A metabolic pathway for bile acid dehydroxylation by the gut microbiome

The human gut microbiota harbour hundreds of metabolic pathways, most of which are encoded by genes that have not yet been identified\(^1\).\(^2\).\(^3\). Their small-molecule products are of interest for three reasons. First, most derive predominantly or exclusively from the microbiota (that is, there is no host source), and many enter the circulation, where they can have effects on peripheral tissues and organ systems. Second, their concentrations are similar to or exceed those of a typical drug; for example, indoxyl sulfate can accumulate in the human host at 130 mg per day\(^9\). Moreover, their concentration ranges are large, typically more than tenfold\(^8\), which could help to explain microbiome-mediated biological differences among people. Finally, of the few high-abundance molecules whose biological functions are well understood, most are ligands for a key host receptor; for example, short-chain fatty acids modulate host immune function via GPR41/GPR43 (refs.\(^{11}\)–\(^{13}\)). Thus, high-abundance, microbiota-derived molecules are responsible for a remarkably broad range of phenotypes conferred on the host by bacteria.

Among these pathways, 7α-dehydroxylation of the primary bile acids cholic acid and chenodeoxycholic acid (CDCA) is particularly notable because the organisms that carry it out are present at very low abundance—an estimated ratio of 1:10\(^{10}\) in a typical gut community\(^9\)–\(^{11}\)—yet they fully process a pool of primary bile acids that reaches concentrations of about 1 mM (ref. \(^{15}\)). Therefore, the flux through this pathway must be very high in the small subset of cells in which it operates, and the low-abundance organisms in the microbiome that perform this transformation have an unusually large impact on the pool of metabolites that enters the host. This pathway’s products—DCA and LCA—are the most abundant secondary bile acids in humans (up to 450–700 μM in caecal contents)\(^{16}\), and are known to be important in three biological contexts: prevention of \(C.\) \textit{difficile} outgrowth\(^{17}\), induction of hepatocellular carcinogenesis\(^{18}\), and modulation of host metabolic and immune responses\(^{17}\)–\(^{19}\). More broadly, DCA, LCA and their derivatives are a major component of the recirculating bile acid pool, representing more than 90% of the pool in the intestine and more than 25% in the gallbladder\(^{11}\). These microbiome-derived bile acids are therefore central to understanding the efficacy of therapeutics that target the bile acid pool and are approved or in clinical use.

The gut microbiota synthesize hundreds of molecules, many of which influence host physiology. Among the most abundant metabolites are the secondary bile acids deoxycholic acid (DCA) and lithocholic acid (LCA), which accumulate at concentrations of around 500 μM and are known to block the growth of \(C.\) \textit{difficile}\(^3\), promote hepatocellular carcinoma\(^2\) and modulate host metabolism via the G-protein-coupled receptor TGR5 (ref. \(^{3}\)). More broadly, DCA, LCA and their derivatives are major components of the recirculating pool of bile acids\(^5\); the size and composition of this pool are a target of therapies for primary biliary cholangitis and nonalcoholic steatohepatitis. Nonetheless, despite the clear impact of DCA and LCA on host physiology, an incomplete knowledge of their biosynthetic genes and a lack of genetic tools to enable modification of their native microbial producers limit our ability to modulate secondary bile acid levels in the host. Here we complete the pathway to DCA and LCA by assigning and characterizing enzymes for each of the steps in its reductive arm, revealing a strategy in which the A–B rings of the steroid core are transiently converted into an electron acceptor for two reductive steps carried out by Fe–S flavoenzymes. Using anaerobic in vitro reconstitution, we establish that a set of six enzymes is necessary and sufficient for the eight-step conversion of cholic acid to DCA. We then engineer the pathway into \(C.\) \textit{sporogenes}, conferring production of DCA and LCA on a nonproducing commensal and demonstrating that a microbiome-derived pathway can be expressed and controlled heterologously. These data establish a complete pathway to two central components of the bile acid pool.

---

1Department of Bioengineering and ChEM-H, Stanford University, Stanford, CA, USA. 2Department of Biochemistry, Albert Einstein College of Medicine, Bronx, NY, USA. 3Department of Chemistry, Indiana University, Bloomington, IN, USA. 4Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. 5Chan Zuckerberg Biohub, San Francisco, CA, USA. 6Present address: Translational Research Department, Daiichi Sankyo RD Novare Co. Ltd, Tokyo, Japan. 7These authors contributed equally: Masanori Funabashi, Tyler L. Grove. ✉ E-mail: steve.almo@einstein.yu.edu; fischbach@fischbachgroup.org
trials for treatment of primary biliary cholangitis and nonalcoholic steatohepatitis.

In 1980 it was shown that the gut bacterium *Clostridium scindens* VPI 12708 carries out the 7α-dehydroxylation of cholic acid to produce DCA20. The knowledge that cholic acid serves as an inducer of 7α-dehydroxylation led to the discovery of a bile-acid-induced operon (termed *bai*) containing eight genes (Fig. 1 and Extended Data Fig. 1). Through heterologous expression and characterization of individual *bai* gene products, enzymes have been attributed to each step of the oxidative arm of the pathway22–26, but the reductive arm remained poorly characterized26. A complete understanding of the pathway would enable efforts to control the composition of the bile acid pool by engineering the microbiome.

