nature chemical biology

ARTICLE

PUBLISHED ONLINE: 3 OCTOBER 2016 | DOI: 10.1038/NCHEMBIO.2189

Functional mining of transporters using synthetic selections

Hans J Genee[1,2], Anne P Bali[1,2], Søren D Petersen[1], Solvej Siedler[1], Mads T Bonde[1], Luisa S Gronenberg[2], Mette Kristensen[1], Scott J Harrison[1] & Morten O A Sommer[1]*

Only 25% of bacterial membrane transporters have functional annotation owing to the difficulty of experimental study and of accurate prediction of their function. Here we report a sequence-independent method for high-throughput mining of novel transporters. The method is based on ligand-responsive biosensor systems that enable selective growth of cells only if they encode a ligand-specific importer. We developed such a synthetic selection system for thiamine pyrophosphate and mined soil and gut metagenomes for thiamine-uptake functions. We identified several members of a novel class of thiamine transporters, PnuT, which is widely distributed across multiple bacterial phyla. We demonstrate that with modular replacement of the biosensor, we could expand our method to xanthine and identify xanthine permeases from gut and soil metagenomes. Our results demonstrate how synthetic-biology approaches can effectively be deployed to functionally mine metagenomes and elucidate sequence–function relationships of small-molecule transport systems in bacteria.

Understanding microbial transporters is important for several biotechnological and medical applications: the development of cell factories for manufacturing of biochemicals requires engineering of nutrient-import and product-secretion systems[1]; the effects of antibiotic treatment and the development of antibiotic resistance are influenced by our understanding of uptake and export in pathogenic species[2]; and for auxotrophic pathogenic species, uptake transporters of essential metabolites constitute potent drug targets[3]. Moreover, recent studies show that in addition to being vital to the individual cell in which they reside, transporters may have an important role in metabolic interactions and codependencies within microbial communities such as the human gut microflora[4].

Despite the importance of microbial membrane transporters for biotechnology and human health, three-quarters of bacterial genes predicted to be transporters in databases have the functional annotation "unknown"[5,6]. Even in the model organism *Escherichia coli*, in which membrane transporters constitute 12.6% of total open reading frames[7], 53% of them currently lack experimental validation[6]. Because the specificity of transporters toward their substrates is highly flexible, the relationship between the sequence and function of transporters is exceptionally complex, making prediction of substrate specificity based on sequence homology extremely difficult. Therefore, sequence-independent methods are needed for broadly identifying and validating substrate-specific transport systems.

Synthetic-biology systems that use small-molecule biosensors such as riboswitches and allosteric transcription factors coupled to screenable marker genes represent promising approaches for noninvasive detection and quantification of intracellular metabolite pools at the single-cell level. When applied to high-throughput screening, biosensor systems have proven to be powerful tools for the elucidation of genotype–phenotype relations, including interrogation of enzyme mutant libraries[8], and for enriching phenotypes with improved metabolic capacity[9–13]. To date, biosensors for at least 251 small-molecule compound classes have been identified in nature[14], and several more have been created synthetically[15].

Here we report a novel method that uses small-molecule biosensors to functionally mine and isolate genes encoding bacterial uptake systems. We demonstrate the method for thiamine (vitamin B1), a key cofactor in all living cells. Using a thiamine pyrophosphate (TPP)-responsive riboswitch as a biosensor, we identified several members of a new class of bacterial thiamine transporters, PnuT. We further demonstrated the generality of the method by deploying a fully synthetic riboswitch responsive to xanthine alkaloids to functionally mine xanthine importers from metagenomes.

RESULTS

A riboswitch-based selection system for thiamine uptake

We used the ThiM19 riboswitch[16] as an *in vivo* sensor for the thiamine-uptake selection system. This riboswitch initiates translation of mRNA in response to high intracellular levels of TPP, the biologically active form of thiamine. The ThiM19 riboswitch is an engineered derivative of the naturally occurring ThiM riboswitch from *E. coli* that reduces translation of thiazole kinase at high TPP levels, thereby acting as a feedback regulator of the TPP salvage pathway[17]. We cloned ThiM19 in a plasmid to regulate expression of an antibiotic-resistance cassette in response to high intracellular levels of TPP in *E. coli* DH10B. This system enabled the construction of a synthetic selection strain whose growth was dependent on excess intracellular TPP (**Fig. 1a**).

Wild-type cells (*E. coli* DH10B) lack the native ThiBPQ transporter for active thiamine uptake. Growth of this strain harboring the thiamine synthetic selection system requires media supplementation with 10 μM thiamine or more under selective conditions (**Fig. 1b**). In the cytoplasm, thiamine is phosphorylated to thiamine monophosphate (TMP) and TPP by the kinases ThiK and ThiL, respectively. We measured intracellular concentrations of thiamines by high-pressure liquid chromatography (HPLC) and found that the endogenous concentration of TPP (synthesized *de novo*) was approximately 1 μM when no extracellular thiamine was added (**Fig. 1c**). When we added 10 μM and 100 μM extracellular thiamine to the growth medium, we measured intracellular TPP levels

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark. [2]Biosyntia ApS, Copenhagen, Denmark. *e-mail: msom@bio.dtu.dk

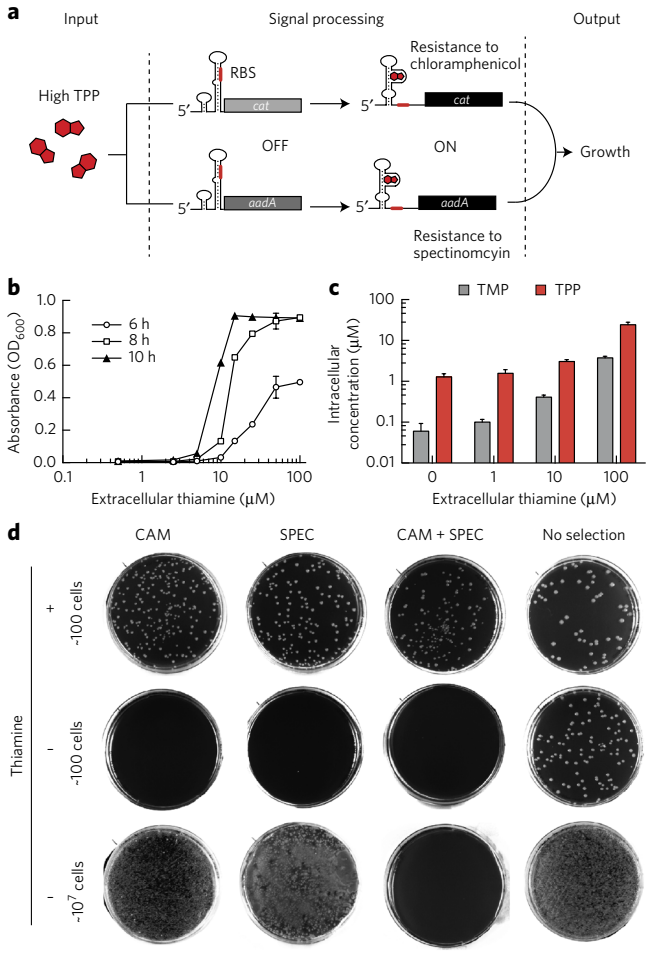

Figure 1 | Synthetic selection system for thiamine uptake. (**a**) The dual ribosome binding site (RBS) selection system controlling chloramphenicol-resistance and spectinomycin-resistance genes (*cat* and *aadA*). Translation of the resistance genes is enabled only after binding of TPP. The dual selection reduces the number of false positives, as false triggering (e.g., by mutation of one riboswitch) will not lead to cell growth. (**b**) Thiamine dose-response curves of the *E. coli* selection strain at three time points. (**c**) Intracellular concentrations of TMP and TPP in *E. coli* Dh10B grown in cultures fed with thiamine at different concentrations. A high extracellular thiamine level is reflected by a higher intracellular TMP/TPP level owing to diffusion and phosphorylation by kinases. Data in **c** and **d** represent averages of biological triplicate experiments, and error bars represent s.d. (n = 3). (**d**) Agar-plate-based growth assays of the *E. coli* selection strain under different selection conditions: chloramphenicol (CAM), spectinomycin (SPEC), chloramphenicol and spectinomycin (CAM + SPEC), and no antibiotics. When approximately 100 cells were plated, thiamine supplementation was required for growth under all selection conditions (top and middle rows). When many cells (>10^7) were plated in the absence of thiamine, the dual selection eradicated false positive growth completely (bottom row). The images shown are representative of biological replicates (n = 3).

of 3 μM and 25 μM, respectively. No improved growth advantage was observed with concentrations greater than 100 μM extracellular thiamine, indicating that the selection system had a dynamic range from approximately 3 μM to 25 μM intracellular TPP.

We addressed the challenges of false positives with a dual-selection design in which two separate TPP riboswitches control two different antibiotic resistance genes conferring resistance to chloramphenicol and spectinomycin, respectively (**Fig. 1a**, **Supplementary Results**, **Supplementary Fig. 1**). After genetic fine-tuning and optimization of selection conditions, the dual-selection system reduced the escape rate approximately 1,000-fold, from more than 1 in 10^4 plated cells under single selection (chloramphenicol or spectinomycin alone) to <1 in 10^7 cells with the dual selection (**Supplementary Fig. 2**). In the absence of added thiamine, dual selection allowed plating of up to ~5 × 10^7 cells on a standard Petri dish (94-mm diameter) without any growth, whereas bacterial lawns emerged when we used a single-selection system (spectinomycin resistance or chloramphenicol resistance alone; **Fig. 1d**).

Metagenomic functional selection of uptake transporter

Having developed a strict selection system with a low escape rate, we next tested its applicability as a tool for functional identification and isolation of bacterial thiamine transporters. To date, only five individual bacterial thiamine transporter proteins have been reported and experimentally studied[18–21], and thus it is likely that unidentified thiamine transporter classes exist. We focused our functional search for thiamine transporters on metagenomic DNA libraries derived from soil and fecal microbiota samples, thereby tapping into the genomes of thousands of different species, including both thiamine-producing and thiamine-requiring bacteria[22,23], the majority of which cannot readily be cultured in the lab. We transformed five separate DNA libraries with a combined size of 2.57 Gbp (**Supplementary Table 1**) expressed as ~2-kb fragments into *E. coli* DH10B harboring the selection system (**Fig. 2a**). From each transformed library, we assayed ~100 million cells on selective agar medium supplemented with zero or 0.5 μM thiamine. At 0.5 μM extracellular thiamine, the intracellular TPP concentration of the selection strain was too low to induce growth (**Fig. 1b**,**c**), and hence acquisition and expression of a functional thiamine uptake system would be necessary for growth. Interestingly, cells transformed with any of the DNA libraries yielded several colonies when they were grown on plates supplemented with 0.5 μM thiamine, whereas only one colony appeared in the absence of thiamine (**Supplementary Table 2**). No colonies emerged from cells transformed with the control library (empty expression vector) on plates supplemented with 0.5 μM thiamine, underlining the low escape rate of the selection system (**Fig. 2b**). We sequenced the metagenomic inserts from 120 colonies (24 from each library) and identified 55 unique metagenomic inserts. Sequence alignments showed that several inserts contained overlapping sequence regions (>500 bp of 100% identity) indicating a common genomic origin (**Supplementary Fig. 3**). Assembly of overlapping inserts generated 37 contigs (hereinafter referred to as CON1–CON37) representing unique genomic segments, which we annotated using BLAST and RAST[24] (**Supplementary Tables 2** and **3**).

