


SCIENTIFIC REPORTS



OPEN

Selection and validation of reference genes for gene expression studies in *Klebsiella pneumoniae* using Reverse Transcription Quantitative real-time PCR

Ana Érika Inácio Gomes¹, Leonardo Prado Stuchi¹, Nathália Maria Gonçalves Siqueira¹, João Batista Henrique¹, Renato Vicentini², Marcelo Lima Ribeiro³ , Michelle Darrieux¹ & Lúcio Fábio Caldas Ferraz¹

For reliable results, Reverse Transcription Quantitative real-time Polymerase Chain Reaction (RT-qPCR) analyses depend on stably expressed reference genes for data normalization purposes. *Klebsiella pneumoniae* is an opportunistic Gram-negative bacterium that has become a serious threat worldwide. Unfortunately, there is no consensus for an ideal reference gene for RT-qPCR data normalization on *K. pneumoniae*. In this study, the expression profile of eleven candidate reference genes was assessed in *K. pneumoniae* cells submitted to various experimental conditions, and the expression stability of these candidate genes was evaluated using statistical algorithms BestKeeper, NormFinder, geNorm, Delta C_t and RefFinder. The statistical analyses ranked *recA*, *rho*, *proC* and *rpoD* as the most suitable reference genes for accurate RT-qPCR data normalization in *K. pneumoniae*. The reliability of the proposed reference genes was validated by normalizing the relative expression of iron-regulated genes in *K. pneumoniae* cells submitted to iron-replete and iron-limited conditions. This work emphasizes that the stable expression of any potential reference candidate gene must be validated in each physiological condition or experimental treatment under study.

Reverse Transcription Quantitative real-time Polymerase Chain Reaction (RT-qPCR) has become the method of choice for gene expression quantification due to its simplicity, reproducibility, high sensitivity and specificity^{1,2}. Relative quantification is the most common method for gene expression analysis, as it analyses the changes in expression of a target gene relative to the expression of a reference gene³.

Although accurate for gene expression analysis, RT-qPCR assays are subjected to several variables that can affect the reliability of the quantification, which include the initial amount of the sample, integrity and quantity of the extracted RNA, primer design and efficiency of cDNA synthesis^{1,4}. The most accepted approach to minimize such variations is to perform a relative normalization, in which the expression level of a target gene is normalized relative to the expression level of an endogenous stably expressed gene, also called an internal control or reference gene⁵⁻⁷. The data normalization aim to correct for variations between cells and tissue types, culture conditions, and experimental treatment, thus allowing gene expression measurements to be compared across the samples^{5,6}. For this reason, choosing an appropriate reference gene is crucial for relative gene expression analysis, since the accuracy of the expression data normalized relies on the stable expression of the reference gene.

A suitable endogenous control should have stable expression within the samples to be compared, regardless of physiological or experimental conditions⁴. It should also be expressed at roughly the same level as the mRNA level of the gene under study. Housekeeping genes are usually chosen for normalization of RT-qPCR data.

¹Laboratório de Biologia Molecular de Microrganismos, Universidade São Francisco, Bragança Paulista, SP, CEP 12916-900, Brazil. ²Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Campinas, SP, CEP 13083-875, Brazil. ³Unidade Integrada de Farmacologia e Gastroenterologia (UNIFAG), Universidade São Francisco, Bragança Paulista, SP, CEP 12916-900, Brazil. Correspondence and requests for materials should be addressed to L.F.C.F. (email: lucio.ferraz@usf.edu.br)

Since these genes are involved in cellular basal metabolism, it is assumed that they are constitutively expressed. However, even housekeeping genes are subjected to expression variations, which imply that the stable expression of any potential reference candidate genes must be validated prior to its utilization on RT-qPCR normalization⁸. According to Vandesompele *et al.*⁷, it is recommended to use at least two reference genes to ensure more accurate and reliable normalization of gene expression analysis. Likewise, it is recommended that any endogenous control be validated in every physiological condition or experimental treatment under study, to ensure its stability in each particular situation⁹.

There is no consensus for an ideal and universal endogenous control for the opportunistic human pathogen *Klebsiella pneumoniae*. This bacterium is an important nosocomial pathogen responsible for wide range infections, including pneumonia, bacteremia, urinary and respiratory tract infections¹⁰. Nowadays, *K. pneumoniae* has become a threat worldwide¹¹, mostly due to the emergence of nosocomial infections caused by multidrug-resistant *K. pneumoniae* isolates and to the spread of invasive infections, such as meningitis, endophthalmitis and liver abscess¹². The virulence determinants of *K. pneumoniae* include the production of capsular polysaccharides, lipopolysaccharides, fimbriae and iron acquisition systems^{12,13}.

In fact, iron is an essential element for most living organisms. In bacteria, iron is needed for growth and cellular metabolism, and it is also considered an important cofactor for regulation of many genes involved in bacterial virulence^{14–16}. However, bacteria face iron scarcity when in contact to mammalian host, where the majority of the iron is intracellular and tightly bound to host iron-binding proteins such as transferrin, lactoferrin and hemoglobin¹⁷.

The predominant strategy employed by pathogens to acquire iron is through the production of siderophores, small iron-chelating molecules that exhibit high affinity for iron¹³. Several siderophores are expressed in *K. pneumoniae* strains, including enterobactin, yersiniabactin, salmochelin, and aerobactin¹². The catecholate siderophore enterobactin has the highest affinity for iron and is considered the most prevalent siderophore-mediated iron acquisition system in *K. pneumoniae*^{12,18,19}. Moreover, enterobactin has an affinity for ferric iron (Fe³⁺) greater than that exhibited by lactoferrin and transferrin and, therefore, can efficiently scavenge iron from the host. Once bound to Fe³⁺, the enterobactin-Fe³⁺ complex is captured and imported into the bacterium through its cognate outer membrane (OM) receptors. Gram-negative bacteria such as *K. pneumoniae* encode the OM receptors FepA, CirA and Fiu for catecholate-type siderophores transport²⁰. FepA receptor display high ligand affinity to ferric enterobactin²¹, while CirA and Fiu exhibit specificity to ferric catecholates and their breakdown products containing ferric iron²². Since the transport of the iron-siderophore complex across the outer membrane is an active process, the siderophore receptors depend on the energy transduction system provided by the inner membrane complex TonB-ExbB-ExbD proteins²³.

Although essential as a nutrient, excess of iron is toxic because of its ability to catalyze Fenton reactions that leads to generation of active species of oxygen¹³. Thus, bacteria have evolved an efficient mechanism for iron acquisition and maintenance of intracellular levels of this element^{14,24}. In *K. pneumoniae*, like many other Gram-negative bacteria, iron homeostasis is regulated by the ferric uptake regulator (Fur)^{25,26}. Fur protein binds ferric iron (Fe²⁺) and other divalent metal cations, and the Fur-Fe²⁺ complex modulates gene expression by binding on consensus DNA sequences, known as Fur boxes, located in the promoter region of the target genes²⁵. Fur also regulates the expression of virulence genes involved in motility, quorum sensing, stress resistance, toxin production, and biofilm formation^{27,28}.

Despite the role of iron in *K. pneumoniae* pathogenicity, studies on the expression of iron-regulated virulence genes in this bacterium can be compromised since systematic investigations to establish reliable reference genes for *K. pneumoniae* have not yet been described. In this respect, we describe in the present study the selection and evaluation of eleven reference candidate genes for RT-qPCR gene expression analysis on *K. pneumoniae*. The expression of the candidate genes was assessed in *K. pneumoniae* under different culture media, non-ideal temperatures and various growth stages and the expression stability of the candidate genes was statistically evaluated with Bestkeeper, NormFinder, geNorm, Delta C_T and RefFinder programs. Finally, the validated reference genes were used to normalize the expression of iron-regulated genes in *K. pneumoniae* submitted to iron-replete and iron-limited conditions.

Results

Selection of candidate reference genes, PCR amplification efficiencies and expression profile.

In this study, eleven genes (*aat*, *ffh*, *glnA*, *gyrA*, *proC*, *recA*, *rho*, *rpoC*, *rpoD*, *rrsH* and *trpS*) were selected and evaluated as potential candidate reference genes for RT-qPCR analyses in *K. pneumoniae*. Functional category, locus number, product name and function of each candidate gene are displayed on Supplementary Table S1.

The selected genes are highly conserved among distinct strains of *K. pneumoniae* from phylogenetic groups KpI, KpII and KpIII. As shown on Supplemental Tables S2 to S12, the nucleotide sequence identity between strains of the same phylogroup ranged from 98% to 100%, whereas sequence alignment between strains of different phylogroups showed sequence identity of 99% (*rrsH* gene, among all strains) to 91% (*aat* gene, KpI versus KpIII strains). The sequence alignments presented a query coverage ranging from 99% to 100% and the alignment of all sequences yielded e-values of 0.0 (data not show).

Table 1 shows the primer sequences specific for each selected candidate gene and the corresponding size of the amplicon for *K. pneumoniae* strain ATCC 10031. The amplification specificity was confirmed by the presence of a single PCR product of expected size on agarose gel electrophoresis (Supplementary Fig. S1). Dissociation-curve analyses revealed only single peaks indicating the absence of primer-dimers formation and nonspecific PCR products (Supplementary Fig. S2). Table 1 also displays that the PCR efficiencies ranged from 83.4% for *proC* to 98.2% for *glnA*, which is within the acceptable range for a reliable real-time PCR quantification. Furthermore, the standard curves revealed acceptable correlation coefficient (R²), thus confirming the reliability of the primer pairs in the RT-qPCR analysis.

