1. Name of applicant	Valentina Stefanetti
2. Contact information	Address: 38, Assisana Street, Perugia Telephone: +39 3393720626 Fax: +39 0755857662 Email: valentina.stefanetti87@gmail.com vale_stefa87@hotmail.it
3. Academic degrees	DVM, PhD
4. Current affiliation	Institute: Microbiology and Infectious Diseases Department: Department of Veterinary Medicine, University of Perugia, Italy
5. Position held	Post doc
6. Name of the host institute	Animal Health Trust, Lanwades Park, Kentford, Newmarket, UK
7. Name and contact information of the scientist from the host institute	Dr. Andrew Waller (andrew.waller@aht.org.uk)
8. Duration and dates of the visit	3 months (from September 2nd until December 1st, 2016)

Detailed report of the activities during the visit

1) Background

Streptococcus equi subspecies *equi* (*S. equi*) is the causative agent of strangles, one of the most frequently diagnosed and feared infectious disease of the horses worldwide. The duration of clinical disease can be several months and so strangles is of substantial economic importance to horse owners. However, despite its economic impact, the only available vaccine causes adverse reactions and is rarely used.

Infection occurs by inhalation or ingestion of the germ, followed by attachment to the cells in the crypt of the tonsil and adjacent lymphoid nodules. The bacteria release enzymes and toxins that damage surrounding cells and initiate inflammation causing fever and rhinitis. The bacteria then translocate within a few hours to the submandibular and retropharyngeal lymph nodes. Although a significant number of neutrophils are recruited into the lymph nodes, they are, however, unable to effectively phagocytose and kill the *S. equi* organisms and this results in an accumulation of extracellular organisms, degenerating neutrophils and necrotic tissue (abscess) in the lymph nodes. Abscesses formed in the retropharyngeal lymph nodes usually rupture into the guttural pouches, leading to the classic mucopurulent nasal discharges observed during strangles. However, incomplete drainage of the abscess material leads to a proportion of horses becoming persistently infected. Residual pus becomes inspissated to form chondroids, which contain live *S. equi* that can remain in the guttural pouch for several years. Development of persistent carriers after an outbreak is now considered an important factor in disease maintenance within a group of horses and spread of the disease to immunologically naïve horses. Moreover, it should be underlined that the selective requirements of lymph node abscess, versus persistence in the guttural pouches, have likely influenced the events of gene loss due to nonsense mutations and deletions and gene gain through the acquisition of mobile genetic elements that have shaped the *S. equi* genome and led to host restriction.

Interestingly, the researchers at the Animal Health Trust are currently using a new generation sequencing technique known as transposon directed insertion-site sequencing (TraDIS) that enables the identification of specific sites at which a transposable element, IS*S1*, inserts within the *S. equi* genome. IS*S1* is encoded on a plasmid named pG^{+}host9, which is transformed into *S. equi* and allowed to randomly insert into the *S. equi* genome. Each insertion of ISS*1* into the *S. equi* genome can potentially disrupt the function of the gene into which it inserts, generating a pool of random IS*S1* insertion mutants (input pool). The intramuscular injection of an input pool of vaccine strain mutants into four Welsh mountain ponies enabled the recovery of surviving mutants from the

injection site that were sequenced by TraDIS to determine if certain mutants dominated the output pool. In particular, approximately 43% of surviving mutants contained the ISS*1* transposon within the putative *S. equi* transcriptional regulators encoded by *fas* (fibronectin/fibrinogen binding/haemolytic activity/streptokinase regulator) and *mga* (multiple virulence gene regulator of Group A Streptococcus (GAS)). The Fas system is encoded by a four-gene locus that encodes three putative proteins (FasBCA) and a sRNA (FasX). Inactivation of *fas* in GAS led to altered production of streptokinase, streptolysin S and fibronectin/fibrinogen-binding proteins. Mga is a global transcriptional regulator that controls the expression of a wide range of genes involved in host cell attachment, immune evasion and carbohydrate metabolism during early stages of infection. The main hypothesis of the present project was that a mutation of Fas and Mga transcriptional regulators can alter the transcription of a series of crucial genes important to the ability of live *S. equi* vaccines to cause adverse reactions. The objective of the present work was to use a transcriptomics approach to identify changes in gene transcription across the *S. equi* genome linked to the loss of function of Fas and Mga.

2) Methodology

Sample preparation

In the present project, we selected two *S. equi* mutant strains from the vaccine library output pools and the vaccine strain *Se*4592 for comparison.

