

RESEARCH ARTICLE

Open Access



# Different genome-wide transcriptome responses of *Nocardioides simplex* VKM Ac-2033D to phytosterol and cortisone 21-acetate

Victoria Yu Shtratnikova<sup>1\*</sup> , Mikhail I. Schelkunov<sup>2,3</sup> , Victoria V. Fokina<sup>4,5</sup> , Eugeny Y. Bragin<sup>4</sup> ,  
Andrey A. Shutov<sup>4,5</sup>  and Marina V. Donova<sup>4,5</sup> 

## Abstract

**Background:** Bacterial degradation/transformation of steroids is widely investigated to create biotechnologically relevant strains for industrial application. The strain of *Nocardioides simplex* VKM Ac-2033D is well known mainly for its superior 3-ketosteroid  $\Delta^1$ -dehydrogenase activity towards various 3-oxosteroids and other important reactions of sterol degradation. However, its biocatalytic capacities and the molecular fundamentals of its activity towards natural sterols and synthetic steroids were not fully understood. In this study, a comparative investigation of the genome-wide transcriptome profiling of the *N. simplex* VKM Ac-2033D grown on phytosterol, or in the presence of cortisone 21-acetate was performed with RNA-seq.

**Results:** Although the gene patterns induced by phytosterol generally resemble the gene sets involved in phytosterol degradation pathways in mycolic acid rich actinobacteria such as *Mycolicibacterium*, *Mycobacterium* and *Rhodococcus* species, the differences in gene organization and previously unreported genes with high expression level were revealed. Transcription of the genes related to KstR- and KstR2-regulons was mainly enhanced in response to phytosterol, and the role in steroid catabolism is predicted for some dozens of the genes in *N. simplex*. New transcription factors binding motifs and new candidate transcription regulators of steroid catabolism were predicted in *N. simplex*.

Unlike phytosterol, cortisone 21-acetate does not provide induction of the genes with predicted KstR and KstR2 sites. Superior 3-ketosteroid- $\Delta^1$ -dehydrogenase activity of *N. simplex* VKM Ac-2033D is due to the *kstDs* redundancy in the genome, with the highest expression level of the gene *KR76\_27125* orthologous to *kstD2*, in response to cortisone 21-acetate. The substrate spectrum of *N. simplex* 3-ketosteroid- $\Delta^1$ -dehydrogenase was expanded in this study with progesterone and its 17 $\alpha$ -hydroxylated and 11 $\alpha$ ,17 $\alpha$ -dihydroxylated derivatives, that effectively were 1(2)-dehydrogenated in vivo by the whole cells of the *N. simplex* VKM Ac-2033D.

(Continued on next page)

\* Correspondence: [vtosha@yandex.ru](mailto:vtosha@yandex.ru)

<sup>1</sup>Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Leninskie gory, h. 1, b. 40, Moscow, Russian Federation 119991

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusion:** The results contribute to the knowledge of biocatalytic features and diversity of steroid modification capabilities of actinobacteria, defining targets for further bioengineering manipulations with the purpose of expansion of their biotechnological applications.

**Keywords:** *Nocardioides simplex*, *Arthrobacter simplex*, *Pimelobacter simplex*, Biocatalysts, Transcriptome, Phytosterol, Cortisone acetate, Progesterone

## Key points

- Role in phytosterol catabolism is predicted for some dozens of *N. simplex* genes
- Four *kstD* genes are induced that provide superior 1(2)-dehydrogenase activity of the strain
- Gene *kstD2* KR76\_27125 is highly over-expressed in response to AcC
- Metabolism of cortisone acetate is not regulated by KstR-repressors
- New candidate transcription regulators of steroid catabolism were predicted

## Background

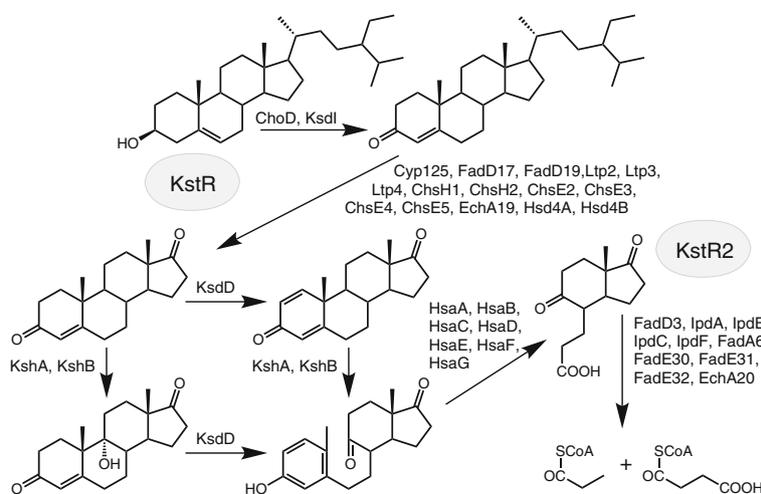
Steroids represent a huge class of specific organic molecules with many of them playing essential roles in all living systems. Diverse bacteria from different ecological niches have been evolved to degrade steroids as sources of carbon and energy, or detoxify exogenic steroids by their structural modifications. Among bacteria, actinobacteria excite a particular interest due to their comprehensive metabolic possibilities and high activity of the enzymatic systems towards various steroids.

Currently, degradation of sterols (such as cholesterol, or phytosterols) by actinobacteria is in the focus of intensive researches due to its exclusive role in

pathogenicity of *Mycobacterium tuberculosis*, and established applications of the non-pathogenic species and their engineered derivatives in biotechnology for production of high-value steroids for the pharmaceutical industry.

Sterol catabolism is a complicated, multi-step process that included degradation of side chain, rings A/B and rings C/D of steroid core oxidation (Fig. 1). This pathway was reported to be controlled by two TetR-type transcriptional repressors, KstR and KstR2 [1–4].

Cholesterol catabolism has been intensively studied in the so-called mycolic acid rich actinobacteria such as pathogenic species: *Mycobacterium tuberculosis* [5], *Rhodococcus* strains [6, 7], as well as in the non-pathogenic species such as *Mycolicibacterium smegmatis* (basonym *Mycobacterium smegmatis*) mc<sup>2</sup>155 [4], *Gordonia cholesterolivorans* [8]. The roles of certain enzymes, mechanisms of their functioning and interactions with substrates were investigated. Cholesterol oxidases (ChOs) and/or 3 $\beta$ -hydroxysteroid dehydrogenases (3-HsDs) were shown to be responsible for modification of 3 $\beta$ -hydroxy-5-ene to 3-keto-4-ene moiety that is considered as a first reaction of sterol degradation [4, 9]. Cytochrome P450 monooxygenases encoded by *cyp125*, *cyp124* and *cyp142* account for sterol hydroxylation at C-26(27) [10, 11], and aldol lyases encoded by *ltp3* and



**Fig. 1** Putative scheme of phytosterol catabolism by *N. simplex* VKM Ac-2033D. “KstR” and “KstR2” in ovals mean regulons

*ltp4* play role in the degradation of C24-branched chain sterols [12]. Acyl-CoA synthetase FadD19 catalyzes the formation of the cholestanoate CoA-thioester [13] that is further oxidized to C-17-ketosteroids by three successive cycles of  $\beta$ -oxidation [4, 14]. 9 $\alpha$ -Hydroxylase (KshAB, consisting of KshA and KshB subunits) and 3-ketosteroid- $\Delta^1$ -dehydrogenase (KstD) [15, 16] are well-known as key enzymes of steroid core degradation whose cooperative action provides the ring B opening (Fig. 1). 3-Hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione (3-HSA) formed is further degraded by the enzymes encoded by *hsaA*, *hsaB*, *hsaC*, *hsaD*, *hsaE*, *hsaF*, *hsaG* [17–19]. The degradation of the last two rings (rings C/D) of the steroid core is catalyzed by the products of the genes of the so-called KstR2-regulon [20]. Unlike mycobacteria, rhodococci and other above-mentioned species related to *Corynebacteriales* order, much less is known on steroid catabolism in actinobacteria belonging to *Propionibacteriales* order such as representatives of *Nocardioide*s genus and related genera. The capability of *Arthrobacter simplex* (syn. *Nocardioide*s *simplex*) to degrade cholesterol has been reported more than 50 years ago and the major intermediate was identified as cholest-4-ene-3-one [21]. Since then mostly 3-ketosteroid- $\Delta^1$ -dehydrogenases (KstDs) from *A. simplex* were studied, the gene coding for these enzymes were identified and the engineered strains were constructed to provide selective production of 1(2)-dehydrosteroids (e.g. [22, 23]).

Recent metagenome research revealed among others the presence of the potent steroid degraders belonging to *Nocardiodaceae* and *Nocardioide*s in different habitats [24]. Dozens of genes associated with steroid catabolism were detected in *Nocardioide*s *dokdonensis* FR1436 isolated from sand sediment [25]. Most the research on the molecular mechanisms of steroid catabolism have been provided by our team for the biotechnologically relevant strain of *Nocardioide*s *simplex* VKM Ac-2033D [26–28]. The data obtained in [26] have been further used for the refolding in situ of a soluble cholesterol oxidase PsChO3 *KR76\_06800* (cholesterol oxidase ChoD) producing cholest-5-en-3-one [29], and for the rational design and protein engineering of cholesterol oxidase (PsChO) *KR76\_21605* producing 3-oxo-4-ene-steroid products thus allowing significant improvement of the enzyme substrate selectivity [30].

Superior 3-ketosteroid- $\Delta^1$ -dehydrogenase (KstD) activity of *Nocardioide*s *simplex* VKM Ac-2033D towards various 3-ketosteroids provides effective production of the pharmaceutical 1(2)-dehydroanalogs [31–34]. It is known that the 1(2)-dehydroanalogs often demonstrate higher therapeutic efficiency and lesser side effects as compared with the corresponding 1(2)-saturated steroid drugs [35, 36]. The *N. simplex* VKM Ac-2033D is also

capable of hydrolyzing steroid esters such as 21-acetylated steroids, reduction of C17 and C20 carbonyl groups of androstanes and pregnanes, respectively [31–33] and able to grow on cholesterol [27].

High KstD activity of *N. simplex* VKM Ac-2033D had been earlier demonstrated towards cortisol [37] and its 6-methylated derivative [38], androst-4-ene-3,17-dione (AD) and its 6-aminomethylated derivatives [34], testosterone [34], androsta-4,9-diene-3,17-dione [39], 21-acetoxy-pregna-4(5),9(11),16(17)-triene-21-ol-3,20-dione and pre-gna-4,9(11)-diene-17 $\alpha$ ,21-diol-3,20-dione acetates [32, 33], 7 $\alpha$ - and 7 $\beta$ -hydroxytestololactone [40], etc. (more information is given in the Supplementary Table S1). In this study, the substrate spectrum of steroids for bioconversion with *N. simplex* was expanded with progesterone (Pr) and its 17 $\alpha$ -hydroxylated and 11 $\alpha$ ,17 $\alpha$ -dihydroxylated derivatives (17 $\alpha$ -OH-Pr and 11 $\alpha$ ,17 $\alpha$ -di-OH-Pr).

The strain had originally been isolated from the Pamir highland soils [41] and identified firstly as *Mycobacterium* sp. 193 [42], then re-classified as *Mycobacterium globiforme* 193 [43], and later as *Arthrobacter globiformis* 193 based on its cultural and morphological properties [44].

In 1976, Prauser had proposed a new genus *Nocardioide*s for the Gram-positive, aerobic, non-sporeforming species, with LL-2,6-diaminopimelic acid in their cell walls and lack of mycolic acids [45]. Accordingly, the coryneform bacteria differed from other *Arthrobacter* species in their cell wall composition such as previously reported *Arthrobacter simplex* had been re-classified as belonging to *Nocardioide*s genus [46]. Currently, more than 90 validly published species are within this genus [47]. The strain *Arthrobacter globiformis* 193 had been re-classified as *Nocardioide*s *simplex* VKM Ac-2033D in 2003 [31].

Intriguingly, Suzuki and Komagata proposed the genus *Pimelobacter* (pimele oil - inhabiting; bacterium - a rod) for the *Arthrobacter* species with LL-2,6-diaminopimelic acid instead of lysine in the peptidoglycan layer with a type strain *Pimelobacter simplex* (syn. *Arthrobacter simplex*) [48]. As a result of taxonomic re-evaluation, *Pimelobacter simplex* was later re-classified as *N. simplex* [49, 50].

This complicated story resulted in the confusing situation when *Arthrobacter simplex*, *Pimelobacter simplex* and *Nocardioide*s *simplex* names are often used as synonyms in the current literature for similar actinobacteria. In our case, a complete genome sequence of *N. simplex* VKM Ac-2033D was presented to NCBI database, and indicated by NCBI database as *Pimelobacter simplex* VKM Ac-2033D [26] under accession no. CP009896, its detailed bioinformatics analysis was carried out in [27].

In our previous study, the putative genes and gene clusters related to the sterol uptake system, aliphatic side chain degradation, A/B- and C/D-ring opening and degradation systems were revealed in the genome of *N. simplex* VKM Ac-2033D [27]. Previously unknown steroid metabolism gene clusters have been predicted. Orthologous clusters were later discovered also in *N. dokdonensis* FR1436<sup>T</sup> [25].

