Proteome, bioinformatic and functional analyses reveal a distinct and conserved metabolic pathway for bile salt degradation in the *Sphingomonadaceae*

Franziska M. Feller[1], Lars Wöhlbrand[2], Johannes Holert[1], Vanessa Schnaars[2], Lea Elsner[1#], William W. Mohn[3], Ralf Rabus[2], and Bodo Philipp[1,4*]

[1] Microbial Biotechnology and Ecology, Institute for Molecular Microbiology and Biotechnology, Westfälische Wilhelms-Universität Münster, Münster, Germany

[2]General and Molecular Microbiology, Institute for Chemistry and Biology of the Marine Environment (ICBM), Carl von Ossietzky University of Oldenburg, Oldenburg, Germany

[3]Department of Microbiology and Immunology, Life Sciences Institute, University of British Columbia, Vancouver, Canada

[4]Applied Ecology and Bioresources, Fraunhofer-Institute for Molecular and Applied Ecology IME, Schmallenberg, Germany

[#]Current address: Medizinisches Labor Wahl, Lüdenscheid, Germany

*Corresponding author: Bodo Philipp

E-Mail: bodo.philipp@uni-muenster.de; Tel.: (+49) 251 8339827; Fax: (+49) 251 8338388; Adress: Institut für Molekulare Mikrobiologie und Biotechnologie, Westfälische Wilhelms-Universität Münster, Corrensstr. 3, 48149 Münster, Germany.

Running title:

Bile salt degradation by *Sphingomonadaceae*

Abstract

Bile salts are amphiphilic steroids with a C_5 carboxylic side chain with digestive functions in vertebrates. Upon excretion, they are degraded by environmental bacteria. Degradation of the bile-salt steroid skeleton resembles the well-studied pathway for other steroids like testosterone, while specific differences occur during side-chain degradation and the initiating transformations of the steroid skeleton. Of the latter, two variants via either $\Delta^{1,4}$- or $\Delta^{4,6}$-3-ketostructures of the steroid skeleton exist for 7-hydroxy bile salts. While the $\Delta^{1,4}$- variant is well-known from many model organisms, the $\Delta^{4,6}$-variant involving a 7-hydroxysteroid dehydratase as key enzyme has not been systematically studied. Here, combined proteomic, bioinformatic and functional analyses of the $\Delta^{4,6}$-variant in *Sphingobium* sp. strain Chol11 were performed. They revealed a degradation of the steroid rings similar to the $\Delta^{1,4}$-variant except for the elimination of the 7-OH as key difference. In contrast, differential production of the respective proteins revealed a putative gene cluster for side-chain degradation encoding a CoA-ligase, an acyl-CoA dehydrogenase, a Rieske monooxygenase, and an amidase, but lacking most canonical genes known from other steroid-degrading bacteria. Bioinformatic analyses predicted the $\Delta^{4,6}$-variant to be widespread among the *Sphingomonadaceae*, which was verified for three type strains which also have the predicted side-chain degradation cluster. A second amidase in the side-chain degradation gene cluster of strain Chol11 was shown to cleave conjugated bile salts while having low similarity to known bile-salt hydrolases. This study signifies members of the *Sphingomonadaceae* remarkably well-adapted to the utilization of bile salts via a partially distinct metabolic pathway.

Importance

This study highlights the biochemical diversity of bacterial degradation of steroid compounds, in particular bile salts. Furthermore, it substantiates and advances knowledge of a variant pathway for degradation of steroids by sphingomonads, a group of environmental bacteria that are well-known for their broad metabolic capabilities. Biodegradation of bile salts is a critical process due to the high input of these compounds from manure into agricultural soils and wastewater treatment plants. In addition, these results may also be relevant for the biotechnological production of bile salts or other steroid compounds with pharmaceutical functions.

Introduction

Bile salts are multifunctional steroidal compounds that act as detergents in the digestion of lipophilic nutrients and exhibit signaling function in vertebrates (1, 2). The amphiphilic character of mammalian bile salts is determined by a carboxylic C_5 side chain at C17 and one to three hydroxy groups on the steroid nucleus. Bile salts are produced from cholesterol in the liver, conjugated to taurine or glycine via amide bonds and excreted into the gastrointestinal tract. In the intestine, free bile salts are released by deconjugation catalyzed by bile-salt hydrolases produced by intestinal bacteria (3). Although most bile salts are reabsorbed (4, 5), about 0.4 – 0.6 g of bile salts are excreted per day by each human (6), adding up to about 18 t of bile salts excreted per year by the population of a city with 100,000 inhabitants.

Upon excretion, bile salts become an energy and carbon source for environmental bacteria (7, 8) and several bile-salt degrading bacteria have been isolated from soils and aquatic habitats (9–12). These include *Rhodococcus jostii* RHA1 (13), *Comamonas testosteroni* CNB-2 and TA441 (14), *Pseudomonas stutzeri* Chol1 (9), *Pseudomonas* sp. strain DOC21 (11), *Azoarcus* sp. strain Aa7 (12), and *Sphingobium* sp. strain Chol11, formerly *Novosphingobium* sp. strain Chol11 (10). Aerobic bile salt degradation proceeds similar to the degradation of other steroids such as cholesterol and can be divided into different phases (Fig 1) (7, 8, 15): 1) Oxidation of the A-ring, 2) side-chain degradation, 3) oxygenolytic cleavage of ring B, and 4) oxygenolytic and hydrolytic degradation of the remaining *seco*-steroid. The first three steps may occur simultaneously (16–18).

In *R. jostii* RHA1, *C. testosteroni* TA441, and *P. stutzeri* Chol1, bile-salt degradation proceeds through the well-elucidated 9,10-*seco* pathway. In phase 1, oxidative reactions

at the A-ring generate intermediates with a $\Delta^{1,4}$-3-keto structure of the steroid skeleton (7, 13, 14). During degradation of the trihydroxy bile-salt model-substrate cholate (I in Fig 1), this leads to formation of $\Delta^{1,4}$-3-ketocholate (III) (16). In phase 2, the bile-salt side-chain is degraded by the successive release of acetyl-CoA and propionyl-CoA (16, 19, 20). In actinobacteria such as *R. jostii* RHA1 acetyl-CoA is predicted to be released by β-oxidation (13). In proteobacteria such as *P. stutzeri* Chol1 acetyl-CoA is released by an aldolase-mediated cleavage reaction and subsequent oxidation of the resulting aldehyde group (19, 21). Both mechanisms of side-chain degradation result in intermediates with C_3 carboxylic side chains (16, 17, 21, 22). In actino- as well as proteobacteria, this C_3 side chain is released as propionyl-CoA by a second cycle of aldolytic cleavage reactions (22–24) resulting in C_{19} steroids, so-called androsta-1,4-diene-3,17-diones (ADDs); in the case of cholate, 7,12β-dihydroxy-ADD (12β-DHADD, IV) is formed (9, 13, 20).

In phase 3, degradation of the steroidal ring system is initiated by the introduction of a hydroxy group at C9 by the monooxygenase KshAB (17), which leads to spontaneous opening of the B-ring driven by the aromatization of ring A. This produces 9,10-*seco* intermediates such as 3,7,12-trihydroxy-9,10-*seco*-androsta-1,3,5-triene-9,17-dione (THSATD, V). Phase 4 starts with the *meta*-cleavage of the aromatic A-ring and hydrolytic cleavage of the former A-ring, which results in *H*-methyl-hexahydro-indanone-propanoate (HIP, VI) intermediates (7, 8). At this stage of degradation, intermediates from differently hydroxylated bile salts are channeled into a common pathway in *P. stutzeri* Chol1 (23). In this process, the former 12-OH is removed and a hydroxy group at former C7 is introduced into 7-deoxy-bile salt derivatives during β-oxidation of the former

B-ring. Further degradation of HIPs proceeds via β-oxidation of the former B-ring and hydrolytic cleavages of rings C and D (25, 26).

In contrast to this well elucidated pathway, degradation of 7-hydroxy-bile salts such as cholate (I in Fig 1) proceeds differently in *Sphingobium* sp. strain Chol11 (10), but can also be divided into the four phases. After the initial formation of Δ^4-3-keto-intermediates such as Δ^4-3-ketocholate (II in Fig 1) in phase 1, the hydroxy group at C7 is eliminated by the 7α-hydroxy steroid dehydratase Hsh2 (27). This leads to the formation of a double bond in the B-ring and to $\Delta^{4,6}$-intermediates such as 12-hydroxy-3-oxo-chol-4,6-dienoate (HOCDA, VII) (10). This variant of the pathway will be referred to as $\Delta^{4,6}$-variant, in contrast to the $\Delta^{1,4}$-variant described above (28). As $\Delta^{4,6}$-derivatives of ADDs such as 12-hydroxy-androsta-1,4,6-triene-3,17-dione (HATD, VIII) can be found in culture supernatants of strain Chol11 growing with cholate (10), side chain degradation seems to be the next phase (phase 2) of degradation. This is initiated by CoA-activation catalyzed by CoA-ligase SclA (29). In contrast to the model organisms using the $\Delta^{1,4}$-variant, *Sphingobium* sp. strain Chol11 growing with cholate produces no intermediates with a shortened side chain that can be found in culture supernatants (10). Interestingly, many genes for side-chain degradation known from other model organisms are missing in strain Chol11 (29, 30). The fact that several KshA homologs and many HIP degradation proteins are encoded in the genome of strain Chol11 suggests that phases 3 and 4, cleavage of the B-ring and further degradation of the *seco*-steroid, proceed similar to the 9,10-*seco*-pathway (29).

To further elucidate bile-salt degradation in strain Chol11 via the $\Delta^{4,6}$-pathway variant, differential proteome analyses of substrate-adapted cells were performed. For this, cholate (I in Fig 1) was used as a model substrate for the $\Delta^{4,6}$-variant. This was

126 compared to the 7-deoxy bile salt deoxycholate (XX in Fig 7), since 7-deoxy bile salts
127 cannot be degraded via $\Delta^{4,6}$-intermediates (27), and therefore require either completely
128 different pathways or variations of one common pathway. As a further reference
129 substrate, 12β-DHADD (IV in Fig 1) was used, because it does not possess a side chain
130 and therefore might reveal proteins that are specific for side-chain degradation.

Results and Discussion

Proteome analyses reveal three gene clusters encoding bile-salt degradation

To assess inducibility of cholate degradation in *Sphingobium* sp. strain Chol11, cholate- and glucose-grown cells were compared. Suspensions of cholate-adapted cells depleted cholate more quickly than those adapted to glucose (Fig S1A). When protein synthesis was inhibited by chloramphenicol, cholate was still completely degraded by cholate-adapted cells, but only a low percentage was depleted by glucose-adapted cells paralled by continuous production of HOCDA (VII in Fig 1) (Fig S1A+B). These findings indicate inducibility of cholate degradation in strain Chol11, with enzymes for A-ring oxidation and 7α-dehydroxylation constitutively produced but those for degradation of the side chain and the steroid nucleus requiring *de novo* synthesis in glucose-grown cells.

Thus, the proteomic profiles of cholate-, deoxycholate-, and 12βDHADD-grown cells were compared to that of glucose-grown cells using two-dimensional difference gel electrophoresis (2D DIGE), whole-cell shotgun proteomics, and analyses of the membrane protein enriched fractions (Tab S1). In total, 44.6% of the 3,550 predicted proteins of strain Chol11 were detected (Fig S2A). Proteins from the COG categories, “inorganic ion transport and metabolism” and “lipid transport and metabolism” showed the highest increase in abundance in cholate-grown cells compared to glucose-grown cells (Fig S2B).

The majority of proteins detected with significantly higher abundances in steroid- versus glucose-grown cells are encoded on chromosome 2 where most predicted steroid-degradation genes in strain Chol11 are located (30). 75 proteins had significantly higher abundances in cells grown on bile salts versus glucose, according to 2D-DIGE analysis.

Most of these proteins are encoded in three gene clusters (Fig 2), of which one (cluster 3) was previously predicted to encode steroid-degradation (29).

The finding, that the same set of proteins is produced in higher quantities during growth with both bile salts and most proteins were also produced for 12β-DHADD degradation indicates that degradation of these steroids generally involves the same proteins. This implies that both 7-hydroxy- and 7-deoxy-bile salts are degraded via the same pathway. Notably, however, a subset of proteins encoded in gene cluster 2 is differentially abundant, depending on the presence of a side chain. This implicates the proteins encoded in gene cluster 2 might be specific for side-chain degradation. Based on proteome and bioinformatic analyses, we compiled a model of bile-salt degradation in strain Chol11 (Fig 3).

Steroid transport

Several TonB-dependent receptor proteins (COG category “inorganic ion transport and metabolism”) were among the membrane proteins with the increases in abundance, including Nov2c232 (gene cluster 2), Nov2c378 (near gene cluster 3), and Nov2c659 (near gene cluster 4). TonB-dependent outer membrane transporter proteins (called TonB-dependent receptor protein) require the accessory proteins TonB, ExbB, and ExbD in the inner membrane to form a functional TonB-system (31). These accessory proteins were identified in the membrane protein-enriched fractions of cells grown on all substrates with similar Mascot scores, indicating constitutive formation (Nov1c1853-1856). Transporters for the uptake of bile salts are not known in Proteobacteria but there are indications that TonB-dependent receptors could be involved. First, they are generally known to participate in the import of complex growth substrates such as lignin

178 degradation compounds (32). Second, in *Novosphingobium tardaugens* NBRC16725, a
179 TonB-dependent receptor was upregulated in estradiol-grown cells, and a corresponding
180 deletion mutant showed reduced growth with estradiol (33).
181 The TonB-dependent receptor Nov2c232 (28 % identity to the TonB-dependent receptor
182 that was deleted in *N. tardaugens* NBRC16725) is encoded in the putative side-chain
183 degradation cluster in close vicinity to a gene encoding a transporter of the Major
184 Facilitator Superfamily (MFS) (Nov2c225), and both were more abundant in cholate- and
185 deoxycholate-grown cells (Fig 4, Table S1). These could be involved in transport of bile
186 salts as well as early degradation intermediates such as Δ^4-3-ketocholate (II in Fig 1) or
187 HOCDA (VII), that are found in strain Chol11 culture supernatants (10, 28). Such
188 transient extracellular accumulation of intermediates (10) is a common for bile-salt
189 degrading Proteobacteria and has been observed during bile-salt degradation in soil
190 (34). Thus, the above transporters could alternatively be involved in intermediate efflux.
191
192 **Phase 1: A-ring oxidation and B-ring dehydratation (blue section in Fig 3)**
193 **3α-Hydroxysteroid dehydrogenase.** The first step of bile-salt degradation is the
194 oxidation of 3-OH to a keto group by a 3α-hydroxy steroid dehydrogenase (3α-HSD) (27,
195 35). A putative 3α-HSD (Nov2c6) is encoded in gene cluster 1 (Figs 4,5). This protein
196 was present in lower abundances in glucose- versus steroid-grown cells (Fig 4).
197 Additional putative 3α-HSDs, Nov2c397 and Nov2c683, are encoded close to gene
198 clusters 3 and 4, respectively, and were not differentially expressed. This as well as
199 Nov2c6 being present in low abundances in glucose-adapted cells is consistent with
200 biotransformation experiments showing that oxidation of the 3-OH is constitutive in strain
201 Chol11.

5β-Δ^4-3-Ketosteroid dehydrogenase. The next step is the introduction of a double bond in the A-ring by 5β-Δ^4-3-ketosteroid dehydrogenase (5β-Δ^4-KSTD, named 5β-Δ^4-KSTD1) (28). 5β-Δ^4-KSTD1 is encoded in cluster 1 (Nov2c19) and is much more abundant in steroid- versus glucose-grown cells (>5.5-fold, Fig 4).

7α-Hydroxysteroid dehydratase. The key enzyme of the $\Delta^{4,6}$-pathway in strain Chol11, the 7α-hydroxysteroid dehydratase Hsh2 (Nov2c400), introduces a double bond in the B-ring by elimination of water (27). The *hsh2* gene is located in close proximity to cluster 3. Hsh2 was previously shown to be active in glucose-grown cells (28), in agreement with its similar abundance in all cells (Fig 4).

