



# Variability of Bacterial Essential Genes Among Closely Related Bacteria: The Case of *Escherichia coli*

Enrique Martínez-Carranza<sup>1</sup>, Hugo Barajas<sup>2</sup>, Luis-David Alcaraz<sup>2</sup>, Luis Servín-González<sup>1</sup>, Gabriel-Yaxal Ponce-Soto<sup>3</sup> and Gloria Soberón-Chávez<sup>1\*</sup>

<sup>1</sup> Departamento de Biología Molecular y Biotecnología, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, Mexico, <sup>2</sup> Departamento de Biología Celular, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico, <sup>3</sup> Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico

## OPEN ACCESS

### Edited by:

Vasco Ariston De Carvalho Azevedo,  
Universidade Federal de Minas  
Gerais, Brazil

### Reviewed by:

Bei-Wen Ying,  
University of Tsukuba, Japan  
Jun Li,  
Zhejiang University of Technology,  
China

### \*Correspondence:

Gloria Soberón-Chávez  
gloria@biomedicas.unam.mx

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 22 February 2018

**Accepted:** 04 May 2018

**Published:** 29 May 2018

### Citation:

Martínez-Carranza E, Barajas H,  
Alcaraz L-D, Servín-González L,  
Ponce-Soto G-Y and  
Soberón-Chávez G (2018) Variability  
of Bacterial Essential Genes Among  
Closely Related Bacteria: The Case  
of *Escherichia coli*.  
Front. Microbiol. 9:1059.  
doi: 10.3389/fmicb.2018.01059

The definition of bacterial essential genes has been widely pursued using different approaches. Their study has impacted several fields of research such as synthetic biology, the construction of bacteria with minimal chromosomes, the search for new antibiotic targets, or the design of strains with biotechnological applications. Bacterial genomes are mosaics that only share a small subset of gene-sequences (core genome) even among members of the same species. It has been reported that the presence of essential genes is highly variable between closely related bacteria and even among members of the same species, due to the phenomenon known as “non-orthologous gene displacement” that refers to the coding for an essential function by genes with no sequence homology due to horizontal gene transfer (HGT). The existence of dormant forms among bacteria and the high incidence of HGT have been proposed to be driving forces of bacterial evolution, and they might have a role in the low level of conservation of essential genes among related bacteria by non-orthologous gene displacement, but this correlation has not been recognized. The aim of this mini-review is to give a brief overview of the approaches that have been taken to define and study essential genes, and the implications of non-orthologous gene displacement in bacterial evolution, focusing mainly in the case of *Escherichia coli*. To this end, we reviewed the available literature, and we searched for the presence of the essential genes defined by mutagenesis in the genomes of the 63 best-sequenced *E. coli* genomes that are available in NCBI database. We could not document specific cases of non-orthologous gene displacement among the *E. coli* strains analyzed, but we found that the quality of the genome-sequences in the database is not enough to make accurate predictions about the conservation of essential-genes among members of this bacterial species.

**Keywords:** essential-genes, core-genome, non-orthologous gene displacement, *Escherichia coli*, whole genome sequence databases

## VARIABILITY OF ESSENTIAL GENES AMONG BACTERIA: A CHALLENGE TO THE CONCEPT OF THE CORE-GENOME

The identification of the gene-products that are essential for the function of a living cell in a particular environment is a crucial task for the understanding of the molecular fundamentals of life (Juhas et al., 2011), and the evolution of the different phylogenetic taxa (Charlebois and Doolittle, 2004). Thus the definition of these indispensable gene-products has been widely studied mainly using bacteria as model organisms, due to their relative simplicity and small genome size (Sleator, 2010; Juhas et al., 2011; Glass et al., 2017). In this mini-review we focus on genes that are defined as indispensable for bacteria growing in rich media, with controlled abiotic factors (i.e., temperature, agitation) which tend to be more conserved than non-essential genes (Ish-Am et al., 2015).

The definition of a bacterial species as an evolutionary coherent biological group is a non-realistic task since bacterial genomes are mosaics that contain segments of different phylogenetic origins (Mira et al., 2004; Fraser et al., 2009; Zhaxybayeva et al., 2009; González-Casanova et al., 2014). Also, bacterial genomes present an enormous genetic variability (Ciccarelli et al., 2006; Hug et al., 2016), and a high incidence of horizontal gene transfer (HGT) (Dagan et al., 2008; Darmoh and Leach, 2014; Soucy et al., 2015). These anomalies in the phylogenetic coherence of bacterial species led to the definition of the concepts of the core-, accessory-, and pan-genome.