Interestingly, an ~600-bp open reading frame with homology to genes annotated as nicotinamide riboside and nicotinamide mononucleotide transporter genes (PnuC) was present in 113 of the 124 (91%) functionally selected inserts. The one insert selected in the absence of thiamine encodes a multidrug-resistance protein that may confer resistance to both chloramphenicol and spectinomycin. The remaining ten inserts encode putative and hypothetical genes (**Supplementary Table 4**), and the mechanism for their selection is unclear. Previous functional metagenomic selections have shown that antibiotic-resistance genes are abundant in both soil and gut microbiota[25,26]. Our finding of only one antibiotic-resistance gene is probably a result of using two types of antibiotics in parallel, demonstrating another important advantage of the dual-selection system when applied to functional metagenomics. Of the 113 inserts encoding putative Pnu transporters, we assembled 26 unique genomic segments encoding homologous Pnu-type transporters (**Fig. 2c**). Members of the Pnu family have previously been shown to transport nicotinamide riboside (PnuC from *E. coli*[27], *Haemophilus influenza*[27], *Neisseria mucosa*[28] and *Salmonella enterica*[29]) and riboflavin (PnuX/RibM from *Corynebacterium glutamicum*[30] and

Streptomyces davawensis[31]), but so far thiamine transport by this family has not, to our knowledge, been experimentally demonstrated[32–34].

To confirm that the selected genes encode thiamine-uptake transporters, we carried out functional growth assays, which showed that growth was dependent on extracellular thiamine (**Fig. 2c**). Dose-response assays demonstrated recovery of growth at extracellular thiamine concentrations as low as 0.1 μM for strains harboring a functionally selected transporter (**Fig. 2d**). HPLC measurement of intracellular thiamine pools showed that the selected clones had substantially higher TMP and TPP levels than the wild-type strain when supplemented with extracellular thiamine (**Fig. 2e,f**), providing evidence that the identified genes encode thiamine transporters.

On average, the sequence similarity between the selected Pnu thiamine transporters (hereinafter referred to as PnuT) and their closest homologs in GenBank is 96.7% at the amino acid level. Yet none of the closely related homologs are annotated as thiamine transporters, although many are annotated with "nicotinamide mononucleotide transport" or "nicotinamide riboside transport" (**Supplementary Table 3**). This illustrates the inadequacy of automated annotation of transporter specificity on the sole basis of sequence similarity and highlights the utility of the functional-selection approach to enable or correct annotation of genomic and metagenomic sequencing data.

Shared synteny was observed for inserts that cluster phylogenetically according to a neighbor-joining tree based on *pnuT* nucleotide sequences (**Fig. 2c**). In ten inserts, all selected from gut microbiome libraries, *pnuT* colocalizes with a gene that matches a thiamine pyrophosphokinase, predicted to phosphorylate thiamine to TPP. Colocalization with specific kinases has also been observed for genes encoding PnuC and PnuX[33], and substrate phosphorylation is proposed to be an essential mechanism for driving import by Pnu transporters, as phosphorylated substrates can no longer bind the transporter and consequently become trapped in the cytoplasm. Because *E. coli* encodes endogenous thiamine kinases, a functional kinase encoded by the insert was not required for functional selection, and 23 selected inserts contain truncated thiamine kinase genes or no thiamine kinase genes (**Fig. 2c**).

In addition to kinases specific to thiamine, another five selected inserts encoded kinases that are predicted to be specific to nicotinamide riboside (**Fig. 2c**). The colocalization of *pnuT* with nicotinamide riboside kinases suggests that such Pnu transporters may transport nicotinamide riboside in addition to thiamine. Such substrate promiscuity could potentially reflect an evolutionary state in which the Pnu transporter has evolved to transport two vitamin precursors. Although transporter specificity has previously been assigned successfully on the basis of genetic colocalization[33], this example demonstrates the importance of experimental methods to complement sequence-based predictions. This is further supported by the observation that for 9 of the 26 inserts, no neighboring genes provided clues to substrate specificity. Also, regulatory motifs such as riboswitches can be a powerful predictor of transporter specificity[33], but scanning of all contigs using RibX[35] identified only one TPP riboswitch, preceding *pnuT* on CON10.

Parts of a gene encoding a TonB-dependent outer-membrane transporter were found in 11 of the 26 inserts (**Fig. 2c** and **Supplementary Table 3**). In Gram-negative bacteria, TonB-dependent outer-membrane transporters are known to actively translocate metal ions, cobalamin and carbohydrates across the outer membrane, as these substrates either are encountered at very low concentrations or are poorly transported through nonspecific porins[36]. It has previously been proposed, on the basis of their colocalization with putative *pnuT* sequences, that TonB-dependent outer-membrane transporters could be involved in the translocation of thiamine across the outer membrane[32]. The functional validation of PnuT provided here strongly supports this hypothesis.

PnuT is abundant in the gut microbiota

We next investigated the phylogenetic origin and estimated the prevalence of *pnuT* in the bacterial kingdom. Our analysis showed that the selected inserts originate from diverse bacterial species and suggested that PnuT is an important thiamine transporter that is widespread in the bacterial kingdom (**Fig. 3a**). We found that selected inserts harbored by the soil microbiome originate predominantly from the phylum Proteobacteria and are closely related to the Pnu transporter from members such as the opportunistic pathogens *Acinetobacter baumannii* (90.3% amino acid sequence identity to CON10 PnuT) and *Pseudomonas auruginosa* (68% amino acid sequence identity to CON9 PnuT), as well as from the industrially relevant *Pseudomonas putida* (69% amino acid sequence identity to CON9 PnuT). *A. baumannii* and the *Pseudomonas* strains do not encode any other known thiamine transporters (**Supplementary Table 11**) and are likely to depend solely on PnuT for thiamine uptake.

PnuT genes isolated from the gut microbiomes are closely related to Pnu transporters of the Bacteroidetes phylum, and related homologs include Pnu transporters of *Parabacteroides* spp., *Odoribacter* and *Alistipes* (98.5–99.5% amino acid sequence identity), as well as opportunistic pathogens including *Bacteroides fragilis* (76.2% amino acid sequence identity to CON11) and *Prevotella copri* (99.5% amino acid sequence identity to CON24) (**Fig. 3a**). Recent sequencing programs have uncovered 383 full genomes of the human gut microbiome[37]. We searched this database for thiamine transporters and found that Bacteroidetes, a dominant phylum in the gut microbiota, does not encode any of the hitherto experimentally validated thiamine transporters. In contrast, every sequenced strain within this phylum harbors a Pnu transporter. The absence of any other known thiamine transporters indicates that PnuT may be the sole thiamine transporter of the Bacteroidetes phylum.

To further investigate the significance of PnuT, we performed a comparative genomics analysis of 1,752 complete bacterial genomes available at MicrobesOnline[38]. Strikingly, 56% of all genomes analyzed (986 of 1,752) did not contain orthologs of any currently validated thiamine transporters. To classify prototrophs and auxotrophs, we mapped genes involved in TPP biosynthesis and salvage. Among the strains without known thiamine transporters, 50% (493 of 986) lacked the capacity for *de novo* TPP synthesis as indicated by the absence of *thiC*, which encodes a protein involved in the committing step of TPP synthesis. This indicates that these species are dependent either on direct thiamine uptake or on salvage of the precursors hydroxyethylthiazole (HET) and hydroxymethylpyrimidine (HMP). Surprisingly, the majority of the genomes that lacked *de novo* synthesis also lacked the *thiE* gene (388 of 493; 79%) needed for TPP generation from HMP and HET. Consequently, about 20% of the bacterial strains, including clinically important pathogens such as *Prevotella multisaccharivorax* and certain species of *Streptococcus pneumoniae* (**Supplementary Fig. 4** and **Supplementary Table 10**), do not contain genes known to be involved in biosynthesis, salvage or import of thiamine. These strains must rely on uncharacterized thiamine biosynthesis, salvage or uptake systems for survival.

For such pathogens, thiamine transporters constitute a direct drug target that could ultimately prevent host colonization and pathogenicity. Indeed, in *Haemophilus influenza*, which lacks biosynthesis genes for the coenzyme nicotinamide adenine dinucleotide (NAD), transport of nicotinamide riboside (an NAD precursor) by PnuC has been shown to be essential to host colonization[3]. Interestingly, we found that many of the strains that are predicted thiamine auxotrophs harbor a Pnu transporter but no other previously validated thiamine transporter homolog (**Fig. 3a** and **Supplementary Table 11**). To test the thiamine-transport capacity of the Pnu transporter from such bacteria, we synthesized and cloned the Pnu genes of four selected clinically important pathogens for which the thiamine-uptake mechanism is presently

