem. Mol. Biol. 122 (4), 253-263 (2010)
3-beta hydroxysteroid dehydrogenase	3(or 17)beta-hydroxysteroid dehydrogenase	NZ_AKCL01000092	Pseudomonas putida SJTE-1 contig000092, whole genome shotgun sequence	Liang,R., J. Bacteriol. 194 (17), 4781-4782 (2012)
3-beta hydroxysteroid dehydrogenase	3(or 17)beta-hydroxysteroid dehydrogenase	NZ_AKCL01000160	Pseudomonas putida SJTE-1 contig000160, whole genome shotgun sequence	Liang,R., J. Bacteriol. 194 (17), 4781-4782 (2012)
3-beta hydroxysteroid dehydrogenase	3(or 17)beta-hydroxysteroid dehydrogenase	NZ_AFMP01000007	Sphingomonas sp. KC8 contig07, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)

Gene/enzyme	Recommended enzyme names (BRENDA, 2017)	Accession n.	Description	Reference/submission information from GenBank
7-alpha-hydroxysteroid dehydrogenase	7alpha-hydroxysteroid dehydrogenase	NZ_AFMP01000007	Sphingomonas sp. KC8 contig07, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
3-alpha-hydroxysteroid dehydrogenase	3alpha-hydroxysteroid 3-dehydrogenase	NZ_AFMP01000031	Sphingomonas sp. KC8 contig31, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NZ_AFMP01000031	Sphingomonas sp. KC8 contig31, whole genome shotgun sequence.	Hu,A. J. Bacteriol. 193 (16), 4266-4267 (2011)
3-beta-hydroxysteroid dehydrogenase	3(or 17)beta-hydroxysteroid dehydrogenase	NC_005027	Rhodopirellula baltica SH 1 chromosome, complete genome.	Wecker BMC Genomics 10, 410 (2009)
3-keto-5alpha-steroid delta(1)-dehydrogenase KstD3	3-oxosteroid 1-dehydrogenase	NC_012490	Rhodococcus erythropolis PR4 DNA, complete genome.	Sekine Environ. Microbiol. 8 (2), 334-346 (2006)
3-ketosteroid delta(1)-dehydrogenase KstD2	3-oxosteroid 1-dehydrogenase	NC_012490	Rhodococcus erythropolis PR4 DNA, complete genome.	Sekine Environ. Microbiol. 8 (2), 334-346 (2006)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_012490	Rhodococcus erythropolis PR4 DNA, complete genome.	Sekine Environ. Microbiol. 8 (2), 334-346 (2006)
3-ketosteroid delta(1)-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_012490	Rhodococcus erythropolis PR4 DNA, complete genome.	Sekine Environ. Microbiol. 8 (2), 334-346 (2006)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008269	Rhodococcus jostii RHA1 plasmid pRHL1, complete sequence.	McLeod Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada

Gene/enzyme	Recommended enzyme names (BRENDA, 2017)	Accession n.	Description	Reference/submission information from GenBank
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008269	Rhodococcus jostii RHA1 plasmid pRHL1, complete sequence.	McLeod Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008268	Rhodococcus jostii RHA1, complete genome.	McLeod Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008268	Rhodococcus jostii RHA1, complete genome.	McLeod Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008268	Rhodococcus jostii RHA1, complete genome.	McLeod Submitted (24-JUL-2006) Microbiology and Immunology, University of British Columbia, 1365-2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada

Gene/enzyme	Recommended enzyme names (BRENDA, 2017)	Accession n.	Description	Reference/submission information from GenBank
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_002944	Mycobacterium avium subsp. paratuberculosis str. k10, complete genome	Li Proc. Natl. Acad. Sci. U.S.A. 102 (35), 12344-12349 (2005)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_002944	Mycobacterium avium subsp. paratuberculosis str. k10, complete genome	Li Proc. Natl. Acad. Sci. U.S.A. 102 (35), 12344-12349 (2005)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_002944	Mycobacterium avium subsp. paratuberculosis str. k10, complete genome	Li Proc. Natl. Acad. Sci. U.S.A. 102 (35), 12344-12349 (2005)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_002944	Mycobacterium avium subsp. paratuberculosis str. k10, complete genome	Li Proc. Natl. Acad. Sci. U.S.A. 102 (35), 12344-12349 (2005)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_016946	Mycobacterium intracellulare ATCC 13950, complete genome.	Kim J. Bacteriol. 194 (10), 2750 (2012)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_016946	Mycobacterium intracellulare ATCC 13950, complete genome.	Kim J. Bacteriol. 194 (10), 2750 (2012)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_016946	Mycobacterium intracellulare ATCC 13950, complete genome.	Kim J. Bacteriol. 194 (10), 2750 (2012)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_016946	Mycobacterium intracellulare ATCC 13950, complete genome.	Kim J. Bacteriol. 194 (10), 2750 (2012)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008596	Mycobacterium smegmatis str. MC2 155 chromosome, complete genome.	Fleischmann Submitted (19-OCT-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008596	Mycobacterium smegmatis str. MC2 155 chromosome, complete genome.	Fleischmann Submitted (19-OCT-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008596	Mycobacterium smegmatis str. MC2 155 chromosome, complete genome.	Fleischmann Submitted (19-OCT-2006) The Institute for Genomic Research, 9712

Gene/enzyme	Recommended enzyme names (BRENDA, 2017)	Accession n.	Description	Reference/submission information from GenBank
				Medical Center Dr, Rockville, MD 20850, USA
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008596	Mycobacterium smegmatis str. MC2 155 chromosome, complete genome.	Fleischmann Submitted (19-OCT-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008611	Mycobacterium ulcerans Agy99 chromosome, complete genome.	Stinear Genome Res. 17 (2), 192-200 (2007)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_008611	Mycobacterium ulcerans Agy99 chromosome, complete genome.	Stinear Genome Res. 17 (2), 192-200 (2007)
3-beta-hydroxy-delta(5)-steroid dehydrogenase	3beta-hydroxy-DELTA5-steroid dehydrogenase	NC_013209	Acetobacter pasteurianus IFO 3283-01 DNA, complete genome.	Azuma Nucleic Acids Res. 37 (17), 5768-5783 (2009)
3-beta-hydroxy-delta(5)-steroid dehydrogenase	3beta-hydroxy-DELTA5-steroid dehydrogenase	NZ_AFSD01000001	Agrobacterium tumefaciens F2 chromosome C c1, whole genome shotgun sequence.	Li J. Bacteriol. 193 (19), 5531 (2011)
3-beta-hydroxy-delta(5)-steroid dehydrogenase	3beta-hydroxy-DELTA5-steroid dehydrogenase	NC_009667	Ochrobactrum anthropi ATCC 49188 chromosome 1, complete sequence.	Chain J. Bacteriol. 193 (16), 4274-4275 (2011)
3-beta-hydroxy-delta(5)-steroid dehydrogenase	3beta-hydroxy-DELTA5-steroid dehydrogenase	NC_007761	Rhizobium etli CFN 42, complete genome.	Gonzalez Proc. Natl. Acad. Sci. U.S.A. 103 (10), 3834-3839 (2006)
3-beta-hydroxy-delta(5)-steroid dehydrogenase	3beta-hydroxy-DELTA5-steroid dehydrogenase	NC_010505	Methylobacterium radiotolerans JCM 2831, complete genome.	Copeland Submitted (19-MAR-2008) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA

Gene/enzyme	Recommended enzyme names (BRENDA, 2017)	Accession n.	Description	Reference/submission information from GenBank
3-beta-hydroxy-delta(5)-steroid dehydrogenase	3beta-hydroxy-DELTA5-steroid dehydrogenase	NZ_KB900701	Bradyrhizobium elkanii USDA 76 BraelDRAFT_scaffold1.1, whole genome shotgun sequence.	Kyrpides Submitted (16-APR-2013) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598-1698, USA
3-beta-hydroxy-delta(5)-steroid dehydrogenase	3beta-hydroxy-DELTA5-steroid dehydrogenase	NC_008783	Bartonella bacilliformis KC583, complete genome.	Hendrix Submitted (18-DEC-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
KsdD-like steroid dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_000962	Mycobacterium tuberculosis H37Rv, complete genome.	Lew Tuberculosis (Edinb) 91 (1), 1-7 (2011)
3-beta-hydroxy-delta(5)-steroid dehydrogenase	3beta-hydroxy-DELTA5-steroid dehydrogenase	NC_007643	Rhodospirillum rubrum ATCC 11170 chromosome, complete genome.	Munk Stand Genomic Sci 4 (3), 293-302 (2011)
3-oxo-5-alpha-steroid 4-dehydrogenase	3-oxo-5alpha-steroid 4-dehydrogenase (NADP+)	NC_009614	Bacteroides vulgatus ATCC 8482, complete genome.	Xu PLoS Biol. 5 (7), E156 (2007)
delta 4, 5-alpha steroid dehydrogenase	3-oxo-5alpha-steroid 4-dehydrogenase (NADP+)	NZ_JH604622	Pseudomonas fragi B25 Scaffold1, whole genome shotgun sequence.	Mei J. Bacteriol. 194 (12), 3276-3277 (2012)
3-oxo-5-alpha-steroid 4-dehydrogenase	3-oxo-5alpha-steroid 4-dehydrogenase (NADP+)	NC_006347	Bacteroides fragilis YCH46 DNA, complete genome.	Kuwahara Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14919-14924 (2004)
NAD(P)H steroid dehydrogenase	3alpha-hydroxysteroid 3-dehydrogenase	NC_003902	Xanthomonas campestris pv. campestris str. ATCC 33913 chromosome, complete genome.	da Silva Nature 417 (6887), 459-463 (2002)

Gene/enzyme	Recommended enzyme names (BRENDA, 2017)	Accession n.	Description	Reference/submission information from GenBank
NAD(P)H steroid dehydrogenase	3alpha-hydroxysteroid 3-dehydrogenase	NC_003902	Xanthomonas campestris pv. campestris str. ATCC 33913 chromosome, complete genome.	da Silva Nature 417 (6887), 459-463 (2002)
3-oxo-5-alpha-steroid 4-dehydrogenase	3-oxo-5alpha-steroid 4-dehydrogenase (NADP+)	NC_004663	Bacteroides thetaiotaomicron VPI-5482 chromosome, complete genome.	Xu Science 299 (5615), 2074-2076 (2003)
Putative steroid dehydrogenase	3beta-hydroxy-DELTA5-steroid dehydrogenase (possibly)	NC_010397	Mycobacterium abscessus chromosome, complete sequence.	Ripoll unsubmitted
3-beta hydroxysteroid dehydrogenase	3(or 17)beta-hydroxysteroid dehydrogenase	NC_004129	Pseudomonas protegens Pf-5, complete genome.	Paulsen Nat. Biotechnol. 23 (7), 873-878 (2005)
7-alpha-hydroxysteroid dehydrogenase	Recommended names (BRENDA)	NC_002695	Escherichia coli O157:H7 str. Sakai chromosome, complete genome.	Bergholz BMC Microbiol. 7, 97 (2007)
3-beta hydroxysteroid dehydrogenase/isomerase family protein	3(or 17)beta-hydroxysteroid dehydrogenase	NC_003997	Bacillus anthracis str. Ames chromosome, complete genome.	Read Nature 423 (6935), 81-86 (2003)
3-ketosteroid-delta-1-dehydrogenase	3-oxosteroid 1-dehydrogenase	NC_010382	Lysinibacillus sphaericus C3-41, complete genome.	Hu J. Bacteriol. 190 (8), 2892-2902 (2008)
7-alpha-hydroxysteroid dehydrogenase	7alpha-hydroxysteroid dehydrogenase	NC_007618	Brucella melitensis biovar Abortus 2308 chromosome I, complete sequence, strain 2308.	Chain Infect. Immun. 73 (12), 8353-8361 (2005)
7-alpha-hydroxysteroid dehydrogenase	7alpha-hydroxysteroid dehydrogenase	NC_009850	Arcobacter butzleri RM4018, complete genome.	Kaakoush Antonie Van Leeuwenhoek 92 (4), 429-441 (2007)

Gene/enzyme	Recommended enzyme names (BRENDA, 2017)	Accession n.	Description	Reference/submission information from GenBank
7-alpha-hydroxysteroid dehydrogenase	7alpha-hydroxysteroid dehydrogenase	NC_011751	Escherichia coli UMN026 chromosome, complete genome.	Touchon PLoS Genet. 5 (1), E1000344 (2009)
7-alpha-hydroxysteroid dehydrogenase	7alpha-hydroxysteroid dehydrogenase	NC_011750	Escherichia coli IAI39 chromosome, complete genome.	Touchon PLoS Genet. 5 (1), E1000344 (2009)
7-alpha-hydroxysteroid dehydrogenase	7alpha-hydroxysteroid dehydrogenase	NC_012039	Campylobacter lari RM2100, complete genome.	Miller Foodborne Pathog. Dis. 5 (4), 371-386 (2008)
7-alpha-hydroxysteroid dehydrogenase	7alpha-hydroxysteroid dehydrogenase	NZ_CP007181	Campylobacter coli RM4661, complete genome.	Wright Submitted (06-FEB-2014) Agricultural Research Service, US Department of Agriculture, 800 Buchanan Street, Albany, CA 94710, USA
7-alpha-hydroxysteroid dehydrogenase	7alpha-hydroxysteroid dehydrogenase	NZ_KB849749	Acinetobacter radioresistens DSM 6976 = NBRC 102413 = CIP 103788 acLrZ-supercont1.9, whole genome shotgun sequence.	Unpublished
7-alpha-hydroxysteroid dehydrogenase	7alpha-hydroxysteroid dehydrogenase	NC_008599	Campylobacter fetus subsp. fetus 82-40, complete genome.	Fouts unpublished
7-alpha-hydroxysteroid dehydrogenase	7alpha-hydroxysteroid dehydrogenase	NC_000915	Helicobacter pylori 26695 chromosome, complete genome.	Raymond PLoS ONE 3 (5), E2259 (2008)
ORF18 (likely 3alpha-dehydrogenase) 1b	3alpha-hydroxysteroid 3-dehydrogenase	AB076368	Comamonas testosteroni TA441 tesH, tesI, ORF17, ORF18 genes for 3-ketosteriod-delta1-dehydrogenase, 3-ketosteriod-delta4(5alpha)-dehydrogenase, hypothetical protein, complete cds.	Horinouchi, M. Appl. Environ. Microbiol. 69 (8), 4421-4430 (2003)

Gene/enzyme	**Recommended enzyme names (BRENDA, 2017)**	**Accession n.**	**Description**	**Reference/submission information from GenBank**
TesR (regulator gene for degradation) (likely Delta1-dehydrogenase)	3-oxosteroid 1-dehydrogenase	AB186487	Comamonas testosteroni TA441 degradation gene cluster (ORF genes and tesR gene), complete cds.	Horinouchi,M. Biochem. Biophys. Res. Commun. 324 (2), 597-604 (2004)
3-ketosteroid-delta 4(5 alpha)-dehydrogenase	3-oxosteroid 1-dehydrogenase	L23428	Comamonas testosteroni ATCC 17410 delta 4, 5-alpha steroid dehydrogenase gene, complete cds.	Florin J. Bacteriol. 178 (11), 3322-3330 (1996)
3-oxosteriod 1-dehydrogenase node-1 length 715095-cov-155.621-id-1	3-oxosteriod 1-dehydrogenase	R. equi ATCC13557 unsubmitted	R.equi ATCC13557 (rast output for spades assembly)	This thesis
17beta hydroxysteriod dehydrogenase	17beta hydroxysteriod dehydrogenase	FN563149.1	Rhodococcus equi 103S chromosome.	Letek et al., PLoS Genet. 2010 Sep 30;6(9):e1001145. doi: 10.1371/journal.pgen.1001145.