Here, by purifying and assaying pathway enzymes under anaerobic conditions, we reconstituted 7α-dehydroxylation in vitro. We then transferred the pathway from its genetically intractable producer *C. scindens* into *C. sporogenes*, conferring production of DCA and LCA on a nonproducing commensal bacterial species. These data establish a complete pathway for two central components of the bile acid pool, and they provide a genetic basis for controlling the bile acid output of the microbiome.

Reconstitution of 7α-dehydroxylation

We first set out to de-orphan the remaining steps in the 7α-dehydroxylation pathway. Because previous studies of the *bai* enzymes involved expressing them individually in *Escherichia coli*, we reasoned that an alternative approach—in which enzymes are purified, mixed and assayed in vitro—could help to delineate the set of enzymes necessary and sufficient for 7α-dehydroxylation. Given that the eight-gene *bai* operon is shared among all known 7α-dehydroxylation strains, we focused our efforts on the enzymes encoded by the operon. We cloned three orthologues of each enzyme, expressed them individually in *E. coli* under microaerobic conditions, and purified them anaerobically as amino-terminal His6 fusions. Using this strategy, we obtained at least one soluble, purified orthologue of each Bai enzyme (Extended Data Fig. 2). When we incubated a mixture of the purified Bai enzymes with cholic acid, nicotinamide adenine dinucleotide (NAD)\(^+\), coenzyme A and ATP under anaerobic conditions and monitored the reaction by liquid chromatography with mass spectrometry (LC–MS), we observed the time-dependent conversion of cholic acid to DCA, indicating that the combination of BaiB, BaiCD, BaiA2, BaiE, BaiF, and BaiH is sufficient for 7α-dehydroxylation; no additional enzymes are required (Fig. 2a, b).

To test our hypotheses regarding the order of steps in the pathway, we performed stepwise reconstitutions in which enzymes were added one at a time and intermediates were allowed to build up at each step in the pathway (Fig. 2c). From these data, we draw two conclusions. First, the six enzymes used in the reconstitution are not just sufficient but also necessary, and the pathway proceeds according to the scheme shown in Fig. 2c. We directly observed mass ions consistent with each of the proposed intermediates, providing direct evidence for the previously proposed portion of the biosynthetic route. (See Supplementary Table 1 and Extended Data Fig. 3 for data supporting our provisional structural assignments; two important limitations are that we do not have authentic standards for all intermediates, and that the ability to distinguish bile acid isomers by LC–MS can be limited.) In spite of its conservation in all known dehydroxylating species, BaiF is dispensable for cholic acid dehydroxylation in vitro. As BaiA is a predicted \(\Delta^2\)-ketosteroid isomerase, it may process a substrate other than cholic acid, probably one with a 4,5- or 5,6-olefin.

Second, to our surprise, the absence of BaiH caused the pathway to stall at the highly oxidized intermediate 3-oxo-4,5-6,7-didehydro-DCA, and its addition resulted in two successive 2\(\epsilon\) reductions to form 3-oxo-DCA. BaiH had previously been proposed to oxidize an alternative substrate, 3-oxo-4,5-dehydro-ursodeoxycholic acid25, so a potential role in the reductive arm of the pathway was unexpected. To explore this finding further, we incubated purified BaiH with synthetic 3-oxo-4,5-6,7-didehydro-DCA; we observed that the enzyme catalyses a 2\(\epsilon\) reduction to 3-oxo-4,5-dehydro-DCA, but does not reduce this intermediate further (Extended Data Fig. 4). Notably, 3-oxo-4,5-dehydro-DCA does not build up in the reconstitution reaction containing BaiH, suggesting that another enzyme present in the mixture catalyses the second reductive step. Hypothesizing that the BaiH homologue BaiCD catalyses the second reductive step, we incubated it with synthetic 3-oxo-4,5-dehydro-DCA, revealing that it reduces this substrate to 3-oxo-DCA (Extended Data Fig. 4). Together, these data show that the pathway uses an unusual redox strategy in which the A and B rings of the steroid core are converted into a highly oxidized intermediate, 3-oxo-4,5-6,7-didehydro-DCA; and that the two key reductive steps are catalysed by two homologous enzymes in the Fe–S flavoenzyme superfamily, BaiF and BaiCD.

Finally, the last step in the pathway—reduction of 3-oxo-DCA to DCA—is carried out by BaiA2, as confirmed by assaying purified BaiA2 alone (Extended Data Fig. 5). Thus, BaiA2 and BaiCD both act twice in the pathway, catalysing its first two and last two redox steps.