A single colony of each *S. equi* strain was grown overnight in 3 mL Todd Hewitt (TH) broth containing 30 µg/mL hyaluronidase (Sigma) at 37 °C in a 5% CO_2-enriched atmosphere, centrifuged, the pellet resuspended in 200 µL Gram positive lysis solution (GenElute, Sigma) containing 50 units/µL mutanolysin and 2×10^6 units/mL lysozyme and incubated for 1 h at 37 °C to allow efficient cell lysis. DNA was then purified using GenElute spin columns according to the manufacturer's instructions (Sigma).

Primer design and PCR reaction

A forward primer was designed to the ISS*1* sequence and a set of several different reverse primers were designed in *mga* and *fas.* The primer sequences are listed in Table 1.

PCR reactions contained 2 µL of 1× PCR buffer, 2 µL of deoxyribonucleotide triphosphate (2 mM) mix, 1.6 µL $MgCl_2$ (2 mM), 0.5 µL forward and reverse primers (100 µM), 0.2 µL Sigma Taq DNA polymerase, and 0.2 µL DNA in a total volume of 20 µL. Reactions were thermocycled for 20 sec

at 95 °C followed by 35 cycles for 30 sec at 95 °C, 30 sec at 55 °C and 1 min at 72 °C, and finally for 7 min at 72 °C.

DNA sequencing

PCR products were purified using a Monarch Gel DNA extraction kit, according to the manufacturer's instructions (Biolabs) and prepared for Sanger sequencing as follows. PCR reaction contained 1 µL of 10 µM forward primer, 1 µL SBDD,1 µL Big Dye, 1 µL of purified PCR product and 2 µL of water in a total volume of 6 µL. Reactions were thermocycled for 30 sec at 96 °C followed by 44 cycles for 4 sec at 92 °C, 4 sec at 55 °C and 75 sec at 60 °C. After a brief spin, 60 µl of 80% isopropanol were added to the samples and centrifuged at 4000 rpm for 30 min. Then 100 µl of 60% isopropanol were added and the samples were left at room temperature for 30 min until the tubes were totally dried and finally 10 ml of HIDI were added and mixed by vortexing briefly until the samples were ready for sequencing. The final sequences were subjected to a BLAST analysis against prokaryotic databases (NCBI) to confirm specific amplification.

Extraction of RNA

Once the mutant strains containing IS*S1* insertions in *fas* and *mga* were identified, RNA was isolated from the two mutants and the wild-type bacteria (*Se*4592 vaccine strain) for transcriptomic sequencing. *S. equi* strains were inoculated onto Todd Hewitt (TH) plates and grown overnight at 37 °C in an atmosphere containing 5% CO_2. A single colony of each strain was inoculated into 10 mL of Todd Hewitt Broth (THB) and grown overnight at 37 °C in an atmosphere containing 5% CO_2. Cultures were diluted 1:10 in fresh THB and grown to an $OD_{600\ nm}$ of 0.3. Two milliliters of each culture was then mixed with 4 mL of RNA protect (Qiagen) and pelleted by centrifugation for 10 min at 5000*g* and then 8000*g* for a further 10 min. Each pellet was resuspended in 200 µL of tris-EDTA buffer (TE) (Fluka) containing 3 mg of lysozyme (Sigma) and 500 units of mutanolysin (Sigma). Total RNA was extracted using RNeasy mini kit, according to the manufacturer's protocol with the some minor alterations. After incubation at room temperature for 45 min with repeated vortexing, 700 µL of RLT buffer from the kit was added and the tube vortexed briefly. The lysis mixture was transferred to a tube containing 0.05 g of acid washed glass beads (Sigma) and vortexed constantly for 5 min. The lysis mixture was centrifuged for 10 sec at 16,100 *g* and the supernatant was removed for RNA extraction with the specific kit.

To remove genomic DNA, all the samples were treated with RNase-free DNase kits (Qiagen). RNA purity and quantity were determined using a Nanodrop ND1000 spectrophotometer (NanoDrop Technologies).

Library construction

RNA was depleted of ribosomal RNA using RiboZero RNA removal kit (Illumina), strictly following the manufacture's instruction. After this step, RNA concentration was measured using the Qubit RNA assay (ThermoFisher Scientific) and the sample stored at −20 °C overnight. Library construction for Illumina sequencing was carried out using NEBNext Ultra Directional RNA Library Preparation Kit (Illumina) following the protocol for use with ribosome depleted RNA. Libraries were quantified by Real Time PCR using SYBER fast Illumina library quantification kit (Kapa Biosystem), no-template controls (NTCs) were included in the run to detect contamination introduced during reaction setup.

