

# Recent advances in characterization of essential genes and development of a database of essential genes

Ya-Ting Liang<sup>1</sup>, Hao Luo<sup>1</sup>, Yan Lin<sup>1</sup>, Feng Gao<sup>1,2,3</sup>

<sup>1</sup> Department of Physics, College of Science, Tianjin University, Tianjin, China

<sup>2</sup> Frontiers Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin, China

<sup>3</sup> SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin, China



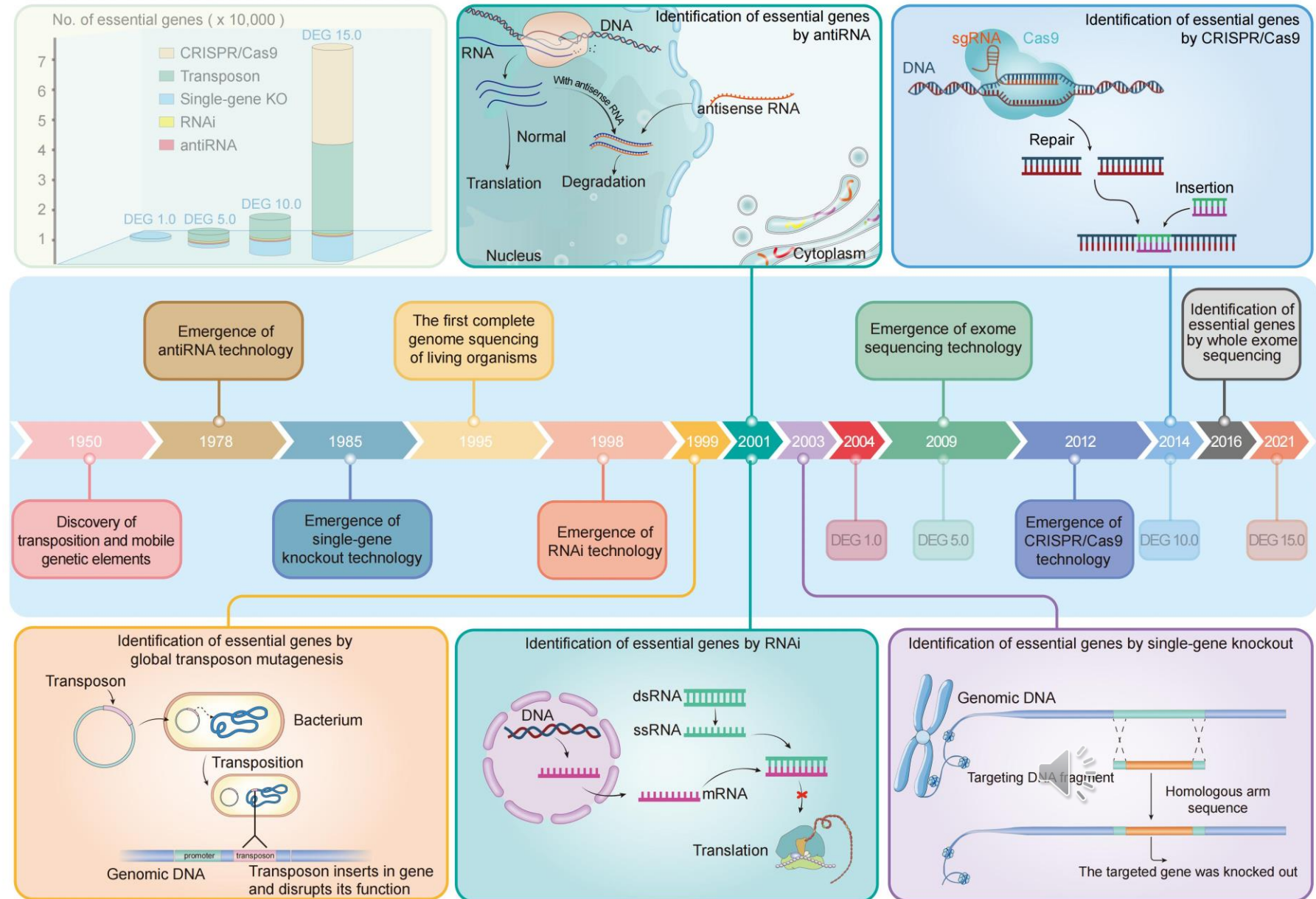
Ya-Ting Liang, Hao Luo, Yan Lin, Feng Gao. 2024. Recent advances in characterization of essential genes and development of a database of essential genes. *iMeta* e157. <https://doi.org/10.1002/imt2.157>

# Introduction to Essential Genes

**Essential Genes:** Genes that are crucial for the survival of an organism. These genes encode proteins and enzymes that play key roles in maintaining basic cellular functions, growth, and development. Impairment or loss of these genes may result in the inability of the organism to survive.

