

Functional Metagenome Mining: A sequence directed strategy for the retrieval of functional enzymes for biocatalysis**

Jack W. E. Jeffries, Natalie Dawson, Christine Orengo, Thomas S. Moody, Derek J. Quinn, Helen C. Hailes and John M. Ward*

Abstract: Biocatalytic reactions are increasingly being used as a sustainable strategy in organic synthesis and it is recognised that there is need for new enzyme discovery. To establish the utility and versatility of a metagenomics approach, metagenomic DNA extracted from the oral cavity was sequenced and used to create an *in silico* contig library. This enables individual open reading frames, operons or all the enzymes of a particular family to be identified and then retrieved from the original DNA by PCR. As proof of principle a lactate dehydrogenase, a malate dehydrogenase and transketolases were identified *in silico*, successfully cloned and assayed. This new enzyme retrieval sequence directed method gives constructive access to metagenomic diversity and importantly improves on the low hit rate experienced when using conventional metagenomic screens.

Over the last 30 years enzymes have increasingly been used in commercial chemical processes.^[1] Enzyme catalysts have distinct advantages over chemical catalysts such as cost, sustainability, low toxicities and the use of moderate reaction conditions. The greatest advantage however, is the wide range of catalytic activities displayed by naturally occurring enzymes and the stereoselectivities that can be achieved. Drawbacks when using enzymes can include a low organic solvent tolerance, narrow pH working ranges, and low thermostability. However these properties along with enhanced stereoselectivities can be engineered using random or directed evolution techniques.^[2]

One of the key requirements for this type of engineering is multiple amino acid sequences. Having many enzymes that catalyze the same reaction, with variation in their primary sequence is a much more effective starting point for the engineering process than any single example of an enzyme family. One successful method for identifying large numbers of novel enzymes has been to mine fully sequenced and annotated

genomes of laboratory cultivable strains. Such sequence directed genome mining lends itself to the retrieval of multiple targets, however a limitation to this method is the number of annotated available species.^[3] With only 0.1-5% of bacteria cultivable in the laboratory, much of the existing diversity of bacteria are inaccessible in this way.^[4] At the same time, with the advances in high throughput sequencing, whole or partial genomes of uncultivable bacteria are being made available for study. The nascent field of metagenomics gives insight into the genomes of previously unstudied bacteria and by extension a potential wealth of new biocatalysts for the generation of small molecules and large bioactive molecules such as polyketides and glycopeptides.^[5]

The use of metagenomics to obtain new enzymes and biocatalysts for industrial applications is steadily growing but still in its infancy.^[6] Efforts to capture enzymes from metagenomic samples have relied upon the creation of physical metagenomics libraries in more tractable bacteria.^[7] Poor transcription and translation, enzymes with activities that are difficult to screen for all lead to a low "hit rate", the number of enzymes discovered compared with the number of clones screened, for physical functional libraries. This low "hit rate" is disappointing especially when considering the wealth of genomic diversity available through metagenomics.^[8]

At the same time high throughput sequencing and *in silico* functional annotation can be used to study microbial communities as they exist *in vivo*. Studies have used this burgeoning amount of sequence data to aid enzyme retrieval. The large numbers of annotated proteins are used to build consensus sequences from which degenerate PCR primers are designed.

Here a strategy is described which uses the large number of individual protein annotations, generated through high throughput sequencing, to design specific primer pairs for the retrieval of enzymes in a way analogous to genome mining. By building an *in silico* metagenomic contiguous read (contig) library from high throughput sequencing data a resource is created that can be quickly and importantly repeatedly mined for enzymes. Together with conventional molecular biology techniques such as Polymerase Chain Reaction (PCR), DNA cloning and protein overexpression, this strategy will now allow constructive access to the wealth and diversity of enzymes that so far have been out of reach. A comparison of our approach with existing functional metagenomics approaches is summarized in Figure 1.

