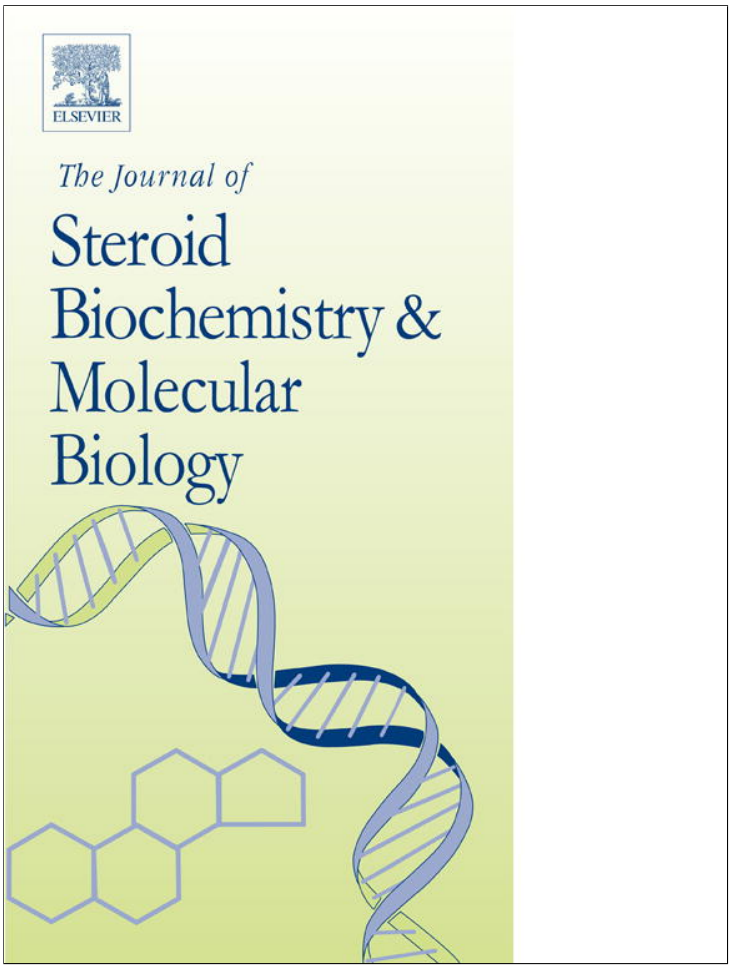

Journal of Steroid Biochemistry & Molecular Biology 138 (2013) 41–53

Contents lists available at SciVerse ScienceDirect

Journal of Steroid Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/jsbmb

Comparative analysis of genes encoding key steroid core oxidation enzymes in fast-growing *Mycobacterium* spp. strains

E.Yu. Bragin[a,b,*], V.Yu. Shtratnikova[a], D.V. Dovbnya[b], M.I. Schelkunov[a,d], Yu.A. Pekov[a], S.G. Malakho[a], O.V. Egorova[a], T.V. Ivashina[b], S.L. Sokolov[b], V.V. Ashapkin[a,c], M.V. Donova[b]

[a] *Center of Innovations and Technologies "Biological Active Compounds and Their Applications", Russian Academy of Sciences, Moscow 119991, Russian Federation*
[b] *G.K.Skryabin Institute of Biochemistry & Physiology of Microorganisms, Russian Academy of Sciences, Pushchino, Moscow Region, Russian Federation*
[c] *A.N.Belozersky Institute of Physico-Chemical Biology, M.V. Lomonosov Moscow State University, Moscow 119991, Russian Federation*
[d] *V. A. Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*

ARTICLE INFO

Article history:
Received 24 October 2012
Received in revised form 28 January 2013
Accepted 24 February 2013

Keywords:
Steroid bioconversion
Mycobacterium
Androst-4-ene-3,17-dione
Androst-1,4-diene-3,17-dione
Phytosterol
Genome sequencing

ABSTRACT

A comparative genome analysis of *Mycobacterium* spp. VKM Ac-1815D, 1816D and 1817D strains used for efficient production of key steroid intermediates (androst-4-ene-3,17-dione, AD, androsta-1,4-diene-3,17-dione, ADD, 9α-hydroxy androst-4-ene-3,17-dione, 9-OH-AD) from phytosterol has been carried out by deep sequencing. The assembled contig sequences were analyzed for the presence putative genes of steroid catabolism pathways. Since 3-ketosteroid-9α-hydroxylases (KSH) and 3-ketosteroid-Δ^1-dehydrogenase (Δ^1 KSTD) play key role in steroid core oxidation, special attention was paid to the genes encoding these enzymes.

At least three genes of Δ^1 KSTD (*kstD*), five genes of KSH subunit A (*kshA*), and one gene of KSH subunit B of 3-ketosteroid-9α-hydroxylases (*kshB*) have been found in *Mycobacterium* sp. VKM Ac-1817D. Strains of *Mycobacterium* spp. VKM Ac-1815D and 1816D were found to possess at least one *kstD*, one *kshB* and two *kshA* genes. The assembled genome sequence of *Mycobacterium* sp. VKM Ac-1817D differs from those of 1815D and 1816D strains, whereas these last two are nearly identical, differing by 13 single nucleotide substitutions (SNPs). One of these SNPs is located in the coding region of a *kstD* gene and corresponds to an amino acid substitution Lys (135) in 1816D for Ser (135) in 1815D.

The findings may be useful for targeted genetic engineering of the biocatalysts for biotechnological application.

Abbreviations: AD, androst-4-ene-3,17-dione; ADD, androsta-1,4-diene-3,17-dione; 9-OH-AD, 9α-hydroxy androst-4-ene-3,17-dione; ORFs, open reading frames; SNP, single nucleotide substitution; CTAB, cetyl trimethyl ammonium bromide; bp, base pair; BWA, Broadband-Wheeler Aligner; FAD, flavin adenine dinucleotide; PWM, position weight matrix; ChoX, cholesterol oxidase; Δ^1 KSTD, 3-ketosteroid-Δ^1-dehydrogenase; KSH, 3-ketosteroid-9α-hydroxylase; KshA, subunit A of 3-ketosteroid-9α-hydroxylase; KshB, subunit B of 3-ketosteroid-9α-hydroxylases; HSD, 3β-hydroxysteroid-dehydrogenase; HsaAB, 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione monooxygenase; HsaC, 2,3-dehydroxyphenyl dioxygenase; HsaD, 4,5:9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oate hydrolase; FadA5, acetyl-CoA acetyltransferase; FadD17, acyl-CoA synthetase; FadD19, acyl-CoA synthetase; FadE26, acyl-CoA dehydrogenase; FadE27, acyl-CoA dehydrogenase; FadE28, acyl-CoA dehydrogenase; Hsd4A, hydroxysteroid dehydrogenase; Hsd4B, 2-enoyl acyl-CoA hydratase; EchA19, enoyl-CoA hydratase; TesB, acyl-CoA thioesterase II; ChoX(D,E), cholesterol oxidase; HsaA, 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione monooxygenase subunit; HsaB, 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione monooxygenase subunit; HsaC, 3,4-dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione 4,5-dioxygenase; HsaD, 4,5:9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oate hydrolase; HsaE, 2-hydroxypenta-2,4-dienoate hydratase; HsaF, 4-hydroxy-2-oxovalerate aldolase; HsaG, acetaldehyde dehydrogenase; Ltp2, lipid-transfer protein 2; Ltp3, lipid-transfer protein 3 (acetyl-CoA acetyltransferase); Ltp4, lipid transfer protein 4 (keto acyl-CoA thiolase); YrbE4A, integral membrane protein; YrbE4B, integral membrane protein; KstR, HTH-type transcriptional repressor; KstR2, HTH-type transcriptional repressor.

* Corresponding author at: Center of Innovations and Technologies "Biological Active Compounds and Their Applications", Russian Academy of Sciences, Gubkina-3-2, Moscow 119991, Russian Federation. Tel.: +7 499 1325694; mobile: +7 906 0987231.

E-mail addresses: bragory@yandex.ru (E.Yu. Bragin), vtosha@yandex.ru (V.Yu. Shtratnikova), anagoge@ibpm.pushchino.ru (D.V. Dovbnya), shelkmike@gmail.com (M.I. Schelkunov), yurapekov@gmail.com (Yu.A. Pekov), _sophie@rambler.ru (S.G. Malakho), ovycherova@mail.ru (O.V. Egorova), ivashina@ibpm.pushchino.ru (T.V. Ivashina), sls@ibpm.pushchino.ru (S.L. Sokolov), ashapkin@genebee.msu.ru (V.V. Ashapkin), donova@ibpm.pushchino.ru (M.V. Donova).

0960-0760/$ – see front matter © 2013 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.jsbmb.2013.02.016

1. Introduction

Natural sterols are, typically, steroid 3β-alcohols with a 5(6)-double bond and aliphatic side chain attached to C17 (Fig. 1, **I**). Being important components of the cell membranes, sterols play a significant role in membrane fluidity, cell differentiation and proliferation [1,2]. Cholesterol is known as animal sterol; sitosterol, stigmasterol, campesterol, brassicasterol are abundant plant sterols; the so called phytosterols are mixtures of plant sterols; ergosterol is a major sterol of yeasts and fungi.

Sterols are considered to be natural substrates for bacteria which utilize them as carbon and energy sources. Recently, the critical role of cholesterol catabolism by *Mycobacterium tuberculosis* was confirmed for pathogenicity and progression of tuberculosis infection [3,4]. Cholesterol-rich domains are also targeted by strictly intracellular bacterial pathogens to penetrate the intracellular compartment [5,6].

On the other hand, natural sterols (mainly, phytosterol of the soya origin) represent available and low-cost raw material for the pharmaceutical industry. The fast-growing soil mycobacteria were reported to carry out degradation of the aliphatic side chain, oxidation of the 3β-hydroxy function and isomerization of the 5(6)-double bond of sterols to form 3-keto-4-ene androstanes which are the key intermediates for the synthesis of pharmaceutical steroids. The most marketed intermediates are androst-4-ene-3,17-dione (AD, Fig. 1, **III**), androsta-1,4-diene-3,17-dione (ADD, Fig. 1, **IV**) and 9α-hydroxy androst-4-ene-3,17-dione (9-OH-AD, Fig. 1, **VII**). These steroids can be produced in a single biotechnological step as the major products from natural sterols (cholesterol, phytosterol, ergosterol) using sterol-transforming actinobacteria, mostly, soil fast-growing mycobacteria [7–12].

The putative metabolic pathway of sterol oxidation by actinobacteria was proposed based on the identification of intermediates [8,11]. The accumulated information allows to present sterol catabolic route by actinobacteria as shown in Fig. 1. In general, sterols oxidation appears to be highly conserved in these microorganisms and involves three major processes: sterol uptake, elimination of aliphatic side chain at C17 and steroid core degradation [7,11,13].