| Genes | Product Name | Forward and Reverse primer sequences (5' > 3') | Amplicon size | R ² | E (%) | Mean C _T ± SD |
|-------------|--|---|---------------|----------------|-------|--------------------------|
| <i>proC</i> | Pyrroline-5-carboxylate reductase | GATTGCCGATATCGTCTTCG GAGACCACCAGCGACTCTTT | 99 bp | 0.989 | 83.4 | 24.91 ± 0.62 |
| <i>glnA</i> | Glutamine synthetase | GAAGGCCGTAACAAAGGTCA TACACATGGTGGAAACGGATG | 97 bp | 0.988 | 98.2 | 23.79 ± 1.19 |
| <i>gyrA</i> | DNA gyrase subunit A | GTGACCCGTCGTACGATTTT GATAATCGGGTCGATGTTGG | 99 bp | 0.987 | 96.6 | 24.45 ± 1.10 |
| <i>recA</i> | Recombinase A | TTAAACAGGCCGAATTCCAG CCGCTTTCTCAATCAGCTTC | 99 bp | 0.989 | 92.1 | 22.40 ± 0.61 |
| <i>rpoD</i> | RNA polymerase sigma factor RpoD | TCCGGTGCATATGATTGAGA ATACGCTCAGCCAGCTCTTC | 105 bp | 0.989 | 87.5 | 25.05 ± 0.80 |
| <i>rho</i> | Transcription termination factor Rho | AACTACGACAAGCCGGAAAA ACCGTTACCACGCTCCATAC | 99 bp | 0.998 | 92.4 | 23.92 ± 0.90 |
| <i>rpoC</i> | DNA-directed RNA polymerase subunit beta' | TATTCTGGTTCCACGCAACA GGATAACAACGGAAACGCACTT | 97 bp | 0.969 | 91.2 | 23.28 ± 0.88 |
| <i>rrsH</i> | 16S ribosomal RNA | GACGATCCCTAGTGGTCTG GTGCAATAT'CCCCACTGCT | 95 bp | 0.957 | 94.2 | 10.60 ± 0.55 |
| <i>aat</i> | Leucyl/phenylalanyl-tRNA-protein transferase | CTGGATAACCAGCAGTATCGTTC GTACATTCCACCTACCAGCGTATT | 106 bp | 0.994 | 84.1 | 38.83 ± 1.55 |
| <i>ffh</i> | Signal recognition particle protein | GCTAAGCCGAAATCATCAA ATGTCGTCGAACTGCTTGAG | 104 bp | 0.986 | 97.2 | 23.93 ± 1.39 |
| <i>trpS</i> | Tryptophanyl-tRNA synthetase II | GCCACTGTAAGCGCTACTC GCCGATAACGTCAGCGTATT | 100 bp | 0.992 | 87.1 | 29.14 ± 1.50 |

Table 1. Selected candidate reference genes, their corresponding product name, primer sequences (annealing temperature of 60 °C), amplicon size in base pairs (bp), their respective PCR amplification efficiencies and the mean C_T values (± standard deviation) assessed in *K. pneumoniae* cells submitted to various experimental conditions and at different phases of growth. R²: Correlation coefficient; E: PCR efficiency (%); S.D.: standard deviation.

The differences in transcript levels between the candidate genes are given as C_T values (mean ± standard deviation) on Table 1. Although the average C_T values ranged from 10 to 38, the majority of the candidate genes presented average C_T values between 22 and 24. *rrsH* (16S ribosomal RNA) was the most abundantly expressed gene in all of the samples (10.60 ± 0.55, mean CT ± standard deviation), whereas *aat* was the least abundantly expressed gene (38.83 ± 1.55). Overall, the selected genes presented minimum cycle variation across all samples, as indicated by the low standard deviation of the mean C_T. Most genes showed cycle variation below 1 cycle and none above 2 cycles. *rrsH* and *recA* presented the least variation on cycle number (standard deviation of 0.55 and 0.61, respectively), and *ffh* and *aat* showed the largest variation on the cycle number (standard deviation of 1.39 and 1.55, respectively).

Expression stability of candidate reference genes. The expression stability of the eleven candidate genes were assessed under various experimental conditions and evaluated using the statistical algorithms BestKeeper, geNorm, NormFinder, Delta C_T and RefFinder analyses.

Of the eleven candidate reference genes initially selected, two genes, *aat* and *trpS*, were excluded after preliminary analyses by BestKeeper and geNorm. According to BestKeeper, genes with standard deviation (SD values) greater than 1 are considered inappropriate as reference gene. BestKeeper analyses revealed that most of the selected candidate genes are suitable to be considered a reference gene since they had SD values lower than 1 (Supplementary Table S13). However, *aat* and *trpS* had the highest variable expression across all cultural conditions, with SD values of 1.56 and 1.50 respectively. Besides a great SD value, *trpS* also presented a coefficient of variation of the C_Ts (CV value) of 5.14% (Table S13) and, therefore, was considered the gene with the most expression variation. Furthermore, geNorm analysis indicated that *aat* and *trpS* presented expression stability value *M* of 1.19 and 1.10, respectively, which are above the threshold of ≤ 1.0 that indicates stability of gene expression. For these reasons, *aat* and *trpS* were considered unsuitable for a reference gene and were not included on further analysis.

The remaining nine genes were submitted to the Pearson correlation coefficient (symbolized by *r*) calculation by BestKeeper and to the subsequent statistical analyses with geNorm, NormFinder and RefFinder programs. Table 2 summarizes the statistical calculations generated by BestKeeper, geNorm, NormFinder, Delta C_T and RefFinder programs and shows the ranking order of the genes with the most stable expression across all different conditions tested.

As stated by BestKeeper program, the higher the coefficient of correlation (*r*), the greater the stability of the gene. As shown in Table 2, BestKeeper indicated *recA* (*r* = 0.849), *rho* (*r* = 0.816) and *glnA* (*r* = 0.814) with the largest *r* value and *p*-value of 0.001. Thus, these genes showed the most stable expression and hence they are the most suitable genes for endogenous control. Although *rrsH* gene showed stable expression in all tested conditions – as indicated by the least variation on C_T number (mean 10.60 ± 0.55 standard deviation, Table 1) – this gene presented a high CV value (4.10%) and the lowest correlation coefficient (*r* = 0.42, *p*-value of 0.106), which resulted in the ninth position to *rrsH* gene in the BestKeeper ranking (Table 2).

| Ranking | BestKeeper | | NormFinder | | GeNorm | | Delta C _T analysis | | Reffinder analysis | |
|---------|-------------|---------------|-------------|----------------------------------|-------------|---------|-------------------------------|------------------|--------------------|----------------------------------|
| | Gene | r (p-value) | Gene | Stability Value (standard error) | Gene | M value | Gene | Average of STDEV | Gene | Geometric mean of ranking values |
| 1 | <i>recA</i> | 0,849 (0.001) | <i>recA</i> | 0,296 (0.082) | <i>rpoC</i> | 0,44 | <i>recA</i> | 0,92 | <i>recA</i> | 1.78 |
| 2 | <i>rho</i> | 0,816 (0.001) | <i>rho</i> | 0,328 (0.085) | <i>proC</i> | 0,44 | <i>rho</i> | 0,92 | <i>rpoD</i> | 2.99 |
| 3 | <i>glnA</i> | 0,814 (0.001) | <i>proC</i> | 0,423 (0.096) | <i>rpoD</i> | 0,49 | <i>rpoD</i> | 1 | <i>rho</i> | 3 |
| 4 | <i>gyrA</i> | 0,739 (0.001) | <i>rrsH</i> | 0,515 (0.109) | <i>rrsH</i> | 0,6 | <i>gyrA</i> | 1,01 | <i>rpoC</i> | 3.13 |
| 5 | <i>ffh</i> | 0,712 (0.002) | <i>gyrA</i> | 0,535 (0.112) | <i>recA</i> | 0,67 | <i>proC</i> | 1,02 | <i>proC</i> | 3.31 |
| 6 | <i>rpoD</i> | 0,687 (0.003) | <i>rpoC</i> | 0,614 (0.124) | <i>rho</i> | 0,78 | <i>rpoC</i> | 1,06 | <i>rrsH</i> | 4.28 |
| 7 | <i>proC</i> | 0,66 (0.005) | <i>rpoD</i> | 0,674 (0.134) | <i>gyrA</i> | 0,86 | <i>rrsH</i> | 1,08 | <i>gyrA</i> | 6.19 |
| 8 | <i>rpoC</i> | 0,541 (0.03) | <i>ffh</i> | 0,76 (0.148) | <i>glnA</i> | 0,92 | <i>glnA</i> | 1,12 | <i>glnA</i> | 7.74 |
| 9 | <i>rrsH</i> | 0,42 (0.106) | <i>glnA</i> | 0,776 (0.15) | <i>ffh</i> | 0,99 | <i>ffh</i> | 1,31 | <i>ffh</i> | 9 |

Table 2. Expression stability ranking of the candidate reference genes according to BestKeeper, NormFinder and geNorm original softwares and the Delta C_T and Reffinder analysis.

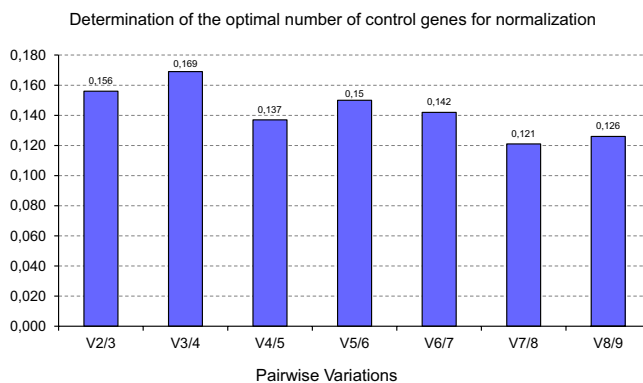


Figure 1. Optimal number of reference genes for normalization of RT-qPCR data indicated by geNorm analysis. GeNorm calculates the pairwise variation (V_n/V_{n+1} , V value) between the normalization factors NF_n/NF_{n+1} to determine the minimum number of reference genes required for accurate normalization. Vandesompele and colleagues⁷ suggest a cut-off threshold of 0.15, below which the inclusion of another reference gene is not required. The $V_{4/5}$ value of 0.137 indicates that the inclusion of the fifth most stable reference gene (*recA*) would have no significant contribution to the accuracy of normalization.

GeNorm analysis of the candidate reference genes rendered an average expression stability M values below the threshold of 1.0, indicating that all tested genes have stable expression across samples (Table 2). The analysis indicated *proC* (0.44), *rpoC* (0.44) and *rpoD* (0.49) with M values below 0.5 and hence they were considered the top most stable genes (Table 2). *rrsH* (0.60), *recA* (0.67) and *rho* (0.78) were estimated to have intermediary M values, while *gyrA* (0.86), *glnA* (0.92) and *ffh* (0.99) exhibited the highest M values among all genes analyzed, but still below 1.0. Accordingly, these genes also had stable expression and are suitable as normalization factor in RT-qPCR analysis.

GeNorm also calculates pairwise variation (V value) to determine the optimal number of reference genes required for accurate normalization of RT-qPCR data, based on the recommended V value threshold of 0.15. As shown in Fig. 1, the V value drops below the recommended 0.15 cut-off only when the fifth most stable reference gene (*recA*) is included ($V_{4/5}$ of 0.137). This result indicates that the top four reference genes (*proC*, *rpoC*, *rpoD* and *rrsH*) would be adequate to ensure accurate normalization of RT-qPCR data, and that the inclusion of *recA* would have no significant contribution to the accuracy of normalization. However, as emphasized by the geNorm authors⁷ the proposed value should not be taken as a too strict threshold but rather is intended to be guidance for determination of the optimal number of reference genes.

