Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries

Stefan M. Gaida1,*, Nicholas R. Sandoval1,*, Sergios A. Nicolaou1, Yili Chen1, Keerthi P. Venkataramanan1 & Eleftherios T. Papoutsakis1

A key limitation in using heterologous genomic or metagenomic libraries in functional genomics and genome engineering is the low expression of heterologous genes in screening hosts, such as *Escherichia coli*. To overcome this limitation, here we generate *E. coli* strains capable of recognizing heterologous promoters by expressing heterologous sigma factors. Among seven sigma factors tested, RpoD from *Lactobacillus plantarum* (*Lpl*) appears to be able of initiating transcription from all sources of DNA. Using the promoter GFP-trap concept, we successfully screen several heterologous and metagenomic DNA libraries, thus enlarging the genomic space that can be functionally sampled in *E. coli*. For an application, we show that screening fosmid-based *Lpl* genomic libraries in an *E. coli* strain with a chromosomally integrated *Lpl rpoD* enables the identification of *Lpl* genetic determinants imparting strong ethanol tolerance in *E. coli*. Transcriptome analysis confirms increased expression of heterologous genes in the engineered strain.

**References**

1 Department of Chemical and Biomolecular Engineering, Molecular Biotechnology Laboratory, Delaware Biotechnology Institute, University of Delaware, 15 Innovation Way, Newark, Delaware 19711, USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to E.T.P. (email: epaps@udel.edu).
It is widely recognized\(^1\) that the microbial diversity present in nature\(^2\)\(^-\)\(^3\), coupled with fast evolution rates\(^4\), results in an enormous stock of genetic material that remains largely unexplored for either fundamental studies or biotechnological applications\(^5\). With fewer than 1% of the organisms having being successfully cultured in the laboratory so far\(^6\), this enormous genetic diversity can be harvested best in the form of metagenomic libraries\(^1\)\(^-\)\(^3\). Effective functional screening of metagenomic libraries can unlock the hidden potential of the genetic diversity in nature and lead to the identification of novel or potent enzymatic activities as well as cellular programs, which could be used to engineer superior strains for biotechnological applications\(^7\). Effective screening of metagenomic libraries is grossly limited by the ability of the organism that is hosting the metagenomic library to express genes from anonymous organisms represented in the library\(^8\). This is largely attributed to the inability of the transcriptional machinery of the host organism to recognize promoters from the metagenome and possibly also translate metagenomic transcripts\(^9\). A similar situation presents expansion of the genomic space that can be explored in the context of genome engineering\(^8\)\(^-\)\(^10\), whereby, so far, the genomic diversity that can be generated and screened for developing useful strains is confined to mutational perturbations of single genomes\(^11\)\(^-\)\(^12\). As a result, novel genes, genetic programs or pathways outside single genomes are largely not accessible.

A larger functional sample space (defined as the fraction of expressed genes from a given genomic space) is a prerequisite for successful activity-based or trait-based screening. Large-insert libraries, such as fosmid-based and bacterial artificial chromosome-based genomic libraries, maximize the genomic space\(^14\), but only if those heterologous genes are expressed in the host. Due to its well-developed genetic toolbox, *Escherichia coli* is the preferred host for screening large-insert genomic libraries. However, expression of heterologous DNA in *E. coli* is limited\(^15\)\(^,\)\(^16\) and depends mostly on the recognition of the foreign promoter by the sigma factor subunits of the RNA-polymerase (RNAP) of *E. coli*\(^11\)\(^,\)\(^16\)\(^,\)\(^17\). Thus, enabling *E. coli* to recognize a larger fraction of heterologous promoters would increase the functional sample space and enable efficient screening of heterologous DNA libraries, a strategy identified as an important goal for enabling efficient screening of metagenomic libraries\(^1\). While a known and long standing issue, no reports have described success to this end, to the best of our knowledge.

Here we report a strategy to enable effective screening of heterologous genomic libraries in *E. coli* by expressing heterologous sigma factors. Our hypothesis is that, when expressing heterologous sigma factors, the core RNAP of the host (here *E. coli*) can be recruited to initiate the transcription from heterologous promoters. We show that the expression of the *Lactobacillus plantarum* (*Lpl*) RpoD increases the functional genomic space in *E. coli*, as quantified by green fluorescence protein (GFP) expression using five heterologous genome-wide, promoter GFP-trap libraries. *Lpl* RpoD increases the GFP+ population in all the five libraries tested, which were constructed from phylogenetically diverse genomes, namely, those of *Lpl*, *Bacillus subtilis* (*Bsu*), *Deinococcus radiodurans* (*Dra*), *Clostridium pasteurianum* (*Cpa*) and *C. acetobutylicum* (*Cac*). Furthermore, we demonstrate that the concept works well for a metagenomic library composed of DNA extracted from soil. As a proof of concept for how such effective screening of heterologous genomic libraries can be applied to generate novel complex traits, we expressed a fosmid-based *Lpl* genomic library with large inserts and then screened for genetic loci imparting ethanol tolerance to *E. coli*. We show increased transcripts from heterologous genes located on a fosmid imparting ethanol tolerance. Ethanol tolerance is a complex trait of industrial importance\(^18\) that has attracted considerable attention\(^19\)\(^-\)\(^22\), and has been also employed as a model phenotype for the tool development\(^11\)\(^,\)\(^19\)\(^-\)\(^22\). *Lpl* is one of the most ethanol, butanol and generally alcohol and solvent-tolerant organisms known\(^26\)\(^-\)\(^29\).

Our strategy can increase the efficiency of genomic library screening to facilitate the discovery of novel genetic elements from otherwise inaccessible genomes.

**Results**

GFP-trap libraries assess recognition of heterologous promoters.

We desired to assess, in a quantitative and high-throughput way, the fraction of heterologous promoters that can be recognized by the *E. coli* RNAP to initiate transcription. To this effect, for each of five phylogenetically diverse genomes, we constructed promoter GFP-trap libraries (Fig. 1b), similar to what was previously described\(^10\). The five genome-wide heterologous libraries were LPL-trap, BSU-trap, DRA-trap, CPA-trap and CAC-trap libraries, which were constructed from the *Lpl*, *Bsu*, *Dra*, *Cpa* and *Cac* genomes, respectively (Table 1). For clarity, we describe the construction and properties of these libraries based on the LPL-trap and LPL\(^{lac}\)-trap libraries. The latter was constructed from the *Lpl* genome as a positive control to quantify transcriptional termination within the genomic fragments, and serves as a validation for the proposed concept (described below and in Supplementary Note 1).

