

Transcriptomic studies and assessment of Yersinia pestis reference genes in various conditions

Lionel Koch, Thomas Poyot, Marine Schnetterle, Sophie Guillier, Estelle Soulé, Flora Nolent, Olivier Gorgé, Fabienne Neulat-Ripoll, Eric Valade, Florent Sebbane, et al.

▶ **To cite this version:**

Lionel Koch, Thomas Poyot, Marine Schnetterle, Sophie Guillier, Estelle Soulé, et al.. Transcriptomic studies and assessment of Yersinia pestis reference genes in various conditions. Scientific Reports, 2019, 9 (1), pp.2501. 10.1038/s41598-019-39072-x . hal-02498186

HAL Id: hal-02498186
https://amu.hal.science/hal-02498186

Submitted on 4 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

SCIENTIFIC REPORTS

OPEN

Transcriptomic studies and assessment of *Yersinia pestis* reference genes in various conditions

Received: 2 August 2018
Accepted: 14 December 2018
Published online: 21 February 2019

Lionel Koch[1,2,3], Thomas Poyot[1], Marine Schnetterle[1,3], Sophie Guillier[1,3], Estelle Soulé[1,3], Flora Nolent[1,3], Olivier Gorgé[1,3], Fabienne Neulat-Ripoll[1,3], Eric Valade[1,2,3], Florent Sebbane[4] & Fabrice Biot[1,3]

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) is a very sensitive widespread technique considered as the gold standard to explore transcriptional variations. While a particular methodology has to be followed to provide accurate results many published studies are likely to misinterpret results due to lack of minimal quality requirements. *Yersinia pestis* is a highly pathogenic bacterium responsible for plague. It has been used to propose a ready-to-use and complete approach to mitigate the risk of technical biases in transcriptomic studies. The selection of suitable reference genes (RGs) among 29 candidates was performed using four different methods (GeNorm, NormFinder, BestKeeper and the Delta-Ct method). An overall comprehensive ranking revealed that 12 following candidate RGs are suitable for accurate normalization: *gmk*, *proC*, *fabD*, *rpoD*, *nadB*, *rho*, *thrA*, *ribD*, *mutL*, *rpoB*, *adk* and *tmk*. Some frequently used genes like *16S RNA* had even been found as unsuitable to study *Y. pestis*. This methodology allowed us to demonstrate, under different temperatures and states of growth, significant transcriptional changes of six efflux pumps genes involved in physiological aspects as antimicrobial resistance or virulence. Previous transcriptomic studies done under comparable conditions had not been able to highlight these transcriptional modifications. These results highlight the importance of validating RGs prior to the normalization of transcriptional expression levels of targeted genes. This accurate methodology can be extended to any gene of interest in *Y. pestis*. More generally, the same workflow can be applied to identify and validate appropriate RGs in other bacteria to study transcriptional variations.

Plague is a fatal disease caused by the Gram-negative bacterium *Yersinia pestis*[1] which has been responsible for approximately 200 million deaths in the past and is still a global public health issue. Between 2010 and 2015, the World Health Organization (WHO) reported more than 3200 cases with more than 500 deaths[2]. In 2017, at least 64 patients succumbed during an outbreak in Madagascar[3]. Plague mostly circulates in rodent populations through the bites of infected fleas. Bubonic plague is the usual presentation in humans after fleabites while primary septicemia without detectable bubos is possible. Pneumonic plague may occur after bubonic and septicemic forms. At this stage, bacteria are transmissible by air, causing deadly primary pneumonic presentations[1]. Thus, life cycle of *Y. pestis* between the mammalian hosts through the flea vectors implies that the bacterium senses the oscillations of its environmental temperature. Therefore, it is not surprising that temperature controls several biological processes playing a key role in the colonization of both the mammalian host and the flea vector[4]. In fact, previous comparative transcriptomic and proteomic studies have shown that many genes are differentially expressed when temperature increases from 21/28 to 37 °C and inversely[5].

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) is a widespread[6] gold standard[7] technique to explore transcriptional variations. It is often the easiest and the most cost effective solution

[1]Institut de Recherche Biomédicale des Armées (IRBA), Brétigny-sur-Orge, France. [2]Ecole du Val de Grace (EVDG), Paris, France. [3]Aix Marseille University, INSERM, SSA, IRBA, MCT, Marseille, France. [4]Inserm, University of Lille, CNRS, CHU Lille, Institut Pasteur de Lille, U1019-UMR8204-CIIL-Center for Infection and Immunity of Lille, Lille, France. Correspondence and requests for materials should be addressed to F.B. (email: fbiot.irba@gmail.com)

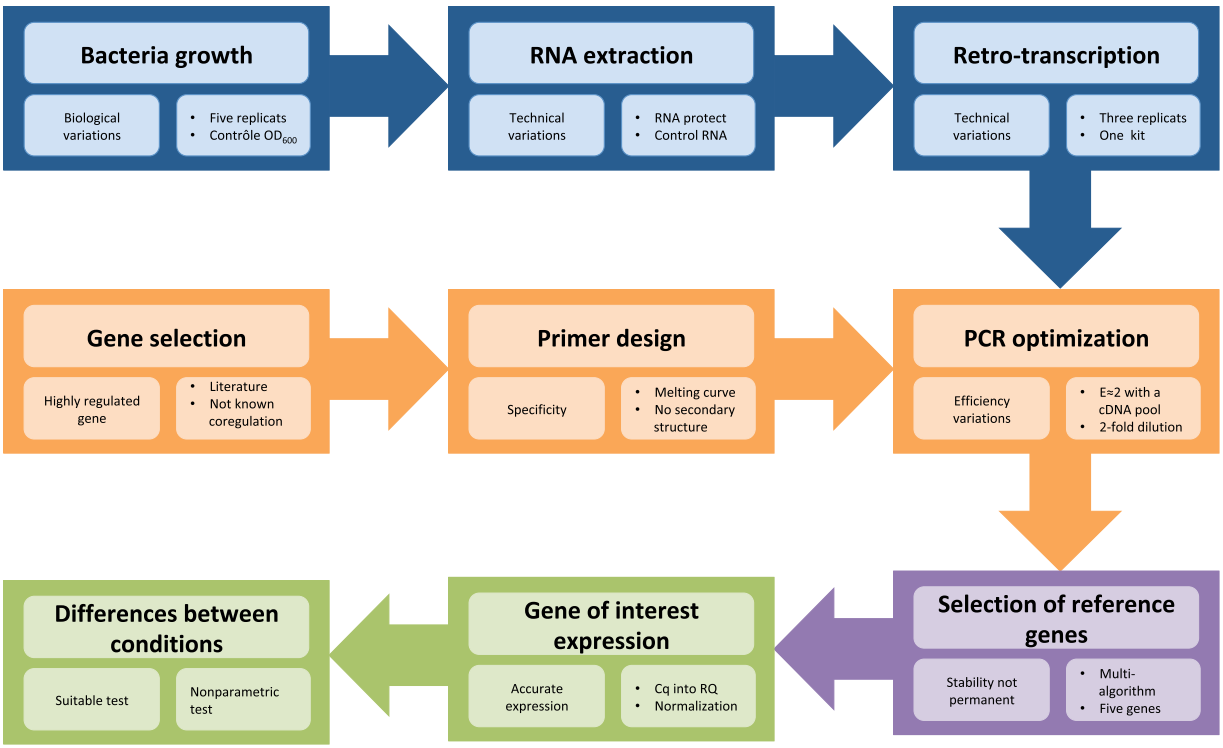


Figure 1. Workflow for validation of RGs and transcriptomic studies. Blue section presents the three steps to obtain suitable cDNA and control its quality. Orange section describes the method to choose pertinent candidates RGs and to obtain suitable PCR conditions from the primers design to the PCR parameters' optimization (E = 2 corresponds to 100% efficiency and E = 1 to 0%). Purple section shows the selection of the appropriate number of stable enough RGs using multiple algorithms. Green section displays the normalization of the gene of interest's expression and the use of the appropriate statistic tests to identify differences.

to measure the genes' expression[8] and to understand complex regulatory networks. Moreover, the expansion of the RNA sequencing (RNA-seq) did not replace the RT-qPCR because these two techniques work complementarily[7]. RNA-seq identifies the most relevant genes and RT-qPCR validates its results[9], especially in the field of environmental and host adaptation[10,11] and antimicrobial response[12].