The core-genome is defined as the set of genes shared by all the individuals that belong to a certain species (and therefore should contain the essential genes particular to that species); the accessory-genome consists of genes that are present only in a fraction of the members of a species, or even just in one individual. Finally, the pan-genome constitutes the complete genetic repertoire of a bacterial species, that is formed by the addition of the core-genome and accessory-genome (Medini et al., 2005; Tettelin et al., 2008; Guimarães et al., 2015). Most genes belonging to the accessory-genome are acquired by HGT, and are supposed to encode adaptive functions that are not essential for the biology of a particular species (Dagan et al., 2008; Darmoh and Leach, 2014; Soucy et al., 2015).

Core-genome genes have been postulated to represent the evolutionary coherent bacterial genetic information and have been used to construct phylogenetic trees (Ciccarelli et al., 2006; Hug et al., 2016). However, it has been reported that core-genome conserved genes like tRNA synthases (Soucy et al., 2015), and the transcriptional regulator CRP (Soberón-Chávez et al., 2017), showed phylogenetic inconsistencies in some bacteria, indicating that even conserved genes might be prone to be inherited by HGT. Furthermore, it has been shown that genes coding for essential functions for the biology of *Azotobacter vinelandii* have been inherited by HGT and thus form part of the accessory-genome of the Pseudomonads (the phylogenetic group that contains *Pseudomonas* and *Azotobacter* species) (González-Casanova et al., 2014).

The inconsistency between the definition of the core-genome and these anomalies in *A. vinelandii* taxonomy have been analyzed using as a framework the population genetics model proposed by Blath et al. (2013, 2015) that has been recently reviewed (Shoemaker and Lennon, 2018). The model proposed by Blath et al. (2013, 2015) states that organisms forming dormant structures such as spores and cysts that persist in the environment for periods much larger than their generation times, give rise to an ancestral genetic pool that temporally escapes from natural selection and other evolutionary forces. It has been postulated that bacteria fulfill the postulates of Blath's model (González-Casanova et al., 2014), since many species form spores or cysts, and phages contain fragments of bacterial genomes that can be considered as forming part of dormant forms (Lang et al., 2012).

The use of comparative genomics for the study of essential genes assumes phylogenetic conservation among related organisms (Grazziotin et al., 2015; Luo et al., 2015; Oren et al., 2015), and hence that these genes form part of the core-genome (Tazzyman and Bonhoeffer, 2015; Zhang et al., 2015). However, the analysis of essential genes by comparative genomics has shown that only a small number of essential gene-sequences are conserved among different bacterial groups (Ciccarelli et al., 2006; Juhas et al., 2011; Bergmiller et al., 2012). For example, a detailed search for conserved essential genes among 147 prokaryotes (130 Bacteria and 17 Archaea) rendered only 35 conserved genes, most of which participate in transcription and translation (Charlebois and Doolittle, 2004). Furthermore, when the number of prokaryotic whole-genome sequences reached 1000 (930 Bacteria and 70 Archaea), it was reported that there were no common gene-sequences that were conserved in all of these organisms, and that only four genes (two coding for proteins and two for RNAs) were conserved among Bacteria (Lagesen et al., 2010). Furthermore, there have also been reports that essential gene-sequences can vary among strains belonging to the same bacterial species (Juhas et al., 2011; Turner et al., 2015; Ibberson et al., 2017).

The reduced number of conserved essential genes among different bacteria has generated the concept of non-orthologous gene displacement that describes a variant form of a pathway in which a certain essential gene is replaced by a functional equivalent with no DNA homology, that differs in its evolutionary origin (Koonin et al., 1996; Charlebois and Doolittle, 2004). To have a less stringent definition of a conserved essential gene, that contemplates the possibility that non-orthologous gene-displacement "erase" some of the sequence conservation of essential genes, the concept of persistent genes has been developed (Fang et al., 2005, 2008; Acevedo-Rocha et al., 2013; Luo et al., 2015).

The implications of the high incidence of non-orthologous gene displacement for the evolution of bacteria has not been addressed in the context of synthetic biology and of the study of essential genes, but is in accordance with the postulates of the Blath et al. (2013, 2015) probabilistic model in population genetics, and the high incidence of HGT.