According to the NormFinder program, genes with the lowest stability values present the slightest change in expression. As shown in Table 2, NormFinder indicated *recA*, *rho* and *proC* as the top ranked genes with stability values of 0.296, 0.328 and 0.423, respectively. *rrsH* (0.515), *gyrA* (0.535), *rpoC* (0.614) and *rpoD* (0.674) had intermediary stability values, while *ffh* (0.760) and *glnA* (0.776) were ranked the least stable genes with the highest stability values. NormFinder also indicated the best combination of two genes, which were *proC* and *rho* with a stability value of 0.185.

The online available RefFinder tool was also used to compare and rank the expression stability of the candidate reference genes. As displayed on Table 2, the recommended comprehensive ranking by RefFinder indicates *recA*, *rpoD* and *rho* as the most suitable genes for endogenous control, with geometric mean of ranking values of 1.78,

| Genes | Product Name | Putative FUR BOX | | |
|-------------|---|-----------------------|---------------------------------------|--------------------|
| | | Location ^a | Sequence ^b | Score ^c |
| <i>cirA</i> | Colicin I receptor and catecholate siderophore receptor | 242 bp | <u>GATAATGAT</u> TACGATTATC | 21.80 |
| <i>iroN</i> | Outer membrane receptor FepA | 48 bp | TG <u>TAA</u> TGATAAT <u>TGTT</u> ATC | 17.98 |
| <i>fiu</i> | Catecholate siderophore receptor Fiu | 107 bp | <u>GCAA</u> ATGATAACT <u>ATTC</u> TT | 16.17 |

Table 3. Putative Fur boxes identified on the promotor region of genes related to iron-acquisition systems. ^aLocation in base pairs (bp) upstream of the start codon. ^bNucleotides identical to the proposed Fur-binding consensus sequence²⁵ are underlined. ^cScore expressed in bits.

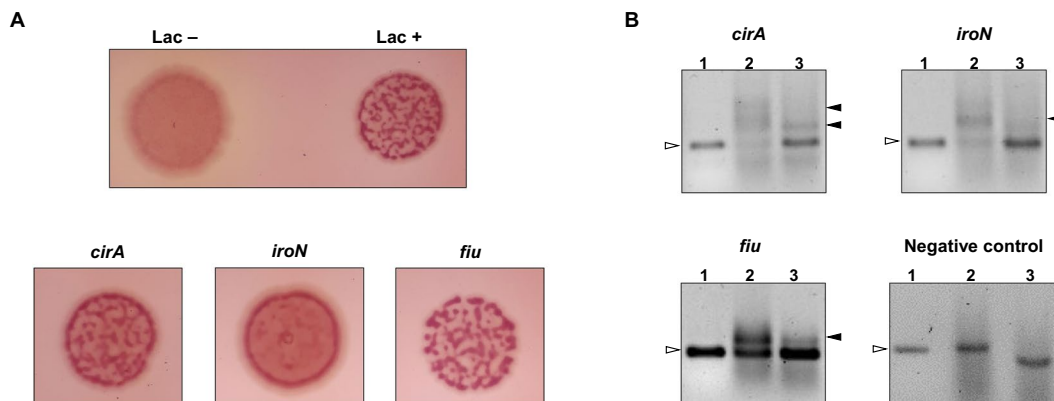


Figure 2. Validation of the putative Fur boxes identified on the upstream region of the genes *cirA*, *iroN* and *fiu* by FURTA (A) and EMSA (B). In (A), Lac⁻ indicates FURTA-negative phenotype, whereas Lac⁺ indicates FURTA-positive phenotype. All putative Fur boxes resulted on red *E. coli* H1717 colonies on MacConkey plates, which were interpreted as FURTA positive results. In (B), EMSA of the DNA fragments containing the putative Fur box of *cirA*, *iroN*, *fiu* and the negative control (DNA fragment without Fur box). Lanes 1, 2 and 3 contained 50 ng of the respective DNA probes. The DNA probes were incubated with 500 nM of His-Fur protein either in the presence of divalent cation (Lanes 2) or under divalent cation-free conditions by adding 2 mM EDTA (Lanes 3). Open arrowheads indicate the free DNA probes, while closed arrowheads indicate the mobility shift corresponding to the Fur/DNA complexes. Full-length gels are presented in Supplementary Fig. S3.

2.99 and 3.0, respectively. Moreover, Delta C_T analysis performed by RefFinder shows *recA*, *rho* and *proC* as the top ranked genes, with average of standard deviation (STDEV) of 0.92, 0.92 and 1.0, respectively.

Supplementary Table S14 compares the BestKeeper, geNorm and NormFinder results calculated by RefFinder with those calculated by their corresponding original softwares. The BestKeeper ranking and stability values were exactly the same, no matter if the analysis were executed by the original BestKeeper software or by RefFinder tool. Concerning geNorm analysis, the only difference observed when the analysis were performed by the original software or by RefFinder tool was a shift between *rpoD* and *proC* on the second and third position on the rank. On the other hand, comparison of the NormFinder results performed by the original software and the RefFinder tool reveals similar results only among the top three ranking genes (Table S14).

Identification of Fur-regulated genes. Bioinformatics analyses were carried out to identify putative Fur-binding sites on the promoter region of genes related to iron-acquisition systems in *K. pneumoniae*. Table 3 shows the putative Fur boxes identified on the upstream region of the genes *cirA*, *iroN* and *fiu*, which encode the catecholate-type siderophore receptors CirA, FepA, and Fiu. To determine whether these new putative Fur boxes are functional, two assays were executed: Fur Titration Assay (FURTA) and DNA Electrophoretic Mobility Shift Assay (EMSA).

On FURTA, *E. coli* strain H1717, which harbors the *lacZ* reporter gene under control of the Fur-regulated *fhuF* gene promoter, was transformed with the putative Fur boxes cloned in a high copy number vector. If the putative Fur boxes were functional, *E. coli* Fur repressor is titrated away from the promoter_{*fhuF*}::*lacZ* gene fusion, thus releasing the expression of *lacZ*. The product of this gene, beta galactosidase, will render red *E. coli* H1717 colonies on MacConkey lactose agar plates under iron-rich condition. If the putative Fur boxes were not functional, *E. coli* Fur repressor remains bound on promoter_{*fhuF*}::*lacZ* gene fusion and, in the absence of beta galactosidase, colorless *E. coli* H1717 colonies will appear on the MacConkey agar plates.

Figure 2A shows that all putative Fur boxes identified on this study rendered red *E. coli* H1717 colonies on MacConkey plates, which confirms that *E. coli* Fur repressor was able to bind the cloned putative Fur boxes in an iron-dependent manner.

To confirm the results observed on FURTA, EMSA was performed to confirm whether *K. pneumoniae* Fur protein directly binds on the putative Fur boxes. As shown on Fig. 2B, the *K. pneumoniae* purified His-Fur

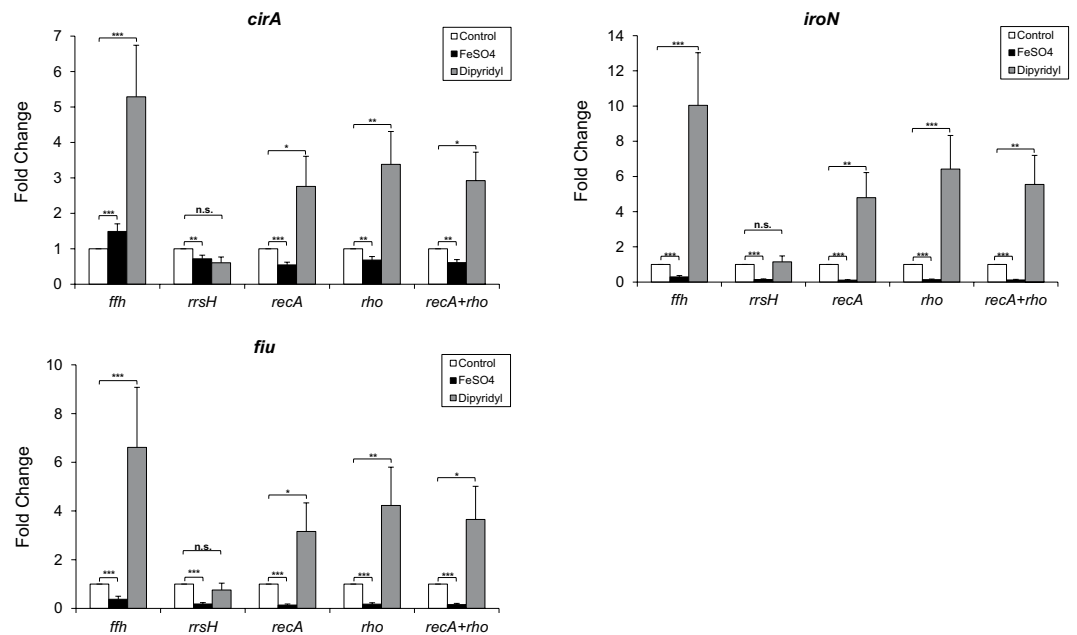


Figure 3. Relative expression of the Fur-regulated genes *cirA*, *iroN* and *fiu* in *K. pneumoniae* cells submitted to iron-replete (FeSO₄) and iron-limited (Dipyridyl) conditions. The expression data were normalized using *ffh*, *rrsH*, *recA* or *rho* as reference genes individually and with the geometric mean of *recA + rho*. Adjusted *p*-value are indicated as following: **p* < 0.05, ***p* ≤ 0.01 and ****p* ≤ 0.001. N.S.: non-significant *p*-value.

protein was able to interact with the DNA fragments containing the putative Fur boxes in the presence of divalent Manganese ions (Mn²⁺), which resulted on a mobility shift of the DNA fragments. Addition of EDTA to the binding reaction mixture abolished the mobility shift of the fragments, indicating that divalent cations are required for Fur interaction with the promoter regions of *cirA*, *iroN* and *fiu* genes. The negative control, which consisted of a 254 base pairs DNA fragment without Fur box sequence, was unable to complex with the *K. pneumoniae* purified His-Fur protein, and a mobility shift was not observed (Fig. 2B).