Table Appendix 0-6: Gene sequences coding for dehydrogenase taken from Kisiela et al., (2012).

Gene/enzyme	**RefSeq ID**	**Description**	**Reference/submission information**
3-ketosteroid-Δ1-dehydrogenase	YP_907615	Mycobacterium ulcerans Agy99	Stinear et al., Genome Res. 17 (2), 192-200 (2007)
3-ketosteroid-Δ1-dehydrogenase	ZP_04751485	Mycobacterium kansasii ATCC 12478	Veyrier et al., Submitted (08-DEC-2008) Department of Medicine, McGill University, A5.156, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4, Canada
3-ketosteroid-Δ1-dehydrogenase	YP_001073300	Mycobacterium sp. JLS	Copeland et al., Submitted (20-FEB-2007) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
3-ketosteroid-Δ1-dehydrogenase	YP_641822	Mycobacterium sp. MCS	Copeland et al., Submitted (12-JUN-2006) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
3-ketosteroid-Δ1-dehydrogenase	YP_940731	Mycobacterium sp. KMS	Copeland et al., Submitted (22-DEC-2006) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
3-ketosteroid-Δ1-dehydrogenase	YP_890167	Mycobacterium smegmatis str. MC2 155	Fleischmann et al., Submitted (19-OCT-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
3-ketosteroid-Δ1-dehydrogenase	YP_956014	Mycobacterium vanbaalenii PYR-1	Copeland et al., Submitted (28-DEC-2006) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
3-ketosteroid-Δ1-dehydrogenase	YP_001132789	Mycobacterium gilvum PYR-GCK	Copeland et al., Submitted (13-APR-2007) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA

Gene/enzyme	RefSeq ID	Description	Reference/submission information
Probable dehydrogenase	YP_001701374	Mycobacterium abscessus	Ripoll et al., PLoS ONE 4 (6), E5660 (2009)
3-ketosteroid-Δ1-dehydrogenase	YP_116669	Nocardia farcinica IFM 10152	Ishikawa et al., Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14925-14930 (2004)
3-ketosteroid-Δ1-dehydrogenase	YP_704476	Rhodococcus jostii RHA1	McLeod et al., Proc. Natl. Acad. Sci. U.S.A. 103 (42), 15582-15587 (2006)
3-ketosteroid-Δ1-dehydrogenase	YP_002781639	Rhodococcus opacus B4	Takarada et al., Submitted (02-APR-2009) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
3-ketosteroid-Δ1-dehydrogenase	ZP_04388196	Rhodococcus erythropolis SK121	Sebastian et al., Submitted (20-APR-2009) J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
3-ketosteroid-Δ1-dehydrogenase	YP_002764184	Rhodococcus erythropolis PR4	Sekine et al., Environ. Microbiol. 8 (2), 334-346 (2006)
3-ketosteroid-Δ1-dehydrogenase	ZP_05478502	Streptomyces sp. AA4	Fischbach et al., Submitted (09-FEB-2009) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
3-ketosteroid-Δ1-dehydrogenase	YP_923969	Nocardioides sp. JS614	Coleman et al., J. Bacteriol. 193 (13), 3399-3400 (2011)
3-ketosteroid-Δ1-dehydrogenase	ZP_06607635	Aeromicrobium marinum DSM 15272	Qin et al., Submitted (10-APR-2009) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3-ketosteroid-Δ1-dehydrogenase	YP_001413067	Parvibaculum lavamentivorans DS-1	Schleheck et al., Arch. Microbiol. 183 (6), 369-377 (2005)
3-ketosteroid-Δ1-dehydrogenase	YP_001696275	Lysinibacillus sphaericus C3-41	Hu et al., J. Bacteriol. 190 (8), 2892-2902 (2008)

Gene/enzyme	RefSeq ID	Description	Reference/submission information
3-ketosteroid-Δ1-dehydrogenase	ZP_01725238	Bacillus sp. B14905	Edwards et al., Submitted (12-JUN-2006) J Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
putative FAD binding domain protein	ZP_05095058	marine gamma proteobacterium HTCC2148	Amann et al., Submitted (16-JUL-2008) J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
Fumarate reductase/succinate dehydrogenase flavoprotein subunit	ZP_01626281	marine gamma proteobacterium HTCC2080	Thrash et al., J. Bacteriol. 192 (14), 3842-3843 (2010)
putative FAD binding domain protein	ZP_05094533	marine gamma proteobacterium HTCC2148	Amann et al., Submitted (16-JUL-2008) J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
fumarate reductase/succinate dehydrogenase flavoprotein domain protein	ZP_03544409	Comamonas testosteroni KF-1	Schleheck et al., Appl. Environ. Microbiol. 70 (7), 4053-4063 (2004)
hypothetical protein	ZP_03607292	Methanobrevibacter smithii DSM 2375	Sudarsanam et al., Submitted (27-OCT-2008) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
short chain dehydrogenase	YP_447755	Methanosphaera stadtmanae DSM 3091	Fricke et al., J. Bacteriol. 188 (2), 642-658 (2006)

Gene/enzyme	RefSeq ID	Description	Reference/submission information
hypothetical protein	ZP_03636608	Holdemania filiformis DSM 12042	Sudarsanam et al., Submitted (19-DEC-2008) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
short-chain dehydrogenase/reductase SDR	YP_003182859	Eggerthella lenta DSM 2243	Lucas et al., Submitted (02-SEP-2009) US DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598-1698, USA
hypothetical protein	ZP_03759971	Clostridium asparagiforme DSM 15981	Sudarsanam et al., Submitted (14-JAN-2009) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
short chain dehydrogenase/reductase	ZP_04672086	Clostridiales bacterium 1_7_47_FAA	Allen-Vercoe et al., Submitted (13-AUG-2008) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
hypothetical protein	ZP_03777187	Clostridium hylemonae DSM 15053	Sudarsanam et al., Submitted (02-FEB-2009) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
hypothetical protein	ZP_02432210	Clostridium scindens ATCC 35704	Sudarsanam et al.,Submitted (07-NOV-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA

Gene/enzyme	RefSeq ID	Description	Reference/submission information
hypothetical protein	ZP_03293133	Clostridium hiranonis DSM13275	Sudarsanam et al., Submitted (03-FEB-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
hypothetical protein	ZP_02236388	Dorea formicigenerans ATCC 27755	Sudarsanam et al., Submitted (08-OCT-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
hypothetical protein	ZP_03463615	[Bacteroides] pectinophilus ATCC43243	Sudarsanam et al., Submitted (13-NOV-2008) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
Oxidoreductase, short chain dehydrogenase/reductase family protein	ZP_01772477	Collinsella aerofaciens ATCC 25986	Sudarsanam et al., Submitted (25-APR-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
hypothetical protein	ZP_03296838	Collinsella stercoris DSM 13279	Sudaranam et al., Submitted (01-OCT-2008) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA

Gene/enzyme	RefSeq ID	Description	Reference/submission information
hypothetical protein	ZP_03799118	Coprococcus comes ATCC 27758	Sudaranam et al., Submitted (18-FEB-2009) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
hypothetical protein	ZP_02027406	Eubacterium ventriosum ATCC27560	Sudaranam et al., Submitted (27-MAR-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
short chain dehydrogenase	ZP_05744752	Lactobacillus antri DSM 16041	Qin et al., Submitted (10-APR-2009) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
oxidoreductase, short chain dehydrogenase/reductase family protein	ZP_03682514	Catenibacterium mitsuokai DSM 15897	Fulton et al., Submitted (16-DEC-2008) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
hypothetical protein	ZP_02075484	Clostridium sp. L2-50	Sudaranam et al., Submitted (08-AUG-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA

Gene/enzyme	RefSeq ID	Description	Reference/submission information
hypothetical protein	ZP_02205767	Coprococcus eutactus ATCC 27759	Sudaranam et al., Submitted (11-SEP-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
3-oxoacyl-(acyl-carrier-protein) reductase	ZP_06113400	Clostridium hathewayi DSM 13479	Weinstock et al., Submitted (28-APR-2009) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
hypothetical protein	ZP_02429777	Clostridium ramosum DSM 1402	Sudaranam et al., Submitted (09-NOV-2007) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA
short-chain dehydrogenase	ZP_04565678	Coprobacillus sp. D7	Ward et al., Submitted (05-FEB-2009) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
short chain dehydrogenase/reductase	YP_001274304	Methanobrevibacter smithii ATCC 35061	Samuel et al., Proc. Natl. Acad. Sci. U.S.A. 104 (25), 10643-10648 (2007)
short-chain dehydrogenase	ZP_05976067	Methanobrevibacter smithii DSM 2374	Weinstock et al., Submitted (30-SEP-2009) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA

Gene/enzyme	RefSeq ID	Description	Reference/submission information
hypothetical protein	YP_637479	Mycobacterium sp. MCS	Copeland et al., Submitted (05-JUN-2006) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
hypothetical protein	YP_936319	Mycobacterium sp. KMS	Copeland et al., Submitted (22-DEC-2006) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
hypothetical protein	YP_001537700	Salinispora arenicola CNS-205	Penn et al., ISME J 3 (10), 1193-1203 (2009)
hypothetical protein	YP_001159497	Salinispora tropica CNB-440	Eustaquio et al., Proc. Natl. Acad. Sci. U.S.A. 106 (30), 12295-12300 (2009)
hypothetical protein	ZP_05478503	Streptomyces sp. AA4	Fischbach et al., Submitted (09-FEB-2009) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
hypothetical protein	YP_116670	Nocardia farcinica IFM 10152	Ishikawa et al., Proc. Natl. Acad. Sci. U.S.A. 101 (41), 14925-14930 (2004)
Probable dehydrogenase	YP_001701375	Mycobacterium abscessus	Ripoll et al., PLoS ONE 4 (6), E5660 (2009)
hypothetical protein	YP_908262	Mycobacterium ulcerans Agy99	Stinear et al., Genome Res. 17 (2), 192-200 (2007)
hypothetical protein	YP_001848529	Mycobacterium marinum M	Stinear et al., Genome Res. 18 (5), 729-741 (2008)
hypothetical protein	ZP_04747468	Mycobacterium kansasii ATCC12478	Veyrier et al., Submitted (08-DEC-2008) Department of Medicine, McGill University, A5.156, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4, Canada

Gene/enzyme	RefSeq ID	Description	Reference/submission information
putative 3-ketosteroid delta(4)(5alpha)-dehydrogenase	YP_002764185	Rhodococcus erythropolis PR4	Sekine et al., Environ. Microbiol. 8 (2), 334-346 (2006)
3-ketosteroid-delta4 dehydrogenase	ZP_04388024	Rhodococcus erythropolis SK121	Sebastian et al., Submitted (20-APR-2009) J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
fumarate reductase/succinate dehydrogenase	YP_003277400	Comamonas testosteroni CNB-2	Ma et al., Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
fumarate reductase/succinate dehydrogenase flavoprotein domain protein	ZP_03544408	Comamonas testosteroni KF-1	Schleheck et al., Submitted (02-JAN-2009) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
hypothetical protein	YP_728803	Ralstonia eutropha H16	Pohlmann et al., Nat. Biotechnol. 24 (10), 1257-1262 (2006)
hypothetical protein	YP_001796921	Cupriavidus taiwanensis LMG 19424	Amadou et al., Genome Res. 18 (9), 1472-1483 (2008)
hypothetical protein	YP_001166106	Novosphingobium aromaticivorans DSM 12444	Copeland et al., Submitted (08-MAY-2007) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
hypothetical protein	YP_001263819	Sphingomonas wittichii RW1	Miller et al., J. Bacteriol. 192 (22), 6101-6102 (2010)
hypothetical protein	YP_001262358	Sphingomonas wittichii RW1	Miller et al., J. Bacteriol. 192 (22), 6101-6102 (2010)
fumarate reductase/succinate dehydrogenase flavoprotein domain	ZP_03265787	Burkholderia sp. H160	Ormeno-Orrillo et al., J. Bacteriol. 194 (24), 6927 (2012)
hypothetical protein	YP_713249	Frankia alni ACN14a	Normand et al., Genome Res. 17 (1), 7-15 (2007)

Gene/enzyme	RefSeq ID	Description	Reference/submission information
hypothetical protein	YP_001135820	Mycobacterium gilvum PYR-GCK	Copeland et al., Submitted (13-APR-2007) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
hypothetical protein	YP_887193	Mycobacterium smegmatis str. MC2 155	Fleischmann et al., Submitted (29-NOV-2006) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
hypothetical protein	YP_001068597	Mycobacterium sp. JLS	Copeland et al., Submitted (20-FEB-2007) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
fumarate reductase/succinate dehydrogenase	YP_003277399	Comamonas testosteroni CNB-2	Ma et al., Appl. Environ. Microbiol. 75 (21), 6812-6819 (2009)
3-ketosteroid-delta1-dehydrogenase	YP_340631	Pseudoalteromonas haloplanktis TAC125	Medigue et al., Genome Res. 15 (10), 1325-1335 (2005)
fumarate reductase/succinate dehydrogenase flavoprotein domain-containing protein	YP_001674315	Shewanella halifaxensis HAW-EB4	Copeland et al., Submitted (12-FEB-2008) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
fumarate reductase/succinate dehydrogenase flavoprotein domain-containing protein	YP_001501980	Shewanella pealeana ATCC 700345	Copeland et al., Submitted (09-OCT-2007) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
fumarate reductase/succinate dehydrogenase flavoprotein, N-	YP_295753	Ralstonia eutropha JMP134	Copeland et al., Submitted (18-AUG-2005) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA

