MICROBIOLOGY

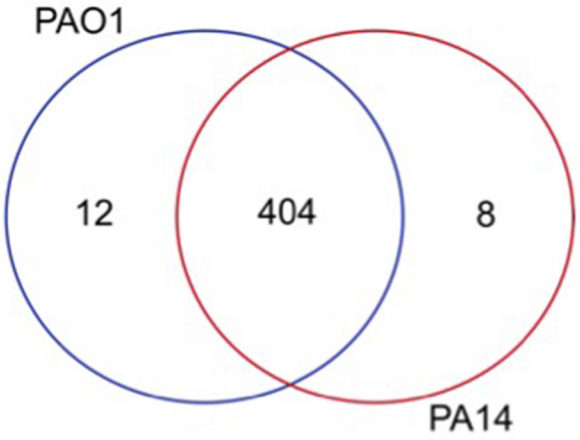
RESEARCH ARTICLE

Martínez-Carranza *et al.*, *Microbiology*

DOI 10.1099/mic.0.000833

Evolution of bacteria seen through their essential genes: the case of *Pseudomonas aeruginosa* and *Azotobacter vinelandii*

Enrique Martínez-Carranza[1], Gabriel-Yaxal Ponce-Soto[2], Luis Servín-González[1], Luis David Alcaraz[3] and Gloria Soberón-Chávez[1,*]



Most of the *P. aeruginosa* genes that have been defined as essential (404/424) are shared by PAO1 and PA14 strains

Graphical abstract

Abstract

Pseudomonas aeruginosa is a metabolically versatile bacterium and also an important opportunistic pathogen. It has a remarkable genomic structure since the genetic information encoding its pathogenicity-related traits belongs to its core-genome while both environmental and clinical isolates are part of the same population with a highly conserved genomic sequence. Unexpectedly, considering the high level of sequence identity and homologue gene number shared between different *P. aeruginosa* isolates, the presence of specific essential genes of the two type strains PAO1 and PA14 has been reported to be highly variable. Here we report the detailed bioinformatics analysis of the essential genes of *P. aeruginosa* PAO1 and PA14 that have been previously experimentally identified and show that the reported gene variability was owed to sequencing and annotation inconsistencies, but that in fact they are highly conserved. This bioinformatics analysis led us to the definition of 348 *P. aeruginosa* general essential genes. In addition we show that 342 of these 348 essential genes are conserved in *Azotobacter vinelandii*, a nitrogen-fixing, cyst-forming, soil bacterium. These results support the hypothesis of *A. vinelandii* having a polyphyletic origin with a Pseudomonads genomic backbone, and are a challenge to the accepted theory of bacterial evolution.

INTRODUCTION

Pseudomonas aeruginosa is a widely distributed, environmental, metabolically versatile bacterium that belongs to γ-Proteobacteria, but is also an opportunistic human pathogen capable of causing severe chronic lung infections

in cystic fibrosis (CF) patients and different type of infections in immune-compromised individuals [1–3].

The first *P. aeruginosa* genome was reported in the year 2000 [4], and today many studies that focused on the study of its genomic characteristics have been published ([5–8] for example). It has been shown that *P. aeruginosa* strains possess a sizeable circular chromosome varying in size between 6 and 7 mega base-pairs (Mbp) averaging 6175 genes per genome. The pan-genome of this species is composed of 16820 non-redundant genes of which 2503 constitute the core-genome, 9108 genes the accessory genome and 5209 genes are unique [5]. An outstanding characteristic of *P. aeruginosa* genomes is their high degree of genomic conservation between environmental and clinical strains, including those isolated from distant geographical points and dates, even though the phenotypic variability between isolates is high [6, 7].

An additional approach to the study of bacterial genomes and their evolution, besides defining the core, accessory and pan-genome, is the definition of the essential genes of each species under different conditions through functional genomics experiments [9]. Essential genes are defined as those that are necessary for a bacterium to grow in a certain media and thus cannot be inactivated by mutagenesis.

More than 21 800 bacterial genome sequences are available in the National Center for Biotechnology Information (NCBI) database, this vast repository of genomes was possible by the improvement in sequencing technologies and of the bioinformatics tools. The expanding amount of genomic information has fuelled comparative genomics to describe essential gene profiles in multiple bacterial genomes both intra- and inter-species [10]. Different reports regarding comparative genomics of essential genes have established the unexpected result that these type of genes are variable among different bacteria [11, 12] and even among strains of the same species [13, 14].

For example, several works that have tried to determine the essential genes of *P. aeruginosa* using different experimental approaches and in different metabolic conditions (Fig. 1a) have concluded that the number of essential genes in the two most studied type strains (PAO1 and PA14) is variable [14–19]. The first analysis of *P. aeruginosa* essential genes was done with the PAO1 strain in 2003 [14], and it estimated a set of around 300 to 400 essential genes, based on the construction of a transposon (Tn) mutagenesis libraries in rich medium, since Tn insertions were not recovered in 678 open reading frames (ORFs), but the authors claimed that some genes might not be mutagenized not because mutations were lethal, but due to the lack of a saturated mutagenesis [14]. In another study, a non-redundant library of *P. aeruginosa* PA14 Tn mutants (PA14 collection) was obtained, and a set of 335 genes were considered as essential in this *P. aeruginosa* strain [15]. In 2013, Skurnik *et al.* characterized the *P. aeruginosa* PA14 essential genes in the mucosal and systemic environments of an infection model both *in vivo* and *in vitro* [16], and 636 PA14 essential genes were proposed. In this work it was also reported that 210 out of the 636 essential genes were absent in the PAO1 genome; thus this result supported the concept that the essential genes were highly variable even among members of the same bacterial species; *P. aeruginosa* in this case [16].

