

DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements

Hao Luo¹, Yan Lin¹, Feng Gao¹, Chun-Ting Zhang^{1,*} and Ren Zhang^{2,*}

¹Department of Physics, Tianjin University, Tianjin 300072, People's Republic of China and ²Center for Molecular Medicine and Genetics, School of Medicine, Wayne State University, Detroit 48201, USA

Received September 16, 2013; Revised October 17, 2013; Accepted October 24, 2013

ABSTRACT

The combination of high-density transposon-mediated mutagenesis and high-throughput sequencing has led to significant advancements in research on essential genes, resulting in a dramatic increase in the number of identified prokaryotic essential genes under diverse conditions and a revised essential-gene concept that includes all essential genomic elements, rather than focusing on protein-coding genes only. DEG 10, a new release of the Database of Essential Genes (available at <http://www.essentialgene.org>), has been developed to accommodate these quantitative and qualitative advancements. In addition to increasing the number of bacterial and archaeal essential genes determined by genome-wide gene essentiality screens, DEG 10 also harbors essential noncoding RNAs, promoters, regulatory sequences and replication origins. These essential genomic elements are determined not only *in vitro*, but also *in vivo*, under diverse conditions including those for survival, pathogenesis and antibiotic resistance. We have developed customizable BLAST tools that allow users to perform species- and experiment-specific BLAST searches for a single gene, a list of genes, annotated or unannotated genomes. Therefore, DEG 10 includes essential genomic elements under different conditions in three domains of life, with customizable BLAST tools.

INTRODUCTION

Delineating a set of essential genomic elements and proteins that make up a living organism helps to understand critical cellular processes that sustain life (1–3). Identification of essential genes is especially useful to

studies of synthetic biology (4), which seeks to make an artificial self-sustainable living cell, with addable gene circuitries that encode desirable traits. Bacterial essential genes, because of their lethality phenotype, are attractive drug targets, and this is especially important for those having multidrug resistance (5).

Reverse genetics (from gene disruption to phenotypic characterization) has been extensively used to experimentally determine essential genes. One standard method is to perform targeted mutagenesis in a particular gene of interest. Classical examples include essential gene determination in *Bacillus subtilis* (6) and *Escherichia coli* (7), in which all protein-coding genes are deleted one by one. This method gives a clear-cut answer on gene lethality, but it is labor-intensive, time-consuming and requires detailed genome annotation. Single-gene knockout screens can overlook genes causing synthetic lethality, which refers to lethal phenotypes caused by genetic interactions of genes that are nonessential when deleted separately (3). Indeed, duplicated genes are less likely to be essential than singletons (8). Another method is to construct a random transposon-insertion library, followed by determination of insertion sites by DNA hybridization (9) or microarray (10), which suffers from some shortcomings including missing low-abundance transcripts, low resolution in locating insertion sites, and narrow ranges in counting probe density. An advantage of global transposon mutagenesis is that it can simultaneously identify essential noncoding elements in addition to protein-coding genes.

The combination of high-density transposon-mediated mutagenesis and high-throughput sequencing has resulted in significant advancements in the study of essential genes (11). This method, however, has in fact been gradually developed for more than 10 years. In 1999, Venter and coworkers first performed Sanger sequencing to determine transposon insertion sites (12), and later, various versions of combining transposon mutagenesis and next-generation sequencing were developed, such as TraDIS (13), INSeq

*To whom correspondence should be addressed. Tel: +1 313 577 0027; Fax: +1 313 577 5218; Email: rzhang@med.wayne.edu
Correspondence may also be addressed to Chun-Ting Zhang. Tel: +86 22 2740 2987; Fax: +86 22 2740 2697; Email: ctzhang@tju.edu.cn

(14), HITS (15), Tn-seq (16) and Tn-seq Circle (17), here collectively referred to as Tn-seq. The application of Tn-seq has allowed for significant advancements in studies on essential genes over the past few years, resulting in (i) a dramatic increase in the number of prokaryotic species with gene essentiality screens; (ii) a revision of the essential-gene concept that includes all essential genomic elements, such as noncoding RNAs, rather than focusing on protein-coding genes only and (iii) gene essentiality screens in a wide array of experimental conditions *in vitro* and *in vivo*, rather than focusing only on rich media in cell culture.

We constructed a database of essential genes (DEG) in 2004 (18), and DEG 5.0 included essential genes of both bacteria and eukaryotes (19). In addition to DEG, other essential gene databases include EGGs (Essential Genes on Genome Scale, <http://www.nmpdr.org/FIG/eggs.cgi>) and OGEE (online gene essentiality database) (20), where the former hosts microbial gene essentiality data experimentally obtained from published genome-scale gene essentiality screens and the latter hosts essential-gene data obtained from large-scale experiments with associated gene features and text-mining results. Because of text-mining results, OGEE has most essential-gene records, while DEG entries are human curated and is the only one supporting BLAST searches. We have constructed DEG 10 to accommodate the quantitative and qualitative advancements in identifying essential genes by genome-wide essentiality screens in recent years, and the following is a summary of new database developments.