Gene/enzyme	RefSeq ID	Description	Reference/submission information
terminal:FAD dependent oxidoreductase			
3-oxosteroid 1-dehydrogenase	YP_728801	Ralstonia eutropha H16	Pohlmann et al., Nat. Biotechnol. 24 (10), 1257-1262 (2006)
3-oxosteroid 1-dehydrogenase	YP_001796919	Cupriavidus taiwanensis LMG 19424	Amadou et al., Genome Res. 18 (9), 1472-1483 (2008)
fumarate reductase/succinate dehydrogenase flavoprotein domain protein	ZP_02893600	Burkholderia ambifaria IOP40-10	Copeland et al., Submitted (03-MAR-2008) US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
fumarate reductase/succinate dehydrogenase flavoprotein subunit	YP_372172	Burkholderia lata	Copeland et al., Submitted (11-OCT-2005) US DOE Joint Genome Institute, 2400 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
putative FAD-binding component for oxidoreductase	YP_002234257	Burkholderia cenocepacia J2315	Holden et al., J. Bacteriol. 191 (1), 261-277 (2009)
3-oxosteroid 1-dehydrogenase	YP_001262337	Sphingomonas wittichii RW1	Miller et al., J. Bacteriol. 192 (22), 6101-6102 (2010)
fumarate reductase/succinate dehydrogenase flavoprotein domain-containing protein	YP_001262329	Sphingomonas wittichii RW1	Miller et al., J. Bacteriol. 192 (22), 6101-6102 (2010)
fumarate reductase/succinate dehydrogenase flavoprotein domain-containing protein	YP_001166117	Novosphingobium aromaticivorans DSM 12444	Copeland et al., Submitted (08-MAY-2007) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA

Gene/enzyme	RefSeq ID	Description	Reference/submission information
fumarate reductase/succinate dehydrogenase flavoprotein domain-containing protein	YP_001263842	Sphingomonas wittichii RW1	Miller et al., J. Bacteriol. 192 (22), 6101-6102 (2010)
fumarate reductase/succinate dehydrogenase flavoprotein domain-containing protein	YP_001166011	Novosphingobium aromaticivorans DSM 12444	Copeland et al., Submitted (08-MAY-2007) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
3-ketosteroid-delta-1-dehydrogenase	YP_640946	Mycobacterium sp. MCS	Copeland et al., Submitted (12-JUN-2006) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
3-ketosteroid-delta-1-dehydrogenase	YP_939840	Mycobacterium sp. KMS	Copeland et al., Submitted (22-DEC-2006) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
fumarate reductase/succinate dehydrogenase flavoprotein domain-containing protein	YP_003301095	Thermomonospora curvata DSM 43183	Chertkov et al., Stand Genomic Sci 4 (1), 13-22 (2011)
3-ketosteroid-delta-1-dehydrogenase	YP_002783051	Rhodococcus opacus B4	Takarada et al., Submitted (02-APR-2009) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
3-ketosteroid-delta-1-dehydrogenase	YP_705733	Rhodococcus jostii RHA1	McLeod et al., Proc. Natl. Acad. Sci. U.S.A. 103 (42), 15582-15587 (2006)
3-oxosteroid 1-dehydrogenase	ZP_04387793	Rhodococcus erythropolis SK121	Sebastian et al., Submitted (20-APR-2009) J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA

Gene/enzyme	RefSeq ID	Description	Reference/submission information
3-ketosteroid-delta-1-dehydrogenase	YP_702446	Rhodococcus jostii RHA1	McLeod et al., Proc. Natl. Acad. Sci. U.S.A. 103 (42), 15582-15587 (2006)
3-ketosteroid-delta-1-dehydrogenase	YP_002779402	Rhodococcus opacus B4	Takarada et al., Submitted (02-APR-2009) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
3-ketosteroid-delta-1-dehydrogenase	YP_001159498	Salinispora tropica CNB-440	Eustaquio et al., Proc. Natl. Acad. Sci. U.S.A. 106 (30), 12295-12300 (2009)
3-ketosteroid-delta-1-dehydrogenase	YP_001537701	Salinispora arenicola CNS-205	Penn et al., ISME J 3 (10), 1193-1203 (2009)
3-ketosteroid-delta-1-dehydrogenase	YP_001159451	Salinispora tropica CNB-440	Eustaquio et al., Proc. Natl. Acad. Sci. U.S.A. 106 (30), 12295-12300 (2009)
3-ketosteroid-delta-1-dehydrogenase	YP_001537652	Salinispora arenicola CNS-205	Penn et al., ISME J 3 (10), 1193-1203 (2009)
3-ketosteroid-delta-1-dehydrogenase	NP_822771	Streptomyces avermitilis MA-4680 = NBRC 14893	Ikeda et al., Nat. Biotechnol. 21 (5), 526-531 (2003)
3-ketosteroid-delta-1-dehydrogenase	ZP_05016500	Streptomyces sviceus ATCC 29083	Fischbach et al., Submitted (25-FEB-2008) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
3-ketosteroid-delta-1-dehydrogenase	ZP_05533430	Streptomyces viridochromogenes DSM 40736	Fischbach et al., Submitted (09-FEB-2009) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
oxidoreductase	YP_003491889	Streptomyces scabiei 87.22	Bignell et al., Mol. Plant Microbe Interact. 23 (2), 161-175 (2010)

Gene/enzyme	RefSeq ID	Description	Reference/submission information
3-ketosteroid-delta-1-dehydrogenase	YP_003115795	Catenulispora acidiphila	Copeland et al., Stand Genomic Sci 1 (2), 119-125 (2009)
3-ketosteroid-delta-1-dehydrogenase	ZP_05001264	Streptomyces sp. Mg1	Fischbach et al., Submitted (25-FEB-2008) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
3-ketosteroid-delta-1-dehydrogenase	ZP_05504907	Streptomyces sp. C	Fischbach et al., Submitted (09-FEB-2009) Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA
succinate dehydrogenase/fumarate reductase flavoprotein subunit	ZP_04025274	Tsukamurella paurometabola DSM 20162	Lucas et al., Submitted (26-FEB-2009) US DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598-1698, USA
succinate dehydrogenase/fumarate reductase flavoprotein subunit	YP_003272100	Gordonia bronchialis DSM 43247	Ivanova et al., Stand Genomic Sci 2 (1), 19-28 (2010)
succinate dehydrogenase/fumarate reductase flavoprotein subunit	ZP_04387329	Rhodococcus erythropolis SK121	Sebastian et al., Submitted (20-APR-2009) J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
3-ketosteroid delta(1)-dehydrogenase	YP_002766394	Rhodococcus erythropolis PR4	Sekine et al., Environ. Microbiol. 8 (2), 334-346 (2006)
3-ketosteroid-delta-1-dehydrogenase	NP_959464	Mycobacterium avium subsp. paratuberculosis K-10	Li et al., Proc. Natl. Acad. Sci. U.S.A. 102 (35), 12344-12349 (2005)

Gene/enzyme	RefSeq ID	Description	Reference/submission information
3-ketosteroid-delta-1-dehydrogenase	YP_879904	Mycobacterium avium 104	Fleischmann et al., Submitted (19-OCT-2006) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
3-ketosteroid-delta-1-dehydrogenase	ZP_05215097	Mycobacterium avium subsp. avium ATCC 25291	Turenne et al., Submitted (22-JAN-2009) Medicine, McGill University Health Centre, 1650 Cedar Ave, Montreal, Quebec H3G 1A4, Canada
3-ketosteroid-delta-1-dehydrogenase	ZP_05227620	Mycobacterium intracellulare ATCC 13950	Turenne et AL., Submitted (30-NOV-2007) McGill University and Genome Quebec Innovation Centre, 740 Avenue Docteur Penfield, Montreal, Quebec H3A 1A4, Canada
3-ketosteroid-delta-1-dehydrogenase	YP_001853283	Mycobacterium marinum M	Stinear et al., Genome Res. 18 (5), 729-741 (2008)

Table Appendix 0-7: BLAST comparison of the genomes R. equi ATCC13557, Sphingomonas sp. KC8 and P. putida SJTE-1.

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1	451	Translation initiation factor 2	60	2037	54.93	30	1481	48.35
5	457	putative DNA-damage-inducible protein F	97	3044	29.44	39	2200	25.44
8	309	tRNA pseudouridine synthase B (EC 4.2.1.70)	60	2035	43.38	30	1477	43.95
11	315	Riboflavin kinase (EC 2.7.1.26) / FMN adenylyltransferase (EC 2.7.7.2)	97	3078	35.89	57	3343	37.67
12	90	SSU ribosomal protein S15p (S13e)	60	2034	55.06	30	1476	51.69
15	447	FIG007959: peptidase, M16 family	73	2479	27.69	30	1384	36.07
18	253	4-hydroxy-tetrahydrodipicolinate reductase (EC 1.17.1.8)	60	2051	33.21	18	695	31.79
26	303	4-hydroxy-tetrahydrodipicolinate synthase (EC 4.3.3.7)	111	3698	35.58	39	2192	35.84
31	906	Cell division protein FtsK	126	4245	44.36	54	3066	44.38
36	163	C-terminal domain of CinA type S	145	4833	33.6	59	3449	40.16
47	297	Diacylglycerol kinase-related protein	30	930	29.21	26	1114	24.58

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
48	543	Alkyldihydroxyacetonephosphate synthase (EC 2.5.1.26)	134	4415	35.88	14	505	35.85
54	346	RecA protein	51	1506	68.11	39	2197	68.75
61	506	tRNA-i(6)A37 methylthiotransferase	110	3657	30.63	59	3417	40.56
66	309	tRNA dimethylallyltransferase (EC 2.5.1.75)	65	2229	35.78	20	834	36.17
67	290	Diaminopimelate epimerase (EC 5.1.1.7)	77	2597	29.33	62	3632	29.83
72	382	Butyryl-CoA dehydrogenase (EC 1.3.99.2)	122	4194	36.86	42	2378	52.66
73	243	SOS-response repressor and protease LexA (EC 3.4.21.88)	30	913	37.5	39	2167	35.59
75	163	Ribonucleotide reductase transcriptional regulator NrdR	94	2981	44.9	34	1818	44.09
76	1314	ATP-dependent helicase hrpA	114	3775	48.34	66	3891	37.61
77	410	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62)	13	473	38.27	66	3822	35.39
79	325	Ribokinase (EC 2.7.1.15)	16	575	44.41	68	3929	32.69
86	310	Hydrogen peroxide-inducible genes activator	79	2677	38.38	42	2448	36.08

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
96	318	UDP-glucose 4-epimerase (EC 5.1.3.2)	3	41	47.68	40	2366	45.83
99	389	NAD-independent protein deacetylase AcuC	79	2709	33.57	51	2906	32.93
100	899	possible acetyl-CoA synthetase	146	4868	28.41	7	155	33.33
110	402	putative transmembrane transport protein	99	3174	30.58	42	2424	29.84
111	460	RNA polymerase sigma factor RpoD	90	2851	59.41	62	3599	57.98
117	103	Deoxyuridine 5'-triphosphate nucleotidohydrolase (EC 3.6.1.23)	79	2655	44.29	39	2097	41.18
125	423	23S rRNA (Uracil-5-) -methyltransferase RumA (EC 2.1.1.-)	60	1975	29.73	43	2479	30.65
130	608	Glutaminase (EC 3.5.1.2)	14	543	39.86	19	746	38.09
131	657	1-deoxy-D-xylulose 5-phosphate synthase (EC 2.2.1.7)	94	2994	39.65	6	80	39.88
139	404	Ribonuclease D (EC 3.1.26.3)	112	3729	30.69	68	3926	31.98
143	356	Uroporphyrinogen III decarboxylase (EC 4.1.1.37)	73	2439	43.18	66	3850	39.77

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
147	117	Transcriptional regulator, GntR family	79	2711	42.67	30	1528	46.55
150	147	Peptide methionine sulfoxide reductase MsrB (EC 1.8.4.12)	34	1162	45.8	66	3833	52.07
160	684	Threonyl-tRNA synthetase (EC 6.1.1.3)	16	568	41.68	66	3779	41.18
164	376	Phosphatidylinositol alpha-mannosyltransferase (EC 2.4.1.57)	66	2268	34.33	8	253	30.26
170	251	FIG000859: hypothetical protein	110	3674	44.98	27	1222	41.53
171	192	Crossover junction endodeoxyribonuclease RuvC (EC 3.1.22.4)	110	3675	39.74	27	1223	39.29
172	202	Holliday junction DNA helicase RuvA	110	3676	37.16	27	1224	47.33
173	378	Holliday junction DNA helicase RuvB	110	3677	56.11	27	1226	52.48
175	552	Protein-export membrane protein SecD (TC 3.A.5.1.1)	105	3299	26.95	26	972	36
176	392	Protein-export membrane protein SecF (TC 3.A.5.1.1)	105	3300	29.01	26	971	25.56

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
178	198	Adenine phosphoribosyltransferase (EC 2.4.2.7)	78	2634	44.44	8	207	45.75
179	779	GTP pyrophosphokinase (EC 2.7.6.5), (p)ppGpp synthetase I	79	2670	40.2	27	1204	36.47
183	427	Histidyl-tRNA synthetase (EC 6.1.1.21)	106	3319	42.73	66	3783	44.18
186	335	Malate dehydrogenase (EC 1.1.1.37)	99	3116	23.79	26	1000	23.83
198	519	Multicopper oxidase	93	2958	41.41	18	625	45.35
202	154	Bona fide RidA/YjgF/TdcF/RutC subgroup	139	4603	30.97	40	2371	39.62
203	342	Radical SAM domain protein	45	1271	35.6	39	2340	36.53
207	310	YpfJ protein, zinc metalloprotease superfamily	60	1978	41.7	26	1016	38.64
208	594	Aspartyl-tRNA synthetase (EC 6.1.1.12) @ Aspartyl-tRNA(Asn) synthetase (EC 6.1.1.23)	110	3673	49.06	39	2188	43.41
222	469	ATPase, AAA family	126	4247	44.32	26	1138	40.36
223	891	Alanyl-tRNA synthetase (EC 6.1.1.7)	115	3855	41.76	40	2370	42.75
224	147	Putative Holliday junction resolvase YqgF	68	2331	31.68	39	2125	29.23
225	484	FIG004453: protein YceG like	34	1114	35.33	42	2406	24.93

