Effect of heritable symbionts on maternally-derived embryo transcripts

Mariana Mateos¹, Nadisha O. Silva¹,², Paulino Ramirez¹, Victor M. Higareda-Alvear³, Rodolfo Aramayo² & James W. Erickson²

Maternally-transmitted endosymbiotic bacteria are ubiquitous in insects. Among other influential phenotypes, many heritable symbionts of arthropods are notorious for manipulating host reproduction through one of four reproductive syndromes, which are generally exerted during early developmental stages of the host: male feminization; parthenogenesis induction; male killing; and cytoplasmic incompatibility (CI). Major advances have been achieved in understanding mechanisms and identifying symbiont factors involved in reproductive manipulation, particularly male killing and cytoplasmic incompatibility. Nonetheless, whether cytoplasmically-transmitted bacteria influence the maternally-loaded components of the egg or early embryo has not been examined. In the present study, we investigated whether heritable endosymbionts that cause different reproductive phenotypes in Drosophila melanogaster influence the mRNA transcriptome of early embryos. We used mRNA-seq to evaluate differential expression in Drosophila embryos lacking endosymbionts (control) to those harbouring the male-killing Spiroplasma poulsonii strain MSRO-Br, the CI-inducing Wolbachia strain wMel, or Spiroplasma poulsonii strain Hyd1; a strain that lacks a reproductive phenotype and is naturally associated with Drosophila hydei. We found no consistent evidence of influence of symbiont on mRNA composition of early embryos, suggesting that the reproductive manipulation mechanism does not involve alteration of maternally-loaded transcripts. In addition, we capitalized on several available mRNA-seq datasets derived from Spiroplasma-infected Drosophila melanogaster embryos, to search for signals of depurination of rRNA, consistent with the activity of Ribosome Inactivating Proteins (RIPs) encoded by Spiroplasma poulsonii. We found small but statistically significant signals of depurination of Drosophila rRNA in the Spiroplasma treatments (both strains), but not in the symbiont-free control or Wolbachia treatment, consistent with the action of RIPs. The depurination signal was slightly stronger in the treatment with the male-killing strain. This result supports a recent report that RIP-induced damage contributes to male embryo death.

Heritable associations between arthropods and endosymbiotic bacteria are widespread and influential to their hosts¹ and communities². With few exceptions³, inheritance is achieved via the mother, generating an asymmetry in the interests of the symbiont regarding the host’s sex, whereby males are generally dead-ends for the symbiont. Consistent with this inequality, numerous heritable bacteria manipulate host reproduction in favour of symbiont-bearing females. Four such reproductive phenotypes have been described⁴. Feminization occurs when genetic males develop and function as females. Parthenogenesis induction occurs in haplo-diploid systems, where unfertilized eggs, which would otherwise develop as males, develop instead into reproductively functional females. In male-killing (or son-killing), the symbiont eliminates infected males to the presumed advantage of surviving infected female siblings⁵. In cytoplasmic incompatibility (CI)⁶, matings between symbiont-infected males and uninfected females result in death of offspring at the embryonic stage. The CI mechanism involves symbiont-mediated damage to the male sperm that is rescued in the presence of a compatible symbiont strain in the egg⁷.

¹Department Wildlife and Fisheries Sciences, Texas A&M University, College Station, TX, USA. ²Department Biology, Texas A&M University, College Station, TX, USA. ³Centro de Ciencias Genómicas (CCG), Universidad Nacional Autónoma de México (UNAM), Cuernavaca, Morelos, Mexico. Correspondence and requests for materials should be addressed to M.M. (email: mmateos@tamu.edu)
Most maternally inherited symbionts are transmitted to a new host through the egg cytoplasm. As such, they may manipulate the female host’s reproductive system or usurp her cellular machinery in order to invade developing oocytes. Infection-induced changes during oogenesis may thus have effects on the composition of an egg (e.g. Wolbachia reduces the maternal transmission of the gypsy endogenous retrovirus). Drosophila melanogaster eggs contain maternal RNAs that are exclusively expressed during early development (prior to embryonic stage 5 or ~2h after egg deposition; AED). Zygotic transcription is silent during this period and therefore maternal mRNAs play a crucial role in early embryonic development. The egg, in a sense, is a point of convergence between the existing host, symbiont, and new host, and consequently could undergo symbiont-induced changes that could lay the foundation for the occupation of the symbiont within the new host. Furthermore, it is possible that infection by a reproductive parasite could cause changes in maternally-derived components that are necessary to induce a reproductive phenotype. As hosts of two independent lineages of maternally transmitted bacteria, Wolbachia and Spiroplasma, members of the genus Drosophila, have emerged as a model system for heritable symbiosis e.g. 38–34.

The wall-less bacterial genus Spiroplasma (class Mollicutes) is generally associated with arthropods and plants, and can reside intra- and extra-cellularly. The nature of Spiroplasma-host associations ranges from pathogenic to mutualistic, although the fitness consequences of the majority of Spiroplasma strains remain unknown. Most strains of Spiroplasma known to date appear to transmit horizontally from the environment or via a vector (e.g. several insect-vectored plant pathogens). A few strains of Spiroplasma, however, are maternally-transmitted by their arthropod hosts. Among these, ~20 species in the genus Drosophila are reported to harbour Spiroplasma35,23,28. A subset of maternally-transmitted Spiroplasma of Drosophila and other arthropods are male killers. All of the Drosophila-associated Spiroplasma male-killing strains that have been genetically characterized to date fall within the Spalsonii clade; one of the four Spalsonii clades that independently invaded the genus Drosophila. The poulsonii clade also contains non-male-killing strains such as “Hyd1” and “Neo”; naturally-occurring defensive mutualists of Drosophila hydei and Drosophila neotestacea, respectively25,37, as well as several strains that lost the ability to kill males in the lab15,28,29.