Sterol oxidation is initiated with 3β-hydroxy-5-ene moiety oxidation with cholesterol oxidases (ChoXs, e.g. ChoD and the analogs) and 3β-hydroxysteroid-dehydrogenases (HSDs) to form cholest-4-ene-3-one (from cholesterol) or the corresponding stenones from other sterols, e.g. sitost-4-ene-3-one from sitosterol (Fig. 1, **II**) [13,14]. In parallel, the alkyl side chain is degraded in accordance with fatty acid β-oxidation route [8]. Numerous genes coding for such enzymes are known [3,15–17]. Steroid core is attacked with 3-ketosteroid-9α-hydroxylase (KSH) and/or 3-ketosteroid-Δ1-dehydrogenase (Δ^1 KSTD). Several ring-degradation genes, *hsaAB* (3-HSA hydroxylase), *hsaC* (2,3-dihydroxyphenyl dioxygenase), *hsaD* (4,9-DSHA-hydrolase) were shown to be required for the growth on cholesterol as a sole carbon source [18].

It is important to note that stenones (Fig. 1, **II**), or any of the 3-keto-intermediates formed at the oxidative degradation of side chain at C-17 can be attacked with KSH and/or Δ^1 KSTD thus forming corresponding 1-dehydro- or 9α-hydroxy-derivatives. The side chain oxidation of these derivatives is similar to that of cholestenone (Fig. 1, **II**) [11,19].

The action of both KSH and Δ^1 KSTD results in the unstable 1,4-dien-9α-hydroxy steroids which are destroyed on a 9(10)-double bond in ring B thus initiating further full degradation of steroid core. In Fig. 1, the forming of unstable 9α-hydroxy-androst-1,4-diene-3,17-dione (Fig. 1, **VIII**) is presented followed by non-enzymatic 9(10)-secosteroid (Fig. 1, **IX**) formation and its further oxidation, but similar mechanisms of steroid core destruction may take place at the earlier steps for the intermediate steroids with partly oxidized side chain. For instance, in *M. tuberculosis* cholesterol side chain and ring degradation occur simultaneously [20].

When inactivating KSH, ADD can be accumulated as a result of both side chain degradation of 1-dehydrogenated intermediates and 1-dehydrogenation of AD [11]. Correspondingly, 9-OH-AD can be accumulated due to the side chain oxidation of 9α-hydroxylated intermediates and/or 9α-hydroxylation of AD when blocking Δ^1 KSTD [19]. When both KSH and Δ^1 KSTD are blocked, AD is forming as a major product from sterols [8,10].

Gene clusters involved in steroid catabolism in *Rhodococcus* and *Mycobacterium* species were elucidated [3–5,15,20–24]. The genes encoding the enzymes of cholesterol catabolic pathways were studied mostly in *M. smegmatis* mc^2 155 and *M. tuberculosis* [3,25,26]. The key enzymes involved in steroid core oxidation in rhodococci were characterized [e.g. 5,22,27].

However, the information on genetic organization of fast-growing soil mycobacteria capable of effective producing C19 steroids from phytosterol remains scarce. The assignment of β-oxidation genes to this pathway is difficult due to redundant genes encoding β-oxidation in actinomycetes [3].

The strains of *Mycobacterium* spp. VKM Ac-1815D, 1816D, 1817D were shown to produce AD, ADD and 9-OH-AD, respectively, as major product from phytosterol [19,28–34]. Under the optimized conditions, almost full phytosterol conversion by the strains was observed with less than 3% of unreacted substrate. The molar yield of AD by *Mycobacterium* sp. VKM Ac-1815D reached 68–72%, while other steroids were accumulated in smaller amounts: ADD – (6–10%), 20-hydroxymethyl pregn-4-ene-3-one (HMP) – (14–16%), 20-hydroxymethyl pregna-1,4-diene-3-one (HMPD) – in minorities. *Mycobacterium* sp. VKM Ac-1816D converted phytosterol mainly to ADD (70–72%) with AD (2–4%), HMPD (14–16%), and HMP in minorities. Major product by *Mycobacterium* sp. VKM Ac-1817D was 9-OH-AD (68–70%) with a mixture of 9α-hydroxylated C_{22}–C_{24}-steroids in smaller amounts (around 20% totally).

In this paper we describe a comparative analysis of the sterol catabolic system genes in these three strains and evaluation of the strain features significant for 3,17-diketosteroid production from phytosterol on the basis of genome data mining. Taking into consideration the key role of KSH and Δ^1 KSTD for the forming AD, ADD and 9-OH-AD from sterols, special attention was paid to genes *kstD*, *kshA* and *kshB* encoding these enzymes.

2. Materials and methods

2.1. Microorganisms

Strains of *Mycobacterium* spp. VKM Ac-1815D (Tax ID 177066), VKM Ac-1816D (Tax ID 183984), VKM Ac-1817D (Tax ID 183985) were obtained from All-Russian Collection of Microorganisms (VKM IBPM RAS). The strains had been originally isolated from soil samples and differed on sterol transforming activity. They were further maintained for more than 30 years and subjected to multiple passages and laboratory selection on the base of sterol transforming activity using conventional methods which did not involve induced mutagenesis or genetic engineering.

2.2. Genomic DNA isolation

Genomic DNA extraction from *Mycobacterium* spp. was carried out as described [35] with the following modifications. The mycobacterial cells were subcultured from agar slants into 65 ml of nutrient medium containing (g/l): $K_2HPO_4{\cdot}3H_2O$ – 0.5, KH_2PO_4 – 0.5, $(NH_4)_2HPO_4$ – 1.5, $MgSO_4{\cdot}7H_2O$ – 0.2, $FeSO_4{\cdot}7H_2O$ – 0.005, $ZnSO_4{\cdot}7H_2O$ – 0.002, glycerol – 10, yeast extract (Difco, USA) – 10 and Tween 80 – 1, on a rotary shaker (180 rpm) at 30 °C for

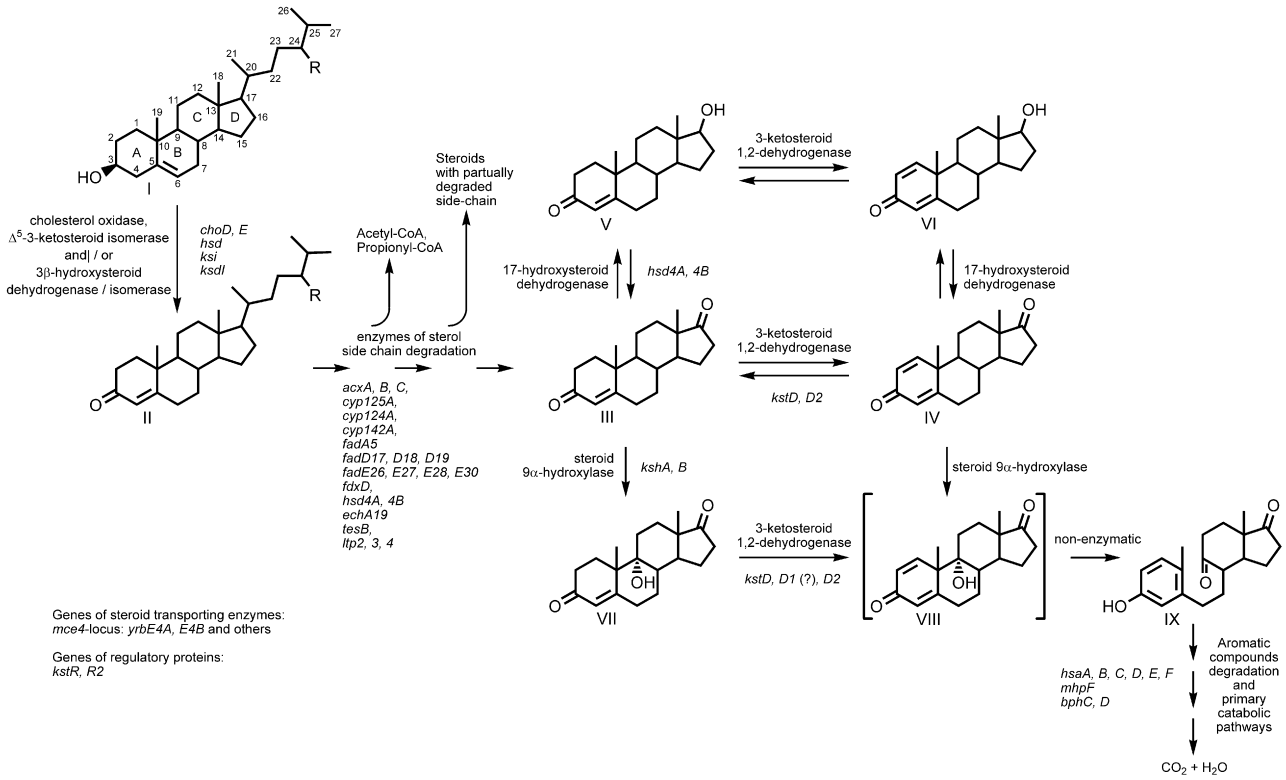

Fig. 1. A proposed pathway of cholesterol catabolism by actinobacteria (adapted from [7,8,11,17,51,52,54]). **I** – R = H, cholesterol, R = C_2H_5, sitosterol; **II** – R = H, cholest-4-ene-3-one; R = C_2H_5, sitost-4-ene-3-one; **III** – androst-4-ene-3,17-dione (AD); **IV** – androsta-1,4-androstadiene-3,17-dione (ADD); **V** – testosterone; **VI** – 1-dehydro-testosterone; **VII** – 9α-hydroxy-androst-4-ene-3,17-dione (9-OH-AD); **VIII** – 9α-hydroxy-androst-1,4-diene-3,17-dione, unstable intermediate; **IX** – 9(10)-secoandrost-1,3,5(10)-triene-9,17-dione. Stenones (**II**) or any of the 3-keto-intermediates formed at the oxidative degradation of side chain at C17 can be attacked with KSHs and/or Δ^1 KSTDs thus forming corresponding 1-dehydro- or 9α-hydroxy-derivatives (not shown) [11,19]. The side chain oxidation of these derivatives is similar to that of stenones (**II**). Side chain and core degradation may occur simultaneously (not shown) [20].

42–48 h. Then, 100 ml of the same medium was inoculated with 10% of the cell suspension obtained and a second-step cultivation was performed for 24 h. The cells from 20 ml broth were harvested by centrifugation at 8000 × *g* for 10 min. Bead-beating disruption was performed with glass beads placed in 50 ml centrifuge tubes containing bacterial cells suspended in 9.5 ml TE buffer (10 mM Tris–HCl, pH 8, 1 mM EDTA) and 1 mg/ml lysozyme. Portions of 0.5 ml 10% SDS and 50 μl proteinase K (20 mg/ml) were added and suspension was incubated for 1 h at 37 °C. Proteins were precipitated with 1.8 ml of 5 M NaCl and 1.5 ml CTAB (0.1 g/ml suspended in 0.7 M NaCl) for 20 min at 65 °C. DNA was purified by sequential extractions of the liquid phase with the chloroform–isoamyl alcohol (24:1, v/v) and phenol–chloroform–isoamyl alcohol (25:24:1, v/v) mixtures and precipitated with 0.6 volume of isopropanol at a room temperature for 15 min. Precipitated DNA was collected by centrifugation (4000 × *g*, 4 °C, 10 min), washed with 70% ethanol, air-dried and re-suspended in TE buffer.

