Metagenomic discovery of novel enzymes and biosurfactants in a slaughterhouse biofilm microbial community

Stephan Thies[1], Sonja Christina Rausch[1], Filip Kovacic[1], Alexandra Schmidt-Thaler[5], Susanne Wilhelm[1], Frank Rosenau[2], Rolf Daniel[3], Wolfgang Streit[4], Jörg Pietruszka[5,6], Karl-Erich Jaeger[1,6]*

[1] Institut für Molekulare Enzymtechnologie, Heinrich-Heine-Universität Düsseldorf, Jülich, Germany

[2] Zentrum für Peptidpharmazeutika, Universität Ulm, Ulm, Germany

[3] Institut für Mikrobiologie und Genetik, Abteilung für Genomik und Angewandte Mikrobiologie & Göttingen Genomics Laboratory, Georg-August-Universität Göttingen, Germany-

[4] Biocenter Klein Flottbek, Department of Microbiology and Biotechnology, University of Hamburg, Hamburg, Germany

[5] Institut für Bioorganische Chemie, Heinrich-Heine-Universität Düsseldorf, Jülich, Germany

[6] Institut für Bio- und Geowissenschaften IBG-1: Biotechnologie, Forschungszentrum Jülich GmbH, Jülich, Germany

Correspondence: K.-E. Jaeger, Institute of Molecular Enzyme Technology, Heinrich-Heine-Universität Düsseldorf, Forschungszentrum Jülich, Wilhelm-Johnen-Straße, D-52426 Jülich, Germany.

E-mail: karl-erich.jaeger@fz-juelich.de

Keywords: metagenome, biosurfactant, hydrolases, slaughterhouse drain, N-acyl amino acid synthases (NAS), N-acyltyrosine

Abstract

DNA derived from environmental samples is a rich source of novel bioactive molecules. The choice of the habitat to be sampled predefines the properties of the biomolecules to be discovered due to the physiological adaptation of the microbial community to the prevailing environmental conditions. We have constructed a metagenomic library in *Escherichia coli* DH10b with environmental DNA (eDNA) isolated from the microbial community of a slaughterhouse drain biofilm consisting mainly of species from the family Flavobacteriaceae. By functional screening of this library we have identified several lipases, proteases and two clones (SA343 and SA354) with biosurfactant and hemolytic activities. Sequence analysis of the respective eDNA fragments and subsequent structure homology modelling identified genes encoding putative N-acyl amino acid synthases with a unique two-domain organisation. The produced biosurfactants were identified by NMR spectroscopy as N-acyltyrosines with N-myristoyltyrosine as the predominant species. Critical micelle concentration and reduction of surface tension were similar to those of chemically synthesised N-myristoyltyrosine. Furthermore, we showed that the newly isolated N-acyltyrosines exhibit antibiotic activity against various bacteria. This is the first report describing the successful application of functional high-throughput screening assays for the identification of biosurfactant producing clones within a metagenomic library.

Introduction

Metagenomics allow to access novel biocatalysts and metabolites from organisms that are not cultivable[1–3]. In sequence-based approaches, genes are detected using DNA probes or degenerate oligonucleotides derived from known genes encoding the protein family of the interest or by homology search of datasets obtained from eDNA deep sequencing[1,2,4]. In contrast, phenotypic screening approaches apply activity-based assays enabling the discovery of so far unknown proteins belonging to completely novel families. Since the environmental conditions shape the microbial diversity, the choice of the respective habitat is essential for successful mining for novel biocatalysts as shown for cold, hot, and halophilic, habitats[5,6]. Furthermore, nutrient availability largely determines the spectrum of enzymes to be identified[7,8].

Regarding phenotypic screening approaches, functional expression of the cognate genes in standard laboratory host strains as well as the availability of efficient screening assays are necessary prerequisites[9,10]. Many biotechnological relevant enzymes including hydrolases and oxidoreductases have already been discovered by metagenomic screenings[7,11,12]. In addition, several secondary metabolites including patellamide D, violaceins, and polytheonamides were successfully isolated in metagenome studies[2,13,14]. However, the discovery of secondary metabolites still remains challenging, probably due to the demand of proper precursor molecules, functionally interacting auxiliary proteins and the necessity for heterologous expression of large gene clusters[15,16]. Another bottleneck is the availability of suitable high throughput screening systems to detect desired natural products[2,17,18]. Nonetheless, it has been estimated that metagenomic screening have enormous potential for exploration of novel secondary metabolites[2,3].

Biosurfactants are biologically produced surface active secondary metabolites which can serve as sustainable alternatives for chemically synthesized surfactants[19]. They comprise a structurally diverse group of amphiphilic compounds consisting of hydrophobic fatty acids or fatty acid derivatives and hydrophilic groups, for example sugars, peptides or amino acids[20]. These biosurfactants can be used as detergents and emulsifiers for a variety of agricultural, pharmaceutical and ecological applications[21]. In the last years, many efforts were made to improve the production of already known biosurfactants[22–24]. Furthermore, recently developed high throughput screening methods allowed for the identification several biosurfactant producing strains from environmental samples by[25–28], however, to our knowledge, the successful screening of metagenomic libraries for biosurfactants was not yet reported[29].

In this study, we present the construction and phenotypic screening for lipases, proteases and hemolysins of a metagenomic library obtained from a biofilm isolated from slaughterhouse drain. We describe the successful application of a functional screening assay allowing for the identification of biosurfactant producing clones within a metagenomic library. A recently constructed expression system[30] allowed for the efficient expression of the biosurfactant biosynthetic enzymes thereby enabling the identification and initial characterization of the produced biosurfactant.

Results

Construction of metagenomic libraries and phylogenetic analysis

A microbial community living in a biofilm attached to a solid surface of a blood bottom drain in the butchery Frenken Vieh- und Fleisch GmbH (Düren, Germany) was used as eDNA source for the construction of metagenomic libraries. Slaughterhouse material was previously proven to allow for laboratory cultivation of different strains with proteolytic, lipolytic, hemolytic, and biosurfactant activities[31–35]. Here, we expected that a microbial community residing in a habitat rich in blood, fats, remnants of animal skin and flesh would be adapted to the efficient degradation of proteins and lipids. The isolated eDNA was used for assessing the phylogenetic diversity of this community by sequence analysis of 16S rRNA genes and the construction of a metagenomic library, which was screened for lipolytic enzymes, proteases, haemolysins and biosurfactants. The biofilm community consisted mainly of *Flavobacterium* sp. and *Chryseobacterium* sp. belonging to the family of Gram-negative, aerobic, rod-shaped Flavobacteriaceae (Figure 1) with a low G+C genomic content (32-38%).

Metagenomic libraries were constructed in *E. coli* DH10b using eDNA isolated from the bacterial biofilm and cloned into the broad host range shuttle vector pEBP18[30]. The respective clones were grown on LB agar containing tributyrin, skim milk or sheep blood for detection of lipolytic, proteolytic, hemolytic and biosurfactant activities, respectively. The library consisted of 1.7×10^5 clones with >85% containing inserts with an average size of 5.2 kb (size range was 2-10 kb) as determined by analysing 48 randomly selected clones.

Screening for bioactive molecules

The functional screenings of the complete library revealed 15 clones with tributyrin-degrading (TB) lipolytic activity (TB94 - TBIH8), 5 clones with skim milk-degrading (SM) proteolytic activity (SM321 - SMVIIIE12) (Tables S1 and S2, Figure S1). Additionally, we have identified 2 clones (SA343, SA354) with blood-degrading hemolytic activity (Figure 2A), which were confirmed as surface active (SA) biosurfactant-producing clones by the atomized oil assay (Figure 2B). Surface active clones did not produce proteases and phospholipases (data not shown), which are typically considered as hemolysins, thus, biosurfactant production most likely caused hemolysis. We could show that clones SA343 and SA354 produced biosurfactant in liquid culture, as confirmed by the grid assay[36] (Figure 2C). To the best of our knowledge, this is the first example of identifying biosurfactant producing clones by screening of a metagenomic library. For further analysis and characterization, approximately 65 mg of crude biosurfactant was produced and isolated from 1 L cultures of *E. coli* DH10b harbouring plasmids pEBP*SA343* or pEBP*SA354*.

Determination of metagenomic DNA sequences

Plasmids from 22 clones with biological activities detected by screening of the metagenomic library were sequenced and their functions and phylogenetic relationships were predicted from sequence similarity to known genes (Table S2). Apparently, genes of enterobacterial origin were enriched during functional screening in *E. coli*, but the 16S rRNA analysis revealed that the microbial community of the slaughterhouse biofilm consisted predominantly of species from Flavobacteriaceae.