The mechanism by which Spiroplasma exerts death of Drosophila male embryos is not fully understood, but several aspects have been elucidated32,15,30–35. First, a functional dosage compensation complex (DCC; also known as the male-specific lethal complex) is required, as mutants of components of this complex that are infected with male-killing Spiroplasma, fail to express male-killing. Secondly, the DCC, which normally acetylates X chromatin in males, mis-localizes to other regions of the nucleus immediately prior to the killing stage. This phenomenon is accompanied by inappropriate histone acetylation, and genome-wide transcriptional misregulation. Thirdly, abnormal massive apoptosis and neural disorganization occurs30,31. The male X chromosome exhibits signs of DNA damage (chromatin bridges and segregation defects), and bridge breakage triggers sex-specific abnormal apoptosis via p53-dependent pathways. Recently Harumoto and Lemaitre identified spaid, a Spiroplasma-encoded gene that appears to be responsible for male killing. Overexpression of spaid in D. melanogaster causes death of males, but not females, and induces massive apoptosis and neural defects, reminiscent of the Spiroplasma-induced male-killing phenomenon in pattern, but in a somewhat delayed fashion (i.e., developmental arrest at embryonic stage 12–13 in Spiroplasma-infected wild-type D. melanogaster vs. at the second larval instar with expression of the spaid transeqe in Spiroplasma-free D. melanogaster). Spaid contains an OTU (ovarian tumor) deubiquitinase domain and ankyrin repeats (ANK). Harumoto and Lemaitre propose that the OTU domain promotes nuclear localization of Spaid (in both female and male embryos), while the ankyrin repeats interact with DCC complex itself or with its associated histone modifications. How this leads to DNA damage and segregation defects of the male X chromosome, as well as to other phenotypes associated with male-killing described above, and which are the host cellular targets of Spaid remain unknown.

In addition to Spaid, a different type of toxin has been reported in Spiroplasma. Ribosome inactivating proteins (RIPs) are plant- (e.g. ricin and saporin) and bacteria-encoded (e.g. Shiga toxins/otoxin) enzymes that cleave an adenine base (hereafter “depurinate”) from a specific position within a motif of the 28S rRNA that is universally conserved across eukaryotes. This motif is known as the sarcin–ricin loop (SRL). Depurination of this site renders the ribosome incapable of protein synthesis36. Several confirmed and predicted RIPs are encoded in the genomes of Spiroplasma strains that associate with Drosophila, including the male-killing strain native to D. melanogaster (MSRO) and the closely-related non-male-killing strains sNeo37,38 and Hyd1. These poulsonii-clade strains are also known for their defensive abilities against certain parasitic wasps and nematodes of Drosophila39,25,22,27,40,41. Evidence of SRL depurination consistent with the action of RIP has been detected in nematodes and wasps exposed to Drosophila harbouring Spiroplasma37,31; and has led to the hypothesis that RIPs play a major role in the Spiroplasma-mediated defence against wasps and nematodes. Depurination of Drosophila ribosomes in Spiroplasma-infected treatments had been detected in larvae, but was mostly restricted to “cell-free hemolymph”, suggesting only ribosomes outside the cell were targets38. Furthermore, despite detecting significant levels of depurinated Drosophila ribosomes in Spiroplasma-infected flies, this phenomenon is generally not accompanied by a significant decrease in intact (i.e., non-depurinated) ribosomes37,39. This, along with the observation that the detected levels of depurination were not associated with larva-to-adult fly mortality, was interpreted as RIP activity having a negligible direct effect on fly fitness17,37. A more recent study, however, revealed that Spiroplasma-mediated depurination of Drosophila ribosomes varies widely by life stage. It is strongest in embryos and old adults, but also not accompanied by a detectable decrease in intact ribosomes. In addition, significant ribosome depurination (but not significant depletion of intact ribosomes) occurs under heterologous expression of two Spiroplasma RIP genes in D. melanogaster, confirming their RIP activity. Their expression was also associated with embryo mortality (male mortality was higher), and with a reduction in fly lifespan and in adult haemocyte number8.

In this study we examined whether Spiroplasma influences the composition of maternally loaded transcripts of host eggs (i.e., prior to fertilization), particularly in a manner that may facilitate male-killing. To do
so we compared the effects of three heritable endosymbionts (a male-killing Spiroplasma, a non-male-killing Spiroplasma, and a CI-inducing Wolbachia) on the composition of mRNA transcripts of early embryos of Drosophila melanogaster. Furthermore, we used data from this study and published RNA-seq data to analyse whether there is a difference in signals of rRNA depurination of host embryo ribosomes in the presence of male-killing and non-male-killing symbionts.

Materials and Methods

Generation of symbiont treatments. We used three infection treatments, one Wolbachia and two Spiroplasma strains, and a symbiont-free control (Fig. 1). Laboratory stocks of D. melanogaster (Canton S strain; CS) that naturally harbour the Wolbachia strain wMel were used to generate the Wolbachia treatment (W−S−). Positive infection for wMel was confirmed based on PCR with Wolbachia-specific primers targeting the wsp gene. The same stock was reared in tetracycline food (final concentration 0.02 g/ml) for two generations, followed by three generations of antibiotic-free food to generate a Wolbachia-free (W−) stock. The W− flies served as the symbiont-free control. The Spiroplasma infection treatments (W−S+) were generated by artificially infecting Wolbachia-free (W−) flies with the strain MSRO-BR (Red 42) native to D. melanogaster or Hyd1 (TEN-104-106) native to D. hydei. Fifteen Wolbachia-free (W−) females (15 lines) were infected per Spiroplasma strain. These artificially infected lines were maintained for 3–5 generations before being used for the experiment. Spiroplasma-infected (W−S+) lines were selected every generation to ensure positive infection status, based on PCR with Spiroplasma-specific 16S ribosomal DNA primers. MSRO treatment lines were backcrossed to Wolbachia-free (W−) CS males every generation, as male-killing by this strain is nearly perfect. A minimum of four infected lines was combined, per replicate, at the start of the experiment, to create a total of three biological replicates for each Spiroplasma treatment. The biological replicates for the Wolbachia treatment and control were maintained as three different populations for four generations prior to the start of the experiment.