- (i) In addition to protein-coding genes, DEG 10 now harbors essential genomic elements, including noncoding RNAs, promoters, regulatory sequences and replication origins (21,22).
- (ii) The number of bacteria with saturated genome-wide gene essentiality screens has nearly tripled, compared with that in DEG 5 (19).
- (iii) DEG 10 contains essential genomic elements determined not only *in vitro* (culture dishes), but also *in vivo* (intact mice) (14), not only for survival but also for pathogenesis (23), not only in rich media, but also in more diverse conditions, such as those required for cholesterol catabolism (24), antibiotic resistance (17), bile acid tolerance (13,25) and bacteriophage infection (26).
- (iv) DEG 10 hosts archaeal essential genes determined from the first gene essentiality screen in an archaeal genome (27).
- (v) DEG 10 hosts both essential and nonessential protein-coding genes.
- (vi) DEG 10 is integrated with customizable BLAST tools that allow users to perform species- and experiment-specific searches for a single gene, a list of genes, annotated or unannotated genomes.

Therefore, DEG 10 (www.essentialgene.org) reflects the progress of the research on essential genes by including essential genomic elements under different conditions in three domains of life, with customizable BLAST tools.

DATABASE NEW DEVELOPMENTS

Increased number of bacterial species with genome-wide essentiality screens

The combination of high-throughput sequencing and high-density transposon mutagenesis has largely accelerated the process in determining essential genes. Compared to DEG 5 (19), the number of bacteria with saturated genome-wide gene essentiality screens has nearly tripled in DEG 10, which has data for 31 bacteria. DEG 10 contains more than 12 000 bacterial essential genes, more than twice the number of those in DEG 5. The figures corresponding to newly added essential genes are highlighted in Table 1.

In addition to essential genes, in fact, nonessential genes can be determined as well in most genome-wide essentiality screens. Single-gene knockout experiments directly determine whether a particular gene is essential or nonessential. Genome-wide transposon mutagenesis determines nonessential genes first, because all recovered mutants only harbor transposon insertions in nonessential genes, while essential genes are, in fact, inferred. Therefore, nonessential genes can be reliably identified by both kinds of approaches. Because information about nonessential genes can be important as well, DEG 10 hosts nonessential genes, which are organized into a sub-database.

Determination of essential noncoding genomic elements

It is increasingly being recognized that bacterial genomes encode large amounts of noncoding RNAs (28). The use of high-density transposon mutagenesis and high-throughput sequencing makes identification of essential noncoding RNAs possible. In the genome of *Caulobacter crescentus*, 428 735 unique Tn5 insertions were generated and mapped in its 4 Mb genome. Therefore, in addition to identifying 480 essential protein-coding genes, 29 tRNAs and eight small noncoding RNAs were also found to be essential (22). In *Mycobacterium tuberculosis*, 36 488 transposon insertions were generated and mapped, and in addition to essential protein-coding genes, 25 nondisruptable genomic segments were found. These segments include 10 tRNAs and the RNA catalytic unit of RNaseP, which is required for tRNA processing (21). In a study with a similar method for the *Salmonella serovars Typhimurium*, 15 noncoding RNAs were found to be essential (29). It is noteworthy that RNaseP was again among the identified essential noncoding RNA, and therefore it is likely to be a widely required noncoding RNA among bacteria.

Mann *et al.* tested the hypothesis that some noncoding RNAs have niche-specific roles in virulence (23). Because increasing evidence suggests sRNAs are involved in pathogenesis, Mann *et al.* first performed RNA-seq to define the sRNA repertoire of *S. pneumoniae*, a causative agent for pneumonia, and identified 89 sRNAs. To examine organ-specific roles in pneumococcal pathogenesis, they generated a pool of pneumococcal mutants by transposon mutagenesis, administered the mutants in organs vital to the progression of pneumococcal diseases, the