Δ^1-3-Ketosteroid dehydrogenase. In the $\Delta^{1,4}$-variant bile-salt degradation pathway, the next step is the introduction of a second double bond in the A-ring at C1 by a Δ^1-3-ketosteroid dehydrogenase (Δ^1-KSTD), which is a structural prerequisite for subsequent cleavage of the B-ring. The formation of $\Delta^{1,4,6}$-intermediates such as HATD (VIII in Fig 1) during growth with bile salts shows, that this reaction also occurs in strain Chol11 (10). Putative Δ^1-KSTDs (36, 37) are Nov2c82 (encoded between gene clusters 1 and 2, not detected in any condition), and Nov2c695 (encoded close to gene cluster 4, about 2- to 3-fold increased abundance in steroid-grown cells) (Fig 4). This expression pattern is in line with the biotransformation assays showing no Δ^1-KSTD activity in glucose-adapted cells. As no $\Delta^{1,4,6}$-intermediates with a side chain were reported for strain Chol11 during growth on bile salts (10, 27), this step might occur after side-chain degradation in strain Chol11.

Oxidation of the 12-OH-group. The previously observed formation of the 3,12-dioxo-chol-4,6-dienoate (DOCDA, XIII in Fig 3) (10, 27) suggests the involvement of a 12α-dehydrogenase in the degradation of 12-hydroxy bile acids in strain Chol11. A similar

226 reaction is catalyzed in *C. testosteroni* by SteA, which is active on 12α-hydroxy steroids
227 without a side chain (38). In *C. testosteroni*, the resulting 12-oxo-steroids are then
228 reduced to the corresponding 12β-steroids by SteB (Fig 5B) before cleavage of the A-
229 ring. The presence of a 12β-OH in the degradation intermediate HATD (VIII in Fig 1)
230 indicates that the reduction to a 12β-OH also takes place in strain Chol11. Homologs to
231 SteA and SteB are encoded in cluster 1 (Nov2c15 and Nov2c16, respectively) and were
232 detected with higher Mascot scores in steroid-grown cells (Fig 4).
233
234 **Phase 2: Side-chain degradation (green section in Fig 3)**
235 **CoA-activation of the side chain.** The steroid-C24-CoA-ligase SclA catalyzing the
236 initial step of side chain degradation in strain Chol11 was previously described (29) and
237 is encoded in the putative side-chain degradation gene cluster 2. SclA abundance,
238 relative to glucose-grown cells, was 3.3-fold higher in cholate- and deoxycholate-grown
239 cells, and 1.3-fold higher in 12β-DHADD-grown cells (Fig 4).
240 **Desaturation of the side chain.** Previous enzyme assays with cell free extract of strain
241 Chol11 indicated that a double bond is introduced into the CoA activated side chain (29).
242 This reaction is typically catalyzed by $\alpha_2\beta_2$-heterotetrameric acyl-CoA dehydrogenases
243 (ACADs) during the degradation of steroids (39). Two predicted ACAD proteins in gene
244 cluster 2 (Nov2c221 and Nov2c222) had 4- to 5-fold increased abundances in cholate-
245 and deoxycholate-grown cells, but not in 12β-DHADD-grown cells (Fig 4). This suggests
246 that they comprise an ACAD involved in side-chain degradation of bile salts, but the
247 location of any double bond formed by this ACAD is unknown.
248 **Further side-chain degradation.** The next step in bile-salt side-chain degradation is
249 typically the addition of water to the double bond (Fig 6B). The enzymes catalyzing this

hydration belong to the MaoC family of the thioesterase superfamily, containing a hotdog fold domain, or the Crotonase family (19, 40–42). These hydratases consist either of a single protein, such as the C_5 side-chain hydratases of *P. stutzeri* Chol1 and *M. tuberculosis* (19, 42), or of two subunits, such as the C_3 side-chain hydratase of *M. tuberculosis* (40). Two proteins from the thiolase superfamily with hotdog fold domains are encoded in cluster 2 (Nov2c219 and Nov2c220) adjacent to the ACAD genes. Nov2c219 was detected in only cholate- and deoxycholate-grown cells, and Nov2c220, only in cholate-grown cells (Fig 4). However, these proteins show only very low similarity (less than 20 % identity) to known side-chain hydratases, which suggests a different function for these proteins.

Homologs to other known side-chain degradation enzymes, such as thiolases and aldolases, are not encoded in cluster 2 (29). In *P. stutzeri* Chol1 and *R. jostii* RHA1, the first cycle of side-chain degradation leads to the release of acetyl-CoA and a shortened C_3-side chain. For the degradation of the C_3-side chain, a second cycle of aldolytic cleavage with similar steps including introduction of a double bond, addition of water and aldolytic cleavage is necessary. For this second cycle, both organisms encode a second set of enzymes (13, 19, 23).

Our proteome and bioinformatic analyses of strain Chol11 revealed no enzymes that could potentially be involved in side-chain cleavage or degradation of the C_3-side chain. This suggests that side-chain degradation in strain Chol11 is a mechanism other than aldolytic or thiolytic cleavage. A so-far unknown alternative mechanism might involve other proteins encoded in cluster 2, including putative hydroxysteroid dehydrogenases, amidases, and a Rieske monooxygenase (Fig. 6).

Phase 3: B-ring cleavage by the monooxygenase KshAB (yellow section in Fig 3)

The first step in the degradation of the steroid nucleus is the cleavage of the B-ring by the KshAB monooxygenase system (9, 43). Five homologs of the oxygenase component, KshA, are encoded on chromosome 2 (Nov2c66, Nov2c228, Nov2c407, Nov2c430, and Nov2c440 with 27 % - 32 % identity to KshA of *P. stutzeri* Chol1) (Fig 4). The numerous B-ring cleaving KshA homologs in the genome of strain Chol11 strongly suggest that steroid nucleus degradation starts with 9,10-*seco* cleavage, although the resulting 9,10-*seco* steroids have so far never been detected in cell-free supernatants of strain Chol11 cultures. Similar multiplicity of KshA homologs is also known from steroid degrading *Rhodococci* (44, 45). Nov2c228, Nov2c407, Nov2c430, and Nov2c440 had increased abundances in steroid-grown cells and therefore are candidates for this reaction. However, Nov2c228 has the lowest similarity to $KshA_{Chol1}$, its encoding gene is localized in the side-chain degradation gene cluster, and the abundance of Nov2c228 was only increased in bile-salt grown cells, but not in 12β-DHADD-grown cells. Thus, a different role for this enzyme appears feasible. In this context, the similarities of the KshA oxygenases and Neverland oxygenases, that are involved in the production of ecdysteroids in arthropods (46), could indicate a wider function for Rieske monooxygenases in steroid metabolism.

Interestingly, there are no distinct homologs of the reductase component KshB encoded in the genome of strain Chol11. This was also reported for *N. tardaugens* NBRC16725 (47). A flavodoxin reductase Novbp123 with 29 % identity to KshB of *P. stutzeri* Chol1 is encoded on plasmid pSb of strain Chol11 (30). This enzyme was detected in the membrane protein-enriched fractions of all tested cells with similar Mascot scores, indicating that its synthesis is not regulated in response to steroid degradation.

Novbp123 is encoded in a gene cluster together with a ferredoxin, a cytochrome c, and several exported and membrane proteins, suggesting that it is involved in membrane electron transport rather than bile-salt degradation specifically (Fig S3). This points at a different, KshB-independent electron shuttling mechanism for the KshA homologs in strain Chol11.

Phase 4: Complete degradation of the 9,10-*seco* intermediates (orange section in Fig. 3).

Most proteins required for the degradation of the 9,10-*seco* degradation intermediates, derived from both cholate and deoxycholate, and the respective HIP intermediates (26) are encoded in gene cluster 3 (Fig 7A). All of these proteins are at least 1.5-fold increased in abundance in steroid-grown cells (Fig. 4). This confirms that degradation of the steroid nucleus proceeds via the 9,10-*seco* pathway. Cluster 3 is very similar to the cluster encoding testosterone degradation in *N. tardaugens* NBRC16725 (47). In both organisms, genes encoding homologs for the reductase component TesA1 of the 9,10-*seco* steroid monooxygenase, TesA1A2, are missing, which could be a further hint at a different electron shuttling mechanism. However, in cluster 3 of strain Chol11, a flavin reductase (Nov2c347) is encoded near the gene for the oxygenase component, TesA2 (Nov2c349), indicating Nov2c347 could serve as a TesA1 substitute. Regarding HIP-degradation, a homolog for the gene encoding HIP-CoA-ligase ScdA (48) is missing in strain Chol11, but the CoA transferase, Nov2c359, could have this function.

Fate of the 12-OH. In *P. stutzeri* Chol1, the 12β-OH is removed during C- and D-ring degradation by the elimination of water catalyzed by Hsh1 and subsequent reduction of the resulting double bond by Sor1 (23) (Fig 5C). In strain Chol11, homologs to Hsh1 and

Sor1 are encoded in cluster 1 near *steA* and *steB* (*nov2c12* and *nov2c13*, respectively, Fig. 5A) and were found in increased abundances in steroid-grown cells (Fig 4). A gene cluster with the same order of genes for the 12-OH transforming enzymes SteA, SteB, Hsh1 and Sor1 is present in *N. tardaugens* NBRC16725 and a similar cluster is present in *C. testosteroni* CNB-2*, P. stutzeri* Chol1, and *Azoarcus* sp. strain Aa7 (Fig 5A), implicating a general role of these enzymes in bile-salt degradation.

Channeling of 7-hydroxy and 7-deoxy bile salts into C- and D- ring degradation. During degradation of HIP-intermediates (such as VI, XV, and XXII in Fig 7), the remainder of the B-ring is degraded via β-oxidation, which requires a hydroxy group at the former C7 (23) (Fig 7B). This hydroxy group is present in 7-hydroxy bile salts such as cholate, but during the degradation of 7-deoxy bile salts, such as deoxycholate, it has to be introduced into the propanoate side chain of the respective HIP-intermediates. This is initiated by the introduction of a double bond by the heteromeric ACAD ScdC1C2 followed by the addition of water by the hydratase ScdD (23, 49, 50).

In $\Delta^{4,6}$-intermediates, the hydroxy group at C7 is eliminated. Thus, bile salt degradation by the $\Delta^{4,6}$-pathway variant results in HIP-like intermediates with a double bond in the propanoate side chain attached to ring C (XV). This is the same intermediate as found during degradation of 7-deoxy bile salts after introduction of a double bond (23), and the needed hydroxy group could be added by the hydratase, Nov2c364, which was detected all steroid grown cells in similar abundance (Fig 4). Although the ACAD reaction catalyzed by ScdC1C2 is not needed for the degradation of cholate via the $\Delta^{4,6}$-variant, both subunits, Nov2c367 and Nov2c361, were found in all steroid-grown cells with similar Mascot scores (Fig 4). Thus, it is possible that the Δ^{6}-double bond had meanwhile been reduced, necessitating the ACAD.

As the hydroxy group eliminated by Hsh2 must again be added at the stage of HIP intermediates, the benefit of the elimination remains unclear. It might be related to the fact that many intermediates of bile-salt degradation are excreted in significant amounts during growth of bile-salt degrading bacteria not only in laboratory cultures but also in soil samples (10, 34). Other bile-salt degrading strains, such as *P. stutzeri* Chol1, that exclusively use the $\Delta^{1,4}$-degradation pathway are unable to utilize $\Delta^{4,6}$-compounds as growth substrates (10, 27). To this end, the dehydration might provide a way for strain Chol11 to secure the individual availability to these important carbon sources in natural habitats.

The $\Delta^{4,6}$-variant is widespread within the family *Sphingomonadaceae*

Prediction of the $\Delta^{4,6}$-variant pathway in steroid degrading bacteria. Apart from strain Chol11, steroid degradation has been reported in other sphingomonads such as *N. tardaugens* NBRC16725 (47, 51) and *Sphingomonas* sp. strain KC8 (52). In addition, *Sphingomonas* sp. strain Chol10 was also isolated as a HOCDA-degrading bacterium (10).

To investigate the prevalence of the $\Delta^{4,6}$-variant pathway within the family *Sphingomonadaceae*, we searched all 398 complete and draft genomes from the genera *Sphingobium, Novosphingobium,* and *Sphingomonas* available from the NCBI RefSeq database for the simultaneous presence of key steroid degradation proteins and Hsh2, the key protein of the $\Delta^{4,6}$-pathway. First, steroid degrading bacteria were predicted using 23 Hidden Markov Models (HMMs) representing ten key proteins of canonical steroid nucleus degradation (53). Based on the presence of seven out of ten key proteins including KshA and TesB, 53 genomes were predicted to encode steroid

degradation. Second, Hsh2 orthologs were determined in these genomes by BLASTp analyses using Hsh2 of *Sphingobium* sp. strain Chol11 (27) as query. Thirty-nine genomes containing both, the steroid-degradation genes and *hsh2,* were found. To further confirm the prediction of these proteins being involved in steroid degradation in these organisms, a reciprocal BLASTp analysis was performed using steroid degradation proteins from *P. stutzeri* Chol1 and $Hsh2_{Chol11}$ as queries (Fig 8). Thirty-eight genomes were confirmed to encode Hsh2 orthologs, and all of them contained orthologs for the majority of key proteins for steroid nucleus degradation. However, only a small subset of side-chain degradation proteins is encoded in most of the *Sphingobium* and *Novosphingobium* genomes. This was in contrast to the *Sphingomonas* genomes where orthologs of the genes encoding CoA-ligase, heteromeric ACAD, heteromeric hydratase, and aldolase for the degradation of the C_3-side chain from *P. stutzeri* Chol1 are present.

This suggests that bile-salt degradation via $\Delta^{4,6}$-intermediates is widely distributed among members of *Sphingomonas, Sphingobium,* and *Novosphingobium*, while the distinct side-chain degradation mechanism proposed for strain Chol11 may occur in most members of *Sphingobium* and *Novosphingobium*.

Bile-salt degradation in strains predicted to use the $\Delta^{4,6}$-variant. To investigate whether strains predicted to degrade bile salts via the $\Delta^{4,6}$-pathway variant, a selection of type strains with complete genome sequences was analysed for bile salt degradation. *Novosphingobium aromaticivorans* F199, *Sphingobium herbicidovorans* MH, and *N. tardaugens* NBRC16725 were tested for the degradation of the 7α-hydroxy bile salts, cholate and chenodeoxycholate, and the 7-deoxy bile salt, deoxycholate (structures in Fig 9). Accumulation of $\Delta^{4,6}$-intermediates was also monitored.

Strains MH, NBRC16725 and F199 grew on all tested bile salts (Fig 9B,D,F), and transiently accumulated the characteristic $\Delta^{4,6}$-intermediates HOCDA (VII in Fig 1) and DOCDA (XIII in Fig 3) (Fig 9C,E,G,I). In addition, during degradation of the 7α-hydroxy bile salts, strain F199 formed the intermediates XXIV and XXV that also have a $\Delta^{4,6}$-structure and lack a side chain according to their UV- and mass spectra, respectively (Fig 9I). This supports our prediction that the $\Delta^{4,6}$-variant is the predominant pathway for bile acid degradation in *Sphingobium,* and *Novosphingobium* strains.

Like strain Chol11, strains NBRC16725, MH, and F119 were able to fully metabolize side-chain bearing bile salts, despite encoding not all sidechain catabolic enzymes proteins known from other bile salt degrading bacteria. A BLASTp analysis revealed the presence of gene clusters with high similarity to gene cluster 2 of strain Chol11 in all three strains (Fig 6A); although the clusters from the *Novosphingobium* strains, NBRC16725 and F119, were notably missing several genes. In particular, homologs of the thiolase superfamily proteins Nov2c219 and Nov2c220, the only candidates for side chain hydratases in strain Chol11, were absent in the genome of NBRC16725. Homologs of SclA, the putative ACAD Nov2c221/Nov2c222, the Rieske monooxygenase Nov2c228, and the putative amidases Nov2c227 and Nov2c229 were found in the predicted side-chain degradation clusters of all three strains.

