Dear Author,

Here are the final proofs of your article. Please check the proofs carefully.

All communications with regard to the proof should be sent to bmcproductionteam2@spi-global.com.

Please note that at this stage you should only be checking for errors introduced during the production process. Please pay particular attention to the following when checking the proof:

- Author names. Check that each author name is spelled correctly, and that names appear in the correct order of first name followed by family name. This will ensure that the names will be indexed correctly (for example if the author's name is 'Jane Patel', she will be cited as 'Patel, J.').
- Affiliations. Check that all authors are cited with the correct affiliations, that the author who will receive correspondence has been identified with an asterisk (*), and that all equal contributors have been identified with a dagger sign (†).
- Ensure that the main text is complete.
- Check that figures, tables and their legends are included and in the correct order.
- Look to see that queries that were raised during copy-editing or typesetting have been resolved.
- Confirm that all web links are correct and working.
- Ensure that special characters and equations are displaying correctly.
- Check that additional or supplementary files can be opened and are correct.

Changes in scientific content cannot be made at this stage unless the request has already been approved. This includes changes to title or authorship, new results, or corrected values.

How to return your corrections

Returning your corrections via online submission:
- Please provide details of your corrections in the online correction form. Always indicate the line number to which the correction refers.

Returning your corrections via email:
- Annotate the proof PDF with your corrections.
- Send it as an email attachment to: bmcproductionteam2@spi-global.com.
- Remember to include the journal title, manuscript number, and your name when sending your response via email.

After you have submitted your corrections, you will receive email notification from our production team that your article has been published in the final version. All changes at this stage are final. We will not be able to make any further changes after publication.

Kind regards,

BioMed Central Production Team 2

BMC Veterinary Research

RESEARCH ARTICLE Open Access

Transcriptomics of *Haemophilus* (*Glässerella*) *parasuis* serovar 5 subjected to culture conditions partially mimetic to natural infection for the search of new vaccine antigens

Álvaro Álvarez-Estrada, César B. Gutiérrez-Martín*, Elías F. Rodríguez-Ferri and Sonia Martínez-Martínez

Abstract

Background: *Haemophilus* (*Glässerella*) *parasuis* is the etiological agent of Glässer's disease in pigs. Control of this disorder has been traditionally based on bacterins. The search for alternative vaccines has focused mainly on the study of outer membrane proteins. This study investigates the transcriptome of *H.* (*G.*) *parasuis* serovar 5 subjected to in vitro conditions mimicking to those existing during an infection (high temperature and iron-restriction), with the aim of detecting the overexpression of genes coding proteins exposed on bacterial surface, which could represent good targets as vaccine candidates.

Results: The transcriptomic approach identified 13 upregulated genes coding surface proteins: TbpA, TbpB, HxuA, HxuB, HxuC, FhuA, FimD, TolC, an autotransporter, a protein with immunoglobulin folding domains, another large protein with a tetratricopeptide repeat and two small proteins that did not contain any known domains. Of these, the first six genes coded proteins being related to iron extraction.

Conclusion: Six of the proteins have already been tested as vaccine antigens in murine and/or porcine infection models and showed protection against *H.* (*G.*) *parasuis*. However, the remaining seven have not yet been tested and, consequently, they could become useful as putative antigens in the prevention of Glässer's disease. Anyway, the expression of this seven novel vaccine candidates should be shown in other serovars different from serovar 5.

Keywords: *Haemophilus* (*Glässerella*) *parasuis*, Glässer's disease, Vaccine antigens, RNA-sequencing, Transcriptome, Iron uptake

Background

Haemophilus (*Glässerella*) *parasuis* is a Gram-negative bacterium which forms part of the microbiota in the upper respiratory tract in pigs. Under certain conditions, such as stress or absence of prior contact, virulent strains can cause a systemic infection resulting in polyserositis, meningitis or arthritis (Glasser's disease) [1]. In addition, *H.* (*G.*) *parasuis* is involved in pneumonias as secondary agent within the porcine respiratory complex disease [2]. Each year *H.* (*G.*) *parasuis* causes significant loss to the swine industry worldwide [1].

Most vaccines used to prevent *H.* (*G.*) *parasuis* infection are bacterins although a minory of them are based on live vaccines. These traditional vaccines present several disadvantages, with the main one being the lack of cross-protection against different serotypes [3]. The use of these vaccines has been gradually replaced by subunit vaccines, whose study has been focused on outer membrane proteins (Omps) among other molecules. However, a huge variability has been still found among isolates from different countries, with substantial variations in MLST profiles, in such a manner that problems

* Correspondence: cbgutm@unileon.es
Microbiology & Immunology Section, Animal Health Department, Faculty of Veterinary Medicine, University of León, León, Spain

with cross-protection remain [4]. Recent advances in genomics, proteomics and transcriptomics have greatly enabled the search for Omps that are more likely to behave as good vaccine antigens [5].

Most bacteria remodel their coating structures inside the host since they need to adapt to new environments that could be potentially harmful to them, such as high temperature, osmolarity, pH or oxidative stress and these changes often involve the synthesis of surface structures that are important virulence factors [6]. It has been speculated that the change from a physiological temperature to a higher one (similar to that hyperthermia measured during Glässer's disease) in the host could also be used by some pathogens as a signal to enter into a persistence state in animals that leads to expression of mechanisms triggered during hyperthermia, used to avoid the host immune response. Some of them may correspond to changes in bacterial surface proteins [7].

Although iron is an essential element for organisms, being required for energy processes and DNA, protein or sugar metabolism; however, the concentration of free iron in the host is not enough to support the growth of bacteria [8]. For this reason, pathogenic bacteria have developed different mechanisms to scanvenge iron from host (siderophores, hemophores or host-molecule-binding proteins), which involve the expression of surface-exposed proteins [9]. In this respect, some reports have already concerned the expression of genes of *H.* (*G.*) *parasuis* to iron-restriction stress [10–12].

The aim of this work was to study the modifications which occur in the transcriptome of *H.* (*G.*) *parasuis* by RNA sequencing, when it is grown in vitro under culture conditions of iron-restriction and temperature stresses. These conditions were selected in order to partially mimic the host environment during natural infection. The transcriptome of bacteria grown under these conditions was compared with that of bacteria grown under optimal in vitro conditions (37 °C and non-iron-restriction stress) for detecting the overexpression of genes coding proteins exposed on the bacterial surface.

Results

Quality control of RNA samples

The RNA integrity from each sample was tested by automated electrophoresis in a Bioanalyzer Agilent 2100. The RNA integrity number (RIN) was not calculated because of the peculiar arrangement of the rRNA peaks from bacteria belonging to genus *Haemophilus* [22], in which the 23S subunit of the rRNA is fragmented in 1.2 and 1.7 kb portions. However, the graphs showed that the RNA present in each sample had A correct integrity (data not shown).

