

# Genome-wide bioinformatics analysis of steroid metabolism-associated genes in *Nocardioides simplex* VKM Ac-2033D

Victoria Y. Shtratnikova<sup>1</sup>  · Mikhail I. Schelkunov<sup>2,3</sup> · Victoria V. Fokina<sup>4</sup> · Yury A. Pekov<sup>1</sup> · Tanya Ivashina<sup>4</sup> · Marina V. Donova<sup>4</sup>

Received: 20 October 2015 / Revised: 4 January 2016 / Accepted: 16 January 2016 / Published online: 1 February 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** Actinobacteria comprise diverse groups of bacteria capable of full degradation, or modification of different steroid compounds. Steroid catabolism has been characterized best for the representatives of suborder *Corynebacterineae*, such as *Mycobacteria*, *Rhodococcus* and *Gordonia*, with high content of mycolic acids in the cell envelope, while it is poorly understood for other steroid-transforming actinobacteria, such as representatives of *Nocardioides* genus belonging to suborder *Propionibacterineae*. *Nocardioides simplex* VKM Ac-2033D is an important biotechnological strain which is known for its ability to introduce  $\Delta^1$ -double bond in various 1(2)-saturated 3-ketosteroids, and perform conversion of  $3\beta$ -hydroxy-5-ene steroids to 3-oxo-4-ene steroids, hydrolysis of acetylated steroids,

reduction of carbonyl groups at C-17 and C-20 of androstanes and pregnanes, respectively. The strain is also capable of utilizing cholesterol and phytosterol as carbon and energy sources. In this study, a comprehensive bioinformatics genome-wide screening was carried out to predict genes related to steroid metabolism in this organism, their clustering and possible regulation. The predicted operon structure and number of candidate gene copies paralogs have been estimated. Binding sites of steroid catabolism regulators KstR and KstR2 specified for *N. simplex* VKM Ac-2033D have been calculated de novo. Most of the candidate genes grouped within three main clusters, one of the predicted clusters having no analogs in other actinobacteria studied so far. The results offer a base for further functional studies, expand the understanding of steroid catabolism by actinobacteria, and will contribute to modifying of metabolic pathways in order to generate effective biocatalysts capable of producing valuable bioactive steroids.

Communicated by M. Kupiec.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00294-016-0568-4) contains supplementary material, which is available to authorized users.

✉ Victoria Y. Shtratnikova  
vtosha@yandex.ru

<sup>1</sup> Department of Bioengineering and Bioinformatics, M.V. Lomonosov Moscow State University, Leninskie Gory, h. 1, b. 73, Moscow 119991, Russian Federation

<sup>2</sup> Institute for Information Transmission Problems, Russian Academy of Sciences, Bolshoy Karetny per. 19, b. 1, Moscow 127051, Russian Federation

<sup>3</sup> A.N. Belozersky Research Institute of Physico-Chemical Biology, M.V. Lomonosov Moscow State University, Leninskiye Gory, h. 1, b. 41, Moscow 119991, Russian Federation

<sup>4</sup> G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Prospekt Nauki, 5, Puschino, Moscow 142290, Russian Federation

**Keywords** Steroid catabolism · Genome-wide analysis · *Nocardioides* · Transcriptional repressors KstR and KstR2

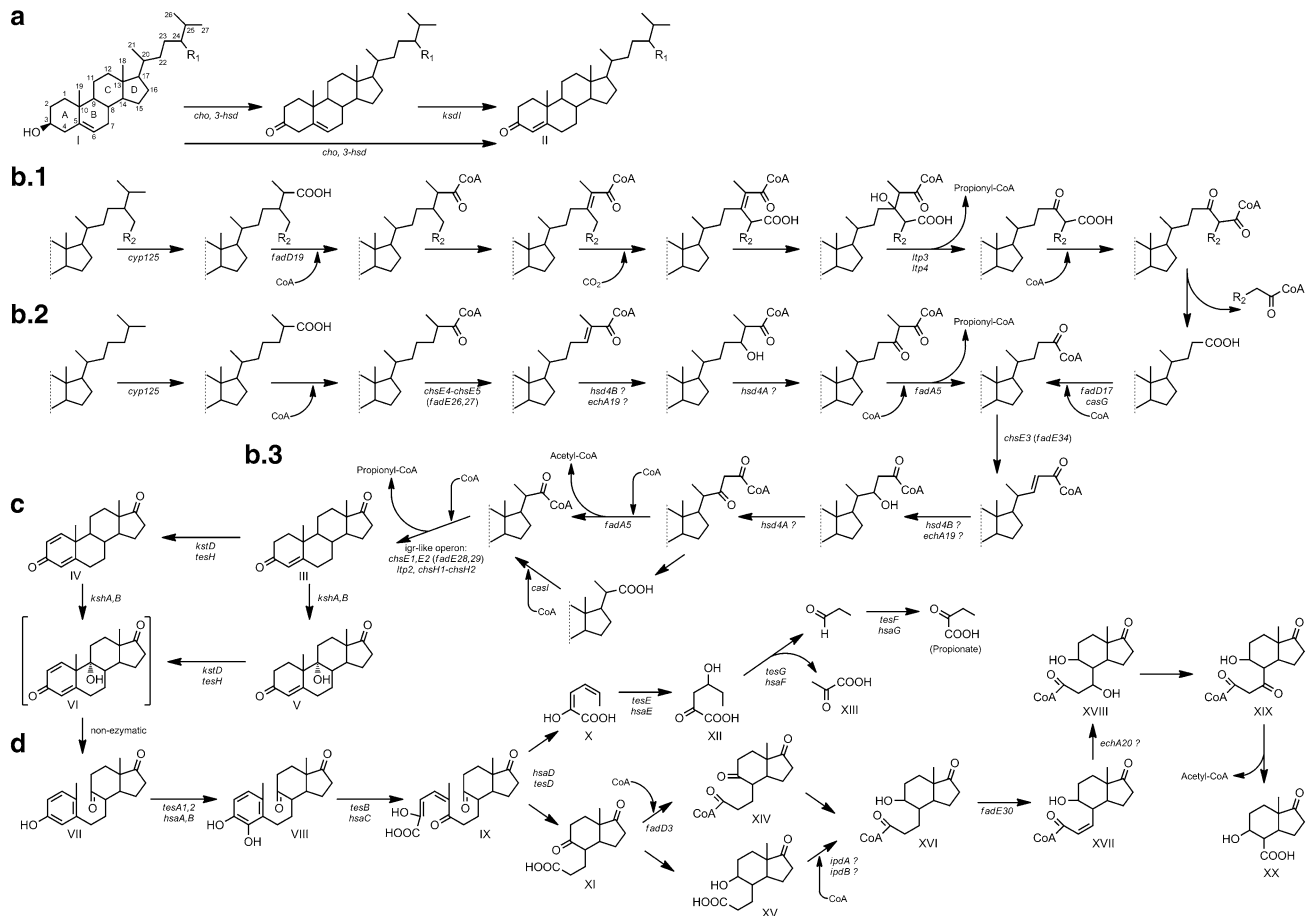
## Introduction

Steroids are biological molecules which fulfill essential physiological functions in higher organisms. These compounds are widespread in nature being synthesized in plants, insects, vertebrates, and lower eukaryotes. Steroid drugs are widely used in different fields of medicine as anti-inflammatory, diuretic, contraceptive, antiandrogenic, progestational, and anticancer agents, male and female sexual hormones, etc. The ability to modify steroids is known for a wide number of microorganisms, and actinobacteria are considered among the most effective biocatalysts.

The phylum *Actinobacteria* is composed of Gram-positive bacteria with an overall high G+C content. These bacteria are known to effectively assimilate hydrophobic compounds and possess diverse enzyme systems with high activity towards different organic compounds, including steroids. Steroid catabolism has been characterized best for the representatives of suborder *Corynebacterineae*, such as *Mycobacteria*, *Rhodococcus*, and *Gordonia*, which characterized by high content of mycolic acids in the cell envelope, the so-called mycolic acid rich organisms (e.g. van

der Geize et al. 2007; Uha et al. 2012; Wilbrink 2011). The known steroid catabolism pathways by these bacteria are presented in Fig. 1 (see “Results and discussion” for detailed explanation). Modification of sterol metabolic pathways by actinobacteria is considered as an effective approach to produce valuable compounds for steroid pharmaceutical industry (Donova and Egorova 2012).

However, the information on genes involved in steroid catabolism in other steroid-transforming actinobacteria, such as representatives of suborder *Propionibacterineae*



**Fig. 1** Biochemical scheme of sterol catabolism. Genes coding respective proteins are denoted. **a** Modification of  $3\beta$ -ol-5-ene to 3-keto-4-ene moiety in A-ring of steroid core; **b** degradation of sterol side chain: **b.1** degradation of the branched side chain to  $C_{24}$ -stenoic acid via aldolic cleavage; **b.2** degradation of unbranched side chain to  $C_{24}$ -stenoic acid via  $\beta$ -oxidation route; **b.3** degradation of  $C_{24}$ -stenoic acid to  $C_{19}$ -steroids; **c** steroid core modifications; **d** steroid core degradation via 9(10)-seco pathway. *I* sterols, *II* stenones.  $R_1 = H$ , cholesterol, cholestenone;  $R_1 = CH_3$ , campesterol, campestenone;  $R_1 = C_2H_5$ , sitosterol, sitostenone;  $R_2 = H$ , campesterol;  $R_2 = CH_3$ , sitosterol; *III* androst-4-ene-3,17-dione (AD), *IV* androsta-1,4-diene-3,17-dione (ADD), *V* 9 $\alpha$ -hydroxy-AD, *VI* unstable 9 $\alpha$ -hydroxy-ADD, *VII* 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione (3-HSA), *VIII* 3,4-dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione (3,4-DHSA), *IX* 4,5-9,10-diseco-3-hydroxy-5,9,17-

trioxoandrosta-1(10),2-diene-4-oic acid (4,9-DSHA), *X*, 2-hydroxyhexa-2,4-dienoic acid (2-HHD), *XI*, 9,17-dioxo-1,2,3,4,10,19-hexanorandrost-5-oic acid (DOHNAA) or 3 $\alpha$ -H-4 $\alpha$ -(3'-propanoic acid)-7 $\alpha\beta$ -methylhexahydro-1,5-indandione (HIP), *XII*, 4-hydroxy-2-oxohexanoic acid, *XIII* pyruvate, *XIV* 9,17-dioxo-1,2,3,4,10,19-hexanorandrost-5-oyl-CoA, *XV* 9-hydroxy-17-oxo-1,2,3,4,10,19-hexanorandrost-5-oic acid or 5-OH-HIP, *XVI* 9-hydroxy-17-oxo-1,2,3,4,10,19-hexanorandrost-6-ene-5-oyl-CoA, *XVII* 7,9-dihydroxy-17-oxo-1,2,3,4,10,19-hexanorandrost-5-oyl-CoA, *XVIII* 9-hydroxy-17-oxo-1,2,3,4,5,6,10,19-octanorandrost-7-oic acid. Adopted from: Fujimoto et al. (1982), van der Geize et al. (2007), Nesbitt et al. (2010), Horinouchi et al. (2012), Wilbrink et al. (2012, 2013), Carere et al. (2013) and Yang et al. (2014, 2015)

belonging to *Nocardioidea* genus is scarce. Meanwhile, there is a growing body of knowledge on the bioconversion of different type steroids by these bacteria evidencing their broad catalytic potential towards steroids (e.g. Yu et al. 2007; Suzuki et al. 2007; Luthra et al. 2015).