Degradation of conjugated bile salts

Strain Chol11 grew on the conjugated bile salts, taurocholate and glycocholate (X and XI, respectively, in Fig. 3, but not on the free amino acids glycine or taurine (Fig 10B,E). The inability to grow on these amino acids agrees with the similar biomass yields on taurocholate and glycocholate versus on unconjugated cholate (27). These results

418 suggests that glycine or taurine are removed by amidases prior to metabolism of the free
419 bile salts. Two candidate amidases are Nov2c227 and Nov2c229), encoded in the
420 predicted side-chain degradation gene cluster, were 2.3- to 8.9-fold more abundant in
421 cholate- and deoxycholate-grown cells.
422 To test their potential role in deconjugation of bile salts, Nov2c227 and Nov2c229, were
423 each heterologously expressed in *E. coli* MG1655. Cell suspensions and cell-free
424 extracts of *E. coli* expressing Nov2c227 transformed both taurocholate and glycocholate
425 to cholate (Fig 10). Cell suspensions and cell-free extracts of *E. coli* expressing the other
426 amidase candidate, Nov2c229, as well as the empty vector control did not substantially
427 catalyze this transformation. Additionally, compounds 2 Da lighter than the conjugated
428 bile salts and cholate were present in almost all assays. *E. coli* possesses a 7α-hydroxy
429 steroid dehydrogenase, which catalyzes the oxidation of the 7α-OH of conjugated and
430 free bile salts (54). Therefore, it is likely that the additional compounds are the 7-keto-
431 derivatives of the conjugated bile salts and cholate (XXVI, XXVII, and XXVIII in Fig 10).
432 Since Nov2c227 cleaves the conjugated bile salts, glycocholate and taurocholate, it was
433 named **B**ile **S**alt **A**midase (Bsa).
434 Interestingly, Bsa contains a TAT signal peptide as predicted by SignalP, indicating that
435 this protein is secreted, and deconjugation of conjugated bile salts takes place in the
436 periplasmatic or extracellular space prior to bile-salt uptake. In contrast to the well-
437 elucidated N-terminal nucleophile family bile-salt hydrolases from probiotic lactic acid
438 bacteria such as *Bifidobacterium longum* (55), Nov2c227 belongs to the large family of
439 amidases. This indicates a different evolutionary origin for bile-salt hydrolases in steroid
440 degradation as opposed to those in intestinal bacteria, despite their common function.
441 Homologs of Bsa are encoded in genomes of many strains of the family

Sphingomonadaceae, e.g. EGO055_026080 from *N. tardaugens* NBRC16725. Furthermore, *C. testosteroni* KF-1 was reported to degrade taurocholate (56) and its genome encodes two Bsa homologs (identity for both 44 %). The respective homologs, ORF25 and ORF26, from model organism *C. testosteroni* TA441 are encoded in its steroid degradation mega-cluster (14, 26). The function of Nov2c229 remains unknown so far.

Conclusion and general discussion

Differential proteome analyses of *Sphingobium* sp. strain Chol11 together with bioinformatic analyses revealed a comprehensive set of candidate proteins for the complete degradation of bile salts. From these data the complete degradation pathway for the steroid nucleus could be deduced, which is a mosaic of unknown reactions (especially side-chain degradation) and reactions known from other steroid-degrading organisms (Fig 3). Apparently, despite the variation in the first steps of 7-hydroxy bile salt degradation in strain Chol11, which leads to the introduction of a double bond in the B-ring by water elimination catalyzed by Hsh2, further degradation of the $\Delta^{4,6}$-intermediates is very similar to the 9,10-*seco* pathway known from other organisms and steroids. However, the $\Delta^{4,6}$-variation produces steroid intermediates that cannot be utilized by other organisms such as *P. stutzeri* Chol1 (10), suggesting the involvement of specialized enzymes in organisms that are able to utilize $\Delta^{4,6}$-intermediates. The prediction and identification of further organisms degrading bile salts via this $\Delta^{4,6}$-variant of the 9,10-*seco*-pathway demonstrates a wide distribution of this variant in the family *Sphingomonadaceae*.

In addition, the interpretation of the proteome analyses in conjunction with bioinformatic analyses resulted in the identification of a side-chain degradation gene cluster encoding key proteins that are presumably responsible for side chain removal by a yet unknown mechanism. The bile-salt hydrolase activity of Bsa as well as the CoA-ester formation by the CoA ligase SclA (29) encoded in this gene cluster further corroborates the involvement of this cluster in bile salt side-chain degradation. The subsequent introduction of a double bond into a yet unknown position is presumably catalyzed by the heteromeric ACAD Nov2c221/222. The absence of several other side-chain degradation genes from the genomes of strain Chol11, *N. tardaugens* NBRC16725, and *N. aromaticivorans* F199 suggests a mechanism different from thiolytic and aldolytic cleavage. Interestingly, homologs of SclA, the putative ACAD Nov2c221/Nov2c222, the Rieske monooxygenase Nov2c228, and the putative amidase Nov2c229 were found in all Sphingomonads confirmed to use the $\Delta^{4,6}$-variant for bile salt degradation. However, the functions of this Rieske monooxygenase and the putative amidase during bile salt side-chain degradation remain unclear. In addition to being highly conserved and encoded near to confirmed side-chain degradation genes in all four tested strains, they seem to be formed exclusively in response to side-chain containing bile salts in strain Chol11. Further investigations regarding this gene cluster using molecular methods are under way. Unfortunately, these analyses are impaired by the complicated genetic modification of strain Chol11.

Interestingly, several members of the family *Sphingomonadacea* are adapted to growth with steroids and bile salts. Moreover, both *Sphingobium* sp. strain Chol11 and *N. tardaugens* NBRC16725 grow only slowly with non-steroidal substrates (51) and some genes for early steps of bile-salt degradation are apparently constitutively induced.

Together with the prevalence of bile-salt degradation in strains that had originally been isolated with xenobiotic compounds (57, 58) indicates that bile-salt degradation may be a conserved property of these organisms and calls attention to its evolutionary origin.

Experimental procedures

Cultivation of bacteria

If not indicated otherwise, *Sphingobium* sp. strain Chol11 (DSM 110934) (10), *S. herbicidovorans* MH (DSM 11019) (58), *N. aromaticivorans* F199 (DSM12444) (59), *N. tardaugens* NBRC16725 (DSM 16702) (51), and *E. coli* MG1655 (DSM 18039) (60) were cultivated in HEPES-buffered medium B (MB) (61). If not indicated otherwise, wildtype strains other than *E. coli* were grown with 1 mM cholate as the sole carbon source, whereas *E. coli* MG1655 was grown with 15 mM glucose. For maintenance, strain Chol11 and *N. tardaugens* NBRC16725 were grown on MB agar with 1 mM cholate, *S. herbicidovorans* MH was grown on CASO agar (Merck Millipore, Burlington, MA, USA), *N. aromaticivorans* F199 was grown on MB agar with 15 mM glucose, and *E. coli* MG1655 was grown on LB agar (62). For cultivation of strains containing pBBR1MCS-5 (63), 20 µg ml^{-1} gentamicin was added. When bile salts were added, gentamicin was omitted. Liquid cultures with volumes up to 5 ml were cultivated in 10 ml test tubes at 200 rpm, larger cultures were cultivated in 500 ml Erlenmeyer flasks without baffles at 120 rpm. All strains were incubated at 30 °C, except for strain maintenance of *E. coli* strains (37 °C). For agar plates, 1.5 % (w/v) Bacto agar (BD, Sparks, USA) was added.

Cholate (≥99 %), deoxycholate (≥97 %), and glycocholate (≥97 %) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Chenodeoxycholate (≥98 %) was purchased from Carl Roth (Karlsruhe, Germany). Taurocholate (>98 %) was purchased from Fluka (Buchs, Switzerland).

Growth experiments

For growth experiments, 3 – 5 ml main cultures were inoculated to a predefined OD_{600} (about 0.02) directly from liquid starter cultures and growth was determined by measurement of OD_{600} (Camspec M107, Spectronic Camspec, Leeds, UK). Growth with bile salts was tested using 1 mM cholate, 1 mM chenodeoxycholate or 1 mM deoxycholate. Growth with conjugated bile salts was tested using 1 mM taurocholate, 1 mM glycocholate, 15 mM glycine or 15 mM taurine. At suitable time points, samples for HPLC-MS measurements were withdrawn.

Starter cultures were grown with 1 mM cholate for strain Chol11, *N. aromaticivorans* F199, and *N. tardaugens* NBRC16725 or 12 mM succinate for *S. herbicidovorans* MH. Starter cultures were incubated over-night for about 15 h.

Biotransformation experiments

Induction of cholate degradation in strain Chol11 was tested using suspensions of cholate- and glucose-grown cells with 1 mM cholate and 10 µg ml^{-1} chloramphenicol to inhibit *de novo* protein synthesis as described (28).

For determining whole-cell biotransformation of various steroid compounds by *E. coli* MG1655 expressing amidase genes, 5 ml starter cultures of *E. coli* MG1655 pBBR1MCS-5 as empty vector control, *E. coli* MG1655 pBBR1MCS-5::*nov2c227* or *E. coli* MG1655 pBBR1MCS-5::*nov2c229* in LB with 20 µg ml^{-1} gentamicin were incubated for about 15 h. Cells were harvested by centrifugation in 2-ml reaction tubes (8,000 x g, ambient temperature, 3 min), washed with MB and resuspended in MB to an OD_{600} of about 1. Cell suspensions were incubated for several days at 30 °C at 200 rpm after addition of either 1 mM taurocholate or 1 mM glycocholate. 30 mM glucose were added to all preparations as carbon source.

For monitoring biotransformations of conjugated bile salts by cell free extracts of *E. coli* MG1655 expressing amidase genes, 50 ml LB with 20 µg ml^{-1} gentamicin were inoculated to an initial OD_{600} of 0.015 with the aforementioned strains of *E. coli* MG1655 and incubated for about 18 h with addition of 0.2 mM isopropyl-β-D-thiogalactopyranosid after about 3 h. Cells were harvested (8,000 x g, 4 °C, 8 min), washed with 10 mM MOPS buffer (pH 7.8) and resuspended in about 2 ml 50 mM MOPS buffer (pH 7.8). Cells were disrupted in 15-ml conical centrifugation tubes by ultrasonication on ice (amplitude 60 %, cycle 0.5, UP200S, Hielscher Ultrasonics, Teltow, Germany) for 8 min with a 1 min break after 4 min. Cell debris was removed by centrifugation (25,000 x g, 4 °C, 30 min) in 2-ml reaction tubes. Cell extracts were used immediately in enzyme assays or stored at -20 °C for later use. Enzyme assays (1 ml) contained 50 mM MOPS (pH 7.8), 1 mM substrate and 100 µl cell extract. Assays were incubated for 30 min – 15 h at 30 °C. Samples of all enzyme assays were subjected to HPLC-MS measurements.

Proteome analyses

Generation of substrate adapted cells. Strain Chol11 was freshly thawed from a cryopreservation culture for each cultivation on MB agar and streaked twice onto plates with 1 mM cholate. After 1 – 4 days, the strain was transferred to new agar plates containing either 1 mM cholate, 1 mM deoxycholate, 2 mM 12β-DHADD or 15 mM glucose. Cells were further transferred twice to the same medium after incubation for 2 days with steroidal substrates or 3 – 4 days with glucose. From these plates, twelve 5 ml MB starter cultures containing the same carbon source as the respective solid media were inoculated and incubated for about 15 h for steroidal compounds or 30 h for glucose. Subsequently, 1 mM bile salt, 2 mM 12β-DHADD or 15 mM glucose were added and the starter cultures were incubated for further 1 – 1.5 h. Starter cultures with

the same substrate were pooled and used for inoculation of twelve 100 ml main cultures in 500-ml Erlenmeyer flasks without baffles containing the same carbon sources at an OD_{600} of 0.015. Main cultures were incubated at 30 °C and 200 rpm and growth was monitored until the half maximal OD_{600} was reached. Cultures were harvested in two 50-ml conical centrifugation tubes by centrifugation (5,525 x g, 4 °C, 30 min) and kept on ice. Cells were washed with 25 ml 100 mM Tris buffer (pH 7.5 with HCl) containing 5 mM $MgCl_2$ and cells of the same culture were pooled. After centrifugation, cells were resuspended in 625 µl of the same buffer, transferred to 2 ml reaction tubes and harvested by centrifugation (14,000 x g, 4 °C, 15 min). After weighing, pellets were snap-frozen using liquid nitrogen and stored at -70 °C.

Profiling of soluble proteins by 2D DIGE and protein identification by MALDI-TOF-MS/MS. Soluble proteins were extracted from cells of strain Chol11 and 2D DIGE performed essentially as described previously (64). Per growth condition, four biological replicate samples were prepared and 50 µg total protein used for minimal labelling with 200 pmol of Lightning SciDye DIGE fluors (SERVA Electrophoresis GmbH, Heidelberg, Germany). Glucose-adapted cells served as reference state and were labelled with Sci5. Protein extracts of the other three (test) states were each labelled with Sci3. The internal standard contained equal amounts of all test and the reference state(s) and was labelled with Sci2. First dimension separation by isoelectric focusing (IEF) was conducted with 24 cm-long IPG strips (pH 3–11 NL; GE Healthcare) run in a Protean i12 system (Bio-Rad, Munich, Germany). The IEF program used was as follows: 50 V for 13 h, 200 V for 1 h, 1,000 V for 1 h, gradual gradient to 10,000 V within 2 h and 10,000 V until 80,000 Vhs were reached. Second dimension separation according to molecular size was done by SDS-PAGE (12.5% gels, v/v) using an EttanDalt*twelve* system (GE Healthcare).

586 Directly after electrophoresis, 2D DIGE gels were digitalized using a CCD camera
587 system (Intas Advanced 2D Imager; Intas Science Imaging Instruments GmbH,
588 Göttingen, Germany) (65). Cropped gel images were analyzed with the SameSpots™
589 software (version 5.0.5.0, TotalLab, Newcastle upon Tyne, UK) and spots with changes
590 in abundance of ≥ |1.5|-fold and an ANOVA p-value of ≤ 1×10^{-4} were accepted as
591 significant (66). Spots of interest were excised from at least two separate, preparative
592 colloidal Coomassie Brilliant Blue stained gels (300 µg protein load) using the EXQuest
593 spot cutter (Bio-Rad), and subsequently washed and tryptically digested as described
594 recently (67).
595 Sample digests were spotted onto Anchorchip steel targets (Bruker Daltonik GmbH,
596 Bremen, Germany) and analyzed with an UltrafleXtreme MALDI-TOF/TOF mass
597 spectrometer (Bruker Daltonik GmbH) as recently described (67). Peptide mass
598 fingerprint (PMF) searches were performed with a Mascot server (version 2.3; Matrix
599 Science, London, UK) against the translated genome of strain Chol11 with a mass
600 tolerance of 25 ppm. Five lift spectra were collected to confirm PMF identification and
601 three additional spectra were acquired of unassigned peaks applying feedback by the
602 ProteinScape platform (version 3.1; Bruker Daltonik GmbH). In case of failed PMF
603 identification, eight lift spectra of suitable precursors were acquired. MS/MS searches
604 were performed with a mass tolerance of 100 ppm. For both, MS and MS/MS searches,
605 Mascot scores not meeting the 95% certainty criterion were not considered significant. A
606 single miscleavage was allowed (enzyme trypsin) and carbamidomethyl (C) and
607 oxidation (M) were set as fixed and variable modifications, respectively.
608 **Shotgun proteomic analysis.** For shotgun analysis, cell pellets of three biological
609 replicate samples per growth condition were suspended in lysis buffer, cells disrupted

and the debris-free fraction reduced, alkylated and subjected to tryptic digest as previously described (67). Obtained peptides were separated by nanoLC (UltiMate 3000; ThermoFisher Scientific, Germering, Germany) using a trap-column (C18, 5 µm bead size, 2 cm length, 75 µm inner diameter; ThermoFisher Scientific) and a 25 cm analytical column (C18, 2 µm bead size, 75 µm inner diameter; ThermoFisher Scientific) applying a 360 min linear gradient (68). The nanoLC eluent was continuously analyzed by an online-coupled ion-trap mass spectrometer (amaZon speed ETD; Bruker Daltonik GmbH) using the captive spray electrospray ion source (Bruker Daltonik GmbH). The instrument was operated in positive mode with a capillary current of 1.3 kV and drygas flow of 3 l min^{-1} at 150°C. Active precursor exclusion was set for 0.2 min. Per full scan MS, 20 MS/MS spectra of the most intense masses were acquired. Protein identification was performed with ProteinScape as described above, including a mass tolerance of 0.3 Da for MS and 0.4 Da for MS/MS searches and applying a target decoy strategy (false discovery rate < 1%).