Upregulation under mimetic conditions (iron-restriction and 41 °C)

The number of genes upregulated under these conditions was 433, of which 154 had a $\log_2 > 10$. Among these 154, there were eight pseudogenes, two genes encoding tRNA and 144 genes encoding proteins (Fig. 1). **F1**
The amino acid sequence of the proteins encoded by the upregulated genes was obtained, and the cellular location of the proteins and their relation to pathogenesis was investigated. Four extracellular proteins (Eps), 17 Omps, 10 periplasmic proteins (Pps), 13 inner membrane proteins (Imps) and 100 cytoplasmic proteins (Cps) were found using CELLO v.2.5. As the main aim of this study was to search the proteins exposed to the cell surface, we verified individually the location of those proteins that CELLO assigned as belonging to the extracellular and Omp fractions and they were found different. Thus, two of the proteins firstly assigned to Omps were found to be Imps, while three others were Pps and two more were Cps. One of the Eps and another protein initially assigned to Omps appeared to be really the same extracellular protein but they were noted as two different proteins in the reference genome because of a point mutation involving the emergence of a stop codon. Therefore, after this correction, four proteins remained assigned to an extracellular localization (Eps), nine as Omps, 15 as Imps, 13 as Pps and 102 as Cps (Fig. 1).

A total of 31 proteins were recognized as being related to pathogenesis and therefore possibly involved in the pathogenesis of Glässer's disease (Fig. 1), and of them, four were Eps, nine were Omps, six were Imps, five were Pps and seven were Cps. All proteins predicted to be located on the bacterial surface were also predicted as being related to pathogenesis.

Table 1 shows the upregulated proteins that were predicted to be located on the bacterial surface (Omps or Eps) and/or related to pathogenesis. The proteins predicted to be located on the bacterial surface or related to pathogenes is that were not identified in GenBank and Uniprot databases were subjected to further studies of sequence homology and searched for the presence of domains of a known function. The findings are shown in Table 2. Additional file 1 summarizes the findings for genes that were upregulated under mimetic conditions with a $\log_2$ (fold change) > 10. **T1** **T2**

Downregulation under mimetic conditions (without iron-restriction and 37 °C)

The number of genes underexpressed under control conditions was 460, of which 187 were selected for having a $\log_2 > 10$. They included four pseudogenes, seven genes encoding tRNA, another gene encoding rRNA and 175 genes encoding proteins (Fig. 2). The CELLO v.2.5 program predicted six Eps, five Omps, 34 Imps, 30 Pps **F2**

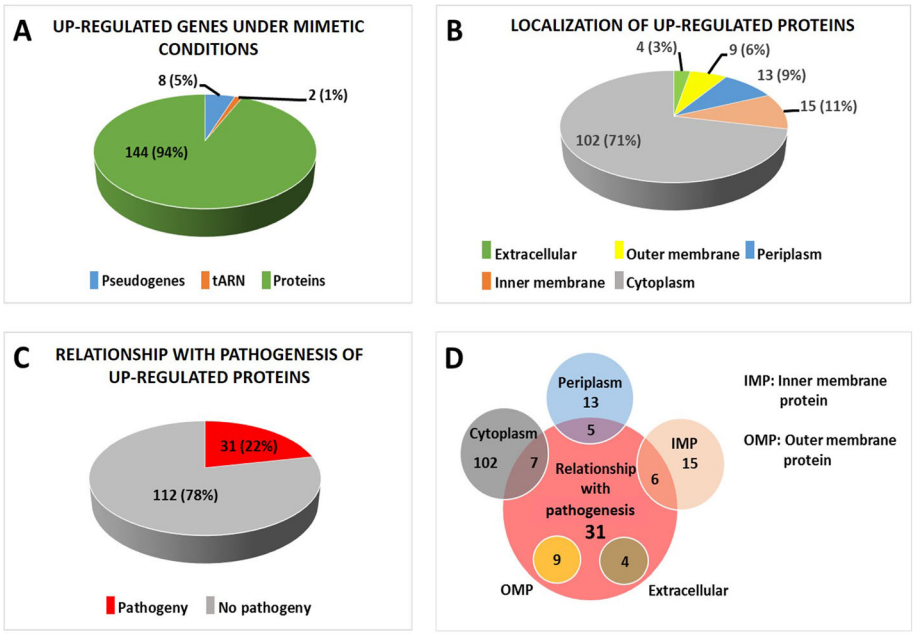


Fig. 1 Genes and proteins upregulated under mimetic conditions. **(a)**: Number of pseudogenes, tRNA and protein-coding genes indicating the percentage of the total number of upregulated genes with $\log_2$ (fold change) > 10 under mimetic conditions. **(b)**: Number of different locations of proteins upregulated under mimetic conditions and percentage of total proteins. Further corrections were taken into account. **(c)**: Number and percentage of proteins related to pathogenesis upregulated under mimetic conditions. Further corrections were considered. **(d)**: Venn diagram representing the relationship between cell localization and the pathogenesis of upregulated proteins. Further corrections were taken into account

and 100 Cps. Those assigned as Omps or Eps were individually rechecked and, after correction, four proteins remained assigned as Omps wihle the number of Cps rose to 101 (Fig. 2). The MP3 server found 34 proteins that were recognized as being related to pathogenesis (Fig. 2), of which five were Eps, three were Omps, 13 were Imps, eight were Pps and five were Cps. Additional file 2 summarizes findings for genes that were downregulated under mimetic conditions with a log2 > 10.

Gene ontology (GO) term enrichment analysis

Among the genes upregulated under mimetic conditions, only the GO term GO:0003676 (nucleic acid binding) was found to be enriched and associated with the presence in 37 upregulated genes (Table 3). With regard to genes downregulated under mimetic conditions, 31 GO terms were classified as being enriched (Table 4).

Discussion

Most bacteria reshape their coating structures inside the host since they need to adapt to a new potentially harmful environment [6]. Although the environment that a microorganism endures inside the host is much more complex than that replicated in the laboratory, the two selected conditions in this study (iron-restriction and temperature higher than 37 °C) attempted to partially simulate the infection in natural conditions. Some reports have been carried out on the changes occurring in the *H.* (*G.*) *parasuis* transcriptome when this bacterium was subjected to these two mimetic conditions of high temperature and iron scarcity [10, 12, 23]. However, the two circumstances in those studies were not tested together because these studies were focused in the understanding of both metabolism and virulence factors but not to search putative candidates that could be used as vaccine antigens.

The bacterial mechanisms used to remove iron from the host need surface-exposed proteins [9], and their expression is induced by a low iron concentration [12]. Among the genes upregulated under mimetic conditions, we detected six genes coding for Eps or Omps related to the obtaining of iron from the host (TbpA, TbpB, HxuA, HxuB, HxuC and FhuA). In a previous study of the transcriptome of *Actinobacillus pleuropneumonaie* exposed to the intraalveolar environment for two hours, upregulation of genes encoding membrane proteins involved in iron uptake were detected [24]. This finding reinforces our results concerning the high probability that the six above mentioned genes will be overexpressed during infection with *H. (G.) parasuis*.