**Engineering the pathway into *C. sporogenes***

Having determined the set of enzymes that are necessary and sufficient for the pathway, we sought to gain genetic control over the pathway as a first step towards engineering the bile acid output of the gut community. We began by attempting to construct a mutation in the *baiCD* gene of the native producer, *C. scindens*, using the ClosTron group II intron system; however, we were unsuccessful owing to an inability to introduce DNA constructs into *C. scindens* by conjugation. As an alternative approach, we considered expressing the *bai* pathway in a gut commensal that is unable to carry out 7α-dehydroxylation; however, with notable exceptions29–32, methods for transferring pathways in *Clostridium* are underdeveloped. To our knowledge, no pathway from the human microbiome has been mobilized from one *Clostridium* species to another.

We selected *C. sporogenes* American Type Culture Collection (ATCC) strain 5579 as the recipient for two reasons: it is related to *C. scindens*, making it likely that ancillary metabolic requirements for the pathway (for example, cofactor biogenesis) would be met; and genetic tools have been developed that enable plasmids to be transformed into *C. sporogenes*33. Our initial attempts to clone the entire eight-gene *bai* operon (*baib–baif*) into an *E. coli*–*C. sporogenes* shuttle vector failed to yield clones harbouring the complete operon. Reasoning that there might be a gene in the cluster that is toxic to *E. coli*, we cloned various fragments of the cluster under the control of different promoters (detailed in Supplementary Table 2), eventually managing to split the cluster into three pieces, each in its own *E. coli*–*C. sporogenes* shuttle vector: *baib–baif* in pMTL83153 (pMF01), *baig* in pMTL8333 (pMF02),
and BaiH–BaiI in pMTL83253 (pMF03) (Fig. 3a and Extended Data Fig. 7). Genes in pMF01 and pMF03 were placed under the control of the spoIIE promoter from C. sporogenes ATCC 15579, which is expressed during the late stages of Clostridium growth, while BaiG in pMF02 was driven by the strong fdx promoter. We conjugated these plasmids sequentially into C. sporogenes to yield strain MF001.

When incubated with cholic acid, MF001 produces DCA in a time-dependent manner, in contrast with a control strain that contained only the transporter (baiG) (Fig. 3b, c), which does not. Additionally, MF001 converts CDCA to LCA (Extended Data Fig. 6). These data suggest that there might be a compensatory CoA hydrolase or that nonenzymatic hydrolysis of the CoA thioester happens to some extent in vivo.

**Identifying branch points in the pathway**

To uncover potential branch points for engineering the biosynthesis of non-native pathway products, we constructed a set of strains in which each of the eight genes was individually deleted (Extended Data Fig. 7). We grew these strains with cholic acid and assayed their culture supernatant for the build-up of intermediates (Fig. 3d). Deletion of genes in the oxidative arm of the pathway resulted in the build-up of early pathway intermediates, as expected. Two exceptions were the BaiF mutant, which produced only choly-CoA; and the BaiF-deficient strain, which generated a small quantity of the final product DCA, suggesting that there might be a compensatory CoA hydrolase or that nonenzymatic hydrolysis of the CoA thioester happens to some extent in vivo.

Intriguingly, the BaiH mutant accumulates a key intermediate in the reductive arm of the pathway, 3-oxo-4,5-6,7-didehydro-DCA (Fig. 3d), supporting our finding that BaiH catalyses the first reductive step in the pathway. Moreover, strains of C. sporogenes expressing BaiG/BaiH and BaiG/BaiCD convert, respectively, 3-oxo-4,5-6,7-didehydro-DCA to 3-oxo-4,5-dehydro-DCA and 3-oxo-4,5-dehydro-DCA to 3-oxo-DCA (Fig. 3e), providing access to intermediates that do not accumulate in a culture of C. scindens. Notably, the fully oxidized and partially reduced intermediates are branch points for the production of *allo* (5α) bile acids.
which have important biological activities, including the induction of regulatory T cells. Thus, gaining genetic control over the pathway by expressing it in an alternative gut microbe provides opportunities for the rational and deliberate control of bile acid metabolism and the production of alternative molecules with distinct biological properties.

Colonizing mice with engineered *C. sporogenes*

Finally, we colonized germ-free mice with MF001 to see whether it would confer the production of pathway products on the host. We included two other experimental groups: as a negative control, germ-free mice monocolonized by *C. sporogenes* plus *baiG* (the bile acid transporter); and as a positive control, germ-free mice monocolonized by wild-type *C. scindens*, a native *bai*-operon-containing (7α-dehydroxylating) strain. As shown in Fig. 4, the engineered *C. sporogenes* plus *baiB–baiI* strain (MF001) conferred production of DCA on the host. The level of production was substantially lower than that observed from *C. scindens*; we suspect that this is because of the need to include additional genes, not yet known, that couple the pathway to a pool of reduced cofactor, increasing flux. Nonetheless, our data...
provide an important starting point for efforts to study and engineer the bile acid pool.

Engineering pathways from the microbiome

Our results reveal the complete bile acid 7α-dehydroxylation pathway, bringing it closer to the level of knowledge we have about endogenous human metabolic pathways. Key features of the pathway might serve as a model for other pathways that produce high-abundance metabolites in the gut (see Supplementary Discussion and Extended Data Fig. 8).