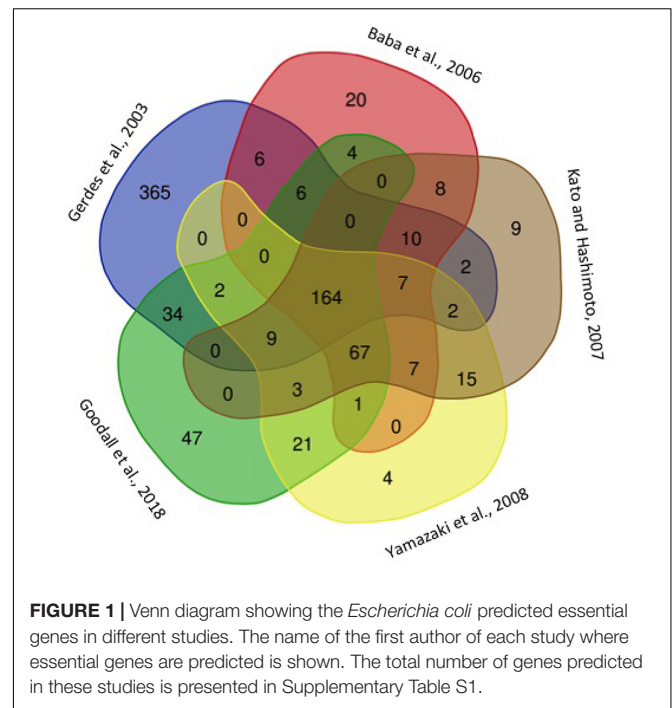
## THE SEARCH FOR BACTERIA WITH REDUCED GENOMES

The isolation of bacterial cells with reduced genomes is an active area of research for synthetic biology (Gil et al., 2004), that has shown to be important not only to define the indispensable genetic information required for a viable cell, but also for biotechnological purposes. This approach has been used to produce strains with suitable characteristics for industrial applications (Moya et al., 2009), and to define targets for the development of new antibiotics (Juhás et al., 2012a). The bacterial cell with the smallest genome that was chemically synthesized is *Mycoplasma mycoides* JCV1-Syn3.0 that contains a 531 kbp genome with a total of 473 genes (48 of them are not essential) (Hutchison et al., 2016; Glass et al., 2017).

In the case of *Escherichia coli* K-12, a strain with several deletions which eliminate 29.7% of its genome was constructed; this strain presents a lower growth rate, morphological changes, and multiple nucleoids (Hashimoto et al., 2005; Supplementary Table S1), while another deletion-derivative with the removal of 35% of its genome, produced fast-growing cultures with a higher cell-density and was engineered to produce increased threonine levels (Lee et al., 2009; Juhás et al., 2014). In addition, *E. coli* K-12 MG1655 derivatives with multiple deletions which eliminate around 15% of the genome, including all IS sequences and transposons, was reported to present a high growth rate and a high electroporation efficiency (Posfai et al., 2006; Supplementary Table S1).

The definition of essential genes was originally pursued by a mutagenesis strategy. They were defined as those in which it was not possible to isolate mutants, to eliminate by deletion, or whose expression could not be silenced using antisense-RNAs (Rusmini et al., 2014). In the case of *E. coli* (Supplementary Table S1), the first attempt to isolate mutants proposed that 620 genes were essential (Gerdes et al., 2003), but the selection of specific deletion-mutants in each open reading frame (ORF) enabled Baba et al. (2006) to reduce this number to 303. Recently a saturated transposon mutagenesis and sequencing-strategy defined 358 essential genes in *E. coli* K-12 strain BW25113 (Goodall et al., 2018); 248 of these essential genes are shared with the Keio collection (Baba et al., 2006) and also with the PEC database (Yamazaki et al., 2008), while 47 genes defined as essential are particular to this saturated-mutagenesis strategy (Goodall et al., 2018; **Figure 1**).

Kato and Hashimoto (2007) reported 268 essential genes in *E. coli* based on the inability to isolate deletions that include any of these genes, and 152 essential genes were identified in this bacterium by the inability to be silenced by antisense-RNAs (Meng et al., 2012; Supplementary Table S1). The lower number of essential genes detected using antisense-RNAs silencing might be because this method still permits a low level of expression in contrast to the mutagenesis strategy. The comparison of the specific essential genes identified in some of these studies (**Figure 1**) shows that only 164 of the total of 4218 *E. coli* genes were identified as essential in all of them (Supplementary Table S2). This result implies that only around 3.8% of the *E. coli* genome encodes for essential functions. However, it seems likely



that the 164 common essential genes are not enough to render a viable *E. coli* derivative, since this number is considerably smaller than the number of genes that have been identified in any of the reported approaches to identify essential genes in this bacterium. Therefore, it is possible that the small number of essential genes that are common in all these studies (**Figure 1** and Supplementary Table S2) are a product of the different experimental approaches taken and of the variability of culture conditions that were used.

Another approach for the definition of essential genes is through comparative genomics and gene conservation, looking for genes that have been reported to be essential in one organism, that are conserved among different bacteria (Arigoni et al., 1998; Glass et al., 2009; Alcaraz et al., 2010; Acevedo-Rocha et al., 2013). The use of comparative genomics in the search of essential genes (Luo et al., 2015) has led to the construction of essential genes databases (Ye et al., 2013; Luo et al., 2014; Peng et al., 2017), and to define essential genes in determined bacterial species, such as *Burkholderia cenocepacia* (Juhás et al., 2012b).

