Genomic and transcriptomic profiling of nematode parasites surviving after vaccine exposure

Guillaume Sallé[1,2,*], Roz Laing[3], James A. Cotton[2], Kirsty Maitland[3], Axel Martinelli[2#], Nancy Holroyd[2], Alan Tracey[2], Matthew Berriman[2], W. David Smith[4], George F. J. Newlands[4], Eve Hanks[3], Eileen Devaney[3], Collette Britton[3]

[1]INRA - U. Tours, UMR 1282 ISP Infectiologie et Santé Publique, Centre de recherche Val de Loire, Nouzilly, France

[2]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United-Kingdom

[3]Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Bearsden Road, Glasgow, United-Kingdom

[4]Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik, Midlothian, United-Kingdom

[#]Present address: Global Station for Zoonosis Control, Global Institution for Collaborative Research and Education (GI-CoRE), Hokkaido University, N20 W10 Kita-ku, Sapporo, Japan

*Corresponding author

Email: gs12@sanger.ac.uk

Abstract

Nematodes are economically important parasites of many livestock species, while others are important human pathogens causing neglected tropical diseases. In both humans and animals, anthelmintic drug administration is the main control strategy, but the emergence of drug-resistant worms, has stimulated the development of alternative control approaches. Among these, vaccination is considered to be a sustainable and cost efficient strategy. Currently, Barbervax® for the ruminant strongylid *Haemonchus contortus* is the only registered subunit vaccine for a helminth parasite, although a vaccine for the human hookworm *Necator americanus* is undergoing clinical trials (HOOKVAC consortium). As both these vaccines comprise a relatively small number of proteins there is potential for selection of nematodes with altered sequence or expression of these antigens. Here we compared the genome and transcriptome of *H. contortus* populations surviving following vaccination with Barbervax® with worms from control animals. Barbervax® antigens are native integral membrane proteins isolated from the brush border of the intestinal cells of the adult parasite and many of them are proteases. Our findings provided no evidence of selective pressure on the specific vaccine antigens in the surviving parasite populations. However, surviving parasites showed increased expression of other proteases and regulators of lysosome trafficking, and displayed up-regulated lipid storage and defecation abilities that may have circumvented the vaccine effect. Despite the minimal evolutionary time monitored within this experiment, genomic data suggest that a hard selective sweep mediated by the vaccine is unlikely. Implications for potential human hookworm vaccines are discussed.

Author Summary

Gastrointestinal nematodes are important parasites of humans and livestock species. Their high adaptive potential allows them to develop resistance to the anthelmintic drugs used to

control them. Vaccines could also select for resistance ultimately leading to vaccination failure. Here, we provide the first characterization of the genetic and transcriptomic composition of a worm population that survived following vaccination and challenge infection in comparison to a control population. Noticeably, no selection pressure was found on the vaccine targets. Instead, surviving parasites demonstrated increased expression of other genes involved in blood meal digestion that could facilitate survival in response to the vaccine.

Introduction

Gastrointestinal nematodes (GIN) are clinically and economically important parasites of humans [1] and livestock species [2]. Human GIN (e.g. hookworm, roundworm and whipworm) infect over one billion people worldwide, resulting in the loss of over four million disability-adjusted life years (DALYs) in 2013 [1]. In ruminants, parasitic nematode infections cost the global livestock industry billions of dollars annually in production losses and treatments [2]. GIN therefore impede both human health and wealth and are an aggravating factor of poverty [3].

Control of veterinary parasites has relied primarily on strategic drug administration strategies [4]. However the increase in anthelminthic resistance, particularly multidrug resistance, threatens the viability of the livestock industry in many regions of the world [2]. Similarly, control of human helminthiases involves large-scale community treatment, which has resulted in reductions in GIN prevalence over the last 30 years [1, 5], but there is also the potential for anthelminthic drug failure [1]. Indeed recent surveys indicated a low or variable cure rate of human hookworm infection after benzimidazole treatment, with no reduction in anaemia in some endemic regions [6, 7].

It is unlikely that novel anthelmintic compounds will be approved at an equivalent pace to the emergence of anthelmintic resistance [8]. Greater research effort is therefore being directed at vaccine development for more sustainable GIN control in both veterinary and human settings [1, 9]. Vaccines may be used alone or combined with drug treatment to reduce the emergence of drug resistance [10]. In comparison to antimicrobial or anthelmintic drugs, there are few examples of the development of resistance to vaccination in eukaryotic pathogens [11]. However, the antigenic complexity and immunoregulatory capacity of nematode parasites makes vaccine development a difficult task [9]. Only two vaccines are currently commercially available: Barbervax® licensed in Australia in 2014 and comprising parasite gut membrane glycoproteins of the ovine GIN *Haemonchus contortus* [12, 13], and Bovilis huskvac®, an irradiated larval vaccine for the cattle lungworm *Dictyocaulus viviparus* [14]. For the human hookworm *Necator americanus*, a phase 1 clinical vaccine trial has been carried out [15].

While L3 larvae of *H. contortus* and *N. americanus* mediate infection via different routes, adult worms of both of these clade V parasites reside in the digestive tract, where they blood feed. Digestion of haemoglobin in haematophagous nematodes requires activity of different proteolytic enzymes, including aspartic, cysteine and metallo-proteases and exopeptidases [16]. The large expansion of protease gene families identified within the genomes of *H. contortus* [17, 18] and *N. americanus* [19] supports the importance of these enzymes in parasite biology. These blood digesting proteases are therefore valid targets for vaccine control. Barbervax® is prepared from gut membrane extracts from *H. contortus* adult worms and contains two major protease fractions, H11 and H-gal-GP [20]. H11 is a family of microsomal aminopeptidases for which five isoforms have been identified so far [21, 22]. H-gal-GP is a 1,000 kDa complex of four zinc metallopeptidases (MEP1-4) and two pepsinogen-like aspartyl proteases (PEP-1 and PEP-2) [23], together with additional

components thought unlikely to be protective (thrombospondin, galectins and cystatin [24]). Vaccination of sheep with H11 or H-gal-GP individually reduced worm burden and faecal egg count by 70% and 95%, respectively [21, 22, 24-26]. Cysteine proteases HmCP-1,4 and 6, enriched from adult *H. contortus* gut membrane provided a lower level of protection [27]. The *N. americanus* vaccine comprises the Na-APR-1 aspartic protease combined with gluthatione-S-transferase-1 (Na-GST-1) which is involved in heme detoxification [1, 28, 29]. Both Barbervax® and the *N. americanus* vaccine induce antibodies that are ingested during parasite blood-feeding and are thought to inhibit enzymatic activity and interfere with gut function [24, 28]. Because its gut-membrane antigens are not exposed to the host immune system during natural infection, Barbervax® relies on the induction of antibodies to "hidden" antigens [24]. Therefore, it is speculated that the Barbervax® proteins are not under purifying selective pressure during natural infection, but whether vaccine-induced immunity exerts any selection on parasite genetics or influences gene expression is currently unknown.

The high level of genetic diversity observed in genomic datasets of *H. contortus* [17] and other helminths underpins their capacity for adaptation and contributes to the evolution of drug resistance [30]. It is clear that pathogens can evolve in response to other interventions, including vaccination, in some cases leading to vaccine escape and failure [11, 31]. Given the limited number of antigens composing the *N. americanus* and *H. contortus* vaccines, strong selection may arise in the field. Here we compare the genomes and transcriptomes of *Haemonchus* adults surviving in Barbervax® vaccinated animals with worms recovered from control animals post challenge infection. Identifying any effects that vaccines may have on the diversity of helminth populations may guide their optimal use in both veterinary and human settings.

Materials and methods

120 **Experimental design and collection of parasite material**

121 Twelve six month old worm-free Texel cross lambs were allocated into groups of six,
122 balanced for sex and weight. One group was injected subcutaneously with two doses of
123 Barbervax® four weeks apart, whilst the second, control group was not vaccinated. All sheep
124 were given a challenge infection of 5,000 *H. contortus* MHco3(ISE) L3 administered *per os*
125 on the same day as the second vaccination. Fecal egg counts (FEC) were monitored twice
126 weekly between days 17 and 29 post-challenge by a McMaster technique with a sensitivity of
127 50 eggs/g. Adult worms were recovered from each sheep at post-mortem 31 days post-
128 challenge and worm volumes recorded. Antibody titres were measured by ELISA, with plates
129 coated with Barbervax® (50 µl per well at 2 µg/ml). Serum samples were serially diluted
130 (from 1/100 to 1/51200) in PBS/0.5% Tween and binding detected using mouse anti-sheep
131 IgG (Clone GT-34, Sigma G2904; 1:2500 dilution) and rabbit anti-mouse IgG-HRP
132 conjugate (Dako P0260; 1:1000 dilution). Antibody titres are expressed as the reciprocal of
133 the end-point dilution resulting in an OD of ≥ 0.1 above the average negative control value.
134 All experimental procedures were approved by the Moredun Research Institute Experiments
135 and Ethics committee and carried out in accordance with the Animals (Scientific Procedures)
136 Act of 1986.
137
138 **Extraction protocol, library preparation and sequencing**
139 DNA and RNA sequencing was carried out on two pools of surviving *H. contortus* adult male
140 worms per sheep according to the available number of worms. Six to twelve male worms
141 were sequenced from each sheep. To avoid any confounding factors from eggs in females or
142 differences in sex ratio between samples, only male worms were used for sequencing. In
143 total, 49 and 52 worms that survived following challenge infection of the Barbervax

144 vaccinated sheep (V group) were analysed while 63 and 60 were picked from control sheep
145 (C group), for DNA and RNA preparations, respectively.
146 Total RNA was extracted from the worms using a standard Trizol (Thermo Fisher Scientific,
147 15596026) protocol and libraries made with the Illumina TruSeq RNA preparation kit. DNA
148 was extracted using the Wizard Genomic DNA Purification kit (Promega, A1120), following
149 the manufacturer's instructions, apart from digestion with proteinase K in TEN buffer with
150 SDS, followed by RNase A (Promega, A7973) treatment. RNA and DNA were sequenced
151 using a HiSeq 2500 platform with v3 chemistry, generating a total 45Gb of DNA sequence.
152
153 **Real-time PCR**
154 Total RNA was extracted from five female worms from the same populations as the
155 sequenced males. 3 µg total RNA was used per oligo(dT) cDNA synthesis (SuperScript® III
156 First-Strand Synthesis System, ThermoFisher, 18080051) with no-reverse transcriptase
157 controls included for each sample. cDNA was diluted 1:100 for RT-qPCR and 1ul added to
158 each reaction. RT-qPCR was carried out following the Brilliant III Ultra Fast SYBR QPCR
159 Master Mix protocol (Agilent Technologies, 600882) and results analysed using MxPro
160 qPCR Software, Version 4.10. Gene expression was normalised to *ama* (HCOI01464300)
161 and *gpd* (HCOI01760600) [32]. Primer sequences are listed in Table S1.
162
163 **Improved *H. contortus* assembly and corresponding gene model**
164 The *H. contortus* MHco3.ISE reference genome assembly used for this study was a snapshot
165 of the latest version as of 14/11/2014. This assembly consists of 6,668 scaffolds with a total
166 assembly length of 332,877,166 bp; of which 22,769,937 bp are sequence gaps. The N50
167 scaffold length is 5,236,391 bp and N90 length is 30,845 bp. Specifically for this project,
168 preliminary gene models were annotated on this assembly by transferring the gene models

from the published (v1.0) genome assembly [17] using RATT [33] with default parameters, and with a *de novo* approach using Augustus v2.6.1 [34] with exon boundary 'hints' from the RNAseq data described in [17], mapped against the new reference genome in the same way as in this previous paper.