The DNA sequences from clones with proteolytic activity (SM321 - SMVIIIE12) showed 65% to 80% identity with genes from the family of Xanthomonadaceae from the class γ-proteobacteria. At least one ORF encoding a putative peptidase was identified in each clone thus supporting the results of the functional screening. Detailed analysis showed that clones SM760 and SMVIIIE12 carried identical fragments with six base pair substitutions in non-coding regions. Both clones resembled a truncated version of clone SM679 with which they share 99% sequence identity. Hence, we have identified four novel putative proteases on the three different fragments, since SM321 harbours two orfs with high similarity to proteolytic enzymes.

Sequence analysis of 15 clones with lipolytic activities revealed that most ORFs showed similarity to genes from α-proteobacteria (TB157, TB303 ORFs 1-3, TB304, TB305, TB306 ORFs 8-9, TB310), β-proteobacteria (TB94, TB307, TB 308, TB312, TB313, TB314, TB I H8) and γ-proteobacteria (TB350). However, genes encoded on eDNA fragments of clones TB303 and TB306 were similar to genes from Flavobacteria in agreement with the phylogenetic analysis of the metagenome-derived eDNA. Each eDNA fragment of the TB clones contained at least one ORF encoding a putative lipolytic enzyme. Accordingly, we have identified 21 novel putative lipase/esterase encoding genes on the eDNA fragments of the 15 lipolytic clones, although some of the sequences were very similar (e.g. TB94 and TB307 are 99% identical).

Clone SA354 contained an 8.8 kb eDNA fragment with parts of the sequence showing up to 88% identity to genomic sequences of the α-proteobacterium *Sinorhizobium*, but none of the six predicted ORFs was similar to known biosurfactant related genes (Figure 3, Table S2). The two ORFs located on the 3.3 kb fragment obtained from clone SA343 showed similarity to SA354 ORFs with 87% sequence identity (*orf*1 of SA343, *orf*5 of SA354) and 79% sequence identity (*orf*2 of SA343, *orf*6 of SA354) (Figure 3A, Table S2). The genes encoded by *orf*1 of SA343 and *orf*5 of SA354 showed high similarity (approximately 70% sequence identity in the respective part of the gene) to lysophospholipid acyl transferase genes (Table S2). Parts of *orf*2 from SA343 and *orf*6 from SA354 revealed significant similarity (73% sequence identity) to peptidases. Although *in silico* prediction suggested protease activity, none of the surface active clones showed activity on skim milk agar plates (data not shown). Therefore, the putative acyl transferase genes were considered as the most promising candidates causing surface activity and hemolytic phenotypes of these clones.

Identification of biosurfactant biosynthesis genes and enzymes

To address the question whether the putative acyl transferase encoding genes are indeed responsible for the observed phenotypes, we constructed a T7-polymerase based expression vector carrying *orf*5 of SA354. Expression of this gene in *E. coli* BL21(DE3) led to the hemolytic and biosurfactant producing phenotypes (Figures 3B and 3C). Conclusively, heterologous production of surface active molecules in *E. coli* is presumably catalysed by a single protein encoded by *orf*5 of SA354. This was a striking result, because several genes organised on a contiguous DNA fragment are usually required for the production of heterologous biosurfactant[22,37].

Homology modelling of the putative acyl transferases encoded by *orf*1 of SA343 and *orf*5 of SA354 revealed similarity of their N-terminal domains with N-acyl amino acid synthase FeeM (PDB code 2G0B)[38] and C-terminal domains with bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) binding protein PA4608 from *Pseudomonas aeruginosa* (PDB code 1YWU)[39] (Figures 3D and S3). Remarkably, such a two domain molecular organisation was not found in any previously described protein, neither by sequence similarity search nor by homology modelling.

As the N-acyl amino acid synthase (NAS) domain is likely responsible for the enzymatic activity of the proteins encoded by *orf*1 of SA343 and *orf*5 of SA354, they were designated as Nas343 and Nas354, respectively. The superposition of homology models of N-terminal domains (residues 1-200) of Nas343 and Nas354 with the structure of FeeM[38] revealed low root-mean-square deviation (RMSD) values (<0.6 Å) as expected for proteins with similar sequence and function. Mainly, these differences can be assigned to the C- and N- terminal parts of each domain, while the main protein folds show no significant differences (Figure 3E). A sequence alignment of Nas343 and Nas354 with homologous N-acyl amino acid synthases classified them into the type I N-acyl amino acid synthase family characterised by three conserved sequence motifs and a putative active site glutamate residue essential for N-acyl amino acid synthesis activity[40] (Figure 3F). In Nas354 and Nas343 we identified Glu103 as the putative active site residues, which showed structural conservation with Glu95 proposed to be the active site residue of FeeM[38] (Figure 3E). The substitution of the putative catalytic Glu103 of Nas354 by a functionally unrelated alanine resulted in loss of hemolytic and surface active phenotypes (Figure S3). These results provide first experimental proof of the function of the conserved glutamate as the catalytic active residue in N-acyl amino acid synthases. Apparently, the catalytic activity of the NAS domain of Nas354 is essential for the hemolytic and biosurfactant producing activities in agreement with previous results which showed that expression of *nas* genes in *E. coli* led to the synthesis of N-acyl derivatives of amino acids[40,41]. The previously proposed mechanism of N-acyl amino acid biosynthesis[38,42] suggests that the expression host *E. coli* provides fatty acids linked to acyl-carrier proteins (ACP) which together with Nas343 and Nas354 carry out the biosynthetic reaction.

Homology models of the C-terminal domains (residues 279-394) of Nas343 and Nas354 superimposed with the structure of the PA4608[39] revealed low RMSD values (<0.5 Å) which indicate a similar function. Residues 200-278 which link the Nas and the c-di-GMP binding domains did not show structural similarity to any known protein structure. Presently, the function of the c-di-GMP binding domain cannot be directly related to the biosurfactant producing activities of Nas354 and Nas343.

Identification and biophysical properties of the biosurfactant

Biosurfactants produced by clones SA343 and SA354 were identified by thin-layer chromatography (TLC) of culture extracts. For both clones, just one hydrophobic compound with an R_f = 0.75 (Figures 4A and S4) was identified. The absence of this compound in the empty vector control indicated that the activities of Nas343 and Nas354 were required for its production. The purified surface-active compound was identified by NMR (see supplementary data) and mass spectrometric (MS) analyses as tyrosine N-acylated with myristic acid (Figure 4B). Traces of other fatty acids (e.g. unsaturated C_{16} fatty acid) bound to tyrosine were also detected by MS but these compounds were not analysed further. We demonstrated that two novel metagenomic tyrosine NASs, Nas343 and Nas354, enable the synthesis of the biosurfactant N-myristoyltyrosine by *E. coli*.

The biophysical properties N-acyltyrosine purified from clone SA354 including the specific rotation of polarized light, the critical micelle concentration (CMC) and the reduction of surface tension were determined and compared to chemically synthesised N-myristoyl-L-tyrosine (Table 1). Similar values for specific rotation of polarized light for both compounds indicated the presence of *L*- rather than *D*-tyrosine as a component of the biosurfactant. As both biologically and chemically synthesized N-acyltyrosine samples showed low water solubility at neutral pH values, tensiometric characterisation was performed at pH 12 (Figure 4B). Both compounds drastically reduced the surface tension between the water/air interphase already at low concentrations and they reached a similar constant minimal surface tension at concentrations above 1 mg/mL (Table 1). The critical micelle concentrations calculated from tensiometric results (Figure 4B) were comparable for chemically and biologically synthesised N-myristoyltyrosine (Table 1). These surface active properties confirm that N-acyltyrosine production is responsible for the observed surface active phenotype of the metagenomic clones SA343 and SA354.

Antimicrobial activity of N-acyltyrosine

N-acyl amino acids were previously identified as compounds with antibiotic activity against the Gram-positive bacterium *Bacillus subtilis*[40,43]; this activity was confirmed for N-acyltyrosine purified from clone SA354 (Figure 5A). In addition, we have demonstrated inhibitory activity of N-acyltyrosine on the growth of the mycobacterium *Corynebacterium glutamicum*, a Gram-positive bacterium with a cell wall composition different from *B. subtilis,* and the Gram-negative bacteria *Chromobacterium violaceum* CV026 and *Sorangium cellulosum* DSM53796 (Myxobacteria) (Figures 5B-D). Growth of the γ-proteobacteria *E. coli* DH5α, *P. aeruginosa* PA01, *Serratia marcescens* DSM12481 and *Pseudomonas putida* KT2440 was not affected (data not shown). Antibiotic activity of N-acyltyrosine towards Gram-negative species and Mycobacteria was not reported so far. Although the underlying mechanism is presently unknown, one can speculate that N-acyltyrosine might cause cell membrane disruption as reported for rhamnolipid biosurfactants[44]. This hypothesis is supported by the observation described here and elsewhere that N-acyltyrosine as well as other biosurfactants lyse red blood cells[45].

Discussion

Screening of metagenomic libraries offers access to novel bioactive molecules, but requires the combination of a suitable host organism with a functional expression system and an effective screening pipeline. Functional elements to be present include promoters, regulatory sequences and ribosomal binding sites as well as the environment for correct protein folding[46,47] resulting in the so called “different host - different hit” effect[48].