## VARIABILITY OF ESSENTIAL GENES IN *E. coli*

The use of *E. coli* as a model organism has been crucial to the fields of molecular biology and synthetic biology (Juhás et al., 2011), so we decided to analyze with some detail the reported essential genes of this bacterium in the genome of different strains. Thus, to determine the existence of non-orthologous gene displacement among different *E. coli* strains we searched for the 303 essential genes defined in the Keio collection in 63 whole genome sequences. We chose these *E. coli* genomes because they are supposedly the best annotated genomes deposited in



the NCBI database (the NCBI accession numbers of the whole genome sequences of these 63 *E. coli* strains are shown in Supplementary Table S3).

The retrieved information from the 63 genomes shows that only four of these genomes (W3110, MG1655, DH1, and BW2952) contained the 303 essential genes described in the Keio collection (see Supplementary Figure S1). We found that 11 genes were absent in more than one of the analyzed genomes and that another 11 genes were absent in only one of these genomes (Supplementary Figure S1 and **Table 1**). Most of the genes that are absent in more than one of the genomes analyzed encode for antitoxins or phage repressor proteins (**Table 1**). Thus it seems likely that these genes are essential in the BW25113 strain used to isolate the mutants of the Keio collection because this strain carries the corresponding toxin or phage that is counteracted by their products, while these toxins and phages seem to be absent in the genomes that lack these supposedly essential genes. Even two K-12 derived strains lack *cohE* that encodes for the  $\epsilon$ 14 prophage repressor (**Table 1**). Also the *waaU* gene, involved in lipopolysaccharide (LPS) biosynthesis, is missing in several of the 63 strains; it is very much likely that strains that lack *waaU* have a different LPS structure than that of strain BW25113.

The *alsK* gene, coding for the allose kinase that is involved in the phosphorylation of the hexose allose is absent in 21 of these 63 genomes. It is difficult to understand why AlsK would be essential for *E. coli* cultured in rich medium. Furthermore, *alsK* deletion mutants constructed by the Datsenko and Wanner (2000) procedure have been reported in strain MG1655 (Xia et al., 2017), showing that AlsK is not an essential protein. Thus, it is clear that *alsK* is not an essential gene as was previously reported (Baba et al., 2006) and coincides with the recently reported results (Goodall et al., 2018).

The data presented in **Table 1** show that some genes are essential only in specific genomic backgrounds and that they do not necessarily encode for fundamental traits for all members of a bacterial species, *E. coli* in this case. Furthermore, these genes are part of the accessory genome by definition, since they are not present in all the *E. coli* strains, and most of them are inherited by HGT (they form part of phages and other mobile genetic elements). These genes represent examples of essential genes (at least in some genetic backgrounds such as the BW25113 strain) that do not form part of the core-genome and highlight the fact that essential genes and genes of the core-genome are not synonymous concepts.

In contrast, all the 11 genes that are absent from only one of the genomes encode for metabolic functions, and in several instances, the activity of the encoded protein plays a central role in a fundamental cell processes, like DNA replication or protein translation or secretion (**Table 1**).

The apparent absence of essential genes in any of the 63 *E. coli* genomes included in this review could be due to non-orthologous gene-displacement, or to sequencing or annotation mistakes as has been reported previously (Charlebois and Doolittle, 2004). Surprisingly, a detailed genome analysis showed that multiple missing essential genes (10) by BLAST search were indeed present, as evidenced by synteny analysis. The BLAST algorithm was not able to detect the genes because of sequencing errors

in the reported genome sequences resulted in frameshifts that resulted in apparent absence of the genes (see Supplementary Figure S2).

The only case where there is a real deletion of any of the essential genes defined by Baba et al. (2006) is the *bcsB* gene (**Table 1**), which forms part of the *bcs* operon that encodes proteins involved in the synthesis of cellulose. The *bcsB* gene is absent as part of a deletion spanning 12 kb of the chromosome of the commensal HS strain (Rasko et al., 2008) which comprises most of the genes of the *bcs* operon. The *bcsB* gene was not found

**TABLE 1** | Essential genes defined in the Keio collection that are absent in the 63 *Escherichia coli* whole genome sequences analyzed (refers to data shown in **Figure 1**).