Another report identified 336 essential genes in PAO1 and 434 in PA14 *P. aeruginosa* strains. Among the PA14 essential genes 20 lacked orthologues in PAO1, thus this study also indicated that the essential genes were variable among strains although in a minor proportion (20/434) [18]. Recent work with PAO1 strain reported condition-specific and general essential genes in six culture media [19]; a total of 551 essential genes were identified, among them 199 are condition-specific and 352 are general essential genes meaning that these genes are necessary for growth in all media tested [19]. The comparison of the gene sequence of the reported essential genes that are common to the four more recent studies [15, 16, 18, 19] shows that there are only 48 genes that are common to all these studies (Fig. 1b), a number that is too low to account for a cell to be viable. Most of the shared 48 essential genes identified in these articles [15, 16, 18, 19] belong to the COG group of translation, ribosomal structure and biogenesis, but they also include genes that participate in different cellular functions (Fig. 1c).

The high variability of essential genes among strains belonging to the same species is intriguing and has been explained as the result of non-orthologous gene displacement [11, 12], but it has also been proposed that this apparent variability might be due, at least in part, to sequencing and annotation errors [12, 20]. In the case of *P. aeruginosa,* the high variability of essential genes [16] is difficult to explain considering the high degree of genome conservation among different isolates [6, 7].

This work aims to perform a comprehensive bioinformatics search for the PA14 putative essential genes that were not detected in the PAO1 genome [16], and the 20 PAO1 essential genes that were apparently absent from PA14 chromosome

Received 17 December 2018; Accepted 26 June 2019; Published 03 July 2019
Author affiliations: [1]Departamento de Biología Molecular y Biotecnología, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Ciudad Universitaria, Apdo. Postal 70228, C. P. 04510, CDMX, Mexico; [2]Institute for Bio- and Geosciences (IBG-2: Plant Sciences), Forschungszentrum Jülich, Wilhelm Johnen Straße, Jülich, Germany; [3]Facultad de Ciencias, Universidad Nacional Autónoma de México, C.P. 04510, CDMX, Mexico.
*Correspondence: Gloria Soberón-Chávez, gloria@biomedicas.unam.mx
Keywords: Bacterial evolution; *Pseudomonas aeruginosa*; Azotobacter vinelandii; essential genes; core-genome.
Abbreviations: CF, cystic fibrosis; HGT, horizontal gene transfer; Mbp, mega base-pairs; NCBI, National Center for Biotechnology Information; ORFs, open reading frames.
One supplementary table is available with the online version of this article.