*Nocardioidea simplex* VKM Ac-2033D is an important biotechnological strain capable of performing one of the most important reactions in the synthesis of valuable steroid hormones, introduction of  $\Delta^1$ -double bond in steroid core of 3-oxosteroids. Other biotechnologically relevant activities of the strain include hydrolysis of acetylated steroids, reduction of carbonyl groups at C-17 and C-20 of androstanes and pregnanes, respectively, as well as degradation of sterols and androstanes (Fokina et al. 2003a, b, 2010; Fokina and Donova 2003; Sukhodolskaya et al. 2010). Despite the long history of the strain application for steroid bioconversions, little is known about the genes encoding key steps of steroid catabolism.

The strain had been originally isolated from soil and in the early works referred as “*Mycobacterium* sp. 193” (Krassilnikov et al. 1959), “*Mycobacterium globiforme*” (Lestrovaya et al. 1965), “*Arthrobacter globiformis*” (e.g. Koshcheyenko et al. 1983; Arinbasarova et al. 1996). It had been later re-identified (Fokina et al. 2003a) as *Nocardioidea simplex* (O’Donnell et al. 1982; Validation list 1983) based on the 16S rRNA gene sequence analysis and results of comparative studies of chemotaxonomic, morphological and physiological characteristics. The strain had been deposited in All-Russian Collection of Microorganisms at the Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences (VKM IBPM RAS) under the No. VKM Ac-2033D.

Recently, we have announced a complete genome sequencing of *N. simplex* VKM Ac-2033D (Shtratnikova et al. 2015). The genome of over 5.6 Mb with high G+C-content (72.66 %) comprised over 5420 protein-coding sequences distributed across a circular chromosome. In this study, we analyzed genome sequences by means of bioinformatics to predict a suite of genes, operon structures and clusters of the genes involved in sterol/steroid metabolism, and their regulation.

Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional repressors, KstR and KstR2 (Kendall et al. 2007, 2010). No other regulators of steroid catabolic pathways in actinobacteria were reported to date. We calculated de novo binding motifs of KstR-regulators in *N. simplex* VKM Ac-2033D, and searched for binding sites (KstR binding sites) in the whole genome to disclose putative candidates for steroid catabolism-associated genes that differ from the known genes.

The knowledge of its genome organization will contribute to the understanding of steroid bioconversion features, and may also extend the biocatalytic applications of the

strain. Genome investigations including comparative analysis of the steroid catabolism-associated genes for other steroid-transforming representatives of the family *Nocardioidea*, or any closely related actinomycetes were hitherto unreported.

## Materials and methods

### Strain, its cultivation and genome sequencing

The strain of *Nocardioidea simplex* VKM Ac-2033D was cultured as described in (Fokina et al. 2003a). Mineral medium was composed of (g/l):  $\text{KH}_2\text{PO}_4$ —0.5,  $\text{K}_2\text{HPO}_4$ —0.5,  $(\text{NH}_4)_2\text{HPO}_4$ —1.5,  $\text{MgSO}_4$ —0.2,  $\text{FeSO}_4$ —0.005,  $\text{ZnSO}_4$ —0.002.  $\beta$ -Sitosterol (S&D Chemicals, England) and cholesterol (Serva, Germany) were added as fine powders (1 g/l). In some experiments, the medium was supplemented with methyl- $\beta$ -cyclodextrin (MCD), 17 g/l, (CAVASOL® W7 M, Wacker-Chemie, Germany). Cultivation was carried out aerobically on a rotary shaker at 30 °C for 120 h. Steroids were assayed as described in (Ivashina et al. 2012).

For RNA extraction the mineral medium was supplemented with (g/l): yeast extract (Difco, USA)—1.0, glycerol—5.0, MCD—3.46. Cholesterol was added to a final concentration of 0.2 g/l. Cells were harvested when the culture reached  $\text{OD}_{600} = 0.81$ .

The sequencing and assembly procedures were described in (Shtratnikova et al. 2015).

### Genome analysis, annotation and operon identifying

The annotation of the genome was carried out using online service XBASE (<http://www.xbase.ac.uk/annotation/>), GenBank tool PGAP and on-line service RAST (<http://rast.nmpdr.org/>). The whole genome sequences of *Nocardioidea* sp. JS614, *Mycobacterium tuberculosis* H37Rv, *Mycobacterium smegmatis* mc<sup>2</sup>155 were used as references for XBASE. The complete genome sequence in GenBank: CP009896. Operons with found genes were calculated with internet-service FgenesB (<http://linux1.softberry.com/berry.phtml?topic=fgenesb&group=programs&subgroup=gfindb>).

### Analysis of homologous relations and comparison with other *Nocardioidea* strains

Orthologous and paralogous relations between genes of *N. simplex* VKM Ac-2033D and other actinobacteria, as well as different *Nocardioidea* strains were analyzed using OrthoMCL 2.09 (Li et al. 2003). For comparison, the genomes of 19 *Nocardioidea* strains were chosen that were

**Table 1** Actinobacterial strains for orthogroups construction

Strain	GenBank accession number
<i>Nocardioides simplex</i> VKM Ac-2033D	CP009896
<i>Mycobacterium tuberculosis</i> H37Rv	AL123456
<i>Mycobacterium smegmatis</i> str. MC2 155	NC_018289
<i>Mycobacterium neoaurum</i> VKM Ac-1815D	CP006936
<i>Mycobacterium</i> sp. VKM Ac-1817D	CP009914
<i>Rhodococcus erythropolis</i> PR4 and its plasmids	NC_012490; NC_007486, NC_007487, NC_007491
<i>Rhodococcus jostii</i> RHA1 and its plasmids	NC_008268; CP000432, NC_008270, NC_008271
<i>Rhodococcus equi</i> ATCC 33707	NZ_CM001149
<i>Rhodococcus opacus</i> B4 and its plasmids	NC_012522; NC_012523; NC_006969; NC_006970; NC_012520; NC_012521
<i>Nocardia farcinica</i> IFM 10152 and its plasmids	NC_006361; NC_006362; NC_006363
<i>Gordonia polyisoprenivorans</i> VH2 and its plasmid	NC_016906; NC_016907
<i>Gordonia neofelifaecis</i> NRRL B-59395	NZ_AEUD00000000 (genomic scaffolds)

well assembled, i.e. a genome was assembled in less than 100 contigs. Proteins of the selected strains (Table 1 for actinobacteria, Table S1 for *Nocardioides*) were clustered into orthologous groups (orthogroups) with an OrthoMCL inflation parameter 1.5. The list of all orthogroups is presented in Table S2 for actinobacteria, list of orthogroups of steroid catabolism genes for *Nocardioides* genus is presented in Table S3. To find out putative genes encoding enzymes of steroid catabolism out of orthogroups, BLASTx search was run (Altschul et al. 1990) in actinobacteria (Table 1) and betaproteobacterium *Comamonas testosteroni*. Reciprocal BLAST was used in several cases for searching genes which corresponded to known genes one-to-one.

### Analysis of transcription factor binding sites

The regions 500 bp upstream plus 50 bp downstream were analyzed with respect to start codons of genes putatively involved in steroid metabolism. Over-represented motifs in these regions using MEME 4.10 (Bailey and Elkan 1994) were searched. Motifs were allowed to be from 8 to 50 bp long. 20 Top-scoring motifs in a form of position-weight matrices (PWM's) were compared with the respective motifs of mycobacterial steroid metabolism regulators KstR (Kendall et al. 2007) and KstR2 (Kendall et al. 2010). Two motifs that were clearly similar to the motifs of KstR and KstR2 of mycobacteria are further referred to as motifs of KstR and KstR2 in *N. simplex* VKM Ac-2033D.

Then, to seek for all sites of KstR and KstR2 in *N. simplex* VKM Ac-2033D genome the regions 500 bp upstream plus 50 bp downstream were scanned with respect to start codons of all *N. simplex* VKM Ac-2033D genes with KstR and KstR2 motifs found in the previous step. The scan was performed by a tool FIMO from MEME suite 4.10 (Grant et al. 2011). Motifs with false discovery rate lower than 0.01 ( $q$  value, estimated by FIMO using

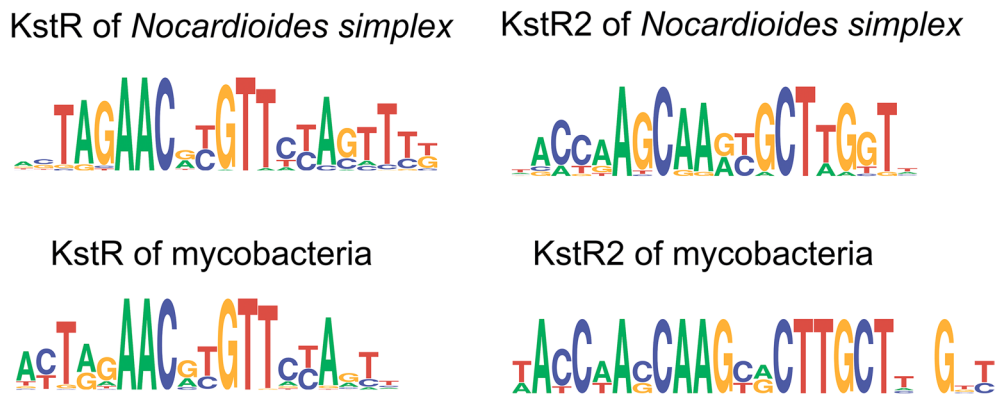
Benjamini-Hochberg technique) were considered as putative binding sites. For comparison, the same regions were also scanned with UGENE 1.13.1 (Okonechnikov et al. 2012) using respective sites of KstR and KstR2 from mycobacteria, considering motifs with score no less than 85 % of maximum possible score. Binding sites which are present in the results of both tools were used for analysis. Determining of the KstRs binding sites was used for operon prediction.