To test the strategy an *in silico* metagenomic contig library of the oral cavity was created and formatted into a BLAST database. As a proof of principle that specific sequences identified from the *in silico* library can be retrieved by PCR, two lactate dehydrogenase (LDH) sequences and a malate dehydrogenase (MDH) sequence were chosen to query the *in silico* metagenomic contig library. Furthermore five transketolases (TKs) were retrieved as exemplars of industrially relevant enzymes.^[9]

[*] Mr J. W. E. Jeffries, Prof. J. M. Ward,
Department of Biochemical Engineering, University College London,
Bernard Katz Building, Gordon Street, London, WC1H 0AH (UK)
E-mail: j.ward@ucl.ac.uk

Dr N. Dawson, Prof. C. Orengo,
Department of Structural and Molecular Biology, University College
London, Gower Street, WC1E 6BT (UK)

Prof. T. S. Moody, Dr D. J. Quinn,
Department of Biocatalysis and Isotope Chemistry, Almac, 20
Seagoe Industrial Estate, Craigavon, Northern Ireland (UK)

Prof. H. C. Hailes
Department of Chemistry, University College London, 20 Gordon
Street, London WC1H 0AJ (UK)

** Funding from the Biotechnology and Biological Sciences Research
council (BBSRC) for J. W. E. J. and ALMAC is gratefully
acknowledged.
Supporting information for this article is given via a link at the end of
this document.

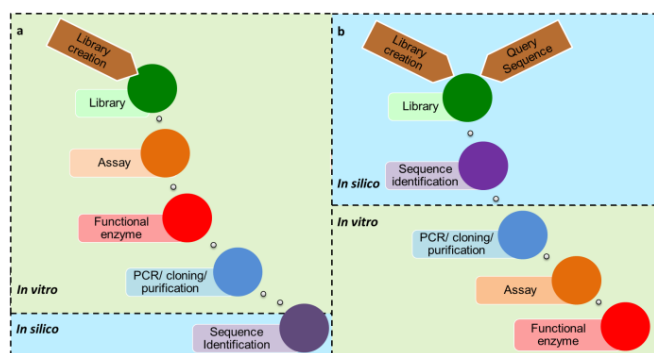


Figure 1. Comparison of the existing functional metagenomics studies and the *in silico* approach developed. **a)** A schematic representation of the current process of enzyme retrieval from physical metagenomics library. The library creation, assay and selection are carried out *in vitro* (area in green) with sequence identification only occurring once a positive clone has been identified (area in blue). **b)** A schematic representation of enzyme retrieval from an *in silico* library using our approach. Once the *in silico* metagenomic library has been created, querying with different driver sequences can identify multiple enzyme families for recovery.

The *in silico* metagenomic contig library was created by sequencing a mixed sample of DNA extracted from the oral cavity (tongue scrapings) of 9 volunteers. Roche 454 sequencing produced 1.1 million reads, which were combined and built into larger contigs using the MIRA assembler. In total 39,971 contigs were generated; the largest was 41.9 kbp the smallest 104 bp with an N50 value of 2416 for the data set. 125,211 Open reading frames (ORFs) and 88,247 protein annotations were marked using metagene mark and the Pfam standalone tool respectively. This contig library was formatted into a BLAST database.

Two LDH amino acid sequences were chosen from Uniprot with which to query the contig library; one each from *Rothia mucalaginosa* (Uniprot id: D2NTCO) and *Streptococcus parasanguinis* (Uniprot id: I1ZLP7), species known to have DNA present in the metagenomic sample. Also chosen was the sequence of a MDH from *Streptomyces venezuelae* (Uniprot id: F2RK50) an organism that does not reside in the mouth but has members of the same genus in the mouth and data set. These enzymes were chosen due to the likelihood of them existing in the dataset and straightforward activity assays.

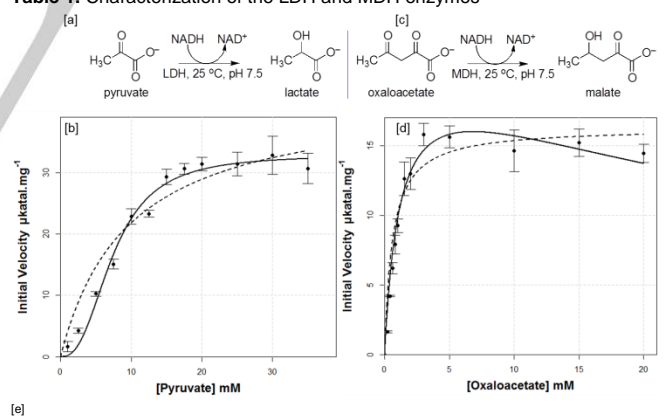
A BLAST search of the library with the LDH sequences from both *Rothia* and *Streptococcus* identified the same 19 contigs. In addition to these 19 contigs, 2 contigs were exclusive to the search with *Rothia* and 2 found only when querying with *Streptococcus*. When querying with the MDH sequence from *Streptomyces* 4 further contigs were identified not matching any found before. In total 27 contigs were identified, 23 found when searching with an LDH sequence, 10 containing full length enzyme sequences. From the 4 contigs found searching with the MDH sequence, only 2 contained full length non redundant sequences.