In this study, *E. coli* DH10b was used as the host for the expression of metagenomics genes encoding both hydrolytic and biosurfactant biosynthetic enzymes. Apparently, this bacterium not only provided suitable cofactors and precursors needed for the synthesis of the target metabolite including compatible acyl carrier proteins required for the synthesis of N-acyl amino acids[38], but also mechanisms for active or passive secretion of the target metabolite into the culture medium. Furthermore, *E. coli* DH10b showed intrinsic resistance to the produced surface-active compound N-acyltyrosine. It should be noted, however, that this γ-proteobacterial expression system exhibited a clear bias towards expression of proteobacterial genes although the microbial community of analysed biofilm consisted mostly Flavobacteria species. Apparently, the functional expression of genes from Flavobacteria is difficult in evolutionary distant *E. coli* as reported previously[49].

Biosurfactants as metabolites with multiple potential applications represent interesting targets to be identified by metagenomic approaches[2,29]. So far, screening methods suitable for the detection of biosurfactant producers in environmental samples were used only to identify novel cultivable bacteria; but they were not applied to screening of metagenomic libraries[26,28]. In this study, we have used a newly developed expression system to construct a metagenomic library[30] which was subsequently subjected to high-throughput screening for biosurfactant producing clones allowing to identify N-acyltyrosine. To the best of our knowledge, this is the first report of a biosurfactant identified by functional screening of a metagenomic library.

N-acyl amino acids have previously been identified as compounds with antibiotic activity towards *B. subtilis*[43]. They were among the first metagenome-derived small molecules identified; however, their surface-active properties were not described so far. They were shown to be synthesized by metagenome-encoded N-acyl amino acid synthases[50], but until now, only few of these enzymes were identified in cultured bacteria[40,51]. The physiological function of N-acyl amino acids and their cognate synthases is still not clear. These enzymes can be incorporated in a metabolic pathway where they catalyse the synthesis of an N-acyl amino acid as an intermediate product[42], but they may appear as single enzymes as well[38]. Besides that, a genetic link of a subset of these enzymes exists to protein sorting systems belonging to the PEP-CTERM exosortase family[51]. Presumably, genes *nas343* and *nas354* are not located in an operon encoding a biosynthetic pathway. Apparently, also heterologous proteins can accept acyl-ACPs provided by *E. coli* as observed for other acyl-ACP-dependent synthetases[42,52]. Although the origin of these genes remains unknown, the high homology of the respective proteins and the generally low overall sequence conservation among NAS type I family members[53] suggest a close evolutionary relationship of the respective source organisms.

The N-acyl amino acid synthases described here revealed a unique molecular organisation of a NAS domain linked to a putative c-di-GMP binding PilZ domain which has not been observed so far. C-di-GMP is a universal secondary messenger regulating multiple cellular processes in bacteria[54]. PilZ domains with wide phylogenetic distribution were found as a single domain or fused to other domains predicted to have regulatory or transport functions[55]. Binding of c-di-GMP to a PilZ domain usually affects the cellular physiology through modulation of protein-protein or protein-DNA interactions, but a direct effect on the catalytic activity of enzymes involved in alginate and cellulose biosynthesis was also shown[56,57]. NasP, a metagenome-derived NAS, was shown to be activated by c-AMP, another widespread bacterial secondary messenger[58]. Therefore, we propose a role of the c-di-GMP binding PilZ domains for the regulation of Nas354 and Nas343 activity; the detailed mechanism needs to be further explored. We observed structural homology of NAS proteins to autoinducer synthetases indicating a function of N-acyl amino acids as cellular messengers[38] thereby assigning NAS proteins a role as putative regulators of bacterial signaling processes.

The production of N-acyl amino acids is biotechnologically relevant[59] with N-acyltyrosines being applied in the cosmetic industries[60,61]. Comparison of chemically synthesized N-myristoyltyrosine and biologically produced N-acyltyrosine revealed similar surfactant properties although heterogeneity in the natural product was observed for fatty acids bound to tyrosine. Such promiscuity towards fatty acids with different chain lengths and saturation grades is common for biosurfactant synthesizing enzymes[24,62] and may result in the formation of mixed surface films or micelles with altered physical properties[63]. Hence, the here described N-acyl amino acids represent an interesting group of biosurfactants with a remarkably simple biosynthesis and a variety of important biotechnological applications.

Materials and Methods

Bacterial strains and growth conditions

The strains and plasmids used in this study are listed in table 2. In general, bacteria were grown at 37°C in nutrient rich liquid LB medium or solid LB agar (1.5% w/v) supplemented with kanamycin, 50 mg/L or ampicillin 100 mg/L, if necessary. *Escherichia coli* cells were transformed by the calcium chloride method or by electroporation[64].

Recombinant DNA methods

Plasmid DNA was isolated using innuPREP Plasmid Mini Kit (Analytik Jena, Jena Germany) according to the manufacturer's instructions. DNA ligation, DNA dephosphorylation, restriction endonuclease digestion, and agarose gel electrophoresis were performed according to standard techniques[64,65]. Restriction enzymes, T4 DNA ligase, and shrimp alkaline phosphatase (SAP) were purchased from Fermentas (St. Leon-Rot, Germany) or from New England Biolabs (Frankfurt am Main, Germany). Metagenomic libraries were prepared in pEBP18 as described before[65]. Gene *nas354* was amplified from pEBP*SA354*, using 5'-ATATCATATGCAAGACACCACGTTACTC-3' and 5'-ATATCTCGAGCTCAGGCTGGGTGGTGTGCA-3' oligonucleotides and *Pfu* DNA polymerase according to standard protocols. pET*nas354* expression plasmid was constructed by inserting the amplified *nas354* gene in the *Nde*I and *Xba*I restriction sites of pET22b vector. Site directed mutagenesis of *nas354* was performed with QuikChange PCR method using 5'-GGCCAGATTGCCGCAGTGTCGGCCTTG-3' and 5'-CAAGGCCGACACTGCGGCAATCTGGCC-3' oligonucleotides and *Pfu* DNA polymerase[66]. Conditions of PCR amplifications were optimized for each primer pair. PCR products were separated on 0.8% (w/v) agarose gels and the desired DNA fragments were purified with the QIAEX II Gel Extraction Kit (QIAGEN, Hilden, Germany) according to manufacturer's instructions.

Sequencing

Metagenomic shotgun plasmid libraries were sequenced by the Gottingen genomics laboratory by a combination of 454 pyrosequencing and Sanger sequencing. The metagenomic plasmid DNA derived from all clones was pooled in equimolar amounts and a 454-shotgun library was generated using the GS FLX Rapid Library Prep kit (Roche Life Science, Mannheim, Germany) following the instructions of the manufacturer. The library was sequenced using a 454 GS-FLX system and Titanium chemistry as recommended by the manufacturer (Roche Life Science). A *de novo* assembly of the reads was performed by using the Roche Newbler assembly software 2.3 (Roche Life Science). The resulting contigs were assigned to the corresponding plasmids by Sanger sequencing of the plasmid inserts ends using BigDye 3.0 chemistry and an ABI3730XL capillary sequencer (Applied Biosystems, Life Technologies GmbH, Darmstadt, Germany). Remaining gaps of the plasmid inserts were closed by PCR-based techniques and Sanger sequencing of the products.

Bioinformatic analysis

Phylogenetic analysis. Bacterial 16S rRNA genes were PCR-amplified according to standard procedures[67] by using the oligonucleotide primer sets as described[68,69] consisting of 616V (AGAGTTTGATYMTGGCTCAG) and 1492R (CGGYTACCTTGTTACGAC) and 140 ng of metagenomic template DNA. The resulting PCR products were ligated into pGEM®-T Easy Vector System (Promega, Mannheim, Germany). Transformed competent *E. coli* DH5α cells were grown overnight on LB agar plates and white colonies were randomly selected. Plasmid DNA was extracted and inserts of the correct size were identified by hydrolysis with *Eco*RI. Plasmid preparations were partially sequenced with the (universal) internal 16S rRNA primer GM1F (CCAGCAGCCGCGGTAAT) by Eurofins MWG Operon. Only 16S rRNA sequences with at least 600 bp (37 sequences) were included in phylogenetic analysis. All sequences were edited by Eurofins MWG Operon and compared with DNA sequences in the public domain through BLAST searches[70]. The 16S rRNA gene sequences were compiled and aligned using ARB software[71]. Alignments were then manually checked and sequences grouped into operational taxonomic units (OTUs) exhibiting 97% similarity. Maximum-likelihood trees and 100 bootstrap replicates were constructed with the OTUs and the closest neighbors using PhyML[72].