### RNA extraction

Total RNA was extracted from 4 ml of bacterial suspension (CFU  $6.8 \times 10^8 \text{ ml}^{-1}$ ) using RNeasy Mini Kit (Qiagen, USA) according to the manufacturer's protocol following by treatment with RNase free DNase I (Thermo Scientific, Lithuania). Total RNA was quantified at 260/280 nm using the NanoPhotometer<sup>TM</sup> P-Class (Implen, Germany) and stored at  $-70^\circ\text{C}$ .

### Real-time quantitative polymerase chain reaction

First-strand cDNA was synthesized using Revert Aid H minus first strand cDNA kit (Thermo Scientific, Lithuania) with 2  $\mu\text{g}$  of total RNA following the manufacturer's instructions. A real-time PCR was performed using a DNA-Technology system (Russia) with qPCRmix-HS SYBR (Evrogen, Russia). The nucleotide sequences of primers used in this study for target genes and reference gene 16S rRNA are listed in Table S4. Each sample was run in duplicate and the experiment was performed in duplicate. Reaction mix of total volume 25  $\mu\text{l}$  contains 2  $\mu\text{l}$  of cDNA, 0.4  $\mu\text{M}$  of each primer and 5  $\mu\text{l}$  of qPCRmix-HS SYBR. The amplification was performed as follows: initial denaturation at  $94^\circ\text{C}$  for 4 min 30 s, followed by 40 cycles of denaturation at  $95^\circ\text{C}$  for 20 s, annealing at  $53^\circ\text{C}$  for 20 s



**Fig. 2** Comparison of KstR and KstR2 binding motifs of *N. simplex* VKM Ac-2033D and mycobacteria. Motifs for mycobacteria are calculated based on sites reported for *M. tuberculosis* and *M. smegmatis*

in (Kendall et al. 2007, 2010). More conservative positions are indicated by higher letters. Diagrams were made by means of Weblogo [(Crooks et al. 2004); weblogo.berkeley.edu]

and extension at 72 °C for 20 s. To confirm that no genomic DNA was present in the RNA samples, cDNA was substituted with 2  $\mu$ l of reaction mixture prepared without the M-MuLV reverse transcriptase. The 16S rRNA gene was used as internal control. The relative gene expression is represented by average crossing point ( $\Delta$ Cp) values. The specificity of the PCR products was verified by electrophoresis on a 1.2 % agarose gel.

## Results and discussion

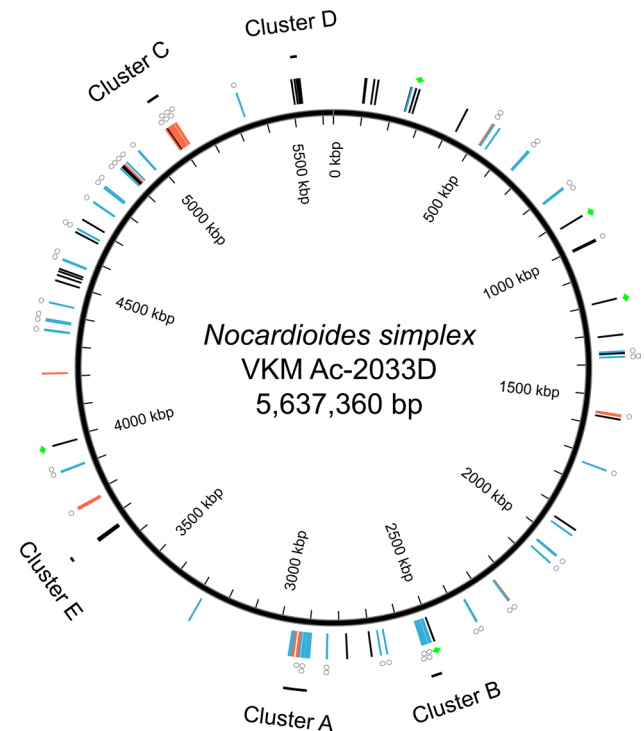
### Determination of binding motifs of KstR and KstR2

Binding motifs of KstR and KstR2, predicted in *N. simplex* VKM Ac-2033D de novo, nearly coincides with those in mycobacteria. Binding preferences of KstRs for *N. simplex* VKM Ac-2033D and for mycobacteria are displayed in Fig. 2. Other overrepresented short sequences near the steroid catabolism-related genes were of low complexity and, therefore, it is unlikely that they correspond to binding sites of some transcription factors.

### General clustering of steroid catabolic gene homologs

During this study, we observed the clustering of predicted operons with orthologs of known genes involved in steroid catabolism. Three large clusters have putative KstRs binding sites: cluster A (KR76\_14160-KR76\_14505), cluster B (KR76\_12190-KR76\_12345), and cluster C (KR76\_25015-KR76\_25200) (Figs. 3, 4).

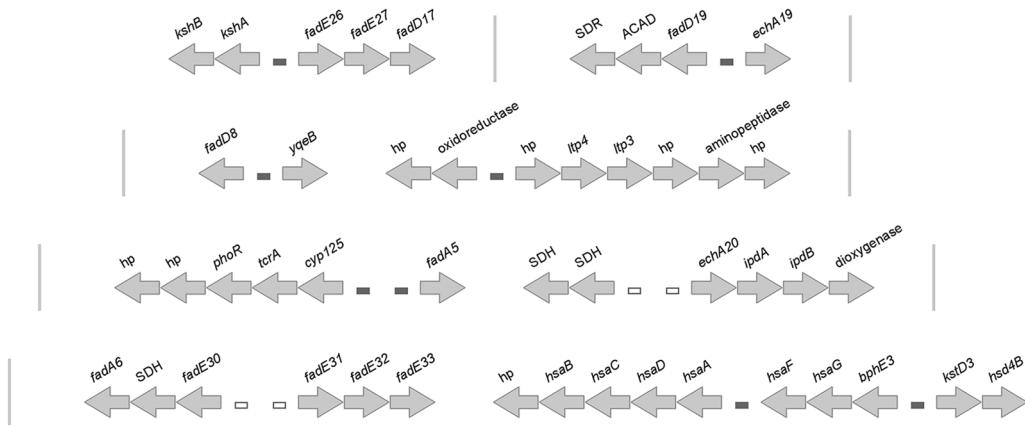
Cluster A (Figs. 3, 4) contained candidate genes related to the degradation of steroid core rings A/B, and presumably, to the oxidation of ring C and 17-acyl side chain. It was very similar to the clusters described for other actinobacteria both in terms of the set and the order



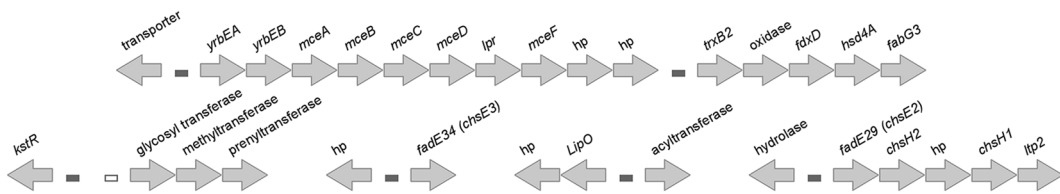
**Fig. 3** General scheme of the genome with orthologs of steroid catabolism genes. Stripes, separate predicted operons. *Black stripes*, predicted operons with no known binding sites. *Blue stripes*, predicted operons with putative KstR binding sites. *Red stripes*, predicted operons with putative KstR2 binding sites. *Green diamonds*, predicted *mce*-loci. A, B, C, D, E, clusters of operons with genes of proteins related to steroid catabolism. Circles, predicted operons with putative binding sites of either KstR or KstR2 and with genes not known to be related to steroid catabolism

of genes: cluster *ro04482-ro04706* in *R. jostii* RHA1 (van der Geize et al. 2007); *MSMEG\_5903-MSMEG\_5943* in *M. smegmatis* (Uhía et al. 2012); *Rv3574-Rv3492c*, which was revealed in *M. tuberculosis* H37Rv and *M.*