From this subset of 12 full length enzyme sequences, three were selected as proof of principle to take forward for PCR from the original DNA used to generate the *in silico* library. Two enzymes were chosen with the highest similarity (>95% identity) to the LDH query sequences from *Rothia* and *Streptococcus* to examine the accuracy of the retrieval from BLAST sequence, through contig to retrieved protein. A third enzyme sequence

with only 46% identity to the MDH from *Streptomyces* was also chosen. The query sequence from *Streptomyces venezuelae* and a contig containing a low similarity sequence were deliberately chosen to test the limits of the BLAST search approach, to see if low similarity sequences identified through BLAST searching retained the activity of the original query sequence.

All three enzymes were retrieved through PCR and were expressed in *Escherichia coli*: *Rothia 2665* (LDH) and *Streptomyces 3443* (MDH) were well expressed and could be successfully purified. *Streptococcus 5169* (LDH) was less well expressed and perhaps for this reason could not be purified. Initial rates of reaction for the purified enzymes for varying substrate concentrations were plotted against the change in substrate concentration. Values for the V_{max} , K_{cat} and K_{cat}/K_m for both enzymes were calculated using Michaelis-Menton kinetics and cooperative and inhibitory equations for the LDH and MDH respectively (Table 1). The oligomeric state of the LDH and MDH were inferred as a homotetramer and homodimer and the kinetic data calculated based on these assumptions.^{[10] [11]} The data was consistent with existing studies on enzymes from related bacteria and fit with the assumption of multimeric states for the enzymes.^[12] The DNA of the enzymes was sequenced and compared to the original query sequences. The enzymes *Rothia 2665* and *Streptococcus 5169* had <98% similarity to the sequences used for BLAST. *Streptomyces 3443* showed 62% identity to the original query sequence, an increase in percentage identity due to the relatively poorer quality of the initial contig sequence. Full alignments of the DNA and amino acid sequences for all enzymes are given in supplementary information.

Table 1. Characterization of the LDH and MDH enzymes



	V_{max} $\mu\text{katal}\cdot\text{mg}^{-1}$	K_m mM	K_{cat} s^{-1}	K_{cat}/K_m $\text{M}^{-1}\cdot\text{s}^{-1}$
LDH 2665 Michaelis-Menton	42.8 (± 3.7)	9.6 (± 2.6)	1.2×10^6	1.3×10^8
LDH 2665 cooperative	32.9 (± 0.9)	7.4 (± 0.4)	9.5×10^5	1.3×10^8
<i>A. viscosus</i> LDH ^[10]	35	0.8	8.8×10^5	1.0×10^9
MDH 3443 Michaelis-Menton	16.2 (± 0.7)	0.6 (± 0.1)	2.2×10^5	3.6×10^8

MDH 3443 inhibition	21.6 (± 2.2)	1.2 (± 0.3)	2.9×10^5	2.4×10^8
<i>S. coelicolor</i> MDH ^[10]	26.7	0.2	1.8×10^3	1.0×10^7

Characterisation of purified enzymes at 25 °C following the decrease in absorbance at 340 nm as NADH is oxidised and substrate is reduced. [a] The reaction catalysed by LDH 2665 reducing pyruvate to lactate. [b] Graph showing initial velocities for LDH 2665 plotted against substrate concentration fit with the Michaelis Menton equation (dashed line), and the Hill equation for co-operativity (solid line). [c] The reaction catalysed by MDH 3443 reducing oxaloacetate to malate. [d] Graph showing initial velocities for MDH 3443 plotted against substrate concentration fit with the Michaelis Menton equation (dashed line) and an equation for substrate inhibition (solid line). [e] Biochemical characteristics for both enzymes calculated from the assay data. Values are also shown of enzymes from closely related species, errors for the values are given in parentheses where available.