Gene name	Assigned function of gene product	Number of strain where the gene is absent*
<i>bcsB</i>	Cellulose biosynthesis	1 (HS)
<i>dxr</i>	Participates in isoprenoid synthesis, catalyzing the conversion of 1-deoxy-D-xylulose-5-phosphate (DXR) to 2-methyl-D-erythritol-4-phosphate (MEP)	1 (CFT073)
<i>dnaA</i>	Chromosomal replication initiator protein DnaA	1 (CFT073)
<i>infC</i>	Factor3 for the initiation of protein translation	1 (P12B)
<i>dnaX</i>	DNA polymerase III/DNA elongation factor III, tau and gamma subunits	1 (P12B)
<i>mukB</i>	Chromosome condensin MukBEF, ATPase and DNA-binding subunit	1 (P12B)
<i>rplE</i>	Ribosomal protein E of the large ribosome subunit	1 (P12B)
<i>serS</i>	Seryl-tRNA synthetase (SerRS)	1 (P12B)
<i>plsB</i>	Membrane-bound glycerol-3-phosphate acyltransferase catalyzes the first step in phospholipid biosynthesis	1 (536)
<i>secA</i>	Couples ATP hydrolysis with protein secretion through the cell membrane	1 (UM146)
<i>trmD</i>	SAM-dependent tRNA m(1)G37 methyltransferase	1 (APEC 01)
<i>chpS</i>	ChpS antitoxin; toxin is ChpB	6
<i>maze</i>	MazE antitoxin; toxin is MazF	7
<i>yelM</i>	YelM antitoxin; toxin is YoeB	15
<i>alsK</i>	D-allose kinase	21
<i>cohE</i>	Ci-like repressor of $\epsilon$ 14 prophage	45
<i>racR</i>	Rac prophage repressor	50
<i>waaU</i>	Lipopolysaccharide core biosynthesis	55
<i>mqsA</i>	MqsA antitoxin; toxin is MqsR	50
<i>yibJ</i>	Putative protein belonging to the "rearrangement hot spot" (RHS) family	24
<i>ydlB</i>	Qin prophage predicted protein	18
<i>dicA</i>	Transcriptional repressor of <i>dicB</i> , Qin prophage	24

\*In the case where the gene is absent only in one strain, the name of the strain is shown in parenthesis.

To perform this search, the encoded protein sequence of each of the 303 *E. coli* K-12 essential genes defined by Baba et al. (2006) was used as a query to perform local alignments (using TBLASTN) against the complete genome sequences of the 63 *E. coli* strains used in this study. In the case where one gene seemed to be absent, it was further searched in the corresponding K-12 locus to determine whether the whole sequence was missing, or whether it was present but having a partial deletion or frameshift mutation compared with the K-12 genome (all data were obtained from the NCBI database). Missing essential genes in single genomes were also searched based on the synteny of the *E. coli* chromosomes using the Artemis program (Carver et al., 2012).

to be essential in the recently reported saturated transposon and sequencing strategy (Goodall et al., 2018).

The high frequency of sequencing errors in these 63 *E. coli* whole genomes shows that the majority of genome sequences that are available in databases are prone for false negatives when searching them with sequence homology tools in the comparative genomics studies of essential genes.

The definition of the core-genome that relies on homologous sequences that are present in all individuals of a particular species cannot be achieved if a high frequency of sequencing and annotation mistakes is populating databases. This sequence genome quality problem represents a relevant drawback in the study of essential genes using the available databases constructed for this purpose (Ye et al., 2013; Luo et al., 2014; Peng et al., 2017), and is an obstacle to evaluate the extent of non-orthologous gene displacement that has occurred in bacterial genomes.

## CONCLUDING REMARKS

The aim of this review is to address the high variability of essential bacterial genes in relation to bacterial evolution, and the role that comparative genomics plays in this field (Arigoni et al., 1998; Glass et al., 2009; Alcaraz et al., 2010; Acevedo-Rocha et al., 2013). We highlight the fact that the sequence conservation of essential genes among different phylogenetic groups of bacteria, and even among the same bacterial species, has been reported to be very low (Charlebois and Doolittle, 2004; Lagesen et al., 2010; Juhas et al., 2011).

The low level of sequence conservation among essential genes has been explained by the existence of a phenomenon called non-orthologous gene displacement that refers to the natural selection of essential functions encoded by genes with no sequence homology. These non-orthologous essential genes hence will not belong to the core-genome, which is the genetic information that has been used to build all bacterial taxonomy. To determine the extent of non-orthologous gene displacement that occurs in different bacterial species is a matter of great importance for the field of bacterial evolution.

In this mini-review we addressed the situation of the 303 essential *E. coli* genes defined by Baba et al. (2006), trying to determine whether non-orthologous gene displacement could be documented in this bacterial species. We show that the essentiality of some genes depends on the presence of other non-essential genes (like antitoxin and phage repressor genes), and that this is an important source of variability for the presence

## REFERENCES

- Acevedo-Rocha, C. G., Fang, G., Schmidt, M., Ussery, D. W., and Danchin, A. (2013). From essential to persistent cells: a functional approach to constructing synthetic life. *Trends Genet.* 29, 273–279. doi: 10.1016/j.tig.2012.11.001
- Alcaraz, L. D., Moreno-Hagelsieb, G., Eguarte, L. E., Souza, V., Herrera-Estrella, L., and Olmedo, G. (2010). Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* 11:332. doi: 10.1186/1471-2164-11-332

of essential genes. Another important source of variation is due to sequencing mistakes that make the essential genes invisible to commonly used bioinformatics tools (Supplementary Figure S2). The problem of sequencing and annotation mistakes in the use of comparative genomics for the study of essential genes has been reported previously (Charlebois and Doolittle, 2004). No case of non-orthologous gene displacement was documented.