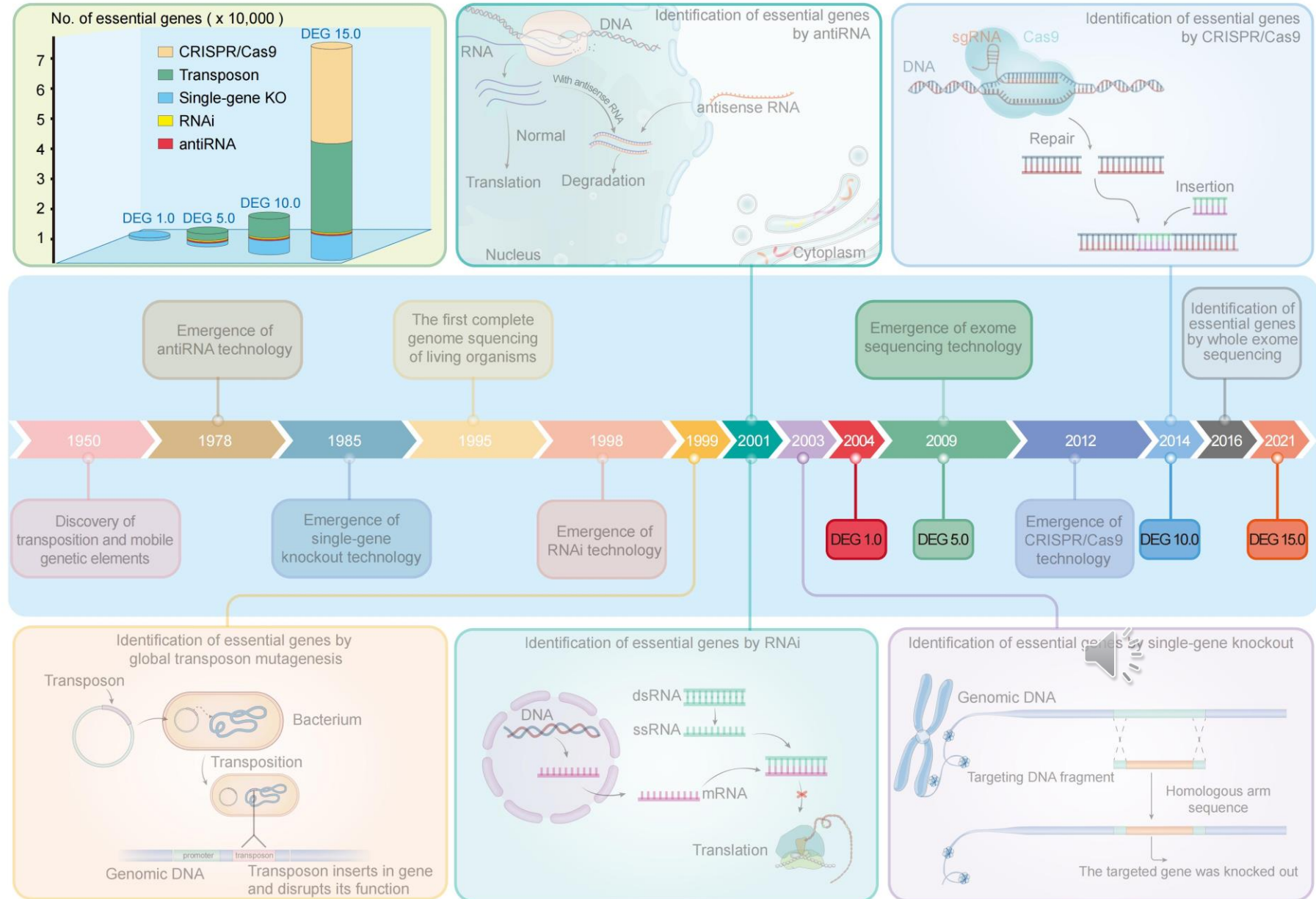
## Experimental detection methods:

RNAi,  
antiRNA,  
CRISPR/Cas9,  
transposon mutagenesis,  
single gene knockout, etc.