RNAseq data handling and differential expression analysis

RNAseq data were mapped onto the reference genome using a gene index built with Bowtie2 [35] and TopHat v2.1.0 [36] with maximal intron length of 50 Kbp and an inner mate distance of 30 bp that identified 48.8% of the reads being mapped unambiguously to a gene feature. Counts of reads spanning annotated gene features were subsequently determined with HTSeq v0.6.0 [37].

To ensure our biological conclusions are not sensitive to details of the statistical methods used, we implemented two different analysis frameworks for the RNA-seq count data, using the DESeq2 v1.12.4 framework [38] and the *voom* function as implemented in the LIMMA package v3.28.21 [39] in R v3.3.1 [40]. Genes found to be significantly differentially expressed (DE, adjusted p-value <5%) by both VOOM and DESeq2 analyses were retained. A gene ontology enrichment analysis was performed using the TopGO package v2.26.0 [41]. Gene identifiers of the vaccine core components, namely MEP-3 [42], MEP-1,2,4, PEP-1 [43] and PEP-2 [23] as well as H11, were retrieved via a BLAST search of their nucleotide sequence against the *H. contortus* MHco3.ISE reference assembly [17] in WormBase ParaSite [44]. The expression levels of candidate housekeeping genes [32] were also retrieved using the gene identifiers associated with their GenBank records (Table 2).

Genomic sequence data mapping and SNP calling procedures

Reads were mapped onto the reference genome using SMALT v0.7.4 (http://www.sanger.ac.uk/science/tools/smalt-0) software with a minimum 90% identity cut-off (-y 0.9), a maximum insert size of 2000bp and a k-mer length of 13 with a step-size of 2 between adjacent k-mers. For computational efficiency and to avoid mapping to multiple scaffolds representing haplotypes of the same loci, only the 140 biggest contigs (those over 100 Kbp in length) were included in the analysis. These spanned 83.89% of the genome assembly (279.24 Mbp). Duplicated and bad quality reads (mapping quality below 20, unmapped reads and reads with an-mapped mate) were filtered out with Samtools v0.1.19 [45], yielding an average of 3.2×10^9 reads (Table 2) for each pool. Local realignment around indels was performed with the GATK software v3.4 [46, 47]. A first set of SNPs was called using mpileup and computing allele counts for each site after filtering out indels with Popoolation2 [48]. The GATK HaplotypeCaller workflow was also implemented for calling variant sites in each pool after specifying a ploidy of 2 and implementing a hard filtering of raw variants (QD < 2, DP > 468.33, FS > 21.72, MQ < 33.07, MQRankSum < -1.76, ReadPosRankSum < -2.16). In addition, SNPs found in less than 11 out of the 12 pools were discarded. Applying this hard filtering framework to every scaffold resulted in a total of 5,291,725 SNPs, and subsequently only variants present in both the GATK and popoolation2 sets of variant calls were retained for analysis (n = 4,807,726 SNPs in total). SNP calling was performed for all 12 available pools of worms, but given the lower amount of genomic data available (a single lane), one sample (V_6) was discarded for further analyses to avoid biases in allele frequency estimates.

Estimation of allele frequencies and site frequency spectra

Maximum-likelihood estimates for allele frequencies were derived following the framework of [49] using a python script available upon request. Only SNPs for which the null hypothesis

of a single allele being present was rejected (*i.e.* showing a p-value below 0.05, in every pool considered) were retained for subsequent analyses (n = 198,045 SNPs). A principal components analysis was run on allele frequency estimates using the *prcomp()* function implemented in the R software.

Search for genomic selection footprints, gene ontology analysis and simulations

Departures from neutrality were investigated by genome-wide estimation of Tajima's D, π, and Watterson's θ using the popoolation2 software [48] and 10kbp wide overlapping sliding windows. Pairwise-F_{ST} statistics between pool of worms were estimated along the genome using the popoolation2 software [48] using a 10Kbp wide sliding window with a 1 Kbp overlap (44,668 tests), or using a gene-wise estimation (6,194 tests). The F_{ST} statistics is a measure of the similarity between allelelic frequencies from different populations [50]. Outlier F_{ST} values are encountered when diversifying selection is acting on particular loci, in contrast to other regions submitted to random evolution, *i.e.* genetic drift. In our setting, pairwise F_{ST} between C samples (C-C) measures the genetic differentiation occurring when pools of worms are undergoing a bottleneck linked to the immune response in unvaccinated sheep, either by chance or because of some immune selection. High F_{ST} between C and V pools (C-V) should point at regions affected by vaccine response exposure. Finally, high F_{ST} between V samples (V-V) may originate from different genetic samplings associated with survival to the vaccine response, *e.g.* different genes at stake. Identification of the regions specifically put under selection by the sheep vaccine response relies on comparison of the observed F_{ST} values to a null distribution of F_{ST} under random drift or the application of empirical thresholds. This latter option was favoured as the former requires prior information on the selection coefficient and population genetic parameters that are unknown. Therefore, to identify the regions specifically put under selection by the sheep vaccine response, we

selected regions that showed an outlying C-V F_{ST} defined as being significantly higher than C-V estimates (at least two standard deviations of the average C-V F_{ST}) and significantly higher than outlying C-C F_{ST} estimates (average C-C plus three standard deviations of the mean). In addition and because we were only interested in the regions specifically constrained by the host vaccinal response, these regions also had to exhibit low differentiation among C samples (less than two standard deviations of the mean).

Genes and their promoter regions (2000 bp up- and down-stream) overlapping windows fulfilling these conditions were retrieved to perform a gene ontology analysis with the GOWINDA software, running 100,000 simulations and reporting GO terms with a false discovery rate below 5% [51]. A neighbour-joining tree based on Reynold's distance [52] between pools was generated using the *nj()* function of the *ape* package.

To assess the impact of the experimental design on the observed F_{ST} values within and among experimental groups, forward simulations were implemented with the simuPOP software [53]. A population of 55,000 individuals was simulated and split into 11 populations of 5,000, 20 to 25% of which randomly surviving to adult stage in *H. contortus* [54]. A marker under selection conferring between 95 and 99% chance of dying to recessive homozygotes and 1% to other genotypes in the vaccinated populations was modelled. This marker also gave a 1% chance of dying to the recessive homozygotes and 0% to other genotypes in the unvaccinated population. A neutral marker was also modelled. Finally, 20 of the surviving individuals were randomly sampled to compute pair-wise F_{ST} estimates based on the virulence-associated and neutral markers. For each starting allele frequency, 10,000 simulations were run.

Results

Vaccination greatly reduces faecal egg counts in vaccinated sheep

Parasitological data confirmed a significant reduction in *H. contortus* infection following Barbervax vaccination. Over the course of the trial, vaccinated sheep (Group V) shed significantly fewer eggs (mean 390 ± 639 eggs per gram faeces (epg), Fig 1A, Table S2) than the control group (Group C) given the same challenge infection dose without prior vaccination (mean 5,914± 2,628 epg), representing a 15-fold decrease (Wilcoxon test, p= 0.002). Vaccinated sheep contained fewer worms, indicated by the significantly lower worm volume collected at necropsy compared to control sheep (2.8 mL ± 1.9 versus 6.7 mL ± 3.5; Table S2). Among the V group, V_5 showed an outlying egg excretion over the course of the trial (1,647 epg at necropsy; upper 95% confidence interval limit of 861 epg estimated after 1,000 bootstraps), suggesting a relatively suboptimal vaccine response in this animal. This is supported by the lower antibody titre of this sheep, relative to the other Barbervax vaccinated animals, at day 28 post-challenge infection (Fig 1B).

Fig 1. Faecal egg counts, worm volume and anti-Barbervax IgG titer of individual animals

Fig 1A Faecal egg counts from each of the 12 sheep in the trial were plotted for each available time point. The plot shows an 15-fold difference in egg excretion between vaccinated and control sheep on day 29 post challenge infection.

Fig 1B Faecal egg count measured at necropsy plotted against respective anti-Barbervax® vaccine IgG titer, showing a negative correlation between vaccine response and egg count.

Limited but unexpected genetic differentiation between worm populations in different hosts

Whole genome sequencing was carried out on groups of male worms recovered from vaccinated or control animals post-challenge infection. Genome sequence data was examined

for any evidence that the vaccination-induced reduction in FEC, and in adult worm burden, had an impact on the genetics of the worm populations recovered. SNP calling and allele frequency based selection procedures yielded an average of 1,855,712 SNPs per pool of worms (Table 1), of which 198,045 sites spanning 130 contigs were polymorphic in every pool. Sample V_6 whose coverage was lower was discarded for population genetic parameter inferences to avoid biases in allele frequency estimates (Table 1).

Table 1. Available transcriptomic and genomic data for each pool of worms

Sample	RNAseq library size (M reads)	DNAseq library size (bp)	DNAseq mean coverage	% of bases with cov.>15	No. SNPs identified	Tajima's *D*
C_1	11,617,210	3,710,647,378	14.44	45.2	1,996,265	-0.36
C_2	11,983,512	3,376,010,836	13.14	38.8	1,855,906	-0.34
C_3	11,397,785	3,675,872,370	14.3	44	2,013,325	-0.3
C_4	8,411,138	3,624,001,304	14.1	43.4	1,958,571	-0.35
C_5	10,485,931	3,782,001,835	14.72	45.5	2,061,175	-0.39
C_6	11,627,840	3,231,531,531	12.57	36.1	1,832,316	-0.31
V_1	11,673,469	2,876,605,295	11.19	29.1	1,592,790	-0.31
V_2	11,147,222	2,943,617,550	11.45	30	1,683,880	-0.35
V_3	11,876,035	2,833,188,035	11.02	27.9	1,609,184	-0.3
V_4	12,298,240	3,394,736,282	13.21	38.9	1,888,056	-0.35
V_5	11,459,627	3,580,556,572	13.93	42.4	1,921,368	-0.29
V_6	6,122,970	138,2085,315	5.38	1.8	n/a	n/a

Tajima's *D* estimates, that measure departure from neutral evolution, were neither significantly different from 0 (expected under neutral mutation-drift equilibrium) nor differed between the two groups of sheep (Table 1). Principal component analysis (PCA) of the SNP data also showed little systematic difference between C and V samples (Figs 2A, 2B). This finding was also supported by estimated genome-wide F_{ST} statistics. We used the genome-wide average F_{ST} between pairs of worm pools to quantify the level of genetic differentiation observed between pools, comparing pairs surviving in vaccinates (V-V comparisons), pairs of control samples (C-C comparisons) and between pairs of vaccine and control pools (C-V comparisons). These statistics showed similar, moderate differentiation among (average C-V F_{ST} 0.059 ± 0.013) and within experimental groups (average C-C F_{ST} 0.054 ± 0.016, average