Analysis of the membrane protein-enriched fraction. Total membrane fractions were prepared from two biological replicates per substrate condition as described (67). The obtained protein content was determined with the RC-DC assay (Bio-Rad) and 8 µg total protein separated by SDS-PAGE gels (~7 cm separation gel). Following staining with Coomassie Brilliant Blue (69), each sample lane was cut into 8 slices and each slice into small pieces of ~1-2 mm^3 prior to washing, reduction, alkylation and tryptic digest (67). Separation and mass determination was performed as described above, using a 100 min linear gradient. Identified proteins (performed as described above) of each slice per sample were compiled using the protein extractor of the ProteinScape platform.

Cloning techniques

Cloning was performed according to standard procedures and as described elsewhere (27).

For expression of amidase genes in *E. coli* MG1655, genes were amplified using the respective primer pairs expfor/exprev (Tab 1) and ligated into vector pBBR1MCS-5. The respective ligation products were transferred to *E. coli* MG1655 by heat shock transformation. Presence and correct ligation of plasmids was confirmed by colony PCR using M13 primers.

Analytical methods

Steroid compounds were analyzed by HPLC-MS. For this, samples were centrifuged (>16,000 x g, ambient temperature, 5 min) to remove cells and particles. Supernatants were stored at -20 °C and centrifuged again prior to measurement. Samples from cell suspension experiments with *E. coli* MG1655 were directly frozen at -20 °C and only centrifuged after thawing to break the cells. HPLC-MS measurements were performed with a Dionex Ultimate 3000 HPLC (ThermoFisher Scientific, Waltham, MA, USA) with an UV/visible light diode array detector and coupled to an ion trap amaZon speed mass spectrometer (Bruker Daltonik, Bremen, Germany) with an electrospray ion source. Compounds were separated over a reversed phase C_{18} column (150 x 3 mm, Eurosphere II, 100-5 C_{18}; Knauer Wissenschaftliche Geräte, Berlin, Germany) at 25 °C. Samples of cell suspensions for testing the induction of steroid degradation in *Sphingobium* sp. strain Chol11 were measured as described by (27), whereas all other samples were measured as described by (28).

Bile salt concentrations were determined as peak area from base peak chromatograms measured in negative ion mode. Intermediates were identified according to retention time, UV- and MS-spectra, and comparison with known compounds. Structures of

658 unknown metabolites were proposed on the basis of retention time as well as UV- and
659 MS-spectra.

660 **Bioinformatic methods**

661 Searches for homologous proteins and determinations of protein similarities were
662 performed using the BLASTp algorithm (70, 71). Protein similarities were calculated from
663 global alignments in the BLAST suite using the Needleman-Wunsch algorithm (70, 72)
664 Protein domains and families were predicted using Interpro (73). For functional
665 annotation of strain Chol11, the eggNOG database (74) was used.
666 For the bioinformatic identification of other sphingomonads using the $\Delta^{4,6}$-variant of the
667 9,10-*seco*-pathway, first a database of putative steroid degraders was set up similar to
668 (75). On 18 October 2018, all complete and draft genomes available for the genera
669 *Sphingobium, Novosphingobium,* and *Sphingomonas* were downloaded from the
670 RefSeq database (version 89). Using 23 Hidden Markov Models (HMMs) (53), these
671 genomes were searched for ten homologs of steroid degradation proteins using HMMER
672 v3 (76) using a maximum *E*-value of 10^{-25}. HMMs for Δ^{1}-KSTD (KstD), KshA, TesA2
673 (=HsaA in *R. jostii* RHA1), TesB (=HsaC), TesE (=HsaE), TesF (=HsaG), TesG (=HsaF),
674 ScdK (=IpdC), ScdL1 (=IpdA), and ScdL2 (=IdpB) were used. Bacteria were predicted
675 to be able to degrade steroids when their genomes encoded homologs of seven out of
676 the eleven steroid degradation key enzymes including KshA and TesA2. With these
677 genomes, a reciprocal BLASTp search (75) was conducted using the key enzyme of the
678 $\Delta^{4,6}$-pathway variant Hsh2 from strain Chol11 (27) as query using *E*-value and identity
679 cutoffs of 10^{-25} and 35 %, respectively. These values were optimized empirically
680 comparing analyses using $Hsh2_{Chol11}$ as well as BaiE from *C. scindens* (UniProt-ID
681 P19412) which has a similar function in a different pathway (77). The results of both

682 analyses were compared and *E*-value and identity cutoffs were chosen to ensure, that
683 proteins were only identified as homologs of one of these dehydratases. All genomes
684 from putative steroid degraders containing Hsh2 homologs were subjected to a
685 reciprocal BLASTp analysis using known steroid degradation proteins from *P. stutzeri*
686 Chol1 as queries. For data analysis and preparation of figures, R (v3.5.1) was used
687 together with the packages circlize (v0.4.8), genoPlotR (v0.8.9), ggplot2 (v3.2.1),
688 ComplexHeatmap (v1.18.1), gplots (v3.0.3), RColorBrewer (v1.1-2), VennDiagram
689 (v1.6.20), ape (v5.3), reshape2 (v1.4.3), tidyverse (v1.3.0), and readxl (v1.3.1).

690 **Acknowledgements**

691 The authors thank Karin Niermann and Kirsten Heuer (both Münster) as well as
692 Christina Hinrichs (Oldenburg) for excellent experimental support and Florentin Schmidt
693 for help with planning of cloning.

694 This work was funded by two grants of the Deutsche Forschungsgemeinschaft (DFG
695 projects PH71/3-2 and INST 211/646-1 FUGG) to BP and a scholarship of the DAAD
696 Stiftung in cooperation with the Prof. Dr. Bingel-Stiftung to FF.

References

1. Maldonado-Valderrama J, Wilde P, MacIerzanka A, MacKie A. 2011. The role of bile salts in digestion. Adv Colloid Interface Sci 165:36–46.
2. Hylemon PB, Zhou H, Pandak WM, Ren S, Gil G, Dent P. 2009. Bile acids as regulatory molecules. J Lipid Res 50:1509–1520.
3. Jones B V., Begley M, Hill C, Gahan CGM, Marchesi JR. 2008. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. Proc Natl Acad Sci USA 105:13580–13585.
4. Hofmann AF, Hagey LR, Krasowski MD. 2010. Bile salts of vertebrates: structural variation and possible evolutionary significance. J Lipid Res 51:226–46.
5. Russell DW. 2003. The enzymes, regulation, and genetics of bile acid synthesis. Annu Rev Biochem 72:137–174.
6. Ridlon JM, Kang D-JJ, Hylemon PB. 2006. Bile salt biotransformations by human intestinal bacteria. J Lipid Res 47:241–259.
7. Philipp B. 2011. Bacterial degradation of bile salts. Appl Microbiol Biotechnol 89:903–915.
8. Olivera ER, Luengo JM. 2019. Steroids as environmental compounds recalcitrant to degradation: genetic mechanisms of bacterial biodegradation pathways. Genes (Basel) 10:512.
9. Philipp B, Erdbrink H, Suter MJF, Schink B. 2006. Degradation of and sensitivity to cholate in *Pseudomonas* sp. strain Chol1. Arch Microbiol 185:192–201.
10. Holert J, Yücel O, Suvekbala V, Kulić Ž, Möller H, Philipp B. 2014. Evidence of distinct pathways for bacterial degradation of the steroid compound cholate suggests the potential for metabolic interactions by interspecies cross-feeding.

Environ Microbiol 16:1424–1440.

11. Merino E, Barrientos A, Rodríguez J, Naharro G, Luengo JM, Olivera ER. 2013. Isolation of cholesterol- and deoxycholate-degrading bacteria from soil samples: evidence of a common pathway. Appl Microbiol Biotechnol 97:891–904.

12. Yücel O, Borgert SR, Poehlein A, Niermann K, Philipp B. 2018. The 7α-hydroxysteroid dehydratase Hsh2 is essential for anaerobic degradation of the steroid skeleton of 7α-hydroxyl bile salts in the novel denitrifying bacterium *Azoarcus* sp. strain Aa7. Environ Microbiol 21:800–813.

13. Mohn WW, Wilbrink MH, Casabon I, Stewart GR, Liu J, van der Geize R, Eltis LD. 2012. Gene cluster encoding cholate catabolism in *Rhodococcus* spp. J Bacteriol 194:6712–6719.

14. Horinouchi M, Hayashi T, Kudo T. 2012. Steroid degradation in *Comamonas testosteroni*. J Steroid Biochem Mol Biol 129:4–14.

15. Wipperman MF, Sampson NS, Thomas ST, T. Thomas S. 2014. Pathogen roid rage: cholesterol utilization by *Mycobacterium tuberculosis*. Crit Rev Biochem Mol Biol 49:269–93.

16. Birkenmaier A, Holert J, Erdbrink H, Möller HM, Friemel A, Schönenberger R, Suter MJF, Klebensberger J, Philipp B. 2007. Biochemical and genetic investigation of initial reactions in aerobic degradation of the bile acid cholate in *Pseudomonas* sp. strain Chol1. J Bacteriol 189:7165–7173.

17. Swain K, Casabon I, Eltis LD, Mohn WW. 2012. Two transporters essential for reassimilation of novel cholate metabolites by *Rhodococcus jostii* RHA1. J Bacteriol 194:6720–6727.

18. Capyk JK, Casabon I, Gruninger R, Strynadka NC, Eltis LD. 2011. Activity of 3-

745 ketosteroid 9α-hydroxylase (KshAB) indicates cholesterol side chain and ring
746 degradation occur simultaneously in *Mycobacterium tuberculosis*. J Biol Chem
747 286:40717–40724.
748 19. Holert J, Jagmann N, Philipp B. 2013. The essential function of genes for a
749 hydratase and an aldehyde dehydrogenase for growth of *Pseudomonas* sp. strain
750 Chol1 with the steroid compound cholate indicates an aldolytic reaction step for
751 deacetylation of the side chain. J Bacteriol 195:3371–3380.
752 20. Szentirmai A. 1990. Microbial physiology of sidechain degradation of sterols. J Ind
753 Microbiol 6:101–115.
754 21. Holert J, Kulić Ž, Yücel O, Suvekbala V, Suter MJF, Möller HM, Philipp B. 2013.
755 Degradation of the acyl side chain of the steroid compound cholate in
756 *Pseudomonas* sp. strain Chol1 proceeds via an aldehyde intermediate. J Bacteriol
757 195:585–595.
758 22. Birkenmaier A, Möller HM, Philipp B. 2011. Identification of a thiolase gene
759 essential for β-oxidation of the acyl side chain of the steroid compound cholate in
760 *Pseudomonas* sp. strain Chol1. FEMS Microbiol Lett 318:123–130.
761 23. Holert J, Yücel O, Jagmann N, Prestel A, Möller HM, Philipp B. 2016. Identification
762 of bypass reactions leading to the formation of one central steroid degradation
763 intermediate in metabolism of different bile salts in *Pseudomonas* sp. strain Chol1.
764 Environ Microbiol 18:3373–3389.
765 24. Barrientos Á, Merino E, Casabon I, Rodríguez J, Crowe AM, Holert J, Philipp B,
766 Eltis LD, Olivera ER, Luengo JM. 2015. Functional analyses of three acyl-CoA
767 synthetases involved in bile acid degradation in *Pseudomonas putida* DOC21.
768 Environ Microbiol 17:47–63.

769 25. Crowe AM, Casabon II, Brown KL, Liu J, Lian J, Rogalski JC, Hurst TE, Snieckus
770 V, Foster LJ, Eltis LD. 2017. Catabolism of the last two steroid rings in
771 *Mycobacterium tuberculosis* and other bacteria. MBio 8:1–16.
772 26. Horinouchi M, Koshino H, Malon M, Hirota H, Hayashi T. 2019. Steroid
773 degradation in *Comamonas testosteroni* TA441: identification of the entire β-
774 oxidation cycle of the cleaved B ring. Appl Environ Microbiol 85:1–17.
775 27. Yücel O, Drees S, Jagmann N, Patschkowski T, Philipp B. 2016. An unexplored
776 pathway for degradation of cholate requires a 7α-hydroxysteroid dehydratase and
777 contributes to a broad metabolic repertoire for the utilization of bile salts in
778 *Novosphingobium* sp. strain Chol11. Environ Microbiol 18:5187–5203.
779 28. Feller FM, Marke G, Drees SL, Wöhlbrand L, Rabus R, Philipp B. 2021. Substrate
780 inhibition of 5β-Δ^4-3-ketosteroid dehydrogenase in *Sphingobium* sp. strain Chol11
781 acts as circuit breaker during growth with toxic bile salts. Front Microbiol
782 12:655312.
783 29. Yücel O, Holert J, Ludwig KC, Thierbach S, Philipp B. 2018. A novel
784 steroidcoenzyme A ligase from *Novosphingobium* sp. strain Chol11 is essential for
785 an alternative degradation pathway for bile salts. Appl Environ Microbiol 84:1–16.
786 30. Yücel O, Wibberg D, Philipp B, Kalinowski J. 2018. Genome sequence of the bile
787 salt-degrading bacterium *Novosphingobium* sp. strain Chol11, a model organism
788 for bacterial steroid catabolism. Genome Announc 6:e01372–17.
789 31. Noinaj N, Guillier M, Barnard TJ, Buchanan SK. 2010. TonB-dependent
790 transporters: regulation, structure, and function. Annu Rev Microbiol 64:43–60.
791 32. Fujita M, Mori K, Hara H, Hishiyama S, Kamimura N, Masai E. 2019. A TonB-
792 dependent receptor constitutes the outer membrane transport system for a lignin-

793 derived aromatic compound. Commun Biol 2:1–10.
794 33. Ibero J, Galán B, Rivero-Buceta V, García JL. 2020. Unraveling the 17β-estradiol
795 degradation pathway in *Novosphingobium tardaugens* NBRC 16725. Front
796 Microbiol 11:588300.
797 34. Mendelski MN, Dölling R, Feller FM, Hoffmann D, Ramos Fangmeier L, Ludwig
798 KC, Yücel O, Mährlein A, Paul RJ, Philipp B. 2019. Steroids originating from
799 bacterial bile acid degradation affect *Caenorhabditis elegans* and indicate potential
800 risks for the fauna of manured soils. Sci Rep 9:11120.
801 35. Oppermann UCT, Maser E. 1996. Characterization of a 3α-hydroxysteroid
802 dehydrogenase/carbonyl reductase from the Gram-negative bacterium
803 *Comamonas testosteroni*. Eur J Biochem 241:744–749.
804 36. Horinouchi M, Kurita T, Yamamoto T, Hatori E, Hayashi T, Kudo T, Hirosawa W-S.
805 2004. Steroid degradation gene cluster of *Comamonas testosteroni* consisting of
806 18 putative genes from *meta*-cleavage enzyme gene *tesB* to regulator gene *tesR*.
807 Biochem Biophys Res Commun 324:597–604.
808 37. Knol J, Bodewits K, Hessels GI, Dijkhuizen L, van der Geize R. 2008. 3-Keto-5α-
809 steroid Δ^1-dehydrogenase from *Rhodococcus erythropolis* SQ1 and its orthologue
810 in *Mycobacterium tuberculosis* H37Rv are highly specific enzymes that function in
811 cholesterol catabolism. Biochem J 410:339–346.
812 38. Horinouchi M, Hayashi T, Koshino H, Malon M, Yamamoto T, Kudo T. 2008.
813 Identification of genes involved in inversion of stereochemistry of a C-12 hydroxyl
814 group in the catabolism of cholic acid by *Comamonas testosteroni* TA441. J
815 Bacteriol 190:5545–5554.
816 39. Thomas ST, Sampson NS. 2013. *Mycobacterium tuberculosis* utilizes a unique

817 heterotetrameric structure for dehydrogenation of the cholesterol side chain.
818 Biochemistry 52:2895–2904.
819 40. Yang M, Guja KE, Thomas ST, Garcia-Diaz M, Sampson NS. 2014. A distinct
820 MaoC-like enoyl-CoA hydratase architecture mediates cholesterol catabolism in
821 *Mycobacterium tuberculosis*. ACS Chem Biol 9:2632–2645.
822 41. Yuan T, Werman JM, Yin X, Yang M, Garcia-Diaz M, Sampson NS. 2021.
823 Enzymatic β-oxidation of the cholesterol side chain in *Mycobacterium tuberculosis*
824 bifurcates stereospecifically at hydration of 3-oxo-cholest-4,22-dien-24-oyl-CoA.
825 ACS Infect Dis acsinfecdis.1c00069.
826 42. Bonds AC, Yuan T, Werman JM, Jang J, Lu R, Nesbitt NM, Garcia-Diaz M,
827 Sampson NS. 2020. Post-translational succinylation of *Mycobacterium*
828 *tuberculosis* enoyl-CoA hydratase EchA19 slows catalytic hydration of cholesterol
829 catabolite 3-oxo-chol-4,22-diene-24-oyl-CoA. ACS Infect Dis 6:2214–2224.
830 43. Petrusma M, Van Der Geize R, Dijkhuizen L. 2014. 3-Ketosteroid 9α-hydroxylase
831 enzymes: Rieske non-heme monooxygenases essential for bacterial steroid
832 degradation. Antonie Van Leeuwenhoek 106:157–172.
833 44. Petrusma M, Hessels G, Dijkhuizen L, van der Geize R. 2011. Multiplicity of 3-
834 ketosteroid-9α-hydroxylase enzymes in *Rhodococcus rhodochrous* DSM43269 for
835 specific degradation of different classes of steroids. J Bacteriol 193:3931–3940.
836 45. Penfield JS, Worrall LJ, Strynadka NC, Eltis LD. 2014. Substrate specificities and
837 conformational flexibility of 3-ketosteroid 9α-hydroxylases. J Biol Chem
838 289:25523–25536.
839 46. Yoshiyama-Yanagawa T, Enya S, Shimada-Niwa Y, Yaguchi S, Haramoto Y,
840 Matsuya T, Shiomi K, Sasakura Y, Takahashi S, Asashima M, Kataoka H, Niwa R.