LPL libraries were constructed from randomly sheared fragments of genomic *Lpl* DNA with an eightfold genomic coverage (Methods). Sequencing of 10 randomly selected inserts confirmed an average insert size of 726 bp (Table 1), purposefully chosen to be smaller than the average gene size in prokaryotes (of about 924 bp (ref. 31)) to maximize the number of DNA fragments that contain promoters that are not followed by transcriptional terminators (Supplementary Note 1). The library insert was fused in front of a promoterless GFP gene (*gfp*, Fig. 1b), which was optimized for translation by incorporating three frame stop codons and a ribosomal binding site in front of the gene. Thus, transcription initiated inside a library insert leads to expression of *gfp* and the resulting green fluorescence is used as a direct measure of transcription from *Lpl* promoters. Flow cytometry (FC) analyzes this fluorescent signal from individual library clones (Fig. 1c) and, thus, the expression profile of the libraries can be acquired in a high-throughput fashion to quantify the fraction of *Lpl* promoters recognized by *E. coli*. Random fragmentation of genomic DNA (gDNA) generates a collection of different inserts containing promoters, terminators as well as DNA of open-reading frames (Supplementary Fig. 1). We first tested the validity of our FC assay by analysing the GFP expression profile of the LPL\(^{lac}\)-trap library (Fig. 2a). Here the isopropylthigalactoside (IPTG)-inducible *E. coli lac* promoter, *P\(^{lac}\)*, is placed upstream of the library insert to initiate transcription leading to GFP expression if no terminator is present in the insert. We performed a simulation based on the LPL-trap and LPL\(^{lac}\)-trap libraries (Supplementary Note 1) and we estimated that 62% of the LPL\(^{lac}\)-trap fragments would lead to GFP expression on IPTG induction. Experimentally, we observed that the fraction of GFP-expressing cells increased steadily to a maximum of 54%, 7 h post induction (Fig. 2a). While lower than predicted (see discussion in Supplementary Note 1), this demonstrates that our FC assay is conservatively valid.

To establish the baseline of *Lpl* promoter expression in unmodified *E. coli*, we determined the fraction of library inserts containing an *Lpl* promoter that is recognized by the native *E. coli* RNAP following the GFP expression profile of the LPL-trap library when co-transformed with the control plasmid (pControl). A maximum of 6.5% of the library population became positive for GFP expression on IPTG induction. Experimentally, we observed that the fraction of GFP-expressing cells increased steadily to a maximum of 54%, 7 h post induction (Fig. 2a). While lower than predicted (see discussion in Supplementary Note 1), this demonstrates that our FC assay is conservatively valid.

To establish the baseline of *Lpl* promoter expression in unmodified *E. coli*, we determined the fraction of library inserts containing an *Lpl* promoter that is recognized by the native *E. coli* RNAP following the GFP expression profile of the LPL-trap library when co-transformed with the control plasmid (pControl). A maximum of 6.5% of the library population became positive for GFP expression on IPTG induction. Experimentally, we observed that the fraction of GFP-expressing cells increased steadily to a maximum of 54%, 7 h post induction (Fig. 2a). While lower than predicted (see discussion in Supplementary Note 1), this demonstrates that our FC assay is conservatively valid.
Figure 1 | Concept and Strategy. (a) A heterologous transcription factor expressed in the cell recruits the native core RNA-polymerase and initiates transcription from heterologous promoters present on genomic library inserts leading to increased transcription. (b) Two promoter GFP-trap libraries, LPL-trap and LPLlac-trap libraries, were constructed to assay transcription from heterologous promoters. A small Lpl genomic fragment of about 726 bp was fused to a gfp reporter gene as a single transcriptional unit. (c) If transcription inside the genomic fragment occurs, the cell expresses GFP, which can be measured in a high throughput fashion via flow cytometry. Transcription and, therefore, GFP expression is initiated by P lac located in front of the genomic insert for the LPLlac-trap library. GFP-positive 7 h post induction (Fig. 2a), indicating that some Lpl promoters are recognized by the native E. coli RNAP.

Heterologous sigma factors enable recognition of foreign promoters. Using the LPL-trap library, we investigated whether expressing the major sigma factor of L. plantarum (RpoD) can increase the fraction of Lpl promoters recognized by E. coli (Fig. 1a). Lpl rpoD is one of only three sigma factor genes in the Lpl genome and whose regulon encompasses 99% of all Lpl genes. The LPL-trap library was co-transformed into E. coli together with the Lpl rpoD expressing plasmid (pLPLrpoD) and the GFP profile of the resulting library population was followed after induction of Lpl rpoD expression. Lpl rpoD expression was verified via PCR with reverse transcription (RT–PCR; Supplementary Note 2; Supplementary Fig. 2). A steady increase of GFP-positive cells was observed (Fig. 2a) reaching, 7 h post induction, a maximum of 23%, which represents a 3.5-fold increase of GFP-positive cells compared with the plasmid control strain (pControl; Fig. 2a). Simulation analysis (Supplementary Note 1) estimates that 25% of all inserts in LPL-trap library carry an Lpl promoter upstream of gfp, which, if recognized by the host, would result in GFP expression. Thus, this 23% fraction compares favourably with the predicted maximum of 25% (see above and Supplementary Note 1) and suggests that most of the Lpl promoters in the library can be recognized by the engineered E. coli strain. To rule out the possibility that the Lpl RpoD initiates transcription from the backbone of the GFP-trap library vector, we tested GFP-trap plasmids that contained random synthetic DNA fragments as inserts, which contained neither a promoter nor a terminator and the GFP population was small (Supplementary Note 3; Supplementary Fig. 3).