It is a challenge for the research in the field of essential genes to identify the phylogenetic origin of essential genes that are encoded by non-orthologous sequences. To perform reliable comparative genomics analysis it is imperative to improve databases aiming for high quality sequence of genomes and assemblies so it is possible to identify interesting features like non-orthologous gene displacement.

## AUTHOR CONTRIBUTIONS

EM-C made a bibliographic search for articles related to the theme of this mini-review and constructed **Figure 1**. GS-C conceived and designed the work. L-DA, G-YP-S, EM-C, and HB performed the database searches. GS-C, LS-G, L-DA, G-YP-S, and HB made substantial contributions for the analysis and interpretation of the information. GS-C, LS-G, EM-C, L-DA, and HB participated in drafting the work and revised it critically. GS-C, LS-G, EM-C, L-DA, G-YP-S, and HB approved the final version of the work, and agreed to be accountable for its content.

## ACKNOWLEDGMENTS

The authors acknowledge the participation of Karla Manzano, Fernando Riveros, Enrique Morett, and Pablo Vinuesa in the preliminary analysis of the 63 *E. coli* whole genome sequences. EM-C is a doctoral student of Programa de Maestría y Doctorado en Ciencias Bioquímicas, Universidad Nacional Autónoma de México (UNAM). This study was performed in partial fulfillment of the requirements of EM-C Ph.D. degree, and he received a fellowship from CONACYT (CVU-483310).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.01059/full#supplementary-material>

- Arigoni, F., Talabot, F., Peitscht, M., Edgerton, M. D., Meldrum, E., Allet, E., et al. (1998). A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.* 16, 851–856. doi: 10.1038/nbt0998-851
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K12 in frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2:0008. doi: 10.1038/msb4100050
- Bergmiller, T., Ackermann, M., and Silander, O. K. (2012). Patterns of evolution conservation of essential genes correlate with the conservability. *PLoS Genet.* 8:e1002803. doi: 10.1371/journal.pgen.1002803