TbpA and TbpB are porcine transferrin binding proteins that show different protection degrees against Gläsers disease [26, 27]. HxuA, HxuB and HxuC correspond to hemophore, transporter and receptor of the heme/hemopexin-binding protein (hxu) operon, respectively

t1.1 **Table 1** Upregulated proteins under mimetic conditions related to pathogenesis and located on the bacterial surface

t1.2	*Locus*[a]	GenBank product	Access number Uniprot	Name Uniprot	Patho-genesis	Location
t1.3	*HAPS_RS00370*	protein TolA	B8F375	|tolA|Cell envelope integrity inner membrane protein TolA	P	IM
t1.4	*HAPS_RS00485*	TonB-dependent receptor	–	–	P	OM
t1.5	*HAPS_RS00735*	hypothetical protein	–	–	P	OM
t1.6	*HAPS_RS00740*[a]	autotransporter domain-containing protein	–	–	P**	OM[a]
t1.7	*HAPS_RS00745*[a]	hypothetical protein	–	–	P**	EX**
t1.8	*HAPS_RS01255*	ABC transporter permease	B8F3P0	|HAPS_0253|ABC-type nitrate/sulfonate/bicarbonate transport permease	P	IM
t1.9	*HAPS_RS01260*	ABC transporter substrate-binding protein	B8F3P1	|HAPS_0254|ABC-type nitrate/sulfonate/bicarbonate transport systems periplasmic components protein	P	CP
t1.10	*HAPS_RS01265*	ABC transporter ATP-binding protein	B8F3P2	|HAPS_0255|ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase	P	IM
t1.11	*HAPS_RS01400*	sulfurtransferase FdhD	–	–	P	CP
t1.12	*HAPS_RS01435*	hypothetical protein	B8F3S4	|HAPS_0289|Uncharacterized protein	P	CP
t1.13	*HAPS_RS01805*	ABC transporter ATPase	B8F3Z5	|HAPS_0364|ATPase components of ABC transporters with duplicated ATPase domains-containing protein	P	PP
t1.14	*HAPS_RS01895*	hypothetical protein	B8F410	|HAPS_0382|Uncharacterized protein	P	EX
t1.15	*HAPS_RS02610*	hypothetical protein	B8F4D5	|purL|Phosphoribosylformylglycinamidine synthase	P	CP
t1.16	*HAPS_RS03735*	fimbrial usher protein	–	–	P	OM
t1.17	*HAPS_RS04480*	hypothetical protein	–	–	P	PP
t1.18	*HAPS_RS04485*	hypothetical protein	B8F5E9	|HAPS_0923|Uncharacterized protein	P	EX
t1.19	*HAPS_RS06520*	hypothetical protein	B8F6H3	|yaaH|Permease, Inner membrane protein yaaH	P	IM
t1.20	*HAPS_RS00370*	protein TolA	B8F375	|tolA|Cell envelope integrity inner membrane protein TolA	P	IM
t1.21	*HAPS_RS00485*	TonB-dependent receptor	–	–	P	OM
t1.22	*HAPS_RS00735*	hypothetical protein	–	–	P	OM
t1.23	*HAPS_RS00740*[a]	autotransporter domain-containing protein	–	–	P**	OM**
t1.24	*HAPS_RS00745*[a]	hypothetical protein	–	–	P**	EX**
t1.25	*HAPS_RS06610*	cell envelope protein TonB	B8F6J2	|tonB|Protein TonB	P	PP
t1.26	*HAPS_RS07630*	iron ABC transporter permease	B8F723	|hmuU|Hemin transport system permease protein HmuU	P	IM
t1.27	*HAPS_RS07950*	hypothetical protein	–	–	P	CP
t1.28	*HAPS_RS09000*	hypothetical protein	B8F7Q0	|HAPS_1850|Uncharacterized protein	P	IM
t1.29	*HAPS_RS10195*	DUF262 domain-containing protein	B8F899	|HAPS_2100|Uncharacterized protein	P	CP
t1.30	*HAPS_RS10530*	TolC family protein tolC	B8F8F2	|tolC|RND efflux system outer membrane lipoprotein/RND superfamily resistance-nodulation-cell division antiporter	P	OM
t1.31	*HAPS_RS10585*	ligand-gated channel	B8F8G2	|hxuC|Heme/hemopexin utilization protein C/outer membrane receptor protein, mostly Fe transport	P	OM
t1.32	*HAPS_RS10590*	ShlB/FhaC/HecB family hemolysin secretion/activation protein	B8F8G3	|hxuB|Heme/hemopexin-binding protein B, hemolysin activation/secretion protein	P	OM
t1.33	*HAPS_RS10595*	hypothetical protein	B8F8G4	|hxuA|Heme/hemopexin-binding protein A (Heme:hemopexin utilization protein A)	P	EX
t1.34	*HAPS_RS10780*	membrane protein	B8F8J9	|HAPS_2219|Possible outer membrane protein/FOG: TPR repeat protein	P	OM
t1.35	*HAPS_RS10800*	transferrin-binding protein-like solute binding protein	–	–	P	OM

Table 1 Upregulated proteins under mimetic conditions related to pathogenesis and located on the bacterial surface *(Continued)*

Locus[a]	GenBank product	Access number Uniprot	Name Uniprot	Patho-genesis	Location
HAPS_RS10805	lactoferrin/transferrin family TonB-dependent receptor	B8F8K4	|tbpA|Transferrin-binding protein 1	P	OM
HAPS_RS11010	transpeptidase	B8F8P5	|HAPS_2267|Uncharacterized protein	P	PP
HAPS_RS11030	transcriptional regulator	B8F8P9	|impA|SOS-response transcriptional repressor	P	CP

P indicates the relationship to pathogenesis; Locations are indicated by: *OM* outer membrane, *EX* extracellular, *IM* inner membrane, *PP* periplasmic, *CP* cytoplasmic. The *locus* in the reference genome (SH0165 strain) is indicated. [a] indicates that this was a single protein recognized as two different ones.

[28]. A protection of 87.5% for HxuC, 62.5% for HxuB and 37.5% for HxuA has been recently showed in mice against *H. (G.) parasuis* [29]. The upregulated gene *HAPS_RS00485*, coding for FhuA protein, a receptor for siderophores, has not been tested as vaccine antigen until date [30]. Curiously, Melnikow et al. [10] did not find *fhuA* among the genes upregulated under iron-restrictive conditions. One possible explanation for the difference between this study and ours could be that the combination of both conditions is required for upregulation of this gene.

Depending on the bacterial species, fever could cause a heat stress [31] that triggered changes involving coating structures [6]. It would be therefore expected that the reprogramming of the transcriptome undergone by the bacterium to resist heat stress could affect some genes encoding for surface-exposed proteins. The upregulated gene *tolC* found in this study encodes an Omp (TolC) lipoprotein that forms a trimeric channel and acts in the transport of several molecules [32]. Li et al. [33] observed that mice immunized with this protein and then challenged with *H.* (*G.*) *parasuis* showed a survival rate of 80%.