2011. The conserved Rieske oxygenase DAF-36/Neverland is a novel cholesterol-metabolizing enzyme. J Biol Chem 286:25756–25762.

47. Ibero J, Galán B, Díaz E, García JL. 2019. Testosterone degradative pathway of *Novosphingobium tardaugens*. Genes (Basel) 10:871.

48. Horinouchi M, Hayashi T, Koshino H, Kudo T. 2006. ORF18-disrupted mutant of *Comamonas testosteroni* TA441 accumulates significant amounts of 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid and its derivatives after incubation with steroids. J Steroid Biochem Mol Biol 101:78–84.

49. Horinouchi M, Hayashi T, Koshino H, Malon M, Hirota H, Kudo T. 2014. Identification of 9α-hydroxy-17-oxo-1,2,3,4,10,19-hexanorandrostan- 5-oic acid in steroid degradation by *Comamonas testosteroni* TA441 and its conversion to the corresponding 6-en-5-oyl coenzyme A (CoA) involving open reading frame 28 (ORF28)- and ORF30-encoded acyl-CoA dehydrogenases. J Bacteriol 196:3598–3608.

50. Horinouchi M, Hayashi T, Koshino H, Malon M, Hirota H, Kudo T. 2014. Identification of 9α-hydroxy-17-oxo-1,2,3,4,10,19-hexanorandrost-6- en-5-oic acid and β-oxidation products of the C-17 side chain in cholic acid degradation by *Comamonas testosteroni* TA441. J Steroid Biochem Mol Biol 143:306–322.

51. Fujii K, Satomi M, Morita N, Motomura T, Tanaka T, Kikuchi S. 2003. *Novosphingobium tardaugens* sp. nov., an oestradiol-degrading bacterium isolated from activated sludge of a sewage treatment plant in Tokyo. Int J Syst Evol Microbiol 53:47–52.

52. Roh H, Chu KH. 2010. A 17β-estradiol-utilizing bacterium, *Sphingomonas* strain KC8: part i - characterization and abundance in wastewater treatment plants.

865 Environ Sci Technol 44:4943–4950.

866 53. Holert J, Cardenas E, Bergstrand LH, Zaikova E, Hahn AS, Hallam SJ, Mohn WW.
867 2018. Metagenomes reveal global distribution of bacterial steroid catabolism in
868 natural, engineered, and host environments. MBio 9:e02345–17.

869 54. Yoshimoto T, Higashi H, Kanatani A, Sheng Lin XU, Nagai H, Oyama H, Kurazono
870 K, Tsuru D. 1991. Cloning and sequencing of the 7α-hydroxysteroid
871 dehydrogenase gene from *Escherichia coli* HB101 and characterization of the
872 expressed enzyme. J Bacteriol 173:2173–2179.

873 55. Foley MH, O'Flaherty S, Barrangou R, Theriot CM. 2019. Bile salt hydrolases:
874 gatekeepers of bile acid metabolism and host-microbiome crosstalk in the
875 gastrointestinal tract. PLOS Pathog 15:e1007581.

876 56. Rösch V, Denger K, Schleheck D, Smits THM, Cook AM. 2008. Different bacterial
877 strategies to degrade taurocholate. Arch Microbiol 190:11–18.

878 57. Balkwill DL, Drake GR, Reeves RH, Fredrickson JK, White DC, Ringelberg DB,
879 Chandler DP, Romine MF, Kennedy DW, Spadoni CM. 1997. Taxonomic study of
880 aromatic-degrading bacteria from deep-terrestrial- subsurface sediments and
881 description of *Sphingomonas aromaticivorans* sp. nov., *Sphingomonas*
882 *subterranea* sp. nov., and *Sphingomonas stygia* sp. nov. Int J Syst Bacteriol
883 47:191–201.

884 58. Zipper C, Nickel K, Angst W, Kohler H-PE. 1996. Complete microbial degradation
885 of both enantiomers of the chiral herbicide mecoprop [(RS)-2-(4-chloro-2-
886 methylphenoxy)propionic acid] in an enantioselective manner by *Sphingomonas*
887 *herbicidovorans* sp. nov. Appl Environ Microbiol 62:4318–4322.

888 59. Fredrickson JK, Brockman FJ, Workman DJ, Li SW, Stevens TO. 1991. Isolation

889 and characterization of a subsurface bacterium capable of growth on toluene,
890 naphthalene, and other aromatic compounds. Appl Environ Microbiol 57:796–803.
891 60. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J,
892 Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden
893 MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of
894 *Escherichia coli* K-12. Science 277:1453–1462.
895 61. Jagmann N, Brachvogel H-P, Philipp B. 2010. Parasitic growth of *Pseudomonas*
896 *aeruginosa* in co-culture with the chitinolytic bacterium *Aeromonas hydrophila*.
897 Environ Microbiol 12:1787–1802.
898 62. Bertani G. 1951. Studies on lysogenesis. I. The mode of phage liberation by
899 lysogenic *Escherichia coli*. J Bacteriol 62:293–300.
900 63. Kovach ME, Elzer PH, Steven Hill D, Robertson GT, Farris MA, Roop RM,
901 Peterson KM, Elzer A' PH, Steven Hill D, Robertson GT, Farris MA, Martin R, Ii R,
902 Peterson KM, Peterson KM. 1995. Four new derivatives of the broad-host-range
903 cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene
904 166:175–176.
905 64. Gade D, Thiermann J, Markowsky D, Rabus R. 2003. Evaluation of two-
906 dimensional difference gel electrophoresis for protein profiling. J Mol Microbiol
907 Biotechnol 5:240–251.
908 65. Strijkstra A, Trautwein K, Roesler S, Feenders C, Danzer D, Riemenschneider U,
909 Blasius B, Rabus R. 2016. High performance CCD camera system for
910 digitalisation of 2D DIGE gels. Proteomics 16:1975–1979.
911 66. Schnaars V, Dörries M, Hutchins M, Wöhlbrand L, Rabus R. 2018. What's the
912 Difference? 2D DIGE Image Analysis by DeCyder™ versus SameSpots™. J Mol

913 Microbiol Biotechnol 28:128–136.
914 67. Koßmehl S, Wöhlbrand L, Drüppel K, Feenders C, Blasius B, Rabus R. 2013.
915 Subcellular protein localization (cell envelope) in *Phaeobacter inhibens* DSM
916 17395. Proteomics 13:2743–2760.
917 68. Wöhlbrand L, Rabus R, Blasius B, Feenders C. 2017. Influence of NanoLC
918 column and gradient length as well as MS/MS frequency and sample complexity
919 on shotgun protein identification of marine bacteria. J Mol Microbiol Biotechnol
920 27:199–212.
921 69. Neuhoff V, Arold N, Taube D, Ehrhardt W. 1988. Improved staining of proteins in
922 polyacrylamide gels including isoelectric focusing gels with clear background at
923 nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250.
924 Electrophoresis 9:255–262.
925 70. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL.
926 2008. NCBI BLAST: a better web interface. Nucleic Acids Res 36:W5–9.
927 71. Altschul SF, W G, W M, Myers EW, J LD. 1990. Basic local alignment search tool.
928 J Mol Biol 215:403–410.
929 72. Needleman SB, Wunsch CD. 1970. A general method applicable to the search for
930 similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453.
931 73. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang
932 HY, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R,
933 Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G,
934 Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi
935 MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-
936 Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE,

Yong SY, Finn RD. 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res 47:D351–D360.

74. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P. 2019. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314.

75. Bergstrand LH, Cardenas E, Holert J, van Hamme JD, Mohn WW. 2016. Delineation of steroid-degrading microorganisms through comparative genomic analysis. MBio 7:1–14.

76. Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol 7:e1002195.

77. Wells JE, Hylemon PB. 2000. Identification and characterization of a bile acid 7α-dehydroxylation operon in *Clostridium* sp. strain TO-931, a highly active 7α-dehydroxylating strain isolated from human feces. Appl Environ Microbiol 66:1107–1113.

Legends to the Figures

Fig 1. Section of cholate degradation via the $\Delta^{1,4}$- and $\Delta^{4,6}$-variants of the 9,10-*seco*-pathway. Solid lines: known reactions, dotted lines: predicted reactions. Blue: A-ring oxidation and 7-OH elimination (phase 1), green: side-chain degradation (phase 2), yellow: B-ring cleavage (phase 3), orange: degradation of the 9,10-*seco*-steroid (phase 4). Compound abbreviations: IV, 12β-DHADD (7α,12β-dihydroxy-androsta-1,4-diene-3,17-dione); V, THSATD (3,7,12-trihydroxy-9,10-*seco*-androsta-1,3,5(10)-triene-9,17-dione); VI, DH-HIP (3',7-dihydroxy-H-methyl-hexahydro-indanone-propanoate); VII, HOCDA (12α-hydroxy-3-oxo-4,6-choldienoic acid); VIII, HATD (12-hydroxy-androsta-1,4,6-triene-3,17-dione); IX, DHSATD (3,12β-dihydroxy-9,10-*seco*-androsta-1,3,5(10),6-tetraene-9,17-dione).

Fig 2. Chromosome 2 of *Sphingobium* sp. strain Chol11 highlighting steroid-specific proteome signatures. Distinct steroid degradation gene clusters are labelled and numbered according to the text. Rings from outside to inside: (i) Localization of genes encoding orthologs to enzymes involved in different phases of steroid degradation based on reciprocal BLASTp analyses similar to (29) (Table S1), (ii-iv) Genes encoding proteins with >|1.5|-fold abundance changes under at least one test condition according to 2D-DIGE with glucose-grown cells as reference, abundance changes for cells grown with (ii) cholate, (iii) deoxycholate or (iv) 12β-DHADD, (v) coding sequences transcribed in clockwise direction and (vi) coding sequences transcribed in counter-clockwise direction.

Fig 3. Proposed model of bile-salt degradation in *Sphingobium* sp. strain Chol11 based on integrated bioinformatic, differential proteome and physiological analyses. Gene

977 locus tags are shortened to their numbers, e.g. 6 for Nov2c6. Color coding: (i) enzymes:
978 red, function demonstrated experimentally; black, function predicted due to reciprocal
979 BLASTp analyses with characterized steroid degradation proteins; grey, function
980 predicted due to chromosomal location and automatic annotation; (ii) intermediates: red,
981 intermediates identified in cell-free supernatants of strain Chol11 cultures degrading
982 cholate; black, intermediates predicted due to pathway prediction; (iii) background: blue,
983 A-ring oxidation and 7-OH elimination (phase 1); green, side-chain degradation (phase
984 2); yellow, B-ring cleavage (phase 3); orange, degradation of the 9,10-*seco*-steroid
985 (phase 4). Compound abbreviations: XIII, DOCDA (3,12-dioxo-4,6-choldienoc acid); XIV,
986 Δ^{22}-DOCTRA (3,17-dioxo-4,6,(22*E*)-choltrienoic acid); XV, Δ^{3}-7-OH-HIP (7-hydroxy-H-
987 methyl-hexahydro-indanone-3-propenoate).

988 **Fig 4.** Differential protein expression in *Sphingobium* sp. strain Chol11 adapted to
989 growth with cholate (CA, I in Fig 1), deoxycholate (DCA, XX in Fig 7), and 12β-DHADD
990 (DHADD, IV in Fig 1) compared to glucose (G) as reference state. Proteins are ordered
991 according to their location on chromosome 2. In most cases, results from only one
992 protein identification and quantitation approach are displayed: 2D-DIGE (black to red
993 gradient: protein abundance fold changes with glucose-grown cells as reference state),
994 shotgun-MS analysis (Mascot scores in blue) or MS analysis of the membrane protein-
995 enriched fraction (Mascot scores in green). In case of protein identification by multiple
996 approaches, priority was for 2D DIGE followed by shotgun analyses. Proteins not
997 identified with any method are shown in grey. The complete dataset can be found in
998 Table S1. Red font: known proteins with experimentally verified functions.

Fig 5. (A) Steroid degradation gene cluster 1 (*nov2c06–20*) of *Sphingobium* sp. strain Chol11 as well as similar clusters from other bile-salt degrading strains. (B) Stereoinversion of the 12-OH of the cholate-degradation intermediate DHADD as found in *C. testosteroni* (38). (C) Removal of the 7-OH of cholate-degradation intermediate DH-HIP, which corresponds to the 12-OH in cholate (23). Compound abbreviations: XVI, 12α-DHADD (7,12α-dihydroxy-androsta-1,4-diene-3,17-dione); XVI, HADT (7-hydroxy-adrosta-1,4-diene-3,12,17-trione); XVIII, 3-hydroxy-H-methyl-hexahydro-indanone-6-propenoate; XIX, 3-hydroxy-H-methyl-hexahydro-indanone-propanoate.

Fig 6. (A) Steroid degradation gene cluster 2 (*nov2c218–232*) of *Sphingobium* sp. strain Chol11 and similar clusters from other bile-salt degrading strains. (B) Bile-salt side chain degradation mechanism in *P. stutzeri* Chol1 (black enzyme names) (19, 21, 23, 24) and *R. jostii* RHA1 (grey enzyme names) (13). Colored: Corresponding enzymes in strain Chol11 (colors correspond to genes in A), bold: experimentally verified, light: suggested by bioinformatic analyses.

Fig 7. (A) Steroid degradation gene cluster 3 (*nov2c344–370*) of *Sphingobium* sp. strain Chol11 compared to the testosterone degradation gene cluster of *N. tardaugens* NBC16725 (47). Connection lines indicate homologs. (B) Proposed degradation of 7-hydroxy and 7-deoxy bile-salts in strain Chol11 (black) as opposed to degradation of 7-hydroxy bile salts via the $\Delta^{1,4}$-variant (grey). Compound abbreviations: XXII, 7-hydroxy-H-methyl-hexahydro-indanone-propanoate.

Fig 8. Distribution of steroid degradation gene orthologs in 38 *Sphingomonas, Sphingobium,* and *Novosphingobium* strains. Heat-map showing the BLAST identities for reciprocal BLASTp hits to Hsh2 from *Sphingobium* sp. strain Chol11 and the bile-salt

degradation proteins from *P. stutzeri* Chol1. Names of strains tested and found to use the $\Delta^{4,6}$-variant pathway are in red. For accession numbers of the genomes see Fig S4.

Fig 9. Degradation of bile salts by selected strains predicted to use the $\Delta^{4,6}$-variant of the steroid degradation pathway. (A) Structures of the tested bile salts and predicted intermediates detected in cell-free culture supernatants. (B,D,F) Growth on 1 mM bile salts. Colors of symbols correspond to colors of substrate names. Values are means of triplicates with standard deviation error bars, except in F where replicates on chenodeoxycholate are plotted separately. (C,E,G) Accumulation of intermediates during degradation of cholate in culture supernatants of the respective strains. See Fig 3 for structures not shown here. HPLC-UV chromatograms at 290 nm of culture supernatants. Steroid compounds were identified by retention time, absorbance spectrum and mass. Compound abbreviations: XXIV, 9,12-dihydroxy-androsta-4,6-diene-3,17-dione; XXV: 12-hydroxy-androsta-4,6-diene-3,17-dione.