Embryo collection. Approximately 40–50 three-day-old virgin females, from each replicate of each treatment, were allowed to mate in cages with Wolbachia-free (W−) CS males during the collection period, and allowed to lay eggs on cornmeal food plates. The initial batch of eggs was discarded to improve the chances of collecting fertilized eggs from the same stage. Thereafter, egg laying was monitored and cornmeal plates were changed approximately every 45 min, so as to collect embryos that were on average ~60–75 min old. Eggs were collected from each replicate with a small brush 6–8 times over a 2-day period for each treatment and the control. The eggs were placed in sterile RNase-free 1.7 ml microtubes, and immediately put on dry ice during the collection period, after which they were transferred to −80 °C for storage.

RNA extraction, library preparation, and sequencing. Three to four biological replicates per condition (MSRO, Hyd1, wMel, and control) were used for the extractions (see Fig. 1). RNA was extracted per collection tube of eggs (mentioned above) with Trizol® Plus RNA Purification System (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. RNA from tubes that belonged to the same biological replicate within a treatment was pooled. All RNA samples were DNase-treated with Ambion® DNA-free (Invitrogen) to remove any DNA contamination. Total RNA was quantified with a NanoDrop® ND-1000 UV spectrophotometer (NanoDrop Technologies, Wilmington, DE), and sample quality and integrity were further tested with the Agilent 2100 BioAnalyzer (Agilent Inc., Santa Clara, CA).

Total RNA was submitted to the Texas AgriLife Genomics and Bioinformatics Services facility for library preparation, sequencing, and demultiplexing. Three biological replicates were subjected to multiplex NuGen

![Figure 1](image-url)
library preparation and pooled into a single sequencing lane of the Illumina (San Diego, CA) Genome Analyzer II platform (single-end; 76 bp read length; italicized samples in Fig. 1). The remaining biological replicates were subjected to the Illumina mRNA TruSeq kit library preparation protocol (12 libraries). Half of the libraries were pooled into one lane of the Illumina HiSeq 2000 platform and the other half were pooled into another lane (single-end; 100 bp read length; see Fig. 1).

**Analyses of differential expression (DE).** Raw reads have been deposited in the NCBI SRA Database under Accession Numbers SRR7279355-SRR7279369 (BioProject PRJNA474708; BioSample SAMN09370647). Command lines, as well as input and output files needed to replicate our analyses, from the different pipelines used are available in Supporting Data Files S1–S4. Raw RNA sequence files were first processed with Trimmomatic (v 0.35) to remove adapters and low quality reads (see Fig. 1). The processed runs were then mapped to the Flybase Drosophila genome (version dmel-all-r6.18) with HiSat2 (v.2.0.5) to obtain bam files. The files containing the mapped read information were then analysed with HTseq (0.6.1) to obtain read counts for both genes and exon regions specified in a corresponding gff file for the dmel-all-r6.18 genome sequence. The resulting count files were then manually formatted into a count matrix suitable for differential expression analysis (Supporting Data File S2).

Differential expression analyses were performed with three different pipelines edgeR v.3.20.5 and Limma + Voom v. limma_3.34.5 and DeSeq2 v.1.18.1. Command line and relevant input files for these analyses are provided in Supporting Data Files S1–S3. All programs conducted their respective tests on genes/exons with 10 or more mapped reads per replicate within at least one of two the treatments being compared. For the edgeR pipeline both the glm and QFglm models were used for DE. For Deseq2, the wald statistical test was used for DE analysis. Before analysis in both Limma-Voom and EdgeR pipeline, two genes with extremely high counts, 16S rRNA (FBgn0013686) and ef-1α (FBgn0284245), were removed. Genes with an adjusted $p$-value (or false discovery rate; FDR) of 0.05 or less were considered to be potentially differentially expressed (DE).