This specific and sensitive approach requires a very particular methodology[13] to provide with accurate and reliable results[14]. Unfortunately, Bustin *et al.*[15] demonstrated that the majority of published studies did not fulfill all requirements even in the highest impact factor revues. Indeed, the whole workflow must be considered[16] and drastic quality controls have to be performed at each step[17]. With highly pathogenic agents as *Y. pestis*, the method of RNA extraction must be carefully chosen[18] for biosafety concerns as well as to preserve the RNA quality, essential for further reactions[19]. The reverse transcription step transforming extracted RNA into cDNA is critical for the accuracy of the qPCR, which measures the relative quantity (RQ) of cDNA[20]. However, considering numerous technical specifications can minimize the variability[21]. The knowledge of the PCR efficiency and the use of controls are necessary to guarantee the quality and the reproducibility of the results[22]. A normalization strategy is essential to mitigate the technical variations for qPCR during the previous steps, and then enabling comparisons between different samples.

The use of reference genes (RGs) is the most commonly accepted method for normalizing[23]. The ratio of the mRNA concentrations of the genes of interest is reported to those of the RGs, which are considered as stably expressed. The stability of this expression is evaluated with different algorithms, which are elaborated in the method section. However, a suitable and universal set of RGs does not exist and every investigator should select and analyze candidate RGs for every experimental conditions[13]. Several sets of candidate genes have already been evaluated for different bacteria[24–26] but to our best knowledge, RGs have not been investigated in *Y. pestis*. This highly regulated pathogen is an excellent model for studying the variation of its genes expression in response to environmental changes. Our experimental parameters have been chosen to maximize the transcription variability in the reputed stable genes, and hence to ensure our choice of RGs in a broad spectrum of physiological conditions.

We identified the most appropriated combination of RGs, which could be used for any transcriptional study in *Y. pestis* including transcriptomic analysis using RNA-seq. We highlight that this methodology could also be a starting point for a similar study in other species. Furthermore we present, through a workflow strictly applying the MIQE (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) recommendations[13] (Fig. 1), a global approach reducing bias for transcriptomic studies in bacteria. These consensual guidelines describe how to perform and interpret qPCR experiments in order to ensure the reliability and the transparency of the results. Lastly, as a proof-of-concept, we used this set of genes to perform an accurate normalization of six

Resistance-Nodulation-Division (RND) efflux pumps gene expression in *Y. pestis* in function of the temperature and the growth phase. These genes have been selected because they have been reported in other bacteria to be implicated in a wide range of physiological aspects as antimicrobial resistance, bacterial stress responses, fitness, colonization and virulence[27].

Results

Expression level of the candidate RGs. A total of 29 candidate RGs were selected (Table 1) on the basis of our laboratory experience and of an extended literature screening of genes designated as "housekeeping" gene in *Yersinia* genus[28,29] and other genes commonly used as RGs in other bacteria species and present in *Y. pestis* genome[24–26,30]. In order to avoid co-regulation induced bias, we chose genes with various functions, distributed throughout the entire chromosome (Fig. S1) and not belonging to the same operon. The quantification cycle (C_q) values from each of the 29 candidate RGs were generated by RT-qPCR using RNA extracted from exponential and stationary-phase of *Y. pestis* grown at 21 °C, 26 °C and 37 °C in Luria-Bertani (LB) broth. Results are represented in a box-and-whiskers plot (Fig. 2). Position and dispersion parameters are represented in Table S1. Analysis showed consequent differences between each selected candidates, indicating a differential expression depending on the experimental conditions and an internal variability. The C_q value medians ranged from 20.95 for *rpoS* to 31.52 for *ribD* except for *16S rRNA* for which the median was found at 10.83. A C_q value interquartile gap was observed from 0.42 [10.65; 11.07] for 16S rRNA to 3.31 [20.47; 23.78] for *rpoA*. A C_q value variation of 1 corresponds to the doubling of RNA quantity and only five genes (*16S rRNA*, *nadB*, *cysG*, *rpoD* and *tmk*) displayed an interquartile gap equal or less than 1. The range varied between 2.3 [27.16; 29.46] for *rpoB* and 6.97 [18.62; 25.59] for *rpoA*. The C_q value of only five genes (*rpoB*, *nadB*, *mutL*, *gmk* and *16S rRNA*) had a range 3, that corresponds to a 8-fold increased RNA quantity. Under all growing conditions, C_q value variations remain important even for genes with the lowest dispersion parameters.

For each temperature (21; 26 and 37 °C), RNAs have been extracted at exponential and stationary phases generating 6 groups. The intra-group variability was observed as drastically decreased with a C_q value interquartile gap between 0.16 (*glnA*; 21 °C; exponential; [25.71; 25.87]) and 1.42 (*cysK*; 37 °C; exponential; [19.10; 20.52]) and a range between 0.56 (*dnaK*; 21 °C; exponential; [21.26; 21.82]) and 4.01 (*gapA*; 26 °C; exponential; [23.14; 27.15]). The C_q value of five genes (*16S rRNA*, *rpoD*, *fabD*, *rho* and *gyrB*) had an interquartile gap equal or less than 0.5 in each sub-group (Fig. S2). Position and dispersion parameters in each sub-group are represented in Table S2.

Evaluation of candidate RGs stability. The position and dispersion parameters provided an overview of the disparity in expression level and internal variation of genes expression. However, these parameters based on absolute data are not sufficient to evaluate the stability of the expression of each candidate RG. Hence, relative expression has to be taken into account by statistical analysis in order to assess gene stability[31]. Four methods (GeNorm[32], BestKeeper[33], NormFinder[34], and the delta-Ct method[35]) have been used to rank the 29 candidate RGs. BestKeeper is the only one, which evaluates the stability of each gene independently of each other. Each algorithm provides a different ranking (Table 2). However, several genes were considered to be stable by different methods with overlapping ranks. Indeed, based on the results from GeNorm[32] and considering that a M-value lower than 0.7 is the cut-off value to select a suitable reference gene, 15 genes were considered stable: *rpoD*/*proC*, *fabD*, *gmk*, *rho*, *thrA*, *adk*, *ribD*, *nadB*, *mutL*, *tmk*, *rpoB*, *tpiA*, *ftsZ* and *secA* (in descending order with regard to stability). With the exception of *secA*, all the above cited genes as well as *cysG*, *hcaT*, *recA*, *cysK* and *gapA* were also found to be suitable using BestKeeper algorithm and based on the authors recommendations (cut-off value of the standard deviation lower than 1)[33]. Although there is no cut-off value or recommended number of genes for the NormFinder[34] algorithm, ranking using this latter algorithm was mainly in accordance with results obtained from the GeNorm algorithm. Indeed, the 15 genes previously identified as suitable RGs by GeNorm occupied the first 16 positions of the NormFinder ranking. Moreover, the 5 most stable genes (*rpoD*/*proC*, *fabD*, *gmk*, and *rho*) were the same according to these two algorithms and the delta-Ct method[35] ranked in a different order. Lastly, in order to determine a global ranking, we used the RefFinder[36] web-based tool to calculate the geometric mean of the four previous rankings. We found that *gmk* is the most stably expressed gene followed by *proC*, *fabD*, *rpoD*, *nadB*, *rho*, *thrA*, *ribD*, *mutL rpoB*, *adk* and *tmk*. Those twelve genes were considered stable enough to be employed as RGs by the four previously detailed methods and therefore were the best candidates for RT-qPCR experiment in *Y. pestis*. These selected genes were among those with the lowest dispersion C_q parameters.