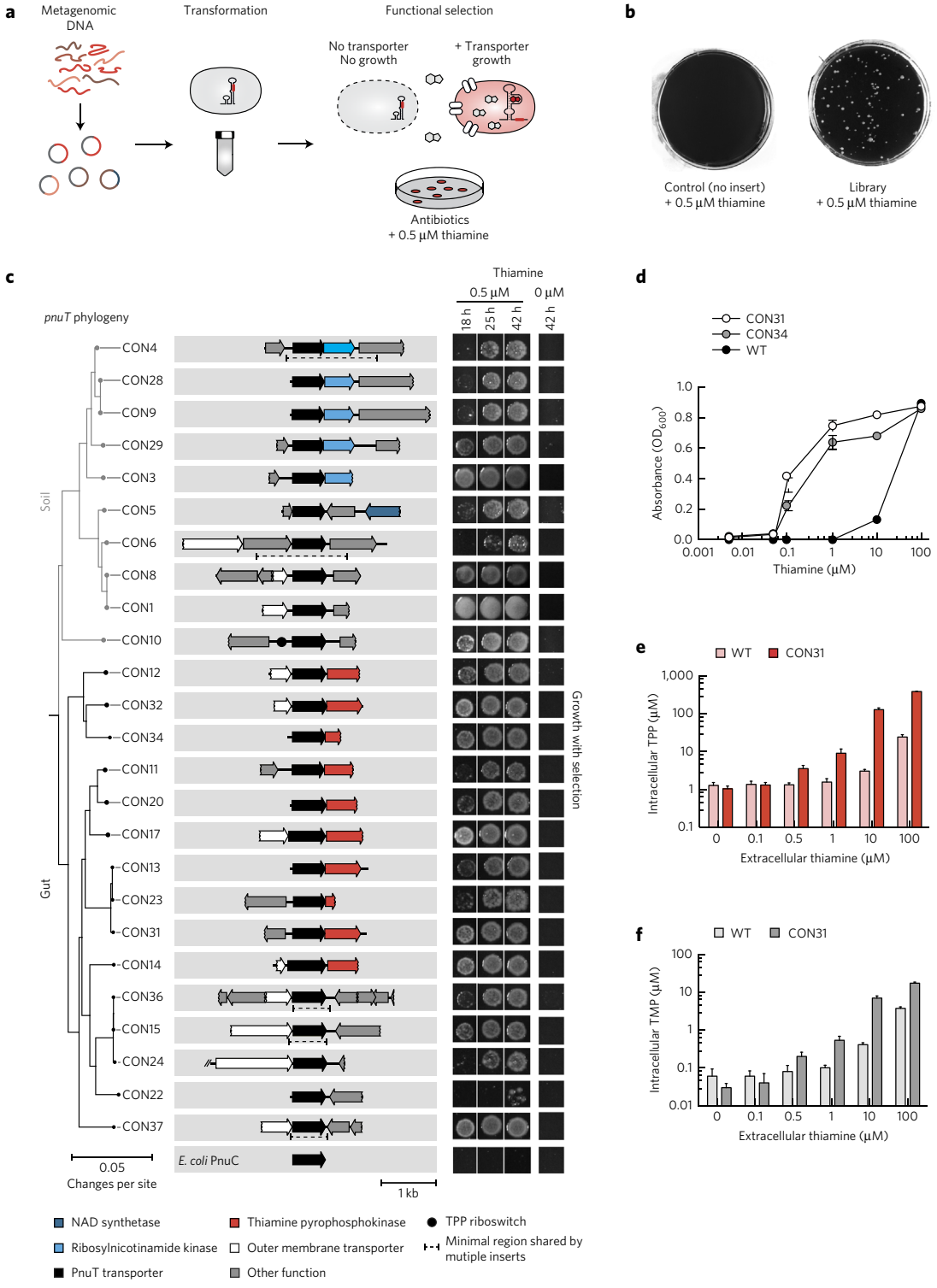

Figure 2 | Functional metagenomic selection of thiamine transporters. (**a**) Total DNA extracted from soil and gut fecal samples (metagenomic DNA) was fragmented into ~2-kb fragments, cloned into an expression vector and transformed into an *E. coli* host strain harboring the thiamine selection system. The cell library was plated on selective growth medium supplemented with low amounts of thiamine. Cells that expressed a thiamine-uptake transporter from the metagenomic DNA insert imported extracellular thiamine and had increased intracellular TPP concentrations, leading to induction of riboswitch-mediated antibiotic resistance. (**b**) Colony formation on selective agar plates after a 40-h incubation of ~25 million cells of the selection strain transformed with either a human fecal gut metagenomic library (right) or a control library (left). (**c**) Overview of functionally selected metagenomic fragments encoding thiamine transporters. Left, neighbor-joining tree based on multiple alignment of the *pnuT* nucleotide sequences. Right, phenotypic response (spot assays) of each selection strain when expressing the metagenomic insert in the presence of no or low amounts of thiamine. Images are representative of three replicates. (**d**) Thiamine dose-response curves of the wild-type strain (WT) and two strains expressing functionally selected thiamine transporters. (**e**,**f**) Intracellular concentrations of TPP (**e**) and TMP (**f**) in a functionally selected strain (CON31) and the wild-type strain at different concentrations of exogenously added thiamine. Data in **d**–**f** represent averages of biological triplicate experiments, and error bars represent s.d. (n = 3).

unknown: *Bacteroides fragilis*, *Streptococcus pneumoniae*, *Prevotella multisaccharivorax* and *Helicobacter pylori*. Of these, thiamine auxotrophy has been experimentally confirmed for *H. pylori*[39]. Thiamine transport by the *H. pylori* Pnu transporter has previously been proposed but has not been experimentally validated[34]. When expressed in *E. coli* containing the thiamine selection system, the Pnu transporters from *B. fragilis*, *P. multisaccharivorax* and *H. pylori* indeed allowed growth under selection conditions, demonstrating transport of thiamine. In contrast, PnuC from *S. pneumoniae* and *E. coli* PnuC did not enable growth (**Fig. 3b**). The functional evidence of thiamine transport by PnuT from *B. fragilis*, *P. multisaccharivorax* and *H. pylori* should provide new targets for drug development against these and related pathogens.

A generalizable transport selection platform

To demonstrate that the concept of synthetic selections to study microbial transport is generalizable to other compounds, we built a new selection system responsive to xanthine alkaloids and performed functional metagenomic selections to identify xanthine uptake transporters. To construct the xanthine alkaloids selection

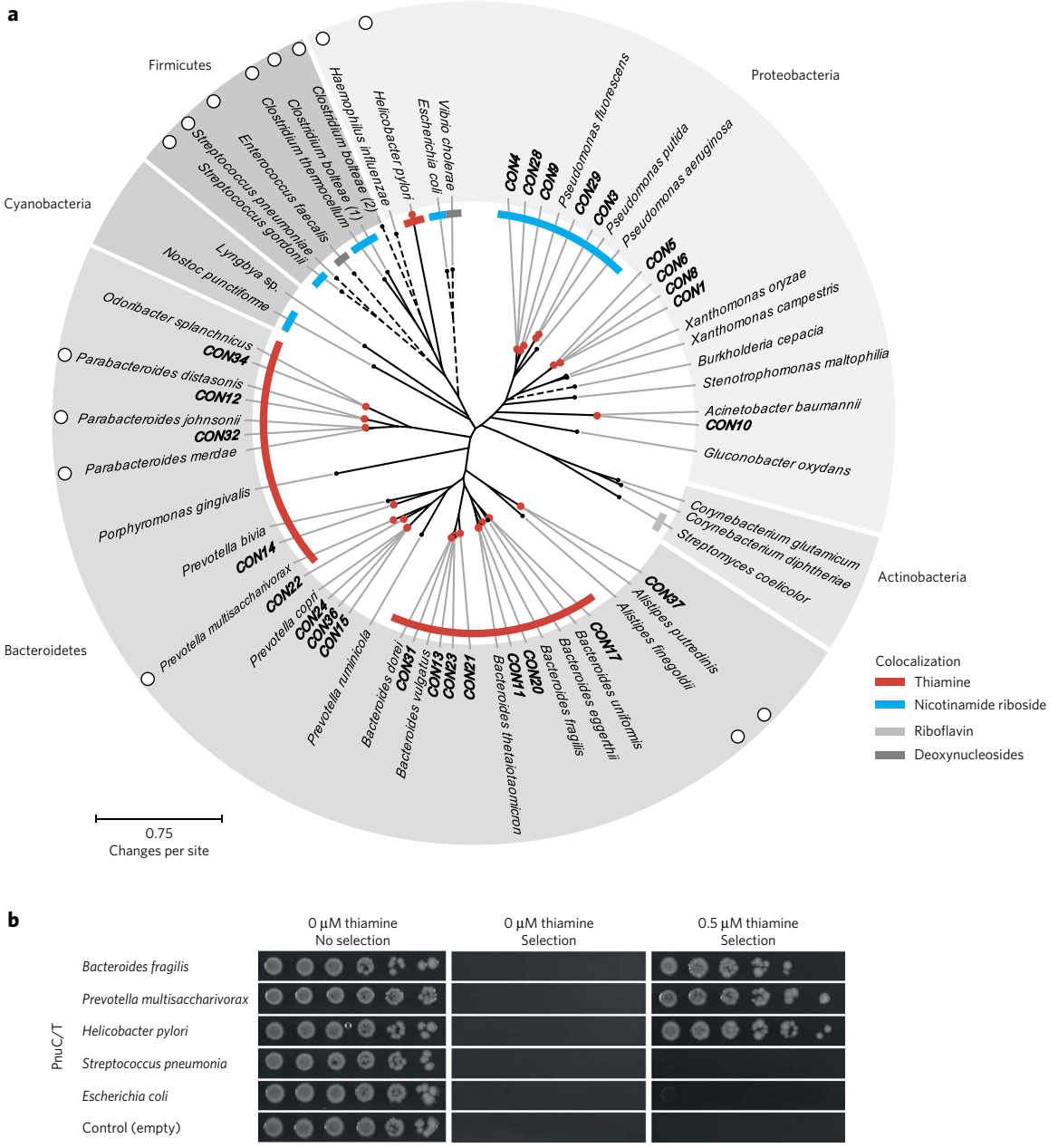

Figure 3 | Phylogenetic and functional relationship of Pnu transporters. (**a**) Phylogenetic relationships of Pnu transporters derived from soil and gut microbiomes (labeled *CON1*–*CON26*) and other selected Pnu transporters available from GenBank, displayed in an unrooted neighbor-joining tree. Colors in the inner circle denote genetic colocalization of Pnu transporters with genes involved in salvage or biosynthesis pathways of specified molecules. Dashed lines indicate the presence of homologs of previously characterized thiamine transporters in the genome; solid lines indicate that no known thiamine transporter is present in genome. Open circles at the outer edge of the tree denote an incomplete thiamine pathway, leading to dependency on exogenous thiamine uptake. Red dots indicate thiamine transporters that were functionally validated in this study. (**b**) Characterization of heterologous Pnu transporters from indicated species expressed in the *E. coli* selection strain. Each strain was serially diluted tenfold and spotted on the specified selective medium. Growth at low thiamine concentrations in selective medium indicated an active thiamine transporter. Shown are representative images of biological replicates ($n = 3$).