312 V-V F_{ST} 0.062 ± 0.017) (Fig 2C). These values suggested some independent evolution within
313 each sheep (Fig 2C), even in the absence of vaccine exposure. Indeed, some level of genetic
314 differentiation occurred between C populations for neutral alleles, *i.e.* not controlling for the
315 vaccine survival, whereas 90% of our simulated F_{ST} estimates in this case fell below 0.01
316 (Fig 3A). While all populations showed higher F_{ST} than expected from the simulations under
317 pure genetic drift, the impact of the vaccine selection could not be differentiated from the
318 selection occurring in control sheep.
319

320 **Fig 2. Genomic structure of worm populations according to their experimental group**
321 Fig 2A. An overview of the folded site frequency spectra, *i.e.* the distribution of minor allele
322 frequency values ranging from 0 to 0.5, within each pool of worms. Allele frequency
323 estimates have been derived as reported in [49] and SNPs with significant allelic frequencies
324 have been selected for each pool. This plot demonstrates an enrichment in intermediate allele
325 frequencies in the V populations consistent with a stronger bottleneck.
326 Fig 2B. Principal components analysis of genotype data with their relative contributions
327 showing individual pool coordinates and experimental group centres of gravity. Ellipses
328 encompassing the variance of each experimental group largely overlap, favouring a lack of a
329 common genomic divergence process operating between experimental groups.
330 Fig 2C. Neighbor-joining tree derived from mean pair-wise F_{ST} estimate between pools. The
331 tree supports the lack of agreement between differentiation occurring among pools and the
332 experimental groups. It also illustrates the independent evolution of pools recovered from the
333 most extremely bottlenecked populations.
334

335 **Fig 3. Results of the simulated pair-wise F_{ST} estimates**

Fig 3A. The dispersion of the average pair-wise F_{ST} estimates within control (red), vaccinated (green) or among groups (blue) for a marker under neutral evolution with initial frequencies of 0.2 and 0.5 and across the various possible virulence-associated marker frequencies.

Fig 3B. The dispersion of the pair-wise F_{ST} estimates based on a marker controlling for survival to the vaccine response.

However, SNP data also provided indications of a stronger bottleneck occurring in the V samples. First, V populations exhibited significantly fewer segregating sites in agreement with a reduced genetic diversity ($p=0.04$, Table 2); these sites were enriched for variants segregating at intermediate allele frequencies in the V populations, also consistent with a loss of rare variants in these pools due to a bottleneck (Fig 2A). Second, V-V and C-V differentiation measured by F_{ST} statistics was significantly higher than between control samples (Student's t test, $p<2.2 \times 10^{-16}$). The slightly higher V-V differentiation could in principal either be a result of stronger drift in these populations or a wide-ranging selective response to vaccination, both of which agree with the stronger bottleneck observed from parasitological measures. Simulations confirmed that this pattern of higher V-V F_{ST} values, and similar F_{ST} for C-C and C-V comparisons is expected (Fig 3B) in the presence of selection due to the vaccine response.

The overall higher differentiation observed within worms surviving vaccination was mostly driven by differences between V_5 (least protected) and a cluster of three samples (V_1, 2 and 3), that all showed the strongest effect of vaccination, and so potentially the strongest genetic bottleneck (Fig 2C). This variation in allele frequency across sites among survivors suggests that the vaccine response did not exert its selection pressure on the same sites across all V samples, *i.e.* survival following vaccine exposure could be mediated by different genomic regions.

Vaccination induces a slight genetic bottleneck but no detectable hard selective sweep

Despite this evidence of different evolutionary pathways between V samples from different hosts, we investigated whether any regions of the genome were consistently different between C and V samples (Fig 4B). We identified 145 windows (Fig 4B, red dots) showing high C-V differentiation (mean F_{ST} 0.096 ± 0.007) but little or no C-C differentiation (mean F_{ST} 0.04 ± 0.006). This pattern is compatible with differentiation mediated by the vaccine response that would affect allelic variance of V in comparison to C. Overall, these windows were on 11 contigs and defined 21 discrete regions putatively under selection following vaccination. These windows tend to show high V-V differentation (mean F_{ST} 0.088 ± 0.019) hence suggesting that any selection applied by the vaccine response may differ between hosts.

Fig 4. Distribution of average pair-wise F_{ST} estimates and differentially expressed genes along the genome

A. Expression fold-change of the differentially expressed genes were plotted against their genomic positions. Dot size correlates their fold-change magnitude and colour corresponds to higher (red) or lower (green) expression in V compared to C worms.

B. Mean pair-wise F_{ST} estimates have been plotted against each site, ordered by contig and position and for each possible pair-wise group comparison, *i.e.* within pools of worms recovered from control sheep (C-C in green), within pools of worms recovered from vaccinated sheep (V-V, in purple) or among groups (C-V, in blue). Bigger red dots represent sites thought to have been differentiated as a result of the vaccine response. Similar differentiation values are observed throughout the considered sites, indicative of a lack of outstanding differentiation occurring on a particular set of loci.

386

387 Using our simulation results, we can estimate the probability of observing a 10kb
388 window with F_{ST} values as extreme as these 145 (the false discovery rate; FDR) under a
389 neutral model of random drift and no selection. The estimated FDR of between 0.0028 to
390 0.0058 (depending on initial allele frequency) would imply between 125 and 259 false-
391 positive windows in the 44,668 tested windows; although only between 12 and 26 false
392 positive windows are expected when linkage disequilibrium between markers within 10 Kbp
393 is taken into account (based on unpublished estimates of linkage in *H. contortus*). This
394 suggests that most, but probably not all, of the observed C-V differentiation results from
395 random processes.

396 The gene ontology analysis run on the full list of genes retrieved from differentiated
397 windows highlighted a few GO terms independently associated with three unique genes
398 (Table S3) involved in different biological processes, *i.e.* necrotic cell death in relation to
399 dUTP metabolic process (HCOI02115600), structural integrity of muscle (HCOI00179400)
400 or hormone activity (HCOI01942200). In addition, an astacin metallopeptidase M12A gene
401 (HCOI01292600) was found among the 47 genes lying within the defined differentiated
402 genomic windows.

403 Overall gene-wise F_{ST} estimates show the same trends as those observed for the
404 genome-wide estimates, *i.e.* equivalent amount of differentiation between populations (mean
405 pair-wise F_{ST} of 0.06; 0.054; 0.063 and standard deviation 0.019; 0.024; 0.026 for C-V, C-C,
406 and V-V estimates, respectively Table S4). Applying the same framework, to select for
407 putatively selected markers, highlighted 11 genes including a Beige/BEACH coding gene
408 (HCOI02096300) orthologous to the LYSosomal Trafficking regulator protein (*lyst-1*) gene
409 in *C.elegans* (Table S4).

410

411 **Transcriptional response of worms to host vaccination is dominated by higher**
412 **expression of proteases and protease inhibitors**

413 We next investigated any changes in *H. contortus* gene expression in worms surviving in
414 vaccinated sheep relative to those surviving in controls. On average 11M (standard deviation
415 of 1.79M) reads were available for each library (Table 2). In PCA of the normalized RNA-
416 seq read counts, the first two axes explained 53% of the total variation, 37% of which was
417 resolved along the 1st axis that separated the experimental groups (Fig S1). Two pools of
418 worms sampled from control sheep, C_4 and C_6, showed atypical behaviour that was
419 resolved along the 2nd PCA axis (Fig S1). These samples were discarded from the dataset for
420 subsequent analyses, resulting in a comparison of 6 V samples and 4 C samples.
421 We found 52 genes significantly differentially expressed (DE; adjusted p-value < 0.05)
422 between the two experimental groups, with six genes exhibiting a fold change >4 and 34
423 genes showing a fold change >2 (Figs 5, S2, Table S5). Adult worm survival following
424 vaccination was associated with an increase in expression of most of the DE genes, *i.e.* 46 out
425 of 52. Among the top six DE genes, the only down-regulated gene was a glycoside hydrolase
426 domain-containing protein (HCOI00569100, Table 2, Fig 5A). Four of the most highly up-
427 regulated genes encoded proteins containing peptidase domains (HCOI01945600,
428 HCOI01283800, HCOI01736400 Table 2, Fig 5A), or a peptidase inhibitor I4 domain
429 (HCOI01549900, Table 2, Fig 5A), while one gene was unannotated (HCOI01623600).
430 Expression of the peptidases (HCOI01945600, HCOI01283800, HCOI01736400) was
431 validated by quantitative RT-PCR in female worms from the same population as the
432 sequenced males, and confirmed a two to three-fold over-expression of each mRNA in
433 worms surviving in vaccinated sheep compared to controls (Fig 5B).
434

435 **Fig 5. Expression level of the top differentially expressed genes within each**
436 **experimental group (5A) and associated correlation with faecal egg count in control**
437 **populations (5B)**

438 A. A boxplot for every gene that exhibited an absolute log-transformed fold change of 2
439 between the experimental conditions.

440 B. Fold change in expression level of selected genes, by qRT-PCR, shown relative to C
441 control population.

442 C. log10-transformed transcript counts measured in control samples plotted against faecal
443 egg count (in eggs/g) at 29 days post-infection for each of the most differentially expressed
444 genes. Regression lines with associated 95% confidence intervals are also provided.

445

446 To test whether the six most DE genes were associated with increased fitness under control
447 conditions, we estimated the correlation between their normalised transcript counts and FEC
448 in the control C worm populations. Spearman correlation ranged from -0.15 to 0.15 with p-
449 value >0.6, rejecting this association. Furthermore these genes exhibited low transcript counts
450 in control C populations (ranging from 1.5 to 8,476 counts for HCOI01623600 and
451 HCOI01283800 respectively, Fig 5C, Table S5), suggesting that their higher expression in
452 worms from group V may be triggered or selected for by the vaccine exposure.