Table 1. Contents in DEG 10

Domain of life	Organism	No. of essential genomic elements		Method ^a	Saturated	Ref.
		Coding	Noncoding			
Bacteria	<i>Acinetobacter baylyi</i>	499		Single-gene knockout	Yes	(50)
	<i>Bacillus subtilis</i>	261	2	Single-gene knockout	Yes	(6,51)
	<i>Bacteroides thetaiotaomicron</i>	325		INSeq	Yes	(14)
	<i>Burkholderia thailandensis</i>	406		Tn-seq	Yes	(52)
	<i>Campylobacter jejuni</i>	233		Transposon mutagenesis	Yes	(53)
	<i>Caulobacter crescentus</i>	480	532	Tn-seq	Yes	(22)
	<i>Escherichia coli</i>	620		Genetic footprinting	Yes	(54)
	<i>Escherichia coli</i>	303		Single-gene knockout	Yes	(7)
	<i>Francisella novicida</i>	396		Tn-seq	Yes	(55)
	<i>Haemophilus influenzae</i>	667		Genetic footprinting	Yes	(56)
	<i>Helicobacter pylori</i>	344		Transposon mutagenesis followed by microarray (MATT)	Yes	(57)
	<i>Mycobacterium tuberculosis</i>	614		Transposon mutagenesis followed by hybridization (TraSH)	Yes	(58)
	<i>Mycobacterium tuberculosis</i>	774		Tn-seq	Yes	(24)
	<i>Mycobacterium tuberculosis</i>	742	35	Tn-seq	Yes	(21)
	<i>Mycoplasma genitalium</i>	382		Transposon mutagenesis followed by Sanger sequencing, Single-gene knockout	Yes	(12,59)
	<i>Mycoplasma pulmonis</i>	321		Transposon mutagenesis followed by Sanger sequencing	Yes	(60)
	<i>Porphyromonas gingivalis</i>	463		Tn-seq	Yes	(42)
	<i>Pseudomonas aeruginosa</i>	335		Transposon mutagenesis followed by genetic footprinting	Yes	(61)
	<i>Pseudomonas aeruginosa</i>	117		Tn-seq	Yes	(17)
	<i>Salmonella entericaserovar Typhi</i>	356		TraDIS	Yes	(13)
	<i>Salmonella enterica Typhimurium</i>	306	15	TraDIS	Yes	(29)
	<i>Salmonella entericaserovar Typhimurium</i>	105		Tn-seq	Yes	(25)
	<i>Salmonella entericaserovar Typhimurium SL1344</i>	353	23	TraDIS	Yes	(29)
	<i>Salmonella entericaserovar Typhi Ty2</i>	358	24	TraDIS	Yes	(29)
	<i>Salmonella typhimurium</i>	490		Insertion-duplication	Yes	(62)
	<i>Shewanella oneidensis</i>	403		Transposon mutagenesis followed by microarray	Yes	(63)
	<i>Sphingomonas wittichii</i>	579	32	Tn-seq	Yes	(64)
	<i>Staphylococcus aureus</i>	302		Antisense RNA; Allelic Replacement Mutagenesis	No	(65-67)
	<i>Staphylococcus aureus</i>	351		Transposon-Mediated Differential Hybridisation (TMDH)	Yes	(68)
	<i>Streptococcus pneumoniae</i>	244		Insertion-duplication, allelic replacement mutagenesis	No	(69,70)
	<i>Streptococcus pneumoniae</i>		72	Tn-seq	Yes	(23)
	<i>Streptococcus sanguinis</i>	218		Single-gene knockout	Yes	(41)
<i>Vibrio cholerae</i>	789		Tn-seq	Yes	(71)	
Archaea	<i>Methanococcus maripaludis</i>	519		Tn-seq	Yes	(27)
Eukaryotes	<i>Arabidopsis thaliana</i>	358		Single-gene knockout	No	(48)
	<i>Aspergillus fumigatus</i>	35		Conditional promoter replacement	No	(72)
	<i>Caenorhabditis elegans</i>	294		RNA interference	No	(73)
	<i>Danio rerio</i>	315		Insertional mutagenesis	No	(74)
	<i>Drosophila melanogaster</i>	376		P-element insertion	No	(75)
	<i>Homo sapiens</i>	2452		Orthologs and literature search	No	(76,77)
	<i>Mus musculus</i>	2136		Single-gene knockout	No	(78)
	<i>Saccharomyces cerevisiae</i>	1110		Single-gene knockout	Yes	(37)
	<i>Schizosaccharomyces pombe</i>	1260		Single-gene knockout	Yes	(38)

Figures in bold denote newly added essential genes.

^aGenetic footprinting is a method that performs transposon mutagenesis followed by PCR to determine transposon insertion sites (79). Tn-seq here collectively refers to a method that uses the next-generation sequencing to determine transposon insertion sites, including, TraDIS, INSeq, HITS, Tn-seq and Tn-seq Circle.

nasopharynx, lungs and bloodstream, and performed deep sequencing in DNA from recovered mutants. Consequently, 28 sRNAs in the lung, 26 in the nasopharynx and 18 in the blood were found to alter fitness in these host niches. Therefore, this study used Tn-seq to assay the role of sRNA in pathogenesis in a niche-specific manner (23).

In addition to noncoding RNAs, other noncoding elements of the genome can be essential as well. These include promoters of some essential protein-coding genes, regulatory sequences and replication origins. Indeed, Christen *et al.* identified 402 essential promoter regions and two essential elements in the replication origin of the *Caulobacter* genome, in addition to 91 essential intergenic sequences with unknown functions (22). Zhang *et al.* identified 35 intergenic elements for optimal growth of *M. tuberculosis* (21). DEG 10 collects the above identified noncoding genomic elements, with annotations from the Rfam database (30), if relevant annotations are available. Because of the apparent essential role of replication origins, DEG 10 also links to DoriC, which is a database of bacterial and archaeal replication origins (31).

