Title
Rapidly moving new bacteria to model-organism status

Authors
Liu, Hualan
Deutschbauer, Adam M

Publication Date
2018-06-01

DOI
10.1016/j.copbio.2017.12.006

Peer reviewed

Rapidly moving new bacteria to model-organism status

Hualan Liu and Adam M. Deutschbauer[1]

Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, MS 977-250, Berkeley, CA 94720

[1]To whom correspondence should be addressed: AMDeutschbauer@lbl.gov
P: 510-486-5473

Abstract

The paradigm of large research communities collectively working on a small number of model bacteria such as *Escherichia coli* and *Bacillus subtilis* is changing. While these classic model bacteria will continue to be important for advanced systems biology and new technology development, we envision that increasingly small research teams will be deeply investigating their own favorite strains, for example as new hosts for metabolic engineering or as key members of a complex microbiome. Given the lack of a research community and the sheer number of possible bacteria to interrogate, the development and application of technologies to rapidly and inexpensively advance these unstudied strains to "model-organism" status is imperative. Here, we discuss the minimal information and tools necessary to develop a new model bacterium and how existing approaches can bring this power into the hands of a single investigator.

Introduction

Microbiology-based solutions have been proposed for many of our most pressing planetary challenges including human health, sustainable agriculture, biomanufacturing, and environmental stewardship [1]. Meeting these challenges will require the molecular genetic investigation of a significant number of bacteria for which no literature exists, for example the newly isolated members of a complex microbiome [2] or the development a new autotrophic host for metabolic engineering of advanced chemicals [3]. The investigation of a new bacterium poses unique challenges, the most obvious being the lack of a research community with available genetic tools, strain collections, and a wealth of accumulated knowledge, such as those that have been developed in *E. coli* and *B. subtilis*. Therefore, translating the potential of these molecularly unstudied bacteria into applications will be hindered unless we accelerate the development of new model bacteria.

In this short review, we present some of the minimum criteria necessary to move a new bacterium to "model-organism" status: (1) An accurate parts-list of genes, proteins, and promoters, (2) a genetic toolbox, (3) accurate, data-driven gene annotations, and (4) computational platforms for data integration and systems-level analyses (**Figure 1**). We describe existing technologies and resources available to meet each of these four criteria with a focus on approaches that can be applied by a single investigator rapidly and at a relatively low cost.

An accurate genomic parts-list

A complete and accurately assembled genome, with precise annotation of the structure and function of its features, is a crucial foundation for the study of any bacterium. The development of next-generation sequencing (NGS) technologies has made bacterial genome sequencing routine. Though these assemblies are often unclosed draft genomes, due to the short read lengths of the Illumina sequencing platform and repeat elements in bacterial genomes, third-generation sequencing technologies that produce longer read lengths, such as PacBio [4], should help to seal the gaps and generate complete genome sequences. This hybrid strategy combining both short-read and long-read sequencing has shown its power for high quality genome assembly [5]. Despite its higher error rate relative to Illumina sequencing, PacBio sequencing data has improved to the point that complete bacterial genome assemblies can be achieved using this platform alone [6], although the cost is still significantly higher than Illumina sequencing. An added benefit of the PacBio platform is the utility of these data to identify methylated bases and thus define all methylated motifs (the methylome) in the genome [7]. In a recent study, the methylome of 230 bacteria and archaea was determined using PacBio sequencing data, and revealed 834 different methylated motifs [8]. As described below, there are downstream benefits to both a complete genome sequence (e.g. ~~for example~~, for normalization of transposon site sequencing data) and identification of methylated motifs (for designing strategies to increase transformation efficiency), therefore we recommend generating PacBio data for any bacteria being advanced to model-organism status.

There are a number of annotation pipelines for calling genes and assigning putative functions including the Integrated Microbial Genome (IMG) [9] and the Rapid Annotation using Subsystem Technology (RAST) systems [10]. Nevertheless, the accuracy of the gene calls and the protein function predictions are less than ideal, in particular for a strain under more in-depth investigation. Furthermore, these annotation servers do not typically identify gene structures beyond the coding sequence, leaving critical information such as promoters, transcriptional start sites (TSSs) antisense transcription, and small regulatory RNAs (sRNA) unknown. Now well-established transcriptomic and proteomic approaches can be applied at low cost to rapidly define the genomic parts-list of an organism. Differential RNA-seq (dRNA-seq) has become the preferred approach for precise mapping of 5’ TSSs in bacterial genomes as this approach can differentiate between primary and degraded/processed transcripts [11]. dRNA-seq requires the preparation and sequencing of two libraries, one treated with terminator exonuclease and one without, and has been used to identify thousands of TSSs in *E. coli* including over 5,000 antisense transcripts [12]. Putative sRNAs can ~~be~~ also be readily mapped using RNA-seq as illustrated in *Synechocystis* sp. PCC 6803 [13] and *Acinetobacter baumannii [14]*.

Another established approach, termed proteogenomics, uses mass spectrometry-based proteomic data, to search against pre-existing protein databases, to identify novel proteins missed by the primary annotation and to modify gene models [15]. Application of this approach to *Mycobacterium smegmatis* validated hundreds of predicted gene models and identified 63 new proteins [16]. Proteogenomics can be coupled to transcriptomics to further refine gene models. For example, in the sulfate-reducing bacterium *Desulfovibrio vulgaris*, a combination of proteomics, tiling microarrays, and RNA-seq was used to identify over 1,000 transcriptional start sites and revise 505 protein annotations including 127 proteins that had been missed by the original genome annotation [17].

Genetic tools and strain collections

The ability to modify the genome is essential for hypothesis testing and for the development of any new bacterial model system. Numerous such approaches have been developed over the years using random strategies (i.e. transposon mutagenesis) or targeted strategies such as gene deletions, CRISPR-based genome editing, and recombineering. In recombineering, linear DNA substrates are introduced into the host cell, which expresses a bacteriophage recombination system [18]. Only very short homology regions are needed for the recombination event. Recombineering is very efficient for site-specific mutagenesis, in-frame gene deletions, and gene replacements [19]. By mining for species-specific phage recombination proteins, a number of groups have adapted recombineering to new systems including *Photorhabdus luminescens* [20] and *Vibrio natriegens* [21].