453 **Table 2. Gene of interest expression levels, fold change and associated p-values**

	Gene ID	Base mean	logFC DESeq2	adj. P DESeq2	logFC VOOM	adj. P VOOM	Correlation with FEC31	WormBase ParaSite Gene description	*C. elegans* orthologue	Candidate Gene Name	Genbank Acc. Number
Top differentially expressed	HCOI00569100	24.21	-2.39	**2.40E-13**	-5.16	**4.55E-03**	0.63 (0.05)	Glycoside hydrolase domain containing protein [U6P060]	n/a	n/a	n/a
	HCOI01945600	2000.03	2.02	**2.33E-16**	2.39	**9.83E-04**	-0.64 (0.05)	Peptidase A1 domain containing protein [U6PP66]	pcl, Bace	n/a	n/a
	HCOI01623600	23.12	2.03	**2.05E-09**	4.21	**6.77E-03**	-0.79 (0.01)	n/a	n/a	n/a	n/a
	HCOI01283800	38840.11	2.15	**3.58E-15**	2.79	**1.28E-03**	-0.76 (0.01)	Peptidase C1A domain containing protein [U6P6R9]	CtsB1	n/a	n/a
	HCOI01549900	1104.78	2.20	**6.42E-16**	2.86	**1.31E-03**	-0.73 (0.02)	Protease inhibitor I4 domain containing protein [U6PNP0]	srp-1,2,3,6,7,8	n/a	n/a
	HCOI01736400	2678.92	2.49	**4.60E-31**	3.01	**7.91E-05**	-0.81 (0.004)	n/a	CtsB1	n/a	n/a
Vaccine Antigen	HCOI01993300	4049.71	0.30	3.09E-01	0.32	3.46E-01	n/a	Propeptide domain containing protein [U6PXI5]	n/a	pep-2	AJ577754.1
	HCOI01993500	13499.65	0.34	2.65E-01	0.35	3.06E-01	n/a	Propeptide and Peptidase A1 domain containing protein [U6PQD5]	n/a	pep-1	AF079402.1
	HCOI00348800	8859.39	0.47	**1.56E-02**	0.51	1.14E-01	n/a	Peptidase M13 domain containing protein [U6NMI3]	n/a	mep-2	AF080117.1
	HCOI01333400	9325.90	0.59	**3.88E-02**	0.62	1.64E-01	n/a	Peptidase M13 domain containing protein [U6PHP6]	nep-9, nep-20	mep-3	AF080172.1
	HCOI02032800	2207.13	0.71	**1.25E-02**	0.90	5.97E-02	n/a	Peptidase M1 domain containing protein [U6PYE0]	T07F10.1	h11	FJ481146.1
	HCOI00308300	18250.90	0.73	**4.82E-04**	0.78	5.85E-02	n/a	Peptidase M13 domain containing protein [U6NME0]		mep-1	AF102130.1
	HCOI00631000	5690.45	0.77	**2.40E-04**	0.81	5.97E-02	n/a			mep-4	AF132519.1
Housekeeping genes	HCOI00909100	5753.25	-0.41	5.29E-01	-0.60	3.93E-01	n/a	Nematode fatty acid retinoid binding domain containing protein [U6NYW0]	n/a	far	CDJ86885.1
	HCOI00117100	1379.12	0.08	9.64E-01	0.07	7.96E-01	n/a	Superoxide dismutase [Cu-Zn] [U6NGP5]	n/a	sod	CDJ80830.1
	HCOI01760600	24868.64	0.08	8.59E-01	0.08	7.92E-01	n/a	Glyceraldehyde-3-phosphate dehydrogenase (inferred by orthology to a human protein) [Source:UniProtKB;Acc:P04406]	n/a	gpd	CDJ92718.1
	HCOI01743600	194.02	0.13	9.13E-01	0.14	7.28E-01	n/a	RNA recognition motif domain containing protein [U6NLP1]	n/a	ncbp	CDJ82645.1
	HCOI01464300	974.31	0.32	3.24E-01	0.35	3.06E-01	n/a	DNA-directed RNA polymerase [U6PFA6]	n/a	ama	CDJ91461.1

454 Interestingly, 14 genes among the 52 DE gene set encoded peptidases or peptidase inhibitors
455 exemplified by the significant enrichment for serine-type (6 out of 78 annotated genes, p=9.6
456 x 10^{-8}) and cysteine-type peptidase (8 out of 98 annotated genes, p=2.8 x 10^{-10}) GO terms
457 (Table S5). This shift toward peptidase activity is also consistent with down-regulation of the
458 gamma interferon-inducible lysosomal thiol reductase (*GILT*, HCOI02049600), which is
459 known to catalyse the reduction of cysteine proteases (Table S5).
460 Higher expression of two genes involved in the anti-microbial response, the *Lys-8* coding
461 gene (HCOI00041100) associated with lysozyme formation, and the anti-microbial peptide
462 theromacin coding gene (HCOI00456500), was also found in worms surviving in vaccinated
463 animals. A proteinase inhibitor (HCOI01591500) and a prolyl-carboxypeptidase coding gene
464 (HCOI01624100) showing 99.6% similarity with contortin 2 (Genbank CAM84574.1,
465 BLASTP, e-value=0) also showed significantly greater expression in the V group (Table S5).

466

467 **Some of the transcriptomic differences could result from the vaccine selection pressure**

468 To reconcile both transcriptomic and genomic data together, average F_{ST} across the coding
469 sequence of the gene and neutrality test statistics were estimated for the set of differentially
470 expressed genes. Most of the differentially expressed genes were not associated with high F_{ST}
471 values (Fig S3A), although two protease coding genes, namely a peptidase C1A coding gene
472 (HCOI00658500) and a cathepsin-like protease (HCOI01736400), were found among the
473 25% most differentiated genes (high C-V F_{ST}) between C and V populations (Fig S3A). In
474 addition, a gene encoding a peptidase A1 (aspartic protease, HCOI01945600) showed a
475 significant contrast in Tajima's *D* estimate between the experimental groups with a pattern
476 consistent with selection occurring in the vaccine survivors (Fig S3B, S3C; p=0.03; D= 0.21
477 and -0.40 in C and V groups, respectively).

478

479 **Vaccine antigen coding genes are not differentially expressed between experimental**
480 **groups**

481 Importantly we found that most of the genes encoding the core components of the
482 Barbervax® vaccine (MEPs, PEPs, Aminopeptidases) were not significantly differentially
483 expressed between V and C worms or where significant, showed slight over-expression in the
484 V worm population (Table 2, Fig 6A). In addition, the SNPs located within or in the vicinity
485 of the vaccine antigen coding-genes, *i.e.* for pep-1, hmCP-1, hmCP-4 and cystatin coding
486 genes showed similar distribution within the two groups (Fig 6B) and thus showed neither
487 evidence for genetic differentiation among groups (Fig 6C; gene-wise F_{ST} estimates available
488 for mep-1 and hmCP-1 only, Table S3) nor for selection (Fig 6D).

489

490 **Fig 6. Expression level, allele frequencies, differentiation and neutrality test estimates**
491 **for the vaccine antigen coding genes**

492 Figure 6A shows the normalized transcript counts for every vaccine antigen coding gene.
493 Figure 6B displays allelic frequencies for every SNP found within vaccine antigen coding
494 genes.

495 Figure 6C demonstrates the relationship between gene-wise F_{ST} estimates and fold change in
496 expression between the two experimental groups. For the purpose of this plot and to unify the
497 color scale across F_{ST} and fold change values, absolute fold change values have been scaled
498 by 20. Figure 6D provides the dispersion of Tajima's D test for the set of vaccine antigen
499 coding genes.

500

Discussion

In comparison to the development of antimicrobial or anthelmintic resistance, vaccine resistance has rarely been reported [11]. These contrasting findings may relate both to the prophylactic use of vaccines, which prevent the spread of resistant mutants among hosts, and the multiplicity of pathways targeted by the host immune response following vaccination [11]. However, highly diverse populations, such as *H. contortus* likely encompass a wide range of genotypes that could be differentially selected, ultimately leading to vaccine resistance through replacement [55-57].

Resistance to all but the newest anthelmintic drugs is common and widespread amongst gastrointestinal nematode parasites of ruminants. Barbervax®, which is specific for *H. contortus*, is the only vaccine registered for a gut dwelling nematode of any host. While this vaccine provides a useful level of protection mediated mainly by reducing pasture contamination, a small proportion of worms do survive vaccination. Here, we investigated whether the genome and transcriptome of these survivors differed from control worms. Whilst this study focuses on a species of veterinary significance, our findings also have relevance to other species, such as the human hookworm *N. americanus*, as vaccines against both parasites target antigens involved in blood meal digestion [24, 29, 58].

Populations exposed to evolutionary forces, such as an effective vaccine, tend to differentiate from each other at the genomic level through changes in allele frequencies, either randomly (drift) or under the influence of environmental constraints favouring particular alleles (selection). The resulting variation in allele frequency among populations can be measured by F_{ST} statistics [50, 59], while departure from neutral mutation-drift equilibrium can be identified by positive or negative values of Tajima's *D* statistic. Our observations suggest that the sheep immune response did not impose a high level of selection pressure on the worm population, with differentiation largely driven by random genetic drift, as indicated by

Tajima's *D* estimates, which were not significantly different from 0. The low level of differentiation both amongst and within groups also support similar population histories between experimental groups. Analysis of the genes encoding the vaccine targets, against which one might have expected selection to operate, showed no departure from neutrality in either experimental group and the genetic differentiation measured for a H-Gal-GP fraction member that passed the quality filter of the analysis, *i.e.* MEP-3 (HCOI00308300), was both low and identical between groups. Several explanations can be proposed. Firstly, the sampling strategy associated with the experimental design may have induced extra drift in the control population therefore resulting in some levels of differentiation within control worm populations, as only ten to 20 individuals of the surviving worm populations per sheep were randomly sampled for sequencing. In addition, even if the overall population size of *H. contortus* is huge [30], it faces a high level of stochasticity at each step of its life cycle, *i.e.* sampling on pasture [60] and host-mediated bottleneck, as recent meta-analysis estimates suggested that 25% of ingested infective larvae ingested fail to establish [54]. However, simulation predictions that take into account both the life cycle and sampling design suggest that F_{ST} estimates based on neutral markers should be around 0, ruling out this hypothesis and perhaps indicating that control populations were actually under slight selection pressure mediated by the sheep immune system.

Furthermore, the short evolutionary time scale (a single generation of selection) may be insufficient to detect differential selection between control worms and those surviving the vaccine. Under this framework, our window filtering strategy should have resulted in 0.28 to 0.58 % of signals falsely declared differentiated, but a maximal power to detect true survival-associated markers of 2%. It is therefore difficult to disentangle drift from selection and the implementation of the "evolve and resequence" approach, in which vaccine selection is applied over multiple generations of evolution of a worm population, could favour the

551 identification of sites undergoing selection [61]. Controlled infection of sheep by *H.*
552 *contortus* followed by vaccination is a reasonably tractable experiment that could offer
553 additional insights and serve as a model for other host-nematode systems.
554 Predictions from the simulation results suggest that observed F_{ST} estimates would be
555 compatible with survival-associated alleles segregating at intermediate frequencies, *i.e.* 0.3. If
556 true, it would suggest that subsequent selection for these putative alleles with intermediate
557 frequencies should be relatively easy in the field, especially because *H. contortus* populations
558 are optimised for the spreading of beneficial alleles, *i.e.* high gene flow and genetic diversity
559 [30]. In a different setting, such as human hookworm, characterized by both sporadic gene
560 flow and fluctuating population size [62], evolution is expected to be mainly driven by
561 random sampling of genetic variants, rendering selection less efficient [63]. Such a moderate
562 allele frequency should however limit its random loss over time. This prediction still remains
563 to be validated after at least one genomic candidate is found and a critical assessment of its
564 genetic variability is performed across a wide range of field samples.
565 Notably, and perhaps counterintuitively, no evidence was found that increased expression of
566 vaccine targets could mediate survival after vaccine exposure. Indeed, both experimental
567 groups exhibited similar levels of vaccine antigen transcripts. In contrast one
568 metallopeptidase and one exopeptidase, belonging to the same functional families [64] as the
569 vaccine MEP (M13 peptidase) and H11 (M1 peptidase), were over-expressed in the vaccine
570 survivors. However, we have no direct evidence that these can compensate for the function of
571 the peptidases within the vaccine. Instead, survival following Barbervax® vaccination was
572 associated with the higher expression of a limited subset of genes, enriched in those coding
573 for peptidases, most of which were cysteine peptidases. Differential tuning of a GILT-like
574 gene, *i.e.* down-regulated in worms surviving the vaccine response, would also support
575 proteolytic function as an important feature for vaccine survival, as this pleiotropic gene is

known to modulate cysteine protease activity and stability [65]. In addition, there was an indication of higher selection pressure on a *lyst-1* orthologue, a regulator of endosomal trafficking in *C. elegans* polarized epithelial cells [66], that may share the same function in *H. contortus* and thus contribute to efficient processing of protein material from the intestinal lumen. This suggests that regulation of the proteolytic pathways in vaccine survivors results in improved survival. While the precise function of cysteine peptidases is hard to infer *in silico*, current knowledge from *in vitro* studies points to their role in the proteolytic cascade responsible for degrading haemoglobin or immunoglobulin G [16]. Perhaps worms that over-express these proteins may either better maintain blood coagulation and digestion, or they may degrade host IgG to evade the vaccine response, or some combination of both. Indeed the vaccine is proposed to disrupt digestion in the worm gut by inactivating the intestinal proteases it targets. Processing of ingested proteins by an alternative proteolytic pathway may improve the survival and/or fecundity of worms suffering dietary restriction. In addition, the over-expression of a myo-inositol-1 phosphate synthase in vaccine survivors may also support this theory as this gene is known to act on lipid storage [67] and in the defecation cycle [68], both critical in the digestion process, and hence impacting worm growth and lifespan.