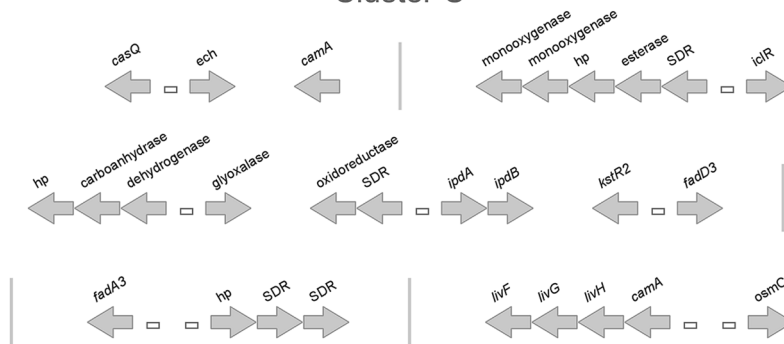
### Cluster A



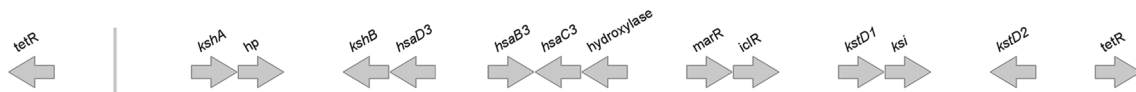
### Cluster B



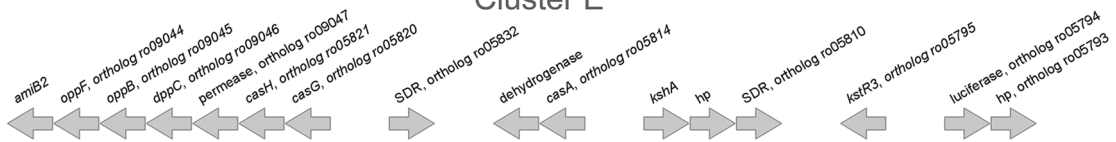
### Cluster C



### Cluster D



### Cluster E



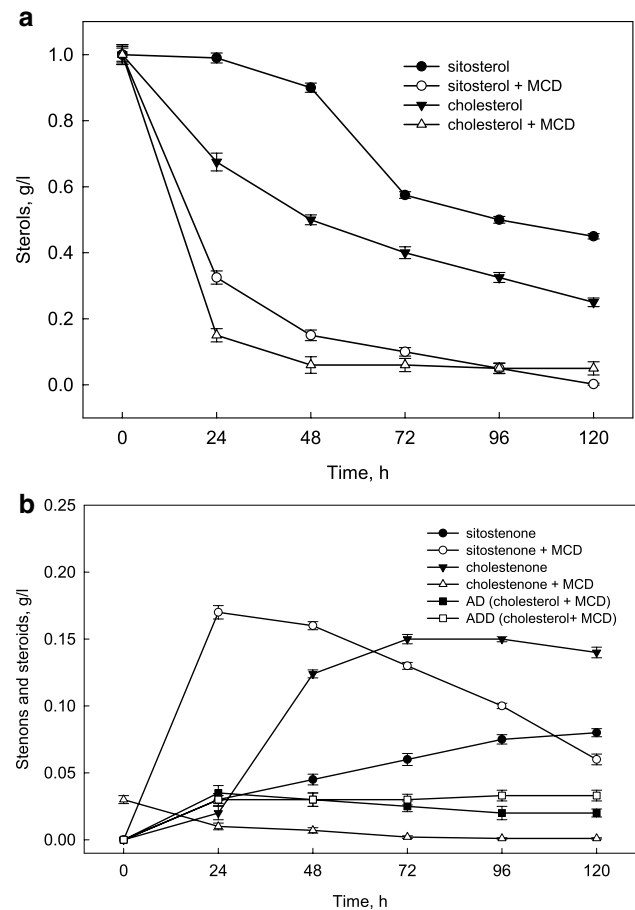
**Fig. 4** Clusters of genes coding for steroid metabolism related proteins. Gene order to be read from *top left to bottom right*. *Vertical lines* designate regions where several genes are omitted. Names of genes belonging to the same operon are written close to each other; neighboring operons are separated by a larger space. *Filled rectangles* denote putative KstR binding sites; *empty rectangles* denote putative binding sites of KstR2

*bovis* Bacillus Calmette-Guérin (BCG) (Wilbrink 2011), as well as in gram-negative bacterium *C. testosteroni* TA441 (Horinouchi et al. 2004). Cluster A matches to a part of *R. jostii* cluster *ro04531-ro04694*. It is noteworthy that the *kshA* gene is absent in the *N. simplex* VKM Ac-2033D genome between the operons *bphE3hsaGF* and *hsaADCB*, but it is present between similar operons in *R. jostii* RHA1 (Wilbrink 2011). Nonetheless, an ortholog of *kshA* having KstRs binding sites is situated in cluster A within the operon which is similar to *MSMEG\_5925\_5927*.

In the *N. simplex* VKM Ac-2033D genome, cluster B (Figs. 3, 4) contains candidate genes related to a side chain degradation pathway. These genes match to the group *ro04482-ro04488* of the main RHA1 cluster (Wilbrink 2011), as well as to the cluster 1 of *M. smegmatis* (Uhiá et al. 2012).

Most genes in clusters A and B have putative KstR binding sites, while the genes of cluster C (Figs. 3, 4) seem to be under the control of KstR2. It was reported that KstR2-regulon encoded genes involved in rings C and D degradation (Casabon et al. 2013b). In *N. simplex* VKM Ac-2033D, the predicted KstR2-regulon is situated in cluster A and comprises several genes matched to the genes of such a regulon in *M. smegmatis* and *M. tuberculosis* (Kendall et al. 2010). A predicted regulon of cluster C has no counterparts among the clusters in the annotated genomes of other actinobacteria. It includes candidate genes of cytochrome P450, oxidoreductases, short-chain dehydrogenase/reductase SDR, 3 $\alpha$ (20 $\beta$ )-hydroxysteroid dehydrogenase, enoyl-CoA hydratases, acetyl-CoA acetyltransferase *fadA3*, and acyl-CoA synthetase *fadD3*. It is possible that the regulon encoding the degradation of rings C/D is split in *N. simplex* VKM Ac-2033D, or the strain has several degradative pathways.

Several clusters do not have putative KstRs binding sites but includes orthologs of the known steroid catabolic genes. Among them, the clusters D (*KR76\_27035-KR76\_27130*) and E (*KR76\_17995-KR76\_18070*) may be of interest (Figs. 3, 4). Their genes could be controlled by non-KstR regulators. A group of predicted operons within cluster D included orthologs of *ksh* subunits, two *kstDs*, gene of steroid delta-isomerase and some predicted transcriptional regulators. Cluster E includes orthologs of genes of cholate degradation cluster of *R. jostii* RHA1 (Mohn et al. 2012).



**Fig. 5** Time-courses of cholesterol, sitosterol (a) and cholest-4-en-3-one and sitost-4-en-3-one (b) during *N. simplex* VKM Ac-2033D growth in sterol-mineral medium

## Sterol utilization

*Nocardioide*s *simplex* VKM Ac-2033D utilized cholesterol and sitosterol as carbon and energy sources (Fig. 5a). No, or poor growth was observed in blanks (mineral medium only, or mineral medium supplemented with MCD only). The rate of consumption was higher in the presence of MCD which is known to enhance bioconversion process (Jadoun and Bar 1993). Androstenedione and androstadienedione were detected as intermediates in low amounts (Fig. 5b). The results confirmed the ability of the strain to degrade sterol molecule.

## Genes encoding sterol uptake system

Diverse actinobacteria contain *mce* transporter operons involved in sterol uptake (Mohn et al. 2008; García et al. 2012). The *Mce4* system in *R. jostii* RHA1 may recognize only side-chains of at least eight carbons (Mohn et al.

2008; García et al. 2012) and included ATPase, two permease subunits (YrbE or Sup), and proteins of unknown function (Mce4ABCDEFHI), closely spaced, transcribed and regulated as one operon under negative regulation by KstR (Mohn et al. 2008). The *mce4* locus expression was induced on cholesterol, but not strongly, probably because of the high basal expression (van der Geize et al. 2007; Mohn et al. 2008; Uhía et al. 2012).

In this study, six predicted *mce* loci were revealed in *N. simplex* VKM Ac-2033D by BLAST. In comparison, two and four *mce* loci had been found in *R. jostii* RHA1 (Mohn et al. 2008), and *M. tuberculosis* H37Rv (Wilbrink 2011), respectively. Many genes of predicted *mce*-loci do not have orthologs in calculated orthogroups, because of the possible recent horizontal transfer of these genes, or their fast evolution. Only one predicted operon (*KR76\_12195-KR76\_12240*) in cluster B (Fig. 4) possesses a putative KstR box (Table S5). Although a predicted operon *KR76\_01200-KR76\_01230* does not possess a KstR binding sites, it may be of interest since it contains the gene *choD* encoding cholesterol oxidase. A similar operon was recently found in *Gordonia cholesterolivorans* with a constitutively transcribed gene encoding ChoOx-2 thus indicating its possible contribution to steroid transport (Drzyzga et al. 2011).

### Cholesterol oxidase and 3-hydroxysteroid dehydrogenase/ $\Delta^5$ -isomerase

The accumulation of cholest-4-en-3-one, or sitost-4-en-3-one was observed during *N. simplex* VKM Ac-2033D incubation with cholesterol, or sitosterol, respectively (Fig. 5); thus evidence modification of 3 $\beta$ -ol-5-ene to 3-keto-4-ene moiety in A-ring of steroid core. This reaction is considered as an initial step of sterols oxidation by actinobacteria (Fig. 1a). In many organisms, the enzymes accounting for 3 $\beta$ -hydroxy dehydrogenation and  $\Delta^{5\rightarrow4}$  isomerization are cholesterol oxidases (ChOs) and/or 3 $\beta$ -hydroxysteroid dehydrogenases (3-HSDs) (Yam et al. 2009). Except for ChoM1 and ChoM2 (Yao et al. 2013), cholesterol oxidases are not critical enzymes for cholesterol utilization in mycobacteria and rhodococci (Griffin et al. 2011; Ivashina et al. 2012; Uhía et al. 2012). Herein, only the cholesterol oxidase ortholog *KR76\_09550* possesses a putative KstR binding site (Table S1), while paralogs of ChoD *KR76\_01200*, *KR76\_06800* and *KR76\_13200* do not. Surprisingly, no genes with high identity to the 3-HSDs known have been revealed in the genome of *N. simplex* VKM Ac-2033D. It should be noted that in some cases 3-HSDs may not be essential for steroid metabolism (Casabon et al. 2013b). Taking into consideration the high efficiency of cholesterol oxidation by *N. simplex* VKM Ac-2033D, one may propose that *KR76\_09550*, or other unidentified enzymes may be

responsible for this functionality in this organism, or this reaction is not controlled by KstR regulators.

Both ChOs and 3-HSDs were shown to have dual functions: they catalyze both the oxidation of 3 $\beta$ -hydroxyl group and  $\Delta^{5\rightarrow4}$  isomerization of 3 $\beta$ -ol-5-ene-steroids. In some proteobacteria (*Comamonas testosteroni*, *Pseudomonas putida*) distinct enzymes were reported to be responsible for isomerization reactions.  $\Delta^5$ -3-Ketosteroid isomerase encoded by *ksi* accounted for double bond isomerization in *C. testosteroni* (Horinouchi et al. 2012) (Fig. 1a). In the *N. simplex* VKM Ac-2033D genome, two copies of candidate genes *ksi* (*ksdI*) of one orthogroup (Table S2) are present, one of them, *KR76\_23530*, having a putative KstR box (Table S5). The putative *ksi* gene *KR76\_27120* identical to that of *A. simplex* (Molnár et al. 1995) was detected in a cluster D (Fig. 4) within the same predicted operon with *kstD*. Most probably, *KR76\_27120* is either not functional, or under the regulation of transcriptional factors differing from KstR. This assumption correlates with a very low expression level of similar genes in *A. simplex* and *S. lividans* (Molnár et al. 1995). The separate *ksi* gene *KR76\_23530* might be a functional one.

### Genes coding for sterol side chain degradation

In mycolic acid rich actinobacteria, the pathway of the side-chain degradation of cholesterol is generally similar to the fatty acids  $\beta$ -oxidation (Capyk et al. 2009b; Nesbitt et al. 2010; Thomas and Sampson 2013) (Fig. 1b.1, b.2). It is initiated by hydroxylation at C26(27), which is catalyzed by the cytochrome P450 monooxygenase Cyp125 (Rosloniec et al. 2009; Capyk et al. 2009b; Wilbrink 2011). As evidenced for *M. tuberculosis* and *M. bovis*, cytochrome Cyp142 is capable of replacing the activity of Cyp125 (Johnston et al. 2010). Cyp125 was shown to be involved in the further oxidation of 26-hydroxycholest-4-en-3-one to cholest-4-en-3-one-26-oic acid (Ouellet et al. 2010). In *M. smegmatis*, the expression of gene *MSMEG\_5995 cyp125* was induced with cholesterol and the promoter had a KstR box (Uhía et al. 2012). In this study, the predicted operon in cluster A includes the gene *KR76\_14335* which is in orthogroup of *cyp125* (Tables S2, S5). Another predicted operon with a putative KstR binding site far from major clusters includes orthologs of *cyp123* (*KR76\_04935*) and *cyp51* (*KR76\_04945*) (Tables S2, S5). We did not find any genes highly identical to *cyp124*, or *cyp142* which had been shown to provide functional redundancy in *M. tuberculosis* (Johnston et al. 2010).