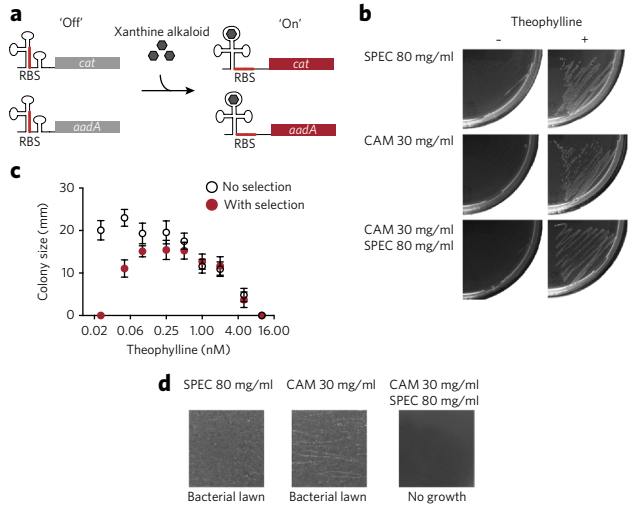

Figure 4 | Riboswitch-based xanthine alkaloid selection system. (**a**) The theophylline ribosome binding site (RBS) coupled to genes for chloramphenicol resistance (*cat*) and spectinomycin resistance (*aadA*). (**b**) Addition of theophylline induces resistance to chloramphenicol (CAM) and spectinomycin (SPEC) individually or in combination. (**c**) Mean diameter of colonies of the selection strain at various theophylline concentrations, with or without selection (30 μg/ml chloramphenicol and 80 μg/ml spectinomycin), after 22 h of growth at 37 °C. Error bars indicate ±s.d. of biological replicates (n = 3). (**d**) Plating of ~10^8 cells with single or dual selection in the absence of inducer ligand. Dual selection eliminated false positive growth at high cell densities. Images in **b** and **d** are representative of biological replicates (n = 3).

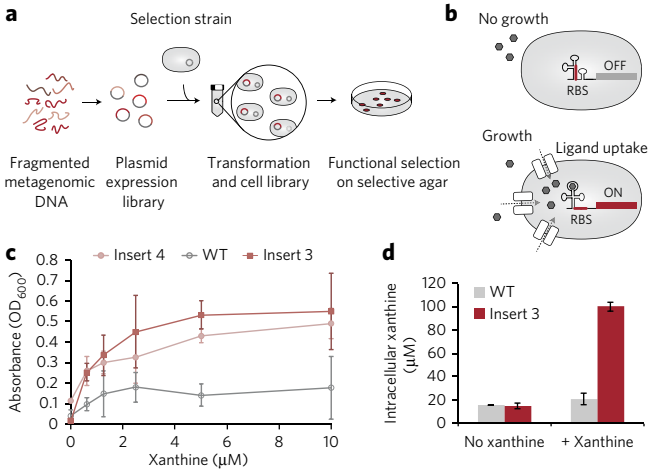

Figure 5 | Functional metagenomic selections identify xanthine uptake transporters. (**a**) Total DNA extracted from soil and gut fecal samples (metagenomic DNA) was fragmented into 1- to 3-kb fragments, cloned into an expression vector and transformed into an *E. coli* host strain harboring the xanthine alkaloid selection system. The cell library was plated on LB agar plates under selective conditions. (**b**) Cells expressing functional xanthine uptake transporters from the metagenomic DNA inserts import extracellular xanthine and have increased intracellular xanthine concentrations, leading to induction of ribosome binding site (RBS)-mediated antibiotic resistance. (**c**) Xanthine dose-response curves for strains expressing functionally selected xanthine transporters (insert 3 and insert 4) and the wild-type strain (WT; harboring the selection plasmid but with no xanthine permease overexpressed), grown under selective conditions. (**d**) Intracellular concentrations of xanthine measured by LC-MS at 0 μM and 400 μM exogenously added xanthine. Error bars in **c** and **d** represent ±s.d. of biological replicates (n = 3).

system, we replaced the TPP riboswitch of the TPP selection system with the synthetic theophylline riboswitch[40] (**Fig. 4a**). In contrast to the TPP riboswitch, which is based on a naturally evolved aptamer, the theophylline riboswitch is developed from the fully synthetic mTCT8-4 aptamer created by *in vitro* selection using the SELEX procedure[41]. When it was transformed into *E. coli*, the resulting strain displayed theophylline-dependent growth under selective conditions (**Fig. 4b**,**c**). As with the TPP selection system, dual selection (using chloramphenicol and spectinomycin in combination) allowed plating of up to 10^8 cells on an agar plate in the absence of inducer without any false positive colonies appearing in 3 d. In contrast, use of a single-selection protocol (chloramphenicol or spectinomycin alone) resulted in a bacterial lawn at high cell densities (**Fig. 4d**). In addition to theophylline, a number of related xanthine alkaloids are known to bind the mTCT8-4 aptamer *in vitro*: theophylline binds with an affinity of ~0.32 μM, whereas xanthine, which is an endogenous metabolite of *E. coli*, binds with an affinity of ~8.5 μM (ref. 41). We supplemented the medium with xanthine but did not observe induction of growth under selective conditions, suggesting that xanthine uptake may be limiting. Although *E. coli* does encode endogenous xanthine importers (*ygfO* and *yicE* (*xanP*)), it has been shown that deregulation is required for significant measurable increases in xanthine uptake[42].

We then explored the capability of the selection strain for mining xanthine uptake functions via functional metagenomic selections (**Fig. 5a**,**b**). Cell libraries consisting of the selection strain transformed with expression plasmids containing fragmented metagenomic inserts were plated on selective agar medium with trace amounts of xanthine. We isolated 42 colonies expressing four unique metagenomic DNA inserts (**Supplementary Table 5**). Interestingly, two isolated sequences (inserts 3 and 4) encoded full-length genes similar to bacterial genes annotated as xanthine permeases, with more than 99% sequence identity at the amino acid level to putative proteins from *Acinetobacter calcoaceticus* and *Prevotella copri*, respectively (**Supplementary Table 5**). Both permeases are predicted members of the nucleobase-ascorbate transporter (NAT/NCS2) family, which includes transporters for cellular uptake of a range of small molecules, including xanthine. The other two inserts encoded multidrug-resistance proteins and were therefore excluded from further analysis.

To test whether the observed growth response was indeed due to uptake of xanthine, we first performed xanthine dose-response experiments focused on the two isolated inserts encoding putative xanthine permeases. Under selective conditions, strains expressing the metagenomic inserts showed substantially improved growth when supplemented with xanthine at concentrations as low as 1 μM in minimal media (**Fig. 5c**). In contrast, no difference in growth was observed for the control strain (selection strain expressing an empty vector). To further analyze the xanthine uptake function, we measured intracellular xanthine pools using HPLC coupled to mass spectrometry (LC-MS). When we supplemented the growth medium with xanthine, we measured a 6.3-fold increase of intracellular xanthine levels (from 15.9 ± 2.3 μM to 100.4 ± 4.0 μM; average and s.d. of three biological replicates) in a strain expressing the putative xanthine permease (**Fig. 5d**). In contrast, in the wild-type strain, intracellular xanthine levels increased by only 33% (from 15.9 ± 0.2 μM to 21.1 ± 5.1 μM). Combined, these data confirm that the functionally selected metagenomic inserts encode xanthine permeases and demonstrate that synthetic biosensors can be implemented for genetic mining of small-molecule transport systems.

DISCUSSION

Our results suggest a powerful experimental strategy based on synthetic biology that enables functional identification of small-molecule transporters with high throughput. Tailoring the system

to new compounds of interest will require implementation of new substrate-specific biosensors to create synthetic selection hosts. Biosensor availability and implementability are therefore critical to the generality of the method. In nature, biosensing is a commonly used strategy to regulate gene expression in accordance with intracellular metabolite pools. To date, more than 17 small-molecule compound classes are known for riboswitches, and 243 compound classes have been identified for allosteric transcription factors[14]. The sensed metabolites include vitamins, amino acids and antibiotics. Because natural biosensors have evolved *in vivo*, they respond to concentrations in physiologically relevant ranges and display high specificity toward their target ligands. This makes them ideal candidates as sensor modules for the development of synthetic selection systems for cellular transporters.

For compounds for which no natural biosensor is known, biosensors can be created synthetically. Ligand-binding domains of riboswitches can be selected *in vitro* from large RNA pools, and selected aptamers can be integrated into expression platforms to enable conditional gene expression. This has led to fully synthetic riboswitches for a handful of compounds[15] including the xanthine alkaloid riboswitch used in this study. For protein-based biosensors, a considerable body of work has been devoted to altering the ligand specificity of existing transcription factors via protein engineering[43]. Although biosensor engineering has proven difficult, recent strategies that combine emerging rational design tools and high-throughput screening are likely to pave the way for more streamlined approaches for non-natural biosensor development[44,45]. Finally, a third strategy is to extend the metabolic network from undetectable compounds to molecules for which a biosensor is available by inserting the necessary enzymatic reaction steps. Considering the available enzymes and known biosensors, this strategy theoretically doubles the number of effector compounds[46].

In the present study we assayed metagenomic fragments 1–3 kb long. This could partly explain why we did not identify bacterial transport functions that rely on complexes of multiple proteins such as the ECF–ThiT transporter (common in the *Bacillus*/*Clostridium* group[33]) or the ThiBPQ thiamine transporter (common among proteobacteria[33]). First, these transport systems are typically encoded by genes with a combined length of more than 3 kb. Second, such protein complexes are not necessarily encoded near each other on the chromosome[33]. Other ancillary genes required for transport function must also be considered. For example, PnuT transporters facilitate diffusion of thiamine across the membrane and require thiamine kinases to drive this import. The *E. coli* DH10B strain used in this study encodes such a kinase (ThiK).

Overexpression of membrane proteins is known to have several adverse effects on the host[47], and it is likely that the promoter and the high copy number of the expression vector (20–50 copies) caused toxicity of certain metagenomic inserts. Indeed, overexpression of the synthesized *pnuC* or *pnuT* open reading frames using a strong ribosome binding site resulted in growth defects, but expression of the genes using a 5′ untranslated region from one of the *pnuT* sequences selected from the metagenome rescued cell growth (Online Methods). This underlines the importance of choosing an appropriate expression cassette. Optimal expression of proteins and especially membrane proteins is in itself an important issue[47], and synthetic selections could be an interesting approach to broadly investigate expression levels that are optimal for transport function.

Traditional transporter assays that rely on the use of radiolabeled compounds or advanced analytical instrumentation are indispensable tools for the determination of accurate biophysical properties such as rate kinetics and selectivity. Our method does not represent a replacement of such vital tools. However, the main shortcoming of traditional methods is the complexity of assays and the consequent low throughput of experiments. On this parameter, our method represents a drastic improvement that allows experimental investigation of transport in ultra-high throughput.