2.3. Sequencing

The samples of genomic DNA were cut with NEBNext dsDNA Fragmentase (New England Biolabs) to fragments of about 400 bp mean length. Libraries of genomic DNA fragments (about 500 bp with adapters) were prepared by the techniques recommended by suppliers of DNA sample preparation kits (Illumina, New England Biolabs).

The sequencing was done on a Genome Analyzer IIx (paired-end 72 nucleotides reads) and on HiSeq 2000 (paired-end 100 nucleotides reads) by the protocols recommended by the manufacturer (Illumina).

2.4. Draft genome assemblies

The short reads obtained in a FASTQ Illumina 1.5+ format were used for the de novo genome assembly with Velvet 1.2 [36]. The contig sequences up to 500 kbp length were obtained.

2.5. Annotation of contigs

An internet-service xBASE (http://www.xbase.ac.uk/annotation/) was used for contigs annotation. The known nucleotide sequences of *M. tuberculosis* H37Rv and *M. smegmatis* mc^2 155 genomes were used as references (Table 1).

2.6. Phylogenetic analysis

The 16S rRNA sequences obtained in this study were compared with GenBank database gene sequences for the relative *Mycobacterium* species. Phylogenetic trees were created in MEGA 5 software [37], using alignment with ClustalW and neighbor-joining algorithm.

2.7. Determination and analysis of genes encoding proteins of steroid catabolism

The confidence levels of different parts of contig sequences were estimated by means of read alignment to contigs with BWA program and Samtools analysis. Only the sequences having maximal confidence levels were further analyzed. The sequences of *M. tuberculosis* H37Rv, *M. smegmatis* mc^2 155, *Rhodococcus erythropolis* PR4, and *R. jostii* RHA1 genomes were used as references for putative steroid catabolism genes identification and

Table 1
Known genes of steroid catabolism.

Name	Function of product	Reference	*R. erythropolis* PR4		*R. jostii* RHA1		*M. tuberculosis* H37Rv		*M. smegmatis* MC^{2}155	
			Locus_tag	GI	Locus_tag	GI	Locus_tag	GI	Locus_tag	GI
Genes encoding enzymes of steroid side-chain oxidation										
acxB	5-Oxoprolinase	[17]			ro10159	111026227				
acxA	5-Oxoprolinase (ATP-hydrolyzing)	[17]			ro10160	111026228				
acxC	Acetone carboxylase gamma subunit	[17]			ro10161	111026229				
	Acetone carboxylase beta subunit	[17]			ro10162	111026230				
fadA5	Acetyl-CoA acetyltransferase (acyl-CoA thiolase)	[56]			ro04678	111021650				
fadA5	Putative acyl-CoA thiolase	[56]			ro05815	111022778	Rv3546	15610682		
fadD17	Acyl-CoA synthetase	[3]	RER_09530	226183557, 226304442	ro04691	111021663, 110821193	Rv3506	81669455, 2924443, 15610642		
fadD18	Acyl-CoA synthetase	[3]					Rv3513c	2924450, 15610649		
fadD19	Acyl-CoA synthetase (steroid-CoA ligase)	[3,15]	RER_09360	226183540, 226304425	ro04689	111021661, 110821191	Rv3515c	81706952, 41352800, 57117119		
fadE26	Acyl-CoA dehydrogenase	[3]	RER_54850	226188089, 226308972	ro04693	111021665, 110821195	Rv3504	15610640, 2924441	MSMEG_5906	118469439, 118468972
fadE30	Acyl-CoA dehydrogenase	[3]	RER_09550	226183559, 226304444	ro04596	111021568, 110821098			MSMEG_0603	118468972, 18170259
	Putative acyl-CoA dehydrogenase		RER_08880	226183492, 226304377						
fadE27	Acyl-CoA dehydrogenase	[3]			ro04692	111021664, 110821194	Rv3505	15610641, 2924442		
fadE28	Short/branched chain acyl-CoA dehydrogenase	[3]			ro04484	4222035	Rv3544c	15610680		
cyp125	Cytochrome P450 125	[4,17]			ro04679	111021651	Rv3545c	15610681		
	Member of cholesterol catabolic gene cluster	[17]			ro04676	111021648				
	Member of cholesterol catabolic gene cluster	[17]			ro04677	111021649				
	Short chain dehydrogenase	[17]			ro04654	111021626				
	Short chain dehydrogenase	[17]			ro04653	111021625				
hsd4A	17β-Hydroxysteroid dehydrogenase	[3]			ro04695	110821197	Rv3502c	2924439		
hsd4B	2-Enoyl acyl-CoA hydratase	[3,17]		158605232	ro04531	111021503, 110821033	Rv3538	57117122, 41352803		
echA19	Enoyl-CoA hydratase	[3,17,21]			ro04688	111021660, 110821190	Rv3516	2924453, 15610652		
tesB	Acyl-CoA thioesterase		RER_33450	226185949	ro06887	11102383	Rv2605c	15609742, 2104308	MSMEG_2938	118468478
ltp2	Branched-chain 3-ketoacyl-CoA thiolase	[3]			ro04488		Rv3540c	15610676		
ltp3	SCPx related 3-ketoacyl-CoA thiolase	[56]			ro04683	111021655	Rv3523	15610659, 1666137		
ltp4	3-Ketoacyl-CoA thiolase	[56]			ro04684	111021656	Rv3522	2924459, 15610658		
fdxD	Ferredoxin	[17]			ro10303	111026367	Rv3503c	15610639		
	Ferredoxin				ro00424	111017446				
	Ferredoxin				ro00378	111017400				

Table 1 (*Continued*)

Name	Function of product	Reference	*R. erythropolis* PR4		*R. jostii* RHA1		*M. tuberculosis* H37Rv		*M. smegmatis* MC^2155	
			Locus_tag	GI	Locus_tag	GI	Locus_tag	GI	Locus_tag	GI
Genes encoding enzymes of steroid core destruction										
kstD1	3-Ketosteroid Δ^1-dehydrogenase	[3]	RER_51050	226187709	ro04532	111021504	Rv3537	1666123, 81345744		
kstD2	3-Ketosteroid Δ^1-dehydrogenase	[3]	RER_12410	226183845						
kstD3	3-Keto-5α-steroid Δ^1-dehydrogenase	[3]	RER_07370	226183341						
ksdD	3-Ketosteroid Δ^1-dehydrogenase	[3]	RER_33600	226185964						
hsd	3β-Hydroxysteroid dehydrogenase/Δ^5-Δ^4-isomerase	[57]					Rv1106c	81669375, 15608246, 2896743		
hsdMS	3β-Hydroxysteroid dehydrogenase/Δ^5-Δ^4-isomerase	[25]							MSMEG_5228	118469383
choD	Cholesterol oxidase	[3]			ro06201	111023164, 110822694	Rv3409c	1449378, 81672536		
	Putative cholesterol oxidase	[3]	RER_04330	7714923	ro04305	111021282, 110820812				
choE	Cholesterol oxidase	[58]	RER_10120	226183616, 226304501						
kshA	3-Ketosteroid 9α-hydroxylase oxygenase subunit	[3]	RER_51130	226187717, 226308600	ro09003	111025785, 110824764	Rv3526	81556699, 15610662, 1666134	MSMEG_5925	322518627
kshA	3-Ketosteroid 9α-hydroxylase oxygenase subunit	[3]	RER_13800	226183984, 226304869	ro04538	111021510, 110821040			MSMEG_2870	118468408, 118169695
kshA2	3-Ketosteroid 9α-hydroxylase oxygenase subunit	[46,47]	RER_07540	226183358, 226304243	ro02490	111019481, 110819011				
	Putative 3-ketosteroid 9α-hydroxylase subunit	[46,47]	RER_09150	226183519, 226304404	ro05811	111022774, 110822304				
kshB	3-Ketosteroid 9α-hydroxylase reductase subunit	[3]	RER_17750	226184379, 226305264	ro09002	111025784, 110824763	Rv3571	81671918	MSMEG_6039	322518314
kshB	3-Ketosteroid 9α-hydroxylase reductase subunit	[3]			ro05833	111022796, 110822326			MT3676	81671918
hsaA	Flavin-dependent monooxygenase, oxygenase subunit	[3]			ro04539	123340107, 111021511, 110821041	Rv3570c	81671917, 1877301, 15610706	MT3675	81671917
hsaB	Flavin-dependent monooxygenase, reductase subunit	[3]			ro04542	123340106, 111021514, 110821044	Rv3567c	81671915, 1877298, 15610703	MT3672	81671915
hsaC, bphC	3,4-DHSA dioxygenase (2,3-dihydroxybiphenyl 1,2-dioxygenase)	[3]	RER_51660	226187770	ro04541	111021513, 110821043, 75532975	Rv3568c	81818382, 1877299	MT3673	81818382
hsaC homolog					ro02488	111019479, 110819009			MSMEG_2891	118468892, 118170179
hsaC homolog					ro05803	111022766, 110822296			MSMEG_6036	118471712, 118172999
hsaC homolog					ro09005	111025787, 110824766				
bphC1					ro08055	510289, 111024840, 110823819				
hsaD, bphD	HOPDA hydrolase	[3]	RER_51680	226187772	ro04540	75467942, 111021512, 110821042	Rv3569c	81671916, 15610705, 1877300	MSMEG_6037	118472122, 118173409
bphD1					ro10136	35764415, 111026204, 110825047			MT3674	1671916

Table 1 (*Continued*)

Name	Function of product	Reference	*R. erythropolis* PR4		*R. jostii* RHA1		*M. tuberculosis* H37Rv		*M. smegmatis* MC²155	
			Locus_tag	GI	Locus_tag	GI	Locus_tag	GI	Locus_tag	GI
hsaE	2-Hydroxypentadienoate hydratase	[3]		158605229	ro04533	111021505, 110821035	Rv3536c	15610672, 1666124		
hsaF	4-Hydroxy-2-ketovalerate aldolase	[3]			ro04535	111021507, 110821037, 122955482	Rv3534c	81345745, 15610670, 1666126		
hsaG, mhpF	Acetaldehyde dehydrogenase	[3]		158605228	ro04534	111021506, 110821036, 123144620	Rv3535c	81670058, 15610671, 1666125		
ksi, ksdI	Putative Δ^5-3-ketosteroid isomerase		RER_13740	226183978, 226304863						
Genes encoding enzymes of steroid transport										
yrbE4a	Conserved integral membrane protein	[3]	RER_17230	226184327, 226305212			Rv3501c	328864220		
yrbE4b	Conserved integral membrane protein	[3]	RER_17240	226184328, 226305213			Rv3500c	328864232		
Genes encoding regulatory proteins related to steroid catabolism										
kstR	Transcriptional regulator	[39]	RER_51040	226187708, 226308591, 6606082	ro04482	123340118	Rv3574	81671921, 15610710, 1877305	MSMEG_6042	322518316
kstR2	HTH-type transcriptional repressor	[39]			ro04598	123045895	Rv3557c	81671909, 15610693, 1877288	MSMEG_6009	322518315

annotation which were performed using NCBI Genome Workbench (http://www.ncbi.nlm.nih.gov/projects/gbench) and BLAST 2.2.25+ [38]. The genes found appear to be determinative in respect to biotechnological value of the strains under study (Table 1).