One of the genes of  $\beta$ -oxidation route, *fadD19*, codes for a steroid-CoA ligase which is essential in *R. rhodochrous* DSM43269 for degradation of the C24-branched sterols (Fig. 1b.1) i.e. sitosterol, stigmasterol and others (Wilbrink et al. 2011). In *M. smegmatis*, FadD19 was shown to be



induced with cholesterol (Uhía et al. 2012). We revealed an ortholog of *fadD19* KR76\_14210 with a putative KstR box situated within a cluster A in the *N. simplex* VKM Ac-2033D genome (Tables S2, S5).

Acyl-CoA-dehydrogenases encoded by *fadE26* (renamed as *chsE4*) and *fadE27* (renamed as *chsE5*) forms a heterotetramer with two dimers of ChsE4-ChsE5 with one active site each (Yang et al. 2015). ChsE4-ChsE5 preferentially catalyzes the oxidation of 3-oxo-cholest-4-en-26-oyl CoA in the first cycle of cholesterol side chain  $\beta$ -oxidation, but complex can catalyze the oxidation of 3-oxo-chole-4-en-24-oyl CoA as well as 3-oxo-4-pregnene-20-carboxyl-CoA (Fig. 1b.3). These genes are situated within the operons which are up-regulated on cholesterol (van der Geize et al. 2007; Uhía et al. 2012). We found one predicted operon with the ortholog of *ChsE4-ChsE5* (KR76\_14175-KR76\_14180) in *N. simplex* VKM Ac-2033D with a putative KstR binding site in the cluster A (Tables S2, S5).

Steroid-24-oyl-CoA synthetase encoded by *fadD17* appears to catalyze the ligation of CoA in process of  $\beta$ -oxidation of steroids with C5 side chain (Fig. 1b). FadD17 was not essential for growth of *M. tuberculosis* on cholesterol (Griffin et al. 2011), but its homolog was up-regulated on cholate in *R. jostii* RHA1 (Mohn et al. 2012). The enzyme was proposed to have additional functions, e.g. catalysis of CoA thioesterification of long-chain fatty acids (Casabon et al. 2014). In the genome of *N. simplex* VKM Ac-2033D, ortholog of *fadD17* KR76\_14185 was detected within cluster A and had a putative KstR box (Tables S2, S5). Gene KR76\_18025 orthologous to gene of isofunctional acyl-CoA-synthetase *CasG* of *R. jostii* RHA1 (Mohn et al. 2012) was found in cluster E without a KstR binding site (Tables S2, S5, S6). Ortholog of gene of steroid-22-oyl-CoA synthetase *CasI* of *R. jostii* RHA1 (Mohn et al. 2012; Casabon et al. 2014) (KR76\_13775) without a KstR box was found in *N. simplex* VKM Ac-2033D to be situated separately from main clusters (Tables S2, S5, S6). Probably, under certain conditions the products of these genes may contribute to the degradation pathways of steroids with short side chains.

FadE34 (ro04483), renamed as ChsE3, catalyzes only the oxidation of 3-oxo-chole-4-en-24-oyl CoA in the second cycle of  $\beta$ -oxidation (Yang et al. 2015; Ruprecht et al. 2015). In the genome of *N. simplex* VKM Ac-2033D its ortholog KR76\_12295 is located in the same cluster B under a putative KstR binding site (Tables S2, S5).

Degradation of C24-branched chain sterols was suggested to occur via aldolytic cleavage (Fig. 1b.1) and requires aldol lyases encoded by *ltp3* and *ltp4* (Wilbrink et al. 2012). In *N. simplex* VKM Ac-2033D, one predicted operon was identified with orthologs of *ltp3* (KR76\_14285) and *ltp4* (KR76\_14280) genes in cluster A with a putative KstR box (Tables S2, S5).

A role in side chain degradation was proposed for enoyl-CoA-hydratases encoded by *echA19* (van der Geize et al. 2007) and *hsd4B* (Uhía et al. 2012) because of their location proximal to other genes coding for side chain degradation, as well as of its up-regulation on cholesterol and a presence a KstR box in the promoter. In the case of *N. simplex* VKM Ac-2033D, ortholog *echA19* KR76\_14215 alone and ortholog *hsd4B* KR76\_14505 in predicted operon with ortholog of *kstD3* are located in cluster A, both predicted operons possess putative KstR binding sites (Tables S2, S5).

The Hsd4A might be involved in the degradation of cholesterol side chain, and also may act as 17 $\beta$ -hydroxysteroid dehydrogenase and D-3-hydroxyacyl-CoA dehydrogenase in the presence of branched fatty acids (Uhía et al. 2012). In the genome of *N. simplex* VKM Ac-2033D, the *hsd4A* ortholog KR76\_12260 was revealed in cluster B within predicted operon with a putative KstR box (Tables S2, S5).

Several studies are devoted to an *igr*-locus of *M. tuberculosis*, which is essential for virulence. The operon consists of lipid transfer protein (*ltp2/Rv3540c*), 2 MaoC-like hydratases (*chsH1/Rv3541c* and *chsH2/Rv3542c*), 2 acyl-CoA dehydrogenases (*fadE29/chsE2/Rv3543c* and *fadE28/chsE1/Rv3544c*), and cytochrome P450 (*cyp125/Rv3545c*) (Thomas et al. 2011). The FadE29 (renamed as ChsE2) and FadE28 (ChsE1) form heterotetramer with two active sites that catalyzes 3-oxo-4-pregnene-20-carboxyl-CoA oxidation (Thomas and Sampson 2013), ChsH1-ChsH2 catalyzes the hydration of a 3-oxo-4,17-pregnadiene-20-carboxyl-CoA (Yang et al. 2014) (Fig. 1b.3). The strain of *N. simplex* VKM Ac-2033D possesses a similar predicted operon KR76\_12325- KR76\_12345 in cluster B with a putative KstR box (Fig. 4). The genes orthologous to *chsE2*, *ltp2*, *chsH1* and *chsH2* were detected within the operon (Tables S2, S5). The product of gene *chsE2* in *M. tuberculosis* was shown to function only in a complex with the product of *chsE1*, and the genes in operon were adjacent to each other (Thomas and Sampson 2013). We could not find any *chsE1* homologs in the genome of *N. simplex* VKM Ac-2033D. It is unclear therefore, whether the product of *chsE2* has another function and operates differently, or it is not functional in *N. simplex* VKM Ac-2033D.

### Genes encoding 3-ketosteroid dehydrogenase and 9 $\alpha$ -hydroxylase

At present, the only steroid core degradation pathway is known for actinobacteria and some proteobacteria, the so-called 9(10)-seco pathway (e.g. Donova and Egorova 2012). It includes the formation of chemically unstable 1,4-diene-9 $\alpha$ -hydroxy structure followed by A-ring aromatization (Fig. 1c) and further multistep core degradation (Fig. 1d). The key enzymes accounting for the

opening of B-ring with 9(10)-seco-steroid formation are 3-ketosteroid  $\Delta^1$ -dehydrogenase (KstD) and 3-ketosteroid-9 $\alpha$ -hydroxylase (Ksh) (Petrusma et al. 2011; Capyk et al. 2011).

3-Ketosteroid  $\Delta^1$ -dehydrogenases (KstDs) are FAD-dependent enzymes which catalyze the *trans*-di-axial elimination of the C-1( $\alpha$ ) and C-2( $\beta$ ) hydrogen atoms of the steroid A-ring (Itagaki et al. 1990). *N. simplex* VKM Ac-2033D and other actinobacteria are capable to introduce a  $\Delta^1$ -double bond into a wide range of 1-saturated 3-ketosteroids, but not 3-hydroxysteroids; their strains contains different number of *kstDs* (McLeod et al. 2006; Knol et al. 2008; Fernández de las Heras et al. 2012; Bragin et al. 2013). In the genome of *N. simplex* VKM Ac-2033D, we revealed two orthologs of *kstD1/ksdD* of *R. erythropolis*, two of *kstD3* and one of *kstD2* (Tables S2, S5). The candidate gene *kstD3 KR76\_14500* is adjacent to *hsd4B* within the predicted operon with a putative KstR binding site in cluster A, while the *kstD1 KR76\_02655* with a putative KstR box is located far from main clusters. These findings might evidence enzymatic redundancy of KstDs in *N. simplex* VKM Ac-2033D which is known as one of the most effective strains capable of insertion of 1(2)-double bond into 3-ketosteroids of both androstane and pregnane series (Fokina et al. 2003a, b, 2010; Fokina and Donova 2003; Sukhodolskaya et al. 2010).

3-Ketosteroid 9 $\alpha$ -hydroxylase (KSH) is a two-component non-heme monooxygenase which is composed of the oxygenase KshA with Rieske domain and flavin-containing ferredoxin reductase KshB (Capyk et al. 2009a). The number of the genes encoding KshA and KshB varied from 1 to 5 in different bacteria, and the number of *kshAs* did not always correspond to the number of *kshBs* (Bragin et al. 2013). In *N. simplex* VKM Ac-2033D, two genes were orthologous to *kshA*, and two genes to *kshB* (Tables S2, S5), one homolog of *kshA* out of orthogroups was detected in cluster E. In the cluster A both genes *kshA* and *kshB* were within the same predicted operon having a putative KstR binding site. Most likely, these genes play a role in sterol catabolism, and their products relate to the core degradation by a 9(10)-seco-pathway.

### Genes encoding further steroid core degradation pathway

The activities of 3-KstDs and KSHs result in the formation of intermediate 3-hydroxy-9,10-secoandrosta-1,3,5,(10)-triene-9,17-dione (3-HSA) (Fig. 1d). As shown for some actinobacteria and proteobacteria, further degradation pathway includes 3-HSA hydroxylation with flavin-dependent monooxygenase (HsaAB in *R. jostii* RHA1) (Dresen et al. 2010), oxygenolytic cleavage of ring A with HsaC (or BphC) (van der Geize et al. 2007) followed by the removal

of 2-hydroxy-hexa-2,4-dienoic acid (HHD) of the 3 $\alpha$ -H-4 $\alpha$ (3'-propanoate)-7 $\beta$ -methylhexahydro-1,5-indanedione (HIP) with HsaD (or BphD) (Lack et al. 2010) (Fig. 1d). Genes encoding the pathway in *R. jostii* RHA1, *M. tuberculosis* (van der Geize et al. 2007; Dresen et al. 2010) and *C. testosteroni* (Horinouchi et al. 2004) were studied. Catabolic operon *hsaADCB*-like in *M. smegmatis* was up-regulated on cholesterol (Uhiá et al. 2012). In *N. simplex* VKM Ac-2033D, the only one predicted operon *hsaADCB* is under a putative *kstR* box in cluster A (Tables S2, S5), orthologs of *R. jostii* RHA1 genes *hsaD3B3C3* without a *kstR* box are located in cluster D.