It is known that most bacterial enzyme systems involved in steroid metabolic pathways are mainly induced with different steroids, but not expressed constitutively [51]. In this study phytosterol and cortisone 21-acetate (AcC) were estimated as inducers. Phytosterol is a mixture of plant sterols with a branched side chain (with  $\beta$ -sitosterol as a main component) (Fig. 2). Unlike natural sterols such as cholesterol, or phytosterol, AcC is a synthetic corticosteroid ester which is often applied for induction of KstD activity in *N. simplex* and related strains [38, 52]. However, neither growth on AcC, no AcC bioconversion with *N. simplex* VKM Ac-2033D has been examined so far. The fundamental questions also are: whether AcC can be metabolized by *N. simplex* and whether the gene network that over-expressed in response to AcC corresponds to that of steroid degradation. To address these issues, we compared differential transcriptomes of *N. simplex* cultured with (and without) phytosterol, and with (and without) AcC.

Differentially expressed genes (DEG) sets in different induction conditions were compared and expression of the predicted steroid catabolic genes and their clusters was evaluated. The operons with highest expression level and new candidate transcription regulators motifs of

steroid catabolism were predicted. The *kstD* gene with superior expression in response to AcC was revealed, and steroid conversion assay confirmed high level of KstD activity towards progesterone (a predicted preferable substrate for the product of gene *kstD2*) and progesterone derivatives. Moreover, the over-expressed genes encoding the enzymes related to the basic catalytic properties of the *N. simplex* VKM Ac-2033D were revealed.

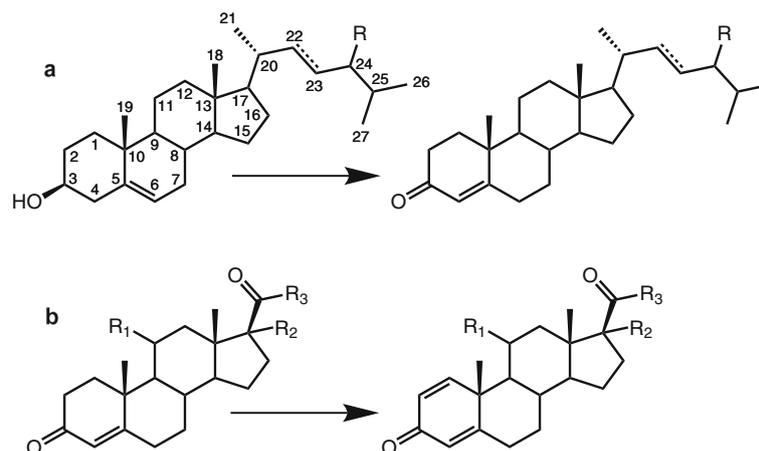
## Results

### Steroid substrates and major bioconversion products

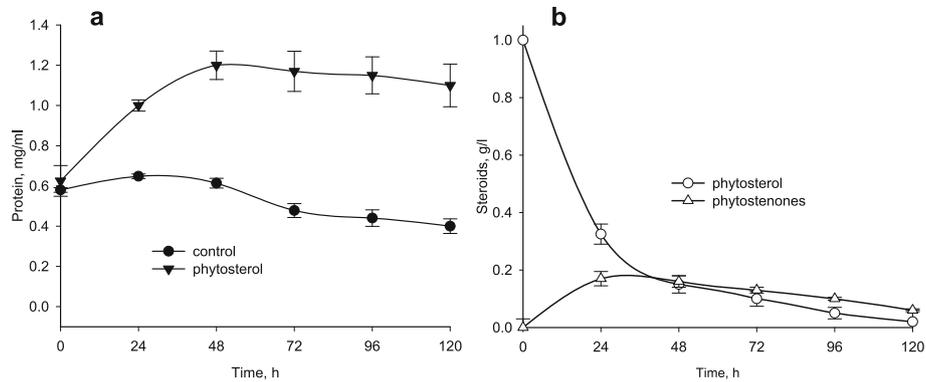
Structures of the steroid substrates used in this work and the bioconversion products detected are presented in Fig. 2. The corresponding 1(2)-dehydroderivatives were identified as major bioconversion products from Pr, 17 $\alpha$ -OH-Pr and 11 $\alpha$ ,17 $\alpha$ -di-OH-Pr and AcC. The main intermediates of phytosterol (comprising a mixture of  $\beta$ -sitosterol, stigmasterol, campesterol, brassicasterol) transformation by *N. simplex* cells were identified as corresponding phytostenones (stigmast-4-en-3-one ( $\beta$ -sitostenone), stigmasta-4,22-dien-3-one, campest-4-en-3-one, brassicast-4-en-3-one) (Supplementary Table S2 and S3). Earlier, campest-4-en-3-one, stigmast-4-en-3-one, and stigmasta-4,22-dien-3-one had been reported as intermediates at bioconversion of campesterol,  $\beta$ -sitosterol, stigmasterol, respectively, with *Arthrobacter simplex* IAM 1660 [21].

### Growth and steroid bioconversion

As shown in Fig. 3a, *N. simplex* VKM Ac-2033D can grow on phytosterol while no growth was observed in the mineral medium with only MCD (control). As



**Fig. 2** Major reactions of steroid bioconversion by *N. simplex*: **a.** Oxidation of 3-hydroxy group of phytosterol constituents: R = C<sub>2</sub>H<sub>5</sub>  $\beta$ -sitosterol/stigmast-4-en-3-one ( $\beta$ -sitostenone); R = C<sub>2</sub>H<sub>5</sub>,  $\Delta^{22,23}$  stigmasterol/stigmasta-4,22-dien-3-one; R = CH<sub>3</sub> campesterol/campest-4-en-3-one, R = CH<sub>3</sub>, brassicasterol/brassicast-4-en-3-one. **b.** R<sub>1</sub> = H, R<sub>2</sub> = H, R<sub>3</sub> = CH<sub>3</sub>, Progesterone (Pr)/1(2)-dehydroprogesterone (DPr); R<sub>1</sub> = H, R<sub>2</sub> = OH, R<sub>3</sub> = CH<sub>3</sub>, 17 $\alpha$ -hydroxyprogesterone (17 $\alpha$ -OH-Pr)/1(2)-dehydro-17 $\alpha$ -hydroxyprogesterone (17 $\alpha$ -OH-DPr); R<sub>1</sub> = OH, R<sub>2</sub> = OH, R<sub>3</sub> = CH<sub>3</sub>, 11 $\alpha$ ,17 $\alpha$ -dihydroxyprogesterone (11 $\alpha$ ,17 $\alpha$ -di-OH-Pr)/1(2)-dehydro-11 $\alpha$ ,17 $\alpha$ -dihydroxyprogesterone (11 $\alpha$ ,17 $\alpha$ -di-OH-DPr); R<sub>1</sub> = O, R<sub>2</sub> = OH, R<sub>3</sub> = CH<sub>2</sub>-O-C(O)-CH<sub>3</sub> 21-acetate of cortisone or cortisone acetate (AcC)/21-acetate of 1(2)-dehydro-cortisone or 21-acetate of prednisone (AcP)



**Fig. 3 a:** Growth of *N. simplex* VKM Ac-2033D on phytosterol. Medium without phytosterol was used in control. **b:** Phytosterol consumption and phytostenones accumulation during strain growth on phytosterol

shown in Fig. 3b, phytosterol was completely degraded by *N. simplex* in approx. 120 h via transit intermediates (phytostenones).

No growth difference was observed when *N. simplex* incubating with AcC in comparison with control (without AcC) (Fig. 4a). Maximum CFU number for AcC which was observed on 24 h was smaller than CFU number in the control:  $17.4 \times 10^8 \pm 2 \times 10^8$  and  $20.8 \times 10^8 \pm 2.8 \times 10^8$  cells/ml, respectively. Unlike phytosterols, AcC was not degraded by *N. simplex*, but converted into the mixture of steroids (21-acetate of prednisone, AcP (main product), prednisone (P), cortisone (C) and their 20-hydroxyderivatives (Fig. 4b)). Quantification of all steroids formed from AcC evidenced the absence of steroid core destruction. Expectedly, no AcP, or any other steroids were detected if no AcC was added (control) (data not shown).

#### High-throughput mRNA sequencing

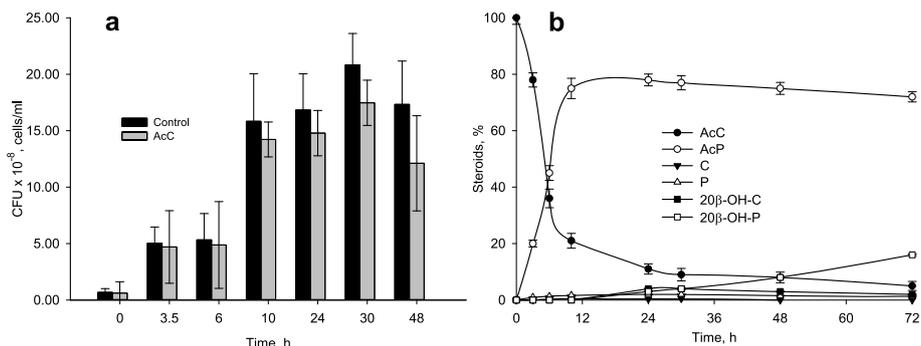
Complete transcriptomes of the *N. simplex* cells grown in the presence, or absence of phytosterol, as well as in the presence, or absence of AcC were obtained. The

total number of reads, the percentage of reads mapping to the rRNA genes, and links to SRA repository in NCBI for experimental variants are provided in the Table 1.

#### Differentially expressed genes (DEG)

DEG analysis showed 91 up- and 42 down-regulated genes in phytosterol plus conditions, while 81 and 31 genes were up- and down regulated in response to AcC, respectively (Table 2). A gene was considered differentially expressed, if its expression increased or decreased more than threefold with  $q$ -value less or equal to 0.01.

In phytosterol induced cells, maximum expression level (59-fold) was observed for gene *KR76\_25195* coding for ABC amino acid transporter; while the minimum level was indicated for *KR76\_00525*, a putative two-component system sensor kinase, at 0.06 of phytosterol minus conditions. When inducing with AcC, maximum over-expression (1207-fold) was determined for the gene *KR76\_27125* encoded KstD; and the minimum level was shown by *KR76\_15900* RidA/YjgF/TdcF/RutC subgroup, at 0.03 of phytosterol minus conditions.



**Fig. 4 a:** Growth of *N. simplex* VKM Ac-2033D in the presence of AcC. Media without AcC was used as a blank. **b:** Dynamics of AcC bioconversion. Data are the average of three replicates. Abbreviations: AcC – cortisone 21-acetate; C - cortisone, 20β-OH-C - 20β-hydroxy cortisone, 20β-OH-P - 20β-hydroxy prednisone

**Table 1** Sequencing statistics and links to reads

Samples	Total number of reads	Percent of reads mapping to the rRNA genes	Link to the reads in SRA
Phytosterol, 1st replicate	10,718,187	2	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX5542009">https://www.ncbi.nlm.nih.gov/sra/SRX5542009</a>
Phytosterol, 2nd replicate	10,138,962	2.9	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX5542008">https://www.ncbi.nlm.nih.gov/sra/SRX5542008</a>
AcC, 1st replicate	8,938,451	3	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX5542005">https://www.ncbi.nlm.nih.gov/sra/SRX5542005</a>
AcC, 2nd replicate	11,883,597	4.4	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX5542004">https://www.ncbi.nlm.nih.gov/sra/SRX5542004</a>
Control, 1st replicate	10,207,198	2.1	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX5542007">https://www.ncbi.nlm.nih.gov/sra/SRX5542007</a>
Control, 2nd replicate	12,237,275	5	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX5542006">https://www.ncbi.nlm.nih.gov/sra/SRX5542006</a>

The steroid induced genes were combined into the operons using bioinformatics methods, supported by the data from the transcriptome analysis. The detailed information about these operons, predicted binding motifs of transcription factors and levels of gene expression is provided in Supplementary Table S4. Induced genes with homology to the known genes of steroid catabolism are depicted in Fig. 5.

Almost two-thirds of the genes that were induced by phytosterol locate in the “typical” steroid metabolism clusters A and B (Table 2), in the newly cluster C that was previously predicted in [27], as well as in the cluster D, which does not have predicted binding motifs for the transcription factors KstR and KstR2.

The only gene that was highly up-regulated both in the presence of phytosterol and AcC, was *KR76\_27125*, *kstD2* coding for KstD. The gene belongs to cluster D. Three genes with unknown functions were significantly down-regulated in phytosterol plus and AcC plus conditions (Supplementary Table S4). No cases of meaningful reverse gene behavior, when one inducer caused significant increase in expression, while another caused significant expression decrease, were detected.

Except for *kstD2*, no significant expression of the known genes putatively involved in the steroid degradation was observed in AcC plus conditions. No pronounced clustering of genes was observed either: the operons that enhanced expression were evenly

distributed throughout the genome (Supplementary Table S4). A number of the AcC-induced genes could be combined into small clusters, specifically, *KR76\_09390* - *KR76\_09460*; *KR76\_16215* - *KR76\_16555*; *KR76\_24345* - *KR76\_24405*; *KR76\_24685* - *KR76\_24800*. These clusters included the genes related to various transporters, amino acid and organic acid metabolism enzymes, cytochromes, transcriptional regulators, dehydrogenases including 3-ketosteroid- $\Delta^1$ -dehydrogenase etc.

#### Expression of steroid catabolism genes orthologs and identification of candidates for new transcription factor regulators of steroid catabolism genes

Our previous study [27] allowed prediction of the KstR and KstR2 binding sites in 71 and 29 operons, respectively. Of these, in phytosterol plus conditions putative KstR-regulated genes of 27 operons (38%) and putative KstR2-regulated genes of 10 operons (34%) increased their expression. In response to AcC, no cases of gene induction with the predicted KstR or KstR2 binding site were observed.