In order to identify the putative genes for KSHA, KSHB and Δ^1 KSTD we have compared candidate sequences with the corresponding reference sets. The sets were constructed based on the known actinobacterial KSHA, KSHB and Δ^1 KSTD protein sequences taken from the NCBI database. Comparisons were performed using Hidden Markov Models with HMMER2 plugin for UGENE (http://ugene.unipro.ru/). Bit score was used as a measure of similarity of a sequence being analyzed to a reference set. The candidate protein was considered as a member of the protein family if its similarity score to a corresponding reference was no less than $M - 2\sigma$, where M is a mean similarity score of reference sequences to a reference set for each protein family and σ is a corresponding standard deviation.

2.8. Identification of regulatory protein binding sites

Respective sequences were checked for a presence of regulatory protein binding sites with the software package UGENE 1.9.0. Positional weight matrices (PWM) were calculated from the known binding sites of transcription factors KstR [21] and KstR2 [39] and used for a search of similar sites in sequence 500 bp length upstream of operons found with the program FgenesB (http://linux1.softberry.com/berry.phtml?topic=fgenesb&group=programs&subgroup=gfindb). Quality parameter (percentage identity) was calculated for each site found. Percentage identity was used as a measure of similarity between sequences under analysis and PWM. Only the sites with percentage identity higher than 85% were further analyzed (default value for UGENE).

2.9. Detection of single nucleotide substitution

In order to detect the single-nucleotide polymorphisms (SNPs), genomes of the closely related strains 1815D and 1816D were mutually aligned with NCBI BLASTn suite. Thus single-nucleotide differences between similar genome sequences were ascertained. In order to eliminate non-significant polymorphisms, the raw short reads obtained were compared to assembled genome sequences with the BWA package. The results were analyzed with Samtools for significance values of each genome nucleotide. Only putative SNPs with maximal significance values were further regarded, about 2% of putative SNPs were discarded. In some cases a single gene sequence was aligned to a number of different homologous genes leading to erroneous detection of multiple SNPs. These cases were detected by an anomalously high density of SNPs and discarded; only solitary SNPs having no neighbor SNPs for at least 100 bp on both sides were further analyzed.

The SNPs revealed were further verified by Sanger sequencing. Respective parts of genomic DNA were amplified by PCR, purified by agarose gel electrophoresis and sequenced on a 3730 DNA Analyzer (Applied Biosystems). Only SNPs confirmed by Sanger sequencing are considered below.

3. Results and discussion

3.1. Draft assemblies of mycobacterial genomes and the annotation of contigs

We used the paired-end reads of 72 and 100 nucleotides to assembly lengthy non-overlapping sequences of genomes (contigs and scaffolds). The sequence coverage was calculated as an average number of times each sequence of genome has been sequenced (the depth of sequencing) and in all cases exceeded 500. This is in

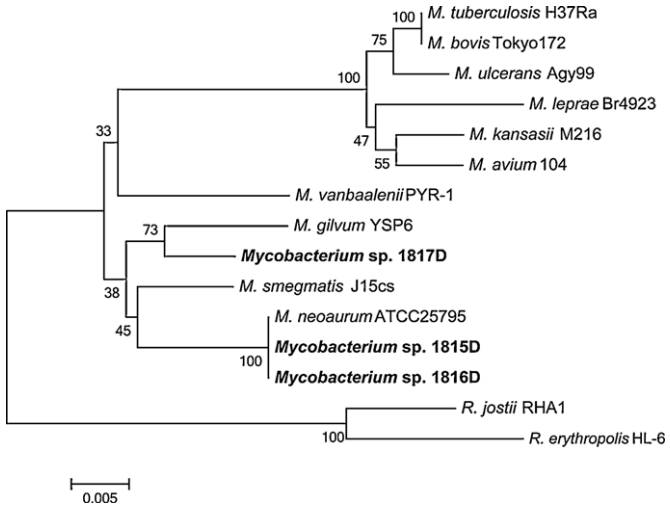

Fig. 2. A phylogenetic dendrogram of 16S rRNA in actinobacteria. Dendrogram of 16S rRNAs from species of *Mycobacterium* and *Rhodococcus*. Bootstrap values, generated from 100 replicates, are shown at the nodes. Scale bar represents nucleotide substitutions per site.

a great excess of coverage values needed for an efficient assembly with most widely used de novo assembly tools [40].

The quality of draft genome assemblies was assessed with N50 statistics, commonly used for such tasks. The N50 of an assembly was calculated as a weighted median of the lengths of its contigs and scaffolds. Thus, the sum of the sequences' lengths of N50 or longer was equal to half the sum of the lengths of all sequences. The results are shown in Table 2. As can be easily seen, the assemblies resulted in N50 of ~343 kbp or larger, thus reconstructing quite long regions of the genomes under study.

The putative genome sizes of *Mycobacterium* spp. VKM Ac-1815D and 1816D, calculated as the sum of the lengths of all contigs, were of 5.4 Mbp, whereas the genome of *Mycobacterium* sp. VKM Ac-1817D appeared to be significantly larger, 6.35 Mbp.

It is known that mycobacterial genomes show considerable size variations. The smallest mycobacterial genomes known belong to the infectious species such as *M. leprae*. Free living mycobacteria, e.g. *M. marinum*, that need a more extensive adaptation potential, usually have larger genome sizes [41].

Expectedly, similar to other actinobacteria, the strains under the study were characterized by GC rich genomes, G + C content of *Mycobacterium* spp. VKM Ac-1815D and 1816D being 66.9%, that of *Mycobacterium* sp. VKM Ac-1817D – 66.2%.

The annotation of genomes by xBASE internet service revealed that the number of genes found depends on the reference genome used (*M. tuberculosis* H37Rv or *M. smegmatis* mc^2 155). Thus, 5287 putative ORFs, including 3230 annotated ones, were found in *Mycobacterium* sp. VKM Ac-1815D when genome sequence of *M. tuberculosis* was used as a reference, whereas these numbers were 5166 and 4136, respectively, when genome sequence of *M. smegmatis* was used as a reference. Similar number of ORF (5179) was found in 1815D strain with an ORF prediction program GeneMarkS [42]. Respective findings for strains 1816D and 1817D are shown in Table 2.

3.2. Phylogenetic relations of the strains

The phylogenetic positions of the strains studied were estimated based on the sequences of their 16S rRNAs (Fig. 2). The strains of *Mycobacterium* spp. VKM Ac-1815D and 1816D were found to be identical to fast growing *M. neoaurum* (with similarity of 100%), whereas *Mycobacterium* sp. VKM Ac-1817D appeared to be more similar to *M. gilvum* and *M. smegmatis* (98.4 and 98.1%, respectively). Recent studies in the field of mycobacterial taxonomy showed that the 16S rRNA sequences exclusively are often not informative enough because of considerable heterogeneity of mycobacterial strains even inside one taxon (e.g. [43]). Therefore, a polyphase approach should be used for further determination of the strains taxonomic position. It could be a subject of special investigation which is not in the frames of this work.

3.3. Genes of sterol catabolism

To find putative genes encoding enzymes of steroid catabolism, we run BLAST search of the translated nucleotide sequences of the contigs to amino acid sequences of known proteins of this type in other actinobacteria (*R. erythropolis*, *R. jostii*, *M. tuberculosis*, *M. smegmatis*). The cutoff limit (*E*-value = 10^{-40}) was chosen so that the chances of false negative (missing of a functional gene) were less than the chances of false positive (erroneous identification of a gene as belonging to the category sought for). Thus the list of genes found should be considered with some restrain. The full lists of these genes (99 in 1815D/1816D and 139 in 1817D) are shown in Supplementary Table 1.

All three strains were found to possess genes encoding the putative enzymes of sterol side chain oxidation (*fadA5*, 5-oxoprolinase (ATP-hydrolyzing), *fadD17*, *fadD19*, *fadE26*, *fadE27*, *fadE28*, short-chain dehydrogenase, 3β-hydroxysteroid dehydrogenase/isomerase, *hsd4A*, *hsd4B*, *echA19*, *tesB*, *ltp2*, *ltp3*, *ltp4*), those of steroid core degradation (*kstD*, *kshA*, *kshB*, *choD*, *hsaA*, *hsaB*, *hsaC*, *hsaD*, *hsaE*, *hsaF*, *hsaG*). The *mce*-locuses which were published to encode steroid transport in actinobacteria [44] were also analyzed including *yrbE4A*, *yrbE4B* genes. Five *mce*-like loci were revealed in 1815D/1816D strains and nine – in 1817D (data not shown).

Some genes found in 1815D/1816D strains seem to be absent in 1817D (*choE*, 5-oxoprolinase, acetone carboxylase beta and gamma subunits, acetyl-CoA acetyltransferase). Nevertheless, total number of the genes encoding enzymes of steroid catabolism in 1817D seems to be significantly higher as compared to 1815D/1816D because most of them are represented by larger families.

3.4. Single nucleotide differences (polymorphisms) between genomes of Mycobacterium spp. VKM Ac-1815D and 1816D

In order to reveal single nucleotide polymorphic sites (SNPs), we have carried out a pairwise alignment of the genomes of the two closely related strains, 1815D and 1816D. The total number of SNPs detected was 13, including 2 located in the intergenic regions, and 11 – in protein coding genes. One of these SNPs was in a putative steroid-catabolism gene *kstD*: a T404 nucleotide of *Mycobacterium* sp. VKM Ac-1815D was substituted with a C nucleotide in *Mycobacterium* sp. VKM Ac-1816D, thus resulting in the replacement of Ser135 with Lys. It is quite possible that this substitution in *kstD* gene accounts for the major differences between the strains. As we have shown previously, *Mycobacterium* sp. VKM Ac-1815D mainly converts sterols (cholesterol, phytosterol, ergosterol) to AD, while a major product of sterols conversion by 1816D is ADD [28,33,34]. Other SNPs are located in the coding sequences of acyl-CoA synthetase *fadD21*, acyl-CoA dehydrogenase *fadE13*, isochorismate synthase, DNA helicase *ErcC3* and some other genes (Supplementary Table 2).

3.5. Genes encoding 3-ketosteroid-9α-hydroxylase and 3-ketosteroid-1-dehydrogenase

Taking into consideration the key role of 3-ketosteroid-9α-hydroxylase (KSHA and KSHB) and 3-ketosteroid-Δ^1-dehydrogenase (Δ^1 KSTD) in steroid core

Table 2
Characteristics of genome assemblies.

Strain	Reads, $\times 10^6$	Putative genome size, kbp	N50, bp	Coverage, fold	GC, %	Number of ORF (XBASE, reference *M. tuberculosis*)[a]		Number of ORF (XBASE, reference *M. smegmatis*)[b]		Number of ORF (GeneMark)
						Total	Annotated	Total	Annotated	
1815D	41	5401	343,484	760	66.9	5287	3230	5166	4136	5179
1816D	30	5401	353,057	555	66.9	5286	3230	5159	4126	5181
1817D	41	6306	386,467	650	66.2	6271	3792	6120	4934	6161

[a] *M. tuberculosis*: genome size – 4,411,532 bp, G + C content – 65.61%, ORFs – 3959 [59].
[b] *M. smegmatis*: genome size – 6,988,209 bp, G + C content – 67.40%, ORFs – 6897 [59].

degradation, we have studied the amino acid sequences of the proteins encoded by putative genes *kshA*, *kshB* and *kstD* in some more details in order to evaluate their probable functional significance. Five genes of putative KSHA proteins were revealed in 1817D, while only two – in 1815D/1816D. Amino acid sequences of all seven putative KSHA proteins were compared to a collection of known KSHA proteins of actinobacteria. All these putative proteins appeared to have high levels of general similarity to the known KSHAs (Table 3). Thus, all genes annotated as *kshA* in 1815D, 1816D and 1817D strains may code for *bona fide* KSHA proteins.