Expression profile of Fur-regulated genes normalized with the candidate reference genes. To evaluate the reliability of the candidate reference genes, the top two most stable reference genes, *recA* and *rho*, were selected to normalize the relative expression levels of the Fur-regulated genes *cirA*, *iroN* and *fiu* in *K. pneumoniae* submitted to iron-replete and iron-limiting conditions. RT-qPCR data were also normalized with *ffh* gene, which was considered the least stable gene, and with *rrsH* gene, which is commonly used as endogenous control. As shown in Fig. 3, normalization of RT-qPCR data using the two most stable reference genes, *recA* and *rho*, revealed upregulation of *cirA*, *iroN* and *fiu* in iron-limiting condition and downregulation in iron-replete condition, when compared to the control condition. Similar results were obtained when the data were normalized with *recA* and *rho* individually and with the *recA + rho* combined as a normalization factor.

On the other hand, normalization with *ffh* and *rrsH* yielded inconsistent results (Fig. 3). When the least stable *ffh* gene was used, *cirA* appears upregulated at both iron-replete and iron-limiting conditions. Moreover, normalization with *ffh* resulted in almost twofold increase in the upregulation of *cirA*, *iroN* and *fiu* in the presence of the iron chelator, compared to the expression profile of these genes normalized with *recA* and *rho*. When *rrsH* was used as normalizer, no significant difference in the expression pattern of *cirA*, *iroN* and *fiu* was observed at iron-limiting conditions (Fig. 3).

Discussion

Klebsiella pneumoniae is an important nosocomial pathogen that has recently become a global threat due to the emergence and spread of hypervirulent and antibiotic-resistant strains presenting elevated morbidity and mortality¹². An important aspect of the studies of *K. pneumoniae* pathogenicity is the expression analysis of virulence-related genes by Reverse Transcription Quantitative real-time Polymerase Chain Reaction (RT-qPCR). One of the most critical steps on RT-qPCR analysis is the choice of the reference gene with which the data will be normalized. The ideal reference gene should have stable expression regardless of the physiological state and with minor variations during experimental conditions. However, there is no ideal reference gene that meets the mentioned criteria. This implies that any potential reference candidate genes must be validated in every physiological condition or experimental treatment under study prior to its utilization on RT-qPCR normalization⁹.

Unfortunately, there is no validated reference gene for *K. pneumoniae* described in the literature. Several genes have been used as normalizer in gene expression analysis on *K. pneumoniae*, although they did not have their expression stability properly validated. For instance, RT-qPCR analysis in *K. pneumoniae* has been normalized by using either standard curve²⁹ or a diversity of genes, such as *uncB*³⁰, *rfaH*³¹, *rpoB*^{32,33}, *rpoD*^{34,35} and ribosomal

RNA coding genes. In fact, the most extensively used reference gene in RT-qPCR analysis in *K. pneumoniae* are the genes encoding the 16S^{36–39} and 23S^{40–42} ribosomal RNAs. However, there are restrictions on the use of a ribosomal RNA for the normalization of messenger RNAs. Firstly, ribosomal RNA genes are not recommended for RT-qPCR analysis due to the high abundance of transcripts from this gene, which hinder the quantification of rare and less abundant mRNA transcripts⁴³. This high abundance also requires the dilution of the cDNA samples prior to RT-qPCR reactions, thus increasing the risks of dilution errors⁷. Secondly, ribosomal and messenger RNAs present distinct lifetimes within the cell, which difficult comparison among them. While messenger RNAs show rapid turn-over according to the physiological conditions of a bacterium, ribosomal RNA is only degraded under certain stress conditions⁴⁴. Thirdly, some studies have suggested that the expression of ribosomal RNA genes may be under some regulatory control^{45–47}, which goes against the constitutive expression characteristic that is essential for any endogenous control.

The lack of suitable reference genes for gene expression analysis in *K. pneumoniae* prompted us to perform a systematic approach to identify and validate reliable reference genes to be used on RT-qPCR analysis in this bacterium.

In this study, we selected and evaluated eleven candidate reference genes (*aat*, *ffh*, *glnA*, *gyrA*, *proC*, *recA*, *rho*, *rpoC*, *rpoD*, *rrsH* and *trpS*) commonly used in different bacterial species as potential reference genes for *K. pneumoniae*. Despite the heterogeneous population of pathogenic *K. pneumoniae*¹¹, the selected candidate genes are highly conserved among strains from the distinct *K. pneumoniae* phylogroups KpI, KpII and KpIII. The expression profile of the candidate genes was assessed in *K. pneumoniae* cells submitted to various experimental conditions and at different phases of growth. Then, the expression stability of the candidate genes was calculated using the statistical algorithms BestKeeper, NormFinder, geNorm, Delta C_T and RefFinder analyses.

The *rrsH* gene encoding the 16S subunit of ribosomal RNA presented stable expression in all conditions tested. However, this gene showed high coefficient of variation and the lowest correlation coefficient among the genes analyzed by BestKeeper program. Besides, the high abundance of transcripts from this gene, indicated by the lowest C_T number among all genes analyzed, hinders the use of this gene as endogenous control. These results suggest that *rrsH* gene should be discarded as reference gene for RT-qPCR analysis in *K. pneumoniae*. Similar conclusions were reached by other authors for other bacteria. For instance, the transcript levels of the *rrs* gene (encoding 16S rRNA) in the lactic acid bacterium *Oenococcus oeni* were 1000-fold higher than the transcript levels of mRNAs⁴³. The authors concluded that this gene should not be chosen as internal control for this bacterium. Takle *et al.*⁴⁸ showed that the 16S gene was not stably expressed in the tested culture conditions, thus suggesting that 16S is inappropriate as a reference gene in gene expression studies in the plant pathogen *Pectobacterium atrosepticum*. According to Nieto *et al.*⁴⁹, the *rrs* gene showed stable expression under the conditions tested, but it is not recommended as a reference gene in the extremophile *Acidithiobacillus ferrooxidans* because its more abundant expression prejudices the measurements of low abundance transcripts. Badejo *et al.*⁵⁰ discarded the *rrs* gene as an internal control for *Mycobacterium gilvum* since the expression of this gene was highly unstable in the experimental conditions analyzed.

Although several studies have indicated that genes encoding ribosomal RNAs are not reliable as internal control, many studies show that these genes can be considered a good reference gene in some bacteria, such as *Moraxella catarrhalis*⁵¹, *Clostridium difficile*⁵² and *Gluconacetobacter diazotrophicus*⁵³. Therefore, before a ribosomal RNA gene be discarded as endogenous control, it is recommended that the expression stability of these genes be tested under various physiological and experimental conditions, and that its elevated expression level be considered and adjusted to avoid possible bias.

The top two most stably expressed genes determined by BestKeeper, NormFinder and Delta C_T analyses were *recA* and *rho*. RefFinder also indicated *recA* as the top stable gene, but ranked *rho* as the third most stable. Besides *recA* and *rho*, *proC* was ranked among the top three most stable genes by NormFinder and geNorm, and *rpoD* was ranked among the top three reference genes by geNorm, Delta C_T and RefFinder programs.

GeNorm's pairwise variation analysis indicated the use of the top four reference genes (*proC*, *rpoC*, *rpoD* and *rrsH*) as the minimum number of genes required to ensure accurate normalization of RT-qPCR data. However, the other statistical programs revealed that *rpoC* is not among the top most stable genes and that *rrsH* was considered unsuitable as reference gene in *K. pneumoniae*. Therefore, we suggest that they should be excluded and, instead, replaced by *recA* and *rho* in the normalization factor.

The BestKeeper, NormFinder and geNorm analyses carried out by RefFinder yielded results with slight differences when compared to the results obtained by the three softwares individually. Although this result seems to validate the RefFinder tool, caution should be taken when employing and interpreting RefFinder outputs. This web-based platform utilizes raw C_T values and does not take into account the efficiency of primers, which may bias the final ranking of the candidate reference genes.

Taken all statistical analysis into consideration we suggest *recA*, *rho*, *rpoD* and *proC* as the most suitable reference genes for normalization of RT-qPCR data in *K. pneumoniae*. Similar to our findings, these reference genes were also indicated as the most suitable internal control genes for RT-qPCR data normalization in other bacteria, such as in *Actinobacillus pleuropneumoniae*⁵⁴, *Erwinia amylovora*⁵⁵, *Gluconacetobacter diazotrophicus*⁵³, *Pectobacterium atrosepticum*⁴⁸, *Pseudomonas aeruginosa*⁵⁶ and *Staphylococcus aureus*⁵⁷. Supplemental Tables S15 to S26 display the primer sequences of the four proposed reference genes for distinct strains of *K. pneumoniae* from phylogroups KpI, KpII and KpIII.

To test the strength and reliability of the proposed reference genes, we applied *recA* and *rho* as reference genes to normalize the relative expression of genes related to iron uptake systems in *K. pneumoniae* cells submitted to iron-replete and iron-limited conditions. We decided to test the proposed reference genes under such conditions because iron availability is crucial for both survival and virulence of pathogenic bacteria.

To acquire iron, *K. pneumoniae* synthesizes and secretes high-affinity iron-chelating molecules known as siderophores. A number of studies have shown that the production of multiple iron uptake systems enhances the

pathogenicity of *K. pneumoniae* clinical isolates and that hypervirulent strains secrete more active siderophore molecules than non-hypervirulent strains^{19,58–60}. Enterobactin, a catecholate-type siderophore, is considered the primary iron acquisition system utilized by *K. pneumoniae* since is almost ubiquitous among classical and hypervirulent strains¹². However, this siderophore is inhibited by the host molecule lipocalin-2 (LCN2). To evade LCN2, some strains of *K. pneumoniae* produce salmochelin, a glycosylated form of enterobactin that it is not neutralized by LCN2¹². In Gram-negative bacteria, including *K. pneumoniae*, siderophore-bound iron is captured by specific outer membrane receptors and it is actively transported into the cytoplasm to be used by the cell²⁰. Salmochelin is captured by the outer membrane receptor FepA specifically encoded by *iroN* gene, while CirA and Fiu receptors, encoded by *cirA* and *fiu*, uptake ferric catecholates and their degradation products containing ferric iron^{20,22}.

In *K. pneumoniae* the iron homeostasis is controlled by the ferric uptake regulator (Fur protein), which acts as a transcriptional repressor of iron-regulated genes. In its classical mechanism of action, Fur complexes with iron cofactor under iron-rich conditions and this complex binds to consensus DNA sequences, named Fur boxes, located in the promoter region of the target genes. Binding of Fur at the promoters prevents the binding of RNA polymerase and the transcription of the target gene²⁷.