Interestingly, the most highly differentially expressed genes show a low level of expression in worms from the control group, suggesting that the vaccine response induced their overexpression in the vaccine survivors or alternatively, that the vaccine selects for natural variation in expression of these genes. Additional transcriptomic evaluation of the offspring of each worm subpopulation, before and after vaccine exposure, would help confirm this observation and distinguish between a genetic effect on gene expression and a regulatory response to vaccine-induced immunity. It would also provide hints about the heritability of this trait.

While the differences in gene expression associated with worm survival in vaccinates indicate that the parasite population may have the potential to ameliorate the vaccine response, we were unable to identify significant genetic changes following a single round of immune selection. The human hookworm vaccine is composed of two antigens and it is probable that alternative pathways may help this species survive the vaccine response by similar mechanisms as described here. As the reported experimental vaccine challenges in dogs [28] resulted in lower reduction in worm burden (in the range of 30%), less selection pressure should be applied on these two targets. It would be interesting to determine whether the 46 orthologues identified in hookworms for our 52 differentially expressed candidates have any association with vaccine survival in this setting.

In conclusion, our data suggest that surviving parasite populations are able to optimize their proteolytic machinery, involving both peptidases and regulators of lysosome trafficking, and display better lipid storage or defecation abilities or both which may enhance survival in the face of a robust vaccine-induced immune response. However, genomic data indicate that the evolution of parasite populations is mostly driven by genetic drift rather than selection for vaccine resistance. Our findings provide encouragement for the future implementation of anti-helminth vaccines and could be extended to other nematode systems where vaccines are much needed. While our experiment, with a single generation of vaccine challenge, has limitations in power to detect selection due to the vaccine response, the absence of detectable selection in this setting is encouraging and suggests that vaccine survival will not evolve quickly. Further validation of these findings should be implemented in a more powerful experiment using an “evolve and resequencing” approach to contrast changes in allele frequencies in vaccinated and unvaccinated populations through time, across multiple generations of vaccine challenge.

Acknowledgements

We thank Stephen Doyle for advice and comments on the manuscript and the biological services staff at MRI for their expert animal care.

References

1. Hotez PJ, Strych U, Lustigman S, Bottazzi ME. Human anthelminthic vaccines: Rationale and challenges. Vaccine. 2016;34(30):3549-55. doi: 10.1016/j.vaccine.2016.03.112. PubMed PMID: 27171753.
2. Kaplan RM, Vidyashankar AN. An inconvenient truth: Global worming and anthelmintic resistance. Veterinary Parasitology. 2012;186(1-2):70-8. doi: 10.1016/j.vetpar.2011.11.048. PubMed PMID: WOS:000303183200010.
3. Rist CL, Garchitorena A, Ngonghala CN, Gillespie TR, Bonds MH. The Burden of Livestock Parasites on the Poor. Trends Parasitol. 2015;31(11):527-30. doi: 10.1016/j.pt.2015.09.005. PubMed PMID: 26604161.
4. McKellar QA, Jackson F. Veterinary anthelmintics: old and new. Trends Parasitol. 2004;20(10):456-61. doi: 10.1016/j.pt.2004.08.002. PubMed PMID: 15363438.
5. Clarke NE, Clements AC, Doi SA, Wang D, Campbell SJ, Gray D, et al. Differential effect of mass deworming and targeted deworming for soil-transmitted helminth control in children: a systematic review and meta-analysis. Lancet. 2017;389(10066):287-97. doi: 10.1016/S0140-6736(16)32123-7. PubMed PMID: 27979381.
6. Keiser J, Utzinger J. Efficacy of current drugs against soil-transmitted helminth infections: systematic review and meta-analysis. JAMA. 2008;299(16):1937-48. doi: 10.1001/jama.299.16.1937. PubMed PMID: 18430913.
7. Soukhathammavong PA, Sayasone S, Phongluxa K, Xayaseng V, Utzinger J, Vounatsou P, et al. Low efficacy of single-dose albendazole and mebendazole against hookworm and effect on concomitant helminth infection in Lao PDR. PLoS Negl Trop Dis. 2012;6(1):e1417. doi: 10.1371/journal.pntd.0001417. PubMed PMID: 22235353; PubMed Central PMCID: PMCPMC3250499.
8. Geary TG, Conder GA, Bishop B. The changing landscape of antiparasitic drug discovery for veterinary medicine. Trends Parasitol. 2004;20(10):449-55. PubMed PMID: WOS:000224257600001.
9. Hewitson JP, Maizels RM. Vaccination against helminth parasite infections. Expert Rev Vaccines. 2014;13(4):473-87. doi: 10.1586/14760584.2014.893195. PubMed PMID: 24606541.
10. Lee BY, Bacon KM, Bailey R, Wiringa AE, Smith KJ. The potential economic value of a hookworm vaccine. Vaccine. 2011;29(6):1201-10. doi: 10.1016/j.vaccine.2010.12.004. PubMed PMID: 21167860; PubMed Central PMCID: PMCPMC3038553.
11. Kennedy DA, Read AF. Why does drug resistance readily evolve but vaccine resistance does not? Proc Biol Sci. 2017;284(1851). doi: 10.1098/rspb.2016.2562. PubMed PMID: 28356449.
12. Bassetto CC, Amarante AF. Vaccination of sheep and cattle against haemonchosis. J Helminthol. 2015;89(5):517-25. doi: 10.1017/S0022149X15000279. PubMed PMID: 25891536.

13. Kearney PE, Murray PJ, Hoy JM, Hohenhaus M, Kotze A. The 'Toolbox' of strategies for managing Haemonchus contortus in goats: What's in and what's out. Vet Parasitol. 2016;220:93-107. doi: 10.1016/j.vetpar.2016.02.028. PubMed PMID: 26995728.

14. McKeand JB. Vaccine development and diagnostics of Dictyocaulus viviparus. Parasitology. 2000;120 Suppl:S17-23. PubMed PMID: 10874707.

15. Brelsford JB, Plieskatt JL, Yakovleva A, Jariwala A, Keegan BP, Peng J, et al. Advances in neglected tropical disease vaccines: Developing relative potency and functional assays for the Na-GST-1/Alhydrogel hookworm vaccine. PLoS Negl Trop Dis. 2017;11(2):e0005385. doi: 10.1371/journal.pntd.0005385. PubMed PMID: 28192438; PubMed Central PMCID: PMCPMC5325600.

16. Williamson AL, Brindley PJ, Knox DP, Hotez PJ, Loukas A. Digestive proteases of blood-feeding nematodes. Trends Parasitol. 2003;19(9):417-23. PubMed PMID: 12957519.

17. Laing R, Kikuchi T, Martinelli A, Tsai IJ, Beech RN, Redman E, et al. The genome and transcriptome of Haemonchus contortus, a key model parasite for drug and vaccine discovery. Genome Biol. 2013;14(8):R88. doi: 10.1186/gb-2013-14-8-r88. PubMed PMID: 23985316; PubMed Central PMCID: PMC4054779.

18. Schwarz EM, Korhonen PK, Campbell BE, Young ND, Jex AR, Jabbar A, et al. The genome and developmental transcriptome of the strongylid nematode Haemonchus contortus. Genome Biol. 2013;14(8):R89. doi: 10.1186/gb-2013-14-8-r89. PubMed PMID: 23985341; PubMed Central PMCID: PMC4053716.

19. Tang YT, Gao X, Rosa BA, Abubucker S, Hallsworth-Pepin K, Martin J, et al. Genome of the human hookworm Necator americanus. Nat Genet. 2014;46(3):261-9. doi: 10.1038/ng.2875. PubMed PMID: 24441737; PubMed Central PMCID: PMCPMC3978129.

20. Smith WD, Pettit D, Smith SK. Cross-protection studies with gut membrane glycoprolein antigens from Haemonchus contortus and Teladorsagia circumcincta. Parasite Immunology. 2001;23(4):203-11. doi: DOI 10.1046/j.1365-3024.2001.00375.x. PubMed PMID: WOS:000168049000005.

21. Munn EA, Smith TS, Smith H, James FM, Smith FC, Andrews SJ. Vaccination against Haemonchus contortus with denatured forms of the protective antigen H11. Parasite Immunol. 1997;19(6):243-8. PubMed PMID: 9364553.

22. Roberts B, Antonopoulos A, Haslam SM, Dicker AJ, McNeilly TN, Johnston SL, et al. Novel expression of Haemonchus contortus vaccine candidate aminopeptidase H11 using the free-living nematode Caenorhabditis elegans. Vet Res. 2013;44:111. doi: 10.1186/1297-9716-44-111. PubMed PMID: 24289031; PubMed Central PMCID: PMCPMC4176091.

23. Smith WD, Skuce PJ, Newlands GF, Smith SK, Pettit D. Aspartyl proteases from the intestinal brush border of Haemonchus contortus as protective antigens for sheep. Parasite Immunol. 2003;25(11-12):521-30. doi: 10.1111/j.0141-9838.2004.00667.x. PubMed PMID: 15053773.

24. Knox DP, Redmond DL, Newlands GF, Skuce PJ, Pettit D, Smith WD. The nature and prospects for gut membrane proteins as vaccine candidates for Haemonchus contortus and other ruminant trichostrongyloids. Int J Parasitol. 2003;33(11):1129-37. PubMed PMID: 13678629.

25. Newton SE, Munn EA. The development of vaccines against gastrointestinal nematode parasites, particularly Haemonchus contortus. Parasitol Today. 1999;15(3):116-22. PubMed PMID: 10322325.

26. LeJambre LF, Windon RG, Smith WD. Vaccination against Haemonchus contortus: performance of native parasite gut membrane glycoproteins in Merino lambs grazing contaminated pasture. Vet Parasitol. 2008;153(3-4):302-12. doi: 10.1016/j.vetpar.2008.01.032. PubMed PMID: 18337013.

27. Knox DP, Smith SK, Redmond DL, Smith WD. Protection induced by vaccinating sheep with a thiol-binding extract of Haemonchus contortus membranes is associated with its protease components. Parasite Immunol. 2005;27(4):121-6. doi: 10.1111/j.1365-3024.2005.00750.x. PubMed PMID: 15910420.