Transcription of nine operons was enhanced in the presence of phytosterol, though the genes do not have predicted binding sites to KstR, or KstR2. Five of them (in which a total of seven genes increased expression significantly) located in cluster D (*KR76\_27035*-*KR76\_27130*).

**Table 2** Differentially expressed genes in gene clusters of *N. simplex* VKM Ac-2033D

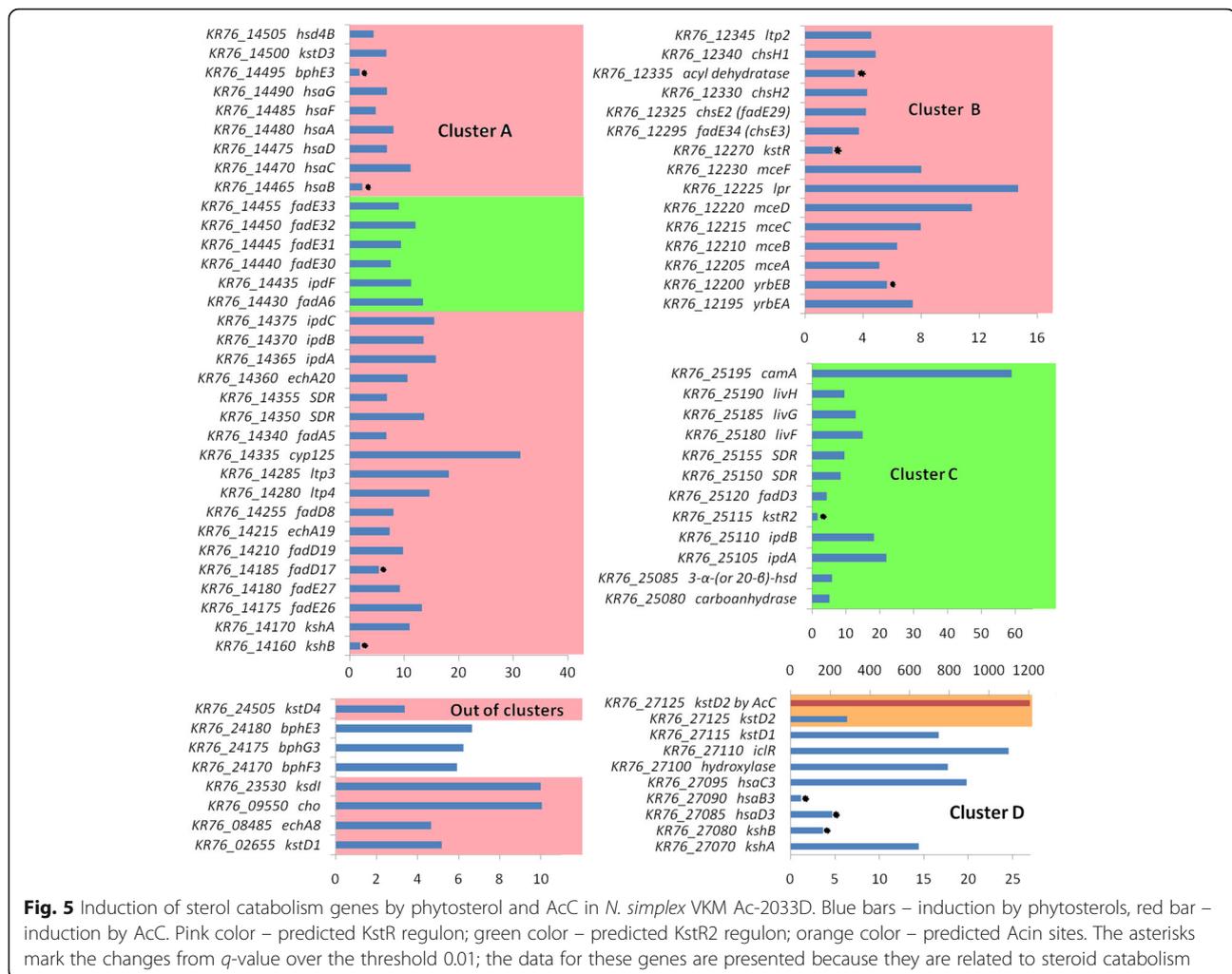
Cluster	Genes	Phytosterol				AcC			
		I	Ave	Max	D	I	Ave	Max	D
A	<i>KR76_14160</i> - <i>KR76_14505</i>	31	10.94	31.24	0	0			0
B	<i>KR76_12190</i> - <i>KR76_12345</i>	12	6.90	14.70	0	0			0
C	<i>KR76_25015</i> - <i>KR76_25200</i>	14	15.10	59.00	0	0			0
D	<i>KR76_27035</i> - <i>KR76_27130</i>	7	13.35	24.53	0	3	415.5	1207.7	0
E	<i>KR76_17995</i> - <i>KR76_18070</i>	0	0.00		0	0			0
Out of clusters		27	7.06	12.44	42	79			32
Total		91	9.94		42	82			32

I - the number of genes that significantly (more than 3 times with  $q$ -value < 0.01) increased their expression

Ave - the average level of induction among the genes that significantly increased expression in the cluster

Max - the maximum level of induction

D - the number of genes that significantly (more than 3 times with  $q$ -value < 0.01) reduced their expression



**Fig. 5** Induction of sterol catabolism genes by phytosterol and AcC in *N. simplex* VKM Ac-2033D. Blue bars – induction by phytosterols, red bar – induction by AcC. Pink color – predicted KstR regulon; green color – predicted KstR2 regulon; orange color – predicted Acin sites. The asterisks mark the changes from  $q$ -value over the threshold 0.01; the data for these genes are presented because they are related to steroid catabolism

To identify candidate motifs for the binding of transcription factors, we used the tool FIMO from the MEME 4.10.0 suite [53]. The search was carried out on the sections 500 bps upstream to 50 bps downstream. Identification was carried out for the operons that (i) increased, (ii) decreased their expression in AcC plus conditions and (iii) down-regulated in phytosterol plus conditions. Several motifs were identified (Supplementary Table S5): three motifs for the operons that increased expression on AcC (Acin1, Acin2 and Acin3) and one motif for the operons that decreased expression on phytosterol (Sitdec1). Nothing was found before the operons that reduced expression on AcC. The operons with these sites are listed in the Supplementary Table S4.

#### Progesterone bioconversion

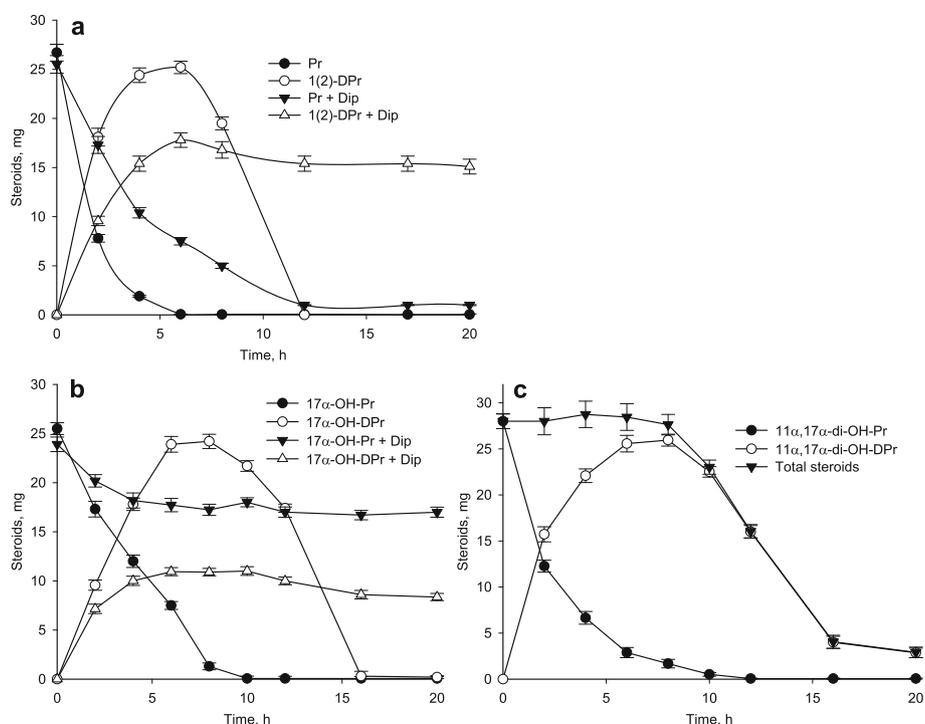
As reported earlier, preferable substrate for KstD2 from *Rhodococcus ruber* is progesterone [54]. The *kstD2* from *N. simplex* has the moderate amino acid sequence

similarity (69%) with that KstD2 from *R. ruber*. In this study, we tested KstD activity of *N. simplex* in vivo towards progesterone (Pr), and its hydroxylated derivatives, such as 17 $\alpha$ -OH-Pr and 11 $\alpha$ ,17 $\alpha$ -di-OH-Pr. As shown in Fig. 6, the corresponding 1(2)-dehydroderivatives were formed as major products from all three substrates. Their accumulation reached maximum level at 10 h, followed by the consequent degradation. Lower degradation was observed in the presence of  $\alpha,\alpha$ -dipyridyl (the inhibitor of iron containing enzymes).

## Discussion

### Sterol degradation

The genetics of sterol (phytosterol and  $\beta$ -sitosterol) degradation by mycolic-acid rich actinobacteria belonging to *Mycolicibacterium* (syn. *Mycobacterium*) (e.g. *M. smegmatis*, *M. neoaurum*), *Mycobacterium* (e.g. *M. tuberculosis*), *Rhodococcus*, *Gordonia* and other genera have been studied intensively during past two decades, while little information is known for the representatives



**Fig. 6 a:** Bioconversion of progesterone (Pr, 0.4 g/l) with and without  $\alpha,\alpha$ -dipyridyl (Dip). **b:** Bioconversion of 17 $\alpha$ -hydroxyprogesterone (17 $\alpha$ -OH-Pr, 1 g/l) with and without  $\alpha,\alpha$ -dipyridyl. **c:** Bioconversion of 11 $\alpha$ ,17 $\alpha$ -di-hydroxyprogesterone (11 $\alpha$ ,17 $\alpha$ -di-OH-Pr, 1 g/l) without  $\alpha,\alpha$ -dipyridyl. In all variants resting AcC-induced *N. simplex* VKM Ac-2033D cells were used

of *Nocardioide*s and related actinobacteria. Similar to many other soil-dwelling bacteria [55] *N. simplex* could be able to utilize abundant plant sterols such as sitosterol which is a core component of plant membranes. Indeed, *N. simplex* VKM Ac-2033D grows on phytosterol as a sole source of carbon and energy as it was confirmed in this study. Phytosterones were detected as major intermediates at phytosterol degradation by the *N. simplex* VKM Ac-2033D cells.

*N. simplex* is capable of full degradation of phytosterol and different other steroids without substituents in steroid core such as androstenedione, or progesterone (Supplementary Table S1). Genes whose products could play role in the reactions of steroid catabolism in *N. simplex* VKM Ac-2033D were predicted (Table 3, Fig. 7) based on their orthology with the known genes and new transcriptomic data.

Bacterial steroid degradation pathway was shown to be controlled by two TetR-type transcriptional repressors named KstR and KstR2 [1–4]. In *Actinobacteria* 3-oxocholest-4-en-26-oic acid and its derivatives may regulate the transcription factor KstR [58], while the effector of KstR2 is HIP-CoA [80]. The reciprocal BLAST with the corresponding regulators in *M. tuberculosis* confirmed the amino acid sequence similarity of KR76\_12270 with KstR and KR76\_25115 with KstR2 [27].

As shown for saprophytic mycobacteria, side chain oxidation and ring B modification are the independent processes [81]. The alkyl side chain of sterols is cleaved according to the mechanism of fatty acid  $\beta$ -oxidation, which is triggered by hydroxylation at C26(27). Further oxidation of the side chain occurs with sequential cleavage of two- and three-carbon fragments. Steroid core degradation seems to carry out through the 9(10)-seco pathway with the B ring opening (see below). After that step sequential destruction of the rings A and B remnants, as well as the rings C and D occurs (Table 3, Fig. 7).

We considered in more details the genes related to the B ring opening. As mentioned above, the only known mechanism of bacterial steroid core degradation is the 9(10)-seco pathway which comprises 1(2)-dehydrogenation catalyzed by KstDs and 9 $\alpha$ -hydroxylation by 3-ketosteroid-9 $\alpha$ -hydroxylase (KshAB) (Fig. 7). The former enzymes encoded by *kstDs* [82], while the Rieske 3-ketosteroid-9 $\alpha$ -hydroxylase consists of KshA monooxygenase and KshB reductase subunits encoding by *kshA* and *kshB*, respectively [71].