In addition to these well-characterized proteins in the databases, another five proteins were predicted to be located on the bacterial surface (Eps or Omps). The largest of them was an autotransporter with a serine protease domain which is annotated as two different genes (*HAPS_RS00740* and *HAPS_RS00745*) in the reference genome (SH0165 strain) because it presents a punctual mutation that triggers off the appearance of a stop codon. This long protein presents a broad sequence homology with the AasP autotransporter of *A. pleuropneumoniae*, which it is upregulated when this bacterium is grown under iron restriction stress [34]. On the other hand, a previous study performed with an *AasP* mutant *in A. pleuropneumoniae* showed that AasP protein is involved in adhesion under iron-restriction conditions [35].

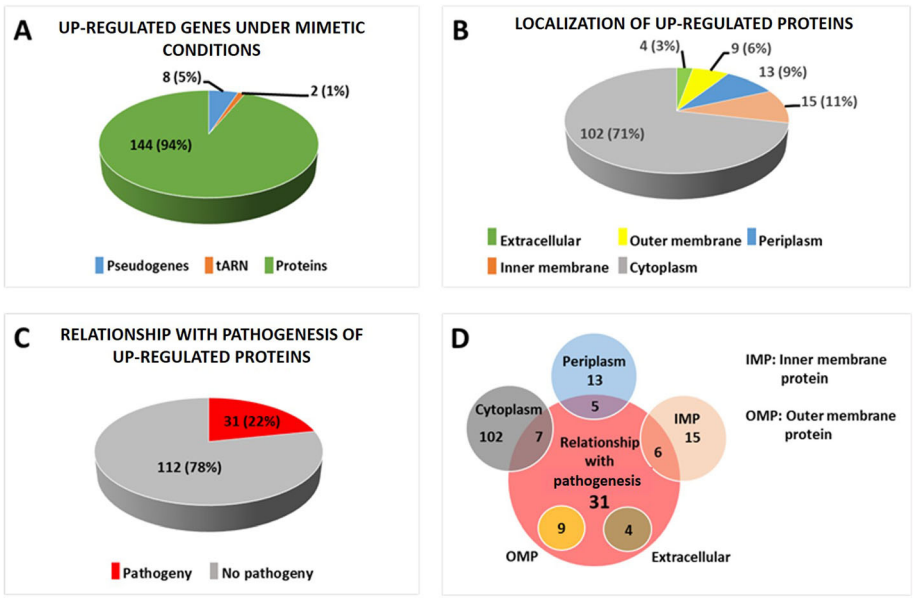
The upregulated gene *HAPS_RS04485* codes a protein that harbors domains with an immunoglobulin-like fold, which are usually present in proteins related to invasion or adhesion in prokaryotes [36]. Among the three remaining surface proteins whose genes were upregulated, two of them (*HAPS_RS00735* and *HAPS_RS01895*) neither presented homology to known proteins nor harbored domains with a known function, while in the remaining one (*HAPS_RS10780*) only a β-barrel and a single tetratricopeptide repeat could be detected.

With regard to the functional enrichment analysis under mimetic conditions to natural infection, only one significant enrichment was observed in the GO term 0003676, corresponding to nucleic acid binding, which was found in 37 upregulated genes. This could be consistent with the important reprogramming, at both transcriptional and translational levels, which the bacteria

Table 2 Findings in upregulated proteins under mimetic conditions related to pathogenesis but not characterized in the databases

Locus	Access number Uniprot	Location	Findings
HAPS_RS00735	–	OM	Nothing was recognized
HAPS_RS00740[a]	–	OM[a]	Half C-terminal Carrier [a]
HAPS_RS00745[a]	–	EX[a]	Half n-terminal auto transporter [a]
HAPS_RS01435	B8F3S4	CP	Nothing was recognized
HAPS_RS01895	B8F410	EX	Nothing was recognized
HAPS_RS04480	–	PP	Domain of the superfamily Glycoside-hydrolase
HAPS_RS04485	B8F5E9	EX	Immunoglobulin-like fold domain
HAPS_RS07950	–	CP	Domain of the Thioredoxin-like superfamily
HAPS_RS09000	B8F7Q0	IM	Homology with permease
HAPS_RS10195	B8F899	CP	ParB-like and HNH nuclease domains

The locus in the reference genome (SH0165 strain) is also indicated. [a] indicates that this was a single protein recognized as two different ones



f2.1 **Fig. 2** Genes and proteins downregulated under mimetic conditions. (**a**): Number of pseudogenes, tRNA and protein-coding genes indicating
f2.2 the percentage of the total number of genes down regulated genes with $\log_2$ (fold change) > 10 under mimetic conditions. **(b)**: Number of
f2.3 different locations of proteins downregulated under mimetic conditions and percentage of total proteins. Further corrections were taken into
f2.4 account. **(c)**: Number and percentage of proteins related to pathogenesis downregulated under mimetic conditions. Further corrections were
f2.5 considered. **(d)**: Venn diagram representing the relationship between cell localization and the pathogenesis of downregulated proteins. Further
f2.6 corrections were taken into account
f2.7

t3.1 **Table 3** List of upregulated genes where GO term GO:0003676 (nucleic acid binding) was found

t3.2 *Locus*	GenBank product	*Locus*	GenBank product
t3.3 *HAPS_RS06635*	tRNA-dihydrouridine synthase	*HAPS_RS06640*	Fis family transcriptional regulator
t3.4 *HAPS_RS07775*	30S ribosomal protein S10	*HAPS_RS07795*	50S ribosomal protein L2
t3.5 *HAPS_RS01040*	LysR family transcriptional regulator	*HAPS_RS07785*	50S ribosomal protein L4
t3.6 *HAPS_RS03460*	tRNA pseudouridine (65) synthase TruC	*HAPS_RS01015*	translation initiation factor IF-2
t3.7 *HAPS_RS07435*	MurR/RpiR family transcriptional regulator	*HAPS_RS07790*	50S ribosomal protein L23
t3.8 *HAPS_RS02800*	transcriptional regulator	*HAPS_RS10775*	ATPase AAA
t3.9 *HAPS_RS02140*	transcriptional regulator	*HAPS_RS06525*	Holliday junction DNA helicase RuvA
t3.10 *HAPS_RS09580*	23S rRNA (guanosine-2′-O-)-methyltransferase	*HAPS_RS07720*	RNA polymerase, beta subunit
t3.11 *HAPS_RS08560*	restriction endonuclease	*HAPS_RS01010*	transcription termination factor NusA
t3.12 *HAPS_RS09560*	30S ribosomal protein S7	*HAPS_RS07780*	50S ribosomal protein L3
t3.13 *HAPS_RS09565*	elongation factor G	*HAPS_RS01395*	elongation factor Ts
t3.14 *HAPS_RS08275*	endonuclease	*HAPS_RS02855*	single-stranded DNA-binding protein
t3.15 *HAPS_RS10985*	Fur family transcriptional regulator	*HAPS_RS00670*	transposase
t3.16 *HAPS_RS02340*	RNA helicase	*HAPS_RS10560*	transcriptional regulator
t3.17 *HAPS_RS04055*	DNA polymerase III subunit epsilon	*HAPS_RS03180*	50S ribosomal protein L25
t3.18 *HAPS_RS01380*	lysine tRNA synthetase	*HAPS_RS04045*	heat-shock protein
t3.19 *HAPS_RS08565*	restriction endonuclease	*HAPS_RS01070*	tRNA s (4) U8 sulfurtransferase
t3.20 *HAPS_RS08440*	transcriptional regulator	*HAPS_RS04180*	formamidopyrimidine-DNA glycosylase
t3.21 *HAPS_RS05455*	DNA-binding protein		