**Power analyses for differential expression.** To identify possible limitations of our experimental design in the detection of DE genes, we performed a statistical power analysis with the method of Hart et al. The parameters (and rationale) used to run the simulations are provided in Table 1 and Fig. 2.
From poly-A-tail-enriched RNA libraries) from two published studies that also compared patterns of gene expression flies, we used the RNA-seq data generated in the present study, as well as RNA-seq data (also derived from Spiroplasma RIP activity in any procedure that relies on reverse transcription (e.g. RNA-seq or reverse-transcription qPCR). RIP-depurinated site in the cDNA and all subsequent PCR amplification steps, can be used to detect evidence of the backbone remains intact)52. When a reverse transcriptase encounters an abasic site, it preferentially adds an adenine in the nascent complementary (c)DNA strand53. This property, which results in an incorrect base at the RIP target position (indicative of intact rRNA) in Spiroplasma mRNA-seq datasets derived from (RIPs), which are encoded in the genomes of several Spiroplasma symbionts: the male-killing Drosophila D. melanogaster. To achieve this, we used mRNA-seq to evaluate differential expression (DE) in mRNAs in the 28S rRNA, consistent with the activity of Ribosome Inactivating Proteins (RIPs), which are encoded in the genomes of several Spiroplasma strains. This assay examined the proportion of adenines (arcsine square root transformed) at the target position (excluding gaps from total number of reads). Source Study was treated as a random effect. Analyses of depurination signal in the sarcin-ricin loop (SRL) of the 28S rRNA of Drosophila. RIP toxins remove a specific adenine present in the sarcin-ricin loop (SRL) of the 28S rRNA leaving an abasic site (i.e., the backbone remains intact)52. When a reverse transcriptase encounters an abasic site, it preferentially adds an adenine in the nascent complementary (c)DNA strand53. This property, which results in an incorrect base at the RIP-depurinated site in the cDNA and all subsequent PCR amplification steps, can be used to detect evidence of RIP activity in any procedure that relies on reverse transcription (e.g. RNA-seq or reverse-transcription qPCR). To examine whether a signal of depurination consistent with RIP activity was detectable in Spiroplasma-infected flies, we used the RNA-seq data generated in the present study, as well as RNA-seq data (also derived from poly-A-tail-enriched RNA libraries) from two published studies that also compared patterns of gene expression from Spiroplasma-infected and Spiroplasma-free D. melanogaster embryos: NCBI Acc. No. PRJDB446950; and PRJNA31837351. The Spiroplasma strain examined by50 and51 is substrain UG (for Uganda) of MSRO. Bowtie2 v.2.1.054 and Samtools 0.1.1955 were used to map and extract, respectively, the reads from the three studies to the 28S rRNA gene of D. melanogaster (Acc. No. NR_133562.1 positions 2591–3970). We verified that the reads that mapped to NR_133562.1:2591–3970 (hereafter “28S reads”) did not belong to other parts of the Drosophila genome or to the genomes of Spiroplasma or Wolbachia, as follows. We used HISAT2 v.2.0.2-beta46 (for Drosophila) and Bowtie2 v. 3.4.254 (for the prokaryotic genomes) to attempt mapping the “28S reads” to the following reference genomes: (1) the D. melanogaster genome (BDGP6); (2) MSRO-UG (NZ_PETG00000000.1); (3) a draft genome of Spiroplasma Hyd1 (unpublished data); and (4) Wolbachia wMel (NC_002978). As expected, the “28S reads” only mapped to (1), and specifically to regions whose annotation terms included “28S rRNA”. To visualize and count the shift from A to T (or other bases) in the source RNA pool, the “28S reads” were mapped again to the NR_133562.1:2591–3970 in Geneious v.11.1.2 (Biomatters Inc., Newark, NJ); “low sensitivity mode”; maximum gap size = 3; iterate up to 25 times). The number of reads containing each of the four bases or a gap at the target site was counted by selecting the position at all the reads to be counted, and recording the counts reported by Geneious under the “Nucleotide Statistics” option (gapped reads were excluded from subsequent analyses). The proportion of reads with an A at the target site (i.e., putatively intact rRNA) was calculated and compared among treatments and replicates. We used a mixed model in JMP Pro 13 (SAS Institute Inc., Cary, NC) to examine the effect of symbiont treatment (fixed) on the proportion of adenines (arcserine square root transformed) at the target position (excluding gaps from total number of reads). Source Study was treated as a random effect. **Results** The present study aimed at determining whether heritable symbionts alter the composition of maternally-loaded mRNAs in D. melanogaster. To achieve this, we used mRNA-seq to evaluate differential expression (DE) in Drosophila embryos lacking endosymbionts (control) to those harbouring one of the following heritable endosymbionts: the male-killing Spiroplasma poulsonii strain MSRO-Br; the CI-inducing Wolbachia strain wMel; and the non-male-killing Spiroplasma poulsonii strain Hyd1 (native to D. hydei). A power analysis was performed to determine the limitations of our experimental design. A secondary goal of this study capitalized on several available mRNA-seq datasets derived from Spiroplasma-infected Drosophila melanogaster embryos, to search for signals of damage (i.e., depurination) to Drosophila rRNA, consistent with the activity of Ribosome Inactivating Proteins (RIPs), which are encoded in the genomes of several Spiroplasma strains. This assay examined the proportion of adenines at the RIP target position (indicative of intact rRNA) in Spiroplasma-infected vs. Spiroplasma-free treatments. **Differential expression.** The number of reads that was obtained, passed QC, and mapped to the Drosophila genome is shown in Table 2. The PCA plot did not reveal any particular grouping of samples according to
Table 2. Number of reads obtained, retained after quality control, and mapped to the reference genome.

| Treatment | Replicate | Total reads Number | Reads after quality control Number | Reads mapped to reference genome Number | % | % |
|-----------|-----------|--------------------|-------------------------------------|-----------------------------------------|----|----|
| Control   | 1_galII   | 38,623,190         | 38,595,804                         | 36,658,295                             | 99.33 | 94.98 |
| Control   | 2_TruSeq_4| 32,414,845         | 32,413,140                         | 31,382,402                             | 99.99 | 96.82 |
| Control   | 6_TruSeq_6| 28,893,141         | 28,891,426                         | 27,923,563                             | 99.99 | 96.65 |
| Control   | 9_TruSeq_7| 31,672,204         | 31,672,204                         | 30,671,362                             | 100.00 | 96.84 |
| Hyd1      | 3_galII   | 39,050,068         | 39,031,083                         | 36,931,211                             | 99.95 | 94.62 |
| Hyd1      | 3_TruSeq_2| 28,074,997         | 28,074,997                         | 27,151,330                             | 100.00 | 96.71 |
| Hyd1      | 10_TruSeq_12| 31,084,062         | 31,084,062                         | 30,070,722                             | 100.00 | 96.74 |
| MSRO      | 2_galII   | 36,838,800         | 36,824,546                         | 34,640,850                             | 99.96 | 94.07 |
| MSRO      | 1_TruSeq_2| 31,841,684         | 31,832,465                         | 30,867,941                             | 99.97 | 96.97 |
| MSRO      | 5_TruSeq_5| 30,537,077         | 30,535,822                         | 29,442,640                             | 100.00 | 96.20 |
| MSRO      | 8_TruSeq_6| 30,124,318         | 30,124,318                         | 29,172,390                             | 100.00 | 96.84 |
| wMel      | 4_TruSeq_4| 32,691,282         | 32,691,282                         | 31,449,013                             | 100.00 | 96.20 |
| wMel      | 7_TruSeq_5| 27,032,887         | 27,032,887                         | 26,073,220                             | 100.00 | 96.45 |
| wMel      | 11_TruSeq_7| 28,550,836         | 28,548,722                         | 27,595,195                             | 99.99 | 96.66 |

Power analyses for differential expression. A plot of the power analyses is shown in Fig. 5 and the rationale for the parameters assumed is provided in Table 1. Assuming a coverage of 20X and a coefficient of variation of 0.3, we should expect to detect: ~100% of genes differentially expressed by log-fold changes ≥3; at least 60% of genes with log-fold changes ≥2; and at least ~40% of genes with log-fold changes ≥1.75. Therefore, these results suggest that our design and data have sufficient power to detect log-fold changes ≥3, but more limited power below that log-fold change value.