The CRISPR-Cas9 system has attracted a lot of attention as a tool for precise and sequence-specific genome editing in eukaryotes, but its application as a genome editing tool in bacteria has been limited due to toxicity issues of Cas9 and/or the relative inefficiency of the nonhomologous end joining pathway in bacteria. Nevertheless, there are a number of recent advances in using CRISPR technology to engineer bacterial genomes and interrogate bacterial physiology ~~{Selle:2015dj}~~[22]. Here, we just highlight two such approaches, and how they have been applied to multiple, genetically diverse bacteria. The first is coupling recombineering with the use of the endonuclease Cpf1 as an alternative to Cas9. Cpf1 is an RNA-guided endonuclease that is both smaller and simpler compared to Cas9, only requiring a single guide RNA for genome targeting. This hybrid approach coupling Cpf1-CRISPR and recombineering has recently been used to edit the genomes of a phylogenetically diverse range of bacteria [23,24]~~[22,23]~~. The second is the use of a catalytically inactive Cas9 to regulate gene expression of target gene(s), typically for repression using CRISPR interference (CRISPRi) [25-27]~~[24-26]~~. In a recent example, CRISPRi was applied to the cyanobacterium *Synechococcus* sp. PCC 7002 to redirect central carbon metabolic flux to increased production of lactate [28] ~~[27]~~.

For any new bacterium, the rapid establishment of genetic systems and genetic parts for engineering is crucial, and recent advances will aid in the streamlining of genetic tool development. One potential roadblock is bacterial restriction modification (RM) systems that can degrade foreign DNA, thereby reducing transformation efficiency. To overcome RM systems, multiple groups have expressed methyltransferases from the target bacterium in *E. coli* to ~~"~~"mimic~~"~~" the host methylation pattern, thereby increasing the transformation efficiency by orders of magnitude [29,30]~~[28,29]~~. A second challenge is that given the myriad of DNA delivery methods, genetic parts, and genetic systems available for strain engineering, the individual testing of all of these parameters can be tedious, time-consuming, and prohibitively costly. Therefore, methods to parallelize the development and testing of genetic parts and methods is crucial to accelerate the development of new model bacteria. To streamline the optimization of DNA delivery by electroporation, a microfluidic-based approach was used to optimize electroporation parameters in three~~3~~ diverse bacteria [31]~~[30]~~. A number of synthetic biology efforts have described high-throughput approaches for characterizing new genetic parts. For example, in the commensal human gut bacterium *Bacteroides thetaiotaomicron*, multiple groups have taken a systematic and high-throughput approach to develop a collection of characterized genetic parts including inducible and constitutive promoters, RBSs, and reporter genes. These tools have been used to modulate the behavior of *B. thetaiotaomicron* in the mouse gut [32]~~[31]~~, image the spatial distribution of several species of *Bacteroides* in the gut [33]~~[32]~~, and regulate the activity of bacterial sialidase expression using a small molecule inducer introduced in

drinking water [34][33]. Lastly, a ‘“magic pool”’ strategy has recently been developed enabling the parallel testing of hundreds of transposon vector constructs in parallel against a target bacterium [35][34]. This approach involves the use of DNA barcodes to mark each transposon vector design, which can then be read-out using next generation sequencing to identify the optimal genetic system. Using this magic pool method, whole genome transposon mutant libraries were rapidly constructed in five5 different genera of bacteria.

Part of the attraction of established model bacteria is the availability of comprehensive strain collections for genome-wide screening or targeted hypothesis testing, for example the gene deletion collections in *E. coli* [36][35] and *B. subtilis* [37][36]. While the resources required for generating such collections are typically outside the capacity of a single researcher, there are alternative approaches including the archiving of transposon mutant collections [38][37]. Traditionally, the bottleneck in creating a large transposon mutant library has been the cost associated with mapping all of the insertion locations for each strain. One solution to this problem is the large-scale rearraying of transposon mutant strains in a microplate format followed by a ‘“smart-pooling”’ strategy to map the identity of all of the strains using next-generation sequencing. This strategy has been applied to a number of bacteria using *B. thetaiotaomicron* [39][38] and *S. oneidensis* [40][39]. While this workflow is greatly facilitated by liquid handling automation, we find that these machines are often available to individual researchers, for example in shared core facilities.

Accurate gene function annotation with functional genomics

Many computational-based gene annotations are incorrect or misleading, particularly when the protein of interest lacks close homology to an experimentally studied relative [41][40]. While we currently cannot achieve the knowledge of a single model bacterium investigated for years by an entire research community, it is possible to make substantial progress on gene function annotation using functional genomics for an entirely unstudied isolate bacterium. While a number of approaches have been developed to tackle the challenge of gene function determination including high-throughput enzymology coupled to metabolomics [42][41], the currently most scalable approaches are those based on high-throughput genetics, in particular the large-scale phenotyping of comprehensive mutant libraries, either as archived single strains [43][42] or in parallel using competitive pooled fitness assays [44][43]. Currently, transposon site sequencing (TnSeq) [45][44] and its many variants [46,47][45,46] are the most commonly used approaches for measuring the phenotypes of thousands of genes in parallel using next-generation sequencing. TnSeq has been applied to a wide range of bacteria and for a number of purposes including the identification of essential genes in cyanobacteria [48][47] and phage sensitivity determinants in *Caulobacter crescentus* [49][48].