t3.22 The locus in the reference genome (SH0165 strain) and GenBank product are indicated

t4.1 **Table 4** GO terms found to be enriched among downregulated genes under mimetic conditions

t4.2 GO term	GO domain	Number of genes
t4.3 GO:0008643 *carbohydrate transport*	BP	16
t4.4 GO:0015980 *energy derivation by oxidation of organic compounds*	BP	17
t4.5 GO:0055114 *oxidation-reduction process*	BP	21
t4.6 GO:0009401 *phosphoenolpyruvate-dependent sugar phosphotransferase system*	BP	13
t4.7 GO:0009401 *phosphoenolpyruvate-dependent sugar phosphotransferase system*	BP	13
t4.8 GO:0006091 *generation of precursor metabolites and energy*	BP	19
t4.9 GO:0045333 *cellular respiration*	BP	14
t4.10 GO:0006099 *tricarboxylic acid cycle*	BP	10
t4.11 GO:0006099 *tricarboxylic acid cycle*	BP	10
t4.12 GO:0009060 *aerobic respiration*	BP	10
t4.13 GO:0044712 *single-organism catabolic process*	BP	17
t4.14 GO:0044699 *single-organism process*	BP	73
t4.15 GO:0005975 *carbohydrate metabolic process*	BP	25
t4.16 GO:0044724 *single-organism carbohydrate catabolic process*	BP	11
t4.17 GO:0044282 *small molecule catabolic process*	BP	12
t4.18 GO:0016052 *carbohydrate catabolic process*	BP	11
t4.19 GO:1901575 *organic substance catabolic process*	BP	18
t4.20 GO:1901476 *carbohydrate transporter activity*	MF	8
t4.21 GO:0015144 *carbohydrate transmembrane transporter activity*	MF	8
t4.22 GO:0044765 *single-organism transport*	BP	23
t4.23 GO:0009056 *catabolic process*	BP	18
t4.24 GO:0005996 *monosaccharide metabolic process*	BP	10
t4.25 GO:1902578 *single-organism localization*	BP	23
t4.26 GO:0071702 *organic substance transport*	BP	17
t4.27 GO:0044723 *single-organism carbohydrate metabolic process*	BP	17
t4.28 GO:0006810 *transport*	BP	26
t4.29 GO:0019318 *hexose metabolic process*	BP	8
t4.30 GO:0051179 *localization*	BP	26
t4.31 GO:0051234 *establishment of localization*	BP	26
t4.32 GO:0044710 *single-organism metabolic process*	BP	49
t4.33 GO:0046365 *monosaccharide catabolic process*	BP	5

t4.34 The number of genes where each GO term was found is indicated. The gene ontology domain (GO domain) to which each GO term belongs is also indicated, *BP*
t4.35 biological process, *MF* molecular function

undergo when exposed to environmental stress [37]. Among these genes, six coded ribosomal subunit binding proteins, four coded rRNA-related enzymes (translational level) and eight coded transcriptional regulators (transcriptional level). Within the upregulated transcriptional regulators, the iron uptake regulatory protein (Fur) must be highlighted. This protein forms a dimer in the presence of Fe^{2+} that represses the expression of genes related to iron uptake. However, when available iron is scarce, this dimer breaks down and transcription of genes related to iron uptake is allowed [38]. The upregulation of a negative regulator of iron acquisition under iron-restricted conditions may be paradoxical effect; however, different findings have been shown over the last years that the role of Fur is not as simple as it had been previously stated although it remains fully valid. There are positively regulated genes for Fe-induced Fur dimers and other positively or negatively regulated genes for Fur when this protein does not form a complex with Fe^{2+}. Some of the genes that are regulated in these ways are related to the response to stressful conditions or to virulence [38]. Therefore, it was not unusual to find *Fur* gene among those upregulated under our stress conditions.

Concerning functional enrichment analysis of downregulated genes under mimetic conditions, the most

enriched terms were related to energy metabolism, redox reactions, or to both of them. Relative to energy metabolism, this finding would make sense with slowing of growth braking expected in a stressful environment, which should result in a general decrease in the metabolism. In the case of redox processes, it might simply be a reflection of the decrease in bacterial growth and/or it could be that the microorganism, when detecting iron deficiency, prefers not to waste energy in the production of compounds that cannot perform their function since the iron is the main limiting factor [9].

As previously mentioned, microorganisms have to face a number of stressors during infection, such as unfavorable temperatures, pH changes or free radicals, which trigger the expression of several proteins known as cellular stress proteins [39]. These proteins behave as chaperones promoting the assembly of other macromolecules and are often called heat shock proteins (Hsp) since it was previously believed that they acted alone against heat stress [40]. Several genes encoding different Hsp were upregulated in our study, such as DnaJ, DnaK, HslO and HslR. Several Hsp behaving as surface antigens were observed, which might seem a contradiction considering that the role played as chaperones should imply an intracellular location. However, many of these proteins, in addition to act as chaperones, also present functions related to pathogenesis that may imply a superficial location [40]. Accordingly, it would be reasonable to study these proteins as vaccine antigens; however, it would not be appropriate to do because Hsp have been highly conserved throughout evolution and there is a high homology between the Hsp from bacterial and those from mammalian origin. For this reason, the use of Hsp as vaccine antigens could trigger autoimmune diseases [39].

Conclusion

Slowing of growth expected in stressful conditions have given rise to the upregulation of 13 *H.* (*G.*) *parasuis* genes coding for proteins located on the bacterial surface. Among them, seven proteins untested to date were detected as vaccine antigens: FhuA (encoded by *HAPS_RS00485*), a fimbrial usher protein (encoded by *HAPS_RS03735*), a long autotransporter (encoded by *HAPS_RS00740* and *HAPS_RS00745*), a protein containing domains with an Ig-like fold (encoded by *HAPS_RS04485*) and other three surface proteins without known function (encoded by *HAPS_RS10780*, *HAPS_RS00735* and *HAPS_RS01895*). These seven novel vaccine candidates could provide protection against Glässer disease, but their effectiveness have to be tried in future studies. Anyway, their expression must be sustained in other serovars different from serovar 5.

Methods

Bacterial strain and growth conditions

H. (*G.*) *parasuis* was grown under (i) in vitro optimal culture conditions (control conditions) and (ii) under in vitro growth conditions partially mimicking the host environment encountered during infection (iron-restriction and temperature stress by raising incubation temperature above 37 °C). The transcriptomes from both growths were compared by RNA sequencing to detect overexpressed genes coding proteins exposed on the bacterial surface.