Database search. Genes encoded on the metagenomic fragments were identified by search for open reading frames with Clone Manager software (Scientific & Educational Software, NC, USA). The homology search was performed using the BLAST algorithm available on the NCBI website[70]. The multiple amino acid sequence alignment was generated with the software Clustal Omega[73].

Homology modelling. The three dimensional structure of Nas343 and Nas354 were modelled on the Phyre II server[74] using the structure of an N-acyl amino acid synthase FeeM (PDB code 2G0B) as a template for the N-terminal domain and the structure of c-di-GMP-binding protein (PDB code 1YWU) as a template for the C-terminal domain. Both template proteins were identified by Phyre II structural homology search algorithms. The UCSF Chimera[75] and the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC[76] were used for structural analysis and visualisation.

Screening for clones with hydrolytic activity

Protease indicator plates consisted of skim milk agar[77] with kanamycin (50 µg/mL). Esterase indicator plates contained LB agar supplemented with 1.5% (v/v) tributyrin, 0.15% (w/v) gum arabic and 50 µg/mL kanamycin[78]. Blood agar base (4% w/v) (Carl Roth, Karlsruhe, Germany) dissolved in water supplemented with 5% (v/v) sterilized sheep blood (Fiebig Nähstofftechnik, Idstein-Niederauroff, Germany) and an appropriate antibiotic was used as hemolysis indicator. Clones with hydrolytic activities were identified by the formation of clear halos surrounding the colonies after growth for 24 h at 37°C on indicator plates followed by storage at 4°C for several days, if necessary. Phospholipase activity was examined with an egg yolk LB agar plate assay[78].

Assays for the detection of biosurfactants

Colonies were grown overnight on LB agar plates supplemented with appropriate antibiotics and inspected for biosurfactant production with the atomized oil assay as previously described using the airbrush “Beginner ESB 100” (Revell) and light paraffin oil (Merck, Darmstadt, Germany)[24,79]. The grid assay for detection of surface active substances in liquid media was accomplished in flat bottom microtiter plates positioned onto a grid of squares with a height of 0.8 mm[24].

Production and purification of N-acyltyrosine biosurfactant

For the production of N-acyltyrosine in liquid medium, *E. coli* DH10b cells carrying plasmids pEBP*SA343* or pEBPSA354 were grown in Erlenmeyer flasks in 1 L LB medium supplemented with kanamycin (50 µg/mL). The expression cultures were inoculated to an optical density (OD) of OD_{580nm} = 0.05 and incubated for 18 h at 37°C under aeration. The culture supernatants were isolated by centrifugation (30 min at 3000*g*) and filtration, acidified with HCl to pH 3 and left overnight at 4°C to allow for precipitation of biosurfactants. Precipitates were isolated by centrifugation for 1 h at 8280*g* at 10°C, dissolved in water, acidified and centrifuged again as described before. The final precipitate was dissolved in acidified water (pH 3) and the biosurfactant was isolated by three successive extractions with an equal volume of ethyl acetate. The organic phase was evaporated and the remaining solid material was dissolved in ethanol or methanol for subsequent TLC and NMR spectroscopic analysis, respectively.

TLC was performed using Alugram Silica SilG or Polygram SIL G/UV plates (Macherey–Nagel) and a mixture of trichloromethane, methanol, and acetic acid at volume ratio of 65:15:3 as the mobile phase. Spots were detected by illumination with UV light at λ = 365 nm or staining with iodine vapour.

Proteins Nas354 and Nas354$_{E103A}$ were expressed in *E. coli* BL21(DE3) from plasmids pET*nas*354 and pET*nas*354$_{E103A}$, respectively. Cells were grown on LB agar or blood agar plates supplemented with 0.4 mM IPTG and 100 mg/L ampicillin.

Determination of optical and surface active properties of N-acyltyrosine

The biosurfactant obtained by ethyl acetate extraction was further purified by column chromatography using Merck silica gel 60 (0.063–0.200 mm). The biosurfactant was dissolved in and eluted from the column with a mixture of petroleum ether, ethyl acetate, and acetic acid at volume ratio of 80:20:5.

The surfactant properties of the purified compounds dissolved in 0.1 M NaOH (pH 12) were tested in a concentration range of 1×10^{-8} – 10 mg/mL with a micro tensiometer (Kibron Inc, Helsinki, Finnland) utilizing the maximum pull force method[80,81]. Each surface tension was measured in triplicate and plotted against the surfactant concentration. The resulting plot was used to determine the cmc and the minimal surface tension using sodium dodecyl sulfate (SDS) as the standard. Optical rotation was measured at 20°C with a PerkinElmer Polarimeter 241 MC against the sodium D-line.

NMR spectroscopy

^{1}H and ^{13}C NMR spectra were recorded at 20°C with a Bruker Avance/DRX 600 spectrometer in $CDCl_3$ with TMS as internal standard. Chemical shifts are given in ppm relative to the Me_4Si (^{1}H: Me_4Si = 0 ppm) or relative to the resonance of the solvent (^{13}C: $CDCl_3$ = 77.0 ppm or ^{13}C: MeOD = 50.4 ppm) (see supplementary data).

GC-MS (ESI/ electron spray ionisation method) was performed with a Hewlett-Packard HP6890 Series-GC-System, 5973 Mass Detector Selective and were recorded by the Competence Center BIOSPEC (Central Institute for Analytics ZEA-3, Forschungszentrum Jülich GmbH, Jülich, Germany).

Chemical synthesis of N-myristoyl-L-tyrosine

N-Myristoyltyrosine was synthesized as described by Aha[82] using 1.45 g (8 mmol) of L-tyrosine suspended with 20 mL dichloromethane into a three-necked flask with reflux condenser under nitrogen. 3.1 mL (24.3 mmol) trimethylsilyl chloride was added and the solvent was heated up to reflux for 10 minutes. Hereafter, 3.3 mL (23.8 mmol) of triethylamine were added and the reaction mixture was allowed to reflux for another 10 minutes. Then, the solvent was cooled down over ice and 1.78 g (7.2 mmol) of myristoyl chloride in 5 mL dichloromethane was added slowly. The mixture was stirred for 22 h at room temperature and the reaction subsequently quenched with ice. The white crystalline product was filtrated and washed with distilled water until the filtrate was not acidic anymore. Pure N-myristoyltyrosine was dried under vacuum for 16 h and analysed via NMR (see supplementary data).

Antibiotic activity assay

Antibiotic activity of isolated biosurfactants was monitored with a modified Kirby Bauer assay[83]. Bacteria obtained from overnight cultures were harvested and dissolved in sterile saline (0.9% w/v NaCl) to a resulting OD_{580nm} = 0.15. This cell suspension was used to inoculate Mueller-Hinton agar plates with sterile cotton buds. 5 µL of purified N-acyltyrosine dissolved in ethanol (10 mg/mL and 50 mg/mL) were applied to sterile disks of Whatman-Paper (5 mm in diameter), the disks were dried on air and applied to the inoculated Mueller-Hinton agar plates. The antibiotic effect of N-acyltyrosine was estimated after overnight incubation of plates at the optimal growth temperature of each strain as reported by the Leibniz-Institute DSMZ – German Collection of Microorganisms and Cell Cultures. Equally treated disks with pure ethanol were used as negative control.

Acknowledgements

This work was supported by the European Commission in the framework of the projects “Marine Metagenomics for new Biotechnological Applications” (MAMBA, Project-ID: 226977) and “Industrial Applications of Marine Enzymes” (INMARE, Project ID: 634486). We thank the Ministry of Innovation, Science and Research of North Rhine-Westphalia and Heinrich-Heine-University Düsseldorf for a scholarship to S.T. within the CLIB2021-Graduate Cluster Industrial Biotechnology.

Author Contributions

K.E.J., S.W., F.R., F.K., J.P. and W.S. conceived and supervised the study. S.T., F.R. and S.W. designed the experiments. S.T., S.C.R., A.S.T. and R.D. performed the experiments. S.W., F.R., F.K. and K.E.J. analyzed data. S.T., S.C.R., F.K. and K.E.J. prepared the figures and wrote the manuscript.

Additional Information

Accession codes: The sequences of eDNA fragments of clones with hydrolytic and hemolytic activities identified in this study are accessible at NCBI GenBank; accession numbers are provided in Supplementary Table S3.

Supplementary information accompanies this paper at http://www.nature.com/srep.

Competing financial interests: The authors declare no competing financial interest.