Determination of essential genes under diverse conditions

The application of high-throughput sequencing makes it possible to determine and quantify contributions of essential genes to organism fitness under conditions that are not practical by using other methods, because of the digital nature of the next-generation sequencing. Therefore, in addition to regular rich medium in cell cultures, in the past several years, bacterial essential genes have been identified under a large number of different conditions, e.g. in intact mice.

One illustrative example is the study on genes required to establish a human gut symbiont (14). Goodman *et al.* first performed the INseq method (transposon mutagenesis followed by next generation sequencing) to identify a set of essential genes for the commensal *B. thetaiotaomicron* *in vitro*. Next, they examined the genes critical for fitness *in vivo*, i.e. in a mammalian gut ecosystem by colonizing bacterial mutants in germ-free mice. By comparing the input (before inoculation) and output (recovered bacteria), 280 genes showed underrepresentation, suggesting them to be critical for *in vivo* fitness. By changing the experimental conditions such as in the presence of human gut-associated bacteria, they identified five adjacent genes that conferred fitness disadvantages during monoassociation of germ-free mice, while showing no impact on bacterial growth *in vitro*, thus highlighting the importance of the *in vivo* context in determining gene essentiality (14).

In a large-scale study, van Opijnen and Camilli performed Tn-seq on the genome of *S. pneumoniae* under 17 *in vitro* conditions (e.g. pH, temperature, antibiotic, heavy metal, stress, nutritional stimulation) and two *in vivo* conditions (carriage and infection), and have identified over 1,800 genotype-phenotype genetic interactions and associated pathways (32). Other condition-specific studies include the identification of essential genes for

bile acid tolerance, a trait required of an enteric bacterium and for carriage of *S. Typhi* in the gall bladder (13,25), resistance to the aminoglycoside antibiotic tobramycin in *Pseudomonas aeruginosa* (17), bacteriophage infection of *S. Typhi* to assess for Vi polysaccharide capsule expression (26) and cholesterol metabolism in *M. tuberculosis* (24). In addition to those determined in rich medium only, DEG 10 harbors condition-specific essential genes as well.

Determination of essential genes in an archaeal genome

Archaea are prokaryotes that constitute a separate domain of life, in addition to bacteria and eukaryotes (33). Some archaea can survive in extreme conditions, such as highly salty or hot environments. Methanogenesis, a process to generate methane, is a specialized anaerobic respiration that requires distinctive biochemical reactions unique to methanogenic archaea, which are responsible for 80% of the methane in greenhouse gas (34).

By using the method of Tn-seq, Sarmiento *et al.* identified essential genes in hydrogenotrophic, methanogenic archaeon *Methanococcus marisnigri* S2, and this was the first genome-wide gene essentiality screen in archaea (27). About 89 000 unique transposon inserts were mapped, and 526 genes were classified as essential in rich medium. Similar to bacteria, many essential genes encode fundamental cellular processes, such as transcription, translation and replication. Some essential genes, however, are unique to the archaeal or methanococcal lineages. For instance, the DNA polymerase PolD is essential, whereas the archaeal homolog of bacterial PolB is not (27).

Determination of essential eukaryotic genes

In contrast to prokaryotic essential genes, which have had a dramatic increase in past years, the number of eukaryotic essential genes, while climbing steadily, does not exhibit a drastic increase, apparently due to the lack of genome-wide mutagenesis strategies. To generate single-gene knockout, however, takes much more effort, and therefore usually requires multi-center collaborations. The aim of the International Knockout Mouse Consortium (IKMC), formed in 2007, is to generate mutant mouse lines with all genes deleted one by one (35). A recent report showed mouse gene deletion mutants have been obtained for ~17 000 of the total 20 000 protein-coding genes (36). Therefore, in the near future, we expect to have a complete set of essential mouse protein-coding genes. With *Saccharomyces cerevisiae* being the first eukaryote to have all of its single-gene deletion mutants generated (37), DEG 10 has added essential genes of *Schizosaccharomyces pombe*, which is the second eukaryote that has a saturated gene deletion study (38).

Customizable BLAST tools

Performing homologous searches with the BLAST program (39) against DEG is common (40–43), and therefore to facilitate this use, we have developed a set of

customizable BLAST tools. Users have the following four options.