For control culture conditions (without iron-restriction and 37 °C), *H.* (*G.*) *parasuis* Nagasaki strain (reference strain of serovar 5 kindly supplied by Kielstein P., Federal Institute for Health Protection of Consumers and Veterinary Medicine, Jena, Germany) was inoculated into 30 ml of PPLO broth (Conda Laboratories, Spain) with 150 μM nicotinamide adenine dinucleotide (NAD, Sigma-Aldrich, Spain) and 0.075% glucose (Sigma-Aldrich), and it was cultured at 37 °C until reaching an optical density of 0.5 at 600 nm (OD_{600}). Three replicates were made under these conditions. For mimetic conditions (iron-restriction and 41 °C), *H.* (*G.*) *parasuis* Nagasaki strain was inoculated into 30 ml of the same broth and was cultured at 41 °C until reaching an OD_{600} of 0.3. Then, iron was restricted by adding 200 μM 2 2′-dipyridyl, and the culture was grown again at 41 °C to an OD_{600} of 0.5. Three replicates were also made.

RNA extraction and sample preparation

When the appropriate OD_{600} was reached for each replicate, the culture was centrifuged at 7000×*g* and 4 °C for 7 min. The supernatants were removed and the pellets were preserved on ice. RNA extraction was performed using the High Pure RNA Isolation kit (Sigma-Aldrich) following the manufacturer's specifications. The DNA-free kit (Thermo-Fisher) was used to remove the contaminating DNA, which was verified by species-specific PCR [13], testing the absence of amplification in samples treated with DNAse. Finally, RNA concentrations were measured in a NanoDrop 1000 (Thermo-Fisher), and samples were stored at − 80 °C. The RNA integrity checking was tested using a Bioanalyzer Agilent 2100 (Agilent Technologies, Spain) from the Laboratory of Instrumental Techniques (University of León, Spain).

Library preparation and Illumina sequencing

Ribosomal RNA (rRNA) was removed from the samples using a Ribo-Zero Magnetic Kit Bacteria (Illuminia, Portugal). Libraries were then prepared following the instructions of the NEBNext Ultra Directional RNA Library Prep kit for Illumina (New England Biolabs, USA). The input of ribosome-depleted RNA to start the protocol was 10 ng quantified by an Agilent 2100 Bioanalyzer

using a RNA 6000 nano LabChip kit (Agilent Technologies, Germany). The fragmentation time used was 15 min. The cDNA libraries obtained were validated and quantified by an Agilent 2100 Bioanalyzer using a DNA7500 LabChip kit (Agilent Technologies). An equimolecular pool of libraries were titrated by quantitative PCR using the Kapa-SYBR FAST qPCR kit for LightCycler480 (Kapa BioSystems, USA), and a reference standard for quantification. The pool of libraries was denatured prior to be seeded on a flowcell at a density of 2 2 pM, where clusters were formed and sequenced with a depth of 10 M using a NextSeq 500 High Output Kit (Illumin) in a 1 × 75 single-read sequencing run on a NextSeq 500 sequencer (Illumin).

Bioinformatic analysis of differential gene expression

Bioinformatic analysis for the RNAseq data was performed by Era7 bioinformatics (Spain) following the protocol described by Trapnell et al. [14]. Quality control of raw readings was performed with the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Alignment to the reference genome was performed using TopHat2 software, but since the genome of the Nagasaki strain has not been completely sequenced, the SH0165 strain genome (GenBank accession nr NC_011852.1), also belonging to serovar 5, was used as the reference genome. Transcripts assembly was performed with the Cufflinks tool and transcripts merge with the Cuffmerge tool. Analysis of differences in gene expression was performed with Cuffdiff. Statistical significance was considered for $p < 0.05$ (corrected by the multiple testing Benjamini-Hochberg method). Only genes with a $\log_2$ (fold change) being > 10 were considered for further analyses.

Studies of differentially expressed genes

Differentially expressed genes were subjected to search in the GenBank database to discard pseudogenes and tRNA genes. GenBank [15] and Uniprot [16] annotation was obtained to genes and proteins differentially expressed. The sequences of the proteins encoded by differentially expressed genes were obtained using QuickGo [17]. Prediction of the protein cellular location was carried out using the CELLO v 2.5 web tool (http://cello.life.nctu.edu.tw/) [18]. Proteins predicted to be related to pathogenesis were searched for using the MP3 web tool (http://metagenomics.iiserb.ac.in/mp3/index.php) [19]. Proteins that were not identified in the GenBank or Uniprot databases but were considered of interest (predicted to be found on the bacterial surface or related to pathogenesis) were subjected to a study using BLASTp in order to find orthologous proteins in related species and using GenBank or InterproScan databases [20] to find domains of a known function. Databases and BLASTp were also used to verify the protein location that CELLO assigned as belonging to the extracellular and Omp fractions.

Genes upregulated under culture mimetic conditions are referred to as "upregulated under mimetic conditions", while those genes upregulated under control conditions are referred to as "downregulated under mimetic conditions". Only genes with $\log_2$ (fold change) was > 10 blank were considered in the analysis of data.

Gene ontology (GO) term enrichment analysis

Functional enrichment analysis in terms of gene ontology (GO) of differentially expressed genes was conducted using the DAVID v 6.7 web server (https://david.ncifcrf.gov/home.jsp) [21]. Only the GO terms containing a minimum of five differentially expressed genes were considered and the GO terms that showed a $p < 0.05$ (corrected by the multiple testing Benjamini-Hochberg method) were considered significantly enriched.

Additional files

Additional file 1: Summary of genes that were upregulated under mimetic conditions with a $\log_2$ (fold change) > 10. Indicated findings in GenBank and Uniprot databases of genes that were upregulated under mimetic conditions with a $\log_2$ (fold change) > 10. P indicates that the encoded protein is related to pathogenesis, and NP indicates that there is no relationship with pathogenesis. The location of the protein is indicated with EX (extracellular), OM (outer membrane), PP (periplasmic), IM (inner membrane) or CP (cytoplasmic). * indicates that they are the same protein (PDF 117 kb).

Additional file 2: Summary of genes that were downregulated under mimetic conditions with a $\log_2$ (fold change) > 10. Indicated findings in GenBank and Uniprot databases of genes that were downregulated under mimetic conditions with a $\log_2$ (fold change) > 10. P indicates that the encoded protein is related to pathogenesis, and NP indicates that there is no relationship with pathogenesis. The location of the protein is indicated with EX (extracellular), OM (outer membrane), PP (periplasmic), IM (inner membrane) or CP (cytoplasmic). * indicates that they are the same protein (PDF 127 kb).