To investigate whether this increased promoter recognition is the specific result of expressing Lpl rpoD or whether overexpression of any sigma factor could have the same effect, expression plasmids containing the major sigma factors from E. coli (pECORσ, expressing the E. coli rpoD), Cac (pCACσ, expressing the Cac sigA) and Bsu (pBSUS, expressing the Bsu sigA) were individually transformed into E. coli together with the

Table 1 | List and features of libraries.

| Library      | Source         | Gram ± | GC%  | Average insert size (bp) | Fold library coverage (95%) |
|--------------|----------------|--------|------|--------------------------|-----------------------------|
| LPL-trap     | Lactobacillus plantarum | +      | 44.5%| ~726                     | 8.0                         |
| Bsu-trap     | Bacillus subtilis     | +      | 43.5%| 1,684                    | 2.9                         |
| CPA-trap     | Clostridium posteuriam | +      | 29.8%| 562                      | 8.2                         |
| DRA-trap     | Deinococcus radiouaruns | +    | 66.6%| 736                      | 3.7                         |
| CAC-trap     | Clostridium acetobutylicium | +    | 30.9%| 267                      | 5.1                         |
| META-trap    | Soil metagenomic DNA | NA     | ~61% | 609                      | NA                          |
| Fos-LPL      | Lactobacillus plantarum | +    | 44.5%| ~35,000                  | 16.5                        |

NA, not applicable.
LPL-trap library and the GFP profiles examined (Fig. 2a). Overexpression of either \textit{E. coli} RpoD or \textit{C. acetobutylicum} SigA resulted in decreased \textit{Lpl} promoter recognition. However, expression of \textit{B. subtilis} SigA together with the LPL-trap library led to increased GFP expression up to 11% (Fig. 2a). While not as substantial an increase as with pLPL\textsubscript{cs}, this suggests that cross recognition of heterologous promoters by different sigma factors is possible; this is pursued further below. We also demonstrated that chromosomal integration of \textit{Lpl} rpoD (Supplementary Note 4; Supplementary Figs 4 and 5) leads to enhanced \textit{Lpl} promoter recognition, although fractionally less than what was achieved by plasmid-based \textit{Lpl} rpoD expression. Furthermore, we found that expression of an alternative \textit{Lpl} sigma factor (RpoN, with only three sigma factors each, like \textit{RpoD} counterpart, \textit{B. subtilis} sigma factors from \textit{D. radiodurans} \textit{Deinococcus radiodurans}, \textit{Lactococcus lactis} \textit{Lactobacillus brevis} (the latter two containing only three sigma factors each, like \textit{Lpl}) in each of these cases, however, no increased \textit{GFP\textsuperscript{+}} population was observed for any of the GFP-trap libraries (Supplementary Fig. 6) showing the unique capabilities of \textit{Lpl} RpoD.

**Cross species promoter recognition.** Metagenomic libraries generated from a mixed population of organisms contain a large genomic diversity of promoters. Thus, effective screening of such libraries requires a recognition of the large set of such diverse promoters. We examined whether \textit{E. coli} transcription from heterologous promoters from multiple organisms (\textit{Lpl}, \textit{Bsu}, \textit{Cpa}, \textit{Dra} and \textit{Cac}) could be initiated by expressing one or a few foreign sigma factors; this would require substantial promoter cross recognition. To do so, we co-transformed a sigma factor expression plasmid (pLPL\textsubscript{cs} or pBSU\textsubscript{cs}) along with an individual library (BSU-trap, CPA-trap, CAC-trap and DRA-trap) and monitored GFP expression (Fig. 2b–e). We observed a large range in the baseline of the \textit{GFP\textsuperscript{+}} population with the control (empty) pControl strain (Fig. 2b–e), likely due to the phylogenetic distances between these organisms (including their G + C content) and the average library insert size (Table 1, discussion in Supplementary Note 6). However, in each case, the strain expressing \textit{Lpl} rpoD showed a large increase in promoter recognition compared to control. Likewise, the strain expressing \textit{Bsu} sigA showed a statistically significant improvement over the control in all libraries. The \textit{Lpl} rpoD expressing strain outperformed the \textit{Bsu} sigA except for equivalent performance on BSU trap.

In addition, we also tested the impact of expressing the main sigma factors from \textit{Deinococcus radiodurans}, \textit{Lactococcus lactis} subsp. \textit{lactis} and \textit{Lactobacillus brevis} (the latter two containing only three sigma factors each, like \textit{Lpl}). In each of these cases, however, no increased \textit{GFP\textsuperscript{+}} population was observed for any of the GFP-trap libraries (Supplementary Fig. 6) showing the unique capabilities of \textit{Lpl} RpoD.

Overexpressing multiple heterologous sigma factors did not provide a benefit in terms of expanding promoter recognition (Supplementary Note 7), and significantly, had a profound negative effect on \textit{E. coli} growth (Supplementary Note 8; Supplementary Fig. 7).

To generalize the concept and its impact, using DNA extracted from soil, we constructed a metagenomic GFP-trap library, which contained prokaryotic and fungal DNA and had a high (61%) G + C content (Methods). The \textit{Lpl} rpoD expressing strain showed the highest \textit{GFP\textsuperscript{+}} population, ninefold higher than the control; the \textit{Bsu} sigA expressing strain was nearly double the control (Fig. 2f). This powerfully demonstrates that the \textit{Lpl} RpoD sigma factor can be used to access a wide range of promoters originating from different species.

**Expression of \textit{Lpl} rpoD alters \textit{E. coli}'s transcriptional program.** Our data (Fig. 2) show that \textit{Lpl} RpoD recruits the \textit{E. coli} RNAP components to initiate transcription from \textit{Lpl} and other heterologous promoters. We observed a negative effect on growth in the \textit{Lpl} rpoD expression strain, similar to what was observed with overexpression of the native RpoD (Supplementary Note 8). This suggests that \textit{Lpl} RpoD competes with native \textit{E. coli} sigma factors to recruit the RNAP complex. One would expect that this competition would affect the native \textit{E. coli} transcriptome. We investigated whether this would affect primarily the regulon of the \textit{Lpl} RpoD counterpart, \textit{E. coli} sigma factor 70, or all sigma factor regulons. Microarray-based transcriptional analysis was carried out by comparing native \textit{E. coli} strain, C600, to a strain overexpressing \textit{Lpl} RpoD. We observed a significant decrease in transcription of genes related to cell envelope, cell motility, chemotaxis, flagella and flagellar biosynthesis, consistent with a previous report (31). Thus, our data demonstrate that \textit{Lpl} RpoD can be used to access a wide variety of promoters generated by a metagenomic library.
Identifying heterologous genetic loci imparting ethanol tolerance.