Inherent in the above genetics strategies is that the phenotype of a mutation provides insight into the function of the gene. If one measures the phenotypes of genes across a wide range of conditions, then the functions of unknown genes can be inferred by the similarity of their phenotypes to genes of known function [43,44][42,43]. To accelerate the assaying of mutant phenotypes, a recent alteration to transposon site sequencing was developed, random barcode TnSeq or RB-TnSeq [47][46], which simplifies the relative measurement of mutant abundance using DNA barcode sequencing. To demonstrate the scalability of this approach, RB-TnSeq was applied to 25 bacteria and thousands of genome-wide fitness assays were performed [50][49]. Using these data, phenotypes for thousands of previously hypothetical proteins were identified, many of which had phenotype profiles that matched those of a gene with a known function.

CRISPR-Cas9 based approaches for large-scale genetics offer some advantages over TnSeq-like approaches including user-guided sequence specificity and the ability to interrogate the phenotypes of essential genes in more depth. In a recent study, CRISPRi was performed against all essential genes in *B. subtilis* and the impact of transcriptional repression of each essential gene was individually assessed by measuring cellular morphology and growth in the presence of different small molecule inhibitors [51][50]. Using these data, the mode-of-action of previously uncharacterized antibiotics could be inferred and new linkages between cell morphology and fitness could be uncovered. CRISPRi also has potential as a high-throughput tool for interrogating the impact of transcriptional repression in bacteria, as recently illustrated in a pooled, competitive growth format in *E. coli* [52][51].

As genetics are applied to an ever increasing number of bacteria, the ability to phenotype these mutants (either as single strains or pooled libraries) will become the bottleneck. While platforms such as phenotype microarrays are powerful for interrogating the fitness of single strains against many growth conditions [53][52], this approach has not been currently scaled genome-wide. In addition, relatively simple growth phenotypes do not capture the environmental complexity of bacteria in their native ecosystems, for example in a plant-bacterium interaction. To illustrate the importance of assaying "natural" conditions, Barczak and colleagues used high-throughput microscopy and automated image analysis to assay the survival of ~26,000 individually arrayed transposon mutants of the human pathogen *Mycobacterium tuberculosis* within macrophages [54][53]. Using these data, they were able to predict and validate a functional relationship between two pathogenicity genes, as well as identify two previously uncharacterized genes who role in infection whose precise roles remain to be elucidated. To extend image analysis of strain libraries to additional bacteria, a new fully automated platform, the Strain Library Imaging Protocol (SLIP), was recently developed for high-throughput single-cell microscopy with arrayed mutant libraries [55][54]. Lastly, microfluidic approaches to encapsulate millions of single cells for high-throughput phenotypic analysis hold great promise, in particular for "*trans* acting" traits that are masked in pooled assays with thousands of mutants, such as metabolite exchange in a small microbial community [56][55].

Data management, analysis, and systems biology

Bacterium-specific databases such as those developed for *E. coli* [57][56], *C. crescentus* [58][57], and *B. subtilis* [59][58] are powerful resources for sharing knowledge and interpreting new datasets, however other strategies must be employed for a new bacterium under investigation. Thankfully, a number of established and developing systems are available for any bacterium that incorporate both genome sequences and functional data for systems-level analyses. The Pathway Tools software package within the Biocyc databases can be used to construct metabolic models overlaid with gene expression and metabolomics data [60][59]. The Department of Energy's Systems Biology Knowledgebase (KBase) can also generate metabolic models from a bacterial genome sequence using Model SEED [61][60], with the added functionality of reconciling the draft metabolic models with growth and genetics data [62][61]. Lastly, functional data including transcriptomics and proteomics can be analyzed within the Integrated Microbial Genomes (IMG) system [9].

Given the increasing scale of DNA sequence and functional genomic data, we envision that most individual researchers without access to large compute power will analyze their own data using cloud computing-based systems such as KBase or Galaxy [63][62]. Beyond computing power, there are a number of added benefits of integrated, sharable platforms for computational

data analyses including archived code for reproducible research [62,64][~~61,63~~], integration of diverse datatypes for systems biology, and the ability to perform comparative analyses across multiple bacteria. Therefore, while comprehensive single organism databases are probably not within reach to the single investigator, ultimately it will not be crucial to have these for the multitude of new bacteria under molecular investigation.

Outlook

In this short review, we highlight some of the key attributes that define a ~~"~~model bacterium~~"~~ and how current approaches can be rapidly applied to advance new bacteria to this status. For the methods discussed here, we view the rapid and streamlined development of genetic systems to virtually any new bacterium as a significant obstacle to be overcome. In addition, we envision that additional technological advances, including metabolomics, high-throughput biochemistry, non-growth phenotyping, and interaction mapping (genetic and protein) will ultimately be expended to a wider range of bacteria, especially if some of these approaches can be coupled to next-generation DNA sequencing. We predict that most of the approaches described in this review will be applied as routinely to a new bacterium as genome sequencing is today and that the data and experimental tools derived from such a concerted effort by the microbiology community will accelerate the usage of bacteria for a number of societal benefits.

Conflict of interest

~~None~~Nothing declared.

Acknowledgements

We thank Morgan Price for helpful comments. This material by ENIGMA – Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://engima.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231.

References and recommended reading

1. Blaser MJ, Cardon ZG, Cho MK, Dangl JL, Donohue TJ, Green JL, Knight R, Maxon ME, Northen TR, Pollard KS, et al.: **Toward a pPredictive uUnderstanding of Earth's mMicrobiomes to aAddress 21st cCentury cChallenges.** *MBio* 2016, **7**:e00714–16.

2. Bai Y, Müller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, Dombrowski N, Münch PC, Spaepen S, Remus-Emsermann M, et al.: **Functional overlap of the Arabidopsis leaf and root microbiota.** *Nature* 2015, **528**:364–369.

3. Claassens NJ, Sousa DZ, Santos Dos VAPM, de Vos WM, van der Oost J: **Harnessing**

the power of microbial autotrophy. *Nat Rev Microbiol* 2016, **14**:692–706.