Here we described the identification and validation of Fur boxes on the promoter region of *cirA*, *iroN* and *fiu*, thus suggesting that the expression of these genes in *K. pneumoniae* is regulated by Fur repressor in an iron-dependent manner. To confirm these results, RT-qPCR analysis were employed to access the expression of *cirA*, *iroN* and *fiu* in *K. pneumoniae* cells subjected to iron-rich and low-iron conditions. When normalized with *recA* and *rho* reference genes, *cirA*, *iroN* and *fiu* appear upregulated and downregulated respectively in iron-replete and iron-restricted conditions. On the other hand, inconsistent results were obtained when *ffh* and *rrsH* were used as normalizers. These results validate the use of *recA* and *rho*, either individually or as a combined normalization factor, as reference genes in *K. pneumoniae* and highlight how the use of unsuitable reference genes can compromise RT-qPCR data normalization.

The differentially expression of *cirA*, *iroN* and *fiu* under iron-rich and low-iron conditions is consistent to the fact that these genes are regulated by Fur repressor and that ferrous iron acts as a corepressor. This is the first report describing the Fur-mediated regulation of *cirA*, *iroN* and *fiu* genes in *K. pneumoniae*, although the iron-dependent regulation of these genes has already been described in other enterobacteria. For instance, in *E. coli* the expression of *cirA*^{61,62}, *fepA*^{61,63,64} and *fiu*⁶¹ are higher under iron-depleted conditions than under iron-replete conditions. Upregulation of *cirA* under iron starvation has also been described in *Salmonella typhimurium*⁶⁵.

It is widely accepted that using at least two reference genes is sufficient to ensure high quality data, and that three reference genes would be ideal. According to the MIQE Guidelines¹, normalization against a single reference gene is only acceptable if the gene presents invariant expression under the experimental tested conditions. Here we showed that *recA* and *rho* were considered, by the majority of the statistical programs, the genes with the most stable expression across all different conditions tested. Moreover, the normalization of RT-qPCR data using either *recA* and *rho* individually or as a combined normalization factor resulted in similar and reliable results. Despite this, to guarantee an accurate normalization of RT-qPCR data we suggest the use of at least two endogenous control among the four proposed reference genes.

In summary, the reliability of RT-qPCR analyses depend on accurate data normalization, which can only be achieved by using validated reference genes. Ideally, the stable expression of any potential reference candidate gene must be tested in every experimental treatment and condition under study. In this study, we recommend *recA*, *rho*, *rpoD* and *proC* as reference genes for gene expression normalization in *K. pneumoniae*. It is worth mentioning that the heterogeneous nature of the *K. pneumoniae* population should be taken into account when using the proposed reference genes. It is advisable to assess the expression stability of the proposed reference genes in a given strain under study, prior to their use on RT-qPCR data normalization. To the best of our knowledge, this is the first systematic study aimed to identify reference genes for RT-qPCR analyses in this bacterium.

Methods

Selection of candidate reference genes and design of primer pairs. Potential candidates for endogenous reference genes for *Klebsiella pneumoniae* were selected based on internal controls commonly used in other bacteria, as described on the literature. The selected candidate genes included: *aat* (leucyl/phenylalanyl-tRNA-protein transferase)⁴⁹, *ffh* (signal recognition particle protein)⁴⁸, *glnA* (glutamine synthetase)⁴⁸, *gyrA* (DNA gyrase subunit A)^{43,48,52,57}, *proC* (pyrroline-5-carboxylate reductase)^{43,48,52,57}, *recA* (recombinase A)^{48,49,53}, *rho* (transcription termination factor Rho)^{22,48,52,53,57}, *rpoC* (DNA-directed RNA polymerase subunit beta)^{49,53}, *rpoD* (RNA polymerase sigma factor RpoD)^{43,53,57}, *rrsH* (16S ribosomal RNA)^{43,48,52,57} and *trpS* (tryptophanyl-tRNA synthetase II)⁴⁹. To minimize the chance of coregulation among the selected genes, the selection was performed so that the candidate genes belonged to different functional categories, such as cell metabolism (*proC* and *glnA*), protein synthesis (*rrsH*, *aat*, *ffh* and *trpS*), DNA replication (*gyrA* and *recA*) and transcription (*rpoC*, *rpoD* and *rho*). To check the homology of the selected candidate genes among different strains of *K. pneumoniae*, we performed BLAST multiple alignment of the nucleotide sequence of each gene on strains from the three phylogroups of *K. pneumoniae*: KpI (strains 1084, ATCC 10031, CG43, JM45, HS11286, KCTC 2242, Kp13, NTUH-K2044 and MGH 78578), KpII (strains ATCC 700603 and HKUOPA4) and KpIII (strain 342).

Primer pairs for each gene were designed based on the entire coding region of the candidate genes. Primers were designed using Primer3 v. 0.4.0⁶⁶ according to the following parameters: primer length of about 20 bases, GC content of 45–60%, melting temperature (T_m value) of 60°C, and amplicon size ranging from 95 to 105 base pair. Prior to the Real-time PCR assays, all primer pairs were tested by conventional PCR followed by agarose gel electrophoresis, in order to check for specificity of PCR amplification. The amplicons were visualized under

UV light and recorded with a digital photodocumentation system (Gel DocTM XR, Biorad). The images were captured and analyzed with the Image LabTM Software version 5.0 (Biorad).

Bacterial strain and growth conditions. *K. pneumoniae* strain ATCC 10031 was cultured in LB liquid medium on a rotary incubator shaker (150 rpm) at 37 °C under aerobic conditions. Bacterial growth was monitored by measuring the optical density of the cultures at a wavelength of 600 nm (O.D._{600nm}), using the GeneQuant Spectrophotometer (GE Healthcare).

The expression of the candidate reference genes was analyzed in *K. pneumoniae* cells submitted to various experimental conditions and at different phases of growth. Regarding the phases of growth, bacterial cells were harvested in the lag (O.D._{600nm} = 0.2), exponential (O.D._{600nm} = 0.6) and stationary (O.D._{600nm} = 1.9) stages. For conditions of thermal stress, bacterial cells at the exponential phase were incubated for 30 minutes at 29 °C (cold shock) or 45 °C (heat shock) and then harvested. The expression of the candidate genes was also analyzed under different availability of iron in the culture medium. In this case, cells at the exponential phase were harvested 30 minutes after the addition of a source of ferrous iron (FeSO₄, Sigma-Aldrich) or an iron chelator (2,2'-Dipyridyl, Sigma-Aldrich) in LB medium to a final concentration of 100 μM. In addition, bacterial cells at the exponential phase were harvested 30 minutes after the addition of MnSO₄ in the culture medium (100 μM, final concentration). The expression of the genes were assessed in the presence of MnSO₄ since this compound is commonly used as a source of divalent ions (Mn²⁺) for studies of Fur-mediated iron regulation, in replace of the highly oxidizable ferrous iron. All culture conditions were done in triplicates. Bacterial cells were harvested by centrifugation, the cell pellets were resuspended in *RNAprotect[®] Bacteria Reagent* for RNA stabilization as described by the manufacturer (Qiagen) and immediately submitted to total RNA extraction.

Total RNA extraction and first strand cDNA synthesis. Total RNA was extracted using *RNeasy Mini Kit* (Qiagen), as recommended by the manufacturer's protocol. An on-column DNase digestion with the RNase-free DNase Set (Qiagen) was included to remove any genomic DNA contamination in RNA samples. RNA integrity was analyzed by agarose gel electrophoresis and its purity and concentration were calculated by measuring the optical density of the samples at 260 and 280 nm using a spectrophotometer.

For single strand cDNA synthesis, 1 μg of high quality purified RNA was reverse transcribed in a 20 μL volume reaction using random hexamers and *ThermoScript[™] RT-PCR System*, according to the manufacturer's instructions (Invitrogen). Prior to use on Real-time PCR assays, the synthesized cDNA samples were diluted with nuclease-free water to a final concentration of 100 ng/μL (roughly a 1:20 dilution).

Real-time PCR assays and verification of PCR amplification efficiency. The Real-time PCR assays were carried out in a 7300 Real-Time PCR System (Applied Biosystems) using the *Platinum[®] SYBR[®] Green qPCR SuperMix-UDG* (Thermo Fisher Scientific), following the manufacturer's instructions. Reaction mixture consisted of 1 μL of diluted cDNA, 400 nM of each primer, ROX reference dye at final concentration of 500 nM and 1x SYBR Green qPCR SuperMix adjusted with nuclease free water to a final volume of 12,5 μL. Due to high abundance of 16S ribosomal RNA subunit, the expression of *rrsH* gene was measured using cDNA diluted 200 fold. The RT-qPCR reactions were done in triplicate for each cDNA sample.

The reactions were initially incubated at 50 °C for 2 minutes for Uracil-DNA glycosylase treatment, followed by denaturation for 2 minutes at 95 °C. After this pretreatment, reactions were subjected to the following thermal cycling conditions: 40 cycles of denaturation at 95 °C for 15 seconds and annealing/extension at 60 °C for 60 seconds. Finally, dissociation (melting) curve analyses were performed to check for nonspecific amplification and/or primer-dimers formation.

The Real-time PCR data were detected and analyzed by the software *7300 Real-Time PCR System Sequence Detection Software v1.4.1* (Applied Biosystems) according to default parameters, which generated the Cycle Threshold (C_T) values for each reaction. The average C_T number of each triplicate Real-time PCR reactions was used on the subsequent statistical analyses.

Standard curves were constructed for every candidate reference gene to determine the PCR amplification efficiency and the regression coefficient (R²) of each pair of primers. Ten-fold serial dilution of genomic DNA from *K. pneumoniae* was used in Real-time PCR reactions and the five-point standard curves were generated by plotting the average C_T numbers versus the logarithm of the amount of template DNA. PCR amplification efficiency (E) of each primer pair was calculated from the linear regression and the slope of the corresponding standard curves, according to the formula: E (%) = [10^(-1/slope) - 1] × 100. The efficiency of each reference gene was considered in all subsequent statistical analysis.

Reference gene expression stability analyses. The expression stability of the selected reference genes was statistically evaluated by three commonly used Microsoft Excel-based softwares: BestKeeper⁶⁷, geNorm⁷ and NormFinder⁶⁸. BestKeeper analyses allow the input of raw C_T values, while geNorm and NormFinder require that the raw C_T values be converted to relative quantification data. To achieve this, the raw C_T values were transformed into relative quantities using the formula 2^(-ΔC_T), in which ΔC_T corresponds to the highest C_T value minus all other C_T value for each reference gene measured across all samples.