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
227	379	Chorismate synthase (EC 4.2.3.5)	40	1205	33.05	45	2595	37.61
228	185	Shikimate kinase I (EC 2.7.1.71)	73	2444	42.5	51	2920	41.1
229	370	3-dehydroquinate synthase (EC 4.2.3.4)	73	2443	42.33	51	2919	42.36
230	141	3-dehydroquinate dehydratase II (EC 4.2.1.10)	17	633	46.32	6	42	45.45
233	188	Translation elongation factor P	35	1177	30.39	18	569	35.91
234	168	Transcription termination protein NusB	94	2986	41.89	38	2057	35.87
236	314	Aspartate carbamoyltransferase (EC 2.1.3.2)	68	2333	50.99	62	3589	53.31
237	460	Dihydroorotase (EC 3.5.2.3)	68	2334	34.07	61	3558	50
239	387	Carbamoyl-phosphate synthase small chain (EC 6.3.5.5)	60	2050	44.16	62	3595	43.73
240	1116	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5)	60	2049	57.58	62	3593	56.08
243	169	Guanylate kinase (EC 2.7.4.8)	79	2663	40.91	54	3114	45.45
245	429	Phosphopantothenoylcysteine decarboxylase (EC 4.1.1.36)	79	2654	41.5	39	2099	43.77

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
		/ Phosphopantothenoylcysteine synthetase (EC 6.3.2.5)						
246	404	S-adenosylmethionine synthetase (EC 2.5.1.6)	68	2298	56.37	26	1090	50
249	339	Magnesium and cobalt transport protein CorA	36	1191	37.35	66	3766	23.98
253	308	Methionyl-tRNA formyltransferase (EC 2.1.2.9)	42	1228	41.33	39	2179	39.47
254	496	Ribosomal RNA small subunit methyltransferase B (EC 2.1.1.-)	42	1227	32.87	51	2828	37.25
263	224	Ribulose-phosphate 3-epimerase (EC 5.1.3.1)	90	2879	46.82	30	1471	49.49
264	337	Diaminohydroxyphosphoribosylaminopyrimidine deaminase (EC 3.5.4.26) / 5-amino-6-(5-phosphoribosylamino)uracil reductase (EC 1.1.1.193)	94	2982	38.63	54	3124	45.68
265	204	Riboflavin synthase eubacterial/eukaryotic (EC 2.5.1.9)	94	2983	47.98	54	3123	40.1

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
266	416	3,4-dihydroxy-2-butanone 4-phosphate synthase (EC 4.1.99.12) / GTP cyclohydrolase II (EC 3.5.4.25)	94	2990	56.29	54	3063	52.38
267	154	6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.78)	94	2985	44.14	54	3121	38.4
269	695	Excinuclease ABC subunit C	122	4184	35.14	21	901	35.99
270	300	FIG000506: Predicted P-loop-containing kinase	108	3415	37.98	51	2816	43.6
273	340	NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)	109	3473	53.03	18	562	52.41
274	409	Phosphoglycerate kinase (EC 2.7.2.3)	68	2294	41.69	18	564	44.58
275	262	Triosephosphate isomerase (EC 5.3.1.1)	60	2040	47.62	39	2176	45.35
280	932	Phosphoenolpyruvate carboxylase (EC 4.1.1.31)	59	1677	36.03	54	3149	47.54
284	250	6-phosphogluconolactonase (EC 3.1.1.31), eukaryotic type	109	3487	29.65	30	1618	28.32

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
286	484	Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49)	124	4212	42.04	30	1617	41.06
288	703	Transketolase (EC 2.2.1.1)	68	2296	42.79	18	561	43.5
290	330	Heme O synthase, protoheme IX farnesyltransferase (EC 2.5.1.-) COX10-CtaB	45	1284	39.69	30	1391	39.16
304	421	Cysteine desulfurase (EC 2.8.1.7), SufS subfamily	59	1655	40.15	66	3883	47.39
306	135	PaaD-like protein (DUF59) involved in Fe-S cluster assembly	146	4883	38.14	7	152	36.27
312	544	ABC transporter ATP-binding protein	53	1528	35.93	54	3183	36.54
317	934	Aconitate hydratase (EC 4.2.1.3) @ 2-methylisocitrate dehydratase (EC 4.2.1.99)	30	942	55.8	66	3798	54.59
320	377	COG0714: MoxR-like ATPases	33	1027	47.68	30	1407	43.65
321	319	hypothetical protein PA3071	33	1028	30.94	30	1408	28.64

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
323	241	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	34	1118	47.9	42	2403	42.02
358	452	FIG016551: Putative peptidase	59	1659	26.01	59	3368	28.99
360	357	Dihydroorotate dehydrogenase (EC 1.3.3.1)	30	960	43.24	66	3876	44.97
362	289	Heme ABC transporter, ATPase component HmuV	60	2011	45.7	30	1492	42.53
367	283	Undecaprenyl-diphosphatase (EC 3.6.1.27)	9	275	35.9	56	3289	31.97
373	386	L-cysteine:1D-myo-inosityl 2-amino-2 -deoxy-alpha-D-glucopyranoside ligase MshC	9	232	38.57	42	2471	32.01
383	384	FIG00995179: hypothetical protein	32	993	30.52	6	97	35.29
385	1193	5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)	17	667	32.73	26	1112	35.94
387	284	Rhodanese domain protein UPF0176, Actinobacterial subgroup	47	1375	32.55	6	53	31.01
390	284	ATP phosphoribosyltransferase (EC 2.4.2.17)	108	3432	30.41	38	2059	32.35
397	342	possible transcriptional regulator	97	3060	41.79	51	2788	42.65

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
404	473	possible diacylglycerol kinase	62	2150	33.08	19	780	39.02
409	318	Twin-arginine translocation protein TatC	68	2353	32.86	21	862	37.36
425	245	Transcriptional regulator, GntR family	60	2060	36.23	54	3119	48.15
433	1199	CobN component of cobalt chelatase involved in B12 biosynthesis	140	4674	38.57	30	1436	38.07
440	401	putative integrase	147	4937	29.18	50	2775	32.36
445	517	TldD family protein, Actinobacterial subgroup	66	2280	31.22	54	3163	27.59
452	544	Apolipoprotein N-acyltransferase (EC 2.3.1.-) in lipid-linked oligosaccharide synthesis cluster	86	2843	38.46	26	1089	28.4
458	712	Para-aminobenzoate synthase, amidotransferase component (EC 2.6.1.85) / Para-aminobenzoate synthase, aminase component (EC 2.6.1.85)	90	2883	42.64	39	2172	49.46
470	147	Transcriptional regulator, AraC family	17	658	38.54	39	2338	37.76
483	632	alkyl sulfatase (EC 3.1.6.-)	33	1013	51.05	34	1720	28.09

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
493	759	DEAD-box ATP-dependent RNA helicase CshA (EC 3.6.4.13)	60	1890	41.59	66	3786	45.76
494	487	6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44)	124	4211	37.03	22	914	32.48
498	342	Pirin, N-terminal:Pirin, C-terminal	142	4770	41.18	54	3111	40.82
507	724	Malate synthase G (EC 2.3.3.9)	86	2819	70.64	56	3221	62.01
514	947	Glycine dehydrogenase [decarboxylating] (glycine cleavage system P protein) (EC 1.4.4.2)	109	3452	60.78	59	3409	39.4
521	135	Glycine cleavage system H protein	77	2561	50.39	59	3406	46.36
525	204	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (EC 2.7.8.5)	122	4185	35.16	51	2916	39.77
527	211	Hypothetical nudix hydrolase YeaB	109	3529	52.78	45	2600	42.57
532	231	Cytidylate kinase (EC 2.7.4.25)	46	1362	41.82	39	2157	42.08
533	322	Pseudouridine synthase family protein	115	3828	43.25	38	2075	38.57
534	238	Segregation and condensation protein B	115	3826	37.06	21	859	38.18

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
535	274	Segregation and condensation protein A	115	3825	30.5	21	858	32.91
536	310	Chromosome (plasmid) partitioning protein ParA	81	2779	47.24	66	3861	44.98
539	555	CTP synthase (EC 6.3.4.2)	59	1629	53.48	26	1156	51.64
542	592	DNA repair protein RecN	60	2055	31.8	26	968	34.97
543	314	NAD kinase (EC 2.7.1.23)	33	1047	36.36	56	3215	36.17
557	491	Aldehyde dehydrogenase (EC 1.2.1.3)	135	4443	41.75	59	3392	53.09
558	250	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	33	1075	39.52	47	2643	41.04
563	395	CAIB/BAIF family protein	45	1336	37.97	60	3513	41.14
579	566	Fumarate hydratase class I, aerobic (EC 4.2.1.2)	108	3361	31.54	20	808	31.24
580	340	Alcohol dehydrogenase (EC 1.1.1.1)	132	4338	32.9	63	3711	34.12
584	159	3-demethylubiquinone-9 3-methyltransferase	136	4503	46.79	56	3267	53.75
603	360	GTP-binding and nucleic acid-binding protein YchF	99	3180	50.55	30	1466	51.09
609	144	Transcriptional regulator, MarR family	45	1354	35.42	68	3917	36.36

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
613	414	FIG00761799: membrane protein	122	4121	28.66	30	1400	26.79
617	335	4-hydroxy-3-methylbut-2-enyl diphosphate reductase (EC 1.17.1.2)	97	3082	49.68	59	3402	52.94
619	398	Exodeoxyribonuclease VII large subunit (EC 3.1.11.6)	109	3493	33.1	30	1591	35.13
620	77	Exodeoxyribonuclease VII small subunit (EC 3.1.11.6)	94	2997	41.82	49	2763	39.29
623	469	Fumarate hydratase class II (EC 4.2.1.2)	108	3410	51.85	54	3029	51.78
629	437	Serine hydroxymethyltransferase (EC 2.1.2.1)	80	2760	54.39	34	1819	54.4
647	165	Transcription elongation factor GreA	60	2047	36.51	62	3592	36.03
650	385	Cystathionine gamma-lyase (EC 4.4.1.1)	112	3726	43.68	66	3800	37.03
652	461	Cystathionine beta-synthase (EC 4.2.1.22)	114	3750	42.77	62	3675	40.66
658	409	Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52)	100	3217	33.73	66	3785	40.22
667	319	Exopolyphosphatase (EC 3.6.1.11)	77	2586	34.5	30	1627	34.37

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
670	429	Enolase (EC 4.2.1.11)	126	4267	58.5	20	813	61.79
672	303	Ferrous iron transport permease EfeU	53	1512	29.41	44	2498	28.33
675	536	ABC transporter ATP-binding protein	140	4670	34.87	61	3532	36.81
676	229	Nucleoside triphosphate pyrophosphohydrolase MazG (EC 3.6.1.8)	51	1478	35.62	59	3439	33.04
677	1226	Transcription-repair coupling factor	30	908	46.25	26	1027	39.31
679	490	N-acetylglucosamine-1-phosphate uridyltransferase (EC 2.7.7.23) / Glucosamine-1-phosphate N-acetyltransferase (EC 2.3.1.157)	81	2769	44.35	42	2463	39.87
680	327	Ribose-phosphate pyrophosphokinase (EC 2.7.6.1)	99	3183	40.58	54	3061	44.73
688	393	Cysteine desulfurase (EC 2.8.1.7)	29	845	33.11	62	3688	38.89
691	198	LSU ribosomal protein L25p	99	3182	32.29	30	1463	35.64
692	193	Peptidyl-tRNA hydrolase (EC 3.1.1.29)	99	3181	42.16	30	1465	37.57
722	431	sensor histidine kinase	108	3367	31.19	29	1288	34.48

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
740	266	SSU rRNA (adenine(1518)-N(6)/adenine (1519)-N(6))-dimethyltransferase (EC 2.1.1.182)	90	2865	36.36	51	2830	39.69
745	280	Putative deoxyribonuclease YcfH	34	1111	37.92	62	3604	34.51
746	529	Methionyl-tRNA synthetase (EC 6.1.1.10)	62	2131	27.32	62	3605	62.4
750	277	rRNA small subunit methyltransferase I	60	1858	36.4	26	1175	37.14
757	97	Arsenate reductase (EC 1.20.4.1)	51	1490	45.92	6	39	51.04
759	133	Redox-sensitive transcriptional activator SoxR	32	998	52.34	30	1527	53.91
770	210	Ribosomal-protein-S5p-alanine acetyltransferase	105	3276	27.27	34	1826	27.84
772	269	UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)	132	4355	34.8	66	3768	38.02
777	146	Large-conductance mechanosensitive channel	60	1966	35.66	30	1473	37.74
779	159	Molybdenum cofactor biosynthesis protein MoaB	30	933	36.03	29	1298	36.89
780	462	Heat shock protein HtrA	59	1751	42.11	57	3327	42.58

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
781	456	Osmosensitive K+ channel histidine kinase KdpD (EC 2.7.3.-)	59	1761	35	51	2818	33.13
783	204	FIG00996993: hypothetical protein	65	2235	31.68	26	1093	32.12
788	55	LSU ribosomal protein L33p @ LSU ribosomal protein L33p, zinc-independent	78	2650	46.94	35	1931	59.09
790	85	SSU ribosomal protein S18p @ SSU ribosomal protein S18p, zinc-independent	65	2214	45.61	42	2398	46.15
801	129	Lactoylglutathione lyase and related lyases	13	489	62.9	66	3801	33.66
806	523	IMP cyclohydrolase (EC 3.5.4.10) / Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)	62	2159	44.78	30	1469	47.97
812	301	Succinyl-CoA ligase [ADP-forming] alpha chain (EC 6.2.1.5)	121	4083	53.36	26	1001	50.66
813	390	Succinyl-CoA ligase [ADP-forming] beta chain (EC 6.2.1.5)	121	4082	41.48	54	3050	38.86
816	825	ATP-dependent DNA helicase UvrD/PcrA	79	2721	39.46	42	2439	38

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
820	531	Glucose-6-phosphate isomerase (EC 5.3.1.9)	60	2026	46.55	59	3430	40.53
842	161	Dihydrofolate reductase (EC 1.5.1.3)	74	2498	37.66	57	3344	40.99
852	330	Rod shape-determining protein MreB	108	3399	37.5	29	1321	36.73
860	406	Flavohemoprotein (Hemoglobin-like protein) (Flavohemoglobin) (Nitric oxide dioxygenase) (EC 1.14.12.17)	104	3266	35.87	25	950	32.44
871	117	HspR, transcriptional repressor of DnaK operon	100	3195	43.1	39	2198	40.54
878	198	FIG00998616: hypothetical protein	7	125	61.24	6	82	37.31
879	352	Transposase	7	130	68.21	44	2573	33.54
881	326	Quinone oxidoreductase (EC 1.6.5.5)	7	132	68.62	66	3854	45.4
882	343	Transcriptional regulator, AraC family	7	126	40	37	1968	26.47
884	455	Transporter, MFS superfamily	7	129	52.8	47	2642	29.52
895	266	alginate biosynthesis regulatory protein AlgR (lytT)	109	3533	38.84	54	3157	37.09
896	410	Autolysis histidine kinase LytS	116	3940	32.73	54	3158	38.94
903	793	Beta-glucosidase (EC 3.2.1.21)	59	1780	29.41	37	1973	35.28