The gut microbiome harbours hundreds of pathways, many of which may modulate host biology, but so far only a few have been the target of engineering. This stands in contrast to natural product pathways from terrestrial and marine microorganisms and plants, which are commonly expressed in heterologous hosts and engineered to generate non-native products. Two technology gaps need to be overcome in order to make microbiome-derived pathways amenable to engineering: first, we need efficient strategies to identify pathways for known metabolites and small-molecule products of orphan gene clusters, and second, we need tools for transferring pathways into bacterial hosts native to the gut and manipulating them to produce novel molecules.

The work described here is a starting point for these efforts. If it can be generalized to other Clostridia species, it could lead to a set of tools for de-orphaning, heterologously expressing, and engineering pathways from the microbiome.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2396-4.

Publisher's note

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

© The Author(s), under exclusive licence to Springer Nature Limited 2020
Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Bacterial strains, culture conditions and bile acids

C. scindens VP112708 and C. sporogenes ATCC 15579 were obtained from the Japan Collection of Microorganisms (JCM) and the American Type Culture Collection (ATCC), respectively. Engineered C. sporogenes strains used here are shown in Supplementary Table 3. They were cultured in TYG (3% w/v tryptone, 2% w/v yeast extract, 0.1% w/v sodium thioglycollate) broth at 37 °C in an anaerobic chamber from Coy Laboratories. E. coli CA434 (HB101/prK24) was cultured at 37 °C in LB broth supplemented with 12 μg ml⁻¹ tetracycline and 100 μg ml⁻¹ carbenicillin. In addition, 20 μg ml⁻¹ chloramphenicol, 100 μg ml⁻¹ spectinomycin or 250 μg ml⁻¹ erythromycin was used for the selection of series of plasmids of pMTL83153, pMTL83353 or pMTL83253 respectively. Plasmids used here are shown in Supplementary Table 2.

Heterologous expression in C. sporogenes

Bacterial cultures were incubated in a Coy anaerobic chamber under an atmosphere consisting of 10% CO₂, 5% H₂ and 85% N₂. Growth media were prereduced by overnight preincubation in the anaerobic chamber. Eight drops of 25 μl of the suspension were pipetted on a TYG agar plate without antibiotics and the plate was incubated anaerobic chamber at 37 °C for 2 days. The bacterial biomass was scraped up and resuspended in 300 μl of PBS. The whole cell suspension was then plated on TYG agar plates supplemented with 250 μg ml⁻¹ d-cycloserine and appropriate antibiotics (15 μg ml⁻¹ thiamphenicol for pMTL83153, 500 μg ml⁻¹ spectinomycin for pMTL83353 or 5 μg ml⁻¹ erythromycin for pMTL83253). After a few days, antibiotic-resistant colonies were picked and restreaked on agar containing the same antibiotic. The resulting clones were confirmed by PCR amplification using appropriate primers (Supplementary Table 4). Multiple plasmids were introduced sequentially, using the same procedure.

Extraction of metabolites

Engineered strains were cultured anaerobically in TYG medium supplemented with appropriate antibiotics from frozen glycerol stocks. We inoculated 10 μl of the overnight culture in 1 ml of TYG medium supplemented with appropriate antibiotics and 1 μM substrate. After 72 h, unless otherwise noted, the culture was extracted with 20% acetone and centrifuged. The supernatant was analysed by LC–MS.

LC–MS analysis of metabolite extracts

Metabolite extracts were analysed using an Agilent 1290 LC system coupled to an Agilent 6530 quadrupole time–of–flight (QTOF) mass spectrometer with a 1.7 μm, 2.1 mm × 50 mm Kinetex C18 column (Phenomenex). Water with 0.05% formic acid (A) and acetone with 0.05% formic acid (B) were used as the mobile phase at a flow rate of 0.35 ml min⁻¹ over a 32-min gradient: 0–1 min, 25%B; 1–25 min, 25–75%B; 25–26 min, 75–100%B; 26–30 min, 100%B; 30–32 min 75–25% B. All data were collected in negative-ion mode.

For detection of CoA conjugates and flavin cofactors, a 1.8 μm, 2.1 mm × 50 mm ZORBAX SB-C18 column (Agilent Technologies) and water with 10 mM ammonium acetate pH 9.0 (A) and acetonitrile (B) was used. A flow rate of 0.3 ml min⁻¹ was used over the 17-min gradient: 0–2 min, 15%B; 2–14 min, 15–50%B; 14–14.1 min 50–95%B, 14.1–17 min, 85%B. All data were collected in positive-ion mode.