An alternative method of enzyme identification was used when searching for transketolase (TK) enzymes. Contigs containing TKs were identified using the Pfam ids for TK C-terminal, N-terminal and pyrimidine binding domains. In total 73 contigs were identified with at least one of the Pfam ids; 12 of the contigs contained full length enzymes. Seven of the 12 were non redundant from which primers were designed for retrieval, 5 of which were successfully retrieved. All 5 of the amino acid sequences had >90% identity to sequences deposited in the NCBI database. None of the sequences in the database have been cloned and assayed having been identified via homology from sequencing data. One of the metagenomics TK, identified as having 90% similarity to a sequence from *Atopobium parvulum*, had an internal stop codon and appeared to be a split domain TK. Sequencing of the PCR product maintained this internal stop codon and SDS gel of the purified enzyme showed two distinct bands corresponding to the computed molecular weight of the two domains Figure 2.

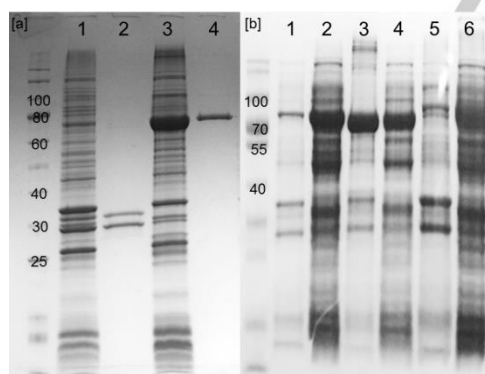


Figure 2. SDS page gel of metagenomic transketolases. [a] Lanes 1-2 show the pellet and soluble purified **TK1**. Two distinct bands are visible corresponding to the predicted molecular weights of the two domains, 30 and 33.7 kDa. Lanes 3-4 show the pellet and soluble purified **TK2** corresponding to a predicted molecular weight of 72.8 kDa. [b] Lanes 1-2 insoluble, soluble fractions of **TK4**. Lanes 3-4 insoluble, soluble fractions of **TK5**. Lanes 5-6 insoluble, soluble fractions of **TK7**. Computed molecular weights for the TKs correspond with the major bands in the soluble fraction; **TK4** - 73 kDa, **TK5** - 71 kDa, **TK7** - 73 kDa.

Clarified lysate of the induced five induced TKs, Figure 2, was incubated with 50 mM of both glycoaldehyde and β -hydroxypyruvate for 30 mins. These initial experiments showed positive activity above background for three out of five TKs following the production of erythrose from the two substrates. Specific activities for the three enzymes were calculated using

this reaction data and protein concentrations calculated using a combination of the Bradford assay and densitometry of protein SDS gels. **TK1**, **TK2** and **TK7** had specific activities of 3.4, 4.8 and 1.2 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ respectively, comparable with the specific activity of an *E.coli* TK 0.65 $\mu\text{mol min}^{-1} \text{mg}^{-1}$.^[13] Of the three, **TK1** and **TK2** were successfully purified and stable and active enough to be characterized (Table 2).

Table 2. Purification and kinetic values for Transketolase

[a]																								
[b]																								
[c]																								
[d]	<table border="1"> <thead> <tr> <th>Enzyme</th> <th>V_{max} $\mu\text{katal.mg}^{-1}$</th> <th>K_m mM</th> <th>K_{cat} s^{-1}</th> <th>K_{cat}/K_m $\text{M}^{-1}.\text{s}^{-1}$</th> </tr> </thead> <tbody> <tr> <td>TK1</td> <td>0.13 ($\pm 3.8 \times 10^{-3}$)</td> <td>3.3 (± 0.47)</td> <td>213</td> <td>6.5×10^4</td> </tr> <tr> <td>TK2</td> <td>0.44 ($\pm 1.3 \times 10^{-2}$)</td> <td>6.5 (± 0.77)</td> <td>1269</td> <td>1.9×10^5</td> </tr> <tr> <td><i>E.coli</i>^[13]</td> <td>0.13</td> <td>35</td> <td>391</td> <td>1.1×10^4</td> </tr> </tbody> </table>				Enzyme	V_{max} $\mu\text{katal.mg}^{-1}$	K_m mM	K_{cat} s^{-1}	K_{cat}/K_m $\text{M}^{-1}.\text{s}^{-1}$	TK1	0.13 ($\pm 3.8 \times 10^{-3}$)	3.3 (± 0.47)	213	6.5×10^4	TK2	0.44 ($\pm 1.3 \times 10^{-2}$)	6.5 (± 0.77)	1269	1.9×10^5	<i>E.coli</i> ^[13]	0.13	35	391	1.1×10^4
Enzyme	V_{max} $\mu\text{katal.mg}^{-1}$	K_m mM	K_{cat} s^{-1}	K_{cat}/K_m $\text{M}^{-1}.\text{s}^{-1}$																				
TK1	0.13 ($\pm 3.8 \times 10^{-3}$)	3.3 (± 0.47)	213	6.5×10^4																				
TK2	0.44 ($\pm 1.3 \times 10^{-2}$)	6.5 (± 0.77)	1269	1.9×10^5																				
<i>E.coli</i> ^[13]	0.13	35	391	1.1×10^4																				