- Blath, J., Eldon, B., Gonzalez-Casanova, A., Kurt, N., and Wilke-Berenguer, M. (2015). Genetic variability under the seed bank coalescent. *Genetics* 200, 921–934. doi: 10.1111/mec.12039
- Blath, J., González-Casanova, A., Kurt, N., and Spano, D. (2013). The ancestral process of long-range seed bank models. *J. Appl. Probab.* 50, 741–759. doi: 10.1239/jap/1378401233
- Carver, T., Harris, S. R., Berriman, M., Parkhill, J., and McQuillan, J. A. (2012). Artemis: an integral platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464–469. doi: 10.1093/bioinformatics/btr703
- Charlebois, L. R., and Doolittle, W. (2004). Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res.* 14, 2469–2477. doi: 10.1101/gr.3024704
- Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287. doi: 10.1126/science.1123061
- Dagan, T., Artzy-Randrup, Y., and Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10039–10044. doi: 10.1073/pnas.0800679105
- Darmoh, E., and Leach, D. R. F. (2014). Bacterial genome instability. *Microbiol. Mol. Biol. Rev.* 78, 1–39. doi: 10.1128/MMBR.00035-13
- Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6640–6645. doi: 10.1073/pnas.120163297
- Fang, G., Rocha, E., and Danchin, A. (2005). How essential are nonessential genes? *Mol. Biol. Evol.* 22, 2147–2156. doi: 10.1093/molbev/msi211
- Fang, G., Rocha, E., and Danchin, A. (2008). Persistence drives gene clustering in bacterial genomes. *BMC Genomics* 9:4. doi: 10.1186/1471-2164-9-4
- Fraser, C., Alm, E. J., Polz, M. F., Sprat, B. G., and Hanege, W. P. (2009). The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323, 741–746. doi: 10.1126/science.1159388
- Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balázs, G., Ravasz, E., Daugherty, M. D., et al. (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG 1655. *J. Bacteriol.* 185, 5673–5684. doi: 10.1128/JB.185.19.5673-5684.2003
- Gil, R., Silva, F. J., Peretó, J., and Moya, A. (2004). Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* 68, 518–537. doi: 10.1128/MMBR.68.3.518-537.2004
- Glass, J. I., Hutchison, I. I. C. A., Smith, H. O., and Venter, J. C. (2009). A system biology tour de force for a near-minimal bacterium. *Mol. Syst. Biol.* 5:330. doi: 10.1038/msb.2009.89
- Glass, J. I., Merryman, C., Wise, K. S., Hutchison, I. I. C. A., and Smith, H. O. (2017). Minimal cells real and imagined. *Cold Spring Harb. Perspect. Biol.* 9:a023861. doi: 10.1101/cshperspect.a023861
- González-Casanova, A., Aguirre-von-Wobeser, E., Espín, G., Servín-González, L., Kurt, N., Spano, D., et al. (2014). Strong seed-bank effects in bacterial evolution. *J. Theor. Biol.* 356, 62–70. doi: 10.1016/j.jtbi.2014.04.009
- Goodall, E. C. A., Robison, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., et al. (2018). The essential genome of *Escherichia coli* K-12. *mBio* 9, e2096-17. doi: 10.1128/mBio.02096-17
- Grazziotin, A. L., Vidal, N. M., and Venancio, T. M. (2015). Uncovering major genomic features of essential genes in Bacteria and a methanogenic Archaea. *FEBS J.* 282, 3395–3411. doi: 10.1111/febs.13350
- Guimarães, L. C., Florczak-Wyspianska, J., de Jesus, L. B., Viana, M. V., Silva, A., Ramos, R. T., et al. (2015). Inside the pan-genome: methods and software overview. *Curr. Genomics* 16, 245–252. doi: 10.2174/1389202916666150423002311
- Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., et al. (2005). Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.* 55, 137–149. doi: 10.1111/j.1365-2958.2004.04386.x
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., and Castelle, C. J. (2016). A new view of the tree of life. *Nat. Microbiol.* 1:16048. doi: 10.1038/nmicrobiol-2016.48
- Hutchison, C. A. I. I., Chuang, R. Y., Niskov, V. N., Assad-García, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351:aad6253. doi: 10.1126/science.aad6253
- Ibberson, C. B., Stacy, A., Fleming, D., Dees, J. L., Rumbaugh, K., Gilmore, M. S., et al. (2017). Co-infecting microorganisms dramatically alter pathogen gene essentiality during polymicrobial infection. *Nat. Microbiol.* 2:17079. doi: 10.1038/nmicrobiol.2017.79
- Ish-Am, O., Kristensen, D. M., and Ruppín, E. (2015). Evolutionary conservation of bacterial essential metabolic genes across all bacterial culture media. *PLoS One* 10:e0123785. doi: 10.1371/journal.pone.0123785
- Juhas, M., Eberl, L., and Church, G. M. (2012a). Essential genes as antimicrobial targets and cornerstones of synthetic biology. *Trends Biotechnol.* 30, 601–607. doi: 10.1016/j.tibtech.2012.08.002
- Juhas, M., Stark, M., Von Mering, C., Lumjaktase, P., Crook, D. W., Valvano, M. A., et al. (2012b). High confidence prediction of essential genes in *Burkholderia cenocepacia*. *PLoS One* 7:e40064. doi: 10.1371/journal.pone.0040064
- Juhas, M., Eberl, L., and Glass, J. I. (2011). Essence of life: essential genes of minimal genomes. *Trends Cell Biol.* 21, 562–568. doi: 10.1016/j.tcb.2011.07.005
- Juhas, M., Reuss, D. R., Zhu, B., and Commichau, F. M. (2014). *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cells factories after one decade of genome engineering. *Microbiology* 160, 2341–2351. doi: 10.1099/mic.0.079376-0
- Kato, J., and Hashimoto, M. (2007). Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* 3:132. doi: 10.1038/msb4100174
- Koonin, E. V., Mushegian, A. R., and Bork, K. P. (1996). Non-orthologous gene displacement. *Trends Genet.* 12, 334–336. doi: 10.1016/0168-9525(96)20010-1
- Lagesen, K., Ussery, D. W., and Wassenaar, T. M. (2010). Genome update: the 1000th genome—a cautionary tale. *Microbiology* 156, 603–608. doi: 10.1099/mic.0.038257-0
- Lang, A. S., Zhaxybayeva, O., and Beatty, J. T. (2012). Gene transfer agents: phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* 10, 472–482. doi: 10.1038/nrmicro2802
- Lee, J. H., Sung, B. H., Kim, M. S., Blattner, F. R., Yoon, B. H., Kim, J. H., et al. (2009). Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. *Microb. Cell Fact.* 8:2. doi: 10.1186/1475-2859-8-2
- Luo, H., Gao, F., and Lin, Y. (2015). Evolutionary conservation analysis between the essential and nonessential genes in bacterial genomes. *Science Rep.* 5:13210. doi: 10.1038/srep13210
- Luo, H., Lin, Y., Gao, F., Zhang, C.-T., and Zhang, R. (2014). DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.* 42, D574–D580. doi: 10.1093/nar/gkt1131
- Medini, D., Donati, C., Tettelin, H., Massignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006
- Meng, J., Kanzaki, G., Meas, D., Lam, C. K., Crummer, H., Tain, J., et al. (2012). A genome-wide inducible phenotypic screen identifies antisense RNA constructs silencing *Escherichia coli* essential genes. *FEMS Microbiol. Lett.* 329, 45–53. doi: 10.1111/j.1574-6968.2012.02503.x
- Mira, A., Pushker, R., Legault, B. A., Moreira, D., and Rodríguez-Valera, F. (2004). Evolutionary relationships of *Fusobacterium nucleatum* based on phylogenetic analysis and comparative genomics. *BMC Evol. Biol.* 4:50. doi: 10.1186/1471-2148-4-50
- Moya, A., Gil, R., Latorre, A., Peretó, J., Garcillán-Barcia, M. P., and De la Cruz, F. (2009). Towards minimal bacterial cells: evolution vs. design. *FEMS Microbiol. Rev.* 33, 225–235. doi: 10.1111/j.1574-6976.2008.00151.x
- Oren, I.-A., Kristensen, D. M., and Ruppín, E. (2015). Evolutionary conservation of bacterial essential metabolic genes across all bacterial culture media. *PLoS One* 10:e0123785. doi: 10.1371/journal.pone.0123785
- Peng, C., Lin, Y., Luo, H., and Gao, F. (2017). A comprehensive overview of online resources to identify and predict bacterial essential genes. *Front. Microbiol.* 8:2331. doi: 10.3389/fmicb.2017.02331
- Posfai, G., Plunkett, I. I. G., Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., et al. (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science* 312, 1044–1046. doi: 10.1126/science.1126439
- Rasko, D. A., Rosovitz, M. J., Myers, G. S., Mongodin, E. F., Fricke, W. F., Gajer, P., et al. (2008). The pangenome structure of *Escherichia coli*: comparative analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* 190, 6881–6893. doi: 10.1128/JB.00619-08