There are three *kshA* orthologs in the genome of *N. simplex* [27]. Transcription of two of them was enhanced in phytosterol induced cells, specifically, KR76\_14170 that located in cluster A (10.9-fold), and KR76\_

**Table 3** *N. simplex* VKM Ac-2033D genes whose products taking part in steroid degradation, predicted

Step	Enzyme annotation	<i>M. tub</i>	<i>R. jostii</i>	<i>C. testosteroni</i>	References	<i>N. simplex</i>	Cluster	Regulon	Phyt
<i>Regulation</i>									
	TetR transcriptional repressor	<i>kstR</i>			[2]	KR76_12270	B	KstR	
	TetR transcriptional repressor	<i>kstR2</i>			[2]	KR76_25115	C	KstR2	
<i>A-ring activation</i>									
I-II	cholesterol oxidase	<i>cho</i>			[56]	KR76_09550		KstR	+
I-II	$\Delta^5$ - $\Delta^4$ -3-ketosteroid isomerase	<i>ksi</i>			[57]	KR76_23530		KstR	+
					[57]	KR76_24125		KstR	
<i>Cholesterol side chain degradation</i>									
III-V	cytochrome P450 125	<i>cyp125</i>			[10, 58]	KR76_14335	A	KstR	+
V-VI	3-oxocholest-4-en-26-oate-CoA ligase	<i>fadD19</i>			[59]	KR76_14210	A	KstR	+
VI-VII	Acyl-coA dehydrogenases	<i>chsE4-5</i> ( <i>fadE26-27</i> )			[60]	KR76_14175- KR76_14180	A	KstR	+
VII-VIII	2-enoyl-CoA hydratase	<i>echA19</i>			[6]	KR76_14215	A	KstR	+
IX-X; XI-XII	aldol lyase	<i>ltp3</i>			[12]	KR76_14285	A	KstR	+
	aldol lyase	<i>ltp4</i>				KR76_14280	A	KstR	+
XII-XIII	steroid-22-oyl-CoA synthetase	<i>fadD17</i>	<i>casI</i>		[61]	KR76_14185	A	KstR	
XIII-XIV	acyl-CoA dehydrogenase	<i>chsE3</i> ( <i>fadE34</i> )			[60, 62]	KR76_12295	B	KstR	+
XIV-XV	3 $\alpha$ ,7 $\alpha$ ,12 $\alpha$ -trihydroxy-5 $\beta$ -chol-22-en-24-oyl-CoA hydratase	<i>hsd4B</i> , <i>Rv3538</i>	<i>casD</i>		[4, 61, 63]	KR76_14505	A	KstR	+
XV-XVI	dehydrogenase	<i>hsd4A</i>			[4]	KR76_12260	B	KstR	
XIV-XVII	3-ketoacyl-CoA thiolase	<i>fadA5</i>	<i>casB</i>		[64]	KR76_14340	A	KstR	+
XVII-XVIII; <i>igr</i> -operon	7 $\alpha$ ,12 $\alpha$ -dihydroxy-23,24-bisnorchola-1,4-dien-22-oyl-CoA dehydrogenase	<i>chsE2</i> ( <i>fadE29</i> )	<i>casL</i>		[65]	KR76_12325	B	KstR	+
	acetyl-CoA acetyltransferase	<i>chsH2</i>	<i>casM</i>		[66, 67]	KR76_12330	B	KstR	+
	acyl dehydratase		<i>casN</i>		[66]	KR76_12335	B	KstR	
	beta-hydroxyacyl-ACP dehydratase	<i>chsH1</i>	<i>casO</i>		[66, 67]	KR76_12340	B	KstR	+
	3-ketoacyl-CoA thiolase	<i>ltp2</i>	<i>casP</i>		[66]	KR76_12345	B	KstR	+
<i>Opening of B-ring</i>									
XVIII-XIX; XX-XXI	$\Delta$ -3-ketosteroid dehydrogenase with preference for 9-OH-AD and testosterone	<i>kstD</i>	<i>kstD1</i>		[54, 68]	KR76_02655 KR76_27115		KstR	+
	$\Delta$ -3-ketosteroid dehydrogenase with preference for progesterone	<i>kstD</i>	<i>kstD2</i>		[69]	KR76_27125	D	Acin1,3	+
	$\Delta$ -3-ketosteroid dehydrogenase with preference for saturated steroid substrates	<i>kstD</i>	<i>kstD3</i>	<i>tesH</i>	[54, 70]	KR76_14500	A	KstR	+
XVIII-XX; XIX-XXI	3-ketosteroid 9 $\alpha$ -hydroxylase, oxygenase subunit	<i>kshA</i>			[71]	KR76_14170 KR76_27070	A	KstR	+
	3-ketosteroid 9 $\alpha$ -hydroxylase, reductase subunit	<i>kshB</i>				KR76_14160 KR76_27080	A	KstR	
<i>eCleavage of A-ring</i>									
XXII-XXIII	flavin-dependent monooxygenase	<i>hsaA</i> <i>hsaB</i>		<i>tesA1A2</i>	[18, 72]	KR76_14480 KR76_14465 KR76_27090	A A D	KstR KstR	+
XXIII-XXIV	extradiol dioxygenase	<i>hsaC</i>		<i>tesB</i>	[72, 73]	KR76_14470	A	KstR	+

**Table 3** *N. simplex* VKM Ac-2033D genes whose products taking part in steroid degradation, predicted (Continued)

Step	Enzyme annotation	<i>M. tub</i>	<i>R. jostii</i>	<i>C. testosteroni</i>	References	<i>N. simplex</i>	Cluster	Regulon	Phyt
						KR76_27095	D		+
XXIV-XXV	carbon-carbon hydrolase	<i>hsaD</i>		<i>tesD</i>	[19, 72]	KR76_14475	A	KstR	+
						KR76_27085	D		
XXV-XXVI	hydratase	<i>hsaE</i>		<i>tesE</i>	[17, 74]	KR76_14495	A	KstR	
						KR76_24180			+
XXVI-Propanal	4-hydroxy-2-oxovalerate aldolase	<i>hsaF</i>		<i>tesG</i>		KR76_14485	A	KstR	+
						KR76_24170			+
Propanal-Propionate	acetaldehyde dehydrogenase, acetylating	<i>hsaG</i>		<i>tesF</i>		KR76_14490	A	KstR	+
						KR76_24175			+
<i>Ring D degradation</i>									
XVII-XVIII	HIP-CoA synthetase	<i>fadD3</i>		<i>scdA</i>	[75, 76]	KR76_25120	C	KstR2	+
XVII-XIX	5-Oxo HIC-CoA oxidase	<i>ipdF</i>		<i>scdG</i>	[20, 76]	KR76_14435	A	KstR2	+
XXIX-XXX	COCHEA-CoA hydrolase, α-subunit	<i>ipdA</i>		<i>scdL1L2</i>	[20, 76]	KR76_14365	A	KstR2	+
						KR76_25105	C	KstR2	+
	COCHEA-CoA hydrolase, β-subunit	<i>ipdB</i>				KR76_14370	A	KstR2	+
						KR76_25110	C	KstR2	+
XXX-XXXI	acyl-CoA dehydrogenase	<i>fadE30</i>		<i>scdC1C2</i>	[76, 77]	KR76_14440	A	KstR2	+
XXXI-XXXII	acyl-CoA hydrolase			<i>scdD</i>	[76]				
XXXII-XXXIII	dehydrogenase, oxidized a hydroxyl in propionate chain	<i>hsd4A</i>		<i>scdE</i>	[76]	KR76_14350	A	KstR2	+
XXXIII-XXXIV	acetyl-CoA acetyltransferase	<i>fadA6</i>		<i>scdF</i>	[20, 76]	KR76_14430	A	KstR2	+
XXXIV-XXXV; XXXVI-XXXVII	5-Oxo HIC-CoA oxidase	<i>ipdF</i>		<i>scdG</i>	[20, 76]	KR76_14435	A	KstR2	+
XXXIV-XXXV; XXXVI-XXXVII	5-OH HIC-CoA reductase	<i>ipdC</i>		<i>scdK</i>	[20, 76]	KR76_14375	A	KstR2	+
<i>Ring C degradation</i>									
XXXVII-XXXVIII	HIEC-CoA hydrolase	<i>echA20</i>		<i>scdY</i>	[20, 76]	KR76_14360	A	KstR2	+
XXXVIII-XXXIX	CoA transferase	<i>ipdA</i>		<i>scdL1L2</i>	[20, 76]	KR76_14365	A	KstR2	+
						KR76_25105	C	KstR2	+
	CoA transferase	<i>ipdB</i>				KR76_14370	A	KstR2	+
						KR76_25110	C	KstR2	+
XL-XLI	acyl-CoA dehydrogenases	<i>fadE31</i> , <i>fadE32</i>	<i>ro04591</i> - <i>ro04593</i>		[61, 78]	KR76_14445 KR76_14450	A	KstR2	+
	acyl-CoA dehydrogenases	<i>fadE33</i>				KR76_14455	A	KstR2	+
XLI-XLII	enoyl-CoA hydratase	<i>echA13</i>		<i>scdN</i>	[20, 76]	KR76_25020	C	KstR2	+

Step – degradation step from Fig. 7

Enzyme annotation – function of enzyme

*M. tub.* – name of gene in *M. tuberculosis* investigations*R. jostii* - name of gene (or locus\_tag) in *R. jostii* investigations*C. testosteroni* - name of gene in *C. testosteroni* investigations*N. simplex* - locus\_tag of ortholog that is candidate gene for this function. Functions are predicted by the annotation, location, predicted regulon that ortholog belongs to and gene expression changes

Cluster – cluster where gene is located in genome [27]

Regulon – repressor whose binding site is predicted in the promoter of operon that ortholog belongs to

Phyt – gene in *N. simplex* VKM Ac-2033D that is induced by phytosterol (+)



(See figure on previous page.)

**Fig. 7** Biochemical scheme of sterol catabolism in *N. simplex* VKM Ac-2033D. Genes coding respective proteins are denoted, locus-tags from *N. simplex* VKM Ac-2033D are predicted, see Table 3. **a** Modification of 3 $\beta$ -ol-5-ene to 3-keto-4-ene moiety in A-ring of steroid core; **b** degradation of the C24-branched sterol side chain to C<sub>19</sub>-steroids; **c** steroid core modifications; **d** steroid core degradation via 9(10)-seco pathway. I - sterols; II - stenones. R = H, campesterol, campestenone; R = CH<sub>3</sub>, sitosterol, stigmast-4-en-3-one ( $\beta$ -sitostenone); XVII - androst-4-ene-3,17-dione (AD), XVIII - androsta-1,4-diene-3,17-dione (ADD), XIX - 9 $\alpha$ -hydroxy-AD, XX - unstable 9 $\alpha$ -hydroxy-ADD, XXI - 3 $\beta$ -hydroxy-9,10-seco-androsta-1,3,5(10)-triene-9,17-dione (3 $\beta$ HSA), XXII - 3,4-dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione (3,4-DHSA), XXIII - 4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oic acid (4,9-DSHA), XXIV - 2-hydroxyhexa-2,4-dienoic acid (2-HHD), XXV - 4-hydroxy-2-oxohexanoic acid, XXVI - 9,17-dioxo-1,2,3,4,10,19-hexanorandrost-5-oic acid (DOHNAA) or 3 $\alpha$ -H-4 $\alpha$ -(3'-propanoate)-7 $\beta$ -methylhexahydro-1,5-indadione (HIP), XXVII - 9,17-dioxo-1,2,3,4,10,19-hexanorandrost-5-oyl-CoA (HIP-CoA) XXVIII - 9-hydroxy-17-oxo-1,2,3,4,10,19-hexanorandrost-5-oic acid or 3 $\alpha$ -H-4 $\alpha$ -(3'-propanoate)-5 $\alpha$ -hydroxy-7 $\beta$ -methylhexahydro-1-indanone (5-OH-HIP), XXIX - 9-hydroxy-17-oxo-1,2,3,4,10,19-hexanorandrost-5-oyl-CoA (5-OH-HIP-CoA), XXX - 9-hydroxy-17-oxo-1,2,3,4,10,19-hexanorandrost-6-ene-5-oyl-CoA (HIPE-CoA), XXXI - 7,9-dihydroxy-17-oxo-1,2,3,4,10,19-hexanorandrost-5-oyl-CoA, XXXII - 9-hydroxy-7,17-dioxo-1,2,3,4,10,19-hexanorandrost-5-oyl-CoA, XXXIII - 9-hydroxy-17-oxo-1,2,3,4,5,6,10,19-octa-norandrost-7-oyl-CoA or 3 $\alpha$ -H-4 $\alpha$ (carboxylCoA)-5 $\alpha$ -hydroxy-7 $\beta$ -methylhexahydro-1-indanone (5-OH-HIC-CoA), XXXIV - 9,17-dioxo-1,2,3,4,5,6,10,19-octa-norandrost-7-oyl-CoA, XXXV - 9-hydroxy-17-oxo-1,2,3,4,5,6,10,19-octa-norandrost-8(14)-en-7-oyl-CoA, XXXVI - 9,17-dioxo-1,2,3,4,5,6,10,19-octa-norandrost-8(14)-en-7-oyl-CoA or 7 $\alpha$ -Methyl-1,5-dioxo-2,3,5,6,7,7a-hexahydro-1H-indene-4-carboxylic acid (HIEC-CoA), XXXVII - 9-oxo-1,2,3,4,5,6,10,19-octanor-13,17-secoandrost-8(14)-ene-7,17-dioic acid-CoA-ester or (R)-2-(2-carboxyethyl)-3-methyl-6-oxocyclohex-1-ene-1-carboxyl-CoA (COCHEA-CoA), XXXVIII - 6-methyl-3,7-dioxo-decane-1,10-dioic acid-CoA ester, XXXIX - 4-methyl-5-oxo-octane-1,8-dioic acid-CoA ester, XL - 4-methyl-5-oxo-oct-2-ene-1,8-dioic acid-CoA ester (MOODA-CoA), XLI - 3-hydroxy-4-methyl-5-oxo-octane-1,8-dioic acid-CoA ester, XLII - 4-methyl-3,5-dioxo-octane-1,8-dioic acid-CoA ester, XLIII - 2-methyl-3-oxo-hexane-1,6-dioic acid-CoA ester, XLIV - succinyl-CoA, XLV - propionyl-CoA. Adopted from: [6, 12, 14, 17, 20, 60, 67, 75, 76, 79]

27070 in cluster D (14-fold), which has no predicted *kstR* binding sites (Table 3). Noteworthy, none of them increased expression in response to AcC.