Signals of depurination. The number of reads mapped to the target site, as well as the proportion of adenosines (representative of non-depurinated rRNA) by treatment replicate, are shown Fig. 6 and Table S1. Symbiont treatment had a significant effect on the proportion of adenosines (F_{4,28} = 29.8; p < 0.0001), whereas source study did not (Wald p = 0.4895). Both MSRO (UG and BR) treatments had a significantly smaller proportion of Ag (mean = 96.78%) than the control (mean = 99.79%) and wMel (mean = 100%) treatments. Hyd1 exhibited intermediate proportions of As (mean = 98.53%; Fig. 6). In other words, we detected a mean depurination of 3.22% for MSRO and 1.47% for Hyd1, and effectively no depurination for the control wMel treatment. Pooling of the two substrains of MSRO (i.e., MSRO-UG and MSRO-BR) into one treatment produced similar results (not shown). The Harumoto et al. dataset, which included two embryo stages and differentiated between male and female embryos, did not exhibit substantially different depurination levels between stages or sex (e.g. mean %A was 98.27 for females and 98.49 for males; see Table S1).

Outcome of mapping RNAseq reads to symbiont genomes. To determine the degree to which poly-A-tailed enriched RNAseq data yield reads assignable to the bacterial symbionts, we used Bowtie2 (v.2.3.4) with default settings to map the reads from each symbiont-infected replicate to the closest symbiont reference genome available. The number of reads per replicate mapped to the symbiont genome ranged from 5 (for a Hyd1 replicate) to 8619 (for a wMel replicate; Table S2), representing a very small fraction of the total reads. The percentage of these putative symbiont-derived reads that mapped to the symbiont ribosomal RNA genes was quite variable (range 3.94–100%; mean 52%). The above numbers of symbiont-derived genes are inadequate for analyses of differential expression (DE) of the symbiont. Collectively these findings suggest that libraries generated via a poly-A-tail enrichment process tend to yield a very low proportion of reads of bacterial origin; a result that is expected given that polyadenylation is not part of the process of bacterial mRNA maturation.
Table 3. The number, identity (Flybase gene number), log fold-change (logFC), adjusted p-value or FDR, of genes identified as differentially expressed (DE) between each treatment and the symbiont-free control, by each of the four methods utilized. Results are shown for analyses of both genes and exons. Only the Hyd1 treatment exhibited significant DE genes. *Not significant at alfa = 0.05.

Discussion

Effects of symbionts on early embryo mRNA composition. The present study aimed to test whether infection by a heritable bacterium strain that kills males during the embryonic stage influences the composition of mRNA transcripts present in the early embryo (ca. 60–75 min post-oviposition), and thus, before the onset of zygotic transcription. Based on several analysis pipelines and assumptions, we did not detect any differential expression in genes among embryos harbouring the male-killing Spiroplasma strain (MSRO-Br), the CI-inducing Wolbachia strain, and the symbiont-free control. These results suggest that these symbionts do not influence the composition of initial maternally-loaded transcripts or their degradation up to the examined stage, unless they do so to a degree below what was detectable by our experimental design (e.g. below a log-fold change of 3; see Power Analyses). Furthermore, it is possible, that such cytoplasmically transmitted symbionts could influence the composition of maternally-loaded proteins, or regulate the protein complement of the early embryo in a transcriptionally-independent manner (e.g. through post-transcriptional or post-translational controls). To our knowledge, these features have not been compared in Spiroplasma-infected vs. uninfected embryos.

Our results showing no effect of the male-killing Spiroplasma on maternally-loaded mRNA composition are consistent with what is known about the male-killing mechanism. A Spiroplasma-encoded protein (Spaid) appears to interact with the DCC, which assembles onto the X chromosome of wild-type males but not females, and is accompanied by DNA damage and abnormal apoptosis. This model does not require "priming" of the egg during oogenesis by the symbiont. Nonetheless, the delayed expression of male death in males flies over-expressing the wild type Spaid (2nd larval instar), compared to those harbouring male-killing Spiroplasma-strain native to D. hydei, yielded equivocal results, with four pipelines revealing no differentially expressed genes, and the other four pipelines revealing an intersection of five DE genes (ect, Osi6, Osi7, FBgn0038339, and FBgn0037099). All of these five genes had a lower expression level in the Hyd1 treatment compared to the other three treatments (Fig. 4). The time course of expression levels reported in FlyBase for these genes, which is based on Wolbachia-infected flies, indicates that early embryos have lower levels than later embryo stages (see Table 4). Based on this pattern, a plausible hypothesis is that the Hyd1 treatment exhibits a developmental delay. To address this, we compared the expression levels among treatments of six genes that should exhibit high levels at the 0–2 h embryo and substantially lower levels at the 2–4 h embryo (see Table 4). Our rationale was that if the Hyd1 treatment were developmentally delayed, we might detect a trend of higher expression levels at these genes compared to the other treatments. Nonetheless, Hyd1 exhibited no such a pattern; except for one of these genes (i.e., FBgn0035955; not significant; Fig. 7). Therefore, collectively, the patterns of expression of Hyd1-infected embryos are inconsistent with a developmental delay.
Available information on the function of the five genes at which the Hyd1 treatment had significantly lower expression provides little insight into the possible causes or consequences. Based on FlyBase, *ectodermal* (ect; FBgn0000451) is expected to have very low expression (FlyBase RPKM = 1) at the stage we examined. It is expressed later (Stages 13–16 ≈ embryo 14–16 h) in several tissue types: foregut, epidermis, trachea, and hindgut, with possible roles in cuticle development and tubular formation (e.g. tracheal tubes)\(^5\). *Osiris 6* (Osi6; FBgn0027527) and *Osiris 7* (Osi7; FBgn0037414) are also expected to have very low expression (FlyBase RPKM = 10 and 5, respectively) at the stage examined; with the highest expression at later embryonic stages (i.e., 14–16 h). Both genes are physically close to each other and belong to the Osiris gene cluster (OSI; FBgg0000612). The OSI family of genes shares a highly conserved protein domain is present, and retains high synteny, in all insects examined to date, and appears to have evolved via gene duplications at the base of the Insecta\(^6\). In *D. melanogaster*, these genes are within the dosage-sensitive *Triploid-lethal* (Tpl) locus; individuals with one or three copies of Tpl die as late embryos or early first instar larvae\(^6\). Although Osiris genes are generally considered of unknown function, based on expression patterns or knock-down experiments, they have been associated with cuticle formation, metamorphosis, digestion, resistance/tolerance to toxins and viruses, wing formation, and phenotypic plasticity (reviewed in Smith, et al.\(^6\)).