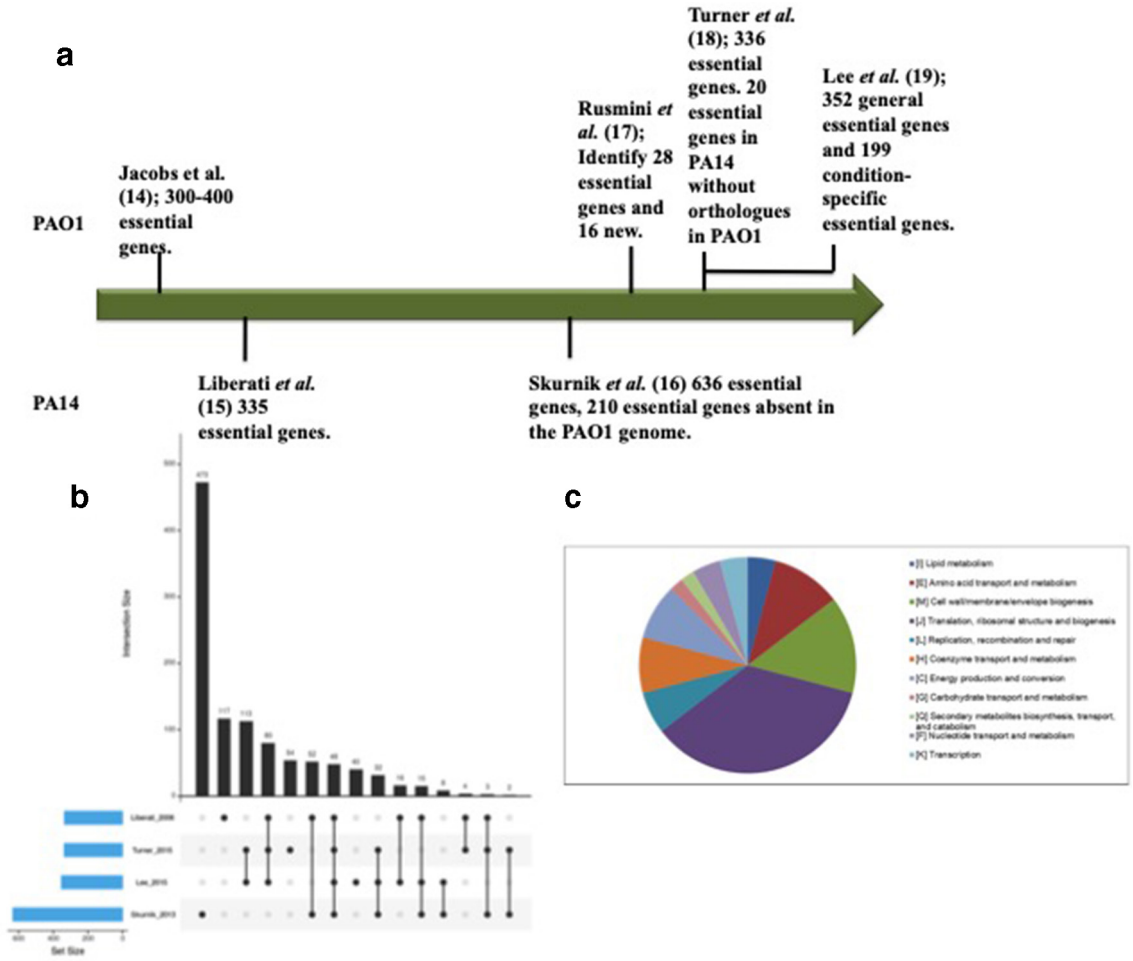


Fig. 1. Reported variability of *P. aeruginosa* essential genes. (a) Timeline of studies made to determine essential genes with different aims and with several experimental approaches in *P. aeruginosa* PAO1 and PA14. (b) UpsetR image of the essential genes reported by Liberati *et al.* [15], Skurnik *et al.* [16], Turner *et al.* [18] and Lee *et al.* [19]. These comparisons show that there are 48 genes that were reported as essential in the four studies. (c) Analysis of the function of the 48 genes that are shared by the four analysed articles using COG classification.

[18]. The results reported here show that the essential genes are highly conserved among these *P. aeruginosa* type strains PAO1 and PA14. Thus it can be concluded that the PAO1 352 general essential genes reported recently [19] are representative of the essential genes of the *P. aeruginosa* species.

It has been reported that the nitrogen-fixing soil bacterium *Azotobacter vinelandii* is very closely related to *P. aeruginosa* [21, 22] even though these bacteria have different biological characteristics. The phenotypic diversity of these two bacterial species with highly conserved genomes has been explained by the polyphyletic origin of *A. vinelandii* that has acquired genetic information essential for its particular biology by horizontal gene transfer (HGT) from bacteria that have a distant phylogenetic relation while conserving a Pseudomonads core-genome [21–23]. To further confirm the close phylogenetic relation of *A. vinelandii* and *P. aeruginosa* we searched, using the same bioinformatics approach, for the 352 *P. aeruginosa* general essential genes [19] in the genome of *A. vinelandii*, and found that only ten of these genes are not present. The unusual *A. vinelandii* genomic constitution provides evidence of the importance of HGT in the evolution of this bacterial species, as has been previously described [23].

METHODS

Searching for orthologous essential genes in the *P. aeruginosa* PA14, PAO1 and *A. vinelandii* DJ genomes

The comparison of the reported essential genes in PAO1 and PA14 [15, 16, 18, 19] shown in Fig. 1 was made using usetR [24]. The function of genes was made using the COG classification [25].

The proteomes we used in this work were retrieved from NCBI: *P. aeruginosa* PAO1 NC_002516, *P. aeruginosa* PA14 NC_008463 and *A. vinelandii* DJ CP001157. The three query

sequences to search proteins in these proteomes were: (I) the 210 essential genes predicted in *P. aeruginosa* PA14 that were reported to be absent in the PAO1 genome [16]; (II) the previously reported set of 336 *P. aeruginosa* PAO1 essential genes for growth in the sputum of CF patients [18]; (III) the 352 general essential genes of *P. aeruginosa* PAO1 recently reported [19]. Also, the PAO1 essential genes of set III were searched in the *A. vinelandii* DJ genome.

The procedure that we used was as follows: the previously reported essential genes in *P. aeruginosa* PAO1 and PA14 (three sets) were retrieved, translated into proteins and used as queries in BLASTP version 2.2.26+searches [26], using the whole proteomes of PAO1, PA14 or *A. vinelandii* DJ as databases in each case, with an e-value cut-off of 1e-5. The sequence hits were considered homologous if they had at least 90 % of the query sequence length and at least 80 % amino acid identity. The search for each protein was performed using the protein sequence encoded by every essential gene in the proteome of the other strain (each reported PA14 essential gene was searched in the PAO1 chromosome and vice versa). The best hit and coverage of each searched protein defined as the homologous candidate was selected and we found that when any of these genes is present in the genome in which we searched, it showed a complete coverage and almost identical sequence. The alignment of the *P. aeruginosa* PAO1 and *A. vinelandii* DJ chromosomes were analysed using and MAUVE version snapshot_2015-02-13 (http://darlinglab.org/mauve) [27] (Fig. 2). To determine the genomic context in *A. vinelandii* of *P. aeruginosa* essential genes that were detected in this soil bacterium, we manually searched for each gene using the aligned chromosomes shown in Fig. 2. The list of 337 *P. aeruginosa* essential genes that were found in the syntenic regions of the chromosome of *A. vinelandii,* is shown in Table S1 (available in the online version of this article).

RESULTS AND DISCUSSION

Analysis of the 210 *P. aeruginosa* PA14 essential genes reported as absent in the PAO1 genome

We found out after searching for the 210 PA14 essential genes that were reported to be absent in the PAO1 genome [16], using the procedure described in Methods, that only eight of these genes were absent (Table 1). Furthermore, the eight identified genes missing from the PAO1 genome might not be essential, but Tn insertions might not be selected in them since most of them are of a small size (between 147 and 390 nucleotides), decreasing the probability of isolating Tn insertions. Thus the previously reported variability of *P. aeruginosa* PAO1 and PA14 essential genes [16, 18] is not sustained by our analysis.