Set of stable RGs for RT-qPCR. According to the MIQE guidelines[13], the assessment of the stability of the RGs for each experimental condition should be performed prior to conduct the transcriptomic study. Furthermore, only one gene may not be sufficient unless its invariant expression has been proven. Vandesompele *et al.*[32] proposed to test up to five candidate RGs and to determine how many are necessary for a correct normalization. Thus, we propose the panel of the five following genes *gmk*, *proC*, *fabD*, *rpoD* and *nadB*, as an universal panel for transcriptional studies in *Y. pestis*. However, even if these five genes had been validated in each condition of temperature and growth phase (data not shown) and could be proposed for all studies, they might not be the best one for a normalization process in specific conditions. Therefore, we performed the RefFinder analysis by grouping the data according to the temperature or the state of growth (Table S3). By picking the five best-ranked genes belonging to the 12 previously validated genes in each condition, an optimal panel can be proposed for each condition (Table 3). These five genes are ranked from the most stable to the less stable in our conditions but should be used together and reevaluated for each new experiment.

Comparison between normalized expression level for the genes of interest. In *Yersinia enterocolitica*, the expression of *acrAB* (YPO3132-33) gene encoding a RND efflux pump has been reported to be regulated

Gene symbole	Locus tag	Gene description	primer 5′-3′	
Candidate reference gene				
16S rRNA	YPOr01	16S ribosomal RNA	Fw	AGAGATGCTAAAGTGCCTTC
			Rv	CCAACATTTCACAACACGAG
glnA	YPO0024	glutamine synthetase	Fw	GCTGAACATGTTTTGACGATGCT
			Rv	TGCTCTTTCCCTTTGGTATCAGT
gmk	YPO0040	guanylate kinase	Fw	TTTTCATCTTGCCGCCATC
			Rv	GATAACTTCTTCGCTATCCTG
tpiA	YPO0085	triosephosphate isomerase	Fw	ATCCGTGACCATATCGCCAAG
			Rv	TTTCAGTGATGCACCGCCAAC
cysG	YPO0158	uroporphyrin III C-methyltransferase	Fw	GAAAAGTTGCTGACCCACGAC
			Rv	TGCACCGACCAACACCAC
rpsL	YPO0200	30S ribosomal protein S12	Fw	ACTAACGGTTTTGAAGTCAC
			Rv	TCTTTAACACGACCGCCAC
rpoA	YPO0234	DNA-directed RNA polymerase, alpha subunit	Fw	CTGAAAACGCCGAACCTG
			Rv	ATGCCTAAAGAAAGACCAC
mutL	YPO0371	DNA mismatch repair protein	Fw	GTTTCTTTACCATTACGCCAAC
			Rv	TGGCGATATCTCTTCATGCTG
thrA	YPO0459	bifunctional aspartokinase I/homoserine dehydrogenase I	Fw	TGTCCGGTTCACTTTCCTTC
			Rv	ATACCCCAACGCCTTAGCTT
dnaK	YPO0468	molecular chaperone DnaK	Fw	AGGCTGTCACTAACCCTCA
			Rv	CGCTGTGCTTCTTCGTCT
ftsZ	YPO0560	cell division protein FtsZ	Fw	GCCATCTCCAGTCCGTTG
			Rv	ATAGTGTTACCCACGGTCTC
secA	YPO0564	protein translocase subunit SecA	Fw	ATCAACCGCATGGAACCTG
			Rv	TTTCGCTAAACGCTCACGGAAC
rpoD	YPO0643	RNA polymerase sigma factor	Fw	AAGACGGTATCAATCAGGTTC
			Rv	ATTCGCCCGCTTCAACAC
proC	YPO0942	pyrroline-5-carboxylate reductase	Fw	TAAAGCCCCAGTTAATGGCCGATG
			Rv	ACCAACTTATCGCTAAAATCGACCTG
tolB	YPO1124	translocation protein TolB	Fw	TCTGATGTCACCCGCTTGGTC
			Rv	AATAACCAATGCTGATTTGCCAC
gyrA	YPO1216	DNA gyrase A	Fw	CAGTGGCAGAATATCCAAC
			Rv	TTGAACAGCACCGACAAC
fabD	YPO1598	malonyl-CoA transacylase	Fw	ATAATCCGGTACGTTGGACTG
			Rv	AGGCCGGTCAATACTTTACCTG
tmk	YPO1605	thymidylate kinase	Fw	CCTTTAGCGGAAAAACTGCGTGA
			Rv	GGCGGCATACAGCATTAATACCT
gapA	YPO2157	glyceraldehyde-3- phosphate dehydrogenase	Fw	GAAACTGCACGTAAGCAC
			Rv	TTAACGCCCATAACGAAC
nadB	YPO2710	L-aspartate oxidase	Fw	GTATTGCCGCTGTTTTCGAC
			Rv	TGGCAATAAATTCAACGGCTTC
hcaT	YPO2904	3-phenylprop- ionic transporter	Fw	CTTAATGGCTTCAACGAC
			Rv	CAGACAGTAAACGAACCAC
cysK	YPO2992	cysteine synthase A	Fw	CGGTCAAGAGATCAAACCTG
			Rv	AATGCCTTCTTCATCCATC
adk	YPO3118	Adenylate kinase	Fw	ATCACGTTAAATTCAACCCAC
			Rv	CTTACGGACAGTCGCTTC
ribD	YPO3183	bifunctional diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-5-phosphoribosylaminouracil reductase	Fw	GAAACGGAATATTAACTCTG
			Rv	CTAGCTCGCATAATCCAC
recA	YPO3307	recombinase A	Fw	TCCAGATCCTCTACGGTG
			Rv	CGCCAGCTTTCTCAATCAG
rpoS	YPO3355	RNA polymerase sigma factor RpoS	Fw	GCCGAGAAATTGGTTTGACAC
			Rv	AACAACGCCTCGATGCTCAG
rpoB	YPO3747	DNA-directed RNA polymerase, beta subunit	Fw	CCTTTAGCGGAAAAACTGCGTGA
			Rv	ACATTTTCTACTAATTGCACCCT