Characterization of 2 purified TKs following the production of erythrose from the reaction of β -hydroxy pyruvate with increasing concentrations of glycoaldehyde. All kinetics calculated based on the formation of erythrose [a] The reaction catalysed by the TK enzyme where a two carbon ketol unit is transferred from β -hydroxy pyruvate irreversibly to glycoaldehyde to produce erythrose and CO_2 . [b] Graph showing initial velocities for **TK1** plotted against substrate concentration fitted with the Michaelis Menton equation. [c] Graph showing initial velocities for **TK2** plotted against substrate concentration fitted with the Michaelis Menton equation. Biochemical characteristics for both enzymes calculated from the assay data compared to a purified *E.coli* TK. Errors are given in parenthesis.

Physical functional metagenomics libraries are a powerful tool for finding novel enzymatic activities however they can have a low "hit rate", due in part to the failure of host systems to transcribe and translate heterologous DNA. In our strategy DNA is first sequenced and analysed *in silico*, then single open reading frames or larger segments of DNA are amplified and can be placed in the correct orientation in front of a strong promoter/transcription system. While *E.coli* is a tractable and well-studied system, not all sequences retrieved from the metagenome will express well in this host. However because our method provides isolated DNA sequences they can be cloned into multiple expression vectors and expressed in different host systems much more easily than larger Fosmid or BAC libraries. Metagenomic clone libraries rely on visualizable outputs for enzyme identification; enzymes which are not readily visualized in some manner are unobtainable using physical strategies. Using our *in silico* approach all types of enzymes can be

identified before any physical manipulation of the DNA takes place. After PCR capture and cloning individual enzyme activity can be followed using the whole range of spectrophotometric and mass spectroscopy techniques as well as colorimetric assays. To date research in retrieving enzymes from metagenomes via PCR have relied upon degenerate primers from consensus sequences to retrieve enzymes. The research presented here uses primers designed for specific non redundant sequences from a specific sequenced metagenome in an effort to increase the number and variety of sequences retrieved. Furthermore using a sequence directed approach gives insight beyond any single open reading frame. The *in silico* approach may also aid in the characterization of an enzyme by providing contextual information about substrate specificity from the annotation of neighboring genes especially within operons.

The *in silico* library from which the enzymes were identified was created from a relatively small data set using technology that has now been surpassed in terms of base pair out-put and read quality. Sequencing of other environmental samples, at larger sequence depth with current technologies, and the creation of multiple *in silico* libraries would allow fast, productive access to the huge enzymatic diversity of chemically/biologically interesting niches. Having shown it is possible to identify and retrieve targeted active enzymes from a metagenomic sample using sequence directed, specific primers, the next step is to retrieve larger numbers of commercially relevant enzymes from multiple environmental samples.

Experimental Section

LDH and MDH Enzymes were assayed at varying concentration of substrate either sodium pyruvate or oxaloacetic acid (Sigma) respectively. NADH concentration was kept constant at 1 mM. The reaction was followed by the decrease in absorbance at 340 nm as NADH was oxidised to NAD⁺ concomitantly with reduction of the substrate. 80 µl of a 1.25 mM NADH solution initiated the reaction of 10 µl of enzyme with 10 µl of substrate, total reaction volume 100 µl. NADH and substrates were prepared in 0.5 M Hepes buffer pH 7.5. Assays were run in 96 well plates in triplicate at 25 °C, pH 7.5 using a fluostar Optima plate reader. Zero enzyme controls were set up as well as an NADH standard across a concentration range of 0.1–1 mM. Substrate range for LDH 1.0, 2.5, 5.0, 7.5, 10.0, 12.5, 15.0, 17.5, 20.0, 25.0, 30.0, 35.0 mM of pyruvate. Substrate range for MDH 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, 3.0, 5.0, 10.0, 15.0, 20.0 mM of oxaloacetic acid. An LDH enzyme concentration of 0.126 mg/ml and MDH concentration of 0.293 mg/ml was ascertained by absorbance at 280 nm. The molecular weight of 144.88 kDa and 78.3 kDa and extinction coefficients for the LDH homotetramer and MDH homodimer respectively, were calculated using the EXPASY protparam tool.