BestKeeper relies on the coefficient of variation (CV [% C_T]) and the standard deviation (SD) of the C_T values (SD [±C_T]) to estimate the expression stability of the candidate genes. According to BestKeeper program, the genes can be ranked from the most stably expressed, exhibiting the lowest SD [±C_T], to the least stably expressed, exhibiting the highest SD [±C_T]. Genes with SD greater than 1 are considered inconsistent by BestKeeper. Furthermore, BestKeeper performs numerous pairwise correlation analyses between each gene analyzed and the geometric mean C_T value of all the candidate genes together. Within each such pairwise correlation the program

calculates the Pearson correlation coefficient, symbolized by r , and the probability p -values⁶⁷. According to BestKeeper, the higher the correlation coefficient (r), the greater the stability of gene expression.

GeNorm ranks the candidate reference genes based on the expression stability value M , which is defined as the average pairwise variation between a particular reference gene and all other reference genes tested. Genes with the lowest M values are considered to have the most stable expression under tested experimental conditions and genes with M values above the threshold of 1.5 are not acceptable as reference genes⁷. In an even more stringent analysis, several studies have established the threshold of M value ≤ 1.0 to identify the most suitable reference genes for RT-qPCR normalization^{48,54,69,70}. Stepwise exclusion of the least stable gene (with the highest M value) allows ranking of the tested genes according to their expression stability in the tested samples.

GeNorm also performs pairwise variation analysis to determine the optimal number of reference genes required for accurate normalization. This is achieved by calculating the pairwise variation value (V_n/V_{n+1} value) between sequential normalization factors containing increasing number reference genes (NF_n/NF_{n+1}). Vandesompele and colleagues⁷ recommend to add additional reference genes to the normalization factor until the V_n/V_{n+1} (V) value reaches the cut-off threshold of 0.15. V value below 0.15 means that the inclusion of an additional reference gene is not necessary since will not improve the normalization accuracy.

NormFinder considers not only the overall expression variation of the candidate reference genes but also the intra- and inter-group expression variations to calculate the stability value for each candidate gene⁶⁸. Therefore, NormFinder ranks the candidates with minimal estimated variation between sample subgroups of the sample set. According to the analysis, the top ranked genes are those with the smallest stability value and the best combination of two reference genes is also indicated.

The expression stability of the selected reference genes was also evaluated by RefFinder⁷¹, available on the website <http://leonxie.esy.es/RefFinder/>. RefFinder is a web-based tool that provides a recommended comprehensive ranking based on the geometric mean of ranking values. RefFinder also integrates BestKeeper, geNorm and NormFinder softwares, and the comparative Delta C_T method⁷² to compare the expression stability of the candidate reference genes. The raw C_T values are directly input into the program to calculate and rank the tested candidate reference genes.

Identification of putative Fur-regulated genes. Since iron is a crucial cofactor for regulation of virulence genes expression in many pathogenic bacteria, we decided to test the strength and reliability of the proposed reference genes on relative expression normalization of iron-regulated genes in *K. pneumoniae* cells submitted to iron-replete and iron-limited conditions. To find iron-regulated genes, bioinformatics analysis were employed on the genomic sequence of *K. pneumoniae* strain ATCC 700721/MGH 78578 (GenBank accession number CP000647.1) to identify putative Fur-binding sites on the promoter region of genes related to iron-acquisition systems in this bacterium. These analyses were conducted according to a theoretical approach described elsewhere⁷³ and adapted to *K. pneumoniae*. In brief, a set of experimentally confirmed Fur boxes from *K. pneumoniae*^{40,42,74} was used to create a position weight matrix (PWM) model. This matrix was employed to search the *K. pneumoniae* genomic region where genes related to iron-acquisition systems are located. A 19bp sliding window was used in this search. Only windows with scores, in the Fur weight matrix, higher than 7 bits were retained in the analysis. Complementary oligonucleotides containing the sequences of the putative Fur boxes were annealed to form double-stranded DNA probes. These DNA probes were used in Fur titration assay (FURTA) and DNA electrophoretic mobility shift assay (EMSA), in order to validate the Fur interactions with the putative Fur boxes.

Fur Titration Assay (FURTA). FURTA was performed with *Escherichia coli* strain H1717 (kindly provided by Prof. Klaus Hantke, University of Tübingen, Germany), as described elsewhere⁷⁵. *E. coli* strain H1717 carries the *lacZ* reporter gene under control of the Fur-regulated *shuF* gene promoter. The promoter *shuF::lacZ* gene fusion is exceptionally sensitive to changes of iron and Fur repressor concentrations. In the presence of iron, *E. coli* Fur repressor binds to the *shuF* promoter region and the *lacZ* reporter gene is not expressed, rendering colorless *E. coli* H1717 colonies on MacConkey lactose agar plates (Lac⁻ phenotype). However, *E. coli* H1717 transformed with multicopy plasmids cloned with Fur binding sites will appear red on the plates (Lac⁺ phenotype), because the high number of newly introduced Fur boxes will cause the dissociation of the repressor Fur from the *shuF* promoter, thus releasing the transcription of the *lacZ* gene.

The double-stranded DNA probes containing the sequences of the putative Fur boxes were cloned into high copy number pGEM[®]-T Easy vector (Promega). The resulting vectors were introduced into the *E. coli* strain H1717 and the transformants were plated onto MacConkey lactose agar containing 100 μ g/mL ampicillin and under iron-rich condition. Plates were incubated for 18 h at 37 °C and the Lac phenotype was recorded. Circular pGEM[®]-T Easy vector without insert (i.e., not cloned with DNA probes) was used as a negative control, whereas vector cloned with the previously validated Fur box of the *K. pneumoniae entC* gene⁴⁰ was used as a positive control.

Expression and purification of *K. pneumoniae* Fur protein. The entire coding region of *fur* gene was PCR amplified from genomic DNA of *K. pneumoniae* with forward primer 5'-GTTCGATATGATGACTGACAACAATACC-3', containing restriction site for *NdeI* (underlined nucleotides) and reverse primer 5'-TATCTCGAGTTATTTTCCACCGCG-3', containing restriction site for *XhoI* (underlined nucleotides). The PCR product was cloned into the expression vector pET-28a (Novagen) at the *NdeI* and *XhoI* sites, and the resulting plasmid was introduced into *E. coli* BL21(DE3).

To overexpress the *K. pneumoniae* histidine-tagged Fur protein (His-Fur), *E. coli* BL21(DE3) cells transformed with the recombinant plasmid were grown in LB medium to an O.D._{600 nm} of 0.4. At this point, isopropyl- β -D-thiogalactopyranoside (IPTG, Sigma-Aldrich) was added to a final concentration of 1 mM and the culture was incubated for 4 h. After IPTG induction, in-culture bacterial cell lysis was promoted by adding

CelLytic™ Express 1 mL Tablets (Sigma-Aldrich) to the culture, followed by incubation for 20 minutes at 37 °C and orbital shaker at 180 rpm. The His-Fur protein was then purified by affinity chromatography under native conditions by adding *HIS-Select Nickel Affinity Gel* (Sigma-Aldrich) to the lysed cell solution, according to the manufacturer's protocol.

Eluted fractions containing the His-Fur were pooled, dialyzed overnight at 4 °C on storage buffer (20 mM Tris-HCl pH 7.8, 1 mM DTT, 0.1 mM MnSO₄ and 10% glycerol v/v) and concentrated using *Pierce® Concentrator* (Thermo Fisher Scientific) with molecular-mass cut off of 10 kDa, following the instructions of the manufacturer. The concentration of the purified His-Fur was determined by the Bradford method and the purity was verified by SDS-PAGE analysis.

DNA Electrophoretic Mobility Shift Assay (EMSA). The purified *K. pneumoniae* Fur protein and DNA fragments containing the putative Fur boxes were used on DNA Electrophoretic Mobility Shift Assays (EMSA). These DNA fragments were obtained by PCR amplifying the pGEM®-T Easy vectors cloned with the putative Fur boxes. Amplifications were done with universal M13 primers and the resulting 285 base pairs long fragments were then used as probes on EMSA. The negative control consisted of a 254 base pairs DNA fragment without Fur box sequence. This fragment was obtained by PCR amplifying a pGEM®-T Easy vector without insert (i.e., vector not cloned with the putative Fur boxes).

EMSA was performed as described elsewhere⁷³, with minor modifications. In brief, 500 nM of His-Fur protein was initially equilibrated for 10 minutes on ice in a 10 µL reaction volume containing 1x binding buffer (10 mM Tris, 50 mM KCl, 1 mM DTT, pH 7.5), 0.5 mM MgCl₂, 0.5 mM MnSO₄ and 2.5% (v/v) glycerol. To this binding reaction, 50 ng of the DNA probes were added and the mixture was incubated for 20 minutes on ice. In addition, EMSA were carried out under divalent cation-free conditions by adding EDTA to a final concentration of 2 mM in the above reaction mixture.

Samples were loaded onto a 2% (w/v) agarose gel and submitted to electrophoresis for 30 minutes at 100 Volts in 1x Bis-Tris borate (pH 7.5) buffer containing 0.1 mM MnSO₄. The agarose gels were stained after electrophoresis by soaking then on an ethidium bromide solution (0.5 µg/mL) for 15 minutes with gentle agitation. The DNA bands were visualized under UV light and recorded with a digital photodocumentation system (Gel Doc™ XR, Biorad). Capture and analysis of the images were done with the Image Lab™ Software version 5.0 (Biorad).

Evaluation of usefulness of the candidate reference genes. To demonstrate the usefulness of the reference genes validated in this study, we selected the top two most stable reference genes, *recA* and *rho*, to normalize the relative expression levels of the Fur-regulated genes *cirA*, *iroN* and *fiu* in *K. pneumoniae* submitted to iron-replete and iron-limiting conditions.

K. pneumoniae cells were grown in LB medium until the exponential phase (O.D._{600nm} = 0.6). At this point, ferrous iron (100 µM of FeSO₄, final concentration) or an iron chelator (100 µM of 2,2'-Dipyridyl, final concentration) were added and after 30 minutes of incubation the cells were harvested by centrifugation. The control condition consisted of *K. pneumoniae* cells grown in LB medium and harvested at the exponential phase (O.D._{600nm} = 0.6). All culture conditions were performed at least twice. The bacterial cell pellets were resuspended in *RNAprotect® Bacteria Reagent* for RNA stabilization and immediately submitted to total RNA extraction. The total RNA extraction and the first strand cDNA synthesis were conducted as above described.