716 28. Loukas A, Bethony JM, Mendez S, Fujiwara RT, Goud GN, Ranjit N, et al. Vaccination with
717 recombinant aspartic hemoglobinase reduces parasite load and blood loss after hookworm infection
718 in dogs. PLoS Med. 2005;2(10):e295. doi: 10.1371/journal.pmed.0020295. PubMed PMID: 16231975;
719 PubMed Central PMCID: PMCPMC1240050.
720 29. Zhan B, Perally S, Brophy PM, Xue J, Goud G, Liu S, et al. Molecular cloning, biochemical
721 characterization, and partial protective immunity of the heme-binding glutathione S-transferases
722 from the human hookworm Necator americanus. Infect Immun. 2010;78(4):1552-63. doi:
723 10.1128/IAI.00848-09. PubMed PMID: 20145100; PubMed Central PMCID: PMCPMC2849424.
724 30. Gilleard JS, Redman E. Genetic Diversity and Population Structure of Haemonchus contortus.
725 Adv Parasitol. 2016;93:31-68. doi: 10.1016/bs.apar.2016.02.009. PubMed PMID: 27238002.
726 31. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after
727 pneumococcal vaccination in the United States. PLoS Pathog. 2007;3(11):e168. doi:
728 10.1371/journal.ppat.0030168. PubMed PMID: 18020702; PubMed Central PMCID:
729 PMCPMC2077903.
730 32. Lecova L, Ruzickova M, Laing R, Vogel H, Szotakova B, Prchal L, et al. Reliable reference gene
731 selection for quantitative real time PCR in Haemonchus contortus. Mol Biochem Parasitol.
732 2015;201(2):123-7. doi: 10.1016/j.molbiopara.2015.08.001. PubMed PMID: 26255779.
733 33. Otto TD, Dillon GP, Degrave WS, Berriman M. RATT: Rapid Annotation Transfer Tool. Nucleic
734 Acids Res. 2011;39(9):e57. doi: 10.1093/nar/gkq1268. PubMed PMID: 21306991; PubMed Central
735 PMCID: PMCPMC3089447.
736 34. Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding
737 in eukaryotes. Nucleic Acids Res. 2004;32(Web Server issue):W309-12. doi: 10.1093/nar/gkh379.
738 PubMed PMID: 15215400; PubMed Central PMCID: PMCPMC441517.
739 35. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
740 2012;9(4):357-9. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286; PubMed Central PMCID:
741 PMCPMC3322381.
742 36. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq.
743 Bioinformatics. 2009;25(9):1105-11. doi: 10.1093/bioinformatics/btp120. PubMed PMID: 19289445;
744 PubMed Central PMCID: PMCPMC2672628.
745 37. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput
746 sequencing data. Bioinformatics. 2015;31(2):166-9. doi: 10.1093/bioinformatics/btu638. PubMed
747 PMID: 25260700; PubMed Central PMCID: PMCPMC4287950.
748 38. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
749 seq data with DESeq2. Genome Biol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PubMed
750 PMID: 25516281; PubMed Central PMCID: PMCPMC4302049.
751 39. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis
752 tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29. doi: 10.1186/gb-2014-15-2-r29.
753 PubMed PMID: 24485249; PubMed Central PMCID: PMCPMC4053721.
754 40. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R
755 Foundation for Statistical Computing; 2016.
756 41. Alexa A, Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology. 2016.
757 42. Smith WD, Smith SK, Pettit D, Newlands GF, Skuce PJ. Relative protective properties of three
758 membrane glycoprotein fractions from Haemonchus contortus. Parasite Immunol. 2000;22(2):63-71.
759 PubMed PMID: 10652118.
760 43. Britton C, Redmond DL, Knox DP, McKerrow JH, Barry JD. Identification of promoter
761 elements of parasite nematode genes in transgenic Caenorhabditis elegans. Mol Biochem Parasitol.
762 1999;103(2):171-81. PubMed PMID: 10551361.
763 44. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite - a comprehensive
764 resource for helminth genomics. Mol Biochem Parasitol. 2016. doi:
765 10.1016/j.molbiopara.2016.11.005. PubMed PMID: 27899279.

766 45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
767 Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. doi:
768 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID:
769 PMCPMC2723002.
770 46. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for
771 variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet.
772 2011;43(5):491-8. doi: 10.1038/ng.806. PubMed PMID: 21478889; PubMed Central PMCID:
773 PMCPMC3083463.
774 47. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome
775 Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.
776 Genome Res. 2010;20(9):1297-303. doi: 10.1101/gr.107524.110. PubMed PMID: 20644199; PubMed
777 Central PMCID: PMCPMC2928508.
778 48. Kofler R, Pandey RV, Schlotterer C. PoPoolation2: identifying differentiation between
779 populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011;27(24):3435-
780 6. doi: 10.1093/bioinformatics/btr589. PubMed PMID: 22025480; PubMed Central PMCID:
781 PMC3232374.
782 49. Lynch M, Bost D, Wilson S, Maruki T, Harrison S. Population-genetic inference from pooled-
783 sequencing data. Genome Biol Evol. 2014;6(5):1210-8. doi: 10.1093/gbe/evu085. PubMed PMID:
784 24787620; PubMed Central PMCID: PMCPMC4040993.
785 50. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining,
786 estimating and interpreting F(ST). Nat Rev Genet. 2009;10(9):639-50. doi: 10.1038/nrg2611. PubMed
787 PMID: 19687804; PubMed Central PMCID: PMCPMC4687486.
788 51. Kofler R, Schlotterer C. Gowinda: unbiased analysis of gene set enrichment for genome-wide
789 association studies. Bioinformatics. 2012;28(15):2084-5. doi: 10.1093/bioinformatics/bts315.
790 PubMed PMID: 22635606; PubMed Central PMCID: PMCPMC3400962.
791 52. Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a
792 short-term genetic distance. Genetics. 1983;105(3):767-79. PubMed PMID: 17246175; PubMed
793 Central PMCID: PMCPMC1202185.
794 53. Peng B, Kimmel M. simuPOP: a forward-time population genetics simulation environment.
795 Bioinformatics. 2005;21(18):3686-7. doi: 10.1093/bioinformatics/bti584. PubMed PMID: 16020469.
796 54. Saccareau M, Sallé G., Robert-Granié C., Duchemin, T., Jacquiet P., Blanchard, A., Cabaret, J.,
797 Moreno, C.R. Meta-analysis of the parasitic phase traits of Haemonchus contortus infection in sheep.
798 Parasites & Vectors. 2017;In press.
799 55. Barnett TC, Lim JY, Soderholm AT, Rivera-Hernandez T, West NP, Walker MJ. Host-pathogen
800 interaction during bacterial vaccination. Curr Opin Immunol. 2015;36:1-7. doi:
801 10.1016/j.coi.2015.04.002. PubMed PMID: 25966310.
802 56. Martcheva M, Bolker BM, Holt RD. Vaccine-induced pathogen strain replacement: what are
803 the mechanisms? J R Soc Interface. 2008;5(18):3-13. doi: 10.1098/rsif.2007.0236. PubMed PMID:
804 17459810; PubMed Central PMCID: PMCPMC2405901.
805 57. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal
806 vaccination. Lancet. 2011;378(9807):1962-73. doi: 10.1016/S0140-6736(10)62225-8. PubMed PMID:
807 21492929; PubMed Central PMCID: PMCPMC3256741.
808 58. Ranjit N, Zhan B, Hamilton B, Stenzel D, Lowther J, Pearson M, et al. Proteolytic degradation
809 of hemoglobin in the intestine of the human hookworm Necator americanus. J Infect Dis.
810 2009;199(6):904-12. PubMed PMID: 19434933.
811 59. Wright S. The genetical structure of populations. Ann Eugen. 1951;15(4):323-54. PubMed
812 PMID: 24540312.
813 60. Cabaret J, Mangeon N, Gruner L. Estimation of uptake of digestive tract strongyle larvae
814 from pasture, using oesophagus fistulated sheep. Vet Parasitol. 1986;19(3-4):315-20. PubMed PMID:
815 3705424.

816 61. Turner TL, Miller PM. Investigating natural variation in Drosophila courtship song by the
817 evolve and resequence approach. Genetics. 2012;191(2):633-42. doi: 10.1534/genetics.112.139337.
818 PubMed PMID: 22466043; PubMed Central PMCID: PMCPMC3374323.
819 62. Hawdon JM, Li T, Zhan B, Blouin MS. Genetic structure of populations of the human
820 hookworm, Necator americanus, in China. Mol Ecol. 2001;10(6):1433-7. PubMed PMID: 11412366.
821 63. Charlesworth B. Fundamental concepts in genetics: effective population size and patterns of
822 molecular evolution and variation. Nat Rev Genet. 2009;10(3):195-205. doi: 10.1038/nrg2526.
823 PubMed PMID: 19204717.
824 64. Rawlings ND, Barrett AJ, Bateman A. MEROPS: the peptidase database. Nucleic Acids Res.
825 2010;38(Database issue):D227-33. doi: 10.1093/nar/gkp971. PubMed PMID: 19892822; PubMed
826 Central PMCID: PMCPMC2808883.
827 65. Rausch MP, Hastings KT. Diverse cellular and organismal functions of the lysosomal thiol
828 reductase GILT. Mol Immunol. 2015;68(2 Pt A):124-8. doi: 10.1016/j.molimm.2015.06.008. PubMed
829 PMID: 26116226; PubMed Central PMCID: PMCPMC4623965.
830 66. de Souza N, Vallier LG, Fares H, Greenwald I. SEL-2, the C. elegans neurobeachin/LRBA
831 homolog, is a negative regulator of lin-12/Notch activity and affects endosomal traffic in polarized
832 epithelial cells. Development. 2007;134(4):691-702. doi: 10.1242/dev.02767. PubMed PMID:
833 17215302.
834 67. Ashrafi K, Chang FY, Watts JL, Fraser AG, Kamath RS, Ahringer J, et al. Genome-wide RNAi
835 analysis of Caenorhabditis elegans fat regulatory genes. Nature. 2003;421(6920):268-72. doi:
836 10.1038/nature01279. PubMed PMID: 12529643.
837 68. Tokuoka SM, Saiardi A, Nurrish SJ. The mood stabilizer valproate inhibits both inositol- and
838 diacylglycerol-signaling pathways in Caenorhabditis elegans. Mol Biol Cell. 2008;19(5):2241-50. doi:
839 10.1091/mbc.E07-09-0982. PubMed PMID: 18287529; PubMed Central PMCID: PMCPMC2366867.

840

841 # Supporting information

842 **S1 Fig. Principal component analysis (PCA) of transcript counts**

843 Pools of worn (worm?) coordinates were plotted against first two components of transcript
844 counts variance. First PCA axis explains 36% of total variance and relates to differences
845 between the two considered experimental groups, *i.e.* worms exposed to the vaccine response
846 (V) or the control group (C).

847

848 **Fig S2. Number of differentially expressed genes found by each of the two implemented**
849 **methods**

850 Total number of significantly differentially expressed genes found by at least one of the two
851 methods (DESeq2, VOOM, or both) are plotted according to their regulation pattern, *i.e* up or

852 down-regulated in the vaccine survivors, to their estimated fold change, i.e. log2FC > 2, 1 or
853 0.