The operon *hsaEGF* with the genes encoding further A-ring degradation was shown to elevate expression during *M. smegmatis* growth on cholesterol (Uhiá et al. 2012). In the genome of *N. simplex* VKM Ac-2033D, two orthologs of *bphE3* of *R. jostii* RHA1 are present (Table S5) which may relate to ring A catabolism (Carere et al. 2013). In our strain, the predicted operon of *hsaFG-bphE3 KR76\_14485-KR76\_14495* having a putative KstR binding site is in the cluster A, another predicted operon with orthologs *bphF3G3E3 (KR76\_24170-KR76\_24180)* without of KstR binding sites is far from main clusters (Tables S2, S5).

It was assumed that the propionate moiety of HIP might be removed via  $\beta$ -oxidation (Fig. 1d) (van der Geize et al. 2007; Wilbrink 2011).

FadD3 is a HIP-CoA synthetase that initiates catabolism of steroid rings C and D in *M. tuberculosis* (Casabon et al. 2013a) (Fig. 1d). Orthologs are found in all sterol catabolizing bacteria. The ortholog of *fadD3 KR76\_25120* with a putative KstR2 box is located within cluster C in the *N. simplex* VKM Ac-2033D genome (Tables S2, S5).

Genes *ipdA* and *ipdB* are supposed to encode a heterodimeric CoA transferase; *R. equi* deletion mutant was impaired in growth on 5-OH-HIP (van der Geize et al. 2011) (Fig. 1d). In *R. jostii* RHA1 genome they are located in the group *ro04649-ro04656* that up-regulated during growth both on cholesterol and on cholate (Mohn et al. 2012). In *N. simplex* VKM Ac-2033D, two copies of both orthologs were detected within two predicted operons in cluster A and in cluster C, and both operons had putative KstR2 binding sites (Tables S2, S5).

Acyl-CoA dehydrogenase FadE30 was shown to be responsible for the further dehydrogenation of 3'-propanoate substituent at C4 of the 5-OH-HIP-CoA (van der Geize et al. 2011) (Fig. 1d). In the *N. simplex* VKM Ac-2033D genome, a predicted operon which includes orthologs of acyl-CoA dehydrogenase (*fadE30 KR76\_14430*, SDH (short-chain dehydrogenase) *KR76\_14435* and acetyl-CoA acetyltransferase (*fadA6 KR76\_14440*) is located in a cluster A. The group of acyl-CoA dehydrogenases encoded by *fadE31*, *fadE32* and *fadE33* belongs to a heterotetrameric subfamily of ACAD.

The products of these genes were shown to be functional only in each other's complexes (Wipperman et al. 2013). Their predicted operon *KR76\_14445-KR76\_14455* is located closely to the above predicted operon in the opposite strand in the *N. simplex* VKM Ac-2033D genome (Fig. 4). Between these operons, two putative KstR2 binding sites are located. A similar group of genes *ro04591-ro04599* in the genome of *R. jostii* RHA1, was up-regulated on cholate, expressed at a high constitutive level on cholesterol and belongs to KstR2 regulon (Mohn et al. 2012).

### Genes encoding transcriptional repressors KstR and KstR2

Transcription of genes involved in steroid metabolism is strongly regulated by transcriptional repressors of TetR-type *kstR* and *kstR2*. KstR was proven to control a large regulon consisted of several dozens of genes (Kendall et al. 2007), while *kstR2* regulated a smaller regulon comprised of 15 genes (Kendall et al. 2010). The repressors may regulate themselves.

In the genome of *N. simplex* VKM Ac-2033D, one ortholog *kstR* *KR76\_12270* is in a cluster B, though its KstR binding site was estimated with above-threshold false discovery rate (*q* value of 0.0295). One ortholog *kstR2* *KR76\_25115* is in a cluster C, and had its own putative KstR2 binding site. Reciprocal BLAST confirms their similarity with *M. tuberculosis* regulators as one-to-one genes. Another TetR-regulator with a putative KstR2 binding site, *KR76\_07630*, is located separately in the genome (Tables S2, S5).

Orthologs *casA* *KR76\_18040*, *kstR3* *KR76\_18060* and *KR76\_27035* are detected in clusters E and D (Tables S2, S5). It is reasonable to propose that they may represent regulators that are distinct from KstR and KstR2, and could play a role in other steroid catabolism in *N. simplex* VKM Ac-2033D.

### Expression of *KR76\_14500*, *KR76\_12295*, and *KR76\_25120* genes in the presence of cholesterol

RT-qPCR experiments were performed to estimate the expression levels of sterol catabolic genes using mRNA from the cells grown in the presence, or absence of cholesterol. Three genes (*KR76\_14500*, *KR76\_12295*, and *KR76\_25120*) were chosen that were localized in three different chromosomal clusters, participated in different stages of sterol catabolic pathway, and predicted to be controlled by transcriptional regulators KstR, or KstR2. All three genes were induced in the presence of cholesterol. The expression of *fadD3* *KR76\_25120* was less responsive to addition of cholesterol ( $\Delta\text{Cp} = 3.6 \pm 0.95$ ). A higher level of expression was observed for *kstD3* *KR76\_14500* and *fadE34/chsE3* *KR76\_12295* ( $\Delta\text{Cp} = 11.8 \pm 0.65$

and  $\Delta\text{Cp} = 8.3 \pm 0.6$ , respectively). For 16S rRNA gene, selected as endogenous control, no significant differences in expression level were observed regardless of the conditions tested.

### Other genes of interest

*N. simplex* VKM Ac-2033D and related organisms are known to be applied for  $\Delta^1$ -dehydrogenation of 1-saturated pregnane steroids, such as cortisol, to produce therapeutically important 1-dehydroanalogs, or valuable intermediates. The  $\Delta^1$ -dehydrogenation by whole-cell catalysts was often accompanied by the reduction of 20-carbonyl group (Aribasaraova et al. 1985; Mahato et al. 1988; Fokina and Donova 2003). In *M. tuberculosis*, *Rv2002* (FabG3), a NADH-dependent  $3\alpha,20\beta$ -hydroxysteroid dehydrogenase is likely to be involved in steroid metabolism (Yang et al. 2003). In the genome of *N. simplex* VKM Ac-2033D, *fabG3* ortholog *KR76\_12265* has a putative KstR box and is located in cluster B (Supplementary material, Tables S1, S2), other paralogs *KR76\_23000* and *KR76\_13560* does not have KstR box. The gene *KR76\_25085*, ortholog of 3-alpha-hydroxysteroid dehydrogenase of *N. farcinica* IFM 10152, was revealed in a cluster C and has a putative KstR box.

A whole genome screening to identify genes with putative binding sites of KstR-regulators in *N. simplex* VKM Ac-2033D predicted operons with the genes encoded acyl-CoA transferases and hydrolases and other transcriptional regulators (*KR76\_19215-KR76\_19220*) that might be involved in the side chain degradation, or steroid core oxidation (Tables S2). Several genes are presumably involved in the metabolism of main nitrogenous bases (*KR76\_11475-KR76\_11490*, *KR76\_19205-KR76\_19210*, *KR76\_22540-KR76\_22555*, *KR76\_23525*, *KR76\_09920-KR76\_09930*, *KR76\_26165-KR76\_26180*), and sulfate metabolism (*KR76\_03905-KR76\_03920*) (Table S2). We could not find genes with a putative KstRs binding sites in *N. simplex* VKM Ac-2033D which matches with additional genes enhancing their expression on cholesterol in *R. jostii* RHA1 (Wilbrink 2011), e.g. genes encoding cycloalkanone catabolism *ro06687-ro06698*, that are suggested to be involved in ring D degradation of HIP (Wilbrink 2011). A few matches were with the genes of KstRs-regulons in *M. smegmatis* (Uhá et al. 2012) (e.g. AraC family transcriptional regulator, succinic semialdehyde dehydrogenase, helicase, peptidase). It is possible that these additional sets of gene are species-, or strain-specific.

The strain *N. simplex* VKM Ac-2033D is capable of utilization of lithocholic acid (data not shown). We also searched genome of *N. simplex* VKM Ac-2033D to find orthologs of genes of cholate catabolism cluster of *R. jostii* RHA1 (Wilbrink 2011; Mohn et al. 2012) (Table S6). The orthologs are genes of main clusters, e.g. *kshA*, *kstD*,

*fadA5* in cluster A, *hsaBCD* in cluster D, several single-copy orthologs in cluster E; several orthologs are located separately from main clusters (e.g., *casI*, *casE*). So, there is no a cholate degradation-like cluster in *N. simplex* VKM Ac-2033D, orthologs of its genes might be involved in degradation of bile acids, in other degradative pathways or be non-functional.

### Comparison with other *Nocardioidea* strains

Known for the ability to utilize ethene, vinyl chloride, propene and butene, fluoroethene, and nicotine, *Nocardioidea* sp. JS614 has been proposed to use as a biocatalyst for production of chiral epoxides (Coleman et al. 2011). *Nocardioidea* sp. CF8 can utilize butane, halogenated hydrocarbons; C2 to C16 *n*-alkanes (Kimbrel et al. 2013). Recently, in a study of a small flowering plant *Arabidopsis thaliana* microbiome (Bai et al. 2015), 17 *Nocardioidea* genomes were sequenced. Unfortunately, nothing is known about the possibility of these strains to utilize cholesterol.

We analyzed differences between the genome of *N. simplex* VKM Ac-2033D and the genomes of *Nocardioidea* strains already sequenced and annotated. No steroid catabolism genes or a few genes, but not a whole set were found in Leaf285, Leaf307, Root122, Soil774, Soil777, Soil805. The set of genes putatively involved in steroid catabolism was found in the strains of *Nocardioidea* spp. CF8, JS614, Root1257, Root140, Root151, Root190, Root224, Root240, Root614, Root682, Root79, Soil796, Soil797 thus allowing to presume that these strains could be able to utilize cholesterol or the corresponding steroids.