Cloning of the bai operon

All amplification by polymerase chain reaction (PCR) was conducted using PrimeSTAR Max DNA polymerase (Takara Bio) according to the manufacturer’s instructions. Sequences of primers for target genes and cloning vectors are in Supplementary Table 4. For the heterologous expression of bai genes under the fixD promoter, pMTL vectors were amplified with primers 1 and 2. For the expression of bai genes under the spoIIIE promoter, pMTL vectors harbouring the spoIIIE promoter were constructed first. pMTL vectors were amplified with primers 1 and 3 to remove the spoIIIE promoter, and the spoIIIE promoter region, which is the 277-base-pair sequence upstream of CLOSPO_01065, was amplified with primers 4 and 5. Then these two PCR fragments were assembled by overlap PCR. The target gene sequences were amplified with primer pairs shown in Supplementary Table 4. PCR fragments were assembled with the amplified fragments of vectors using a Gibson assembly kit (New England Bio Labs). E. coli Stbl4 competent cells (Invitrogen) were transformed with the assembled plasmids by electroporation and transformants were confirmed by PCR. Positive clones containing the assembled plasmids were cultivated, with plasmids obtained by miniprep and verified by sequencing.

Cloning of bai operon genes

To increase the probability of assembling a complete bai operon, we cloned the genes encoding baiB, baiA2, baiCD, baiE, baiF, and baiH from C. scindens VP112708, Clostridium hylemonae and Clostridium hiranonis using the primers in Supplementary Table 5 and the KOD Xtremae Hot Start PCR kit (Millipore) according to the manufacturer’s protocol. Each PCR-amplified gene contains ligation-independent cloning (LIC) sites that are complementary to the pSGC vector. PCR products were purified with the Agencourt Ampure XP PCR clean-up kit (Beckman Coulter) according to the manufacturer’s protocol. The pSGC vector was prepared for LIC by linearization with the restriction enzyme Bsal. LIC sites were installed by adding T4 DNA polymerase (NEB) to 10 μl of linearized plasmid in a 50 μl reaction containing 2.5 mM GTP, 1× NEB buffer 2, and 1× bovine serum albumin (BSA) for 1 h at 22 °C. T4 DNA polymerase was heat-inactivated by incubation at 75 °C for 20 min. PCR products (in a volume of 2 μl) were treated with 50 μl of T4 DNA polymerase in a 10 μl reaction containing 2.5 mM CTP, 1× NEB buffer 2 and 1× BSA for 1 h at 22 °C. T4 DNA polymerase was heat-inactivated by incubation at 75 °C for 20 min. The LIC reaction was assembled by mixing 15 ng of digested vector DNA with roughly 40 ng of digested PCR product; the reaction mixture was then incubated at 22 °C for 10 min. A 30 μl aliquot of DH10B cells (NEB) was transformed with 2 μl of the LIC reaction mixture using standard bacterial transformation protocols. This cloning procedure adds a His₆ tag to the amino terminus of each protein with the following sequence: MHHHHHHSSGVDLGTENLYFQS. All final constructs were sequence verified (Genescript).

Expression and purification of BaiH and BaiCD

BL21 (DE3) cells containing the pPH151 plasmid were transformed with the pSGC plasmid containing either BaiCD or BaiH. The transformants were selected on an LB/agar plate containing 50 μg ml⁻¹ kanamycin and 34 μg ml⁻¹ chloramphenicol. A single colony was used to inoculate 20 ml of LB overnight culture containing the above antibiotics. The
overnight culture was used to inoculate 21 of Studier's autoinduction media (ZYP-5052 supplemented with 1 mM flavin mononucleotide and 200 μM FeCl₃) housed in a 21 Pyrex media bottle. Cultures were grown with constant aeration using a sparging stone attached to a pressurized, 0.22-μm filtered air source, all in a water bath maintained at 37°C. After 5 h, aeration was stopped and the culture was placed in an ice bath for 1 h. The culture was returned to a 22°C water bath and light aeration was resumed. After 5 min, cysteine was added to a final concentration of 600 μM. The culture was grown at 22°C for roughly 20 h before being harvested by centrifugation at 10,000 g. Cell pellets were flash frozen and stored in liquid N₂ until purification. All subsequent steps were carried out in an MBraun anaerobic chamber maintained at less than 0.1 ppm oxygen (MBraun, Stratham, NH). Plastics were brought into the chamber and allowed to sit for two weeks before use. All solvents and buffer stocks were degassed by sparging with argon gas for 4 h before being taken into the chamber.

In a typical purification, roughly 30 g of BaiCD or BaiH cell paste was resuspended in 30 ml of lysis buffer containing 50 mM HEPES, pH 7.5, 300 mM KC1, 4 mM imidazole, 10 mM 2-mercaptoethanol (BME), 10% glycerol, 1 mM flavin mononucleotide (FMN), 1 mM flavin adenine dinucleotide (FAD) and 1% Triton-X305. The resuspension was subjected to 50 rounds of sonic disruption (80% output, 3-s pulse on, 12-s pulse off) at 4°C. The lysate was cleared by centrifugation at 4°C for 1 h at 15,000 g. The supernatant was loaded with an ÄKTA express fast protein liquid chromatography (FPLC) system onto a 5 ml fast-flow HisTrap column (GE Healthcare Life Sciences) equilibrated in lysis buffer lacking FMN, FAD and Triton-X305. The column was washed with 10 column volumes of lysis buffer before elution with 5 ml of buffer containing 50 mM HEPES, pH 7.5, 300 mM KC1, 300 mM imidazole, 10 mM BME and 10% glycerol. The fractions containing protein, based on absorbance at 280 nm, were pooled and reconstituted with iron and sulfur as described. The reconstituted proteins were then passed over a HiPrep 16/60 Sephacryl S-200 HR column equilibrated in 20 mM HEPES, pH 7.5, 300 mM KC1, 5 mM DTT and 10% glycerol. The proteins were concentrated to roughly 1 ml with a Vivaspin 20 concentrator (Sartorius Stedium Biotech). The protein concentration was estimated by A₅₆₀ using the extinction coefficient calculated on the basis of its corresponding amino-acid sequence.