MiSeq DNA Sequencing

The samples were sequenced on an Illumina MiSeq DNA sequencer, using a MiSeq reagent kit V3 (Illumina). Genes controlled by the regulator were identified by virtue of decreased or increased transcription in the mutant strains (*fas* and *mga*) relative to wild-type (*Se*4592 vaccine). The bioinformatics pipeline required for the analysis of these data was already established at the Animal Health Trust.

3) Results and Discussion

The presence and size of the amplified products were confirmed by electrophoresis on 1% agarose gels, as shown in Figure 1. All gel-purified PCR products were sequenced using gene specific primers. The final sequences were submitted to a BLAST analysis to verify specific amplification. After alignments in the EMBL GenBank database, the sequences were found to be SEQ_0231 and SEQ_0302, confirming that strains were mutants in *mga* and *fas*, respectively.

Once the mutants in *fas* and *mga* were identiifed, RNA was isolated from the two mutants and the wild-type bacteria (*Se*4592 vaccine strain) to begin the transcriptomic assay.

After extraction and DNase treatment, the RNA OD_{260nm}/OD_{280nm} ratio ranged from 1.9 to 2.1, indicating that the RNA preparation had worked well for all the samples. Treating the sample with DNase is an important step because contaminating DNA can lead to inaccurate RNA quantification, interfere with rRNA removal and negatively affect RNA-seq library prep and sequencing. Additionally, accurate RNA quantification is necessary to use the correct amount of Ribo-Zero rRNA Removal Solution. After this step a Qubit Flourometer was used to quantify the yield of Ribo-Zero-treated RNA (Table 2).

Sequencing libraries were prepared, purified and quantified. Library quantification was performed by amplifying a set of six pre-diluted DNA standards and diluted library samples by qPCR. The average C_q score for each DNA standard was plotted against $\log_{10}$ (concentration in pM) to generate a standard curve, that had a slope= -3.492; R^2= 0.999; efficiency=93.4%. The concentrations of diluted library samples were calculated against the standard curve, using absolute quantification. The three samples of the present project had the following concentrations: *fas* mutant at 173 nM, *mga* mutant at 178 nM and the *Se*4592vaccine strain at 76 nM. 4 pM were then sequenced on an Illumina MiSeq sequencer. After the run, RNA sequencing datasets in FastQ format underwent quality control analysis using FASTQC, confirming a good quality of the reads.

The indexed sequencing reads from each isolate were mapped against the reference genome of *Se*4592 using Bowtie2, and transcriptomes were reconstructed using Cufflinks. The transcriptomes of the mutant strains were compared to the transcriptome of *Se*4592 using Cuffdiff. A genome browser screenshot is shown in figure 2.

I have listed only those genes that showed a statistically significant difference ($P< 0.001$) in transcription between the vaccine strain, and the *mga* or *fas* mutants. Nine genes were differentially transcribed in the *mga* mutant and 7 genes in the *fas* mutant (Table 3). These results are very interesting because show that a mutation of *fas* and *mga* alters the transcription of a series of genes, that deserve further investigation.

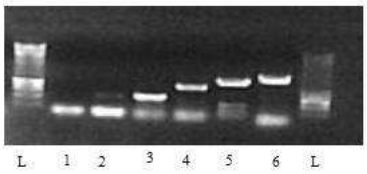
Unfortunately, due to a shortage of time, it was not possible to perform replicates of this experiment or to confirm transcriptome findings by qPCR assay. However, both of these steps will be conducted in the near future at the AHT through a new PetPlan Charitable Trust grant-funded project. These data will improve our knowledge of the genes controlled by these important regulators and will shed light on the processes that underlie the formation of adverse vaccine reactions towards improving the safety of strangles vaccines.

Table 1.

ID	**PRIMER SEQUENCES 5'-3'**
ISS*1*_F	ccacgaatagaaggactgtc
Mga_R1	atgaataatagctttaggagcc
Mga_R2	gctctgccattactatcatctg
Mga_R3	caagcgctttaaggttgacaagc
Mga_R4	tctatccagctcagcaaggactg
Mga_R5	agatatcagcatgctcacagcc
Mga_R6	agtagcttcagagtcgatgcg
Fas_R1	ctactagctgcgatagtg
Fas_R2	aagtgccagtttaatgtctg
Fas_R3	tccataggtttccagattagc
Fas_R4	tggtgcgcaaggatagctcag
Fas_R5	gacccagatgagcagcatgag
Fas_R6	tgtctcaaacggtagtgatcac
Fas_R7	gagagaccatacaagaatcacc
Fas_R8	caaaacttaactcccatatc