4. Rhoads A, Au KF: **PacBio sSequencing and iIts aApplications.** *Genomics Proteomics Bioinformatics* 2015, **13**:278–289.

5. Wick RR, Judd LM, Gorrie CL, Holt KE: **Unicycler: rResolving bacterial genome assemblies from short and long sequencing reads.** *PLoS Comput Biol* 2017, **13**:e1005595.

6. Ray J, Waters RJ, Skerker JM, Kuehl JV, Price MN, Huang J, Chakraborty R, Arkin AP, Deutschbauer A: **Complete Ggenome Ssequence of *Cupriavidus basilensis* 4G11, iIsolated from the Oak Ridge Field Research Center sSite.** *Genome Announc* 2015, **3**:e00322–15.

7. Pirone-Davies C, Hoffmann M, Roberts RJ, Muruvanda T, Timme RE, Strain E, Luo Y, Payne J, Luong K, Song Y, et al.: **Genome-wide methylation patterns in *Salmonella enterica* Subsp. *enterica* sSerovars.** *PLoS One* 2015, **10**:e0123639.

8. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, Froula J, Kang DD, Malmstrom RR, Morgan RD, et al.: **The eEpigenomic lLandscape of pProkaryotes.** *PLoS Genet*. 2016, **12**:e1005854.

9. Chen I-M.A, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, Ratner A, Huang J, Andersen E, Huntemann M, et al.: **IMG/M: integrated genome and metagenome comparative data analysis system.** *Nucleic Acids Res*. 2017, **45**:D507–D516.

10. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, et al.: **The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST).** *Nucleic Acids Res*. 2014, **42**:D206–14.

11. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, et al.: **The primary transcriptome of the major human pathogen *Helicobacter pylori*.** *Nature* 2010, **464**:250–255.

12. Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM, Storz G: **Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*.** *J. Bacteriol*. 2015, **197**:18–28.

13. Xu W, Chen H, He C-L, Wang Q: **Deep sequencing-based identification of small regulatory RNAs in sSynechocystis sp. PCC 6803.** *PLoS One* 2014, **9**:e92711.

14. Weiss A, Broach WH, Lee MC, Shaw LN: **Towards the complete small RNome of *Acinetobacter baumannii*.** *Microb Genom* 2016, **2**:e000045.

15. Nesvizhskii AI: **Proteogenomics: concepts, applications and computational strategies.** *Nat Methods* 2014, **11**:1114–1125.

16. Potgieter MG, Nakedi KC, Ambler JM, Nel AJM, Garnett S, Soares NC, Mulder N, Blackburn JM: **Proteogenomic aAnalysis of *Mycobacterium smegmatis* uUsing**

hHigh rResolution mMass sSpectrometry. *Front Microbiol* 2016, **7**:427.

17. Price MN, Deutschbauer AM, Kuehl JV, Liu H, Witkowska HE, Arkin AP: **Evidence-based annotation of transcripts and proteins in the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough.** *J. Bacteriol.* 2011, **193**:5716–5727.

18. Ellis HM, Yu D, DiTizio T, Court DL: **High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides.** *Proc. Natl. Acad. Sci. U. S. A.* 2001, **98**:6742–6746.

19. Pines G, Freed EF, Winkler JD, Gill RT: **Bacterial rRecombineering: gGenome eEngineering via pPhage-bBased hHomologous rRecombination.** *ACS Synth Biol* 2015, **4**:1176–1185.

20. Yin J, Zhu H, Xia L, Ding X, Hoffmann T, Hoffmann M, Bian X, Müller R, Fu J, Stewart AF, et al.: **A new recombineering system for Photorhabdus and Xenorhabdus.** *Nucleic Acids Res.* 2015, **43**:e36–e36.

21. Lee HH, Ostrov N, Gold MA, Church GM: **Recombineering in *Vibrio natriegens*.** *bioRxiv* 2017, doi:10.1101/130088.

22. Selle K, Barrangou R: **Harnessing CRISPR-Cas systems for bacterial genome editing.** *Trends Microbiol* 2015, **23**:225–232.

23. Jiang Y, Qian F, Yang J, Liu Y, Dong F, Xu C, Sun B, Chen B, Xu X, Li Y, et al.: **CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*.** *Nat Commun* 2017, **8**:15179.

24. Yan M-Y, Yan H-Q, Ren G-X, Zhao J-P, Guo X-P, Sun Y-C: **CRISPR-Cas12a-aAssisted rRecombineering in bBacteria.** *Appl Environ Microbiol* 2017, **83**:e00947–17.

25. Yao L, Cengic I, Anfelt J, Hudson EP: **Multiple gGene rRepression in Cyanobacteria uUsing CRISPRi.** *ACS Synth Biol* 2016, **5**:207–212.

26. Choudhary E, Thakur P, Pareek M, Agarwal N: **Gene silencing by CRISPR interference in mycobacteria.** *Nat Commun* 2015, **6**:6267.

27. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA: **Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.** *Cell* 2013, **152**:1173–1183.

28. Gordon GC, Korosh TC, Cameron JC, Markley AL, Begemann MB, Pfleger BF: **CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the *Cc*yanobacterium *Synechococcus* sp. strain PCC 7002.** *Metab. Eng.* 2016, **38**:170–179.

29. Monk IR, Tree JJ, Howden BP, Stinear TP, Foster TJ: **Complete bBypass of rRestriction sSystems for mMajor *Staphylococcus aureus* lLineages.** *MBio* 2015, **6**:e00308–15.

30. Zhang G, Wang W, Deng A, Sun Z, Zhang Y, Liang Y, Che Y, Wen T: **A mimicking-of-**

DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria. *PLoS Genet.* 2012, **8**:e1002987.

31. Garcia PA, Ge Z, Moran JL, Buie CR: **Microfluidic screening of electric fields for electroporation.** *Sci Rep* 2016, **6**:21238.