It is likely that the 210 PA14 genes were reported as absent in PAO1 because of the bioinformatics analysis that was performed, which was based on the annotation and name given to a gene in the database and not to the direct search of the gene or protein sequences [16].

P. aeruginosa PAO1 genes essential for growth in CF sputum are almost all present in the PA14 genome

We searched for the 336 PAO1 genes that have been described as essential for growth in CF sputum [18] in the PA14 genome, and identified a set of 12 genes that are missing in this strain (Table 2). These genes are part of a gene cluster that encodes proteins that participate in O-antigen lipopolysaccharide (LPS) biosynthesis that is not present in PA14 and thus are not essential in PA14 since this strain has a different gene set for LPS biosynthesis [28]. Again in this analysis, we observed a very low variability of essential genes (12/336) among *P. aeruginosa* PAO1 and PA14.

Analysis of the 352 *P. aeruginosa* PAO1 general essential genes in the PA14 genome

A set of 352 general essential genes were proposed [19] after an analysis of the genes in *P. aeruginosa* PAO1 strain that were required for growth in several culture conditions. These 352 genes were not mutated after saturating Tn mutagenesis and selection in different culture conditions [19]. We hypothesized that if these 352 genes really represented general *P. aeruginosa* essential genes, all these genes that were identified in the PAO1 genome would be conserved in the PA14 chromosome. In agreement with this hypothesis we found that only 2 out of the 352 genes were absent in the PA14 strain (Table 3). One of them is involved in O-antigen synthesis (*wbpB*) and the other is a pyocin S2 immunity protein (*imm2*). Thus it seems that in both cases they are really non-essential genes in the PA14 strain, since LPS genes are variable between different strains [28] and PA14 does not produce this pyocin so the immunity protein is not necessary. If the two missing genes in the PA14 are not considered as general essential, it can be suggested that most if not all of the 350 genes detected [19] represent the essential genes set of *P. aeruginosa*.

General essential genes in *P. aeruginosa* are well conserved in its close relative *A. vinelandii*

It has been postulated based on the phylogenetic analysis of 16S rRNA gene sequences and conserved protein alignments, that *A. vinelandii* should be considered as a member of the genus *Pseudomonas* [21, 22]. However, when a large set of proteins is considered (1455 families) *A. vinelandii* was postulated to be a closely related genera, but being an out-group of *Pseudomonas* [29]. These apparent inconsistencies might be due to the fact that in the article where 1455 genes were considered to build the phylogenetic tree of *Pseudomonas* and *A. vinelandii*, they included genes that were directly inherited by vertical gene transfer by both genera, as those coding for 16S rRNA, and the 25 conserved proteins, but also genes inherited by HGT such as those encoding the enzymes for nitrogen fixation [29]. In any scenario, it is interesting that the biological characteristics of *A. vinelandii* diverge significantly from those of *Pseudomonas* and particularly from *P. aeruginosa* (its closest *Pseudomonas* relative according to some reports) [21, 22]. *Pseudomonas* are metabolically versatile cosmopolitan bacteria and *P. aeruginosa* is also an opportunistic pathogen, while *A. vinelandii* is a nitrogen-fixing soil