Initial experiments on the Ts were carried out with clarified lysate. 30 µl of lysate containing the induced Tk enzymes was incubated for 5 min with 25 µl of a stock solution of 28.8 mM Thiamine di phosphate (ThDP) and 108 mM MgCl₂. After 5 min 145 µl of 103 mM β-hydroxy pyruvate and 100 µl of 150 mM Glycoaldehyde was added to initiate the reactions. Triplicate reactions were allowed to proceed for 30 min at 25 °C when 50 µl of the reactions were taken and quenched in 250 µl of a 0.5 % trifluoroacetic acid (TFA) solution. Production of erythrose was followed by HPLC at 210 nm on a Dionex 500 system using a C18 Aminex column at 60 °C, mobile phase 0.1 % v/v TFA flow rate 0.6 ml min⁻¹, retention time for erythrose 11.40 min. Protein concentration of the lysate was

calculated using the Bradford assay and the TK contribution calculated using densitometry of protein SDS gels. Purified transketolase enzymes were assayed on a smaller scale than the lysate. 10 µl of purified enzyme was incubated for 5 min with 10 µl of 48 mM ThDP and 180 mM MgCl₂. After incubation for 5 min 90 µl of 110 mM β-hydroxy pyruvate and 90 µl of glycoaldehyde at a range of concentrations, 1–45 mM, was added to initiate the reaction. Reactions were run at 25 °C in triplicate for 5 min with 50 µl samples being taken and quenched in 250 µl of a 0.5% TFA solution. Production of erythrose was followed by HPLC as described above. Enzyme concentrations of 3.9 and 2.5 mg/ml for **TK1** and **TK2** respectively were calculated by absorbance at 280 nm. Molecular weight and extinction coefficients for the TKs were calculated using the EXPASY protparam tool, a combined molecular weight of 63.7 kDa for **TK1** and 72.8 kDa for **TK2** were computed.

Keywords: Functional Metagenomics • Polymerase Chain Reaction • Enzymes • Biocatalysis •