FBgn0038339 (CG6118) contains a BTB/POZ domain, and appears to be involved in regulation of transcription by RNA polymerase II. FBgn0037099 (CG7173) contains a Kazal domain (i.e., a type of serine proteinase inhibitor). Both of these genes are expected to have very low expression (FlyBase RPKM = 1 and 0, respectively) at the stage we examined, with peak expression levels occurring in the 14–16 h embryo.

Assuming that the significant results for the five genes are repeatable, they imply more perturbation of gene expression by Hyd1 than by the other symbiont strains; a finding that could reflect that *D. melanogaster* is outside
the fundamental niche of Hyd1, as also suggested by its lower vertical transmission efficiency in this host (80%64), and by the difficulty of maintaining this artificial association in the lab (personal observation and Kageyama, et al65). Reports of additional phenotypic effects of Hyd1 in D. melanogaster have been anecdotal or contradictory. Concerning adult longevity, Kageyama, et al.65 reported substantially shorter adult life spans of both males and females, whereas Silva, et al.66 detected little to no adult mortality. The study of Silva, et al.66 is expected to more closely represent the conditions and specific D. melanogaster (i.e., Canton S) and Hyd1 (i.e., sub-strain TEN104–106) genetic backgrounds of the present transcriptomics study. Concerning fly fecundity, both Kageyama, et al.65 and Hutchence, et al.64 report anecdotal evidence of lower female fecundity in Hyd1-infected D. melanogaster than their uninfected counterparts; Silva, et al.66 did not examine this trait. Consistent with the detrimental effect of Hyd1 on fecundity of D. melanogaster, the microarray-based study of Hutchence, et al64 (focused on adults), revealed downregulation of a cluster of genes involved in egg production and fertilization. Nonetheless, in general, Hutchence, et al.64 found that D. melanogaster adults infected with Hyd1 exhibited less perturbation of gene expression than those infected by MSRO-Br and Spiroplasma poulsonii strain NSRO64 (a close relative of MSRO that also kills males and naturally occurs in Drosophila nebulosa). The apparently different patterns of gene expression perturbation by Hyd1 between our study and that of Hutchence, et al.64 could be due to the different fly stages examined (embryos vs. adults), different experimental tools (RNA-seq vs. microarrays), or other factors (e.g. all of Hutchence et al.’s treatments were also infected with Wolbachia).

The lack of an effect of wMel, a CI-inducing strain, on D. melanogaster early embryo mRNA composition, is not unexpected given what is known about the CI mechanism. Essentially, wMel encodes two contiguous protein coding genes (cifA and cifB)17. Each of these proteins is capable of causing CI when expressed in the male germline, whereas only cifA (when expressed in embryos) is able to rescue CI18. Our results suggest that wMel does not “prime” the egg for rescue by manipulating maternally-loaded transcripts or their degradation.

Depurination patterns. Our study capitalized on the presence of (non-target) rRNA reads in RNAseq datasets prepared by poly-A tail enrichment, to estimate levels of depurination at the RIP target position of the sarcin-ricin loop (SLR) in the 28SrRNA of eukaryotes. We acknowledge that the levels of depurination estimated from mRNAseq data may be downwardly biased because rRNA depurinated by RIPS is highly prone to hydrolysis of the sugar-phosphate backbone at the lesion site67. Furthermore, the freezing to which these samples were subjected between collection and library preparation might have decreased the detectability of depurination53. In addition, it is not known whether the polyA-RNA enrichment protocol used prior to library preparation, which is aimed at depleting rRNA in the sample, could bias representation of depurinated vs. intact rRNA. Only one study has compared inferences of ribosome depurination from mRNAseq vs. qPCR assays in the same system. Based on mRNAseq, Hamilton et al.17 detected ~3.8% depurination (i.e., ~96.2% adenine) in the nematode Howardula infesting adult Drosophila neotestacea harbouring the non-male-killing Spiroplasma sNeo strain. Comparatively, using the qPCR approach, the abundance of depurinated template representing RIP-induced depurination increased ~20-fold, whereas the levels of intact nematode rRNA were reduced ~six-fold in the presence of Spiroplasma17. Notwithstanding the potential biases, the consistent finding of no depurination in Spiroplasma-free treatments vs. depurination in Spiroplasma-present treatments across three independent studies, lends credibility to the mRNAseq-based approach we employed for inference of depurination.

The significantly lower proportion of adenines at the RIP target site for the Spiroplasma treatment and the control is consistent with the action of a Spiroplasma-encoded RIP, and supports recent findings that: (1) MSRO-UG infection causes ribosome depurination in D. melanogaster embryos; (2) that heterologous expression of a Spiroplasma RIP1 and RIP2 genes (separately) also causes ribosome depurination.

Figure 5. Power Analyses. Estimation of power for alfa = 0.05, at five different effect values (log-fold change: 1.5–4); number of replicates (n = 3 and 4), coverage values (X = 10 and 20), and coefficient of variation (c.v. = 0.2, 0.3, 0.4).
in *D. melanogaster* embryos (measured by a qPCR approach); and (3) that the degree of ribosome depurination (“RIP activity” sensu Garcia-Arraez, et al.38) in the different *Spiroplasma* treatments or RIP transgene constructs is positively correlated with embryo mortality38. Importantly, despite detecting significantly higher levels of depurinated template with the qPCR approach, none of the qPCR assays revealed significantly lower levels of intact template38. This suggests that detectable depletion of intact template in the qPCR approach is not a pre-requisite for detecting a phenotypic effect of depurination.