32. Mimee M, Tucker AC, Voigt CA, Lu TK: **Programming a human commensal bacterium, *Bacteroides thetaiotaomicron*, to sense and respond to stimuli in the murine gut microbiota.** *Cell Syst* 2015, **1**:62–71.

33. Whitaker WR, Shepherd ES, Sonnenburg JL: **Tunable expression tools enable single-cell strain distinction in the gut microbiome.** *Cell* 2017, **169**:538–546.e12.

34. Lim B, Zimmermann M, Barry NA, Goodman AL: **Engineered regulatory systems modulate gene expression of human commensals in the gut.** *Cell* 2017, **169**:547–558.e15.

35. Liu H, Price M, Waters RJ, Ray J, Carlson HK, Lamson JS, Chakraborty R, Arkin AP, Deutschbauer AM: **Magic pools: parallel assessment of transposon delivery vectors in bacteria**. *bioRxiv* 2017, doi:10.1101/158840.

36. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: **Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection**. *Mol. Syst. Biol.* 2006, **2**:2006 0008.

37. Koo B-M, Kritikos G, Farelli JD, Todor H, Tong K, Kimsey H, Wapinski I, Galardini M, Cabal A, Peters JM, et al.: **Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*.** *Cell Syst* 2017, **4**:291–305.e7.

38. Ramage B, Erolin R, Held K, Gasper J, Weiss E, Brittnacher M, Gallagher L, Manoil C: **Comprehensive arrayed transposon mutant library of *Klebsiella pneumoniae* outbreak strain KPNIH1.** *J. Bacteriol.* 2017, doi:10.1128/JB.00352-17.

39. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI: **Identifying genetic determinants needed to establish a human gut symbiont in its habitat.** *Cell Host Microbe* 2009, **6**:279–289.

40. Baym M, Shaket L, Anzai IA, Adesina O, Barstow B: **Rapid construction of a whole-genome transposon insertion collection for *Shewanella oneidensis* by Knockout Sudoku.** *Nat Commun* 2016, **7**:13270.

41. Schnoes AM, Brown SD, Dodevski I, Babbitt PC: **Annotation error in public databases: misannotation of molecular function in enzyme superfamilies.** *PLoS Comput Biol* 2009, **5**:e1000605.

42. Sévin DC, Fuhrer T, Zamboni N, Sauer U: **Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in *Escherichia coli*.** *Nat Methods* 2017, **14**:187–194.

43. Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R, Lee S, Kazmierczak KM, Lee

KJ, Wong A, et al.: **Phenotypic landscape of a bacterial cell**. *Cell* 2011, **144**:143–156.

44. Deutschbauer A, Price MN, Wetmore KM, Shao W, Baumohl JK, Xu Z, Nguyen M, Tamse R, Davis RW, Arkin AP: **Evidence-based annotation of gene function in *Shewanella oneidensis* MR-1 using genome-wide fitness profiling across 121 conditions.** *PLoS Genet* 2011, **7**:e1002385.

45. van Opijnen T, Bodi KL, Camilli A: **Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms**. *Nat Methods* 2009, **6**:767–772.

46. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, et al.: **Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants**. *Genome Res* 2009, **19**:2308–2316.

47. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, et al.: **Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons.** *MBio* 2015, **6**:e00306–15.

48. Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP, Deutschbauer A, Golden SS: **The essential gene set of a photosynthetic organism.** *Proc Natl Acad Sci U S A* 2015, **112**:E6634–43.

49. Christen M, Beusch C, Bösch Y, Cerletti D, Flores-Tinoco CE, Del Medico L, Tschan F, Christen B: **Quantitative sSelection aAnalysis of bBacteriophage φCbK sSusceptibility in *Caulobacter crescentus*.** *J Mol Biol* 2016, **428**:419–430.

50. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, Esquivel Z, et al.: **Deep aAnnotation of pProtein fFunction across dDiverse bBacteria from mMutant pPhenotypes**. *bioRxiv* 2016, doi:10.1101/072470.

51. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, Hawkins JS, Lu CHS, Koo B-M, Marta E, et al.: **A cComprehensive, CRISPR-based fFunctional aAnalysis of eEssential gGenes in bBacteria.** *Cell* 2016, **165**:1493–1506.

52. Wang T, Guo J, Guan C, Wu Y, Liu B, Xie Z, Zhang C, Xing X-H: **Pooled CRISPR interference screens enable high-throughput functional genomics study and elucidate new rules for guide RNA library design in *Escherichia coli***. *bioRxiv* 2017, doi:10.1101/129668.

53. Lee WC, Goh KL, Loke MF, Vadivelu J: **Elucidation of the mMetabolic nNetwork of *Helicobacter pylori* J99 and Malaysian cClinical sStrains by pPhenotype mMicroarray.** *Helicobacter* 2017, **22**:e12321.

54. Barczak AK, Avraham R, Singh S, Luo SS, Zhang WR, Bray M-A, Hinman AE, Thompson M, Nietupski RM, Golas A, et al.: **Systematic, multiparametric analysis of *Mycobacterium tuberculosis* intracellular infection offers insight into coordinated virulence.** *PLoS Pathog* 2017, **13**:e1006363.

55. Shi H, Colavin A, Lee TK, Huang KC: **Strain lLibrary iImaging pProtocol for high-throughput, automated single-cell microscopy of large bacterial collections**

arrayed on multiwell plates. *Nat Protoc* 2017, **12**:429–438.

56. Terekhov SS, Smirnov IV, Stepanova AV, Bobik TV, Mokrushina YA, Ponomarenko NA, Belogurov AA, Rubtsova MP, Kartseva OV, Gomzikova MO, et al.: **Microfluidic droplet platform for ultrahigh-throughput single-cell screening of biodiversity.** *Proc Natl Acad Sci U S A* 2017, **114**:2550–2555.

57. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, et al.: **The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12.** *Nucleic Acids Res.* 2017, **45**:D543–D550.