We desired to demonstrate that increased transcription of genes from the ethanol tolerant \textit{L. plantarum} facilitated by the expression of \textit{Lpl} \textit{rpoD} could be employed to select \textit{Lpl} genetic loci enabling a selectable trait. To this effect, we expressed in wild-type (WT) \textit{E. coli} MG1655 the \textit{Lpl} fosmid library FosLp with an insert size of about 35 kbp, as well as the control fosmid FosC in tandem with \textit{Lpl} \textit{rpoD} from plasmid pLPL\textsigma{}-2 (Table 1; pLPL\textsigma{}-2 differs from pLPL\sigma{} in origin of replication due to compatibility issues with the fosmid library). Cultures of the library MG1655(FosLP/pLPL\textsigma{}-2), as well as of the control strain MG1655(FosC/pLPL\textsigma{}-2), were screened in parallel via a serial-enrichment strategy consisting of alternating rounds of \textsim{}9\% (v/v)-ethanol exposure and recovery phases of 12–24 h (Fig. 4a). Following an initial cultivation for 7 h with IPTG induction to express \textit{Lpl} \textit{rpoD} and, thus, most of the \textit{Lpl} genes, ethanol was added to \textsim{}9\% v/v concentration. After 12 h of exposure to ethanol, 10 ml of cultures were transferred into fresh media and allowed to recover for 12–24 h until the next ethanol exposure. Serial dilutions were plated after each phase to isolate individual clones and growth as well as ethanol concentrations were monitored throughout the serial enrichment. Two culture replicates of the control strain MG1655(FosC/pLPL\textsigma{}-2) died out after the second or third exposure. Clones with increased survival were isolated after the third exposure phase from the MG1655(FosLP/pLPL\textsigma{}-2) culture. Among 18 clones tested (15 of which showed increased tolerance, 2 sequenced and found unique) for survival in 7\% v/v ethanol (Methods), clone 10T31 (Fos10T31) consistently showed the highest survival. Sequencing Fos10T31 revealed a 30-kb insert spanning the \textit{Lpl} chromosomal locus 3,176,475-3,206,327 (NCBI Ref-Seq accession code NC_004567) containing several annotated genes with a potential role in ethanol tolerance. Ethanol tolerance is a complex phenotype involving several potential mechanisms, including mechanisms involving stress proteins, molecular pumps, DNA-repair proteins, altered membrane properties and energy metabolism\textsuperscript{18,20,23}. Here we found genes encoding transporters (\textit{araP}, \textit{lp}_3563 and \textit{lp}_3565), two membrane proteins (\textit{lp}_3575 and \textit{lp}_3577), proteins associated with energy metabolism (\textit{lox} and \textit{poxA}), as well as two proteins (\textit{catalase (kat)} and a heat-shock protein (\textit{clpL})) involved in stress response.

To verify that the observed ethanol tolerance is not the result of spontaneous mutations introduced during the ethanol exposure, Fos10T31 and FosC were retransformed into \textit{E. coli} MG1655 lacZ::\textit{rpoD} (\textit{rpoD} integration was to ensure stable strains for tolerance assays) never exposed to ethanol. Ethanol survival of \textit{E. coli} MG1655 lacZ::\textit{rpoD} expressing Fos10T31 was 5.7-fold and 13.9-fold higher after a 24 and 48 h exposure to 7\% (v/v) ethanol, respectively, than the control strain expressing FosC (Fig. 4b). Thus, the observed ethanol tolerance is a specific effect of the
genes on Fos10T31. Furthermore, the effect of Fos10T31 was tested in the WT MG1655. Without Lpl rpoD expression, the increase in tolerance of the Fos10T31 strain compared with the FosC strain to 7% ethanol was reduced to 1.3-fold and 1.4-fold (statistically insignificant, t-test P-value > 0.05) at 24 and 48 h, respectively. Thus, Lpl rpoD expression greatly increases the screening sensitivity to identify beneficial genomic loci. Note that the absolute survival in MG1655 lacZ::rpoD for both Fos10T31 and FosC is lower than in the WT strain. This reflects the impact of Lpl rpoD expression strain compared with the control. To test, we randomly selected multiple 6-kb regions on the E. coli genome and documented the transcription profile from the Lpl RpoD expression strains compared with control strains (pre-normalized read density output from RNAseq experiments, Supplementary Fig. 8). While the quantitative output of RNAseq reads varies, the two strain types show extremely similar transcription patterns with respect to where transcription is initiated and terminated. That is, the relative abundance of the messenger RNAs (mRNAs) was significantly altered between the two strains, but the same mRNA transcripts were present in both populations.

Discussion
We demonstrated that, for the first time, the transcription machinery of E. coli (and in principle of any other host organism) can be engineered to recognize a large set of heterologous specific Lpl genes that impart tolerance are identified with further subcloning, they can be expressed off native E. coli promoters without the need for Lpl rpoD expression in final strain.

RNA-seq shows expression of Lpl genes is enabled by Lpl RpoD. The data of Figs 2 and 3 show that Lpl RpoD enables superior expression of genes from Lpl libraries, and that increased expression of Lpl genes from Fos10T31 imparts ethanol tolerance. To confirm and detail these conclusions, we performed strand-specific RNA-seq transcriptome analysis of Lpl genes in E. coli. Four cultures were grown in LB in biological duplicates: MG1655(FosC), MG1655(Fos10T31), MG1655 lacZ::rpoD(FosC) and MG1655 lacZ::rpoD(Fos10T31). These cultures were induced for 6 h with IPTG at which time samples were taken for RNA-seq analysis. RNA was used to make strand-specific RNA-seq libraries, which were sequenced using HiSeq 2500 (Illumina). Reads were aligned to both the E. coli MG1655 and Lpl WCF51 genomes and normalized reads per kilobase per million (RPKM) calculated. A negligible number of reads aligned to the Lpl genome from samples with FosC (with the exception of MG1655 lacZ::rpoD(FosC), where reads aligned to Lpl rpoD, as expected). Out of the 29 Lpl genes (most of which are coded on the negative strand) on the Fos10T31 insert, 18 were expressed higher (based on coding strand RPKM, q-value <0.05) in MG1655 lacZ::rpoD(Fos10T31) compared with MG1655(Fos10T31) (Fig. 5a). Only one gene, Lpl araP, was expressed higher in MG1655. Compared with MG1655, three times as many Fos10T31 genes in MG1655 lacZ::rpoD had above average RPKM (19.6) values; this is consistent with Lpl-trap data where 3.5 times as many clones were GFP positive in the strain expressing Lpl RpoD (Fig. 2a). Lpl RpoD enables transcription from both coding and noncoding strands of the Fos10T31 insert (Fig. 5b). This is not unusual and takes place natively in most prokaryotic genomes as has been recently reported\(^3,34\). These data also will make possible to select a subset of Fos10T31 genes (those with the highest differential expression between the two strains; Fig. 5b) for screening to identify those responsible for the tolerant phenotype.