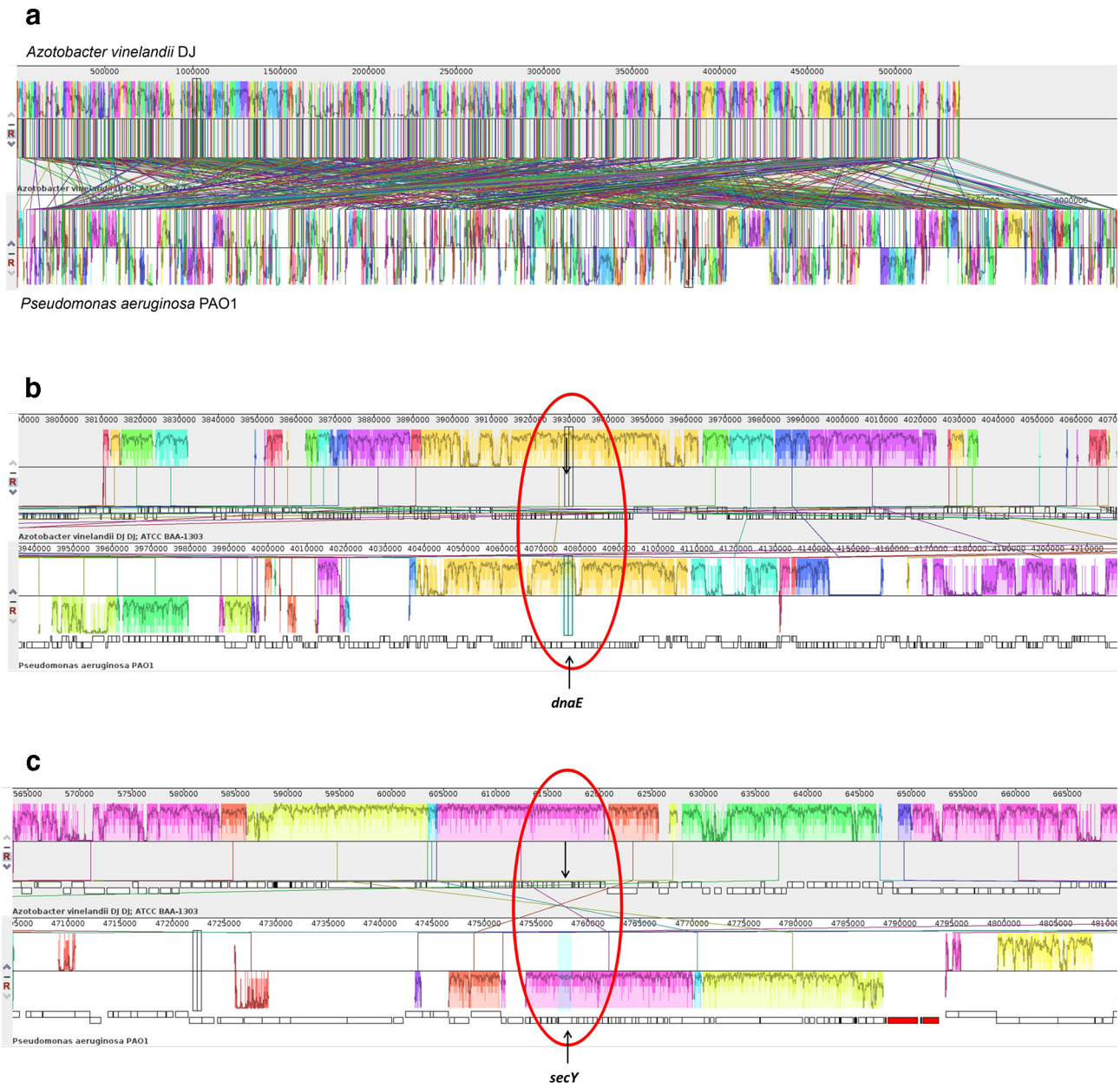


Fig. 2. Conservation of small synthenic regions of *P. aeruginosa* PAO1 and *A. vinelandii* DJ chromosomes. MAUVE alignment of the entire *P. aeruginosa* PAO1 and *A. vinelandii* DJ chromosomes (lines join chromosome segments that are both homologous and syntenic between these genomes) (a). MAUVE alignment of the chromosomal region of *P. aeruginosa* PAO1 and *A. vinelandii* DJ chromosomes where *dnaE* (b) and secY (c) are encoded (colours show regions that are homologous in the two compared chromosomes).

bacterium that forms cysts and exhibits the highest respiratory rate of any living cell.

We consider that it is questionable to reclassify *A. vinelandii* as a new *Pseudomonas* species having biological characteristics that are widely different, solely based on the finding that part of its gene phylogenies show that *A. vinelandii* and *Pseudomonas* share a unique common ancestor. Furthermore, the comparison of all *A. vinelandii* genes with the pan-genome of Pseudomonads showed that almost half of the *A. vinelandii* genes encode proteins showing a high identity of 67 % or higher with a *Pseudomonas* homologue [23], but that the other half of its genome that is not derived from *Pseudomonas* contains genes encoding traits that are important for its biology. Some examples of genes inherited by *A. vinelandii* through HGT from distantly related bacterial genera that are important for biology of this bacterium include those encoding for alkyl-resorcinols synthesis, a lipid that forms part of the cyst envelope, a gene encoding an alginate lyase that is essential for cyst germination, and some of the genes coding for its flagellum synthesis, among others

Table 1. Set of essential genes reported in *P. aeruginosa* PA14 [16] that we found to be missing in *P. aruginosa* PAO1. The best hit and E-value in each sequence analysis is shown

Query data sequences					Best hit data		
Gene annotation	**Function**	**COG**	**Query sequence length (aa)**	**E-value**	**Best hit length sequence (aa)**	**Best hit gene anotation**	**Sequence proportion of alignment: query part/total hit sequence**
PA14_59110*	Hypothetical protein	Function unknown	93	Does not apply	No hits found	No hits found	Does not apply
PA14_58980	Hypothetical protein	Function unknown	74	1.6	337	PA3567	21/337
PA14_58940	Hypothetical protein	Function unknown	49	2.3	923	PA0572	20/923
PA14_31250	Hypothetical protein	Function unknown	130	0.4	142	*ohr*	40/142
PA14_31080	peptidase – peptidase_S26 superfamily	Conjugal transfer protein	191	0.003	179	PA1303	109/179
PA14_15530	*uhpA* – two component system response regulator	Signal trasduction mechanisms	91	0.73	137	PA3289	28/137
PA14_15470	*merP* – Mercury resistance system periplasmic binding protein	Defense mechanisms	91	5e-6	792	PA3920	80/792
PA14_14310*	Prob. transcriptional regulator, ArsR family	Transcriptional regulators	101	Does not apply	No hits found	No hits found	Does not apply

*ORFs without any homology in the PAO1 genome.

[23]. This unusual *A. vinelandii* genomic constitution has been postulated to mean that this soil bacterium has a Pseudomonads genomic backbone that includes rRNA and at least 25 conserved proteins [21, 22], but also a high proportion of genes inherited by HGT that includes genes that encode for important *A. vinelandii* biological characteristics [23]. Thus it has been claimed that this bacterium has a polyphyletic origin [23].