Broadest substrate range without clear substrate preference was reported for KshA5 of *Rhodococcus rhodochrous* DSM 43269 [83]. The enzyme showed activity towards different  $\Delta^4$ ,  $\Delta^{1,4}$ , 5 $\alpha$ -H and 5 $\beta$ -H steroids, steroids with bulky aliphatic side chains and was found to be an only KshA enzyme of the DSM 43269 strain that exhibited activity towards 5 $\alpha$ -androstane-17 $\beta$ -ol-3-one and cortisol [83]. The closest ortholog of *kshA5* of *R. rhodochrous* DSM43269 is *KR76\_27070* in *N. simplex* VKM Ac-2033D (identity 62%). Its product could be involved in progesterone (and other steroids) degradation by *N. simplex* via 9 $\alpha$ -hydroxylation of their 1(2)-dehydrogenated intermediates. Unlike *R. rhodochrous* DSM 43269, the *N. simplex* VKM Ac-2033D does not degrade cortisol (Supplementary Table S1). It is possible that the product of *KR76\_27070* is not active towards cortisol. The 3-keto-4-ene-steroid substrates substituted in positions 6, 7, 9, and 11 are considered more protected than unsubstituted 3-keto-4-ene-steroids from the action of the key enzymes involved in the degradation the steroid core; such 'protected' 3-keto-4-ene-steroids are widely used as the substrates in bioconversions with *N. simplex* [34, 37, 40].

The amino acid alignment and conserved amino acid residues [71] of expressed KshAs of *N. simplex* are shown in Supplementary Figure S1. N240 in the catalytic domain is replaced with D240. Similar replacement was earlier observed for the KshA of *Mycolicibacterium* sp. VKM Ac-1817D (syn. *Mycobacterium* sp. VKM Ac-1817D) [84]. Since *N. simplex* VKM Ac-2033D is not capable of degrading 11-functionalised steroids (e.g.

cortisol, prednisolone, cortisone), this replacement might have altered the catalytic site in *N. simplex*.

It is well known that the number of oxygenase (KshA) and reductase (KshB) subunits of 9 $\alpha$ -hydroxylase can differ [83]. For instance, five genes coding for KshA and two *kshB* genes were revealed in the genome of *Mycobacterium* sp. VKM Ac-1817D. Interestingly, when *N. simplex* VKM Ac-2033D growing on phytosterol, two *kshAs* and one *kshB* were up-regulated [84].

The domain structure of KshB of *N. simplex* is similar to KshB from *R. erythropolis* strain SQ1 [85] (Supplementary Fig. S2), with two replacements in a NAD binding domain of *KR76\_14160* that might indicate disturbed functionality of *KR76\_14160* consistent with its lower induction level (1.9-fold) (Table 4, Fig. 5).

With regard to KstDs, five candidate *kstD* genes were earlier revealed in the genome of *N. simplex* that are divided into three orthogroups, specifically, *kstD1*, *kstD2* and *kstD3* [27].

KstD1 of *Rhodococcus erythropolis* PR4 and *R. ruber* shows substrate preference for 9-OH-AD and testosterone [54]. The orthologs of *kstD1* are *KR76\_02655* and *KR76\_27115* which increased their expression in phytosterol plus conditions.

KstD2 of *R. ruber* showed maximum preference for progesterone [54], while KstD2 of *Gordonia neofelifaecis* NRRL B-59395 transformed progesterone and cholest-4-en-3-one [69]. We tested KstD activity of *N. simplex* in vivo towards Pr, and its hydroxylated derivatives, such as 17 $\alpha$ -OH-Pr and 11 $\alpha$ ,17 $\alpha$ -di-OH-Pr. The corresponding 1(2)-dehydroderivatives were formed (Fig. 6). Lower degradation was observed in the presence of  $\alpha,\alpha$ -dipyridyl (the inhibitor of the iron containing enzymes, such as KshA). We suppose that progesterone bioconversion

**Table 4** Differentially expressed genes encoding KstDs and KshA/B

Locus_tag	Ortholog in Mtb	Localization	Product	q-value Phyt/C <sup>a</sup>	Phyt/C	q-value AcC/C <sup>b</sup>	AcC/C
KR76_14170	<i>kshA</i>	Cluster A	3-ketosteroid-9 $\alpha$ -hydroxylase	2.24E-55	10.9	1	1.3
KR76_27070	<i>kshA</i>	Cluster D	3-ketosteroid-9 $\alpha$ -hydroxylase	3.15E-59	14.4	1	1
KR76_14160	<i>kshB</i>	Cluster A	3-ketosteroid-9 $\alpha$ -hydroxylase	0.7	1.9	1	1.2
KR76_27080	<i>kshB</i>	Cluster D	3-ketosteroid-9 $\alpha$ -hydroxylase	1	3.7	1	1
KR76_02655	<i>kstD1</i>	Out of clusters	3-ketosteroid- $\Delta^1$ -dehydrogenase	1.01E-07	5.2	1	1.4
KR76_27115	<i>kstD1</i>	Cluster D	3-ketosteroid- $\Delta^1$ -dehydrogenase	1.4E-131	16.7	1	1.2
KR76_27125	<i>kstD2</i>	Cluster D	3-ketosteroid- $\Delta^1$ -dehydrogenase	1.44E-14	6.4	0	1207.7
KR76_01140	<i>kstD3</i>	Out of clusters	3-ketosteroid- $\Delta^1$ -dehydrogenase	1	0.3	1	0.3
KR76_14500	<i>kstD3</i>	Cluster A	3-ketosteroid- $\Delta^1$ -dehydrogenase	1.12E-12	6.7	1	1

<sup>a</sup>Phyt/C – expression changes on Phytosterol with respect to control

<sup>b</sup>AcC/C – expression changes on AcC with respect to control

might be carried out by the *kstD2* KR76\_27125 product. Amino acid sequence of KstD2 KR76\_27125 of *N. simplex* differ from consensus KstD2 [86] in some residues (Supplementary Fig. S3). Role of such substitution is not clear at the moment.

KstD3 of *R. ruber* preferred saturated steroid substrates (such as 5 $\alpha$ -androstane-17 $\beta$ -ol-3-one) followed by progesterone [54]; KstD3 of *Gordonia neofelifaecis* showed higher activity to 16 $\alpha$ ,17 $\alpha$ -epoxyprogesterone > AD > cholest-4-en-3-one > progesterone [69]. The expression of *kstD3* ortholog KR76\_14500 in *N. simplex* was increased on phytosterol 6.7-fold and was not increased in response to AcC.

Thus, effective 1(2)-dehydrogenation of 3-ketosteroids by *N. simplex* VKM Ac-2033D (Supplementary Table S1) is due to the presence of five *kstDs*, of which four genes are over-expressed in response to phytosterol and one (*kstD2*) is induced with AcC, but its transcription level is highest among all genes (more than 1200-fold).

Therefore, the genome-wide transcriptome analysis confirmed the transcriptional activation of the gene clusters related to the 9(10)-seco pathway of steroid catabolism in phytosterol induced *N. simplex* cells.

#### Cortisone 21-acetate conversion

Unlike phytosterol, AcC could not be utilized by *N. simplex* VKM Ac-2033D as a carbon and energy source (Fig. 4a), but was converted mainly to its 1(2)-dehydroanalog - AcP (Fig. 4b). Probably, the presence of 11-oxo group in the molecule of AcC (or 11 $\beta$ -hydroxyl group in cortisol) hinders the action of 9 $\alpha$ -hydroxylase (KshAB). These results are in accordance with the previously reported *N. simplex* activity towards cortisol and its derivatives [34, 37], and correlate with the data presented in Table 4 demonstrating quite different gene expression under phytosterol and AcC induction.

Along with the KstD activity, slight deacetylase and 20 $\beta$ -reductase activities were observed to form cortisone

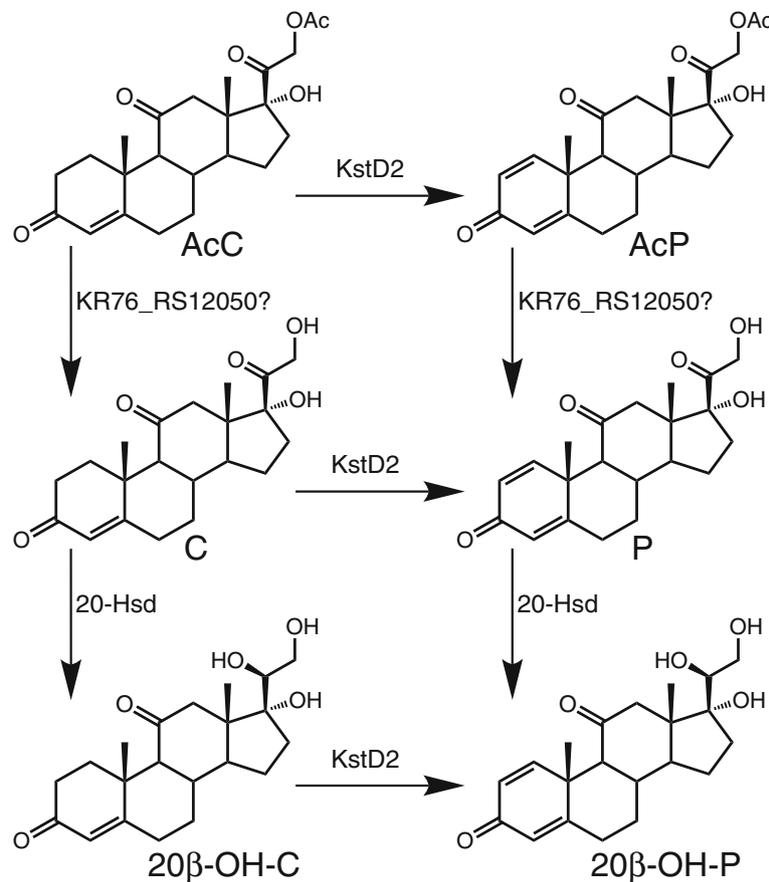
(C), prednisone (P), and 20 $\beta$ -reduced derivatives (Fig. 4b).

As shown earlier, efficiency of the 1(2)-dehydrogenation by *N. simplex* VKM Ac-2033D did not depend on the presence of acetyl group in the steroid inducer [31]. The order of deacetylation and 1(2)-dehydrogenation reactions can differ depending on the induction conditions (Fig. 8). In the AcC induced cells, 1(2)-dehydrogenation preceded deacetylation that coincides with the highest level of *kstD2* expression (Table 4), while in the absence of AcC induction, the reverse order of reactions was observed (data not shown).

Both cytosol-dissolved and membrane-associated constitutive steroid esterases were earlier reported for *N. simplex* [32]. Among the genes whose transcription is enhanced in response to AcC, no inducible esterases were detected. However, the analysis of the complete transcriptome revealed KR76\_12050 acyl-CoA thioesterase II with a high level of constitutive expression. Suggesting that the product of this gene might be responsible for 21-deacetylation we check a domain composition of the corresponding protein using InterProScan [87]. Since no membrane-bound domains were detected in this protein, the cytosolic enzymes are most likely responsible for the hydrolysis of 21-acetyl group in AcC.

Searching of the genes related to 20-reductase activity was carried out with protein BLAST. In *N. simplex* genome, three homologs of *fabG3* coding for 3 $\alpha$ ,20 $\beta$ -hydroxysteroid dehydrogenase [88] were revealed. Of these, expression of KR76\_13560 is greatly enhanced in response to AcC (13.7-fold) (Supplementary Table S4). Product of this gene is similar (amino acid sequence) by 56% with the cortisone-20 $\beta$ -reductase from *Streptomyces hydrogenans* [89].

FabG of *Bacillus megaterium* was identified as the enzyme with 20 $\beta$ -HSD activity [90] that reduce the  $\beta$ -ketoacyl group to a  $\beta$ -hydroxy group. Only one of homologs in *N. simplex* VKM Ac-2033D, namely, KR76\_



**Fig. 8** *N. simplex* VKM Ac-2033D AcC bioconversion scheme. Designations: AcP – 21-acetate of 1(2)-dehydro-cortisone or 21-acetate of prednisone; C – cortisone; P – prednisone; 20β-OH-C – 20β-hydroxy cortisone; 20β-OH-P – 20β-hydroxy prednisone

14350, increased its expression in the presence of phytoosterol (Supplementary Table S4). None of these genes was induced by AcC.

Probably, the product of the gene *KR76\_13560* may be responsible for the 20β-reducing activity of *N. simplex*. As shown earlier by Medentsev et al. [91], steroid 1(2)-dehydrogenation by *Arthrobacter grobiformis* (syn. *N. simplex*) did not depend on the level of NAD(P) H, but the resulting redox equivalents transferred to the respiratory chain to generate transmembrane potential ( $\Delta\mu\text{H}^+$ ). Hypothetically, accumulated reduced equivalents may stimulate constitutive NADH-dependent 20β-HSD activity.

#### Transcription factors binding motifs

Totally, about one third part of the operons was induced in which the predicted binding motifs for KstR and KstR2 were found earlier [27]. We have three possible explanations for such a low percentage. i) Some of the operons may be repressed by other transcription factors aside from KstR and KstR2. Even after KstR and KstR2 unbind, those other factors might continue to repress

the operons. ii) Some of the predicted KstR and KstR2 binding sites could be false positive predictions. iii) If two operons are directed outward (“← →”), the intergenic region between them is small and a KstR or KstR2 site is present there, we supposed that it regulates both operons, while in fact it would regulate one of them. The relative contributions of these three effects were not estimated in this study.