The RT-qPCR reactions were carried out in triplicates in a 7300 Real-Time PCR System using the *Platinum® SYBR® Green qPCR SuperMix-UDG* in reaction mixture as previously described. Primers for *cirA*, *iroN* and *fiu* genes were designed with Primer3 v. 0.4.0 using the entire coding region of the selected genes. The primers were designed in order to have about 20 bases of length, melting temperature of 60 °C, and amplicon size ranging from 95 to 105 base pair (see Supplementary Table S27). Before the Real-time PCR assays, primers for *cirA*, *iroN* and *fiu* genes were tested by conventional PCR followed by agarose gel electrophoresis (Supplementary Fig. S4). RT-qPCR data were normalized with the top two most stable reference genes, *recA* and *rho*, individually and with a normalization factor calculated as the geometric mean of the expression levels of the two genes. To emphasize the need of using suitable endogenous genes, *ffh* gene, which was considered the least stable gene, was also used in the normalization of the expression levels of the Fur-regulated genes. In addition, RT-qPCR data were normalized with *rrsH* gene, which encodes 16S ribosomal RNA subunit, since this gene is commonly used as endogenous control in RT-qPCR experiments. The relative expression levels of the target genes were calculated according to the comparative critical threshold ($\Delta\Delta C_T$) method⁷⁶. GraphPad Prism 7.00 (GraphPad Software, Inc.) was used for the statistical analyses. The differences on the expression levels were evaluated by Student's *t* test with correction for multiple tests (Holm-Sidak method). The differences were considered statistically significant with *p*-values ≤ 0.05.

Data Availability. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

References

- Bustin, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of molecular endocrinology* **29**, 23–39 (2002).
- Nolan, T., Hands, R. E. & Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nature protocols* **1**, 1559–1582, <https://doi.org/10.1038/nprot.2006.236> (2006).
- PFALL, M. W. In *Real-time PCR Vol. I* (ed Dorak, M. T.) 63–82 (Taylor and Francis Group, 2006).
- Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry* **55**, 611–622, <https://doi.org/10.1373/clinchem.2008.112797> (2009).
- Derveaux, S., Vandesompele, J. & Hellems, J. How to do successful gene expression analysis using real-time PCR. *Methods* **50**, 227–230, <https://doi.org/10.1016/j.ymeth.2009.11.001> (2010).

6. Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes and immunity* **6**, 279–284, <https://doi.org/10.1038/sj.gene.6364190> (2005).
7. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology* **3**, RESEARCH0034 (2002).
8. Schmittgen, T. D. & Zakrajsek, B. A. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *Journal of biochemical and biophysical methods* **46**, 69–81 (2000).
9. Guenin, S. *et al.* Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *Journal of experimental botany* **60**, 487–493, <https://doi.org/10.1093/jxb/ern305> (2009).
10. Podschun, R., Sievers, D., Fischer, A. & Ullmann, U. Serotypes, hemagglutinins, siderophore synthesis, and serum resistance of Klebsiella isolates causing human urinary tract infections. *J Infect Dis* **168**, 1415–1421 (1993).
11. Holt, K. E. *et al.* Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. *Proc Natl Acad Sci USA* **112**, E3574–3581, <https://doi.org/10.1073/pnas.1501049112> (2015).
12. Paczosa, M. K. & Meccas, J. Klebsiella pneumoniae: Going on the Offense with a Strong Defense. *Microbiology and molecular biology reviews: MMBR* **80**, 629–661, <https://doi.org/10.1128/MMBR.00078-15> (2016).
13. Li, B., Zhao, Y., Liu, C., Chen, Z. & Zhou, D. Molecular pathogenesis of Klebsiella pneumoniae. *Future microbiology* **9**, 1071–1081, <https://doi.org/10.2217/fmb.14.48> (2014).
14. Andrews, S. C., Robinson, A. K. & Rodriguez-Quinones, F. Bacterial iron homeostasis. *FEMS microbiology reviews* **27**, 215–237 (2003).
15. Litwin, C. M. & Calderwood, S. B. Role of iron in regulation of virulence genes. *Clinical microbiology reviews* **6**, 137–149 (1993).
16. Schaible, U. E. & Kaufmann, S. H. Iron and microbial infection. *Nat Rev Microbiol* **2**, 946–953, <https://doi.org/10.1038/nrmicro1046> (2004).
17. Parrow, N. L., Fleming, R. E. & Minnick, M. F. Sequestration and scavenging of iron in infection. *Infect Immun* **81**, 3503–3514, <https://doi.org/10.1128/IAI.00602-13> (2013).
18. Bachman, M. A. *et al.* Klebsiella pneumoniae yersiniabactin promotes respiratory tract infection through evasion of lipocalin 2. *Infect Immun* **79**, 3309–3316, <https://doi.org/10.1128/IAI.05114-11> (2011).
19. Koczura, R. & Kaznowski, A. Occurrence of the Yersinia high-pathogenicity island and iron uptake systems in clinical isolates of Klebsiella pneumoniae. *Microbial pathogenesis* **35**, 197–202 (2003).
20. Miethke, M. & Marahiel, M. A. Siderophore-based iron acquisition and pathogen control. *Microbiology and molecular biology reviews: MMBR* **71**, 413–451, <https://doi.org/10.1128/MMBR.00012-07> (2007).
21. Jalal, M. A. & van der Helm, D. Purification and crystallization of ferric enterobactin receptor protein, FepA, from the outer membranes of Escherichia coli UT5600/pBB2. *FEMS Lett* **243**, 366–370 (1989).
22. Nikaido, H. & Rosenberg, E. Y. Cir and Fiu proteins in the outer membrane of Escherichia coli catalyze transport of monomeric catechols: study with beta-lactam antibiotics containing catechol and analogous groups. *J Bacteriol* **172**, 1361–1367 (1990).
23. Braun, V. Energy-coupled transport and signal transduction through the gram-negative outer membrane via TonB-ExbB-ExbD-dependent receptor proteins. *FEMS microbiology reviews* **16**, 295–307 (1995).
24. Wandersman, C. & Delepelaire, P. Bacterial iron sources: from siderophores to hemophores. *Annual review of microbiology* **58**, 611–647, <https://doi.org/10.1146/annurev.micro.58.030603.123811> (2004).
25. Escolar, L., Perez-Martin, J. & de Lorenzo, V. Opening the iron box: transcriptional metalloregulation by the Fur protein. *J Bacteriol* **181**, 6223–6229 (1999).
26. Fillat, M. F. The FUR (ferric uptake regulator) superfamily: diversity and versatility of key transcriptional regulators. *Archives of biochemistry and biophysics* **546**, 41–52, <https://doi.org/10.1016/j.abb.2014.01.029> (2014).
27. Carpenter, B. M., Whitmire, J. M. & Merrell, D. S. This is not your mother's repressor: the complex role of fur in pathogenesis. *Infect Immun* **77**, 2590–2601, <https://doi.org/10.1128/IAI.00116-09> (2009).
28. Troxell, B. & Hassan, H. M. Transcriptional regulation by Ferric Uptake Regulator (Fur) in pathogenic bacteria. *Frontiers in cellular and infection microbiology* **3**, 59, <https://doi.org/10.3389/fcimb.2013.00059> (2013).
29. Zhu, H., Liu, H. J., Ning, S. J. & Gao, Y. L. The response of type 2 quorum sensing in Klebsiella pneumoniae to a fluctuating culture environment. *DNA and cell biology* **31**, 455–459, <https://doi.org/10.1089/dna.2011.1375> (2012).
30. Li, D. W. *et al.* Properties and expression of a multidrug efflux pump AcrAB-KocC from Klebsiella pneumoniae. *Biological & pharmaceutical bulletin* **31**, 577–582 (2008).
31. Chiang, M. K., Lu, M. C., Liu, L. C., Lin, C. T. & Lai, Y. C. Impact of Hfq on global gene expression and virulence in Klebsiella pneumoniae. *PLoS One* **6**, e22248, <https://doi.org/10.1371/journal.pone.0022248> (2011).
32. Choi, M. J. & Ko, K. S. Loss of hypermucoviscosity and increased fitness cost in colistin-resistant Klebsiella pneumoniae sequence type 23 strains. *Antimicrobial agents and chemotherapy* **59**, 6763–6773, <https://doi.org/10.1128/AAC.00952-15> (2015).
33. Srinivasan, V. B. *et al.* Functional characterization of a novel outer membrane porin KpnO, regulated by PhoBR two-component system in Klebsiella pneumoniae NTUH-K2044. *PLoS One* **7**, e41505, <https://doi.org/10.1371/journal.pone.0041505> (2012).
34. Johnson, J. G. & Clegg, S. Role of MrkJ, a phosphodiesterase, in type 3 fimbrial expression and biofilm formation in Klebsiella pneumoniae. *J Bacteriol* **192**, 3944–3950, <https://doi.org/10.1128/JB.00304-10> (2010).
35. Yang, J. *et al.* Transcriptional activation of the mrkA promoter of the Klebsiella pneumoniae type 3 fimbrial operon by the c-di-GMP-dependent MrkH protein. *PLoS One* **8**, e79038, <https://doi.org/10.1371/journal.pone.0079038> (2013).
36. Balestrino, D., Haagensen, J. A., Rich, C. & Forestier, C. Characterization of type 2 quorum sensing in Klebsiella pneumoniae and relationship with biofilm formation. *J Bacteriol* **187**, 2870–2880, <https://doi.org/10.1128/JB.187.8.2870-2880.2005> (2005).
37. De Araujo, C., Balestrino, D., Roth, L., Charbonnel, N. & Forestier, C. Quorum sensing affects biofilm formation through lipopolysaccharide synthesis in Klebsiella pneumoniae. *Research in microbiology* **161**, 595–603, <https://doi.org/10.1016/j.resmic.2010.05.014> (2010).
38. Ruzin, A., Immermann, F. W. & Bradford, P. A. Real-time PCR and statistical analyses of acrAB and ramA expression in clinical isolates of Klebsiella pneumoniae. *Antimicrobial agents and chemotherapy* **52**, 3430–3432, <https://doi.org/10.1128/AAC.00591-08> (2008).
39. Ruzin, A., Visalli, M. A., Keeney, D. & Bradford, P. A. Influence of transcriptional activator RamA on expression of multidrug efflux pump AcrAB and tigecycline susceptibility in Klebsiella pneumoniae. *Antimicrobial agents and chemotherapy* **49**, 1017–1022, <https://doi.org/10.1128/AAC.49.3.1017-1022.2005> (2005).
40. Lin, C. T. *et al.* Fur regulation of the capsular polysaccharide biosynthesis and iron-acquisition systems in Klebsiella pneumoniae CG43. *Microbiology* **157**, 419–429, <https://doi.org/10.1099/mic.0.044065-0> (2011).
41. Wang, Z. C., Huang, C. J., Huang, Y. J., Wu, C. C. & Peng, H. L. FimK regulation on the expression of type 1 fimbriae in Klebsiella pneumoniae CG43S3. *Microbiology* **159**, 1402–1415, <https://doi.org/10.1099/mic.0.067793-0> (2013).
42. Wu, C. C. *et al.* Fur-dependent MrkHI regulation of type 3 fimbriae in Klebsiella pneumoniae CG43. *Microbiology* **158**, 1045–1056, <https://doi.org/10.1099/mic.0.053801-0> (2012).
43. Desroche, N., Beltramo, C. & Guzzo, J. Determination of an internal control to apply reverse transcription quantitative PCR to study stress response in the lactic acid bacterium Oenococcus oeni. *Journal of microbiological methods* **60**, 325–333, <https://doi.org/10.1016/j.mimet.2004.10.010> (2005).