Continued

Gene symbole	Locus tag	Gene description	primer 5′-3′	
rho	YPO3867	rho termination factor	Fw	AAGAAGTAACCGAGATGCAAC
			Rv	TTTCAGCGACTTGAACGTGAC
gyrB	YPO4094	DNA gyrase B	Fw	GACTATCGCCCGCATGGAAC
			Rv	ACCTCTTTATTGCCCAAATCCTC
Gene of interest				
/	YPO0421	multidrug efflux protein	Fw	CAGTATTACATTGGTGCTC
			Rv	TATCCGTACCATAAGGCAG
/	YPO1001	integral membrane efflux protein	Fw	ATGACCGCTATTTCGTTTATCCTC
			Rv	ATGCCGCCAAATACCGTG
mdtB	YPO2848	multidrug efflux system subunit	Fw	ACAAGACAACGGGCTGATTC
			Rv	CTACTGCCGGATCTTTGAGG
acrD	YPO3043	aminoglycoside/multidrug efflux system	Fw	CGTCGATGCCTTTGGTTCACA
			Rv	CGCACTGACAATATCACTCGT
acrB	YPO3133	multidrug efflux protein	Fw	CGTACCCAGAAAGTACTGAACCA
			Rv	CCGGTGTTCTGACCTTGACCA
/	YPO3482	multidrug efflux protein	Fw	GTGATGTGGATACCCGCTCT
			Rv	CCGTAAATCCAGCGAGTTGT

Table 1. Primers for candidate reference genes and for *acrB* and the five putative RND efflux pumps genes of *Yersinia pestis*.

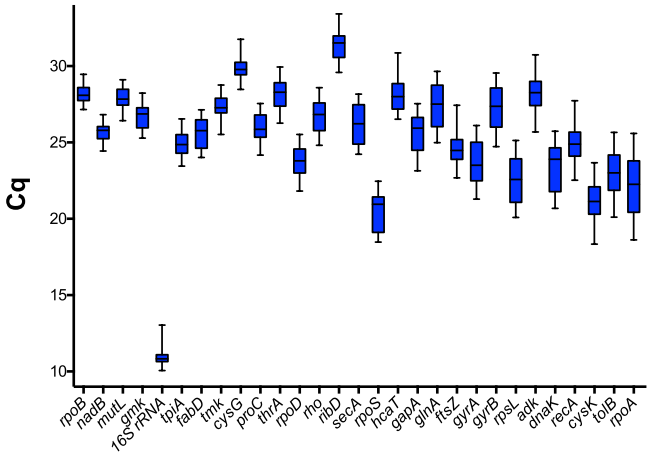


Figure 2. Expression level for the 29 candidate reference genes in all samples. Expression level was represented by the C_q values of each gene, ranked from the smallest to the largest range. Ebox indicated the 25th and 75th percentiles and the whiskers caps represented the maximum and minimum values. The line across the boxes indicated the median.

by temperature and growth phase[37]. In *Y. pestis*, in addition to *acrAB*, five putative RND efflux pumps genes have been identified: YPO0420-21, YPO1000-01, YPO2847-48-49, YPO3043 and YPO3482-83. We assessed the influence of the temperature and the growth phase on their expression.

Based on the expression of five ranked most stable genes *gmk*, *proC*, *fabD*, *rpoD* and *nadB* given by RefFinder according to the GeNorm method, the pairwise variation (V_n/V_{n+1}) was calculated between the normalization factors NF_n and NF_{n+1} to determine the optimal number of required RGs. The genes inclusion has been stopped when the pairwise variation reached the cut-off value below 0.15, in our case after the fourth ranked gene. Therefore, *gmk*, *proC*, *fabD* and *rpoD* have been chosen as RGs (Fig. 3). The normalized gene of interest's Relative Quantity (RQ) position and dispersion data have been reported in Table S4 and represented in Fig. 4. Due to the non-normal distribution of the data the non-parametric Kruskal and Wallis test followed by a Holm-Bonferroni procedure (α risk set at 5%) have been used.

The results obtained showed that the regulation of *acrAB* in *Y. pestis* is function of temperature and state of growth and that RND efflux pumps genes expression is sensitive to the same stimuli. However, these trends that have been observed only from transcriptomic data do not allow us to extrapolate any phenotypic change or make any biological interpretation. For example, locus tag YPO1001 gene (Fig. 4B) transcription diminished with temperature with a 2.5-fold median decrease at the exponential phase and a 2.2-fold median decrease at the exponential phase between 21 °C and 37 °C. At the opposite, locus tag YPO3482 gene (Fig. 4F) showed a moderate increase in function of the temperature (only significant at the stationary phase). The other genes expression

Ranking	GeNorm		NormFinder		BestKeeper		Delta Ct		RefFinder	
	Gene	M value	Gene	Stab. value	Gene	Std dev	Gene	Av. STDEV	Gene	Score
1	***rpoD*** | ***proC***	0.313	***gmk***	0.213	*16S rRNA*	0.36	***gmk***	0.76	***gmk***	2.38
2			***fabD***	0.289	***nadB***	0.42	***fabD***	0.77	***proC***	3.08
3	***fabD***	0.451	***proC***	0.346	***rpoB***	0.44	***proC***	0.80	***fabD***	3.83
4	***gmk***	0.479	***rpoD***	0.411	***tmk***	0.51	***rho***	0.83	***rpoD***	3.85
5	***rho***	0.508	***rho***	0.419	***mutL***	0.51	***rpoD***	0.84	***nadB***	5.05
6	***thrA***	0.532	***nadB***	0.429	*cysG*	0.54	***nadB***	0.84	***rho***	6.32
7	***adk***	0.547	***thrA***	0.430	*tpiA*	0.62	***thrA***	0.85	***thrA***	8.01
8	***ribD***	0.560	***ribD***	0.448	***gmk***	0.64	***ribD***	0.86	***ribD***	8.24
9	***nadB***	0.572	***adk***	0.473	***ribD***	0.69	***adk***	0.86	***mutL***	8.41
10	***mutL***	0.583	***mutL***	0.519	***proC***	0.69	***mutL***	0.89	***rpoB***	9.32
11	***tmk***	0.611	*secA*	0.586	***rpoD***	0.75	*secA*	0.90	***adk***	9.60
12	***rpoB***	0.631	*ftsZ*	0.626	*ftsZ*	0.77	*gyrB*	0.93	***tmk***	9.78
13	*tpiA*	0.651	***tmk***	0.641	*hcaT*	0.81	*ftsZ*	0.96	*16S rRNA*	11.18
14	*ftsZ*	0.671	*gyrB*	0.642	***thrA***	0.82	***rpoB***	0.97	*tpiA*	12.16
15	*secA*	0.690	***rpoB***	0.644	***adk***	0.87	*tpiA*	0.97	*ftsZ*	12.72
16	*gyrB*	0.707	*tpiA*	0.649	***rho***	0.89	***tmk***	0.97	*secA*	14.14
17	*gapA*	0.725	*gapA*	0.677	*recA*	0.90	*gapA*	0.99	*cysG*	15.13
18	*glnA*	0.745	*gyrA*	0.772	***fabD***	0.91	*glnA*	1.02	*gyrB*	15.94
19	*gyrA*	0.762	*cysG*	0.775	*cysK*	0.94	*gyrA*	1.03	*gapA*	17.70
20	*rpsL*	0.782	*glnA*	0.785	*gapA*	0.96	*cysG*	1.05	*hcaT*	19.68
21	*tolB*	0.798	*hcaT*	0.799	*rpoS*	1.01	*tolB*	1.08	*gyrA*	20.08
22	*dnaK*	0.813	*recA*	0.807	*secA*	1.17	*hcaT*	1.08	*glnA*	20.45
23	*cysG*	0.829	*tolB*	0.817	*tolB*	1.19	*recA*	1.09	*recA*	21.32
24	*recA*	0.844	*16S rRNA*	0.857	*gyrB*	1.21	*rpsL*	1.09	*tolB*	21.98
25	*hcaT*	0.858	*rpsL*	0.871	*gyrA*	1.28	*16S rRNA*	1.11	*rpsL*	24.08
26	*16S rRNA*	0.874	*dnaK*	0.880	*dnaK*	1.29	*dnaK*	1.11	*dnaK*	24.94
27	*rpoA*	0.908	*rpoA*	1.295	*glnA*	1.33	*rpoA*	1.42	*cysK*	25.41
28	*cysK*	0.953	*cysK*	1.380	*rpsL*	1.38	*cysK*	1.55	*rpoS*	26.75
29	*rpoS*	1.017	*rpoS*	1.794	*rpoA*	1.78	*rpoS*	1.89	*rpoA*	27.49