References

1. Ferrer, M., Martínez-Abarca, F. & Golyshin, P. N. Mining genomes and ‘metagenomes’ for novel catalysts. *Curr. Opin. Biotechnol.* **16,** 588–93 (2005).
2. Piel, J. Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. *Annu. Rev. Microbiol.* **65,** 431–53 (2011).
3. Liebl, W. *et al.* Alternative hosts for functional (meta)genome analysis. *Appl. Microbiol. Biotechnol.* **98,** 8099–109 (2014).
4. Guazzaroni, M.-E., Silva-Rocha, R. & Ward, R. J. Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. *Microb. Biotechnol.* **8,** 52–64 (2015).
5. López, G. *et al.* A novel thermoalkalostable esterase from *Acidicaldus sp.* strain USBA-GBX-499 with enantioselectivity isolated from an acidic hot springs of Colombian Andes. *Appl. Microbiol. Biotechnol.* **98,** 8603–16 (2014).
6. Tchigvintsev, A. *et al.* The environment shapes microbial enzymes: five cold-active and salt-resistant carboxylesterases from marine metagenomes. *Appl. Microbiol. Biotechnol.* **99,** 2165–78 (2015).
7. Steele, H. L., Jaeger, K.-E., Daniel, R. & Streit, W. R. Advances in recovery of novel biocatalysts from metagenomes. *J. Mol. Microbiol. Biotechnol.* **16,** 25–37 (2009).
8. Beazley, M. J. *et al.* Microbial community analysis of a coastal salt marsh affected by the Deepwater Horizon oil spill. *PLoS One* **7,** e41305 (2012).
9. Schloss, P. D. & Handelsman, J. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biol.* **6,** 229 (2005).
10. Streit, W. R. & Schmitz, R. A. Metagenomics - the key to the uncultured microbes. *Curr Opin Microbiol* **7,** 492–498 (2004).
11. Lee, M. H. & Lee, S.-W. Bioprospecting potential of the soil metagenome: novel enzymes and bioactivities. *Genomics Inform.* **11,** 114–20 (2013).
12. Chow, J. *et al.* The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases. *PLoS One* **7,** e47665 (2012).
13. Iqbal, H. A., Feng, Z. & Brady, S. F. Biocatalysts and small molecule products from metagenomic studies. *Curr. Opin. Chem. Biol.* **16,** 109–16 (2012).
14. Wilson, M. C. & Piel, J. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem. Biol.* **20,** 636–47 (2013).
15. Zhang, H., Boghigian, B. A., Armando, J. & Pfeifer, B. A. Methods and options for the heterologous production of complex natural products. *Nat. Prod. Rep.* **28,** 125–51 (2011).
16. Loeschcke, A. *et al.* TREX: a universal tool for the transfer and expression of biosynthetic pathways in bacteria. *ACS Synth. Biol.* **2,** 22–33 (2013).
17. Kennedy, J. *et al.* Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. *J. Appl. Microbiol.* **111,** 787–799 (2011).
18. Charlop-Powers, Z., Milshteyn, A. & Brady, S. F. Metagenomic small molecule discovery methods. *Curr. Opin. Microbiol.* **19,** 70–5 (2014).
19. Marchant, R. & Banat, I. M. Biosurfactants: a sustainable replacement for chemical surfactants? *Biotechnol. Lett.* **34,** 1597–1605 (2012).

20. Soberón-Chávez, G. & Maier, R. in *Biosurfactants* (ed. Soberón-Chávez, G.) **20,** 1–11 (Springer Berlin Heidelberg, 2011).

21. Fracchia, L. *et al.* in *Biosurfactants: Production and Utilization—Processes, Technologies, and Economics* (eds. Kosaric, N. & Sukan, F.) 245–68 (CRC Press Taylor & Francis Group, 2014). doi:10.1201/b17599-6

22. Wittgens, A. *et al.* Growth independent rhamnolipid production from glucose using the non-pathogenic *Pseudomonas putida* KT2440. *Microb. Cell Fact.* **10,** 80 (2011).

23. Saharan, B., RK, S. & Sharma, D. A review on biosurfactants: fermentation, current developments and perspectives. *Genet. Eng. Biotechnol. Journal,* **2011,** 1–14 (2012).

24. Thies, S. *et al.* Heterologous production of the lipopeptide biosurfactant serrawettin W1 in *Escherichia coli*. *J. Biotechnol.* **181,** 27–30 (2014).

25. Belcher, R. W., Huynh, K. V, Hoang, T. V & Crowley, D. E. Isolation of biosurfactant-producing bacteria from the Rancho La Brea Tar Pits. *World J. Microbiol. Biotechnol.* **28,** 3261–7 (2012).

26. Hamed, B. S., Smii, L., Ghram, A. & Maaroufi, A. Screening of potential biosurfactant-producing bacteria isolated from seawater biofilm. *African J. Biotechnol.* **11,** 14153–158 (2012).

27. Bodour, A. A., Drees, K. P. & Maier, R. M. Distribution of biosurfactant-producing bacteria in undisturbed and contaminated arid Southwestern soils. *Appl. Environ. Microbiol.* **69,** 3280–7 (2003).

28. Kügler, J. H. *et al.* Glycolipids produced by *Rouxiella* sp. DSM 100043 and isolation of the biosurfactants via foam-fractionation. *AMB Express* **5,** 82 (2015).

29. Sachdev, D. P. & Cameotra, S. S. Biosurfactants in agriculture. *Appl. Microbiol. Biotechnol.* **97,** 1005–16 (2013).

30. Troeschel, S. C. *et al.* Novel broad host range shuttle vectors for expression in *Escherichia coli, Bacillus subtilis* and *Pseudomonas putida*. *J. Biotechnol.* **161,** 71–9 (2012).

31. Borges, W. da S., Cardoso, V. L. & Resende, M. M. de. Use of a greasy effluent floater treatment station from the slaughterhouse for biosurfactant production. *Biotechnol. Appl. Biochem.* **59,** 238–44 (2012).

32. Doddapaneni, K. K. *et al.* Purification and characterization of a solvent and detergent-stable novel protease from *Bacillus cereus*. *Microbiol. Res.* **164,** 383–90 (2009).

33. Ramani, K., Chockalingam, E. & Sekaran, G. Production of a novel extracellular acidic lipase from *Pseudomonas gessardii* using slaughterhouse waste as a substrate. *J. Ind. Microbiol. Biotechnol.* **37,** 531–5 (2010).

34. Tatineni, R. *et al.* Purification and characterization of an alkaline keratinase from *Streptomyces* sp. *Bioresour. Technol.* **99,** 1596–602 (2008).

35. Kuo, Y.-H., Hsu, H.-C., Chen, Y.-C., Liang, T.-W. & Wang, S.-L. A novel compound with antioxidant activity produced by *Serratia ureilytica* TKU013. *J. Agric. Food Chem.* **60,** 9043–7 (2012).

36. Chen, C.-Y., Baker, S. C. & Darton, R. C. The application of a high throughput analysis method for the screening of potential biosurfactants from natural sources. *J. Microbiol. Methods* **70,** 503–10 (2007).

37. Lee, Y.-K. *et al.* Cloning of srfA operon from *Bacillus subtilis* C9 and its expression

in *E. coli*. *Appl. Microbiol. Biotechnol.* **75,** 567–72 (2007).

38. Van Wagoner, R. M. & Clardy, J. FeeM, an N-acyl amino acid synthase from an uncultured soil microbe: structure, mechanism, and acyl carrier protein binding. *Structure* **14,** 1425–35 (2006).
39. Ramelot, T. A. *et al.* NMR structure and binding studies confirm that PA4608 from *Pseudomonas aeruginosa* is a PilZ domain and a c-di-GMP binding protein. *Proteins* **66,** 266–71 (2007).
40. Brady, S. F., Chao, C. J. & Clardy, J. Long-chain N-acyltyrosine synthases from environmental DNA. *Appl. Environ. Microbiol.* **70,** 6865–70 (2004).
41. Rondon, M. R. *et al.* Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Env. Microbiol* **66,** 2541–2547 (2000).
42. Brady, S. F., Chao, C. J. & Clardy, J. New natural product families from an environmental DNA (eDNA) gene cluster. *J. Am. Chem. Soc.* **124,** 9968–9 (2002).
43. Brady, S. F. & Clardy, J. Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *J. Am. Chem. Soc.* **122,** 12903–04 (2000).
44. Sotirova, A. *et al.* The importance of rhamnolipid-biosurfactant-induced changes in bacterial membrane lipids of *Bacillus subtilis* for the antimicrobial activity of thiosulfonates. *Curr. Microbiol.* **65,** 534–41 (2012).
45. Desai, J. D. & Banat, I. M. Microbial production of surfactants and their commercial potential. *Microbiol. Mol. Biol. Rev.* **61,** 47–64 (1997).
46. Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol* **22,** 346–353 (2004).
47. Sørensen, H. P. & Mortensen, K. K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol* **115,** 113–128 (2005).
48. Gabor, E., Liebeton, K., Niehaus, F., Eck, J. & Lorenz, P. Updating the metagenomics toolbox. *Biotechnol. J.* **2,** 201–6 (2007).
49. Chen, S., Kaufman, M. G., Bagdasarian, M., Bates, A. K. & Walker, E. D. Development of an efficient expression system for *Flavobacterium* strains. *Gene* **458,** 1–10 (2010).
50. Brady, S. F. & Clardy, J. Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water. *J. Nat. Prod.* **67,** 1283–6 (2004).
51. Craig, J. W., Cherry, M. A. & Brady, S. F. Long-chain N-acyl amino acid synthases are linked to the putative PEP-CTERM/exosortase protein-sorting system in Gram-negative bacteria. *J. Bacteriol.* **193,** 5707–15 (2011).
52. Thies, S. *et al.* Heterologous production of the lipopeptide biosurfactant serrawettin W1 in Escherichia coli. *J. Biotechnol.* **181,** 27–30 (2014).
53. Brady, S. F. & Clardy, J. N-acyl derivatives of arginine and tryptophan isolated from environmental DNA expressed in *Escherichia coli*. *Org. Lett.* **7,** 3613–6 (2005).
54. Sondermann, H., Shikuma, N. J. & Yildiz, F. H. You've come a long way: c-di-GMP signaling. *Curr. Opin. Microbiol.* **15,** 140–6 (2012).
55. Amikam, D. & Galperin, M. Y. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* **22,** 3–6 (2006).
56. Merighi, M., Lee, V. T., Hyodo, M., Hayakawa, Y. & Lory, S. The second

messenger bis-(3’-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **65,** 876–95 (2007).