Search for motifs for the binding of candidate factors revealed several recurring motifs before operons that strengthened, or weakened their expression.

Two sites (Acin1, Acin3) were found before *KR76\_27125*, *kstD2*, with a maximum observed level of induction (Fig. 5). On the other side of these sites the transcriptional regulator gene *KR76\_27130* belonging to the TetR family is located whose expression was induced 10-fold. The increased expression of this gene and the absence of the predicted binding sites for the transcription factors KstR and KstR2 in this area suggest that *KR76\_27130* could be involved in the regulation of *kstD2* induction and that the predicted motifs belong to it. Recently, the TetR family transcriptional factor *KR76\_*

27130 was confirmed to be a steroid catabolism regulator named SRTF1 (Steroid Responsive Transcriptional Factor 1) [92]. Its binding site coincided with the predicted Acin1 motif.

Other genes annotated as transcription factors that induced by AcC, specifically *KR76\_27330* (transcriptional regulator, MerR family) and *KR76\_26400* (transcriptional regulator, PadR family), both are located in an operon with predicted Acin3 site; and *KR76\_24405* (putative transcriptional regulator) is in an operon with the predicted Acin1 site. However, the found Acin and Sitdec1 motifs are hardly enough to explain the change in the expression of all the operons upon the addition of phytosterol, or AcC. It seems that there would be another transcription factor with a small number of motifs that could not be detected by MEME tools. Post-transcriptional regulation of mRNA from these operons also cannot be fully excluded.

Nevertheless, the calculations and predictions indicate the most likely candidate genes that can serve as a target for future biotechnological research.

In the future, it is planned to expand the experiments by studying of a range of new inducers and increasing the number of replicates. New inducers will clarify the role of genes that are homologous to already known steroid catabolism genes but do not have binding motifs with its confirmed regulators. It is interesting to use as inducers substances that are intermediates at various stages of steroid catabolism, for example, androstenedione or androstendione. This will make it possible to determine the complexes of genes that are activated at later stages and to compare the catabolism of steroids in *N. simplex* VKM Ac-2033D with catabolism in more studied organisms. Knockout and proteomic experiments are also promising, confirming the role of calculated transcription factors.

## Conclusion

The gene network involved in phytosterol degradation by *N. simplex* VKM Ac-2033D is generally similar to those described for mycobacteria and related actinobacteria, but differs mainly by the existence of a cluster that has its own system of regulation.

In this study, we have generated experimental and computational evidences for the functionality of the clusters A, B, C, D those were identified by the bioinformatics analysis of the genome [27]. While the clusters A and B have been previously described for other bacteria, the clusters C and D were detected only by the analyzing of the *N. simplex* VKM Ac-2033D genome. Probably, the cluster D has its own previously unreported regulation system. We predicted a role in phytosterol catabolism for some dozens of genes of *N. simplex* VKM Ac-2033D; the

genes taking part in AcC bioconversion were also suggested.

Our analysis suggests that the bioconversion of cortisone acetate by *N. simplex* is not regulated by the KstR-repressor.

The 3-ketosteroid- $\Delta^1$ -dehydrogenase activity of *N. simplex* VKM Ac-2033D is firstly shown for the progesterone structure. The extremely high level of 1(2)-dehydrogenation of 3-ketosteroids is explained by the multiplicity and high level of induction of 3-ketosteroid- $\Delta^1$ -dehydrogenases (*kstD2* *KR76\_27125* in the case of AcC). A number of genes that can be responsible for the increased 20 $\beta$ -reducing activity have been found, however, in the case of different inducers, transcription of different genes is enhanced.

The knowledge of the genome organization, the functionality of the existing sterol catabolism genes and their possible involvement in steroid oxidation process contributes to our understanding of steroid bioconversion by actinobacteria, as well as to the expansion of the application field of the *N. simplex* VKM Ac-2033D in biotechnology.

## Materials and methods

### Reagents

Phytosterol (total sterols content – 95.47%;  $\beta$ -sitosterol – 42.39%, stigmasterol – 26.08%, campesterol – 23.48%, brassicasterol – 3.52%) was obtained from Jiangsu Spring Fruit Biological Products Co., Ltd., China. Pregn-4-ene-3,20-dione (progesterone, Pr), pregn-4-en-17 $\alpha$ -ol-3,20-dione (17 $\alpha$ -hydroxy-progesterone, 17 $\alpha$ -OH-Pr), pregn-4-ene-11 $\alpha$ ,17 $\alpha$ -diol-3,20-dione (11 $\alpha$ ,17 $\alpha$ -dihydroxy-progesterone, 11 $\alpha$ ,17 $\alpha$ -di-OH-Pr) were purchased from Sigma (USA). 21-Acetoxy-pregn-4-ene-17 $\alpha$ ,21-diol-3,11,20-trione (21-acetate of cortisone, cortisone 21-acetate, AcC) was purchased from Steraloids (USA), yeast extract from Difco (USA), soya peptone from HiMedia (India), randomly methylated  $\beta$ -cyclodextrin (MCD) from Wacker Chemie (Germany). Other materials and solvents were of analytical grade and purchased from domestic commercial suppliers.

### Microorganism

A strain of *Nocardioides simplex* VKM Ac-2033D was obtained from the All-Russian Collection of Microorganisms (VKM). For RNA extraction, *N. simplex* VKM Ac-2033D was cultured as described in [27] on the mineral medium (control) supplemented with (g/l): yeast extract - 1.0, glycerol - 5.0, MCD - 3.46, and the same medium supplemented with phytosterol (0.5 g/L), or AcC (0.2 g/L).

Cultivation was carried out aerobically on a rotary shaker (200 rpm) at 30 °C for 16 h. For steroid bioconversion experiments, the cells were harvested by

centrifuge (4 °C, 8000×g, 30 min) and washed with sterile saline as described in [31].

#### Growth estimations

Growth on phytosterol, or AcC as a sole source of carbon was estimated as described in [27] on the mineral medium supplemented with MCD 17 (g/L) (control), phytosterol (1 g/L), or AcC (1 g/L).

Because of low phytosterol and AcC solubilities, optical or dry weight measurements were inaccurate, and the biomass grown upon phytosterol plus or AcC plus conditions was assayed by protein content or CFU estimation, respectively. The protein concentration was determined by the Lowry method with preliminary alkaline hydrolysis of biomass by adding 0.9 ml of 0.5 N NaOH to a 0.1 ml sample. For colony-forming units (CFU) counting, some broth samples were serially diluted with saline under vigorous agitation and plated on the solid LB medium. The experiments were carried out in no less than three replicates.

#### Bioconversion of steroids by the resting cells, isolation and identification of steroids

Progesterone, 17 $\alpha$ -hydroxy-progesterone, and 11 $\alpha$ ,17 $\alpha$ -dihydroxy-progesterone (0.4, 1.0, and 1.0 g/l, respectively) were tested as the substrates for bioconversion by resting whole cells of *N. simplex* (0.4, 0.2 and 0.2 g/l, respectively). Biomass was prepared as described above (in 2.2). Bioconversions were carried out in 0.02 M Naphosphate buffer, pH 7.0, aerobically (200 rpm) at 30 °C in 750-ml shake flasks containing 100 ml buffered media with MCD (3.5–8.0 g/l). Concentration of  $\alpha$ , $\alpha$ -dipyridyl when added was 0.4 g/l. The experiments were carried out in triplicates.

Steroid metabolites were isolated by preparative TLC [34]. Chromatographic purity of the compounds was monitored by TLC. For TLC, DC-Fertigfolien ALU-GRAM<sup>®</sup> SIL G / UV<sub>254</sub>10 × 20 cm plates (Macherey-Nagel, Germany) were used. Steroids were developed in benzene:acetone 3:1 (v/v) or 4:1 (v/v) and visualized under UV-light (254 nm) in the chemscope CN-6 (VilberLourmat, Germany).

HPLC analyses of steroid bioconversions were performed on a 5  $\mu$ m C18 reverse phase column Symmetry 4.6 × 250 mm with a guard column Symmetry C18, 5  $\mu$ m, 3.9 × 20 mm (Waters, Ireland). Mobile phase consisted of CH<sub>3</sub>CN:H<sub>2</sub>O:CH<sub>3</sub>COOH (40:60:0.01, v/v); at 50 °C, with a flow rate of 1 ml min<sup>-1</sup> and UV detection at 240 nm were used. Peaks were compared to authentic progesterone (P), 17 $\alpha$ -hydroxy-progesterone (17 $\alpha$ -OH-P), and 11 $\alpha$ ,17 $\alpha$ -dihydroxy-progesterone (11 $\alpha$ ,17 $\alpha$ -di-OH-P) (an external standard calibration method).

The structure of some steroids was confirmed by mass spectrometry (MS) and <sup>1</sup>H-NMR spectroscopy. MS

spectra of the reported steroids were recorded on a Finnigan LCQ Advantage MAX quadrupole ion trap mass spectrometer (Thermo Electron San. Jose, USA) in positive ions [M + H]<sup>+</sup>, at evaporator temperature 350 °C, capillar – 170 °C. MS/MS spectra were obtained using normalized collision energy (Normalized Collision Energy™) ranging from 20 to 40%. Collection and processing of data were performed using the Xcalibur software. Proton nuclear magnetic resonance (<sup>1</sup>H-NMR) spectra were recorded at 400 MHz with a Bruker Avance 400 spectrometer. Chemical shifts were measured relative to a solvent signal.

#### Isolation of mRNA and high-throughput sequencing

For RNA isolation, the cells were taken in the exponential growth phase (16 h), harvested by centrifuge at 8000×g for 10 min and immediately ground in a porcelain mortar under liquid nitrogen. Total RNA was isolated using Qiagen RNeasy mini kit according to the instructions of the supplier (Qiagen, Netherlands). The resulting mixture was treated with DNase I. The ribosomal RNA was depleted using Ribo-Zero rRNA Removal Kit (Bacteria) according to the protocols of the supplier (Epicentre, USA).

Sample preparation of mRNA for high-throughput sequencing was made with TruSeq RNA Sample prep kit v.2 according to the protocols of the manufacturer (Illumina, USA). Sequencing was conducted on HiSeq 2000 (50-nucleotide single-read run) according to the protocols of the manufacturer (Illumina, USA). The experiments were carried out in duplicate.

Real-time PCR validation was performed with using the AriaMx Realtime PCR system (Agilent, Richardson, TX, USA) with an M-439 kit (Eva Green I) (Syntol, Moscow, Russian Federation). The nucleotide sequences of primers used in this study for target and reference genes are listed in Supplementary Table S6. The amplification was performed as follows: 95 °C 5 min (1 cycle), 95 °C 10 s, 54 °C 15 s, 72 °C 30 s (40 cycles). Relative gene expression levels were calculated using the double delta Cq method [93].

#### Computational analyses

The differential expression analysis was performed by Rockhopper 2.03 [94]. Rockhopper does not require the mapping of a whole read (mapping part of a read is enough). It was thus unnecessary to trim the reads before processing, and the raw reads were not trimmed. The reads were mapped by Rockhopper in the strand specific mode, with the seed length of 20 bp, allowing at most 1 mismatch between the mapped part of a read and the genome. The options “compute operons” (-y) and “identify transcript boundaries” (-t) were switched on. A gene was considered differentially expressed

between the conditions phytosterol plus and phytosterol minus or AcC plus and AcC minus, if its expression increased or decreased more than threefold with  $q$ -value less or equal to 0.01.

To check the percentage of reads corresponding to rRNA, the reads were mapped to the rRNA genes of *N. simplex* by CLC Assembly Cell 4.2 (<https://www.qiagenbioinformatics.com/products/clc-assembly-cell/>); the reads were required to map with at least 98% sequence similarity.

The operon predictions made by Rockhopper were supplemented with the operon predictions made by FgenesB [95]. *M. tuberculosis* H37Rv was used as the “closest organism” in FgenesB. If Rockhopper and FgenesB provided different predictions for a particular operon, we chose the shortest prediction.

The predictions of KstR and KstR2 binding sites were taken from [27]. The search for membrane-associated domains in the protein corresponding to the gene with the locus tag *KR76\_12050* was performed by the InterProScan [87] web server (<https://www.ebi.ac.uk/interpro/interproscan.html>) on August 5, 2018.

Alignments of the KshA, KshB and KstD proteins were done with CLC Genomics Workbench, v. 7.5.1.

Genes whose products could play role in the biochemical reactions were predicted based on the annotation, homology with the known enzymes (using orthogroups), gene expression changes, location and predicted regions to which the orthologs belong.

Searching of the genes related to 20-reductase activity was carried out with protein BLAST for *fabG3* coding for 3 $\alpha$ ,20 $\beta$ -hydroxysteroid dehydrogenase [88] and *fabG* of *Bacillus megaterium* [90].

### Search of transcription factors binding motifs

To find out binding motifs of transcription factors that may be responsible for the change in operon expression, the following protocol was used:

1. Regions spanning from 500 bps upstream to 50 bps downstream of start codons of the first genes in the operons that changed their expression were taken.
2. In these regions, potential binding motifs were de novo predicted by MEME 4.10.0 [96]. The minimum allowed motif width was set to 8 bps, the maximum to 50 bps. Any number of motifs was allowed in each of the regions (the option “-mod anr”). Only motifs with  $e$ -value, less or equal to 0.1 were considered further.
3. Potentially, the found motifs may correspond to low complexity sequences, or promoters. To identify and discard such motifs, we took the regions spanning from 500 bps upstream to 50 bps downstream of start codons of all *N. simplex* genes

and searched for the motifs corresponding to all the motifs found in the previous paragraph. The search was done by the tool FIMO from the MEME 4.10.0 suite [53] with the  $q$ -value threshold for reported motifs of 0.05. If a motif had 1000 sites or more, it was deemed unrelated to a transcription factor.