Study on *kshA* in *M. tuberculosis* has revealed one *kshA* homolog in the genome [45]. The multiple *kshA* genes, encoding the oxygenase component of 3-ketosteroid 9α-hydroxylase, have been found in *Rhodococcus erythropolis* SQ1 and *R. rhodochrous* DSM43269 [22,45,47]. Interestingly, only deletion of all *kshA*s resulted in the total blockage of AD degradation. The presence of several *kshA* homologues has been suggested to facilitate adaptation of soil bacteria to unstable environmental conditions [22].

To elucidate evolutionary relationships between different putative KSHA proteins in 1817D and 1815D/1816D we have composed a dendrogram comparing them to known KSHA proteins of different actinobacteria (*Mycobacterium*, *Rhodococcus*, *Gordonia*, *Cupriavidus*, *Aeromicrobium* (*A. marinum*) and *Amycolicicoccus* (*Am. subflavius*)) (Fig. 3).

The protein encoded by gene *kshA_1* of 1815D and 1816D strains (1815D_1) is nearly identical (99.2% of amino acids identity) to a known KSHA protein of *M. neoaurum* (Fig. 3). This is in a good agreement with the results of the 16S rRNA similarity study described above. The sequence of protein encoded by *kshA_1* gene of 1817D strain (1817D_1) is quite similar to that of 1815D_1 (86.3%). These two proteins form a cluster of related sequences together with KSHA proteins of *M. smegmatis*, *M. tuberculosis* and *M. parascrofulaceum*. The protein encoded by the gene *kshA_2* of 1817D strain (1817D_2) is associated with the same cluster, though its similarity to other members of this cluster is lower (66.6% identity to the nearest neighbor, KSHA of *M. parascrofulaceum*). The proteins encoded by the genes *kshA_3* and *kshA_5* of 1817D strain (1817D_3 and 1817D_5) are most similar to KSHA of *M. avium* (83% and 75.4%) and form another cluster together with KSHA proteins of *M. smegmatis*, *M. avium*, *M. intracellulare* and *M. colombiense*. The last two proteins encoded by the gene *kshA_2* of 1815D/1816D (1815D_2) and the gene *kshA_4* of 1817D (1817D_4) form a still another separate cluster. Their mutual similarity is significantly higher than the similarity to a nearest reference protein (80.6 and 60.9%, respectively). Thus, multiple KSHA proteins in the strains 1815D, 1816D and 1817D probably have various evolutionary histories.

The KSH enzymes of *M. tuberculosis*, *R. rhodochrous* and *R. erythropolis* SQ1 are known to be the two-subunit terminal oxygenases, composed of a Rieske-domain containing oxygenase (KSHA) and a class IA ferredoxin reductase (KSHB) [47]. In a detailed study of KSHA structure the conservative Rieske domain has been shown to consist of amino acid residues 24–153, where cysteine residues 67 and 86 are responsible for one Fe^{2+} ion binding, and histidine residues 69 and 89 – for another. The catalytic domain of KSHA (amino acid residues 154–374) contains a number of conservative amino acid residues, namely, Val 176, Asp 178, His 181, His 186, Gln 204, Tyr 232, Met 238, Asn 240, Asn 257, Phe 301, Asp 304 and Trp 308 [48].

We have compared respective parts of proteins encoded by putative *kshA* and *kshB* genes of our strains with amino acid sequences of the above mentioned conservative domains of *M. tuberculosis* proteins KSHA and KSHB. The key amino acid residues in the sequences of Rieske domains were the same in all our putative KSHA proteins.

Some amino acid residues differences in the substrate binding part of the catalytic domain were revealed. Instead of conservative Val 176 there is an Asn in KSHA_2 of 1815D/1816D and KSHA_4 of 1817D and an Ile in KSHA_2 of 1817D, instead of Asn240 – Asp in KSHA_3 and KSHA_5 of 1817D. The putative proteins KSHA_1 of 1815D/1816D and KSHA_1 of 1817D seem to be most close relatives of *M. tuberculosis* KSHA. Both amino acids Val and Ile are nonpolar, thus their exchange in KSHA_2 of 1817D probably does not affect the substrate specificity of the enzyme. Quite conversely, the exchange between a polar uncharged amino acid Asn and in KSHA_3

Table 3
Similarity scores of candidate KSHA, KSHB and KSTD proteins.

Proteins	*M*	2σ	*Mycobacterium* sp. VKM Ac-1817D										*Mycobacterium* spp. VKM Ac-1815D, 1816D			
			1		2		3		4		5		1		2	
			Ss	*E*-v	Ss	*E*-v	Ss	*E*-v	Ss	*E*-v	Ss	*E*-v	Ss	*E*-v	Ss	*E*-v
KSHA	*924.98*	*211.8*	**9710**	0.0	**879.6**	0.0	**932.0**	0.0	**856.0**	0.0	**898.2**	*E*-162	**9605**	0.0	**853.7**	*E*-157
KSHB	*877.5*	*105.4*	**8526**	0.0	**736.7**	*E*-141	319.5	5*E*-066	–	–	–	–	**8122**	*E*-165	121.5	7*E*-042
KSTD	*1032.0*	*288.3*	**11674**	0.0	**890.1**	0.0	**884.8**	0.0	**700.2**	*E*-129	461.4	9*E*-075	**11831**	0.0	–	–

Dimensionless similarity scores (Ss) and *E*-values (*E*-v) of BLAST alignment of candidate proteins to respective reference proteins are presented. The scores meeting the criterion (Ss ≥ M-2σ) are shown in bold.

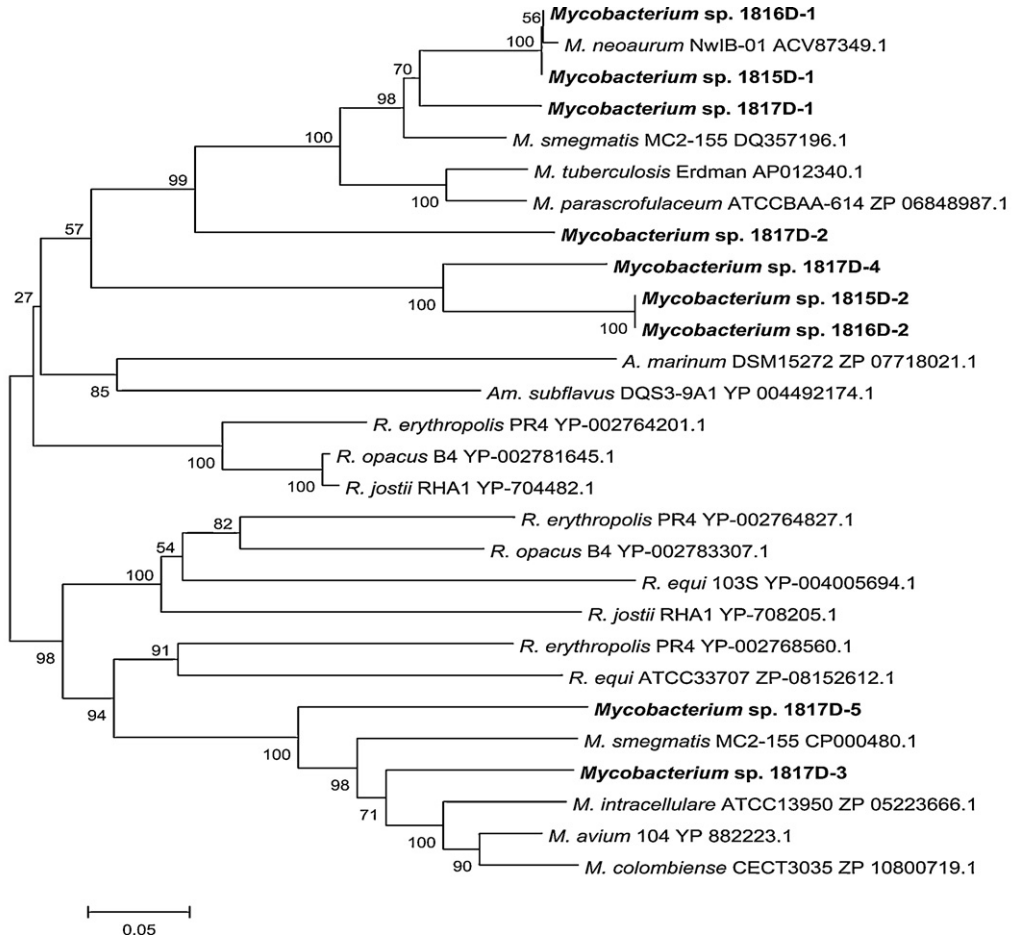

Fig. 3. A phylogenetic dendrogram of 3-ketosteroid 9α-hydroxylase component A (KSHA) in actinobacteria. Phylogenetic tree was constructed on the basis of 3-ketosteroid 9α-hydroxylase component A (KSHA) candidate protein sequences of strains *Mycobacterium* spp. VKM Ac-1815D, 1816D and 1817D (*E*-value < e^{-40}) and known KSHA proteins of *Rhodococcus*, *Mycobacterium*, *Aeromicrobium* (*A. marinum*) and *Amycolicicoccus* (*Am. subflavius*). Multiple KSHA candidate proteins from each of *Mycobacterium* spp. 1815D, 1816D and 1817D strains are numbered. Bootstrap values, generated from 100 replicates, are shown at the nodes. Scale bar represents amino acid substitutions per site.

and a negatively charged one (Asp) in KSHA_5 of 1817D probably is not without effect on the substrate specificity of enzyme.

Based on the BLAST analysis results we have selected three proteins of 1817D and two proteins of 1815D/1816D as most promising candidates for the role of *bona fide* KSHB proteins (Table 3). The evolutionary relationships of different putative proteins to the reference KSHB proteins of actinobacteria are presented on a dendrogram (Fig. 4). KSHB_1 proteins of both 1815D/1816D and 1817D show similarity to KSHB of *M. smegmatis* (77.3 and 85.6%, respectively) and form a separate cluster together with this reference protein. The putative protein KSHB_2 of 1817D is most similar to KSHB proteins of *Rhodococcus* spp.

It is known that there are three domains in KSHB containing a number of conservative amino acid residues, namely, a flavin binding domain RxYSL, a NAD binding domain GGIGITP or AGSGITP, and a 2Fe2S-cluster binding domain $Cx_4Cx_2Cx_{29}C$ [48–50]. As shown in Table 4, all these domains are present in putative KSHB proteins of 1815D, 1816D and 1817D strains, proving them to be *bona fide* KSHB enzymes. The sequence of these domains in proteins encoded by genes *kshB_3* of 1817D and *kshB_2* of 1815D is least similar to the reference.