57. Morgan, J. L. W., McNamara, J. T. & Zimmer, J. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat. Struct. Mol. Biol.* **21,** 489–96 (2014).
58. Clardy, J. & Brady, S. F. Cyclic AMP directly activates NasP, an N-acyl amino acid antibiotic biosynthetic enzyme cloned from an uncultured beta-proteobacterium. *J. Bacteriol.* **189,** 6487–9 (2007).
59. Pinazo, A., Pons, R., Pérez, L. & Infante, M. R. Amino acids as raw Material for biocompatible surfactants. *Ind. Eng. Chem. Res.* **50,** 4805–17 (2011).
60. Doering, T., Traeger, A. & Kainz, S. Self-tanning composition having a UV protective effect. (2006).
61. Breyfogle, L. E. & Finley, R. A. Personal-care composition comprising oil-soluble solid sunscreens. *WO Patent 2,010,108,085* (2012). at <http://patentscope.wipo.int/search/en/WO2010108085>
62. Dwivedi, D. *et al.* Antimycobacterial serratamolides and diacyl peptoglucosamine derivatives from *Serratia sp*. *J. Nat. Prod.* **71,** 637–41 (2008).
63. Nagarajan, R. Molecular theory for mixed micelles. *Langmuir* **1,** 331–341 (1985).
64. Sambrook, J. & Russell, D. W. *Molecular cloning: a laboratory manual*. (Cold Spring Harbor Laboratory Press, 2001).
65. Troeschel, S. C., Drepper, T., Leggewie, C., Streit, W. R. & Jaeger, K.-E. Novel tools for the functional expression of metagenomic DNA. *Methods Mol. Biol.* **668,** 117–39 (2010).
66. Papworth, C., Bauer, J., Braman, J. & Wright, D. Original Table of Contents. *Strategies* **9,** 3–3 (1996).
67. Lane, D. J. in *Nucleic Acid Techniques in Bacterial Systematics* (eds. Stackebrandt, E. & Goodfellow, M.) 125–175 (John Wiley & Sons, 1991). at <http://ci.nii.ac.jp/naid/10008470323/en/>
68. Brosius, J., Dull, T. J., Sleeter, D. D. & Noller, H. F. Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *J Mol Biol* **148,** 107–127 (1981).
69. Kane, M. D., Poulsen, L. K. & Stahl, D. A. Monitoring the enrichment and isolation of sulfate-reducing bacteria by using oligonucleotide hybridization probes designed from environmentally derived 16S rRNA sequences. *Appl Env. Microbiol* **59,** 682–686 (1993).
70. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–402 (1997).
71. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–71 (2004).
72. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
73. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).
74. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4,** 363–71 (2009).
75. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory

research and analysis. *J. Comput. Chem.* **25,** 1605–12 (2004).

76. Schrödinger, L. *The PyMOL Molecular Graphics System, Version~1.3r1*. (2010).

77. Pelzer, A. *et al.* Subtilase SprP exerts pleiotropic effects in *Pseudomonas aeruginosa*. *Microbiologyopen* **3,** 89–103 (2014).

78. Jaeger, K.-E. & Kovacic, F. Determination of lipolytic enzyme activities. *Methods Mol. Biol.* **1149,** 111–34 (2014).

79. Burch, A. Y., Shimada, B. K., Browne, P. J. & Lindow, S. E. Novel high-throughput detection method to assess bacterial surfactant production. *Appl. Environ. Microbiol.* **76,** 5363–72 (2010).

80. Harkins, W. D. & Jordan, H. F. A method for the determination of surface and interfacial tension from the maximum pull on a ring. *J. Am. Chem. Soc.* **52,** 1751–1772 (1930).

81. Padday, J. F., Pitt, A. R. & Pashley, R. M. Menisci at a free liquid surface: surface tension from the maximum pull on a rod. *J. Chem. Soc. Faraday Trans. 1 Phys. Chem. Condens. Phases* **71,** 1919 (1975).

82. Aha, B. S. Biologisch abbaubare Tenside aus nachwachsenden Rohstoffen: N-Acylaminosäuren - Synthesen und Tensideigenschaften. (Bergische Universität Wuppertal, 2010). at <http://elpub.bib.uni-wuppertal.de/servlets/DocumentServlet?id=360>

83. Bauer, A. W., Kirby, W. M., Sherris, J. C. & Turck, M. Antibiotic susceptibility testing by a standardized single disk method. *Am. J. Clin. Pathol.* **45,** 493–6 (1966).

84. Kunst, F. *et al.* The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390,** 249–56 (1997).

85. McClean, K. H. *et al.* Quorum sensing and *Chromobacterium violaceum*: exploitation of violacein production and inhibition for the detection of N-acylhomoserine lactones. *Microbiology* **143,** 3703–3711 (1997).

86. Abe, S., Takayama, K.-I. & Kinoshita, S. Taxonomical studies on glutamic acid-producing bacteria. *J. Gen. Appl. Microbiol.* **13,** 279–301 (1967).

87. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189,** 113–130 (1986).

88. Durfee, T. *et al.* The complete genome sequence of *Escherichia coli DH10B*: insights into the biology of a laboratory workhorse. *J. Bacteriol.* **190,** 2597–606 (2008).

89. Woodcock, D. M. *et al.* Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Res.* **17,** 3469–3478 (1989).

90. Holloway, B. W., Krishnapillai, V. & Morgan, A. F. Chromosomal genetics of *Pseudomonas*. *Microbiol. Rev.* **43,** 73–102 (1979).

91. Nelson, K. E. *et al.* Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4,** 799–808 (2002).

92. Labrum, E. L. & Bunting, M. I. Spontaneous and induced color-variation of the HY strain of *Serratia marcescens*. *J. Bacteriol.* **65,** 394–404 (1953).

93. Winkler, U. & Timmis, K. Pleiotropic mutations in Serratia marcescens which increase the synthesis of certain exocellular proteins and the rate of spontaneous prophage induction. *Mol. Gen. Genet.* **124,** 197–206 (1973).

94. Schneiker, S. *et al.* Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* **25,** 1281–9 (2007).

Figure legends

Figure 1. Dendrogram of the 16S rRNA genes identified in the bacterial community of a slaughterhouse biofilm. Phylogenetic relationship of bacterial 16S rRNA gene sequences to the closest known relatives was determined by Maximum-Likelihood analysis. The newly identified sequences were grouped into OTUs (similarity >97%) and are marked in bold. The percentage of bootstrap re-samplings is ≥88 and is indicated at nodes. Numbers in parentheses indicate the amount of sequences associated with the OTU. The scale bar represents the expected number of changes per nucleotide position.

Figure 2. Phenotypic screening of clones from a slaugtherhouse metagenomic library. A) Hemolytic activities detected as halos on blood agar plates. Surface activities detected by B) halo formation with the atomized oil assay and C) an optical distortion of the grid caused by altered surface tension of culture supernatants in microtiter plates. Two *E. coli* DH10b clones expressing genes from environmental DNA fragments SA343 and SA354 were compared with a control strain carrying the empty vector pEBP18 (EV).