# Development of the DEG Database

Established and maintained by the Bioinformatics Center of Tianjin University (TUBIC) since 2004, the DEG database has accumulated information on essential genes identified at the whole-genome level in over 60 prokaryotic species and more than 10 eukaryotic species. It is the earliest established, most comprehensive, and most influential database of essential genes.



# Introduction to the DEG Database


**(A) DEG database**

DEG Bacteria Archaea Eukaryotes Non-coding Citation Download

Essential genes are those indispensable for the survival of an organism, and therefore are considered a foundation of life. DEG hosts records of currently available essential genomic elements, such as protein-coding genes and non-coding RNAs, among bacteria, archaea and eukaryotes. Essential genes in a bacterium constitute a minimal genome, forming a set of functional modules, which play key roles in the emerging field, synthetic biology.

DEG database  
Source: <https://tubic.org/deg>

Function module interface



**(B) Brief overview of data in the DEG database**

No.	Organism	Essential genes	Conditions	Reference
1	<i>Bacillus subtilis</i> 168	271	Rich medium	Kobayashi K, et al (2003) Essential <i>Bacillus subtilis</i> genes. Proc Natl Acad Sci U S A, 100:4678-83.
2	<i>Staphylococcus aureus</i> N315	302	Rich medium	Ji Y, et al (2001) Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. Science, 292:2569.
3	<i>Vibrio cholerae</i> N16961	779	Rich medium	Cameron DE, Jiratch MUMekalanon J (2008) A defined transposon mutant library and its use in identifying motility genes in <i>Vibrio cholerae</i> . PNAS, 105:3537-3541.
4	<i>Haemophilus influenzae</i> Rd KW02	642	Rich medium	Akerley BL et al (2002) A genome-scale analysis for identification of genes required for growth or survival of <i>Haemophilus influenzae</i> . Proc Natl Acad Sci U S A 99:965-71.
5	<i>Haemophilus influenzae</i> Rd KW02	642	Rich medium	Akerley BL et al (2002) A genome-scale analysis for identification of genes required for growth or survival of <i>Haemophilus influenzae</i> . Proc Natl Acad Sci U S A 99:965-71.
6	<i>Streptococcus pneumoniae</i>	244	Rich medium	Tharass J, et al (2002) Identification of 113 conserved essential genes using a high-throughput gene disruption system in <i>Streptococcus pneumoniae</i> . Nucleic Acids

**(C) Detailed information page**

Organism In Bacteria Search **BLAST** Browse

Organism	<i>Bacillus subtilis</i> 168
RefSeq	NC_000964
Essential Genes	271
Non-essential Genes	3955
Method	
Medium	Rich medium
Pubmed	12682299
Reference	Kobayashi K, et al (2003) Essential <i>Bacillus subtilis</i> genes. Proc Natl Acad Sci U S A, 100:4678-83.

**(D) Search by Sequence (BLAST)**

Organism in DEG *Bacillus subtilis* 168 (DEG1001)

BLASTN  
BLASTP  
BLASTX

BLAST

Nucleotide sequence/Amino acid sequence...

0 / 10000

BLAST Clear

**(E) Distribution of essential genes on leading or lagging strand and by cellular location**

Distribution of essential genes on leading or lagging strand

lagging: 19 (7.01%)

leading: 252 (92.99%)

Distribution of essential genes by cellular location

Unknown: 14 (5.17%)

Cytoplasmic Membrane: 38 (14.02%)

Cell wall: 2 (0.74%)

Cytoplasmic: 217 (80.07%)