Our results suggest that MSRO substrain Brazil (MSRO-BR) causes comparable levels of depurination to the Uganda substrain, which is consistent with its strong male-killing effect43,66, and with the identical content of RIP-encoding genes (unpublished data). Interestingly, the male-killer MSRO generally exhibited a higher signal of depurination than the non-male-killer Hyd1 (MSRO range 1–8%; Hyd1 range 1–<3%; Fig. 6). This difference could be the result of differences in titers of MSRO vs. Hyd1 at the embryonic stage. Densities of these strains in *D. melanogaster* Canton-S background at the adult stage are lower for Hyd1 than for MSRO-Br, but only immediately after adult eclosion66; densities at the embryonic stage have not been compared. Alternatively, it is possible that the RIP genes encoded in the MSRO genome are more actively expressed or secreted, or are more efficient than the putative RIP genes detected in the genome of Hyd139. Levels of fly ribosome depurination in the presence of sNeo, the

| DE Genes in this study | FlyBase report for wMel infected flies (RPKM)* |
|------------------------|-----------------------------------------------|
|                        | 0–2h embryo | 2–4h embryo | 4–6h embryo | 6–8h embryo |
| FBgn0038339 (CG6118)   | 1           | 1           | 1           | 3           |
| FBgn0000451 (ect)     | 1           | 1           | 1           | 5           |
| FBgn0027527 (Osr6)    | 10          | 7           | 8           | 31          |
| FBgn0037414 (Osr7)    | 5           | 3           | 4           | 16          |
| FBgn0037099 (CG7173)  | 0           | 0           | 1           | 2           |

Table 4. RPKM counts reported in FlyBase for the five DE genes found in our study and for six additional genes, “Developmental Markers” whose expression is higher in the 0–2 h embryo than in later stages. Gene name or Annotation Symbol in parenthesis. “Note: RPKM values reported in FlyBase are not expected to be directly comparable to the normalized count values reported in the present study.

Figure 6. Presence of male-killing *Spiroplasma* (MSRO) results in a significant signal of depurination. Top Panel. Percent of adenines (i.e., the expected base in the absence of depurination) at the target site of Ribosome Inactivating Proteins (RIPs). Different letters indicate significant Posthoc tests (at P-value < 0.04). Bottom Panel. Coverage (number of reads) that mapped to the target depurination sites. Jitter points (replicates) and box plots per treatment (symbiont and source study) are shown.
non male-killing strain native to *D. neotestacea*, have not been assayed in embryos, but a qPCR assay of *D. neotestacea* ovaries indicated a small, albeit non significant, signal of depurination\(^17\). Similarly, an mRNA-seq experiment of adult *D. neotestacea* revealed 0.4% depurination (i.e., \(\%A = 99.6\%\)) in the presence of sNeo\(^17\). Therefore, the three members of the *poulsonii* clade examined to date (MSRO, Hyd1, and sNeo) encode RIPs capable of depurination of *Drosophila* ribosomes, but the male-killing strain exhibits the highest levels depurination, a phenomenon that appears to contribute to the male-killing mechanism\(^38\). The general patterns, however, indicate that *Spiroplasma* RIPs are particularly efficient at depurinating the ribosomes of natural enemies of *Drosophila* (i.e., parasitic wasps and nematodes), leading to the hypothesis that RIP-induced depurination plays an important role in defence mechanism\(^17,37\). The higher depurination levels of fly ribosomes detected in embryos (and ovaries) compared to other fly stages could be due to greater exposure of ribosomes to RIP prior to cellularization, after which *Spiroplasma* becomes effectively extra-cellular\(^38\). The relatively high levels of depurination in old adults appears to be the result of higher *Spiroplasma* densities at that stage\(^38\). Additional roles have been attributed to RIPs, which could contribute to the male-killing phenotype. For example, several RIPs are reported to cause DNA damage\(^68,69\). Thus, one or more *Spiroplasma*-encoded RIP might directly contribute to the DNA damage reported during the process of male-killing\(^30\).

**Conclusions**

This study employed a transcriptomics approach to examine whether cytoplasmically-transmitted bacteria, including reproductive manipulators that strongly impact survival of the embryonic stage of *Drosophila*, influence composition maternally-loaded mRNAs. The results revealed that mRNA composition does not differ significantly among the embryos harbouring the reproductive parasites *Spiroplasma* MSRO and *Wolbachia* wMel and those lacking endosymbionts. Only the symbiont *Spiroplasma* Hyd1, which does not manipulate reproduction,
appeared to alter expression levels of a handful of genes (5–12), but not all analytical approaches supported this finding. Our power analyses indicated that our experimental design should be able to detect most of the genes exhibiting ≥ 3-log-fold change in expression among treatments. Collectively, our results suggest that these cytoplasmically-transmitted bacteria do not alter the composition of mRNAs of the early embryo, and are thus unlikely to use this mechanism to exert their reproductive phenotypes. Capitalizing on several transcriptomics datasets, this study also detected signals of Spiroplasma-induced damage to ribosomes in early Drosophila embryos, with greater damage caused by the male-killer (MSRO) than the non-male-killer (Hyd1), consistent with recent results implicating this mechanism in male killing.

Data Availability

The datasets generated during and/or analysed during the current study are available as Supporting Data at NCBI's Sequence Read Archive (SRA) database under Accession Numbers SRR7279355–SRR7279369 (BioProject PRJNA474708; BioSample SAMN09370647).