Similar to mycolic acid rich actinobacteria, *N. simplex* VKM Ac-2033D harbors different steroid catabolic pathways encoded by distinct clusters of genes under predicted KstR and KstR2-regulation. The strain does not possess a whole set of enzymes for cleavage of propyl moiety of C23-C21-side chain (none orthologs of *chsE1* (*fadE28*) have been found). Nonetheless, the strain is able to degrade cholesterol. It is reasonable to assume an existence of unknown enzymes, or additional mechanisms of sterol side chain degradation which may comprise genes of a new predicted regulon. This suggestion needs a proof from transcriptomic, proteomic and enzymologic studies.

Genes described in this investigation could be used for gene modifying in biotechnological attempts to improve industrial properties of microorganisms in steroid production.

### Conclusions

In this work, the organization of steroid metabolism associated genes in the genome of important biotechnological

strain, *Nocardioidea simplex* VKM Ac-2033D, was first studied. Predicted genes related to sterol uptake system, aliphatic side chain degradation at C17, A/B-, and C/D-ring degradation systems are distributed across one circular chromosome and form several major clusters. The distinctive features of the strain genome include the presence of a predicted regulon of KstR2 having no analogs in other annotated genomes of actinobacteria, an absence of genes encoding 3-hydroxysteroid dehydrogenase and absence of *chsE1* (*fadE28*) orthologs. Future transcriptomic studies are of significance for clearer understanding of the candidate genes' importance for steroid modifications by the organism.

The presented study advances our understanding of steroid metabolism by actinobacteria. The results suggest that steroid catabolic pathways are generally conserved in actinobacteria, but apparently additional system for steroid oxidation may exist. Obviously, biocatalytic potential of *Nocardioidea* and relative actinobacteria toward steroids is not fully disclosed, and the knowledge of steroid catabolic pathways in these organisms will open novel perspectives at the creation of effective biocatalysts for steroid pharmaceutical industry.

**Acknowledgments** VS acknowledges the financial support of her work from Russian Foundation for Basic Research (mol\_a No. 14-04-31089). VF, TI and MD are grateful to Russian Science Foundation for a support of their work (Grant No. 14-24-00169). The authors are thankful to Dmitry Dovbnya (IBPM RAS) and Andrey Shutov (IBPM RAS) for assistance in figure preparation.

### References

- Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410. doi:10.1016/S0022-2836(05)80360-2
- Arinbasarova AY, Karpov AV, Fokina VV et al (1996) Kinetic characteristics of 1-en-dehydrogenation of 6 $\alpha$ -methylhydrocortisone by cells of *Arthrobacter globiformis* 193. *Enzyme Microb Technol* 19:501–506. doi:10.1016/S0141-0229(96)80001-C
- Arinbasarova AY, Medentsev AG, Akimenko VK et al (1985) Redox reactions in hydrocortisone transformation by *Arthrobacter globiformis* cells. *J Steroid Biochem* 23:307–312
- Bai Y, Müller DB, Srinivas G et al (2015) Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* 528:364–369. doi:10.1038/nature16192
- Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28–36
- Bragin EY, Shtratnikova VY, Dovbnya DV et al (2013) Comparative analysis of genes encoding key steroid core oxidation enzymes in fast-growing *Mycobacterium* spp. strains. *J Steroid Biochem Mol Biol* 138:41–53. doi:10.1016/j.jsbmb.2013.02.016
- Capyk JK, Casabon I, Gruninger R et al (2011) Activity of 3-ketosteroid 9 $\alpha$ -hydroxylase (KshAB) indicates cholesterol side chain and ring degradation occur simultaneously in *Mycobacterium tuberculosis*. *J Biol Chem* 286:40717–40724. doi:10.1074/jbc.M111.289975
- Capyk JK, D'Angelo I, Strynadka NC, Eltis LD (2009a) Characterization of 3-ketosteroid 9 $\alpha$ -hydroxylase, a Rieske oxygenase in the

- cholesterol degradation pathway of *Mycobacterium tuberculosis*. J Biol Chem 284:9937–9946. doi:10.1074/jbc.M900719200
- Capyk JK, Kalscheuer R, Stewart GR et al (2009b) Mycobacterial cytochrome p450 125 (cyp125) catalyzes the terminal hydroxylation of c27 steroids. J Biol Chem 284:35534–35542. doi:10.1074/jbc.M109.072132
- Carere J, McKenna SE, Kimber MS, Seah SYK (2013) Characterization of an aldolase-dehydrogenase complex from the cholesterol degradation pathway of *Mycobacterium tuberculosis*. Biochemistry 52:3502–3511. doi:10.1021/bi400351h
- Casabon I, Crowe AM, Liu J, Eltis LD (2013a) FadD3 is an acyl-CoA synthetase that initiates catabolism of cholesterol rings C and D in actinobacteria: role of FadD3 in cholesterol catabolism. Mol Microbiol 87:269–283. doi:10.1111/mmi.12095
- Casabon I, Swain K, Crowe AM et al (2014) Actinobacterial acyl coenzyme A synthetases involved in steroid side-chain catabolism. J Bacteriol 196:579–587. doi:10.1128/JB.01012-13
- Casabon I, Zhu S-H, Otani H et al (2013b) Regulation of the KstR2 regulon of *Mycobacterium tuberculosis* by a cholesterol catabolite. Mol Microbiol 89:1201–1212. doi:10.1111/mmi.12340
- Coleman NV, Wilson NL, Barry K et al (2011) Genome sequence of the ethene- and vinyl chloride-oxidizing actinomycete *Nocardioide* sp. strain JS614. J Bacteriol 193:3399–3400. doi:10.1128/JB.05109-11
- Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190. doi:10.1101/gr.849004
- Donova MV, Egorova OV (2012) Microbial steroid transformations: current state and prospects. Appl Microbiol Biotechnol 94:1423–1447. doi:10.1007/s00253-012-4078-0
- Dresen C, Lin LY-C, D'Angelo I et al (2010) A flavin-dependent monooxygenase from *Mycobacterium tuberculosis* involved in cholesterol catabolism. J Biol Chem 285:22264–22275. doi:10.1074/jbc.M109.099028
- Drzyzga O, de las Heras LF, Morales V et al (2011) Cholesterol degradation by *Gordonia cholesterivorans*. Appl Environ Microbiol 77:4802–4810. doi:10.1128/AEM.05149-11
- Fokina VV, Donova MV (2003) 21-Acetoxy-pregna-4(5),9(11),16(17)-triene-21-ol-3,20-dione conversion by *Nocardioide* simplex VKM Ac-2033D. J Steroid Biochem Mol Biol 87:319–325. doi:10.1016/j.jsbmb.2003.10.002
- Fokina VV, Sukhodolskaya GV, Baskunov BP et al (2003b) Microbial conversion of pregna-4,9(11)-diene-17 $\alpha$ ,21-diol-3,20-dione acetates by *Nocardioide* simplex VKM Ac-2033D. Steroids 68:415–421. doi:10.1016/S0039-128X(03)00043-6
- Fokina VV, Sukhodolskaya GV, Gulevskaya SA et al (2003a) The 1(2)-dehydrogenation of steroid substrates by *Nocardioide* simplex VKM Ac-2033D. Microbiology 72:24–29. doi:10.1023/A:1022265720470
- Fokina VV, Sukhodolskaya GV, Savinova TS et al (2010) Microbial 1-dehydrogenation of 6-aminomethyl substituted androstenedione. J Biotechnol 150:208. doi:10.1016/j.jbiotec.2010.09.021
- Fujimoto Y, Chen CS, Gopalan AS, Sih CJ (1982) Microbial degradation of the phytosterol side chain. II. Incorporation of [<sup>14</sup>C]-NaHCO<sub>3</sub> onto the C-28 position. J Am Chem Soc 104:4720–4722. doi:10.1021/ja00381a056
- García JL, Uhía I, Galán B (2012) Catabolism and biotechnological applications of cholesterol degrading bacteria: cholesterol degradation. Microb Biotechnol 5:679–699. doi:10.1111/j.1751-7915.2012.00331.x
- Van der Geize R, Grommen AWF, Hessels GI et al (2011) The steroid catabolic pathway of the intracellular pathogen *Rhodococcus equi* is important for pathogenesis and a target for vaccine development. PLoS Pathog 7:e1002181. doi:10.1371/journal.ppat.1002181
- Van der Geize R, Yam K, Heuser T et al (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 104:1947–1952. doi:10.1073/pnas.0605728104
- Grant CE, Bailey TL, Noble WS (2011) FIMO: scanning for occurrences of a given motif. Bioinforma Oxf Engl 27:1017–1018. doi:10.1093/bioinformatics/btr064
- Griffin JE, Gawronski JD, Dejesus MA et al (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog 7:e1002251. doi:10.1371/journal.ppat.1002251
- de las Heras LF, van der Geize R, Drzyzga O et al (2012) Molecular characterization of three 3-ketosteroid- $\Delta$ 1-dehydrogenase isoenzymes of *Rhodococcus ruber* strain Chol-4. J Steroid Biochem Mol Biol 132:271–281. doi:10.1016/j.jsbmb.2012.06.005
- Horinouchi M, Hayashi T, Kudo T (2012) Steroid degradation in *Comamonas testosteroni*. J Steroid Biochem Mol Biol 129:4–14. doi:10.1016/j.jsbmb.2010.10.008
- Horinouchi M, Kurita T, Yamamoto T et al (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. doi:10.1016/j.bbrc.2004.09.096
- Itagaki E, Matsushita H, Hatta T (1990) Steroid transhydrogenase activity of 3-ketosteroid- $\Delta$ 1-dehydrogenase from *Nocardia corallina*. J Biochem (Tokyo) 108:122–127
- Ivashina TV, Nikolayeva VM, Dovbnya DV, Donova MV (2012) 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. J Steroid Biochem Mol Biol 129:47–53. doi:10.1016/j.jsbmb.2011.09.008
- Jadoun J, Bar R (1993) Microbial transformations in a cyclodextrin medium. Part 4. Enzyme vs microbial oxidation of cholesterol. Appl Microbiol Biotechnol 40:477–482. doi:10.1007/BF00175734
- Johnston JB, Ouellet H, Ortiz de Montellano PR (2010) Functional redundancy of steroid C26-monooxygenase activity in *Mycobacterium tuberculosis* revealed by biochemical and genetic analyses. J Biol Chem 285:36352–36360. doi:10.1074/jbc.M110.161117
- Kendall SL, Burgess P, Balhana R et al (2010) Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2. Microbiol Read Engl 156:1362–1371. doi:10.1099/mic.0.034538-0
- Kendall SL, Withers M, Soffair CN et al (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*: transcriptional repressor controlling a large lipid metabolism regulon in mycobacteria. Mol Microbiol 65:684–699. doi:10.1111/j.1365-2958.2007.05827.x
- Kimbrel JA, Chang J, Arp DJ, Sayavedra-Soto LA (2013) The draft genome sequence of *Nocardioide* sp. strain CF8 reveals the scope of its metabolic capabilities. Genome Announc. doi:10.1128/genomeA.00439-13
- Knol J, Bodewits K, Hessels GI et al (2008) 3-Keto-5 $\alpha$ -steroid  $\Delta$ (1)-dehydrogenase from *Rhodococcus erythropolis* SQ1 and its orthologue in *Mycobacterium tuberculosis* H37Rv are highly specific enzymes that function in cholesterol catabolism. Biochem J 410:339–346. doi:10.1042/BJ20071130
- Koshcheyenko KA, Turkina MV, Skryabin GK (1983) Immobilization of living microbial cells and their application for steroid transformations. Enzyme Microb Technol 5:14–21. doi:10.1016/0141-0229(83)90057-1
- Krassilnikov N, Skryabin GK, Aseeva I, Korsunskaya L (1959) The 1,2-dehydrogenation of hydrocortisone by *Mycobacterium* sp. 193 cells. Dokl Biol Sci Sect 128:1063–1065
- Lack NA, Yam KC, Lowe ED et al (2010) Characterization of a carbon-carbon hydrolase from *Mycobacterium tuberculosis*