4. If a motif had less than 1000 sites, it was considered potentially corresponding to some transcription factor.

This protocol was applied separately three times: for the operons that (i) increased, or (ii) decreased their expression in AcC conditions, and (iii) decreased their expression in phytosterol conditions.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12896-021-00668-9>.

**Additional file 1: Supplementary Table S1.** Steroid bioconversion by *N. simplex* VKM Ac-2033D. **Supplementary Table S2.** MS-characteristics of the steroid substrates and *N. simplex* bioconversion products. **Supplementary Table S3.** <sup>1</sup>H-NMR spectra. **Supplementary Table S5.** New candidate motifs for the binding of transcription factors for steroid catabolism regulation in *N. simplex* VKM Ac-2033D. **Supplementary Table S6.** Real-time qPCR.

**Additional file 2: Supplementary Table S4.** DEGs and operons.

**Additional file 3: Supplementary Figure S1-S3.** Alignments of KshA, KshB, KstD [pdf]. **Supplementary Figure S1.** KshA alignment. **Supplementary Figure S2.** KshB alignment. **Supplementary Figure S3.** KstD alignment.

### Abbreviations

MCD: methyl- $\beta$ -cyclodextrin; AcC: cortisone 21-acetate; KstD: 3-ketosteroid- $\Delta^1$ -dehydrogenase; DEG: differentially expressed gene(s)

### Acknowledgements

This work has been performed in the frames of ERA CoBioTech program on the Project: Synthetic biology for industrial production of steroids (Acronym: Syntheroids) and was supported exclusively by the Ministry of Science and Higher Education of Russian Federation (unique project identifier RFME F158818X0008). Authors are very grateful to Maria Logacheva (Moscow State University, Moscow) for the high-throughput sequencing.

### Authors' contributions

VS designed research, made sequencing libraries, analyzed the data; MS provided bioinformatics analysis; VF performed microbiological works; MD administrated and coordinated the project; EB isolated RNA; VF and AS conducted biochemical and chemical analyses; VS, MS, VF, MD wrote the manuscript; EB, VF and AS draw figures. All authors read and approved the manuscript.

### Funding

ERA CoBioTech program on the Project: Synthetic biology for industrial production of steroids (Acronym: Syntheroids); Ministry of Science and Higher Education of Russian Federation (unique project identifier RFME F158818X0008).

### Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files as Supplementary Table S3. Reads are in <https://www.ncbi.nlm.nih.gov/sra/SRX5542009>, <https://www.ncbi.nlm.nih.gov/sra/SRX5542008>, <https://www.ncbi.nlm.nih.gov/sra/SRX5542005>, <https://www.ncbi.nlm.nih.gov/sra/SRX5542004>, <https://www.ncbi.nlm.nih.gov/sra/SRX5542007>, <https://www.ncbi.nlm.nih.gov/sra/SRX5542006>.

**Ethics approval and consent to participate**

not applicable.

**Consent for publication**

not applicable.

**Competing interests**

The authors declare that they have no conflict of interest.

**Author details**

<sup>1</sup>Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Leninskie gory, h. 1, b. 40, Moscow, Russian Federation 119991.

<sup>2</sup>Skolkovo Institute of Science and Technology, Nobelya str., 3, Moscow, Russian Federation 121205. <sup>3</sup>Institute for Information Transmission Problems, Russian Academy of Sciences, Bolshoy Karetny per., h. 19, b. 1, Moscow, Russian Federation 127994. <sup>4</sup>G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Federal Research Center "Pushchino Center for Biological Research of the Russian Academy of Sciences", pr. Nauki, 5, Pushchino, Moscow Region, Russian Federation 142290. <sup>5</sup>Pharmins, Ltd., R&D, Institutskaya str, 4, Pushchino, Moscow Region, Russian Federation 142290.

Received: 18 September 2020 Accepted: 14 December 2020

Published online: 13 January 2021

**References**

- Kendall SL, Withers M, Soffair CN, Moreland NJ, Gurcha S, Sidders B, et al. 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.* 2007;65:684–99.
- Kendall SL, Burgess P, Balhana R, Withers M, ten Bokum A, Lott JS, et al. Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2. *Microbiology.* 2010;156:1362–71.
- Uhía I, Galán B, Medrano FJ, García JL. Characterization of the KstR-dependent promoter of the gene for the first step of the cholesterol degradative pathway in *Mycobacterium smegmatis*. *Microbiology.* 2011;157:2670–80.
- Uhía I, Galán B, Kendall SL, Stoker NG, García JL. Cholesterol metabolism in *Mycobacterium smegmatis*: cholesterol pathway. *Environ Microbiol Rep.* 2012;4:168–82.
- Griffin JE, Gawronski JD, Dejesus MA, loerger TR, Akerley BJ, Sassetti CM. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *Plos Pathog.* 2011;7:e1002251.
- van der Geize R, Yam K, Heuser T, Wilbrink MH, Hara H, Anderton MC, et al. 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.* 2007;104:1947–52.
- McLeod MP, Warren RL, Hsiao WWL, Araki N, Myhre M, Fernandes C, et al. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A.* 2006;103:15582–7.
- Drzyzga O, Fernández de las Heras L, Morales V, Navarro Llorens JM, Perera J. Cholesterol degradation by *Gordonia cholesterolivorans*. *Appl Environ Microbiol.* 2011;77:4802–10.
- Ivashina TV, Nikolayeva VM, Dovbnya DV, Donova MV. 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.* 2012;129:47–53.
- Capyk JK, Kalscheuer R, Stewart GR, Liu J, Kwon H, Zhao R, et al. Mycobacterial cytochrome P450 125 (Cyp125) catalyzes the terminal hydroxylation of C27 steroids. *J Biol Chem.* 2009;284:35534–42.
- Johnston JB, Ouellet H, Ortiz de Montellano PR, et al. *J Biol Chem.* 2010;285:36352–60.
- Wilbrink MH, van der Geize R, Dijkhuizen L. Molecular characterization of *Itp3* and *Itp4*, essential for C24-branched chain sterol-side-chain degradation in *Rhodococcus rhodochrous* DSM 43269. *Microbiology.* 2012;158(Pt\_12):3054–62.
- Casabon I, Swain K, Crowe AM, Eltis LD, Mohn WW. Actinobacterial acyl coenzyme a synthetases involved in steroid side-chain catabolism. *J Bacteriol.* 2014;196:579–87.
- Nesbitt NM, Yang X, Fontán P, Kolesnikova I, Smith I, Sampson NS, et al. A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol. *Infect Immun.* 2010;78:275–82.
- Capyk JK, Casabon I, Gruninger R, Strynadka NC, Eltis LD. Activity of 3-ketosteroid 9 $\alpha$ -hydroxylase (KshAB) indicates cholesterol side chain and ring degradation occur simultaneously in *Mycobacterium tuberculosis*. *J Biol Chem.* 2011;286:40717–24.
- Bragin EY, Shtratnikova VY, Schelkunov MI, Dovbnya DV, Donova MV. Genome-wide response on phytosterol in 9-hydroxyandrostenedione-producing strain of *Mycobacterium* sp. VKM ac-1817D. *BMC Biotechnol.* 2019;19:39.
- Carere J, McKenna SE, Kimber MS, Seah SYK. Characterization of an aldolase-dehydrogenase complex from the cholesterol degradation pathway of *Mycobacterium tuberculosis*. *Biochemistry.* 2013;52:3502–11.
- Dresen C, Lin LY-C, D'Angelo I, Tocheva EI, Strynadka N, Eltis LD. A flavin-dependent monooxygenase from *Mycobacterium tuberculosis* involved in cholesterol catabolism. *J Biol Chem.* 2010;285:22264–75.
- Lack NA, Yam KC, Lowe ED, Horsman GP, Owen RL, Sim E, et al. Characterization of a carbon-carbon hydrolase from *Mycobacterium tuberculosis* involved in cholesterol metabolism. *J Biol Chem.* 2010;285:434–43.
- Crowe AM, Casabon I, Brown KL, Liu J, Lian J, Rogalski JC, et al. Catabolism of the last two steroid rings in *Mycobacterium tuberculosis* and other bacteria. *mBio.* 2017;8. <https://doi.org/10.1128/mBio.00321-17>.
- Nagasawa M, Bae M, Tamura G, Arima K. Microbial transformation of sterols. Part II: cleavage of sterol side chains by microorganisms. *Agric Biol Chem.* 1969;33:1644–50.
- Choi K-P, Murooka Y, Molnár I. Secretory overproduction of *Arthrobacter simplex* 3-ketosteroid-delta-1-dehydrogenase by *Streptomyces lividans* with a multi-copy shuttle vector. *Appl Microbiol Biotechnol.* 1995;43:1044–9.
- Zhang H, Tian Y, Wang J, Li Y, Wang H, Mao S, et al. Construction of engineered *Arthrobacter simplex* with improved performance for cortisone acetate biotransformation. *Appl Microbiol Biotechnol.* 2013;97:9503–14.
- Holert J, Cardenas E, Bergstrand LH, Zaikova E, Hahn AS, Hallam SJ, et al. Metagenomes reveal global distribution of bacterial steroid catabolism in natural, engineered, and host environments. *mBio.* 2018;9:e02345–17.
- Kwak MJ, Kwon SK, Kim JF. Complete genome sequence of the sand-sediment actinobacterium *Nocardioides dokdonensis* FR1436<sup>T</sup>. *Stand Genomic Sci.* 2017;12. <https://doi.org/10.1186/s40793-017-0257-z>.
- Shtratnikova VY, Schelkunov MI, Pekov YA, Fokina VV, Logacheva MD, Sokolov SL, et al. Complete genome sequence of steroid-transforming *Nocardioides simplex* VKM Ac-2033D. *Genome Announc.* 2015;3:e01406–14.
- Shtratnikova VY, Schelkunov MI, Fokina VV, Pekov YA, Ivashina T, Donova MV. Genome-wide bioinformatics analysis of steroid metabolism-associated genes in *Nocardioides simplex* VKM Ac-2033D. *Curr Genet.* 2016;62:643–56.
- Shtratnikova VY, Schelkunov MI, Fokina VV, Bragin EY, Lobostova TG, Shutov AA, et al. Genome-wide transcriptome profiling provides insight on cholesterol and lithocholate degradation mechanisms in *Nocardioides simplex* VKM Ac-2033D. *Genes.* 2020;11:1229.
- Qin H-M, Wang J-W, Guo Q, Li S, Xu P, Zhu Z, et al. Refolding of a novel cholesterol oxidase from *Pimelobacter simplex* reveals dehydrogenation activity. *Protein Expr Purif.* 2017;139:1–7.
- Qin H-M, Zhu Z, Ma Z, Xu P, Guo Q, Li S, et al. Rational design of cholesterol oxidase for efficient bioresolution of the cholestane skeleton substrates. *Sci Rep.* 2017;7. <https://doi.org/10.1038/s41598-017-16768-6>.
- Fokina VV, Sukhodolskaya GV, Baskunov BP, Turchin KF, Grinenko GS, Donova MV. Microbial conversion of pregna-4,9(11)-diene-17 $\alpha$ ,21-diol-3,20-dione acetates by *Nocardioides simplex* VKM Ac-2033D. *Steroids.* 2003;68:415–21.
- Fokina VV, Donova MV. 21-Acetoxy-pregna-4(5),9(11),16(17)-triene-21-ol-3,20-dione conversion by *Nocardioides simplex* VKM Ac-2033D. *J Steroid Biochem Mol Biol.* 2003;87:319–25.
- Fokina VV, Sukhodolskaya GV, Gulevska SA, Gavrish EY, Evtushenko LI, Donova MV. The 1(2)-dehydrogenation of steroid substrates by *Nocardioides simplex* VKM Ac-2033D. *Microbiology.* 2003;72:24–9.
- Sukhodolskaya G, Fokina V, Shutov A, Nikolayeva V, Savinova T, Grishin Y, et al. Bioconversion of 6-(N-methyl-N-phenyl) aminomethyl androstane steroids by *Nocardioides simplex*. *Steroids.* 2017;118:9–16.
- Mutafova B, Mutafov S. Microbial transformations of plant origin compounds as a step in preparation of highly valuable pharmaceuticals. *J Drug Metab Toxicol.* 2016;7:204–15.