References

1. Duron, O. & Hurst, G. D. Arthropods and inherited bacteria: from counting the symbionts to understanding how symbionts count. BMC Biol. 11, 45, https://doi.org/10.1186/1741-7007-11-45 (2013).
2. Kraft, L. J., Kopco, J., Harmon, J. P. & Oliver, K. M. Aphid symbionts and endogenous resistance traits mediate competition between rival parasitoids. PLoS One 12, e0180729 (2017).
3. Watanabe, K., Yukihito, F., Matsuura, Y., Fukatsu, T. & Noda, H. Intrasperm vertical symbiont transmission. Proc Natl Acad Sci USA 111, 7433–7437, https://doi.org/10.1073/pnas.1402476111 (2014).
4. Veneti, Z., Bentley, J. K., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Veneti, Z., Venet,
65. Kageyama, D.
66. Silva, N. O., Guenther, L. L., Xie, J. & Mateos, M. Infection densities of three
68. Bolognesi, A.
67. Endo, Y. & Tsurugi, K. Mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes.
59. Rosa, J. B., Metzstein, M. M. & Ghabrial, A. S. An
52. Walsh, M. J., Dodd, J. E. & Hautbergue, G. M. Ribosome-inactivating proteins: potent poisons and molecular tools.
51. Hart, S. N., Therneau, T. M., Zhang, Y., Poland, G. A. & Kocher, J. P. Calculating sample size estimates for RNA sequencing data.
42. Jeyaprakash, A. & Hoy, M. A. Long PCR improves DNA amplification:
44. Dobson, S. L. Transfection of
46. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements.
47. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data.
45. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data.
48. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.
46. Dobson, S. L. Transfection of Wolbachia pipientis into Drosophila embryos. Current Protocols in Microbiology 5, 3A.4.1–3A.4.11, https://doi.org/10.1002/9780470172925.mc03a04s05 (2007).
47. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169, https://doi.org/10.1093/bioinformatics/btu638 (2015).
48. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140, https://doi.org/10.1093/bioinformatics/btp616 (2010).
49. Law, C. W., Alhamdooosh, M., Su, S., Smyth, G. K. & Rüchel, M. E. RNA-seq analysis is easy as 1-2-3 with limma. Genome and edgeR, F1000Research 5, 1408 (2016).
50. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550, https://doi.org/10.1186/s13059-014-0550-0 (2014).
51. Hart, S. N., Therneau, T. M., Zhang, Y., Poland, G. A. & Kocher, J. P. Calculating sample size estimates for RNA sequencing data. J. Comput. Biol. 20, 970–978, https://doi.org/10.1089/cmb.2012.0283 (2013).
52. Walsh, M. J., Dodd, J. E. & Hauertbergue, G. M. Ribosome-inactivating proteins: potent poisons and molecular tools. Virulence 4, 774–784, https://doi.org/10.4161/viru.26399 (2013).
53. Melchior, W. B. Jr. & Tolleson, W. H. A functional quantitative polymerase chain reaction assay for shiga, stx1 and stx2, and related ribosome-inactivating proteins. Anal. Biochem. 396, 204–211, https://doi.org/10.1016/j.ab.2009.09.024 (2010).
54. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359, https://doi.org/10.1038/nmeth.1923 (2012).
55. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
56. Graveley, B. R. et al. The developmental transcriptome of Drosophila melanogaster. Nature 471, 473–479, https://doi.org/10.1038/nature09715 (2011).
57. Gutzwiler, F. et al. Dynamics of Wolbachia pipientis gene expression across the Drosophila melanogaster life cycle. G3 (Bethesda) 5, 2843–2856, https://doi.org/10.1534/g3.115.021931 (2015).
58. Raha, D., Nguyen, Q. D. & Garen, A. Molecular and developmental analyses of the protein encoded by the Drosophila gene ectodermal. Dev. Genet. 11, 310–317, https://doi.org/10.1002/dvg.1020110410 (1990).
59. Rosa, J. B., Metzstein, M. M. & Ghafari, A. S. An ichor-dependent apical extracellular matrix regulates seamless tube shape and integrity. PLoS Genet. 14, e1007146, https://doi.org/10.1371/journal.pgen.1007146 (2018).
60. Shah, N., Dorer, D. R., Moriyama, E. N. & Chrisiens, A. C. Evolution of a large, conserved, and syntenic gene family in insects. G3 (Bethesda) 2, 313–319, https://doi.org/10.1534/g3.111.001412 (2012).
61. Misof, B. et al. Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767, https://doi.org/10.1126/science.1257570 (2014).
62. Dorer, D. R., Rudnick, J. A., Moriyama, E. N. & Chrisiens, A. C. A family of genes clustered at the Tripl-e-lethal locus of Drosophila melanogaster has an unusual evolutionary history and significant syntenic with Anopheles gambiae. Genetics 165, 613–621 (2003).
63. Smith, C. R., Morandin, C., Noureddine, M. & Pant, S. Conserved roles of Osiris genes in insect development, polymorphism and protection. J. Evol. Biol. 31, 516–529, https://doi.org/10.1111/jeb.12328 (2018).
64. Butchence, K. J., Fischer, B., Paterson, S. & Hurst, G. D. D. How do insects react to novel inherited symbionts? A macroarray analysis of Drosophila melanogaster response to the presence of natural and introduced. Spiroplasma. Mol. Ecol. 26, 950–958 (2011).
65. Kageyama, D. et al. Prevalence of a non-male-killing Spiroplasma in natural populations of Drosophila hydei. Appl. Environ. Microbiol. 72, 6667–6673 (2006).
66. Silva, N. O., Guenther, L. L., Xie, J. & Mateos, M. Infection densities of three Spiroplasma strains in the host Drosophila melanogaster. Biosystems 97, 83–93, https://doi.org/10.1016/S1389-1399(06)00173-7 (2007).
67. Endo, Y. & Tsurugi, K. Mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. Nucleic Acids Symp. Ser. 188, 187–190 (1986).
68. Bolognesi, A. et al. Endocytosis and intracellular localisation of type 1 ribosome-inactivating protein saporin-66. J. Biol. Regul. Homeost. Agents 26, 97–109 (2012).
69. Belthold-Vargas, L. R. et al. Type 1 ribosome-inactivating proteins—Entomotoxic, oxidative and genotoxic action on Anticarsia gemmatalis (Hubner) and Spodoptera frugiperda (JE Smith) (Lepidoptera: Noctuidae). J. Insect Physiol. 55, 51–58 (2009).
Author Contributions
Conceived and designed study: N.O.S., M.M., R.A., J.W.E., Conducted the experiments: N.O.S., M.M., Analysed the data: M.M., N.O.S., P.R., V.M.H.A., R.A., Interpreted results: M.M., N.O.S., P.R., V.M.H.A., R.A., J.W.E., Drafted the manuscript: M.M., N.O.S., P.R., V.M.H.A., Reviewed and edited the manuscript: M.M., N.O.S., P.R., V.M.H.A., R.A., J.W.E.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-45371-0.

Competing Interests: The authors declare no competing interests.

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

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

© The Author(s) 2019