- (i) To perform BLAST search for a single gene. The major improvement for this option is that users now can perform species-specific BLAST search, in addition to having the option to change *P*-values. The output is unprocessed BLAST raw results.
- (ii) To perform BLAST search for a list of genes. Users can submit a list of protein or DNA sequences, and the BLAST output will be organized and processed to generate an XML file that is parsed by the Biopython module (44). The output includes how many genes among the queried gene set have DEG homologs, and how many homologous genes in DEG are found. All homologous genes are clickable by linking to corresponding alignments. The above function can also be done in a species-specific manner.
- (iii) To perform BLAST search for annotated genome sequences. Because of the increasing pace of genome sequencing, in many cases users need to analyze whole-genome sequences. By using this option, users can submit a whole-genome sequence or scaffold, with annotation information, i.e. either in the GenBank format or by uploading Protein Table Files (PTT format).
- (iv) To perform BLAST search for unannotated genome sequences. If users need to analyze whole-genome sequences that have not been annotated, DEG is integrated with two gene-finding programs, Zcurve (45) and Glimmer (46), for gene identification. Protein-coding genes are first identified by Zcurve or Glimmer, and then BLAST searches are performed against DEG. For both Options 3 and 4, the output is processed and organized to convey information on the number of homologs in DEG and in queried genomes, with linking to alignments. The XML files and resulting webpages are stored for 7 days on the server, and can be retrieved as needed.

With the aforementioned new tools, users can perform BLAST searches for single genes, multiple genes, annotated genomes or unannotated genomes with filters to restrict the search to a subset of species or experiments with desirable *P*-values.

FUTURE PERSPECTIVE

Recent breakthroughs in sequencing technology, i.e. the next-generation sequencing that parallelizes the process to sequence millions of reads concurrently, have fundamentally changed many areas of biological research, and the research on essential genes is no exception. Significant advancements have been made in essential-gene studies, for example, the concept of the essential gene has been revised to include all essential genomic elements, rather than focusing on protein-coding genes only. It is not difficult to envision that in the near future, genome-wide gene essentiality screens will be performed in a large number of

bacteria and archaea, under increasingly diverse experimental conditions, and will result in dramatic increases in identified prokaryotic essential genomic elements. The accumulation of essential-gene information will be particularly helpful in identifying bacterial drug targets (47) and in constructing the minimal genome in studies of synthetic biology (4). Without breakthroughs in genome-wide mutagenesis technology, however, there will likely be no dramatic increases in identified eukaryotic essential genes. Nevertheless, it is expected that single-gene knockout projects for the model organisms, such as mice and *Arabidopsis thaliana* (48), will soon be completed. It is increasingly being recognized that mammalian genomes have highly complex transcriptomes (49), and therefore we would expect that some eukaryotic noncoding elements, such as long noncoding RNAs, will be identified as essential. DEG will continue to incorporate newly discovered essential genomic elements in a timely manner to keep pace with this rapidly developing field.

FUNDING

Startup fund from Wayne State University (to R.Z.). Funding for open access charge: In part by the National Natural Science Foundation of China [31171238, 30800642 and 31200991]; Program for New Century Excellent Talents in University [No. NCET-12-0396].

Conflict of interest statement. None declared.

REFERENCES

1. Koonin, E.V. (2000) How many genes can make a cell: the minimal-gene-set concept. *Annu. Rev. Genomics Hum. Genet.*, **1**, 99–116.
2. Koonin, E.V. (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.*, **1**, 127–136.
3. Gerdes, S., Edwards, R., Kubal, M., Fonstein, M., Stevens, R. and Osterman, A. (2006) Essential genes on metabolic maps. *Curr. Opin. Biotechnol.*, **17**, 448–456.
4. Lartigue, C., Glass, J.I., Alperovich, N., Pieper, R., Parmar, P.P., Hutchison, C.A. 3rd, Smith, H.O. and Venter, J.C. (2007) Genome transplantation in bacteria: changing one species to another. *Science*, **317**, 632–638.
5. Galperin, M.Y. and Koonin, E.V. (1999) Searching for drug targets in microbial genomes. *Curr. Opin. Biotechnol.*, **10**, 571–578.
6. Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P. *et al.* (2003) Essential *Bacillus subtilis* genes. *Proc. Natl Acad. Sci. USA*, **100**, 4678–4683.
7. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006 0008.
8. Chen, W.H., Trachana, K., Lercher, M.J. and Bork, P. (2012) Younger genes are less likely to be essential than older genes, and duplicates are less likely to be essential than singletons of the same age. *Mol. Biol. Evol.*, **29**, 1703–1706.
9. Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E. and Holden, D.W. (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science*, **269**, 400–403.
10. Mazurkiewicz, P., Tang, C.M., Boone, C. and Holden, D.W. (2006) Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. *Nat. Rev. Genet.*, **7**, 929–939.