**(F) Scatter plot of KEGG enrichment**

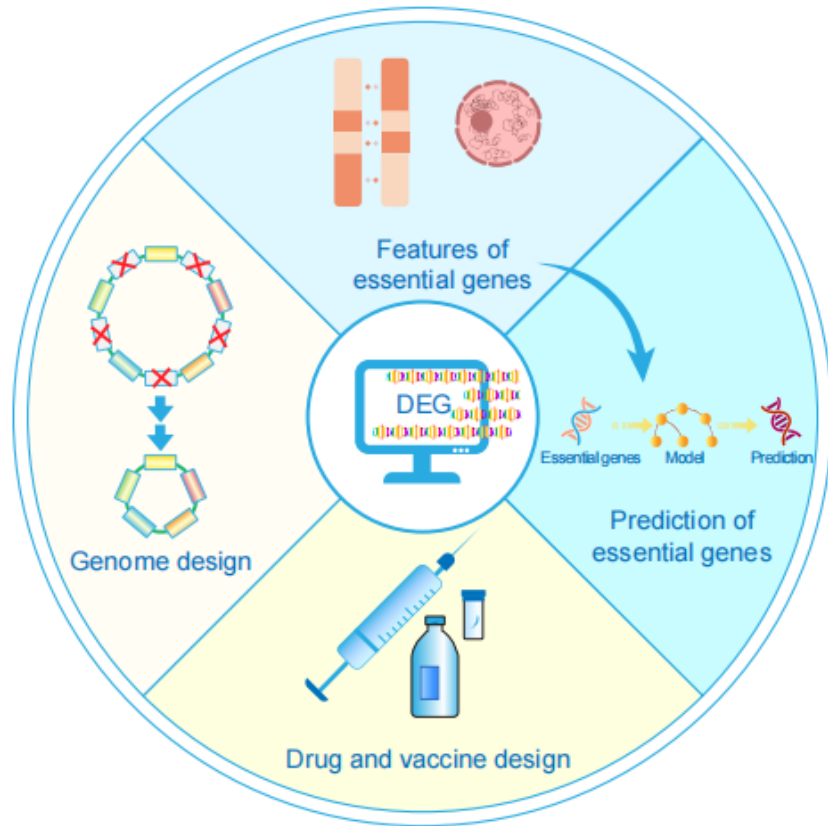
Scatter plot showing GeneRatio (x-axis, 0 to 0.5) versus enrichment (y-axis, 1e-10 to 1e-1). The plot shows a dense cluster of points at low GeneRatio and high enrichment, with a few points extending to higher GeneRatio values.

**(G) Scatter plot of GO enrichment**

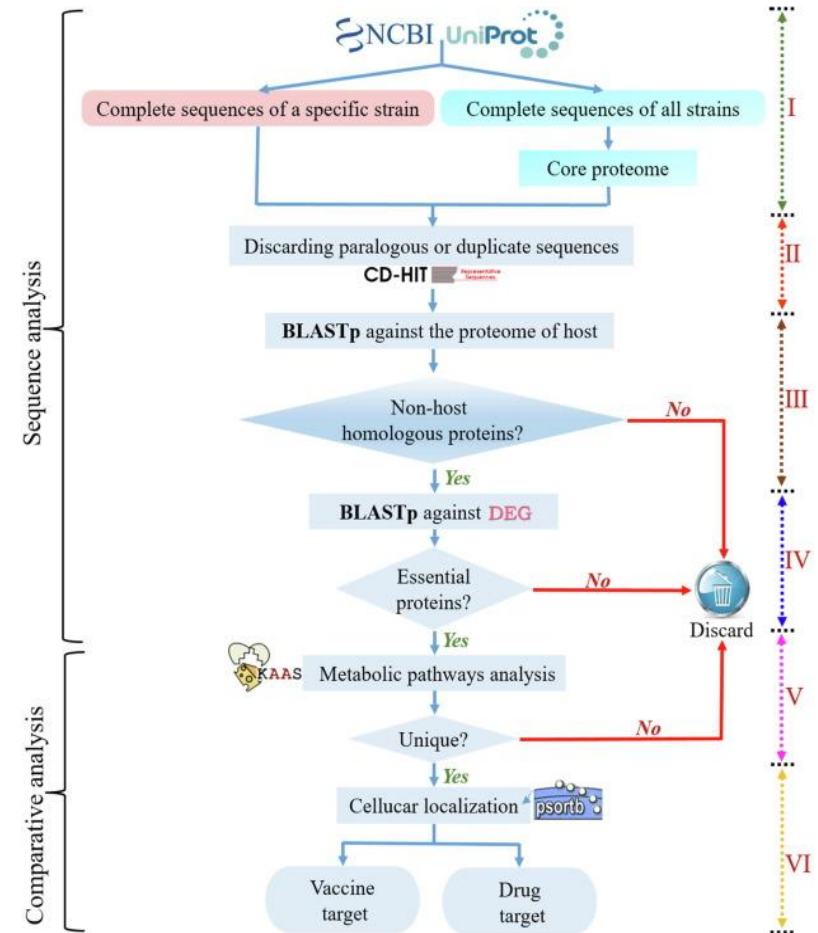
Scatter plot showing GeneRatio (x-axis, 0 to 1) versus enrichment (y-axis, 1e-5 to 1e-1). The plot shows a dense cluster of points at low GeneRatio and high enrichment, with a few points extending to higher GeneRatio values. Key GO terms are labeled: Biological process, Molecular function, and Cellular component.