In vitro reconstitution of bile acid pathway
Six assays each contained 50 mM HEPES pH 7.5, 50 mM KC1, 200 μM NAD, 100 μM CoA and 200 μM ATP. In addition, each assay contained 0.1 mM of between one and six of the following enzymes: BaiB from C. scindens, BaiA2 from C. scindens, BaiD from C. hiranonis, BaiE from C. hylemonae and BaiH from C. scindens. All reactions were initiated with the addition of cholic acid and incubated at 22°C for 30 min before being quenched by addition of an equal volume of 100 mM H₂SO₄ at the designated times. Each assay was performed in triplicate. Product formation was monitored by LC–MS as described above.

**Bile acid pathway reconstitution kinetics**
To determine the rate of DCA production by the in vitro pathway, we carried out assays with 50 mM HEPES pH 7.5, 50 mM KC1, 200 μM NAD, 100 μM CoA and 200 μM ATP, plus 0.1 mM of each of BaiB from C. scindens, BaiA2 from C. scindens, BaiD from C. hiranonis, BaiE from C. hylemonae and BaiH from C. scindens. Reactions were initiated by adding cholic acid and incubated at 22°C. Samples of the reaction were removed and mixed with an equal volume of 100 mM H₂SO₄ at the designated times. Each assay was performed in triplicate. Product formation was monitored by LC–MS as described above.

**Kₘ assay for BaiCD**
Kinetic parameters for BaiCD from C. hiranonis were determined in assays that contained 0.45 μM enzyme, 50 mM HEPES pH 7.5, 50 mM KC1 and 500 μM NADH. Reactions were incubated at 22°C before being initiated with 3-oxo-4,5-dehydro-deoxycholic acid. Concentrations of substrate were varied between 3.91 μM and 500 μM. We removed 20 μl of samples and mixed them with an equal volume of 100 mM H₂SO₄ to stop the reaction. Product formation was determined by LC–MS as above. Reactions were performed in triplicate and data were fit to the Michaelis–Menten equation by the least-squares method.

**Kₘ assay for BaiH**
Kinetic parameters for BaiH from C. scindens were determined in assays that contained 0.45 μM enzyme, 50 mM HEPES pH 7.5, 50 mM KC1 and 500 μM NADH. Reactions were incubated at 22°C before being initiated with 3-oxo-4,5,6,7-didehydro-deoxycholic acid. Concentrations of substrate were varied between 0.78 μM and 100 μM. We removed 20 μl of samples and mixed them with an equal volume of 100 mM H₂SO₄ to stop the reaction. Product formation was determined by LC–MS as above. Reactions were performed in triplicate and data were fit to the Michaelis–Menten equation by the least-squares method.

**Colonization of germ-free mice**
Engineered C. sporogenes strains or C. scindens ATCC 15579 were cultured anaerobically in TYG medium (supplemented with appropriate antibiotics) or BHI medium from frozen glycerol stocks. Of the overnight cultures, 50 μl were used to inoculate 5 ml of fresh growth medium containing 50 μM cholic acid. After 48 h, bacterial cells were pelleted by centrifugation, washed twice with PBS, resuspended in 25% glycerol solution, and stored at 80°C. Germ-free C57BL/6 mice...
male, roughly 8 weeks of age, n = 6 or 7 per group) were obtained from Taconic Biosciences (Hudson, NY), and colonies were maintained in gnotobiotic isolators in accordance with the Administrative Panel on Laboratory Care (APLAC) of the Stanford Institutional Animal Care and Use Committee (IACUC). The mice were maintained on a Teklad Custom Diet (TD.180755, with 0.5% cholic acid). Mice were gavaged once daily for three successive days using the following procedure: glycerol stocks of engineered strains of C. sporogenes or C. scindens ATCC 15579 were thawed to room temperature, and mice were inoculated by oral gavage with roughly 200 µl of the thawed glycerol stock (about 1 × 10^7 colony-forming units). Six days after colonization, faecal pellets were collected and mice were euthanized humanely by CO2 asphyxiation. Caecal contents were also collected, snap-frozen in liquid nitrogen, and stored at −80 °C.