In general, the results are in good agreement with high 9α-hydroxylase activity expressed by 1817D strain which provide the accumulation of 9-OH-AD as a major product from phytosterol [32]. Probably, the presence of five KSHAs in 1817D, their diversity and the specific features revealed are of importance for steroid 9α-hydroxylation by the strain.

1(2)-Dehydrogenation, involving 3-ketosteroid-1-dehydrogenase (Δ^1 KSTD) is another key reaction of steroid core degradation (Fig. 1). As known, Δ^1 KSTD is a FAD-containing protein with a N-terminal conservative FAD-dependent domain [50]. As found by a BLAST analysis, five putative Δ^1 KSTD proteins in 1817D and one in 1815D/1816D are similar to the reference Δ^1 KSTD proteins. A dendrogram analysis of putative Δ^1 KSTD proteins shows that those encoded by *kstD_1* of 1815D/1816D and *kstD_1* of 1817D are quite similar to each other and to Δ^1 KSTD of *M. smegmatis* (Fig. 5). The sequences of Δ^1 KSTD_2 and Δ^1 KSTD_3 of 1817D are less similar. Thus, different types of Δ^1 KSTDs seem to exist in 1817D.

The similarity of amino acid sequences encoded in genes *kstD_4* and *kstD_5* of 1817D to the known Δ^1 KSTD proteins appears to be beyond the cutoff limit (Table 3). Nevertheless, all putative proteins, including those two, contain a FAD-binding domain with the presence of all conservative amino acid residues (Fig. 6) and conservative amino acid residues essential for Δ^1 KSTD activity [52], namely Y362, Y122 and T500 in 1815/1816D and corresponding residues in 1817D KstD proteins. One may conclude that, even if the proteins encoded by *kstD_4* and *kstD_5* genes of 1817D are not Δ^1 KSTDs, they are the enzymes with similar FAD-dependent activity.

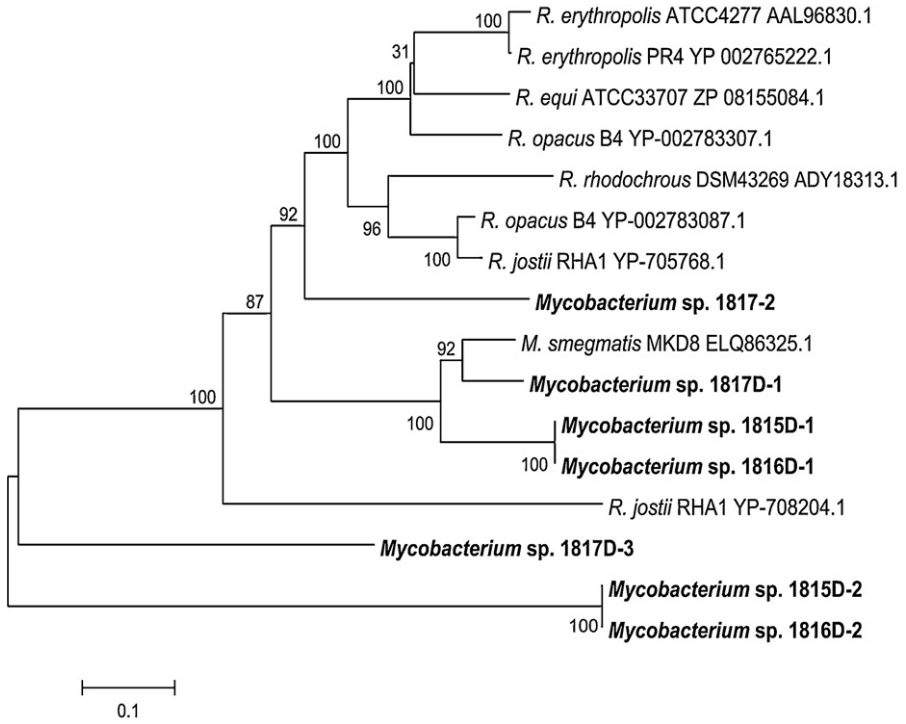

Fig. 4. A phylogenetic dendrogram of 3-ketosteroid 9α-hydroxylase component B (KSHB) in actinobacteria. Phylogenetic tree was constructed on the basis of 3-ketosteroid 9α-hydroxylase component B (KSHB) candidate protein sequences of strains *Mycobacterium* spp. VKM Ac-1815D, 1816D and 1817D (*E*-value < e^{-40}) and known KSHB proteins of *Rhodococcus* and *Mycobacterium* species. Multiple KSHB candidate proteins from each of *Mycobacterium* spp. 1815D, 1816D and 1817D strains are numbered. Bootstrap values, generated from 100 replicates, are shown at the nodes. Scale bar represents amino acid substitutions per site.

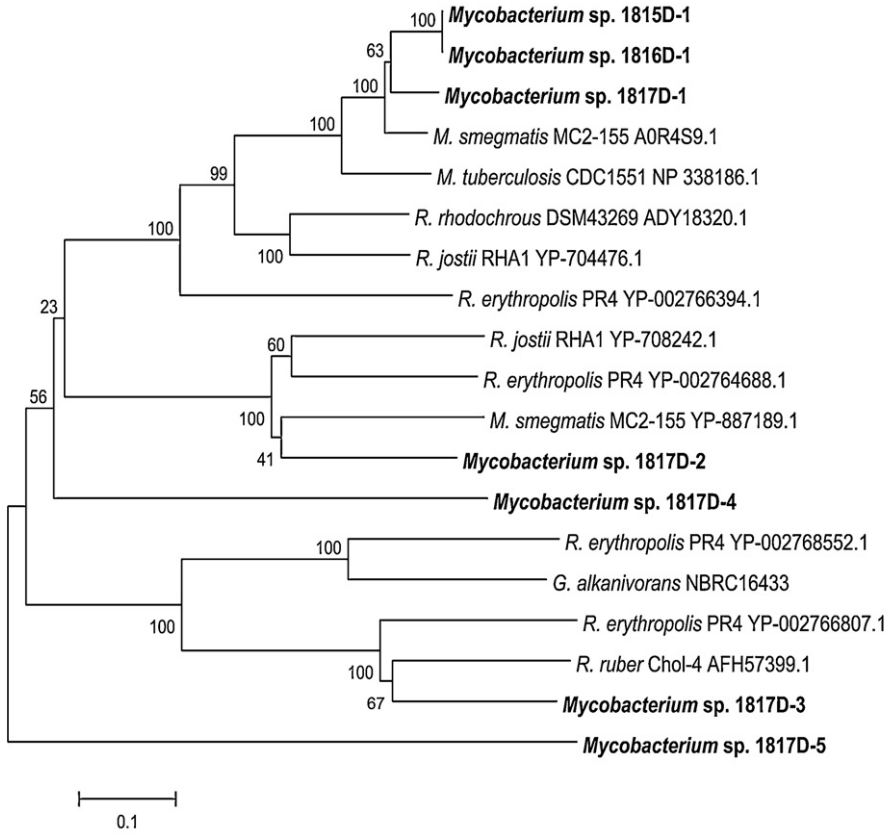

Fig. 5. A phylogenetic dendrogram of 3-ketosteroid-Δ^1-dehydrogenases (Δ^1 KSTD) in actinobacteria. Phylogenetic tree was constructed on 3-ketosteroid-1-dehydrogenase (Δ^1 KSTD) candidate protein sequences of strains *Mycobacterium* spp. VKM Ac-1815D, 1816D and 1817D (*E*-value < e^{-40}) and known Δ^1 KSTDs from *Rhodococcus*, *Mycobacterium* and *Gordonia* (*G. alkanivorans*). Multiple Δ^1 KSTD candidate proteins from *Mycobacterium* sp. 1817D are numbered. Bootstrap values, generated from 100 replicates, are shown at the nodes. Scale bar represents amino acid substitutions per site.

Table 4
Conservative domains of KSHB proteins.

Protein	Flavin binding domain, **RxYSL**	NAD binding domain, **G(A)GIGITP**	2Fe2S-cluster binding domain, **C**x_4**C**x_2**C**x_{29}**C**
M. tuberculosis H37Rv	SVA**RCYSL**CSSP	LLLA**AGSGITP**IMSI	FS**C**REGH**C**GA**C**ACTLRAGKVNMGVNDVLEQQDLDEGLILA**C**QSRPESDSVEVTYDE
1817D_1	SVA**RCYSL**SSSP	LLLA**AGSGITP**MMAI	FS**C**REGH**C**GA**C**AVLKKSGDVEMKINDVLEPSDLEEGLILG**C**QATPVSDSVEVTYDE
1817D_2	SVA**RCYSL**ASSP	LLWA**AGSGITP**VMSI	YS**C**KEGQ**C**GS**C**AARVVRGDVDMAACDILEPDDLADGVILG**C**QARPVSDDIHIEF
1817D_3	HR**RCYS**MSSSP	VTFA**GGSGITP**VFSL	SS**C**ETGS**C**GT**C**MAQVVEGSARMVNNDALDDDEVAEGWVVT**C**QALPTSRTVRVVYE
1815D_1	SVA**RCYSL**SSSP	LLLA**AGSGITP**MMAI	FS**C**REGH**C**GA**C**AVLMRKGDVEMEINDVLEPSDLDEGLILA**C**QALPTSDSVEVTYDEX
1815D_2	YIR**RTYSI**CTST	VAIA**AGSGITP**VLSA	YA**C**MGGA**C**GT**C**MATLTSGTVTMDQNFALSTEQVRTGHILT**C**QSRPTSATVGVDFDAX

Conservative amino acids are indicated with bold large font.

Table 5
Quality parameters of putative binding sites of *kstR* and *kstR2*.

Strain	Gene/operon	*kstR*	*kstR2*
1815D	*kshA*_1	96.9097	Not found
	*kshA*_2	Not found	Not found
	*kshB*_1	90.9514, 88.893	Not found
	kstD	85.4163	Not found
	kstR	98.2432	Not found
	kstR2	Not found	97.8336
1816D	*kshA*_1	96.9097	Not found
	*kshA*_2	Not found	Not found
	*kshB*_1	90.9514, 88.893	Not found
	kstD	85.4163	Not found
	kstR	98.2432	Not found
	kstR2	Not found	97.8336
1817D	*kshA*_1	96.0313	87.529
	*kshA*_2	89.7714	87.529
	*kshA*_3, *fadD*17-*kstD*_3-*kshA*_4, *kshA*_5	Not found	Not found
	*kshB*_1	88.0146, 93.8194	Not found
	*kshB*_2	Not found	Not found
	*kstD*_1	85.4163	91.0235
	*kstD*_2, *kstD*_4	Not found	Not found
	kstR	99.1216	Not found
	kstR2	Not found	100

The Δ^1 KSTD isoenzymes can play distinct functions in actinobacteria. In *Rhodococcus erythropolis* SQ1 the presence of three Δ^1 KSTDs has been reported [50,51]. Preferred substrates for Δ^1 KSTD1 and Δ^1 KSTD2 enzymes were AD and 9-OH-AD, and the deletion of both *kstD1* and *kstD2* genes resulted in stoichiometric accumulation of 9-OH-AD from AD [51]. The third enzyme, Δ^1 KSTD3 expressed highest Δ^4 – dehydrogenase activity towards A-ring saturated androstane steroids [52]. Recently three different genes coding for Δ^1 KSTDs have been also found in the genome of *Rhodococcus ruber* Chol-4 [52]. Only one of these isoenzymes (Δ^1 KSTD2) is shown to be essential for AD catabolic pathway, while Δ^1 KSTD3 is involved in cholesterol catabolism, and Δ^1 KSTD1 has no effect on steroid oxidation by the strain.