Table 2. Stability values and ranking of 29 candidate reference genes. Ranking are based on results from GeNorm, NormFinder, BestKeeper and Delta-Ct. M value: stability value calculated from GeNorm; Stab. Value: calculated from NormFinder; std dev: standard deviation calculated from BestKeeper, Av. STDEV: average standard deviation calculated from Delta-Ct method. Score calculate from RefFinder; In bold, the 12 genes validated in every conditions.

State of growth	All conditions	Exponential	Stationary	21 °C	26 °C	37 °C
Recommended reference genes	*gmk*	*fabD*	*nadB*	*rho*	*gmk*	*gmk*
	proC	*gmk*	*fabD*	*gmk*	*rho*	*proC*
	fabD	*nadB*	*gmk*	*nadB*	*fabD*	*nadB*
	rpoD	*proC*	*rpoD*	*mutL*	*nadB*	*mutL*
	nadB	*rpoD*	*proC*	*proC*	*proC*	*thrA*
State of growth	**Exponential 21 °C**	**Exponential 26 °C**	**Exponential 37 °C**	**Stationary 21 °C**	**Stationary 26 °C**	**Stationary 37 °C**
Recommended reference genes	*gmk*	*gmk*	*proC*	*rho*	*rho*	*rho*
	rho	*fabD*	*gmk*	*fabD*	*rpoB*	*nadB*
	fabD	*rho*	*rpoD*	*gmk*	*gmk*	*fabD*
	nadB	*nadB*	*adk*	*nadB*	*nadB*	*gmk*
	mutL	*rpoB*	*fabD*	*mutL*	*thrA*	*rpoD*

Table 3. Five recommended reference genes in function of the state of growth. The five recommended reference genes correspond to the five best-ranked genes belonging to the 12 previously validated genes in each condition. They are recommended in exponential or stationary phase regardless the temperature and for each temperature regardless the state of growth. Five genes are also proposed for each specific condition of temperature and state of growth.

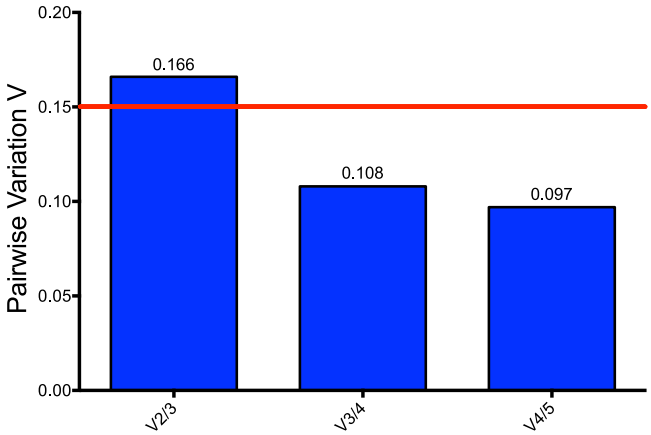


Figure 3. Pairwise variation calculated by GeNorm algorithm. The pairwise variation (V_n/V_{n+1}) was calculated between the normalization factors NF_n and NF_{n+1} to determine the optimal number of reference genes for accurate RT-qPCR normalization. The cut-off value was 0.15.

levels (Fig. 4A,C,D,E) showed less clear patterns. Despite the lack of formal hypothesis concerning the gene regulation, some statistically significant variations in RND efflux pumps expression have been demonstrated to depend on temperature and growth phase. Bearing in mind that the total efflux capacity of the bacterium is dependent of multiples pumps, these variations probably have an important role on the physiology of *Y. pestis* during the different steps of its life cycle. This role should be evaluated through further studies, where the same methodology could be applied.

Effect of normalizing with non-validated RGs. To illustrate the impact of the choice of the most appropriate RGs, we performed two alternative normalizations of the gene expression data. On one hand, we used the expression of the *16S RNA* gene, the most commonly used normalizer[30], and on the other hand we used the geometrical mean of the expression of our five last ranked genes, namely *rpsL*, *dnaK*, *cysK*, *rpoS* and *rpoA*. The results are presented respectively in Figs S3 and S4. By these two normalizations, the number of statistically significant results dramatically dropped from 32 to 14 and 16 respectively. Furthermore, three differences have been found as statistically significant with the *16S RNA* normalization, but not with our RGs.

Discussion

The selection and validation of RGs are the most critical steps for transcriptomic studies[13]. For the highly pathogenic bacterium *Y. pestis*, a set of validated RGs might be proposed for exploring gene expression in many experimental conditions. However our results do not support the established practices and are in accordance with Bustin *et al*. who demonstrated that publications using RT-qPCR have deficiencies in the validation of experimental designs even in high impact factor journals[15]. Indeed, we observed that several genes such as *16S RNA*, *gyrA*, *recA* or *rpoA*, which are commonly used to normalize the expression of targeted genes[30] are unsuitable for RT-qPCR in *Y. pestis*. We showed that growth conditions increased internal variability even in the most stable expressed genes and that the use of not-suitable RGs exposes the author to the risk of misinterpretation of results. It emphasizes the absolute necessity of choosing the right RGs to increase the signal-to-noise ratio and to validate results in the specific experimental condition. To our best knowledge, this is the first study validating RGs in *Y. pestis* on various growth conditions. More generally, we highlight that the previous validation of RGs before performing transcriptomic studies on prokaryotes is not standard enough while this step should be the prior requirement. This methodology is more conducted on eukaryotes, reason why all the algorithms we used were developed for humans, animals or plants.

The case of *16S RNA* deserves to be more scrutinized because it is the most used gene on RT-qPCR in bacteria[30]. Even if this gene was found as the most stable gene by the BestKeeper algorithm, it did not reach the minimum requirements to be accepted by the other algorithms. In our study, the normalization using *16S RNA* expression did not allow us to highlight as many differences as our validated standardization and even more identified differences that most probably do not exist. The suitability of *16S RNA* as a RG was already questioned[38]. Moreover, while the MIQE recommendations state that the abundances should show a strong correlation with the total amounts of mRNA present in the samples[13], *16S RNA* has a very high level of transcription with about a 10^5-fold ratio between its transcripts and the targeted genes mRNA. For these reasons, *16S RNA* cannot be recommended as a suitable RG at least under our conditions. This illustrates the necessity of performing validation tests for each experimental condition and using several algorithms to perform this test.