854

855 **Fig S3. Differentiation and neutrality test estimates for the differentially expressed**
856 **genes**

857 Figure S3A shows the relationship between gene-wise F_{ST} estimates and fold change in
858 expression between the two experimental groups. For the purpose of this plot and to unify the
859 color scale across F_{ST} and fold change values, absolute fold change values have been scaled
860 by 20. Figure S3B provides the dispersion of neutrality test statistics, namely pi, theta and
861 Tajima's D for the set of differentially expressed genes. Fold change absolute value
862 correlated with the dot size while its magnitude is indicated by colors ranging from green
863 (down-regulated in vaccine survivors) to red (over-expressed in vaccine survivors).

864

865 **S1 Table. List of primer sequences used for qPCR validation**

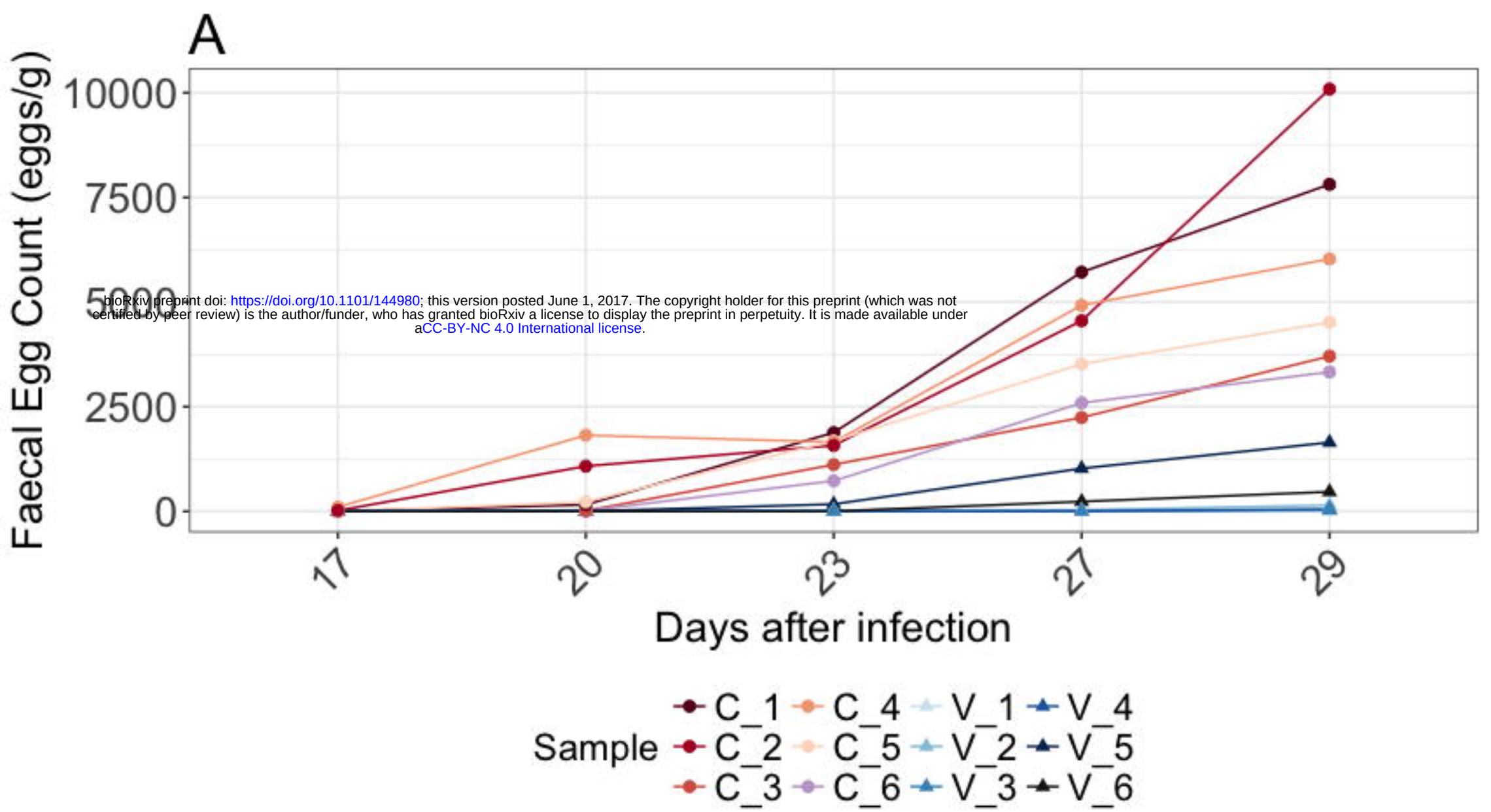
866 **S2 Table. Faecal egg count and worm volumes recovered at necropsy**

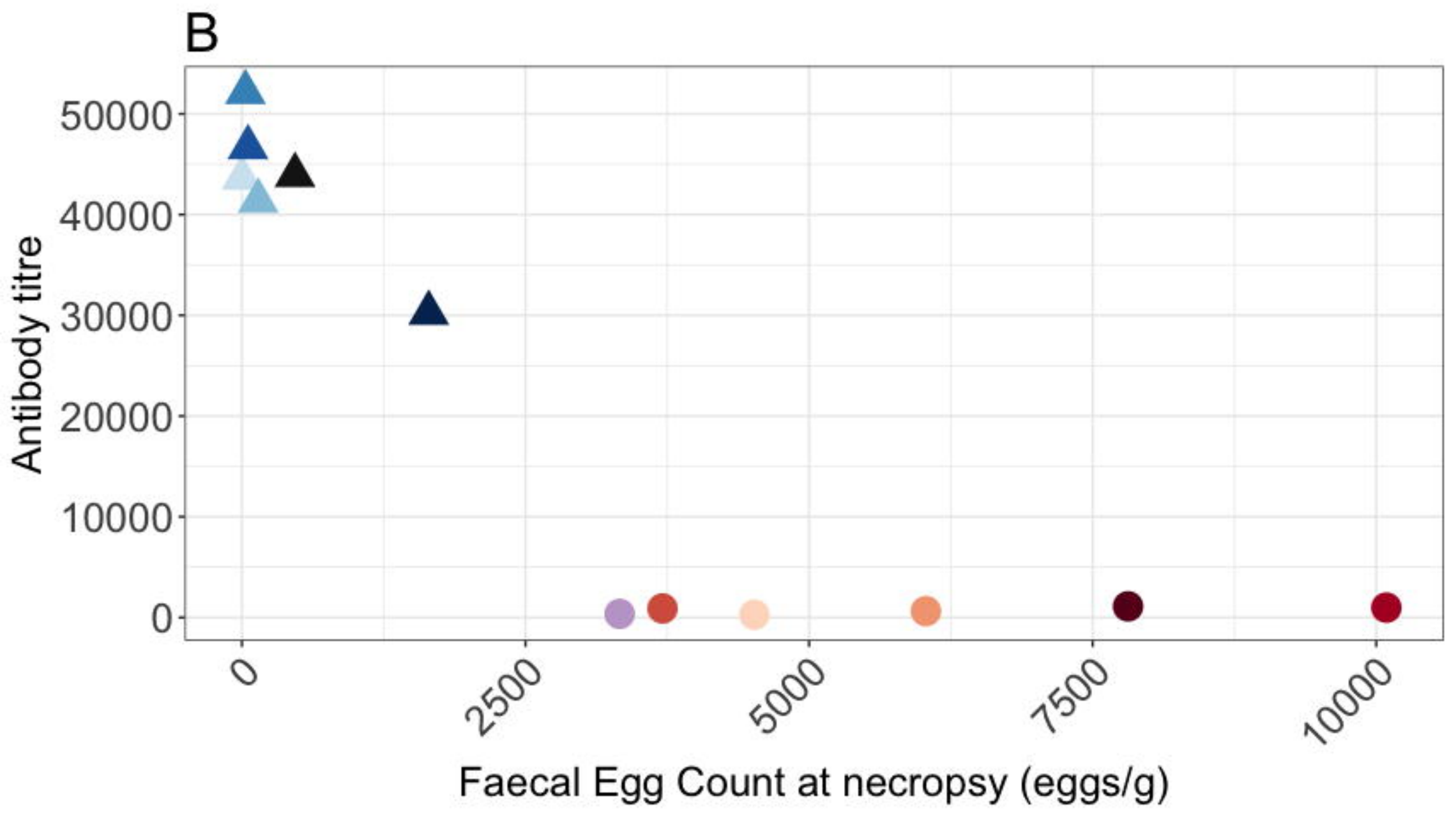
867 **S3 Table. Significantly enriched GO terms associated with the 145 differentiated**
868 **windows**

869 **S4 Table. Gene-wise average pair-wise F_{ST} estimates**

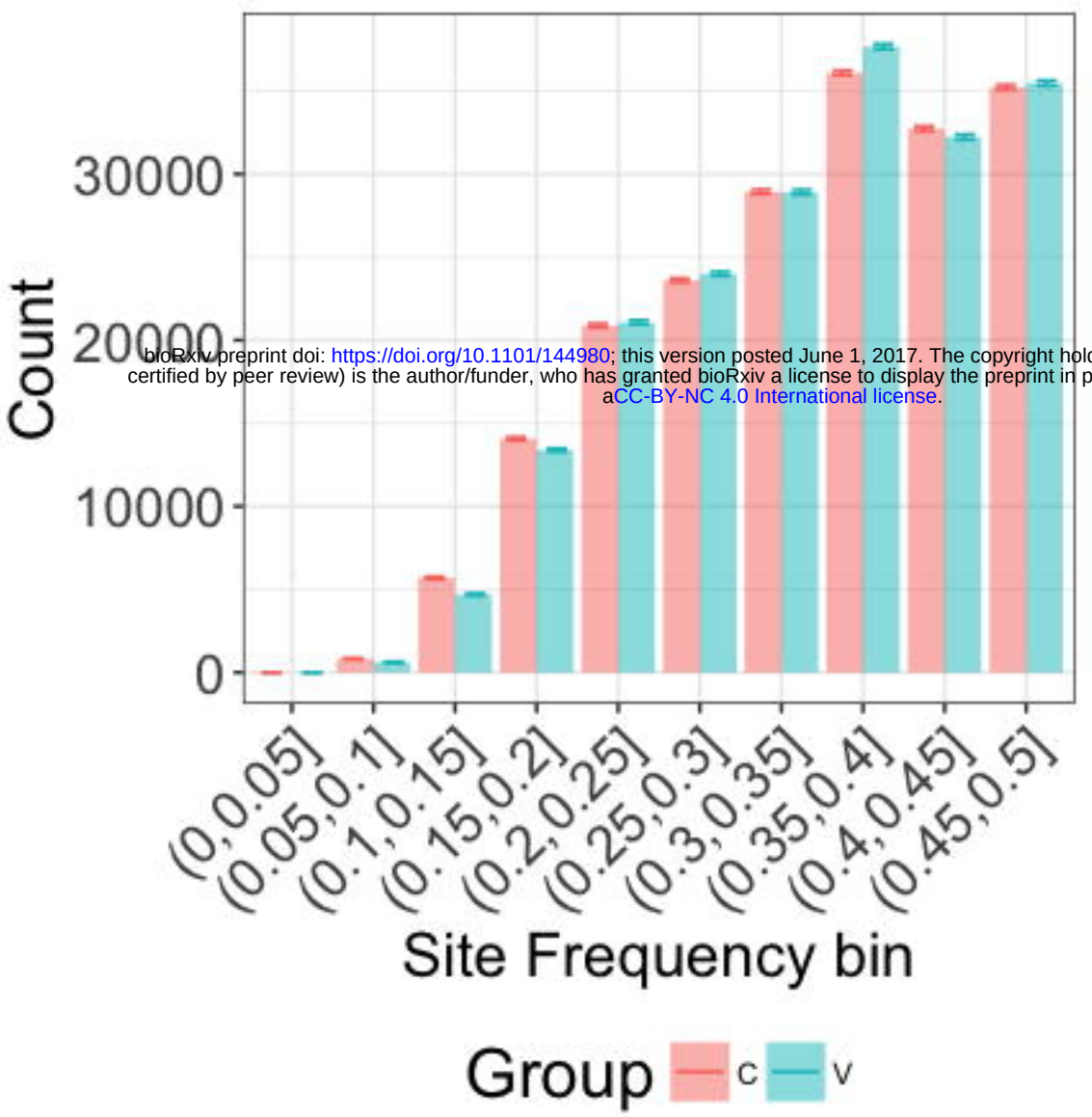
870 **S5 Table. Complete list of differentially expressed genes**