We also examined the change in the E. coli transcriptome when the Lpl RpoD was expressed from a single chromosomal copy. Using our strand-specific RNAseq data, we compared the transcription profiles of MG1655(Fos10T31) and MG1655 lacZ::rpoD(Fos10T31). Similar to above, we observed differential expression of 41% of E. coli genes (1767 genes, t-test P-value P<0.05). Of those, 20% (346 genes) had greater than twofold change in either direction (118 upregulated and 228 down-regulated). To address whether expression of Lpl RpoD results in different sites of transcription of foreign DNA, we asked whether there were a large number of novel and different transcripts observed in the Lpl RpoD expression strain compared with the control. To test, we randomly selected multiple 6-kb regions on the E. coli genome and documented the transcription profile from the Lpl RpoD expression strains compared with control strains (pre-normalized read density output from RNAseq experiments, Supplementary Fig. 8). While the quantitative output of RNAseq reads varies, the two strain types show extremely similar transcripts and transcription patterns with respect to where transcription is initiated and terminated. That is, the relative abundance of the messenger RNAs (mRNAs) was significantly altered between the two strains, but the same mRNA transcripts were present in both populations.
promoters, thus leading to increased expression of heterologous genes. This was successfully exploited for a function-based screening of a fosmid library to identify Lpl determinants imparting ethanol tolerance in E. coli. The ethanol tolerance imparted by the genes on Fos10T31 expressed in MG1655 lacZ::rpoD compares very favourably with other reported tolerant strains in the literature (Supplementary Table 1). For example, Uchiyama et al. compared very favourably with other reported tolerant lacZ::rpoD imparted by the genes on Fos10T31 expressed in MG1655 initiating transcription (only 1 and 4% were GFP positive), even if the consensus sequences are identical. Here we focused on overcoming limitations in heterologous gene expression at the transcription level. Methods have been reported to overcome limitations at the translation level, such as directed evolution of ribosomal protein S1 to increase translation initiation or the introduction of transfer RNAs to accommodate rare codons. Weenvision expression of heterologous sigma factors as a tool for the functional screening of metagenomic and heterologous genomic libraries. Desired genes and operons, once discovered, can then be optimized via codon optimization, synthetic regulatory structure, choice of strain and so on, but to be found, they must be first transcribed.

**Methods**

**Strain and growth conditions.** Strains, plasmids and oligonucleotides are listed in Supplementary Tables 2, 3 and 4, respectively. Cultures were performed in lysogeny broth (LB, containing 10 g l\(^{-1}\) NaCl, 10 g l\(^{-1}\) bacto tryptone and 5 g l\(^{-1}\) yeast extract) with antibiotics (100 μg ml\(^{-1}\) ampicillin, 35 μg ml\(^{-1}\) chloramphenicol for plasmids and 12.5 μg ml\(^{-1}\) for fosmids, 50 μg ml\(^{-1}\) kanamycin) as required at 37°C with shaking at 220 r.p.m.

**Construction of destination plasmids.** To streamline the cloning of multiple sigma factors as well as the construction of promoter GFP-trap libraries from a variety of organismal or metagenomic DNA (mgDNA), a two-plasmid expression system was constructed. Two compatible plasmids, a low-copy plasmid for sigma

Figure 5 | Strand-specific RNA-seq analysis to assess expression of L. plantarum genes in E. coli. (a) Reads per kilobase per million (RPKM mapped sequence reads) of transcripts on Fos10T31 in the control strain MG1655 (blue bars) and MG1655 lacZ::rpoD (green bars). Genes listed in order of position on Fos10T31. Dot indicates differential expression (q-value < 0.05), double dot indicates q-value < 0.01. (b) Representative raw read alignment of transcripts from the Fos10T31 insert sequence. Most genes are coded on the negative strand. Scales for all alignments are identical. Red reads are for transcripts from genes on the negative strand of the Lpl DNA, and blue is for transcripts from genes on the positive strand. Compared with the control strain, the impact of Lpl-RpOD expression in E. coli is profound in generating transcripts from both the strands of Lpl DNA.
factor expression and a high copy plasmid for the promoter GFP-trap libraries, were constructed as destination vectors to utilize the in vitro recombinolation Gateway technology (Invitrogen) at Carlsbad, CA, USA. The desired combination of sigma factor and genomic library can be cloned in the two plasmid expression system without cumbersome genetic manipulation. A detailed description of the construction of the destination plasmids is found in other sections below. (Construction of the sigma factor expression plasmids' and 'Construction of the promoter GFP-trap destination vectors'.) Briefly, gfp was first cloned into the multiple cloning site (MCS) of pUC19 (New England Biolabs (NEB), Ipswich, MA, USA) after a LR recombination cassette (Invitrogen) was introduced upstream of gfp. The resulting plasmid was designated as pL-UC-LR-GFP. Destinatory plasmid pL-DNA was generated by removing Pjac from pUC-LR-GFP. These two high copy plasmids were used to construct the promoter GFP-trap libraries (see below). Cloning an LR recombination cassette in pACYC-Duet (Novagen Merck KGaA, Darmstadt, Germany) generated the low-copy destination plasmid pACYC-LR for sigma factor expression (see below).

**Construction of sigma factor expression plasmids.** Sigma factor coding genes were first amplified from gDNA of the respective organism. The sigma factor expression set comprises the following plasmids: pLPL (expressing Lpl pRod (lp.1962)); pLPLpRoN (expressing Lpl pRod (lp.0787)); pECoR (expressing E. coli pRod (bp.8067)); pBSuR (expressing B. subtilis sigA (BSU25200)); pCaCoR (expressing C. acetobutylicum sigA (CA.1380)); pDARsig (expressing Deinococcus radiodurans RodD (DR.0996)); pLALsig (expressing Lactococcus lactis subsp. lactis RodD (L1239)); pLPLsig (expressing Lactococcus plantarum RodD); pLPLsig2 (expressing RodD from pLPLsig) generated by recombination with pDEST14 comprises: pLPLsig2-2 (expressing Lpl pRod), pBsuR sig2 (expressing B. subtilis sigA) and pCaCoR sig2 (expressing C. acetobutylicum sigA). The pENTR-gus vector, which contains a promoterless gus gene from *Arabidopsis thaliana*, was recombined with pACYC-LR and pDEST4 to create the corresponding entry and pDEST-recombinant. Expression of sigma factors was verified via RT-PCR (Supplementary Note 2).

The destination vector for sigma factor overexpression was based on the low copy plasmid pACYC-Duet (Novagen) that was converted to a destination plasmid by introducing an LR-recombinosome cassette amplified from pDEST40 (Invitrogen) with the primer pair DEU-F and DEST-R. A restriction site overhang (Supplementary Table 4) and cloned via restriction recombination with pENTR-gus as per the manufacturer’s suggestions. The generated destination plasmid, designated as pControl, was used as a control plasmid.