The need for suitable RGs does not limit to RT-qPCR experiments but it may be extend to the recent advances in molecular biology as reverse transcription digital PCR (RT-dPCR). Through "The Digital MIQE Guidelines", Huggett *et al*., who signed the MIQE Guidelines in 2009, emphasized the necessity of justifying the number and the choice of RGs[39] for RT-dPCR. Even if absolute quantification obtained by dPCR have been proven to be more precise and reproducible than qPCR, these two technologies have been recently qualified as complimentary to obtain the best results for any experiment[40].

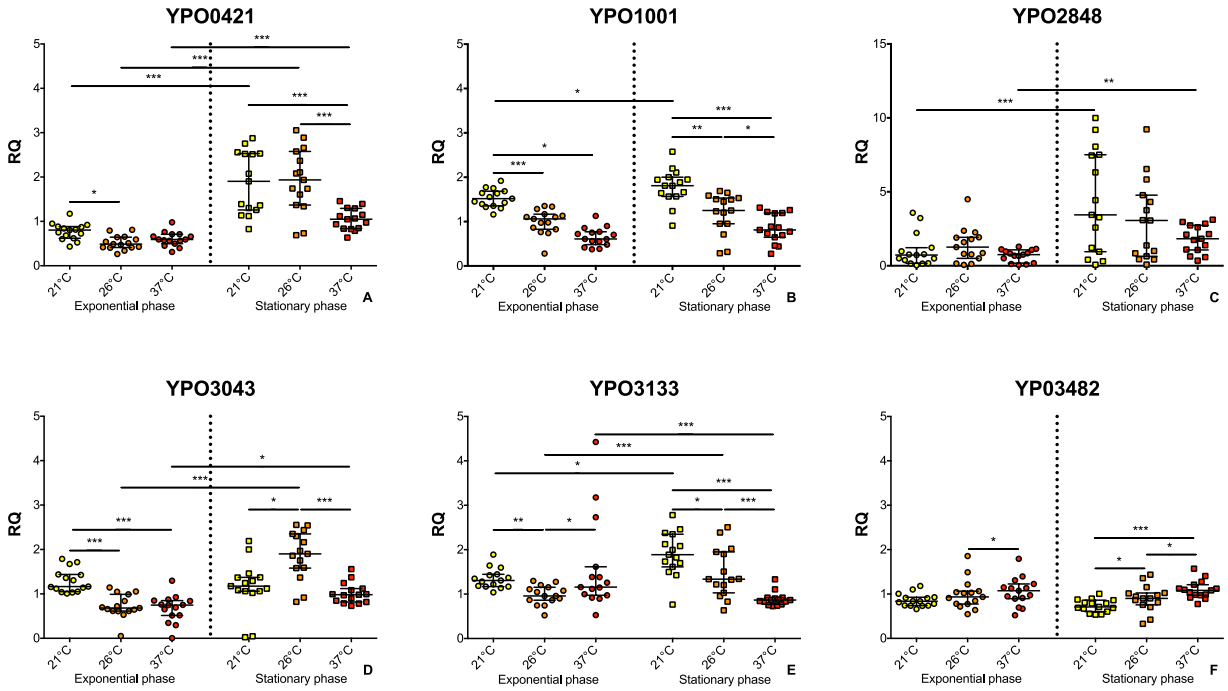


Figure 4. Expression of the gene of RND efflux pumps in function of temperature and state of growth. All data are presented as the relative expression levels obtained by normalization of the relative quantity for the genes of interest (A: YPO0421; B: YPO1001; C: YPO2848; D: YPO3043; E: YPO3133; F: YPO3482). Median and interquartile gap were shown. Statistical analysis was performed using Kruskal Wallis test followed by a Holm-Bonferroni procedure: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

In conclusion, this methodology allowed us to observe slight variations of efflux pump gene's expression in *Y. pestis*. These differences could not be highlighted in previous transcriptomic studies using microarray expression data in *Y. pestis*[41–43]. The risk of bias is limited by using a validated set of RGs to normalize RT-qPCR data and following the MIQE recommendations with rigorous quality controls. This experimental workflow will authorize us to explore not only the interplay between the different active efflux systems by measuring their expression under various conditions but more widely all the transcriptomic changes in *Y. pestis* during its complete cycle of life.

Methods

All steps are designed to follow the MIQE guidelines[13]. Graphical representation has been made using Prism® 6 (GraphPad Software, Inc.).

Bacterial strain and growth conditions. All experiments with live agents were realized in a Biosafety Level 3 laboratory. *Yersinia pestis* CO92 strain[44] was cultivated on Trypticase soja plates (Biomérieux) at 26 °C for 24 hours. Bacteria were grown overnight at experimental temperatures (21 °C, 26 °C or 37 °C) in a Luria-Bertani (LB) Lennox broth (Becton-Dickinson), then diluted in a fresh LB Lennox broth at OD_{600} of 0.010 and grown for an additional time until the culture reaches an OD_{600} around 0.600 [0.582; 0.650] for the exponential phase and an OD_{600} around 2.00 [1.815; 2.198] for the stationary phase (Table S5). A spectrophotometer BioPhotometer® Plus (Eppendorf) was used for OD determination. The MIQE guidelines[13] identified the sample acquisition as the first potential source of experimental variability in RT-qPCR experiment. In order to control sample-to-sample variation, we processed five independent biological replicates for each condition (30 in total).

RNA isolation and cDNA preparation. At each defined growth's state, RNA was stabilized with the RNAprotect® Cell Reagent (Qiagen). Bacteria were lysed and RNA was purified with the RNeasy® Lipid Tissue Mini Kit (Qiagen). Samples were also treated with the RNase-Free DNase Set (Qiagen) for 10 min in the column. Extraction products have been stored at −80 °C. RNA concentration [150.8 ng/µL; 375.1 ng/µL] and purity was determined in duplicate using a NanoDrop 2000 spectrophotometer (Thermo Scientific) with 1.5 µL of product. Both $OD_{260/280}$ [2.13 and 2.20] and $OD_{260/230}$ [1.80; 2.49] ratios around 2 were considered as "pure" according to the manufacturer's recommendations. RNA integrity has been verified using a Bioanalyzer® 2100 (Agilent) with a RNA Nano Chip (Agilent) and the RNA integrity number [9.8; 10] showed nucleic acid quality close to the maximal value[45] (Table S5). 1 µg RNA in a total reaction volume of 50 µL was used to prepare cDNA using the Reverse Transcription Core kit 500 RT-RTCK-05 (Eurogentec) with random nonameric primers. The initial step lasted 10 min at 25 °C and was followed by the RT step 30 min at 48 °C and the inactivation step 5 min at 95 °C. cDNA were stored at −20 °C. Because the RT step is known to be an important source of external variation[21], three separated reverse transcriptions were performed for each extraction product (90 in total) and only one manufactured kit was used for all the procedure. A pool of cDNA was constituted by adding the same quantity of each

RNA extract. This pool has been used as an internal calibrator for the qPCR experiment. A no-RT control with no reverse transcriptase has been performed to validate the absence of genomic DNA contamination.

Selection of candidate RGs, genes of interest and primer design. For each identified gene, a pair of primers was designed using the Oligo7 primer analysis software (DBA Oligo Inc.)[46] based on the published genome[44] (Table 3). The genes of interest were those coding all potential RND efflux pumps in *Y. pestis*, based on a homology study of the AcrB protein[47,48] using the COG database[49]. The specificity of the primers were verified *in silico* using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) then *in vitro* as well as the lack of primer dimers by the presence of a single melting curve point after a qPCR, showing a single product for each pair of primers[50].