Murine faecal samples (roughly 30 mg) or intestinal contents (roughly 100 mg) were preweighed into a 2 ml screwtop tube containing six 6-mm ceramic beads (Precellys CK28 Lysing Kit). Then, 300 µl or 1 ml of a mixture of ice-cold acetonitrile, methanol and water (4/4/2, v/v/v) was added to each tube, and samples were homogenized by vigorous shaking using a Qiagen Tissue Lyser II at 25/s for 10 min. The resulting homogenates were centrifuged for 15 min at 14,000g at 4 °C, and 100 µl of the supernatant was combined with 100 µl of an internal standard solution (2 µM d4-cholic acid in H2O). The resulting mixtures were filtered through a Durapore PVDF 0.22-μm membrane using Ultrafree centrifugal filters (Millipore, UFC30GV00), and 5 µl was analysed by LC–MS as described above in the Methods section ‘LC–MS analysis of metabolite extracts’.

### Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

### Data availability
Mass spectrometry data that support our findings have been deposited in MassIVE (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) under accession code MSV000085048. Source data are provided with this paper.

41. Leppik, R. A. Improved synthesis of 3-keto, 4-ene-3-keto, and 4,6-diene-3-keto bile acids. Steroids 41, 475–484 (1983).
42. Lanz, N. D. RlmN and AtsB as models for the overproduction and characterization of radical SAM proteins. Methods Enzymol. 516, 125–152 (2012).

### Acknowledgements
We thank C. T. Walsh, D. Dodd, C. O’Loughlin and members of the Fischbach and Almo laboratories for helpful comments on the manuscript. This work was supported by National Institutes of Health (NIH) grants DP1 DK113598 (to M.A.F.), R01 DK110174 (to M.A.F.), PO1 HL147623 (to M.A.F.), P01 GM118393-01 (to S.C.A.), US4 GM093342 (to S.C.A.), US4 GM094662 (to S.C.A.) and DP2 HD101401-01 (to C.G.), the Chan–Zuckerberg Biohub (to M.A.F.), a Howard Hughes Medical Institute (HHMI)–Simons Faculty Scholars Award (to M.A.F.), an Investigators in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Foundation (to M.A.F.), and the Price Family Foundation (to S.C.A.).

### Author contributions
M.F., T.L.G., S.C.A. and M.A.F. conceived and designed the experiments. M.F. developed the system for gene-cluster expression in Clostridium, and M.F., C.G. and Y.V. performed the bacterial genetics experiments. T.L.G. expressed and purified enzymes and set up biochemical reconstitution experiments. M.F. analysed the data from biochemical and microbiological experiments by LC–MS. M.E.M. and L.C.B. synthesized bile acid intermediates. M.W. and S.H. performed and analysed mouse experiments. M.F., T.L.G., M.W., S.C.A. and M.A.F. analysed data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

### Competing interests
M.A.F. is a co-founder and director of Federation Bio.

### Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2396-4.
Correspondence and requests for materials should be addressed to S.C.A. or M.A.F.

### Peer review information
Nature thanks Pieter Dorrestein and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | Previously proposed pathway for 7α-dehydroxylation of cholic acid in C. scindens VPI12708. See main text for details and a summary of the previous literature. CA, cholic acid.
**Extended Data Fig. 2** | Purification of recombinant Bai proteins. a, SDS–PAGE analysis of purified Bai proteins after Ni-affinity and size-exclusion purification, visualized by Coomassie blue staining. The image was generated using a Bio-Rad Gel Doc Universal Hood II Molecular Imager. MWM, molecular weight marker; 1, BaiB from *C. scindens*; 2, BaiB from *C. hylemonae*; 3, BaiCD from *C. hiranonis*; 4, BaiE from *C. scindens*; 5, BaiE from *C. hiranonis*; 6, BaiA2 from *C. scindens*; 7, BaiF from *C. hylemonae*; 8, BaiH from *C. scindens*; 9, Bai from *C. scindens*; 10, BaiI from *C. hiranonis*. b, Ultraviolet–visible spectra of BaiCD from *C. hiranonis* (24 μM, left) and BaiH from *C. scindens* (13 μM, right). Features at 370 nm and 450 nm are indicative of flavin bound to BaiCD and BaiH, and are partially obscured by the presence of a [4Fe–4S] cluster, which has broad absorbance between 300 nm and 700 nm. c, The presence of FMN and FAD is confirmed by mass spectrometry. Experiments in a–c were repeated independent twice, with similar results.
Extended Data Fig. 3 | Bile acid standards. **a**, For each compound in the study for which we have an authentic standard, we show an EIC of the authentic standard and the experimentally observed compound. Because the data shown here were collected from samples run at different times, a drift in retention time may be responsible for the peak pairs that do not have identical retention times. **b**, We observed a drift in retention time in the LC–MS data collected for the experiment shown in Fig. 2c. For two representative compounds from that data set, we show an EIC of the experimentally observed compound and an authentic standard run contemporaneously, showing that the retention times remain consistent with our peak assignments.
Extended Data Fig. 4 | Kinetic parameters for BaiCD and BaiH.  