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
912	341	ATPase	17	690	42.77	51	2869	34.65
920	308	putative oxidoreductase	33	1055	43.96	26	1018	43.69
921	271	Enoyl-CoA hydratase (EC 4.2.1.17)	134	4429	37.9	37	1962	39.63
926	193	Copper metallochaperone, bacterial analog of Cox17 protein	63	2173	33.02	26	1182	38.89
930	300	ABC transporter, ATP-binding protein	77	2576	40.99	68	3933	39.62
932	502	Acetyl-coenzyme A carboxyl transferase alpha chain (EC 6.4.1.2) / Acetyl-coenzyme A carboxyl transferase beta chain (EC 6.4.1.2)	33	1081	39.06	42	2384	37.97
935	335	Putative oxidoreductase YncB	43	1240	39.22	21	903	48.33
944	400	NADH:flavin oxidoreductases, Old Yellow Enzyme family	7	154	32.68	39	2212	38.54
947	458	Deoxyribodipyrimidine photolyase (EC 4.1.99.3)	100	3194	36.4	51	2910	39.66
949	221	Transcriptional regulator, GntR family	8	229	32.09	44	2560	38.35
950	282	Fumarylacetoacetate hydrolase family protein	50	1428	31.25	44	2559	43.63

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
956	242	FKBP-type peptidyl-prolyl cis-trans isomerase	116	3964	42.34	30	1402	37.96
957	433	Citrate synthase (si) (EC 2.3.3.1)	121	4074	57.25	51	2855	55.97
960	217	Pyridoxamine 5'-phosphate oxidase (EC 1.4.3.5)	110	3593	48.95	45	2588	42.02
962	373	Phosphoserine aminotransferase (EC 2.6.1.52)	46	1365	21.56	26	1144	23.28
974	301	NADH pyrophosphatase (EC 3.6.1.22)	125	4225	36.73	26	1149	36.53
996	483	Nitrate/nitrite transporter	30	963	29.19	62	3702	29.17
1000	393	Sulfur carrier protein adenylyltransferase ThiF	100	3190	45.34	39	2096	45.16
1006	546	putative ATP-dependent RNA helicase	68	2313	45.29	14	508	41.71
1017	225	RNA polymerase sigma-E factor	9	248	32.72	21	869	35.45
1018	167	Cys-tRNA(Pro) deacylase YbaK	109	3544	51.57	13	491	42.41
1019	342	luciferase family protein	147	4946	50.32	7	119	53.27
1032	946	Protein export cytoplasm protein SecA ATPase RNA helicase (TC 3.A.5.1.1)	60	1840	46.95	61	3560	48.93
1040	484	Adenosylhomocysteinase (EC 3.3.1.1)	68	2308	49.81	61	3573	62.21
1045	509	Amino acid permease	13	475	26.64	54	3037	40.76
1048	461	Phosphomannomutase (EC 5.4.2.8)	79	2656	35.45	54	3128	36.53

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1060	396	4-carboxymuconolactone decarboxylase (EC 4.1.1.44)	60	1802	44.92	54	3127	42.86
1061	277	Pca regulon regulatory protein PcaR	60	1808	42.11	26	1129	26.85
1062	404	Acetyl-CoA acetyltransferase (EC 2.3.1.9)	146	4878	47.45	39	2280	42.14
1063	274	possible sugar phosphate isomerase/ epimerase	13	467	44.72	52	2955	29.5
1065	365	FIG00995642: hypothetical protein	13	468	30.17	44	2542	25.85
1066	268	oxidoreductase of aldo/keto reductase family, subgroup 1	135	4458	38.1	52	2951	38.57
1068	297	Quinate/shikimate 5-dehydrogenase I delta (EC 1.1.1.25)	17	634	49.64	66	3847	33.7
1076	428	benzoate MFS transporter BenK	2	5	36.3	30	1539	32.54
1079	301	dTDP-Rha:A-D-GlcNAc-diphosphoryl polyprenol, A-3-L-rhamnosyl transferase WbbL	3	35	29.68	54	2981	28.99
1080	282	dTDP-4-dehydrorhamnose reductase (EC 1.1.1.133)	94	2968	26.69	54	2985	42.51

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1087	170	Phosphoribosylaminoimidazole carboxylase catalytic subunit (EC 4.1.1.21)	79	2703	59.74	54	3053	64.63
1088	422	Phosphoribosylaminoimidazole carboxylase ATPase subunit (EC 4.1.1.21)	79	2702	35.21	54	3052	34.81
1100	304	Thiosulfate sulfurtransferase, rhodanese (EC 2.8.1.1)	65	2241	31.02	57	3333	33
1123	503	Glycerol kinase (EC 2.7.1.30)	109	3542	56.54	26	965	51.75
1125	401	Transcriptional regulator, GntR family domain / Aspartate aminotransferase (EC 2.6.1.1)	78	2644	31.87	54	2990	30.84
1145	584	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	121	4077	48.44	51	2872	49.05
1146	259	Succinate dehydrogenase iron-sulfur protein (EC 1.3.99.1)	121	4078	42.74	51	2871	39.52
1147	445	D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4)	132	4381	30.43	62	3608	32.22
1149	348	Tryptophanyl-tRNA synthetase (EC 6.1.1.2)	60	1873	35.33	39	2107	48.35

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1155	438	O-acetylhomoserine sulfhydrylase (EC 2.5.1.49) / O-succinylhomoserine sulfhydrylase (EC 2.5.1.48)	14	536	54.44	51	2812	57.31
1158	290	Methylenetetrahydrofolate dehydrogenase (NADP+) (EC 1.5.1.5) / Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9)	29	786	46.5	35	1947	45.36
1161	182	tRNA (cytidine(34)-2'-O)-methyltransferase (EC 2.1.1.207)	70	2386	51.97	8	212	36.42
1163	346	Putative membrane protein YeiH	135	4469	37.29	62	3678	32.38
1171	103	rhodanese-related sulfurtransferase	70	2389	30.49	65	3739	38.3
1173	1094	DNA polymerase III alpha subunit (EC 2.7.7.7)	4	54	42.17	46	2640	39.11
1194	526	GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2)	109	3498	51.81	30	1581	54.21
1197	489	Inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205)	109	3497	56.2	51	2827	58.2

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1207	101	Heat shock protein 60 family co-chaperone GroES	60	1824	40.62	26	1095	51.04
1210	446	Large subunit naph/bph dioxygenase	80	2752	32.5	44	2530	30.49
1211	123	3-phenylpropionate dioxygenase ferredoxin subunit	7	127	32.89	31	1655	43.4
1213	190	Biphenyl dioxygenase subunit beta (EC 1.14.12.18) (Biphenyl 2,3-dioxygenase)	2	8	25.64	44	2529	34.13
1230	347	TsaD/Kae1/Qri7 protein, required for threonylcarbamoyladenosine t(6)A37 formation in tRNA	90	2854	44.54	54	3107	43.38
1231	166	Ribosomal-protein-S18p-alanine acetyltransferase (EC 2.3.1.-)	99	3156	32.26	59	3422	44.74
1232	225	TsaB protein, required for threonylcarbamoyladenosine (t(6)A) formation in tRNA	59	1675	33.52	59	3423	52.63

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1234	182	TsaE protein, required for threonylcarbamoyladenosine t(6)A37 formation in tRNA	65	2232	43.37	61	3571	35
1236	409	Alanine racemase (EC 5.1.1.1)	78	2638	35.43	6	48	35.56
1237	488	NAD(P)HX epimerase / NAD(P)HX dehydratase	65	2233	36.46	43	2478	39.66
1240	621	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16)	81	2767	44.71	42	2464	45.66
1244	443	Phosphoglucosamine mutase (EC 5.4.2.10)	60	2041	45.62	34	1832	46.36
1245	169	SSU ribosomal protein S9p (S16e)	60	1868	54.55	6	73	50
1246	148	LSU ribosomal protein L13p (L13Ae)	60	1869	52.82	6	74	45.86
1258	381	Butyryl-CoA dehydrogenase (EC 1.3.99.2)	140	4684	44.62	12	384	38.74
1259	297	tRNA pseudouridine synthase A (EC 4.2.1.70)	33	1083	34.75	62	3616	35.71
1260	181	LSU ribosomal protein L17p	90	2943	50.52	56	3231	47.27
1261	355	DNA-directed RNA polymerase alpha subunit (EC 2.7.7.6)	90	2942	42.99	56	3232	45.42

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1262	202	SSU ribosomal protein S4p (S9e)	90	2941	47.09	34	1815	41.01
1263	138	SSU ribosomal protein S11p (S14e)	90	2940	60.63	56	3233	62.1
1264	123	SSU ribosomal protein S13p (S18e)	90	2939	57.26	56	3234	56.56
1266	86	Translation initiation factor 1	126	4242	69.01	19	744	58.9
1269	182	Adenylate kinase (EC 2.7.4.3)	59	1676	42.47	3	22	43.98
1270	441	Preprotein translocase secY subunit (TC 3.A.5.1.1)	90	2937	43.25	19	750	41.83
1271	147	LSU ribosomal protein L15p (L27Ae)	90	2936	39.46	19	751	44.67
1272	60	LSU ribosomal protein L30p (L7e)	90	2935	38.6	19	752	43.86
1273	214	SSU ribosomal protein S5p (S2e)	90	2934	53.01	19	753	51.95
1274	136	LSU ribosomal protein L18p (L5e)	90	2933	54.26	19	754	61.29
1275	180	LSU ribosomal protein L6p (L9e)	90	2932	47.68	19	755	52.84
1276	133	SSU ribosomal protein S8p (S15Ae)	90	2931	50	19	756	46.21
1278	189	LSU ribosomal protein L5p (L11e)	90	2929	58.52	19	758	59.22
1279	106	LSU ribosomal protein L24p (L26e)	90	2928	38.83	19	759	42.59
1280	123	LSU ribosomal protein L14p (L23e)	90	2927	68.03	19	760	64.23

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1281	91	SSU ribosomal protein S17p (S11e)	90	2926	63.51	19	761	52.05
1282	79	LSU ribosomal protein L29p (L35e)	90	2925	42.37	19	762	49.12
1283	139	LSU ribosomal protein L16p (L10e)	90	2924	54.87	19	763	57.46
1284	266	SSU ribosomal protein S3p (S3e)	90	2923	53.54	19	764	52.38
1285	135	LSU ribosomal protein L22p (L17e)	90	2922	57.14	19	765	45.54
1286	94	SSU ribosomal protein S19p (S15e)	90	2921	64.13	19	766	61.96
1287	278	LSU ribosomal protein L2p (L8e)	90	2920	64.6	19	767	49.82
1288	102	LSU ribosomal protein L23p (L23Ae)	90	2919	48.1	19	768	43.33
1289	222	LSU ribosomal protein L4p (L1e)	90	2918	46.72	19	769	32.67
1290	218	LSU ribosomal protein L3p (L3e)	90	2917	53.4	19	770	48.54
1291	102	SSU ribosomal protein S10p (S20e)	90	2916	60	19	771	56.44
1306	352	Probable phenylacetic acid degradation NADH oxidoreductase paaE (EC 1.-.-.-)	80	2753	32.86	22	907	32.63
1308	196	Transcriptional regulator, TetR family	97	3069	41.94	62	3681	50.88
1309	256	Glutamate racemase (EC 5.1.1.3)	100	3191	35.35	34	1823	36.87
1320	397	Translation elongation factor Tu	90	2903	66.75	19	772	71.75

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1321	701	Translation elongation factor G	122	4169	58.61	19	773	61.77
1322	157	SSU ribosomal protein S7p (S5e)	90	2913	62.6	19	774	56.41
1323	125	SSU ribosomal protein S12p (S23e)	90	2912	71.54	19	775	73.98
1331	533	Methylcrotonyl-CoA carboxylase carboxyl transferase subunit (EC 6.4.1.4)	122	4193	51.95	38	2022	49.16
1338	319	Alcohol dehydrogenase (EC 1.1.1.1)	77	2579	45.28	26	1071	37
1343	1320	DNA-directed RNA polymerase beta' subunit (EC 2.7.7.6)	90	2911	50.7	11	378	51.77
1344	1163	DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)	90	2910	56.95	11	379	58.35
1354	257	Conserved hypothetical integral membrane protein YrbE1A	108	3426	31.71	54	3144	34.72
1355	349	Methionine ABC transporter ATP-binding protein	108	3425	38.93	39	2279	40.73
1356	129	LSU ribosomal protein L7/L12 (P1/P2)	90	2909	52.8	6	25	54.47
1357	186	LSU ribosomal protein L10p (P0)	90	2908	33.93	6	26	36.42

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1358	238	LSU ribosomal protein L1p (L10Ae)	90	2907	52	6	27	57.4
1359	145	LSU ribosomal protein L11p (L12e)	90	2906	55.14	6	28	57.45
1360	266	Transcription antitermination protein NusG	90	2905	43.68	6	29	42.02
1366	302	Probable protease htpX homolog (EC 3.4.24.-)	35	1165	29.74	26	994	40.36
1369	233	2-heptaprenyl-1,4-naphthoquinone methyltransferase MenG (EC 2.1.1.163)	68	2346	39.83	59	3407	37.86
1381	540	ATP-dependent DNA helicase RecQ	115	3807	38.9	66	3776	42.94
1389	393	Probable low-affinity inorganic phosphate transporter	122	4179	38.84	57	3318	41.88
1414	325	Cytochrome c-type biogenesis protein CcsA/ResC	116	3956	34.65	28	1265	31.76
1434	326	Porphobilinogen synthase (EC 4.2.1.24)	8	223	48.75	13	410	53.68
1436	335	Porphobilinogen deaminase (EC 2.5.1.61)	53	1519	40.97	54	3108	42.57
1445	267	Pyrroline-5-carboxylate reductase (EC 1.5.1.2)	73	2460	43.17	20	839	35.34
1457	95	hypothetical protein	60	1983	40	28	1258	35.82