- [1] a) B. G. Davis, V. Boyer, *Nat. Prod. Rep.* **2001**, *18*, 618–640; b) A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, B. Witholt, *Nature* **2001**, *409*, 258–268. c) M. T. Reetz, *J. Am. Chem. Soc.* **2013**, *135*, 12480–12496.
- [2] a) U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–194; b) J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, et al., *Science* **2010**, *329*, 309–13; c) J. Strafford, P. Payongsri, E. G. Hibbert, P. Morris, S. S. Batth, D. Steadman, M. E. B. Smith, J. M. Ward, H. C. Hailes, P. A. Dalby, *J. Biotechnol.* **2012**, *157*, 237–245.
- [3] a) T. Furuya, K. Kino, *Appl. Microbiol. Biotechnol.* **2010**, *86*, 991–1002; b) J. Yin, P. D. Straight, S. Hrvatin, P. C. Dorrestein, S. B. Bumpus, C. Jao, N. L. Kelleher, R. Kolter, C. T. Walsh, *Chem. Biol.* **2007**, *14*, 303–312; c) C. Vergne-Vaxelaire, F. Bordier, A. Fossey, M. Besnard-Gonnet, A. Debar, A. Mariage, V. Pellouin, A. Perret, J. L. Petit, M. Stam, et al., *Adv. Synth. Catal.* **2013**, *355*, 1763–1779.
- [4] a) J. Handelsman, M. R. Rondon, S. F. Brady, J. Clardy, R. M. Goodman, *Chem. Biol.* **1998**, *5*, R245–R249; b) V. Torsvik, L. Øvreås, T. F. Thingstad, *Science* **2002**, *296*, 1064–6.
- [5] a) J.-S. Gong, Z.-M. Lu, H. Li, Z.-M. Zhou, J.-S. Shi, Z.-H. Xu, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 6603–11; b) H. A. Iqbal, Z. Feng, S. F. Brady, *Curr. Opin. Chem. Biol.* **2012**, *16*, 109–116; c) J. J. Banik, S. F. Brady, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17273–17277.
- [6] a) P. D. Schloss, J. Handelsman, *Curr. Opin. Biotechnol.* **2003**, *14*, 303–310; b) B. Temperton, S. J. Giovannoni, *Curr. Opin. Microbiol.* **2012**, *15*, 605–612.
- [7] a) Y. F. Yeh, S. C. Y. Chang, H. W. Kuo, C. G. Tong, S. M. Yu, T. H. D. Ho, *Gene* **2013**, *519*, 360–366; b) M. Allgaier, A. Reddy, J. I. Park, N. Ivanova, P. D'Haeseleer, S. Lowry, R. Sapra, T. C. Hazen, B. A. Simmons, J. S. Vanderghaynst, P. Hugenholtz, *PLoS One* **2010**, *5*, e8812.
- [8] a) E. M. Gabor, W. B. L. Alkema, D. B. Janssen, *Environ. Microbiol.* **2004**, *6*, 879–86; b) J. Handelsman, *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 669–685; c) G. Fuchu, Y. Ohtsubo, M. Ito, R. Miyazaki, A. Ono, Y. Nagata, M. Tsuda, *Appl. Microbiol. Biotechnol.* **2008**, *79*, 627–632; d) S. Bayer, C. Birkemeyer, M. Ballschmitter, *Appl. Microbiol. Biotechnol.* **2011**, *89*, 91–98.
- [9] a) J. Shaeri, R. Wohlgemuth, J. M. Woodley, *Org. Process Res. Dev.* **2006**, *10*, 605–610. b) J. Shaeri, I. Wright, E. B. Rathbone, R. Wohlgemuth, J. M. Woodley, *Biotechnol. Bioeng.* **2008**, *101*, 761–767.
- [10] A. Feldman-Salit, S. Hering, H. L. Messiha, N. Veith, V. Cojocar, A. Sieg, H. V. Westerhoff, B. Kreikemeyer, R. C. Wade, T. Fiedler, *J. Biol. Chem.* **2013**, *288*, 21295–306.

- [11] D. Mikulášová, M. Kollárová, M. Miginiac-Maslow, P. Decottignies, J.-P. Jacquot, E. Kutejová, N. Mernik, I. Egyudová, R. Musrati, T. Horecká, *FEMS Microbiol. Lett.* **1998**, *159*, 299–305.
- [12] a) A T. Brown, C. P. Christian, R. L. Eifert, *J. Bacteriol.* **1975**, *122*, 1126–1135; b) Y.-D. Ge, Z.-Y. Cao, Z.-D. Wang, L.-L. Chen, Y.-M. Zhu, G.-P. Zhu, *Biosci. Biotechnol. Biochem.* **2010**, *74*, 2194–2201.
- [13] E. G. Hibbert, T. Senussi, S. J. Costelloe, W. Lei, M. E. B. Smith, J. M. Ward, H. C. Hailes, P. a. Dalby, *J. Biotechnol.* **2007**, *131*, 425–432.

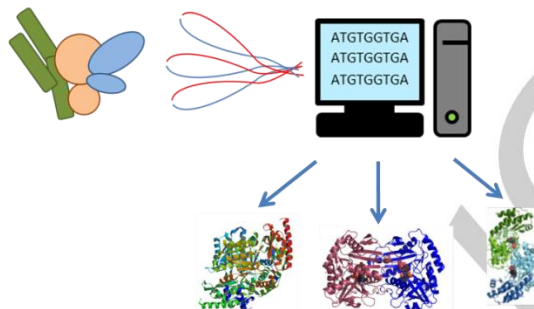
WILEY-VCH

Entry for the Table of Contents (Please choose one layout)

Layout 1:

COMMUNICATION

A combination of high throughput sequencing and bioinformatic tools are used to generate an *in silico* metagenomic library. Such a library can grant fast, productive access to the wealth of enzyme diversity promised by metagenomics.



*Author(s), Corresponding Author(s)**

Page No. – Page No.

Title

Layout 2:

COMMUNICATION

*Author(s), Corresponding Author(s)**

Page No. – Page No.

Title

Text for Table of Contents