11. Barquist,L., Boinett,C.J. and Cain,A.K. (2013) Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol.*, **10**, 1161–1169.
12. Hutchison,C.A., Peterson,S.N., Gill,S.R., Cline,R.T., White,O., Fraser,C.M., Smith,H.O. and Venter,J.C. (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. *Science*, **286**, 2165–2169.
13. Langridge,G.C., Phan,M.D., Turner,D.J., Perkins,T.T., Parts,L., Haase,J., Charles,I., Maskell,D.J., Peters,S.E., Dougan,G. *et al.* (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res.*, **19**, 2308–2316.
14. Goodman,A.L., McNulty,N.P., Zhao,Y., Leip,D., Mitra,R.D., Lozupone,C.A., Knight,R. and Gordon,J.I. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe*, **6**, 279–289.
15. Gawronski,J.D., Wong,S.M., Giannoukos,G., Ward,D.V. and Akerley,B.J. (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc. Natl Acad. Sci. USA*, **106**, 16422–16427.
16. van Opijnen,T., Bodi,K.L. and Camilli,A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods*, **6**, 767–772.
17. Gallagher,L.A., Shendure,J. and Manoil,C. (2011) Genome-scale identification of resistance functions in Pseudomonas aeruginosa using Tn-seq. *MBio*, **2**, e00315–e00310.
18. Zhang,R., Ou,H.Y. and Zhang,C.T. (2004) DEG: a database of essential genes. *Nucleic Acids Res.*, **32**, D271–D272.
19. Zhang,R. and Lin,Y. (2009) DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res.*, **37**, D455–D458.
20. Chen,W.H., Minguez,P., Lercher,M.J. and Bork,P. (2012) OGEE: an online gene essentiality database. *Nucleic Acids Res.*, **40**, D901–D906.
21. Zhang,Y.J., Ioerger,T.R., Huttenhower,C., Long,J.E., Sassetti,C.M., Sacchetti,J.C. and Rubin,E.J. (2012) Global assessment of genomic regions required for growth in Mycobacterium tuberculosis. *PLoS Pathog.*, **8**, e1002946.
22. Christen,B., Abeliuk,E., Collier,J.M., Kalogeraki,V.S., Passarelli,B., Coller,J.A., Fero,M.J., McAdams,H.H. and Shapiro,L. (2011) The essential genome of a bacterium. *Mol. Syst. Biol.*, **7**, 528.
23. Mann,B., van Opijnen,T., Wang,J., Obert,C., Wang,Y.D., Carter,R., McGoldrick,D.J., Ridout,G., Camilli,A., Tuomanen,E.I. *et al.* (2012) Control of virulence by small RNAs in Streptococcus pneumoniae. *PLoS Pathog.*, **8**, e1002788.
24. Griffin,J.E., Gawronski,J.D., Dejesus,M.A., Ioerger,T.R., Akerley,B.J. and Sassetti,C.M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.*, **7**, e1002251.
25. Khatiwara,A., Jiang,T., Sung,S.S., Dawoud,T., Kim,J.N., Bhattacharya,D., Kim,H.B., Ricke,S.C. and Kwon,Y.M. (2012) Genome scanning for conditionally essential genes in Salmonella enterica Serotype Typhimurium. *Appl. Environ. Microbiol.*, **78**, 3098–3107.
26. Pickard,D., Kingsley,R.A., Hale,C., Turner,K., Sivaraman,K., Wetter,M., Langridge,G. and Dougan,G. (2013) A genomewide mutagenesis screen identifies multiple genes contributing to Vi capsular expression in Salmonella enterica serovar Typhi. *J. Bacteriol.*, **195**, 1320–1326.
27. Sarmiento,F., Mrazek,J. and Whitman,W.B. (2013) Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Methanococcus maripaludis. *Proc. Natl Acad. Sci. USA*, **110**, 4726–4731.
28. Weinberg,Z., Perreault,J., Meyer,M.M. and Breaker,R.R. (2009) Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature*, **462**, 656–659.
29. Barquist,L., Langridge,G.C., Turner,D.J., Phan,M.D., Turner,A.K., Bateman,A., Parkhill,J., Wain,J. and Gardner,P.P. (2013) A comparison of dense transposon insertion libraries in the Salmonella serovars Typhi and Typhimurium. *Nucleic Acids Res.*, **41**, 4549–4564.