The latest version, DEG 15.0, includes essential gene datasets obtained experimentally from 78 groups of bacteria, 35 groups of eukaryotes, and 2 groups of archaea. It incorporates modules for leading/lagging strands, subcellular localization distribution, KEGG and GO enrichment analysis.

# Applications of the DEG Database

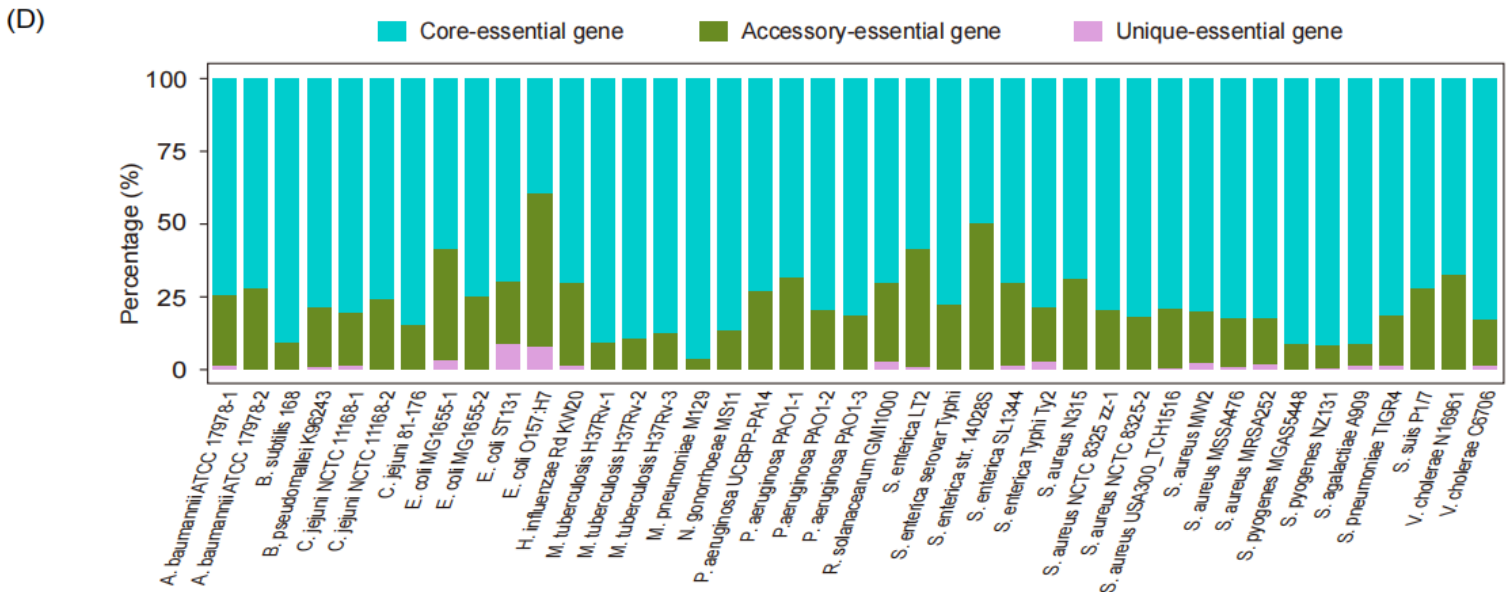
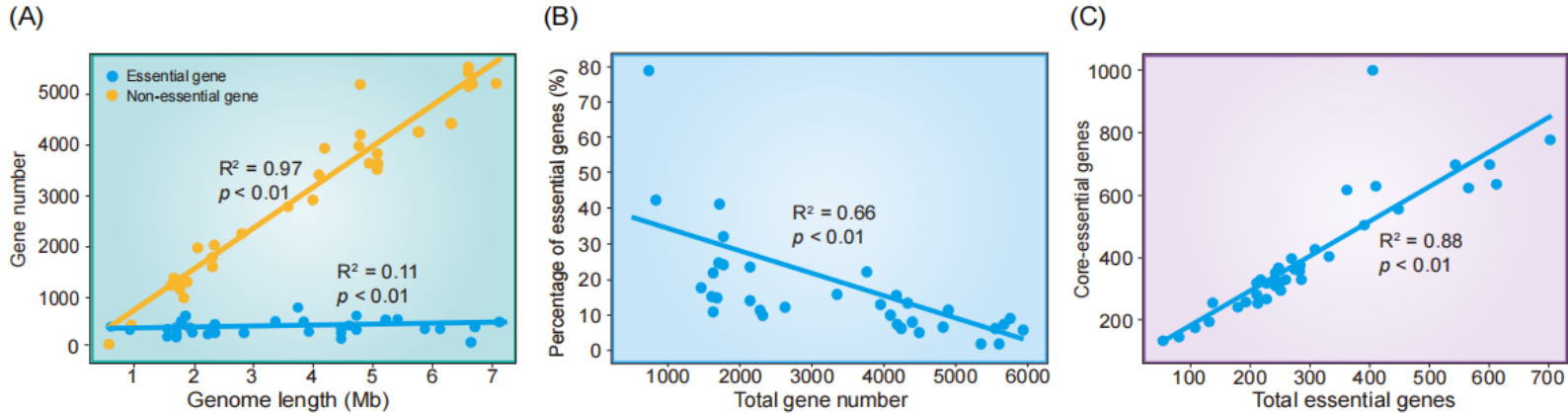