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1460	445	Glycosyltransferase MshA involved in mycothiol biosynthesis (EC 2.4.1.-)	66	2273	38.16	49	2760	36.84
1479	368	UDP-N-acetylenolpyruvoylglucosamine reductase (EC 1.1.1.158)	34	1129	34.76	38	2095	36.22
1487	296	Formyltetrahydrofolate deformylase (EC 3.5.1.10)	60	1817	59.65	56	3206	47.54
1498	287	UDP-galactose-lipid carrier transferase (EC 2.-.-.-)	48	1380	33.78	39	2189	37.24
1511	231	Urea ABC transporter, ATPase protein UrtE	63	2182	42.99	54	3022	35.45
1512	1196	Urea carboxylase (EC 6.3.4.6)	122	4191	48.97	54	2996	49.08
1520	295	probable 3-oxoacyl-(acyl carrier protein) reductase	60	1910	36.65	59	3489	39.78
1527	468	Dihydrolipoamide dehydrogenase (EC 1.8.1.4)	121	4081	41.2	19	781	44.99
1533	316	Transcriptional regulator, AsnC family	77	2556	41.24	6	49	40.43
1539	542	Heat shock protein 60 family chaperone GroEL	60	1823	61.64	26	1096	58.79

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1545	263	CDP-diacylglycerol--serine O-phosphatidyltransferase (EC 2.7.8.8)	60	2000	32.12	8	205	42.99
1548	726	FIG00999703: hypothetical protein	60	2043	40.35	29	1294	42.61
1565	206	Peptide deformylase (EC 3.5.1.88)	42	1229	39.33	62	3613	40.94
1567	267	Exodeoxyribonuclease III (EC 3.1.11.2)	11	393	35.96	21	877	38.43
1573	284	Hydroxymethylpyrimidine phosphate kinase ThiD (EC 2.7.4.7)	62	2119	32.92	34	1833	39.92
1578	331	transcriptional regulator, AraC family	78	2613	37.5	52	2956	37.27
1582	250	Thiazole biosynthesis protein ThiG	73	2469	51.56	6	43	53.04
1584	356	Glycine oxidase ThiO (EC 1.4.3.19)	97	3088	30.47	12	400	26.21
1585	231	Thiamin-phosphate pyrophosphorylase (EC 2.5.1.3)	62	2120	30.69	18	567	31.1
1590	963	DNA topoisomerase I (EC 5.99.1.2)	30	917	43.99	34	1827	37.49
1601	273	Phosphoserine phosphatase (EC 3.1.3.3)	49	1416	35.19	54	3143	29.27
1608	663	Acetyl-coenzyme A synthetase (EC 6.2.1.1)	60	2027	51.57	39	2224	53.07

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1612	468	L-Proline/Glycine betaine transporter ProP	59	1722	52.07	42	2442	46.75
1613	403	Na+/H+ antiporter NhaA type	110	3596	41.36	54	3046	40.41
1618	223	Possible membrane-anchored thioredoxin-like protein	116	3957	39.39	61	3566	40.48
1619	227	Endonuclease III (EC 4.2.99.18)	110	3560	43.88	18	694	38.34
1623	225	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	90	2887	30.48	30	1537	29.3
1625	311	Chromosome initiation inhibitor	10	304	32.38	30	1557	40.78
1630	153	Bona fide RidA/YjgF/TdcF/RutC subgroup	29	805	34.09	30	1560	61.84
1631	260	4-oxalocrotonate decarboxylase (EC 4.1.1.77)	138	4561	30.29	30	1559	48.68
1645	774	Multimodular transpeptidase-transglycosylase (EC 2.4.1.129) (EC 3.4.-.-)	73	2472	36.97	30	1621	35.71
1648	188	ThiJ/PfpI family protein	11	376	42.13	39	2098	24.43
1655	459	FAD dependent oxidoreductase	76	2522	32.97	14	506	43.88

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1669	489	Catalase (EC 1.11.1.6)	90	2944	55.56	54	3117	54.26
1680	628	Acylamino-acid-releasing enzyme	86	2837	28.18	34	1814	29.18
1686	133	Arsenate reductase (EC 1.20.4.1)	11	383	32.33	44	2512	28.44
1701	344	Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11)	33	1085	36.02	28	1249	45.43
1702	422	Aspartokinase (EC 2.7.2.4)	115	3856	49.39	59	3388	47.87
1713	568	2-isopropylmalate synthase (EC 2.3.3.13)	109	3490	55.48	11	313	25.56
1717	253	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	33	1059	40.4	1	2	43.31
1745	207	Transcriptional regulator, GntR family	135	4481	40	47	2693	42.86
1754	195	Recombination protein RecR	118	4014	48.44	62	3614	41.67
1755	108	FIG000557: hypothetical protein co-occurring with RecR	117	4013	39.81	62	3683	31.19
1759	433	Cyclopropane-fatty-acyl-phospholipid synthase (EC 2.1.1.79)	79	2734	36.99	51	2911	38.58

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1768	213	regulatory protein, LuxR:Response regulator receiver	122	4183	35.45	22	915	33.77
1772	314	transcriptional regulator, AraC family	135	4465	38.57	51	2879	38.95
1777	718	DNA polymerase III subunits gamma and tau (EC 2.7.7.7)	117	4012	37.58	62	3684	38.34
1805	305	glutamyl-Q-tRNA synthetase	60	2018	43.21	30	1566	47.47
1806	406	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)	108	3431	46.19	68	3923	45.74
1807	167	ATP:Cob(I)alamin adenosyltransferase (EC 2.5.1.17)	60	1836	37.8	30	1568	47.62
1809	123	ATP synthase epsilon chain (EC 3.6.3.14)	81	2770	33.67	39	2288	37.78
1810	483	ATP synthase beta chain (EC 3.6.3.14)	81	2771	61.42	39	2289	61.52
1811	327	ATP synthase gamma chain (EC 3.6.3.14)	81	2772	38.56	39	2290	32.82
1812	548	ATP synthase alpha chain (EC 3.6.3.14)	81	2773	53.02	39	2291	56.78
1813	275	ATP synthase delta chain (EC 3.6.3.14)	81	2774	39.02	39	2292	39.74
1816	260	ATP synthase F0 sector subunit a	81	2776	27.91	21	896	32.14

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1819	220	TsaC protein (YrdC domain) required for threonylcarbamoyladenosine t(6)A37 modification in tRNA	115	3824	30.05	59	3436	35.78
1822	360	Peptide chain release factor 1	100	3188	46.81	66	3784	47.7
1823	78	LSU ribosomal protein L31p @ LSU ribosomal protein L31p, zinc-dependent	73	2452	56.94	8	191	39.13
1824	681	Transcription termination factor Rho	77	2584	56.99	66	3852	55.86
1827	438	Homoserine dehydrogenase (EC 1.1.1.3)	59	1710	38.44	54	3089	42.14
1829	551	Arginyl-tRNA synthetase (EC 6.1.1.19)	73	2454	26.74	21	855	40.11
1838	618	Sulfate adenylyltransferase subunit 1 (EC 2.7.7.4)	60	1879	47.52	51	2945	48.9
1839	309	Sulfate adenylyltransferase subunit 2 (EC 2.7.7.4)	60	1880	50.83	51	2944	49.83
1850	259	Enoyl-CoA hydratase (EC 4.2.1.17)	29	836	66.67	30	1531	58.17
1851	304	3-hydroxyisobutyrate dehydrogenase (EC 1.1.1.31)	60	1989	46.26	30	1530	54.36

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1852	359	3-hydroxyisobutyryl-CoA hydrolase (EC 3.1.2.4)	140	4685	39.88	30	1532	42.18
1853	383	Butyryl-CoA dehydrogenase (EC 1.3.99.2)	29	837	43.28	30	1533	57.07
1855	502	Methylmalonate-semialdehyde dehydrogenase (EC 1.2.1.27)	97	3072	57.46	30	1534	64.26
1865	416	N-acetyl-L,L-diaminopimelate deacetylase (EC 3.5.1.47)	4	67	48.82	26	997	42.96
1873	343	putative phosphotransferase	130	4319	33.22	13	408	33.44
1874	208	Transcriptional regulator, AsnC family	135	4476	36	62	3669	33.54
1875	366	Pyruvate dehydrogenase E1 component alpha subunit (EC 1.2.4.1)	116	3868	35.09	42	2418	34.97
1876	334	Pyruvate dehydrogenase E1 component beta subunit (EC 1.2.4.1)	116	3867	46.99	42	2419	45.1
1879	477	NAD(P) transhydrogenase subunit beta (EC 1.6.1.2)	45	1331	45.76	34	1843	47.59
1880	109	NAD(P) transhydrogenase alpha subunit (EC 1.6.1.2)	45	1332	45.35	34	1841	39.22

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
1881	372	NAD(P) transhydrogenase alpha subunit (EC 1.6.1.2)	45	1333	49.45	34	1840	43.94
1899	460	Glycolate dehydrogenase (EC 1.1.99.14), subunit GlcD	134	4416	39.25	56	3217	39.01
1950	293	Predicted L-lactate dehydrogenase, Fe-S oxidoreductase subunit YkgE	134	4414	21.77	56	3219	24.8
1969	260	Alkanesulfonates transport system permease protein	53	1566	52.87	30	1546	26.83
1972	455	nitrilotriacetate monooxygenase component A	11	355	37.16	49	2739	26.51
1977	710	Protease II (EC 3.4.21.83)	113	3738	45.06	8	162	40.88
1995	226	Phosphoribosylformylglycinamidine synthase, glutamine amidotransferase subunit (EC 6.3.5.3)	109	3504	29.26	18	551	59.02
1998	476	C4-dicarboxylate-transport transmembrane protein DctA	110	3648	48.05	20	784	45.05
2008	527	Amidophosphoribosyltransferase (EC 2.4.2.14)	33	1077	35.93	26	1075	44.69

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2009	360	Phosphoribosylformylglycinamidine cyclo-ligase (EC 6.3.3.1)	50	1470	44.67	39	2180	43.14
2014	506	Cytochrome d ubiquinol oxidase subunit I (EC 1.10.3.-)	60	1972	31.84	39	2265	32.11
2015	341	Cytochrome d ubiquinol oxidase subunit II (EC 1.10.3.-)	60	1971	32.65	39	2266	35.4
2031	302	Phosphate transport system permease protein PstA (TC 3.A.1.7.1)	79	2695	35.04	42	2458	36.54
2032	259	Phosphate transport ATP-binding protein PstB (TC 3.A.1.7.1)	138	4569	57.65	42	2457	58.59
2035	222	Phosphate transport system regulatory protein PhoU	79	2692	28.23	42	2456	31.67
2039	381	tRNA dihydrouridine synthase B (EC 1.-.-.-)	62	2157	38.34	59	3446	37.95
2051	69	Mobile element protein	59	1712	59.26	26	1171	68.29
2057	403	POSSIBLE OXIDOREDUCTASE	53	1549	25.96	34	1690	42.03
2079	305	Methylisocitrate lyase (EC 4.1.3.30)	28	750	42.35	29	1315	33.33

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2098	404	Serine phosphatase RsbU, regulator of sigma subunit	45	1307	37.14	11	368	36.27
2110	339	transcriptional regulator, AraC family protein	134	4408	30.52	17	545	33.51
2111	395	Aspartate aminotransferase (EC 2.6.1.1)	106	3323	40.21	6	64	43.4
2145	351	ABC transporter (iron.B12.siderophore.hemin) , permease component	17	623	43.75	30	1493	39.46
2156	637	Methionine ABC transporter ATP-binding protein	66	2265	35.28	54	3186	37.08
2157	145	Transcriptional regulator, MarR family	128	4298	36.11	25	937	39.24
2161	393	Valine--pyruvate aminotransferase (EC 2.6.1.66)	60	2016	34.97	6	51	44.09
2162	418	Phosphoribosylamine--glycine ligase (EC 6.3.4.13)	62	2161	46.59	30	1592	45.5
2168	465	Two-component system response phosphate sensor kinase, PhoR	79	2688	38.05	61	3572	38.74
2170	105	Quaternary ammonium compound-resistance protein SugE	50	1436	39.22	34	1760	42.31

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2204	255	Conserved hypothetical integral membrane protein YrbE1A	45	1317	31.67	39	2278	34.65
2216	254	3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) @ 17hydroxysteroid dehydrogenase type 10 (HSD10)-like	29	839	54.9	51	2892	60.94
2245	530	Long-chain fatty-acid-CoA ligase (EC 6.2.1.3), Mycobacterial subgroup FadD3	122	4196	35.34	39	2335	44.68
2286	410	CDP-4-dehydro-6-deoxy-D-glucose 3-dehydratase (EC 4.2.1.-)	116	3894	32.91	8	268	47.74
2304	153	Histone acetyltransferase HPA2 and related acetyltransferases	79	2713	30.94	35	1939	45.05
2306	333	Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94)	122	4114	41.92	54	3106	44.04
2307	369	D-alanine--D-alanine ligase (EC 6.3.2.4)	60	1846	32.49	21	848	34.46
2310	322	Thiamine-monophosphate kinase (EC 2.7.4.16)	94	2987	36.54	38	2055	35.14

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2312	226	Uracil-DNA glycosylase, family 1	59	1769	42.53	26	1086	45.74
2319	176	Ribosomal RNA small subunit methyltransferase D (EC 2.1.1.-)	100	3189	38.1	42	2443	42.5
2320	153	Phosphopantetheine adenylyltransferase (EC 2.7.7.3)	73	2490	39.74	49	2765	40.14
2324	232	Ribonuclease III (EC 3.1.26.3)	59	1748	40.47	43	2481	38.16
2327	567	Potassium-transporting ATPase A chain (EC 3.6.3.12) (TC 3.A.3.7.1)	122	4122	47.16	20	823	48.29
2328	699	Potassium-transporting ATPase B chain (EC 3.6.3.12) (TC 3.A.3.7.1)	122	4123	61.07	20	822	61.57
2329	217	Potassium-transporting ATPase C chain (EC 3.6.3.12) (TC 3.A.3.7.1)	122	4124	36.65	20	821	37.5
2331	227	DNA-binding response regulator KdpE	122	4126	50.23	20	820	42.86
2332	477	Amino-acid carrier protein AlsT	94	2962	42.83	34	1831	34.21
2338	1201	Chromosome partition protein smc	116	4006	45.73	26	1154	41.38

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2340	498	Signal recognition particle receptor protein FtsY (=alpha subunit) (TC 3.A.5.1.1)	73	2476	46.43	62	3634	46.74
2342	113	Nitrogen regulatory protein P-II	77	2603	59.82	6	59	63.39
2343	846	[Protein-PII] uridylyltransferase (EC 2.7.7.59)	59	1650	26.49	56	3229	27.04
2344	527	Signal recognition particle, subunit Ffh SRP54 (TC 3.A.5.1.1)	59	1719	47.36	28	1255	48.53
2348	294	Inner membrane protein translocase component YidC, long form	81	2783	36.78	39	2119	38.37
2350	312	Aldo-keto reductase	144	4798	31.08	51	2902	42.05
2356	263	probable ABC drug resistance transporter, permease component	12	442	27.52	51	2890	30.17
2361	157	Thiol peroxidase, Bcp-type (EC 1.11.1.15)	111	3696	40.3	21	872	37.86
2367	535	ATPase component BioM of energizing module of biotin ECF transporter	60	1812	43.94	57	3328	48.64
2368	373	Alcohol dehydrogenase (EC 1.1.1.1)	59	1623	37.37	33	1678	37.87