a, Michaelis–Menten analysis of the conversion of 3-oxo-4,5-dehydro-DCA to 3-oxo-DCA by BaiCD. Reaction mixtures contained 0.45 μM BaiCD and 1 mM NADH, with the substrate concentration varying between 15 μM and 500 μM.  
b, Michaelis–Menten analysis of the conversion of 3-oxo-4,5,6,7-didehydro-DCA to 3-oxo-4,5-dehydro-DCA by BaiH. Reaction mixtures contained 0.45 μM BaiH and 1 mM NADH, with the substrate concentration varying between 3 μM and 100 μM.  
Data indicate the average product level ± 1 s.d. (three biological replicates).
Extended Data Fig. 5 | Biochemical analysis of 3-oxo-DCA reduction by BaiA2. Combined EICs showing the conversion of 3-oxo-DCA to DCA by recombinant BaiA2. This experiment was performed once.
Extended Data Fig. 6 | 7α-dehydroxylation of CDCA in vivo. Combined EICs showing the conversion of CDCA to LCA by a C. sporogenes strain harbouring the complete bai operon on three plasmids (MF001) versus a control strain of C. sporogenes harbouring the transporter baiG (MF012). The strains were cultivated with 1 μM cholic acid for 72 h; an acetone extract of the culture supernatant was analysed by high-performance LC (HPLC)/MS. The single asterisk indicates isoLCA; the peak indicated by the double asterisk is provisionally assigned as isoCDCA. This experiment was performed once.
Extended Data Fig. 7 | Constructs for expressing the bai operon and portions thereof in C. sporogenes. Each of the plasmids has replication origins (origin and repH) for E. coli and Clostridium, the traJ gene to enable conjugal plasmid transfer, and an antibiotic-resistance gene (catP, aad9 or ermB). The bai genes were introduced into these plasmids under the control of the fdx or spoIE promoter. For the genetic analysis of baiCD and baiH function, pMTL83153-based plasmids were used.
Extended Data Fig. 8 | Metabolic logic of the 7α-dehydroxylation pathway. Highly oxidized metabolic intermediates as anaerobic electron acceptors. In the first half of the 7α-dehydroxylation pathway, two successive two-electron oxidations set up a vinylogous dehydration of the 7-hydroxyl, yielding the highly oxidized intermediate 3-oxo-4,5,6,7-didehydro-DCA. In the second half of the pathway, three successive two-electron reductions reduce this molecule to DCA, resulting in a net two-electron reduction. The first two of these reductions are carried out by Fe–S flavoenzymes, which comprise a suite of four cofactors that enable them to convert two-electron inputs to a one-electron manifold. The previously proposed pathway is shown in Extended Data Fig. 1.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- **n/a**
- **Confirmed**

- [ ] The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- [ ] A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [ ] The statistical test(s) used AND whether they are one- or two-sided
  
  *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- [ ] A description of all covariates tested
- [ ] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- [ ] A full description of the statistical parameters including central tendency (e.g. mean) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- [ ] For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  
  *Give P values as exact values whenever suitable.*
- [ ] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [ ] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [ ] Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

*Our web collection on [statistics for biologists](https://www.nature.com/) contains articles on many of the points above.*

Software and code

Data collection

- Agilent MassHunter

Data analysis

- Agilent MassHunter, GraphPad Prism 7

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](https://www.nature.com/) for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

There are no restrictions on data availability. Mass spectrometry data are publicly available in the MassIVE repository.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- [x] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)
Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

| n/a | Involved in the study |
|-----|-----------------------|
| [ ] | Antibodies            |
| [ ] | Eukaryotic cell lines |
| [ ] | Palaeontology         |
| [ ] | Animals and other organisms |
| [ ] | Human research participants |
| [x] | Clinical data        |

Data exclusions

| n/a | Involved in the study |
|-----|-----------------------|
| [ ] | No data were excluded from the manuscript. In some figure panels, a representative example is shown (e.g., an individual set of chromatograms) in order to highlight a given result. |

Replication

| n/a | Involved in the study |
|-----|-----------------------|
| [ ] | Reproducibility was verified by performing independent biological replicates. All attempts at replication were successful. |

Randomization

| n/a | Involved in the study |
|-----|-----------------------|
| [ ] | Samples were not randomized. They were run consecutively on the LC-MS; instrument drift within each sample set was minimal. |

Blinding

| n/a | Involved in the study |
|-----|-----------------------|
| [ ] | Researchers were not blinded during these experiments. Data were analyzed using software that calculates metabolite levels in an objective manner, so blinding is less relevant in this setting. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| [ ] | Antibodies            |
| [ ] | Eukaryotic cell lines |
| [ ] | Palaeontology         |
| [ ] | Animals and other organisms |
| [ ] | Human research participants |
| [x] | Clinical data        |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| [ ] | ChIP-seq               |
| [ ] | Flow cytometry         |
| [x] | MRI-based neuroimaging |

Animals and other organisms

Policy information about studies involving animals, ARRIVE guidelines recommended for reporting animal research

Laboratory animals

We used germ-free C57BL6 mice, male, 8 +/- 2 weeks of age.

Wild animals

The study did not involve wild animals.

Field-collected samples

The study did not involve field-collected samples.

Ethics oversight

The study protocol was approved by A-PLAC, the Stanford IACUC.

Note that full information on the approval of the study protocol must also be provided in the manuscript.