- Rusmini, R., Vecchiotti, D., Macchi, R., Vidal-Aroca, F., and Bertoni, G. (2014). A shotgun antisense approach to the identification of novel essential genes in *Pseudomonas aeruginosa*. *BMC Microbiol.* 14:24. doi: 10.1186/1471-2180-14-24
- Shoemaker, W., and Lennon, J. (2018). Evolution with a seed bank: the population genetic consequences of microbial dormancy. *Evol. Appl.* 11, 60–75. doi: 10.1111/eva.12557
- Sleator, R. D. (2010). The story of *Mycoplasma mycoides* JCVI-syn1.0: the forty million dollars microbe. *Bioeng. Bugs* 1, 229–230. doi: 10.4161/bbug.1.4.12465
- Soberón-Chávez, G., Alcaraz, L. D., Morales, E., Ponce-Soto, G. Y., and Servín-González, L. (2017). The transcriptional regulators of the CRP family regulate different essential bacterial functions and can be inherited vertically and horizontally. *Front. Microbiol.* 8:959. doi: 10.3389/fmicb.2017.00959
- Soucy, S. M., Huang, J., and Gogarten, J. P. (2015). Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* 16, 472–482. doi: 10.1038/nrg3962
- Tazzyman, S. J., and Bonhoeffer, S. (2015). Why There Are No Essential Genes on Plasmids. *Mol. Biol. Evol.* 32, 3079–3088.
- Tettelin, H., Riley, D., Cattuto, C., and Medini, C. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 12, 472–475. doi: 10.1016/j.mib.2008.09.006
- Turner, K. H., Wessel, A. K., Palmer, G. C., Murray, J. L., and Whiteley, M. (2015). Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4110–4115. doi: 10.1073/pnas.1419677112
- Xia, T., Sriram, N., Lee, S. A., Altman, R., Urbauer, J. L., Altman, E., et al. (2017). Glucose consumption in carbohydrate mixtures by phosphotransferase-system mutants of *Escherichia coli*. *Microbiology* 163, 866–877. doi: 10.1099/mic.0.000480
- Yamazaki, Y., Niki, H., and Kato, J. (2008). Profiling of *Escherichia coli* chromosome database. *Methods Mol. Biol.* 416, 385–389. doi: 10.1007/978-1-59745-321-9\_26
- Ye, Y.-N., Hua, Z.-G., Huang, J., Rao, N., and Guo, F.-B. (2013). CEG: a database of essential gene clusters. *BMC Genomics* 14:769. doi: 10.1186/1471-2164-14-769
- Zhang, X., Peng, C., Zhang, G., and Gao, F. (2015). Comparative analysis of essential genes in prokaryotic genomic islands. *Sci. Rep.* 5:12561. doi: 10.1038/srep12561
- Zhaxybayeva, O., Swithersb, K. S., Lapierrrec, P., Fournierb, G. P., Bickhartb, D. M., DeBoyd, R. T., et al. (2009). On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5865–5870. doi: 10.1073/pnas.0901260106

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Martínez-Carranza, Barajas, Alcaraz, Servín-González, Ponce-Soto and Soberón-Chávez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.