Sigma factors from different organisms were PCR amplified from gDNA with restriction site overhangs (Supplementary Table 4) and cloned via restriction enzyme cloning in the MCS of pUC19 under the control of Pjac. Using this plasmid, the sigma factors together with the pCam were amplified again with the primer pair pUC19-for and pUC19-rev. The amplified sigma factors were TOPO-TA cloned into pCR8/GW/TOPO (Invitrogen), creating pLPLsigR. The expression vector of Lpl sig was a 767-bp fragment generated by in vitro recombination with pENTR-gus as per the manufacturer’s suggestion. The recombinant plasmid, designated as pControl, was used as a control plasmid.

**Construction of the promoter GFP-trap libraries.** gDNA was isolated using the ChargeSwitch gDNA Mini Bacteria Kit (Invitrogen) according to the manufacturer’s instructions from a Lactobacillus plantarum WCFS1 culture grown to an optical density at 600 nm (OD600) of 1. A fraction of 25 μl of the DNA was sheared by 30 passages through a 50-μl microsyringe and used as a substrate for the double digestion of both destination plasmids was verified by restriction digest and sequencing. Primers used to construct these plasmids are listed in Supplementary Table 4.

**Construction of the promoter GFP-trap libraries.** gDNA was isolated using the ChargeSwitch gDNA Mini Bacteria Kit (Invitrogen) according to the manufacturer’s instructions from a Lactobacillus plantarum WCFS1 culture grown to an optical density at 600 nm (OD600) of 1. A fraction of 25 μl of the DNA was sheared by 30 passages through a 50-μl microsyringe and used as a substrate for the double digestion of both destination plasmids was verified by restriction digest and sequencing. Primers used to construct these plasmids are listed in Supplementary Table 4.

**Construction of the promoter GFP-trap libraries.** gDNA was isolated using the ChargeSwitch gDNA Mini Bacteria Kit (Invitrogen) according to the manufacturer’s instructions from a Lactobacillus plantarum WCFS1 culture grown to an optical density at 600 nm (OD600) of 1. A fraction of 25 μl of the DNA was sheared by 30 passages through a 50-μl microsyringe and used as a substrate for the double digestion of both destination plasmids was verified by restriction digest and sequencing. Primers used to construct these plasmids are listed in Supplementary Table 4.

**Construction of the promoter GFP-trap libraries.** gDNA was isolated using the ChargeSwitch gDNA Mini Bacteria Kit (Invitrogen) according to the manufacturer’s instructions from a Lactobacillus plantarum WCFS1 culture grown to an optical density at 600 nm (OD600) of 1. A fraction of 25 μl of the DNA was sheared by 30 passages through a 50-μl microsyringe and used as a substrate for the double digestion of both destination plasmids was verified by restriction digest and sequencing. Primers used to construct these plasmids are listed in Supplementary Table 4.

**Construction of the promoter GFP-trap libraries.** gDNA was isolated using the ChargeSwitch gDNA Mini Bacteria Kit (Invitrogen) according to the manufacturer’s instructions from a Lactobacillus plantarum WCFS1 culture grown to an optical density at 600 nm (OD600) of 1. A fraction of 25 μl of the DNA was sheared by 30 passages through a 50-μl microsyringe and used as a substrate for the double digestion of both destination plasmids was verified by restriction digest and sequencing. Primers used to construct these plasmids are listed in Supplementary Table 4.

**Construction of the promoter GFP-trap libraries.** gDNA was isolated using the ChargeSwitch gDNA Mini Bacteria Kit (Invitrogen) according to the manufacturer’s instructions from a Lactobacillus plantarum WCFS1 culture grown to an optical density at 600 nm (OD600) of 1. A fraction of 25 μl of the DNA was sheared by 30 passages through a 50-μl microsyringe and used as a substrate for the double digestion of both destination plasmids was verified by restriction digest and sequencing. Primers used to construct these plasmids are listed in Supplementary Table 4.

**Construction of the promoter GFP-trap libraries.** gDNA was isolated using the ChargeSwitch gDNA Mini Bacteria Kit (Invitrogen) according to the manufacturer’s instructions from a Lactobacillus plantarum WCFS1 culture grown to an optical density at 600 nm (OD600) of 1. A fraction of 25 μl of the DNA was sheared by 30 passages through a 50-μl microsyringe and used as a substrate for the double digestion of both destination plasmids was verified by restriction digest and sequencing. Primers used to construct these plasmids are listed in Supplementary Table 4.
fragments of about 35 kb (selected from an agarose gel) were ligated into the CopyControl Fosmid Library Production Kit (Epicentre) according to the manufacturer’s recommendations. DNA fragments were first end-repaired, and fragments of about 35 kb (selected from an agarose gel) were ligated into the library. The fosmid library was constructed as per the manufacturer’s recommendations using the supplied control DNA, a 42-kb fragment of the human X-chromosome.

Because the fosmid library was constructed in a restriction and methylation negative cloning strain (Epich1030-1T8), fosmid DNA (extracted from the fosmid library) was electroporated into a restriction negative but methylation positive cloning strain (NEB Saph). Transformations were pooled and outgrown as described earlier (see ‘Construction of the promoter GFP-trap destination vectors’). On the basis of the size of a serial dilution plated after pooling, the library was estimated to contain 1.2 × 10^8 fosmids. Fosmid DNA, extracted from this library, was then used to transform the screening strain E. coli MG1655 containing pLPL-2. Here a library size of 6 × 10^6 clones was achieved. The control fosmid FosC was also methylated in NEB Sapha cells and introduced into E. coli MG1655 containing pLPL-2 to generate the appropriate control strain.

Chromosomal integration of Lpl rpoD. The lambda red system21 was used to integrate Lpl rpoD into the lac locus of the WT E. coli strain MG1655. First, a kanamycin resistance gene (KanR) flanked by two FRT sites was amplified from pKD4 (FRT-Kan-for and FRT-Kan-rev primer pair) and inserted via restriction digest cloning with SacI and EcoRI into the MCS of pUC19 (NEB) yielding plasmid pKD4 (FRT-Kan-for and FRT-Kan-rev primer pair) and inserted upstream of the KanR in pUC-Kan via restriction digest cloning with KpnI and SacI yielding pUC-pKD-kan.