Continued application of the presented method should permit deep mining of transporters of entire microbial communities and allow exhaustive elucidation of transporters of individual organisms. In addition, performing functional selections of mutant libraries of individual transporters will allow researchers to evolve new substrate specificities and investigate sequence–function relationships. Ultimately, such applications will help close the gaps in our understanding of microbial small-molecule transport.

Received 26 January 2016; accepted 28 July 2016;
published online 3 October 2016

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The complete list of GenBank accession codes from this study can be found in **Supplementary Tables 3** and **4**.

References

1. Kell, D.B., Swainston, N., Pir, P. & Oliver, S.G. Membrane transporter engineering in industrial biotechnology and whole cell biocatalysis. *Trends Biotechnol.* **33**, 237–246 (2015).
2. Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406**, 775–781 (2000).
3. Herbert, M. *et al.* Nicotinamide ribosyl uptake mutants in Haemophilus influenzae. *Infect. Immun.* **71**, 5398–5401 (2003).
4. Degnan, P.H., Barry, N.A., Mok, K.C., Taga, M.E. & Goodman, A.L. Human gut microbes use multiple transporters to distinguish vitamin B_{12} analogs and compete in the gut. *Cell Host Microbe* **15**, 47–57 (2014).
5. Arai, M., Okumura, K., Satake, M. & Shimizu, T. Proteome-wide functional classification and identification of prokaryotic transmembrane proteins by transmembrane topology similarity comparison. *Protein Sci.* **13**, 2170–2183 (2004).
6. Ren, Q., Chen, K. & Paulsen, I.T. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.* **35**, D274–D279 (2007).
7. Ren, Q. & Paulsen, I.T. Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes. *J. Mol. Microbiol. Biotechnol.* **12**, 165–179 (2007).
8. Michener, J.K. & Smolke, C.D. High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metab. Eng.* **14**, 306–316 (2012).
9. van Sint Fiet, S., van Beilen, J.B. & Witholt, B. Selection of biocatalysts for chemical synthesis. *Proc. Natl. Acad. Sci. USA* **103**, 1693–1698 (2006).
10. Yang, J. *et al.* Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. *Nat. Commun.* **4**, 1413 (2013).
11. Raman, S., Rogers, J.K., Taylor, N.D. & Church, G.M. Evolution-guided optimization of biosynthetic pathways. *Proc. Natl. Acad. Sci. USA* **111**, 17803–17808 (2014).
12. Binder, S. *et al.* A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome Biol.* **13**, R40 (2012).
13. Siedler, S. *et al.* SoxR as a single-cell biosensor for NADPH-consuming enzymes in *Escherichia coli*. *ACS Synth. Biol.* **3**, 41–47 (2014).
14. Novichkov, P.S. *et al.* RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* **14**, 745 (2013).
15. Wittmann, A. & Suess, B. Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett.* **586**, 2076–2083 (2012).
16. Muranaka, N., Sharma, V., Nomura, Y. & Yokobayashi, Y. An efficient platform for genetic selection and screening of gene switches in *Escherichia coli*. *Nucleic Acids Res.* **37**, e39 (2009).
17. Winkler, W., Nahvi, A. & Breaker, R.R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).
18. Rodionov, D.A. *et al.* A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.* **191**, 42–51 (2009).
19. Webb, E., Claas, K. & Downs, D. thiBPQ encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphate in Salmonella typhimurium. *J. Biol. Chem.* **273**, 8946–8950 (1998).
20. Jeanguenin, L. *et al.* Comparative genomics and functional analysis of the NiaP family uncover nicotinate transporters from bacteria, plants, and mammals. *Funct. Integr. Genomics* **12**, 25–34 (2012).

21. Rodionova, I.A. *et al.* Genomic distribution of B-vitamin auxotrophy and uptake transporters in environmental bacteria from the Chloroflexi phylum. *Environ. Microbiol. Rep.* **7**, 204–210 (2015).
22. LeBlanc, J.G. *et al.* Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr. Opin. Biotechnol.* **24**, 160–168 (2013).
23. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
24. Aziz, R.K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
25. Sommer, M.O., Dantas, G. & Church, G.M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131 (2009).
26. Forsberg, K.J. *et al.* The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).
27. Sauer, E., Merdanovic, M., Mortimer, A.P., Bringmann, G. & Reidl, J. PnuC and the utilization of the nicotinamide riboside analog 3-aminopyridine in Haemophilus influenzae. *Antimicrob. Agents Chemother.* **48**, 4532–4541 (2004).
28. Jaehme, M., Guskov, A. & Slotboom, D.J. Crystal structure of the vitamin B3 transporter PnuC, a full-length SWEET homolog. *Nat. Struct. Mol. Biol.* **21**, 1013–1015 (2014).
29. Grose, J.H. *et al.* Assimilation of nicotinamide mononucleotide requires periplasmic AphA phosphatase in Salmonella enterica. *J. Bacteriol.* **187**, 4521–4530 (2005).
30. Vogl, C. *et al.* Characterization of riboflavin (vitamin B 2) transport proteins from *Bacillus subtilis* and *Corynebacterium glutamicum*. *J. Bacteriol.* **189**, 7367–7375 (2007).
31. Hemberger, S. *et al.* RibM from Streptomyces davawensis is a riboflavin/roseoflavin transporter and may be useful for the optimization of riboflavin production strains. *BMC Biotechnol.* **11**, 119 (2011).
32. Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. & Gelfand, M.S. Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *J. Biol. Chem.* **277**, 48949–48959 (2002).
33. Gelfand, M.S. & Rodionov, D.A. Comparative genomics and functional annotation of bacterial transporters. *Phys. Life Rev.* **5**, 22–49 (2008).
34. Jaehme, M. & Slotboom, D.J. Structure, function, evolution, and application of bacterial Pnu-type vitamin transporters. *Biol. Chem.* **396**, 955–966 (2015).
35. Abreu-Goodger, C. & Merino, E. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res.* **33**, W690–W692 (2005).
36. Schauer, K., Rodionov, D.A. & de Reuse, H. New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? *Trends Biochem. Sci.* **33**, 330–338 (2008).
37. Methé, B.A. *et al.* A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
38. Dehal, P.S. *et al.* MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.* **38**, D396–D400 (2010).
39. Nedenskov, P. Nutritional requirements for growth of *Helicobacter pylori*. *Appl. Environ. Microbiol.* **60**, 3450–3453 (1994).
40. Lynch, S.A. & Gallivan, J.P. A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res.* **37**, 184–192 (2009).
41. Jenison, R., Gill, S., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. *Science* **263**, 1425–1429 (1994).
42. Karatza, P. & Frillingos, S. Cloning and functional characterization of two bacterial members of the NAT/NCS2 family in *Escherichia coli*. *Mol. Membr. Biol.* **22**, 251–261 (2005).
43. Mahr, R. & Frunzke, J. Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol.* **100**, 79–90 (2016).
44. Taylor, N.D. *et al.* Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **13**, 177–183 (2016).
45. Goler, J.A., Carothers, J.M. & Keasling, J.D. Dual-selection for evolution of in vivo functional aptazymes as riboswitch parts. *Methods Mol. Biol.* **1111**, 221–235 (2014).
46. Libis, V., Delépine, B. & Faulon, J.-L. Expanding biosensing abilities through computer-aided design of metabolic pathways. *ACS Synth. Biol.* http://dx.doi.org/10.1021/acssynbio.5b00225 (2016).
47. Wagner, S. *et al.* Consequences of membrane protein overexpression in *Escherichia coli*. *Mol. Cell. Proteomics* **6**, 1527–1550 (2007).

Acknowledgments

We thank Y. Yokobayashi (University of California, Davis, California, USA) for providing plasmid pLacthiM19tetA-gfpuv, J. Gallivan (Emory University, Atlanta, Georgia, USA) for providing plasmid pSKD314, D. Paiva (Technical University of Denmark, Kongens Lyngby, Denmark) for metagenomic libraries, R. Lavallee for technical HPLC support, and C. Munck for critical reading of the manuscript. This study was funded by the Novo Nordisk Foundation and the European Union Seventh Framework Programme (FP7-KBBE-2013-7 single stage) under grant agreement no. 613745, Promys. H.J.G. acknowledges additional financial support from Novozymes A/S.

Author contributions

H.J.G. and M.O.A.S. conceived the study. H.J.G. and S.D.P. developed the thiamine functional selection system, and H.J.G. and A.P.B. performed functional metagenomic selections for thiamine uptake. H.J.G., M.K. and L.S.G. developed the HPLC assay for measurement of thiamines. H.J.G. and A.P.B. cloned heterologous genes. H.J.G. and M.T.B. developed the xanthine alkaloid selection system, and H.J.G. performed functional metagenomic selections and analysis. H.J.G. and S.S. performed dose-response characterizations of selected xanthine importers. S.J.H. developed the LC-MS method for xanthine alkaloids and performed measurements of samples prepared by S.S. and H.J.G. H.J.G. wrote the manuscript with contributions from all other authors.

Competing financial interests

The authors declare competing financial interests: details are available in the online version of the paper.

Additional information

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to M.O.A.S.

ONLINE METHODS

Materials and general considerations. Plasmid DNA and PCR products were purified using a QIAprep Spin Miniprep kit and QIAquick PCR purification kit, respectively (Qiagen). Gel extractions were performed using a GeneJET gel extraction kit (Fermentas). Thiamine–HCl, thiamine monophosphate–HCl, thiamine pyrophosphate–HCl, caffeine, theophylline, theobromine, xanthine and all antibiotics used in this study were purchased from Sigma-Aldrich. Xanthine stock solutions were prepared as 250 mM xanthine in 1 M NaOH. Synthetic oligonucleotides were purchased from Integrated DNA Technologies. *Escherichia coli* DH10B was used for all experiments. For all incubations in liquid media, cells were grown with shaking at 250 r.p.m. at 37 °C. Strains were stored at −80 °C in a 15% (vol/vol) glycerol solution. Plasmid manipulations were performed using USER cloning as previously described[48]. Sequences of all vectors constructed were verified by DNA sequencing (Beckman Coulter, UK). Cell cultures were performed in Luria broth (LB), M9 minimal media supplemented with 1% glucose, 0.1 g/l L-leucine, 0.1 g/l isoleucine, 0.1 g/l valine, and trace elements or modified rich MOPS medium without thiamine (mrMOPS) (**Supplementary Table 6**).