44. Deutscher, M. P. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic acids research* **34**, 659–666, <https://doi.org/10.1093/nar/gkj472> (2006).
45. Gralla, J. D. Escherichia coli ribosomal RNA transcription: regulatory roles for ppGpp, NTPs, architectural proteins and a polymerase-binding protein. *Mol Microbiol* **55**, 973–977, <https://doi.org/10.1111/j.1365-2958.2004.04455.x> (2005).
46. Vandecasteele, S. J., Peetermans, W. E., Merckx, R. & Van Eldere, J. Quantification of expression of Staphylococcus epidermidis housekeeping genes with Taqman quantitative PCR during *in vitro* growth and under different conditions. *J Bacteriol* **183**, 7094–7101, <https://doi.org/10.1128/JB.183.24.7094-7101.2001> (2001).
47. Wagner, R. Regulation of ribosomal RNA synthesis in E. coli: effects of the global regulator guanosine tetraphosphate (ppGpp). *Journal of molecular microbiology and biotechnology* **4**, 331–340 (2002).
48. Takle, G. W., Toth, I. K. & Brurberg, M. B. Evaluation of reference genes for real-time RT-PCR expression studies in the plant pathogen Pectobacterium atrosepticum. *BMC plant biology* **7**, 50, <https://doi.org/10.1186/1471-2229-7-50> (2007).
49. Nieto, P. A., Covarrubias, P. C., Jedlicki, E., Holmes, D. S. & Quatrini, R. Selection and evaluation of reference genes for improved interrogation of microbial transcriptomes: case study with the extremophile Acidithiobacillus ferrooxidans. *BMC molecular biology* **10**, 63, <https://doi.org/10.1186/1471-2199-10-63> (2009).
50. Badejo, A. C., Badejo, A. O., Shin, K. H. & Chai, Y. G. A gene expression study of the activities of aromatic ring-cleavage dioxygenases in Mycobacterium gilvum PYR-GCK to changes in salinity and pH during pyrene degradation. *PLoS One* **8**, e58066, <https://doi.org/10.1371/journal.pone.0058066> (2013).
51. Hays, J. P. The evaluation of putative endogenous control housekeeping genes for real-time polymerase chain reaction expression studies in Moraxella catarrhalis. *Diagnostic microbiology and infectious disease* **65**, 323–326, <https://doi.org/10.1016/j.diagmicrobio.2009.06.021> (2009).
52. Metcalf, D., Sharif, S. & Weese, J. S. Evaluation of candidate reference genes in Clostridium difficile for gene expression normalization. *Anaerobe* **16**, 439–443, <https://doi.org/10.1016/j.anaerobe.2010.06.007> (2010).
53. Galisa, P. S. *et al.* Identification and validation of reference genes to study the gene expression in Gluconacetobacter diazotrophicus grown in different carbon sources using RT-qPCR. *Journal of microbiological methods* **91**, 1–7, <https://doi.org/10.1016/j.mimet.2012.07.005> (2012).
54. Nielsen, K. K. & Boye, M. Real-time quantitative reverse transcription-PCR analysis of expression stability of Actinobacillus pleuropneumoniae housekeeping genes during *in vitro* growth under iron-depleted conditions. *Applied and environmental microbiology* **71**, 2949–2954, <https://doi.org/10.1128/AEM.71.6.2949-2954.2005> (2005).
55. Kaluzna, M., Kuras, A. & Pulawska, J. Validation of reference genes for the normalization of the RT-qPCR gene expression of virulence genes of Erwinia amylovora in apple shoots. *Scientific reports* **7**, 2034, <https://doi.org/10.1038/s41598-017-02078-4> (2017).
56. Savli, H. *et al.* Expression stability of six housekeeping genes: A proposal for resistance gene quantification studies of Pseudomonas aeruginosa by real-time quantitative RT-PCR. *Journal of medical microbiology* **52**, 403–408, <https://doi.org/10.1099/jmm.0.05132-0> (2003).
57. Theis, T., Skurray, R. A. & Brown, M. H. Identification of suitable internal controls to study expression of a Staphylococcus aureus multidrug resistance system by quantitative real-time PCR. *Journal of microbiological methods* **70**, 355–362, <https://doi.org/10.1016/j.mimet.2007.05.011> (2007).
58. Russo, T. A., Olson, R., MacDonald, U., Beanan, J. & Davidson, B. A. Aerobactin, but not yersiniabactin, salmochelin, or enterobactin, enables the growth/survival of hypervirulent (hypermucoviscous) Klebsiella pneumoniae *ex vivo* and *in vivo*. *Infect Immun* **83**, 3325–3333, <https://doi.org/10.1128/IAI.00430-15> (2015).
59. Russo, T. A. *et al.* Aerobactin mediates virulence and accounts for increased siderophore production under iron-limiting conditions by hypervirulent (hypermucoviscous) Klebsiella pneumoniae. *Infect Immun* **82**, 2356–2367, <https://doi.org/10.1128/IAI.01667-13> (2014).
60. Russo, T. A. *et al.* Hypervirulent K. pneumoniae secretes more and more active iron-acquisition molecules than “classical” K. pneumoniae thereby enhancing its virulence. *PLoS One* **6**, e26734, <https://doi.org/10.1371/journal.pone.0026734> (2011).
61. McHugh, J. P. *et al.* Global iron-dependent gene regulation in Escherichia coli. A new mechanism for iron homeostasis. *J Biol Chem* **278**, 29478–29486, <https://doi.org/10.1074/jbc.M303381200> (2003).
62. White-Ziegler, C. A., Malhowski, A. J. & Young, S. Human body temperature (37degrees C) increases the expression of iron, carbohydrate, and amino acid utilization genes in Escherichia coli K-12. *J Bacteriol* **189**, 5429–5440, <https://doi.org/10.1128/JB.01929-06> (2007).
63. Bleuel, C. *et al.* TolC is involved in enterobactin efflux across the outer membrane of Escherichia coli. *J Bacteriol* **187**, 6701–6707, <https://doi.org/10.1128/JB.187.19.6701-6707.2005> (2005).
64. Higgs, P. I., Larsen, R. A. & Postle, K. Quantification of known components of the Escherichia coli TonB energy transduction system: TonB, ExbB, ExbD and FepA. *Mol Microbiol* **44**, 271–281 (2002).
65. Heithoff, D. M. *et al.* Coordinate intracellular expression of Salmonella genes induced during infection. *J Bacteriol* **181**, 799–807 (1999).
66. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology* **132**, 365–386 (2000).
67. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnology letters* **26**, 509–515 (2004).
68. Andersen, C. L., Jensen, J. L. & Orntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research* **64**, 5245–5250, <https://doi.org/10.1158/0008-5472.CAN-04-0496> (2004).
69. Botteldoorn, N. *et al.* Real-time reverse transcription PCR for the quantification of the mntH expression of Salmonella enterica as a function of growth phase and phagosome-like conditions. *Journal of microbiological methods* **66**, 125–135, <https://doi.org/10.1016/j.mimet.2005.11.003> (2006).
70. Spinsanti, G. *et al.* Selection of reference genes for quantitative RT-PCR studies in striped dolphin (Stenella coeruleoalba) skin biopsies. *BMC molecular biology* **7**, 32, <https://doi.org/10.1186/1471-2199-7-32> (2006).
71. Xie, F., Xiao, P., Chen, D., Xu, L. & Zhang, B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant molecular biology*, <https://doi.org/10.1007/s11103-012-9885-2> (2012).
72. Silver, N., Best, S., Jiang, J. & Thein, S. L. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC molecular biology* **7**, 33, <https://doi.org/10.1186/1471-2199-7-33> (2006).
73. Ferraz, L. F. *et al.* Ferric iron uptake genes are differentially expressed in the presence of copper sulfides in Acidithiobacillus ferrooxidans strain LR. *Antonie van Leeuwenhoek* **99**, 609–617, <https://doi.org/10.1007/s10482-010-9533-2> (2011).
74. Cheng, H. Y. *et al.* RmpA regulation of capsular polysaccharide biosynthesis in Klebsiella pneumoniae CG43. *J Bacteriol* **192**, 3144–3158, <https://doi.org/10.1128/JB.00031-10> (2010).
75. Stojiljkovic, I., Baumler, A. J. & Hantke, K. Fur regulon in gram-negative bacteria. Identification and characterization of new iron-regulated Escherichia coli genes by a fur titration assay. *Journal of molecular biology* **236**, 531–545, <https://doi.org/10.1006/jmbi.1994.1163> (1994).
76. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods* **25**, 402–408, <https://doi.org/10.1006/meth.2001.1262> (2001).

Acknowledgements

This study was funded by the São Paulo Research Foundation (FAPESP), grant#2008/11365-1. AEIG was supported by a scholarship from São Paulo Research Foundation (FAPESP), grant#2010/18328-4. LPS was supported by a scholarship from the Coordination for the Improvement of Higher Education Personnel (CAPES, Ministry of Education of Brazil). NMGS received a scholarship from the National Council for Scientific and Technological Development (CNPq, Ministry of Science, Technology, Innovations and Communications of Brazil).

Author Contributions

L.F.C.F. and A.E.I.G. conceived and designed the experiments. A.E.I.G., L.P.S., N.M.G.S. and J.B.H. executed the experiments and the analysis. R.V. conducted the bioinformatics analysis. M.L.R. and M.D. contributed with reagents/materials/analysis tools. L.F.C.F. wrote the manuscript and coordinated its revision. M.L.R. and M.D. assisted with critical revision of the manuscript. All authors revised and approved the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-27420-2>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018