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2383	449	NADP-specific glutamate dehydrogenase (EC 1.4.1.4)	99	3133	56.5	51	2790	30.08
2387	375	Xanthine and CO dehydrogenases maturation factor, XdhC/CoxF family	128	4294	47.25	39	2267	35.26
2400	558	ABC transporter ATP-binding protein	99	3132	56.43	59	3453	55.95
2402	1622	NAD-specific glutamate dehydrogenase (EC 1.4.1.2), large form	31	978	41.58	30	1334	38.86
2421	862	Membrane alanine aminopeptidase N (EC 3.4.11.2)	33	1042	28.14	45	2586	28.83
2430	457	Cell division trigger factor (EC 5.2.1.8)	29	784	30.52	62	3630	31.18
2433	426	ATP-dependent Clp protease ATP-binding subunit ClpX	29	782	63.29	62	3626	62.06
2437	275	Formate dehydrogenase chain D (EC 1.2.1.2)	53	1585	29.64	13	409	34.43
2450	688	Phosphate acetyltransferase (EC 2.3.1.8)	101	3228	37.62	54	3136	31.83

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2460	408	O-acetylhomoserine sulfhydrylase (EC 2.5.1.49) / O-succinylhomoserine sulfhydrylase (EC 2.5.1.48)	33	1076	44.94	8	247	50.13
2467	405	Phosphoribosylglycinamide formyltransferase 2 (EC 2.1.2.-)	59	1723	65.38	56	3284	55.74
2476	422	transcriptional regulator, LuxR family	51	1500	42.19	29	1325	45.76
2491	430	Adenylosuccinate synthetase (EC 6.3.4.4)	65	2223	54.01	26	1141	43.93
2500	309	Cobalt-zinc-cadmium resistance protein CzcD	23	711	39.12	21	906	31.87
2505	239	DedA protein	94	3012	38.3	26	1172	25.95
2529	140	Ribosome-associated heat shock protein implicated in the recycling of the 50S subunit (S4 paralog)	53	1578	33.08	56	3241	34.48
2534	851	ClpB protein	97	3102	57.49	30	1414	57.02
2541	526	O-succinylbenzoic acid--CoA ligase (EC 6.2.1.26)	114	3771	36.94	24	925	40.4
2556	263	Short chain oxidoreductase	14	522	34.72	47	2667	38.7

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2564	299	Luciferase-like monooxygenase superfamily	11	336	32.61	38	2041	41.61
2572	190	Heat shock protein GrpE	60	2054	32.35	26	980	29.59
2573	614	Chaperone protein DnaK	60	2053	55.97	26	1181	59.97
2587	495	Uncharacterized iron-regulated membrane protein; Iron-uptake factor PiuB	63	2176	28.78	30	1585	27.35
2598	368	Peptide chain release factor 2	59	1687	48.31	57	3348	44.41
2603	158	tmRNA-binding protein SmpB	60	2059	46.81	39	2193	48.61
2607	334	Proline iminopeptidase (EC 3.4.11.5)	69	2363	44.86	39	2158	36.49
2612	538	Phosphoglucomutase (EC 5.4.2.2)	139	4596	61.51	62	3682	28.86
2621	358	Putative metal chaperone, involved in Zn homeostasis, GTPase of COG0523 family	60	1960	29.55	30	1435	29.21
2622	210	Cobalamin biosynthesis protein BluB @ 5,6-dimethylbenzimidazole synthase, flavin destructase family	50	1462	41.21	30	1495	36.32

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2624	138	Molybdenum cofactor biosynthesis protein MoaE	60	1891	33.83	51	2918	34.35
2625	337	Molybdenum cofactor biosynthesis protein MoaC / Molybdenum cofactor biosynthesis protein MoaB	60	1893	51.97	39	2169	51.33
2627	350	Molybdenum cofactor biosynthesis protein MoaA	34	1109	35.74	56	3214	38.8
2630	269	NAD synthetase (EC 6.3.1.5)	63	2207	51.75	54	3060	27.38
2639	215	(2Fe-2S)-binding domain protein	116	4004	47.48	66	3903	41.22
2648	188	NADPH:quinone oxidoreductase	122	4142	50.44	8	296	40.87
2652	729	Ribonucleotide reductase of class Ib (aerobic), alpha subunit (EC 1.17.4.1)	110	3638	28.04	20	806	26.32
2653	336	Ribonucleotide reductase of class Ib (aerobic), beta subunit (EC 1.17.4.1)	110	3636	22.38	20	804	24.12

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2659	550	Cytochrome c oxidase polypeptide I (EC 1.9.3.1)	45	1277	47.35	30	1392	46.01
2660	390	Phosphoserine phosphatase (EC 3.1.3.3)	65	2243	40.1	20	833	42.4
2674	267	Electron transfer flavoprotein, beta subunit	121	4064	31.43	54	3049	31.17
2675	319	Electron transfer flavoprotein, alpha subunit	121	4065	43.71	54	3048	43.69
2678	250	1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51)	108	3388	31.47	59	3418	33.56
2680	373	tRNA-specific 2-thiouridylase MnmA	125	4235	35.26	66	3765	42.02
2682	704	DNA ligase (EC 6.5.1.2)	116	4008	39.57	39	2100	40.81
2686	100	Aspartyl-tRNA(Asn) amidotransferase subunit C (EC 6.3.5.6) @ Glutamyl-tRNA(Gln) amidotransferase subunit C (EC 6.3.5.7)	108	3398	31.18	39	2128	35.63
2687	495	Aspartyl-tRNA(Asn) amidotransferase subunit A (EC 6.3.5.6)	108	3397	48.85	39	2129	49.09

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
		@ Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.7)						
2692	223	Putative two-component system response regulator	90	2874	33.33	34	1771	38.1
2694	505	Aspartyl-tRNA(Asn) amidotransferase subunit B (EC 6.3.5.6) @ Glutamyl-tRNA(Gln) amidotransferase subunit B (EC 6.3.5.7)	108	3396	41.72	39	2130	39.52
2701	650	Acetolactate synthase large subunit (EC 2.2.1.6)	60	2003	45.24	20	835	46.09
2702	168	Acetolactate synthase small subunit (EC 2.2.1.6)	60	2002	42.04	20	836	38.36
2703	338	Ketol-acid reductoisomerase (EC 1.1.1.86)	60	2001	56.88	20	837	59.39
2704	531	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	51	1486	37.45	26	1143	39.88
2705	338	3-isopropylmalate dehydrogenase (EC 1.1.1.85)	33	1088	38.95	8	245	40.56

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2712	478	3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)	33	1091	60.09	26	1068	58.95
2713	204	3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)	33	1090	47.12	26	1070	44.04
2714	243	DNA-binding protein HU / low-complexity, AKP-rich domain	16	562	40.23	62	3622	43.1
2716	740	Polyphosphate kinase (EC 2.7.4.1)	77	2587	41.02	39	2183	45.05
2719	229	Dihydrolipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex (EC 2.3.1.168)	121	4080	43.28	26	1003	45.56
2721	507	Cytosol aminopeptidase PepA (EC 3.4.11.1)	109	3445	38.24	51	2834	43.52
2722	372	Aminomethyltransferase (glycine cleavage system T protein) (EC 2.1.2.10)	77	2562	40.06	59	3405	36.93
2723	368	Branched-chain amino acid aminotransferase (EC 2.6.1.42)	140	4669	38.53	62	3611	55.65

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2725	370	Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase (EC 2.4.2.21)	50	1457	40.53	26	1012	42.51
2729	117	probable iron binding protein from the HesB_IscA_SufA family	90	2896	41.38	21	876	40.74
2732	333	Cytochrome c oxidase polypeptide II (EC 1.9.3.1)	45	1276	26.42	30	1393	26.07
2735	551	Ubiquinol--cytochrome c reductase, cytochrome B subunit (EC 1.10.2.2)	60	1866	27.19	8	210	25
2738	181	Cytochrome c oxidase polypeptide III (EC 1.9.3.1)	45	1279	35.42	30	1388	36.46
2740	352	Anthranilate phosphoribosyltransferase (EC 2.4.2.18)	90	2884	38.37	39	2171	40.88
2758	463	2-keto-3-deoxy-D-arabino-heptulosonate-7-phosphate synthase II (EC 2.5.1.54)	35	1170	52.5	29	1314	51.95
2764	345	Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29)	94	2996	35	49	2764	32.35

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2765	327	5,10-methylenetetrahydrofolate reductase (EC 1.5.1.20)	68	2309	35.42	26	1111	36.39
2775	335	rRNA small subunit methyltransferase H	60	1856	42.55	38	2084	37.8
2777	588	Cell division protein FtsI [Peptidoglycan synthetase] (EC 2.4.1.129)	60	1854	31.53	38	2086	27.61
2778	548	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase (EC 6.3.2.13)	60	1853	37.58	38	2087	42.23
2779	507	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-alanine ligase (EC 6.3.2.10)	60	1852	36	38	2088	35.76
2780	365	Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13)	60	1851	36.54	38	2089	38.1
2781	504	UDP-N-acetylmuramoylalanine--D-glutamate ligase (EC 6.3.2.9)	60	1850	34.55	38	2090	34.64

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2783	382	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase (EC 2.4.1.227)	60	1848	32.81	38	2092	30.89
2784	499	UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8)	60	1847	36.46	38	2093	37.93
2786	410	Cell division protein FtsZ (EC 3.4.24.-)	60	1843	50	21	851	52.26
2787	242	COG1496: Uncharacterized conserved protein	97	3100	38.56	20	797	41.61
2788	207	Hypothetical protein YggS, proline synthase co-transcribed bacterial homolog PROSC	73	2459	39.71	66	3870	35.82
2800	303	Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)	97	3099	37.01	62	3661	40.39
2804	311	RarD protein	86	2821	36.59	26	991	39.54
2811	451	Threonine dehydratase (EC 4.3.1.19)	142	4726	38.69	39	2162	35.28
2840	189	Carbonic anhydrase (EC 4.2.1.1)	45	1272	45.26	25	944	35.58
2849	552	Exoenzymes regulatory protein AepA precursor	78	2620	28.01	39	2246	34.59

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2850	538	Dipeptide-binding ABC transporter, periplasmic substrate-binding component (TC 3.A.1.5.2)	128	4284	26.23	39	2142	29.63
2859	454	Glutamine synthetase type I (EC 6.3.1.2)	114	3774	31.34	66	3827	31.06
2882	236	Transcriptional regulator, GntR family	108	3436	38.1	42	2431	43.4
2917	269	Monoglyceride lipase (EC 3.1.1.23)	76	2529	31.25	49	2735	35.78
2919	115	PhnB protein	114	3766	33.05	51	2854	34.15
2935	344	Transcription termination protein NusA	60	2038	36.31	30	1483	42.65
2936	182	FIG000325: clustered with transcription termination protein NusA	60	2039	35.71	30	1484	34.44
2940	204	transcriptional regulator, TetR family	116	3988	43.28	38	1985	41.18
2942	529	NADH-ubiquinone oxidoreductase chain N (EC 1.6.5.3)	122	4149	38.98	18	586	40.17
2943	527	NADH-ubiquinone oxidoreductase chain M (EC 1.6.5.3)	122	4150	36.51	18	585	37.12

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2944	644	NADH-ubiquinone oxidoreductase chain L (EC 1.6.5.3)	122	4151	43.39	18	584	41.18
2945	100	NADH-ubiquinone oxidoreductase chain K (EC 1.6.5.3)	122	4152	36.84	18	583	44.79
2946	277	NADH-ubiquinone oxidoreductase chain J (EC 1.6.5.3)	122	4153	32.89	18	582	30.94
2947	182	NADH-ubiquinone oxidoreductase chain I (EC 1.6.5.3)	122	4154	43.94	18	581	37.5
2948	435	NADH-ubiquinone oxidoreductase chain H (EC 1.6.5.3)	122	4155	37.42	18	580	45.99
2950	689	NADH-ubiquinone oxidoreductase chain F (EC 1.6.5.3)	122	4157	46.5	18	578	50.63
2952	252	NADH-ubiquinone oxidoreductase chain C (EC 1.6.5.3)	122	4159	43.37	18	574	37.66
2953	191	NADH-ubiquinone oxidoreductase chain B (EC 1.6.5.3)	122	4160	50	18	573	56.64

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
2954	123	NADH ubiquinone oxidoreductase chain A (EC 1.6.5.3)	122	4161	36.45	18	572	36.59
2965	582	Prolyl-tRNA synthetase (EC 6.1.1.15), bacterial type	110	3663	43.24	43	2484	41.87
2969	494	Multidrug efflux pump P55	7	144	32.65	7	128	35.38
2970	409	Siroheme synthase / Precorrin-2 oxidase (EC 1.3.1.76) / Sirohydrochlorin ferrochelatase (EC 4.99.1.4) / Uroporphyrinogen-III methyltransferase (EC 2.1.1.107)	30	965	40.17	39	2220	40.66
2973	197	Cob(I)alamin adenosyltransferase (EC 2.5.1.17)	50	1464	39.01	30	1437	40.93
2988	511	Cobyric acid synthase (EC 6.3.5.10)	50	1459	45.57	38	1974	45.75
2989	286	Methionine aminopeptidase (EC 3.4.11.18)	59	1649	50.2	6	61	48.75
3001	263	2-aminoethylphosphonate ABC transporter permease protein II (TC 3.A.1.9.1)	74	2502	34.93	14	518	36.69
3007	620	Flp pilus assembly protein TadB	33	1030	27.18	30	1457	29.28

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3008	457	Type II/IV secretion system ATP hydrolase TadA/VirB11/CpaF, TadA subfamily	141	4694	40.16	39	2285	49.61
3016	388	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (EC 1.17.7.1)	106	3318	50.44	35	1937	48.26
3017	411	Membrane-associated zinc metalloprotease	59	1641	28.07	8	196	27.69
3018	386	1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.267)	59	1642	43.83	8	197	46.72
3020	355	Ribosomal RNA large subunit methyltransferase N (EC 2.1.1.-)	106	3315	38.07	54	3021	39.7
3026	287	Phosphatidate cytidylyltransferase (EC 2.7.7.41)	59	1643	42.61	8	198	38.94
3027	186	Ribosome recycling factor	59	1645	43.78	8	200	42.53
3028	243	Uridine monophosphate kinase (EC 2.7.4.22)	59	1646	47.66	8	201	52.12
3029	275	Translation elongation factor Ts	59	1647	35.13	8	202	39.6
3030	277	SSU ribosomal protein S2p (SAe)	59	1648	50	8	203	53.39
3037	374	Rossmann fold nucleotide-binding protein Smf possibly involved in DNA uptake	42	1230	36.27	34	1825	37.26