Construction of thiamine transport functional selection system. The thiamine transport selection system is based on a single plasmid encoding two copies of the $thiMN_{15}$#19 thiamine pyrophosphate (TPP) riboswitch[16] regulating the expression of a chloramphenicol acyl transferase gene (*cat*) and an aminoglycoside adenyltransferase gene (*aadA*) conferring resistance to chloramphenicol and spectinomycin, respectively (**Supplementary Fig. 1a**). The plasmid backbone encoding the dual-selection system contains an ampicillin marker gene (*ampR*) and a p15A origin of replication supporting 10–12 copies of the plasmid per cell. The p15A origin is compatible with the pZE21 cloning vector[49], which was used for metagenomic expression libraries. To avoid intraplasmid homologous recombination caused by the presence of two identical sequences (promoters and $thiMN_{15}$#19 riboswitch upstream of *cat* and *aadA*), the elements essential for plasmid propagation (*ampR* and the p15A origin) were positioned between the identical sequences.

Three factors were assumed to be important for the functional selection system: (i) The mRNA levels of *cat* and *aadA*; (ii) the 5′ sequences of *cat* and *aadA*, as the $thiMN_{15}$#19 riboswitch structure extends into the open reading frame in its native configuration[16]; and (iii) the direction of *ampR*, as transcriptional read-through can critically affect gene expression of one of the marker genes. We constructed a set of plasmids addressing the above-mentioned factors by varying all three parameters (i.e., promoter sequence, N-terminal leader sequence and *ampR* gene directionality). This was achieved by a combinatorial PCR-based USER cloning approach supported by computational design using the AMUSER web server[48]. Specifically, six fragments containing uracil at the 5′ ends for subsequent cloning were generated by PCR (**Supplementary Table 7**). Fragments 1 and 2 encoding *cat*, *aadA* and the p15A origin of replication were generated by primers oGEN66/oGEN65 and oGEN68/oGEN67, respectively. Fragment 1 contained a short leader sequence (L2) and fragment 2 contained a long leader sequence (L1), which were added during PCR via the primer tails. The *ampR* gene was amplified with primers encoding the *cat*/*aadA* promoters in the primer tails. Three fragments (3, 4 and 5) were generated by PCR using primer pairs oGEN73/oGEN73, oGEN71/oGEN71 and oGEN72/oGEN72, respectively (**Supplementary Fig. 1a**). The primers encode three promoters of different strengths (strong, BBa_23100; medium, BBa_23105; and weak, BBa_23117) and were furthermore designed such that mutations in the promoter regions could easily arise during PCR, resulting in even more promoter variations. Finally, fragment 6, which was generated using primers oGEN70/oGEN69, encoded the two $thiMN_{15}$#19 riboswitches. Uracil excision of each of the fragments generated overhangs facilitating seamless assembly of fragments 1 and 2 with 6 and of 3, 4 and 5 with 6, forming a circular plasmid of four fragments in total—for example, 1-6-3-6 or 2-6-5-6 (**Supplementary Fig. 1a**). The overhangs of fragments 1 and 2 allowed for assembly in both directions. Hence, mixing all six fragments offered 12 different possible constructs. With the possibility of promoter mutations in fragments 3, 4 and 5, multiple possible construct designs arise from a single-step assembly of all fragments. After assembly, the mix was transformed into *E. coli* DH10B cells. Sanger sequencing of 50 randomly picked transformants yielded 17 unique constructs of different composition (strains EcGEN31–47) (**Supplementary Fig. 1b**).

Phenotypic screening of thiamine selection constructs. The 17 strains, named EcGEN31–47, carrying unique dual-selection constructs were cultured in liquid M9 minimal media in microtiter plates (200 µl per well) for 21 h at 37 °C with shaking (300 r.p.m.). The media was supplemented with ampicillin (50 µg/ml) and various concentrations of chloramphenicol or spectinomycin in combination with either 0 or 100 µM thiamine. The concentrations of added antibiotics corresponded to 0, 1, 4, 8, 16, 100, 400 and 800 times the *E. coli* MG1655 minimum inhibitory concentrations (MIC) of chloramphenicol (4 µg/ml) and spectinomycin (16 µg/ml)[50]. The growth of strains EcGEN31–47 was measured by optical density at 630 nm (OD_{630}) using an Elx808 plate reader (BioTek) after 13 and 21 h of incubation (**Supplementary Fig. 1b**). The experiments were performed in biological quadruplicate. Overnight cultures were inoculated from fresh colonies at a dilution resulting in a starting OD_{630} of 0.015 for all experiments. We found promoter strength and leader sequence to be critical parameters for a functional assay, and only a subset of the constructs resulted in a thiamine-dependent phenotype in a practical range of antibiotics (**Supplementary Fig. 1b**).

Optimizing growth conditions for functional selections. Growth selections on agar plates were chosen over liquid enrichment selections because agar-plate-based selection allows for direct separation and isolation of single clones, which simplifies downstream analysis and genotyping. In addition, because liquid-based selections enrich phenotypes with high fitness, there is a risk that a few selected clones will quickly dominate the population and consequently dilute other selected clones with lower fitness. Together, these factors favor agar-based functional selection for applications that involve gene mining.

In addition to optimizing the genetic design of the selection plasmid, we fine-tuned the growth conditions for agar-based selections by adjusting the composition of the growth medium and testing the level of selection pressure (i.e., concentrations of chloramphenicol and spectinomycin). Because complex medium such as LB medium contains trace amounts of thiamine, which would affect the selection, we used a modified rich MOPS medium (mrMOPS) that did not contain any thiamine (**Supplementary Table 6**) and which produced substantially more reproducible growth profiles than, for example, M9 minimal medium.

An ideal selection platform allows growth of all cells in the presence of inducer, with practically no growth in the absence of inducer. To determine the appropriate levels of chloramphenicol and spectinomycin in the growth medium, we spotted 10× serial dilutions of a saturated overnight culture of the selection strain (EcGEN46; **Supplementary Fig. 1b**) on fresh, pre-dried mrMOPS agar plates supplemented with 50 µg/ml ampicillin and a matrix of concentrations of spectinomycin (ranging from 0 to 500 µg/ml) and chloramphenicol (ranging from 0 to 500 µg/ml). At 50 µg/ml spectinomycin and 14 µg/ml chloramphenicol, we observed severely reduced growth at high cell densities in the absence of supplemented thiamine, whereas the presence of 100 µM thiamine fully rescued growth for all cells after 40 h of growth. At this concentration we further characterized the emergence of false positives in the absence of added thiamine. A single colony of the selection strain was inoculated and grown overnight to saturation in mrMOPS medium. Volumes ranging from 1 to 100 µl were plated on selective agar growth medium in standard Petri dishes (94-mm diameter) corresponding to 10^6–10^8 cells in total or ~1.4×10^4 to 1.4×10^6 cells per square centimeter (**Supplementary Fig. 2**) and incubated at 37 °C. The plates were freshly prepared (within less than 2 d of plating) and pre-dried (1–2 h in a LAF bench) immediately before plating of the culture. When we plated many cells (~0.7×10^8 or 50 µl of an overnight culture) on agar medium containing either only chloramphenicol or spectinomycin alone, we observed high rates of false positives after just 1 d of incubation (**Fig. 1**, **Supplementary Fig. 2**). However, the combination of the two antibiotics resulted in clean plates for at least 3 d. The results were consistent over multiple biological replicates ($n = 5$). Consequently, we can conclude that the dual-selection setup enables functional selection of single cells from entire populations with an escape rate of less than 1 in 0.5×10^8. In a study conducted in parallel to this one, Raman *et al.*[11] demonstrated a

comparable effect with respect to reducing unintended growth with a similar dual-selection setup using a transcription factor as biosensor.

Thiamine dose-response curves. To characterize the effect of extracellular thiamine on growth of the selection strain (EcGEN46) under selective conditions, we inoculated a fresh colony into mrMOPS containing ampicillin (50 μg/ml) and grew the culture to mid-exponential phase. Using this, we inoculated a new culture to an initial OD_{630} of 0.015 in 200 μl mrMOPS containing ampicillin (50 μg/ml) and chloramphenicol (14 μg/ml) and supplemented thiamine at concentrations ranging from 0 to 100 μM. The culture was incubated for 21 h at 37 °C with shaking (250 r.p.m.) and OD_{630} was measured continuously (every 5 min) using an Elx808 plate reader (BioTek). The experiments were performed in biological triplicate.

Measurements of intracellular thiamine moieties by HPLC. To measure intracellular thiamine, TMP and TPP, we inoculated a fresh colony into mrMOPS containing ampicillin (50 μg/ml) and allowed it to grow to mid-exponential phase. From this, a new culture was inoculated at an initial OD_{630} of 0.015 into 10 ml of mrMOPS media containing ampicillin (50 μg/ml) and thiamine at concentrations of 0, 0.01, 0.1, 1, 10, and 100 μM, respectively. After 22 h of incubation at 37 °C with shaking (250 r.p.m.), we measured OD_{630}, collected cells by centrifugation and washed them twice in ice-cold PBS. We extracted intracellular metabolites by suspending the cell pellet in 1 ml of ice-cold methanol. We collected the supernatant and further pre-concentrated the extracted metabolites using a SpeedVac (Svant SC210A, Thermo Scientific) at room temperature. The resulting extracts were resuspended in 100 μl of water. We then measured individual thiamine compounds using a modified thiochrome HPLC method as described previously[51]. Briefly, thiamine, TMP and TPP were derivatized to fluorescent thiochromes via the following procedure: 50 μl of cell extracts were added to 100 μl of 4 M potassium acetate and mixed by pipetting. Then, 50 μl of freshly prepared 3.8 M potassium ferricyanide in 7 M NaOH was added. The solution was mixed by pipetting and finally quenched by the addition of 50 μl of fresh 0.06% H_2O_2 in saturated KH_2PO_4. The derivatization mix was stable at 4 °C for up to 3 months[52]. Samples were injected into a Hypersil GOLD column (15 cm × 2.1 mm, 3 μm) (Thermo reference no. 25003-152130) and eluted with a gradient of 10–30% methanol (H_2O, 50% to 30%) and 40% 4 mM tetrabutyl ammonium hydrogen sulfate in 10 mM K_2HPO_4 (pH 7.0). The flow rate was 0.8 ml/min, and the column temperature was held at 35 °C. Fluorescence emission was measured at 444 nm after excitation at 365 nm. Concentrations of thiamine, TMP and TPP were estimated on the basis of comparison to standard curves. For calculations of intracellular concentrations of thiamine, we assumed an OD-specific total cell volume of 3.4 μL per ml of culture per OD_{600} (ref. 53).