Figure 3. Identification of genes responsible for biosurfactant activity of clones SA343 and SA354. A) Organisation of putative genes identified on eDNA fragments from pEBP*SA343* and pEBP*SA354*. B) Hemolytic activity on blood agar plates, and C) surface activity on LB agar plates determined with the atomized oil assay of *E. coli* BL21(DE3) expressing *nas354* and carrying an empty vector (EV, pET22b). D) Unique two domain architecture of Nas354 (*orf*5 of eDNA fragment 354) modelled with Phyre II server. The model of Nas343 resembles the one of Nas354 (see Figure S2). In green is shown the N-terminal NAS domain modelled using the crystal structure of FeeM (PDB

code 2G0B) and in orange is shown the C-terminal domain modelled using the c-di-GMP-binding protein from *Pseudomonas aeruginosa* (PDB code 1YWU). E) Superimposition of modelled NAS domains of Nas343 (red), Nas354 (green) and FeeM structure (blue) showing the high structural similarity of the three proteins. Structural conservation of the putative catalytic active site glutamate residues shown as sticks is indicated by the red circle. F) Amino acid sequence alignment of NAS domains (residues 1-200) from Nas343 and Nas354 with known type I Nas proteins, FeeM[38], NasY1 and NasY2[40]. Conserved sequence motifs characteristic for Nas type I proteins are indicated by red boxes. The putative active site glutamate residues are highlighted in yellow. Residues similar and identical in at least four proteins are indicated by grey and black background, respectively.

Figure 4: Comparison of chemically synthesized N-myristoyltyrosine with biologically produced N-acyltyrosine. A) Thin layer chromatographic analysis showing similar migration distances (indicated by an arrow) of pure chemically synthesized N-myristoyltyrosine and N-acyltyrosine isolated from culture supernatant of *E. coli* DH10b carrying pEBP*SA354*. Start and solvent fronts are indicated by one and two asterisks, respectively. The TLC plate was stained with iodine vapour. B) Surface tension measurements with chemically synthesized N-myristoyltyrosine and biologically produced N-acyltyrosine show similar properties. Both compounds were dissolved in aqueous solution of NaOH (0.1 M, pH 12). Solid and dotted lines represent trend lines generated with Microsoft Excel2010 to illustrate changes in surface tension dependent on concentration of chemically synthesized N-myristoyltyrosine and biologically produced N-acyltyrosine, respectively. Values represent the mean of three

measurements. The structure of the predominant compound of the isolated biosurfactant identified as N-myristoyltyrosine is shown in the coffered box. The structural data were obtained by NMR analysis of biosurfactant purified from liquid cultures of *E. coli* DH10b carrying pEBP*SA343* or pEBP*SA354*.

Figure 5. The effect of N-acyltyrosine on the growth of different bacterial strains. A Kirby-Bauer assay was performed with biologically produced N-acyltyrosine at concentrations indicated to the left. Whatman paper discs soaked with N-acyltyrosine dissolved in ethanol were applied to Mueller-Hinton agar plates streaked with: A) *Bacillus subtilis* TEB1030, B) *Sorangium cellulosum* DSM53796, C) *Corynebacterium glutamicum* ATCC13032, D) *Chromobacterium violaceum* CV026. E) Atomized oil assay on Mueller-Hinton agar plates indicates zone of diffusion of N-acyltyrosine under test conditions.

Tables

Table 1 Physical properties of N-acyltyrosines

	N-acyl tyrosine biologically produced	**N-myristoyl tyrosine chemically synthesized**
R_f-value	0.75	0.75
specific rotation of polarized light	α]D20°C = +2.1	[α]D20°C = +2.034
CMC (pH 12)	0.114 % (w/v)	0.063 % (w/v)
Min. surface tension (pH 12)	34 mN/m	30 mN/m

Table 2: Strains and plasmids used in this study

Strains	**Genotype**	**Reference**
Bacillus subtilis TEB1030 Marburg 168	wild-type	[84]
Chromobacterium violaceum CV026	variant of ATCC31532 deficient for violacein production	[85]
Corynebacterium glutamicum ATCC13032	wild-type	[86]
Escherichia coli BL21(DE3)	*F- ompT hsdSB* (*rB-mB-*) *gal dcm* (DE3)	[87]
Escherichia coli DH10b	*F- mcrA* Δ(*mrr-hsdRMS-mcrBC*) *Φ80lacZΔM15 ΔlacX74 endA1 recA1 deoR* Δ(*ara,leu*)*7697 araD139 galU galK nupG rpsL λ-*	[88]
Escherichia coli DH5α	F- φ80*lacZ*ΔM15 Δ(lacZYA-*arg*F) U169 *recA*1 *end*A1 *hsdR17* (rk-,mk+*) phoA sup*E44 λ- *thi*-1 *gyrA*96 *relA*1	[89]
Pseudomonas aeruginosa PAO1	wild-type	[90]
Pseudomonas putida KT2440	wild-type	[91]
Serratia marcescens DSM12481	wild-type	[92,93]
Sorangium cellulosum DSM53796	wild-type	[94]
Plasmids	**Genotype**	**Source/reference**
pEBP18	*amyE' 'amyE* ColE1 *cos gfpmut3** Km^R Cm^R MCS *minR'* P_{T7} P_{xyl}	[30]
pEBP-*SM321 – SMVIIIE12*	pEBP18 containing fragments of metagenomic DNA of proteolytic clones listed in the table S1	This work
pEBP-*TB94 – TB I H8*	pEBP18 containing fragments of metagenomic DNA from lipolytic clones listed in the table S1	This work
pEBP-*SA343*	pEBP18 containing a 3.3 kb fragment of metagenomic DNA inclusive *nas343* gene	This work
pEBP-*SA354*	pEBP18 containing a 8.8 kb fragment of metagenomic DNA inclusive *nas354* gene	This work
pET22b	*ColE1* P_{T7}, φ10*lacIq*, Amp^R	EMD Millipore
pET*nas354*	pET22b containing the *nas354* gene	This work
pET*nas354*^{E103A}	pET22b containing the *nas354* gene with Glu103 substituted by Ala	This work
pGEM®-T Easy	Amp^R, *lacZ*	Promega

Figure 1

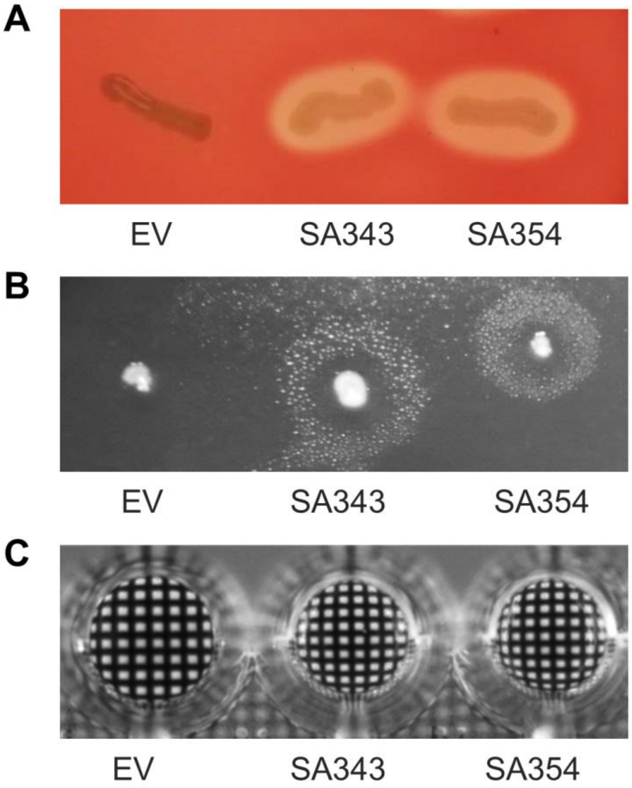
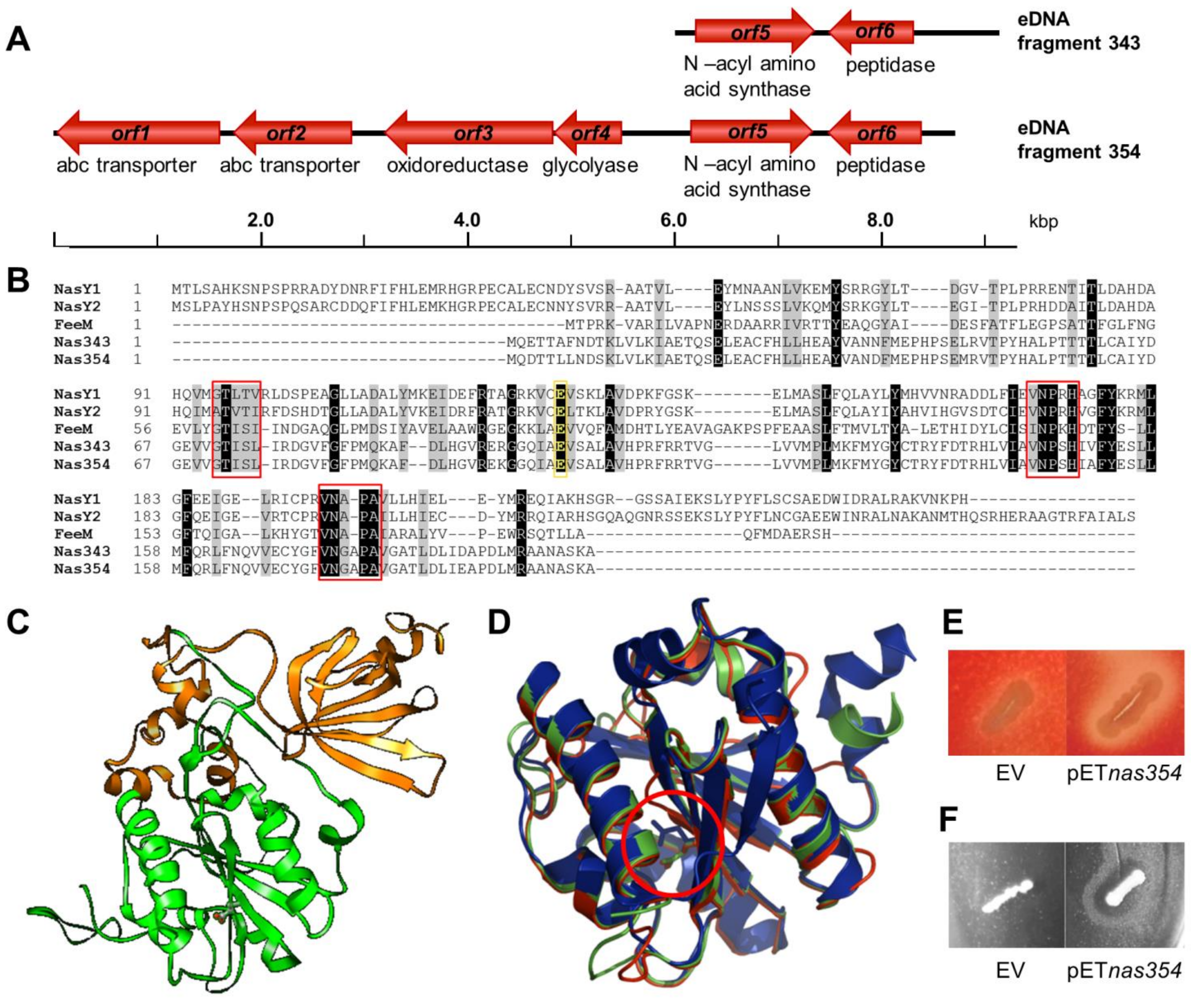