From this plasmid, a knock-in cassette was assembled (pUC9-for and pUC19-rev primer pair) composed of the Lpl rpoD and the downstream KanR flanked by the homologous regions of pUC9 to the lac locus of E. coli. This cassette was electrotransformed into E. coli MG1655 containing the helper plasmid pKD20, carrying the lambda red system and plated on the LB plates containing kanamycin. Colony PCR (lac-for and lacZ-rev primer pair) was performed to select clones with successful integration of the cassette. Clones with a confirmed integration were transformed with the helper plasmid pCP20 and cured of the kanR via FRT-mediated excision. The resulting markerless pLPL-trap library was verified via colony PCR (lac-for and lacZ-rev primer pair) and sequencing. Expression of Lpl rpoD was checked via RT–PCR (Supplementary Note 4).

Flow cytometry and fluorescent activated cell sorting (FACS). GFP-trap libraries (1 μg DNA) were transformed into sigma factor genes (hosted by E. coli NEB 10beta for plasmid-based sigma factor genes or E. coli MG1655 for integration-based genes) and grown to late exponential phase (OD600 ≥ 0.8–1.0) and stored at −80 °C for long-term storage. Plates were incubated at 37 °C to ensure the maintained library coverage. Library cultures were started with a 2-mL frozen stock or 4% overnight culture and grown in 500-mL baffled flask containing 100 mL of culture. Cultures were incubated for 24 and 48 h with closed caps to avoid ethanol recovery phases was carried out until a clear difference in growth between the induced cultures and the control culture was observed (Fig. 4a). Serial dilutions were plated after each exposure phase to isolate clones.

Ethanol tolerance assay for strain characterization. Cultures of individual clones grown overnight were used to inoculate 50-mL falcon tubes containing 10 mL LB with IPTG. These tubes were outgrown with open caps for 7 h to express Lpl rpoD (and thus Lpl genes) before 0.7 μL of culture was replaced by 100% ethanol. The cultures were incubated for 24 and 48 h with closed caps to avoid ethanol recovery phases. Colony-forming units before and after ethanol exposure were determined by plating a serial dilution and the specific survival rate of a strain was calculated as c.f.u.(t) = 24 or 48 h/c.f.u.(t = 0) × 100%.

Strand-specific RNA-seq and analysis. MG1655::FosC, MG1655::LPL-trap + FosC, MG1655 + FosT1031 and MG1655::LPL + FosT1031 (all in duplicate) were grown in LB broth and induced with 1 μM IPTG for 6 h at which time 10-mL culture was sampled for RNA isolation. RNA isolation samples were centrifuged at 4,000 relative centrifugal force (RCF) at 4 °C for 10 min, and the supernatant was stored at −80 °C. Ten μg of RNA sample was reverse transcribed with the RNA-seq kit (Qiagen) had RNA DNA-free kit (Life) and was enriched for mRNA with the MicroExon Select (Ambion) thrice, according to the manufacturer’s protocols. The ScriptSeq v2 (Illumina) kit was used to construct the RNAseq libraries. The fragment length of the libraries was checked via Bioanalyzer (Agilent, Santa Clara, CA, USA) before sequencing.

Deep sequencing using HiSeq2500 (Illumina) with a 75-bp read length resulted in individual library sequence files. Files were processed to remove barcodes, trim adaptors and obtain read counts. Rockhopper software58 was used to align raw reads with the E. coli MG1655 genome (accessions from National Center for Biotechnology Information) and to do differential expression analysis (at q-value < 0.05). Integrated Genome Viewer was used for visualizing read alignments49.

Gene expression analysis. pLPL-2 and pControl were transformed into WT E. coli MG1655 and cultured in parallel as described above. Growth was followed by OD measurements and cell pellets for microarray probe generation were collected throughout the time course and stored at −85 °C. Probe generation, microarray hybridization and analysis were performed as described45. Briefly, 25 μg of total RNA were primed with 3 μL of random hexamer primers (5 μg/μL); Roche Molecular Biochemicals, Indianapolis, IN, USA) at 70 °C for 10 min, then transcribed reversed transcribed at 37 °C overnight with 200 U SuperScript III R.Tase (Invitrogen) and 360 μL of aminoacyl labeling mix (dATP, dCTP, dGTP, dTTP and aa-dUTP) containing a 2:3:2:2:1 ratio of Cy3 and Cy5 dyes (Amersham CyDye Mono-Reactive Dye, GE Healthcare, Pittsburgh, PA, USA) following the manufacturer’s recommendations. Two hundred and fifty ng of each labelled cDNA (Cy3 and Cy5) were hybridized using the Gene Expression Hybridization Kit (Agilent) according to the manufacturer’s instructions. Fifteen K. e. coli Gene Expression Microarrays (Agilent) were hybridized with probes generated from three time points (Fig. 3a) for each of the two biological replicate experiments, whereby technical replicates were performed as dye-swap experiments. Normalization was carried out using the Bioconductor package56,47, and genes with a significantly altered expression were identified via one-class time course SAM analysis49.

References
1. National Research Council (US). Committee on Metagenomics: Challenges and Functional Applications (National Academies Press, 2007).
2. Curtis, T. P. & Sloan, W. T. Exploring microbial diversity - a vast below.
3. Schmeisser, C., Steele, H. & Streit, W. R. Metagenomics, biotechnology with non-culturable microbes. Appl. Microbiol. Biotechnol. 75, 953–962 (2006).
4. Hairston, N. G., Ellner, S. P., Geber, M. A., Yoshida, T. & Fox, J. A. Rapid evolution and the convergence of ecological and evolutionary time. Ecol. Lett. 11, 1114–1127 (2008).
5. Papoutsakis, E. T. Exploring the capabilities of the geobiosphere’s microbial genome. AlChE J. 59, 688–698 (2013).
6. Knetisch, A., Bowien, S., Whitel, G., Gottschalk, G. & Daniel, R. Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. Appl. Environ. Microbiol. 69, 3048–3050 (2003).
7. Fernandez-Arrojo, L., Guazzaroni, M. E., Lopez-Cortes, N., Beloqui, A. & Ferrer, M. Metagenomic era for biodiversity identification. Curr. Opin. Biotechnol. 21, 725–733 (2010).

© 2015 Macmillan Publishers Limited. All rights reserved.
8. Nicolau, S. A., Gaida, S. M. & Papoutsakis, E. T. Coexpressing genomic libraries (CoGel) identify interactions among distantly located genetic loci by developing complex microbial phenotypes. *Nucleic Acids Res.* **39**, 2011.