Functional metagenomic selections for thiamine transporters. Metagenomic expression libraries (**Supplementary Table 1**) were constructed as described previously[25]. Briefly, the procedure involves (i) isolation of total DNA from soil or fecal matter, (ii) fragmentation of extracted DNA by sonication into pieces of 1–3 kb and (iii) blunt-end cloning into an expression plasmid. In this study the cloning vector pZE21 (ref. 49) was used for all libraries. Libraries AB95D01, 57SDb01 and 57SDb03 were obtained from previously published studies, and 'GranjaWorker' and 'GranjaPig' were prepared from 5 g of fecal matter obtained from a human and a pig, respectively, as described previously[25].

All plasmid libraries were harbored and amplified in *E. coli* TOP10 cells (Thermo Fisher). From each library culture, plasmids were prepped from a 5-mL overnight culture, and 400 ng of purified plasmid DNA of each library was transformed into 50 μl of electrocompetent *E. coli* EcGEN46 (thiamine selection strain) by electroporation using standard protocols for 1-mm electroporation cuvettes. Cells were recovered in 1 ml of SOC medium for 1 h at 37 °C. All libraries resulted in >1 × 10^6 c.f.u., estimated by plating of 1 μl of recovery cultures on LB agar media containing kanamycin (50 μg/ml) and ampicillin (50 μg/ml). The remaining recovered cells were washed twice in mrMOPS to remove any thiamine from the SOC media and inoculated into 10 mL of mrMOPS containing kanamycin (50 μg/mL) and ampicillin (50 μg/mL) and grown overnight. Overnight cultures were plated in volumes of 10, 25 and 50 μl, corresponding to a total of 0.5 × 10^8 –1 × 10^8 c.f.u., on fresh, pre-dried mrMOPS agar plates containing kanamycin (50 μg/mL), chloramphenicol (14 μg/mL), spectinomycin (50 μg/mL) supplemented with or without 0.5 μM thiamine. As a negative control, an overnight culture of the transport selection strain (EcGEN46) carrying the pZE21 vector (no metagenomic insert) was plated in parallel. The plates were incubated at 37 °C for 48 h. For all metagenomic libraries, several colonies appeared on media supplemented with 0.5 μM thiamine, whereas plates with no thiamine were empty except for one colony that appeared from the soil library (AB95D01). Not a single colony appeared for the negative control strain (no metagenomic insert), demonstrating that 0.5 μM extracellular thiamine is not sufficient to increase the intracellular levels of TPP of the background strain to levels that activate the riboswitch selection system.

Sequencing and analysis of metagenomic inserts. We picked 120 clones (24 from each library) functionally selected from agar plates with 0.5 μM thiamine and streaked them on mrMOPS selective plates containing kanamycin (50 μg/mL), chloramphenicol (14 μg/mL), spectinomycin (50 μg/mL), and 0.5 μM thiamine to ensure nonheterogeneous clones. From these, single colonies were picked and inoculated into liquid mrMOPS medium containing kanamycin (50 μg/mL) and ampicillin (50 μg/mL) and grown overnight. Selected metagenomic inserts harbored by these clones were sequenced using Sanger sequencing (Beckman Coulter, UK). Primers GEBO29 and oGEN44 were used for bidirectional sequencing, and gaps were closed using primer walking (**Supplementary Table 7**). Quality trimming and assembly of reads were performed using the deFUME web server[54] and manually curated using CLC Main Workbench 7 (Qiagen). A multiple alignment of the nucleotide sequences of the 120 inserts using CLC Main Workbench showed that several inserts were 100% identical and that 55 inserts were unique. Multiple identical inserts had arisen, as the cell libraries were grown overnight, allowing unique clones to be amplified in the population. The multiple sequence alignment further showed that several of the 55 unique inserts were 100% identical over stretches of sequence regions of >500 bp, indicating a common genomic origin (**Supplementary Fig. 3**). The common area shared by multiple overlapping inserts is in principle similar to a truncation analysis and points to which sequence region is required for functional selection.

Experimental validation of functionally selected thiamine transporters. To determine whether functionally selected metagenomic inserts encoded thiamine uptake systems, we first assayed the isolated clones on selective growth agar medium in the presence or absence of thiamine. Three single colonies of each functionally selected and restreaked clone were picked and inoculated into mrMOPS medium containing ampicillin (50 μg/mL) and kanamycin (50 μg/mL) and grown overnight. 5 μL of 20× serial dilutions of the overnight cultures were then spotted on mrMOPS agar containing kanamycin (50 μg/mL), chloramphenicol (14 μg/mL), and spectinomycin (50 μg/mL) supplemented with or without 0.5 μM thiamine, and growth was monitored after 18, 25 and 42 h of incubation at 37 °C. To test the dose response to extracellular thiamine, we analyzed two clones harboring functionally selected metagenomic inserts encoding PnuT transporters (CON31 and CON34) according to the procedure described above ("Thiamine dose-response curves"). To validate that growth was indeed a result of increased intracellular levels of TPP, we measured intracellular thiamine, TMP and TPP after the extraction and HPLC procedure described above ("Measurements of intracellular thiamines by HPLC"). These experiments were performed in biological triplicate.

Sequence alignment and phylogenetic analysis. CLC Main Workbench (CLC Bio) was used for all sequence alignments and phylogenetic tree constructions. To investigate the phylogenetic relationships of the identified *pnuT* genes (**Fig. 2c**), we aligned the nucleotide sequences of the *pnuT* open reading frames (gap open cost 10.0, gap extension cost 1.0, end gap cost 'cheap') and computed an unrooted phylogenetic tree from the alignment using the neighbor-joining algorithm[55] with bootstrap analysis (1,000 replicates). Phylogenetic distribution of functionally selected transporters (**Fig. 3a**) was assessed by multiple alignment of the amino acid sequences of all identified PnuT proteins including closely and distantly related Pnu transporters from a number of full-genome-sequenced bacteria available from GenBank. An unrooted

doi:10.1038/nchembio.2189

phylogenetic tree was computed from the alignment using the neighbor-joining algorithm[55] with bootstrap analysis (1,000 replicates). Sequence identities were calculated by pairwise or multiple sequence alignment. For genes encoding Pnu transporters present in full-genome-sequenced bacteria, colocalization to thiamine, nicotinamide riboside, riboflavin or deoxynucleotide salvage or biosynthesis genes was assessed using the 'TreeBrowser' feature available from MicrobesOnline[38].

Comparative genomics of thiamine biosynthesis, salvage, and transport. Only few bacterial thiamine transporters have been experimentally validated; these include ThiT (YuaJ) from *Lactococcus lactis*[18] (ThiT is the thiamine-binding integral membrane protein of an ECF-type ABC transporter) and the ABC transporters ThiBPQ from *Escherichia coli* and *Salmonella typhimurium*[19]. Additionally, bioinformatic predictions based on colocalization with thiamine biosynthesis genes and thiamine regulatory elements have led to predictions of a number of hypothetical thiamine transporters, of which two have recently been experimentally validated: NiaP from *Thermus thermophilus*[20] and ThiXYZ from *Chloroflexus aurantiacus*[21]. ThiB and ThiT are widely distributed among bacteria, and homologs are generally assumed to facilitate transport of thiamine. ThiXYZ is proposed to primarily transport various forms of pyrimidines including the thiamine precursor hydroxymethylpyrimidine (HMP). NiaP is a member of the large major facility transporter superfamily and is typically known to transport nicotinate[56]. The wide span of substrates transported by the same families of transporters make homology-based predictions of substrate specificity using these few queries nearly impossible.

To investigate the prevalence of genes encoding thiamine transport, biosynthesis and salvage, we used the "Phylogenetic profiling" feature at MicrobesOnline[38]. This feature allows gene searches against 1,752 complete bacterial genomes based on predicted functional annotations (TIGRFAM, COG and more) as well as searches based on similarity (BLAST). **Supplementary Table 9** lists all genes searched for and the search parameters used. For most genes we found the TIGRFAM functional assignment to be the most accurate search parameter. When searching for three-component transport systems (for example, ABC transporters) we used only the substrate-binding domain. Searches for orthologs of the thiamine transporters NiaP from *Thermus thermophilus*[20] and ThiXYZ from *Chloroflexus aurantiacus*[21] were based on similarity by BLAST (using a ≥200 bit score cutoff) as the COG or TIGRFAM annotations of these describe the entire transporter family, most of which are not thiamine transporters. The results of the search, performed on 2 February 2015, are provided in **Supplementary Table 11**.

In bacteria, *de novo* synthesis of thiamine pyrophosphate (TPP) starts with the formation of hydroxyethylthiazole phosphate (HET-P) by ThiS, ThiG, ThiF, ThiH/ThiO and hydroxymethylpyrimidine phosphate (HMP-P) by ThiC. Next HMP-P is phosphorylated to HMP-PP by ThiD and joined with HET-P by ThiE, forming thiamine monophosphate, which is phosphorylated to TPP (**Supplementary Fig. 4**). Some bacteria rely on salvage from HMP and HET, which requires ThiD, ThiE and ThiM. Bacteria with incomplete biosynthetic pathways are auxotrophs and rely on either salvage of HMP and/or HET or direct uptake of thiamine. Bacteria lacking the salvage enzymes depend solely on thiamine uptake. We used absence (no hit in genome according to our search parameters) of the committing enzyme ThiC as an indicator for the lack of *de novo* synthesis and absence of ThiE as an indicator for lack of salvage capabilities. By doing so, we probably underestimated the true number of thiamine auxotrophs, as genomes that encode, for example, ThiE might not encode ThiM or ThiD required for HET and HMP phosphorylation, respectively, and consequently would not encode all genes necessary for salvage.

Thiamine transporters among full-genome-sequenced human gut microbiome species. Protein sequences of experimentally validated thiamine transporters were compiled from *Thermus thermophilus* (NiaP)[20], *Chloroflexus aurantiacus* (ThiY)[21], *Lactococcus lactis* (ThiT)[18], *Escherichia coli*, *Salmonella typhimurium* (ThiB)[19], *Bacteroides fragilis* and *Helicobater pylori* (PnuT) and used as Blastp queries against the 383 bacterial complete reference genomes of the human gut microbiome[37] at http://hmpdacc.org/. Presence of a transporter was determined based on an *E*-value cutoff of 1e–20.