Fig 10. Utilization of conjugated bile salts by *Sphingobium* sp. strain Chol11 and transformation of conjugated bile salts by the amidases Nov2c227 and Nov2c229. (A) Structures of conjugated bile salts and 7-keto metabolites. (B) Growth of strain Chol11 on 1 mM glycocholate (green squares) or glycine (yellow diamonds). (C) Transformation of glycocholate by whole cells for 11 days, and (D) by cell-free extracts for 15 h. In C, D, F and G: black trace, *E. coli* MG1655 pBBR1MCS-5 (empty vector control); light green, *E. coli* MG1655 pBBR1MCS-5::*nov2c227*; and dark green, *E. coli* MG1655 pBBR1MCS-5::*nov2c229*. (E) Growth of strain Chol11 on 1 mM taurocholate (green squares) and taurine (yellow diamonds). (F) Transformation of taurocholate by whole cells for 4 days and (G) by cell-free extracts for 15 h. Error bars indicate standard deviation (n=3).

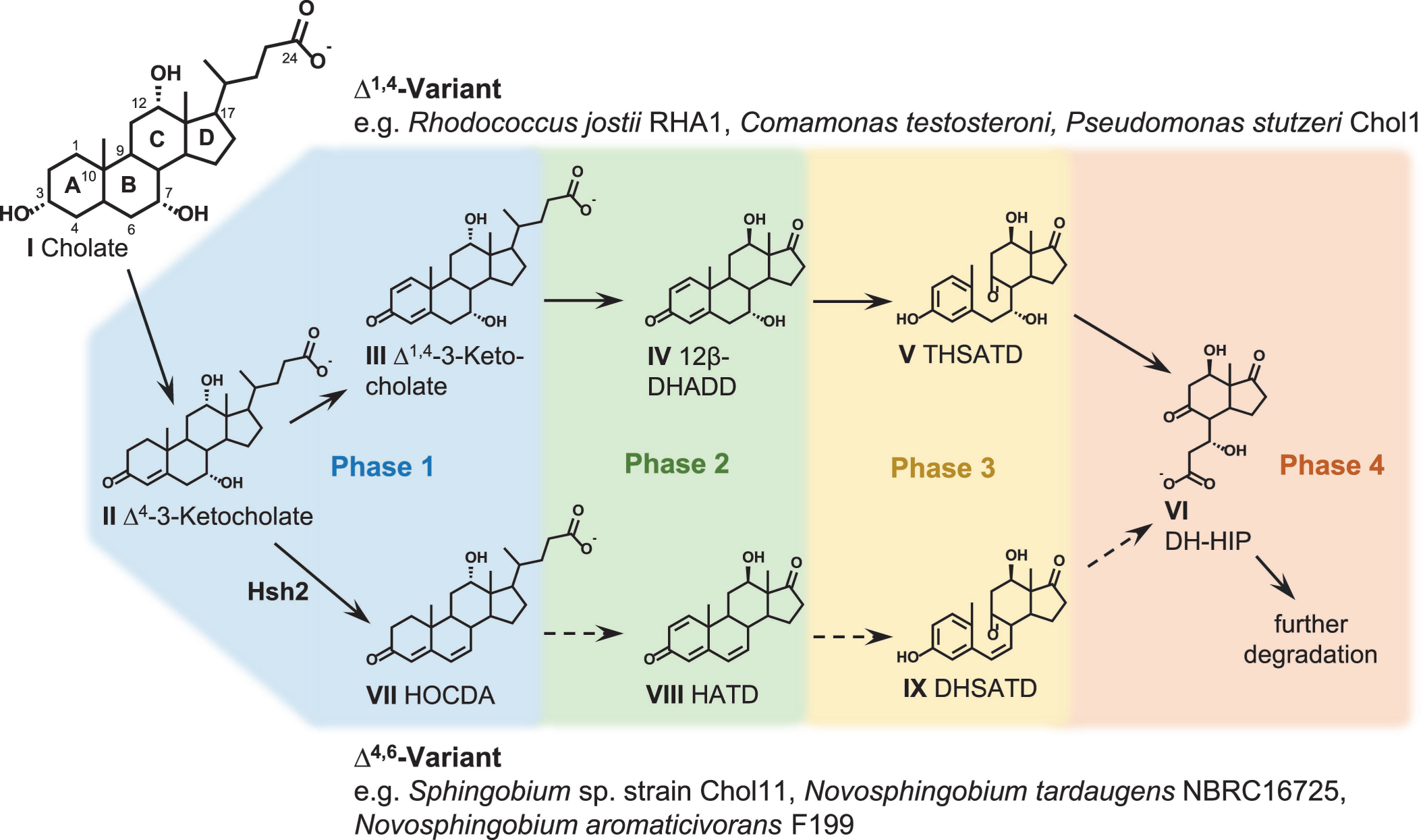
HPLC-MS base peak chromatograms measured in negative ion mode. Steroid compounds were identified by retention time, absorbance spectrum and mass. Masses are indicated for the respective deprotonated acids.

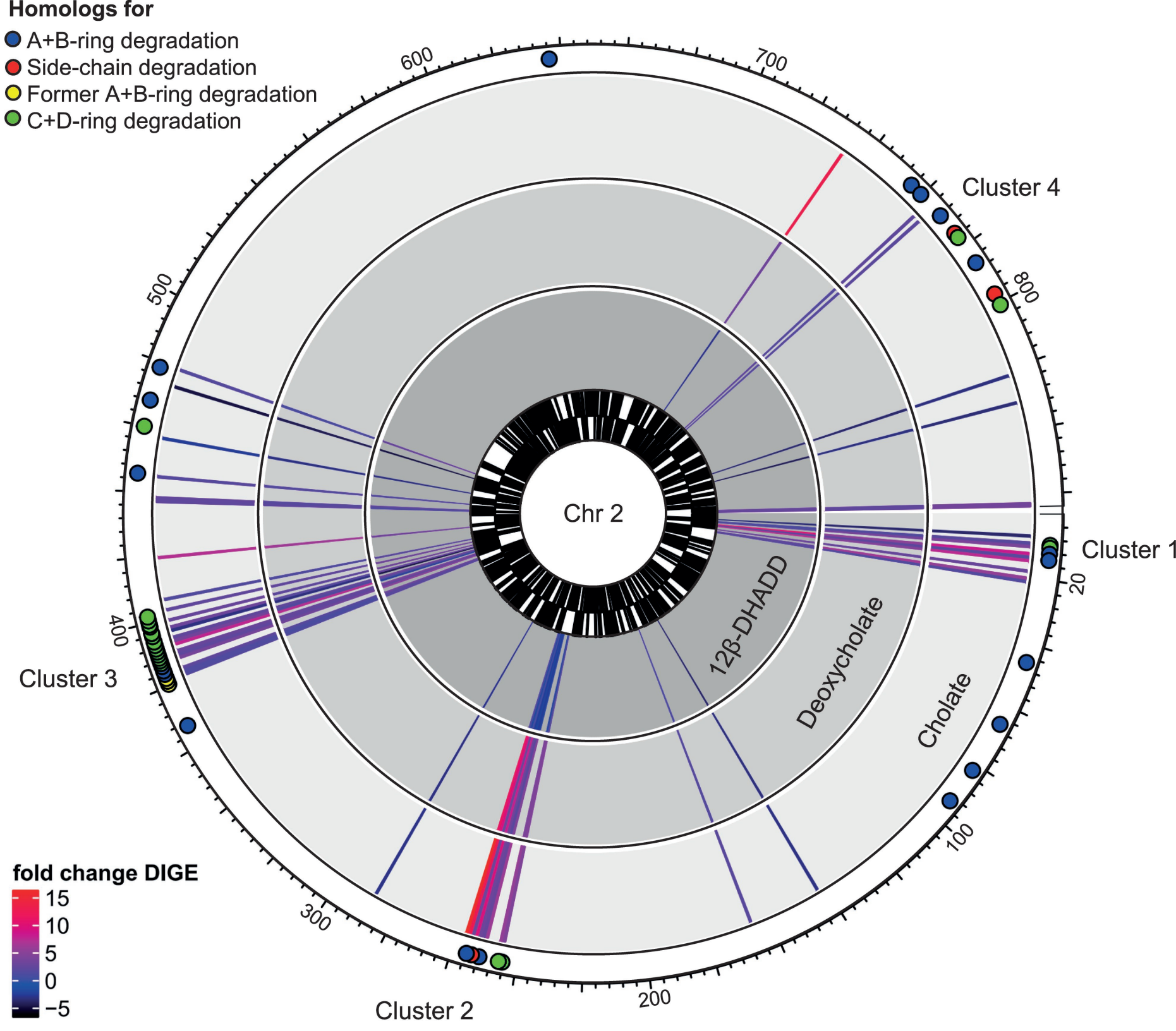
Tables

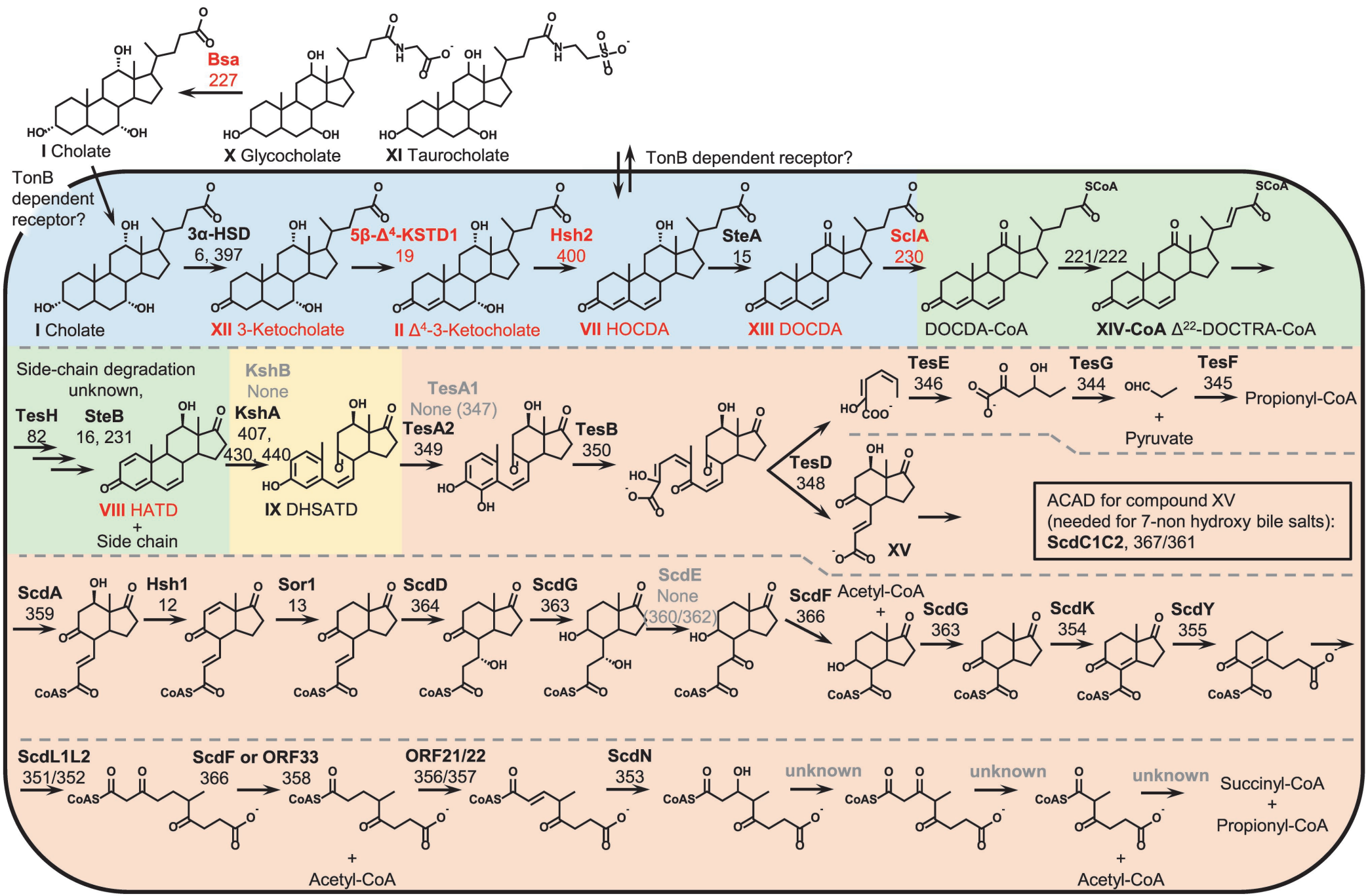
Tab 1. Primers used for construction of plasmids for heterologous expression.*

Name	Sequence	Restriction site
expfor_Chol11_nov2c227	AGACTCGAGGGCTAAGCCCTTCGACAAGC	XhoI
exprev_Chol11_nov2c227	AGATCTAGATTAAACCGCCGGACGCCTG	XbaI
expfor_Chol11_nov2c229	AGACTCGAGAGAAAATCAACAAGATCGC	XhoI
exprev_Chol11_nov2c229	AGATCTAGATTATTTGAGGTCAAAGACCT	XbaI
M13 for (-43)	AGGGTTTTCCCAGTCACGACGTT	
M13 rev (-49)	GAGCGGATAACAATTTCACACAGG	

* Underlined: restriction sites


I Cholate
Δ1,4-Variant
e.g. Rhodococcus jostii RHA1, Comamonas testosteroni, Pseudomonas stutzeri Chol1
II Δ4-3-Ketocholate
III Δ1,4-3-Keto-cholate
IV 12β-DHADD
V THSATD
VI DH-HIP
further degradation
Hsh2
VII HOCDA
VIII HATD
IX DHSATD
Phase 1
Phase 2
Phase 3
Phase 4
Δ4,6-Variant
e.g. Sphingobium sp. strain Chol11, Novosphingobium tardaugens NBRC16725, Novosphingobium aromaticivorans F199


Homologs for
A+B-ring degradation
Side-chain degradation
Former A+B-ring degradation
C+D-ring degradation
Chr 2
12β-DHADD
Deoxycholate
Cholate
Cluster 1
Cluster 2
Cluster 3
Cluster 4
20
100
200
300
400
500
600
700
800
fold change DIGE
15
10
5
0
−5