In order to study the presence in *A. vinelandii* of *P. aeruginosa* essential genes, we searched for the 352 general essential genes of *P. aeruginosa* previously reported for PAO1 (19) in the genome of *A. vinelandii* DJ. We found that 342 of the PAO1 general essential genes have at least 90 % of the query sequence length and at least 80 % amino acid identity, and that only ten genes were not present in the *A. vinelandii* genome, but two of them (*wbpB* and *imm2*, Table 4) were particular to the PAO1 strain and cannot be considered as general essential *P. aeruginosa* genes as discussed above. Moreover, two out of the eight general *P. aeruginosa* genes that are absent in *A. vinelandii* (*nadE* and PA0092, Table 4) are not essential genes, since Tn-insertion mutants have been isolated in the PAO1 strain [14]. These analyses led us to define 348 general essential *P. aeruginosa* genes.

Furthermore, considering the other six PAO1 essential genes that were not found in the *A. vinelandii* DJ chromosome, we found that the PA4674 gene encodes the HigA antitoxin (Table 4), therefore as the toxin is not present in the *A. vinelandii* genome, this gene cannot be consider as an essential gene in *A. vinelandii*. The same seems to be the case for the *P. aeruginosa* PA0906 gene that confers protection to a lysis phenotype and resembles a phage or pyocin repressor, and *prtN* that encodes a repressor of pyocin genes [30] (Table 4). In summary, of the 348 general putative *P. aeruginosa* essential genes (considering that *nadE* and PA0092 are not truly essential and *wbpB* and *imm2* are not generally essential) only six genes cannot be detected in the *A. vinelandii* genome, but three (PA4674, PA0906 and *prtN*) seem to be conditionally essential depending on the presence of a toxin, a phage or a pyocin. The remaining three genes are annotated as hypothetical proteins (PA0442, PA4405 and PA4685) and their function remains to be determined.

The extraordinarily high degree of essential gene conservation between *A. vinelandii* and *P. aeruginosa,* which is high even if these two bacterial species were both members of the *P. aeruginosa* species, strongly argues in favour of the proposed polyphyletic origin of *A. vinelandii*, having a genomic backbone derived through vertical lineage from *Pseudomona*s in which genes inherited by HGT from distantly phylogenetic bacteria have been incorporated and are responsible for encoding fundamental traits for its biology [23].

The general structure of the *A. vinelandii* chromosome compared with the *P. aeruginosa* chromosome shows that its *Pseudomonas* genomic backbone is interdispersed within regions that have no homology nor synteny (Fig. 2a). The

Table 2. Essential genes reported in *P. aeruginosa* PAO1 (18) that we found to be missing in the PA14 strain

Query data sequences					Best hit data		
Gene annotation	**Function**	**COG**	**Query sequence length (aa)**	**E-value**	**Best hit length sequence (aa)**	**Best hit gene anotation**	**Sequence proportion of alignment: query part/ total hit sequence**
wbpJ	Probable glycosyl transferase	Cell wall/LPS/ envelope	413	0.21	209	PA14_31750	40/209
wbpI	UDP-N-acetylglucosamine 2-epimerase/ polysaccharide biosynthetic process	Cell wall/LPS/ envelope	354	1e-41	378	PA14_23370	309/378
wbpH	Probable glycosyltransferase/O antigen biosynthetic process	Cell wall/LPS/ envelope	373	2e-9	381	PA14_71910	296/381
wbpG	LPS biosynthesis protein/lipopolysaccharide biosynthetic process	Cell wall/LPS/ envelope	377	3.2	305	PA14_57720	52/305
hisF2	Imidazoleglycerol-phosphate synthase, cyclase subunit/O antigen biosynthetic process	Amino acid transport and metabolism	251	3e-46	256	PA14_67880/*hisF1*	227/256
wzx	O-antigen translocase/O antigen biosynthetic process	Cell wall/LPS/ envelope	411	0.054	688	PA14_21450	71/688
wzy	B-band O-antigen polymerase/lipopolysaccharide biosynthetic process	Cell wall/LPS/ envelope	438	3.8	399	*ftsW*	131/399
wbpE	UDP-2-acetamido-2-dideoxy-D-ribo-hex-3-uluronic acid transaminase/O antigen biosynthetic process	Cell wall/LPS/ envelope	359	2e-48	382	PA14_18370	344/382
wbpD	UDP-2-acetamido-3-amino-2,3-dideoxy-D-glucuronic acid N-acetyltransferase/ polysaccharide biosynthetic process	Cell wall/LPS/ envelope	191	2e-08	258	PA14_14700	136/258
wbpB	UDP-2-acetamido-2-deoxy-D-glucuronic acid 3-dehydrogenase/O antigen biosynthetic process	Cell wall/LPS/ envelope	316	0.009	347	PA14_02930	200/347
wbpA	UDP-N-acetyl-D-glucosamine 6-dehydrogenase/ polysaccharide biosynthetic process	Cell wall/LPS/ envelope	436	9e-54	422	PA14_23380	410/422

interdisperse *A. vinelandii* chromosome regions that are not syntenic to *P. aeruginosa* cannot be considered as genomic islands [31] since they encode for genes that are fundamental for its biology and their GC content is not different from the rest of the genome [23]. Furthermore, there are only around ten reported *A. vinelandii* genomic islands (http://www.pathogenomics.sfu.ca/islandviewer/accession/NC_012560.1/) and they do not coincide with the chromosomal regions that show homology and synteny with *P. aeruginosa*. Furthermore, we found that almost all the 342 *A. vinelandii* genes that

Table 3. General essential genes reported *P. aeruginosa* PAO1 (19) that we found to be missing in *P. aeruginosa* PA14