58. Lasker K, Schrader JM, Men Y, Marshik T, Dill DL, McAdams HH, Shapiro L: **CauloBrowser: A systems biology resource for Caulobacter crescentus.** *Nucleic Acids Res* 2016, **44**:D640–5.

59. Michna RH, Zhu B, Mäder U, Stülke J: **SubtiWiki 2.0--an integrated database for the model organism *Bacillus subtilis*.** *Nucleic Acids Res* 2016, **44**:D654–62.

60. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, et al.: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases.** *Nucleic Acids Res* 2016, **44**:D471–80.

61. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: **High-throughput generation, optimization and analysis of genome-scale metabolic models.** *Nat Biotechnol* 2010, **28**:977–982.

62. Arkin AP, Stevens RL, Cottingham RW, Maslov S, Henry CS, Dehal P, Ware D, Perez F, Harris NL, Canon S, et al.: **The DOE Systems Biology Knowledgebase (KBase)**. *bioRxiv* 2016, doi:10.1101/096354.

63. Goecks J, Nekrutenko A, Taylor J, Galaxy Team: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol* 2010, **11**:R86.

64. Grüning BA, Rasche E, Rebolledo-Jaramillo B, Eberhard C, Houwaart T, Chilton J, Coraor N, Backofen R, Taylor J, Nekrutenko A: **Jupyter and Galaxy: eEasing entry barriers into complex data analyses for biomedical researchers.** *PLoS Comput Biol* 2017, **13**:e1005425.

~~1. Blaser MJ, Cardon ZG, Cho MK, Dangl JL, Donohue TJ, Green JL, Knight R, Maxon ME, Northen TR, Pollard KS, et al.: **Toward a Predictive Understanding of Earth's Microbiomes to Address 21st Century Challenges.** *MBio* 2016, **7**:e00714–16.~~

~~2. Bai Y, Müller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, Dombrowski N, Münch PC, Spaepen S, Remus-Emsermann M, et al.: **Functional overlap of the Arabidopsis leaf and root microbiota.** *Nature* 2015, **528**:364–369.~~

~~3. Claassens NJ, Sousa DZ, Santos Dos VAPM, de Vos WM, van der Oost J: **Harnessing the power of microbial autotrophy.** *Nat Rev Microbiol* 2016, **14**:692–706.~~

4. Rhoads A, Au KF: **PacBio Sequencing and Its Applications.** *Genomics Proteomics Bioinformatics* 2015, **13**:278–289.

5. Wick RR, Judd LM, Gorrie CL, Holt KE: **Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads.** *PLoS Comput Biol* 2017, **13**:e1005595.

6. Ray J, Waters RJ, Skerker JM, Kuehl JV, Price MN, Huang J, Chakraborty R, Arkin AP, Deutschbauer A: **Complete Genome Sequence of Cupriavidus basilensis 4G11, Isolated from the Oak Ridge Field Research Center Site.** *Genome Announc* 2015, **3**:e00322–15.

7. Pirone-Davies C, Hoffmann M, Roberts RJ, Muruvanda T, Timme RE, Strain E, Luo Y, Payne J, Luong K, Song Y, et al.: **Genome-wide methylation patterns in Salmonella enterica Subsp. enterica Serovars.** *PLoS One* 2015, **10**:e0123639.

8. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, Froula J, Kang DD, Malmstrom RR, Morgan RD, et al.: **The Epigenomic Landscape of Prokaryotes.** *PLoS Genet.* 2016, **12**:e1005854.

9. Chen I-MA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, Ratner A, Huang J, Andersen E, Huntemann M, et al.: **IMG/M: integrated genome and metagenome comparative data analysis system.** *Nucleic Acids Res.* 2017, **45**:D507–D516.

10. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, et al.: **The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST).** *Nucleic Acids Res.* 2014, **42**:D206–14.

11. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, et al.: **The primary transcriptome of the major human pathogen Helicobacter pylori.** *Nature* 2010, **464**:250–255.

12. Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM, Storz G: **Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli.** *J. Bacteriol.* 2015, **197**:18–28.

13. Xu W, Chen H, He C-L, Wang Q: **Deep sequencing-based identification of small regulatory RNAs in Synechocystis sp. PCC 6803.** *PLoS One* 2014, **9**:e92711.

14. Weiss A, Broach WH, Lee MC, Shaw LN: **Towards the complete small RNome of Acinetobacter baumannii.** *Microb Genom* 2016, **2**:e000045.

15. Nesvizhskii AI: **Proteogenomics: concepts, applications and computational strategies.** *Nat Methods* 2014, **11**:1114–1125.

16. Potgieter MG, Nakedi KC, Ambler JM, Nel AJM, Garnett S, Soares NC, Mulder N, Blackburn JM: **Proteogenomic Analysis of Mycobacterium smegmatis Using High Resolution Mass Spectrometry.** *Front Microbiol* 2016, **7**:427.

17. Price MN, Deutschbauer AM, Kuehl JV, Liu H, Witkowska HE, Arkin AP: **Evidence-based annotation of transcripts and proteins in the sulfate-reducing bacterium Desulfovibrio vulgaris Hildenborough.** *J. Bacteriol.* 2011, **193**:5716–5727.

18. Ellis HM, Yu D, DiTizio T, Court DL: **High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides.** *Proc. Natl. Acad. Sci. U.S.A.* 2001, **98**:6742–6746.

19. Pines G, Freed EF, Winkler JD, Gill RT: **Bacterial Recombineering: Genome Engineering via Phage-Based Homologous Recombination.** *ACS Synth Biol* 2015, **4**:1176–1185.

20. Yin J, Zhu H, Xia L, Ding X, Hoffmann T, Hoffmann M, Bian X, Müller R, Fu J, Stewart AF, et al.: **A new recombineering system for Photorhabdus and Xenorhabdus.** *Nucleic Acids Res.* 2015, **43**:e36–e36.