Figure 1. DNA ladder 1Kb (L); all the numbers indicate the different reverse primers used for *mga* and *fas* , respectively. For *mga* the reverse primers that gave the bands shown oh the gel were Mga_R3, Mga_R4, Mga_R5, Mga_R6. For *fas* reverse primers were Fas_R5, Fas_R6, Fas_R7, Fas_R8.

mga



fas

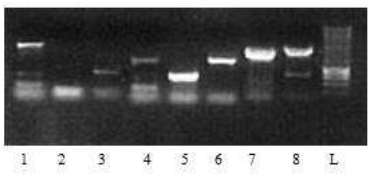
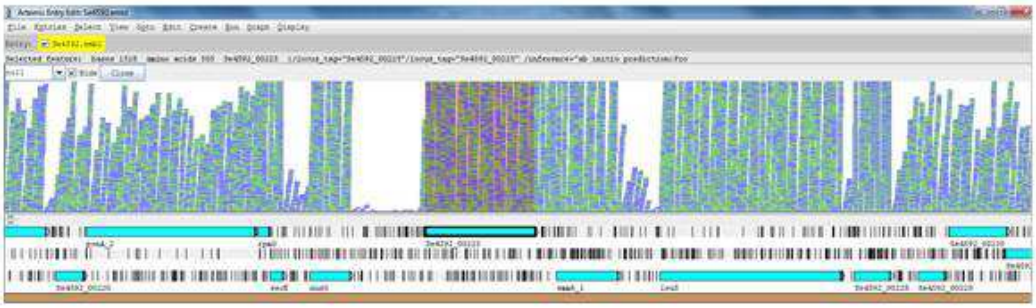


Table 2

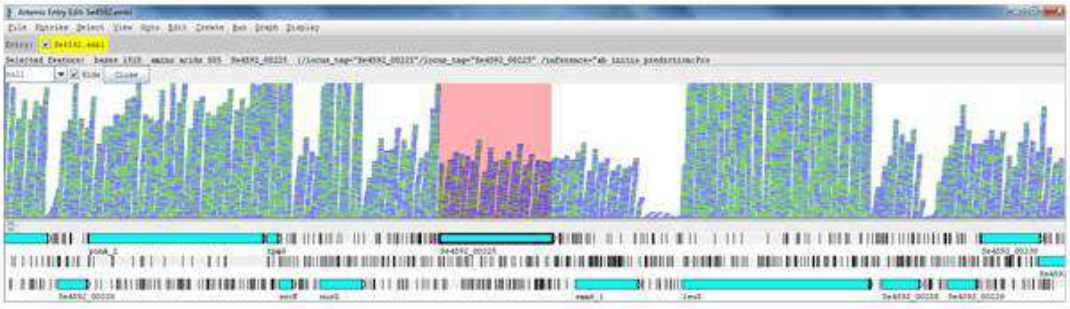
Qubit ™ Fluorometer						
Sample	Conc in the Qubit		ul used	Dilution	Sample concentration	
Mutant in fas	144	ng/mL	1	200	28.9	ug/mL
Mutant in mga	144	ng/mL	1	200	28.8	ug/mL
Se4592 Vaccine	158	ng/mL	1	200	31.6	ug/mL