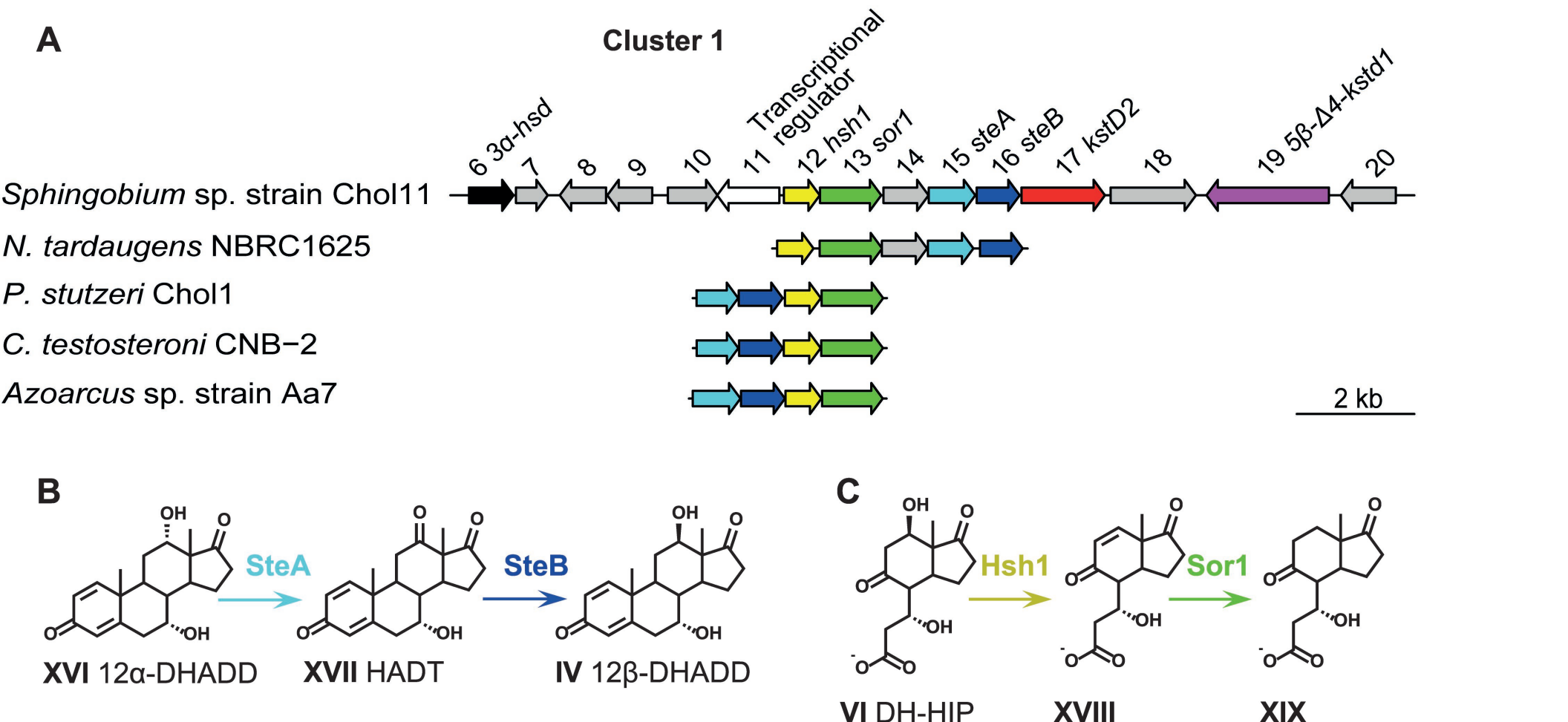

Bsa
227
I Cholate
X Glycocholate
XI Taurocholate
TonB dependent receptor?
TonB dependent receptor?
3α-HSD
6, 397
5β-Δ4-KSTD1
19
Hsh2
400
SteA
15
ScIA
230
221/222
SCoA
SCoA
I Cholate
XII 3-Ketocholate
II Δ4-3-Ketocholate
VII HOCDA
XIII DOCDA
DOCDA-CoA
XIV-CoA Δ22-DOCTRA-CoA
Side-chain degradation unknown,
TesH
82
SteB
16, 231
VIII HATD
+
Side chain
KshB
None
KshA
407, 430, 440
IX DHSATD
TesA1
None (347)
TesA2
349
TesB
350
TesD
348
XV
TesE
346
TesG
344
TesF
345
Propionyl-CoA
+
Pyruvate
ACAD for compound XV
(needed for 7-non hydroxy bile salts):
ScdC1C2, 367/361
ScdA
359
Hsh1
12
Sor1
13
ScdD
364
ScdG
363
ScdE
None
(360/362)
ScdF
366
Acetyl-CoA
+
ScdG
363
ScdK
354
ScdY
355
ScdL1L2
351/352
ScdF or ORF33
366
358
ORF21/22
356/357
ScdN
353
unknown
unknown
unknown
+
Acetyl-CoA
+
Acetyl-CoA
Succinyl-CoA
+
Propionyl-CoA

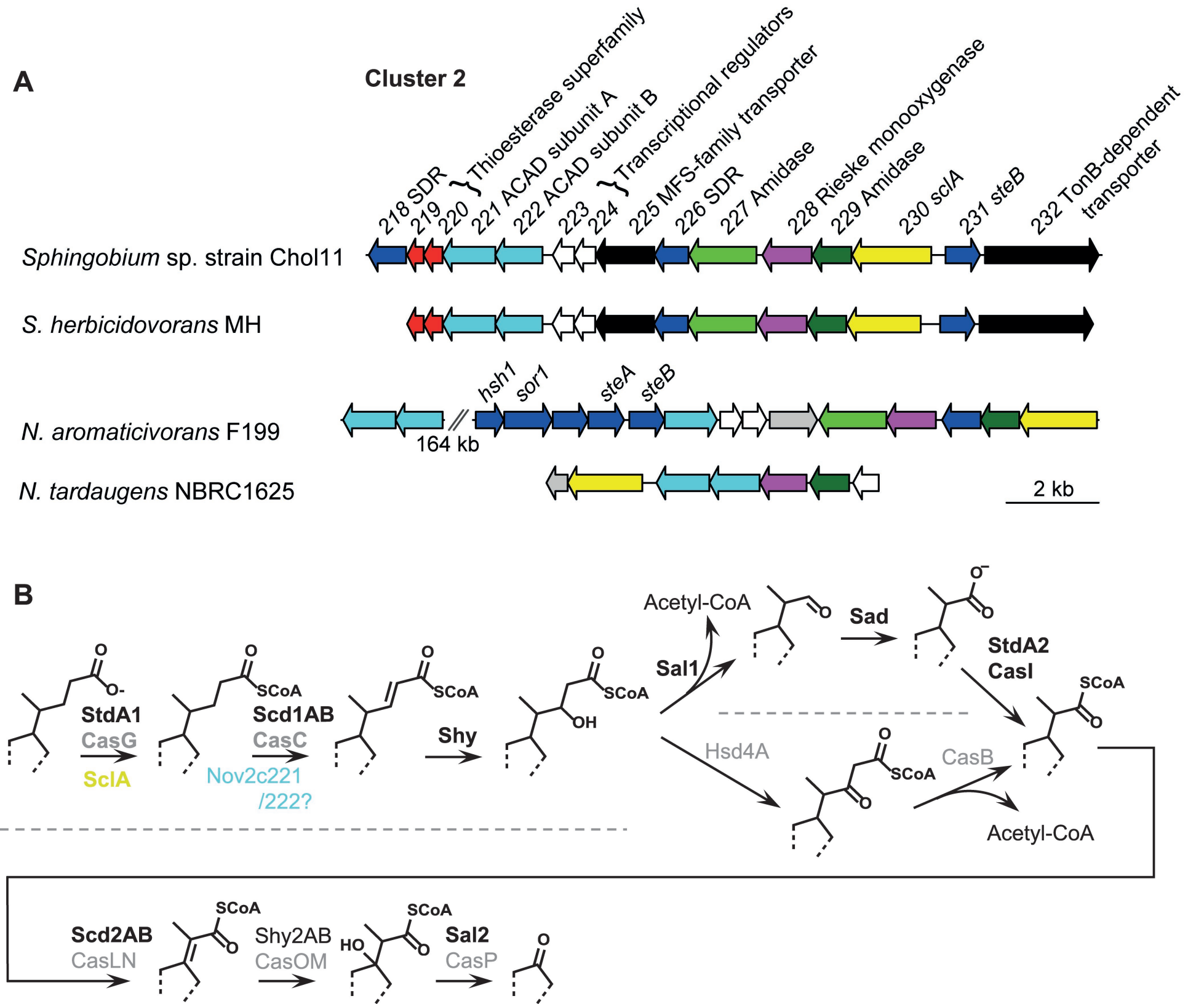
Locus Tag	Cluster	Name or predicted function	Phase of degradation	CA	DCA	DHADD	G
Nov2c6	1	3α-HSD	A-ring oxidation	[illegible]	[illegible]	[illegible]	[illegible]
Nov2c7	1	unknown	unknown	-2.7 / 507.4	-2.7 / 446.1	-2.5 / 291.7	- / 145.2
Nov2c8	1	unknown	unknown	134.5	173.2	141.6	0.0
Nov2c9	1	12β-HSD	Hydroxy group transformation	2.1	1.8	1.3	-
Nov2c10	1	Monooxygenase	unknown	4.2	3.7	2.0	-
Nov2c11	1	Transriptional regulator	Regulation				
Nov2c12	1	Hsh1	Hydroxy group transformation	7.9	5.1	9.7	-
Nov2c13	1	Sor1	Hydroxy group transformation	2.2	2.4	2.2	-
Nov2c14	1	3α-HSD	A-ring oxidation	7.3	4.6	6.9	-
Nov2c15	1	SteA	Hydroxy group transformation	838.5	699.6	757.8	135.3
Nov2c16	1	SteB	Hydroxy group transformation	651.8	521.5	586.5	187.3
Nov2c17	1	5α-Δ^4-KSTD	A-ring oxidation	3.1	2.9	3.2	-
Nov2c18	1	Monooxygenase	unknown				
Nov2c19	**1**	**5β-Δ^4-KSTD1**	**A-ring oxidation**	5.5	6.1	2.6	-
Nov2c20	1	unknown	unknown	1.8	1.7	2.1	-
Nov2c66	none	KshA	B-ring cleavage	0.0	0.0	0.0	176.0
Nov2c82	none	Δ^1-KSTD	A-ring oxidation				
Nov2c150	none	TonB-dependent receptor	Transport	2.1	1.9	1.3	-
Nov2c218	2	7α-HSD	Hydroxy group transformation	365.1	378.4	0.0	0.0
Nov2c219	2	Thioesterase	unknown	300.1	301.4	0.0	0.0
Nov2c220	2	Thioesterase	unknown	60.5	0.0	0.0	0.0
Nov2c221	2	ACAD subunit A	Side-chain degradation	4.4	4.5	1.9	-
Nov2c222	2	ACAD subunit B	Side-chain degradation	5.5	4.3	-1.1	-
Nov2c223	2	Transriptional regulator	Regulation	222.5	164.3	0.0	156.5
Nov2c224	2	Transriptional regulator	Regulation	133.8	102.8	0.0	0.0
Nov2c225	2	MFS transporter	Transport	262.0	169.8	0.0	0.0
Nov2c226	2	7α-HSD	Hydroxy group transformation	5.9	4.4	-1.2	-
Nov2c227	**2**	**Bsa**	**Degradation of conjugated bile salts**	3.0	2.3	-1.6	-
Nov2c228	2	KshA	B-ring cleavage?	3.4	3.5	-1.2	-
Nov2c229	2	Amidase	unknown	8.9	7.8	-1.5	-
Nov2c230	**2**	**SclA**	**Side chain**	3.3	3.3	1.3	-
Nov2c231	2	12β-HSD	AB ring	11.7	8.8	-1.0	-
Nov2c232	2	TonB-dependent receptor	Transport	15.0	10.5	1.6	-
Nov2c344	3	TesG	Steroid nucleus degradation	1.8	1.6	2.3	-
Nov2c345	3	TesF	Steroid nucleus degradation	1.7	1.5	2.3	-
Nov2c346	3	TesE	Steroid nucleus degradation	2.5	2.0	3.3	-
Nov2c347	3	Flavin reductase	(Steroid nucleus degradation)	263.1	227.1	280.3	0.0
Nov2c348	3	TesD	Steroid nucleus degradation	526.8	425.8	505.4	333.2
Nov2c349	3	TesA2	Steroid nucleus degradation	3.8	3.4	4.8	-
Nov2c350	3	TesB	Steroid nucleus degradation	2.6	2.7	2.8	-
Nov2c351	3	ScdL1	Steroid nucleus degradation	3.1	2.9	4.0	-
Nov2c352	3	ScdL2	Steroid nucleus degradation	1026.7	937.5	1069.0	390.0
Nov2c353	3	ScdN	Steroid nucleus degradation	7.5	7.1	-1.4	-
Nov2c354	3	ScdK	Steroid nucleus degradation	3.7	1.8	1.9	-
Nov2c355	3	ScdY	Steroid nucleus degradation	2.7	2.6	3.4	-
Nov2c356	3	ORF21	Steroid nucleus degradation	882.2	854.4	957.3	287.2
Nov2c357	3	ORF22	Steroid nucleus degradation	-2.2 / 674.5	-2.3 / 584.3	-3.4 / 576.3	- / 194.4
Nov2c358	3	ORF33	Steroid nucleus degradation	3.3	2.2	3.0	-
Nov2c359	3	CoA-transferase	Steroid nucleus degradation	245.5	212.7	290.0	0.0
Nov2c360	3	ScdE	Steroid nucleus degradation	3.0	2.4	3.4	-
Nov2c361	3	ScdC2	Steroid nucleus degradation	503.4	475.5	423.3	0.0
Nov2c362	3	Alcohol dehydrogenase	unknown	260.9	155.9	323.8	0.0
Nov2c363	3	ScdG	Steroid nucleus degradation	2.6	2.5	3.4	-
Nov2c364	3	ScdD	Steroid nucleus degradation	663.2	603.9	590.0	322.0
Nov2c365	3	Steroid-Δ^5-isomerase	unknown	402.5	314.9	374.1	312.9
Nov2c366	3	ScdF	Steroid nucleus degradation	1.8	1.6	2.2	-
Nov2c367	3	ScdC1	Steroid nucleus degradation	1285.2	1196.5	1304.5	398.8
Nov2c378	none	TonB-dependent receptor	Transport	7.0	6.0	3.5	-
Nov2c397	none	3α-HSD	A-ring oxidation	786.8	710.2	630.9	647.7
Nov2c400	**none**	**Hsh2**	**Elimination of 7α-OH**	388.7	247.4	160.1	162.5
Nov2c401	none	TonB-dependent receptor	Transport	2.1	2.0	1.8	-
Nov2c407	none	KshA	B-ring cleavage	966.3	888.2	928.0	565.9
Nov2c430	none	KshA	B-ring cleavage	0.0	54.9	140.7	0.0
Nov2c440	none	KshA	B-ring cleavage	944.6	855.4	0.0	578.0
Nov2c659	4	TonB-dependent receptor	Transport	12.0	3.4	-2.0	-
Nov2c683	4	3α-HSD	A-ring oxidation	767.3	674.6	598.9	426.3
Nov2c687	4	12α-HSD	Hydroxy group transformation	639.0	513.7	252.9	0.0
Nov2c695	4	Δ^1-KSTD	A-ring oxidation	704.3	672.9	569.0	234.7