21. Lee HH, Ostrov N, Gold MA, Church GM: **Recombineering in Vibrio natriegens**. *bioRxiv* 2017, doi:10.1101/130088.

22. Jiang Y, Qian F, Yang J, Liu Y, Dong F, Xu C, Sun B, Chen B, Xu X, Li Y, et al.: **CRISPR-Cpf1 assisted genome editing of Corynebacterium glutamicum.** *Nat Commun* 2017, **8**:15179.

23. Yan M-Y, Yan H-Q, Ren G-X, Zhao J-P, Guo X-P, Sun Y-C: **CRISPR-Cas12a-Assisted Recombineering in Bacteria.** *Appl Environ Microbiol* 2017, **83**:e00947–17.

24. Yao L, Cengic I, Anfelt J, Hudson EP: **Multiple Gene Repression in Cyanobacteria Using CRISPRi.** *ACS Synth Biol* 2016, **5**:207–212.

25. Choudhary E, Thakur P, Pareek M, Agarwal N: **Gene silencing by CRISPR interference in mycobacteria.** *Nat Commun* 2015, **6**:6267.

26. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA: **Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.** *Cell* 2013, **152**:1173–1183.

27. Gordon GC, Korosh TC, Cameron JC, Markley AL, Begemann MB, Pfleger BF: **CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium Synechococcus sp. strain PCC 7002.** *Metab. Eng.* 2016, **38**:170–179.

28. Monk IR, Tree JJ, Howden BP, Stinear TP, Foster TJ: **Complete Bypass of Restriction Systems for Major Staphylococcus aureus Lineages.** *MBio* 2015, **6**:e00308–15.

29. Zhang G, Wang W, Deng A, Sun Z, Zhang Y, Liang Y, Che Y, Wen T: **A mimicking-of-DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria.** *PLoS Genet.* 2012, **8**:e1002987.

30. Garcia PA, Ge Z, Moran JL, Buie CR: **Microfluidic Screening of Electric Fields for Electroporation.** *Sci Rep* 2016, **6**:21238.

31. Mimee M, Tucker AC, Voigt CA, Lu TK: **Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota.** *Cell Syst* 2015, **1**:62–71.

32. Whitaker WR, Shepherd ES, Sonnenburg JL: **Tunable Expression Tools Enable Single-Cell Strain Distinction in the Gut Microbiome.** *Cell* 2017, **169**:538–546.e12.

33. Lim B, Zimmermann M, Barry NA, Goodman AL: **Engineered Regulatory Systems Modulate Gene Expression of Human Commensals in the Gut.** *Cell* 2017, **169**:547–558.e15.

34. Liu H, Price M, Waters RJ, Ray J, Carlson HK, Lamson JS, Chakraborty R, Arkin AP, Deutschbauer AM: **Magic pools: parallel assessment of transposon delivery vectors in bacteria**. *bioRxiv* 2017, doi:10.1101/158840.

35. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: **Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection**. *Mol. Syst. Biol.* 2006, **2**:2006 0008.

36. Koo B-M, Kritikos G, Farelli JD, Todor H, Tong K, Kimsey H, Wapinski I, Galardini M, Cabal A, Peters JM, et al.: **Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis.** *Cell Syst* 2017, **4**:291–305.e7.

37. Ramage B, Erolin R, Held K, Gasper J, Weiss E, Brittnacher M, Gallagher L, Manoil C: **Comprehensive arrayed transposon mutant library of Klebsiella pneumoniae outbreak strain KPNIH1.** *J. Bacteriol.* 2017, doi:10.1128/JB.00352-17.

38. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI: **Identifying genetic determinants needed to establish a human gut symbiont in its habitat.** *Cell Host Microbe* 2009, **6**:279–289.

39. Baym M, Shaket L, Anzai IA, Adesina O, Barstow B: **Rapid construction of a whole-genome transposon insertion collection for Shewanella oneidensis by Knockout Sudoku.** *Nat Commun* 2016, **7**:13270.

40. Schnoes AM, Brown SD, Dodevski I, Babbitt PC: **Annotation error in public databases: misannotation of molecular function in enzyme superfamilies.** *PLoS Comput Biol* 2009, **5**:e1000605.

41. Sévin DC, Fuhrer T, Zamboni N, Sauer U: **Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in Escherichia coli.** *Nat Methods* 2017, **14**:187–194.

42. Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R, Lee S, Kazmierczak KM, Lee KJ, Wong A, et al.: **Phenotypic landscape of a bacterial cell**. *Cell* 2011, **144**:143–156.

43. Deutschbauer A, Price MN, Wetmore KM, Shao W, Baumohl JK, Xu Z, Nguyen M, Tamse R, Davis RW, Arkin AP: **Evidence-based annotation of gene function in Shewanella oneidensis MR-1 using genome-wide fitness profiling across 121 conditions.** *PLoS Genet.* 2011, **7**:e1002385.

44. van Opijnen T, Bodi KL, Camilli A: **Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms**. *Nat Methods* 2009, **6**:767–772.

45. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, et al.: **Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants**. *Genome Res* 2009, **19**:2308–2316.

46. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, et al.: **Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons.** *MBio* 2015, **6**:e00306–15.

47. Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP, Deutschbauer A, Golden SS: **The essential gene set of a photosynthetic organism.** *Proc. Natl. Acad. Sci. U.S.A.* 2015, **112**:E6634–43.

48. Christen M, Beusch C, Bösch Y, Cerletti D, Flores-Tinoco CE, Del Medico L, Tschan F, Christen B: **Quantitative Selection Analysis of Bacteriophage φCbK Susceptibility in Caulobacter crescentus.** *J Mol Biol* 2016, **428**:419–430.

49. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, Esquivel Z, et al.: **Deep Annotation of Protein Function across Diverse Bacteria from Mutant Phenotypes**. *bioRxiv* 2016, doi:10.1101/072470.

50. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, Hawkins JS, Lu CHS, Koo B-M, Marta E, et al.: **A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria.** *Cell* 2016, **165**:1493–1506.