Figure 2



```
NasY1   1  MTLSAHKSNPSPRRADYDNRFIFHLEMRHGRPECALECNDYSVSR-AATVL----EYMNAANLVKEMYSRRGYLT----DGV-TPLPRRENTITLDAHDA
NasY2   1  MSLPAYHSNPSPQSARCDDQFIFHLEMKHGRPECALECNNYSVRR-AATVL----EYLNSSSLVKQMYSRKGYLT----EGI-TPLPRHDDAITLDAHDA
FeeM    1  ----------------------------------------MTPRK-VARILVAPNERDAARRIVRTTYEAQGYAI----DESFATFLEGPSATTFGLFNG
Nas343  1  ----------------------------------MQETTAFNDTKLVLKIAETQSELEACFHLLHEAYVANNFMEPHPSELRVTPYHALPTTTTLCAIYD
Nas354  1  ----------------------------------MQDTTLLNDSKLVLKIAETQSELEACFHLLHEAYVANDFMEPHPSEMRVTPYHALPTTTTLCAIYD

NasY1   91 HQVMGTLTVRLDSPEAGLLADALYMKEIDEFRTAGRKVCEVSKLAVDPKFGSK--------ELMASLFQLAYLYMHVVNRADDLFIEVNPRHAGFYKRML
NasY2   91 HQIMATVTIRFDSHDTGLLADALYVKEIDRFRATGRKVCELTKLAVDPRYGSK--------ELMASLFQLAYIYAHVIHGVSDTCIEVNPRHVGFYKRML
FeeM    56 EVLYGTISI-INDGAQGLPMDSIYAVELAAWRGEGKKLAEVVQFAMDHTLYEAVAGAKPSPFEAASLFTMVLTYA-LETHIDYLCISINPKHDTFYS-LL
Nas343  67 GEVVGTISL-IRDGVFGFPMQKAF--DLHGVRERGGQIAEVSALAVHPRFRRTVG------LVVMPLMKFMYGYCTRYFDTRHLVIAVNPSHIVFYESLL
Nas354  67 GEVVGTISL-IRDGVFGFPMQKAF--DLHGVREKGGQIAEVSALAVHPRFRRTVG------LVVMPLMKFMYGYCTRYFDTRHLVIAVNPSHIAFYESLL

NasY1  183 GFEEIGE--LRICPRVNA-PAVLLHIEL---E-YMREQIAKHSGR--GSSAIEKSLYPYFLSCSAEDWIDRALRAKVNKPH-----------------
NasY2  183 GFQEIGE--VRTCPRVNA-PAILLHIEC---D-YMRRQIARHSGQAQGNRSSEKSLYPYFLNCGAEEWINRALNAKANMTHQSRHERAAGTRFAIALS
FeeM   153 GFTQIGA--LKHYGTVNA-PAIARALYV---P-EWRSQTLLA---------------QFMDAERSH--------------------------------
Nas343 158 MFQRLFNQVVECYGFVNGAPAVGATLDLIDAPDLMRAANASKA-------------------------------------------------------
Nas354 158 MFQRLFNQVVECYGFVNGAPAVGATLDLIEAPDLMRAANASKA-------------------------------------------------------
```

Figure 3

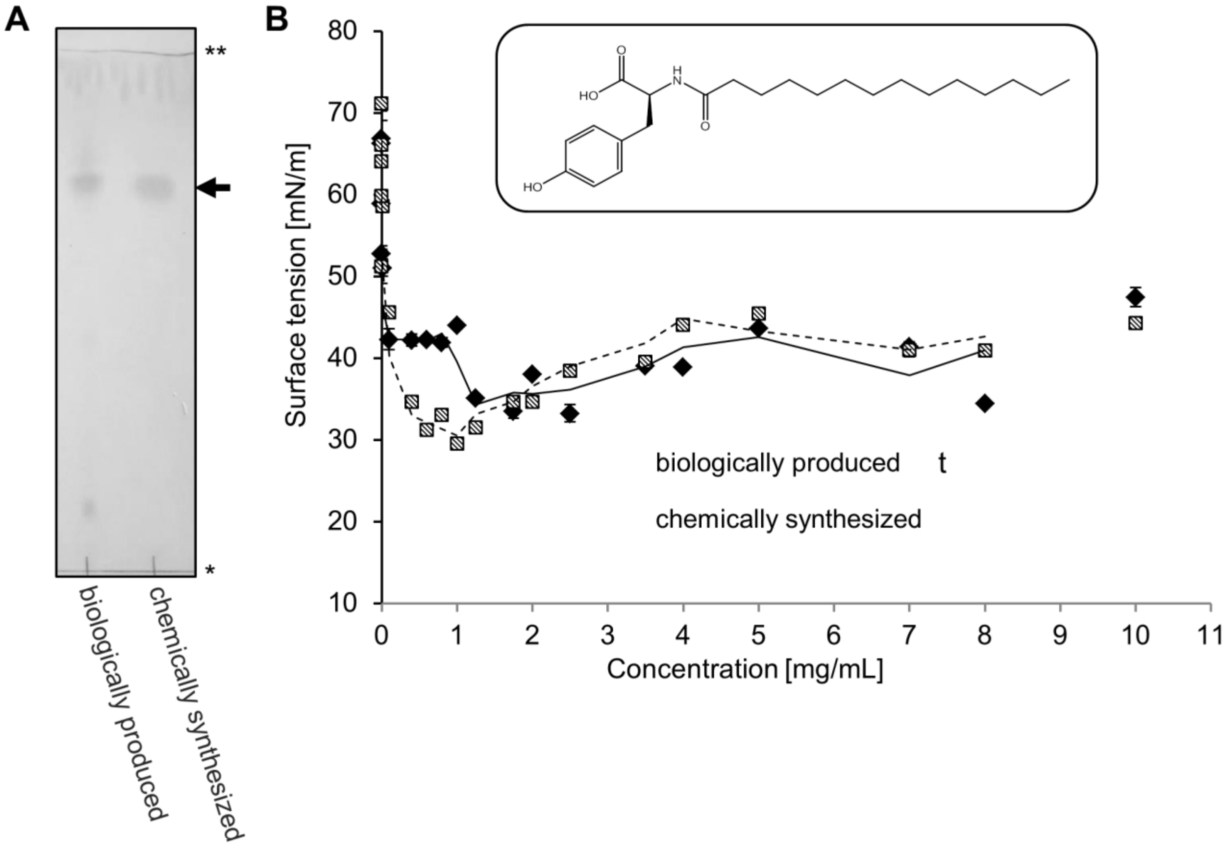

A
**
*
biologically produced
chemically synthesized
B
O
HO
H
N
O
HO
80
70
60
50
40
30
20
10
Surface tension [mN/m]
biologically produced t
chemically synthesized
0
1
2
3
4
5
6
7
8
9
10
11
Concentration [mg/mL]

Figure 4