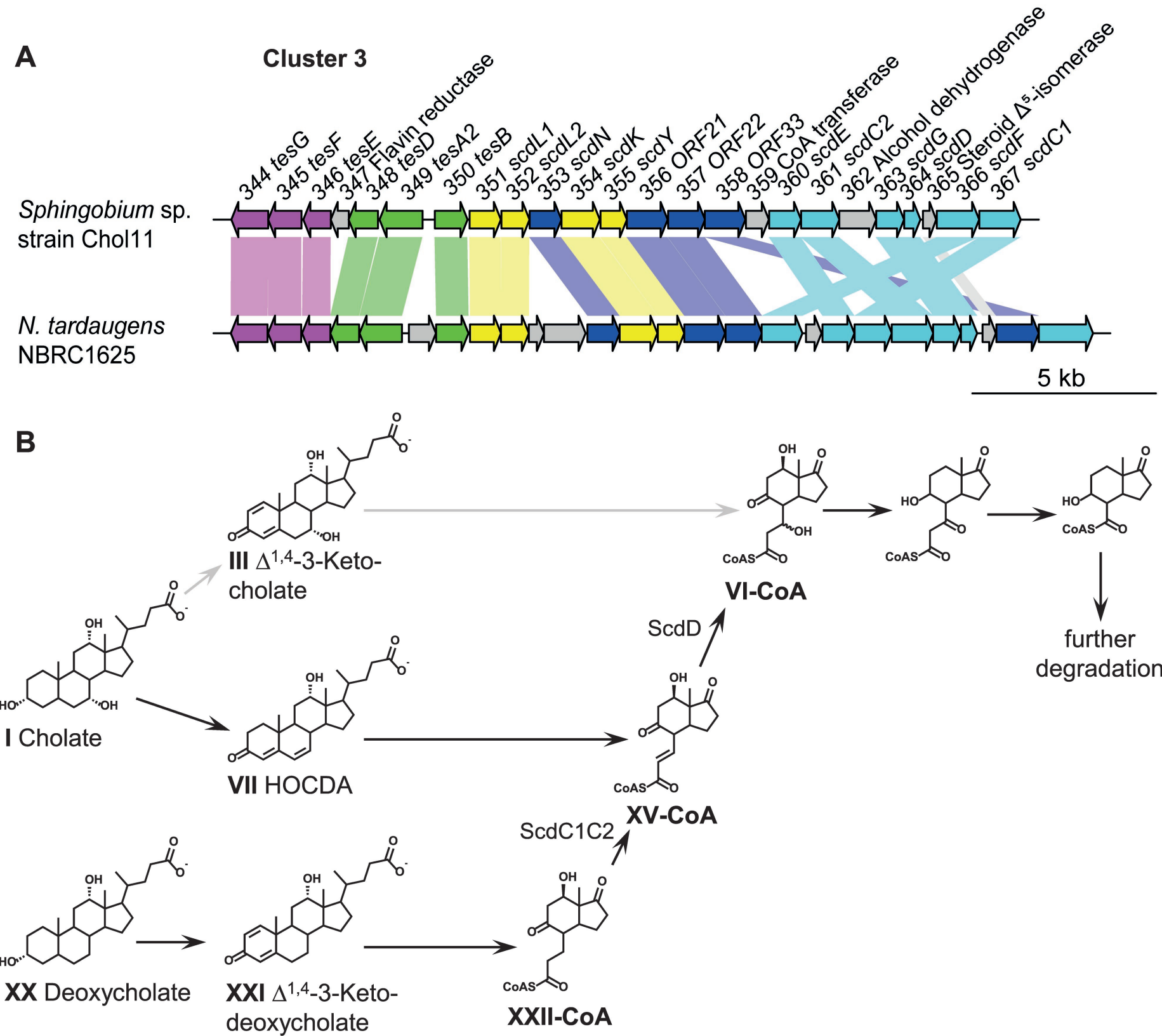
2D-DIGE
15.0
10.0
5.0
0.0
-5.0

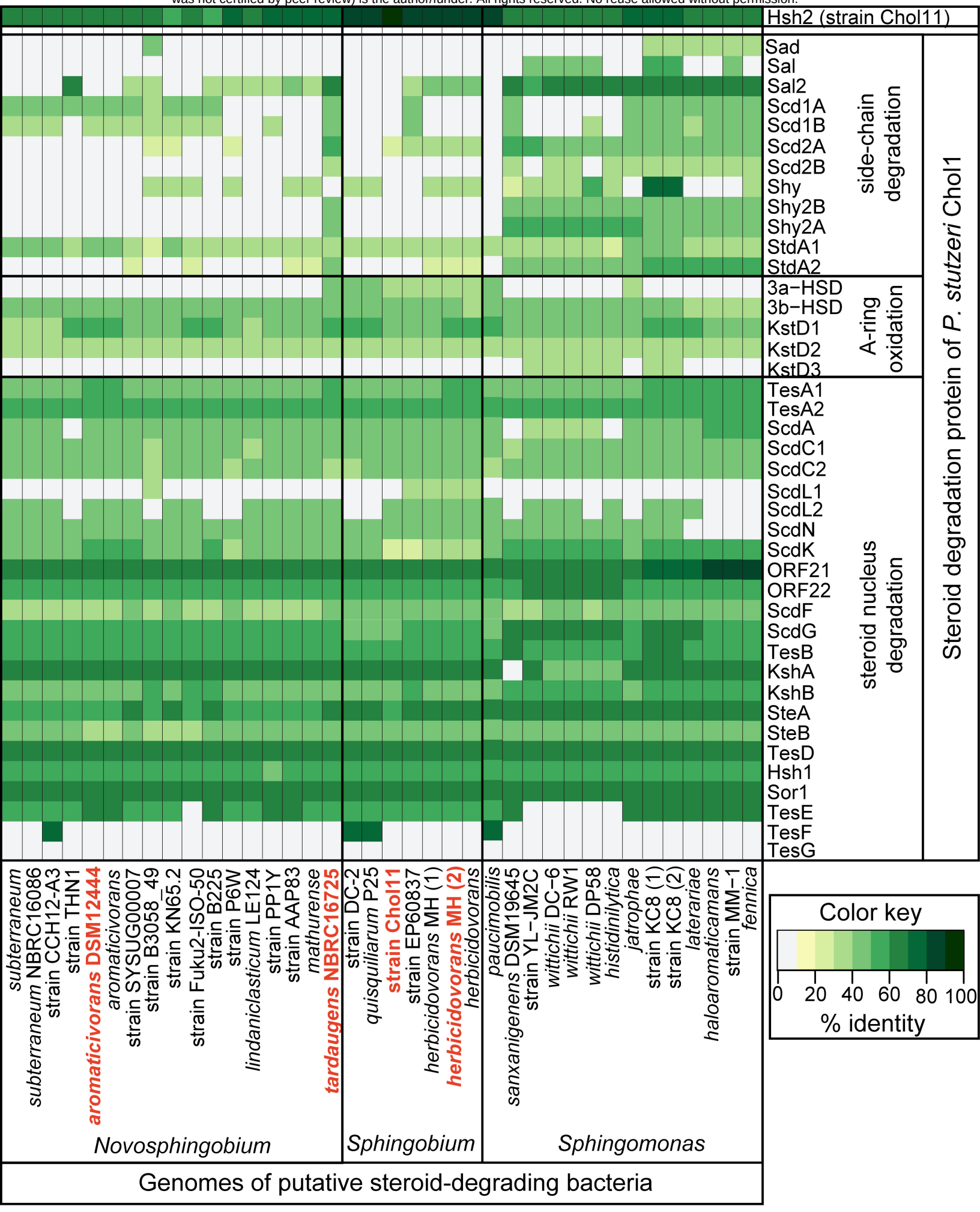
Shotgun MS
0.0
100.0
500.0
1000.0
1500.0

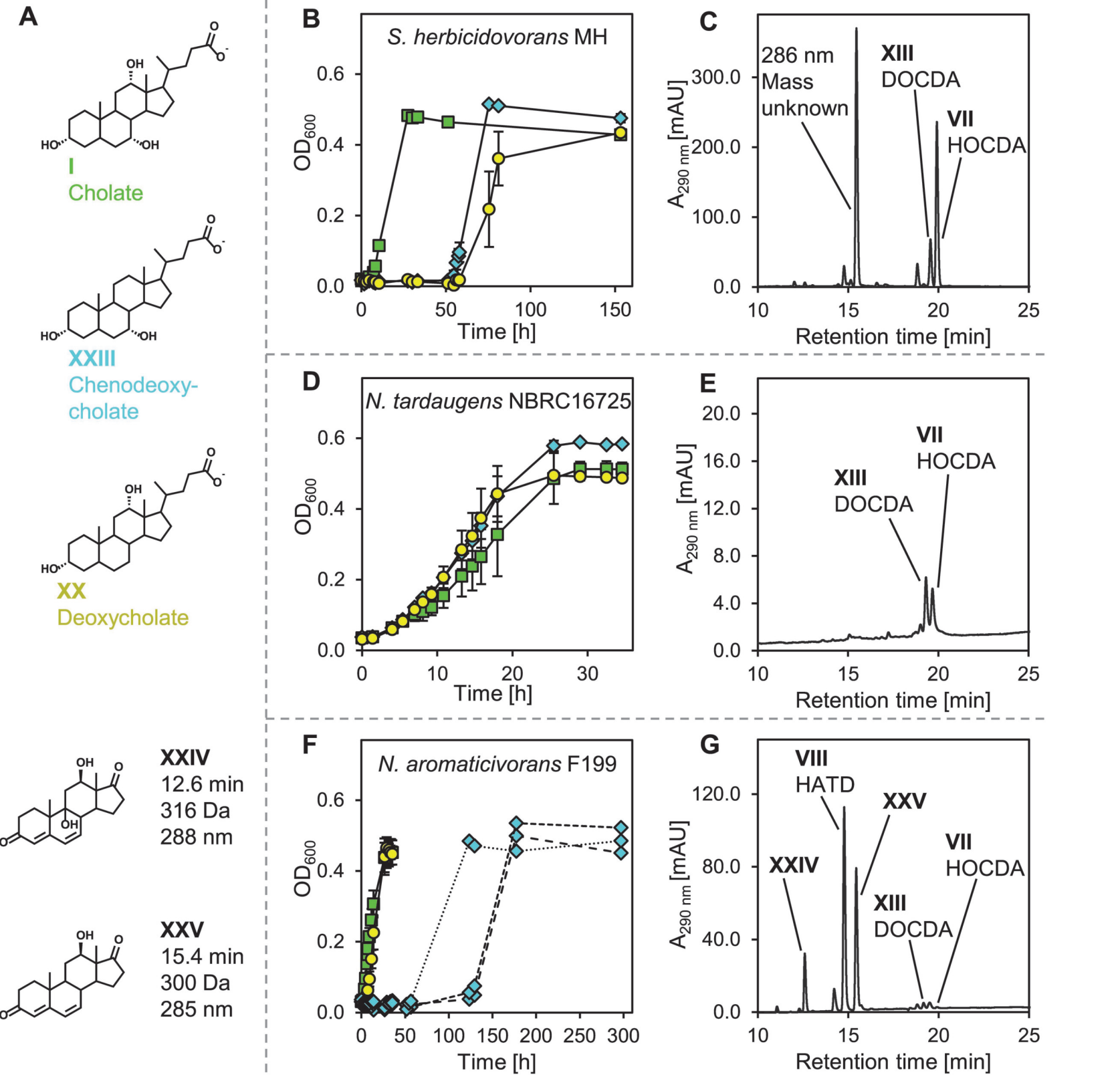
Membrane
0.0
150.0
300.0

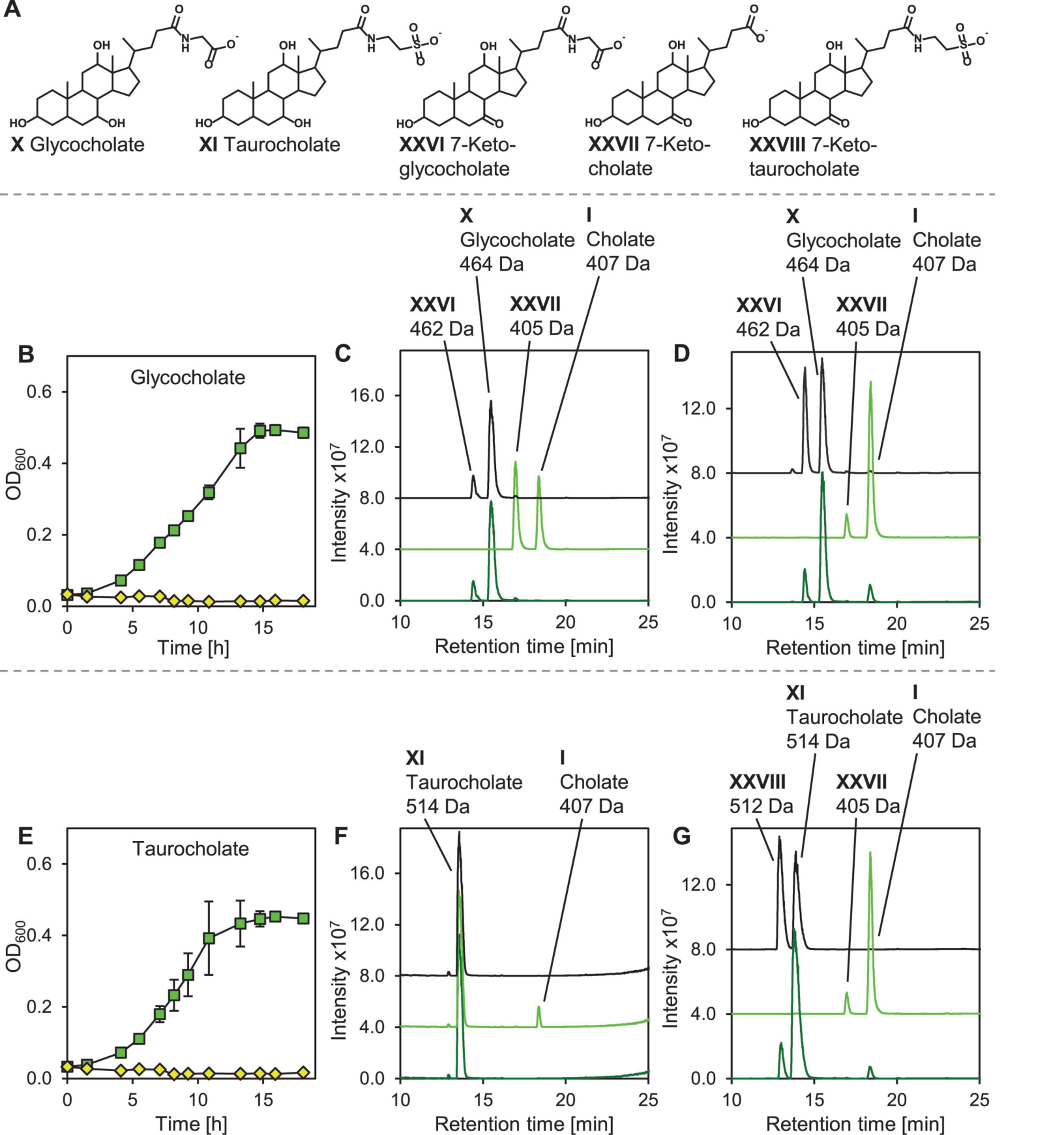

A
Cluster 1
6 3α-hsd
7
8
9
10
11 Transcriptional regulator
12 hsh1
13 sor1
14
15 steA
16 steB
17 kstD2
18
19 5β-Δ4-kstd1
20
Sphingobium sp. strain Chol11
N. tardaugens NBRC1625
P. stutzeri Chol1
C. testosteroni CNB−2
Azoarcus sp. strain Aa7
2 kb
B
SteA
SteB
XVI 12α-DHADD
XVII HADT
IV 12β-DHADD
C
Hsh1
Sor1
VI DH-HIP
XVIII
XIX


A
Cluster 2
218 SDR
219
220
Thioesterase superfamily
221 ACAD subunit A
222 ACAD subunit B
Transcriptional regulators
223
224
225 MFS-family transporter
226 SDR
227 Amidase
228 Rieske monooxygenase
229 Amidase
230 sclA
231 steB
232 TonB-dependent transporter
Sphingobium sp. strain Chol11
S. herbicidovorans MH
hsh1
sor1
steA
steB
N. aromaticivorans F199
164 kb
N. tardaugens NBRC1625
2 kb
B
StdA1
CasG
SclA
Scd1AB
CasC
Nov2c221
/222?
Shy
Sal1
Acetyl-CoA
Sad
StdA2
CasI
Hsd4A
CasB
Acetyl-CoA
Scd2AB
CasLN
Shy2AB
CasOM
Sal2
CasP
SCoA
OH
HO


A
Cluster 3
344 tesG
345 tesF
346 tesE
347 Flavin reductase
348 tesD
349 tesA2
350 tesB
351 scdL1
352 scdL2
353 scdN
354 scdK
355 scdY
356 ORF21
357 ORF22
358 ORF33
359 CoA transferase
360 scdE
361 scdC2
362 Alcohol dehydrogenase
363 scdG
364 scdD
365 Steroid Δ5-isomerase
366 scdF
367 scdC1
Sphingobium sp. strain Chol11
N. tardaugens NBRC1625
5 kb
B
III Δ1,4-3-Keto-cholate
VI-CoA
further degradation
I Cholate
VII HOCDA
ScdD
XV-CoA
ScdC1C2
XX Deoxycholate
XXI Δ1,4-3-Keto-deoxycholate
XXII-CoA




A
I
Cholate
XXIII
Chenodeoxy-
cholate
XX
Deoxycholate
B
S. herbicidovorans MH
OD600
Time [h]
C
286 nm
Mass
unknown
XIII
DOCDA
VII
HOCDA
A290 nm [mAU]
Retention time [min]
D
N. tardaugens NBRC16725
OD600
Time [h]
E
VII
HOCDA
XIII
DOCDA
A290 nm [mAU]
Retention time [min]
XXIV
12.6 min
316 Da
288 nm
XXV
15.4 min
300 Da
285 nm
F
N. aromaticivorans F199
OD600
Time [h]
G
VIII
HATD
XXV
XXIV
XIII
DOCDA
VII
HOCDA
A290 nm [mAU]
Retention time [min]


A
X Glycocholate
XI Taurocholate
XXVI 7-Keto-glycocholate
XXVII 7-Keto-cholate
XXVIII 7-Keto-taurocholate
B
Glycocholate
OD600
Time [h]
C
X Glycocholate 464 Da
I Cholate 407 Da
XXVI 462 Da
XXVII 405 Da
Intensity x10^7
Retention time [min]
D
X Glycocholate 464 Da
I Cholate 407 Da
XXVI 462 Da
XXVII 405 Da
Intensity x10^7
Retention time [min]
E
Taurocholate
OD600
Time [h]
F
XI Taurocholate 514 Da
I Cholate 407 Da
Intensity x10^7
Retention time [min]
G
XI Taurocholate 514 Da
I Cholate 407 Da
XXVIII 512 Da
XXVII 405 Da
Intensity x10^7
Retention time [min]