Query data sequences					Best hit data		
Gene annotation	**Function**	**COG**	**Query sequence length (aa)**	**E-value**	**Best hit length sequence (aa)**	**Best hit gene anotation**	**Sequence proportion of alignment: Query part/ total hit sequence**
imm2	pyocin S2 immunity protein/bacteriocin immunity	Adaptation/protection/defense mechanisms	87	1.0	2352	PA14_21020	25/2352
wbpB	UDP-2-acetamido-2-deoxy-D-glucuronic acid 3-dehydrogenase/O antigen biosynthetic process	Putative enzymes; cell wall/ LPS/capsule	316	0.009	347	PA14_02930	200/347

Table 4. General essential *P. aeruginosa* genes that are absent in the *A. vinelandii* DJ genome

Query data sequences					Best hit data		
Gene annotation	**Function**	**COG**	**Query sequence length (aa)**	**E-value**	**Best hit length sequence (aa)**	**Best hit gene anotation**	**Sequence proportion of alignment: query part/total hit sequence**
PA0092*	Protein secretion by the type VI secretion system/toxin-antitoxin pair type II binding	Adaptation/protection/defense mechanisms	94	4.0	405	*glcF*	61/405
PA0442	Hypothetical protein	Hypothetical/unclassified	38	3.5	725	*glcB*	22/725
prtR	transcriptional regulator PrtR/negative regulation of secondary metabolite biosynthetic process	Transcriptional regulators	256	0.076	129	Avin_04120	69/129
PA0906	lysis phenotype repressor, AlpR/cellular response to DNA damage stimulus/ negative regulation of transcription, DNA-templated	Transcriptional regulators	237	0.033	406	Avin_15780	82/406
imm2	Pyocin S2 immunity protein / bacteriocin immunity	Adaptation/protection/defense mechanisms	87	0.61	242	Avin_38330	28/242
wbpB	UDP-2-acetamido-2-deoxy-D-glucuronic acid 3-dehydrogenase/O antigen biosynthetic process	Putative enzyme; cell wall/LPS/ envelope	316	3e-7	369	Avin_50290	127/369
PA4405	Hypothetical protein	Hypothetical/unclassified	131	0.27	461	*cysS*	48/461
PA4674	Antitoxin HigA/toxin-antitoxin pair type II binding	Hypothetical/unclassified	101	3e-17	99	Avin_27690	82/99
PA4685	Hypothetical protein	Hypothetical/unclassified	231	0.38	303	Avin_39900	64/303
*nadE**	NH3-dependent NAD synthetase/ cofactor biosynthetic process/NAD biosynthetic process	Amino acid transport and metabolism	275	5e-12	556	Avin_11920	191/556

*Transposon-insertion mutants have been reported in these genes in the PAO1 strain [14].

are orthologous to the general essential *P. aeruginosa* genes are encoded within regions that have a conserved synteny with this *Pseudomonas* species (Fig. 2b, c show two examples of syntenic regions that encode essenial genes). The complete list of the 337 genes detected in the regions of chromosome synteny and their function described by their COG classification, are shown in Table S1. The five putative essential genes that are not encoded in regions of genome synteny are *lptD lptE, dxs, yidC* and PA0728.

CONCLUDING REMARKS

Several works have reported that essential genes can be highly variable among species and even among strains of the same species. In this context *P. aeruginosa* PAO1 and PA14 strains have been used as an example of this high variability [16, 18]. However, we have shown in this work that the general essential genes of *P. aeruginosa* are extremely well conserved between the PAO1 and PA14 strains, since only 12 PAO1 genes out of the 352 proposed general essential genes for this strain were absent in the PA14 chromosome, and 8 of the 636 PA14 proposed essential genes were absent from the PAO1 genome (Tables 1–3). This high degree of conservation of essential genes is in agreement with the high degree of genomic conservation of *P. aeruginosa* strains [5–7].

The high degree of conservation of essential genes among different *P. aeruginosa* strains is in agreement with their phylogeny [5–7], but it is expected that the essential genes of two species belonging to different bacterial genera that have widely different biology will show very little conservation. This divergent biology of *P. aeruginosa* and *A. vinelandii* can be appreciated considering that the former is an ubiquitous environmental bacterium and also an opportunistic pathogen, while the latter is a nitrogen-fixing soil bacterium that is able to form cysts and presents the highest rate of respiration among all living cells [23]. Unexpectedly considering their diverse phenotypic characteristics, the comparison of the general essential genes of *P. aeruginosa* with *A. vinelandii,* which is described in this work, shows that they are extremely well conserved (Table 4). Furthermore, the general structure of the chromosomes of these bacteria shows only stretches of homologous genes having synteny (Fig. 2) and these small conserved regions encode most of the 342 essential genes that are orthologous to *P. aeruginosa* general essential genes (see Table S1 for a list of the shared 337 *A. vinelandii* genes that

are encoded in the regions with synteny with *P. aeruginosa* PAO1 chromosome). These results reinforce the previously established hypothesis for the origin of *A. vinelandii* as a bacterium with a Pseudomonads genomic backbone inherited vertically, with multiple genes inherited by HGT [23], which have been incorporated interspersed in the chromosome and that encode for fundamental traits for the biology of this soil bacterium.

The existence of bacteria with polyphyletic origins [23, 32, 33] shows that the evolution of bacteria has distinct characteristics that are not shared by other organisms, as their ability to incorporate through HGT genetic information that is essential for their biology.