3.6. Transcription factors binding sites

To evaluate the presence of binding sites for transcription repressor proteins KstR, specifically involved in the control of cholesterol utilization, we have analyzed 500 bp sequences of DNA immediately preceding the operons, containing the genes *kshA*, *kshB* and *kstD*. Besides, the data on the presence of such binding sites before their own genes (*kstR* and *kstR2*) probably reflecting existence of negative feedback regulation are also given in Table 5.

No significant differences concerning the presence of KstR-binding sites and their qualities were found between the strains 1815D and 1816D. A putative *kshA* gene which does not contain such binding sites, as well as a gene containing a single strong site, was found in both strains. There are two KstR-binding sites before *kshB* gene, thus indicating a possibility of its stable repression. A single relatively weak site is present before the *kstD* gene. Thus, *kstD* must be actively expressed. The AD/ADD accumulation in these strains is supposed to be a result of strong repression of *kshB*. Indeed, the strains of 1815D and 1816D are capable of effective accumulating of AD and ADD, respectively, from phytosterol, and demonstrated poor growth on AD and ADD as sole source of carbon and energy thus indicating the absence, or lower level of 9α-hydroxylase activity [31,33].

Several *kshA* genes were found in the 1817D strain, with only two of them having KstR-binding sites. Of the two copies of *kshB* in this strain, only one possessed a KstR-binding site. This finding can explain a high level of *kshAB* expression resulting in 9-OH-AD accumulation in this strain [32,53]. On the other hand, only one of several *kstD* genes in this strain has KstR-binding sites. One may suggest that this is a cause of rapid 9-OH-AD conversion into unstable 9-OH-ADD followed by its further degradation as it was

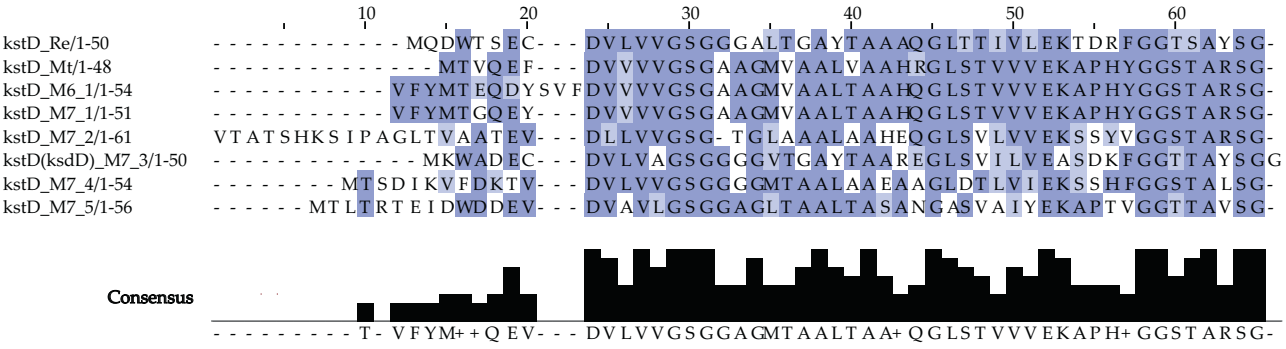

Fig. 6. Alignment of FAD-binding domains of Δ^1 KSTD proteins. Alignment of known and putative Δ^1 KSTD FAD-binding domain amino acid sequences (GenBank accession numbers are in parentheses): Rv3537 from *Mycobacterium tuberculosis* (CAB05041) (*kstD_Mt*); Δ^1 KSTD1 from *Rhodococcus erythropolis* SQ1 (AAF19054) (*kstD_Re*); putative Δ^1 KSTD from 1815D; putative Δ^1 KSTD from 1816D; multiple putative Δ^1 KSTDs from 1817D. Consensus sequence is under alignment. Height of columns in consensus corresponds with number of identical amino acid residues. Alignment was generated using the MUSCLE v3.0 [55].

reported earlier [53]. The cytosolic Δ^1 KSTD enzyme (dimer, Mw 58 kDa) has been isolated. It converts AD to ADD and express activity also towards 9-OH-AD.

4. Conclusions

In this work full genome sequences of three strains of *Mycobacterium* spp. were produced as sets of lengthy contigs and scaffolds (N50 larger than 340 kbp). As a matter of fact, two of the strains studied, namely, *Mycobacterium* spp. VKM Ac-1815D and 1816D, possess identical genomes, and therefore may well be regarded as one and the same strain with few differences, e.g. some single nucleotide variations. It should be pointed out that one of these few SNPs, located in a single-copy *kstD* gene, may be the cause of major practical difference between VKM Ac-1815D and 1816D: production of AD and ADD, respectively, as major products from sterols. On the other hand, the strain of *Mycobacterium* sp. VKM Ac-1817D has a quite different and larger genome, as compared to those of VKM Ac-1815D and 1816D, and obviously represents a different species of mycobacteria.

All three strains studied possess multiple genes encoding enzymes of steroid catabolism, including those of side chain cleavage, steroid core destruction and steroid transportation. It is worth a special note that different members of steroid catabolism gene families in these actinobacteria seem to have various evolutionary histories and probably have various physiological roles.

Different sets of steroid catabolism genes are probably the main reasons explaining variable biocatalytic capabilities of these strains. Therefore these genes appear to be most obvious and promising targets for genetic manipulations purposing creation of novel industrial strains. On the other hand, one must take into consideration not only the genes themselves but also their regulatory sequences. Our data on KstR and KstR2 repressor proteins binding sites as well as the accumulated information on the subject may serve a good illustration to this view.

Accession codes. All sequence data obtained in this work were presented to NCBI. The URL are: Mycobacterium sp. VKM Ac-1815D – http://www.ncbi.nlm.nih.gov/nuccore/ANBI00000000, *Mycobacterium* sp. VKM Ac-1816D – http://www.ncbi.nlm.nih.gov/nuccore/AOHQ00000000, *Mycobacterium* sp. VKM Ac-1817D – http://www.ncbi.nlm.nih.gov/nuccore/AOHR00000000.

Acknowledgements

This work was partially supported by the Russian Foundation for Basic Research (grants no. 11-04-12034 ofi_m and 13-04-01551_a) and "Molecular and Cell Biology" Program of the Presidium RAS.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsbmb.2013.02.016.

References

[1] V. Piironen, D.G. Lindsay, T.A. Miettinen, J. Toivo, A.M. Lampi, Plant sterols: biosynthesis, biological function and their importance to human nutrition, Journal of the Science of Food and Agriculture 80 (2000) 939–966.
[2] P. Fernandes, J.M.S. Cabral, Phytosterols: applications and recovery methods, Bioresource Technology 98 (2007) 2335–2350.
[3] R. van der Geize, K. Yam, T. Heuser, M.H. Wilbrink, H. Hara, M.C. Anderton, E. Sim, L. Dijkhuizen, J.E. Davies, W.W. Mohn, L.D. Eltis, A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages, Proceedings of the National Academy of Sciences of the United States of America 104 (2007) 1947–1952.
[4] J.C. Chang, M.D. Miner, A.K. Pandey, W.P. Gill, N.S. Harik, C.M. Sassetti, D.R. Sherman, *igr* genes and *Mycobacterium tuberculosis* cholesterol metabolism, Journal of Bacteriology 191 (2009) 5232–5239.
[5] R. van der Geize, A.W. Grommen, G.I. Hessels, A.A. Jacobs, L. Dijkhuizen, The steroid catabolic pathway of the intracellular pathogen *Rhodococcus equi* is important for pathogenesis and a target for vaccine development, PLoS Pathogens 7 (2011) e1002181.
[6] P. Goluszko, B. Nowicki, Membrane cholesterol: a crucial molecule affecting interactions of microbial pathogens with mammalian cells, Infection and Immunity 73 (2005) 7791–7796.
[7] M.V. Donova, O.V. Egorova, Microbial steroid transformations: current state and prospects, Applied Microbiology and Biotechnology 94 (2012) 1423–1447.
[8] A. Szentirmai, Microbial physiology of sidechain degradation of sterols, Journal of Industrial Microbiology and Biotechnology 6 (1990) 101–115.
[9] M.V. Donova, Transformation of steroid compounds by actinobacteria, Applied Biochemistry and Microbiology 43 (2007) 1–14.
[10] A. Malaviya, J. Gomes, Androstenedione production by biotransformation of phytosterols, Bioresource Technology 99 (2008) 6725–6737.
[11] J.L. García, I. Uhía, B. Galán, Catabolism and biotechnological applications of cholesterol degrading bacteria, Microbial Biotechnology 5 (2012) 679–699.
[12] W. Ismail, Y.-R. Chiang, Oxic and anoxic metabolism of steroids by bacteria, Journal of Bioremediation & Biodegradation S1-001 (2011) 1–13.
[13] I. Uhía, B. Galán, S.L. Kendall, N.G. Stoker, J.L. García, Cholesterol metabolism in *Mycobacterium smegmatis*, Environmental Microbiology Reports 4 (2012) 168–182.
[14] T. Ivashina, V. Nikolayeva, D. Dovbnya, M. Donova, Cholesterol oxidase ChoD is not a critical enzyme accounting for oxidation of sterols to 3-keto-4-ene steroids in fast-growing *Mycobacterium* sp. VKM Ac-1815D, Journal of Steroid Biochemistry and Molecular Biology 129 (2012) 47–53.
[15] M.H. Wilbrink, M. Petrusma, L. Dijkhuizen, R. van der Geize, FadD19 of *Rhodococcus rhodochrous* DSM43269, a steroid-coenzyme a ligase essential for degradation of C-24 branched sterol side chains, Applied and Environment Microbiology 77 (13) (2011) 4455–4464.
[16] N.M. Nesbitt, X. Yang, P. Fontan, I. Kolesnikova, I. Smith, N.S. Sampson, E. Dubnau, A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstanedione from cholesterol, Infection and Immunity 78 (2010) 275–282.
[17] M.H. Wilbrink, Microbial sterol side chain degradation in Actinobacteria, Dissertation, The University of Groningen, 2011. http://dissertations.ub.rug.nl/faculties/science/2011/m.h.wilbrink/
[18] K.C. Yam, Characterization of HsaC and HsaD, an oxygenase and a hydrolase in the cholesterol catabolic pathway of *Mycobacterium tuberculosis*, Dissertation, The University of British Columbia, 2011. http://hdl.handle.net/2429/37060
[19] M. Donova, S. Gulevskaya, D. Dovbnya, I. Puntus, *Mycobacterium* sp. mutant strain producing 9α-hydroxy androstenedione from sitosterol, Applied Microbiology and Biotechnology 67 (2005) 671–678.
[20] J.K. Capyk, I. Casabon, R. Gruninger, N.C. Strynadka, L.D. Eltis, Activity of 3-ketosteroid 9α-hydroxylase (KshAB) indicates cholesterol side chain and ring degradation occur simultaneously in *Mycobacterium tuberculosis*, Journal of Biological Chemistry 286 (47) (2011) 40717–40724.
[21] S.L. Kendall, M. Withers, C.N. Soffair, N.J. Moreland, S. Gurcha, B. Sidders, R. Frita, A. ten Bokum, G.S. Besra, J.S. Lott, N.G. Stoker, A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*, Molecular Microbiology 65 (2007) 684–699.
[22] M. Petrusma, L. Dijkhuizen, R. van der Geize, *Rhodococcus rhodochrous* DSM 43269 3-ketosteroid-9α-hydroxylase, a two-component iron-sulfur-containing monooxygenase with subtle steroid substrate specificity, Applied and Environment Microbiology 75 (2009) 5300–5307.
[23] K.Z. Rosłoniec, M.H. Wilbrink, J.K. Capyk, W.W. Mohn, M. Ostendorf, R. van der Geize, L. Dijkhuizen, L.D. Eltis, Cytochrome P450 125 (CYP125) catalyzes C26-hydroxylation to initiate sterol side chain degradation in *Rhodococcus jostii* RHA1, Molecular Microbiology 74 (2009) 1031–1043.
[24] S.T. Thomas, B.C. VanderVen, D.R. Sherman, D.G. Russell, N.S. Sampson, Pathway profiling in *Mycobacterium tuberculosis*: elucidation of cholesterol-derived catabolite and enzymes that catalyze its metabolism, Journal of Biological Chemistry 286 (2011) 43668–43678.
[25] I. Uhía, B. Galán, V. Morales, J.L. García, Initial step in the catabolism of cholesterol by *Mycobacterium smegmatis* mc^2 155, Environmental Microbiology 13 (2011) 943–959.
[26] K.C. Yam, I. D'Angelo, R. Kalscheuer, H. Zhu, J.X. Wang, V. Snieckus, L.H. Lu, P.J. Converse, W.R. Jacobs, N. Strynadka, L.D. Eltis, Studies of a ring-cleaving dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of *Mycobacterium tuberculosis*, PLoS Pathogens 5 (2009) e1000344.
[27] K.C. Yam, S. Okamoto, J.N. Roberts, L.D. Eltis, Adventures in *Rhodococcus* – from steroids to explosives, Canadian Journal of Microbiology 57 (2011) 155–168.
[28] M. Donova, D. Dovbnya, A. Kalinichenko, A. Arinbasarova, A. Morozova, L. Vagabova, K. Koshcheyenko, The method of androsta-1,4-diene-3,17-dione production, Patent RU N 2039824 (1995).
[29] M. Donova, D. Dovbnya, A. Kalinichenko, A. Arinbasarova, A. Morozova, L. Vagabova, K. Koshcheyenko, The method of 9α-hydroxy androst-4-ene-3,17-dione production, Patent RU N2077590 (1997).
[30] M. Donova, D. Dovbnya, A. Kalinichenko, A. Arinbasarova, A. Morozova, L. Vagabova, K. Koshcheyenko, The method of androst-4-en-3,17-dione production, Patent RU N20779258 (1997).