30. Gardner,P.P., Daub,J., Tate,J., Moore,B.L., Osuch,I.H., Griffiths-Jones,S., Finn,R.D., Nawrocki,E.P., Kolbe,D.L., Eddy,S.R. *et al.* (2011) Rfam: Wikipedia, clans and the “decimal” release. *Nucleic Acids Res.*, **39**, D141–D145.
31. Gao,F., Luo,H. and Zhang,C.T. (2013) DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. *Nucleic Acids Res.*, **41**, D90–D93.
32. van Opijnen,T. and Camilli,A. (2012) A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res.*, **22**, 2541–2551.
33. Woese,C.R., Kandler,O. and Wheelis,M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA*, **87**, 4576–4579.
34. Reeceburgh,W.S. (2007) Oceanic methane biogeochemistry. *Chem Rev.*, **107**, 4865–4869.
35. Ringwald,M., Iyer,V., Mason,J.C., Stone,K.R., Tadepally,H.D., Kadin,J.A., Bult,C.J., Eppig,J.T., Oakley,D.J., Briois,S. *et al.* (2011) The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium. *Nucleic Acids Res.*, **39**, D849–D855.
36. Dolgin,E. (2011) Mouse library set to be knockout. *Nature*, **474**, 262–263.
37. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Veronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., Andre,B. *et al.* (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, **418**, 387–391.
38. Kim,D.U., Hayles,J., Kim,D., Wood,V., Park,H.O., Won,M., Yoo,H.S., Duhig,T., Nam,M., Palmer,G. *et al.* (2010) Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. *Nat. Biotechnol.*, **28**, 617–623.
39. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
40. Holman,A.G., Davis,P.J., Foster,J.M., Carlow,C.K. and Kumar,S. (2009) Computational prediction of essential genes in an unculturable endosymbiotic bacterium, Wolbachia of Brugia malayi. *BMC Microbiol.*, **9**, 243.
41. Xu,P., Ge,X., Chen,L., Wang,X., Dou,Y., Xu,J.Z., Patel,J.R., Stone,V., Trinh,M., Evans,K. *et al.* (2011) Genome-wide essential gene identification in Streptococcus sanguinis. *Sci. Rep.*, **1**, 125.
42. Klein,B.A., Tenorio,E.L., Lazinski,D.W., Camilli,A., Duncan,M.J. and Hu,L.T. (2012) Identification of essential genes of the periodontal pathogen Porphyromonas gingivalis. *BMC Genomics*, **13**, 578.
43. Juhas,M., Stark,M., von Mering,C., Lumjaktase,P., Crook,D.W., Valvano,M.A. and Eberl,L. (2012) High confidence prediction of essential genes in Burkholderia cenocepacia. *PLoS One*, **7**, e40064.
44. Cock,P.J., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B. *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.
45. Guo,F.B., Ou,H.Y. and Zhang,C.T. (2003) ZCURVE: a new system for recognizing protein-coding genes in bacterial and archaeal genomes. *Nucleic Acids Res.*, **31**, 1780–1789.
46. Delcher,A.L., Bratke,K.A., Powers,E.C. and Salzberg,S.L. (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, **23**, 673–679.
47. Agüero,F., Al-Lazikani,B., Aslett,M., Berriman,M., Buckner,F.S., Campbell,R.K., Carmona,S., Carruthers,I.M., Chan,A.W., Chen,F. *et al.* (2008) Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat. Rev. Drug Discov.*, **7**, 900–907.
48. Meinke,D., Muralla,R., Sweeney,C. and Dickerman,A. (2008) Identifying essential genes in Arabidopsis thaliana. *Trends Plant Sci.*, **13**, 483–491.
49. Mercer,T.R., Gerhardt,D.J., Dinger,M.E., Crawford,J., Trapnell,C., Jeddelloh,J.A., Mattick,J.S. and Rinn,J.L. (2012) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.*, **30**, 99–104.
50. de Berardinis,V., Vallenet,D., Castelli,V., Besnard,M., Pinet,A., Cruaud,C., Samair,S., Lechaplais,C., Gyapay,G., Richez,C. *et al.* (2008) A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADPI. *Mol. Syst. Biol.*, **4**, 174.