Abbreviations

Cps: Cytoplasmic proteins; Eps: Extracellular proteins; Fur: Ferric uptake regulatory protein; GO: Gene ontology; Hsp: Heat shock proteins; Imps: Inner membrane proteins; Omps: Outer membrane proteins; Pps: Periplasmic proteins; RNAseq: RNA-sequencing

Funding

This work was supported by project AGL2011–23195 from the Spanish Ministry of Economy and Competitivity. AAE was a recipient of long-predoctoral fellowships from this Ministry financed by the project AGL2011–23195.

Availability of data and materials

All data generated or analyzed during the current study are included in this published article and supplementary information files.
The datasets supporting the conclusions of this article are included within the article and additional files.

Authors' contributions

AAE and SMM performed the bacterial cultures and the RNA sampling; AAE performed the study of the differentially expressed genes; AAE wrote the paper; SMM, CBGM and EFRF contributed to critical review of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable: *H. parasuis* Nagasaki strain did not require any administrative or ethical permissions to use it.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Received: 10 July 2018 Accepted: 12 October 2018

References

1. Nedbalcova K, Satran P, Jaglic R, Sondirasova R, Kucerova Z. *Haemophilus parasuis* and Glässer's disease in pigs: a review. Veterinar Med. 2006;51:168–79.
2. Li J, Wang S, Li C, Wang C, Liu Y, Wang G, He X, Hu L, Liu Y, Cui M, Bi C, Shao Z, Wang X, Xiong T, Cai X, Huang L, Weng C. Secondary *Haemophilus parasuis* infection enhances highly pathogenic porcine reproductive and respiratory syndrome virus (HP-PRRSV) infection-mediated inflammatory responses. Vet Microbiol. 2017;204:35–42. https://doi.org/10.1016/j.vetmic.2017.03.035.
3. Liu H, Xue Q, Zeng Q, Zhao Z. *Haemophilus parasuis* vaccines. Vet Immunol Immunopathol. 2016;180:53–8. https://doi.org/10.1016/j.vetimm.2016.09.002.
4. Ruiz A, Oliveira S, Torremorell M, Pijoan C. Outer membrane proteins and DNA profiles in strains of *Haemophilus parasuis* recovered from systemic and respiratory sites. J Clin Microbiol. 2001;39:1757–62.
5. Sette A, Rappuoli R. Reverse vaccinology: developing vaccines in the era of genomics. Immunity. 2010;33:530–41. https://doi.org/10.1016/j.immuni.2010.09.017.
6. Darwin AJ. Stress relief during host infection: the phage shock protein response supports bacterial virulence in various ways. PLoS Pathog. 2013;9:e1003388. https://doi.org/10.1371/journal.ppat.1003388.
7. González Plaza JJ, Hulak N, Zhumadilov Z, Akilzhanova A. Fever as an important resource for infectious diseases research. Intractable Rare Dis Res. 2016;5:97–102. https://doi.org/10.5582/irdr.2016.01009.
8. del Río ML, Gutiérrez Martín CB, Rodríguez Barbosa JI, Navas J, Rodríguez Ferri EF. Identification and characterization of the TonB region and its role in transferrin-mediated iron acquisition in *Haemophilus parasuis*. FEM Immunol Med Immunopathol. 2005;45:75–86.
9. Parrow NL, Fleming RE, Minnick MF. Sequestration and scavenging of iron in infection. Infect Immun. 2013;81:3503–14. https://doi.org/10.1128/IAI.00602-13.
10. Melnikow E, Dornan S, Sargent C, Duszenko M, Evans G, Gunkel N, Selzer PM, Ullrich HJ. Microarray analysis of *Haemophilus parasuis* gene expression under *in vitro* conditinos mimicking the *in vivo* environment. Vet Microbiol. 2005;110:255–63.
11. Metcalf DS, MacInnes JI. Differential expression of *Haemophilus parasuis* genes in response to iron restriction and cerebrospinal fluid. Can J Vet Res. 2007;71:181–8.
12. Xie Q, Jin H, Luo R, Wan Y, Chu J, Zhou H, Shi B, Chen H, Zhou R. Transcriptional responses of *Haemophilus parasuis* to iron-restriction stress *in vitro*. Biometals. 2009;22:907–16.
13. Oliveira S. Diagnostic notes *Haemophilus parasuis*. J Swine Health Prod J Swine Health Prod. 2007;15:99–103 http://www.aasv.org/shap.html. Accessed 21 Nov 2017.
14. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2012;31:46–53. https://doi.org/10.1038/nbt.2450.
15. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res. 2005;33(Database issue):D34–8. https://doi.org/10.1093/nar/gki063.
16. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45:D158–69. https://doi.org/10.1093/nar/gkw1099.
17. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for gene ontology searching. Bioinformatics. 2009;25:3045–6. https://doi.org/10.1093/bioinformatics/btp536.
18. Lee S, Lee B, Kim D. Prediction of protein secondary structure content using amino acid composition and evolutionary information. Prot Struct Funct Bioinform. 2005;62:1107–14. https://doi.org/10.1002/prot.20821.
19. Gupta A, Kapil R, Dhakan DB, Sharma VK. MP3: a software tool for the prediction of pathogenic proteins in genomic and metagenomic data. PLoS One. 2014;9:e93907. https://doi.org/10.1371/journal.pone.0093907.
20. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 2017;45:D190–9. https://doi.org/10.1093/nar/gkw1107.
21. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57. https://doi.org/10.1038/nprot.2008.211.
22. Song XM, Forsgren A, Janson H. Fragmentation heterogeneity of 23S ribosomal RNA in *Haemophilus* species. Gene. 1999;230:287–93 http://www.ncbi.nlm.nih.gov/pubmed/10216268. Accessed 21 Nov 2017.
23. Hill CE, Metcalf DS, JI MI. A search for virulence genes of Haemophilus parasuis using differential display RT-PCR. Vet Microbiol. 2003;96:189–202 http://www.ncbi.nlm.nih.gov/pubmed/14519336. Accessed 21 Nov 2017.
24. Goethe R, Gonzáles OF, Lindner T, Gerlach GF. A novel strategy for protective Actinobacillus pleuropneumoniae subunit vaccines: detergent extraction of cultures induced by iron restriction. Vaccine. 2000;19:966–75 http://www.ncbi.nlm.nih.gov/pubmed/11115723. Accessed 21 Nov 2017.
25. Bello-Ortí B, Howell KJ, Tucker AW, Maskell DJ, Aragon V. Metatranscriptomics reveals metabolic adaptation and induction of virulence factors by *Haemophilus parasuis* during lung infection. Vet Res. 2015;46:102.
26. Frandoloso R, Martínez S, Rodríguez-Ferri EF, García-Iglesias MJ, Pérez-Martínez C, Martínez-Fernández B, Gutiérrez Martín CB. Development and characterization of protective *Haemophilus parasuis* subunit vaccines based on native proteins with affinity to porcine transferrin and comparison with other subunit and commercial vaccines. Clin Vaccine Immunol. 2011;18:50–8. https://doi.org/10.1128/CVI.00314-10.
27. Frandoloso R, Martínez-Martínez S, Calmettes C, Fegan J, Costa E, Curran D, Yu RH, Gutiérrez-Martín CB, Rodríguez-Ferri EF, Moraes TF, Schryvers AB. Nonbinding site-directed mutants of transferrin binding protein B exhibit enhanced immunogenicity and protective capabilities. Infect Immun. 2015;83:1030–8. https://doi.org/10.1128/IAI.02572-14.
28. Morton DJ, Seale TW, Madore LL, VanWagoner TM, Whitby PW, Stull TL. The haem-haemopexin utilization gene cluster (hxuCBA) as a virulence factor of *Haemophilus influenzae*. Microbiology. 2007;153(Pt 1):215–24. https://doi.org/10.1099/mic.0.2006/000190-0.
29. Wen Y, Yan X, Wen Y, Cao S, He L, Ding L, Zhang L, Zhou P, Huang X, Wu R, Wen X. Immunogenicity of the recombinant HxuCBA proteins encoded by *hxuCBA* gene cluster of *Haemophilus parasuis* in mice. Gene. 2016;591:478–83. https://doi.org/10.1016/j.gene.2016.07.001.
30. Mikael LG, Pawelek PD, Labrie J, Sirois M, Coulton JW, Jacques M. Molecular cloning and characterization of the ferric hydroxamate uptake (fhu) operon in *Actinobacillus pleuropneumoniae*. Microbiology. 2002;148(Pt 9):2869–82. https://doi.org/10.1099/00221287-148-9-2869.
31. Hasday JD, Singh IS. Fever and the heat shock response: distinct, partially overlapping processes. Cell Stress Chaperones. 2000;5:471–80 http://www.ncbi.nlm.nih.gov/ pubmed/11189454. Accessed 21 Nov 2017.
32. Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. Nature. 2000;405:914–9. https://doi.org/10.1038/35016007.
33. Li M, Cai R-J, Song S, Jiang Z-Y, Li Y, Gou H-C, Chu PP, Li CL, Qiu HJ. Evaluation of immunogenicity and protective efficacy of recombinant outer membrane proteins of *Haemophilus parasuis* serovar 5 in a murine model. PLoS One. 2017;12:e0176537. https://doi.org/10.1371/journal.pone.0176537.
34. Deslandes V, Nash JHE, Harel J, Coulton JW, Jacques M. Transcriptional profiling of *Actinobacillus pleuropneumoniae* under iron-restricted conditions. BMC Genomics. 2007;8:72. https://doi.org/10.1186/1471-2164-8-72.
35. Tegetmeyer HE, Fricke K, Baltes N. An isogenic *Actinobacillus pleuropneumoniae* AasP mutant exhibits altered biofilm formation but retains virulence. Vet Microbiol. 2009;137:392–6. https://doi.org/10.1016/j.vetmic.2009.01.026.
36. Bodelón G, Palomino C, Fernández LÁ. Immunoglobulin domains in *Escherichia coli* and other enterobacteria: from pathogenesis to applications in antibody technologies. FEMS Microbiol Rev. 2013;37:204–50. https://doi.org/10.1111/j.1574-6976.2012.00347.x.