[31] O. Egorova, S. Gulevskaya, A. Puntus, A. Filonov, M. Donova, Mutants of *Mycobacterium* sp. producing androstenedione, Journal of Chemical Technology and Biotechnology 77 (2002) 141–147.
[32] M. Donova, D. Dovbnya, G. Sukhodolskaya, S. Khomutov, V. Nikolayeva, I. Kwon, K. Han, Microbial conversion of sterol-containing soybean oil production waste, Journal of Chemical Technology and Biotechnology 80 (2005) 55–60.
[33] O. Egorova, V. Nikolayeva, G. Sukhodolskaya, M. Donova, Transformation of C19-steroids and testosterone production by sterol-transforming strains of *Mycobacterium* spp., Journal of Molecular Catalysis B: Enzymatic 5 (2009) 198–203.
[34] D. Dovbnya, O. Egorova, M. Donova, Microbial side-chain degradation of ergosterol and its 3-substituted derivatives: a new route for obtaining of deltanoids, Steroids 75 (2010) 653–658.
[35] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, NY, 1989.
[36] D.R. Zerbino, E. Birney, Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs, Genome Research 18 (2008) 821–829.
[37] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: Molecular Evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Molecular Biology and Evolution 28 (2011) 2731–2739.
[38] S. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, Journal of Molecular Biology 215 (1990) 403–410.
[39] S. Kendall, P. Burgess, R. Balhana, M. Withers, A. ten Bokum, J.S. Lott, C. Gao, I. Uhia-Castro, N.G. Stoker, Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2, Microbiology 156 (2010) 1362–1371.
[40] Y. Lin, J. LI, H. Shen, L. Zhang, C.J. Papasian, H.W. Deng, Comparative studies of *de novo* assembly tools for next-generation sequencing technologies, Bioinformatics 27 (2011) 2031–2037.
[41] T.M. Wassenaar, J. Bohlin, T.T. Binnewies, D.W. Ussery, Genome comparison of bacterial pathogens, Genome Dynamics 6 (2009) 1–20.
[42] J. Besemer, A. Lomsadze, M. Borovsky, GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions, Nucleic Acids Research 29 (2001) 2607–2618.
[43] M.F. Schinsky, R.E. Morey, A.G. Steigerwalt, M.P. Douglas, R.W. Wilson, M.M. Floyd, W.R. Butler, M.I. Daneshvar, B.A. Brown-Elliott, R.J. Wallace, M.M. McNeil, D.J. Brenner, J.M. Brown, Taxonomic variation in the *Mycobacterium fortuitum* third biovariant complex: description of *Mycobacterium boenickei* sp. *nov*., *Mycobacterium houstonense* sp. *nov*., *Mycobacterium neworleansense* sp. *nov*. and *Mycobacterium brisbanense* sp. *nov*. and recognition of *Mycobacterium porcinum* from human clinical isolates, International Journal of Systematic and Evolutionary Microbiology 54 (2004) 1653–1667.
[44] W.W. Mohn, R. van der Geize, G.R. Stewart, S. Okamoto, J.L. Dijkhuizen, L.D. Eltis, The actinobacterial *mce4* locus encodes a steroid transporter, Journal of Biological Chemistry 283 (2008) 35368–35374.
[45] Y. Hu, R. van der Geize, G.S. Besra, S.S. Gurcha, A. Liu, M. Rohde, M. Singh, A. Coates, 3-Ketosteroid 9α-hydroxylase is an essential factor in the pathogenesis of *Mycobacterium tuberculosis*, Molecular Microbiology 75 (2010) 107–121.
[46] R. van der Geize, G.I. Hessels, M. Nienhuis-Kuiper, L. Dijkhuizen, Characterization of a second *Rhodococcus erythropolis* SQ1 3-ketosteroid 9α-hydroxylase activity comprising a terminal oxygenase homologue, KshA2, active with oxygenase-reductase component KshB, Applied and Environment Microbiology 74 (2008) 7197–7203.
[47] M. Petrusma, G. Hessels, L. Dijkhuizen, R. van der Geize, Multiplicity of 3-ketosteroid-9α-hydroxylase enzymes in *Rhodococcus rhodochrous* DSM43269 for specific degradation of different classes of steroids, Journal of Bacteriology 193 (15) (2011) 3931–3940.
[48] J.K. Capyk, I. D'Angelo, N.C. Strynadka, L.D. Eltis, Characterization of 3-ketosteroid 9α-hydroxylase, a Rieske oxygenase in the cholesterol degradation pathway of *Mycobacterium tuberculosis*, Journal of Biological Chemistry 284 (2009) 9937–9946.
[49] R. van der Geize, G.I. Hessels, R. van Gerwen, R. van der Meijden, L. Dijkhuizen, Molecular and functional characterization of *kshA* and *kshB*, encoding two components of 3-ketosteroid-9α-hydroxylase, a class IA monooxygenase, in *Rhodococcus erythropolis* strain SQ1, Molecular Microbiology 45 (2002) 1007–1018.
[50] R. van der Geize, G.I. Hessels, L. Dijkhuizen, Molecular and functional characterization of the *kstD2* gene of *Rhodococcus erythropolis* SQ1 encoding a second 3-ketosteroid delta(1)-dehydrogenase isoenzyme, Microbiology 148 (2002) 3285–3292.
[51] J. Knol, K. Bodewits, G.I. Hesseles, L. Dijkhuizen, R. van der Geize, 3-Keto-5a-steroid D^1-dehydrogenase from *Rhodococcus erythropolis* SQ1 and its orthologue in *Mycobacterium tuberculosis* H37Rv are highly specific enzymes that function in cholesterol catabolism, Biochemical Journal 410 (2008) 339–346.
[52] L. Fernández de las Heras, R. van der Geize, O. Drzyzga, J. Perera, J.M. Navarro Llorens, Molecular characterization of three 3-ketosteroid-Δ^1-dehydrogenase isoenzymes of *Rhodococcus ruber* strain Chol-4, Journal of Steroid Biochemistry and Molecular Biology 132 (3-5) (2012) 271–281.
[53] G. Sukhodolskaya, V. Nikolayeva, S. Khomutov, M. Donova, Steroid-1-dehydrogenase activity of *Mycobacterium* sp. VKM Ac-1817D strain producing 9α-hydroxy androst-4-ene-3,17-dione from sitosterol, Applied Microbiology and Biotechnology 74 (2007) 867–873.
[54] K. Rosłoniec, Steroid transformation by *Rhodococcus* strains and bacterial cytochrome P450 enzymes, Dissertation, University of Groningen, 2010. http://dissertations.ub.rug.nl/faculties/science/2011/k.z.du.plessis.roslo/
[55] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32 (2004) 1792–1797.
[56] M.H. Wilbrink, R. van der Geize, L. Dijkhuizen, Molecular characterization of *ltp3* and *ltp4*, essential for C24-branched chain sterol-side-chain degradation in *Rhodococcus rhodochrous* DSM 43269, Microbiology 158 (2012) 3054–3062.
[57] X. Yang, E. Dubnau, I. Smith, N.S. Sampson, Rv1106c from *Mycobacterium tuberculosis* is a 3β-hydroxysteroid dehydrogenase, Biochemistry 46 (2007) 9058–9067.
[58] J. Navas, B. Gonzalez-Zorn, N. Ladron, P. Garrido, J.A. Vazquez-Boland, Identification and mutagenesis by allelic exchange of *choE*, encoding a cholesterol oxidase from the intracellular pathogen *Rhodococcus equi*, Journal of Bacteriology 183 (16) (2001) 4796–4805.
[59] L. Li, J.P. Bannantine, Q. Zhang, A. Amonsin, B.J. May, D. Alt, N. Banerji, S. Kanjilal, V. Kapur, The complete genome sequence of *Mycobacterium avium* subspecies paratuberculosis, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 12344–12349.