51. Commichau, F.M., Pietack, N. and Stulke, J. (2013) Essential genes in *Bacillus subtilis*: a re-evaluation after ten years. *Mol. Biosyst.*, **9**, 1068–1075.
52. Baugh, L., Gallagher, L.A., Patrapuvich, R., Clifton, M.C., Gardberg, A.S., Edwards, T.E., Armour, B., Begley, D.W., Dieterich, S.H., Dranow, D.M. *et al.* (2013) Combining functional and structural genomics to sample the essential Burkholderia structome. *PLoS One*, **8**, e53851.
53. Metris, A., Reuter, M., Gaskin, D.J., Baranyi, J. and van Vliet, A.H. (2011) In vivo and in silico determination of essential genes of *Campylobacter jejuni*. *BMC Genomics*, **12**, 535.
54. Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balazsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyrpides, N.C., Anderson, I., Gelfand, M.S. *et al.* (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.*, **185**, 5673–5684.
55. Gallagher, L.A., Ramage, E., Jacobs, M.A., Kaul, R., Brittnacher, M. and Manoil, C. (2007) A comprehensive transposon mutant library of *Francisella novicida*, a bioweapon surrogate. *Proc. Natl Acad. Sci. USA*, **104**, 1009–1014.
56. Akerley, B.J., Rubin, E.J., Novick, V.L., Amaya, K., Judson, N. and Mekalanos, J.J. (2002) A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc. Natl Acad. Sci. USA*, **99**, 966–71.
57. Salama, N.R., Shepherd, B. and Falkow, S. (2004) Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. *J. Bacteriol.*, **186**, 7926–7935.
58. Sasseti, C.M., Boyd, D.H. and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.*, **48**, 77–84.
59. Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A. 3rd, Smith, H.O. and Venter, J.C. (2006) Essential genes of a minimal bacterium. *Proc. Natl Acad. Sci. USA*, **103**, 425–430.
60. French, C.T., Lao, P., Loraine, A.E., Matthews, B.T., Yu, H. and Dybvig, K. (2008) Large-scale transposon mutagenesis of *Mycoplasma pulmonis*. *Mol. Microbiol.*, **69**, 67–76.
61. Liberati, N.T., Urbach, J.M., Miyata, S., Lee, D.G., Drenkard, E., Wu, G., Villanueva, J., Wei, T. and Ausubel, F.M. (2006) An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc. Natl Acad. Sci. USA*, **103**, 2833–2838.
62. Knuth, K., Niesalla, H., Hueck, C.J. and Fuchs, T.M. (2004) Large-scale identification of essential *Salmonella* genes by trapping lethal insertions. *Mol. Microbiol.*, **51**, 1729–1744.
63. Deutschbauer, A., Price, M.N., Wetmore, K.M., Shao, W., Baumohl, J.K., Xu, Z., Nguyen, M., Tamse, R., Davis, R.W. and Arkin, A.P. (2011) Evidence-based annotation of gene function in *Shewanella oneidensis* MR-1 using genome-wide fitness profiling across 121 conditions. *PLoS Genet.*, **7**, e1002385.
64. Roggo, C., Coronado, E., Moreno-Forero, S.K., Harshman, K., Weber, J. and van der Meer, J.R. (2013) Genome-wide transposon insertion scanning of environmental survival functions in the polycyclic aromatic hydrocarbon degrading bacterium *Sphingomonas wittichii* RW1. *Environ. Microbiol.*
65. Ko, K., Lee, J., Song, J., Baek, J., Oh, W., Chun, J. and Yoon, H. (2006) Screening of essential Genes in *Staphylococcus aureus* N315 using comparative genomics and allelic replacement mutagenesis. *J. Microbiol. Biotechnol.*, **16**, 623–632.
66. Ji, Y., Zhang, B., Van Horn, S.F., Warren, P., Woodnutt, G., Burnham, M.K. and Rosenberg, M. (2001) Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science*, **293**, 2266–2269.
67. Forsyth, R.A., Haselbeck, R.J., Ohlsen, K.L., Yamamoto, R.T., Xu, H., Trawick, J.D., Wall, D., Wang, L., Brown-Driver, V., Froelich, J.M. *et al.* (2002) A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol. Microbiol.*, **43**, 1387–1400.
68. Chaudhuri, R.R., Allen, A.G., Owen, P.J., Shalom, G., Stone, K., Harrison, M., Burgis, T.A., Lockyer, M., Garcia-Lara, J., Foster, S.J. *et al.* (2009) Comprehensive identification of essential *Staphylococcus aureus* genes using Transposon-Mediated Differential Hybridisation (TMDH). *BMC Genomics*, **10**, 291.
69. Thanassi, J.A., Hartman-Neumann, S.L., Dougherty, T.J., Dougherty, B.A. and Pucci, M.J. (2002) Identification of 113 conserved essential genes using a high-throughput gene disruption system in *Streptococcus pneumoniae*. *Nucleic Acids Res.*, **30**, 3152–3162.
70. Song, J.H., Ko, K.S., Lee, J.Y., Baek, J.Y., Oh, W.S., Yoon, H.S., Jeong, J.Y. and Chun, J. (2005) Identification of essential genes in *Streptococcus pneumoniae* by allelic replacement mutagenesis. *Mol. Cells*, **19**, 365–374.
71. Cameron, D.E., Urbach, J.M. and Mekalanos, J.J. (2008) A defined transposon mutant library and its use in identifying motility genes in *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA*, **105**, 8736–8741.
72. Hu, W., Sillaots, S., Lemieux, S., Davison, J., Kauffman, S., Breton, A., Linteau, A., Xin, C., Bowman, J., Becker, J. *et al.* (2007) Essential gene identification and drug target prioritization in *Aspergillus fumigatus*. *PLoS Pathog.*, **3**, e24.
73. Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M. *et al.* (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, **421**, 231–237.
74. Amsterdam, A., Nissen, R.M., Sun, Z., Swindell, E.C., Farrington, S. and Hopkins, N. (2004) Identification of 315 genes essential for early zebrafish development. *Proc. Natl Acad. Sci. USA*, **101**, 12792–12797.
75. Spradling, A.C., Stern, D., Beaton, A., Rhem, E.J., Laverty, T., Mozden, N., Misra, S. and Rubin, G.M. (1999) The Berkeley *Drosophila* Genome Project gene disruption project: single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics*, **153**, 135–177.
76. Georgi, B., Voight, B.F. and Bucan, M. (2013) From mouse to human: evolutionary genomics analysis of human orthologs of essential genes. *PLoS Genet.*, **9**, e1003484.
77. Liao, B.Y. and Zhang, J. (2008) Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc. Natl Acad. Sci. USA*, **105**, 6987–6992.
78. Liao, B.Y. and Zhang, J. (2007) Mouse duplicate genes are as essential as singletons. *Trends Genet.*, **23**, 378–381.
79. Smith, V., Botstein, D. and Brown, P.O. (1995) Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. *Proc. Natl Acad. Sci. USA*, **92**, 6479–6483.