Quantitative PCR. qPCR was performed in a LightCycler® 480 (Roche) using the LightCycler® 480 SYBR Green I Master (Roche) in 96 wells white PCR microplate and a sealing film (Fisher scientific). The reaction mixture, which was 20 µL in total volume, contained 10 µL of LightCycler® 480 SYBR Green I Master 2 × (Roche), an equal quantity depending on the reaction of forward and reverse primer at 10 µmol/L and 1 µL of cDNA at the appropriate dilution. The mixture was filled to 20 µL by PCR grade water. The PCR program was started from one step of pre-incubation at 95 °C for 5 min, followed by 45 amplification cycles. Each cycle was started by a denaturation step at 95 °C for 10 s and followed by an annealing step at a specific temperature for a specific time depending on the reaction. The amplification step lasted 5 s at 72 °C. The last step was a melting curve analysis. For each run, 2 No Template Controls (NTC) have been performed. The quantity of primers and cDNA as well as the annealing time and temperature were determined to optimize the PCR efficiency (Table S6) as explained below.

Determination of the optimal quantitative real-time PCR conditions and the reaction efficiency. We choose to express data obtained from the LightCycler® 480 Software release 1.5.1.62 (Roche) as C_q values[51] calculated by the 2(-Delta Delta C(T)) method[52]. PCR conditions for each candidate reference gene were determined to optimize the PCR efficiency by the generation of a three-point standard curve based on a two-fold dilution series of the cDNA pool. The efficiency was calculated from the slope of the standard curve generated for each run in the following equation $E = 10^{(-1/slope)}$, where $E = 2$ corresponds to 100% efficiency and $E = 1$ to 0%[22]. The efficiency for a specific dilution ratio was obtained from the standard curve passing by the two-fold upper and lower dilution. Because low PCR efficiency and high differences between PCR efficiencies can lead to dramatic quantification error, the E value should be as close as 2 as possible and homogenous for all the studied genes[53]. A value between 1.900 and 2.100 (90 and 110% respectively) is usually considered as suitable[26]. PCR parameters including the dilution of the cDNA sample were determined for each candidate reference gene to reach an efficiency of the reaction as close as possible from 2. All PCR parameters are reported in Table S6. The calculated efficiencies for the candidate RGs varied from 1.966 to 2.025 (96.6% to 102.5%) and from 1.957 to 2.007 (95.7% to 100.7%) for the genes of interest. The mean for the candidate reference genes was 1.986 with a standard deviation of 0.013 and respectively 1.989 and 0.026 for the genes of interest showing a high PCR efficiency level (close to 100%) and a good homogeneity between the different reactions. It is important to note that efficiency close to 2 is very important to obtain because some of algorithms as BestKeeper and Delta-Ct methods as well as the RefFinder web-tool use non-transformed data. Efficiency too different to 100% could induce some bias. The internal variability, estimated by calculating the difference of C_q values between two identical and representative samples represented by the pool previously described was always lower than 0.30.

Evaluation and validation of candidate RGs stability. In order to avoid inter-run variability[51], the samples were tested in one unique run in addition of four wells containing cDNA pool and two wells of negative control. No replicate has been performed at this step. The stability of reference gene candidates has been evaluated using four methods: Genorm[32], NormFinder[34], BestKeeper[33] and the comparative Delta-Ct method[35] combined in the online package RefFinder[36].

- The GeNorm algorithm calculates for each candidate reference gene the pairwise variation with all other reference gene candidates. The control gene stability measurement M is defined as the average pairwise variation of one gene with all the other candidates. With the stepwise exclusion of the highest M value, the two most stable genes can be selected. Vandesompele *et al*. recommended to use at least the three most stable genes and to stepwise include the next most stable gene until the inclusion has no significant effect anymore[32].
- The NormFinder algorithm combines the intra and inter-group variations to estimate the systematic error, which is induced by the use of one reference gene candidate. Andersen *et al*. assumed that the average of the candidate genes expression would show less systematic variation than the individual candidate genes and advised to use several genes for normalization[34].
- The BestKeeper algorithm discards the less stable candidate RGs by calculating the standard deviation (SD) and a coefficient of variance of the derived crossing points. Pfaffl *et al*. advised to use the geometric mean of the best candidate RGs to normalize the expression data of the genes of interest. All those best genes are combined into an index. The Pearson correlation coefficient (r) and the probability p are calculated to describe the relation between the index and each candidate[33].
- For Silver *et al*., the Delta-Ct algorithm is based on the comparison of the expression of all possible pairs of genes within each pair-wise sample. The stability of the candidate RGs is estimated from the range and the percentiles of this difference and used to rank the candidate RGs[35].

For using GeNorm and NormFinder, expression data had to be preliminary transformed into relative quantities (RQ) by taking into account the previously described efficiency according to the equation $RQ = E^{\Delta Ct(control\text{-}sample)}$. On the opposite side, BestKeeper and Delta-Ct methods used non-transformed data. On

the RefFinder[36] algorithm, Xie *et al*. used the C_q data of each sample for each gene and processed them with the four previous algorithms. The global comprehensive ranking of the candidate RGs is based on the geometric mean of the ranking from the four algorithms. Expression data are not transformed and efficiency is supposed to be very close to 2.