636 37. Starosta AL, Lassak J, Jung K, Wilson DN. The bacterial translation stress
637 response. FEMS Microbiol Rev. 2014;38:1172–201. https://doi.org/10.1111/
638 1574-6976.12083.
639 38. Carpenter BM, Whitmire JM, Merrell DS. This is not your mother's repressor:
640 the complex role of fur in pathogenesis. Infect Immun. 2009;77:2590–601.
641 https://doi.org/10.1128/IAI.00116-09.
642 39. Ghazaei C. Role and mechanism of the Hsp70 molecular chaperone
643 machines in bacterial pathogens. J Med Microbiol. 2017;66:259–65. https://
644 doi.org/10.1099/jmm.0.000429.
645 40. Henderson B, Allan E, Coates ARM. Stress wars: the direct role of host and
646 bacterial molecular chaperones in bacterial infection. Infect Immun. 2006;74:
647 3693–706. https://doi.org/10.1128/IAI.01882-05.

Author Query Form

Journal: BMC Veterinary Research

Title: Transcriptomics of Haemophilus (Glässerella) parasuis serovar 5 subjected to culture conditions partially mimetic to natural infection for the search of new vaccine antigens

Q1 **Authors: Álvaro Álvarez-Estrada, César B. Gutiérrez-Martín, Elías F. Rodríguez-Ferri, Sonia Martínez-Martínez**

Article: 1647

Dear Authors,

During production of your paper, the following queries arose. Please respond to these by annotating your proofs with the necessary changes/additions. If you intend to annotate your proof electronically, please refer to the E-annotation guidelines. We recommend that you provide additional clarification of answers to queries by entering your answers on the query sheet, in addition to the text mark-up.

Query No.	Query	Remark
Q1	Author names: Please confirm that the author names are presented accurately and in the correct sequence (given names/initials, family name). Author 1: Given name: Álvaro Family name: Álvarez-Estrada Author 2: Given name: César Given name: B. Family name: Gutiérrez-Martín Author 3: Given name: Elías Given name: F. Family name: Rodríguez-Ferri Author 4: Given name: Sonia Family name: Martínez-Martínez	
Q2	Affiliations: As per standard instruction, “city” is required for affiliations; however, this information is missing in affiliation “1”. Please check if the provided city is correct and amend if necessary.	
Q3	Footnote “**” is indicated in the text, however, no corresponding footnote explanation/detail was provided. Please provide its corresponding information.	
Q4	Please confirm if figures 1 and 2 are captured correctly.	
Q5	URL: Please check that the following URLs are working. If not, please provide alternatives: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/	
Q6	URL: Please check that the following URLs are working. If not, please provide alternatives: http://cello.life.nctu.edu.tw/	
Q7		

Query No.	Query	Remark
	URL: Please check that the following URLs are working. If not, please provide alternatives: http://metagenomics.iiserb.ac.in/mp3/index.php	
Q8	URL: Please check that the following URLs are working. If not, please provide alternatives: https://david.ncifcrf.gov/home.jsp	
Q9	Query: Authors' contributions: As per standard instruction, the statement "All authors read and approved the final manuscript." is required in the "Authors' contributions" section. Please note that this was inserted at the end of the paragraph of the said section. Please check if appropriate.	
Q10	Reference [25] was provided in the reference list; however, this was not mentioned or cited in the manuscript. As a rule, all references given in the list of references should be cited in the main body. Please provide its citation in the body text.	
Q11	URL: Please check that the following URLs are working. If not, please provide alternatives: http://www.aasv.org/shap.html	
Q12	URL: Please check that the following URLs are working. If not, please provide alternatives: http://www.ncbi.nlm.nih.gov/pubmed/14519336.%20Accessed%2021%20Nov%202017	
Q13	URL: Please check that the following URLs are working. If not, please provide alternatives: http://www.ncbi.nlm.nih.gov/pubmed/11115723	
Q14	URL: Please check that the following URLs are working. If not, please provide alternatives: http://www.ncbi.nlm.nih.gov/	