Funding information
E.M.-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 for the PhD degree and received a fellowship from CONACYT (CVU-483310).

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Moradali MF, Ghods S, Rehm BHA. *Pseudomonas aeruginosa* Lifestyle: A Paradigm for Adaptation, Survival, and Persistence. *Front Cell Infect Microbiol* 2017;7:39.
2. Wiens JR, Vasil AI, Schurr MJ, Vasil ML. Iron-regulated expression of alginate production, mucoid phenotype, and biofilm formation by *Pseudomonas aeruginosa*. *MBio* 2014;5:1–12.
3. Lister PD, Wolter DJ, Hanson ND. Antibacterial-resistant *Pseudomonas aeruginosa*: clinical impact and complex regulation of chromosomally encoded resistance mechanisms. *Clin Microbiol Rev* 2009;22:582–610.
4. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P *et al*. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 2000;406:959–964.
5. Mosquera-Rendón J, Rada-Bravo AM, Cárdenas-Brito S, Corredor M, Restrepo-Pineda E *et al*. Pangenome-wide and molecular evolution analyses of the *Pseudomonas aeruginosa* species. *BMC Genomics* 2016;17:1–14.
6. Klockgether J, Cramer N, Wiehlmann L, Davenport CF, Tümmler B. *Pseudomonas aeruginosa* genomic structure and diversity. *Front Microbiol* 2011;2:1–18.
7. Valot B, Guyeux C, Rolland JY, Mazouzi K, Bertrand X *et al*. What it takes to be a *Pseudomonas aeruginosa*? The core genome of the opportunistic pathogen updated. *PLoS One* 2015;10:126468.
8. Rouli L, Merhej V, Fournier P-E, Raoult D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect* 2015;7:72–85.
9. Luo H, Lin Y, Gao F, Zhang C-T, Zhang R. Deg 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res* 2014;42:D574–D580.
10. Luo H, Gao F, Lin Y. Evolutionary conservation analysis between the essential and nonessential genes in bacterial genomes. *Sci Rep* 2015;5:1–8.
11. Acevedo-Rocha CG, Fang G, Schmidt M, Ussery DW, Danchin A. From essential to persistent genes: a functional approach to constructing synthetic life. *Trends in Genetics* 2013;29:273–279.
12. Juhas M, Reuß DR, Zhu B, Commichau FM. *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiology* 2014;160:2341–2351.
13. Gallagher LA, Shendure J, Manoil C. Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. *MBio* 2011;2:1–8.
14. Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E *et al*. Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 2003;100:14339–14344.
15. Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E *et al*. An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc Natl Acad Sci U S A* 2006;103:2833–2838.
16. Skurnik D, Roux D, Aschard H, Cattoir V, Yoder-Himes D *et al*. A comprehensive analysis of in vitro and in vivo genetic fitness of *Pseudomonas aeruginosa* using high-throughput sequencing of transposon libraries. *PLoS Pathog* 2013;9:e1003582.
17. Rusmini R, Vecchietti D, Macchi R, Vidal-Aroca F, Bertoni G. A shotgun antisense approach to the identification of novel essential genes in *Pseudomonas aeruginosa*. *BMC Microbiol* 2014;14:24–29.
18. Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M. Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. *Proc Natl Acad Sci U S A* 2015;112:4110–4115.
19. Lee SA, Gallagher LA, Thongdee M, Staudinger BJ, Lippman S *et al*. General and condition-specific essential functions of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 2015;112:5189–5194.
20. Charlebois RL, Doolittle WF. Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res* 2004;14:2469–2477.
21. Özen AI, Ussery DW. Defining the *Pseudomonas* genus: where do we draw the line with *Azotobacter*? *Microb Ecol* 2012;63:239–248.
22. Rediers H, Vanderleyden J, De Mot R. *Azotobacter vinelandii*: a *Pseudomonas* in disguise? *Microbiology* 2004;150:1117–1119.
23. González-Casanova A, Aguirre-von-Wobeser E, Espín G, Servín-González L, Kurt N *et al*. Strong seed-bank effects in bacterial evolution. *J Theor Biol* 2014;356:62–70.
24. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. Upset: visualization of intersecting sets. *IEEE Trans Vis Comput Graph* 2014;20:1983–1992.
25. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science* 1997;278:631–637.
26. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z *et al*. Gapped blast and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
27. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394–1403.
28. Whitfield C, Trent MS. Biosynthesis and export of bacterial lipopolysaccharides. *Annu Rev Biochem* 2014;83:99–128.
29. Setubal JC, dos Santos P, Goldman BS, Ertesvåg H, Espin G *et al*. Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *J Bacteriol* 2009;191:4534–4545.
30. Matsui H, Sano Y, Ishihara H, Shinomiya T. Regulation of pyocin genes in *Pseudomonas aeruginosa* by positive (*prtN*) and negative (*prtR*) regulatory genes. *J Bacteriol* 1993;175:1257–1263.
31. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G *et al*. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* 2017;45:W30–W35.
32. Mira A, Pushker R, Legault BA, Moreira D, Rodríguez-Valera F. Evolutionary relationships of *Fusobacterium nucleatum* based on phylogenetic analysis and comparative genomics. *BMC Evol Biol* 2004;4:50–17.
33. Zhaxybayeva O, Swithers KS, Lapierre P, Fournier GP, Bickhart DM *et al*. On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc Natl Acad Sci U S A* 2009;106:5865–5870.

Edited by: D. Grainger and D. Lee