Application to the study of efflux pumps. In order to study the expression of the genes of interest, we used the same methodology for the primer design and for the optimization of the qPCR than those developed for the candidate RGs. However, the RGs have been determined by using only the GeNorm algorithm with a cut-off value at 0.15. Indeed, this algorithm is the only one, which takes into account the PCR efficiency, indicates the number of references gene to use and provides a cut-off control value. The RQ of the genes of interest have been normalized using the geometrical mean of the RQ of the selected RGs. Statistical analysis has been performed by employing R® software (The R foundation).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Prentice, M. B. & Rahalison, L. Plague. *Lancet Lond. Engl.* **369**, 1196–1207 (2007).
2. Bertherat, E. Plague around the world, 2010–2015. *Wkly. Epidemiol. Rec.* **91**, 89–104 (2016).
3. Tsuzuki, S. *et al*. Dynamics of the pneumonic plague epidemic in Madagascar, August to October 2017. *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* **22** (2017).
4. Schotthoefer, A. M. *et al*. Effects of temperature on the transmission of *Yersinia Pestis* by the flea, *Xenopsylla Cheopis*, in the late phase period. *Parasit. Vectors* **4**, 191 (2011).
5. Pieper, R. *et al*. Integral and peripheral association of proteins and protein complexes with *Yersinia pestis* inner and outer membranes. *Proteome Sci.* **7**, 5 (2009).
6. Huggett, J. F., O'Grady, J. & Bustin, S. qPCR, dPCR, NGS – A journey. *Biomol. Detect. Quantif.* **3**, A1–A5 (2015).
7. Costa, C., Giménez-Capitán, A., Karachaliou, N. & Rosell, R. Comprehensive molecular screening: from the RT-PCR to the RNA-seq. *Transl. Lung Cancer Res.* **2**, 87–91 (2013).
8. Nonis, A., De Nardi, B. & Nonis, A. Choosing between RT-qPCR and RNA-seq: a back-of-the-envelope estimate towards the definition of the break-even-point. *Anal. Bioanal. Chem.* **406**, 3533–3536 (2014).
9. Fang, Z. & Cui, X. Design and validation issues in RNA-seq experiments. *Brief. Bioinform.* **12**, 280–287 (2011).
10. Chen, S., Thompson, K. M. & Francis, M. S. Environmental Regulation of *Yersinia*Pathophysiology. *Front. Cell. Infect. Microbiol.* **6** (2016).
11. Brogaard, L. *et al*. Concurrent host-pathogen gene expression in the lungs of pigs challenged with *Actinobacillus pleuropneumoniae*. *BMC Genomics* **16**, 417 (2015).
12. Biot, F. V. *et al*. Interplay between three RND efflux pumps in doxycycline-selected strains of *Burkholderia thailandensis*. *PloS One* **8**, e84068 (2013).
13. Bustin, S. A. *et al*. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).
14. Bustin, S. A. *et al*. The need for transparency and good practices in the qPCR literature. *Nat. Methods* **10**, 1063–1067 (2013).
15. Bustin, S. & Nolan, T. Talking the talk, but not walking the walk: RT-qPCR as a paradigm for the lack of reproducibility in molecular research. *Eur. J. Clin. Invest.* **47**, 756–774 (2017).
16. Remans, T. *et al*. Reliable gene expression analysis by reverse transcription-quantitative PCR: reporting and minimizing the uncertainty in data accuracy. *Plant Cell* **26**, 3829–3837 (2014).
17. Derveaux, S., Vandesompele, J. & Hellemans, J. How to do successful gene expression analysis using real-time PCR. *Methods San Diego Calif* **50**, 227–230 (2010).
18. Phongsisay, V., Perera, V. N. & Fry, B. N. Evaluation of eight RNA isolation methods for transcriptional analysis in *Campylobacter jejuni*. *J. Microbiol. Methods* **68**, 427–429 (2007).
19. Fleige, S. & Pfaffl, M. W. RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Aspects Med.* **27**, 126–139 (2006).
20. Ståhlberg, A., Håkansson, J., Xian, X., Semb, H. & Kubista, M. Properties of the Reverse Transcription Reaction in mRNA Quantification. *Clin. Chem.* **50**, 509–515 (2004).
21. Bustin, S. *et al*. Variability of the reverse transcription step: practical implications. *Clin. Chem.* **61**, 202–212 (2015).
22. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
23. Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.* **6**, 279–284 (2005).
24. Peng, S., Stephan, R., Hummerjohann, J. & Tasara, T. Evaluation of three reference genes of *Escherichia coli* for mRNA expression level normalization in view of salt and organic acid stress exposure in food. *FEMS Microbiol. Lett.* **355**, 78–82 (2014).
25. Alqarni, B., Colley, B., Klebensberger, J., McDougald, D. & Rice, S. A. Expression stability of 13 housekeeping genes during carbon starvation of *Pseudomonas aeruginosa*. *J. Microbiol. Methods* **127**, 182–187 (2016).
26. Kałużna, M., Kuras, A. & Puławska, J. Validation of reference genes for the normalization of the RT-qPCR gene expression of virulence genes of *Erwinia amylovora* in apple shoots. *Sci. Rep.* **7**, 2034 (2017).
27. Li, X.-Z., Plésiat, P. & Nikaido, H. The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clin. Microbiol. Rev.* **28**, 337–418 (2015).
28. Laukkanen-Ninios, R. *et al*. Population structure of the *Yersinia pseudotuberculosis* complex according to multilocus sequence typing. *Environ. Microbiol.* **13**, 3114–3127 (2011).
29. Kalia, V. C. & Kumar, P. Genome Wide Search for Biomarkers to Diagnose *Yersinia* Infections. *Indian J. Microbiol.* **55**, 366–374 (2015).
30. Rocha, D. J. P., Santos, C. S. & Pacheco, L. G. C. Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. *Antonie Van Leeuwenhoek* **108**, 685–693 (2015).
31. Gutierrez, L. *et al*. The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnol. J.* **6**, 609–618 (2008).
32. Vandesompele, J. *et al*. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034 (2002).
33. 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. *Biotechnol. Lett.* **26**, 509–515 (2004).

34. Andersen, C. L., Jensen, J. L. & Ørntoft, 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 Res.* **64**, 5245–5250 (2004).
35. 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 Mol. Biol.* **7**, 33 (2006).
36. Xie, F., Xiao, P., Chen, D., Xu, L. & Zhang, B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.*, https://doi.org/10.1007/s11103-012-9885-2 (2012).
37. Raczkowska, A., Trzos, J., Lewandowska, O., Nieckarz, M. & Brzostek, K. Expression of the AcrAB Components of the AcrAB-TolC Multidrug Efflux Pump of *Yersinia enterocolitica* Is Subject to Dual Regulation by OmpR. *PloS One* **10**, e0124248 (2015).
38. Valihrach, L. & Demnerova, K. Impact of normalization method on experimental outcome using RT-qPCR in *Staphylococcus aureus*. *J. Microbiol. Methods* **90**, 214–216 (2012).
39. Huggett, J. F. *et al*. The digital MIQE guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. *Clin. Chem.* **59**, 892–902 (2013).
40. Taylor, S. C., Laperriere, G. & Germain, H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci. Rep.* **7** (2017).
41. Han, Y. *et al*. Comparative transcriptomics in *Yersinia pestis*: a global view of environmental modulation of gene expression. *BMC Microbiol.* **7**, 96 (2007).
42. Vadyvaloo, V., Jarrett, C., Sturdevant, D. E., Sebbane, F. & Hinnebusch, B. J. Transit through the flea vector induces a pretransmission innate immunity resistance phenotype in *Yersinia pestis*. *PLoS Pathog.* **6**, e1000783 (2010).
43. Schrimpe-Rutledge, A. C. *et al*. Comparative omics-driven genome annotation refinement: application across *Yersiniae*. *PloS One* **7**, e33903 (2012).
44. Parkhill, J. *et al*. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527 (2001).
45. Schroeder, A. *et al*. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **7**, 3 (2006).
46. Rychlik, W. OLIGO 7 primer analysis software. *Methods Mol. Biol. Clifton NJ* **402**, 35–60 (2007).
47. Udani, R. A. & Levy, S. B. MarA-like regulator of multidrug resistance in *Yersinia pestis*. *Antimicrob. Agents Chemother.* **50**, 2971–2975 (2006).
48. Lister, I. M., Raftery, C., Mecsas, J. & Levy, S. B. *Yersinia pestis* AcrAB-TolC in antibiotic resistance and virulence. *Antimicrob. Agents Chemother.* **56**, 1120–1123 (2012).
49. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
50. Ririe, K. M., Rasmussen, R. P. & Wittwer, C. T. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.* **245**, 154–160 (1997).
51. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, R19 (2007).
52. 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 San Diego Calif* **25**, 402–408 (2001).
53. Ramakers, C., Ruijter, J. M., Deprez, R. H. L. & Moorman, A. F. M. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **339**, 62–66 (2003).

Acknowledgements

This work was supported by the French Ministry of Defense (grant no. PDH-2-NRBC-4-B-4109).

Author Contributions

L.K. and F.B. conceived and supervised the project. L.K., T.P. and F.B. designed the experiments and developed data analysis procedures. L.K., T.P., S.B. and F.B. wrote the paper. L.K., T.P., M.S., S.G., F.N. and E.S. performed all the experiments. L.K., S.B. and F.B. analyzed and interpreted the data. O.G., F.N.-R. and E.V. provided support for the experiments.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-39072-x.

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) 2019