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3038	503	MG(2+) CHELATASE FAMILY PROTEIN / ComM-related protein	78	2607	46.05	39	2123	41.68
3042	213	Ribonuclease HII (EC 3.1.26.4)	59	1634	46.63	34	1834	49.73
3043	258	Signal peptidase I (EC 3.4.21.89)	59	1749	29.06	30	1399	28.35
3044	114	LSU ribosomal protein L19p	59	1715	58.49	28	1251	55.08
3046	325	Sulfatase modifying factor 1 precursor (C-alpha-formyglycine- generating enzyme 1)	121	4093	30.99	38	1984	57.42
3047	813	Transcription accessory protein (S1 RNA-binding domain)	53	1572	62.79	30	1475	44.05
3050	233	tRNA (Guanine37-N1) -methyltransferase (EC 2.1.1.31)	59	1716	44.4	28	1252	43.17
3051	189	16S rRNA processing protein RimM	59	1717	26.88	28	1253	25.32
3053	149	SSU ribosomal protein S16p	59	1718	40	28	1254	44.44
3068	407	L-lactate dehydrogenase (EC 1.1.2.3)	60	2062	38.86	39	2245	50
3083	107	4Fe-4S ferredoxin, iron-sulfur binding	59	1615	52.73	56	3242	49.38

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3085	261	Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases	33	1022	35.75	30	1404	32.46
3094	435	Vanillate O-demethylase oxygenase subunit (EC 1.14.13.82)	134	4425	30.18	47	2651	32.2
3125	389	FIG01001855: hypothetical protein	11	346	24.47	49	2736	41.37
3131	272	FIG003003: hypothetical protein	147	4894	50.7	56	3265	46.82
3150	314	LysR-family transcriptional regulator	8	218	36.55	25	940	44.68
3153	185	Ribonuclease HI (EC 3.1.26.4)	122	4138	36.36	59	3400	37.14
3156	348	2,3-butanediol dehydrogenase, R-alcohol forming, (R)- and (S)-acetoin-specific (EC 1.1.1.4)	97	3027	40.45	19	719	37.5
3159	254	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	79	2739	40.56	20	791	44.49
3162	479	Aldehyde dehydrogenase (EC 1.2.1.3); Probable coniferyl aldehyde dehydrogenase (EC 1.2.1.68)	73	2486	44.18	33	1675	36.53

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3180	426	Tyrosyl-tRNA synthetase (EC 6.1.1.1)	90	2899	29.64	26	1029	42.68
3186	475	Argininosuccinate lyase (EC 4.3.2.1)	53	1516	43.08	66	3771	44.81
3187	400	Argininosuccinate synthase (EC 6.3.4.5)	110	3556	41.52	54	3081	43.19
3189	313	Ornithine carbamoyltransferase (EC 2.1.3.3)	109	3546	45.54	8	223	54.38
3190	407	Acetylornithine aminotransferase (EC 2.6.1.11)	115	3848	46.45	8	224	42.37
3191	308	Acetylglutamate kinase (EC 2.7.2.8)	79	2657	46.05	35	1950	45.49
3192	413	Glutamate N-acetyltransferase (EC 2.3.1.35) / N-acetylglutamate synthase (EC 2.3.1.1)	60	1839	38.96	61	3563	37.75
3193	342	N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38)	90	2895	41.79	11	341	32.86
3196	830	Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)	16	563	33.22	27	1195	33.92
3197	357	Phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20)	16	564	46.47	27	1196	43.84
3198	267	FIG011178: rRNA methylase	65	2216	28.57	51	2813	34.03
3199	130	LSU ribosomal protein L20p	16	565	56.41	27	1198	48.31

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3200	65	LSU ribosomal protein L35p	16	566	35.94	27	1199	44.07
3201	174	Translation initiation factor 3	16	567	52.52	54	3140	51.8
3205	991	Excinuclease ABC subunit A	90	2946	57.2	40	2344	57.45
3215	711	Excinuclease ABC subunit B	34	1103	55.89	54	3044	55.1
3217	420	Dephospho-CoA kinase (EC 2.7.1.24)	97	3109	35.82	66	3846	27.32
3227	913	DNA polymerase I (EC 2.7.7.7)	45	1297	38.43	54	3168	34.61
3236	445	Pyruvate kinase (EC 2.7.1.40)	116	3980	41.53	27	1215	42.53
3237	493	Glutamate synthase [NADPH] small chain (EC 1.4.1.13)	73	2440	36.62	56	3288	39.87
3238	1546	Glutamate synthase [NADPH] large chain (EC 1.4.1.13)	73	2441	42.28	56	3286	43.5
3240	315	Prolipoprotein diacylglyceryl transferase (EC 2.4.99.-)	75	2510	24.89	20	795	24.73
3241	268	Tryptophan synthase alpha chain (EC 4.2.1.20)	45	1254	39.33	42	2385	35.82
3242	388	Tryptophan synthase beta chain (EC 4.2.1.20)	45	1255	61.1	42	2386	62.63

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3243	270	Indole-3-glycerol phosphate synthase (EC 4.1.1.48)	90	2885	41.73	39	2170	43.68
3245	534	Anthranilate synthase, aminase component (EC 4.1.3.27)	90	2881	44.62	39	2174	42.08
3246	327	FIG074102: hypothetical protein	35	1167	43.83	66	3905	39.83
3262	269	UDP-glucose dehydrogenase (EC 1.1.1.22)	8	211	39.15	66	3793	43.81
3271	245	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	127	4269	38.07	44	2556	36.03
3279	291	Aldo-keto reductase	66	2262	30.9	38	2028	30.46
3282	293	N-formylglutamate deformylase (EC 3.5.1.68)	60	1943	40	30	1514	37.29
3304	418	tRNA-guanine transglycosylase (EC 2.4.2.29)	105	3297	38.44	8	305	40.93
3313	388	Glutaryl-CoA dehydrogenase (EC 1.3.99.7)	45	1335	46.48	38	2038	46.61
3319	153	Predicted transcriptional regulator of sulfate adenylyltransferase, Rrf2 family	106	3306	30.51	28	1272	31.82
3320	321	C-5 sterol desaturase (EC 1.3.-.-)	77	2588	37.67	54	3101	39.07
3323	444	putative integral membrane efflux protein	78	2619	32.21	34	1902	32.09

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3331	344	Ribonuclease BN (EC 3.1.-.-)	114	3761	37	45	2593	26.98
3343	145	tRNA-specific adenosine-34 deaminase (EC 3.5.4.-)	109	3502	50.78	38	2068	48.06
3345	317	Arogenate dehydrogenase (EC 1.3.1.43)	46	1363	38.72	54	3038	31.97
3361	142	putative phenylacetic acid degradation protein	53	1571	33.33	26	974	32.33
3369	701	Periplasmic aromatic aldehyde oxidoreductase , molybdenum binding subunit YagR	145	4834	39.62	20	793	33.33
3392	1162	helicase (Snf2/Rad54 family)	110	3606	31.46	13	477	37.58
3399	420	Quinone oxidoreductase (EC 1.6.5.5)	49	1417	44.62	39	2299	46.99
3409	315	Glycosyltransferase	23	718	34.2	54	3068	40.85
3411	332	dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)	108	3383	30.16	54	2984	47.06
3412	439	Permeases of the major facilitator superfamily	17	629	41.1	38	2026	38.11
3414	194	dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13)	57	1600	35.87	54	2983	36.31
3418	180	Superoxide dismutase [Mn] (EC 1.15.1.1)	108	3412	44.63	62	3637	38.15
3419	860	Valyl-tRNA synthetase (EC 6.1.1.9)	108	3442	42.15	8	172	44.84

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3420	477	Dihydrofolate synthase (EC 6.3.2.12) @ Folylpolyglutamate synthase (EC 6.3.2.17)	33	1080	29.87	42	2383	33.96
3422	140	Nucleoside diphosphate kinase (EC 2.7.4.6)	106	3314	48.09	51	2836	43.18
3425	104	LSU ribosomal protein L21p	99	3147	45.1	26	1066	47
3426	66	LSU ribosomal protein L27p	99	3148	50.85	26	1065	59.57
3427	487	GTP-binding protein Obg	99	3149	44.19	26	1061	42.12
3428	367	Glutamate 5-kinase (EC 2.7.2.11)	99	3150	43.9	26	1060	41.33
3430	155	Glyoxalase family protein	142	4728	42.67	38	2077	46.31
3438	462	D-serine/D-alanine/glycine transporter	109	3526	46.68	30	1538	38.32
3439	428	Gamma-glutamyl phosphate reductase (EC 1.2.1.41)	62	2148	49.14	8	177	49.51
3442	247	Nicotinate-nucleotide adenylyltransferase (EC 2.7.7.18)	62	2147	34.29	8	178	27.46
3443	139	Ribosomal silencing factor RsfA (former Iojap)	62	2146	33.66	8	179	31.11
3451	257	Aquaporin Z	116	4000	43.4	42	2440	43.28
3454	621	Translation elongation factor LepA	59	1750	53.31	57	3320	53.05

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3455	311	FIG00994303: hypothetical protein	132	4349	35.23	54	3152	37.8
3462	295	Sulfate transport system permease protein CysT	76	2538	45.31	34	1807	44.4
3463	269	Sulfate transport system permease protein CysW	76	2537	45.68	34	1808	45.54
3464	316	Sulfate and thiosulfate import ATP-binding protein CysA (EC 3.6.3.25)	76	2536	54.09	34	1809	60.7
3472	248	Phosphoadenylyl-sulfate reductase [thioredoxin] (EC 1.8.4.8) / Adenylyl-sulfate reductase [thioredoxin] (EC 1.8.4.10)	28	756	35.78	39	2216	35.9
3473	570	Ferredoxin--sulfite reductase, actinobacterial type (EC 1.8.7.1)	17	671	27.96	39	2218	26.87
3474	396	Hypothetical radical SAM family enzyme in heat shock gene cluster, similarity with CPO of BS HemN-type	73	2466	34.87	26	1177	34.38

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3478	254	Ribosomal RNA small subunit methyltransferase E (EC 2.1.1.-)	68	2319	32.37	27	1189	29.78
3482	334	Phosphate starvation-inducible protein PhoH, predicted ATPase	62	2125	56.11	59	3415	47.1
3483	181	Metal-dependent hydrolase YbeY, involved in rRNA and/or ribosome maturation and assembly	62	2126	38.94	59	3414	29.73
3485	306	GTP-binding protein Era	81	2782	46.43	66	3858	49.12
3490	268	Undecaprenyl diphosphate synthase (EC 2.5.1.31)	59	1644	40.25	8	199	46.09
3497	149	Zinc uptake regulation protein ZUR	45	1293	32.22	59	3420	30.89
3502	430	Deoxyguanosinetriphosphate triphosphohydrolase (EC 3.1.5.1)	30	953	36.29	21	854	47.26
3503	644	DNA primase (EC 2.7.7.-)	90	2852	35.6	62	3598	36.34
3513	389	3-oxoacyl-[acyl-carrier-protein] synthase, KASII (EC 2.3.1.179)	34	1116	41.28	42	2405	44.44

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3514	104	Acyl carrier protein	34	1117	40.28	42	2404	40.3
3515	291	Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)	34	1119	31.84	42	2402	30.94
3527	329	Adenosylcobinamide-phosphate synthase (EC 6.3.1.10)	50	1461	42.11	38	1975	39.51
3529	323	Cytochrome oxidase biogenesis protein Surf1, facilitates heme A insertion	45	1281	22.89	38	2005	26.61
3530	136	Low molecular weight protein tyrosine phosphatase (EC 3.1.3.48)	34	1130	39.85	26	1163	36.89
3545	435	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62)	68	2318	36.41	56	3238	38
3546	389	8-amino-7-oxononanoate synthase (EC 2.3.1.47)	86	2826	37.82	56	3235	38.95
3547	232	Dethiobiotin synthetase (EC 6.3.3.3)	86	2829	36.36	56	3237	34.88
3549	530	Alkyl hydroperoxide reductase protein F (EC 1.6.4.-)	16	597	77.3	62	3620	36.13

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3550	188	Alkyl hydroperoxide reductase protein C (EC 1.6.4.-)	16	599	54.35	42	2449	36.14
3564	367	Biotin synthase (EC 2.8.1.6)	86	2825	33.33	54	2998	36.92
3572	344	Quinolinate synthetase (EC 2.5.1.72)	110	3691	34.45	39	2137	39.88
3573	526	L-aspartate oxidase (EC 1.4.3.16)	59	1756	39.93	54	3167	42.17
3574	287	Quinolinate phosphoribosyltransferase [decarboxylating] (EC 2.4.2.19)	103	3238	41.22	39	2138	44.4
3580	372	glycosyl transferase, group 1 family protein	60	1814	33.59	11	374	35.14
3592	414	Histidinol dehydrogenase (EC 1.1.1.23)	108	3433	38.59	38	2058	42.79
3594	205	Imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19)	72	2424	42.03	54	3031	51.98
3597	212	Imidazole glycerol phosphate synthase amidotransferase subunit (EC 2.4.2.-)	72	2425	39.72	54	3032	39.63
3599	245	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (EC 5.3.1.16)	72	2427	34.44	54	3033	39.26

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
		/ Acting phosphoribosylanthranilate isomerase (EC 5.3.1.24)						
3601	258	Imidazole glycerol phosphate synthase cyclase subunit (EC 4.1.3.-)	72	2428	61.39	54	3034	62.14
3602	116	Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19)	68	2349	53.12	26	1077	51.06
3604	188	Transcriptional regulator, TetR family	30	883	48.17	14	515	51.11
3608	399	Beta-lactamase class C and other penicillin binding proteins	110	3591	43.08	26	1067	26.01
3616	346	Aliphatic amidase AmiE (EC 3.5.1.4)	137	4543	80.47	34	1810	27.44
3621	486	Methylated-DNA--protein-cysteine methyltransferase (EC 2.1.1.63)	99	3166	36.87	39	2294	39.24
3638	299	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	11	378	53.15	44	2539	40.94

Gene	**Length**	**Function**	**Contig**	**Gene**	**percent identity (%)**	**Contig**	**Gene**	**percent identity (%)**
***R. equi* ATCC13557**			***P. putida* SJTE-1**			***Sphingomonas* sp. KC8**		
3709	324	Transcriptional regulator containing an amidase domain and an AraC-type DNA-binding HTH domain	29	820	39.37	54	3077	34.88