871

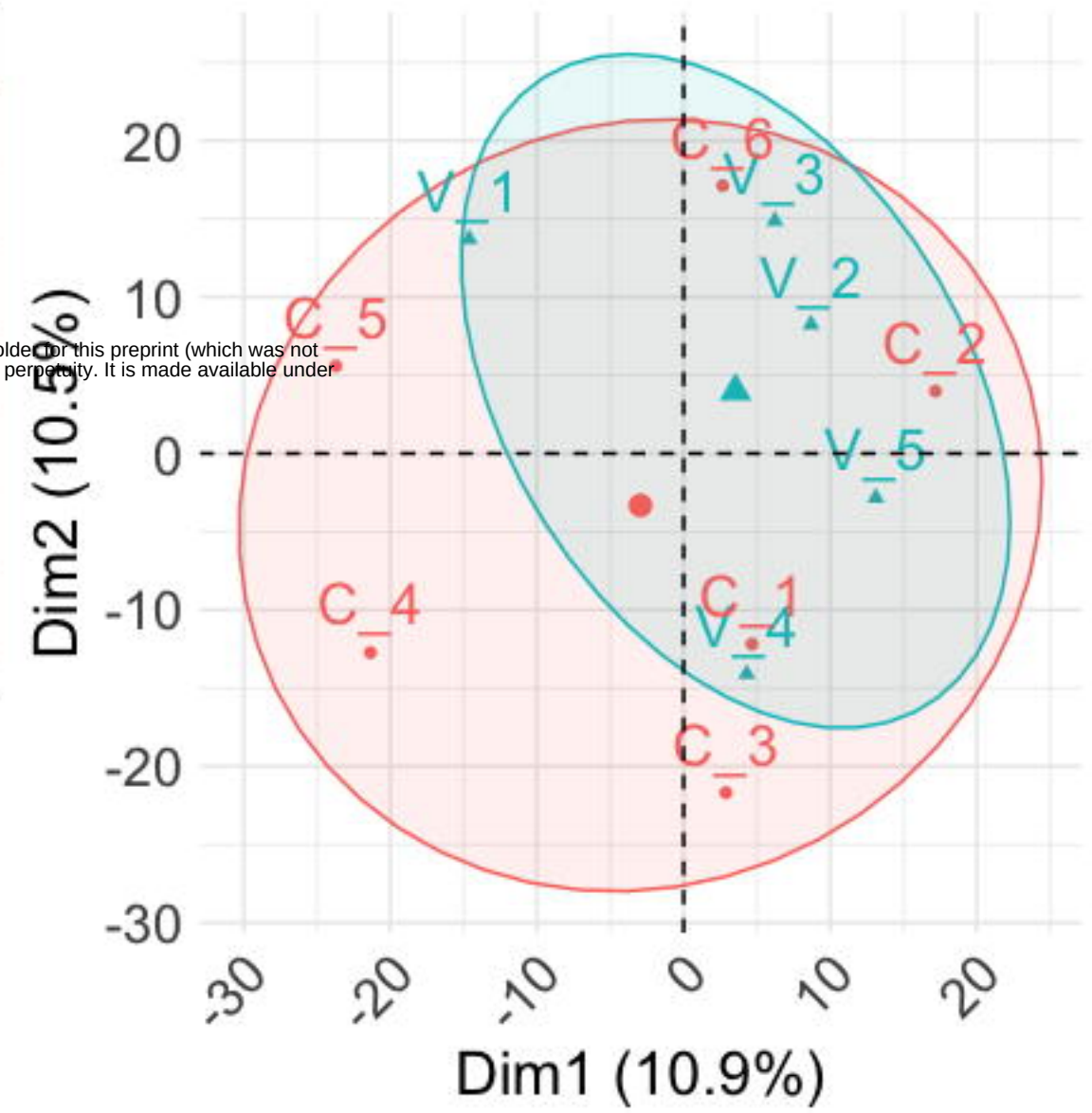




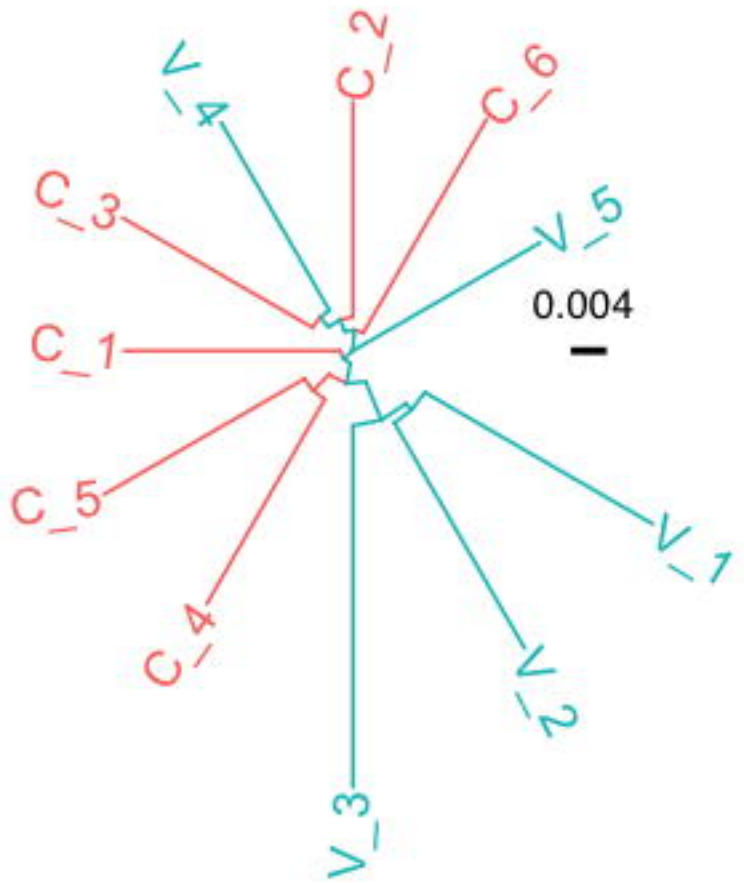
A



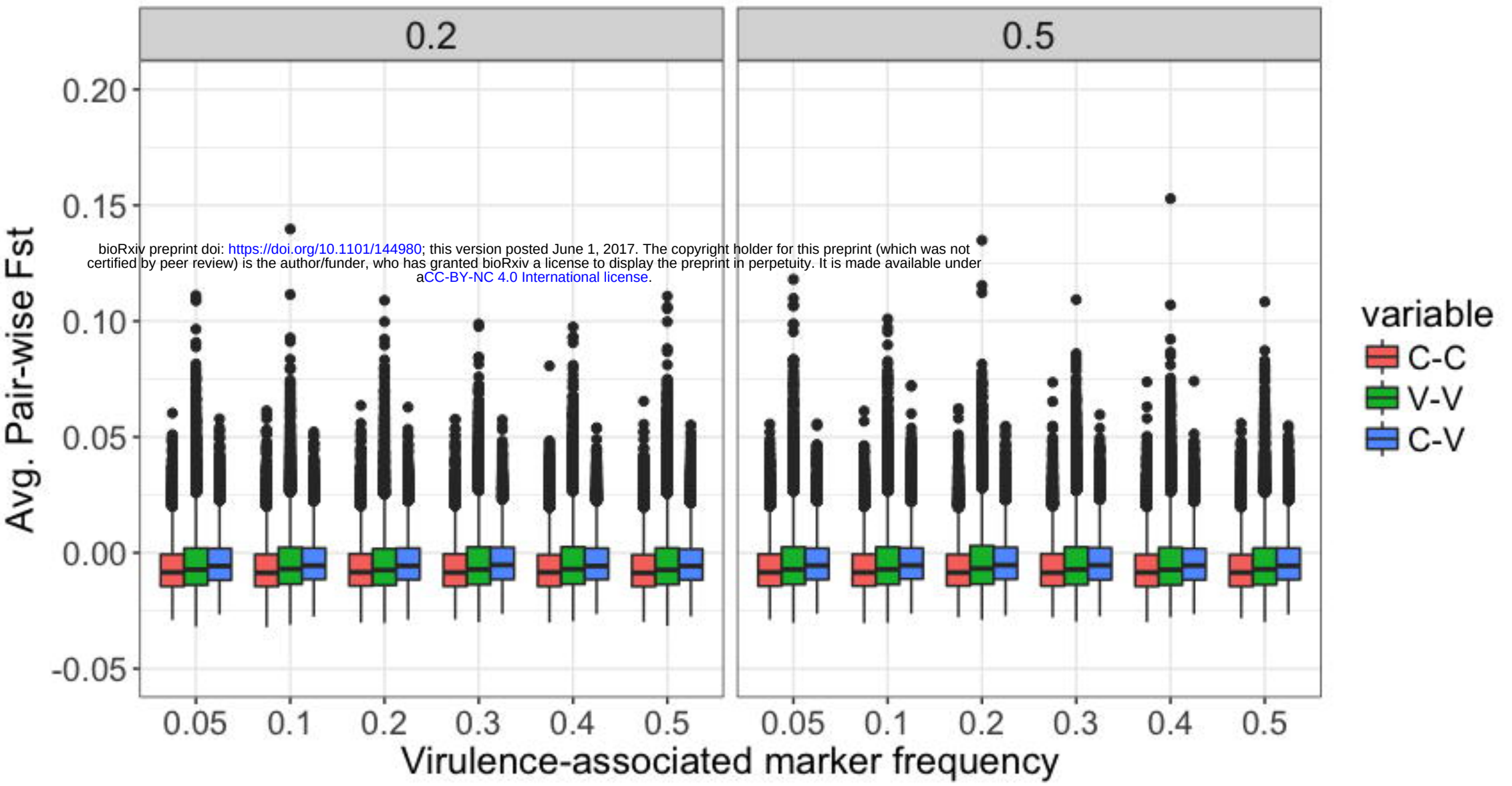
B



C



A



B