51. Wang T, Guo J, Guan C, Wu Y, Liu B, Xie Z, Zhang C, Xing X-H: **Pooled CRISPR interference screens enable high-throughput functional genomics study and elucidate new rules for guide RNA library design in Escherichia coli**. *bioRxiv* 2017, doi:10.1101/129668.

52. Lee WC, Goh KL, Loke MF, Vadivelu J: **Elucidation of the Metabolic Network of Helicobacter pylori J99 and Malaysian Clinical Strains by Phenotype Microarray.** *Helicobacter* 2017, **22**:e12321.

53. Barczak AK, Avraham R, Singh S, Luo SS, Zhang WR, Bray M-A, Hinman AE, Thompson M, Nietupski RM, Golas A, et al.: **Systematic, multiparametric analysis of Mycobacterium tuberculosis intracellular infection offers insight into coordinated virulence.** *PLoS Pathog.* 2017, **13**:e1006363.

54. Shi H, Colavin A, Lee TK, Huang KC: **Strain Library Imaging Protocol for high-throughput, automated single-cell microscopy of large bacterial collections arrayed on multiwell plates.** *Nat Protoc* 2017, **12**:429–438.

55. Terekhov SS, Smirnov IV, Stepanova AV, Bobik TV, Mokrushina YA, Ponomarenko NA, Belogurov AA, Rubtsova MP, Kartseva OV, Gomzikova MO, et al.: **Microfluidic droplet platform for ultrahigh-throughput single-cell screening of biodiversity.** *Proc. Natl. Acad. Sci. U.S.A.* 2017, **114**:2550–2555.

56. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, et al.: **The EcoCyc database: reflecting new knowledge about Escherichia coli K-12.** *Nucleic Acids Res.* 2017, **45**:D543–D550.

57. Lasker K, Schrader JM, Men Y, Marshik T, Dill DL, McAdams HH, Shapiro L: **CauloBrowser: A systems biology resource for Caulobacter crescentus.** *Nucleic Acids Res.* 2016, **44**:D640–5.

58. Michna RH, Zhu B, Mäder U, Stülke J: **SubtiWiki 2.0--an integrated database for the model organism Bacillus subtilis.** *Nucleic Acids Res.* 2016, **44**:D654–62.

59. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, et al.: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases.** *Nucleic Acids Res.* 2016, **44**:D471–80.

60. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: **High-throughput generation, optimization and analysis of genome-scale metabolic models.** *Nat. Biotechnol.* 2010, **28**:977–982.

61. Arkin AP, Stevens RL, Cottingham RW, Maslov S, Henry CS, Dehal P, Ware D, Perez F, Harris NL, Canon S, et al.: **The DOE Systems Biology Knowledgebase (KBase).** *bioRxiv* 2016, doi:10.1101/096354.

62. Goecks J, Nekrutenko A, Taylor J, Galaxy Team: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol.* 2010, **11**:R86.

63. Grüning BA, Rasche E, Rebolledo-Jaramillo B, Eberhard C, Houwaart T, Chilton J, Coraor N, Backofen R, Taylor J, Nekrutenko A: **Jupyter and Galaxy: Easing entry barriers into complex data analyses for biomedical researchers.** *PLoS Comput Biol* 2017, **13**:e1005425.

* of special interest

** of outstanding interest

* Reference #8. This is the first study to measure DNA methylation genome-wide in a large number of prokaryotes. The authors performed PacBio DNA sequencing on 230 diverse bacteria and archaea and identified 834 different methylated motifs among the studied microorganisms.

* Reference #20. In this study, the authors mined the genome of *Photorhabdus* to identify three3 proteins comprising a lambda Red-like system. They then adapted these proteins for use in an efficient recombineering system in *Photorhabdus luminescens* to express a large endogenous secondary metabolite gene cluster.

* Reference #298. Describes the construction of a set of *E. coli* strains that mimic the adenine methylation pattern of *Staphylococcus aureus*, thus enabling high efficiency DNA transformation into multiple lineages of this pathogen.

** Reference #310. This study describes a microfluidic devise for optimizing electroporation parameters and demonstrates its utility for both Gram positive and Gram negative bacteria.

** Reference #332. In this work, the authors use a high-throughput cloning and characterization strategy to create an expression toolbox in Bacteroides that spans multiple orders of magnitude. They subsequently use these expression tools to image different Bacteroides strains in the mouse gut.

** Reference #343. The authors describe the development of a tetracycline-regulated system for inducible gene expression in Bacteroides. They use this system to monitor the ***in situ*** activity of an enzymatic activity within mice.

* Reference #421. In this study, a high-throughput pipeline of protein purification and untargeted metabolomics was used to screen the enzymatic activity of over 1,000 poorly annotated ***E. coli*** proteins. The authors were able to experimentally validate the previously unknown enzymatic activities of 12 enzymes.

** Reference #5049. This work represents the largest effort to date to functionally annotate bacterial proteins using high-throughput genetics.

** Reference #510. This study is a large-scale phenotypic characterization of all essential genes in *Bacillus subtilis* using CRISPRi. In addition to identifying the mode-of-action of an uncharacterized antibiotic, they also used high-throughput imaging to discover new associations between cell growth and morphology.

** Reference #543. In this study, thousands of *Mycobacterium tuberculosis* mutants were screened for their growth and survival in macrophages using high-content imaging. From these data, the authors identified a link between protein secretion and lipid production in mediating pathogenesis.

* Reference #621. Presents the U.S. Department of Energy's System Biology Knowledgebase, an open-source computational platform for systems-level analyses of genomics data for microorganisms, metagenomes, and plants.

Figure legend

Figure 1. Developing new model bacteria. High-throughput tools can be rapidly applied to transition diverse bacteria to new model systems. In the panels at the bottom, we highlight the criteria for a new model bacterium and provide example approaches that can be applied to meet each of these criteria.