Figure 2

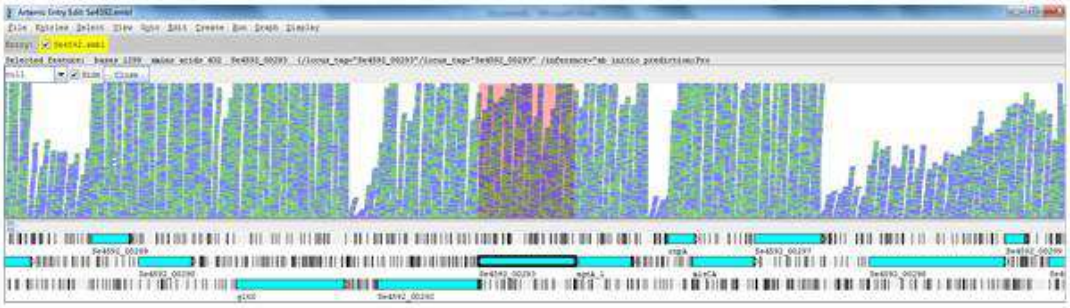
*Se*4592 Vaccine



Mga



*Se*4592 Vaccine



Fas

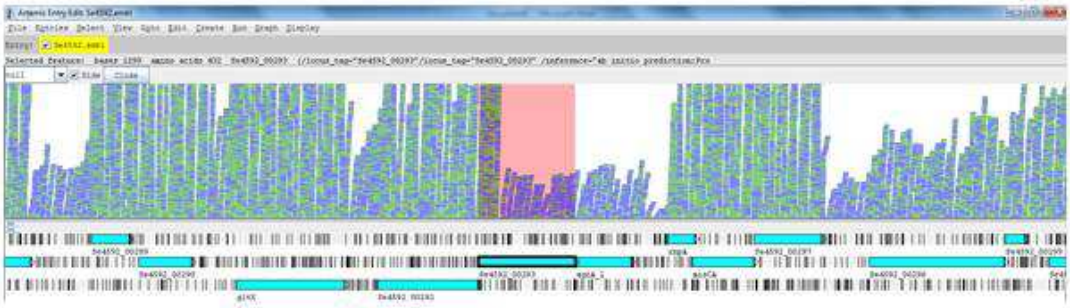


Table 3

test_id	locus	sample_1	sample_2	value_1	value_2	log2(fold_	fold change	gene	product
Se4592_00294	null:247840-248581	4592vaccine	fas	392,879	47,7287	-3,04116	0,121484477	fasC	response regulator protein,Accessory gene regulator protein A,putative two-component response-regulatory protein
Se4592_00456	null:416820-417855	4592vaccine	fas	524	99,1374	-2,40206	0,189193511	hrcA	heat-inducible transcription repressor
Se4592_01322	null:1290229-1290940	4592vaccine	fas	440,409	84,8949	-2,37509	0,192763772		mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase family protein,Muramidase-2 precursor,N-acetylmuramo
Se4592_01382	null:1348108-1348762	4592vaccine	fas	3920,49	703,204	-2,47902	0,179366355	yecS	transport system membrane protein,Inner membrane amino-acid ABC transporter permease protein yecS,amino acid
Se4592_01383	null:1348789-1349656	4592vaccine	fas	4846,82	743,161	-2,70529	0,153329606	yckB	putative ABC transporter cystine-binding lipoprotein precursor,hypothetical protein,cystine transporter subunit,lysin
Se4592_01443	null:1407175-1408144	4592vaccine	fas	1694,93	376,104	-2,17202	0,221899429	acoA	Pyruvate dehydrogenase E1 component alpha subunit,Acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit
Se4592_01546	null:1519874-1520345	4592vaccine	fas	1882,2	11104	2,56058	5,899479333	sigV	RNA polymerase sigma factor protein

test_id	locus	sample_1	sample_2	value_1	value_2	log2(fold_	fold change	gene	product
Se4592_00225	null:171222-172740	4592vaccine	mga	1805,94	84,5149	-4,4174	0,046798288		Mga-like regulatory protein,M protein trans-acting positive regulator (MGA) PRD domain
Se4592_00226	null:173042-173885	4592vaccine	mga	1569,55	54,9955	-4,83489	0,035039024	emm6_1	cell surface-anchored protein,M protein,serotype 6 precursor,Uncharacterized conserved protein,chromosome segregation
Se4592_00228	null:177161-177653	4592vaccine	mga	15534,5	146,31	-6,73031	0,009418391	Se18.9	anti-phagocytic factor H binding protein
Se4592_00255	null:208576-209116	4592vaccine	mga	193,688	1577,77	3,02608	8,145935732	bioY	BioY family protein,Biotin ECF transporter S component BioY,BioY family
Se4592_00261	null:213153-214407	4592vaccine	mga	4690,6	83,2026	-5,817	0,017738157	icaA_1	hyaluronan synthase,Poly-beta-1,6-N-acetyl-D-glucosamine synthase,N-glycosyltransferase
Se4592_00887	null:829384-830476	4592vaccine	mga	982,366	140,114	-2,80966	0,142629122		collagen-like surface-anchored protein SclF,LPXTG cell wall anchor domain,Collagen triple helix repeat
Se4592_00968	null:918174-919329	4592vaccine	mga	63,5668	7,7737	-3,0316	0,122291825	fhaB	Mac family protein,Mac 1 hemagglutinin,Ribonucleases G and E,LPXTG cell wall anchor cell surface-anchored protein,Filame
Se4592_00969	null:919414-920764	4592vaccine	mga	50,13	5,82641	-3,10499	0,116226012	fhaB	cell surface-anchored protein,Filamentous hemagglutinin,Ribonucleases G and E
Se4592_01860	null:1827632-1829162	4592vaccine	mga	118,718	15,8117	-2,90848	0,133187048		collagen-like surface-anchored protein SclI,M protein, serotype 49 precursor,LPXTG cell wall anchor domain,Collagen triple