Cloning of specific Pnu transporters from known organisms. The open reading frames for *pnuC* and *pnuT* genes of *Bacteroides fragilis* NCTC 9343, *Streptococcus pneumoniae* CDC0288-04, *Prevotella multisaccharivorax* PPPA20 and *Helicobacter pylori* 26695 were synthesized without codon changes and delivered as linear DNA fragments (Invitrogen). The *pnuC* gene from *E. coli* DH10B was amplified from genomic DNA and included as a control. First, all genes were inserted into the multiple cloning site of pZE21 by Gibson assembly[57] using the Gibson Assembly Master Mix (New England BioLabs) according to the manufacturer's instructions. Primers for Gibson assembly were designed using the j5 software[58]. The ligation mixtures were transformed into 50 µl of electrocompetent *E. coli* EcGEN46 (thiamine transport selection strain) by standard electroporation procedures, and transformants were plated on LB agar medium containing ampicillin (50 µg/mL) and kanamycin (50 µg/mL). Sequencing of 8 clones of each construct showed that 38 of 40 clones contained missense mutations in the *pnuC* or *pnuT* gene and that only two constructs were successful. The clones encoding correct constructs exhibited severely impaired growth. Together these observations suggest that expression levels of the transporters were too high and caused growth defects. In order to use a more appropriate expression level, we changed the 5′ untranslated region (5′ UTR) from 'GAATTCATTAAAGAGGAGAAAGGTACCATG' to 'GAATTCATTAAAGAGGAGAAAGGTACCGGGCCCCCCCCTCGAGGTCCAGGCGATGTTCAACGTGCTGGCCAACGTAACGGTGCGGTTTTAGGCGATG'. The latter 5′ UTR was obtained from one of the functionally selected metagenomic inserts (CON37_insert58_5_UTR). We repeated the cloning by Gibson assembly and transformed the ligation mix into the selection strain (as described above), and we verified successful clones of all constructs by Sanger sequencing using primers GEBO29 and oGEN44 (**Supplementary Table 7**).

Characterization of specific Pnu transporters from known organisms. For characterization of the synthesized and cloned Pnu transporters' ability to transport thiamine when expressed in *E. coli* DH10B, overnight cultures in mrMOPS media containing ampicillin (50 µg/mL) and kanamycin (50 µg/mL) inoculated from single colonies of sequence-verified clones were serially diluted (10×) and 5 µL of each dilution was spotted on mrMOPS media containing kanamycin (50 µg/mL), chloramphenicol (20 µg/mL) and spectinomycin (50 µg/mL) and various concentrations of thiamine (0.000, 0.005, 0.050, 0.500, and 5.000 µM). As a control, the cells were additionally spotted on mrMOPS containing kanamycin (50 µg/mL) and no thiamine. Growth was monitored after 1 and 2 d of incubation at 37 °C. The experiments were performed in biological triplicate.

Construction of the xanthine alkaloid sensor strain. The xanthine alkaloid dual-selection plasmid pGEN7 was constructed based on pGEN28 from strain EcGEN34 (**Supplementary Table 8**). The two TPP riboswitches were replaced with two theophylline riboswitches (variant 12.1 (ref. 40)) by PCR-based USER cloning as described previously[48]. Following assembly, the mix was transformed into chemically competent *E. coli* DH10B cells by standard procedures, yielding the xanthine-alkaloid-responsive strain EcGEN94. Constructed plasmids were verified by Sanger sequencing (Beckman Coulter, UK).

Characterization of colony formation of xanthine alkaloid selection system. Plate-based phenotypic characterization of the xanthine-alkaloid-responsive strain (EcGEN94) was performed on freshly prepared LB agar plates containing ampicillin (50 µg/mL), chloramphenicol (30 µg/mL), and spectinomycin (80 µg/mL), supplemented with or without theophylline or xanthine at varying concentrations. Single colonies of the strain were picked and inoculated in liquid LB medium containing ampicillin (50 µg/mL) and grown to an OD_{600} of 0.3–0.5, corresponding to ~1.5×10^6 to 3.5×10^6 cells per 100 µL. Subsequently an aliquot of the culture was serially diluted with LB medium to a final concentration of ~300 cells per 100 µL. 100 µL from this solution was plated on each plate using glass beads. All plates were incubated in darkness at 37 °C for up to 72 h, and colony size was measured by imaging and analysis using Colony DocIt (UVP). To test the efficiency of the dual selection for reducing the emergence of false positives, we plated up to 1.3×10^8 EcGEN94 cells in biological triplicates on LB agar media containing spectinomycin (80 µg/mL) and

chloramphenicol (30 μg/mL) (dual selection). As a control, plating on medium containing only chloramphenicol (30 μg/mL) or spectinomycin (80 μg/mL) ('single selection') was performed in biological triplicate for each antibiotic. Each plate was incubated for 72 h.

Functional metagenomic selections for xanthine uptake transporters. 400 ng of plasmid DNA from metagenomic expression libraries AB95D01, 57SDb01, and 57SDb03 (**Supplementary Table 1**) was transformed into 50 μl of electrocompetent EcGEN94 as described in "Functional metagenomic selections for thiamine transporters." All library transformations produced >1 × 10^6 transformants as estimated by plating of 1 μl of recovery cultures on LB agar media containing kanamycin (50 μg/mL) and ampicillin (50 μg/mL). The remaining recovered cells were inoculated into 9 mL of LB media containing kanamycin (50 μg/mL) and ampicillin (50 μg/mL) and grown overnight. From each library, volumes of 50 and 100 μL of an overnight culture corresponding to 0.5 × 10^8 to 1 × 10^8 cells were plated on fresh, pre-dried LB agar plates containing kanamycin (50 μg/mL), chloramphenicol (30 μg/mL), and spectinomycin (80 μg/mL). As a negative control, an overnight culture of the selection strain carrying the pZE21 vector (no metagenomic insert) (EcGEN1) was plated in parallel. The plates were incubated at 37 °C for up to 69 h.

Sequencing and analysis of xanthine metagenomic inserts. Clones functionally selected from agar plates were picked and streaked on LB agar selective plates containing kanamycin (50 μg/mL), chloramphenicol (30 μg/mL), and spectinomycin (80 μg/mL) to ensure nonheterogeneous clones. Antibiotic-resistant colonies were picked and inoculated into liquid LB medium containing kanamycin (50 μg/mL) and ampicillin (50 μg/mL) and grown overnight. Selected metagenomic inserts harbored by clones identified from the functional selections were then sequenced and processed as described for the thiamine based selections. The sequences of insert 3 and insert 4 encoding novel xanthine permeases were deposited at GenBank (**Supplementary Tables 3** and **4**).

Xanthine dose-response curves. For each strain (EcGEN1, EcGEN71 and EcGEN75), three individual colonies were picked and inoculated in liquid M9 medium containing ampicillin (50 μg/mL) and kanamycin (50 μg/mL) and grown to mid-exponential phase. From these, new cultures were inoculated at an initial OD_{630} of 0.015 into 200 μl of M9 media in a microtiter plate containing kanamycin (50 μg/mL), ampicillin (50 μg/mL) and xanthine at concentrations of 0, 0.625, 1.25, 2.5, 5, 10, and 1,000 μM with or without chloramphenicol (20 μg/mL). The cultures were grown in an ELx808 Absorbance Microplate Reader from BioTek, and OD_{630} was measured every 5 min.

Extraction of intracellular xanthine. For each strain (EcGEN1 and EcGEN75), three individual colonies were picked and inoculated in M9 containing ampicillin (50 μg/mL) and kanamycin (50 μg/mL) and grown to mid-exponential phase. From these, new cultures were inoculated in shake flasks at an initial OD_{600} of 0.05 into 20 mL of M9 media containing kanamycin (50 μg/mL), ampicillin (50 μg/mL) and 0 μM or 400 μM xanthine, respectively. 2-mL samples for the determination of intracellular concentrations were taken at OD_{600} ~ 1. After 5 min of centrifugation at 0 °C the supernatant was discarded and the cell pellet was washed in 1 mL of ice-cold M9 medium without xanthine. The cell pellet was resuspended in 100 μL of water and stored at −80 °C. The cells were thawed on ice and sonicated at room temperature for 10 min in a Transsonic TI-H-15 water bath (Elma Schmidbauer, Germany) at 45 kHz. Cell debris was centrifuged for 10 min at 0 °C and the supernatant was collected and used for LC-MS analysis.

LC-MS analysis. LC-MS measurements were carried out on a Dionex UltiMate 3000 UHPLC (Thermo Fisher Scientific, San Jose, CA) connected to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Reagents used for LC-MS were all of analytical grade. For UHPLC we used a Discovery HS F5-5, 15 cm × 4.6 mm, 5-μm column. The temperature was 35 °C and the flow rate was 1 mL/min with a mobile phase of 100% formic acid (0.1%) for 1 min followed by a linear gradient of 100% formic acid (0.1%)/0% acetonitrile (0.1%) to 5% formic acid (0.1%)/95% acetonitrile (0.1%) over 5 min. This gradient was held for 1 min, after which it was changed immediately to 100% formic acid (0.1%) and 0% acetonitrile (0.1%) and held for 3 min. The sample was passed on to the MS equipped with a heated electrospray ionization source (HESI) in positive-ion mode with nitrogen as the nebulizer gas (60 AU). The cone and probe temperature were 325 °C and 380 °C, respectively. Probe gas flow was 20 AU and spray voltage was 3,500 V. Time between scans was 100 ms. A standard curve was prepared using 1, 2.5, 5, 10, 15 and 25 μM xanthine dissolved in water.

48. Genee, H.J. *et al.* Software-supported USER cloning strategies for site-directed mutagenesis and DNA assembly. *ACS Synth. Biol.* **4**, 342–349 (2015).
49. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
50. Chaudhuri, R.R. *et al.* Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. *PLoS One* **5**, e8801 (2010).
51. Schyns, G. *et al.* Isolation and characterization of new thiamine-deregulated mutants of *Bacillus subtilis* **187**, 8127–8136 (2005).
52. Bontemps, J. *et al.* Determination of thiamine and thiamine phosphates in excitable tissues as thiochrome derivatives by reversed-phase high-performance liquid chromatography on octadecyl silica. *J. Chromatogr.* **307**, 283–294 (1984).
53. Volkmer, B. & Heinemann, M. Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLoS One* **6**, e23126 (2011).
54. van der Helm, E., Geertz-Hansen, H.M., Genee, H.J., Malla, S. & Sommer, M.O.A. deFUME: dynamic exploration of functional metagenomic sequencing data. *BMC Res. Notes* **8**, 328 (2015).
55. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
56. Jaehme, M. & Slotboom, D.J. Diversity of membrane transport proteins for vitamins in bacteria and archaea. *Biochim. Biophys. Acta* **1850**, 565–576 (2015).
57. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
58. Hillson, N.J., Rosengarten, R.D. & Keasling, J.D. j5 DNA assembly design automation software. *ACS Synth. Biol.* **1**, 14–21 (2012).

doi:10.1038/nchembio.2189