- (1) Artificial genome design and construction
- (2) Drug and vaccine design
- (3) Analysis of the features of essential gene
- (4) Prediction of essential genes



Reverse drug/vaccinology design process

# Statistical Analysis of the DEG Data



- The number of essential genes in different species remains within a certain range and does not proportionally change with genome length.
- Essential genes and core genes largely overlap.



# Conclusion

- Advances in experimental technology have facilitated the acquisition of essential gene data for multiple species.
- With the continuous accumulation of essential gene data, databases like DEG have been established.
- The DEG database plays an important role in various fields such as artificial genome design and construction, drug and vaccine design, analysis of essential gene features, and prediction of essential genes.
- The overlap between essential genes and core genes may play a more fundamental and critical role for organisms.
- Essentiality of genes is not an absolute concept; it depends on specific environments and contexts. The classification of essential genes may be an important direction for future research.



Ya-Ting Liang, Hao Luo, Yan Lin, Feng Gao. 2024. Recent advances in characterization of essential genes and development of a database of essential genes. *iMeta* e157. <https://doi.org/10.1002/imt2.157>

**iMeta:** Integrated meta-omics to change the understanding of the biology and environment

**WILEY**



**“iMeta”** is an open-access Wiley partner journal launched by scientists of the Chinese Academy of Sciences. iMeta aims to promote metagenomics, microbiome, and bioinformatics research by publishing original research, methods, or protocols, and reviews. The goal is to publish high-quality papers (Top 10%, IF > 15) targeting a broad audience. Unique features include video submission, reproducible analysis, figure polishing, APC waiver, and promotion by social media with 500,000 followers. Three issues were released in [March](#), [June](#), and [September](#) 2022.



Society: <http://www.imeta.science>

Publisher: <https://wileyonlinelibrary.com/journal/imeta>

Submission: <https://mc.manuscriptcentral.com/imeta>



[office@imeta.science](mailto:office@imeta.science)



[iMeta](#)



[iMetaScience](#)



[iMetaScience](#)