- involved in cholesterol metabolism. *J Biol Chem* 285:434–443. doi:[10.1074/jbc.M109.058081](https://doi.org/10.1074/jbc.M109.058081)
- Lestrovaya N, Nazaruk M, Skryabin GK (1965) The dehydrogenation and reduction of the a ring of delta-4-3-ketosteroids by cell-free extracts of *Mycobacterium globiforme* 193. *Dokl Biol Sci Sect* 163:768–770
- Luthra U, Khadpekar S, Trivedi A et al (2015) Biotransformation of 4-androstene-3, 17-dione to androst-1, 4-diene-3, 17-dione by *Nocardioideis simplex*. *World J Pharm Pharm Sci* 4:1935–1943
- Mahato SB, Banerjee S, Podder S (1988) Oxidative side-chain and ring fission of pregnanes by *Arthrobacter simplex*. *Biochem J* 255:769–774
- McLeod MP, Warren RL, Hsiao WWL et al (2006) The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A* 103:15582–15587. doi:[10.1073/pnas.0607048103](https://doi.org/10.1073/pnas.0607048103)
- Mohn WW, van der Geize R, Stewart GR et al (2008) The actinobacterial mce4 locus encodes a steroid transporter. *J Biol Chem* 283:35368–35374. doi:[10.1074/jbc.M805496200](https://doi.org/10.1074/jbc.M805496200)
- Mohn WW, Wilbrink MH, Casabon I et al (2012) Gene cluster encoding cholate catabolism in *Rhodococcus* spp. *J Bacteriol* 194:6712–6719. doi:[10.1128/JB.01169-12](https://doi.org/10.1128/JB.01169-12)
- Molnár I, Choi KP, Yamashita M, Murooka Y (1995) Molecular cloning, expression in *Streptomyces lividans*, and analysis of a gene cluster from *Arthrobacter simplex* encoding 3-ketosteroid-delta 1-dehydrogenase, 3-ketosteroid-delta 5-isomerase and a hypothetical regulatory protein. *Mol Microbiol* 15:895–905
- Nesbitt NM, Yang X, Fontán P et al (2010) A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol. *Infect Immun* 78:275–282. doi:[10.1128/IAI.00893-09](https://doi.org/10.1128/IAI.00893-09)
- Okonechnikov K, Golosova O, Fursov M, the UGENE team, (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28:1166–1167. doi:[10.1093/bioinformatics/bts091](https://doi.org/10.1093/bioinformatics/bts091)
- Ouellet H, Guan S, Johnston JB et al (2010) *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one: CYP125A1 in cholesterol metabolism. *Mol Microbiol* 77:730–742. doi:[10.1111/j.1365-2958.2010.07243.x](https://doi.org/10.1111/j.1365-2958.2010.07243.x)
- O'Donnell AG, Goodfellow M, Minnikin DE (1982) Lipids in the classification of *Nocardioideis*: reclassification of *Arthrobacter simplex* (Jensen) lochhead in the genus *Nocardioideis* (Prauser) emend. *Arch Microbiol* 133:323–329
- Petrusma M, Hessels G, Dijkhuizen L, van der Geize R (2011) Multiplicity of 3-ketosteroid-9 $\alpha$ -hydroxylase enzymes in *Rhodococcus rhodochrous* DSM43269 for specific degradation of different classes of steroids. *J Bacteriol* 193:3931–3940. doi:[10.1128/JB.00274-11](https://doi.org/10.1128/JB.00274-11)
- Rosloniec KZ, Wilbrink MH, Capyk JK et al (2009) Cytochrome P450 125 (CYP125) catalyses C26-hydroxylation to initiate sterol side-chain degradation in *Rhodococcus jostii* RHA1. *Mol Microbiol* 74:1031–1043. doi:[10.1111/j.1365-2958.2009.06915.x](https://doi.org/10.1111/j.1365-2958.2009.06915.x)
- Ruprecht A, Maddox J, Stirling AJ et al (2015) Characterization of novel acyl coenzyme A dehydrogenases involved in bacterial steroid degradation. *J Bacteriol* 197:1360–1367. doi:[10.1128/JB.02420-14](https://doi.org/10.1128/JB.02420-14)
- Shtratnikova VY, Schelkunov MI, Pekov YA et al (2015) Complete genome sequence of steroid-transforming *Nocardioideis simplex* VKM Ac-2033D. *Genome Announc*. doi:[10.1128/genomeA.01406-14](https://doi.org/10.1128/genomeA.01406-14)
- Sukhodolskaya GV, Fokina VV, Savinova TS et al (2010) Combined chemical and microbiological synthesis of exemestane from sitosterol. *J Biotechnol* 150:189. doi:[10.1016/j.jbiotec.2010.08.493](https://doi.org/10.1016/j.jbiotec.2010.08.493)
- Suzuki K, Konno R, Shimzu T et al (2007) A fermentation product of phytosterol including campestenone reduces body fat storage and body weight gain in mice. *J Nutr Sci Vitaminol (Tokyo)* 53:63–67
- Thomas ST, Sampson NS (2013) *Mycobacterium tuberculosis* utilizes a unique heterotetrameric structure for dehydrogenation of the cholesterol side chain. *Biochemistry* 52:2895–2904. doi:[10.1021/bi4002979](https://doi.org/10.1021/bi4002979)
- Thomas ST, VanderVen BC, Sherman DR et al (2011) Pathway profiling in *Mycobacterium tuberculosis*: elucidation of cholesterol-derived catabolite and enzymes that catalyze its metabolism. *J Biol Chem* 286:43668–43678. doi:[10.1074/jbc.M111.313643](https://doi.org/10.1074/jbc.M111.313643)
- Uhía I, Galán B, Kendall SL et al (2012) Cholesterol metabolism in *Mycobacterium smegmatis*: cholesterol pathway. *Environ Microbiol Rep* 4:168–182. doi:[10.1111/j.1758-2229.2011.00314.x](https://doi.org/10.1111/j.1758-2229.2011.00314.x)
- Validation list (1983) Validation of the publication of new names and new combinations previously effectively published outside the IJSB: List No. 12. *Int J Syst Bacteriol* 33:896–897. doi:[10.1099/00207713-33-4-896](https://doi.org/10.1099/00207713-33-4-896)
- Wilbrink MH (2011) Microbial sterol side chain degradation in Actinobacteria. University Library Groningen, Groningen
- Wilbrink MH, van der Geize R, Dijkhuizen L (2012) Molecular characterization of *ltp3* and *ltp4*, essential for C24-branched chain sterol-side-chain degradation in *Rhodococcus rhodochrous* DSM 43269. *Microbiology* 158:3054–3062. doi:[10.1099/mic.0.059501-0](https://doi.org/10.1099/mic.0.059501-0)
- Wilbrink MH, Petrusma M, Dijkhuizen L, van der Geize R (2011) FadD19 of *Rhodococcus rhodochrous* DSM43269, a steroid-coenzyme A ligase essential for degradation of C-24 branched sterol side chains. *Appl Environ Microbiol* 77:4455–4464. doi:[10.1128/AEM.00380-11](https://doi.org/10.1128/AEM.00380-11)
- Wipperman MF, Yang M, Thomas ST, Sampson NS (2013) Shrinking the FadE proteome of *Mycobacterium tuberculosis*: insights into cholesterol metabolism through identification of an  $\alpha_2\beta_2$  heterotetrameric acyl coenzyme A dehydrogenase family. *J Bacteriol* 195:4331–4341. doi:[10.1128/JB.00502-13](https://doi.org/10.1128/JB.00502-13)
- Yam KC, D'Angelo I, Kalscheuer R et al (2009) Studies of a ring-cleaving dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of *Mycobacterium tuberculosis*. *PLoS Pathog* 5:e1000344. doi:[10.1371/journal.ppat.1000344](https://doi.org/10.1371/journal.ppat.1000344)
- Yang M, Guja KE, Thomas ST et al (2014) A distinct MaoC-like enoyl-CoA hydratase architecture mediates cholesterol catabolism in *Mycobacterium tuberculosis*. *ACS Chem Biol* 9:2632–2645. doi:[10.1021/cb500232h](https://doi.org/10.1021/cb500232h)
- Yang M, Lu R, Guja KE et al (2015) Unraveling cholesterol catabolism in *Mycobacterium tuberculosis*: ChsE4-ChsE5  $\alpha_2\beta_2$  acyl-CoA dehydrogenase initiates  $\beta$ -oxidation of 3-oxo-cholest-4-en-26-oyl CoA. *ACS Infect Dis* 1:110–125. doi:[10.1021/id500033m](https://doi.org/10.1021/id500033m)
- Yang JK, Park MS, Waldo GS, Suh SW (2003) Directed evolution approach to a structural genomics project: Rv2002 from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci* 100:455–460. doi:[10.1073/pnas.0137017100](https://doi.org/10.1073/pnas.0137017100)
- Yao K, Wang F-Q, Zhang H-C, Wei D-Z (2013) Identification and engineering of cholesterol oxidases involved in the initial step of sterols catabolism in *Mycobacterium neoaurum*. *Metab Eng* 15:75–87. doi:[10.1016/j.ymben.2012.10.005](https://doi.org/10.1016/j.ymben.2012.10.005)
- Yu C-P, Roh H, Chu K-H (2007) 17 $\beta$ -estradiol-degrading bacteria isolated from activated sludge. *Environ Sci Technol* 41:486–492