9. Lynch, M. D., Warnecke, T. & Gill, R. T. SCALeX: multiscale analysis of library enrichment. *Nat. Methods* **4**, 87–93 (2007).

10. Reyes, L. H., Almario, M. P., Winkler, J., Orozco, M. M. & Kao, K. C. Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in *Escherichia coli*. *Metab. Eng.* **14**, 579–590 (2012).

11. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1563–1568 (2006).

12. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).

13. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. & Stephanopoulos, G. Engineering the combinatorial library of *Escherichia coli* sigma-dependent genes, promoters, and sigma factor selectivity. *J. Bacteriol.* **187**, 1591–1603 (2005).

14. Nicolau, S. A., Gaida, S. M. & Papoutsakis, E. T. A comparative view of metabolite and substrate stress and tolerance in microbial biproduct: from biofuels and chemicals, to biocatalysis and bioremediation. *Metab. Eng.* **12**, 307–319 (2010).

15. Goodarzi, H. et al. Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*. *Mol. Syst. Biol.* **6**, 378 (2010).

16. Nicolau, S. A., Gaida, S. M. & Papoutsakis, E. T. Exploring the combinatorial genomic space in *Escherichia coli* for ethanol tolerance. *Biotechnol. J.* **7**, 1337–1345 (2012).

17. Woodruff, L. B., Boyle, N. R. & Gill, R. T. Engineering improved ethanol production in *Escherichia coli* with a genome-wide approach. *Metab. Eng.* **17**, 1–11 (2013).

18. Woodruff, L. B. et al. Genome-scale identification and characterization of ethanol tolerance genes in *Escherichia coli*. *Metab. Eng.* **15**, 124–133 (2013).

19. Chen, T. J., et al. Significant rewiring of the transcriptome and proteome of an *Escherichia coli* strain harboring a tailored exogenous global regulator lrp. *PloS ONE* **7**, e37126 (2012).

20. Zingaro, K. A. & Papoutsakis, E. T. Toward a semisynthetic stress response system to engineer microbial soluble tolerance. *MBio* **3**, e00308–e00312 (2012).

21. Zingaro, K. A. & Terry Papoutsakis, E. *GroE*SL overexpression impacts *Escherichia coli* carbon metabolism and ethanol tolerance. *J. Bacteriol.* **194**, 19–31 (2012).

22. Couto, J. A., Pina, C. & Hogg, T. Enhancement of apparent resistance to ethanol in *Lactobacillus hilgardii*. *Biotechnol. Lett.* **19**, 487–490 (1997).

23. Knoshaug, E. P. & Zhang, M. Butanol tolerance in a selection of *Bacillus subtilis* strains to lyophilisation and stress environmental conditions. *Microbiol. Lett.* **23**, 53–61 (2004).

24. Uchiyama, T. & Miyazaki, K. Functional metagenomics for enzyme discovery: strategies to efficient screening. *Curr. Opin. Biotechnol.* **20**, 616–622 (2009).

25. Nicolaos, P. et al. Condition-dependent transcriptome reveals high-level regulatory architecture in *Escherichia coli*. *Science* **335**, 1103–1106 (2012).

26. Bernstein, J. R., Butler, T., Shen, C. R. & Liao, J. C. Directed evolution of ribosominal protein S1 for enhanced translational efficiency of high GC Rhodopseudomonas palustris DNA in *Escherichia coli*. *J. Biol. Chem.* **282**, 18929–18936 (2007).

27. Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6**, 494–500 (1995).

28. Baca, A. M. & Hol, W. G. J. Overcoming codon bias: a method for high-level overexpression of *Plasmid* and other AT-rich parasite genes in *Escherichia coli*. *Int. J. Parasitol.* **30**, 113–118 (2000).

29. Clarke, L. & Carbon, J. Colony bank containing synthetic Col El Hybrid plasmids representative of entire *Escherichia coli* Genome. *Cell* 9, 91–99 (1976).

30. Knoshaug, E. P. & Zhang, M. Butanol tolerance in a selection of *Bacillus subtilis* strains to lyophilisation and stress environmental conditions. *Microbiol. Lett.* **23**, 53–61 (2004).

31. Xu, L. et al. Functional metagenomics for enzyme discovery: strategies to efficient screening. *Curr. Opin. Biotechnol.* **20**, 616–622 (2009).

32. Stevens, M. J. A., Molenaar, D., de Jong, A., De Vos, W. M. & Kleerebezem, M. Gene-expression screening of environmental metagenome libraries for isolation of microorganisms. *Appl. Environ. Microbiol.* **66**, 2541–2547 (2000).

33. Galen, E. M., Alkema, W. B. & Jansen, D. B. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ. Microbiol.* **6**, 879–886 (2004).

34. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. & Stephanopoulos, G. Engineering the combinatorial genomic space in *Escherichia coli* for ethanol tolerance. *Biotechnol. J.* **7**, 1337–1345 (2012).

35. Uchiyama, T. & Miyazaki, K. Functional metagenomics for enzyme discovery: strategies to efficient screening. *Curr. Opin. Biotechnol.* **20**, 616–622 (2009).

36. Todt, T. J. et al. Genome-wide prediction and validation of sigma70 promoters in *Lactobacillus plantarum* WPCFS1. *PLoS ONE* **7**, e45097 (2012).

37. Jones, S. W. et al. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**, 2541–2547 (2000).

38. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Exploring the combinatorial library of *Escherichia coli* sigma-dependent genes, promoters, and sigma factor selectivity. *J. Bacteriol.* **187**, 1591–1603 (2005).

39. Nicolaos, P. et al. Condition-dependent transcriptome reveals high-level regulatory architecture in *Escherichia coli*. *Science* **335**, 1103–1106 (2012).

40. Bernstein, J. R., Butler, T., Shen, C. R. & Liao, J. C. Directed evolution of ribosomal protein S1 for enhanced translational efficiency of high GC *Rhodopseudomonas palustris* DNA in *Escherichia coli*. *J. Biol. Chem.* **282**, 18929–18936 (2007).

41. Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6**, 494–500 (1995).

42. Nicolaos, P. et al. Condition-dependent transcriptome reveals high-level regulatory architecture in *Escherichia coli*. *Science* **335**, 1103–1106 (2012).

43. Bernstein, J. R., Butler, T., Shen, C. R. & Liao, J. C. Directed evolution of ribosomal protein S1 for enhanced translational efficiency of high GC *Rhodopseudomonas palustris* DNA in *Escherichia coli*. *J. Biol. Chem.* **282**, 18929–18936 (2007).

44. Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6**, 494–500 (1995).