36. Costa S, Zappaterra F, Summa D, Semeraro B, Fantin G.  $\Delta 1$ -dehydrogenation and C20 reduction of cortisone and hydrocortisone catalyzed by *Rhodococcus* strains. *Molecules*. 2020;25:2192.
37. Sukhodolskaya GV, Donova MV, Nikolaeva VM, Koshcheyenko KA, Dovbnya DV, Khomutov SM, et al. Method of the producing 1(2)-dehydroderivatives of 4-delta-3-ketosteroids. RU Patent 2156302. 2000.
38. Fokina VV, Karpov AV, Sidorov IA, Andriushina VA, Arinbasarova AY. The influence of  $\beta$ -cyclodextrin on the kinetics of 1-en-dehydrogenation of 6 $\alpha$ -methylhydrocortisone by *Arthrobacter globiformis* cells. *Appl Microbiol Biotechnol*. 1997;47:645–9.
39. Sukhodolskaja GV, Savinova TS, Fokina VV, Shutov AA, Nikolaeva VM, Lukashev NV, et al. Microbiological method of producing 1,2-dehydrogenated derivatives of delta-4-3-keto-steroids of androstane family in aqueous organic media. RU Patent 2447154. 2012.
40. Lobastova TG, Khomutov SM, Shutov AA, Donova MV. Microbiological synthesis of stereoisomeric 7( $\alpha/\beta$ )-hydroxytestolactones and 7( $\alpha/\beta$ )-hydroxytestolactones. *Appl Microbiol Biotechnol*. 2019;103:4967–76.
41. Zvyagintseva IS, Skryabin GK. Dehydrogenation of steroids by mycobacteria. *Biol Bull Acad Sci USSR*. 1964;4:525–32.
42. Krassilnikov N, Skryabin G, Aseeva I, Korsunskaya L. The 1,2-dehydrogenation of hydrocortisone by *Mycobacterium* sp. 193 cells. *Dokl Biol Sci Sect*. 1959; 128:1063–5.
43. Lestrovaya N, Nazaruk M, Skryabin G. 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*. 1965;163:768–70.
44. Arinbasarova AY, Koshcheyenko KA. Covalent binding of cells with activated silica gel. *Prikl Biokhim Mikrobiol*. 1980;16:854–61.
45. Prauser H. *Nocardioidea*, a new genus of the order *Actinomycetales*. *Int J Syst Bacteriol*. 1976;26:58–65.
46. O'Donnell AG, Goodfellow M, Minnikin DE. Lipids in the classification of *Nocardioidea*: reclassification of *Arthrobacter simplex* (Jensen) Lochhead in the genus *Nocardioidea* (Prauser) emend. O'Donnell et al. as *Nocardioidea simplex* comb. nov. *Arch Microbiol*. 1982;133:323–9.
47. Goodfellow M, Whitman WB, Bergey DH. *Bergey's manual of systematic bacteriology*. Vol. 5. The Actinobacteria. 2. Ed. New York: Springer; 2012.
48. Suzuki K-I, Komagata K. *Pimelobacter* gen. Nov., a new genus of coryneform bacteria with LL-diaminopimelic acid in the cell wall. *J Gen Appl Microbiol*. 1983;29:59–71.
49. Collins MD, Dorsch M, Stackebrandt E. Transfer of *Pimelobacter tumescens* to *Terrabacter* gen. Nov. as *Terrabacter tumescens* comb. nov. and of *Pimelobacter jensenii* to *Nocardioidea* as *Nocardioidea jensenii* comb. nov. *Int J Syst Bacteriol*. 1989;39:1–6.
50. Yoon J-H, Park Y-H. The genus *Nocardioidea*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. *The prokaryotes*. New York, NY: Springer New York; 2006. p. 1099–113. [https://doi.org/10.1007/0-387-30743-5\\_44](https://doi.org/10.1007/0-387-30743-5_44).
51. Rohman A, Dijkstra BW. The role and mechanism of microbial 3-ketosteroid  $\Delta 1$ -dehydrogenases in steroid breakdown. *J Steroid Biochem Mol Biol*. 2019; 191:105366.
52. Catroux G, Fournier J-C, Blachère H. Importance de la forme cristalline de l'acétate de cortisone pour sa déshydrogénation en C-1 par *Arthrobacter simplex*. *Can J Biochem*. 1968;46:537–42.
53. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinforma Oxf Engl*. 2011;27:1017–8.
54. Guevara G, Fernández de las Heras L, Perera J, Navarro Llorens JM. Functional differentiation of 3-ketosteroid  $\Delta 1$ -dehydrogenase isozymes in *Rhodococcus ruber* strain Chol-4. *Microb Cell Factories*. 2017;16. <https://doi.org/10.1186/s12934-017-0657-1>.
55. Bergstrand LH, Cardenas E, Holert J, Van Hamme JD, Mohn WW. Delineation of steroid-degrading microorganisms through comparative genomic analysis. *mBio*. 2016;7:e00166.
56. Devi S, Kanwar SS. Cholesterol oxidase: source, properties and applications. *Insights Enzyme Res*. 2017;01:5–17.
57. Cho H-S, Choi G, Choi KY, Oh B-H. Crystal structure and enzyme mechanism of  $\Delta^5$ -3-ketosteroid isomerase from *Pseudomonas testosteroni*. *Biochemistry*. 1998;37:8325–30.
58. García-Fernández E, Medrano FJ, Galán B, García JL. Deciphering the transcriptional regulation of cholesterol catabolic pathway in *Mycobacteria*: identification of the inducer of KstR repressor. *J Biol Chem*. 2014;289:17576–88.
59. Wilbrink MH, Petrusma M, Dijkhuizen L, van der Geize R. FadD19 of *Rhodococcus rhodochrous* DSM43269, a steroid-coenzyme a ligase essential for degradation of C-24 branched sterol side chains. *Appl Environ Microbiol*. 2011;77:4455–64.
60. Yang M, Lu R, Guja KE, Wiperman MF, St Clair JR, Bonds AC, et al. Unraveling cholesterol catabolism in *Mycobacterium tuberculosis*: ChsE4-ChsE5  $\alpha 2\beta 2$  acyl-CoA dehydrogenase initiates  $\beta$ -oxidation of 3-oxo-cholesterol-4-en-26-oyl CoA. *ACS Infect Dis*. 2015;1:110–25.
61. Mohn WW, Wilbrink MH, Casabon I, Stewart GR, Liu J, van der Geize R, et al. Gene cluster encoding cholate catabolism in *Rhodococcus* spp. *J Bacteriol*. 2012;194:6712–9.
62. Ruprecht A, Maddox J, Stirling AJ, Visaggio N, Seah SYK. Characterization of novel acyl coenzyme a dehydrogenases involved in bacterial steroid degradation. *J Bacteriol*. 2015;197:1360–7.
63. Wiperman MF, Sampson NS, Thomas ST. Pathogen roid rage: cholesterol utilization by *Mycobacterium tuberculosis*. *Crit Rev Biochem Mol Biol*. 2014; 49:269–93.
64. Schaefer CM, Lu R, Nesbitt NM, Schiebel J, Sampson NS, Kisker C. FadA5 a thiolase from *Mycobacterium tuberculosis*: a steroid-binding pocket reveals the potential for drug development against tuberculosis. *Structure*. 2015;23: 21–33.
65. Thomas ST, Sampson NS. *Mycobacterium tuberculosis* utilizes a unique heterotetrameric structure for dehydrogenation of the cholesterol side chain. *Biochemistry*. 2013;52:2895–904.
66. Thomas ST, VanderVen BC, Sherman DR, Russell DG, Sampson NS. Pathway profiling in *Mycobacterium tuberculosis*: elucidation of cholesterol-derived catabolite and enzymes that catalyze its metabolism. *J Biol Chem*. 2011;286: 43668–78.
67. Yang M, Guja KE, Thomas ST, Garcia-Diaz M, Sampson NS. A distinct MaoC-like enoyl-CoA hydratase architecture mediates cholesterol catabolism in *Mycobacterium tuberculosis*. *ACS Chem Biol*. 2014;9:2632–45.
68. Horinouchi M, Hayashi T, Kudo T. Steroid degradation in *Comamonas testosteroni*. *J Steroid Biochem Mol Biol*. 2012;129:4–14.
69. Zhang Q, Ren Y, He J, Cheng S, Yuan J, Ge F, et al. Multiplicity of 3-ketosteroid  $\Delta 1$ -dehydrogenase enzymes in *Gordonia neofelifaecis* NRRL B-59395 with preferences for different steroids. *Ann Microbiol*. 2015;65:1961–71.
70. Horinouchi M, Kurita T, Yamamoto T, Hatori E, Hayashi T, Kudo T. 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*. 2004;324:597–604.
71. Capyk JK, D'Angelo I, Strynadka NC, Eltis LD. Characterization of 3-ketosteroid 9 $\alpha$ -hydroxylase, a Rieske oxygenase in the cholesterol degradation pathway of *Mycobacterium tuberculosis*. *J Biol Chem*. 2009;284:9937–46.
72. Horinouchi M, Koshino H, Malon M, Hirota H, Hayashi T. Steroid degradation in *Comamonas testosteroni* TA441: identification of metabolites and the genes involved in the reactions necessary before D-ring cleavage. *Appl Environ Microbiol*. 2018;84. <https://doi.org/10.1128/AEM.01324-18>.
73. Yam KC, D'Angelo I, Kalscheuer R, Zhu H, Wang J-X, Snieckus V, et al. Studies of a ring-cleaving dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of *Mycobacterium tuberculosis*. *PLoS Pathog*. 2009;5:e1000344.
74. Horinouchi M, Hayashi T, Koshino H, Kurita T, Kudo T. Identification of 9,17-dioxo-1,2,3,4,10,19-hexanorandrostane-5-oic acid, 4-hydroxy-2-oxohexanoic acid, and 2-hydroxyhexa-2,4-dienoic acid and related enzymes involved in testosterone degradation in *Comamonas testosteroni* TA441. *Appl Environ Microbiol*. 2005;71:5275–81.
75. Casabon I, Crowe AM, Liu J, Eltis LD. 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*. 2013;87:269–83.
76. Horinouchi M, Koshino H, Malon M, Hirota H, Hayashi T. Steroid degradation in *Comamonas testosteroni* TA441: identification of the entire  $\beta$ -oxidation cycle of the cleaved B ring. *Appl Environ Microbiol*. 2019;85. <https://doi.org/10.1128/AEM.01204-19>.
77. van der Geize R, Grommen AWF, Hessel G, Jacobs AAC, Dijkhuizen L. The steroid catabolic pathway of the intracellular pathogen *Rhodococcus equi* is important for pathogenesis and a target for vaccine development. *Plos Pathog*. 2011;7:e1002181.
78. Wiperman MF, Yang M, Thomas ST, Sampson NS. 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*. 2013;195:4331–41.

79. Fujimoto Y, Chen CS, Gopalan AS, Sih CJ. 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.* 1982;104:4720–2.
80. Casabon I, Zhu S-H, Otani H, Liu J, Mohn WW, Eltis LD. Regulation of the KstR2 regulon of *Mycobacterium tuberculosis* by a cholesterol catabolite. *Mol Microbiol.* 2013;89:1201–12.
81. Dovbnya DV, Egorova OV, Donova MV. Microbial side-chain degradation of ergosterol and its 3-substituted derivatives: a new route for obtaining of deltanoids. *Steroids.* 2010;75:653–8.
82. Itagaki E, Matushita H, Hatta T. Steroid transhydrogenase activity of 3-ketosteroid-delta 1-dehydrogenase from *Nocardia corallina*. *J Biochem (Tokyo).* 1990;108:122–7.
83. Petrusma M, Hessels G, Dijkhuizen L, van der Geize R. Multiplicity of 3-ketosteroid-9 $\alpha$ -hydroxylase enzymes in *Rhodococcus rhodochrous* DSM43269 for specific degradation of different classes of steroids. *J Bacteriol.* 2011;193:3931–40.
84. Bragin EY, Shtratnikova VY, Dovbnya DV, Schelkunov MI, Pekov YA, Malakho SG, et al. Comparative analysis of genes encoding key steroid core oxidation enzymes in fast-growing *Mycobacterium* spp. strains. *J Steroid Biochem Mol Biol.* 2013;138:41–53.
85. van der Geize R, Hessels GI, van Gerwen R, van der Meijden P, Dijkhuizen L. Molecular and functional characterization of kshA and kshB, encoding two components of 3-ketosteroid 9 $\alpha$ -hydroxylase, a class IA monooxygenase, in *Rhodococcus erythropolis* strain SQ1. *Mol Microbiol.* 2002;45:1007–18.
86. van der Geize R, Hessels GI, Dijkhuizen L. Molecular and functional characterization of the *kstD2* gene of *Rhodococcus erythropolis* SQ1 encoding a second 3-ketosteroid  $\Delta$ 1-dehydrogenase isoenzyme. *Microbiology.* 2002;148:3285–92.
87. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30:1236–40.
88. Yang JK, Park MS, Waldo GS, Suh SW. Directed evolution approach to a structural genomics project: Rv2002 from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci.* 2003;100:455–60.
89. Ghosh D, Wawrzak Z, Weeks CM, Duax WL, Erman M. The refined three-dimensional structure of 3 $\alpha$ ,20 $\beta$ -hydroxysteroid dehydrogenase and possible roles of the residues conserved in short-chain dehydrogenases. *Structure.* 1994;2:629–40.
90. Gerber A, Milhim M, Hartz P, Zapp J, Bernhardt R. Genetic engineering of *Bacillus megaterium* for high-yield production of the major teleost progestogens 17 $\alpha$ ,20 $\beta$ -di- and 17 $\alpha$ ,20 $\beta$ ,21 $\alpha$ -trihydroxy-4-pregnen-3-one. *Metab Eng.* 2016;36:19–27.
91. Medentsev AG, Arinbasarova AY, Koshcheyenko KA, Akimenko VK, Skryabin GK. Regulation of 3-ketosteroid-1-en-dehydrogenase activity of *Arthrobacter globiformis* cells by a respiratory chain. *J Steroid Biochem.* 1985;23:365–8.
92. Grazon C, Baer RC, Kuzmanović U, Nguyen T, Chen M, Zamani M, et al. A progesterone biosensor derived from microbial screening. *Nat Commun.* 2020;11. <https://doi.org/10.1038/s41467-020-14942-5>.
93. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-Delta Delta C(T))</sup> method. *Methods San Diego Calif.* 2001;25:402–8.
94. McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA, et al. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* 2013;41:e140.
95. Solovyev V, Salamov A. Automatic annotation of microbial genomes and metagenomic sequences. In: *Metagenomics and its applications in agriculture, biomedicine and environmental studies.* Hauppauge: Nova Science Publisher's; 2011. p. 61–78.
96. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* 1994;2:28–36.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

